

Topics in Biodiversity and Conservation

Roseli Pellens  
Philippe Grandcolas *Editors*

# Biodiversity Conservation and Phylogenetic Systematics

Preserving our evolutionary heritage  
in an extinction crisis

 Springer

# **Topics in Biodiversity and Conservation**

Volume 14

More information about this series at <http://www.springer.com/series/7488>

Roseli Pellens • Philippe Grandcolas  
Editors

# Biodiversity Conservation and Phylogenetic Systematics

Preserving our evolutionary heritage in an  
extinction crisis

With the support of Labex BCDIV and ANR BIONEOCAL



Springer Open

### *Editors*

Roseli Pellens  
Institut de Systématique, Evolution,  
Biodiversité, ISYEB – UMR 7205  
CNRS MNHN UPMC EPHE,  
Muséum National d'Histoire Naturelle  
Sorbonne Universités  
Paris, France

Philippe Grandcolas  
Institut de Systématique, Evolution,  
Biodiversité, ISYEB – UMR 7205  
CNRS MNHN UPMC EPHE,  
Muséum National d'Histoire Naturelle  
Sorbonne Universités  
Paris, France

ISSN 1875-1288

ISSN 1875-1296 (electronic)

Topics in Biodiversity and Conservation

ISBN 978-3-319-22460-2

ISBN 978-3-319-22461-9 (eBook)

DOI 10.1007/978-3-319-22461-9

Library of Congress Control Number: 2015960738

Springer Cham Heidelberg New York Dordrecht London

© The Editor(s) (if applicable) and The Author(s) 2016. The book is published with open access at SpringerLink.com.

Chapter 15 was created within the capacity of an US governmental employment. US copyright protection does not apply.

**Open Access** This book is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

All commercial rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

# Foreword

The desperate, and seemingly inexorably worsening, state of biodiversity on Earth is arguably not a consequence of conscious choices. That is much of the problem. There are undoubtedly cases, and many of them, in which people made explicit decisions to forgo the variety of life naturally present in a given area in favour of some alternative benefit (e.g. agricultural activity, energy production, housing). However, by and large, the global losses of species, and the reductions in the abundances and distributions of increasingly the majority of others, are the outcome of outright ignorance of the impacts of anthropogenic activities, of underestimation or misunderstanding of the impacts of those activities, and, perhaps most significantly, a host of individual decisions which whilst independently perhaps quite rational have led to a combined pressure on biodiversity that is far from what it can sustain.

The field of conservation biology has done much to highlight the status and trends in biodiversity, but especially the need for active and explicit choices as to its future. Frustrating as is their failure to date to be realized, the establishment of baselines and targets for biodiversity at regional, national and global scales is the logical framework within which decisions can properly be made as to what environmental changes and management actions are and are not carried forward, and with what consequences. The ‘agony of choice’ needs to be a real choice, albeit the agony may not always be avoided.

Key to determining baselines and targets, and what choices to make, is deciding which metric to use to discriminate between different outcomes, and particularly to compare those of current actions with alternatives. This book provides a cogent argument for the use of phylogenetic diversity as a key metric – that is, measures of biodiversity that capture evolutionary history – and phylogenetic systematics as a core organizing principle. It highlights the benefits and constraints of such an approach, explores the ways in which it can be implemented, and describes a rich diversity of applications. This is the most comprehensive compilation of cutting-edge contributions on this topic to date, provides many valuable insights, and a ‘go to’ source of understanding. The intention to help improve the global condition of biodiversity is apparent throughout.

Biological conservation has oft been hampered by those who have maintained that priorities for action should only be established using approaches that are easily understood by the general public. The same demand has not been made in many other arenas of human endeavor (e.g. medicine, nuclear power), and neither should it constrain biological conservation. That said, there does remain a substantial challenge of encouraging an informed citizenry around the justification and goals of using a phylogenetic diversity approach, and gaining their support. Only by so doing will there be a genuine chance of aligning the multitude of biodiversity-critical decisions being made each and every day across the continents and oceans.

Environment and Sustainability Institute  
University of Exeter, Exeter, UK

Kevin J. Gaston

# Acknowledgements

This book could not exist without the participation of 38 research scientists who took part in the elaboration of these chapters. We thank each of them for their generosity in sharing their experience to make this common work. We are particularly indebted to Dan Faith, for his always prompt and insightful suggestions. Without his help this book would have been quite different. Every chapter was peer reviewed, and reviewers' suggestions and critical thinking helped to improve all manuscripts and greatly contributed to our present understanding of phylogenetic diversity and its importance for biodiversity conservation. So, we thank Anni Arponen, Lisa Ballance, Sven Buerki, Maram Caesar, Marcel Cardillo, Pedro Cardoso, Ben Collen, Rob Cowie, Jorge Crisci, Mike Crisp, Jonathan Davies, Dan Faith, Jon Fjeldså, Felix Forest, Carlos Gonzalez-Orozco, Shan Huang, Jussi Laitila, Romain Julliard, James Justus, Matjaz Kuntner, Frédéric Legendre, Rafael Loyola, Laura J. May-Collado, Daniel Rafael Miranda-Esquivel, Claudia Moreno, Annemarie Ohler, Fabio Pardi, Stéphane Prigent, Carlo Ricotta, Samuel Scheiner, Mike Steel, Andreas Spiller, Géraldine Veron, and Kristen Williams, who kindly acted as referees (often anonymously) for the chapters. The idea of making this book was born during the preparation of a symposium entitled "Phylogenetic Tools for Conservation" that Roseli Pellens organized for the 30th meeting of the Willi Hennig Society at São José do Rio Preto, State of São Paulo, Brazil. We thank Fernando Barbosa Noll and Dalton de Souza Amorim, the meeting organizers, for including this subject in their agenda. We thank the Labex "Diversités biologiques et culturelles: origines, évolution, interactions, devenir" (BCDIV, Jean-Denis Vigne) and the ANR BIONEOCAL (PG) for the financial support that permitted to make this publication open access. We are also very grateful to David Hawksworth, the series editor, and the people at Springer, for their kind help along the preparation of this volume.





# Contents

<b>Phylogenetics and Conservation Biology: Drawing a Path into the Diversity of Life .....</b>	<b>1</b>
Roseli Pellens and Philippe Grandcolas	
<b>Part I Questions</b>	
<b>The Value of Phylogenetic Diversity .....</b>	<b>19</b>
Christopher Lean and James Maclaurin	
<b>The PD Phylogenetic Diversity Framework: Linking Evolutionary History to Feature Diversity for Biodiversity Conservation .....</b>	<b>39</b>
Daniel P. Faith	
<b>Reconsidering the Loss of Evolutionary History: How Does Non-random Extinction Prune the Tree-of-Life?.....</b>	<b>57</b>
Kowiyou Yessoufou and T. Jonathan Davies	
<b>Phylogenetics and Conservation in New Zealand: The Long and the Short of It .....</b>	<b>81</b>
Steven A. Trewick and Mary Morgan-Richards	
<b>What Is the Meaning of Extreme Phylogenetic Diversity? The Case of Phylogenetic Relict Species.....</b>	<b>99</b>
Philippe Grandcolas and Steven A. Trewick	
<b>Part II Methods</b>	
<b>Using Phylogenetic Dissimilarities Among Sites for Biodiversity Assessments and Conservation .....</b>	<b>119</b>
Daniel P. Faith	

<b>Phylogenetic Diversity Measures and Their Decomposition: A Framework Based on Hill Numbers.....</b>	<b>141</b>
Anne Chao, Chun-Huo Chiu, and Lou Jost	
<b>Split Diversity: Measuring and Optimizing Biodiversity Using Phylogenetic Split Networks.....</b>	<b>173</b>
Olga Chernomor, Steffen Klaere, Arndt von Haeseler, and Bui Quang Minh	
<b>The Rarefaction of Phylogenetic Diversity: Formulation, Extension and Application .....</b>	<b>197</b>
David A. Nipperess	
<b>Support in Area Prioritization Using Phylogenetic Information.....</b>	<b>219</b>
Daniel Rafael Miranda-Esquivel	
<b>Assessing Hotspots of Evolutionary History with Data from Multiple Phylogenies: An Analysis of Endemic Clades from New Caledonia .....</b>	<b>237</b>
Roseli Pellens, Antje Ahrends, Peter M. Hollingsworth, and Philippe Grandcolas	
<b>Part III Applications</b>	
<b>Representing Hotspots of Evolutionary History in Systematic Conservation Planning for European Mammals.....</b>	<b>265</b>
Anni Arponen and Laure Zupan	
<b>Priorities for Conservation of the Evolutionary History of Amphibians in the Cerrado.....</b>	<b>287</b>
Débora Leite Silvano, Paula Hanna Valdujo, and Guarino Rinaldi Colli	
<b>Global Spatial Analyses of Phylogenetic Conservation Priorities for Aquatic Mammals .....</b>	<b>305</b>
Laura J. May-Collado, Carlos Zambrana-Torrelío, and Ingi Agnarsson	
<b>Metapopulation Capacity Meets Evolutionary Distinctness: Spatial Fragmentation Complements Phylogenetic Rarity in Prioritization .....</b>	<b>319</b>
Jessica K. Schnell and Kamran Safi	
<b>Patterns of Species, Phylogenetic and Mimicry Diversity of Clearwing Butterflies in the Neotropics .....</b>	<b>333</b>
Nicolas Chazot, Keith R. Willmott, André V.L. Freitas, Donna Lisa de Silva, Roseli Pellens, and Marianne Elias	

**Conservation of Phylogenetic Diversity in Madagascar’s  
Largest Endemic Plant Family, Sarcolaenaceae ..... 355**  
Anaëlle Soulebeau, Roseli Pellens, Porter P. Lowry II, Xavier Aubriot,  
Margaret E.K. Evans, and Thomas Haevermans

**The Future of Phylogenetic Systematics in Conservation  
Biology: Linking Biodiversity and Society ..... 375**  
Roseli Pellens, Daniel P. Faith, and Philippe Grandcolas

**Index..... 385**



# Contributors

**Ingi Agnarsson** Department of Biology, University of Vermont, Burlington, VT, USA

**Antje Ahrends** Royal Botanic Garden Edinburgh, Edinburgh, UK

**Anni Arponen** Department of Biosciences, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland

**Xavier Aubriot** Institut de Systématique, Evolution, Biodiversité, ISYEB – UMR 7205 CNRS MNHN UPMC EPHE, Muséum National d'Histoire Naturelle, Sorbonne Universités, Paris, France

Department of Life Sciences, Natural History Museum, London, UK

**Anne Chao** Institute of Statistics, National Tsing Hua University, Hsin-Chu, Taiwan

**Nicolas Chazot** Institut de Systématique, Evolution, Biodiversité, ISYEB – UMR 7205 CNRS MNHN UPMC EPHE, Muséum National d'Histoire Naturelle, Sorbonne Universités, Paris, France

**Olga Chernomor** Center for Integrative Bioinformatics Vienna, Max F. Perutz Laboratories, University of Vienna, Medical University of Vienna, Vienna, Austria

Bioinformatics and Computational Biology, Faculty of Computer Science, University of Vienna, Vienna, Austria

**Chun-Huo Chiu** Institute of Statistics, National Tsing Hua University, Hsin-Chu, Taiwan

**Guarino Rinaldi Colli** Departamento de Zoologia, Universidade de Brasília, Brasília, DF, Brazil

**T. Jonathan Davies** Department of Biology, McGill University, Montréal, Québec, Canada

**Donna Lisa de Silva** Institut de Systématique, Evolution, Biodiversité, ISYEB – UMR 7205 CNRS MNHN UPMC EPHE, Muséum National d'Histoire Naturelle, Sorbonne Universités, Paris, France

**Marianne Elias** Institut de Systématique, Evolution, Biodiversité, ISYEB – UMR 7205 CNRS MNHN UPMC EPHE, Muséum National d'Histoire Naturelle, Sorbonne Universités, Paris, France

**Margaret E.K. Evans** Institut de Systématique, Evolution, Biodiversité, ISYEB – UMR 7205 CNRS MNHN UPMC EPHE, Muséum National d'Histoire Naturelle, Sorbonne Universités, Paris, France

Laboratory of Tree Ring Research, University of Arizona, Tucson, AZ, USA

**Daniel P. Faith** AMRI, The Australian Museum, Sydney, NSW, Australia

**André V.L. Freitas** Departamento de Zoologia and Museu de História Natural, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, São Paulo, Brazil

**Philippe Grandcolas** Institut de Systématique, Evolution, Biodiversité, ISYEB – UMR 7205 CNRS MNHN UPMC EPHE, Muséum National d'Histoire Naturelle, Sorbonne Universités, Paris, France

**Thomas Haevermans** Institut de Systématique, Evolution, Biodiversité, ISYEB – UMR 7205 CNRS MNHN UPMC EPHE, Muséum National d'Histoire Naturelle, Sorbonne Universités, Paris, France

**Peter M. Hollingsworth** Royal Botanic Garden Edinburgh, Edinburgh, UK

**Lou Jost** EcoMinga Foundation, Baños, Tungurahua, Ecuador

**Steffen Klaere** Department of Statistics, School of Biological Sciences, University of Auckland, Auckland, New Zealand

**Christopher Lean** Philosophy Programme, Australian National University, Canberra, Australia

**Porter P. Lowry II** Institut de Systématique, Evolution, Biodiversité, ISYEB – UMR 7205 CNRS MNHN UPMC EPHE, Muséum National d'Histoire Naturelle, Sorbonne Universités, Paris, France

Missouri Botanical Garden, St. Louis, MO, USA

**James Maclaurin** Department of Philosophy, University of Otago, Dunedin, New Zealand

**Laura J. May-Collado** Department of Biology, University of Vermont, Burlington, VT, USA

**Bui Quang Minh** Center for Integrative Bioinformatics Vienna, Max F. Perutz Laboratories, University of Vienna, Medical University of Vienna, Vienna, Austria

**Daniel Rafael Miranda-Esquivel** Escuela de Biología, Universidad Industrial de Santander, Santander, Colombia

**Mary Morgan-Richards** Ecology Group, Institute of Agriculture and Environment, Massey University, Palmerston North, New Zealand

**David A. Nipperess** Department of Biological Sciences, Macquarie University, NSW, Australia

**Roseli Pellens** Institut de Systématique, Evolution, Biodiversité, ISYEB – UMR 7205 CNRS MNHN UPMC EPHE, Muséum National d'Histoire Naturelle, Sorbonne Universités, Paris, France

**Kamran Safi** Department of Migration and Immuno-ecology, Max Planck Institute for Ornithology, Radolfzell, Germany

Department of Biology, University of Konstanz, Konstanz, Germany

**Jessica K. Schnell** Department of Migration and Immuno-ecology, Max Planck Institute for Ornithology, Radolfzell, Germany

Department of Biology, University of Konstanz, Konstanz, Germany

**Débora Leite Silvano** Laboratório de Zoologia – Universidade Católica de Brasília, Brasília, DF, Brazil

**Anaëlle Soulebeau** Institut de Systématique, Evolution, Biodiversité, ISYEB – UMR 7205 CNRS MNHN UPMC EPHE, Muséum National d'Histoire Naturelle, Sorbonne Universités, Paris, France

**Steven A. Trewick** Ecology Group, Institute of Agriculture and Environment, Massey University, Palmerston North, New Zealand

**Paula Hanna Valdujo** Laboratório de Ecologia da Paisagem WWF-Brasil, Brasília, DF, Brazil

**Arndt von Haeseler** Center for Integrative Bioinformatics Vienna, Max F. Perutz Laboratories, University of Vienna, Medical University of Vienna, Vienna, Austria

Bioinformatics and Computational Biology, Faculty of Computer Science, University of Vienna, Vienna, Austria

**Keith R. Willmott** McGuire Center for Lepidoptera and Biodiversity, Florida Museum of Natural History, University of Florida, Gainesville, FL, USA

**Kowiyou Yessoufou** Department of Environmental Sciences, University of South Africa, Florida, Gauteng, South Africa

**Carlos Zambrana-Torrel** EcoHealth Alliance, New York, NY, USA

**Laure Zupan** Laboratoire d'Ecologie Alpine, UMR-CNRS 5553, Université J. Fourier, Grenoble cedex 9, France





## About the Editors

**Roseli Pellens** is a research engineer in Macroecology at Muséum National d'Histoire naturelle, Paris (Institut de Systématique, Evolution, Biodiversité). She moved from a doctorate of geography and ecology at the Universidade Federal do Rio de Janeiro, where she studied the effect of forest fragmentation in the Atlantic rainforest, to the studies of systematics and evolution in Paris Museum. Presently, she is in charge of developing macroecological studies with collectives of highly skilled systematists and phylogeneticists through the use of big systematic data sets from museum collections. These studies are aimed at answering questions about the major patterns of biodiversity distribution and proposing strategies for biodiversity conservation.

**Philippe Grandcolas** is a senior scientist CNRS. Originally trained as an ecologist, he worked in a laboratory of ethology (Université de Rennes) and then in a laboratory of systematics (Muséum national d'Histoire naturelle, Paris). His research agenda focused on the improvement of knowledge integration and logical reasoning in the framework linking phylogenetics, taxonomy and evolutionary biology. His past and present responsibilities and memberships in national committees, international societies, CBD and GBIF echo his research agenda, for favouring the development and integration of systematics in the field of life sciences. He is presently the Head of the Institut de Systématique, Evolution, Biodiversité. This institute is a laboratory of the Muséum national d'Histoire naturelle, the Centre National de la Recherche Scientifique, the Université Pierre et Marie Curie and the Ecole Pratique des Hautes Etudes, and it comprises most of the systematists in France.

# Phylogenetics and Conservation Biology: Drawing a Path into the Diversity of Life

Roseli Pellens and Philippe Grandcolas

**Abstract** In the midst of a major extinction crisis, the scientific community is called to provide criteria, variables and standards for defining strategies of biodiversity conservation and monitoring their results. Phylogenetic diversity is one of the variables taken in account. Its consideration in biodiversity conservation stemmed from the idea that species are not equal in terms of evolutionary history and opened a completely new line of investigation. It has turned the focus to the need of protecting the Tree of Life, i.e. the diversity of features resulting from the evolution of Life on Earth. This approach is now recognized as a strategy for increasing options for future needs and values as well as for increasing the potential of biodiversity diversification in a future environment. Since its introduction in biodiversity conservation thinking much has been developed in order to compose our conceptual understanding of the importance of protecting the Tree of Life. The aim of this book is to contribute to the ongoing international construction of strategies for reducing biodiversity losses by exploring several approaches for the conservation of phylogenetic diversity. We hope that this concentrated effort will contribute to the emergence of new solutions and attitudes towards a more effective preservation of our evolutionary heritage. The chapters of this book are organized around three main themes: questions, methods and applications, providing a condensed updated picture of the state of the art and showing that either conceptually or methodologically phylogenetic diversity has everything to be on the global agenda of biodiversity conservation.

**Keywords** Tree-of-life • Sixth mass extinction • Evolutionary heritage • Biodiversity monitoring • Essential biodiversity variable

---

R. Pellens (✉) • P. Grandcolas

Institut de Systématique, Evolution, Biodiversité, ISYEB – UMR 7205 CNRS MNHN  
UPMC EPHE, Muséum National d'Histoire Naturelle, Sorbonne Universités,  
45 rue Buffon, CP 50, 75005 Paris, France  
e-mail: [pellens@mnhn.fr](mailto:pellens@mnhn.fr); [pg@mnhn.fr](mailto:pg@mnhn.fr)

© The Author(s) 2016

R. Pellens, P. Grandcolas (eds.), *Biodiversity Conservation and Phylogenetic Systematics*, Topics in Biodiversity and Conservation 14,  
DOI 10.1007/978-3-319-22461-9\_1

During the last centuries and more dramatically in the last four decades, natural habitats were destroyed at rates much higher than ever observed in human history. All biomes were affected, but those located in tropical regions were more impacted, particularly because policies for the development and appropriation of these territories were emphasized during this period. Nonetheless, the massive transformation of these landscapes to give place to crops and towns multiplied species' losses and vulnerability at incredible rates (Millennium Ecosystem Assessment 2005), mostly due to the fact that most of world's biodiversity is concentrated around the tropics (Gaston 2000). In addition to habitat destruction and fragmentation, natural ecosystems were also submitted to high levels of pollution, overexploitation of forestry and fishery resources, invasive species, and to the effects of climate changes mainly provoked by man-induced greenhouse gas emissions. As a result, a high number of species were already extinct and others have suffered severe populations declines (Mace et al. 2005), with many advancing at high speed to higher categories of threat every year (e.g., Hoffmann et al. 2010). So, recent scenarios integrating main extinction drivers suggest that rates of extinction are likely to rise by at least a further order of magnitude over the next few centuries (Mace et al. 2005; Pereira et al. 2010; Barnosky et al. 2012; Proença and Pereira 2013).

This critical situation is now recognized as the “sixth mass extinction”, i.e. the sixth period in the history of life in which more than three-quarters of the living species is lost in a short geological interval (Barnosky et al. 2011). Compared to the first “big five”, this extinction period has the peculiarity of being caused mainly by the way of living of one single species, the humans. Counteracting this trend is perhaps the biggest ethic, political and scientific challenge of our times (Sarkar 2005), as the time for action is short, funds for biodiversity conservation are far from below the real needs (e.g., McCarthy et al. 2012), uncertainties are enormous (Forest et al. 2015), and the solution of conflicts with main-trend ways-of-living and main patterns of distribution and consumption (e.g., Lenzen et al. 2012) often takes much longer than habitat destruction.

In the race to combat extinctions, there is urgency for increasing conservation worldwide. The scientific community is pressed to provide criteria in order to define priorities, as well as for indicating variables and standards that allows for monitoring the evolution of biodiversity in the light of these strategies (Hoffmann et al. 2010; Pereira et al. 2010, 2013; Mace et al. 2010, 2014). Traditionally, biodiversity conservation was based on species counts, valuing sites in terms of species richness, number of endemics and number of threatened species (Myers et al. 2000; Myers 2003; Kier et al. 2009). However, in spite of its generalized use, this kind of data can be very heterogeneous making very difficult comparisons across taxonomic groups, along time and among sites, as species richness can be influenced by many factors, going from the species concept to the spatial scale and sampling effort (see Gaston 1996 for an overview on this subject). Similarly, in spite of the great interest of Red Lists of species' threats, such as that from IUCN (International Union for Conservation of Nature), to indicate imminent risks of extinction, concentrating conservation-limited resources on threatened species can be very risky and these limits must be considered (Possingham et al. 2002). Moreover, measures based on species counts also have the limitation of considering all species as equals, being

blind to particular functional roles in the ecosystem, to associations in communities, or to their evolutionary history.

The contribution of phylogenetic systematics to this debate stemmed from this idea that species are not equal and from the possibility of characterization in terms of evolutionary history (Vane-Wright et al. 1991; Faith 1992). Systematics addresses the interrelatedness of organisms in terms of shared inherited and original features (Hennig 1966; Eldredge and Cracraft 1980; Wiley 1981). This old but recently revived science moved from describing and classifying the living beings in the eighteenth century to macro-evolutionary biology in the twentieth century with modern phylogenetics (O'Hara 1992). Phylogenies are trees of history, showing both the species relationships and the evolution of sets of characters. They are the basis for organizing and retrieving all current knowledge about biodiversity, either structural or functional in an evolutionary context.

The consideration of phylogenetic systematics in biodiversity conservation opened a completely new line of investigation as it has turned the focus to the need of protecting the Tree of Life, i.e. the diversity of features resulting from the evolution of Life on Earth (Mace et al. 2003; Purvis et al. 2005; Mace and Purvis 2008; MacLaurin and Sterelny 2008; Forest et al. 2015). Since its introduction in biodiversity conservation thinking much has been developed in order to compose our present conceptual understanding of the importance of protecting the Tree of Life. Several methodological issues were developed and refined; the input of phylogenetic diversity in comparison with species richness was assessed in different ways; several studies attempting to prioritize species and areas for conservation were developed; the relationship between the losses of evolutionary history with extinctions was studied in different contexts; and different new concepts emerged (see Table 1).

## Glossary

**Biodiversity:** is a very inclusive term formed by contraction of “biological diversity.” In this book, we use this term to express the variety of life, often willing to express the integrative definition of the Convention on Biological Diversity in which “Biological diversity” means “the variability among living organisms from all sources including, *inter alia*, terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are part; this includes diversity within species, between species and of ecosystems”.

**Evolutionary history:** the chronicle of the process whereby the diversity of life is built.

**Phylogenetic Systematics:** the scientific discipline describing and naming the different organisms, assessing their relatedness in the Tree of Life and proposing subsequent classifications. Species phylogenetic relationships are assessed on the basis of originally shared characters modified during evolution.

**Tree of life:** an old metaphor to describe the interrelatedness of all organisms (living and extinct), based on their evolutionary history.

**Table 1** Some examples of studies linking phylogenetic systematics and biodiversity conservation

Problems	Examples
Development of methods and measures to assess taxonomic or evolutionary distinctiveness or phylogenetic diversity	Vane-Wright et al. 1991; May 1990; Faith 1992; Posadas et al. 2001; Pavoine et al. 2005; Redding and Mooers 2006; Isaac et al. 2007; Steel et al. 2007; Hartmann and Steel 2007; Lozupone and Knight 2005; Rosauer et al. 2009; Cadotte and Davies 2010; Chao et al. 2010
Comparison of phylogenetic measures	Schweiger et al. 2008; Davies and Cadotte 2011; Pio et al. 2011
Comparison of phylogenetic diversity to traditional measures	Polasky et al. 2002; Rodrigues and Gaston 2002; Rodrigues et al. 2005, 2011; Hartmann and André 2013
Inclusion of phylogenetics in systematic conservation planning	Walker and Faith 1994; Arponen 2012
Prioritization of areas for the conservation of evolutionary history	Posadas et al. 2001; Lehman 2006; McGoogan et al. 2007; López-Osorio and Miranda-Esquivel 2010; Forest et al. 2007; Buerki et al. 2015; Pollock et al. 2015; Zupan et al. 2014
Prioritization of species	Weitzman 1998; Isaac et al. 2007; Kuntner et al. 2011; Redding et al. 2015
Relationship between extinctions and the loss of phylogenetic diversity	Nee and May 1997; Purvis 2008; Davies et al. 2008; Fritz et al. 2009; Fritz and Purvis 2010; Magnuson-Ford et al. 2010; Jono and Pavoine 2012; Yessoufou et al. 2012; Davies 2015; Faith 2015; Gudde et al. 2013; Huang and Roy 2015
Climate change and the loss of phylogenetic diversity	Faith and Richards 2012; Thuiller et al. 2011, 2015
Phylogenetic and functional diversity	Safi et al. 2011; Huang et al. 2012
Cost of conserving phylogenetic diversity	Weitzman 1998; Nunes et al. 2015
Development of key concepts related to biodiversity conservation that integrates phylogenetic diversity	Evolutionary heritage (Mooers et al. 2005)
	Phylogenetic diversity and option values (Faith 1992; Steel et al. 2007; Forest et al. 2007)
	Ecosystem services (Faith et al. 2010)
	Key biodiversity areas for conservation (Brooks et al. 2015)
	Phylogenetic planetary boundaries and tipping points (Faith et al. 2010)

Please note that these are leading marks: most of these researches approached more than one of these problems

The main aim of this book is to contribute to the ongoing international search for reducing biodiversity losses in this critical period for life on Earth by exploring several approaches for the conservation of phylogenetic diversity. As shown in Table 1, the universe of problems to be prospected in this subject is quite large and could not fit in a single volume. In spite of that, here we provide a condensed updated picture of the state of the art showing that either conceptually or methodologically phylogenetic diversity has everything to be on the global agenda of biodiversity conservation. This book is organized around three main themes: questions, methods and applications. We hope that this concentrated effort will contribute to

the emergence of new solutions and attitudes towards a more effective preservation of our evolutionary heritage.

## Questions

This first section is composed of chapters addressing some central questions concerning the links between biodiversity conservation and phylogenetic systematics. The first, and perhaps the most important of these questions, concerns the nature of the role of phylogenetic systematics in conservation efforts. How do we value the Tree of Life? Why to use aspects of phylogeny in preference to other biodiversity variables? These questions are explored by Lean and Maclaurin in chapter “[The Value of Phylogenetic Diversity](#)”. They develop the idea that phylogenetic diversity plays a unique role in underpinning conservation endeavor and represents the foundation of a general measure of biodiversity. In a synthesis about the reasons and the types of values that should guide biodiversity conservation and qualify a general biodiversity measure, they propose that phylogeny is the only basis for large-scale conservation prioritization. They justify this argument by showing that phylogeny is the only guide for maximizing feature diversity (*sensu* Faith 1992) across many different taxa, and also is the best way to hedge our bets against uncertainties related to environmental changes and to human’s future needs and values.

### Glossary

**PD or Faith’s PD:** is the measure of phylogenetic diversity created by Faith (1992). Specifically it is the sum of the lengths of all phylogenetic branches (from the root to the tip) spanned by a set of species. In this book, we refer to PD or Faith’s PD to indicate this measure.

**Phylogenetic diversity:** all over this book we use this term in very large sense, independently of the measure, willing to express the differences between organisms due to their evolutionary history, and so captured by a phylogeny. It can be used to express the uniqueness of one species or the representativeness of a set of organisms, according to several different measures.

**Evolutionary distinctiveness** (Isaac et al. 2007) or **Evolutionary distinctness:** is here used to indicate measures destined to assess the phylogenetic diversity of each species, independently if it is based on topology or branch length. Contrarily to PD, where the contribution of a species may vary from one set to another depending on the other species occurring in it, with measures of evolutionary distinctiveness each species has an invariable value.

**Taxonomic distinctiveness** (Vane-Wright et al. 1991): like in the case of Evolutionary distinctiveness, it is used to express measures designed to assess the phylogenetic diversity of species, but this definition is restricted to those measures based on tree topology.

If the way we value phylogenetic diversity is central for any justifications for including phylogeny in conservations efforts, an equally important consideration must be the choice of the measure that adequately captures the aspects of phylogenetic diversity that are important for conservation. Lean and MacLaurin propose that this measure should maximize feature diversity. However, there are very few studies comparing the performance of the measures under such criteria (Redding and Mooers 2006; Schweiger et al. 2008; Pio et al. 2011). Dan Faith (chapter “[The PD Phylogenetic Diversity Framework: Linking Evolutionary History to Feature Diversity for Biodiversity Conservation](#)”) addresses this question through the comparison of PD (Faith 1992), in relation to several measures of Evolutionary Distinctiveness (ED) in the context of priority setting for conservation. The core of Dan’s analysis is complementarity (marginal gains and losses of PD or feature diversity), an attribute intrinsic to PD’s algorithm, but lacking in ED measures. Here he shows that PD complementarity allows the identification of sets of species with maximum PD, whereas ED indices are unable to reliably identify such diverse sets.

The next contribution deals with the loss of phylogenetic diversity with extinction. Are there phylogenetic signals in extinctions? What is the role of extrinsic and intrinsic factors in extinctions, and what is the role of phylogeny in data exploration and analysis (Grandcolas et al. 2010)? Are extinction drivers similar to different groups of organisms? What is the role of evolutionary models in the patterns observed? These questions are here explored by Yessoufou and Davies (chapter “[Reconsidering the Loss of Evolutionary History: How Does Non-random Extinction Prune the Tree-of-Life?](#)”). They first review the main extinction drivers, showing that the most relevant might be quite different among vertebrates, invertebrates and plants. By exploring how non-random extinction prunes the Tree of Life under different models of evolution, they call our attention to the fact that the model of evolution is likely to be a key explanatory of the loss of evolutionary history. They also argue that more branches are likely to be lost from the Tree of Life under the speciation model of evolution.

Many of our considerations about the conservation of the Tree of Life are based on our knowledge of a micro-fraction of the living world, given that we often focus on organisms that are very close to human eyes, like vertebrates, vascular plants, and a few emblematic insects. Likewise, most of the phylogenies used to this purpose are based on molecular data, very often on very small sets of short gene sequences. An advantage of molecular data for phylogenetic inference is provision of a standardized set of characters, often reflecting the main patterns of relationship of the species in a group of organisms. However, the extent to which these genes portions evolve and reflect the evolution of other traits is seldom well studied. Such an issue is central to arguments that phylogenetic diversity links to general feature diversity. These problems are explored by Steve Trewick and Mary Morgan-Richards (chapter “[Phylogenetics and Conservation in New Zealand: The Long and the Short of It](#)”). With examples of the phylogenetic position (as assessed through molecular data) of some legendary organisms from New Zealand such as Kākāpō, takahē and tuatara, they shake some established views about the extent molecular branch length reflects other extraordinary ecological, morphological or behavioral



traits. Going further, they turn our lenses to the microscopic life that is much more deeply branched in the Tree of Life. Taking the example of marine sponges, they show that a single sponge provides an environment that can host several distinct microbial communities (microbiomes) and so preserve organisms from more than 40 phyla all branched much deeper than vertebrates and plants. At reading this chapter, we are guided to a more inclusive perspective of biodiversity and we can find more reasons for protecting Kākāpō, takahē, tuatara, marine sponges and... microbes.

Relict species are often presented as examples of important species for the conservation of phylogenetic diversity. Everyone has heard about Coelacanth and Platypus as examples of unique evolutionary histories. In spite of this, the concept of relict species is still plagued with misleading ideas and uses, potentially causing misunderstandings for the use of phylogenetic diversity in general. Philippe Grandcolas and Steve Trewick (chapter “[What Is the Meaning of Extreme Phylogenetic Diversity? The Case of Phylogenetic Relict Species](#)”) aim at freeing the concept from these problems, and use the extreme case of relict species to explore the nature and the use of phylogenetic diversity. The study of relicts helps understanding that early-branching species that make high values of phylogenetic diversity (the “unique PD” of Forest et al. 2015) are not necessarily evolutionarily “frozen”. Their conservation is not only aimed at retaining Life’s diversity but also at keeping evolutionary potential. It is also worth-mentioning that such species have often been empirically shown to have special extinction risks, highlighting again the important role of phylogenetic diversity in conservation biology.

## Methods

In this section we introduce the set of contributions dealing with methodology *sensu stricto*. It starts with two papers dealing with different possibilities of applications and extensions of the PD framework in community assessments, area comparisons and long-term monitoring of biodiversity changes. In chapter “[Using Phylogenetic Dissimilarities Among Sites for Biodiversity Assessments and Conservation](#)”, Dan Faith details one possible extension of the PD family of measures, the Environmental Dissimilarity (*ED*) methods. While PD assumes that shared ancestry accounts for shared features among taxa, *ED* attempts to account for shared features through shared habitat/environment among taxa, thus including those shared features not explained by shared ancestry. With some graphical examples Dan shows how *ED* works. Further, he synthesizes a set of *ED*-based measures. These include *ED* complementarity measures designed with the similar aim of calculating and predicting features gains and losses as we gain or lose areas in conservation planning. He concludes by indicating that *ED* methods appear to offer a robust framework for global assessments and for long-term monitoring of biodiversity change.

In chapter “[Phylogenetic Diversity Measures and Their Decomposition: A Framework Based on Hill Numbers](#)”, Anne Chao, Chun-Huo Chiu and Lou Jost

develop a set of tools for integrating species abundances in PD calculations. This proposition enlarges the range of applications of the PD framework, making it a very useful tool for monitoring changes in biodiversity and warning about important changes in abundance before species become actually extinct. This framework is based on Hill numbers, describing the “effective number of species” found in a sample or region. Here Chao et al. provide a rich overview of abundance-based diversity measures and their phylogenetic generalizations, the framework of Hill numbers, phylogenetic Hill numbers and related phylogenetic diversity measures. They also review the diversity decomposition based on phylogenetic diversity measures and present the associated phylogenetic similarity and differentiation. With a real example, they illustrate how to use phylogenetic similarity (or differentiation) profiles to assess phylogenetic resemblance or difference among multiple assemblages either in space or time.

Phylogenetic reconstructions often result in different near-optimal alternative trees, particularly due to conflicting information among different characters. What do we do as conservation biologists when the phylogenetic reconstruction leads to multiple trees with conflicting signals? This problem is here addressed by a contribution by Olga Chernomor et al. (chapter “[Split Diversity: Measuring and Optimizing Biodiversity Using Phylogenetic Split Networks](#)”) with a proposition of combining the concepts of phylogenetic diversity and split networks in a single concept of phylogenetic split diversity. They show how split diversity works and design its application and the computation solution in biodiversity optimization for some well-known problems of taxon selection and reserve selection, exploring how to include taxon viability and budget in this kind of analysis.

The extent to which sampling effort might influence the rank of conservation priorities is long recognized as a central issue in selecting areas for conservation (Mace and Lande 1991; Mckinney 1999; Régnier et al. 2009), but has so far remained practically untouched in the study of conservation of phylogenetic diversity. Here we have the opportunity to present three different approaches to this problem. The convergence of these independent studies shows the importance of this subject and the recognition of the urgency of searching for solutions. In chapter “[The Rarefaction of Phylogenetic Diversity: Formulation, Extension and Application](#)”, David Nipperess deals with this question in the PD framework by further developing the rarefaction of PD first proposed by Nipperess and Matsen (2013). Here he provides a detailed formulation for the exact analytical solution for expected (mean) Phylogenetic Diversity for a given amount of sampling effort in which whole branch segments are selected under rarefaction. In addition, he extends this framework to show how the initial slope of the rarefaction curve ( $\Delta PD$ ) can be used as a flexible measure of phylogenetic evenness, phylogenetic beta-diversity or phylogenetic dispersion, depending on the unit of accumulation.

In chapters “[Support in Area Prioritization Using Phylogenetic Information](#)” and “[Assessing Hotspots of Evolutionary History with Data from Multiple Phylogenies: An Analysis of Endemic Clades from New Caledonia](#)”, the question of resampling and support of the dataset for defining priority areas is studied in the framework of evolutionary distinctiveness (ED). In chapter “[Support in Area Prioritization Using](#)

**Phylogenetic Information**", Daniel Rafael Miranda-Esquivel develops one scheme to verify the support for area ranking using a jackknife resampling strategy. In this proposition, one can evaluate the more adequate index and the support of the area ranking with different probability values when deleting phylogenies, and/or areas and/or species. In chapter "**Assessing Hotspots of Evolutionary History with Data from Multiple Phylogenies: An Analysis of Endemic Clades from New Caledonia**", we and our collaborators Antje Ahrends and Pete Hollingsworth, propose a scheme for solving the problem of sampling bias in datasets with phylogenies coming from independent and so, non-standardized, spatial sampling. We use the rarefaction of phylogenies to assess the role of the number of phylogenies, of species richness and of the influence of individual phylogenies on site's scores. And then we design a resampling strategy using multiple phylogenies to verify the stability of the results. This method is applied to the case of New Caledonia, a megadiverse island with all locations equally rich in microendemic species and where phylogenetic diversity is especially helpful to determine conservation priorities among sites.

## Applications

This last section is composed by contributions exploring the application of phylogenetic diversity methods in study cases. These studies are deliberately diverse in approaches of the use and applications of phylogenetic diversity, and of measures, spatial scales, geographic locations and taxonomic groups as well. It starts with two analyses integrating the conservation of evolutionary history in systematic conservation planning, a field of conservation biology that deals with conservation prioritization taking in account multiple factors, and in which we can define and revise pre-established criteria and goals (Margules and Pressey 2000; Ball et al. 2009; Moilanen et al. 2009; Kukkala and Moilanen 2013).

In chapter "**Representing Hotspots of Evolutionary History in Systematic Conservation Planning for European Mammals**" Arponen and Zupan use the Zonation software for spatial prioritization to prioritize areas for conservation of the evolutionary history of mammals in Europe. With an analysis at continental and at the scale of each European country, they show that: (a) a strategy focusing only on species richness would miss some areas with important levels of evolutionary history, mainly in regions with medium or low values of species richness; (b) the present system of protected areas performs worse than random selections for protecting the evolutionary history of mammals; and (c) a strategy to protect mammals at the continental scale would be much more effective than separated strategies for each country, although from a political point of view this last one is likely to be more feasible.

In the following contribution, Silvano et al. (chapter "**Priorities for Conservation of the Evolutionary History of Amphibians in the Cerrado**") use a Gap Analysis to evaluate the protection status of 82 anuran species endemic from Brazilian Cerrado and to define priority areas for their conservation. Their results indicate an alarming

situation in which 39 (48 %) endemic and restricted range species are completely unprotected, among them species with very high ED values, and other 43 (52 %) are gap species with less than 20 % of their targets met. The priority areas for the conservation of these species mostly occupy the central portion of the biome, a region that already suffered major habitat destruction, and are forecast to undergo important habitat loss if economic scenario remains unchanged.

The following triad of studies explores the integration of species threat and phylogenetic diversity. It starts with the research of May-Collado, Zambrana-Torrelío and Agnarsson (chapter “[Global Spatial Analyses of Phylogenetic Conservation Priorities for Aquatic Mammals](#)”) dealing with the prioritization of areas for conservation of 127 marine mammals worldwide. Here they use the EDGE (Isaac et al. 2007) and HEDGE (Steel et al. 2007) measures to provide the first spatial analysis for phylogenetic conservation priorities incorporating threat information at global scale. By assessing conservation under “pessimistic” and “optimistic” IUCN extinction scenarios they show how fragile is the world system of protected areas to conserve the evolutionary distinctiveness of marine mammals. They identified 22 Conservation Priority Areas all over the world and showed that only 11.5 % of them overlap with existing marine protected areas. Their results complete prior findings on conservation prioritization for marine mammals, providing a helpful tool for the Conservation of Biological Diversity plan to protect 10 % of world’s marine and coastal regions by 2020.

In the next contribution, Jessica Schnell and Kamran Safi (chapter “[Metapopulation Capacity Meets Evolutionary Distinctness: Spatial Fragmentation Complements Phylogenetic Rarity in Prioritization](#)”) design a framework to predict threat status of Data Deficient and Least Concern species. They propose to combine evolutionary distinctiveness with metapopulation capacity derived from habitat isolation. Here they apply this framework to terrestrial mammals endemic of oceanic islands worldwide, and show that balancing between extinction risks associated to island’s isolation and potential loss of evolutionarily unique species can be very useful to characterize conservation status of island endemic species. Based on it they show that islands such as Guadalcanal, Isle of Pines, Madagascar and Ngqela Sule are very representative for reducing the extinction of mammals with high ED values.

In chapter “[Patterns of Species, Phylogenetic and Mimicry Diversity of Clearwing Butterflies in the Neotropics](#)”, Chazot et al. explore the patterns of distribution of several features of diversity of three genera of ithomiine butterflies in Neotropical Region. Ithomiine display Müllerian mimetism and numerically dominate many butterfly assemblages across the Neotropics, probably conditioning the distribution of other species that interact with them in positive or negative way. So, the loss of ithomiine species in local assemblages may strongly influence the vulnerability of butterfly assemblages. Here they show that, on the one hand, the pattern of distribution of phylogenetic diversity, species richness, and mimicry diversity are highly congruent within genera, and, in a lesser extent, across genera. On the other hand, the potential loss of species due to disruption of mimicry rings, as captured by a measure of vulnerability designed in this study, are not evenly distributed across genera presenting peaks in areas completely distinct of those observed to the other

features. This is a good example of the “agony of choice” of Vane-Wright et al. (1991) illustrating the difficulty of finding an optimal solution in situations in which several parameters account for the existing biodiversity.

We close this section with a note of optimism. The analysis of Soulebeau et al. (chapter “[Conservation of Phylogenetic Diversity in Madagascar’s Largest Endemic Plant Family, Sarcolaenaceae](#)”) shows that the system of protected areas of Madagascar is likely to protect all lineages and 97 % of the phylogenetic diversity of Sarcolaenaceae, the largest endemic plant family of this island. This result is particularly important because neither Sarcolaenaceae nor phylogenetic diversity were specifically considered in the conception or in the recent expansion of Madagascar’s network of protected area (Kremen et al. 2008), showing that a large system of protected area may capture much more biodiversity components and features than originally expected.

For concluding, in the last chapter we – Roseli Pellens, Dan Faith and Philippe Grandcolas – describe the recent transformations of phylogenetic systematics in the light of new facilities of molecular sequencing and data analysis, and discuss its impacts in biological conservation. We finish by exploring the possibility of defining “planetary boundaries” for biodiversity on the basis of phylogenetic diversity, and its important role in linking biodiversity into broader societal perspectives and needs.

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Arponen A (2012) Prioritizing species for conservation planning. *Biodivers Conserv* 21(4):875–893. doi:[10.1007/S10531-012-0242-1](#)
- Ball IR, Possingham HP, Watts M (2009) Chapter 14: Marxan and relatives: software for spatial conservation prioritisation. In: Moilanen A, Wilson KA, Possingham HP (eds) *Spatial conservation prioritisation: quantitative methods and computational tools*. Oxford University Press, Oxford, pp 185–195
- Barnosky AD, Matzke N, Tomiya S et al (2011) Has the Earth’s sixth mass extinction already arrived? *Nature* 471(7336):51–57. doi:[10.1038/Nature09678](#)
- Barnosky AD, Hadly EA, Bascompte J et al (2012) Approaching a state shift in Earth’s biosphere. *Nature* 486(7401):52–58. doi:[10.1038/Nature11018](#)
- Brooks TM, Cuttelod A, Faith DP, Garcia-Moreno J, Langhammer P, Perez-Espona S (2015) Why and how might genetic and phylogenetic diversity be reflected in the identification of key biodiversity areas? *Philos Trans R Soc B* 370(1662). doi: [10.1098/Rstb.2014.0019](#)
- Buerki S, Callmander MW, Bachman S et al (2015) Incorporating evolutionary history into conservation planning in biodiversity hotspots. *Philos Trans R Soc B* 370(1662). doi: [10.1098/Rstb.2014.0014](#)
- Cadotte MW, Davies TJ (2010) Rarest of the rare: advances in combining evolutionary distinctiveness and scarcity to inform conservation at biogeographical scales. *Divers Distrib* 16(3):376–385. doi:[10.1111/J.1472-4642.2010.00650.X](#)

- Chao A, Chiu CH, Jost L (2010) Phylogenetic diversity measures based on Hill numbers. *Philos Trans R Soc B* 365(1558):3599–3609. doi:[10.1098/Rstb.2010.0272](https://doi.org/10.1098/Rstb.2010.0272)
- Davies TJ (2015) Losing history: how extinctions prune features from the tree of life. *Philos Trans R Soc B* 370(1662). doi:[10.1098/Rstb.2014.0006](https://doi.org/10.1098/Rstb.2014.0006)
- Davies TJ, Cadotte MW (2011) Quantifying biodiversity: does it matter what we measure? In: Zachos FE, Habel JC (eds) *Biodiversity hotspots: distribution and protection of conservation priority areas*. Springer, Berlin, pp 43–60
- Davies TJ, Fritz SA, Grenyer R et al (2008) Phylogenetic trees and the future of mammalian biodiversity. *Proc Natl Acad Sci U S A* 105:11556–11563
- Eldredge N, Cracraft J (1980) *Phylogenetic patterns and the evolutionary process: method and theory in comparative biology*. Columbia University Press, New York
- Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61(1):1–10
- Faith DP (2015) Phylogenetic diversity, functional trait diversity and extinction: avoiding tipping points and worst-case losses. *Philos Trans R Soc B* 370(1662). doi:[10.1098/Rstb.2014.0011](https://doi.org/10.1098/Rstb.2014.0011)
- Faith DP, Richards Z (2012) Climate change impacts on the tree of life: changes in phylogenetic diversity illustrated for *Acropora* corals. *Biology* 1:906–932
- Faith DP, Magallon S, Hendry AP et al (2010) Ecosystem services: an evolutionary perspective on the links between biodiversity and human well-being. *Curr Opin Environ Sustain* 2(1–2):66–74. doi:[10.1016/J.Cosust.2010.04.002](https://doi.org/10.1016/J.Cosust.2010.04.002)
- Forest F, Grenyer R, Rouget M et al (2007) Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature* 445(7129):757–760
- Forest F, Crandall KA, Chase MW, Faith DP (2015) Phylogeny, extinction and conservation: embracing uncertainties in a time of urgency. *Philos Trans R Soc B* 370(1662). ArtN 20140002. doi:[10.1098/Rstb.2014.0002](https://doi.org/10.1098/Rstb.2014.0002)
- Fritz SA, Purvis A (2010) Phylogenetic diversity does not capture body size variation at risk in the world's mammals. *Proc R Soc B Biol Sci* 277(1693):2435–2441. doi:[10.1098/Rspb.2010.0030](https://doi.org/10.1098/Rspb.2010.0030)
- Fritz SA, Bininda-Emonds ORP, Purvis A (2009) Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics. *Ecol Lett* 12(6):538–549. doi:[10.1111/J.1461-0248.2009.01307.X](https://doi.org/10.1111/J.1461-0248.2009.01307.X)
- Gaston KJ (1996) Species richness: measures and measurements. In: Gaston KJ (ed) *Biodiversity: a biology of numbers and differences*. Blackwell Science, Oxford, pp 77–113
- Gaston KJ (2000) Global patterns in biodiversity. *Nature* 405:220–227
- Grandcolas P, Nattier R, Legendre F, Pellens R (2010) Mapping extrinsic traits such as extinction risks or modelled bioclimatic niches on phylogenies: does it make sense at all? *Cladistics* 26:1–5
- Gudde RM, Joy JB, Mooers AO (2013) Imperilled phylogenetic endemism of Malagasy lemuriformes. *Divers Distrib* 19(7):664–675. doi:[10.1111/Ddi.12023](https://doi.org/10.1111/Ddi.12023)
- Hartmann K, André J (2013) Should evolutionary history guide conservation? *Biodivers Conserv* 22:449–458
- Hartmann K, Steel MA (2007) Phylogenetic diversity: from combinatorics to ecology. In: Gascuel O, Steel MA (eds) *Reconstructing evolution: new mathematical and computational advances*. Oxford University Press, Oxford
- Hennig W (1966) *Phylogenetic systematics*. University of Illinois Press, Urbana
- Hoffmann M, Hilton-Taylor C, Angulo A et al (2010) The impact of conservation on the status of the world's vertebrates. *Science* 330(6010):1503–1509. doi:[10.1126/Science.1194442](https://doi.org/10.1126/Science.1194442)
- Huang DW, Roy K (2015) The future of evolutionary diversity in reef corals. *Philos Trans R Soc B* 370(1662). ArtN 20140010. doi:[10.1098/Rstb.2014.0010](https://doi.org/10.1098/Rstb.2014.0010)
- Huang S, Stephens PR, Gittleman JL (2012) Traits, trees and taxa: global dimensions of biodiversity in mammals. *Proc R Soc B Biol Sci* 279(1749):4997–5003. doi:[10.1098/rspb.2012.1981](https://doi.org/10.1098/rspb.2012.1981)
- Isaac NJB, Turvey ST, Collen B, Waterman C, Baillie JEM (2007) Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PLoS ONE* 3:e296
- Jono CMA, Pavoine S (2012) Threat diversity will erode mammalian phylogenetic diversity in the near future. *PLoS ONE* 7(9). ARTN e46235. doi:[10.1371/journal.pone.0046235](https://doi.org/10.1371/journal.pone.0046235)



- Kier G, Kreft H, Lee TM et al (2009) A global assessment of endemism and species richness across island and mainland regions. *Proc Natl Acad Sci U S A* 106(23):9322–9327
- Kremen C, Cameron A, Moilanen A et al (2008) Aligning conservation priorities across taxa in Madagascar, a biodiversity hotspot, with high-resolution planning tools. *Science* 320:222–226
- Kukkala AS, Moilanen A (2013) Core concepts of spatial prioritisation in systematic conservation planning. *Biol Rev* 88(2):443–464. doi:[10.1111/Brv.12008](https://doi.org/10.1111/Brv.12008)
- Kuntner M, May-Collado LJ, Agnarsson I (2011) Phylogeny and conservation priorities of afrotherian mammals (Afrotheria, Mammalia). *Zool Scr* 40(1):1–15
- Lehman SM (2006) Conservation biology of Malagasy Strepsirrhines: a phylogenetic approach. *Am J Phys Anthropol* 130:238–253
- Lenzen M, Moran D, Kanemoto K et al (2012) International trade drives biodiversity threats in developing nations. *Nature* 486(7401):109–112. doi:[10.1038/Nature11145](https://doi.org/10.1038/Nature11145)
- López-Osorio F, Miranda Esquivel DR (2010) A phylogenetic approach to conserving Amazonian biodiversity. *Conserv Biol* 24(5):1359–1366
- Lozupone C, Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* 71(12):8228–8235
- Mace GM, Lande R (1991) Assessing extinction threats: toward a reevaluation of IUCN threatened species categories. *Conserv Biol* 5:148–157
- Mace GM, Purvis A (2008) Evolutionary biology and practical conservation: bridging a widening gap. *Mol Ecol* 17(1):9–19
- Mace GM, Gittleman JL, Purvis A (2003) Preserving the tree of life. *Science* 300:1707–1709
- Mace GM, Masundire H, Baillie JEM (2005) Millennium ecosystem assessment: current states and trends – biodiversity. Island Press, Washington, DC
- Mace GM, Cramer W, Diaz S et al (2010) Biodiversity targets after 2010. *Curr Opin Environ Sustain* 2(1–2):3–8. doi:[10.1016/J.Cosust.2010.03.003](https://doi.org/10.1016/J.Cosust.2010.03.003)
- Mace GM, Reyers B, Alkemade R et al (2014) Approaches to defining a planetary boundary for biodiversity. *Global Environ Chang* 28:289–297. doi:[10.1016/J.Gloenvcha.2014.07.009](https://doi.org/10.1016/J.Gloenvcha.2014.07.009)
- MacLaurin J, Sterelny K (2008) What is biodiversity? The University of Chicago Press, Chicago
- Magnuson-Ford K, Mooers AO, RiouxPaquette S, Steel MA (2010) Comparing strategies to preserve evolutionary diversity. *J Theor Biol* 266:107–116
- Margules CR, Pressey RL (2000) Systematic conservation planning. *Nature* 405:243–253
- May RM (1990) Taxonomy as destiny. *Nature* 347:129–130
- McCarthy DP, Donald PF, Scharlemann JPW et al (2012) Financial costs of meeting global biodiversity conservation targets: current spending and unmet needs. *Science* 338(6109):946–949. doi:[10.1126/Science.1229803](https://doi.org/10.1126/Science.1229803)
- McGoogan K, Kivell T, Hutchison M et al (2007) Phylogenetic diversity and the conservation biogeography of African primates. *J Biogeogr* 34(11):1962–1974
- McKinney ML (1999) High rates of extinction and threat in poorly studied taxa. *Conserv Biol* 13:1273–1281
- Millennium Ecosystem Assessment (2005) Ecosystems and human well-being: biodiversity synthesis. World Resources Institute, Washington, DC
- Moilanen A, Kujala H, Leathwick J (2009) The zonation framework and software for conservation prioritization. In: Moilanen A, Wilson KH, Possingham HP (eds) Spatial conservation prioritization: quantitative methods and computational tools. Oxford University Press, Oxford, pp 196–210
- Mooers AO, Heard SB, Chrostowski E (2005) Evolutionary heritage as a metric for conservation. In: Purvis A, Gittleman JL, Brooks T (eds) Phylogeny and conservation, vol 8, Conservation biology. Cambridge University Press, London, pp 120–138
- Myers N (2003) Biodiversity hotspots revisited. *Bioscience* 53(10):916–917
- Myers N, Mittermeier RA, Mittermeier CG et al (2000) Biodiversity hotspots for conservation priorities. *Nature* 403:853–858
- Nee S, May RM (1997) Extinction and the loss of evolutionary history. *Science* 278:692–694

- Nipperess DA, Matsen FA (2013) The mean and variance of phylogenetic diversity under rarefaction. *Methods Ecol Evol* 4(6):566–572. doi:[10.1111/2041-210x.12042](https://doi.org/10.1111/2041-210x.12042)
- Nunes LA, Turvey ST, Rosindell J (2015) The price of conserving avian phylogenetic diversity: a global prioritization approach. *Philos Trans R Soc B* 370(1662). doi:[10.1098/Rstb.2014.0004](https://doi.org/10.1098/Rstb.2014.0004)
- O'Hara RJ (1992) Telling the tree: narrative representation and the study of evolutionary history. *Biol Phil* 7:135–160
- Pavoine S, Ollier S, Dufour AB (2005) Is the originality of a species measurable? *Ecol Lett* 8:579–586
- Pereira HM, Leadley PW, Proenca V et al (2010) Scenarios for global biodiversity in the 21st century. *Science* 330(6010):1496–1501. doi:[10.1126/Science.1196624](https://doi.org/10.1126/Science.1196624)
- Pereira HM, Ferrier S, Walters M et al (2013) Essential biodiversity variables. *Science* 339(6117):277–278. doi:[10.1126/Science.1229931](https://doi.org/10.1126/Science.1229931)
- Pio DV, Broennimann O, Barraclough TG, Reeves G, Rebelo AG, Thuiller W, Guisan A, Salamin N (2011) Spatial predictions of phylogenetic diversity in conservation decision making. *Conserv Biol* 25(6):1229–1239. doi:[10.1111/J.1523-1739.2011.01773.X](https://doi.org/10.1111/J.1523-1739.2011.01773.X)
- Polasky S, Csuti B, Vossler CA, Meyers SM (2002) A comparison of taxonomic distinctness versus richness as criteria for setting conservation priorities for North American birds. *Biol Conserv* 97:99–105
- Pollock LJ, Rosauer DF, Thornhill AH et al (2015) Phylogenetic diversity meets conservation policy: small areas are key to preserving eucalypt lineages. *Philos Trans R Soc B* 370(1662). Unsp 20140007. doi:[10.1098/Rstb.2014.0007](https://doi.org/10.1098/Rstb.2014.0007)
- Posadas P, Miranda Esquivel DR, Crisci JV (2001) Using phylogenetic diversity measures to set priorities in conservation: an example from Southern South America. *Conserv Biol* 15(5):1325–1334
- Possingham HP, Andelman S, Burgman MA et al (2002) Limits to the use of threatened species lists. *Trends Ecol Evol* 17:503–507
- Proença V, Pereira HM (2013) Comparing extinction rates: past, present, and future. *Encycl Biodivers* 2:167–176
- Purvis A (2008) Phylogenetic approaches to the study of extinction. *Ann Rev Ecol Evol* 39:301–319. doi:[10.1146/Annurev-Ecolsys-063008-102010](https://doi.org/10.1146/Annurev-Ecolsys-063008-102010)
- Purvis A, Gittleman JL, Brooks T (eds) (2005) *Phylogeny and conservation*, vol 39, *Conservation biology* 8. Cambridge University Press, Cambridge
- Redding DW, Mooers AO (2006) Incorporating evolutionary measures into conservation prioritization. *Conserv Biol* 20(6):1670–1678
- Redding DW, Mooers AO, Sekercioglu CH, Collen B (2015) Global evolutionary isolation measures can capture key local conservation species in Nearctic and Neotropical bird communities. *Philos Trans R Soc B* 370: 20140013
- Régner C, Fontaine B, Bouchet P (2009) Not knowing, not recording, not listing: numerous unnoticed Mollusk extinctions. *Conserv Biol* 23:1214–1221
- Rodrigues ASL, Gaston KJ (2002) Maximising phylogenetic diversity in the selection of networks of conservation areas. *Biol Conserv* 105:103–111
- Rodrigues AS, Brooks TM, Gaston KJ (2005) Integrating phylogenetic diversity in the selection of priority areas for conservation: does it make a difference? In: Purvis A, Gittleman JL, Brooks T (eds) *Phylogeny and conservation*. Cambridge University Press, Cambridge, pp 101–119
- Rodrigues ASL, Grenyer R, Baillie JEM et al (2011) Complete, accurate, mammalian phylogenies aid conservation planning, but not much. *Philos Trans R Soc B* 366(1578):2652–2660. doi:[10.1098/Rstb.2011.0104](https://doi.org/10.1098/Rstb.2011.0104)
- Rosauer D, Laffan SW, Crisp MD, Donnellan SC, Cook LG (2009) Phylogenetic endemism: a new approach for identifying geographical concentrations of evolutionary history. *Mol Ecol* 18(19):4061–4072. doi:[10.1111/J.1365-294x.2009.04311.X](https://doi.org/10.1111/J.1365-294x.2009.04311.X)
- Safi K, Cianciaruso MV, Loyola RD et al (2011) Understanding global patterns of mammalian functional and phylogenetic diversity. *Philos Trans R Soc B* 366(1577):2536–2544. doi:[10.1098/Rstb.2011.0024](https://doi.org/10.1098/Rstb.2011.0024)



- Sarkar SK (2005) Biodiversity and environmental philosophy: an introduction. Cambridge studies in philosophy and biology. Cambridge University Press, Cambridge
- Schweiger O, Klotz S, Durka W, Kühn I (2008) A comparative test of phylogenetic diversity indices. *Oecologia* 257:485–495
- Steel M, Mimoto A, Mooers AO (2007) Hedging our bets: the expected contribution of species to future phylogenetic diversity. *Evol Bioinforma* 3:237–244
- Thuiller W, Lavergne S, Roquet C, Boulangeat I, Lafourcade B, Araujo MB (2011) Consequences of climate change on the tree of life in Europe. *Nature* 470(7335):531–534. doi:[10.1038/Nature09705](https://doi.org/10.1038/Nature09705)
- Thuiller W, Maiorano L, Mazel F et al (2015) Conserving the functional and phylogenetic trees of life of European tetrapods. *Philos Trans R Soc B* 370(1662). ArtN 20140005. doi:[10.1098/Rstb.2014.0005](https://doi.org/10.1098/Rstb.2014.0005)
- Vane-Wright RI, Humphries CJ, Williams PH (1991) What to protect? Systematics and the agony of choice. *Biol Conserv* 55:235–254
- Walker PA, Faith DP (1994) Procedures for conservation evaluation based on phylogenetic diversity. *Biodivers Lett* 2(5):132–139
- Weitzman ML (1998) The Noah's Ark problem. *Econometrica* 66(6):1279–1298. doi:[10.2307/2999617](https://doi.org/10.2307/2999617)
- Wiley EO (1981) *Phylogenetics. The theory and practice of phylogenetic systematics*. Wiley-Liss, New York
- Yessoufou K, Daru BH, Davies J (2012) Phylogenetic patterns of extinction risk in the eastern Arc ecosystems, an African biodiversity hotspot. *PLoS ONE* 7(10):e47082
- Zupan L, Cabeza M, Maiorano L et al (2014) Spatial mismatch of phylogenetic diversity across three vertebrate groups and protected areas in Europe. *Biodivers Conserv* 20(6):674–685. doi:[10.1111/Ddi.12186](https://doi.org/10.1111/Ddi.12186)

# **Part I**

## **Questions**

# The Value of Phylogenetic Diversity

Christopher Lean and James Maclaurin

**Abstract** This chapter explores the idea that phylogenetic diversity plays a unique role in underpinning conservation endeavour. The conservation of biodiversity is suffering from a rapid, unguided proliferation of metrics. Confusion is caused by the wide variety of contexts in which we make use of the idea of biodiversity. Characterisations of biodiversity range from all-variety-at-all-levels down to variety with respect to single variables relevant to very specific conservation contexts. Accepting biodiversity as the sum of a large number of individual measures results in an empirically intractable framework. However, large-scale decisions cannot be based on biodiversity variables inferred from local conservation imperatives because the variables relevant to the many systems being compared would be incommensurate with one another. We therefore need some general conception of biodiversity that would make tractable such large-scale environmental decision-making. We categorise the large array of strategies for the measurement of biodiversity into four broad groups for consideration as general measures of biodiversity. We compare common moral justifications for the conservation of biodiversity and conclude that some form of instrumental value is the most plausible justification for biodiversity conservation. Although this is often interpreted as a reliance on option value, we opt for a broadly consequentialist characterisation of biodiversity conservation. We conclude that the best justified general measure of biodiversity will be some form of phylogenetic diversity.

**Keywords** Biodiversity • Measurement • Surrogacy • Consequentialism • Justification

---

C. Lean (✉)

Philosophy Programme, Australian National University, Canberra, Australia  
e-mail: [christopher.hunter.lean@gmail.com](mailto:christopher.hunter.lean@gmail.com)

J. Maclaurin

Department of Philosophy, University of Otago, Dunedin, New Zealand  
e-mail: [James.Maclaurin@otago.ac.nz](mailto:James.Maclaurin@otago.ac.nz)

## Introduction

It is not surprising that there is a bewildering array of tools available to those who would measure biodiversity. There are of course countless respects in which organisms and ecosystems vary. More importantly, there are many types of scientific projects which exploit different aspects of biodiversity. In *What is biodiversity?* (2008) Maclaurin and Sterelny argue that, although it began as an idea primarily of interest to conservation biologists, there are now many areas of the life sciences in which biodiversity plays an ontological, explanatory or predictive role.

Moreover, within conservation biology the role of biodiversity has become complex. When biodiversity was first envisaged in the 1980s it was intended as a new organising principle for conservation. In many respects it was to be a replacement for the old idea that conservation was fundamentally about preserving species and the even older idea that it is essentially about preserving wilderness (Nash 1990). But alongside this idea of biodiversity as an overarching goal of conservation, our new understanding of the effects of diversity on ecology, genetics, and morphology allows us to harness particular aspects of biodiversity to achieve specific conservation goals. So now biodiversity takes its place both as a goal for policymakers and as a tool for conservation biologists. In both contexts, biodiversity is difficult to measure. For this reason, much of the growth in biodiversity metrics has been in the development of new and more effective biodiversity surrogates.

In this complex theoretical and methodological landscape, is phylogenetic diversity just one more tool to be used as and when appropriate? In this chapter, we focus on conservation biology and argue that phylogenetic diversity plays a unique role in underpinning conservation endeavour.

In the first section we argue that the conservation of biodiversity is suffering from a rapid, unguided proliferation of metrics. These various measures will be categorized by what they aim to pick out and preserve. We then scrutinise the justification for various types of measures as fundamental principles underpinning large-scale conservation (we explain why ‘large-scale’ in the next section) and argue that this role is best performed by phylogenetic diversity.

## A Maze of Measures

Our current understanding of biodiversity is a mess. It is a fortunate, productive, and useful mess but a mess none the less. This can be traced to the lack of a guiding set of standards from which to assess the value of proposed biodiversity measures. Although measures are tested, the testing has often been piecemeal across conservation biology and related disciplines leading to conflicts over whether a metric has been proved. An example is the debate between Ross Crozier et al. (2005) and Dan Faith and Andrew Baker (2006) over assessing conservation schemes which use phylogenetic diversity for data sets that include systematized taxa without

phylogenies. While Crozier et al. claim that this study is a “proof of concept”, what they take to be an examination of phylogenetic diversity’s applicability to conservation projects in the field, Faith and Baker claim that such examinations were already conducted a decade ago! The lack of a guiding set of standards has resulted in difficulty compiling and comparing measurement procedures in an environment in which new measures are proliferating. It is noted that “in the last decade more than two measures of Phylogenetic Diversity or Functional Diversity were proposed, each year!” (Cianciaruso 2011). This has resulted in measurement options for biodiversity increasing without a clear way of choosing between them. This proliferation of varied, uncategorized measures is referred to by Faith and Baker as the “curse of biodiversity informatics” or “bio-miss-informatics” (Faith and Baker 2006).

The proposed measures of biodiversity are of course, not limited to phylogenetic diversity. There are measures aimed at describing biodiversity using many different accounts of functions, abundance measures, ecosystem services, and hybrids of all of the above. The description of these measures is inconsistent throughout biology because; “The vocabulary used to classify indices is continuously evolving and differs between evolutionary and ecological studies, leading to potential confusion when a term is employed without a clear definition or reference” (Pavoine and Bonsall 2011). Biodiversity particularly suffers from ambiguity regarding biological features scientists and policymakers are referring to when they say an ecosystem has high biodiversity.

Individuals and groups have tried to build consensus around which features are worthy of measurement. One recent attempt to collect an index of measures that are fundamental to biodiversity notes that; “a key obstacle is the lack of consensus about what to monitor” (Pereira et al. 2013, p. 277). The authors propose a set of “Essential Variables of Biodiversity”. These include:

- Genetic composition e.g. allelic diversity
- Species populations e.g. Abundances and distributions
- Species traits e.g. phenology
- Community composition e.g. taxonomic diversity
- Ecosystem structure e.g. habitat structure
- Ecosystem function e.g. nutrient retention

Each of these “variables” can be measured using multiple (sub-) variables. For example ecosystem function in their account includes nutrient retention in a community. This would include the cycling of Nitrogen, Carbon, and Phosphorous through a community, amongst other important nutrients. Biological features such as species traits not only need to be individuated but there are also numerous different mathematical measures for that trait description to decide between. All these variables, their sub-variables, and the different measurement procedures for the sub-variables are understood as actual measures of biodiversity (although for any real ecosystem the majority of these variables will be unanalysed). To what then do we refer when we talk of biodiversity as a conservation goal? According to these authors, we refer to the sum of all these ‘essential’ aspects of biological diversity.

This permissive and conciliatory view of biodiversity, while at first seeming attractive, is problematic as a guide to conservation. Accepting biodiversity as the sum of a large number of individual measures results in an empirically intractable framework. Large-scale conservation requires prioritisation of effort and resources across disparate ecosystems. The many available biodiversity measures make such decisions difficult. In all ecosystems there will be incompletely analysed variables. So either policymakers and conservationists accept that many assessments of biodiversity are incommensurate with one another or they must subscribe to schemes for weighting the various measures. In practice, the relative weighting of the many variables will often be treated as equal but there is an open question as to whether we *should* treat each variable as equal. Should ecosystem biomass be treated as equally important as plant trait disparity? If not then we will have to agree on a seemingly arbitrary rubric of relative weights for the various features being measured. In short, the retention of such a large swath of essential measures creates problems for the practice of conservation.

We accept that the many measures representing the diversity of biological systems can be relevant to particular contexts in conservation and their accuracy and utility can be assessed through experimentation or modelling (Pereira et al. critically assess measures through their “scalability, temporal sensitivity, feasibility, and relevance”, p. 277). But as a whole, the use of biodiversity as a foundational tool in conservation biology suffers from a glut of information that is hard to integrate in a useable way. Those who agree with Michael Soulé’s (1985, p. 727) well-worn description of conservation biology as a crisis discipline, are likely to think such confusion can only get in the way of efficient decision-making. Biodiversity should be a useful concept across disciplines and sites.

Local conservation imperatives often point to particular biodiversity variables to which we should pay attention, e.g. focus on genetic diversity is crucial in trying to bring a single species back from the brink of extinction. However, not all conservation is local. Governments and NGOs must prioritise conservation strategies applied to different ecosystems and applied at different scales, e.g. governments must weigh the conservation value of: conserving endangered species, developing national parks, regulating fisheries, and decreasing carbon emissions.<sup>1</sup> Such large-scale decisions cannot be based on biodiversity variables inferred from local conservation imperatives because the variables relevant to the many systems being compared would be incommensurate with one another. For the reasons noted above, it is impractical to interpret biodiversity in such large-scale contexts as the sum of all the biodiversity variables of all the systems being compared. We therefore need some general or fundamental conception of biodiversity that would make tractable such large-scale environmental decision-making. In what follows, we shall refer to this as a *general measure of biodiversity*.

---

<sup>1</sup>Of course some of these are not purely conservation decisions, but all rest to some important extent upon judgements about the value of natural systems.

## ***One of Many Biodiversities***

In thinking about large-scale differences in biodiversity, we often employ a concept of biodiversity which is very broad. Sarkar et al. claim biodiversity is “diversity at every level of taxonomic, structural, and functional organization of life” (Sarkar et al. 2006). The Convention on Biological Diversity (CBD) proposes that biodiversity is “diversity within species, between species, and of ecosystems” (CBD 1992). According to such definitions, any mathematical measure that categorizes biological difference and preferentially organizes that difference is a measure of biodiversity (including many unimportant and unused metrics e.g. diversity of spottiness as quantified by the number of non-contiguous circular patterns averaged over the members of a species).

This broad characterisation of biodiversity has permitted a range of targets of measurement such as species richness, species diversity, ecosystem function, species function, population relations, ecosystem diversity, biomass, genetic diversity, phylogenetic diversity, and many more. In what follows we collect these measures into broad categories and assess each as the basis for a general measure of biodiversity. We begin by tackling a couple of red herrings.

## **Measures We Rule Out**

A general measure of biodiversity must be capable of guiding large-scale and long-term conservation effort. We think this rules out two types of biodiversity measures: biodiversity surrogates and measures based on ecosystem services. Both are, of course, important tools in conservation, but for the reasons set out below, they cannot underpin a general measure of biodiversity.

## ***Surrogates of Biodiversity***

As noted above, most of the growth in biodiversity metrics has been in the development of new surrogates for biodiversity, i.e. measures of features whose presence is correlated with high biodiversity. If biodiversity measurement is to succeed as a large-scale goal of conservation, then we must be able to assess the success of biodiversity surrogates and we can only do that if we understand what it is that these metrics are surrogates for. Sarkar et al. (2006) argue that “general biodiversity is too diffuse a term to be precisely defined”. The best we can do is to agree to “some convention or consensus about what constitutes the relevant features of biodiversity in a given context”. We think this ‘nothing but surrogates’ view of biodiversity measurement, in effect, risks giving up on the idea of biodiversity as an overarching goal for conservation. Crucially this convention-based view on how we should

characterise biodiversity appears not to rest on underlying principles for the assessment of the conventions underpinning such a consensus on biodiversity measurement.

On our view, a general measure of biodiversity must be definable (or at least capable of clear characterisation) and it must be a feature of biological systems that we can practically assess across clades and ecosystems. This is essential if such a measure is to assist us in forging large-scale conservation policy. Moreover, it must not itself be a surrogate for some further more basic characteristic of living systems that can also be measured across clades and ecosystems.

### *Anthropogenic Variables*

The idea of ecosystem services as a foundation for a general measure of biodiversity is fraught with difficulty. This is partly because the whole idea of ecosystem services is at best very open ended. The Millennium Ecosystem Assessment report (2005) defines ecosystem services as “benefits people obtain from ecosystems”. Despite gallant attempts to assess the global value of ecosystem services in dollar terms (e.g. Costanza et al. 1997), many of the psychological and social benefits are difficult to measure even at small scales and, as a group, the benefits people obtain from ecosystems seem incommensurate with one another (Boyd and Banzhaf 2007). Moreover, while ecosystem services are usually interpreted as inventories of current benefits to humanity, conservation is inherently forward-looking and it is even more difficult to accurately assess the benefits that species and ecosystems will provide to our descendants. Indeed, even if we could agree on a reliable set of measures and agree on a way to aggregate them, many environmental ethicists and many members of the public would balk at the idea that only human interests need be taken into account in conservation decision-making (see for example Stone 1972). So although ecosystem services are an important driver of conservation effort, we think this tool is too limited to form a plausible basis for a general measure of biodiversity.

The idea of biodiversity should capture the diverse features of life not the diverse interests of people. While we grant to Reyes et al. (2012) that there is ‘functional overlap’ between these two features of biological systems we agree with Faith (2012) that ecosystem services and biodiversity are distinct. It is in the interests of humanity to preserve biodiversity, but this fact does not warrant defining biodiversity in terms of current human needs and interests. Moreover, there is practical utility in keeping these ideas separate. Differentiating between ecosystem services and biodiversity has allowed research into whether these features co-vary and what biodiversity targets yield ecosystem services (Benayas et al. 2009; Mace et al. 2012; Worm et al. 2006). In certain cases we may want to prioritize the maintenance or reinstatement of ecosystem services. Differentiating the services from the diversity serves to distinguish such conservation that focuses squarely on the economic and social needs of human populations.



## The Main Candidates

As noted in the previous section, current broad characterisations of biodiversity permit a range of targets of measurement including species richness, species diversity, ecosystem function, species function, population relations, ecosystem diversity, biomass, genetic diversity, phylogenetic diversity, and many more. In this section, for the sake of manageability, we categorise that large array of strategies into four broad groups for consideration as general measures of biodiversity.

### *Species Diversity and Species Richness*

Species diversity is an intuitively simple concept that has yielded numerous mathematical explications combining species richness, the number of species in an area, species evenness, and the relative abundance of species (see Maurer and MacGill 2011). Species richness is extremely common as a measure of biodiversity, partly due to its relative ease of discovery. It is a key variable from which many diversity metrics are constructed influencing the output of species diversity, functional, genetic, and phylogenetic measures. It is, in many contexts, a good indicator of biodiversity. Holmes Rolston goes as far to claim that species richness *is* biodiversity as “(s)pecies are a more evident, mid-range, natural kind” as opposed to other proposed units of biodiversity like genetic diversity or ecosystem diversity (p. 402, Rolston 2001).

Species richness is usually supplemented with other information as just counting the species present gives limited insight into the dynamics of an assemblage. Often species richness is combined with species evenness to create many of the common species diversity measures.<sup>2</sup> This is based on the idea that, given a species richness in an area, species diversity increases when the populations have more even abundances and vice versa. Information theory has provided the most common indices of species diversity, the Shannon evenness and the Simpson evenness indices. Other measures include: Hill’s Indices, Hurlbert’s “Interspecific encounter Index”, Rao’s “Quadratic Entropy” Index, and Fager’s Indices (See Justus 2011; Maurer and Macgill 2011).

While there is a range of ways that species diversity is calculated there is one feature common to these measures. Measures of species richness and diversity are blind to each individual species’ identity. No species is treated as being more valuable than any other. This assumption is directly rejected by measures that prioritize species by any of their individual features including morphology, genetics, or phylogeny.

---

<sup>2</sup>For a sceptical take on the success of such measures see Justus (2011).

## ***Function and Morphology***

Functional diversity, as it is commonly used, is a subset of trait diversity. Functional traits are commonly morphological traits differentiated by the effects the trait has on an ecosystem (Petchey and Gaston 2006). Some ecologists have rejected the need to associate ‘functional’ traits to ecosystem effects and treat function diversity as a synonym of morphology. Evan Weiher (2011) in his summary of functional diversity measures states, “Some have suggested the term ‘functional diversity’ be restricted to measures of trait diversity that affect the functions of ecosystems (Tilman et al. 2001; Petchey and Gaston 2006). We should be wary of unnecessarily restrictive definitions for terms that are conceptual, general, or useful” (pg. 175). He further notes that general morphological trait space can be differentiated without reference to a schematic for differentiating traits. The dizzying range of mathematical measures for dividing morphological space include: distance measures, dendrogram-based measures, variance-based measures including abundances, trait evenness, convex hull mathematics to measure trait volume, and graph theory (See Weiher 2011).

## ***Genetic Diversity***

Genetic diversity is considered by many to be the lowest level of a nested hierarchy of diversity comprising of genetic diversity, species diversity, and community diversity (Culver et al. 2011). Culver et al. suggest that genetic variation is “the essence of all biodiversity” (p. 208). Genetic barcoding of populations has become increasingly common due to the efficiency of new sampling techniques and the increase in computational power. Clearly, there will in the future be more genetic information available to researchers that will aid, not just our understanding of genetic difference, but also our assessments of other forms of diversity such as species diversity and phylogenetic diversity. Despite its clear practical importance, it is implausible that genetic diversity should underpin a general measure of biodiversity. This is partly because genes vary greatly in their effects so that the amount of raw genetic difference between two populations tells you relatively little about the extent to which they differ functionally and ecologically. It is also partly due to the undoubted importance of non-genetic factors in both ecology and evolution (Laland et al. 1999; West-Eberhard 2003; Jablonka and Lamb 2005).

## ***Phylogenetics and Phylogenetic Diversity***

Phylogenetic inference recreates the branching structure of evolutionary relationships between species via cladistic analysis from molecular and morphological data in the form of discrete character states or distance matrices of pairwise

dissimilarities (Vandamme 2009). The computational models used differ both in methodology and epistemological grounding; prominent methods include Maximum Parsimony, Maximum Likelihood, and Bayesian Methodologies. Phylogenetic distance measures aim to quantify the relatedness of groups of species. As the phylogenetic tree represents the evolutionary relations between species it can also be used to calculate how distinct these species are relative to the tree in which they are nested. Methods differ in the way they characterize distance and uniqueness. Some do it in terms of speciation events and others in terms of change in genomes between species. Following Velland et al. (2011), we distinguish two types of fundamentally different measures of phylogenetic diversity (p. 196):

**Node-based trees** represent only topology. They are based only on information about speciation events and so we can infer from them only facts about relatedness. Such measures include: Taxonomic Distinctness (Vane-Wright et al. 1991) and Species Originality (Nixon and Wheeler 1992).

**Distance-based trees** include topological information as well as branch length. Branch length either represents the accumulation of evolutionary change or alternatively the passage of time. Such measures include: PD (Faith 1992, 1994); Originality of Species within a Set (Pavoine et al. 2005); Pendant Edge<sup>3</sup> (Altschul and Lipman 1990) and Species Evolutionary History (Redding and Mooers 2006).

Both groups of methods represent speciation and its creation of distinct evolutionary trajectories and both provide, with varying degrees of success, a means to prioritize the conservation of phylogeny and therefore of species that are particularly distinct in their features and history.

## *The Roles of Phylogenetic Diversity*

Although the role of phylogenetic diversity in conservation biology is open-ended, extant uses can be categorised into three distinct groups.

### (i) *Phylogenetic Diversity as a tool for prediction and explanation*

Conservation is only possible when we have a good understanding of the dynamics of communities and ecosystems. Although we often think of this in ecological terms, evolution is an important contributing factor. In such contexts the measurement of phylogenetic diversity can help us distinguish these component forces at work. For example, all else being equal, we expect species that are closely related to be both morphologically similar and similar in the functional roles that they play in the ecosystems in which they are found. So we can use phylogenetic diversity to predict functional similarity. Such studies allow

---

<sup>3</sup>Note “Pendant Edge” is a recent name (e.g. Redding, and Mooers 2006; Velland et al. 2011) given to the idea introduced but not named in Altschul and Lipman’s original very brief discussion note.

us to detect cases that stand in need of special explanation. These are cases where functional diversity is either higher (over-dispersion) or lower (functional diversity deficit) than expected (see for example Webb et al. 2002). The appropriateness of particular metrics will depend upon the explanatory or predictive target, although we note that common metrics show strong correlation with one another in many circumstances (Vellend et al. 2011, p. 207).

(ii) *Phylogenetic diversity as a surrogate*

Phylogenetic diversity has been employed as a surrogate for a wide variety of valuable features of ecological communities and ecosystems. For example, Srivastava et al. (2012) argue that phylogeny largely determines interactions among species, and so could help predict the cascade of extinctions through ecological networks and hence the way in which those extinctions impact ecosystem function. So, on this account, phylogenetic diversity is at least a surrogate for ecosystem function.

Forest et al. (2007) find a stronger correlation between phylogenetic diversity and feature diversity than between species diversity and feature diversity. So they recommend that we employ phylogenetic diversity, rather than species diversity, as a surrogate for feature diversity. Faith et al. (2010) argue that we should recognise phylogenetic diversity as a surrogate for features of value to human well-being:

We argue that an evolutionary perspective is essential for developing a better understanding of the links between biodiversity and human well-being. We outline the services provided by evolutionary processes, and propose a new term, ‘evosystem services’, to refer to these many connections to humans. (Faith D.P. et al. 2010, p. 66)

(iii) *Phylogenetic diversity as a conservation goal*

The third context in which one might employ phylogenetic diversity is as a goal of conservation. There are certainly examples of phylogenetically orientated conservation. The *Edge of Existence Programme* ([www.edgeofexistence.org](http://www.edgeofexistence.org)), run by the Zoological Society of London, focuses explicitly on the conservation of species that are endangered and phylogenetically distinct. There are many other conservation programmes that take phylogenetic diversity into account (e.g. WWF’s Global 200). That said, phylogenetic diversity is not as widely used in conservation as it might be (Winter et al. 2012, p. 1). This is partly for methodological reasons:

Phylogenetic diversity has long been incorporated in planning tools, but it has not yet had much impact on conservation planning. Applications face limitations of available data on phylogenetic pattern. (Sarkar et al. 2006)

It is also partly due to scepticism about the correlations claimed above:

In our opinion, the justification for preserving phylogenetic diversity as a proxy for functional diversity or evolutionary potential has so far largely failed. Our current knowledge of the benefits to the (future) functioning of ecosystems and securing evolutionary potential remains equivocal. (Winter et al. 2012, p. 4)

Clearly there is limited employment of phylogenetic diversity as goal for large-scale conservation decision-making. There is also some scepticism about

our empirical and philosophical justification for such uses. In the final section of this chapter it is this question about justification to which we turn.

## **Moral Justifications for a General Measure of Biodiversity?**

We have argued that large-scale conservation decision-making would benefit from agreement on a general measure of biodiversity, one that is not tied to particular projects or contexts. We have set out a group of broad categories of measurement strategies with the aim of determining whether one of these might furnish an appropriate general measure. In this section, we set out a similarly broad brush taxonomy of philosophical justifications for the conservation of biodiversity with the aim of determining whether any of those available might provide a justification for conservation based on a general measure of biodiversity and hence might provide us with a basis for inference about the nature of such a general measure. We will argue that the best justification is one that respects the plurality of human and non-human interests in biodiversity as well as uncertainty about how best to secure those interests and about future changes both in the environment and in human affairs.

Philosophical justifications for the conservation of biodiversity come in many forms but all such arguments fall into one of four categories.

### ***Intrinsic Value***

The idea that biodiversity has intrinsic value is enshrined in the Convention on Biodiversity. It is also a central tenet of deep ecology (Naess 1986). Despite its common currency, intrinsic value is capable of multiple interpretations which causes considerable confusion in moral reasoning (O'Neill 1992 p. 119). At least two interpretations are plausible in the current context.

One is the idea that biodiversity has intrinsic value in the sense that it has value over and above its instrumental value. This interpretation is further dependent on what we count as 'instrumental'. If we tie instrumental value to narrow economic purposes, then there seems to be considerable non-instrumental value in biodiversity. If we tie it to a broader set of psychological benefits (provided by recreation, aesthetic appreciation etc.) then the domain of non-instrumental value seems correspondingly smaller and more difficult to characterise.

A second interpretation is that biodiversity has intrinsic value in the sense that it is valuable independently of the valuations of valuers. It does after all seem that the biosphere would remain a locus of value even if some selective extinction event caused the demise of humanity or even the extirpation of all species capable of reasoning about value. But value in this sense seems almost impossible to quantify precisely because it cannot be tied to evaluative judgements made by economic

actors or by environmental stakeholders. The best we seem to be able to say is that some people, when asked, assent to the existence of such value.

Intrinsic value is controversial as a justification for the conservation of biodiversity for two reasons. Firstly, there is philosophical controversy about whether such forms of value exist (Norton 1984, p. 145). Secondly, as it is independent of human projects and human values, it is unclear how it should be measured and hence, how it should be conserved. There seems no way in principle of choosing between variety of ecosystems, variety of species, variety of form and function or variety in genetic make-up etc. as loci for biodiversity's intrinsic value. On the other hand, if intrinsic value is only a justification for the conservation of biodiversity in the very broad sense (set out at the end of section "[Measures we rule out](#)"), that will leave us no further along the path in the project of understanding or employing a practical general measure of biodiversity.

### *Human Emotional Responses to the Natural World*

It is also claimed that biodiversity is valuable because the psychological makeup of human beings causes them to feel an intimate connection with the natural world which might be expressed variously in emotions such as love of, or respect for, nature. The idea that such emotional responses are a result of our evolved psychology was promoted by Wilson (1984) and Kellert and Wilson (1993). We note that the so-called Biophilia Hypothesis has received limited support in the literature (Simaika and Samways 2010 p. 903), but let us assume for the moment that we do share a common innate love of nature.

There are two important problems with grounding conservation in common emotional responses. Firstly, such responses are not always reliable guides to rational action. There is after all some fundamental fact about human beings that also causes them to see cigarettes as valuable. We don't think that this implies that we should 'conserve' cigarettes, because we don't think that this common emotional response is adaptive. Human beings feel positively disposed toward all sorts of things that are not actually good for us. But if we must then judge the adaptiveness of our feelings toward biodiversity, it seems that conservation justified thereby would not be a consequence of our feelings towards biodiversity, but rather of the utility of biodiversity to human populations (to which we turn shortly). Secondly, people clearly differ a great deal in the extent to which they feel positive emotions toward biodiversity (Einarsson 1993). If a general measure of biodiversity is to be inferred from emotional responses to biodiversity, then it seems that we will either have to discount the responses of outliers or average across a relatively large range of responses.

Finally, this style of justification for conservation suffers from the same problems as conservation based on intrinsic value. Even if it were true that almost everyone attached the same equally strong positive emotion to the conservation of the biosphere, it is hard to see how we could turn universal love of nature into a practically applicable general measure of biodiversity. For these reasons, we think it

implausible that common emotional responses to nature will justify general measure of biodiversity.

## ***Instrumental Value***

The benefits conferred by biodiversity on humanity (and indeed on other species) are themselves diverse (aesthetic, ecological, economic, epistemic etc.). Moreover, as Elliott Sober (1986) so eloquently points out, species differ a great deal in their apparent instrumental value. The great majority of species have small geographic ranges, do not perform unique ecological functions within their ecosystems and are not currently of important economic or psychological value to human populations. So Sober asks whether these facts justify the ‘rational attrition’ of species whose instrumental value is very small or unknown. This question about whether we should conserve ‘unremarkable species’ is closely related to the question of whether we should employ a general measure of biodiversity which would see us conserve species and ecosystems over and above those currently known to be of important instrumental value.

The strongest reason for conservation based on a general measure of biodiversity is that preferences or circumstances are likely to change so as to make valuable some proportion of the species in question. It is true that we have at times been overenthusiastic in our predictions about the possible future value of biodiversity such as the claims about the future value of bioprospecting in the Convention on Biodiversity (for more detail, see Maclaurin and Sterelny 2008, pp. 164–7). It is also true that a great deal of economic value resides in ecosystems that have low diversity, viz farms. That said, there has been huge growth in our appreciation for, and enjoyment of, natural variety through ecotourism, national parks, eco-sanctuaries etc. As noted in section “[Measures we rule out](#)”, there is also evidence that biodiversity is correlated with a wide range of ecosystem services. Furthermore, we should be careful not to base our predictions about future value on current categories. Just as ecotourism and bioprospecting are relatively recent ideas, we may in future discover new types of endeavour which place the value of extant species in a new light. In short, there is a *prima facie* reason for conservation based on a general measure of biodiversity, namely that we hedge our bets against an uncertain future. This idea was originally proposed by McNeely et al. (1990) as an instance of option value,<sup>4</sup> but the use of option value in this context has been controversial. Option value is an idea imported from economics. It is essentially a willingness-to-pay measure—the additional amount a person would pay for some amenity over and above its current value in consumption to maintain the option of having that amenity available for the future (van Kooten and Bulte 2000, p. 295). Although one of us has previously

---

<sup>4</sup>This idea has been championed particularly by Dan Faith. For excellent discussions of the option value represented by biodiversity see Faith (1992, 1994, 2013).

expressed enthusiasm for the option value idea (Maclaurin and Sterelny 2008, section 8.4) we now think that the answer lies elsewhere.

The crucial problem with option value is that it ties the value of biodiversity to judgements about value made by ordinary people (consumers in the economist's terms). Clearly actual assessments of such option value will be difficult (Norton 1988). Even if we could assess such judgements, human beings are not good at reasoning about risk and they have limited biological knowledge. So it might be that people's actual assessments about the option value in natural systems would be very poor guides to the likely effects of conservation on future human communities or on future ecosystems. If we hedge our bets to maximise future outcomes then we should do so based on our best information about the probability of such outcomes rather than on the estimates that consumers might make about such outcomes.

In light of these issues, the value of biodiversity is better analysed as an instance of consequentialism, broadly applied. We should conserve biodiversity, not because people want to, but because doing so will on average lead to better outcomes for people and human communities of perhaps more broadly for moral patients (organisms capable of experiencing suffering).<sup>5</sup>

However, even the consequentialist interpretation faces an important objection developed at length in chapter 6 of Maier (2012). It might be objected that our uncertainty about future states of the biosphere and future goals and preferences of people implies that conservation based on a general measure of biodiversity is as likely to produce net harm as it is net benefit (after all, the species we are conserving include many whose effects on human populations are currently unknown).

There are of course instances in which diversity works against us, as when we are threatened by a diversity of pathogens. That said, ours is an extremely successful species with an extremely broad niche. We have become adept at harnessing a great variety of features of the natural world to an astounding variety of ends. The number of species that pose a serious threat to humanity is a vanishingly small proportion of the total species count. Moreover, a great number of weeds and pests are not harmful in their native habitat, but only become harmful when that habitat is radically disturbed or when they are introduced by humans into other ecosystems (Baker 1974).

We therefore think it implausible that conserving unremarkable species will on average produce more harm than benefit. Put another way—were possible, at the press of a button, to destroy all those species and biological communities not known to be of special value to humanity, we think it would be irrational to do so. Humanity (and perhaps other sufficiently sentient species) would almost certainly be worse off. So where we cannot assess the likely payoff for conserving an individual unremarkable species, it is nonetheless rational to assume that that payoff will be positive. This does not of course tell us anything about how large such a payoff will be and we acknowledge that there is an interesting and difficult question about weighing the benefits of such conservation against the opportunity cost of forgoing alter-

---

<sup>5</sup>Although not explicitly consequentialist and still somewhat confusingly called option value, the approach taken by Faith (2013, p. 72) is similar to the current proposal.



native projects (e.g. if we used conservation funding to fight diseases or conservation land to grow more food for burgeoning populations in poor countries). However, we note that this problem of assessing opportunity costs is a global one, affecting all aspects of public policy and hence too large a topic to treat here. Our purpose is to determine how we should in general rank and assess biological systems as candidates for conservation. We leave it to others to determine how what proportion of total human effort ought to be spent on conservation.

## Phylogenetic Diversity as a General Measure of Biodiversity

We have argued that the best general justification for the conservation of biodiversity comes from its instrumental value. We also note that there are many types of such value and that the consequences of conservation focused on instrumental value in general are inherently uncertain. The nature and location of aesthetic, recreational, and other cultural values will inevitably be subject to disagreement. Moreover, we are not in possession of the full facts about the ways in which existing species and ecosystems can benefit (or harm) us and we know even less about the effects that conserved species and ecosystems will have on us and our descendants in the future. Can we harness this uncertainty as a means of developing a general measure of biodiversity?

We have argued that, leaving aside species whose value is currently well understood e.g. charismatic megafauna, economically important crops etc., we are warranted in spending some amount of time and effort in the large-scale conservation of biodiversity via some general measure. So we should conserve at least some of Sober's unremarkable species on the grounds that they might be valuable in some respect, but we cannot predict which respect that will be. This implies that a general measure of biodiversity should not aim at conserving particular features, but rather at conserving a maximal variety of features.

While it is sensible under some circumstances to measure variety of features or of functions, characterisation of overall biological diversity (of the sort attempted by Numerical Taxonomy) fails on philosophical grounds. It is not possible to capture differences in morphology<sup>6</sup> across the whole range of biological form because the idea of the occupation of morphospace makes sense only where we can anchor the dimensions of some particular morphospace to actual biological characteristics of closely related species (Maclaurin and Sterelny 2008, p. 15). The idea of a global morphospace is logically untenable because, as Goodman (1972, p. 437) argues, similarity and difference only make sense if we have some antecedent means of

---

<sup>6</sup>Note that in treating this problem is essentially about morphology, we are running form and function together. This is because we think that, were we to measure all biological form and all biological function, the two groups of characteristics would intersect at the level of physiological traits. So any attempt to develop an overall measure of functional diversity will face the same problems that must be overcome in the development of an overall measure of morphological diversity.

specifying the properties (or in the case of a morphospace, the dimensions) to be analysed. In taxonomy this almost always results in a focus on homologies. So in most cases the measurement of actual morphological diversity is best achieved by anchoring our analysis to actual differences in groups of related species, because only relatively closely related species differ in ways that make the analysis of morphospace tractable.<sup>7</sup>

So while broad difference in form and function is what the moral argument tells us to conserve, it cannot be measured directly in a way that would benefit large-scale conservation decision-making. Nonetheless, we can develop a general measure of biodiversity by exploiting the evolutionary processes that cause functional and morphological divergence within lineages. Both measures of species diversity and of phylogenetic diversity exploit evolution in just this way. If studies like those of Forest et al. (2007) are right, a general measure of biodiversity should be based on phylogenetic diversity, as that will best maximise feature diversity. We therefore conclude that phylogenetic diversity ought to play a fundamental role in conservation biology as the foundation of a general measure of biodiversity. That said, we noted in section “[A maze of measures](#)” that there are many measures of phylogenetic diversity. If conserving phylogeny is justified as a means of hedging our bets against uncertainty, this may help us to wrangle the current diversity in measures of phylogenetic diversity discussed earlier.

Variety in topological measures of phylogenetic diversity reflects the fact that phylogeny is complex. Species do not always bifurcate cleanly. Lineages reticulate and so on (Dagan and Martin 2006). Does this imply that, at large scales, phylogenetic diversity is undefined? We first note that such difficult cases are the exception rather than the rule at least across most of the phylogenetic tree. Secondly there are modifications of standard accounts of phylogenetic diversity designed to account for such phenomena as polytomies (see for example May 1990). Clearly over-dispersion studies (see the above discussion of Webb et al. 2002) are at least based on the assumption that it is possible to make large scale phylogenetic comparisons between very different systems. We cannot, in principle, construct a theoretical morphospace that contains humans and fungi and tardigrades, but we can compare their phylogeny. However, there is an important caveat. Large-scale phylogenetic diversity is tractable using topological measures of phylogenetic diversity and time-based distance measures, but it less obviously so for trait-based distance measures of phylogenetic diversity.

The more we incorporate form and function into a measure of phylogenetic diversity, the less plausible it is to think that you can compare phylogenetic diversity in this very rich sense between distantly related clades. Use of distance-based trees incorporating information about character evolution for such purposes requires the further assumptions (1) that there is a fact of the matter as to what we should count

---

<sup>7</sup>See for example the very wide variety of morphospaces discussed in McGee (1999, 2007). Indeed, it is notable that discussion of “convergent evolution in theoretical morphospace” (2007, pp. 90–2) actually focusses on a theoretical morphospace that models diversity in a single clade, namely the bryozoans (McKinney and Raup 1982).

as a character and (2) that all characters across all clades are of equal significance or contribute equally to biodiversity. To make this more concrete, we would have to assume that there is a fact of the matter as to how many characters contribute to the evolution of human cognition and that the biodiversity represented by the evolution of human cognition is of the same magnitude as the evolution of an equivalent number of characters in some other clade(s) for some other purpose(s).

## Conclusion

We have argued that uncertainty about the application of the current maze of measures of biodiversity results, in part, from uncertainty about our reasons for conserving biodiversity in general. This is problematic for decisions about large-scale conservation, particularly where such conservation includes species and ecosystems whose instrumental value is currently unknown. We have argued that, in such cases, use of a general measure of biodiversity is justified on the grounds that it will best hedge our bets against current and future uncertainty about the location of instrumental value and the needs and preferences of human populations. If we are right, a general measure of biodiversity should aim at the maximisation of feature diversity. The most effective and tractable such measure will be one based on phylogenetic diversity.

**Acknowledgements** Thanks to Daniel Faith, Philippe Grandcolas, James Justus, Arne Mooers, Roseli Pellens, Kim Sterelny, Ted Toadvine and the many others who provided helpful suggestions.

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Altschul S, Lipman D (1990) Equal animals. *Nature* 348:493–494
- Baker H (1974) The evolution of weeds. *Annu Rev Ecol Syst* 5:1–24
- Benayas JMR et al (2009) Enhancement of biodiversity and ecosystem services by ecological restoration: a meta-analysis. *Science* 325(5944):1121–1124
- Boyd J, Banzhaf S (2007) What are ecosystem services? The need for standardized environmental accounting units. *Ecol Econ* 63:616–626
- Cianniaruso MV (2011) Beyond taxonomical space: large-scale ecology meets functional and phylogenetic diversity. *Front Biogeogr* 3(3):87–90
- Costanza R et al (1997) The value of the world's ecosystem services and natural capital. *Nature* 387:253–260
- Crozier RH, Dunnett LJ, Agapow PM (2005) Phylogenetic biodiversity assessment based on systematic nomenclature. *Evol Bioinforma* 1:11–36

- Culver M, Fitak R, Herrmann H (2011) Genetic methods for biodiversity assessment. In: Magurran AE, McGill BJ (eds) *Biological diversity: frontiers in measurement and assessment*. Oxford University Press, Oxford, pp 175–193
- Dagan T, Martin W (2006) The tree of one percent. *Genome* 7:118
- Einarsson N (1993) All animals are equal but some are cetaceans: conservation and culture conflict. In: Milton K (ed) *Environmentalism: the view from anthropology*. Routledge, London, pp 71–82
- Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61:1–10
- Faith DP (1994) Phylogenetic pattern and the quantification of organismal biodiversity. *Philos Trans R Soc Lond B* 345:45–58
- Faith DP (2012) Biodiversity and ecosystem services: similar but different. *Bioscience* 62(9):785–785
- Faith DP (2013) Biodiversity and evolutionary history: useful extensions of the PD phylogenetic diversity assessment framework. *Ann N Y Acad Sci* 1289:69–89
- Faith DP, Baker AM (2006) Phylogenetic diversity (PD) and biodiversity conservation: some bioinformatics challenges. *Evol Bioinforma* 2:121–128
- Faith DP et al (2010) Ecosystem services: an evolutionary perspective on the links between biodiversity and human well-being. *Curr Opin Environ Sustain* 2:66–74
- Forest F et al (2007) Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature* 445:757–760
- Goodman N (1972) Seven strictures on similarity. In: Goodman N (ed) *Problems and projects*. Bobbs Merrill, Indianapolis, pp 437–447
- Jablonska E, Lamb MJ (2005) *Evolution in four dimensions: genetic, epigenetic, behavioural, and symbolic variation in the history of life*. MIT Press, Cambridge, MA
- Justus J (2011) A case study in concept determination: ecological diversity. In: Brown B, de Laplante K, Peacock K (eds) *Handbook of the philosophy of ecology*. Elsevier, New York, pp 147–168
- Kellert SR, Wilson EO (1993) *The biophilia hypothesis*. Island Press, Washington, DC
- Laland KN, Odling-Smee FJ, Feldman MW (1999) Evolutionary consequences of niche construction and their implications for ecology. *Proc Natl Acad Sci U S A* 96:10242–10247
- Mace GM, Norris K, Fitter AH (2012) Biodiversity and ecosystem services: a multilayered relationship. *Trends Ecol Evol* 27(1):19–26
- Maclaurin J, Sterelny K (2008) *What is biodiversity?* The University of Chicago Press, Chicago
- Maier DS (2012) *What's so good about biodiversity?* Springer, Dordrecht
- Maurer BA, McGill BJ (2011) Measurement of species diversity. In: Magurran AE, McGill BJ (eds) *Biological diversity: frontiers in measurement and assessment*. Oxford University Press, Oxford, pp 55–65
- May R (1990) Taxonomy as destiny. *Nature* 347:129–130
- McGhee G (1999) *Theoretical morphospace: the concept and its applications*. Columbia University Press, New York
- McGhee G (2007) *The geometry of evolution: adaptive landscapes and theoretical morphospaces*. Cambridge University Press, Cambridge
- McKinney FK, Raup DM (1982) A turn in the right direction: simulation of erect spiral growth in the bryozoans *Arquimedes* and *Bugula*. *Paleobiology* 8:101–112
- McNeely JA, Miller KR, Reid WV, Mittermeier RA, and Werner TB (1990) *Conserving the world's biological diversity*. International Union for conservation of nature and natural resources.
- Millennium Ecosystem Assessment (2005) *Ecosystems and human well-being: biodiversity synthesis*. World Resources Institute, Washington, DC
- Naess A (1986) The deep ecology movement: some philosophical aspects. *Philos Inq* 8:10–31
- Nash R (1990) *The rights of nature: a history of environ ethics*. Primavera Press, Leichhardt
- Nixon KC, Wheeler Q (1992) Measures of phylogenetic diversity. In: Novacek MJ, Nixon KC (eds) *Extinction and phylogeny*. Columbia University Press, New York, pp 216–234
- Norton BG (1984) Environmental ethics and weak anthropocentrism. *Environ Ethics* 6:131–148

- Norton BG (1988) Commodity, amenity, and morality: the limits of quantification in valuing biodiversity. In: Wilson EO (ed) *Biodiversity*. National Academy Press, Washington, DC, pp 200–205
- O'Neill J (1992) The varieties of intrinsic value. *Monist* 75:119–137
- Pavoine S, Bonsall MB (2011) Measuring biodiversity to explain community assembly: a unified approach. *Biol Rev* 86:792–812
- Pavoine S, Ollier S, Dufour AB (2005) Is the originality of a species measurable? *Ecol Lett* 8:579–586
- Petchey OL, Gaston KJ (2006) Functional diversity: back to basics and looking forward. *Ecol Lett* 9(6):741–758.
- Pereira HM et al (2013) Essential biodiversity variables. *Science* 339(6117):277–278
- Redding DW, Mooers AO (2006) Incorporating evolutionary measures into conservation prioritisation. *Conserv Biol* 20:1670–1978
- Reyers B et al (2012) Finding common ground for biodiversity and ecosystem services. *BioScience* 62(5):503–507
- Rolston H (2001) Biodiversity and endangered species. In: Jamieson D (ed) *A companion to environmental philosophy*. Blackwell, Oxford, pp 403–415
- Sarkar S et al (2006) Biodiversity conservation planning tools: present status and challenges for the future. *Annu Rev Environ Resour* 31:123–159
- Simaika J, Samways M (2010) Biophilia as a universal ethic for conserving biodiversity. *Conserv Biol* 24(3):903–906
- Sober E (1986) Philosophical problems for environmentalism. In: Norton B (ed) *The preservation of species*. Princeton University Press, Princeton, pp 173–194
- Soulé ME (1985) What is conservation biology? *BioScience* 35:727–734
- Srivastava DS et al (2012) Phylogenetic diversity and the functioning of ecosystems. *Ecol Lett* 15:637–648
- Stone C (1972) Should trees have standing? William Kaufmann, Los Altos
- Tilman D et al (2001) Diversity and productivity in a long-term grassland experiment. *Science* 294:843–845
- United Nations Convention on Biological Diversity (1992)
- van Kooten GC, Bulte EH (2000) *The economics of nature: managing biological assets*. Blackwell, Oxford
- Vandamme AM (2009) Basic concepts of molecular evolution. In: Salemi M, Vandamme AM, Lemey P (eds) *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing*. Cambridge University Press, Cambridge, pp 3–30
- Vane-Wright RI, Humphries CJ, Williams PH (1991) What to protect – systematics and the agony of choice. *Biol Conserv* 55:235–254
- Vellend M et al (2011) Measuring phylogenetic biodiversity. In: Magurran AE, McGill BJ (eds) *Biological diversity: frontiers in measurement and assessment*. Oxford University Press, Oxford, pp 193–206
- Webb CO et al (2002) Phylogenies and community ecology. *Annu Rev Ecol Syst* 33:475–505
- Weihel E (2011) A primer of trait and functional diversity. In: Magurran AE, McGill BJ (eds) *Biological diversity: frontiers in measurement and assessment*. Oxford University Press, Oxford, pp 175–193
- West-Eberhard MJ (2003) *Developmental plasticity and evolution*. Oxford University Press, Oxford
- Wilson EO (1984) *Biophilia*. Harvard University Press, Cambridge, MA
- Winter M, Devictor V, Schweiger O (2012) Phylogenetic diversity and the nature of conservation: where are we? *Trends Ecol Evol* 28:199–204
- Worm B et al (2006) Impacts of biodiversity loss on ocean ecosystem services. *Science* 314:787–790

# The PD Phylogenetic Diversity Framework: Linking Evolutionary History to Feature Diversity for Biodiversity Conservation

Daniel P. Faith

**Abstract** Feature diversity refers to the relative number of different features represented among species or other taxa. As a storehouse of possible future benefits to people, it is an important focus for biodiversity conservation. The PD phylogenetic diversity measure provides a way to measure biodiversity at the level of features. PD assumes an evolutionary model in which shared features are explained by shared ancestry. This avoids philosophical and practical weaknesses of the conventional interpretation of biodiversity as based on some measure of pair-wise differences among taxa. The link to features also provides a family of PD-based calculations that can be interpreted as if we are counting-up features of taxa. The range of feature diversity calculations assists comparisons of methods, and helps overcome the current lack of review and synthesis of the variety of proposed methods for integrating evolutionary history into biodiversity conservation. One family of popular indices is based on the evolutionary distinctiveness (ED) measure. These indices all have the limitation that complementarity, reflecting degree of phylogenetic overlap among taxa, is not properly taken into account. Related indices provide priorities or other scores for geographic areas, but do not effectively combine complementarity, probabilities of extinction, and measures of restricted-range. PD-based measures can overcome these problems. Applications include the identification of key biodiversity sites of global significance for biodiversity conservation.

**Keywords** Option value • Endemism • IPBES • Philosophy of science • IUCN

## Introduction

This book addresses important concepts, methods, and applications related to the increasingly important role of evolutionary history in biodiversity conservation. The preservation of the rich heritage represented by the evolutionary history of taxa is a

---

D.P. Faith (✉)

AMRI, The Australian Museum, Sydney, NSW 2010, Australia

e-mail: [dan.faith@austmus.gov.au](mailto:dan.faith@austmus.gov.au)

© The Author(s) 2016

R. Pellens, P. Grandcolas (eds.), *Biodiversity Conservation and Phylogenetic Systematics*, Topics in Biodiversity and Conservation 14,  
DOI 10.1007/978-3-319-22461-9\_3

natural conservation goal (e.g. Mooers and Atkins 2003). This fundamental relationship between evolutionary history and conservation goals traces back at least to the IUCN (1980) proposal that taxonomically distinctive species may deserve greater conservation priority. At about the same time, Soulé (1980), in his book, *Conservation biology: an evolutionary-ecological perspective*, articulated a broad evolutionary perspective for conservation, and argued that “reduction of the biological diversity of the planet is the most basic issue of our time.”

The term “phylogenetic diversity” is relevant to these biodiversity conservation perspectives. The term can be traced back to the introduction of the “PD” phylogenetic diversity index (Faith 1992a, b, 1994a). PD was designed as a simple measure of the degree of representation of evolutionary history (by a given set of taxa). Faith (2002) summarised the basic definition and rationale for PD: “representation of “evolutionary history” (Faith 1994b) encompassing processes of cladogenesis and anagenesis is assumed to provide representation of the feature diversity of organisms. Specifically, the phylogenetic diversity (PD) measure estimates the relative feature diversity of any nominated set of species by the sum of the lengths of all those phylogenetic branches spanned by the set.”

That summary mentions species, but Faith (1992a, b) in fact applied PD from the outset not only to phylogenies whose tips were species, but also to phylogenetic pattern among genetic haplotypes or populations, in order to set spatial priorities to conserve within species genetic diversity (see also Faith et al. 2009). The common element across these levels is the inference of underlying diversity, where the units of variation are features or traits of taxa. This link to “features” reflects the attempt, through PD calculations, to address a fundamental concern of biodiversity conservation - unknown variation, with unknown future values. Faith (1992a, b) suggested that the interpretation of phylogenetic diversity as a measure of feature diversity helps to clarify its link to conservation values: “Diversity is seen as important as the raw material for adapting to change (McNeely et al. 1990), and so provides what McNeely et al. (1990) and others call ‘option value’: a safety net of biological diversity for responding to unpredictable events or needs. The diversity of features represented by a subset of species provides option value in ensuring not only that one or more members of the subset can adapt to changing conditions, but also that society may be able to benefit (e.g. economically) from features of these species in response to future needs.”

Examples of these benefits include many from bioprospecting. For example, Smith and Wheeler (2006) have used phylogeny to assess potential for new discoveries of piscine venoms. Pacharawongsakda et al. (2009) have applied PD to help find natural products from microbes. Another interesting example is found in the study of Saslis-Lagoudakisa et al. (2012). Phylogenetically-related plants have provided a key medical component, discovered independently in the plants found in three different regions.

This perspective accords well with the IUCN (1980) argument for conservation of diversity in order to ensure benefits “for present and future use”. Reid and Miller (1989) echoed these ideas in their early paper, “Keeping options alive: the scientific basis for conserving biodiversity” (see also Wilson 1992; McNeely 1988; Faith 1992a, b). The Millennium Ecosystem Assessment (MA 2005) summarised this

general link between biodiversity and option values: “Biodiversity loss is important in its own right because biodiversity has cultural values, because many people ascribe intrinsic value to biodiversity, and because it represents unexplored options for the future (option values)”.

Option value therefore reflects not only the unknown future benefits from known elements of biodiversity, but also the unknown benefits from unknown elements. The Millennium Ecosystem Assessment (2005) also called for “a ‘calculus’ of biodiversity, so that gains and losses at the level of biodiversity option values can be quantified”. These ideas are echoed in the conceptual framework for the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES; UNEP 2013) which says that values “include bequest value – in other words, the preservation of nature for future generations – or the option values of biodiversity as a reservoir of yet-to-be discovered uses from known and still unknown species and biological processes, or as a constant source, through evolutionary processes, of novel biological solutions to the challenges of a changing environment.”

The PD measure is an attempt to make inferences about “features” as units of variation, including features that are not yet known to science. Faith (1994a, b) characterised PD as one case of a general framework for biodiversity assessment that uses pattern-process models to link objects and lower-level units. In general, the biodiversity units are the things we would like to count up, and the objects contain various units. Typically, many units remain unobserved/unknown, and a pattern-process model defines relationships among the objects, enabling inference of the relative numbers of units represented by different sets of objects (Faith 1994a, b). Thus, PD provides the specific case where species (or haplotypes or populations) are the objects, features are the units, and the pattern-process inferential model is based on evolutionary processes of cladogenesis and anagenesis, manifested in phylogenetic pattern.

The link from phylogeny to feature diversity has supported the wide application of PD. For example, Huang et al. (2012) advocated the use of PD in conservation based on their finding that it provides a much stronger link to “trait diversity”, relative to species. Jono and Pavoine’s (2012) study of threat diversity as a determinant of the extinction risk in mammals assessed the consequences of species declines used PD with the rationale that it “is becoming a key criterion in conservation studies because it can reflect the variety of unique or rare features of a species.”

This rationale has extended to application of PD within ecosystems, where the conservation/management goals focus on maintaining ecosystem functions and services. For example, Cadotte and Davies (2010) argued that “maximizing the preservation of PD will also tend to maximize the preservation of feature diversity, including unmeasured, but ecologically important traits” (see also Gravel et al. 2012).

Studies also link PD, feature diversity, and option values. For example, Larsen et al. (2012) argued that “it is difficult to provide a robust proxy for ‘option value’ – the potential value to society – as these values are not yet realized”, and concluded that “a compelling argument can be made that maximizing the retention of phylogenetic diversity (PD) should also maximize option value, as well as diversification and adaptation of the species in a future of climatic change”. The influential study of Forest et al. (2007) also highlighted the importance of PD as a link to feature



diversity. They explored PD and option value based on an estimated phylogenetic tree and the geographic distribution of angiosperm genera found in the Cape hotspot of South Africa. Forest et al. (2007) concluded that, if we did not know about the medicinal, food, and other useful features of these plants, then preserving sets of species with high PD would be a good way to preserve these unknown features and their associated benefits.

PD now is regarded as “a leading measure in quantifying the biodiversity of a collection of species” (Bordewich and Semple 2012) and as “a resonant symbol of the current biodiversity crisis” (Davies and Buckley 2011), with important applications at both regional/global (e.g. Forest et al. 2007) and within-ecosystem scales (e.g. Cadotte et al. 2009). At the same time, PD must be acknowledged as just one of many biodiversity measures that are based on aspects of evolutionary history (see other chapters in this book). Unfortunately, there is no existing comprehensive review and synthesis covering all these measures. For example, Diniz Filho et al. (2013) recently concluded that “we do not even have a comprehensive and integrative approach to using phylogenies in biodiversity conservation.” Similarly, a recent review of past studies on the topic of evolutionary history and conservation (Winter et al. 2013) argued that there is little basis for distinguishing among the large number of existing phylogenetic indices (see also Devictor et al. 2010).

Partly, the existence of a gap in review and synthesis is not surprising; this area of research is evolving rapidly. The PD measure is applied in various sub-disciplines, highlighting distinctions between within-ecosystem versus global scales, microbial versus macrobial, and taxonomic levels ranging from populations to species and higher taxa (e.g., May-Collado and Agnarsson 2011; Lozupone and Knight 2005; Jono and Pavoine 2012; Jetz et al. 2014).

The other obstacle to synthesis is that, while some attempts at review and synthesis have been made, most have been incomplete or unsuccessful. Notably, philosophers of science have become keenly interested in the science of phylogeny and biodiversity conservation, but have not yet shed much light on the problem (for discussion, see Faith 2013). Philosophers so far largely have focussed on one possible unifying conceptual model of biodiversity. This model traces back to Weitzman’s (1992) general framework for biodiversity, based on the idea of objects, and measures of difference between pairs of objects. The biodiversity of a given set of objects then is reflected, not in a list of the different objects, but in the amount of difference represented by the set. Weikard (2002), following Weitzman’s object-differences framework, argued that “an operational concept of diversity must rely on some measure of dissimilarity between appropriately defined objects.” Maclaurin and Sterelny (2008), in their book, “What is biodiversity?”, and Morgan (2010) also saw this approach as a core framework for characterising biodiversity (the Lean and Maclaurin chapter “The Value of Phylogenetic Diversity”, also takes this as their starting point).

This approach assumes that we can decide on the definition of meaningful differences among the initial objects, and most authors have acknowledged that it is hard to choose among many possible notions of difference. This has not helped in developing a synthesis for phylogenetic measures of diversity. Winter et al. (2013) incor-

rectly interpreted “phylogenetic diversity” as any measure derived from a nominated between-species phylogenetic distance. Their conclusion, that there is little basis for distinguishing among different phylogenetic indices, highlighted well the problems in choosing among different notions of differences. Unfortunately, Winter et al. did not recognize PD as distinctive in avoiding arbitrary notions of difference, and instead using a model-based measure of feature diversity and option values.

A more recent study, by Kelly et al. (2014), acknowledged the feature diversity interpretation of PD, but surprisingly failed to acknowledge its pattern-process model, in which shared ancestry explains shared features. An implication of that model, emphasised from the outset, was that PD will fail to account for convergently-derived features, and that these may be captured by an alternative pattern process model (see Faith 1992a, b, 1996, 2015). The failure to recognise these key lessons from the early work left Kelly et al. destined to merely re-discover the already well-established point that convergences will not be accounted for by PD, rather than making any real progress towards evaluation and synthesis (and perhaps exploring the alternative pattern-process model).

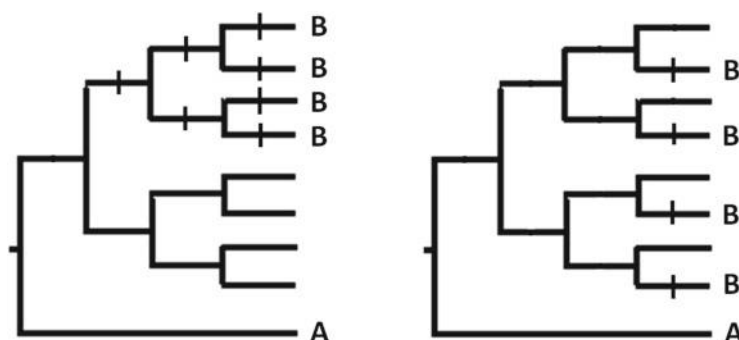
Lack of comparisons and synthesis has made it difficult to interpret some otherwise useful studies. This problem is well illustrated in the recent study by Pio et al. (2014), where “PD” is used to refer to any diversity measure linked in any way to phylogeny. They refer to a variety of published studies on the performance of “PD”, but the reader cannot know when this refers to true PD and when it refers to some other measure. Pio et al. go on to apply the actual PD method in their analyses, but without reference to that as the Faith (1992a) PD method.

Beyond the confusion in terms, there remains a genuine need to compare methods and develop synthesis. The pattern-process model approach that is the basis for PD can help in two ways. First, we can use the PD family of calculations to better recognise that there are many inter-linked, related, indices (dissimilarity, endemism, etc) rather than lots of indices that can be called “diversity” measures (for related discussion, see Sarkar 2008). In the next section, I briefly consider the PD’s counting-up of features as one way to integrate other possible calculations that can be based on those counts. I then turn to the second way that PD’s pattern-process model can help. Here, I will evaluate alternative measures, including those outside PD framework, by examining how well they can be interpreted under the PD features model.

## Calculations and Comparisons

### *Simple Calculations Based on PD*

Many possible calculations can be based on counting-up features within the PD framework. As examples, complementarity, endemism, and dissimilarities between objects all can be calculated. In principle, every index conventionally defined in ecology at the species level has its counterpart for other biodiversity units. Counting-up the total number of features (as units) represented by a set of taxa



**Fig. 1** For each tree, the tick marks correspond to loss in PD if each species from area B is lost. The tick marks show how much PD is uniquely represented by that area. PD endemism sees the scenario on the left as implying greater endemism of area B, compared to the scenario on the right. The  $W_e$  method cannot distinguish between the two scenarios because it ignores a critical aspect of phylogenetic context, called complementarity

remains the core measure of “diversity”, but the other calculations capture other aspects – for example, expected change in biodiversity as a result of extinction.

Useful PD calculations for biodiversity comparisons among geographic localities include PD-dissimilarities between places or samples (see Lozupone and Knight 2005) and PD-endemism (Faith et al. 2004; illustrated in Fig. 1). Another useful calculation is “expected PD”, based on estimated probabilities of extinction. Here, species’ estimated extinction probabilities indicate amounts of “expected PD loss” (discussed further below; see also Faith 2008, 2013). All these calculations operate as if we are applying the standard species-based measures at the features level. Thus, these newer calculations make sense, given the interpretation of PD as counting-up features.

This interpretation has helped to justify other recent proposed extensions of PD. One important case is the integration of abundance information. Faith and Richards (2012) noted that a PD-based Hill numbers framework (Chao et al. 2010; see also Chao et al. chapter “[Phylogenetic Diversity Measures and Their Decomposition: A Framework Based on Hill Numbers](#)”) can be interpreted as an application of the standard species-level Hill numbers calculation, but with evolutionary features (as indicated by PD) substituted for species. Thus, the basic PD evolutionary model provides a simple justification for a phylogenetic measure integrating abundance information.

## Complementarity: A Key PD Attribute

Interpretation of PD as counting-up features extends the fundamental species-level measure of “complementarity” to the features level. A taxon complements others in representing additional evolutionary history (Faith 1994a, b), as depicted in the

branches of the estimated phylogeny. The degree of complementarity reflects the relative number of additional features contributed by that species. For example, given some subset of species that are well-protected, and two species in that taxonomic group that are endangered, the priority for conservation investment may depend on the relative gains in feature diversity (the complementarity values) expected for each species.

Given the importance of complementarity, particularly when dealing with complex conservation issues, it is worth comparing PD with some published phylogenetic calculations. Calculating PD naturally requires that phylogenetic overlap among taxa be taken into account, so that branches – and corresponding features – are not multi-counted. Often, when PD is not applied correctly, the result is a misleading multiple-counting of features. For example, Perez-Losada et al. (2002) incorrectly calculated PD values for sets of freshwater crab species. They simply added up the PD values for individual taxa to produce the overall score for the set of taxa. Consequently, their measure, in multi-counting branches, did not correspond to a valid calculation of PD. Similarly, a study by Vamosi and Wilson (2008), using the term “EH” to refer to evolutionary history, stated that “the combined EH of all the angiosperm orders and families was estimated at 35,244 million years by summing the ages of the separate clades over the angiosperm phylogeny.” Their “combined EH” measure, in multi-counting branches, did not correspond to an estimate of PD. PD calculations would have better captured their intention to assess loss of traits/features.

### ***Calculations Using Phylogenetic Distinctiveness Fail to Integrate Complementarity***

More complex calculations have used measures of phylogenetic or taxonomic “distinctiveness”. These values, calculated for individual taxa, are then to be combined to score sets of taxa or areas. The problem for all popular variants of this approach – whether the terminal taxa (or tips for the tree) are individuals, populations, or places, is that the scores for the taxa do not add up to the proper scores for sets of taxa.

In an early example of such an approach (López-Osorio and Miranda-Esquivel 2010), an area received a score equal simply to the sum of individual scores of member species. López-Osorio and Miranda-Esquivel (2010) used 50 phylogenies covering multiple taxonomic groups in the Amazon, and integrated this phylogenetic information into conservation priority setting in order to “establish conservation priorities for Amazonia’s areas of endemism on the basis of measures of evolutionary distinctiveness”. “Taxonomic rarity” was to be indicated by species that are members of a small number of groups on the cladogram. López-Osorio and Miranda-Esquivel (2010) used an approach suggested by Posadas et al. (2001), which extends the W Index of Vane-Wright et al. (1991). The W index assigns to each species a value that is inversely related to the count of the number of groups on

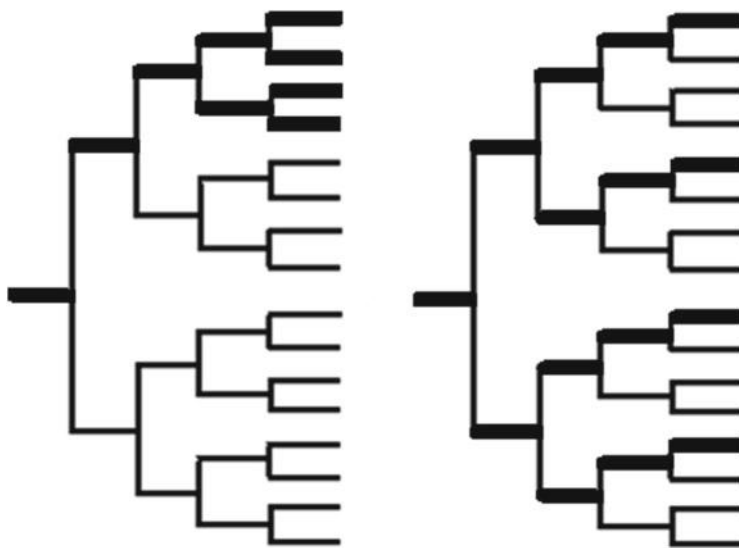
the phylogenetic tree for which the species is a member. Thus, a species that is taxonomically (phylogenetically) distinctive will have a high  $W$  value reflecting its relatively few close relatives. The key index derived from  $W$  is the  $W_e$  index (each  $W$  value is divided by the number of areas with that species, yielding  $W_e$ ). An area receives a score, equal to the sum of the  $W_e$  values of its species. This is to indicate a degree of endemism that integrates phylogeny.

Faith et al. (2004) compared those measures to the phylogenetic diversity measure, PD, and its associated calculations. Faith et al. argued that the  $W_e$  indices for areas differ from PD in not considering the degree of phylogenetic overlap/non-overlap among species (phylogenetic complementarity), and so may fail to effectively represent evolutionary history in priority sets of species or areas. A simple example of the problem is illustrated in Fig. 1. The  $W_e$  method cannot distinguish between the scenarios, yet the PD-endemism value differs for the two.

A family of relatively new measures, while based on PD, also does not fully account for complementarity. ED (“evolutionary distinctiveness”; Isaac et al. 2007; see also Collen et al. 2011) divides up the total PD among all species on the given phylogeny. This provides a fixed score for each species, reflecting its contribution to the total evolutionary history (PD). A species receives a partial credit for each ancestral branch. Thus, ED appears to capture the idea of complementarity among species. However, a key limitation is apparent when species ED scores are combined to provide scores for areas or for sets of priority species. Here, the ED approach does not take phylogenetic complementarity among the species into account. For example, consider the phylogenetic tree in Fig. 2. Based on summed ED scores, we cannot distinguish between an area with four closely related species and an area with four distantly related species; yet the scenario on the right corresponds to higher PD.

Such limitations may be critical in assessing diversity within communities or assemblages. In this context, phylogenetic diversity may be predictive of functionality or productivity (Cadotte et al. 2009). Dalerum (2013) set out to investigate the possible correspondence between phylogenetic diversity and functional diversity for assemblages of large carnivores. While Dalerum referred to “phylogenetic diversity” and to “PD”, in fact, their study used ED, not PD. Dalerum calculated ED for each species and then “estimated the ED of each assembly as the sum of the ED of contributing species.” As the simple example of Fig. 2 shows, this summed ED score will not correspond to the total PD. Unfortunately, the Dalerum study therefore provides little useful evidence for the claimed relationship between phylogenetic and functional diversity in assemblages of large terrestrial carnivores.

These same issues arise for regional or global studies. An interesting study by Daru et al. (2013) on mangroves “identified biogeographic regions that are relatively species-poor but rich in evolutionary history.” While the study presented results referring to loss of “mangrove phylogenetic diversity”, in fact, the measure used was based on ED calculations. Daru et al. argued for the significance of the finding that “areas with a high proportion of species experiencing global declines correspond to areas of unique evolutionary history” arguing that “the loss of currently threatened species might still have a disproportionate impact on mangrove



**Fig. 2** Two drawings of a hypothetical phylogenetic tree. For this simple tree, the ED value is the same for every species. Given the unit length branches, it is  $1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} = 1.94$ . Dark branches in each case indicate the PD represented by the species in an area. On the *left*, the area has four closely related species and on the *right*, the area has four distantly related species – and higher total PD. The PD on the *left* is 9 units, compared to a much higher PD of 15 on the *right*

phylogenetic diversity regionally”. This conclusion was based on apparent “overlap between regions in which species are undergoing declines and regions rich in evolutionarily distinct species.” Unfortunately, their use of a sum of species’ ED values as the regional indicator of phylogenetic diversity loss provides only weak evidence. To see this I again consider Fig. 1. For both trees, the sum of the ED values for the four species found in area B is the same. Thus, ED cannot distinguish between the large PD loss when the species are phylogenetically clumped, and the smaller PD loss when the species are phylogenetically dispersed (as in Fig. 1, left). Again, the PD loss corresponding to an area loss is not well-indicated by total ED, because phylogenetic complementarity is ignored.

A contrasting study is that of Abellán et al. (2013), who found that most of the highly evolutionarily distinct and vulnerable taxa were not covered by any national parks. Critically, while distinctiveness was noted, their proposed solution was based on priorities for areas providing increased PD. They concluded that “when additional conservation areas were selected maximizing the number of unrepresented species, the variation in PD could be very high, and as a consequence, depending on the group and the number of areas added, they could preserve much less evolutionary history than when they were specifically selected to maximize PD.”

The weakness of summed ED scores resembles the limitations of the López-Ororio and Miranda-Esquivel method. This kind of problem seems to link to a long-standing idea that we simply might add up scores for individual taxa, perhaps with

some distinctiveness “weighting”. For example, Gotelli and Chao (2013), in the *Encyclopedia of Biodiversity*, claim that we can calculate “PD” by appropriately weighting the species and then applying conventional species indices such as richness: “The concept of traditional diversity can therefore be extended to consider differences among species.... Differences among species can be based directly on their evolutionary histories, either in the form of taxonomic classification (referred to as taxonomic diversity) or phylogeny (referred to as phylogenetic diversity (PD)) ... weighting each species by a measure of its ...phylogeny.”

The relationship between ED and PD has been investigated previously for calculations that use probabilities of extinction. An EDGE score (Isaac et al. 2007) simply multiplies extinction probability by ED- evolutionary distinctiveness (a score that gives each species some partial credit for ancestral branches). Naturally, that arbitrary partial credit and multiplication is not a particularly good way to determine changing expectations about the diversity that persists as the status of species changes. Faith (2008) showed how the arbitrary partial credit and multiplication in EDGE-type methods does not take phylogenetic complementarity into account, and so will not do a good job in determining conservation priorities delivering high expected PD. Faith also suggested that such priorities can be set by directly looking at expected PD gains and losses. May-Collado and Agnarsson (2011) and Kuntner et al. (2011) also concluded that the PD methods are better in achieving the goal of phylogeny-based conservation than EDGE.

These results are relevant to an interesting study by Safi et al. (2013), who set out to “identify regions of the world where priority species are concentrated, much like the original definition of the biodiversity hotspot.” They identified those regions/countries having the “highest accumulation of top mammal species ranked in terms of their EDGE score” and argued that “Conservation resources would therefore be best allocated among the countries in these regions to protect mammal species with the highest EDGE scores.”

Unfortunately, this may be a weak guideline for the efficient use of limited conservation resources. Their study recalls the issues raised by the use of ED methods in the Daru et al. study, where a given ED score could correspond either to phylogenetically clumped species and a large PD loss (as in Fig. 1, left), or phylogenetically dispersed species and smaller PD loss (Fig. 1, right). Once again, the potential PD loss arising from a given area loss is not well-indicated by a summation of ED (or EDGE values), because phylogenetic complementarity is ignored.

Recent extensions of the ED methods provide some important modifications to take into account species’ range extent and abundance; however, these interesting innovations may suffer similar problems to those described above. Cadotte et al. (2010) introduced one important extension by taking into account numbers of individuals of a given species in a community or ecosystem. The rationale, analogous to that of conventional ED, is that individuals differ in their representation of evolutionary history or phylogenetic diversity, and can receive partial “credit” for a given ancestral branch. Given that PD has been linked to ecosystem functioning (e.g. Cadotte et al. 2008, 2009), the loss of some individuals (e.g. those from species with few individuals and uniquely representing some long branches) should set off alarm



bells if we want to maintain ecosystem functions. Cadotte et al. argue that their measure “can be used by managers to identify individuals, and by extension species, whose loss corresponds to the greatest loss of evolutionary information. If, as has been proposed, evolutionary history captures functional diversity necessary for ecosystem processes and services (e.g. see Cadotte et al. 2008), minimizing this loss of evolutionary diversity might maximize the preservation of ecosystem function.”

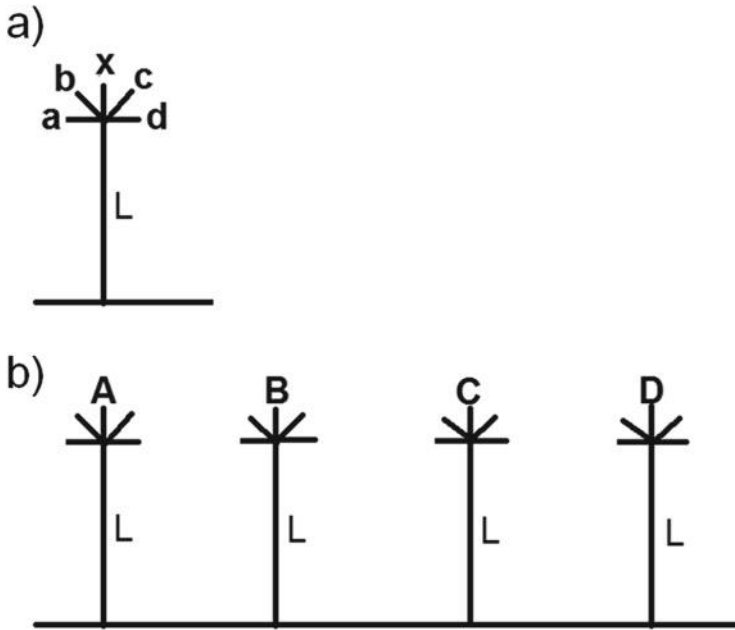
Their basic measure, AED<sub>i</sub>, follows the partitioning logic of ED; here, it records the share of all branches credited to any individual of species *i*. A problem is that, when AED<sub>i</sub> values are summed over individuals, complementarity once again is ignored. This implies that the score for a set of individuals (say, those lost under a nominated management regime) cannot be a reliable indicator of potential PD loss – yet it is PD that matters, given its link to functions. We can see the problem by adapting the example of Fig. 1, imagining that the terminal branches represent individuals. The AED scores for the set of four individuals on the left (marked with B) is the same as that on the right; yet, the loss of PD feature diversity and perhaps functional diversity is much greater in the scenario on the left. Consequently, there seems to be no justification for Cadotte et al.’s claim that AED can be “used by managers to identify individuals, whose loss corresponds to the greatest loss of evolutionary information. ... minimizing this loss of evolutionary diversity might maximize the preservation of ecosystem function.” For a single individual, AED<sub>i</sub> may be a useful index, but if a management strategy potentially impacts numerous individuals, AED will not provide a good comparative index of PD loss.

A measure similar to AED is the “biogeographically weighted evolutionary distinctiveness” metric (BED or BEDT; Cadotte and Davies 2010). BED extends ED by also partitioning the credit among (for example) the grid cells occupied by each species in a region. In this way, range extent information for species is incorporated along with phylogenetic distinctiveness. For species *i*, BED<sub>i</sub> is a weighted sum of the ancestral branch lengths. Each length is weighted by the inverse of the sum, over all descendent species of the branch, of the number of cells occupied by the descendent species (if each descendent species is found in just one cell, then BED<sub>i</sub> is the ED of species *i*). The BEDT score for a cell is the sum of the BED<sub>i</sub> scores for all species *i* found in the cell. Thus, restricted range species that also uniquely represent deep branches will count a lot in the overall scores for grid cells or other areas.

As an example, in Fig. 3, suppose that we can only protect one area. Which is best? For the Area (1) in Fig. 3a, the BEDT score is BED<sub>a</sub> + BED<sub>b</sub> + BED<sub>c</sub> + BED<sub>d</sub>. The BED<sub>i</sub> for each of these four member species (a, b, c, d) is the same, and is equal to  $m/1 + L/5$ . Here, the length *L* is divided by 5 because a, b, c, d, and x each are found in one area; thus, the sum of the number of cells occupied is 5. The BEDT score equals 4 times  $(m/1 + L/5)$ , or  $4m + 4(L/5)$ .

For the Area (2) in Fig. 3b, the BED<sub>i</sub> for each of the four member species again is the same, and equal to  $m/1 + L/5$ . The length *L* again is divided by 5 because A and the four sister species each are found in one area. The BEDT score for Area (2) is BED<sub>A</sub> + BED<sub>B</sub> + BED<sub>C</sub> + BED<sub>D</sub>, or  $4m + 4(L/5)$ . BEDT therefore makes no distinction between the two areas. In contrast, the PD offered by Area (2) is much greater. Thus, BED fails to detect a huge gain in raw PD (and in restricted range PD)





**Fig. 3** Portions of hypothetical phylogenetic trees occurring in two areas. (a) Area (1) uniquely has species a, b, c, d which are on small branches of length  $m$ , and are at the end of a long branch of length  $L$ . Species  $x$  is not found in Area (1), but uniquely occurs in some other area. (b) Area (2) uniquely has species A, B, C, D, which are on small branches of length  $m$ , and are at ends of different long branches of length  $L$ . For each member species, four other sister species on small branches of length  $m$  all uniquely occur in some other area

that could be achieved through protection of the area in Fig. 3b. BED (and the related method of Tucker et al. 2012), is not effective for setting conservation priorities that reflect both phylogenetic diversity and range-restrictedness. I conclude that there is little justification for Cadotte et al.'s conclusion that "Metrics such as BEDT, which combines evolutionary diversity and rarity into a single measure of diversity, may allow a more holistic approach to conservation prioritization."

I noted above that PD gives priority to Area 2 in Fig. 3b, because it offers almost 4 times as much PD. However, this basic PD calculation does not take range rarity into account. Weighted PD-endemism or "PE" (the sum of branches represented in an area, each inverse-weighted by its range, expressed as number of cells; Rosauer et al. 2009) also gives priority to Area 2, because it scores Area 1 with a PE score of  $4m + L/2$ , and Area 2 with a higher PE score of  $4m + 4(L/2)$ .

PE has an interesting property analogous to ED, in that a given cell receives proportional credit for a branch (analogous to the basic ED score where a species gets proportional credit for branches). PE performs well in the example above; however, it shares a weakness of ED, when combined with probabilities and summed-up to provide overall scores. To see this, I consider a recent study of the phylogeny of Malagasy lemuriformes (Gudde et al. 2013). This study set out to identify places

with a concentration of threatened phylogenetic distinctive and rare species. Here, the PE measure was combined with probabilities of extinction. Their “imperilled phylogenetic endemism” (IPE) index is the sum over all branches of branch length times its probability of extinction (product of extinction probabilities of all descendants) times the inverse of its range-extent.

Gudde et al. (2013) claimed to “quantify where on the landscape at-risk evolutionary history is concentrated.” However, their “imperilled phylogenetic endemism” (IPE) index appears to have the weakness that it could highlight places that have no threatened branches at all. As a revealing example, suppose that area A has 20 species, all of IUCN “least concern” (see IUCN 2006, 2012). Suppose that this corresponds to a low probability of extinction of 0.025 (for methods and discussion, see Mooers et al. 2008; Faith and Richards 2012). Each species is found in only ten areas. Suppose that area B has five species, all IUCN “critically endangered” (probability of extinction assumed to be a higher 0.4). Each species is found in 50 areas, but all are found together in this one area. Suppose also that each species is at the end of a branch of some unit length. Also, for simplicity, I will ignore deeper branches (assuming that all species have numerous secure sisters).

IPE in this simple case is equal to the product of the number of branches, the probability of extinction and the inverse of the number of cells containing a given branch. Application of IPE gives area A the higher priority; the IPE score equals 20 times 0.025 times 1/10 or 0.05. IPE gives area B the lower priority; the IPE score equals 5 times 0.4 times 1/50 or 0.04. Application of IPE therefore would ignore the opportunity to save, with a reserve based around area B, five critically endangered species. Instead, IPE would give preference to an area with 20 non-threatened species! This reveals the key limitation of the approach. IPE is supposed to reflect a concentration of range restricted, threatened species. Gudde et al. (2013) argued that “our mapping does indeed quantify where at risk PD is concentrated”. However, IPE, in the example above, actually quantified where *not-at-risk* PD was concentrated!

This weakness of IPE is similar to that of EDGE (see above and Faith 2008). Both methods suffer the weakness that phylogenetic overlap of species is not effectively taken into account. For EDGE type assessments, an existing probabilistic PD approach (Witting and Loeschcke 1995) performs better (Faith 2008; see also May-Collado and Agnarsson 2011; Kuntner et al. 2011). In the final section, I examine the prospects for using this “expected PD” approach to address some conservation assessment problems that have been unsuccessfully treated by the ED type methods.

The PE measure is relevant to another study that attempts to integrate range extent and threat information into PD assessments. In their global study on conservation of phylogenetic diversity of birds, Jetz et al. (2014) devised a measure related to ED to provide scores for regions or areas. Their “EDR” score for a species is simply the ED value divided by the range (number of occupied cells) of the species. Total EDR for a given region then is the summed EDR of all species occurring in the region. Jetz et al. ask, “Under an objective of minimizing global PD loss, how do ED and EDR perform as metrics for a rule-based approach to taxon- and

area-based conservation priority setting?” They claim that EDR indicates high priority conservation areas. However, this modified ED score, when summed to produce EDR area scores, again will not reflect PD (Fig. 2), nor amount of PD that would be lost (Fig. 1).

An alternative, incorporating range information, is a modification of PE.

A threatened-PE (TPE) area score only counts up threatened branches (e.g. those having only threatened descendents; see also Faith 2015). If the range-extents of many species are declining, TPE may be an effective simple index to monitor over time. The TPE of an area will increase if more of its species/branches are threatened or if range extent decreases for some of its species.

## Prospects

In the examples presented above, assessments of sets of taxa (and/or areas) focussed on two related goals. One was the assessment of losses in PD (as in Fig. 1) and the other was assessment of gains in PD (as in Fig. 2). Regarding gains, it is apparent that some indices may fail to record a large gain in PD, because they do not detect the degree to which a set of taxa is spread out phylogenetically. Regarding losses, some indices may miss a large loss in PD because they do not take into account the fact that a set of taxa are clumped phylogenetically. The latter case is a particularly important one, given that these scenarios may correspond to “phylogenetic tipping points”, where long, deeper, branches of the phylogeny are lost (see Faith et al. 2010; Faith and Richards 2012),

The theme of PD gains and losses is a critical one also for the conservation assessment of geographic areas. For species/taxon priorities, the expected PD methods have advantages over the ED and EDGE approaches for estimating expected gains or expected losses (Faith 2008). The application of expected PD by Jono and Pavoine (2012), noted above, provided an example of such an effective assessment of PD expected gains or losses. We also need effective estimates of the expected PD gains or expected PD losses for entire areas or regions.

Expected PD will have advantages over other methods for assessments of areas. For example, the study of Safi et al. (2013), discussed above, highlighted the importance of identifying regions having a concentration of threatened species and phylogenetic diversity. However, they focussed on the “highest accumulation of top mammal species ranked in terms of their EDGE score.” Similarly, Gudde et al. (2013) set out to identify places with a concentration of threatened phylogenetically distinctive and rare species. Both studies, while identifying important assessment issues for the future, unfortunately applied methods that do not fully integrate the principle of phylogenetic complementarity. The expected PD framework may provide an effective way to address such assessment goals.

The identification of Key Biodiversity Areas (KBAs) is one important context for future work of this kind. KBAs are defined as sites of global significance for biodiversity conservation: “contributing significantly to the global persistence of

biodiversity” (see [http://www.iucn.org/about/work/programmes/gpap\\_home/gpap\\_biodiversity/gpap\\_wcpabiodiv/gpap\\_pabiodiv/key\\_biodiversity\\_areas/](http://www.iucn.org/about/work/programmes/gpap_home/gpap_biodiversity/gpap_wcpabiodiv/gpap_pabiodiv/key_biodiversity_areas/); Foster et al. 2012). KBAs typically are identified based on the presence of globally threatened (and/or geographically restricted) species. However, a gap exists in defining and identifying KBAs at the genetic and phylogenetic levels. Expected PD calculations could fill this gap in providing information about both expected gains and expected losses.

As an example, we could examine the gain in expected PD, if a given KBA were protected (probabilities of extinction transformed to some small value). This would be useful in revealing a concentration of threatened PD. On the other hand, we could examine the loss in expected PD if the area was lost (received no protection). This would be useful, in contrast to the IPE measure of Gudde et al. (2013), in revealing areas that have geographically restricted elements of threatened PD. Future work may examine how these basic calculations of expected gains and losses can be used in combination to defined priorities for KBAs and other geographic areas as conservation foci.

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Abellán P, Sánchez-Fernández D, Picazo F, Millán A, Lobo JM, Riber I (2013) Preserving the evolutionary history of freshwater biota in Iberian. *Natl Parks Biol Conserv* 162:116–126
- Bordewich M, Semple C (2012) Budgeted nature reserve selection with biodiversity feature loss and arbitrary split systems. *J Math Biol* 64:69–85
- Cadotte MW, Cardinale BJ, Oakley TH (2008) Evolutionary history and the effect of biodiversity on plant productivity. *Proc Natl Acad Sci U S A* 105(44):17012–17017
- Cadotte MW, Cavender-Bares J, Tilman D, Oakley TH (2009) Using phylogenetic, functional and trait diversity to understand patterns of plant community productivity. *PLoS One* 4(5):e5695
- Cadotte MW, Davies TJ (2010) Rarest of the rare: advances in combining evolutionary distinctiveness and scarcity to inform conservation at biogeographical scales. *Divers Distrib* 16:376–385
- Cadotte MW, Davies TJ, Regetz J, Kembel SW, Cleland EE, Oakley TH (2010) Phylogenetic diversity metrics for ecological communities: integrating species richness, abundance and evolutionary history. *Ecol Lett* 13:96–105
- Chao A, Chiu C-H, Jost L (2010) Phylogenetic diversity measures based on Hill numbers. *Philos Trans R Soc B* 365:3599–3609
- Collen B, Turvey ST, Waterman C et al (2011) Investing in evolutionary history: implementing a phylogenetic approach for mammal conservation. *R Soc Philos Trans Biol Sci* 366:2611–2622
- Dalerum F (2013) Phylogenetic and functional diversity in large carnivore assemblages. *Proc R Soc B* 280:20130049
- Daru BH, Yessoufou K, Mankga LT, Davies TJ (2013) A global trend towards the loss of evolutionarily unique species in mangrove ecosystems. *PLoS One* 8(6):e66686
- Davies TJ, Buckley LB (2011) Phylogenetic diversity as a window into the evolutionary and biogeographic histories of present-day richness gradients for mammals. *Philos Trans R Soc B* 366:2414–2425

- Devictor V, Mouillot D, Meynard C, Jiguet F, Thuiller W, Mouquet N (2010) Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: the need for integrative conservation strategies in a changing world. *Ecol Lett* 13:1030–1040
- Diniz-Filho JA, Loyola RD, Raia P, Mooers AO, Bini LM (2013) Darwinian shortfalls in biodiversity conservation. *Trends Ecol Evol* 28:689–695
- Faith DP (1992a) Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61:1–10
- Faith DP (1992b) Systematics and conservation: on predicting the feature diversity of subsets of taxa. *Cladistics* 8:361–373
- Faith DP (1994a) Phylogenetic pattern and the quantification of organismal biodiversity. *Philos Trans R Soc B* 345:45–58
- Faith DP (1994b) Phylogenetic diversity: a general framework for the prediction of feature diversity. In: Forey PL, Humphries CJ, Vane-Wright RI (eds) *Systematics and conservation evaluation*. Clarendon, Oxford, pp 251–268
- Faith DP (1996) Conservation priorities and phylogenetic pattern. *Conserv Biol* 10:1286–1289
- Faith DP (2008) Threatened species and the preservation of phylogenetic diversity (PD): assessments based on extinction probabilities and risk analysis. *Conserv Biol* 22:1461–1470
- Faith DP (2013) Biodiversity and evolutionary history: useful extensions of the PD phylogenetic diversity assessment framework. *Ann N Y Acad Sci* 1289:69–89
- Faith DP (2015) Phylogenetic diversity and extinction: avoiding tipping points and worst-case losses from the tree of life. *Proc R Soc Lond B* 370:20140011.
- Faith DP, Richards ZT (2012) Implications of climate change for the tree of life. *Biology* 1:906–932
- Faith DP et al (2004) Integrating phylogenetic diversity, complementarity, and endemism for conservation assessment. *Conserv Biol* 18:255–261
- Faith DP et al (2009) The cladistic basis for the Phylogenetic Diversity (PD) measure links evolutionary features to environmental gradients and supports broad applications of microbial ecology's "phylogenetic beta diversity" framework. *Int J Mol Sci* 10:4723–4741
- Faith DP, Magallón S, Hendry AP, Conti E, Yahara T, Donoghue MJ (2010) Evosystem services: an evolutionary perspective on the links between biodiversity and human-well-being. *Curr Opin Environ Sustain* 2:66–74
- Forest F et al (2007) Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature* 445:757–760
- Foster MN, Brooks TM, Cuttelod A, de Silva N, Fishpool LDC, Radford EA, Woodley S (2012) The identification of sites of biodiversity conservation significance: progress with the application of a global standard. *J Threatened Taxa* 4:2733–2744
- Gotelli NJ, Chao A (2013) Measuring and estimating species richness, species diversity, and biotic similarity from sampling data. In: Levin SA (ed) *Encyclopedia of biodiversity*, vol 5, 2nd edn. Academic, Waltham, pp 195–211
- Gravel D et al (2012) Phylogenetic constraints on ecosystem functioning. *Nat Commun* 3:1117
- Gudde RM, Joy JB, Mooers AO (2013) Imperilled phylogenetic endemism of Malagasy lemuriformes. *Divers Distrib* 19:1–12
- Huang S et al (2012) Traits, trees and taxa: global dimensions of biodiversity in mammals. *Proc R Soc B* 279:4997–5003
- Isaac NJB et al (2007) Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PLoS One* 2:e296
- IUCN (1980) World conservation strategy: living resource conservation for sustainable development. International Union for Conservation of Nature and Natural Resources (IUCN), Gland, Switzerland.
- IUCN Red List Standards and Petitions Subcommittee (2006) Guidelines for using the IUCN red list categories and criteria. IUCN, Gland, Switzerland. Available from <http://www.iucn.org/webfiles/doc/SSC/RedList/RedListGuidelines.pdf>
- IUCN (2012) IUCN red list of threatened species. Available: <http://www.redlist.org>
- Jetz W et al (2014) Global distribution and conservation of evolutionary distinctness in birds. *Curr Biol* 24:919–930

- Jono CMA, Pavoine S (2012) Threat diversity will erode mammalian phylogenetic diversity in the near future. *PLoS One* 7(9), e46235. doi:[10.1371/journal.pone.0046235](https://doi.org/10.1371/journal.pone.0046235)
- Kelly S, Grenyer R, Scotland RW (2014) Phylogenetic trees do not reliably predict feature diversity. *Divers Distrib* 20:600–612
- Kuntner M, May-Collado LJ, Agnarsson I (2011) Phylogeny and conservation priorities of afrotherian mammals (Afrotheria, Mammalia). *Zool Scr* 40:1–15
- Larsen FW, Turner WR, Brooks TM et al (2012) Conserving critical sites for biodiversity provides disproportionate benefits to people. *PLoS One* 7, e36971
- López-Orsorio F, Miranda-Esquivel DR (2010) A phylogenetic approach to conserving Amazonian biodiversity. *Conserv Biol* 24:1359–1366
- Lozupone C, Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* 71:8228–8235
- Maclaurin J, Sterelny K (2008) What is biodiversity? The University of Chicago Press, Chicago
- May-Collado LJ, Agnarsson I (2011) Phylogenetic analysis of conservation priorities for aquatic mammals and their terrestrial relatives, with a comparison of methods. *PLoS One* 6(7), e22562. doi:[10.1371/journal.pone.0022562](https://doi.org/10.1371/journal.pone.0022562)
- McNeely JA (1988) Economics and biological diversity: developing and using economic incentives to conserve biological resources. IUCN, Gland, Switzerland, 232 pp
- McNeely JA, Miller KR, Reid WV, Mittermeier RA, Werner TB (1990) Conserving the World's biological diversity. International Union for Conservation of Nature and Natural Resources/World Resources Institute/Conservation International/World Wildlife Fund/US World Bank, Gland
- McNeely JA et al (2005) Biodiversity. In: Chopra K et al (eds) *Ecosystems and human well-being*, vol 3, Policy responses. Millennium Ecosystem Assessment, Island Press, Washington, DC, pp 119–172
- Millennium Ecosystem Assessment (2005) *Ecosystems and human well-being: biodiversity synthesis*. World Resources Institute, Washington, DC
- Mooers AO, Atkins R (2003) Indonesia's threatened birds: over 500 million years of evolutionary heritage at risk. *Anim Conserv* 6:183–188
- Mooers AO et al (2008) Converting endangered species categories to probabilities of extinction for phylogenetic conservation prioritization. *PLoS One* 3, e3700. doi:[10.1371/journal.pone.0003700](https://doi.org/10.1371/journal.pone.0003700)
- Morgan GJ (2010) Evaluating Maclaurin and Sterelny's conception of biodiversity in cases of frequent, promiscuous lateral gene transfer. *Biol Philos* 25:603–621
- Pacharawongsakda E, Yokwai S, Ingsriswang S (2009) Potential natural product discovery from microbes through a diversity-guided computational framework. *Appl Microbiol Biotechnol* 82:579–586
- Perez-Losada M, Jara CG, Bond-Buckup G, Crandall KA (2002) Conservation phylogenetics of Chilean freshwater crabs *Aegla* (Anomura, Aeglidae): assigning priorities for aquatic habitat protection. *Biol Conserv* 105:345–353
- Pio DV et al (2014) Climate change effects on animal and plant phylogenetic diversity in southern Africa. *Glob Chang Biol* 20:1538–1549
- Posadas P, Miranda Esquivel DR, Crisci JV (2001) Using phylogenetic diversity measures to set priorities in conservation: an example from Southern South America. *Conserv Biol* 15:1325–1334
- Reid WV, Miller K (1989) Keeping options alive: the scientific basis for conserving biodiversity. World Resources Institute, Washington, DC, 1989 – Science – 128 pp
- Rosauer D, Laffan SW, Crisp MD, Donnellan SC, Cool LG (2009) Phylogenetic endemism: a new approach for identifying geographical concentrations of evolutionary history. *Mol Ecol* 18:4061–4072
- Safi K, Armour-Marshall K, Baillie JEM, Isaac NJB (2013) Global patterns of evolutionary distinct and globally endangered amphibians and mammals. *PLoS One* 8(5), e63582
- Sarkar S (2008) From ecological diversity to biodiversity. In: Hull D, Ruse M (eds) *The Cambridge companion to the philosophy of biology*. Cambridge University Press, Cambridge

- Saslis-Lagoudakisa CH, Savolainen V, Williamson EM, Forest F, Wagstaff SJ, Baral SR, Watson MF, Pendry CA et al (2012) Phylogenies reveal predictive power of traditional medicine in bioprospecting. *Proc Natl Acad Sci* 109:15835–15840
- Soulé ME (1980) *Conservation biology: an evolutionary-ecological perspective*. Sinauer Associates, Sunderland, 395 pp
- Smith WL, Wheeler WC (2006) Venom evolution widespread in fishes: a phylogenetic road map for the bioprospecting of piscine venoms. *J Hered* 97:206–217
- Tucker CM, Cadotte MW, Davies TJ, Rebelo AG (2012) The distribution of biodiversity: linking richness to geographical and evolutionary rarity in a biodiversity hotspot. *Conserv Biol* 26:593–601
- UNEP (2013) Report of the second session of the Plenary of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. IPBES/2/17, Nairobi, Kenya.
- Vamosi JC, Wilson JR (2008) Nonrandom extinction leads to elevated loss of angiosperm evolutionary history. *Ecol Lett* 11:1047–1053
- Vane-Wright RI, Humphries CJ, Williams PH (1991) What to protect—systematics and the agony of choice. *Biol Conserv* 55:235–254
- Weikard H (2002) Diversity functions and the value of biodiversity. *Land Econ* 78:20–27
- Weitzman ML (1992) On diversity. *Q J Econ* 107:363–405
- Wilson EO (1992) *The diversity of life*. Norton, New York
- Winter M, Devictor V, Schweiger O (2013) Phylogenetic diversity and nature conservation: where are we? *Trends Ecol Evol* 28:199–204
- Witting L, Loeschcke V (1995) The optimization of biodiversity conservation. *Biol Conserv* 71:205–207

# Reconsidering the Loss of Evolutionary History: How Does Non-random Extinction Prune the Tree-of-Life?

Kowiyou Yessoufou and T. Jonathan Davies

**Abstract** Analysing extinction within a phylogenetic framework may seem counter-intuitive because extinction is *a priori* a non-heritable trait. However, extinction risk is correlated with other traits, such as body size, that show a strong phylogenetic signal. Further, there has been much effort in identifying key traits important for diversification, and recent evidence has demonstrated that the processes of speciation and extinction may be inextricably linked. A phylogenetic approach also allows us to quantify the impact of extinction, for example, as the loss of branches from the tree-of-life. Early work suggested that extinctions might result in little loss of evolutionary history, but subsequent studies indicated that non-random extinctions might prune more of the evolutionary tree. Loss of phylogenetic diversity might have ecosystem consequences because functional differences between species tend to be correlated with the evolutionary distances between them. Here we explore how extinction prunes the tree-of-life. Our review indicates that the loss of evolutionary history under non-random extinction (the emerging pattern in extinction biology) might be less pronounced than some previous studies have suggested. However, the loss of functional diversity might still be large, depending on the evolutionary model of trait change. Under a punctuated model of evolution, in which trait differences accrue in bursts at speciation, the number of branches lost is more important than their summed lengths. We suggest that evolutionary models need to be incorporated more explicitly into measures of phylogenetic diversity if we are to use phylogeny as a proxy for functional diversity.

**Keywords** Extinction risk • Phylogenetic comparative methods • Punctuated evolution • Phylogenetic diversity • Feature diversity

---

K. Yessoufou

Department of Environmental Sciences, University of South Africa,  
Florida, Gauteng, South Africa  
e-mail: [kowiyouyessoufou1@gmail.com](mailto:kowiyouyessoufou1@gmail.com)

T.J. Davies (✉)

Department of Biology, McGill University, Montréal, Québec, Canada  
e-mail: [j.davies@mcgill.ca](mailto:j.davies@mcgill.ca)



## Introduction

There is mounting evidence that we are entering a sixth mass extinction (Millennium Ecosystem Assessment 2005), and the future of biodiversity is at risk due to the high rates at which biological diversity – species, habitats, evolutionary diversity – is being eroded. Species are experiencing unprecedented pressures across their ranges owing to global change, including increased invasion success of aliens (Winter et al. 2009), habitat destruction (Vitousek et al. 1997; Haberl et al. 2007), climate change and climate variability (Willis et al. 2008, 2010). Consequently, approximately 30 % of assessed species are currently categorised as threatened by the IUCN, and a greater proportion may be committed to extinction in the near future (Thomas et al. 2004). Current rates of species loss might be 1,000–10,000 times greater than past extinction rates (Pimm et al. 1995; Millennium Ecosystem Assessment 2005) with particularly elevated rates in tropical biomes (Vamosi and Vamosi 2008), known for their unique life-form diversity. At the ecosystem level, with the loss of species, we also lose their contributions to overall ecosystem functioning and services. The loss of ecosystem services is of particular concern because human survival relies strongly on key services such as food production, plant pollination, medicinal plants, clean water, clean air, nutrient cycling, carbon sequestration, climate stability, recreation, tourism, etc. – which are provided by a well functioning system of biological diversity.

It is well established that human activities can drive extinctions within a short period of time (Baillie et al. 2004; Mace et al. 2005a). Because human population has increased exponentially over the last centuries, and is expected to reach nine billion by 2050 ([www.un.org/esa/population/publications/longrange2/2004worldpop2300reportfinalc.pdf](http://www.un.org/esa/population/publications/longrange2/2004worldpop2300reportfinalc.pdf)), pressure on natural ecosystems is also predicted to increase, yet at the same time there will be an even greater demand for the ecosystem services provided by biologically diverse natural systems. As a result, the rate of species extinction is projected to rise by at least a further order of magnitude over the next few hundred years (Mace et al. 2005b), potentially decreasing the provisioning of ecosystem services at a time when demand is growing. Understanding how the ongoing extinction crisis will impact the provisioning of critical ecosystem services is therefore a matter of urgency.

Quantifying the ecosystem contributions of individual species is a major challenge. Current estimates of global diversity vary by over an order of magnitude (see e.g. May 2010), with the vast majority of species (86 % and 91 % of terrestrial and oceanic diversity, respectively) remaining unknown to science (Mora et al. 2011). An in-depth understanding of species ecologies is therefore impractical for most of life; at best, we might be able to infer their placement on the tree-of-life. Whilst there is now a general consensus on the positive link between biodiversity and ecosystem function (Hooper et al. 2012), there has been growing evidence suggesting that evolutionary history provides a more informative measure of biological diversity than traditional metrics based upon richness and abundance (e.g. Faith 1992; Faith et al. 2010; Davies and Cadotte 2011; see also Srivastava et al. 2012 for a comprehensive review). It is suggested that evolutionary history might better capture

functional diversity including unmeasured or hard-to-measure traits (Crozier 1997; Faith 2002). As such, phylogeny provides a unique framework that captures both known (Forest et al. 2007; Saslis-Lagoudakis et al. 2011) and unknown ecosystem services (Faith et al. 2010). Understanding how the current extinction crisis will prune the tree-of-life is therefore critical for ensuring a continued provisioning of the ecosystem services upon which we rely, but for which we might lack detailed ecological knowledge of underlying process or mechanism (Faith et al. 2010).

There has been growing effort to incorporate species evolutionary histories into conservation decision-making (e.g. Purvis et al. 2000a, 2005; Isaac et al. 2007, 2012; Faith 2008). This effort has been facilitated by the rapid rise in analytical tools, and the availability of large comprehensive phylogenetic trees for well studied taxonomic groups such as mammals (Bininda-Emonds et al. 2007), birds (McCormack et al. 2013), amphibians (Pyron and Wiens 2013), and flowering plants (e.g. Davies et al. 2004). Here, we review recent insights from phylogenetic studies of extinction risk, and re-examine how extinctions impact the tree-of-life.

## Speciation and Extinction as Two Natural Processes

Extant species represent just a small fraction of all the species that have ever lived (Jablonski 1995; May et al. 1995; Niklas 1997). This standing biodiversity is the net difference between cumulative speciation and extinction over the evolutionary history of life on Earth. Both the processes of speciation and extinction are therefore intrinsic parts of Earth's natural history. Much effort has gone into exploring geographic and taxonomic patterns of diversity, looking to answer why some regions and some taxa are more species-rich than others. Recent debate has contrasted explanations based upon ecological limits and times for speciation (e.g. see Rabosky and Lovette 2008). Comparisons between sister taxa, which are by definition of equal age, allow us to control for time for speciation, and thus differences in richness must reflect either variation in speciation or extinction rates (Barraclough et al. 1998). Such comparisons have shown that diversification rates have been higher in more tropical lineages (Davies et al. 2004; Rolland et al. 2014), but that higher tropical species richness is most likely a product of both faster rates and longer times for speciation (Jansson and Davies 2008). However, high diversification might be explained by high speciation rates, low extinction rates or a combination of both, and until recently, it has not been possible to reliably disentangle the two.

Unraveling the processes of extinction and speciation remains a major challenge (Benton and Emerson 2007). The fossil record is often thought to provide the most reliable documentation of speciation and extinction, yet the cumulative fossil record suggests that speciation rate increases inexorably through time (Raup 1991; Nee 2006; Benton and Emerson 2007), whereas there is growing evidence suggesting that species accumulate in bursts, and speciation rates decline over time (Simpson 1953; Schluter 2000; Gavrilets and Vose 2005; Scantlebury 2013). Phylogeny provides an alternative tool for reconstructing evolutionary process (Harvey et al.

1994). Nee et al. (1994) illustrated how extinction rates could be estimated from phylogenetic trees (but see Rabosky 2010), but assumed constant rates model. New methods, for example, BiSSE (Maddison et al. 2007) and GeoSSE (Goldberg et al. 2011), relax this assumption, and allow us to estimate extinction and speciation rates simultaneously, for example, with the gain or loss of particular character states (BiSSE) or shifts in geographic distributions (GeoSSE). Phylogeny-based analysis of diversification provides some limited evidence for increasing speciation through time (e.g. Barraclough and Vogler 2002; Linder et al. 2003; Turgeon et al. 2005), but again, a scenario of rapid radiation followed by a decline in speciation rate over time appears to be more common (Harmon et al. 2003; Shaw et al. 2003; Kadereit et al. 2004; Machordom and Macpherson 2004; Morrison et al. 2004; Williams and Reid 2004; Xiang et al. 2005; Kozak et al. 2006; Weir 2006; Phillimore and Price 2008; Scantlebury 2013). This pattern could be linked to a density-dependent model of ecological opportunity and/or reflect punctual mass extinctions (e.g. Yessoufou et al. 2014) that open up new niche space for subsequent radiations (Crisp and Crook 2009). Recently, using phylogenetic information on the Cape Floristic Region, Davies et al. (2011) suggested that the processes of speciation and extinction may be inextricably linked.

Speciation and extinction are part of life's natural history, and to achieve equilibrium in standing diversity, speciation must equal extinction (Raup 1986). Even the classic MacArthur and Wilson (1963, 1967) model of island biogeography suggests that species richness is a dynamic equilibrium between immigration, speciation and extinction. However, today this balance is increasingly biased towards extinction (Millennium Ecosystem Assessment 2005), and we risk moving towards a new low-diversity state as it is not possible to manipulate speciation rates to match current losses (Barraclough and Davies 2005). Whilst there is increasing evidence that evolutionary processes can occur over ecological time scales (Kettlewell 1972; Endler 1986; Kinnison and Hendry 2001; Ashley et al. 2003), speciation can take a longer time to complete, whereas extinctions are occurring over much shorter time spans (Barraclough and Davies 2005). Even for the most famous examples of rapid speciation, such as Lake Victoria cichlids, diversification rates are estimated over 100's to 1000's of years, and evidence of 'reverse-speciation' indicates that speciation might not have been complete (Seehausen 2006). By contrast, rates of extinction are now estimated at many times background rates (Vitousek et al. 1997; Butchart et al. 2004), and are occurring over 10's to 100's of years.

## Shifting the Balance Towards a Low-Diversity Earth

### *Extinction Trends*

Whilst the scale of current species loss parallels that of mass extinction events in the paleontological past (May et al. 1995; Millennium Ecosystem Assessment 2005), unlike past extinctions which were caused by abiotic factors such as asteroid strikes,

volcanic eruptions, and natural climate shifts, the current crisis is driven largely by human activities, and is perhaps the first mass extinction event that can be attributed to a biotic cause. Current estimates indicate that 10–30 % of mammals, amphibians and birds are threatened with extinction (Millennium Ecosystem Assessment 2005). Taxonomic groups are not, however, equally at risk of extinction. Among terrestrial vertebrates, amphibians have the highest proportion at at-risk species, with at least a third of ~6600 known amphibians threatened with extinction (Wake and Vredenburg 2008). It is estimated that 12 % and 20 % of continental birds and mammals, respectively, have already been lost (Wilson 1992), but with a higher rate of loss observed on islands (Lohle and Eschenbach 2011). In fish, of the ~2,000 species that have been assessed 21 % are considered at risk of extinction (IUCN 2010). Our knowledge of extinction risks in invertebrates is much poorer; however, of the 1.3 million known invertebrates, less than 10,000 species have been assessed, of which 30 % are threatened (IUCN 2010).

In plants, extinction trends appear to be even more alarming, but estimates need to be interpreted carefully. For example, over 70 % of Red-listed species of flowering plants are classified as at risk of extinction (category VU or higher) (IUCN 2010). This proportion is much higher than that reported for vertebrate groups (22 %), but as yet only a very small fraction of total plant diversity has been assessed (~13,000 of >300,000 species), and a trend towards focusing on some of the most obviously vulnerable species might bias our estimates of threat upwards. For clades with more complete sampling, such as cycads, the proportion of threatened species remains high (>80 %), but perhaps this ancient group that peaked in diversity in the Jurassic–Cretaceous (Jones 2002; Taylor et al. 2009) when dinosaurs roamed the Earth, is not representative of current seed plant diversity. One recent attempt to estimate the true proportion of threatened species within angiosperms using a statistical model to correct for sampling bias – the sampled Red List – has suggested that the percent of at-risk plant species might actually be more comparable to that for mammals (<http://threatenedplants.myspecies.info/>).

The spatial congruence in taxonomic richness across taxonomic groups has been well described globally (Grenyer et al. 2006), with the richest areas of the world found in highly productive environments at low latitudes and in mountainous regions (Orme et al. 2005). Similarly, there is a geographical pattern in the distribution of rare and threatened taxa, which has been shown at the global scale for vertebrates (e.g. Grenyer et al. 2006), and at various scales for plants (e.g. Zhang and Ma 2008; Davies et al. 2011; Daru et al. 2013). However, hotspots of richness and rarity or threat do not necessarily coincide (Grenyer et al. 2006). For example, vertebrate richness peaks on the Neotropical mainland, but bird rarity concentrates on oceanic island archipelagos, the diversity of rare mammal species peaks on continental shelf islands and rare amphibian species are more centered on continental landmasses (Grenyer et al. 2006). The variation in geographical patterns of rarity may be partially linked to differences in relative dispersal ability across taxa. Spatial variation in extinction risk additionally reflects differences in the distribution of threats facing each group. For example, invasive species and overexploitation are key threats for birds whereas overexploitation is the major driver of species loss in mammals

(Baillie et al. 2004), and climate change, pollution and transmissible diseases are important in amphibians (Stuart et al. 2004).

## ***Extinction Drivers: Animals Versus Plants***

### **Extrinsic Versus Intrinsic Factors**

Explaining why some species appear predisposed to higher extinction risk than others is an important goal for conservation research (McKinney 1997). The five main extinction drivers include habitat loss, climate change, increased pollution, resources over-exploitation and invasive species (Millennium Ecosystem Assessment 2005), and all are linked directly or indirectly to anthropogenic pressures. These drivers parallel Jarred Diamond's 'evil quartet' (Diamond 1984, 1989), but with the more recent addition of climate change, and Diamond additionally included the possibility of extinction cascades in which secondary extinctions follow the loss of key species, for example, due to the disruption of ecosystem processes. We can further simplify this list into extrinsic (e.g. climate change) and intrinsic factors (e.g. ecological traits such as population density and species life-history traits such as body size and gestation length) (Cardillo et al. 2005). Extrinsic factors might help explain geographic variation in extinction risk, whereas intrinsic factors might better explain taxonomic patterns; however, highest risk is where driver intensity associated with extrinsic factors overlaps with species intrinsic vulnerability. In addition, species are increasingly likely to be exposed to multiple drivers, and this will likely further exacerbate risk of extinction (Brook et al. 2008).

### **Extinction Drivers in Animals**

Correlates of extinction risk in animal kingdom have been explored extensively using data from the IUCN Red List (Bennett and Owens 1997; Russell et al. 1998; Purvis et al. 2000a, b; Cardillo 2003; Cooper et al. 2008) with particular attention to mammals (Russell et al. 1998; Cardillo et al. 2005, 2008; Isaac et al. 2007; Huang et al. 2012), perhaps the best-studied higher taxonomic group. Across studies, high extinction risk is generally associated with large body size, long generation times and small geographic range sizes (Bennett and Owens 1997; Russell et al. 1998; Purvis et al. 2000a; Cardillo 2003; Fisher and Owens 2004; Cooper et al. 2008). Conversely, species at low risk of extinction are small, reproduce rapidly, and have a wide niche breadth.

We know, for example, that mammals that are at risk of extinction are, on average, an order of magnitude heavier than non-threatened species (IUCN 2003). The size-selectivity of extinction risk is not unique to the current extinction crisis; past mass extinction events, such as that of the late Pleistocene, were also biased towards larger species (Martin 1967; Johnson 2002). During the late-Pleistocene – early-

Holocene extinction event, there was a mass extinction of much of the mammalian megafauna, resulting in a loss of several complete ecological guilds and their predators (Cione et al. 2003). Size selectivity in extinction risk has been long-recognised (e.g. Pimm 1991; Lawton 1995; Pimm et al. 1988; Cardillo and Bromham 2001; Johnson 2002), and there are many potential explanations. Large-sized mammals might be more extinction-prone because of generally lower average population densities (Damuth 1981), putting them at greater risk from stochastic population dynamics. High risk in large bodied mammals might also reflect the negative correlation between intrinsic rates of population increase and body mass (Fenschel 1974), and thus longer recovery times following population declines. There might also be an increased propensity for humans to exploit larger species (Bodmer et al. 1997; Jerzolimski and Peres 2003). The relationship between species traits and extinction risk, however, is not straightforward, because of the complex interaction between intrinsic and extrinsic drivers, and different clades might have very different predictors (e.g. see Cardillo et al. 2008).

Cardillo and colleagues (2005) demonstrated that risk in small-sized mammals (<3 kg) was largely determined by extrinsic factors including the size and location of geographical ranges. However, the predisposing factors in the larger size class include both intrinsic species properties (e.g. population density, neonatal mass and litters per year) and extrinsic factors. Such fine-scaled analyses can help address whether extinctions are linked to 'bad genes' or 'bad luck' (Raup 1993; Bennett and Owens 1997). For mammals, it appears that extinction in small bodied species is more likely a case of bad luck, driven by extrinsic factors. For larger bodied species, bad genes, that is, genes controlling intrinsic traits such as body size and life history are additional aggravating factors promoting extinctions.

Compared to vertebrates, the distribution and drivers of extinction risk in invertebrate communities has been poorly explored. However, a recent study estimated that one-fifth of invertebrate species may be threatened with extinction, with freshwater species at particular high risk (Collen et al. 2012). Collen and colleagues suggested that the greater threat to freshwater species was predominantly driven by agricultural pollution and dam construction, invasive species and waterborne diseases. More generally, and perhaps unsurprisingly, species that are less mobile and with limited geographic ranges, such as freshwater mollusks, tend to be at higher risk (Collen et al. 2012). In marine ecosystems, however, the market values of some invertebrates correlate strongly with their risk of extinction, e.g. invertebrate species considered luxury seafood (Purcell et al. 2014), providing an exception to the general trend for greater threat to be observed in larger-sized species.

The phylogenetic distribution of extinction risk in mammals has also been of much interest. In mammals, it has been suggested that species subtending from longer phylogenetic branches, and thus representing greater unique evolutionary history, are at higher risk of extinction (Russell et al. 1998; Purvis et al. 2000a). This pattern matches to Wilson's (1961) 'taxon cycle', which predicts that older species would have higher extinction probabilities as species expand and contract in their geographical distributions over their evolutionary lifetimes. Although, as originally described, the taxon cycle referred to the distribution of species on islands (ants on

islands in Melanesia), the concept has been extended to include species on continents (e.g. see Ricklefs and Bermingham 2002). Alternatively, it might simply echo the pattern of historical extinctions, in which older species represent survivors of once more diverse clades (Purvis et al. 2000a). However, the precise relationship between extinction risk and evolutionary age remains debated (Verde et al. 2013). Further, patterns of extinction risk in plants appears to show an opposite trend, with higher risk associated with young species in species rich (Schwartz and Simberloff 2001; Meijaard et al. 2007) and more rapidly diversifying clades (Davies et al. 2011), suggesting that predictors of extinction in plants might be very different to those for mammals.

### Extinction Drivers in Plants

Species extinction in the plant kingdom is predominantly the result of habitat loss, for example through deforestation. Tropical forests, which cover less than 7 % of the world land area, contain over 50 % of global biodiversity (Dirzo and Raven 2003), but these unique habitats are being destroyed at unprecedented rates (Laurance 1991; Achard et al. 2002) as a result of rapid human population growth and economic development. In tropical Asia and Africa, over 40 % of the primary forests is already lost (Wright 2005). This drastic reduction in forest cover has had a devastating impact on plant diversity (Sodhi and Brook 2006). Although there is some evidence that, globally, recent rates of deforestation are slowing, we likely owe a large extinction debt due to the time lag between habitat loss and species losses predicted from the reduction in area. Thus, even should we be successful in preserving the remaining forest cover, many species might still be predicted to be lost over the following decades as habitats return to a new, lower diversity, equilibrium state. This extinction event will likely be exacerbated by the effects of ongoing climate change as local climate conditions shift and species are forced to either adapt to new conditions or track climate space (Willis et al. 2008).

Plant responses to environmental change are difficult to predict. With warming, plants might adapt by shifting their phenologies – the timing of life history strategies – for example flowering earlier and losing leaves later (Parmesan 2007). Recent work indicates significant phylogenetic conservatism in flowering phenology (Davies et al. 2013), suggesting that there might be some evolutionary constraints to species adaptive responses. If the velocity of climate change is high, species may not have the necessary time to adjust their phenological responses. Alternatively, species might track suitable climates, for example by shifting their distribution northwards or towards higher elevations (Sandel et al. 2011). Species already restricted to high elevation biomes might then be particularly vulnerable as increased warming may result in the reduction of suitable habitat and, at the extremes, complete habitat loss. In the biodiversity hotspots of the Eastern Arc Mountains of East



Africa, species at higher elevations already tend to be more threatened, perhaps reflecting recent climate shifts (Yessoufou et al. 2012). Species that are unable to adapt their phenology or track climate through space will be most vulnerable to extinction. In data from Thoreau's woods in Concord, MA, spanning 100 years, it is already possible to detect declines in populations among species that have failed to shift their phenologies to match recent climate change (Willis et al. 2008, 2010). These data also revealed phylogenetic structure in species responses, suggesting evolutionary conservatism not only in flowering times, but also plasticity in flowering times (see also Davies et al. 2013).

As for animals, there has been much work aimed at identifying intrinsic life-history traits that predispose some plant species towards extinction (Sodhi et al. 2008). However, investigating the correlates of extinction risk within the plant kingdom has proven somewhat more challenging, as key traits frequently differ between studies (Walker and Preston 2006; Sodhi et al. 2008). In addition, traits explain only a small proportion of the variation in extinction risk and, with the exception of geographic range size, we have yet to reveal any single strong correlate equivalent to, for example, body size in mammals. Life-history traits that have been found to correlate with plant extinction include pollination syndrome (e.g. wind or animal mediated), sexual system, habit, height, and dispersal mode (Sodhi et al. 2008). For tropical angiosperms, these traits can explain ~10 % of extinction risk (Sodhi et al. 2008), whereas equivalent models of intrinsic drivers for mammals can explain up to 30 % of the variation in extinction risk (Cardillo et al. 2008). However, even for mammals, explanatory power tends to be lower when exploring predictors across disparate clades (Cardillo et al. 2008), reflecting clade specific sensitivities to different drivers. Perhaps, therefore, it is unsurprising that in flowering plants, a group containing up to 500,000 species, predictive models are often poor.

An alternative avenue of exploration has considered the importance of evolutionary history in models of extinction risk (Sodhi et al. 2008; Davies et al. 2011). In plants, there is increasing evidence that a species evolutionary history might be more important than its life history in explaining extinction risk. As mentioned above, threatened terrestrial plants generally fall within species-rich clades (Schwartz and Simberloff 2001; Pilgrim et al. 2004) that represent recent radiations (Davies et al. 2011). However, when we look at the distribution of extinction risks across plant families, species-poor and especially monotypic families also appear to contain species at higher risk of extinction (Vamosi and Wilson 2008). It is therefore possible that mechanistic explanations for variation in extinction risk differ between old and young clades. Old and species-poor families may represent remnants of once more diverse clades, with species vulnerabilities associated with intrinsic life history traits and long generation times, as in mammals. In contrast, extinction risk in younger, still diversifying clades, may be more closely linked to the speciation process, with high extinction risk more closely associated with traits driving speciation, such as small geographic range size and short generation times.



## *The Importance of Phylogeny in Conservation*

### **Why We Need to Evaluate Extinction Risk within a Phylogenetic Framework**

Phylogenetic approaches are now well accepted in many ecological disciplines. Phylogenetic methods are also increasingly commonplace in extinction biology (see Purvis 2008). The necessity of employing a phylogenetic framework for exploring a non-evolving trait such as risk of extinction has been questioned (Grandcolas et al. 2011). Reasons for doing so are multifold. First, as we have discussed above, many drivers of extinction risks can be linked to phylogenetically conserved traits, such as body mass (Cardillo et al. 2005, 2008) and phenology (Willis et al. 2008, 2010). Therefore, phylogenetic comparative methods, such as independent contrasts (Felsenstein 1985) or phylogenetic regression are important because species cannot be considered as statistically independent (see Purvis 2008 for further discussion). Second, species evolutionary history might itself be an important predictor of extinction risk, for example, with higher risks associated with either more evolutionarily distinct lineages (Purvis et al. 2000a; Mace et al. 2003) or centres of diversification (Davies et al. 2011), depending on the clade and taxonomic scale. Third, by considering extinction within a phylogenetic framework, we can quantify directly its impacts on the tree-of-life as the loss of phylogenetic diversity (PD) (Purvis et al. 2000a; Mace et al. 2003). This measure of evolutionary heritage provides a useful conservation metric, typically measured in millions of years, it is easily comprehensible, and simple to calculate for particular regions or taxa (Mooers et al. 2005). Although, there remain practical obstacles in the implementation of phylogenetic approaches for conservation planning, there is now increasing appreciation of the importance of including an evolutionary perspective within conservation goals, as illustrated by the Zoological Society of London's EDGE of existence programme (<http://www.edgeofexistence.org/>) that emphasises the conservation of evolutionary distinct and threatened species (Isaac et al. 2007).

### **Practical Contribution of Phylogeny to Conservation**

The practical contribution of phylogeny to conservation actions has recently been discussed (Cardillo and Meijaard 2012; Winter et al. 2013). In part, the conservation value of the phylogenetic approach is in its ability to guide pre-emptive actions towards identifying and prioritizing the most at-risk species. For example, by identifying species with traits or in regions that predispose them to high risk of extinction, we can identify species that are not yet at risk of extinction but which might become threatened in the near future if current extinction drivers increase in intensity or geographic extent. Cardillo et al. (2006) referred to such species as having high 'latent risk' of extinction. Given limited conservation funding, focusing efforts on species with high latent risk might make economic sense as it is likely to be more

cost effective to prevent species declines before they begin versus reestablishing viable populations for species that have already suffered declines and may have lost much of their natural range. Preserving intact habitats will almost always be easier and cheaper than returning transformed habitats to their natural states.

A justification for placing emphasis on the preservation of phylogenetic diversity per se is that phylogenetic diversity captures feature diversity (Faith 1992; Crozier 1997; see also section “[Feature diversity and evolutionary models of character change](#)”), and thus preserving the set of species that maximizes phylogenetic diversity also maximizes the possibility of having the right set of features in an uncertain future. Forest et al. (2007) provided an example of the utility of phylogenetic diversity in the Cape Floristic Region of South Africa by demonstrating that preserving the phylogenetic diversity of the flora would maximize future options for the benefit of society through a continued provisioning of key ecosystem services. To date, empirical examples of conservation actions implemented explicitly to protect phylogenetic diversity are rare; however, one recent effort spearheaded by the Zoological Society of London’s EDGE programme specifically aims at focusing conservation attention on evolutionary distinct species at risk of extinction. These EDGE species are distinct not only in the history of their evolutionary past, but perhaps also in the functional roles they might fill within ecosystems. The extinction of EDGE species might therefore result in the loss of important ecosystem functions and services for which we have no species substitute. Some EDGE species (e.g. elephants and pandas) are well known, but many others (e.g. Chinese giant salamanders and the peculiar long-beaked echidnas) have been overlooked by traditional conservation strategies (see Isaac et al. 2007, 2012).

Critically, the utility of phylogenetic metrics and methods in conservation biology relies upon the accuracy of the underlying phylogenetic topology and, if we are interested in capturing feature diversity, the evolutionary model of character change along the branches of the tree, a point we explore further in the following sections.

## Extinction and the Loss of Evolutionary History

### *Phylogenetic Structure in Extinction Risks*

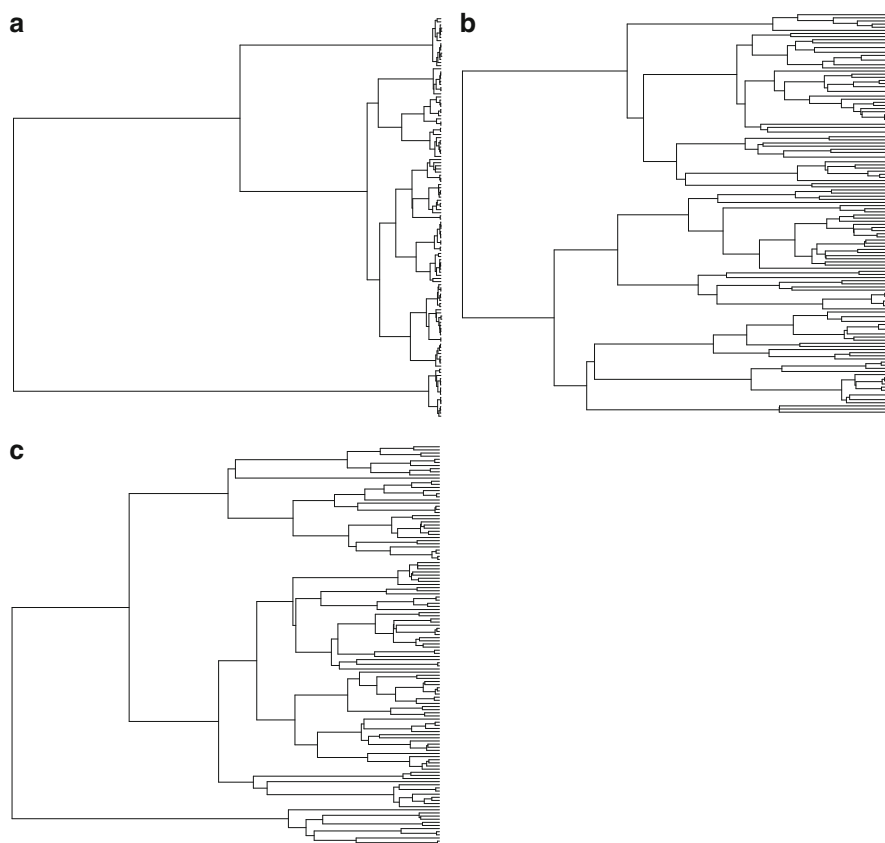
We have discussed above how the process of extinction is non-random with respect to species traits and geography. For example, extinction will tend to remove large-bodied species with slow life histories and narrow niches, and species in regions with high intensity of extinction drivers. Because many of the traits linked to extinction risk (e.g. body size, generation time, dispersal ability etc.) demonstrate phylogenetic conservatism (Fritz and Purvis 2010), such that they tend to be clustered on the phylogeny, extinctions will also tend to cluster on the phylogeny. Whereas evidence for trait-based explanations for plant extinctions is mixed (Freville et al. 2007; Bradshaw et al. 2008; Sodhi et al. 2008; Davies et al. 2011; Daru et al. 2013),

phylogenetic selectivity in extinction risk might also result from a geographical pattern in the drivers of extinction, for example, range elevation might determine a species vulnerability to climate change (Sandel et al. 2011). If closely related species also tend to have close geographical proximities, perhaps reflecting shared habitat preferences or the geographical process of speciation, they will then also be exposed to similar intensity of extinction drivers. There is an increasing weight of evidence suggesting that extinction risk is generally more clustered on a phylogeny than expected by chance (Bennett and Owens 1997; Purvis et al. 2000a; Schwartz and Simberloff 2001), a pattern also observed within the fossil record. Extinction will thus prune the tree-of-life non-randomly. However, how this non-random pruning might impact the loss of evolutionary history has been a subject of recent debate.

### *Quantifying the Loss of Evolutionary History*

Extinction prunes species from the tips of the tree-of-life, resulting in the loss of terminal branches. In a frequently cited paper, Nee and May (1997) used simulations to explore the expected loss of evolutionary history (quantified as the summed branch lengths from the tree-of-life) under various extinction intensities. Perhaps surprisingly, they found that up to 80 % of the tree would remain under even extreme extinction scenarios in which 95 % of species were lost. However, their simulations were unrealistic in two regards. First, they assumed extinction events were random – the field-of-bullets model, in which extinction is independent of species' traits and thus also phylogeny. If extinctions are clustered on a phylogeny, we might also lose the internal branches of the tree that connect them, and thus experience a greater overall loss of phylogenetic diversity (Russell et al. 1998; Purvis et al. 2000a). Second, their expectation was derived assuming a phylogeny based on a coalescent model, which generates a highly unrealistic distribution of branching times, with most branches clustered towards the present (see Fig. 1a), and does not fit to most empirical estimates of phylogenies. Importantly, coalescent trees tend to be 'tip-heavy' such that most branching events are short and clustered towards the present (tips of the tree). Therefore, under this model, most extinctions remove only short terminal branches from the tree, and most major lineages survive even extreme pruning of tips. Empirical phylogenies tend to have a very different distribution of branching times (e.g. Rabosky and Lovette 2008; see also Fig. 1b, c for pure birth and birth-death tree). Mooers et al. (2012) explore further how tree shape impacts the expected loss of phylogenetic diversity. The phylogenetic non-random distribution of extinction risk and the shape of empirical phylogenies might therefore suggest that we risk losing a disproportionate amount of evolutionary history from the tree-of-life.

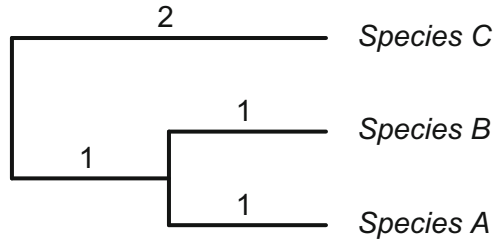
A suite of empirical studies were to follow on from the early work of Nee and May, and emphasized both the phylogenetically non-random nature of species' extinctions and a greater than random loss of phylogenetic diversity (e.g. Purvis et al. 2000a; Purvis 2008; Vamosi and Wilson 2008). A link between non-random



**Fig. 1** Comparison of branching times for different tree reconstruction models of size 128 tips. **a** Coalescent model in which branching clusters towards present; **b** pure birth model in which all lineages have an equal probability of splitting ( $b=1.0$ ) and no lineages go extinct ( $d=0$ ); **c** birth-death model in which lineages have equal rates of splitting and extinction (birth=1.0, death=0.2)

extinction and greater than random loss of phylogenetic diversity seemed intuitive; if two sister species are lost to extinction, not only do we lose the unique phylogenetic diversity captured in the branches from which they subtend, but we also lose the ancestral branch that is shared between them (see Fig. 2). However, in a more recent study, again using simulations, but this time assuming both a more realistic model of diversification and a range of phylogenetic signal in extinction probabilities, Parhar and Mooers (2011) suggested that the loss of phylogenetic diversity under phylogenetically non-random extinctions was more or less indistinguishable from random (see also Heard and Mooers 2000). Seemingly, the observation of phylogenetic signal in extinction risks and the non-random loss of phylogenetic diversity are not necessarily connected directly.

Observations for greater than random losses of phylogenetic diversity that have been inferred for many clades under realistic extinction scenarios likely reflect the

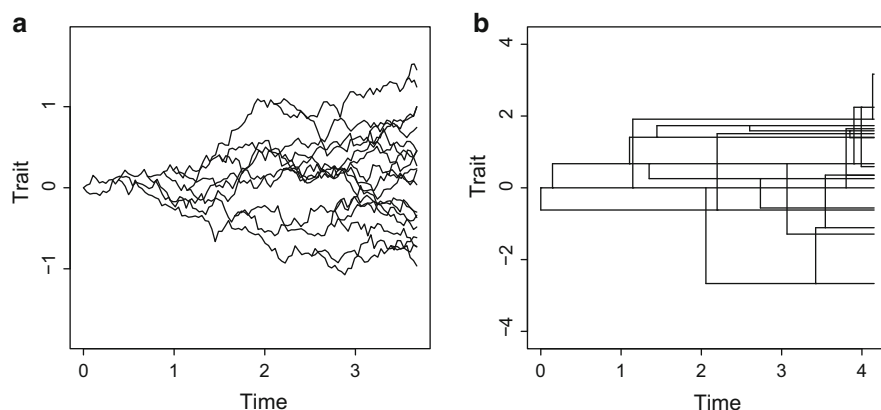


**Fig. 2** Ultrametric phylogenetic tree with three tips (A, B and C) and four branches with lengths in millions of years (Myrs). If tip taxa A and C become extinct, we lose two branches and 3 Myrs of evolutionary history from the tree. If sister taxa A and B become extinct, for example, because they share a phylogenetically conserved trait that predisposes them to high risk, we also lose 3 Myrs of evolutionary history, but this time three branches are lost from the phylogeny

particularities of phylogenetic tree topology in combination with a tendency for more extinction prone species to fall within species poor clades (Heard and Mooers 2000; von Euler 2001; Parhar and Mooers 2011). There does seem to be a general trend within some clades for threatened species to be overrepresented in species-poor clades (e.g. in mammals, Purvis et al. 2000b and birds, Bennett and Owens 1997). In plants, patterns appear mixed. As discussed above, there is some evidence suggesting an opposite trend to vertebrates, with a greater proportion of threatened plant species falling within species-rich clades (Schwartz and Simberloff 2001; Lozano and Schwartz 2005), and less evolutionary distinct lineages (Davies et al. 2011). Globally, however, species poor, and especially monotypic plant families, again appear to be more threatened, and their extinction would also result in a disproportionate loss of evolutionary history (Vamosi and Wilson 2008).

### ***Feature Diversity and Evolutionary Models of Character Change***

Underpinning the theoretical arguments for maximizing the preservation of phylogenetic diversity is the assumption that it captures feature diversity (i.e. variance in measured ecological and morphological traits), and thus selecting the set of taxa to maximize phylogenetic diversity will also maximize feature diversity (Faith 1992; Crozier 1997). Many biological traits demonstrate significant phylogenetic signal (Blomberg et al. 2003) and therefore this assumption might be broadly valid. However, the relationship between phylogenetic diversity, which is measured in millions of years, and feature diversity is not straightforward, but assumes a linear divergence between species over time, for example, as might be modeled under a Brownian motion process, in which trait variance increases in proportion with time, but for which evidence is mixed. Frequently, traits demonstrate much weaker phylogenetic signal than assumed by a strict Brownian motion model (e.g. Kamilar and Cooper 2013). Although there are a large number of alternative models of



**Fig. 3** Simulations showing accumulation of trait variance over time assuming a Brownian motion model of trait evolution **a** in which variance increases in proportion to time, versus a punctuated model of trait evolution **b** in which trait change occurs in bursts at speciation, and a pure-birth process of phylogenetic branching (see also Ingram 2011; Davies 2015)

evolutionary change, including the Ornstein-Uhlenbeck model which approximates stochastic evolution with stabilizing selection (Hansen 1997) and the early burst model that might characterize adaptive radiations (Harmon et al. 2010), we here (see Davies and Yessoufou 2013; Davies 2015) compare the potential loss of phylogenetic diversity under two models with very different assumptions: (1) a model of phylogenetic gradualism as represented by Brownian Motion (Fig. 3a), and (2) a punctuated model of evolution in which trait differences accumulate in bursts at speciation (Fig. 3b).

To date, the model of evolution has rarely been considered explicitly within the conservation phylogenetics literature (e.g. Owens and Bennett 2000). However, if traits evolve following a speciational model – as may be the case for body mass in mammals (Mattila and Bokma 2008) – where trait evolution occurs in bursts at speciation, each individual branch would capture similar feature diversity, and as such, the number of branches might be of equal, or greater conservation value than their summed lengths. Furthermore, because nonrandom extinction may target deeper branches in the tree-of-life (Mckinney 1997; Purvis et al. 2000a; Purvis 2008), we would predict a disproportionate loss of branches without necessarily a concomitant loss of total summed branch lengths (Fig. 2). Non-random extinction might therefore have a greater impact on number of branches lost than on the sum of their branch lengths – which has been the focus of most studies to date.

Using a dated phylogenetic tree for Primates, Carnivora and Artiodactyla, we (Davies and Yessoufou 2013) combined simulations and empirical extinction risk data from the IUCN Red List of threatened species (<http://www.iucnredlist.org/>) to explore the loss of phylogenetic diversity under two alternative evolutionary models. First, following standard practice, we calculated the expected loss of PD assuming a gradual model of evolution. Second, we also calculated the equivalent loss of diversity under a speciational model of evolution (in which all branches are assigned

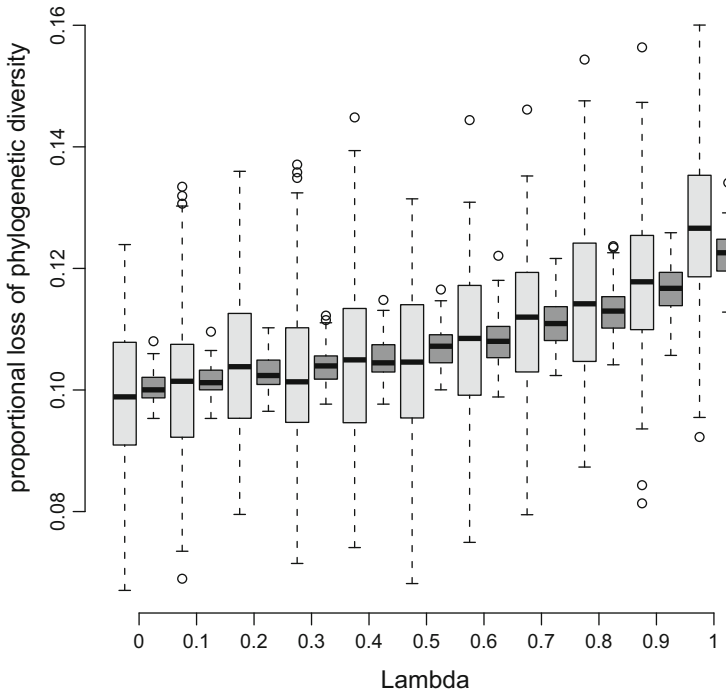
equal weights) following the approach of Witting and Loeschcke (1995). Extinction categories were first converted into extinction probabilities,  $p(\text{ext})$ , following Mooers et al. (2008) and assuming IUCN designations projected to 50 years. We then compared observed losses to expectations from the same distribution of  $p(\text{ext})$ , but randomly assigned to species at the tips of the phylogeny (100 replicates). Last, we explored the relationship between phylogenetic signal, estimated using Pagel's (1999) Lambda, and the loss of evolutionary history by evolving traits along the branches of simulated phylogenetic trees. Here, we assume a birth–death tree ( $b=0.2$ ,  $d=0$ , size  $n=240$ ), in contrast to the unrealistic coalescent trees used by Nee and May (1997). Based on the simulated trait values, a constant fraction of species (the top 25 %, as this broadly matches the proportion of threatened mammal species in the IUCN Red List) were then assigned high risk of extinction ( $p(\text{ext})=0.75$ ).

Our results reveal that under a speciation model of evolution, non-random extinction prunes more branches from the tree-of-life (see also Fig. 2), but that the loss of summed branch lengths (Faith's PD) does not depart significantly from random expectation (Davies and Yessoufou 2013). Although there is a weak trend for greater loss of phylogenetic diversity (PD) and number of branches lost with increasing phylogenetic signal in extinction risk, there is large variance in PD loss under random pruning such that observed losses typically overlap to a greater extent with the null distribution. In contrast, there is much less variance in the number of pruned branches such that random extinctions of equivalent intensity would prune similar number of branches. Therefore, observed number of branches loss more often falls outside the null distribution from randomizations (Fig. 4).

## Conclusion

There is an increasing call for prioritizing efforts towards the conservation of phylogenetic diversity (Mace et al. 2003; Forest et al. 2007; Davies et al. 2008). Implicit within this conservation agenda is an assumption that species diverge in their ecological and morphological traits more or less linearly through time, and thus that the evolutionary distance between species captures their functional differences. We (Davies and Yessoufou 2013) explored scenarios where this assumption is violated, and feature diversity occurs in bursts at speciation, matching to a punctuated model of trait evolution. Our results illustrate that projected extinctions might prune more branches from the tree-of-life than predicted from the same number of extinctions randomly distributed across the phylogeny; however, the loss of summed branch length might be no greater than expected by chance.

We do not suggest that punctuated evolution is necessarily a better model of trait change, but rather we emphasise the need for a more explicit consideration of evolutionary models if our aim is to maximize feature diversity. Recent advances in comparative methods have allowed comparisons between alternative evolutionary models, and frequently find strict Brownian motion to be a poor fit to observed trait



**Fig. 4** Results from simulated extinctions with varying levels of phylogenetic clustering (Lambda) across 100 random birth–death trees (see Fig. 1c) assuming  $p(\text{ext})=0.75$  for the top 25 % of species. *Light grey boxes* = expected loss of PD for empirical branch lengths (assuming phylogenetic gradualism or a Brownian motion model of trait change); *dark grey boxes* = expected loss of PD assuming equal branch lengths (matching to a punctuated model of trait change). Simulations with Lambda=0 are equivalent to random extinctions. This figure is similar to that in Davies and Yessoufou (2013), but presents a new set of stochastic simulations

data (e.g. Blomberg et al. 2003; O’Meara et al. 2006; Harmon et al. 2010). It remains possible that Brownian motion might still best capture aggregate species differences even when individual traits diverge from a Brownian motion model, assuming traits are evolving independently or when selective regimes fluctuate over time (Felsenstein 1988). However, this expectation has rarely been evaluated using empirical data.

Finally, we note that our understanding of the distribution of phylogenetic diversity across space and among communities might also be informed by further consideration of evolutionary models. For example, traditional metrics of phylogenetic diversity tend to correlate very closely with species richness (Rodrigues et al. 2005), although it is possible to identify regions of greater or lower phylogenetic diversity than predicted from species richness alone, for example, by looking at residual variation (e.g. Forest et al. 2007; Davies et al. 2008). The covariation between evolutionary history and species richness might exhibit very different properties under alternative evolutionary models, but as far as we are aware, there have not yet been any equivalent studies exploring such models in geographical space.



**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Achard F, Eva HD, Stibig HJ, Mayaux P, Gallego J, Richards T, Malingreau JP (2002) Determination of deforestation rates of the world's humid tropical forests. *Science* 297:999–1002
- Ashley MV, Wilson MF, Pergams ORW, O'Dowd DJ, Gende SM, Brown JS (2003) Evolutionarily enlightened management. *Biol Conserv* 111:115–123
- Baillie JEM, Hilton-Taylor C, Stuart SN (2004) A global species assessment. IUCN, Gland, Switzerland
- Barracough TG, Davies TJ (2005) Predicting future speciation. In: Purvis A, Gittleman JL, Brooks T (eds) *Phylogeny and conservation*. Cambridge University Press, Cambridge, pp 400–418
- Barracough TG, Vogler AP (2002) Recent diversification rates in North American tiger beetles estimated from a dated mtDNA phylogenetic tree. *Mol Biol Evol* 19:1706–1716
- Barracough TG, Nee S, Harvey PH (1998) Sister-group analysis in identifying correlates of diversification – comment. *Evol Ecol* 12:751–754
- Bennett PM, Owens IPF (1997) Variation in extinction risk among birds: chance or evolutionary predisposition. *Proc R Soc Lond B* 264:401–408
- Benton MJ, Emerson BC (2007) How did life become so diverse? The dynamics of diversification according to the fossil record and molecular phylogenetics. *Palaeontology* 50:23–40
- Bininda-Emonds ORP, Cardillo M, Jones KE, MacPhee RDE, Beck RMD, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A (2007) The delayed rise of present-day mammals. *Nature* 446:507–512
- Blomberg SP, Garland T, Ives AR (2003) Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57:717–745
- Bodmer RE, Eisenberg JF, Redford KH (1997) Hunting and the likelihood of extinction of Amazonian mammals. *Conserv Biol* 11:460–466
- Bradshaw CJA, Giam X, Tan HTW, Brook BW, Sodhi NS (2008) Threat or invasive status in legumes is related to opposite extremes of the same ecological and life history attributes. *J Ecol* 96:869–883
- Brook BW, Sodhi NS, Bradshaw CJA (2008) Synergies among extinction drivers under global change. *Trends Ecol Evol* 23:453–460
- Butchart SHM, Stattersfield AJ, Bennun LA, Shutes SM, Akcakaya HR, Baillie JEM, Stuart SN (2004) Measuring global trends in the status of biodiversity: red list indices for birds. *PLoS Biol* 2:e383
- Cardillo M (2003) Biological determinants of extinction risk: why are smaller species less vulnerable? *Anim Conserv* 6:63–69
- Cardillo M, Bromham L (2001) Body size and risk of extinction in Australian mammals. *Conserv Biol* 15:1435–1440
- Cardillo M, Meijaard E (2012) Are comparative studies of extinction risk useful for conservation? *Trends Ecol Evol* 27:167–171
- Cardillo M, Mace GM, Jones KE, Bielby J, Bininda-Emonds ORP, Sechrest W, Orme CDL, Purvis A (2005) Multiple causes of high extinction risk in large mammal species. *Sciences* 309:1239–1241
- Cardillo M, Mace GM, Gittleman JL, Purvis A (2006) Latent extinction risk and the future battlegrounds of mammal conservation. *Proc Natl Acad Sci U S A* 103:4157–4161

- Cardillo M, Mace GM, Gittleman JL, Jones KE, Bielby J, Purvis A (2008) The predictability of extinction: biological and external correlates of decline in mammals. *Proc R Soc B* 275:1441–1448
- Cione AL, Tonni EP, Soibelzon L (2003) The Broken zig-zag: late Cenozoic large mammal and tortoise extinction in South America. *Rev Mus Argent Cienc Nat* 5:1–19
- Collen B, Böhm M, Kemp R, Baillie JEM (2012) Spineless: status and trends of the world's invertebrates. Zoological Society of London, London. ISBN 978-0-900881-68-8
- Cooper N, Bielby J, Thomas HG, Purvis A (2008) Macroecology and extinction risk correlates of frogs. *Glob Ecol Biogeogr* 17:211–221
- Crisp MD, Cook LG (2009) Explosive radiation or cryptic mass extinction? Interpreting signatures in molecular phylogenies. *Evolution* 63:2257–2265
- Crozier RH (1997) Preserving the information content of species: genetic diversity, phylogeny, and conservation worth. *Annu Rev Ecol Syst* 28:243–268
- Damuth J (1981) Population density and body size in mammals. *Nature* 290:699–700
- Daru BH, Yessoufou K, Mankga LT, Davies TJ (2013) A global trend towards the loss of evolutionarily unique species in mangrove ecosystems. *PLoS One* 8(6):e66686
- Davies TJ (2015) Losing history: how extinctions prune features from the tree-of-life. *Philos Trans R Soc B* 370. doi:[10.1098/rstb.2014.0006](https://doi.org/10.1098/rstb.2014.0006)
- Davies TJ, Cadotte MW (2011) Quantifying biodiversity – does it matter what we measure? In: Zachos FE, Habel JC (eds) *Biodiversity hotspots*. Springer, Heidelberg, pp 43–60
- Davies TJ, Yessoufou K (2013) Revisiting the impacts of non-random extinction on the tree-of-life. *Biol Lett* 9:20130343
- Davies TJ, Barraclough TG, Chase MW, Soltis PS, Soltis DE, Savolainen V (2004) Darwin's abominable mystery: insights from a supertree of the angiosperms. *Proc Natl Acad Sci U S A* 101:1904–1909
- Davies TJ, Fritz SA, Grenyer R, Orme CD, Bielby J, Bininda-Emonds OR, Cardillo M, Jones KE, Gittleman JL, Mace GM, Purvis A (2008) Phylogenetic trees and the future of mammalian biodiversity. *Proc Natl Acad Sci USA* 105:11556–11563
- Davies TJ, Smith GF, Bellstedt DU, Boatwright JS, Bytebier B, Cowling RM, Forest F, Harmon LJ, Muasya AM, Schrire BD, Steenkam Y, Van der Bank M, Savolainen V (2011) Extinction risk and diversification are linked in a plant biodiversity hotspot. *PLoS Biol* 9:e1000620
- Davies TJ, Wolkovich EM, Kraft NJB, Salamin N, Allen JM, Ault TR, Betancourt JL, Bolmgren K, Cleland EE, Cook BI, Crimmins TM, Mazer SJ, McCabe GJ, Pau S, Regetz J, Schwartz MD, Travers SE (2013) Phylogenetic conservatism in plant phenology. *J Ecol* 101:1520–1530
- Diamond J (1984) Normal extinctions of isolated populations. In: Nitecki MH (ed) *Extinctions*. University of Chicago Press, Chicago, pp 191–246
- Diamond J (1989) Overview of recent extinctions. In: Pearl M, Western D (eds) *Conservation for the twenty-first century*. Oxford University Press, New York, pp 37–41
- Dirzo R, Raven PJ (2003) Global state of biodiversity and loss. *Annu Rev Environ Resour* 28:137–167
- Endler JA (1986) *Natural selection in the wild*. Princeton University Press, Princeton, USA
- Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61:1–10
- Faith DP (2002) Quantifying biodiversity: a phylogenetic diversity. *Conserv Biol* 16:248–252
- Faith DP (2008) Threatened species and the potential loss of phylogenetic diversity: conservation scenarios based on estimated extinction probabilities and phylogenetic risk analysis. *Conserv Biol* 22:1461–1470
- Faith DP, Magallón S, Hendry AP, Conti E, Yahara T, Donoghue MJ (2010) Ecosystem services: an evolutionary perspective on the links between biodiversity and human well-being. *Curr Opin Environ Sustain* 2:66–74
- Felsenstein J (1985) Phylogenies and the comparative methods. *Am Nat* 125:1–15
- Felsenstein J (1988) Phylogenies and quantitative characters. *Annu Rev Ecol Syst* 19:445–471
- Fenschel T (1974) Intrinsic rate of natural increase: the relationship with body size. *Oecologia* 14:317–332

- Fisher DO, Owens IPF (2004) The comparative method in conservation biology. *Trends Ecol Evol* 19:391–398
- Forest F, Grenyer R, Rouget M, Davies J, Cowling RM, Faith DP, Balmford A, Manning JC, Proches S, Van der Bank M, Reeves G, Hedderson TAJ, Savolainen V (2007) Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature* 445:757–760
- Freville H, McConway K, Dodd M, Silvertown J (2007) Prediction of extinction in plants: interactions of extrinsic threats and life history traits. *Ecology* 88:2662–2672
- Fritz SA, Purvis A (2010) Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conserv Biol* 24:1042–1051
- Gavrilets S, Vose A (2005) Dynamic patterns of adaptive radiation. *Proc Natl Acad Sci U S A* 102:18040–18045
- Goldberg EM, Lancaster LT, Ree RH (2011) Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Syst Biol* 60:451–465
- Grandcolas P, Nattier R, Legendre F, Pellens R (2011) Mapping extrinsic traits such as extinction risks or modelled bioclimatic niches on phylogenies: does it make sense at all? *Cladistics* 27:181–185
- Grenyer R, Orme CDL, Jackson SF, Thomas GH, Davies RG, Davies TG, Jones KE, Olson VA, Ridgely RS, Rasmussen PC, Ding T-S, Bennett PM, Blackburn TM, Gaston KJ, Gittleman JL, Owens IPF (2006) Global distribution and conservation of rare and threatened vertebrates. *Nature* 444:93–96
- Haberl H, Erb KH, Krausmann F, Gaube V, Bondeau A, Plutzer C, Gingrich S, Lucht W, Fischer-Kowalski M (2007) Quantifying and mapping the human appropriation of net primary production in earth's terrestrial ecosystems. *Proc Natl Acad Sci U S A* 104:12942–12947
- Hansen TF (1997) Stabilizing selection and the comparative analysis of adaptation. *Evolution* 51:1341–1351
- Harmon LJ, Schulte JA, Larson A, Losos JB (2003) Tempo and model of evolutionary radiation in iguanian lizards. *Science* 301:961–964
- Harmon LJ, Losos JB, Davies TJ, Gillespie RG, Gittleman JL, Jennings WB, Kozak KH, McPeck MA, Moreno-Roark F, Near TJ, Purvis A, Ricklefs RE, Dolph S, Schulte JA II, Seehausen O, Sidlauskas BL, Torres-Carvajal O, Weir JT, Mooers AO (2010) Early bursts of body size and shape evolution are rare in comparative data. *Evolution* 64:2385–2396
- Harvey PH, May RM, Nee S (1994) Phylogenies without fossils. *Evolution* 48:523–529
- Heard SB, Mooers AO (2000) Phylogenetically patterned speciation rates and extinction risks change the loss of evolutionary history during extinctions. *Proc R Soc Lond B* 267:613–620
- Hooper DU, Adair EC, Cardinale BJ, Byrnes JEK, Hungate BA, Matulich KL, Gonzalez A, Duffy EJ, Gamfeldt L, O'Connor MI (2012) A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature* 486:105–108
- Huang S, Gittleman JG, Davies TJ (2012) How global extinctions impact regional biodiversity in mammals. *Biol Lett* 8:222–225
- Ingram T (2011) Speciation along a depth gradient in a marine adaptive radiation. *Proc R Soc B* 278:613–618
- Isaac NJB, Turvey ST, Collen B, Waterman C, Baillie JEM (2007) Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PLoS One* 2:e296
- Isaac NJB, Redding DW, Meredith HM, Safi K (2012) Phylogenetically-informed priorities for amphibian conservation. *PLoS One* 7:e43912
- IUCN (2003) IUCN red list of threatened species. IUCN, Gland, Switzerland
- IUCN (2010) IUCN sampled red list index for plants. <http://www.kew.org/science-conservation/kew-biodiversity/plants-at-risk/index.htm>.
- Jablonski D (1995) Extinctions in the fossil record. In: Lawton JH, May RM (eds) *Extinction rates*. Oxford University Press, Oxford, pp 25–44
- Jansson R, Davies TJ (2008) Global variation in diversification rates of flowering plants: energy versus climate change. *Ecol Lett* 11:173–183

- Jerozolinski A, Peres CA (2003) Bringing home the biggest bacon: a cross-site analysis of the structure of hunter-kill profiles in neotropical forests. *Biol Conserv* 111:415–425
- Johnson CN (2002) Determinants of loss of mammal species during the late Quaternary “mega-fauna” extinction: life history and ecology, but not body size. *Proc R Soc B* 269:2221–2228
- Jones DL (2002) Cycads of the world: ancient plants in today’s landscape. Smithsonian Institution Press, Washington, DC
- Kadereit JW, Greibler EM, Comes HP (2004) Quaternary diversification in European alpine plants: pattern and process. *Philos Trans R Soc B* 359:265–274
- Kamilar JM, Cooper N (2013) Phylogenetic signal in primate behaviour, ecology and life history. *Philos Trans R Soc B* 368:20120341
- Kettlewell MG (1972) The evolution of melanism. Oxford University Press, Oxford
- Kinnison MT, Hendry AP (2001) The pace of modern life II: from rates of contemporary micro-evolution to pattern and process. *Genetica* 112:145–164
- Kozak KH, Weisrock DW, Larson A (2006) Rapid lineage accumulation in a non-adaptive radiation: phylogenetic analysis of diversification rates in eastern North American woodland salamanders (Plethodontidae: *Plethodon*). *Proc R Soc B* 273:539–546
- Laurance WF (1991) Ecological correlates of extinction proneness in Australian tropical rainforest mammals. *Conserv Biol* 5:79–89
- Lawton JH (1995) Population dynamic principles. In: Lawton JH, May RM (eds) *Extinction rates*. Oxford University Press, Oxford, pp 147–163
- Linder HP, Eldenäs P, Briggs BG (2003) Contrasting patterns of radiation in African and Australian Restionaceae. *Evolution* 57:2688–2702
- Lohle C, Eschenbach W (2011) Historical bird and terrestrial mammal extinction rates and causes. *Divers Distrib* 18:84–91
- Lozano FD, Schwartz MW (2005) Patterns of rarity and taxonomic group size in plants. *Biol Conserv* 126:146–154
- MacArthur RH, Wilson EO (1963) An equilibrium theory of insular zoogeography. *Evolution* 17:373–387
- MacArthur RH, Wilson EO (1967) The theory of island biogeography. Princeton University Press, Princeton, USA
- Mace GM, Gittleman JL, Purvis A (2003) Preserving the tree of life. *Science* 300:1707–1709
- Mace G, Masundire H, Baillie JEM (2005a) Biodiversity. In: Hassan R, Scholes R, Ash N (eds) *Ecosystems and human well-being: current state and trends: findings of the condition and trends working group*. Island Press, Washington, DC, pp 77–122
- Mace GM, Baillie JEM, Masundire H, Ricketts TH, Brooks TM et al (2005b). Biodiversity. In: *The millennium ecosystem assessment: current status and trends: findings of the conditions and trends working group. Ecosystems and human well-being*. Island Press, Washington, DC, pp 53–98
- Machordom A, Macpherson E (2004) Rapid radiation and cryptic speciation in squat lobsters of the genus *Munida* (Crustacea, Decapoda) and related genera in the southwest Pacific: molecular and morphological evidence. *Mol Phylogenet Evol* 33:259–279
- Maddison WP, Midford PE, Otto SP (2007) Estimating a binary character’s effect on speciation and extinction. *Syst Biol* 56:701–710
- Martin PS (1967) Prehistoric overkill. In: Martin PS, Wright HE (eds) *Pleistocene extinctions: the search for a cause*. Yale University Press, New Haven, pp 75–120
- Mattila TM, Bokma F (2008) Extant mammal body masses suggest punctuated equilibrium. *Proc R Soc B* 275:2195–2199
- May RM (2010) Tropical arthropod species, more or less? *Science* 329(5987):41–42
- May RM, Lawton JH, Stork NE (1995) Assessing extinction rates. In: Lawton JH, May RM (eds) *Extinction rates*. Oxford University Press, Oxford, pp 1–24
- McCormack JE, Harvey MG, Faircloth BC, Crawford NG, Glenn TC, Brumfield RT (2013) A phylogeny of birds based on over 1500 loci collected by target enrichment and high-throughput sequencing. *PLoS One* 8:e54848

- McKinney ML (1997) Extinction vulnerability and selectivity: combining ecological and paleontological views. *Annu Rev Ecol Syst* 28:495–516
- Meijaard E, Sheil D, Marshall J, Nasi R (2007) Phylogenetic age is positively correlated with sensitivity to timber harvest in Bornean mammals. *Biotropica* 40:76–85
- Millennium Ecosystem Assessment (2005) Ecosystems and human well-being: synthesis. Island Press, Washington, DC, 137 pp
- Mooers AØ, Heard SB, Chrostowski E (2005) Evolutionary heritage as a metric for conservation. In: Purvis A, Brooks TL, Gittleman JL (eds) *Phylogeny and conservation*. Oxford University Press, Oxford, pp 120–138
- Mooers AØ, Faith DP, Maddison WP (2008) Converting endangered species categories to probabilities of extinction for phylogenetic conservation prioritization. *PLoS One* 3:e3700
- Mooers A, Gascuel O, Stadler T, Li H, Steel M (2012) Branch lengths on birth–death trees and the expected loss of phylogenetic diversity. *Syst Biol* 61:195–203
- Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B (2011) How many species are there on Earth and in the ocean? *PLoS Biol* 9:e1001127
- Morrison CL, Rios R, Duffy JE (2004) Phylogenetic evidence for an ancient rapid radiation of Caribbean sponge-dwelling snapping shrimps (*Synalpheus*). *Mol Phylogenet Evol* 30:563–581
- Nee S (2006) Birth-death models in macroevolution. *Annu Rev Ecol Syst* 37:1–17
- Nee S, May RM (1997) Extinction and the loss of evolutionary history. *Science* 278:692–694
- Nee S, Holmes EC, May RM, Harvey PH (1994) Extinction rates can be estimated from molecular phylogenies. *Proc R Soc Lond B* 344:77–82
- Niklas KJ (1997) *The evolutionary biology of plants*. University of Chicago Press, Chicago
- O'Meara BC, Ane' C, Sanderson MJ, Wainwright PC (2006) Testing for different rates of continuous trait evolution using likelihood. *Evolution* 60:922–933
- Orme CDL, Davies RG, Burgess M, Eigenbrod F, Pickup N, Olson VA, Webster AJ, Ding T-S, Rasmussen PC, Ridgely RS, Stattersfield AJ, Bennett PM, Blackburn TM, Gaston KJ, Owens IPF (2005) Global hotspots of species richness are not congruent with endemism or threat. *Nature* 436:1016–1019
- Owens IPF, Bennett PM (2000) Quantifying biodiversity: a phenotypic perspective. *Conserv Biol* 14:1014–1022
- Pagel M (1999) Inferring the historical patterns of biological evolution. *Nature* 401:877–884
- Parhar RK, Mooers AØ (2011) Phylogenetically clustered extinction risks do not substantially prune the tree of life. *PLoS One* 6:e23528
- Parmesan C (2007) Influences of species, latitudes and methodologies on estimates of phenological response to global warming. *Glob Chang Biol* 13:1860–1872
- Phillimore AB, Price TD (2008) Density-dependent cladogenesis in birds. *PLoS Biol* 6:e71
- Pilgrim ES, Crawley MJ, Dolphin K (2004) Patterns of rarity in the native British flora. *Biol Conserv* 120:161–170
- Pimm SL (1991) *The balance of nature?* University of Chicago Press, Chicago
- Pimm SL, Jones HL, Diamond J (1988) On the risk of extinction. *Am Nat* 132:757–785
- Pimm SL, Russell GJ, Gittleman JL, Brooks TM (1995) The future of biodiversity. *Science* 269:347–350
- Purcell SW, Polidoro BA, Hamel J-F, Gamboa RU, Mercier A (2014) The cost of being valuable: predictors of extinction risk in marine invertebrates exploited as luxury seafood. *Proc R Soc B* 281:20133296
- Purvis A (2008) Phylogenetic approaches to the study of extinction. *Annu Rev Ecol Syst* 39:301–319
- Purvis A, Agapow PM, Gittleman JL, Mace GM (2000a) Nonrandom extinction and the loss of evolutionary history. *Science* 288:328–330
- Purvis A, Gittleman JL, Cowlshaw G, Mace GM (2000b) Predicting extinction risk in declining species. *Proc R Soc Lond B* 267:1947–1952

- Purvis A, Gittleman JL, Brooks TM (2005) Phylogeny and conservation. Cambridge University Press, Cambridge
- Pyron RA, Wiens JJ (2013) Large-scale phylogenetic analyses reveal the causes of high tropical amphibian diversity. *Proc R Soc B* 280:20131622
- Rabosky DL (2010) Extinction rates should not be estimated from molecular phylogenies. *Evolution* 64–6:1816–1824
- Rabosky DL, Lovette IJ (2008) Explosive evolutionary radiations: increasing extinction or decreasing speciation through time? *Evolution* 62:1866–1875
- Raup DM (1986) Biological extinction in Earth history. *Science* 231:1528–1533
- Raup D (1991) Extinction: bad genes or bad luck? *New Sci* 131:46–49
- Raup DM (1993) Extinction: bad genes or bad luck? Oxford University Press, Oxford
- Ricklefs RE, Bermingham E (2002) The concept of the taxon cycle in biogeography. *Glob Ecol Biogeogr* 11:353–361
- Rodrigues ASL, Brooks TM, Gaston KJ (2005) Integrating phylogenetic diversity in the selection of priority areas for conservation: does it make a difference? In: Purvis A, Gittleman JL, Brooks T (eds) *Phylogeny and conservation*. Cambridge University Press, Cambridge, pp 101–119
- Rolland J, Condamine FL, Jiquet F, Morlon H (2014) Faster speciation and reduced extinction in the tropics contribute to the mammalian latitudinal diversity gradient. *PLoS Biol* 12:e1001775
- Russell GJ, Brooks TM, McKinney MM, Anderson CG (1998) Present and future taxonomic selectivity in birds and mammal extinctions. *Conserv Biol* 12:1365–1376
- Sandel B, Arge L, Dalsgaard B, Davies RG, Gaston KJ, Sutherland WJ, Svenning J-C (2011) The influence of late quaternary climate-change velocity on species endemism. *Science* 334:660–664
- Saslis-Lagoudakis, CH, Kligard, BB, Forest F, Francis L, Savolainen V, Williamson EM, Hawkins JA (2011) The use of phylogeny to interpret cross-cultural patterns in plant use and guide medicinal plant discovery: an example from *Pterocarpus* (Leguminosae). *PLoS ONE* 6(7):e22275
- Scantlebury DP (2013) Diversification rates have declined in the Malagasy herpetofauna. *Proc R Soc B* 280:20131109
- Schluter D (2000) *The ecology of adaptive radiation*. Oxford University Press, Oxford
- Schwartz MW, Simberloff D (2001) Taxon size predicts rarity in vascular plants. *Ecol Lett* 4:464–469
- Seehaussen O (2006) Conservation: losing biodiversity by reverse speciation. *Curr Biol* 9:R334–R337
- Shaw AJ, Cox CJ, Goffinet B, Buck WR, Boles SB (2003) Phylogenetic evidence of a rapid radiation of pleurocarpous mosses (Bryophyta). *Evolution* 57:2226–2241
- Simpson G (1953) *The major features of evolution*. Columbia University Press, New York, 434 p
- Sodhi NS, Brook BW (2006) *Southeast Asian biodiversity in crisis*. Cambridge University Press, Cambridge
- Sodhi NS, Koh LP, Peh KS-H, Tan HTW, Chazdon RL, Corlett RT, Lee TM, Colwell RK, Brook BW, Sekercioglu CH, Bradshaw CJA (2008) Correlates of extinction proneness in tropical angiosperms. *Divers Distrib* 14:1–10
- Srivastava DS, Cadotte MW, MacDonald AAM, Marushia RG, Mirotchnick N (2012) Phylogenetic diversity and the functioning of ecosystems. *Ecol Lett* 15:637–648
- Stuart SN, Chanson JS, Cox NA, Young BE, Rodrigues ASL, Fischman DL, Waller RW (2004) Status and trends of amphibian declines and extinctions worldwide. *Science* 306:1783–1786
- Taylor TN, Taylor EL, Krings M (2009) *Paleobotany: the biology and evolution of fossil plants*. Academic, Burlington
- Thomas CD, Cameron A, Green RE, Bakkenes M, Beaumont LJ, Collingham YC, Erasmus BF, De Siqueira MF, Grainger A, Hannah L, Hughes L, Huntley B, Van Jaarsveld AS, Midgley GF, Miles L, Ortega-Huerta MA, Peterson AT, Phillips OL, Williams SE (2004) Extinction risk from climate change. *Nature* 427:145–148
- Turgeon J, Stoks R, Thum RA, Brown JM, McPeck MA (2005) Simultaneous quaternary radiations of three damselfly clades across the Holarctic. *Am Nat* 165:E78–E107



- Vamosi JC, Vamosi SM (2008) Extinction risk escalates in the tropics. *PLoS One* 3:e3886
- Vamosi JC, Wilson JR (2008) Nonrandom extinction leads to elevated loss of angiosperm evolutionary history. *Ecol Lett* 11:1047–1053
- Verde Arregoitia LD, Blomberg SP, Fischer DO (2013) Phylogenetic correlates of extinction risk in mammals: species in older lineages are not at greater risk. *Proc R Soc B* 280:20131092
- Vitousek PM, Mooney HA, Lubchenco J, Melillo JM (1997) Human domination of earth's ecosystems. *Science* 277:494–499
- Von Euler F (2001) Selective extinction and rapid loss of evolutionary history in the bird fauna. *Proc R Soc B* 268:127–130
- Wake DB, Vredenburg VT (2008) Are we in the midst of the sixth mass extinction? A view from the world of amphibians. *Proc Natl Acad Sci U S A* 105:11466–11473
- Walker KJ, Preston CD (2006) Ecological predictors of extinction risk in the flora of lowland England, UK. *Biodivers Conserv* 15:1913–1942
- Weir J (2006) Divergent timing and patterns of species accumulation in lowland and highland Neotropical birds. *Evolution* 60:842–855
- Williams ST, Reid DG (2004) Speciation and diversity on tropical rocky shores: a global phylogeny of snails of the genus *Echinolittorina*. *Evolution* 58:2227–2251
- Willis CG, Ruhfel B, Primack RB, Miller-Rushing AJ, Davis CC (2008) Phylogenetic patterns of species loss in Thoreau's woods are driven by climate change. *Proc Natl Acad Sci U S A* 105:17029–17033
- Willis CG, Ruhfel BR, Primack RB, Miller-Rushing AJ, Losos JB, Davis CC (2010) Favourable climate change response explains non-native species' success in Thoreau's Woods. *PLoS One* 5:e8878
- Wilson EO (1961) The nature of the taxon cycle in the Melanesian ant fauna. *Am Nat* 95:169–193
- Wilson EO (1992) The diversity of life. Norton WW & Co, New York
- Winter M, Schweigera O, Klotz S, Nentwig W, Andriopoulos P, Arianoutsou M, Basnou C, Delipetrou P, Didžiulis V, Hejda M, Hulme PE, Lambdon PW, Pergl J, Pyšek P, Roy DB, Kühn I (2009) Plant extinctions and introductions lead to phylogenetic and taxonomic homogenization of the European flora. *Proc Natl Acad Sci U S A* 106:21721–21725
- Winter M, Vincent D, Oliver S (2013) Phylogenetic diversity and nature conservation: where are we? *Trends Ecol Evol* 28:199–204
- Witting L, Loeschcke V (1995) The optimization of biodiversity conservation. *Biol Conserv* 71:205–207
- Wright SJ (2005) Tropical forests in a changing environment. *Trends Ecol Evol* 20:553–560
- Xiang Q-Y, Manchester SR, Thoma DT, Zhang W, Fan C (2005) Phylogeny, biogeography, and molecular dating of cornelian cherries (*Cornus*, Cornaceae): tracking tertiary plant migration. *Evolution* 59:1685–1700
- Yessoufou K, Daru BH, Davies TJ (2012) Phylogenetic patterns of extinction risk in the eastern arc ecosystems, an African biodiversity hotspot. *PLoS One* 7:e47082
- Yessoufou K, Bamigboye SO, Daru BH, Van der Bank M (2014) Evidence of constant diversification punctuated by a mass extinction in the African cycads. *Ecol Evol* 4:50–58
- Zhang Y-B, Ma K-P (2008) Geographic distribution patterns and status assessment of threatened species in China. *Biodivers Conserv* 17:1783–1798

# Phylogenetics and Conservation in New Zealand: The Long and the Short of It

Steven A. Trewick and Mary Morgan-Richards

**Abstract** Phylogenetic trees represent the evolutionary relationships of taxa at the branch tips. Although long branches in a tree can arise because a taxon has no close relatives, they can also result from other processes; care is needed when inferences are made from the shape of a phylogeny. New Zealand has many endangered species and some biologists infer high evolutionary distinctiveness of these endemics. Although there is evidence that some New Zealand birds are phylogenetically distinct using them as a calibration of continental drift vicariance has been misleading. In reptiles, extensive conservation resources have been devoted to management of tuatara, in part due to their phylogenetic distinctiveness as sister to all lizards and snakes. The lack of extant diversity in the tuatara lineage could indicate that this line will contribute little to biodiversity in the future, in contrast to New Zealand squamates that have radiated to occupy diverse habitats. All life on earth has a common ancestor so phylogenetic distinctiveness of any organism must be viewed in the context of the whole. A logical extension of building conservation strategy this way is a focus on microscopic life because microbes encompass far more diversity than do eukaryotes. Furthermore, this diversity can be captured in microbiomes such as soils and marine sponges that include many species and many phyla. To achieve true phylogenetic representation of life on earth requires conservation of ecosystems. Although large animals and plants are traditionally chosen as flagship species, a more impartial approach might focus on microbes that underpin ecosystem function.

**Keywords** Evolutionary potential • Kākāpō • Tuatara • Sponges • Microbiomes

---

S.A. Trewick (✉) • M. Morgan-Richards  
Ecology Group, Institute of Agriculture and Environment, Massey University,  
Private Bag 11–222, Palmerston North, 4442 New Zealand  
e-mail: [s.trewick@massey.ac.nz](mailto:s.trewick@massey.ac.nz); [m.morgan-richards@massey.ac.nz](mailto:m.morgan-richards@massey.ac.nz)



## Introduction

Various described as a diversity hotspot, Gondwanan remnant and paradise lost to invasive species (e.g. Daugherty et al. 1993; Gibbs 2006; Lee et al. 2006), New Zealand presents enormous challenges for conservation (DOC 2000). Key is the question of how to prioritise management effort and funding (e.g. Cullen 2012; Walker et al. 2012) and amongst the available tools for prioritisation is phylogenetics (Margules and Pressey 2000; Purvis et al. 2005; Rolland et al. 2011). Here we consider just two aspects of phylogeny in conservation with reference to New Zealand biota. First we examine the implications of long branches in phylogenetic trees and the biological information they might contain. We highlight the role of taxon sampling in the identification of long branches and the biological significance of phylogenetic distinctiveness. We then consider a broader view of phylogenetic diversity including microorganisms that are rarely considered in conservation planning (Nee 2004a, b). As the fountain of phylogenetic diversity, microbial diversity, which also underpins ecological diversity and ecosystem function, provides scaling for conservation that is not influenced by size, scarcity and marketable appeal. We argue that the logical extension of a strict application of conservation prioritisation using phylogenetic distinctiveness must result in a focus on unicellular organisms that are not traditionally emphasised. Using data from marine sponges we provide an example of a micro-environment that is rich in phylogenetic diversity. Diversity-rich microbiomes may be the much-needed foci for conservation of higher order biodiversity.

## Long Branches and Their Biological Meaning

An avowed objective of conservation is the maintenance of maximum evolutionary potential (Avice 2005). But as it is not feasible to confidently predict which lineages will be successful in the future, not least because much that happens in biology is subject to stochasticism. Retaining maximum evolutionary history might be an alternative and better, or at least achievable strategy. In this context, taxa at the tips of long branches attract special attention although a similar level of investment in representatives of speciose clades is also required to conserve the history represented by those lineages too.

On the face of it taxa on long branches appear to represent long evolutionary history. But what is a long branch and what information does it carry (or not carry) about the past?

Long branches on phylogenetic trees result from one of three processes:

1. The lineage might have evolved without lineage splitting increasing species diversity. This would involve each new species replaces its immediate ancestor in succession.
2. The branch/lineage experienced an accelerated rate of molecular evolution in relation to all others, at the locus providing the (presumed) phylogenetic signal.
3. The clade that includes the taxon in question has been extensively pruned so that near relatives have been removed.

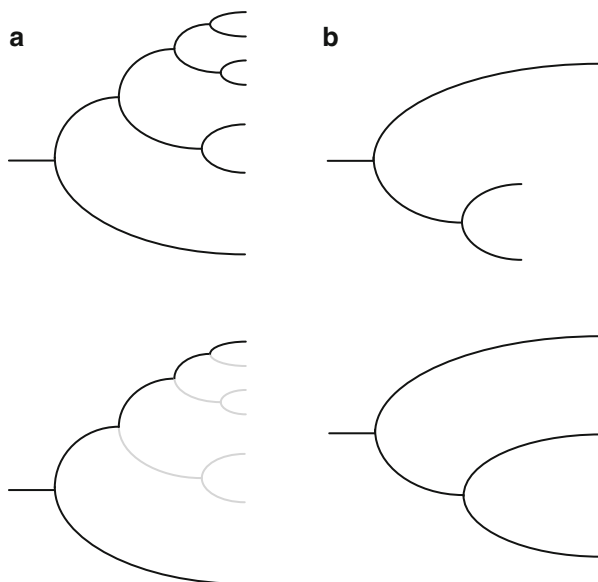
Processes 1 and 2 could in themselves constitute evidence of distinctive unusual evolutionary mechanisms that demand conservation; however this would depend on verification. For 1, a detailed fossil record would be required to refute the alternative and more likely hypothesis that the lineage evolved via lineage splitting (Gould and Eldredge 1993) but has been subject to extinction (as in 3) (Vaux et al. 2015). For 2, analysis of other gene sequences would be needed to identify if rate variation was consistent across the genome or was due to gene-specific positive selection. If rate variation is locus-specific it is highly likely that resulting data are not tree-like, and hence phylogenetically misleading though interesting in other ways.

Process 3 can be further subdivided by the cause of the deficiency of closely related taxa. The absence of close relatives could result simply from experimental failure to sample extant species that are more closely related, or might represent extinction of other members of the clade at any time in the past. These alternatives can be readily tested by inclusion of all plausible extant relatives in phylogenetic analyses. Where a “clade” is truly represented by a singleton (i.e. no closer relatives exist on the planet), then the sister group corollary has to be considered. Every lineage exists as a sister to another lineage or clade so that taxa at the tips of long branches are not intrinsically more important in evolutionary terms than those on short branches. This can be readily demonstrated by the simple expedient of pruning an existing data set (Fig. 1a).

The role of variation in rates of molecular evolution in producing long branches can be determined from the underlying data. In ideal circumstances, if phylogenetic reconstruction has used appropriate models of DNA evolution and informative outgroups, trees with long branches resulting from rate acceleration are expected to look quite different from those that simply lack near relatives (Fig. 1b). Phylogenetic trees inferred from molecular data use sampling at time zero (the present) so it is expected that sequences will change subject to some local rate variation around a mean for a given taxon group, gene etc. with a relatively small variance (see Bromham and Penny 2003). Thus, typically, a phylogeny that is subject to local rate variation will appear unbalanced; branch tips will not be adjacent or nearly so (Fig. 1b). An obvious situation in which local branch rate might result in a long branch and/or phylogenetic misplacement of the node, exists when genes used for tree estimation are under positive/diversifying selection in some taxa, but are constrained in others.

The relative length of a branch in a phylogenetic tree might be used to direct conservation strategy in three distinct ways.

1. Species on long naked branches in phylogenies that include the appropriate sample of extant taxa can be taken as important representation of groups that were once more diverse, and that represent evolutionary potential that is different from the sister clade.
2. Species on long branches for which there is phylogenetic evidence of lineage specific acceleration of molecular evolution can be taken as representing interesting genomes with unusual genetic properties. A long branch of this type might result from genome-wide rate increase (compared to sister group) or locus-specific effects and represent specific adaptive traits.



**Fig. 1** Phylogenetic trees illustrate the evolutionary relationships of species. **a** Influence of sampling on apparent cladogenesis. Pruning branches (grey) from the top phylogeny results in an apparent long branch for the remaining clade singleton (bottom). **b** Long branches where (top) unbalanced branch lengths result from different rates of molecular evolution at the gene used to make the tree (or wrong outgroup), and (bottom) equal rates of molecular evolution but different rates of speciation

3. Taxa on short branches nested within a clade, but accompanied by other character information on their distinctiveness (morphology, behaviour, habitat type) could be important representatives of evolutionarily innovative lineages.

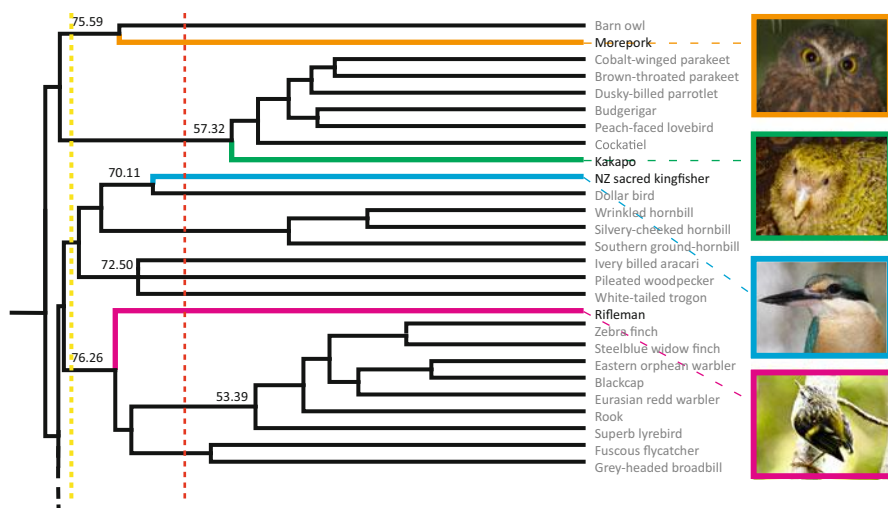
For large organisms such as birds and mammals and many plant groups it is relatively easy to know how complete is taxon sampling amongst extant biota. In most cases existing taxonomy and checklists provide strong indicators. However, for smaller organisms, classification is often incomplete, taxa are not described and there are many instances of misclassification because character analysis has been lacking. Thus the significance of branch length is tempered by other information and the most phylogenetically diverse types of life on earth are severely under-represented.

### ***Birds on Long Branches***

Our understanding of bird evolution has been advanced rapidly through the use of molecular phylogenies that have demonstrated that birds began to diversify before the K/Pg boundary (Cretaceous/Palaeogene, formerly K/T; about 65 million years

ago) (e.g. Penny and Phillips 2004). This finding countered opinions established on a formerly deficient fossil record that extinction of dinosaurs and associated fauna at K/Pg provided the impetus for subsequent bird diversification. Much of this phylogenetic work has centred on analysis of mitochondrial genome data (mitogenomics; e.g., Pratt et al. 2009; Morgan-Richards et al. 2008; Slack et al. 2007; Gibbs and Penny 2010), although multilocus nuclear data have started to be generated from high throughput DNA sequencing (NGS) and advanced bioinformatics (e.g., Hackett et al. 2008; Jetz et al. 2012; McCormack et al. 2013). Recent analyses have focused on teasing out the timing of lineage formation using calibration with fossils or other information. Naturally sampling has been directed at representation of maximum putative taxonomic diversity, especially at the level of orders, and within this, families. A curious artefact of this approach is a sampling bias reflecting not biology but researcher location. For instance, in the analysis of Pacheco et al. (2011) there are many New Zealand birds at the tips of long branches. New Zealand birds are included as representatives of four orders; Strigiformes (owls), Psittaciformes (parrots), Coraciiformes (rollers and their relatives) and Passeriformes (song birds), and three of these represent lineages estimated to have diverged before the K/Pg boundary (Fig. 2).

On the face of it, this is exciting evidence that New Zealand harbours ancient bird lineages that could be seen as consistent with the hypothesis that the continen-



**Fig. 2** New Zealand birds on long branches. Part of the mitogenomic phylogeny of modern birds redrawn from Pacheco et al. (2011), featuring clades arising from the deepest nodes in the tree. The New Zealand species are indicated on the relevant branches; Morepork/Ruru (*Ninox novaeseelandiae*) orange, kākāpō (*Strigops habroptilus*) green, NZ sacred Kingfisher/Kōtare (*Todiramphus sanctus*) blue, rifleman/ tītipounamu (*Acanthisitta chloris*) pink. Numbers at nodes are estimated ages in millions of years (Pacheco et al. 2011). Vertical yellow and red dashed lines indicate timing of Gondwana/Zealandia separation and K/Pg boundary respectively (Images © Sabines's Sunbird, Mno1f, Fir0002, digika (respectively) – Wikimedia Commons)

tal crust of New Zealand has maintained deep phylogenetic diversity since isolation of Zealandia from Gondwana (Trewick et al. 2007; Landis et al. 2008). However, including more bird species in the analysis and information about the distributions of closely related species (within the same genus and family) refutes an inference of Gondwana origin for most of these. For example morepork/ruru (*Ninox novaeseelandiae*) and NZ sacred kingfisher/kōtāre (*Todiramphus sanctus*) are species also found outside of New Zealand (Trewick and Gibbs 2010; Goldberg et al. 2011). In further analyses, the rifleman/ tītipounamu (*Acanthisitta chloris*) does remain sister to the rest of the Passerine clade but the dates inferred are more recent than plate tectonic separation (~40 MYA, Jarvis et al. 2014). An analytical problem associated with long branches in phylogenetic trees is the tendency for them to be drawn to the basal nodes. This “long branch attraction” is an artefact of repeated nucleotide substitution resulting in character convergence by chance, such that shared derived characters states are not available to counter the effect (see Bergsten 2005). Thus caution is always required when making inferences from long branches that appear to have phylogenetically deep origins.

When biogeographic history is used to calibrate molecular clocks the impression of ancient origins of lineages can be exacerbated. For instance Wright et al. (2008) studied parrot evolution and used the timing of Zealandia/Gondwana breakup (~80 mya) to calibrate their molecular clock analysis. This approach rested on the assumption that continental drift resulted in the origin of the lineage leading to kākāpō (see Crisp et al. 2011). This is an appealing idea because the shared stratigraphic history of Zealandia and Gondwana is well known (Campbell and Hutching 2011), and the kākāpō (*Strigops habroptilus*) shows many derived traits not seen in other parrots (e.g. flightless, lek breeding, nocturnal). As a result of this calibration kākāpō and another native New Zealand parrot genus (*Nestor*) were placed on a branch with its node at about 80 mya, apparently supporting the idea of an ancient New Zealand origin of Strigopoidea (Wright et al. 2008). The reasoning is however circular (Waters and Craw 2006), and the underlying assumption clearly falsified. Wright et al. (2008) themselves noted that some over-sea dispersal of parrot ancestors was required to reconcile all parts of their biogeographic analysis. There is separate direct evidence falsifying the hypothesis that Strigopoidea originated through ancient breakup of Gondwana and Zealandia; the existence of a distinct species of *Nestor* on the geologically young volcanic Norfolk Island (~900 km north of NZ) until European time. Clearly birds in this lineage retained the ability to move substantial distances over water (Christidis and Boles 2008).

More recent analyses using multiple fossil calibrations outside the parrots indicate ancestry of this order (Psittaciformes) is probably more recent than both Gondwana/Zealandia breakup and the K/Pg, (Pacheco et al. 2011; White et al. 2011; Schweizer et al. 2011; Jarvis et al. 2014). Analyses retain the New Zealand Strigopoidea as sister to other extant parrots, but inferences about the timing of evolution of the “unique” traits associated with the tip species (alpine kea, temperate kākā, flightless kākāpō) have little to do with the age of the lineage. Neither the evolution of flightlessness in kākāpō nor the current exclusivity of their phylogenetic branch to New Zealand can be attributed to the base of the lineage; flightless-

ness might have evolved anytime since formation of the lineage and extinction of other members of this lineage that once existed outside New Zealand could have occurred at any time in the past (see Fig. 1a).

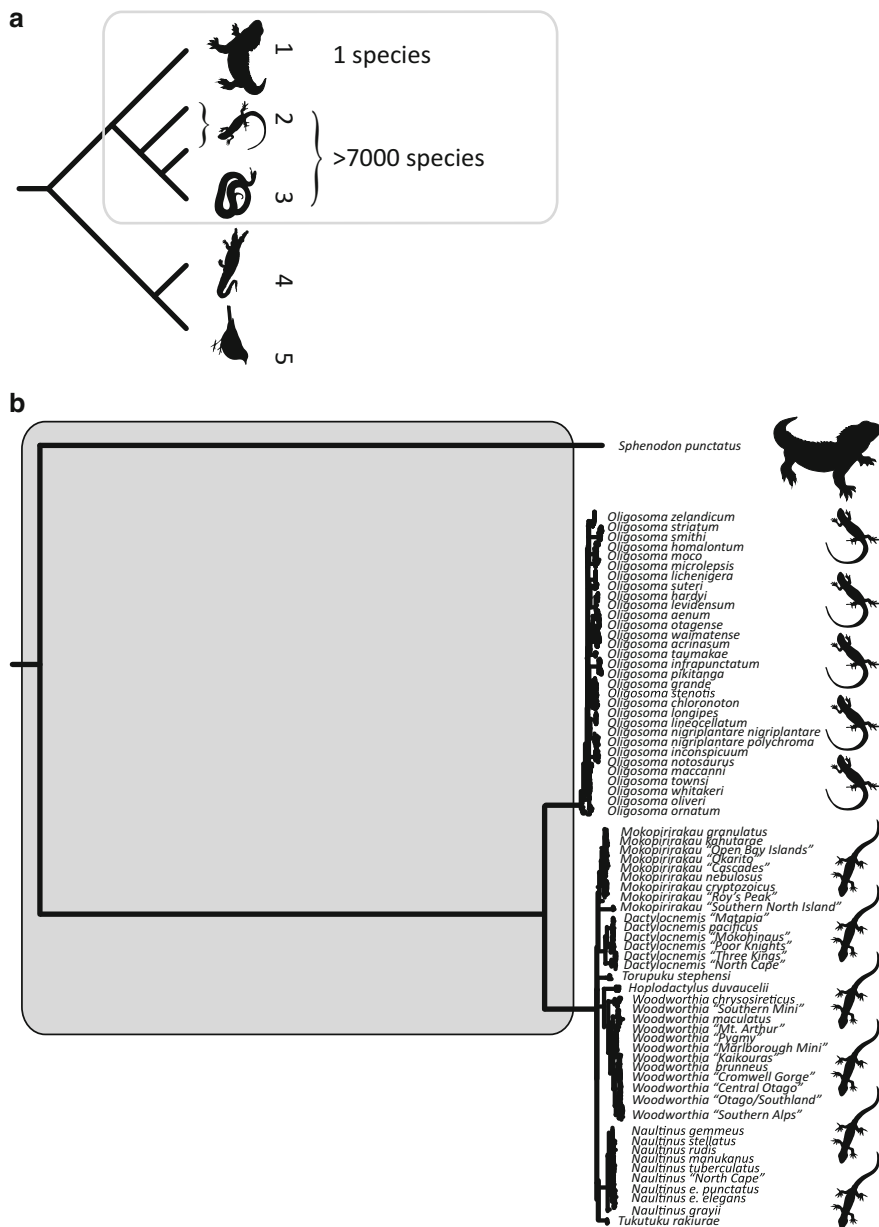
Fossil parrot bones have recently been identified in New Zealand dating to between 16 and 19 million years ago (Worthy et al. 2011). These have some morphological features in common with the genus *Nestor* (kākā and kea) that are not shared with living Australian parrots. There is, however, no available analysis testing the plausibility of alternative systematic classification, and current evidence does not preclude the former existence of *Nestor*-like parrots (Strigopoidae) in past Australia or Antarctica; both are large landmasses that could have supported supposedly New Zealand bird lineages.

Kōtare (NZ Sacred kingfisher) and Morepork/Ruru (owl) at the tips of long branches (Pacheco et al. 2011) can readily be shown to offer spurious information about the New Zealand biota. Both species also occur outside New Zealand, and have many near relatives around the world. Thus, where a lineage is represented by high diversity, the implications of long branches can be readily assessed, but truly sparse lineages (in the extant biota) remain open to question. In contrast, short edges are readily understood, but as morphological and behavioural evolution is not clock-like, species with numerous unusual characteristics might have unexpected close relatives. For example, the extinct New Zealand eagle (*Harpagornis moorei*) was the largest eagle known in the world although it shared a common ancestor with the Australian Little eagle (*Hieraaetus morphonoides*) just a few million years ago (Bunce et al. 2005). Similarly, the takahē (Rallidae, *Porphyrio hochstetteri*) is flightless and the largest of its family, yet is closely related to a common flying species (Trewick 1997; Garcia-R et al. 2014).

## ***On a Reptilian Limb***

The native New Zealand biota includes only two lineages of scaled reptiles (Squamata), diplodactylid geckos and lygosomine skinks, but it also harbours one other lepidosaurian lineage that is missing from the rest of the world (Rhynchocephalia) (Fig. 3a). So although only two of the four reptilian orders are found in New Zealand, the diversity does span an unparalleled phylogenetic scale for this group of vertebrates. Furthermore, New Zealand species diversity is high but only in some parts of the tree (Fig. 3b).

The tuatara (*Sphenodon punctatus*) is very clearly out on a phylogenetic limb and naturally this has resulted in much research interest on its ecology (Towns et al. 2007; Mitchell et al. 2010), reproduction (Cree et al. 1992; Cree et al. 1995; Miller et al. 2009), parasites, immunology (Miller et al. 2007; Godfrey et al. 2010), phylogeography (Hay et al. 2009) and conservation (Daugherty et al. 1990). The node uniting *Sphenodon* with the geckos and skinks may date to Triassic time (>200 mya), although that does not mean that geckos or skinks or *Sphenodon* originated then. In terms of phylogenetic sampling, molecular data for New Zealand lepidos-



**Fig. 3** **a** Cladogram of Lepidosauria (grey box outline) and related lineages 1. Rhynchocephalia; 2. Lizards; 3. Snakes; 4. Crocodiles; 5. Birds. **b** Phylogenetic tree for all New Zealand Lepidosauria inferred using Maximum Likelihood analysis of mtDNA ND2 sequences from Genbank with GTR model (RAxML- Stamatakis 2006) comprising the single extant species of Rhynchocephalia (Hay et al. 2009), ~36 species of gecko (Nielsen et al. 2011), and ~40 species of skink (Chapple et al. 2009). The grey filled box indicates the region of the tree that provides no phylogenetic or biogeographic information about the lineages

sauria is rich, with substantial studies that have drawn on many years of expert searching (Chapple et al. 2009; Nielsen et al. 2011). There is little likelihood that major lineages are missing from the analysis through failed sampling. The sampling of likely relatives from outside New Zealand is also probably now sufficient to provide confidence that the gecko and skink radiations are each monophyletic and endemic.

As a representative of a lineage (Rhynchocephalia) that has been otherwise pruned out, *Sphenodon* does have important conservation status because a single species extinction would result in the loss of the entire lineage, not just in New Zealand but across the globe. In contrast, New Zealand skinks and geckos would have to undergo extinction of numerous species before their respective stem lineages were lost, and even then they would be lost only in New Zealand; related skinks and geckos elsewhere would retain the evolutionary potential of the group.

But this sort of thing must have been happening since the dawn of life on earth, and in terms of taxonomic, biogeographic, ecological, and metabolic diversity, the rhynchocephalids are already extinct. *Sphenodon* is sadly a museum piece that tells us as little about evolution of reptiles as it tells us about New Zealand biogeography. *Sphenodon* does say something about the influence of extinction on biodiversity but gives only a tentative hint of the role of natural selection in this process. The global demise of rhynchocephalia (Jones 2008) corresponds with diversification of squamates, and though it is tempting to see evolutionary cause and effect, there is currently no strong evidence for this (Evans and Jones 2010). However, in New Zealand, extant geckos and skinks appear to have diversified from the Miocene onwards (Chapple et al. 2009; Nielsen et al. 2011), whereas *Sphenodon* did not (or lost diversity as fast as it gained it). Even though there is tantalising evidence that an ancestor of the tuatara might have been present in New Zealand in the Miocene (Jones et al. 2009), there is no evidence for *Sphenodon* diversification. Even amongst extensive Holocene fossils, that include representatives of many vertebrate species extinguished soon after arrival of humans, there is no additional *Sphenodon* diversity (Hay et al. 2008).

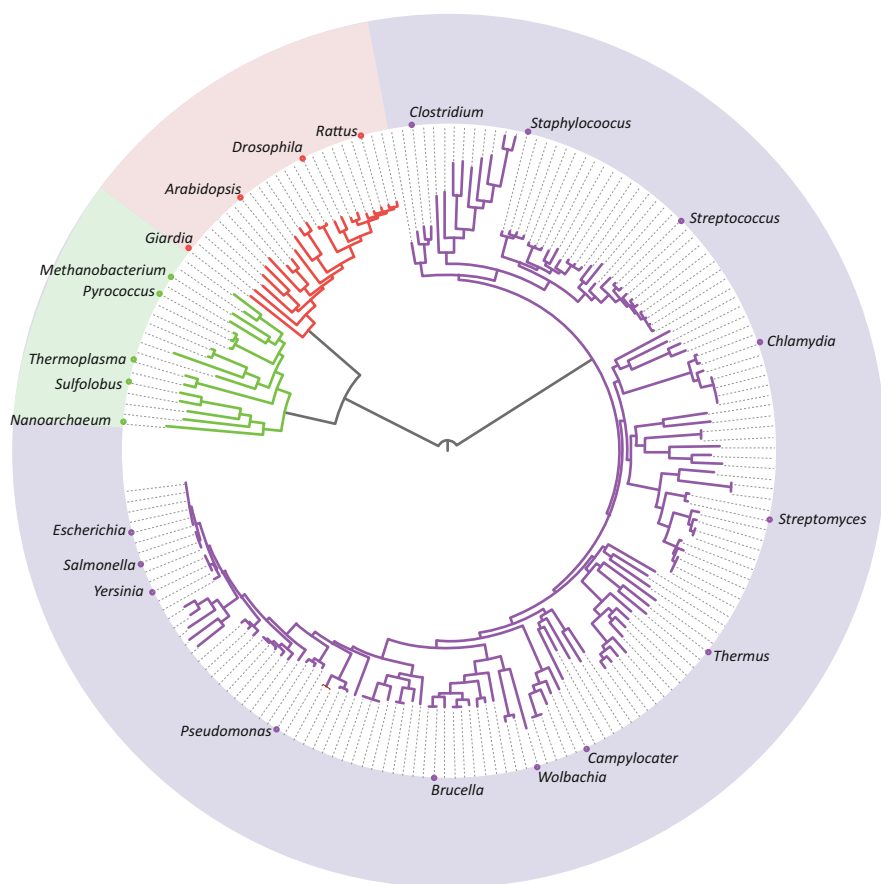
Because it is already rare and restricted to habitat-modified offshore islands, *Sphenodon* conservation does not capture broad ecosystem diversity, although it is host to an endangered species of tick (Miller et al. 2007). Conversely, the gecko and skink lineages occupy diverse habitats in forested and open situations from coast-line to alpine zone; preservation of either or both of those lineages would result in conservation of ecological diversity across New Zealand. New Zealand geckos are biologically interesting because of their viviparous mode of reproduction and ability to occupy alpine habitat; traits that are unique to the lineage and thus of conservation significance.

## Long Branches and Phylogenetic Diversity

A measure of a species' expected contribution to genetic, or evolutionary distinctiveness is derived from its position in a phylogeny that can be used to place a value on that taxon (see chapters in this book). And as all life on earth has a common



ancestor (Darwin 1859; Theobald 2010), we can consider the phylogenetic value of any species in the context of the whole (Fig. 4). This view of life based on DNA sequences of full genomes reveals that phylogenetic diversity is dominated by microscopic organisms and conservation of any visible life (fungi, plants, animals) preserves very little evolutionary distinctiveness (Fig. 4; Ciccarelli et al. 2006). Thus, as a starting point in the application of phylogenetics to conservation we should put equal resources into maintaining diversity within the three major lineages (or long branches): Bacteria, Archaea and Eukarya. However, the only species we know sufficiently well to recognise a decline and have knowledge to remedy are eukaryotes. In addition, it is the habitats provided by multicellular organisms that



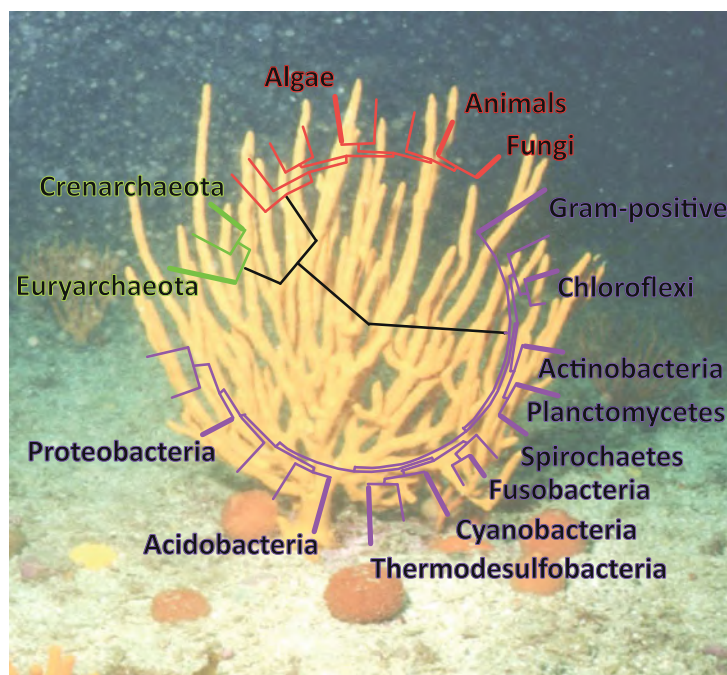
**Fig. 4** Phylogenetic diversity on Earth is dominated by microscopic organisms, as revealed by the tree of life based on 31 universal protein families (Redrawn from Ciccarelli et al. 2006). Branch lengths give an indication of the extent of diversity and lineage age. Note the very shallow branches among popular large creatures (*red* *clade*). Some of the more widely known microbes are labelled but every branch represents a distinct taxon

we have invested resources into studying; habitats that host numerous interesting lineages of Bacteria and Archaea (Eckburg et al. 2005).

Thousands of low-abundance taxa account for most of the observed phylogenetic diversity in any environment. This “rare biosphere” contains a large proportion of phylogenetic diversity and represents an enormous contribution to genetic distinctiveness and evolutionary innovation (Sogin et al. 2006; Nee 2004a). After Anton von Leeuwenhoek first looked at bacteria in lake water and material scraped from his teeth in the seventeenth century, our understanding and appreciation of the distribution and abundance of microorganisms advanced relatively slowly. It is now accelerating rapidly as technological developments allow us to obtain and analyse large amounts of DNA data directly from environmental samples containing large numbers of taxa (Lozupone and Knight 2008). Indeed the current state of technology means that microbial genomes are tractable objects for whole genome sequencing. We will soon know whether the 4957 bacterial taxa found in soil of a commercial apple orchard (Shade et al. 2012) is species rich (but phylogenetically restricted) compared to a marine plankton net sample with 189 species of zooplankton (Machida et al. 2009), or human skin with more than 205 species of bacteria from 19 phyla (Grice et al. 2009). Microbial phenotype arrays allow the gathering of far more precise ecological detail about bacteria than is available for eukaryotes (Bochner 2008). There is also emerging evidence of additional fundamental types of life on Earth (Zakaib 2011).

As an example of the known unknowns, consider New Zealand sponges. Sponges are multicellular (visible) marine animals of the phylum Porifera. In coastal water around New Zealand 733 species of sponges have been recorded from 20 orders (Kelly et al. 2006). As with much of the New Zealand fauna (see Trewick and Morgan-Richards 2009), about 95 % of these are endemic to the region at the species level. However, in themselves these species contribute little directly to global diversity because other closely related species exist elsewhere. Generally sponges are not endangered, although special regions of high diversity that exist in hydrothermal areas and on seamounts are under pressure from benthic trawling (Kelly et al. 2006, and see Gianni 2004).

Nevertheless conservation of any sponge species or even population contributes much more; sponges are home to distinct microbial communities (microbiomes) so the total number of phyla preserved might reach more than 40. Sponges host rich microorganism communities and with next generation DNA sequencing data the number of known bacterial phyla in sponges has recently increased (Webster et al. 2010; Schmitt et al. 2012). Although many of the detected phyla are formally described, such as the Algae, Fungi, Actinobacteria, Chloroflexi (Green non-sulfur bacteria), Cyanobacteria, Nitrospira, and Proteobacteria (Fig. 5), several new ones have also been discovered in sponges (Turque et al. 2010; Webster et al. 2010; Schmitt et al. 2012). A single sponge provides an environment that protects an impressive array of phylogenetic diversity (Taylor et al. 2007). So how can we best conserve the phylogenetic diversity harboured inside sponges? Will one species or one geographic region suffice?



**Fig. 5** Each sponge is home to a community of microscopic life that encompasses the range of known phylogenetic diversity on Earth. Here the major phyla found within a sponge microbiome are named on the tree of life. The sponge pictured is *Raspailia topsenti*, one of five sympatric sponge species studied by Schmitt et al. (2011) from New Zealand coastal waters (Phylogeny redrawn from Ciccarelli et al. 2006. Image © Katie Dowle)

Samples from eight locations around the world detected 2567 bacterial taxa representing 22 phyla living inside sponges (Schmitt et al. 2012), while three species of Australian sponge held a total of 2996 bacteria taxa from 36 phyla (Webster et al. 2010). Different sponge species from the same environment possess distinct symbiotic communities. Some components of their bacterial communities appear to be passed from parent to offspring while other components are acquired from the surrounding seawater (Webster et al. 2010; Schmitt et al. 2012). Thus, although a few bacteria are found in all sponges the majority are either host or region specific. For example tropical sponges have microbial communities that are more similar to each other than to the communities in subtropical sponges.

Schmitt et al. (2011) collected five sponge species from a single bay on the coast of New Zealand. By focusing on just the bacteria that are members of the phylum Chloroflexi they compared species diversity between sponges with either high or low microbial abundance, and contrasted this with Chloroflexi diversity in the surrounding seawater. Fifty-eight species of Chloroflexi were recorded from inside the sponges, but only three species in the seawater (Schmitt et al. 2011). About half these taxa were new to science. Ecologically important roles and specific associa-

tions of Chloroflexi bacteria were inferred for the sponge species with high microbial abundance as the majority of their bacteria fell into sponge-specific and sponge-coral phylogenetic lineages (Schmitt et al. 2011). Thus any single sponge species houses plenty of phylogenetic diversity but if we want to conserve all lineages that are restricted to sponges, we need to conserve more than one sponge species.

## Phylogenetic Extremities

Is it feasible to prioritize for conservation the phylogenetic extremities of life? In fact there is no need to because microbes and peculiar multicellular organisms such as kākāpō, takahē and tuatara are intimately linked. A kākāpō could not function if it and its physical environment were stripped of all microbes. In this respect Kākāpō, like marine sponges are loose assemblies of disparate genomes. Ecosystem function tends to be viewed at the macroscopic scale, but this is only because the tools to visualise the much more extensive and complex underworld have only recently become available. Most, if not all, ecosystem processes are mediated by microscopic life.

## Conclusions

Kākāpō do not need a long phylogenetic branch (thought they are on one) to justify their conservation; their distinctive ecological traits are sufficient reason to protect them. However, it is readily demonstrated, if not quantified, that an environment capable of sustaining a viable population of this species would also sustain many other taxa from soil bacteria to trees. Similarly, takahē (Aves, Rallidae, *Porphyrio*) deserve protection because of their unusual ecological traits representing evolutionary adaptations lost elsewhere in the world through recent extinction, though takahē are on a much shorter branch from their shared ancestor with a common living species, than is the kākāpō. Species' radiations such as geckos need a quite different strategy that does not rely on long-branch status, to maintain their diversity, unusual traits and associated communities. However, conservation of the substantive part of diversity of life on Earth will benefit from next generation sequencing and emerging bioinformatics tools that can identify assemblages of deeply divergent lineages within definable, manageable biomes. Microbiomes are not well understood, and therefore we are not well placed to determine which environments are home to the greatest phylogenetic diversity. Until we have comparative data, we must strive to maximize retention of ecosystem diversity on Earth, from human guts to forest soils, parrot feathers to rocky shores. To maximise conservation of evolutionary potential on Earth we need to pay more attention to our planet's microbial diversity

and in doing so maintain ecosystem process to the benefit of the large appealing species that are so popular.

*Organic life beneath the shoreless waves  
Was born and nurs'd in ocean's pearly caves;  
First forms minute, unseen by spheric glass,  
Move on the mud, or pierce the watery mass;  
(From The Temple of Nature. Erasmus Darwin 1802)*

**Acknowledgements** We are grateful for the assistance of Ananias Escalante for access tree image from Pacheco et al. 2011. Phoenix group members (evolves.massey.ac.nz) provided illuminating discussion. The manuscript was improved by comments from two reviewers and the editors. The authors were assisted by the Marsden Fund (Grant no. 12-MAU-008), administered by the Royal Society of New Zealand.

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Avice JC (2005) Phylogenetic units and currencies above and below the species level. In: Purvis A, Gittleman JL, Brooks T (eds) *Phylogeny and conservation*. Cambridge University Press, Cambridge, pp 76–100
- Bergsten J (2005) A review of long-branch attraction. *Cladistics* 21:163–193
- Bochner BR (2008) Global phenotypic characterization of bacteria. *FEMS Microbiol Rev* 33:191–205
- Bromham L, Penny D (2003) The modern molecular clock. *Nat Rev Genet* 4:216–224
- Bunce M et al (2005) Ancient DNA provides new insights into the evolutionary history of New Zealand's extinct giant eagle. *PLoS Biol* 3(1):e9
- Campbell H, Hutching G (2011) In search of ancient New Zealand. Penguin Group, Auckland
- Chapple DG, Ritchie PA, Daugherty CH (2009) Origin, diversification, and systematics of the New Zealand skink fauna (Reptilia: Scincidae). *Mol Phylogenet Evol* 52:470–487
- Christidis L, Boles WE (2008) *Systematics and taxonomy of Australian birds*. CSIRO Publishing, Canberra, p 200
- Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P (2006) Toward automatic reconstruction of a highly resolved tree of life. *Science* 311:1283–1287
- Cree A, Cockrem JF, Guillelte LJ Jr (1992) Reproductive cycles of male and female tuatara (*Sphenodon punctatus*) on Stephens Island, New Zealand. *J Zool Lond* 226:199–217
- Cree A, Thompson MB, Daugherty CH (1995) Tuatara sex determination. *Nature* 375:543
- Crisp M, Trewick SA, Cook LG (2011) Hypothesis testing in biogeography. *Trends Ecol Evol* 26:66–72
- Cullen R (2012) Biodiversity protection prioritization: a 25 year review. *Wildl Res* 40:108–116
- Darwin C (1859) *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*, 1st edn. John Murray, London
- Daugherty CH, Cree A, Hay JM, Thompson MB (1990) Neglected taxonomy and continuing extinction of tuatara (*Sphenodon*). *Nature* 347:177–179
- Daugherty CH, Gibbs GW, Hitchmough RA (1993) Mega-island or micro-continent? New Zealand and its fauna. *Trends Ecol Evol* 8:437–442

- DOC (2000) The New Zealand biodiversity strategy. Department of Conservation (DOC); Ministry for the Environment (MfE), Wellington, p 163
- Eckburg PB, Bik EM, Bernstein CN et al (2005) Diversity of the human intestinal microbial flora. *Science* 308(5728):1635–1638
- Evans SE, Jones MEH (2010) The origin, early history and diversification of lepidosauromorph reptiles. Chapter 2 in *New aspects of mesozoic biodiversity, Lecture notes in earth sciences*, vol 132, Springer, pp 27–44
- García-R JC, Gibb G, Trewick SA (2014) Deep global evolutionary radiation in birds: diversification and trait evolution in the cosmopolitan bird family Rallidae. *Mol Phylogenet Evol* 81:96–108
- Gianni M (2004) High seas bottom trawl fisheries and their impacts on the biodiversity of vulnerable deep-sea ecosystems. IUCN The World Conservation Union, Gland
- Gibbs GC (2006) Ghosts of Gondwana—the history of life in New Zealand. Craig Potton, Nelson
- Gibbs GC, Penny D (2010) Two aspects along the continuum of pigeon evolution: a South-Pacific radiation and the relationship of pigeons within Neoaves. *Mol Phylogenet Evol* 56:698–706
- Godfrey SS, Moore JA, Nelson NJ, Bull CM (2010) Social network structure and parasite transmission in a territorial reptile, the tuatara (*Sphenodon punctatus*). *Int J Parasitol* 40:1575–1585
- Goldberg J, Trewick SA, Powlesland RG (2011) Population structure and biogeography of *Hemiphaga* pigeons (Aves: Comumbidae) on islands in the New Zealand region. *J Biogeogr* 38:285–298
- Gould SJ, Eldredge N (1993) Punctuated equilibrium comes of age. *Nature* 366:223–227
- Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC, NISC Comparative Sequencing Program, Bouffard GG et al (2009) Topographical and temporal diversity of the human skin microbiome. *Science* 324(5931):1190–1192
- Hackett SJ, Kimball RT, Reddy S et al (2008) A phylogenomic study of birds reveals their evolutionary history. *Science* 320:1763–1768
- Hay JM, Subramanian S, Miller CD, Mohandesan E, Lambert DM (2008) Rapid molecular evolution in a living fossil. *Trends Genet* 24:106–109
- Hay JM, Sarre SD, Lambert DM, Allendorf FW, Daugherty CH (2009) Genetic diversity and taxonomy: a reassessment of species designation in tuatara (*Sphenodon*: Reptilia). *Conserv Genet* 11:1063–1081
- Jarvis ED et al (2014) Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346:1320–1331
- Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO (2012) The global diversity of birds in space and time. *Nature* 491:444–448
- Jones MEH (2008) Skull shape and feeding strategy in *Sphenodon* and other Rhynchocephalia (Diapsida: Lepidosauria). *J Morphol* 269:945–966
- Jones MEH, Tennyson AJD, Worthy JP, Evans SE, Worthy TH (2009) A sphenodontine (Rhynchocephalia) from the Miocene of New Zealand and palaeobiography of the tuatara (*Sphenodon*). *Proc R Soc* 76:1385–1390, London B
- Kelly M, Edwards AR, Wilkinson MR, Alvarez B, de Cook SC, Bergquist PR, Buckeridge JS, Campbell H, Reiswig HM, Valentine C (2006) Phylum porifera sponges. In: Gordon DP (ed) *The New Zealand Inventory of Biodiversity. Volume 1. Kingdom Animalia: Radiata, Lophotrochozoa, and Deuterostomia*. Canterbury University Press, Christchurch
- Landis CA, Campbell HJ, Begg JG, Mildenhall DC, Paterson AM, Trewick SA (2008) The waipounamu erosion surface: questioning the antiquity of the New Zealand land surface and terrestrial fauna and flora. *Geol Mag* 145:173–197
- Lee WG, Allen RB, Tompkins DM (2006) Paradise lost – the last major colonization. In: Allen RB, Lee WG (eds) *Biological invasions in New Zealand*, vol 186. Springer, Berlin, pp 1–13
- Lozupone C, Knight R (2008) Species divergence and the measurement of microbial diversity. *FEMS Microbiol Rev* 32(4):557–578

- Machida RJ, Hashiguchi Y, Nishida M, Nishida S (2009) Zooplankton diversity analysis through single-gene sequencing of a community sample. *BMC Genomics* 10:438
- Margules CR, Pressey RL (2000) Systematic conservation planning. *Nature* 405:243–253
- McCormack JE, Harvey MG, Faircloth BC, Crawford NG, Glenn TC et al (2013) A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. *PLoS One* 8(1):e54848. doi:[10.1371/journal.pone.0054848](https://doi.org/10.1371/journal.pone.0054848)
- Miller HC, Conrad AM, Barker SC, Daugherty CH (2007) Distribution and phylogenetic analyses of an endangered tick, *Amblyomma sphendonti*. *N Z J Zool* 34(2):97–105
- Miller HC, Moore JA, Nelson NJ, Daugherty CH (2009) Influence of MHC genotype on mating success in a free-ranging reptile population. *Proc R Soc Lond B* 276:1695–1704
- Mitchell NJ, Allendorf FW, Keall SN, Daugherty CH, Nelson NJ (2010) Demographic effects of temperature-dependent sex determination: will tuatara survive global warming? *Glob Chang Biol* 16:60–72
- Morgan-Richards M, Trewick SA, Bartosch-Härlid A, Kardailsky O, Phillips MJ, McLenachan PA, Penny D (2008) Bird evolution: testing the Metaves clade with six new mitochondrial genomes. *BMC Evol Biol* 8:20
- Nee S (2004a) More than meets the eye: earth's real biodiversity is invisible. *Nature* 429:804–805
- Nee S (2004b) Extinction, slime and bottoms. *PLoS Biol* 2(8):e272
- Nielsen SV, Bauer AM, Jackman TR, Hitchmough RA, Daugherty CH (2011) New Zealand geckos (Diplodactylidae): cryptic diversity in a post-Gondwanan 3 lineage with trans-Tasman affinities. *Mol Phylogenet Evol* 59:1–22
- Pacheco MA, Battistuzzi FU, Lentino M, Aguilar RF, Kumar S, Escalante AA (2011) Evolution of modern birds revealed by mitogenomics: timing the radiation and origin of major orders. *Mol Biol Evol* 28:1927–1942
- Penny D, Phillips MJ (2004) The rise of birds and mammals: are microevolutionary processes sufficient for macroevolution? *Trends Ecol Evol* 19:516–522
- Pratt RC, Gibbs GC, Morgan-Richards M, Phillips MJ, Hendy MD, Penny D (2009) Towards resolving deep Neoaves phylogeny: data, signal enhancement and priors. *Mol Biol Evol* 26(2):313–326
- Purvis A, Gittleman JL, Brooks T (2005) *Phylogeny and conservation*. Cambridge University Press, Cambridge
- Rolland J, Cadotte MW, Davies J, Devictor V, Lavergne S et al (2011) Using phylogenies in conservation: new perspectives. *Biol Lett* 8:692–694
- Schmitt S, Deines P, Behnam F, Wagner M, Taylor MW (2011) Chloroflexi bacteria are more diverse, abundant, and similar in high than in low microbial abundance sponges. *FEMS Microbiol Ecol* 78:497–510
- Schmitt S, Tsai P, Bell J, Fromont J, Ilan M, Lindquist N, Perez T, Rodrigo A, Schupp PJ, Vacelet J, Webster N, Hentschel U, Taylor MW (2012) Assessing the complex sponge microbiota – core, variable, and species-specific bacterial communities in marine sponges. *ISME J* 6:564–576
- Schweizer M, Seehausen O, Hertwig ST (2011) Macroevolutionary patterns in the diversification of parrots: effects of climate change, geological events and key innovations. *J Biogeogr* 38:2176–2194
- Shade A, Hogan CS, Klimowicz AK, Linske M, McManus PS, Handelsman J (2012) Culturing captures members of the soil rare biosphere. *Environ Microbiol* 14:2247–2252
- Slack KE, Delsuc F, McLenachan PA, Arnason U, Penny D (2007) Resolving the root of the avian mitogenomic tree by breaking up long branches. *Mol Phylogenet Evol* 42:1–13
- Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM, Herndl GJ (2006) Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proc Natl Acad Sci U S A* 103:12115–12120
- Stamatakis A (2006) RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690



- Taylor MW, Radax R, Steger D, Wagner M (2007) Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiol Mol Biol Rev* 71:295–347
- Theobald DL (2010) A formal test of the theory of universal common ancestry. *Nature* 465(7295):219–222
- Towns DR, Parrish GR, Tyrrell CL, Ussher GT, Cree A, Newman DG, Whitaker AH, Westbrooke I (2007) Responses of tuatara (*Sphenodon punctatus*) to removal of introduced Pacific rats from islands. *Conserv Biol* 21:1021–1031
- Trewick SA (1997) Flightless NESS and phylogeny amongst endemic rails (Aves: Rallidae) of the New Zealand region. *Philos Trans R Soc Lond B* 352:429–446
- Trewick SA, Gibbs GC (2010) Vicars, tramps and assembly of the New Zealand avifauna: a review of molecular phylogenetic evidence. *Ibis* 152:226–253
- Trewick SA, Morgan-Richards M (2009) New Zealand biology. In: Gillespie R, Clague E (eds) *Encyclopedia of islands*. University of California Press, Berkeley
- Trewick SA, Paterson AM, Campbell HJ (2007) Hello New Zealand. *J Biogeogr* 34:1–6
- Turque AS, Batista D, Silveira CB, Cardoso AM, Vieira RP et al (2010) Environmental shaping of sponge associated archaeal communities. *PLoS One* 5(12):e15774. doi:[10.1371/journal.pone.0015774](https://doi.org/10.1371/journal.pone.0015774)
- Vaux F, Trewick SA, Morgan-Richards M (2015). Lineages, splits and divergence challenge whether the terms anagenesis and cladogenesis are necessary. *Biol J Linnean Soc.* doi:[10.1111/bij.12665](https://doi.org/10.1111/bij.12665)
- Walker S, Stephens RTT, Overton JMC (2012) A unified approach to conservation prioritisation, reporting and information gathering in New Zealand. *N Z J Ecol* 36:243–251
- Waters JM, Craw D (2006) Goodbye Gondwana? New Zealand biogeography, geology, and the problem of circularity. *Syst Biol* 55:351–356
- Webster NS, Taylor MW, Behnam F, Luecker S, Rattei T, Whalan S, Horn M, Wagner M (2010) Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. *Environ Microbiol* 12:2070–2082
- White NE, Phillips MJ, Gilbert MTP, Alfaro-Nunez A, Willerslev E, Mawson PR, Spencer PBS, Bunce M (2011) The evolutionary history of cockatoos (Aves: Psittaciformes: Cacatuidae). *Mol Phylogenet Evol* 59:615–622
- Worthy TH, Tennyson AJD, Scofield RP (2011) An early Miocene diversity of parrots (Aves, Strigopidae, Nestorinae) from New Zealand. *J Vertebr Paleontol* 31:1102–1116
- Wright TF, Schirtzinger EE, Matumoto T, Eberhard JR et al (2008) A multilocus molecular phylogeny of the parrots (Psittaciformes): support for a Gondwanan origin during the cretaceous. *Mol Biol Evol* 25:2141–2156
- Zakaib GD (2011) The challenge of microbial diversity: out on a limb. *Nature* 476:20–21



# What Is the Meaning of Extreme Phylogenetic Diversity? The Case of Phylogenetic Relict Species

Philippe Grandcolas and Steven A. Trewick

**Abstract** A relict is a species that remains from a group largely extinct. It can be identified according both to a phylogenetic analysis and to a fossil record of extinction. Conserving a relict species will amount to conserve the unique representative of a particular phylogenetic group and its combination of potentially original characters, thus lots of phylogenetic diversity. However, the focus on these original characters, often seen as archaic or primitive, commonly brought erroneous ideas. Actually, relict species are not necessarily old within their group and they can show as much genetic diversity as any species. A phylogenetic relict species can be geographically or climatically restricted or not. Empirical studies have often shown that relicts are at particular risks of extinction. The term relict should not be used for putting a misleading emphasis on remnant or isolated populations. In conclusion, relict species are extreme cases of phylogenetic diversity, often endangered and with high symbolic value, of important value for conservation.

**Keywords** Geological extinction • Genetic diversity • Species age • Endemism • Remnant

## Introduction

Why does phylogenetic diversity (or evolutionary distinctiveness) dramatically matter for biodiversity conservation? The answer to this question first posed by Vane-Wright et al. (1991) and Faith (1992) is often illustrated with examples of

---

P. Grandcolas (✉)

Institut de Systématique, Evolution, Biodiversité, ISYEB – UMR 7205 CNRS MNHN  
UPMC EPHE, Muséum National d'Histoire Naturelle, Sorbonne Universités,  
45 rue Buffon, CP 50, 75005 Paris, France  
e-mail: [pg@mnhn.fr](mailto:pg@mnhn.fr)

S.A. Trewick

Ecology Group, Institute of Agriculture and Environment, Massey University,  
Private Bag 11-222, Palmerston North, 4442 New Zealand  
e-mail: [s.trewick@massey.ac.nz](mailto:s.trewick@massey.ac.nz)

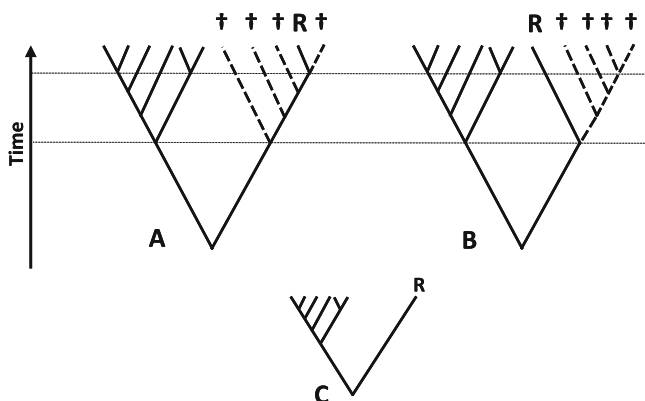
emblematical and unique species. Such exemplar species that speak to everyone from layperson to scientist, include the Coelacanth fish, the Tuatara squamate, the Kiwi bird, the *Platypus* mammal, the Ginkgo tree, etc. All these species are said to be relict, because they represent groups that are mostly extinct (Grandcolas et al. 2014). The message is that these species should be cared for, because their extinction would cause a loss of information about distinct sections of life on Earth and their evolution. Generally, this powerful message is naively extended to characterize the place where these species are found, implying that the biota as a whole is a kind of Noah's ark, globally worthy of consideration for conservation biology (see for example, Gibbs 2006 for the case of New Zealand, or Thorne 1999 for Asia).

To our knowledge, everyone agrees with these views and even the most hard-hearted companies or governments would difficulty take responsibility for destroying such emblematical "survivors". The public message in endorsing this destruction would be that they are the fools that spoil unique multimillion year antiques, even worse than to break a Vase de Soissons into thousands of pieces or to lacerate a delicate and wonderfully conserved Da Vinci painting. Even if very consensual, such emotive views about relicts and biodiversity conservation are still often presented informally, which prevent them to be fully scientific, i.e. theoretically justified, measurable and repeatable.

If then we try to set aside the emotional aspects of these views about relicts, what remains for conservation biology as a rational argument? Do relicts actually represent invaluable species for conservation purposes and why? Are they particularly exotic cases that do not account for most situations encountered by land managers or are they extreme cases of common situations? To answer these questions, we need to carefully define relicts with phylogenetic and paleontological tools. The properties of such characterizations need to be explored regarding the most important issues in conservation biology.

## What Then, Is a Relict Species?

By definition, a relict is something that remains from an entity that has mostly disappeared (Merriam-Webster 2014; Lincoln et al. 1982). In evolutionary biology, a relict species remains of a group that is mainly extinct (Grandcolas et al. 2014; Fig. 1). The basis for this inference is the observation that a species stands alone on a long phylogenetic branch, by comparison with a larger sister-group, because of extinctions that occurred since the emergence of the stem group (Fig. 1). Formally, identifying a relict species requires comparison of sister-groups with different species numbers and characterization of extinction rates using phylogenetic tools on molecular trees (e.g., Ricklefs 2007; Rabosky 2006). This is the notion of phylogenetic relict species is distinct from geographical or environmental or climatic relict species where the relict state is defined according to spatial restriction supposedly arising from extinction of relatives in other parts of the geographical or ecological space (Habel and Assmann 2010; Hampe and Jump 2011). Here we will focus on



**Fig. 1** Two different clades with a relict species “R” remaining after species extinctions (†) in the *right* part of the clade. In the clade on the *left* (a), the relict is among the most recent species as indicated by the position on the time axis (dotted line) while in the clade on the *right* (b), the relict is among the most ancient species. It must be reminded that in most cases with lack of fossil record, the clade would look like the third one on the *bottom* of the figure (c), with the relict species alone on a long branch whose age is difficult to evaluate (c is like clade a or b)

the concept of phylogenetic relict species that is more relevant to phylogenetic diversity. We then briefly consider the concept of geographic or climatic relict species to pinpoint when it has some added value for conservation purposes. These geographical or climatic relict populations could be better called “remnants” in order to avoid confusion with relict species.

When dealing with presumptive relicts in a tree, it is first necessary to check that such long phylogenetic branches are not artifacts generated by problems of tree construction. The most common problem of this kind is long-branch attraction when the analysis of molecular data tends to draw together long branches because of sampling deficiency or fast-evolving molecular markers (Bergsten 2005). A species may be placed onto such an artificially long branch when the inference procedure does not find closely related species, either because they are lacking in the taxon sample of the analysis or because selected DNA sequences have diverged much faster, erasing the information of relatedness. Such naked branches frequently become artificially long because they fall to the base of the reconstructed tree. This problem can look trivial but could occur more and more frequently when phylogenetic analyses are performed at community level within the framework of metagenomics: local and community-focused sampling will not necessarily ensure a reasonable taxonomic coverage and could generate more artifacts than traditional and taxonomy-focused phylogenetic studies.

Second, two theoretical cases have been distinguished among phylogenetic relict species: species that survived an extinction event depleting their group and species belonging to groups that never speciated much (Table 1). Simpson (1944) named them numerical and phylogenetic relicts, respectively. Actually, real situations are inevitably a mix of these two theoretical cases; even in small clades, the relict

**Table 1** Theoretical characteristics of the different kinds of relicts with reference to the evolutionary process involved, the criterion of characterization and the origin of the deep and long branch. Any real situation is actually a combination of the two first theoretical cases and of the third one to different extents. The third case, the geographical or climatic relict, is not necessarily a relict *sensu stricto* but merely a remnant, if not positioned on a deep long branch

Kinds of relict	Evolutionary process	Criterion	Deep long branch
“Numerical” relict	Extinction	Fossil record	Built on extinction
“Phylogenetic” relict	Low speciation	Molecular rate	Built on time
Geographical or climatic “relict”	Area restriction	Fossil record or distribution	Not necessarily

remains from a larger group because extinction rates are never totally zero. Estimating the degree to which long branches have been generated by extinction or by evolutionary stasis requires a combination of data from different research fields. A long molecular branch, whatever its origin, will be most often diagnosed as the result of extinction by methods of “lineage through time” plots (Ricklefs 2007; Quental and Marshall 2010). Paleontological evidence is needed in addition to molecular trees. If a group is known to have been much more speciose in the past, it strongly indicates that the relict actually remains from a much larger and extinct group. From this criterion comes the famous term “living fossil” coined by Darwin (1876) himself: these “like fossils, connect to a certain extent orders at present widely sundered in the natural scale.” Living fossil is however a misleading term because it could lead to the belief that relicts remain globally similar to related fossil taxa through some type of generalized evolutionary stasis (e.g., Eldredge 1987; Eldredge et al. 2005; Parsons 2005). Evolutionary stasis is exceedingly difficult to diagnose since we can always expect to unveil differentiation when we observe more characters in the so-called living fossil and therefore to discard the stasis hypothesis. Actually, none of the classic relicts has ever been found similar to early fossil relatives after closer investigation, therefore refuting the idea of a generalized evolutionary stasis. For example, the venom in *Platypus* is not archaic but totally original, neither squamate nor mammal-type (O’Brien 2008), the coelacanth fish is originally modern in its reproduction mode, being ovoviviparous (Casane and Laurenti 2013). The term “panchronic” (e.g., Janvier 2007) has also been used in this way with the same wrong assumption that relict taxa did not evolve.

Operationally, identifying relict most frequently relies on the phylogenetic criterion because many groups have scanty paleontological records. To what extent this is helpful and meaningful, given the limitations of “lineage through time” plots (Quental and Marshall 2010; Crisp and Cook 2009; Dowle et al. 2013) is unclear. The results obtained in macroevolutionary analyses are always reconstructions from the past, based on incomplete samples and await confirmation by more studies; proposal of a relict species requires a dedicated search for auxiliary evidence for extinction, including an improved fossil record (Grandcolas et al. 2014).

Given this generally accepted definition (e.g., Simpson 1944; Brooks and McLennan 1991), a relict is a species that will show a high phylogenetic diversity score according to any metric (Rodrigues et al. 2005). This is a species that is on a relatively long branch that separated from the remainder of the clade under consideration (i.e. the relict and its sister-group). Therefore, both the position of the species in the phylogenetic topology and the amount of divergence are remarkable. Conserving a relict will contribute to preservation of a species with unique phylogenetic information and with many distinctive (say autapomorphic) characters.

## What a Relict Species Is Not?

Most of the problems with the concept of relicts come from associated concepts that are not formally part of this definition. Because they are survivors, relicts are often misleadingly considered as “missing links”, “living ancestors” or “primitive or basal taxa.” These three last notions are based on a still common misunderstanding of most basic issues in phylogenetics and evolutionary biology (Crisp and Cook 2005). They are based on the fallacious generalization to the whole species of phylogenetic results obtained on very small and biased samples of characters, suggesting that a species would be globally “intermediate” or “primitive.” In a classic case of circular reasoning, a few remarkably “primitive” characters observed in a living species are traditionally considered to have originated very deep in the Tree of Life and are misleadingly considered diagnostic of the globally primitive state of this species and vice versa. The assumption is that searching for other characters would necessarily show that they are also in a primitive state. This assumption is naive because there are no reasons to assume that billions of phenotypic or genomic characters in the same species have all been subject to a global evolutionary stasis. In addition, this assumption has never been empirically met when such species are studied further.

For example, *Mastotermes darwiniensis*, the sister group to all other termites is present today only in Australia but found worldwide in the fossil record. It has profoundly archaic wing venation, egg laying and female genitalia, but it also shows an amazingly derived and multiflagellate spermatozoid (Legendre et al. 2008; Abe et al. 2000). The small tree *Amborella trichopoda* that is endemic to New Caledonia is considered to be the sister group to all flowering plants (Soltis et al. 2002). It has very often been used as a proxy for the ancestral state of many phenotypic traits (e.g., Friedman and Ryerson 2009). However, its mitochondrial genome is amazingly modern and composite, resulting from many horizontal transfers from diverse organisms (Bergthorsson et al. 2004; Rice et al. 2013). There is no organism where all characters are primitive or intermediate like a living ancestor. According to the principles of phylogenetics, it is well recognized that an ancestor with all characters plesiomorphic is therefore by nature paraphyletic and could not be characterized or identified by even only one apomorphy (Nelson 1970; Engelmann and Wiley 1977). In addition, and from a semantic point of view, the term “basal” is nonsensical since

two sister-groups are the same rank, one cannot be basal to the other (Krell and Cranston 2004).

An interesting and neglected characteristic of a relict having survived extinctions is that it is not necessarily a “deep-branching” or “old” species (Fig. 1a); the species could have branched either recently or deeply within a group of which most members of which are already gone (Grandcolas et al. 2014). A molecular phylogenetic study, based only on extant taxa, will not be able to distinguish the age of the species from lineage age (Fig. 1c), unless it permits the discovery of genetic diversity within the crown group species (i.e. several species that were previously confused or several haplotypes within the same species) which will allow the distinction from the stem group. This possibility has been recently illustrated by exemplary studies bearing on famous relict taxa: the coelacanth fishes (Inoue et al. 2005), the cycad plants (Nagalingum et al. 2011), and the gymnosperms as a whole (Crisp and Cook 2011). In these cases, the extant species have been dated as recently differentiated in very old clades that mostly went extinct long ago. Therefore, conserving a relict does not conserve an ancestor or a particular stage of an old evolutionary history but a unique combination of character states representing a larger but mainly extinct group.

## Are Relict Species Evolutionarily Frozen?

We mentioned that taking relicts as living ancestors is an obviously fallacious inference, but this point of view has also been formulated in less exaggerated and misleading terms. For example, relicts have often been considered to have lower evolutionary rates, being in somehow evolutionarily frozen (e.g., Amemiya et al. 2010), which would explain why they did not speciate giving rise to a large group. Parsons (2005) defended the idea that those relicts that live in very specialized and stable niches (e.g., hypersaline biota) would not be subjected to many biotic interactions, preventing any further adaptive change. The same kind of reasoning has been applied to other supposedly narrow niches (Ricklefs 2005), from caves (Gibert and Deharveng 2002; Assmann et al. 2010), deep-sea vents (Van Dover et al. 2002) and oceanic islands (Cronk 1992). The rationale is that the relict is subjected to little diversifying selection in a stable niche, so there is little anagenetic change in the lineage. Darwin (1876: 83–84) himself expressed this for what he called living fossils: “they have endured to the present day, from having inhabited a confined area, and from having been exposed to less varied, and therefore less severe, competition.” Adopting this view, some biologists have questioned the evolutionary value and potential of relicts (e.g., Erwin 1991; Myers and Knoll 2001; Mace and Purvis 2008). Some also doubted the extent to which phylogenetic diversity is an all-purpose criterion to measure the importance of species (Winter et al. 2013): phylogenetic diversity may indicate which species are evolutionarily unique, but does it indicate also which species have evolutionary potential and ability to evolve and to adapt further in a changing world, or both? What use is there for conserving a relict informing about past evolution if it represents the living dead, unable to adapt and soon extinct when facing the next environmental changes?

If all relicts really proved to be frozen or genetically depauperate and unable to evolve, their conservational value would be actually highly decreased. However, we should suspect that these wide generalizations may not always be correct and a brief literature review readily shows that they do not correspond to many real situations. For example, several case studies have shown that coelacanth fishes (Holder et al. 1999; Casane and Laurenti 2013), and also horseshoe crabs (Avisé et al. 1994) have a polymorphic genetic structure, even in spite of their globally conserved genome. Recent studies also showed in some relict species, the tuataras or the *Cercidiphyllum* trees, two other emblematical relict species, the same pattern of mutational and retention of population genetic diversity as in other species (Hay et al. 2008; Qi et al. 2012). More generally, Casane and Laurenti (2013) also warned not to use raw population genetic diversity as a proxy for documenting evolutionary rate or stasis, as it depends on the population size or on the selection forces that could hide high mutation rate. The computation of evolutionary rates also strongly depends on the scale of sampling through generations.

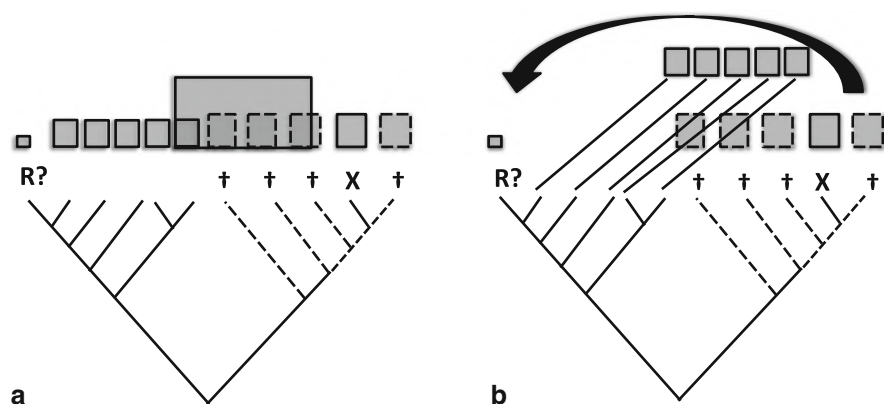
Conserving a relict species is therefore not just a way to save high organismic diversity and it is not at odds with retaining potential for future evolution. Recent empirical and theoretical studies have pointed out that conserving a phylogenetically diverse set of species could be a bet-hedging strategy that allows retention of species with most diverse characteristics of any kind, including evolutionary potential in the short term, i.e. adaptiveness (Faith 1992 chapter 3; Forest et al. 2007; Steel et al. 2007; Davies et al. 2008; Davies and Cadotte 2011; Fjeldså et al. 2012; Lean and Maclaurin chapter “The Value of Phylogenetic Diversity”). Actually, this potential should better be measured and evaluated in each case and not assumed from a priori conceptions of evolutionary stasis. Predicting the evolutionary potential of species on the long term (potential for speciating and radiating) is another issue, actually not feasible from any of their present characteristics (Barracough and Davies 2005; Winter et al. 2013). We should however remember that the evolutionary record of hundreds of millions of years told us many cases of strong diversifications in groups that were first strongly depleted (e.g., Neoaves, Eutherian mammals, etc.) Even if we cannot predict the future of a present relict species in thousands or millions of years, we should at least consider that it is actually not necessarily closed in terms of potential for surviving and diversifying, given what we know from the past histories of other several groups.

## Is There a Geographical or a Climatic Component to the Notion of Relictness?

In the earliest papers on the subject, species were considered relicts according to an inexplicit mixture of several components: taxonomic (now phylogenetic), climatic (e.g., the famous glacial relicts) and geographic (Darwin 1876; Simpson 1944; Darlington 1957; Holmquist 1962; Cronk 1992). All were considered because some species show relict features under each of these criteria, being both phylogenetic and geographic relicts. More recently, Parsons (2005) has drawn some interesting

inferences about the reasons why relicts (that he also called living fossils) can be geographically limited or not. A typical and often cited case is the tree *Ginkgo biloba*, found today only in a region of China and an amazing phylogenetic relict of the large group of Ginkgoales well known from the Cretaceous fossil record (Zhou 2009). This coincidence of criteria is however not always the case and some conspicuous phylogenetic relicts are quite widely distributed, including the horseshoe crab (e.g., Selander et al. 1970) and some tropical bird species (Fjeldså 1994). It appears then that the geographic or environmental criterion is secondary. Sometimes it fits, sometimes it does not, and all relict species need first to be documented on a phylogenetic basis. A “remnant” species strongly restricted geographically (typically isolated or peripheral) is not necessarily a relict that is isolated by extinction of its closest relatives. Phylogenetic or genetic studies could infer other less expected scenarios. The related group of the remnant could have been affected by both extinction and increase of neighboring distributions or the remnant may have originated after a dispersal event from a large distribution source (Fig. 2).

The traditional view of geographical restriction still expressed by various authors also considers the territories harboring one or several famous relicts as antique refuges (Gibbs 2006; Heads 2009). Generally, this biogeographic reasoning is quite circular, justifying the presence of relicts by the old geological age of the deep basement and considering it as a Noah’s Ark (without consideration for more recent and decisive paleogeographic events such as land submersion, major climatic changes, etc.) and vice versa, without searching for independent biological evidence (Waters



**Fig. 2** Two theoretical examples showing how the assumption that a geographically restricted or peripheral species is a relict can be falsified. These examples should be examined first with respect to distribution areas only (*upper* part of the figure), and then with consideration for the phylogenetic tree and extinctions events († and *spotted lines*, *lower* part). In both cases, R? was falsely believed to be a relict on a geographical basis alone (most peripheral and smallest distribution area) while the actual relict X was not identified as such. In the first case (a), the species X was not detected as a relict in the lineage because the distribution area of a neighboring species increased since the extinction of relatives. In the second case (b), the species R? was the most peripheral and isolated one because of a dispersal event from the zone where all the other species of the group were located including the species that went extinct



and Craw 2006). Actually, the fact that a relict remains from a larger group that is mainly extinct is a specific circumstance that makes difficult any inference of local permanence because relatives have disappeared and cannot therefore inform about the evolution of the distribution (Grandcolas et al. 2014).

Being a phylogenetic and a geographical relict at the same time is a frequent case and it brings even more concern for conservation. The relict species is then not only evolutionarily unique but also particularly vulnerable in case of local disturbance because of its limitation to a reduced area. As emphasized by Rodrigues et al. (2005), if this relict occurs in a small area that is not species-rich, it is even more potentially endangered because conservation actions will not be undertaken for other reasons.

There is however a recent trend to define as relicts some narrowly distributed or isolated populations when the species distribution is fragmented, even if the evolutionary status or phylogenetic position of these populations is either not known or presumptively not “deep-branching” and even if there is no evidence that this isolation was caused by the extinction of some related populations (Habel and Assmann 2010; Hampe and Jump 2011). Perhaps, dispersal caused the fragmented distribution (Fig. 2)? Population and conservation biologists wish to point out that such fragmented, isolated or remnant populations are worthy of further investigation or consideration for conservation (e.g., Laurance and Bierregaard 1997). In our opinion, this trend is confusing and brings polysemy within the term “relict.” Surviving extinctions as evidenced by phylogeny or the fossil record for whole species is not the same as having a decreasing or a fragmented distribution area for some populations within a species even if it involves some genetic differentiation (see even Watson 2002, who applies “relict” to the species present in forest fragments prior to fragmentation). Such short-lived population changes are likely frequent and dynamic, as shown by paleoenvironmental studies. One could assume that geographically relict populations are phylogenetically relict species in *statu nascendi*, but there is not (yet) evidence for that. We should be patient and wait for a few thousands of years at least before making our judgement... This is the reason why we proposed that the term relict should only be employed for phylogenetic relict species only. The so-called climatic or geographical relicts should then be better called “remnants” and qualifying the target category (population, forest fragment, etc.), for example, a climatic remnant population or a geographic remnant population (see Eriksson 2000 for a clarification and their possible functional importance).

From the point of view of conservation biology, this clarification is clearly needed and it permits distinction between two different cases. A phylogenetic and possibly geographical relict species must be considered for conservation since it contributes to organismic phylogenetic diversity and is possibly geographically or ecologically vulnerable. A climatic remnant population may just increase local diversity by the presence of one more species and, more significantly, may contribute to inform about interesting historical or ecological processes of distributional changes (Hampe and Jump 2011). The remnant population is neither necessarily deeply rooted into the history of the lineage nor remaining from a larger set. This

has to be eventually documented. Other cases of small distribution areas (for example, newly established populations of expanding species) are still another case, neither relict species nor remnant populations, but population isolates.

## **Relictness: A Relative Notion and the Need for Formal Analyses**

The term relict is generally employed either for emblematic taxa or for well-defined situations where taxonomic, phylogenetic and paleontological characterizations are established from previous studies and publications. This definition generally embodies the relict species, its very large sister-group, the paleontological record and the geographic restriction if any. For example, the relict *Amborella trichopoda* is the sister-group of all other flowering plants and it is only found in New Caledonia (Soltis et al. 2002). But relictness is not a particular state of a character that can be distinguished unambiguously from other and different states. Like most other class-level characterizations such as rarity, specialization or endemism (e.g., Rabinowitz 1981; Futuyma and Moreno 1988; Anderson 1994), relictness is a relative and composite situation that needs to be established by comparison, and in every case by phylogenetic comparison. Strictly speaking, we should not say *Amborella trichopoda* is “a relict” but *Amborella trichopoda* is “a relict species for flowering plants.” This comparison is based on topology and branch lengths that depend on the taxon and character samples used to build the tree. This way, within flowering plants, there are many groups that stand isolated on long branches and could be called relicts, such as *Welwitschia*, *Ephedra*, etc., actually hundred of species among hundreds of thousands of plant lineages (for examples see Jacobson and Lester 2003; Dilcher et al. 2005). As it is set by comparison between sister-groups within a phylogenetic tree, characterization as a relict will depend on the taxon sample used in that tree. For non-phylogeneticists, this could sound like a limitation of this notion that makes it less useful. Actually, a statement of relictness needs to be based on a formal phylogenetic analysis conducted on a given set of taxa. Depending on the tree obtained, a gap analysis can show that one or several branches have exceptional lengths and originate deep in the tree. These branches and their terminal taxa can be named relict taxa. This is required to implement the phylogenetic diversity criterion for conservation, characterizing the extreme case of relicts at the same time.

## **Relicts and Ecosystem Functioning**

Macroevolutionary studies of this type might appear to be far removed from the real nowadays’ world where ecosystems must function and populations must be viable to be conserved. Actually, historical and functional views are not opposed or

disconnected (Brooks and McLennan 1991; Grandcolas 1998). Current research (e.g., Elias et al. 2013) in the framework of community phylogenetics (Ricklefs and Latham 1992; Webb et al. 2002) shows that trophic webs have a phylogenetic structure. Phylogenetic niche conservatism mitigated by exploitative competition means that related species can have similar resource use (Cadotte et al. 2008; but see Mouquet et al. 2012). In this theoretical framework, a relict is then expected to exploit a unique niche, a prediction consistent with some of the adaptive explanations cited above (e.g., Parsons 2005), that relicts can be highly specialized (but inconsistent with relicts as generalists).

Therefore, maximizing phylogenetic diversity for conservation can be expected to select for species whose resource use is unique (Srivastava et al. 2012; Winter et al. 2013). In cases where relicts are found in a very stable and specialized habitat harboring small communities, this original resource use might implicate a key ecosystem service (e.g., Gibert and Deharveng 2002). At the extreme, structuring ecological communities by conserving species on the basis of phylogenetic diversity should select against loss of function in communities, by retaining species with lower niche overlap even if ecological redundancy is decreased.

## Relict Species and Present Extinction Risks

Relict species are therefore extreme cases of phylogenetic diversity and conserving them is of outstanding interest. In addition, they are not the living dead some people see them, which would have no viable populations or be unable to evolve or diversify again, as pictured by some people. In terms of conservation biology, however, we should not only consider whether they are valuable in themselves for conservation but also if they are at higher present extinction risk because of global change and human activities. As detailed by Yessoufou and Davies (chapter “[Reconsidering the Loss of Evolutionary History: How Does Non-random Extinction Prune the Tree-of-Life?](#)”), statistical studies suggest that species-poor, monotypic families, small genera and old groups in mammals, birds and plants – in other words, potentially relicts – are all more prone to extinction (Gaston and Blackburn 1997; Russell et al. 1998; Purvis et al. 2000; Meijaard et al. 2008; Vamossi and Wilson 2008; López-Pujol and Ren 2010). The causes of this situation probably lie in heritable phenotypic traits associated with long branches in these groups (Grandcolas et al. 2011). Even if these studies are biased by focusing on a few well-known groups (mammals, birds and plants) and by using proxies as red lists or meta-analyses for estimating extinction risks, they undoubtedly showed that present extinction could potentially have pernicious effects that were not suspected a priori (Nee and May 1997), by destroying proportionally more evolutionarily unique species. These results require more attention and future analyses should turn toward identifying the phenotypic characters that increase present vulnerability. It should not be assumed however that modern and past extinction risks are the same. The reasoning can be inverted; relicts are successful survivors from past geological times that could resist

any present global change, unless global change is fundamentally different from previous extinction crises.

Relicts are not only worthwhile to conserve by themselves because they are evolutionarily unique. They can also be at higher present extinction risks for phenotypic reasons that remain to explore in every case. Independently from any phenotypic effect, geographical or climatic relictiness and therefore a small distribution area can also be a source of vulnerability in itself.

## Relict Species and Conservation Biology: A Final Appraisal

Relict species, even if not all famous and rooted in very deep histories such as *Platypus* or *Ginkgo*, have been used as a powerful metaphor for explaining the use of phylogenetic diversity in the framework of conservation biology. We have seen that this is appropriate since relicts do represent an extreme case of phylogenetic diversity (Rodrigues et al. 2005). Relicts help understanding that some species can have a unique and decisive historical value, beyond strictly numerical considerations involving species counts or metrics measurements. From this qualitative point of view, phylogenetic diversity has already been given a lot of consideration (contra Winter et al. 2013; but see Rosauer and Mooers 2013). A growing body of research also shows that relict species are probably at higher present risk of extinction, which qualifies them for conservation planning from both perspectives.

Unfortunately, the metaphor has also been a vehicle for several misconceptions, that relicts are also living ancestors, basal taxa, or missing links. Even if these most outdated ideas are extirpated, there remains the tendency of some modern conservation biologists to erroneously conceive relicts as old species with poor evolutionary potential.

One important message of this chapter is therefore to explain why this later conception cannot be generalized or taken as true a priori. When dealing with relicts and phylogenetic diversity in general, it must always be recalled that the present diversity is the result of the balance between past speciation and past extinction. This way, relicts remain from larger groups partly extinct. The consequence is that any computation of their age will be strongly biased if the past occurrence of extinct species is not taken into account. The age of the relict species could be equated naively with the age of the crown group and the base of the branch, when it might actually be quite recent. In addition, the evolutionary rate of the relict lineage should be measured and not just assumed to be generally low by focusing on a minority of emblematic phenotypic characters that remained stable over long time periods.

Conserving organismic diversity requires consideration for “the whole real guts of evolution – which is, how do you come to have horses, and tigers, and things” (Waddington (1967) quoted by Eldredge and Cracraft (1980)). But such a historical view is not at odds with conserving a functional world and a world still keeping some evolutionary potential. There are not two different worlds, the one with the

animals of the zoo and the other with balanced trophic relationships and resilience to global environmental changes.

The other confusion to avoid is that relicts are not simply geographic or ecological remnants. Part of a population can remain in a habitat patch after ecosystem fragmentation without being evolutionarily relict. Using the term “relict” to put emphasis on any isolate or remnant biological entity is unhelpful and confusing.

The metaphor of relicts is not only useful to explain the scientific importance of phylogenetic diversity but also has added political value for the development of public conservation planning. Because of its emblematic value, a relict is potentially a flag species whose presence in a location could help promote conservation. Because of their importance, the position and the characteristics of such relict taxa must be even more accurately specified. We should focus on knowing better to conserve better.

**Acknowledgements** We gratefully thank Jon Fjeldså, Mike Crisp, Dan Faith and Roseli Pellens for their comments on the manuscript, which greatly helped clarifying the ideas presented here. We also took benefit of reading other chapters of the book that provided rich reviews about various aspects of phylogenetic diversity and that allowed more maturation than just cross-referencing. Finally, we are also grateful to many people who commented on previous versions of our thoughts about relicts in other contexts.

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Abe T, Bignell DE, Higashi M (eds) (2000) Termites: evolution, sociality, symbioses, ecology. Kluwer Academic Publishers, Dordrecht
- Amemiya CT, Powers TP, Prohaska SJ, Grimwood J, Schmutz J, Dickson M, Miyake T, Schoenborn MA, Myers RM, Ruddle FH, Stadler PF (2010) Complete HOX cluster characterization of the coelacanth provides further evidence for slow evolution of its genome. *Proc Natl Acad Sci U S A* 107:3622–3627
- Anderson S (1994) Area and endemism. *Q Rev Biol* 69:451–471
- Assmann T, Casale A, Drees C, Habel JC, Matern A, Schuldt A (2010) Review: the dark side of relict species biology: cave animals as ancient lineages. In: Habel JC, Assmann T (eds) *Relict species: phylogeography and conservation biology*. Springer, Berlin, pp 91–103
- Avise JC, Nelson WS, Sugita H (1994) A speciation history of “living fossils”: molecular evolutionary patterns in horseshoe crabs. *Evolution* 48:1986–2001
- Barracough TG, Davies TJ (2005) Predicting future speciation. In: Purvis A, Gittleman JL, Brooks T (eds) *Phylogeny and conservation*. Cambridge University Press, Cambridge, pp 400–418
- Bergsten J (2005) A review of long-branch attraction. *Cladistics* 21:163–193
- Bergthorsson U, Richardson AO, Young GJ, Goertzen LR, Palmer JD (2004) Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*. *Proc Natl Acad Sci U S A* 101:17747–17752
- Brooks DR, McLennan DA (1991) *Phylogeny, ecology, and behavior: a research program in comparative biology*. The University of Chicago Press, Chicago

- Cadotte MW, Cardinale BJ, Oakley TH (2008) Evolutionary history and the effect of biodiversity on plant productivity. *Proc Natl Acad Sci* 105:17012–17017
- Casane D, Laurenti P (2013) Why coelacanths are not ‘living fossils’. *BioEssays* 35:332–338
- Crisp MD, Cook LG (2005) Do early branching lineages signify ancestral traits? *Trends Ecol Evol* 20:122–128
- Crisp MD, Cook LG (2009) Explosive radiation or cryptic mass extinction? Interpreting signatures in molecular phylogenies. *Evolution* 63:2257–2265
- Crisp MD, Cook LG (2011) Cenozoic extinctions account for the low diversity of extant gymnosperms compared with angiosperms. *New Phytol* 192(4):997–1009
- Cronk QCB (1992) Relict floras of Atlantic islands: patterns assessed. *Biol J Linn Soc* 46:91–103
- Darlington PJ (1957) *Zoogeography: the geographical distribution of animals*. Wiley, New York
- Darwin CR (1876) *The origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*, 6th edn. John Murray, London, with additions and corrections
- Davies TJ, Cadotte MW (2011) Quantifying biodiversity: does it matter what we measure? In: Zachos FE, Habel JC (eds) *Biodiversity hotspots*. Springer, pp 43–60
- Davies TJ, Fritz SA, Grenyer R, Orme CDL, Bielby J, Bininda-Emonds ORP, Cardillo M, Jones KE, Gittleman JL, Mace G, Purvis A (2008) Phylogenetic trees and the future of mammalian biodiversity. *Proc Natl Acad Sci U S A* 105:11556–11563
- Dilcher DL, Bernardes-De-Oliveira ME, Pons D, Lott TA (2005) Welwitschiaceae from the lower cretaceous of northeastern Brazil. *Am J Bot* 92:1294–1310
- Dowle EJ, Morgan-Richards M, Trewick SA (2013) Molecular evolution and the latitudinal biodiversity gradient. *Heredity* 110:501–510
- Eldredge N (1987) Part 5. Living fossils. Introduction. In: Eldredge N (ed) *The natural history reader in evolution*. Columbia University Press, New York, pp 137–139
- Eldredge N, Cracraft J (1980) *Phylogenetic patterns and the evolutionary process: method and theory in comparative biology*. Columbia University Press, New York
- Eldredge N, Thompson JN, Brakefield PM, Gavrillets S, Jablonski D, Jackson JBC, Lenski RE, Lieberman BS, McPeck MA, Miller W (2005) The dynamics of evolutionary stasis. *Paleobiology* 31:133–145
- Elias M, Fontaine C, van Veen FJ (2013) Evolutionary history and ecological processes shape a local multilevel antagonistic network. *Curr Biol* 23:1355–1359
- Engelmann GF, Wiley EO (1977) The place of ancestor-descendant relationships in phylogeny reconstruction. *Syst Zool* 26:1–11
- Eriksson O (2000) Functional roles of remnant plant populations in communities and ecosystems. *Global Ecol Biogeogr* 9:443–449
- Erwin TL (1991) An evolutionary basis for conservation strategies. *Science* 253:750–752
- Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61(1):1–10
- Fjelds  J (1994) Geographical patterns for relict and young species of birds in Africa and South America and implications for conservation priorities. *Biodivers Conserv* 3:207–226
- Fjelds  J, Bowie RCK, Rahbek C (2012) The role of mountain ranges in the diversification of birds. *Annu Rev Ecol Syst* 43:249–265
- Forest F, Grenyer R, Rouget M, Davies TJ, Cowling RM, Faith DP, Balmford A, Manning JC, Proches S, van der Bank M, Reeves G, Hedderson TAJ, Savolainen V (2007) Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature* 445:757–760
- Friedman WE, Ryerson KC (2009) Reconstructing the ancestral female gametophyte of angiosperms: insights from *Amborella* and other ancient lineages of flowering plants. *Am J Bot* 96:129–143
- Futuyma DJ, Moreno (1988) The evolution of ecological specialization. *Annu Rev Ecol Syst* 19:207–233
- Gaston K, Blackburn T (1997) Evolutionary age and risk of extinction in the global avifauna. *Evol Ecol* 11:557–565

- Gibbs G (2006) *Ghosts of Gondwana. The history of life in New Zealand*. Craig Potton Publishing, Nelson
- Gibert J, Deharveng L (2002) Subterranean ecosystems: a truncated functional biodiversity. *Bioscience* 52:473–481
- Grandcolas P (1998) Phylogenetic analysis and the study of community structure. *Oikos* 82:397–400
- Grandcolas P, Nattier R, Legendre F, Pellens R (2011) Mapping extrinsic traits such as extinction risks or modelled bioclimatic niches on phylogenies: does it make sense at all? *Cladistics* 27:181–185
- Grandcolas P, Nattier R, Trewick SA (2014) Relict species: a relict concept? *Trends Ecol Evol* 29:655–663
- Habel JC, Assmann T (eds) (2010) *Relict species: phylogeography and conservation biology*. Springer, Berlin
- Hampe A, Jump AS (2011) Climate relicts: past, present, future. *Annu Rev Ecol Evol Syst* 42:313–333
- Hay JM, Subramanian S, Millar CD, Mohandesan E, Lambert DM (2008) Rapid molecular evolution in a living fossil. *Trends Genet* 24:106–109
- Heads M (2009) Globally basal centres of endemism: the Tasman-coral sea region (south-west Pacific), Latin America and Madagascar/South Africa. *Biol J Linn Soc* 96:222–245
- Holder MT, Erdmann MV, Wilcox TP, Caldwell RL, Hillis DM (1999) Two living species of coelacanths? *Proc Natl Acad Sci U S A* 96:12616–12620
- Holmquist C (1962) The relict concept – is it a merely zoogeographical conception? *Oikos* 13:262–292
- Inoue JG, Miya M, Venkatesh B, Nishida M (2005) The mitochondrial genome of Indonesian coelacanth *Latimeria menadoensis* (Sarcopterygii: Coelacanthiformes) and divergence time estimation between the two coelacanths. *Gene* 349:227–235
- Jacobson KM, Lester E (2003) A first assessment of genetic variation in *Welwitschia mirabilis* hook. *J Hered* 94:212–217
- Janvier P (2007) Living primitive fishes and fishes from deep time. In: David J, McKenzie APF, Colin JB (eds) *Fish physiology*, vol 26. Academic Press, San Diego, pp 1–51
- Krell FT, Cranston PS (2004) Which side of the tree is more basal? *Syst Entomol* 29:279–281
- Laurance WF, Bierregaard RO (eds) (1997) *Tropical forest remnants. Ecology, management, and conservation of fragmented communities*. Chicago University Press, Chicago
- Legendre F, Whiting MF, Bordereau C, Canello EM, Evans TA, Grandcolas P (2008) The phylogeny of termites (Dictyoptera: Isoptera) based on mitochondrial and nuclear genes: implications for the evolution of the worker and pseudergate castes, and foraging behaviors. *Mol Phylogenet Evol* 48:615–627
- Lincoln RJ, Boxshall GA, Clark PF (1982) *A dictionary of ecology, evolution and systematics*. Cambridge University Press, Cambridge
- López-Pujol J, Ren MX (2010) China: a hot spot of relict plant taxa. In: Rescigno V, Maletta S (eds) *Biodiversity hotspots*. Nova Publishers, Hauppauge, pp 123–137
- Mace GM, Purvis A (2008) Evolutionary biology and practical conservation: bridging a widening gap. *Mol Ecol* 17:9–19
- Meijaard E, Sheil D, Marshall AJ, Nasi R (2008) Phylogenetic age is positively correlated with sensitivity to timber harvest in Bornean mammals. *Biotropica* 40:76–85
- Merriam-Webster.com. Retrieved 3 Mar 2014, from <http://www.merriam-webster.com/dictionary/relict>
- Mouquet N, Devictor V, Meynard CN, Munoz F, Bersier L-F, Chave J, Couteron P, Dalecky A, Fontaine C, Gravel D, Hardy OJ, Jabot F, Lavergne S, Leibold M, Mouillot D, Münkemüller T, Pavoine S, Przinig A, Rodrigues ASL, Rohr RP, Thébaud E, Thuiller W (2012) Ecophylogenetics: advances and perspectives. *Biol Rev* 87(4):769–785
- Myers N, Knoll AH (2001) The biotic crisis and the future of evolution. *Proc Natl Acad Sci U S A* 98(10):5389–5392

- Nagalingum NS, Marshall CR, Quental TB, Rai HS, Little DP, Mathews S (2011) Recent synchronous radiation of a living fossil. *Science* 334:796–799
- Nee S, May RM (1997) Extinction and the loss of evolutionary history. *Science* 278:692–694
- Nelson GJ (1970) Outline of a theory of comparative biology. *Syst Zool* 19:373–384
- O'Brien SJ (2008) The *Platypus* genome unraveled. *Cell* 133:953–955
- Parsons PA (2005) Environments and evolution: interactions between stress, resource inadequacy and energetic efficiency. *Biol Rev* 80:589–610
- Purvis A, Agapow PM, Gittleman JL, Mace GM (2000) Nonrandom extinction and the loss of evolutionary history. *Science* 288:328–330
- Qi X-S, Chen C, Comes HP, Sakaguchi S, Liu Y-H, Tanaka N, Sakio H, Qiu Y-X (2012) Molecular data and ecological niche modelling reveal a highly dynamic evolutionary history of the East Asian tertiary relict *Cercidiphyllum* (Cercidiphyllaceae). *New Phytol* 196:617–630
- Quental TB, Marshall CR (2010) Diversity dynamics: molecular phylogenies need the fossil record. *Trends Ecol Evol* 25:434–441
- Rabinowitz D (1981) Seven forms of rarity. In: Synge H (ed) *The biological aspects of rare plant conservation*. Wiley, Chichester, pp 205–217
- Rabosky DL (2006) Likelihood methods for detecting temporal shifts in diversification rates. *Evolution* 60:1152–1164
- Rice DW, Alverson AJ, Richardson AO, Young GJ, Sanchez-Puerta MV, Munzinger J, Barry K, Boore JL, Zhang Y, dePamphilis CW, Knox EB, Palmer JD (2013) Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. *Science* 342:1468–1473
- Ricklefs RE (2005) Small clades at the periphery of passerine morphological space. *Am Nat* 165:651–659
- Ricklefs RE (2007) Estimating diversification rates from phylogenetic information. *Trends Ecol Evol* 22:601–610
- Ricklefs RE, Latham RE (1992) Intercontinental correlation of geographical ranges suggests stasis in ecological traits of relict genera of temperate perennial herbs. *Am Nat* 139:1305–1321
- Rodrigues AS, Brooks TM, Gaston KJ (2005) Integrating phylogenetic diversity in the selection of priority areas for conservation: does it make a difference? In: Purvis A, Gittleman JL, Brooks T (eds) *Phylogeny and conservation*. Cambridge University Press, Cambridge, pp 101–119
- Rosauer DF, Mooers AO (2013) Nurturing the use of evolutionary diversity in nature conservation. *Trends Ecol Evol* 28(6):322–323
- Russell GJ, Brooks TM, McKinney MM, Anderson CG (1998) Present and future taxonomic selectivity in bird and mammal extinctions. *Conserv Biol* 12:1365–1376
- Selander RK, Yang SY, Lewontin RC, Johnson WE (1970) Genetic variation in Horseshoe Crab (*Limulus-Polyphemus*), a phylogenetic relic. *Evolution* 24:402–414
- Simpson GG (1944) *Tempo and mode in evolution*. Columbia University Press, New York
- Soltis DE, Soltis PS, Albert VA, Oppenheimer DG, dePamphilis CW, Ma H, Frohlich MW, Theissen G, Grp FGPR (2002) Missing links: the genetic architecture of flower and floral diversification. *Trends Plant Sci* 7:22–31
- Srivastava DS, Cadotte MW, MacDonald AAM, Marushia RG, Mirotnick N (2012) Phylogenetic diversity and the functioning of ecosystems. *Ecol Lett* 15:637–648
- Steel M, Mimoto A, Mooers AO (2007) Hedging our bets: the expected contribution of species to future phylogenetic diversity. *Evol Bioinforma* 3:237–244
- Thorne RF (1999) Eastern Asia as a living museum for archaic angiosperms and other seed plants. *Taiwania* 44:413–422
- Vamosi JC, Wilson JR (2008) Nonrandom extinction leads to elevated loss of angiosperm evolutionary history. *Ecol Lett* 11:1047–1053
- Van Dover CL, German CR, Speer KG, Parson LM, Vrijenhoek RC (2002) Marine biology – evolution and biogeography of deep-sea vent and seep invertebrates. *Science* 295:1253–1257
- Vane-Wright RI, Humphries CJ, Williams PH (1991) What to protect? Systematics and the agony of choice. *Biol Conserv* 55:235–254



- Waddington CH (1967) Comment made during discussion of paper by Dr. Eden. In: Moorehead PS, Kaplan MM (eds) *Mathematical challenges to the Neo-Darwinian interpretation of evolution*. Wistar Institute Press, Philadelphia, p 14
- Waters JM, Craw D (2006) Goodbye Gondwana? New Zealand biogeography geology, and the problem of circularity. *Syst Biol* 55:351–356
- Watson DM (2002) A conceptual framework for studying species composition in fragments, islands and other patchy ecosystems. *J Biogeogr* 29:823–834
- Webb CO, Ackerly DD, McPeck MA, Donoghue MJ (2002) Phylogenies and community ecology. *Annu Rev Ecol Syst* 33:475–505
- Winter M, Devictor V, Schweiger O (2013) Phylogenetic diversity and nature conservation: where are we? *Trends Ecol Evol* 28:199–204
- Zhou ZY (2009) An overview of fossil Ginkgoales. *Palaeoworld* 18:1–22

## **Part II**

# **Methods**

# Using Phylogenetic Dissimilarities Among Sites for Biodiversity Assessments and Conservation

Daniel P. Faith

**Abstract** The PD phylogenetic diversity measure provides a measure of biodiversity that reflects variety at the level of features, among species or other taxa. PD is based on a simple model which assumes that shared ancestry explains shared features. PD provides a family of calculations that operate as if we were directly counting up features of taxa. PD-dissimilarity or phylogenetic beta diversity compares the branches/features represented by two different areas. We also can consider a companion model, which shifts the focus to shared habitat/environment among taxa as the explanation of shared features, including those features not explained by shared ancestry and PD. That model means that PD-dissimilarities, among sampled and unsampled sites, can be predicted using a regression method applied to distances in an environmental-gradients space. However, PD-based conservation planning requires more than the dissimilarities among all sites, in order to make decisions informed by gains and losses of branches/features. The companion model also suggests how to transform dissimilarities to provide these needed estimates. This *ED* (“*Environmental Diversity*”) method out-performs other suggested strategies for analysis of dissimilarities, including the Ferrier et al. method and the Arponen et al. method. The global biodiversity observation network (GEO BON) can use the *ED* method for inferences of biodiversity change that include loss of phylogenetic diversity.

**Keywords** Environmental diversity • Phylogenetic beta diversity • *ED* complementarity • Conservation planning • Biodiversity monitoring

---

D.P. Faith (✉)

AMRI, The Australian Museum, Sydney, NSW 2010, Australia

e-mail: [dan.faith@austmus.gov.au](mailto:dan.faith@austmus.gov.au)

© The Author(s) 2016

R. Pellens, P. Grandcolas (eds.), *Biodiversity Conservation and Phylogenetic Systematics*, Topics in Biodiversity and Conservation 14,  
DOI 10.1007/978-3-319-22461-9\_7

119

## Introduction

This book addresses important concepts, methods, and applications related to the role of evolutionary history in biodiversity conservation. In the chapter “[The PD Phylogenetic Diversity Framework: Linking Evolutionary History to Feature Diversity for Biodiversity Conservation](#)” (Faith 2015a), I reviewed the reasons why we want to conserve evolutionary history. An important rationale is that the tree of life is a storehouse of variation among taxa, and so provides possible future benefits for humans (for discussion, see Faith et al. 2010). I also reviewed the justifications for a specific biodiversity measure. It interprets the degree of representation of evolutionary history as a phylogenetic measure of biodiversity, or “phylogenetic diversity”. This measure of phylogenetic diversity, called “PD” (Faith 1992a, b) is justified as a useful biodiversity measure through its link to “feature diversity”. Feature diversity represents biodiversity “option values” – the term we use to refer to all those potential future benefits for humans – and so is well-justified as a target for biodiversity conservation. Forest et al. (2007) provide a good exemplar study, illustrating how PD links to feature diversity and to food, medicine, and other benefits to humans.

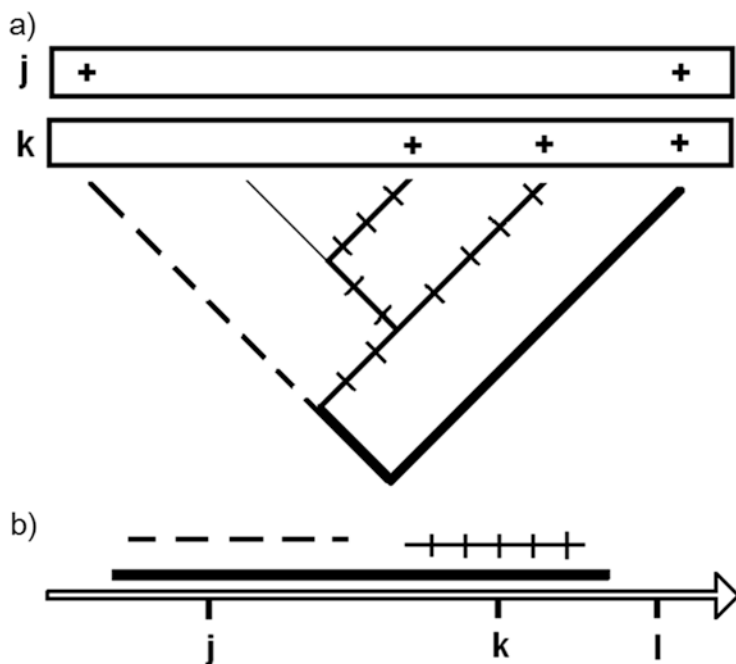
Faith (2002) summarised the link between evolutionary history, PD, and features as follows: “representation of “evolutionary history” (Faith 1994) encompassing processes of cladogenesis and anagenesis is assumed to provide representation of the feature diversity of organisms. Specifically, the phylogenetic diversity (PD) measure estimates the relative feature diversity of any nominated set of species by the sum of the lengths of all those phylogenetic branches spanned by the set...”

The calculation of the PD for a given subset of species (sampled from a phylogenetic tree) is quite simple. It is given by the minimum total length of all the phylogenetic branches required to connect all those species on the tree. However, calculation of PD is attempting something that is not all that simple – an inference of the relative feature diversity of that subset of species. The basis for this inference is an evolutionary model in which branch lengths reflect evolutionary changes, and shared ancestry accounts for shared features (Faith 1992a, b). The model implies that PD, in effect, counts-up the relative number of features represented by a given subset of species (or other taxa, including populations within a species); any subset of species that has greater PD will be expected to have greater feature diversity.

In chapter “[The PD Phylogenetic Diversity Framework: Linking Evolutionary History to Feature Diversity for Biodiversity Conservation](#)”, I described another important implication of the link to feature diversity: PD provides, not one single measure, but a set of calculations interpretable at the level of features of taxa. This helps guide the assessment of the phylogenetic diversity gains and losses from changing probabilities of extinction of species (or other taxa). This PD “calculus” also can help with the conservation problem addressed in this paper: assessing PD gains and losses when we gain or lose geographic areas. PD has long been integrated into conservation planning for areas (Walker and Faith 1994). However, the work so far has largely ignored the problem of geographic knowledge gaps; we do

not know about the phylogenetic diversity represented in every area in a given region. Consequently, for conservation planning, we have to estimate or model these missing quantities, using spatial models incorporating predictive environmental variables.

One pathway for such predictions can take advantage of a part of the PD calculus called “PD-dissimilarities” or “phylogenetic beta diversity” (Fig. 1a; see also Lozupone and Knight 2005; Ferrier et al. 2007; Nipperess et al. 2010; Swenson 2011). PD-dissimilarities can be interpreted as compositional dissimilarities, based



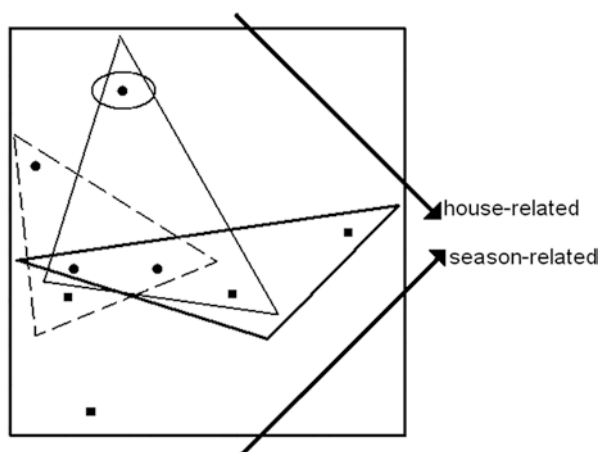
**Fig. 1** (a) A hypothetical phylogenetic tree with 5 taxa. Along the top, the presence of the taxa in two sites, j and k, is shown by + marks. The *dashed-line* branches indicate features only represented in j; *hatched* branches indicate features only represented in k; *bold* branches indicate features represented in both; the *thin* branch indicates features in neither. The presence absence version of Bray-Curtis type PD-dissimilarity between sites j and k counts the number of features in j, not k (length of dashed branches) plus the number of features in k, not j (length of hatched branches), divided by the sum of the total number of features found in each (length of *dashed* plus length of *bold* branches, plus length of hatched plus length of *bold* branches). Other PD-dissimilarity measures combine these counts in other ways. (b) A hypothetical environmental gradient (*hollow-line*) with positions of sites, j, k, and l. Suppose that positions of sites along this gradient reflect their features. Sites with a given feature are found in a corresponding part of the gradient. This clumping is called a “unimodal” response. Above the gradient is the hypothetical unimodal distribution of the branches and corresponding features/branches from 1a. Under the unimodal response model, the features in both j and k, for example, form the *bold line* segment. This unimodal relationship means that the Bray-Curtis type PD-dissimilarity has the most robust link to distances along environmental gradients (or in environmental space; for discussion, see Faith et al. 1987). For further information, also see Faith et al. (2009)

on the branches/features represented at the different sites (a site represents all branches that are ancestral to any of its member species). These calculations are “community-based” approaches in that they compare areas based on the set of elements (the community) found in each area. We can think of the standard compositional dissimilarity measures conventionally applied at the species level as simply re-cast at the level of features, through the PD model (Fig. 1a; for discussion, see Faith 2013).

Spatial predictions can use a form of regression in which PD-dissimilarities between sites are explained and predicted by the known environmental distances between sites. Thus, we can predict the PD-dissimilarity between two un-sampled sites, given their environmental difference. Generalized dissimilarity modelling (GDM; Ferrier 2002; Ferrier et al. 2004, 2007; see also Faith and Ferrier 2002), an extension of matrix regression, is useful for these predictions. GDM realistically allows for a very general monotonic, curvilinear, relationship between increasing environmental distance and compositional dissimilarity. It is also robust in allowing for variation in the rate of compositional change at different positions along environmental gradients. GDM was developed for species-level dissimilarities, but has been extended to the prediction of PD-dissimilarities (Ferrier et al. 2007; Faith et al. 2009; Rosauer et al. 2013).

There are several ways to calculate a PD-dissimilarity (see Fig. 1a, b). The choice of the PD-dissimilarity measure for such analyses can be guided by another critical model, which makes additional assumptions about how features link to environmental variables. To understand the nature of this model, it is important to note that Faith (1992a, b; see also Faith 1996) was careful to point out that PD’s shared-ancestry/shared-features model provides a general prediction about feature diversity, but naturally does not apply to all possible features. This early work proposed that a companion model also can account for shared features, including those that are not explained by shared ancestry (e.g. those features that are convergent, arising independently on the phylogenetic tree). Here, a pattern among species describing shared habitat or environment explains shared branches/features (Fig. 1b; Faith 1989, 1996, 2015b; Faith et al. 2009). Figure 1b illustrates how shared habitat or environment explains shared features: the sites sharing particular branches or features form clumps or clusters in the environmental space (see also Fig. 2). I will refer to this as *unimodal response* (analogous the well-known unimodal response of species to environmental gradients; see e.g. Faith et al. 1987). This unimodal relationship (Fig. 1b) means that the Bray-Curtis type PD-dissimilarity has the most robust link to distances along environmental gradients (or in environmental space; for discussion, see Faith et al. 1987).

This simple model arguably deserves to make a greater contribution towards our understanding of biodiversity methods. For example, an under-appreciation of this companion model has meant that some workers (Kelly et al. 2014) still naively characterise PD as intended to account for all features, including those convergently derived. Similarly, the role of this model in explaining habitat-driven feature diversity has been neglected in the development of functional trait diversity measures (discussed in Faith 2015b). In this paper, I discuss another good reason to consider



**Fig. 2** Bray-Curtis type PD-dissimilarities can be used in robust ordination methods to recover key gradients. A re-drawing of the gradient space from Rintala et al. (2008; see also Faith et al. 2009) for microbial communities in house dust and a microbial phylogenetic tree. *Dots* versus *squares* correspond to samples from two different buildings (for details of sampling see Rintala et al.). *Arrows* at the *right side* indicate major gradients revealed by the ordination. A sample locality represents the branch corresponding to a given family if the locality has one or more descendants of that branch. The two-dimensional space shows unimodal response for four branches (*Acidaminococcaceae*, *Aerococcaceae*, *Enterobacteriaceae*, *Acetobacteraceae*). For further information, see Faith et al. (2009)

this shared-habitat/shared-features model: it can fill a critical gap in our attempts to effectively use PD-dissimilarities for biodiversity assessments.

We can predict the Bray-Curtis type PD-dissimilarities from environmental distances using a GDM regression. However, this is a mixed blessing. We produce PD-dissimilarities for all pairs of sites, but a difficulty is that these dissimilarities do not directly tell us what we want to know for conservation planning – the total phylogenetic diversity represented by a given subset of areas, or the gain or loss in PD if an site is gained or lost. To fill this gap, we need to convert the pairwise dissimilarities into inferences about PD representation and/or gains and losses. I will show how the shared-habitat/shared-features model can guide this analysis.

While there are several natural candidate approaches for taking this extra analysis step (each extends methods applied to species-level dissimilarities), surprisingly, there is no established, accepted method. One proposed approach, based on the unimodal response model, is the *ED* (“*environmental diversity*”) method (defined below; see also Faith and Walker 1996a, b, c), which has for some time been linked to GDM and species-level dissimilarities (Faith and Ferrier 2002). Faith et al. (2009) proposed the application of *ED* to the predicted dissimilarities from phylogenetic GDM analyses, but there are no worked examples exploring this approach. Another attractive method, linked strongly to the GDM approach, is the Ferrier et al. (2004) index. This measure modifies the *ED* approach and has been applied for species-level dissimilarities. A closely related method is that of Arponen et al. (2008).

Both of these have commonalities with *ED*, but the similarities and differences – and the strengths and weaknesses – among these alternative candidate measures has not been explored and documented (for related discussion, see Ferrier and Drielsma 2010).

Given this fundamental gap in building the complete toolbox of PD calculations for conservation, and given the lack of synthesis among candidate methods, this chapter will proceed as follows. I first show how the same model of shared-environment/shared-features that justifies the choice among possible PD-dissimilarity measures (Fig. 1a, b), also justifies the choice of the *ED* method. I then present a sample application of *ED* to PD-dissimilarities. I also present a simple graphical description of *ED* in the one dimensional case, which clarifies how *ED* estimates representation and gains and losses. I then use this graphical representation to reveal key properties of the alternative methods, suggesting critical weaknesses of the Ferrier et al. and Arponen et al. methods. I finish on a positive note, pointing to future work, including expanding the range of calculations useful for conservation assessment based on *ED*.

## How the *ED* Method Converts PD-Dissimilarities to Estimates of Gains and Losses

“ED” refers to a specific family of “environmental diversity” calculations (Faith and Walker 1996a, b, c; Faith 2003; Faith et al. 2003, 2004). *ED* typically uses an environmental gradients space, derived using species compositional dissimilarities and ordination methods (Faith and Walker 1996a, b, c). *ED* has been implemented as a surrogate strategy in biodiversity conservation-planning software that evaluates nominated sets of localities or finds best sites to add to an existing set. For example, *ED* provided the first integration of ‘costs’ into regional biodiversity planning based on comparing gains or ‘*ED*-complementarity’ values to marginal costs to facilitate trade-offs, balancing biodiversity conservation and other needs of society (Faith et al. 1996).

In order to understand the applicability of *ED* to PD-dissimilarities, we have to consider *ED*’s assumptions and then examine a simple example analysis. I referred above to unimodal response (Fig. 1b) and the shared-habitat/shared-features companion model to PD’s shared-ancestry/shared-features model. *ED* explicitly builds on this general unimodal response of species (or other elements) to environmental gradients (for background, see Austin 1985; Faith et al. 1987). *ED*’s environmental space typically is derived using compositional dissimilarities (including those estimated GDM) and ordination methods (for review, see Faith et al. 2004). The dissimilarities, the ordination methods and GDM all are relatively robust approaches under a general model of unimodal responses to environmental gradients (Fig. 1b; Faith et al. 1987; Faith and Walker 1996a; Ferrier et al. 2009).

The unimodal response model not only guides the inference of an environmental space using ordination methods (Faith et al. 1987), but also defines how *ED* methods



should effectively sample that environmental space in order to capture biodiversity. *ED* is based on the idea that many different species (or other elements of biodiversity) respond to similar environmental gradients, and exhibit a general unimodal response at different positions along those gradients (Fig. 1b). It follows that effective representation of these gradients (say, by a proposed set of protected areas) should deliver good representation of the various species or phylogenetic branches.

The assumption of a general unimodal response model directly leads to the use of p-median (and related) criteria for *ED*'s estimation of the number of species represented by a given set of localities in the environmental space or ordination. A p-median criterion is based on a sum of the distances in an environmental space. Each distance in this summation is that between a hypothetical point ('demand point') in the space and its nearest site (among all sites in some selected subset). The selected sites, for example, might be nominated protected-area localities. *ED* is defined based on this calculation. The 'continuous' version of *ED* refers to the case where the demand points are hypothetical points distributed uniformly throughout the continuous environmental space. Faith and Walker (1994, 1996a) demonstrated that, under a simple unimodal response model, species representation will be maximised by a selected set of sites if and only if it satisfies this continuous p-median criterion. Note that the *ED* score, because it counts un-represented species based on a sum of distances, is numerically *small* when the number of represented species is *large* (see example calculations below and in Faith and Walker 1996a). The *ED* surrogates approach therefore provides a rationale for interpreting high environmental diversity for a set of localities as implying high biodiversity for the set (see Beier and Albuquerque 2015).

I referred above to the p-median *and related* criteria. *ED* is not defined by any a priori choice of the p-median criterion. Instead, the various *ED* calculations emerge from the assumption of an underlying unimodal response model. In the simple case, unimodal response implies that features are effectively counted up when we apply calculations linked to the p-median; in other cases, the model implies calculations that are modifications of the simple p-median. Simple *ED* variants include weighting of demand points when species richness varies over the space (Faith and Walker 1996a; Faith et al. 2004), and creating an extended environmental space ('extended polytope'; Faith and Walker 1994, 1996a, b; Faith et al. 2004; see also Hortal et al. 2009). These options modify the parameters used in calculating the p-median. In a later section, I will consider an *ED* variant that departs from p-median in order to capture expected diversity or persistence.

When extended to features and branches from a phylogeny, the unimodal response model supports an expectation that *ED* is compatible with Bray-Curtis type PD-dissimilarities. Does this unimodal model (as idealised in Fig. 1b) apply when the elements are branches or features? Certainly, this relationship can be expected, given that PD-dissimilarity operates as if it is a standard Bray Curtis dissimilarity, but applied to features, not species. The robust ordination of such dissimilarities should produce general unimodal responses, as in the species-level case (Faith et al. 1987).

PD and PD-dissimilarities are commonly applied to molecular phylogenetic trees and microbial community data; here, PD analyses overcome the typical absence of defined microbial species. However, there has not been any clear model linking branches to gradients in such studies. Faith et al. (2009) presented an example documenting unimodal response of branches based on a gradient space for microbial communities, sampled in house dust (Fig. 2; Rintala et al. 2008). In Fig. 2, arrows at the right side indicate major gradients revealed by the ordination of the PD-dissimilarities. The solid dots in the space indicate different communities or sample localities. A sample locality represents the branch corresponding to a given family if the locality has one or more descendants of that branch in the phylogeny (for details see Rintala et al. 2008; Faith et al. 2009).

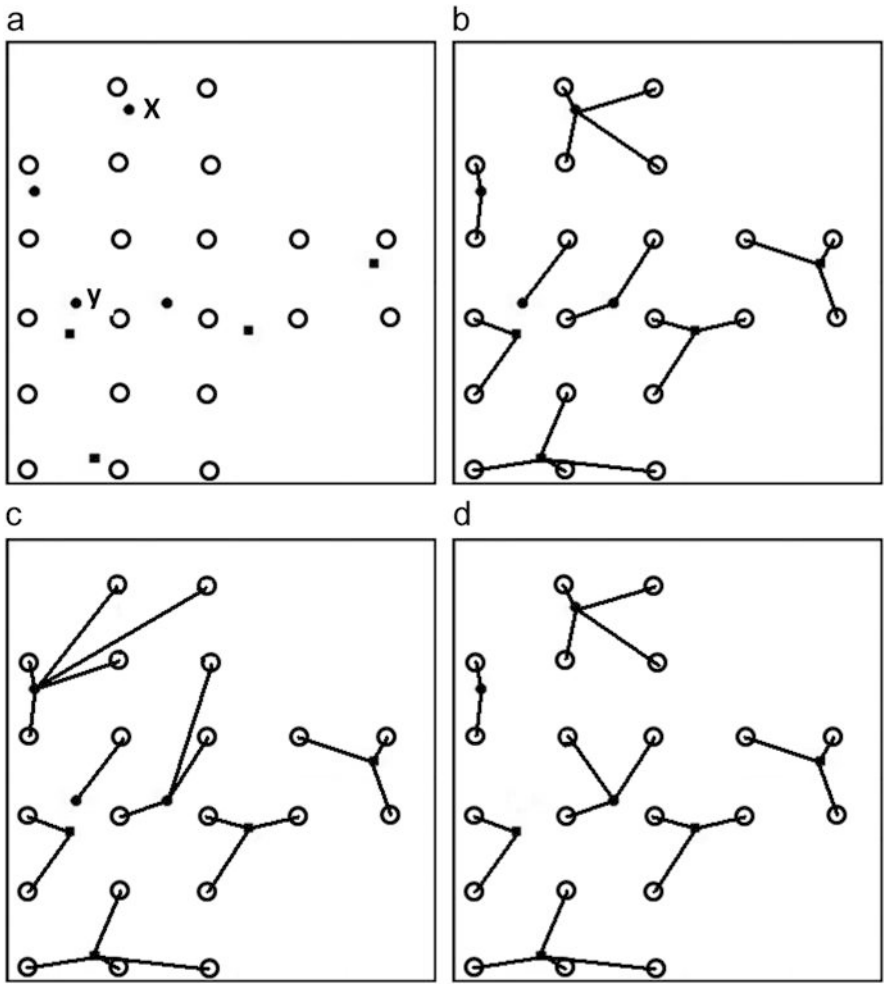
For the ordination space of Rintala et al., Faith et al. (2009) showed that all but 3 of the 56 phylogenetic branches (corresponding to identified families) have a clear unimodal response in the gradients space. Here, a response was recorded as unimodal only if a simple shape could enclose all sample sites representing the given branch (and not include any other sites). This unimodal response for phylogenetic features or branches is a critical property: it provides theoretical justification for GDM on PD-dissimilarities and it accords with the assumptions of the *ED* (environmental diversity) method.

Extending this example, I now will illustrate the application of the *ED* method to the PD-based environmental space of Rintala et al. (Fig. 3). In Fig. 3a, the space (from Fig. 2) is filled with *ED* “demand points”. In Fig. 3b, the *ED* value is calculated as the sum of the distances from each demand point to its nearest sample/site. In Fig. 3c, sample site x is assumed lost and *ED* is re-calculated. In Fig. 3d, alternatively, sample site y is lost and *ED* is re-calculated. We can see from the plots that the loss of sample x clearly results in a greater sum of distances. The loss of sample/site x would imply much greater loss of phylogenetic diversity compared to loss of sample/site y, as indicated by the amount of change in the sums of distances (Fig. 3c,d). This result corresponds to the intuition that sample x, in filling a larger gap in the space relative to sample y, is likely to uniquely represent more features.

## A Simple Graphical Description of *ED* for the Single Gradient Case

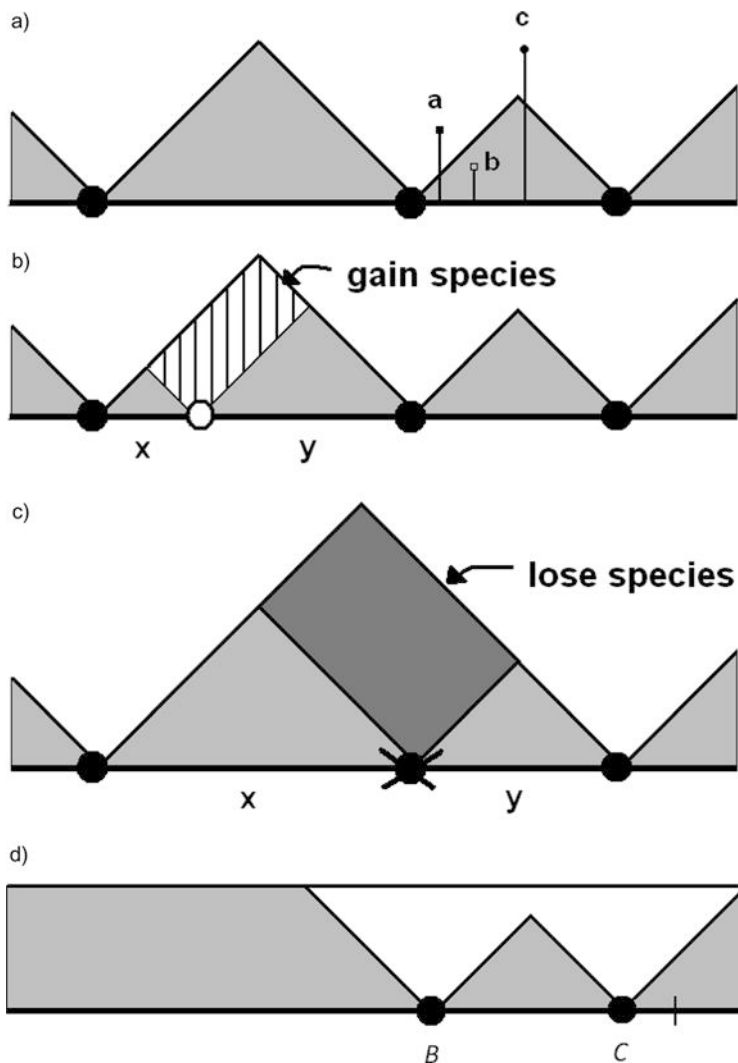
The example in Figs. 2 and 3 illustrated how sites or samples that fill a large gap in environmental space are likely to uniquely represent more branches or features. We can see why *ED counts up* branches or features by looking at a simple one-dimensional gradient and graphical representation of *ED* calculations, which illustrates the link from the counting-up property to *ED* calculations of gains and losses as sites are gained or lost.

Suppose we have an ordination with one gradient (say, a GDM transformation of a climate-related variable; Fig. 4a). Demand points occur continuously along the



**Fig. 3** *ED* analyses for the ordination space based on PD-dissimilarities, from Rintala et al. (Fig. 2). *Black dots* are samples as in Fig. 2 and two of the samples are labelled, *x* and *y*. *Hollow dots* are *ED* demand points. A small number of demand points, uniformly covering the range of samples in the space, are used here to illustrate the method. (a) Ordination space showing samples and demand points. (b) Line segments connect each demand point to its nearest sample, among all samples in a defined subset. The *ED* value is the sum of these distances. Here the subset includes all samples. (c) Sample site *x* is lost from the subset, and *ED* is re-calculated based on the new line segments. (d) Sample site *y* is lost and *ED* is re-calculated based on the new line segments

gradient and define the centers of distribution for features or branches. These features are assumed to have a uniform distribution of range-extents along the gradient (Faith and Walker 1996a). Graphically, the height to the top of the gray area above any demand point (Fig. 4a) reflects the number of features centered at that point that are *not* overlapped by any of the selected sites; these would be features having a



**Fig. 4** (a) A single environmental gradient (*thick black line*) and three selected sites (*black dots*). Each hypothetical branch/lineage, centred at a demand point, graphically is represented in the figure by a point above its demand point, at a vertical distance equal to one-half of its distribution extent on the gradient. Branch/lineage points in the figure are *gray* if no selected site overlaps with the range-extent of the branch/lineage. Branch/lineage 'a' would be captured by the middle site only, branch/lineage 'b' is not sampled by any sites as its extent is too small; it is therefore coloured gray. Branch/lineage 'c' is captured by two sites. The height to the top of the *gray* area above any demand point reflects the total number of branch/lineages centered at that point that are not overlapped by any selected sites. *ED* is the sum of the resulting triangular *gray* areas. When richness varies along the gradient, the corresponding weights on demand points can be interpreted as if we are calculating a volume when counting-up unrepresented branch/lineages to obtain the *ED* score. (b) If the hollow-circle site is added to the selected set indicated by the *black dots*, the *ED* value (number of branch/lineages not represented) will be reduced by the amount equal to the *white-striped* area. This *ED*-complementarity equals  $x \cdot y / 2$ , where  $x$  and  $y$  are distances from the *hollow circle*

range-extent too small to overlap with the nearest site. This number corresponds to the demand point's total contribution to the *ED* value; it indicates the total number of features at that demand point not covered by the selected sites. These demand point contributions form the triangle-shaped gray zones (Fig. 4a), whose total area equals the sum, over all demand points, of the distance from the demand point to its nearest selected site. In this single gradient case, *ED* is simply calculated as the sum of the triangular gray areas. This sum corresponds to the p-median value for the set of selected sites. This link from features to the p-median criterion nicely illustrates how *ED* counts-up features.

The counting-up property is the basis for measures of *ED*-complementarity. An *ED*-complementarity value estimates the number of features gained (lost) when a site is added to (removed from) a set of selected sites (Fig. 4b, c). In this simple single-gradient case, the *ED*-complementarity of a site equals  $\frac{1}{2}$  times the product of its distances to its left and right nearest neighbours (Fig. 4b, c). These basic calculations can be modified by introduction of additional assumptions such as the maximum extent of features along the gradient (Fig. 4d).

The link from the basic unimodal response model to *ED*'s counting-up property provides a basis for comparing *ED* to other methods for transforming dissimilarities to estimates of degree of representation of biodiversity by subsets of sites. The graphical representation will be useful for these comparisons of methods.

## Properties of the Ferrier et al. formula

Ferrier et al (2004) proposed a formula to convert pairwise dissimilarities into “an overall estimate of the proportion of species represented” (e.g. in a set of protected areas). Ferrier et al. predicted “the proportion of species represented (p)” as:

---

**Fig. 4** (continued) site to its left and right nearest neighbours. (c) Removal of the crossed-out site from the selected set (*black dots*) means that the *ED* index of number of branch/lineages not-represented increases by the amount equal to the *dark-gray* area. *ED*-complementarity again equals  $x \cdot y / 2$ , where  $x$  and  $y$  are distances from the crossed-out site to left and right neighbours. (d) A gradient and two selected sites (*black dots*), *B* and *C*, illustrating *ED* options. Branch/lineage extent along the gradient is assumed to not exceed some maximum value. Consequently, selected site, *B*, does not serve demand points along the gradient that are too far away to have any branch/lineages with extent less than or equal to the maximum value that at the same time overlap with *B*. All demand points further away contribute the maximum value to *ED*'s measure of number of branch/lineages not represented. The maximum-value line here is drawn extending across the gradient. The *white* area therefore represents the number of branch/lineages represented by the two selected sites, and the *gray* area corresponds to the number of branch/lineages not represented. The diagram also illustrates another *ED* option. The set of demand points on the right hand side is extended (beyond some initial gradient boundary shown by the *tick mark*) so that selection of site *C* on its own now would imply the capture of the same number of branch/lineages as selection of site *B*.

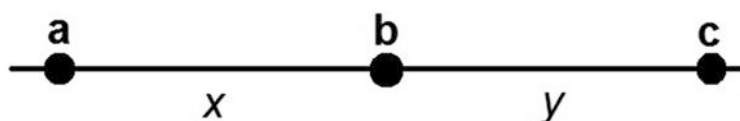
$$p = \frac{\sum_{i=1}^n \left[ \frac{\sum_{j=1}^n ((1-d_{ij}) s_j) / \sum_{j=1}^n (1-d_{ij})}{\sum_{j=1}^n (1-d_{ij})} \right]^z r_i}{\sum_{i=1}^n \frac{r_i}{\sum_{j=1}^n (1-d_{ij})}}$$

where  $n$  is the number of grid cells in a study area,  $r_i$  is the relative richness of each cell and  $d_{ij}$  is the compositional dissimilarity between each pair of cells  $i$  and  $j$ . Further, the state of habitat in each cell (e.g., 1=protected and 0=unprotected) is given by  $s_j$ . The power term,  $z$ , is interpreted as analogous to that in species-area curves (Ferrier et al. 2004).

Ferrier et al. (2004) drew on “principles of the “environmental diversity” (*ED*) approach proposed originally by Faith and Walker (1996a) as a means of assessing the representativeness of protected areas within a continuous environmental or biological space.” Both  $p$  and *ED* intend to convert dissimilarities into a measure of representativeness (e.g., of a subset of sites), but the similarities and differences between the two methods have not been investigated. Allnutt et al. (2008) re-derived the Ferrier et al. measure, and noted the need for comparison with the existing *ED* method: “in contrast to the approach described here, under the *ED* method (Faith and Walker 1996a, b, c), the amount of biodiversity estimated to be retained would depend more on how spread out intact sites are in environmental space, and less on the proportion of habitat retained in any part of this space. Further work is necessary to compare these alternatives in detail.”

Allnutt et al. also noted a concern that was raised in my review of their paper, “Another existing approach to calculate the biodiversity retained, given GDM outputs and habitat state values, is the *ED* method (Faith and Walker 1996; see also Faith et al. 2004). A reviewer of this paper noted that the Ferrier et al. formula relies on the sum of the distances (or similarities) from any site to all the intact sites. A consequence is that selection of additional intact sites will have an attraction to any concentrations (in space) of sites – even allowing further, identical, intact sites to be selected in order to minimise this sum, rather than properly choosing a distant site as a new intact site. In contrast, the *ED* method sees the amount of biodiversity retained as dependent on how spread out the intact sites are in space. Future work may compare these alternatives.”

The graphical presentation of a one-dimensional gradient reveals a critical difference between the two methods. Suppose we have sites along a single environmental gradient as our environmental space (Fig. 5), and there are  $s$  sites at point a, two sites at point b and 1 site at point c. Suppose that one intact site is at point b, and an additional intact site can be located at point c or at point b. We can compare the two scenarios by calculating the numerator of the Ferrier et al. formula (the denominator does not vary). We let  $r_i = 1$  for convenience.



**Fig. 5** A single environmental gradient with  $s$  sites at point  $a$ , two sites at point  $b$  and 1 site at point  $c$ . Distances between sites are given by  $x$  and  $y$ . One intact site is at point  $b$ , and an additional intact site can be located at point  $c$  or at point  $b$

Application of the Ferrier et al. formula will select a duplicate intact site at point  $b$  (over a wide range of values of  $s$  and choice of distances between sites). Suppose, for example, that  $z = .25$ ;  $s = 5$ ;  $x = .4$ ;  $y = .4$ . Then, selecting an additional site at point  $b$  provides a contribution towards  $p$  equal to 4.1, while selecting a site at point  $c$  provides a contribution towards  $p$  equal to only 3.8 (calculations available on request from the author). In contrast, *ED* would select the site at  $c$ , which does increase representation of biodiversity, under the general unimodal model.

It appears that the Ferrier et al. formula for  $p$  can over-estimate the amount of biodiversity that is represented. Put another way, if we started with all sites, the loss of the only site located at point  $c$  along the gradient is seen as less serious than the loss of a duplicate site at point  $b$ . This miss-estimation can have serious consequences for biodiversity conservation; for example, a country could wrongly receive credit for what is in fact a *reduction* in representation of biodiversity.

The Ferrier et al. index was recently applied and recommended by Zerger et al. (2013) as a strategy for building “continental biodiversity information capability”. Given the potential failure of this index to properly assess representativeness, and gains and losses, under our plausible general model, they perhaps incorrectly conclude that “The methodology described by Ferrier et al. (2004) and Allnutt et al. (2008) also allows estimation of the proportion of species expected to be retained in any defined region of interest”. While Zerger et al. refer to species-level analyses, this poor estimation of represented biodiversity will extend to the phylogenetic diversity case, given the direct correspondence of the species and PD/features calculations.

## Maximization of Complementary Richness (MCR)

Similar problems arise for another method that has some similarities to *ED*. Arponen et al. (2008) introduced the ‘maximization of complementary richness’ (MCR) method, described by the authors as the first “successful community-level strategy”. Arponen et al. developed their approach based on an assumption of unimodal responses for species centred at different positions in environmental space. It is logical, therefore, to assess whether their method succeeds in counting-up species or features under this unimodal model.

Arponen et al. did not report the similarities of MCR to the *ED* methods. Without proper comparisons and contrasts with *ED*, it remains unclear whether MCR offers

advantages over the similar *ED* calculations. Their MCR method shares with *ED* several useful properties, including a similar unimodal model, an ordination space, variants of p-median, plus *ED*'s GDM and richness-weighting options (for discussion of *ED* options, see Faith and Walker 1994; Faith and Ferrier 2002; Faith et al. 2003, 2004).

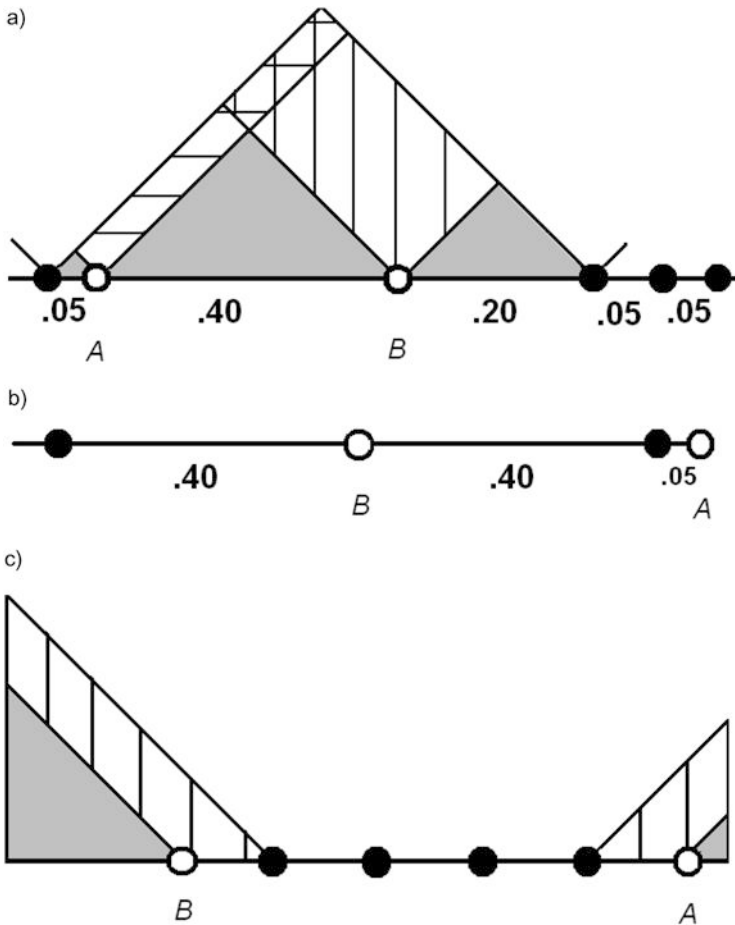
Arponen et al. claimed that MCR has unique properties, but some of these in fact also are shared by *ED*. For example, Arponen et al. (2008, p. 1438) claimed MCR is “different from the previous use of ordinations”, because, in using richness weighting and GDM, it “accounts for gradients in species richness and non-constant turnover rates of community composition”. However, the existing *ED* framework already uses these options (see Faith et al. 2004). Further, MCR, like *ED*, uses points described as “demand” points, served by one or more selected sites. In fact, both methods seek to minimise the degree to which species at demand points are *not* covered by selected sites. Although Arponen et al. describe MCR as *maximising* a summation of ‘ $C_i$ ’ values (and each  $C_i$  value is to reflect the degree to which demand point  $i$  is covered by selected sites), each  $C_i$  equals one minus a product term. Thus, MCR is *minimising* the sum of product terms, and so minimising the degree to which demand points are not covered by selected sites. This property again matches *ED* methods.

Similarities aside, there are critical differences between the two methods. Simple examples will highlight the fact that MCR does have some novel properties relative to *ED* – but these properties de-grade the counting-up property that surely is critical to any truly “successful community-level strategy”.

Novel properties of MCR's basic selection criterion are well-revealed in the simple case where species richness is assumed equal at all sites. MCR then uses the product of a demand point's dissimilarities to all selected sites, and seeks to minimise the sum, over demand points, of these products. Single-gradient scenarios (Fig. 6a) highlight weaknesses of this calculation. Suppose there are two candidate sites for selection, A and B. Selection depends on which site most reduces the MCR product score. Note that when a demand point becomes a selected site, it makes no contribution to the sum of products (as its distance to itself is 0, making its product contribution equal to 0). Selecting site A removes its large product ( $= .05 \times 0.60 \times 0.65 \times 0.70 = 0.014$ ) from the product sum (Fig. 6a). Also, it reduces the product score for non-selected sites (site B), with a reduction equal to  $(1 - 0.4)$  times the previous product value for B of  $(0.45 \times 0.20 \times 0.25 \times 0.30 = 0.007)$ , yielding a reduction of 0.004. Thus, selecting site A reduces the score by about 0.018 ( $0.014 + 0.004$ ). In contrast, selecting site B implies removal of a product term equal to 0.007 (see above), and a reduction in the A product contribution of  $(1 - 0.4)$  times  $0.0137 = 0.008$ . Thus, selecting site B reduces the MCR score by only 0.015, and MCR selects site A.

We also can ask whether site A or B is best to lose (smallest features loss), assuming all sites initially are protected. Loss of B would add a new term to the MCR product sum equal to  $0.45 \times 0.40 \times 0.20 \times 0.25 \times 0.30 = 0.003$ . Loss of A would add a larger term ( $0.05 \times 0.40 \times 0.60 \times 0.65 \times 0.70 = 0.005$ ). MCR prefers to retain site A and lose site B. MCR prefers site A over site B, whether adding or removing sites – yet this does not accord with MCR's own model of random distributions of features in the environmental space.





**Fig. 6** (a) A hypothetical gradient (for example, from GDM) with selected sites (*solid circles*), and two candidate sites for selection, A and B (*hollow circles*). Numbers along gradient are distances between sites. *ED*-complementarity of site B (areas with *vertical stripes*) is 0.045, while that for site A (areas with *horizontal stripes*) is only 0.015, reflecting its close proximity to an already-selected site. *ED* prefers site B, reflecting the greater count in number of branch/lineages gained. In contrast, *MCR*, to minimise its product score, selects site A. For *MCR*, selecting site B reduces the *MCR* product score by only 0.015, while selecting site A reduces the score by a higher value of about 0.018. For *MCR*, the greatest reduction in the product score implies the greatest branch/lineages gain, and so *MCR* prefers site A. For further information, see text. (b) Given two candidate sites (*hollow circles*) and already-selected sites (*solid circles*), *MCR* assigns a higher preference weight to site A, reflecting the large distance from A to the selected site at the other end of the gradient. *ED* identifies site B as the site that would fill the largest gap and provide the greatest gain in branch/lineages representation. (c) There are two candidate sites for selection, A and B (*hollow circles*). *ED*-complementarity values of A and B are shown by respective *striped areas*. Site B, selected by *ED*, provides more new branch/lineages. However, *MCR* cannot distinguish between the two sites

*ED* correctly prefers site *B*, in accord with the unimodal model and counting-up property. *ED*-complementarity for the gain of site *B* (vertical striped area; Fig. 6a) is 0.045, while that for site *A* (horizontal stripes) is only 0.015, reflecting the site's close proximity to an already-selected site. The difference is 0.03, and is the same value when determining the best site to lose, illustrating how *ED* provides a consistent counting-up of features in comparing the two sites under different scenarios. Thus, site *B*, filling a large gap, is expected to contribute more features (Fig. 6a).

This example highlights general MCR weaknesses: a site can be wrongly preferred because MCR is misled by the site's many large dissimilarities to other sites. Arponen et al. attempted to overcome one weakness of their core selection criterion – possible near-duplication of previously selected sites – by applying a down-weighting of those candidate sites close to already-selected sites. The weighting, equal to the product of the site's dissimilarity to all selected sites, does not solve this problem. For example, a site very close to an already-selected site, nevertheless may receive higher weight because it is so far away from other selected sites (Fig. 6b).

MCR's failure to identify gaps is exacerbated by its use of actual sites as demand points (so mimicking 'discrete *ED*'; Faith and Walker 1996a). MCR consequently cannot take into account portions of the environmental space that do not have recorded sites. An example shows how *ED*, but not MCR, will give an edge site deserved priority (Fig. 6c), countering Arponen et al.'s claim that a particular advantage of MCR is that it gives priority to sites on the edge of environmental space.

*ED* succeeds, and MCR fails, in counting-up features under the basic unimodal model. While *ED* successfully has incorporated, in a consistent way, useful options relating to richness, extent of space, GDM, and other options, the MCR calculations degrade the counting-up of features. This contrast between MCR and *ED* has important implications for applications. Suppose we interpret the example (Fig. 6a) as a planning decision, in which the best site, *A* or *B*, will be removed from protection for non-conservation uses. MCR prefers to give away the site (*B*) implying a *greater* features loss. Thus, MCR would be a poor basis for the systematic conservation planning required to reduce rates of biodiversity loss; use of MCR in such conservation planning could inadvertently increase the rate of biodiversity loss. I conclude that MCR, like the Ferrier et al. method, will not provide an effective way to analyse PD-dissimilarities for assessments of PD representation and calculation of gains and losses.

## Discussion

*ED* provides an effective strategy to analyse PD-dissimilarities among areas, and make inferences as if we are counting up branches or features. While well-justified through the link to feature diversity, application of *ED* to date has been frustrated by a lack of synthesis about alternative methods, including inconsistent use of names for methods and miss-representation of basic properties. Araújo et al. (2001, 2003, 2004), Hortal et al. (2009), and Arponen et al. (2008) all have incorrectly characterised "*ED*" as a method using *only* environmental data. Hortal et al. (2009) claimed

to have evaluated the continuous *ED* method of Faith and Walker (1996a), but in fact used a quite different method (see Faith 2011). Recently, Beier and Albuquerque (2015) found strong support for *ED* as a biodiversity surrogate.

The comparison in this study of *ED* to other proposed methods helps to clarify key properties. *ED*, Ferrier et al.'s *p*, and the MCR method share important desirable properties for biodiversity assessment; they transform dissimilarities in order to infer useful information, including the amount of biodiversity represented by subsets of sites. All three methods are based to some degree on the idea of unimodal response. However, among these candidate approaches, *ED* seems to best reflect the plausible underlying model in which elements of biodiversity have general unimodal response to environmental space.

This chapter has attempted to provide some long-overdue comparisons among existing proposed methods, but it is important to note that more comparative evaluations are needed. In the interest of synthesis, I highlight several other methodological issues requiring study.

## ***Hierarchical Clustering***

Faith (2013) recently reviewed the prospects for another strategy, based on a hierarchical clustering of the PD-dissimilarities among sites or samples (including those predicted by GDM). Faith and Walker (1996a), in discussing dissimilarities defined at the species level, had argued that “a robust hierarchical clustering method designed for biotic distribution data, such as flexible-UPGMA with Bray-Curtis dissimilarities, is likely to produce a hierarchy where distances along branches between areas indeed reflect the relative number of species differences.” Faith (2013) suggested an extension of this idea: “This rationale extends to PD-dissimilarities in such a hierarchical clustering, distances along branches between samples reflect the relative difference in the PD of the samples. ....the PD method can be applied to a hierarchy of samples, just as it is applied to a hierarchy (phylogeny) of species. Various PD calculations can be applied to the hierarchies of sites/samples that are based on PD-dissimilarities among samples or sites.” Faith (2013) referred to this method as “PDh”, as it uses the PD calculus, but is applied to a samples/sites hierarchy. The PDh value for a subset of samples/sites indicates the PD of the subset. It is noteworthy that that the suggested hierarchical clustering approach for PDh is a method (Belbin et al. 1993) designed to be compatible with an environmental space and unimodal response.

## ***Persistence Versus Representativeness***

I argued above that Ferrier et al. perhaps inaccurately characterised their formula as estimating “the proportion of species represented”, and I questioned the conclusion of Zenger et al. (2013) that the method of Ferrier et al. (2004) and Allnutt et al.

(2008) “allows estimation of the proportion of species expected to be retained in any defined region of interest.” These problems naturally extend from species-level to the features defined by PD-dissimilarities. Both Allnutt et al. (2008) and Ferrier et al. (2009) have suggested that the Ferrier et al. method contrasts with *ED* because it is intended to address expected persistence, and not just representation. While it seems doubtful that a measure that performs poorly in assessing representation will do well in assessing overall persistence, more work is needed to evaluate whether the Ferrier et al. method provides useful information about biodiversity persistence.

On a positive note, the persistence and the representation goals do not have to be addressed by different frameworks. One *ED* variant, departing from *p*-median, captures expected diversity or persistence in a “probabilistic *ED*” method:

...when we assign probabilities (of expected features persistence or ‘presence’) to sites ... the *p*-median, which strictly depends on nearest neighbours, is relaxed, and the total estimated diversity now depends on summation over ordered nearest neighbours (Faith et al. 2004).

These probabilities form the analogue to the state or condition of habitat in each site  $j$ , given by  $s_j$ , in the Ferrier et al. formula. Given the advantages of *ED* over *p* in the basic representation case, the “probabilistic *ED*” method deserves investigation as an alternative way to integrate state or condition of habitat in sites, for analysis of persistence.

## ***Simulation Methods***

These variants highlight the idea that the critical ingredient of the *ED* framework is unimodal response, reflecting the shared-habitat/shared-features model. Indeed, once we have an environmental space, under this model, we can simulate the sets of branches/features that would correspond, for example, to a nominated subset of sites. Faith et al. (2003) used this approach to map the distributions in geographic space of the hypothetical elements (species or features). This “biodiversity viability analysis” (BVA) uses this spatial information for each element for various biodiversity assessments. Thus, BVA translates information about any inferred element from ordination space to its implied distribution in geographic space (taking advantage of the link that environmental data for all areas provides from ordination space to geographic space). Mokany et al. (2011) provide a method that mimics the *ED*/BVA generation of hypothetical species (or other elements) based on unimodal response and related models. However, their method loses some useful information that BVA/*ED* derives from explicitly sampling from the environmental space under the unimodal response model. Further work is needed to evaluate these methods.

## GEO BON

Future applications may require this full range of *ED* calculations. *ED* is one candidate biodiversity assessment strategy in a new global program for monitoring the status of biodiversity. The Group on Earth Observations Biodiversity Observation Network (GEO BON; Andrefouet et al. 2008) has been developed as a mechanism for gathering and sharing observations regarding biodiversity change. GEO BON is to enhance cooperation among countries to understand changes in biodiversity by monitoring its state and trends. One monitoring strategy in GEO BON will use repeated observations, over time, of changes in the state or condition of sites (e.g., based on remote sensing data). These observations then are integrated with spatial biodiversity models that act as the ‘lens’ for inferences about the corresponding changes in biodiversity (Andrefouet et al. 2008; Faith et al. 2009; Ferrier 2011). The *ED* approach can provide such a biodiversity lens, using available environmental data, genetic, phylogenetic and species data covering multiple taxonomic groups, and GDM to include unsampled sites. In simple applications, *ED* complementarity values can be calculated when localities are judged as newly degraded (or newly protected). Alternatively, the estimates of condition from remote sensing may be interpreted as fractional species losses for localities, calling for methods such as probabilistic *ED*.

One of the GEO BON working groups is tasked with implementing these monitoring strategies to applications assessing change in phylogenetic diversity, over multiple taxonomic groups (including microbial diversity). *ED* methods applied to analyses of PD-dissimilarities (including those describing within-species genetic variation) appear to offer a robust flexible framework for assessments of biodiversity change at this important level of biodiversity.

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Allnutt TF, Ferrier S, Manion G, Powell GVN, Ricketts TH, Fisher BL, Harper GJ et al (2008) A method for quantifying biodiversity loss and its application to a 50-year record of deforestation across Madagascar. *Conserv Lett* 1:173–181
- Andrefouet S, Costello MJ et al (2008) The GEO biodiversity observation network: concept document. GEO – Group on Earth Observations, Geneva, [http://www.earthobservations.com/documents/cop/bi\\_geobon/200811\\_geobon\\_concept\\_document.pdf](http://www.earthobservations.com/documents/cop/bi_geobon/200811_geobon_concept_document.pdf)
- Araújo MB, Densham P, Lampinen R et al (2001) Would environmental diversity be a good surrogate for species diversity? *Ecography* 24:103–110
- Araújo MB, Densham PJ, Humphries CJ (2003) Predicting species diversity with ED: the quest for evidence. *Ecography* 26:380–383
- Araújo MB, Densham PJ, Williams PH (2004) Representing species in reserves from patterns of assemblage diversity. *J Biogeogr* 31:1037–1050
- Arponen A, Moilanen A, Ferrier S (2008) A successful community-level strategy for conservation prioritization. *J Appl Ecol* 45:1436–1445

- Austin MP (1985) Continuum concept, ordination methods, and niche theory. *Ann Rev Ecol Syst* 16:39–61
- Beier P, Albuquerque F (2015) Environmental diversity is a reliable surrogate for species representation. *Conserv Biol* doi:[10.1111/cobi.12508](https://doi.org/10.1111/cobi.12508)
- Belbin L, Faith DP, Milligan G (1993) A comparison of two approaches to beta-flexible clustering. *Multivar Behav Res* 27:417–433
- Faith DP (1989) Homoplasy as pattern: multivariate analysis of morphological convergence in Anseriformes. *Cladistics* 5:235–258
- Faith DP (1992a) Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61:1–10
- Faith DP (1992b) Systematics and conservation: on predicting the feature diversity of subsets of taxa. *Cladistics* 8:361–373
- Faith DP (1994) Phylogenetic pattern and the quantification of organismal biodiversity. *Philos Trans R Soc Lond Ser B* 345:45–58
- Faith DP (1996) Conservation priorities and phylogenetic pattern. *Conserv Biol* 10:1286–1289
- Faith DP (2002) The role of the phylogenetic diversity measure, PD, in bio-informatics: getting the definition right. *Evol Bioinformatics Online* 2006(2):1–7
- Faith DP (2003) Environmental diversity (ED) as surrogate information for species-level biodiversity. *Ecography* 26:374–379
- Faith DP (2011) Attempted tests of the surrogacy value of the ED environmental diversity measures highlight the need for corroboration assessment of surrogacy hypotheses. *Ecol Indic* 11:745–748
- Faith DP (2013) Biodiversity and evolutionary history: useful extensions of the PD phylogenetic diversity assessment framework. *Ann NY Acad Sci* 1289:69–89
- Faith DP (2015a) The PD phylogenetic diversity framework: linking evolutionary history to feature diversity for biodiversity conservation. In: Pellens R, Grandcolas R (eds) *Biodiversity Conservation and Phylogenetic Systematics: preserving our evolutionary heritage in an extinction crisis*. Springer, Dordrecht
- Faith DP (2015b) Phylogenetic diversity, functional trait diversity and extinction: avoiding tipping points and worst-case losses from the tree of life. *Proc Royal Soc London B* 370:20140011
- Faith DP, Ferrier S (2002) Linking beta diversity, environmental variation, and biodiversity assessment. *Science [Online]* 295:636 <http://www.sciencemag.org/cgi/eletters/295/5555/636#504>
- Faith DP, Walker PA (1996a) Environmental diversity: on the best-possible use of surrogate data for assessing the relative biodiversity of sets of areas. *Biodivers Conserv* 5:399–415
- Faith DP, Walker PA (1996b) Integrating conservation and development: effective trade-offs between biodiversity and cost in the selection of protected areas. *Biodivers Conserv* 5:431–446
- Faith DP, Walker PA (1996c) Integrating conservation and development: incorporating vulnerability into biodiversity-assessment of areas. *Biodivers Conserv* 5:417–429
- Faith DP, Minchin PR, Belbin L (1987) Compositional dissimilarity as a robust measure of ecological distance. *Vegetation* 69:57–68
- Faith DP, Walker PA, Ive JR et al (1996) Integrating conservation and forestry production: exploring trade-offs between biodiversity and production in regional land-use assessment. *For Ecol Manag* 85:251–260
- Faith DP, Carter G, Cassis G et al (2003) Complementarity, biodiversity viability analysis, and policy-based algorithms for conservation. *Environ Sci Pol* 6:311–328
- Faith DP, Ferrier S, Walker PA (2004) The ED strategy: how species-level surrogates indicate general biodiversity patterns through an ‘environmental diversity’ perspective. *J Biogeogr* 31:1207–1217
- Faith DP et al (2009) The cladistic basis for the phylogenetic diversity (PD) measure links evolutionary features to environmental gradients and supports broad applications of microbial ecology’s “phylogenetic beta diversity” framework. *Int J Mol Sci* 10:4723–4741

- Faith DP, Magallón S, Hendry AP, Conti E, Yahara T, Donoghue MJ (2010) Ecosystem services: an evolutionary perspective on the links between biodiversity and human-well-being. *Curr Opin Environ Sustain* 2:66–74
- Ferrier S (2002) Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? *Syst Biol* 51:331–363
- Ferrier S (2011) Extracting more value from biodiversity change observations through integrated modelling. *Bioscience* 61:96–97
- Ferrier S, Drielsma M (2010) Synthesis of pattern and process in biodiversity conservation assessment: a flexible whole-landscape modelling framework. *Divers Distrib* 16:386–402
- Ferrier S, Powell GVN, Richardson KS et al (2004) Mapping more of terrestrial biodiversity for global conservation assessment. *Biogeosciences* 54:1101–1109
- Ferrier S, Manion G, Elith J, Richardson KS (2007) Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Divers Distrib* 13:252–264
- Ferrier S, Faith DP, Arponen A, Drielsma M (2009) Community-level approaches to spatial conservation prioritisation. In: Moilanen A, Wilson KA, Possingham HP (eds) *Systematic conservation planning*. Oxford University Press, Oxford, pp 94–109
- Forest F et al (2007) Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature* 445:757–760
- Hortal J, Araújo MB, Lobo JM (2009) Testing the effectiveness of discrete and continuous environmental diversity as a surrogate for species diversity. *Ecol Indic* 9:138–149
- Kelly S, Grenyer R, Scotland RW (2014) Phylogenetic trees do not reliably predict feature diversity. *Divers Distrib* 20:600–612
- Lozupone C, Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* 71:8228–8235
- Mokany K, Harwood TD, Overton JM, Barker GM, Ferrier S (2011) *Ecol Lett* 14:1043–1051
- Nipperess DA, Faith DP, Barton K (2010) Resemblance in phylogenetic diversity among ecological assemblages. *J Veg Sci* 21:809–820
- Rintala H et al (2008) Diversity and seasonal dynamics of bacterial community in indoor environment. *BMC Microbiol* 8:56
- Rosauer D, Ferrier S, Williams KJ, Manion G, Keogh S, Laffan SW (2013) Phylogenetic generalised dissimilarity modelling: a new approach to analysing and predicting spatial turnover in the phylogenetic composition of communities. *Ecography* 37:21–32
- Swenson NG (2011) Phylogenetic beta diversity metrics, trait evolution and inferring the functional beta diversity of communities. *PLoS ONE* 6(6):e21264. doi:[10.1371/journal.pone.0021264](https://doi.org/10.1371/journal.pone.0021264)
- Walker PA, Faith DP (1994) DIVERSITY-PD: procedures for conservation evaluation based on phylogenetic diversity. *Biodivers Lett* 2:132–139
- Zerger A, Williams KJ, Nicholls M, Belbin L, Harwood T, Bordas V, Ferrier S, Perkins G (2013) Biodiversity profiling: components of a continental biodiversity information capability, vol 2, Environmental Information Program Publication. Bureau of Meteorology, Canberra. 72 pp

# Phylogenetic Diversity Measures and Their Decomposition: A Framework Based on Hill Numbers

Anne Chao, Chun-Huo Chiu, and Lou Jost

**Abstract** Conservation biologists need robust, intuitive mathematical tools to quantify and assess patterns and changes in biodiversity. Here we review some commonly used abundance-based species diversity measures and their phylogenetic generalizations. Most of the previous abundance-sensitive measures and their phylogenetic generalizations lack an essential property, the replication principle or doubling property. This often leads to inconsistent or counter-intuitive interpretations, especially in conservation applications. Hill numbers or the “effective number of species” obey the replication principle and thus resolve many of the interpretational problems. Hill numbers were recently extended to incorporate phylogeny; the resulting measures take into account phylogenetic differences between species while still satisfying the replication principle. We review the framework of phylogenetic diversity measures based on Hill numbers and their decomposition into independent alpha and beta components. Both additive and multiplicative decompositions lead to the same classes of normalized phylogenetic similarity or differentiation measures. These classes include multiple-assemblage phylogenetic generalizations of the Jaccard, Sørensen, Horn and Morisita-Horn measures. For two assemblages, these classes also include the commonly used *UniFrac* and *PhyloSør* indices as special cases. Our approach provides a mathematically rigorous, self-consistent, ecologically meaningful set of tools for conservationists who must assess the phylogenetic diversity and complementarity of potential protected areas. Our framework is applied to a real dataset to illustrate (i) how to use phylogenetic diversity profiles to completely convey species abundances and phylogenetic information among species in an assemblage; and (ii) how to use phylogenetic similarity (or differentiation) profiles to assess phylogenetic resemblance or difference among multiple assemblages.

**Keywords** Diversity • Diversity decomposition • Hill numbers • Phylogenetic diversity • Replication principle • Species diversity

---

A. Chao (✉) • C.-H. Chiu  
Institute of Statistics, National Tsing Hua University, Hsin-Chu 30043, Taiwan  
e-mail: [chao@stat.nthu.edu.tw](mailto:chao@stat.nthu.edu.tw)

L. Jost  
EcoMinga Foundation, Baños, Tungurahua, Ecuador



## Introduction

Many of the most pressing and fundamental questions in biodiversity conservation require robust and sensible measures for quantifying and assessing changes in biodiversity. Many environmental and monitoring projects also require objective and meaningful similarity (or differentiation) measures to compare the diversities of multiple assemblages and their degree of complementarity in order to best conserve genetic, species, and ecosystem diversity. An enormous number of diversity measures and related similarity (or differentiation) indices have been proposed, not only in ecology but also in genetics, economics, information science, linguistics, physics, and social sciences, among others. See Magurran (2004) and Magurran and McGill (2011) for overviews.

In traditional species diversity measures, all species are considered to be equally different from each other; only species richness and abundances are involved. There are two general approaches: parametric and non-parametric (Magurran 2004). Parametric approaches assume a particular species abundance distribution (such as the lognormal or gamma) or a species rank abundance distribution (such as the negative binomial or log-series), and then use the parameters (e.g., Fisher's alpha) of the distribution to quantify diversity. However, these methods often do not perform well and the results are un-interpretable unless the "true" species abundance distribution is known (Colwell and Coddington 1994; Chao 2005). The parametric model also does not permit meaningful comparison of assemblages with different abundance distributions. For example, a log-normal abundance model cannot be compared to an assemblage whose abundance distribution follows a gamma distribution. Non-parametric methods make no assumptions about the distributional form of the underlying species abundance distribution. The most widely used abundance-sensitive non-parametric measures have been the Shannon entropy and the Gini-Simpson index. These two measures, along with species richness were integrated into a class of measures called generalized entropies (Havrdra and Charvat 1967; Daróczy 1970; Patil and Taillie 1979; Tsallis 1988; Keylock 2005), which will be briefly reviewed in this chapter.

How to quantify abundance-based species diversity in an assemblage has been one of the most controversial issues in community ecology (e.g. Hurlbert 1971; Routledge 1979; Patil and Taillie 1982; Purvis and Hector 2000; Jost 2006, 2007; Jost et al. 2010). There have also been intense debates on the choice of diversity partitioning schemes; see Ellison (2010) and the Forum that follows it. Surprisingly, all authors in that forum achieved a consensus on the use of Hill numbers, also called "effective number of species", as the best choice to quantify abundance-based species diversity. Hill numbers are a mathematically unified family of diversity indices (differing among themselves only by a parameter  $q$ ) that incorporate species richness and species relative abundances. They were first used in ecology by MacArthur (1965, 1972), developed by Hill (1973), and recently reintroduced to ecologists by Jost (2006, 2007).

Hill numbers obey the *replication principle* or *doubling property*, an essential mathematical property that capture biologists' notion of diversity (MacArthur 1965; Hill 1973). This property requires that if we have  $N$  equally diverse, equally large assemblages with no species in common, the diversity of the pooled assemblage must be  $N$  times the diversity of a single group. In other words, they are linear with respect to addition of equally-common species. We will review different versions of this property later. Classical diversity measures, such as Shannon entropy and the Gini-Simpson index, do *not* obey this principle and can lead to inconsistent or counter-intuitive interpretations, especially in conservation applications (Jost 2006, 2007). Hill numbers resolve many of the interpretational problems caused by classical diversity indices. Diversity measures that obey the replication principle yield self-consistent assessment in conservation applications, have intuitively-interpretable magnitudes, and can be meaningfully decomposed. In this chapter, Hill numbers are adopted as a general framework for quantifying and partitioning diversities.

Pielou (1975, p. 17) was the first to notice that traditional abundance-based species diversity measures could be broadened to include phylogenetic, functional, or other differences between species. We here concentrate on phylogenetic differences, though our framework can also be extended to functional traits (Tilman 2001; Petchey and Gaston 2002; Weiher 2011). For conservation purposes, an assemblage of phylogenetically divergent species is more diverse than an assemblage consisting of closely related species, all else being equal. Phylogenetic differences among species can be based directly on their evolutionary histories, either in the form of taxonomic classification or well-supported phylogenetic trees (Faith 1992; Warwick and Clarke 1995; McPeck and Miller 1996; Crozier 1997; Helmus et al. 2007; Webb 2000; Webb et al. 2002; Pavoine et al. 2010; Ives and Helmus 2010, 2011; Vellend et al. 2011; Cavender-Bares et al. 2009, 2012 among others). Three special issues in *Ecology* were devoted to integrating ecology and phylogenetics; see McPeck and Miller (1996), Webb et al. (2006), and Cavender-Bares et al. (2012) and papers in each issue. Phylogenetic diversity measures are especially relevant for conservation applications, since they quantify the amount of evolutionary history preserved by the assemblage; see Lean and MacLaurin (chapter “[The Value of Phylogenetic Diversity](#)”).

The most widely used phylogenetic metric is Faith's phylogenetic diversity (PD) (Faith 1992) which is defined as the sum of the branch lengths of a phylogenetic tree connecting all species in the target assemblage. As shown in Chao et al. (2010), Faith's PD can be regarded as a phylogenetic generalization of species richness. The rarefaction formula for Faith's PD was developed by Nipperess and Matsen (2013) and Nipperess (chapter “[The Rarefaction of Phylogenetic Diversity: Formulation, Extension and Application](#)”). Recently, Chao et al. (2015) derived an integrated sampling, rarefaction, and extrapolation methodology to compare Faith's PD of a set of assemblages. Like species richness, Faith's PD does not consider species abundances. For some conservation applications, the mere presence or absence of a species is all that matters, or all that can be determined from the available data. In those cases, Faith's PD is a good measure of phylogenetic diversity. However, there

are important advantages to incorporating abundance information into phylogenetic diversity measures for conservation. For example, some human impacts can result in the phylogenetic simplification of an ecosystem, reducing the population shares of phylogenetically distinct species relative to typical species. An abundance-based measure can catch this effect before it leads to actual extinctions.

Ecosystem simplification may be worthy of conservation concern even if it does not lead to extinctions of focal organisms. Often, the focal organisms for conservation represent a tiny fraction of the ecosystem's biomass or richness. Each focal species will be tied to a web of non-focal species whose abundances are not usually monitored (e.g., insects). All else being equal, a more equitable distribution of the abundances of focal organisms will be able to support a more diverse, robust and stable set of non-focal species. Faith (chapter "[Using Phylogenetic Dissimilarities Among Sites for Biodiversity Assessments and Conservation](#)") rightly argues that phylogenetic diversity is a good proxy for functional diversity. Therefore an ecosystem with a more equitable distribution of abundance across phylogenetic lineages should also exhibit greater functional complexity (per interaction between individuals) than an ecosystem whose phylogenetically unusual elements are rare. If we have to prioritize such ecosystems, the more phylogenetically equitable one, which thoroughly integrates diverse lineages, should be preferred. In addition to being more resistant to lineage extinctions, a complex, well-integrated ecosystem may be worth preserving in and of itself, above and beyond its component species; conservation is not just about species. Evolution may take a different course in ecosystems whose members are constantly surprised by their interactions compared with an ecosystem whose interactors are highly predictable. These conservation goals – robustness against extinction of distinctive lineages, and preservation of well-integrated ecosystems with unique future option values – require phylogenetic diversity measures that incorporate species importance values.

Rao's quadratic entropy  $Q$  (Rao 1982), a generalization of the Gini-Simpson index, was the first diversity measure that accounts for both phylogeny and species abundances. The phylogenetic entropy  $H_p$  (Allen et al. 2009) extends Shannon entropy to incorporate phylogenetic distances among species. Since Shannon entropy and the Gini-Simpson index do not obey the replication principle, neither do their phylogenetic generalizations. These generalizations will therefore have the same interpretational problems as their parent measures; see Chao et al. (2010, their Supplementary Material) for examples.

Chao et al. (2010) extended Hill numbers and related similarity measures to incorporate phylogeny. The new phylogenetic Hill numbers obey a generalized replication principle. Their measures were subsequently extended by Faith and Richards (2012) and Faith (2013). Both the original Hill numbers and their phylogenetic generalizations facilitate diversity decomposition (Jost 2007; Chiu et al. 2014). As with the original Hill numbers, both additive and multiplicative decompositions of phylogenetic Hill numbers lead to the same classes of similarity (or differentiation) measures. Hill numbers therefore provide a unified framework to quantify both abundance-based and phylogenetic diversity.

In this chapter, we first briefly review the classic abundance-based species diversity measures (section "[Generalized Entropies](#)") and their phylogenetic

generalizations (section “[Phylogenetic generalized entropies](#)”) for an assemblage. Then we focus on the framework of Hill numbers (section “[Hill numbers and the replication principle](#)”), phylogenetic Hill numbers (section “[Phylogenetic Hill numbers and related measures](#)”) and related phylogenetic diversity measures. We also discuss the replication principle and its phylogenetic generalization (section “[Replication principle for phylogenetic diversity measures](#)”). For multiple assemblages, we review the diversity decomposition based on phylogenetic diversity measures (section “[Decomposition of phylogenetic diversity measures](#)”). The associated phylogenetic similarity and differentiation measures are then presented (section “[Normalized phylogenetic similarity measures](#)”). We use a real example for illustration (section “[An example](#)”). Our practical recommendations are provided in section “[Conclusion](#)”.

## Classic Measures and Their Phylogenetic Generalizations

### *Generalized Entropies*

The species richness of an assemblage is a simple count of the number of species present. It is the most intuitive and frequently used measure of biodiversity, and is a key metric in conservation biology (MacArthur and Wilson 1967; Hubbell 2001; Magurran 2004). However, it does not incorporate any information about the abundances of species, and it is a very hard number to estimate accurately from small samples (Colwell and Coddington 1994; Chao 2005; Gotelli and Colwell 2011).

Shannon entropy is a popular classical abundance-based diversity index and has been used in many disciplines. Shannon entropy is

$$H_{Sh} = -\sum_{i=1}^S p_i \log p_i, \quad (1a)$$

where  $S$  is the number of species in the assemblage, and the  $i$ th species has relative abundance  $p_i$ . Shannon entropy gives the uncertainty in the species identity of a randomly chosen individual in the assemblage. Another popular measure is the Gini-Simpson index,

$$H_{GS} = 1 - \sum_{i=1}^S p_i^2, \quad (1b)$$

which gives the probability that two randomly chosen individuals belong to different species. These two abundance-sensitive measures, along with species richness, can be united into a single family of *generalized entropy*:

$${}^qH = \left(1 - \sum_{i=1}^S p_i^q\right) / (q-1). \quad (1c)$$

The parameter  $q$  determines the sensitivity of the measure to the relative frequencies of the species. When  $q=0$ ,  ${}^qH$  becomes  $S-1$ ; When  $q$  tends to 1,  ${}^qH$  tends to Shannon entropy. When  $q=2$ ,  ${}^qH$  reduces to the Gini-Simpson index. This family was found many times in different disciplines (Havrdra and Charvat 1967; Daróczy 1970; Patil and Taillie 1979; Tsallis 1988; Keylock 2005). There are many other families of generalized entropies, notably the Rényi entropies (Rényi 1961).

Although the traditional abundance-sensitive generalized entropies and their special cases have been useful in many disciplines (e.g., see Magurran 2004), they do not behave in the same intuitive linear way as species richness. In ecosystems with high diversity, mass extinctions hardly affect their values (Jost 2010). They also lead to logical contradictions in conservation biology, because they do not measure a conserved quantity (e.g., under a given conservation plan, the proportion of “diversity” lost and the proportion preserved can *both* be 90 % or more); see Jost (2006, 2007) and Jost et al. (2010). Thus, changes in their magnitude cannot be properly compared or interpreted. Also, the main measure of similarity in the additive approach for traditional measures, the within-group or “alpha” diversity divided by the total or “gamma” diversity, does not actually quantify the compositional similarity of the assemblages under study. This ratio can be arbitrarily close to unity (supposedly indicating high similarity) even when the assemblages being compared have no species in common. Finally, these measures each use different units (e.g., the Gini-Simpson index is a probability whereas Shannon entropy is in units of information), so they cannot be compared with each other. All these problems are consequences of their failure to satisfy the replication principle. Hill numbers obey the replication principle and resolve all these problems; see section “Hill numbers and the replication principle”.

## Phylogenetic Generalized Entropies

The classic measures reviewed in section “Generalized Entropies” were extended to incorporate phylogenetic distance between species. As mentioned in the Introduction and will be shown in section “Phylogenetic Hill numbers and related measures”, Faith’s PD can be regarded as a phylogenetic generalization of species richness.

Rao’s *quadratic entropy* takes account of both phylogeny and species abundances (Rao 1982):

$$Q = \sum_{i,j} d_{ij} p_i p_j, \quad (2a)$$

where  $d_{ij}$  denotes the phylogenetic distance (in years since divergence, number of DNA base changes, or other metric) between species  $i$  and  $j$ , and  $p_i$  and  $p_j$  denote the relative abundance of species  $i$  and  $j$ . This index measures the average phylogenetic distance between any two individuals randomly selected from the assemblage. Rao’s  $Q$  represents a phylogenetic generalization of the Gini-Simpson index because

in the special case of no phylogenetic structure (all species are equally related to one another),  $d_{ii}=0$  and  $d_{ij}=1$  ( $i \neq j$ ), it reduces to the Gini-Simpson index.

The *phylogenetic entropy*  $H_p$  is a generalization of Shannon's entropy to incorporate phylogenetic distances among species (Allen et al. 2009):

$$H_p = -\sum_i L_i a_i \log a_i \quad (2b)$$

where the summation is over all branches of a rooted phylogenetic tree,  $L_i$  is the length of branch  $i$ , and  $a_i$  denotes the summed relative abundance of all species descended from branch  $i$ .

For ultrametric trees, Faith's PD, Allen et al.'s  $H_p$ , and Rao's  $Q$  can be united into a single parametric family of *phylogenetic generalized entropies* (Pavoine et al. 2009):

$${}^q I = \left( T - \sum_i L_i a_i^q \right) / (q-1). \quad (2c)$$

Here,  $L_i$  and  $a_i$  are defined in Eq. (2b) and  $T$  is the age of the root node of the tree. Then  ${}^0 I$  = Faith's PD minus  $T$ ;  ${}^1 I$  is identical to Allen et al.'s entropy  $H_p$  given in Eq. (2b); and  ${}^2 I$  is identical to Rao's quadratic entropy  $Q$  given in Eq. (2a). In the special case that  $T=1$  (the tree height is normalized to unit length) and all branches have unit length, then the phylogenetic generalized entropy reduces to the classical generalized entropy defined in Eq. (1c), with species relative abundances  $\{p_1, p_2, \dots, p_s\}$  as the tip-node abundances.

The abundance-sensitive ( $q>0$ ) phylogenetic generalized entropies provide useful information, but they do not obey the replication principle and thus have the same interpretational problems as their parent measures. This motivated Chao et al. (2010) to extend Hill numbers to phylogenetic Hill numbers, which obey the replication principle; see section “[Phylogenetic Hill numbers and related measures](#)”.

## Hill Numbers and Their Phylogenetic Generalizations

### *Hill Numbers and the Replication Principle*

Pioneering work by Kimura and Crow (1964) in genetics and MacArthur (1965) in ecology showed that the Shannon and Gini-Simpson measures can be easily converted to “effective number of species” (i.e., the number of equally abundant species that are needed to give the same value of the diversity measure), which use the same units as species richness. Shannon entropy can be converted by taking its exponential, and the Gini-Simpson index can be converted by the formula  $1/(1-H_{GS})$ . Hill (1973) integrated species richness and the converted Shannon and Gini-Simpson

measures into a class of diversity measures called “*Hill numbers*” of order  $q$ , or the “*effective number of species*”, defined as

$${}^qD = \left( \sum_{i=1}^S p_i^q \right)^{1/(1-q)}, \quad q \geq 0, q \neq 1. \quad (3a)$$

This measure is undefined for  $q=1$ , but its limit as  $q$  tends to 1 exists and gives

$${}^1D = \lim_{q \rightarrow 1} {}^qD = \exp \left( - \sum_{i=1}^S p_i \log p_i \right) = \exp(H_{Sh}). \quad (3b)$$

The relationship between Hill number of order  $q$  ( $q \neq 1$ ) and the generalized entropy can be expressed as

$${}^qD = \left[ 1 - (q-1)({}^qH) \right]^{1/(1-q)}. \quad (3c)$$

When  $q=0$ , the species abundances do not count at all and  ${}^0D=S$  is obtained. When  $q=1$ , the species are weighed in proportion to their frequencies, and the measure  ${}^1D$  (in Eq. (3b)) can be interpreted as the effective number of common or “typical” species (i.e., species with typical abundances) in the assemblage. When  $q=2$ , abundant species are favored and rare species are discounted; the measure  ${}^2D$  becomes the inverse Simpson concentration. The measure  ${}^2D$  can be interpreted as the effective number of dominant or very abundant species in the assemblage. In general, if  ${}^qD=x$ , then the diversity of order  $q$  of this community is the same as that of an idealized reference community with  $x$  equally abundant species. All Hill numbers are in units of “species”. It is thus possible to plot them on a single graph as a continuous function of the parameter  $q$ . This *diversity profile* characterizes the species-abundance distribution of an assemblage and provides complete information about its diversity. The steepness of its slope graphically illustrates the degree of dominance in the assemblage. An example is given in section “[An example](#)”.

Hill numbers differ fundamentally from Shannon entropy and the Gini-Simpson index in that they obey the replication principle. Hill (1973) proved a weak version of the doubling property: if two completely distinct assemblages (i.e., no species in common) have identical relative abundance distributions, then the Hill number doubles if the assemblages are combined with equal weights. Chiu et al. (2014, their Appendix B) recently proved a strong version of the doubling property: if two completely distinct assemblages have identical Hill numbers of order  $q$  (relative abundance distributions may be different, unlike the weak version), then the Hill number of the same order doubles if the two assemblages are combined with equal weights. Species richness is a Hill number (with  $q=0$ ) and obeys both versions of the doubling property, but most other diversity indices do not obey even the weak version. Because Hill numbers obey this replication principle, changes in their magnitude have simple interpretations, and the ratio of alpha diversity to gamma diversity

accurately reflects the compositional similarity of the communities. The replication principle is best known in economics, where it has long been recognized as an important property of concentration and diversity measures (Hannah and Kay 1977). In ecology, the doubling property has been extensively discussed by many authors (MacArthur 1965, 1972; Hill 1973; Whittaker 1972; Routledge 1979; Peet 1974; Jost 2006, 2007, 2009; Ricotta and Szeidl 2009; Jost et al. 2010) and has been extended to phylogenetic measures (Chao et al. 2010); see below.

## Phylogenetic Hill Numbers and Related Measures

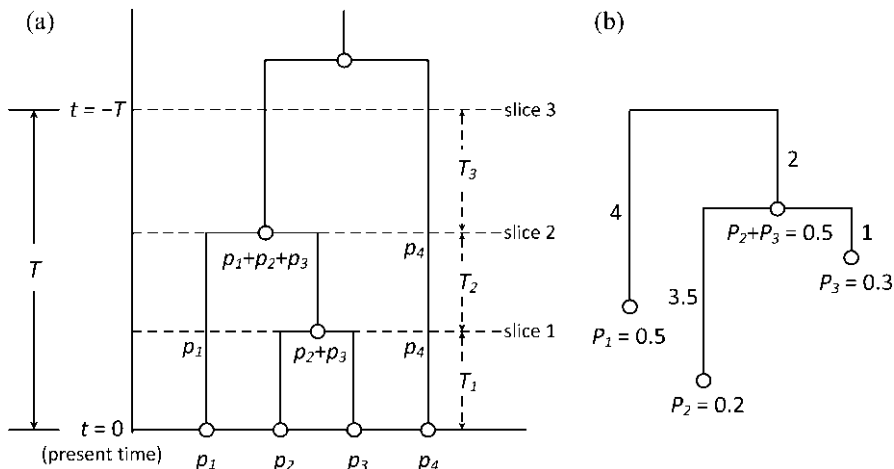
When the branch lengths are proportional to divergence time, all branch tips are the same distance from the root (the first node). Such trees are called “ultrametric” trees. We first discuss the phylogenetic diversity measures for ultrametric trees. The phylogenetic Hill numbers developed by Chao et al. (2010) for an ultrametric tree can be intuitively explained as the Hill number of a time-average of a tree’s generalized entropy over some evolutionary time interval of interest. Suppose the phylogenetic tree for an assemblage is calibrated to some relative or absolute timescale. We can slice this phylogenetic tree at any time  $t$  in the past; see the left panel of Fig. 1 (reproduced from Chao et al. 2010) for illustration and details about how to deal with shared lineages. The number of lineages at that time is the number of branch cuts, and the relative importance of each of these lineages for the present-day assemblage is the sum of the relative abundances of the branch’s descendants in the present-day assemblage. Using these relative importance values, we can calculate the generalized entropy of order  $q$  for the slice. The mean of these entropies, beginning at time  $-T$  (i.e.,  $T$  years before present) and continuing until the present, is converted to a Hill number using Eq. (3c). This is the *phylogenetic Hill number*, which conveys information about the shape of the tree over the time interval of interest. Chao et al. (2010) symbolize it as  ${}^q\bar{D}(T)$ , and also refer to it as the *mean phylogenetic diversity of order  $q$  over  $T$  years* (or simply the *mean diversity* for the interval  $[-T, 0]$ ):

$${}^q\bar{D}(T) = \left\{ \sum_{i \in \mathbf{B}_T} \frac{L_i}{T} a_i^q \right\}^{1/(1-q)} = \frac{1}{T} \left\{ \sum_{i \in \mathbf{B}_T} L_i \left( \frac{a_i}{T} \right)^q \right\}^{1/(1-q)}, \quad q \geq 0, q \neq 1; \quad (4a)$$

$${}^1\bar{D}(T) = \lim_{q \rightarrow 1} {}^q\bar{D}(T) = \exp \left[ - \sum_{i \in \mathbf{B}_T} \frac{L_i}{T} a_i \log a_i \right], \quad (4b)$$

where  $\mathbf{B}_T$  is the set of all branches in the time interval  $[-T, 0]$ ,  $L_i$  is the length of branch  $i$  in the set  $\mathbf{B}_T$ , and  $a_i$  is the total relative abundance descended from branch  $i$ . The mean diversity  ${}^q\bar{D}(T)$  is interpreted as “the effective number of equally abundant and *equally distinct* lineages all with branch lengths  $T$  during the time





**Fig. 1** (a) A hypothetical ultrametric rooted phylogenetic tree with four species. Three different slices corresponding to three different times are shown. For a fixed  $T$  (not restricted to the age of the root), the nodes divide the phylogenetic tree into segments 1, 2 and 3 with duration (length)  $T_1$ ,  $T_2$  and  $T_3$ , respectively. In any moment of segment 1, there are four species (i.e. four branches cut); in segment 2, there are three species; and in segment 3, there are two species. The mean species richness over the time interval  $[-T, 0]$  is  $(T_1 / T) \times 4 + (T_2 / T) \times 3 + (T_3 / T) \times 2$ . In any moment of segment 1, the species relative abundances (i.e. node abundances correspond to the four branches) are  $\{p_1, p_2, p_3, p_4\}$ ; in segment 2, the species relative abundances are  $\{g_1, g_2, g_3\} = \{p_1, p_2+p_3, p_4\}$ ; in segment 3, the species relative abundances are  $\{h_1, h_2\} = \{p_1+p_2+p_3, p_4\}$ . (b) A hypothetical non-ultrametric tree. Let  $\bar{T}$  be the weighted (by species abundance) mean of the distances from root node to each of the terminal branch tips.  $\bar{T} = 4 \times 0.5 + (3.5 + 2) \times 0.2 + (1 + 2) \times 0.3 = 4$ . Note  $\bar{T}$  is also the weighted (by branch length) total node abundance because  $\bar{T} = 0.5 \times 4 + 0.2 \times 3.5 + 0.3 \times 1 + 0.5 \times 2 = 4$ . Conceptually, the ‘branch diversity’ is defined for an assemblage of four branches: each has, respectively, relative abundance  $0.5 / \bar{T} = 0.125$ ,  $0.2 / \bar{T} = 0.05$ ,  $0.3 / \bar{T} = 0.075$  and  $0.5 / \bar{T} = 0.125$ ; and each has, respectively, weight (i.e. branch length) 4, 3.5, 1 and 2. This is equivalent to an assemblage with 10.5 equally weighted ‘branches’: there are four branches with relative abundance  $0.5 / \bar{T} = 0.125$ ; 3.5 branches with relative abundance  $0.2 / \bar{T} = 0.05$ ; one branch with relative abundance  $0.3 / \bar{T} = 0.075$  and two branches with relative abundance  $0.5 / \bar{T} = 0.125$  (This figure is reproduced from Fig. 1 of Chao et al. 2010)

interval from  $T$  years ago to the present”. Here “equally distinct” also implies that the phylogenetic distance between any two species is  $T$ , so lineages are completely distinct (i.e., there are no shared branches).

The phylogenetic Hill numbers are invariant to the units used to measure branch lengths. When all lineages are completely distinct, the measure  ${}^q\bar{D}(T)$  reduces to

the Hill numbers  ${}^qD = \left( \sum_i a_i^q \right)^{1/(1-q)}$ . This includes the special case that  $T$  tends to

zero, i.e., the case that we ignore phylogeny and only consider the present-day community. This shows that the framework based on Hill numbers provides a unified approach to integrate abundances and phylogeny. Also, here we have a simple idealized reference tree to understand the value of  ${}^q\bar{D}(T) = z$  for an arbitrary tree: the

mean phylogenetic diversity of the tree over the time period  $[-T, 0]$  is the same as the diversity of an idealized assemblage consisting of  $z$  equally abundant and equally distinct lineages all with branch length  $T$ .

For  $q=0$ , when  $T$  is chosen as the age of the root node, we have  ${}^0\bar{D}(T) = \text{Faith's PD} / T$ , which can be interpreted as *lineage richness*. Faith's PD can thus be regarded as a phylogenetic generalization of species richness. We can roughly interpret  ${}^1\bar{D}(T)$  as the effective number of *common lineages*, and  ${}^2\bar{D}(T)$  as the effective number of *dominant lineages* in the time period  $[-T, 0]$ . When  $T$  is chosen as the age of the root node, a simple relationship exists between phylogenetic entropy  $H_p$  (Allen et al. 2009) and the measure  ${}^1\bar{D}(T)$ :

$${}^1\bar{D}(T) = \exp(H_p / T). \quad (4c)$$

For  $q=2$ , when  $T$  is chosen as the age of the root node, there is a simple relationship between our measures and the widely used Rao's quadratic entropy  $Q$  (Chao et al. 2010):

$${}^2\bar{D}(T) = \frac{1}{1 - Q / T}. \quad (4d)$$

The *branch or phylogenetic diversity*  ${}^qPD(T)$  of order  $q$  during the time interval from  $T$  years ago to the present is defined as the product of  ${}^q\bar{D}(T)$  and  $T$ . It quantifies the amount of evolutionary history on the system over the interval  $[-T, 0]$ , or “the effective total branch-length” (Chao et al. 2010):

$${}^qPD(T) = T \times {}^q\bar{D}(T) = \left\{ \sum_{i \in \mathbf{B}_T} L_i \left( \frac{a_i}{T} \right)^q \right\}^{1/(1-q)}. \quad (5a)$$

$${}^1PD(T) = \lim_{q \rightarrow 1} {}^qPD(T) = \exp \left[ - \sum_{i \in \mathbf{B}_T} L_i \frac{a_i}{T} \log \left( \frac{a_i}{T} \right) \right]. \quad (5b)$$

If  $q=0$ , and  $T$  is age of the root node, then  ${}^0PD(T)$  reduces to Faith's PD, regardless of branching pattern or abundances. As explained by Chao et al. (2010), we could imagine that all the branch segments in the interval  $[-T, 0]$  form a single assemblage with relative abundance set  $\{a_i/T; i \in \mathbf{B}_T\}$ . In this assemblage, for each  $i$  there are  $L_i$  “branches” with relative abundance  $a_i/T$ . Then the Hill number of order  $q$  for this assemblage is exactly the branch diversity  ${}^qPD(T)$  given in Eq. (5a). Dividing this Hill number by  $T$ , we obtain  ${}^q\bar{D}(T)$  given in Eq. (4a). Note in our framework that  ${}^qPD(T)$  is truly a class of Hill numbers (“the effective number of lineage-years”), whereas  ${}^q\bar{D}(T)$  (“the effective number of lineages”) denotes a (generalized) mean of Hill numbers. See Faith and Richards (2012) and Faith (2013) for extensions of the measure  ${}^qPD(T)$ .

Unlike previous phylogenetic diversity measures developed in the literature,  ${}^q\bar{D}(T)$  and  ${}^qPD(T)$  depend explicitly on two parameters, the abundance sensitivity

parameter  $q$  and the time perspective (or time-depth) parameter  $T$ . The reasons we need this time-depth parameter and our suggestion to choose a perspective time are given as follows.

1. When we compare the phylogenetic diversities of several assemblages based on the measures  ${}^q\bar{D}(T)$  and  ${}^qPD(T)$ , all measures should refer to the same time periods to make meaningful comparisons. That is, the time-depth  $T$  should be kept as the same for all assemblages. Therefore, a parameter is required to specify the time-depth.
2. The choice of time perspective should reflect an investigator's aims and facilitate comparisons with other studies. We suggest that at least two selected time perspectives should be included:  $T=0$ , and  $T$ =the age of the root node of a phylogenetic tree connecting all species in the study. For the case of  $T=0$ , the phylogeny is ignored and the diversity profile reduces to the profile in the present-day assemblage based on the ordinary Hill numbers. If we choose  $T$  to be the age of the oldest node in the tree, we recover some of the standard measures of phylogenetic diversity (see Eqs. (4c) and (4d)).
3. As suggested in Chiu et al. (2014), other time perspectives can be selected, such as  $T$ =the age of the node at which the group of interest diverges from the rest of the species. This choice of  $T$  is independent of the species actually sampled, so it allows statistically robust comparisons across investigations and regions (unlike the conventional choice of  $T$  as the root node of the tree containing the species actually observed). This choice also provides an accurate measure of the proportion of a taxonomic group's evolutionary history preserved in a given assemblage. Another choice is the time of the most recent common ancestor of all taxa alive today. Other choices may be made, depending on the purpose of an investigation. The formula in Chiu et al. (2014, p. 42) can be used to convert phylogenetic diversity from one temporal perspective to another.

To see how the measures vary with  $q$  and time perspective  $T$ , we recommend using two types of profiles to completely characterize phylogenetic tree information and species abundances as described below. See section “An example” for examples. (1) The first type of diversity profile is obtained by plotting  ${}^qPD(T)$  or  ${}^q\bar{D}(T)$  as a function of order  $q$  as  $q$  varies from 0 to about 3 or 4 (beyond which there is usually little change), for some selected values of temporal perspective  $T$ . For this type of profile,  ${}^qPD(T)$  and  ${}^q\bar{D}(T)$  have similar patterns as  $T$  is fixed, so it is sufficient to plot the profile only for one measure. (2) The second type of diversity profile is obtained by plotting  ${}^qPD(T)$  and  ${}^q\bar{D}(T)$  as functions of  $T$  separately for  $q=0$ , 1, and 2. This profile shows the effect of time-depth or evolution change on our diversity measures.

For the second type of profile,  ${}^qPD(T)$  and  ${}^q\bar{D}(T)$  generally exhibit different patterns (the profile of  ${}^q\bar{D}(T)$  is decreasing with  $T$  whereas the profile of  ${}^qPD(T)$  for  $q=0$  (Faith's PD) is always increasing, and for  $q>0$  is generally increasing up to a certain point, so the profiles for both measures are informative. The parameter  $q$  gives the sensitivity of the two measures to present-day species relative abundances. As in the ordinary Hill numbers, the measures with  $q=2$  favor more abundant

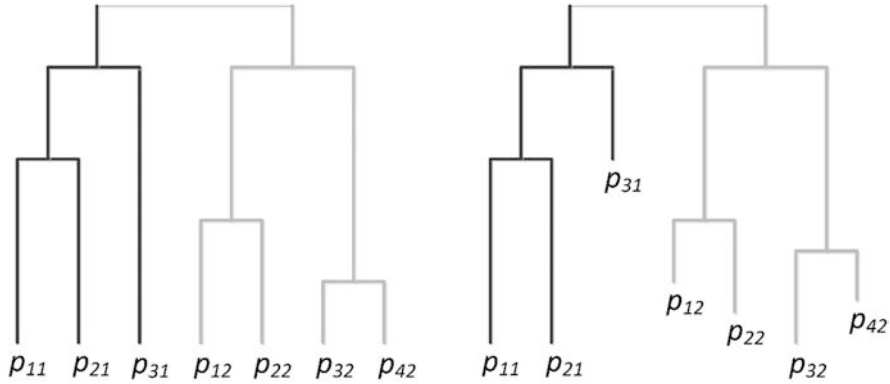
species, so they are useful in ecological studies to examine the phylogenetic relationships of the dominant species in a set of assemblages, or those examining functional diversity. The measures of  $q=0$  emphasizes rare species, so they are useful when abundance information is not necessarily relevant (e.g., when ecologists try to identify past episodes of differentiation, or for some conservation biology applications). The measures with  $q=1$  weigh species according to their frequencies and can be used in most applications when neither dominant nor rare species should be favored.

When the measure of evolutionary change is typically based on the number of nucleotide base changes at a selected locus, or the amount of functional or morphological differentiation from a common ancestor, the branches of the resulting tree will then be uneven, so the tree is non-ultrametric. In this case, Chao et al. (2010) showed that the time parameter  $T$  in all formulas should be replaced by the *mean base change* or *mean branch length*  $\bar{T}$ , the mean of the distances from the tree base to each of the terminal branch tips (i.e., the mean evolutionary change per species over the interval of interest). See the right panel of Fig. 1 for an illustrative example. Let  $B_{\bar{T}}$  denote the set of branches connecting all focal species, with mean branch length  $\bar{T}$ . Then we can express  $\bar{T}$  as  $\bar{T} = \sum_{i \in B_{\bar{T}}} L_i a_i$ . The diversity of a non-ultrametric tree with mean evolutionary change  $\bar{T}$  is the same as that of an ultrametric tree with time parameter  $\bar{T}$ . Therefore, the diversity formulas for a non-ultrametric tree are obtained by replacing  $T$  by  $\bar{T}$  in Eqs. (4a), (4b), (5a), and (5b). The resulting measures are denoted respectively as  ${}^q\bar{D}(\bar{T})$ ,  ${}^1\bar{D}(\bar{T})$ ,  ${}^qPD(\bar{T})$  and  ${}^1PD(\bar{T})$ ; see Chao et al. (2010) for details. When we compare the phylogenetic diversity based on the measures  ${}^q\bar{D}(\bar{T})$  and  ${}^qPD(\bar{T})$  for several non-ultrametric trees, all measures should refer to the same mean base change  $\bar{T}$  to make meaningful comparisons.

## Replication Principle for Phylogenetic Diversity Measures

The replication principle was generalized to a phylogenetic version in Chao et al. (2010). Suppose there are  $N$  equally large and completely phylogenetically distinct assemblages (no shared lineages across assemblages, though lineages within an assemblage may be shared); see Fig. 2 (reproduced from Chiu et al. 2014) for an illustrative example. Suppose these assemblages have the same phylogenetic Hill number  $X$ . If these assemblages are pooled, then the pooled assemblages must have a phylogenetic Hill number  $N \times X$ . In the proof of this replication principle, Chao et al. (2010) assumed that these  $N$  assemblages have the same mean branch lengths. Here we relax this assumption and allow assemblages to have different mean branch lengths. (In the special case of ultrametric trees, this means that we allow different time perspectives for different assemblages.)

Suppose in assemblage  $k$ , the mean branch length is  $\bar{T}_k$ , and the branch set is  $B_{\bar{T}_k, k}$  (we omit  $\bar{T}_k$  in the subscript and just use  $B_k$  in the following proof for notational simplicity) with branch lengths  $\{L_{ik}; i \in B_k\}$  and the corresponding nodes



**Fig. 2** Replication Principle for two completely phylogenetically distinct assemblages with totally different structures. *Left panel:* Assemblage 1 (black) includes three species with species relative abundances  $\{p_{11}, p_{21}, p_{31}\}$  for the three tips. Assemblage 2 (grey) includes four species with species relative abundances  $\{p_{12}, p_{22}, p_{32}, p_{42}\}$  for the four tips. The diversity of the pooled tree is double of that of each tree as long as the two assemblages are completely phylogenetically distinct as shown (no lineages shared between assemblages, though lineages within an assemblage may be shared) and have identical mean diversities (i.e., phylogenetic Hill number). *Right panel:* The same is valid for two completely phylogenetically distinct non-ultrametric assemblages (This figure is reproduced from Fig. 1 of Chiu et al. 2014)

abundances  $\{a_{ik}; i \in \mathbf{B}_k\}$ ,  $k = 1, 2, \dots, N$ . Assume that all assemblages have the same phylogenetic Hill numbers  ${}^q\bar{D}(\bar{T}_k) = X$ , implying  $\sum_{i \in \mathbf{B}_k} L_{ik} a_{ik}^q = X^{1-q} \bar{T}_k$  for all  $k = 1, 2, \dots, N$ . When the  $N$  trees are pooled with equal weight for each tree, each node abundance  $a_{ik}$  in the pooled tree becomes  $a_{ik}/N$ , and the mean branch length becomes  $\bar{T} = (1/N) \sum_{k=1}^N \bar{T}_k$ . Then the phylogenetic Hill number of order  $q$  for the pooled assemblage becomes

$$\begin{aligned} {}^q\bar{D}(\bar{T}) &= \left\{ \sum_{k=1}^N \sum_{i \in \mathbf{B}_k} \frac{L_{ik}}{\bar{T}} \left( \frac{a_{ik}}{N} \right)^q \right\}^{1/(1-q)} = \left\{ \frac{1}{N^q} \sum_{k=1}^N \frac{1}{\bar{T}} \sum_{i \in \mathbf{B}_k} L_{ik} a_{ik}^q \right\}^{1/(1-q)} \\ &= \left\{ \frac{1}{N^q} \sum_{k=1}^N \frac{\bar{T}_k}{\bar{T}} X^{1-q} \right\}^{1/(1-q)} = \left\{ N^{1-q} \times X^{1-q} \right\}^{1/(1-q)} = N \times X. \end{aligned} \quad (6)$$

This proves a stronger version of the replication principle for phylogenetic Hill numbers. Note the mean branch length in the pooled assemblage is the average of individual mean branch lengths. For example, if  ${}^q\bar{D}(\bar{T}_1 = 2) = {}^q\bar{D}(\bar{T}_2 = 6) = 10$ , then in an effective sense, there are ten lineages with mean branch length 2 in Assemblage 1 and there are ten lineages with mean branch length 6 in Assemblage 2. The replication principle implies that there are 20 lineages in the pooled tree with mean branch length 4. Since  ${}^qPD(\bar{T}_k) = {}^q\bar{D}(\bar{T}_k) \times \bar{T}_k$ , the replication principle for the phylogenetic diversity  ${}^qPD(\bar{T})$  does need the assumption that all assemblages

have the same mean branch lengths ( $\bar{T}_1 = \bar{T}_2 = \dots = \bar{T}_N$ ). The proof is parallel and thus omitted.

## Decomposition of Phylogenetic Diversity Measures

Decomposition of species richness and its phylogenetic analogues into within- and between-group (alpha and beta) components is widely used (Whittaker 1972; Faith et al. 2009). However, these take no notice of abundance differences between sites. Conservationists using these measures cannot distinguish a site whose species are equally abundant from a site with the same species but with a highly skewed abundance distribution whose most phylogenetically distinctive species are rare. The former site would be a better bet for conservation. These considerations, and others, motivate the development of decomposition theory for abundance-based phylogenetic diversity measures. The decomposition also leads to abundance-sensitive measures of phylogenetic similarity and complementarity.

When there are  $N$  assemblages, the phylogenetic Hill numbers  ${}^q\bar{D}(T)$  (Eqs. 4a and 4b) and phylogenetic diversity  ${}^qPD(T)$  (Eqs. 5a and 5b) of the pooled assemblage can be multiplicatively decomposed into independent alpha and beta components (Chiu et al. 2014). We briefly describe the decomposition of the measure  ${}^q\bar{D}(T)$  here for the ultrametric case, and only summarize the decomposition of the measure  ${}^qPD(T)$ . The extension to the non-ultrametric case for both measures is obtained by simply replacing all  $T$  in the formulas with the mean branch length  $\bar{T}$  of the pooled assemblage.

To begin the partitioning, a pooled tree is constructed for the  $N$  assemblages. Assume that there are  $S$  species in the present-day assemblage (i.e., there are  $S$  tip nodes). For any tip node  $i$ , let  $z_{ik}$  denote any measure of species importance of the  $i$ th species in the  $k$ th assemblage,  $i = 1, 2, \dots, S$ ,  $k = 1, 2, \dots, N$ . The measure  $z_{ik}$  is referred to as “abundance” for simplicity, although it can be absolute abundances, relative abundances, incidence, biomasses, cover areas or any other importance measure. Define  $z_{+k} = \sum_{i=1}^S z_{ik}$  (i.e., the “+” sign in  $z_{+k}$  denotes a sum over the tip nodes only) as the current size of the  $k$ th assemblage. Let  $z_{++} = \sum_{k=1}^N z_{+k}$  be the total abundance in the present-day pooled assemblage.

Now consider the phylogenetic tree in the time interval  $[-T, 0]$ , and in the pooled assemblage define  $\mathbf{B}_T$  and  $L_i$  as in section “Phylogenetic Hill numbers and related measures”. We extend the definition of  $z_{ik}$  to include all nodes and their corresponding branches by defining  $z_{ik}$  for all  $i \in \mathbf{B}_T$  as the total abundances descended from branch  $i$ . (Here the index  $i$  can correspond to both tip-node and internal node; if  $i$  is a tip-node, then  $z_{ik}$  represents data of the current assemblage as defined in the preceding paragraph.) As shown in Fig. 2 of Chiu et al. (2014), the diversity for each individual assemblage can be computed from the pooled tree structure, and only the node abundances vary with assemblages.

In the pooled assemblage, the node abundance for branch  $i$  ( $i \in \mathbf{B}_T$ ) is  $z_{i+} = \sum_{k=1}^N z_{ik}$  with branch relative abundance  $z_{i+}/z_{++}$ , so the phylogenetic gamma diversity of order  $q$  can be calculated from Eq. (4a) as

$${}^q\bar{D}_\gamma(T) = \frac{1}{T} \left\{ \sum_{i \in \mathbf{B}_T} L_i \left( \frac{z_{i+} / z_{++}}{T} \right)^q \right\}^{1/(1-q)}, q \geq 0, q \neq 1. \quad (7a)$$

The limit when  $q$  approaches unity exists and is equal to

$${}^1\bar{D}_\gamma(T) = \lim_{q \rightarrow 1} {}^q\bar{D}_\gamma(T) = \exp \left[ - \sum_{i \in \mathbf{B}_T} \frac{L_i}{T} \left( \frac{z_{i+}}{z_{++}} \right) \log \left( \frac{z_{i+}}{z_{++}} \right) \right]. \quad (7b)$$

The gamma diversity is the effective number of equally abundant and equally distinct lineages all with branch lengths  $T$  in the pooled assemblage.

Chiu et al. (2014) derived the following phylogenetic alpha diversity for  $q \geq 0$  and  $q \neq 1$ :

$${}^q\bar{D}_\alpha(T) = \frac{1}{TN} \left\{ \sum_{i \in \mathbf{B}_T} L_i \sum_{k=1}^N \left( \frac{z_{ik} / z_{++}}{T} \right)^q \right\}^{1/(1-q)} \quad (8a)$$

For  $q=1$ , we have

$${}^1\bar{D}_\alpha(T) = \lim_{q \rightarrow 1} {}^q\bar{D}_\alpha(T) = \exp \left[ - \sum_{i \in \mathbf{B}_T} L_i \sum_{k=1}^N \frac{z_{ik} / z_{++}}{T} \log \frac{z_{ik} / z_{++}}{T} - \log(NT) \right] \quad (8b)$$

The alpha diversity is interpreted as the effective number of equally abundant and equally distinct lineages all with branch lengths  $T$  in an individual assemblage. When normalized measures of species importance (like relative abundance or relative biomass) are used to quantify species importance, we have  $z_{++} = N$  in Eqs. (8a) and (8b). The alpha formula then reduces to a generalized mean of the local diversities with the following property: if all assemblages have the same diversity  $X$ , the alpha diversity is also  $X$  (Jost 2007). For non-normalized measures of species importance, like absolute abundance or biomass, this property does not hold. This is because when species absolute abundances are compared, for example, a three-species assemblage with absolute abundances  $\{2, 5, 8\}$  will not be treated as identical as another three-species assemblage with absolute abundances  $\{200, 500, 800\}$ . However, these two assemblages are treated as identical when only relative abundances are compared.

Chiu et al. (2014) proved that the phylogenetic gamma Hill number (Eqs. 7a and 7b) is always greater than or equal to the phylogenetic alpha Hill number (Eqs. 8a

and 8b) for all  $q \geq 0$  regardless of species abundances and tree structures. Based on a multiplicative partitioning, the phylogenetic beta diversity is the ratio of gamma diversity to alpha diversity:

$${}^q\bar{D}_\beta(T) = \frac{{}^q\bar{D}_\gamma(T)}{{}^q\bar{D}_\alpha(T)}, \quad q \geq 0. \quad (9)$$

When the  $N$  assemblages are identical in species identities and species abundances, then  ${}^q\bar{D}_\beta(T) = 1$  for any  $T$ . When the  $N$  assemblages are completely phylogenetically distinct (no shared lineages), then  ${}^q\bar{D}_\beta(T) = N$ , no matter what the diversities or tree shapes of the assemblages. The measure  ${}^q\bar{D}_\beta(T)$  thus quantifies the effective number of completely phylogenetically distinct assemblages in the interval  $[-T, 0]$ . As proved by Chiu et al. (2014), the phylogenetic beta diversity  ${}^q\bar{D}_\beta(T)$  is always between unity and  $N$  for any given alpha value, implying alpha and beta components are unrelated (or independent) for both measures,  ${}^q\bar{D}(T)$  and  ${}^qPD(T)$ ; see Chao et al. (2012) for a rigorous discussion of un-relatedness and independence of two measures. When all lineages in the pooled assemblage are completely distinct (no lineages shared) in the interval  $[-T, 0]$ , the phylogenetic alpha, beta and gamma Hill numbers reduce to those based on ordinary Hill numbers. This includes the limiting case in which  $T$  tends to zero, so that phylogeny is ignored.

Parallel decomposition can be made for the phylogenetic diversity  ${}^qPD(T)$ , and we summarize the following relations:  ${}^qPD_\gamma(T) = {}^q\bar{D}_\gamma(T) \times T$  and  ${}^qPD_\alpha(T) = {}^q\bar{D}_\alpha(T) \times T$ . Under a multiplicative partitioning scheme, we have  ${}^qPD_\beta(T) = {}^qPD_\gamma(T) / {}^qPD_\alpha(T) = {}^q\bar{D}_\beta(T)$ , i.e., the beta components from partitioning the phylogenetic Hill numbers  ${}^q\bar{D}(T)$  and phylogenetic diversity  ${}^qPD(T)$  are identical, implying the interpretation and the corresponding similarity or differentiation measures (in the next section) are also identical. Thus, it is sufficient to focus only on the measure  ${}^q\bar{D}_\beta(T)$ , which will be referred to as the *phylogenetic beta diversity* or beta component for simplicity.

For each of the two measures,  ${}^q\bar{D}(T)$  and  ${}^qPD(T)$ , alpha and gamma diversities obey the replication principle. Then the beta diversity formed by taking their ratio is replication-invariant (Chiu et al. 2014). That is, when assemblages are replicated, the beta diversity does not change. Therefore, when we pool equally-distinct subtrees, such as pooling equally-ancient subfamilies, the beta diversity is unchanged by pooling the subfamilies if all subfamilies show the same beta diversity (“consistency in aggregation”).

We now give the phylogenetic beta diversities for the special cases of  $q=0, 1$  and 2.

- (a) When  $q=0$ , we have  ${}^0\bar{D}_\beta(T) = L_\gamma(T) / L_\alpha(T)$ , where  $L_\gamma(T)$  denotes the total branch length of the pooled tree (the gamma component of Faith’s PD) and  $L_\alpha(T)$  denotes the average length of individual trees (the alpha component of Faith’s PD).
- (b) When  $q=1$ , the phylogenetic beta diversity of order 1 is



$${}^1\bar{D}_\beta(T) = \exp \left[ \left( H_{P,\gamma} - H_{P,\alpha} \right) / T + \sum_{k=1}^N \left( \frac{z_{+k}}{z_{++}} \right) \log \left( \frac{z_{+k}}{z_{++}} \right) + \log N \right], \quad (10a)$$

where  $H_{P,\gamma}$  and  $H_{P,\alpha}$  denote respectively the gamma and alpha phylogenetic entropy. When the species importance measure  $z_{ik}$  represents the  $i$ th species relative abundance in the  $k$ th current-time assemblage, then  $z_{+k} = 1$ ,  $z_{++} = N$ ,  $z_{+k} / z_{++} = 1 / N$ . In this special case, we have  ${}^1\bar{D}_\beta(T) = \exp \left[ \left( H_{P,\gamma} - H_{P,\alpha} \right) / T \right]$ . Thus an additive decomposition for phylogenetic entropy  $H_P$  holds (Pavoine et al. 2009; Mouchet and Moullot 2011), as for ordinary Shannon entropy (Jost 2007).

(c) When  $q=2$ , the phylogenetic beta diversity can be expressed as

$${}^2\bar{D}_\beta(T) = \frac{\sum_{i \in B_T} L_i \sum_{k=1}^N z_{ik}^2}{\sum_{i \in B_T} L_i z_{i+}^2}.$$

In the special case of  $z_{+k} = 1$ ,  $z_{++} = N$ , this phylogenetic beta diversity of order 2 can be linked to quadratic entropy as

$${}^2\bar{D}_\beta(T) = \left( 1 - Q_\gamma / T \right)^{-1} / \left( 1 - Q_\alpha / T \right)^{-1}, \quad (10b)$$

where  $Q_\gamma$  and  $Q_\alpha$  denote respectively the gamma and alpha quadratic entropy. The above formula is also applicable to non-ultrametric trees by replacing all  $T$  with  $\bar{T}$ , the mean branch length in the pooled assemblage; see Chiu et al. (2014, Appendix C) for a proof.

## Normalized Phylogenetic Similarity Measures

For traditional abundance-based diversity, the most commonly used similarity measures include  $N$ -assemblage generalizations of the Jaccard et al. (1966) and Morisita-Horn (Morisita 1959) measures. The latter three measures were integrated into a class of  $C_{qN}$  measures by Chao et al. (2008). Jost (2006, 2007), Chao et al. (2008, 2012), and Chiu et al. (2014) have demonstrated that all the above measures are monotonic transformations of beta diversity based on the ordinary Hill numbers. This is an advantage of using the framework of Hill numbers: a direct link exists between diversity and similarity (or differentiation) among assemblages.

Chiu et al. (2014) extended this framework by proposing four classes of similarity (or differentiation) measures that are monotonic functions of phylogenetic beta diversity. The basic idea is that the phylogenetic beta diversity, a ratio of gamma and alpha phylogenetic Hill numbers, is independent of alpha and measures the pure differentiation among assemblages. The phylogenetic beta component always lies

in the range  $[1, N]$  for any measures of species importance and all orders  $q \geq 0$ . Since the range depends on  $N$ , the phylogenetic beta diversity cannot be used to compare phylogenetic differentiation among assemblages across multiple regions with different numbers of assemblages. To remove the dependence on  $N$ , several transformations can be used to transform the phylogenetic beta component onto  $[0, 1]$  to measure local overlap, regional overlap, homogeneity and turnover. We give a summary of these four transformations below and tabulate formulas and the relationship with previous measures in Table 1 for the two most important classes. The formulas for the special cases for  $q=0, 1$  and  $2$  are also displayed there.

1. *A class of branch overlap measures from a local perspective:*

$$\bar{C}_{qN}(T) = \frac{N^{1-q} - \left[ {}^q\bar{D}_\beta(T) \right]^{1-q}}{N^{1-q} - 1}. \quad (11a)$$

This gives the effective average proportion of shared branches in an individual assemblage. This class of similarity measures extends the  $C_{qN}$  overlap measure derived in Chao et al. (2008) to a phylogenetic version. The corresponding differentiation measure  $1 - \bar{C}_{qN}(T)$  quantifies the effective average proportion of non-shared branches in an individual assemblage.

- (1a) For  $q=0$ , this similarity measure is referred to as the “phylo-Sørensen”  $N$ -assemblage overlap measure because for  $N=2$ , it reduces to the measure *PhyloSør* (phylo-Sørensen) developed by Bryant et al. (2008) and Ferrier et al. (2007).
- (1b) For  $q=1$ , this measure  $\bar{C}_{1N}(T)$  is called the “phylo-Horn”  $N$ -assemblage overlap measure because it extends Horn (1966) two-assemblage measure to incorporate phylogenies for  $N$  assemblages.
- (1c) For  $q=2$ ,  $\bar{C}_{2N}(T)$  is called the “phylo-Morisita-Horn”  $N$ -assemblage similarity measure because it extends Morisita-Horn measure (Morisita 1959) to incorporate phylogenies for  $N$  assemblages. The differentiation measure  $1 - \bar{C}_{2N}(T)$  when the species importance measure is relative abundances reduces to the measure proposed by de Bello et al. (2010). However, their measure is valid only for ultrametric trees (p. 7 of de Bello et al. 2010). Here, the measure can be applied to non-ultrametric trees to obtain

$$1 - \bar{C}_{2N}(\bar{T}) = \frac{1 - \left[ 1 / {}^2\bar{D}_\beta(\bar{T}) \right]}{1 - 1/N} = \frac{Q_\gamma - Q_\alpha}{(1 - 1/N)(\bar{T} - Q_\alpha)}, \quad (11b)$$

where  $Q_\gamma$  and  $Q_\alpha$  are respectively gamma and alpha quadratic entropy, and  $\bar{T}$  is the mean branch length in the pooled assemblage. A general form for any species importance measure (including absolute abundances) is

**Table 1** Two major classes of phylogenetic similarity measures based on the transformations of phylogenetic beta diversity when species importance measures are incidences (for  $q = 0$ ), relative abundances or absolute abundances (for  $q = 1$  and 2). The corresponding differentiation measures are the one-complements of the similarity measures. When all lineages are completely distinct (this includes  $T \rightarrow 0$ , ignoring phylogeny), these phylogenetic measures reduce to the corresponding non-phylogenetic versions. All measures can also be applied to non-ultrametric trees if  $\bar{T}$  is substituted for  $T$

Order	Species importance measure	Phylo-local-overlap	Phylo-regional-overlap
$q = 0$		$\bar{C}_{q^N}(T) = \frac{\left[1 / {}^q \bar{D}_\beta(T)\right]^{q-1} - (1/N)^{q-1}}{1 - (1/N)^{q-1}}$	$\bar{U}_{q^N}(T) = \frac{\left[1 / {}^q \bar{D}_\beta(T)\right]^{1-q} - (1/N)^{1-q}}{1 - (1/N)^{1-q}}$
	Incidences	Phylo-Sørensen (= <i>PhyloSfor</i> for $N = 2$ ) $\frac{N - L_j(T) / L_\alpha(T)}{N - 1}$	Phylo-Jaccard (= <i>1-UniFrac</i> for $N = 2$ ) $\frac{L_\alpha(T) / L_j(T) - 1 / N}{1 - 1 / N}$
$q = 1$	Relative abundances	Phylo-Horn $1 - \frac{H_{P,j} - H_{P,\alpha}}{T \log N}$	Phylo-Horn $1 - \frac{H_{P,j} - H_{P,\alpha}}{T \log N}$
	Absolute abundances	$\frac{H_{P,\alpha} - H_{P,j} - T \sum_{k=1}^N \frac{Z_{+k}}{Z_{++}} \log \left( \frac{Z_{+k}}{Z_{++}} \right)}{T \log N}$	$\frac{H_{P,\alpha} - H_{P,j} - T \sum_{k=1}^N \frac{Z_{+k}}{Z_{++}} \log \left( \frac{Z_{+k}}{Z_{++}} \right)}{T \log N}$

$q=2$	Relative abundances	Phylo-Morisita-Horn	Phylo-regional-overlap
	Absolute abundances	$1-\frac{Q_\gamma-Q_\alpha}{(1-1/N)(T-Q_\alpha)}$ $1-\frac{\sum_{i \in B_\gamma} L_i \sum_{m \neq k}^N (z_{im}-z_{ik})^2}{(N-1) \sum_{i \in B_\gamma} L_i \sum_{k=1}^N z_{ik}^2}$	$1-\frac{Q_\gamma-Q_\alpha}{(N-1)(T-Q_\gamma)}$ $1-\frac{\sum_{i \in B_\gamma} L_i \sum_{m \neq k}^N (z_{im}-z_{ik})^2}{(N-1) \sum_{i \in B_\gamma} L_i z_{i+}^2}$

1.  $z_{ik}$ : species abundance (or any species importance measure) descended from branch  $i$  in the  $k$ th assemblage,  $z_{i+} = \sum_{k=1}^N z_{ik} \cdot z_{i+k} = \sum_{i=1}^S z_{ik}$  (assemblage size of the  $k$ th assemblage in the present-day assemblage with species richness  $S$ ), and  $z_{++} = \sum_{i=1}^S \sum_{k=1}^N z_{ik}$  (the total abundance in the present-day pooled assemblage),  $B_\gamma$ : the set of all branches in the time interval  $[-T, 0]$ ,  $L_i$ : length of branch  $i$

2.  $q=0$ ,  $L_\gamma(T)$  and  $L_\alpha(T)$ : gamma and alpha total branch length (i.e., Faith's PD)

3.  $q=1$ ,  $H_{P\gamma}$  and  $H_{P\alpha}$ : gamma and alpha phylogenetic entropy

4.  $q=2$ ,  $Q_\gamma$  and  $Q_\alpha$ : gamma and alpha quadratic entropy

$$1 - \bar{C}_{2N}(\bar{T}) = \frac{\sum_{i \in \mathbf{B}_{\bar{T}}} L_i \sum_{m > k}^N (z_{im} - z_{ik})^2}{(N-1) \sum_{i \in \mathbf{B}_{\bar{T}}} L_i \sum_{k=1}^N z_{ik}^2}. \quad (11c)$$

The above expression shows that the similarity index  $\bar{C}_{2N}(\bar{T})$ , as in all other abundance-sensitive similarity measures, is unity if and only if  $z_{ij} = z_{ik}$  (i.e., species importance measures are identical for any node  $i$  in the branch set and for any two assemblages  $j$  and  $k$ ). This reveals that the similarity index  $\bar{C}_{2N}(\bar{T})$  quantifies the node-by-node resemblance among the  $N$  abundance sets  $\{z_{ik}; i \in \mathbf{B}_{\bar{T}}, k = 1, 2, \dots, N\}$  from a local perspective. See Fig. 2 of Chiu et al. (2014) for a simple example of the framework.

2. A class of branch overlap measures from a regional perspective:

$$\bar{U}_{qN}(T) = \frac{[1 / {}^q \bar{D}_{\beta}(T)]^{1-q} - (1/N)^{1-q}}{1 - (1/N)^{1-q}} \quad (12a)$$

This class of measures quantifies the effective proportion of shared branches in the pooled assemblage. The corresponding differentiation measure  $1 - \bar{U}_{qN}(T)$  quantifies the effective average proportion of non-shared branches in the pooled assemblage.

- (2a) For  $q=0$ , this measure is called the “phylo-Jaccard”  $N$ -assemblage measure because for  $N=2$  the measure  $1 - \bar{U}_{02}(T)$  reduces to the Jaccard-type *UniFrac* measure developed by Lozupone and Knight (2005) and the PD-dissimilarity measure developed by Faith et al. (2009).
- (2b) For  $q=1$ , this measure is identical to the “phylo-Horn”  $N$ -assemblage overlap measure  $\bar{C}_{1N}(T)$ ; see Table 1.
- (2c) For  $q=2$ , we refer to the measure  $\bar{U}_{2N}(T)$  as a “phylo-regional-overlap” measure. When the species importance measure is relative abundance, we have the following formula for non-ultrametric trees:

$$1 - \bar{U}_{2N}(\bar{T}) = \frac{N - {}^2 \bar{D}_{\beta}(\bar{T})}{N-1} = \frac{Q_{\gamma} - Q_{\alpha}}{(N-1)(\bar{T} - Q_{\gamma})},$$

where  $\bar{T}$  denotes the mean branch length in the pooled assemblage. A general form for any species importance measure (including absolute abundances) is

$$1 - \bar{U}_{2N}(\bar{T}) = \frac{\sum_{i \in \mathbf{B}_{\bar{T}}} L_i \sum_{m > k}^N (z_{im} - z_{ik})^2}{(N-1) \sum_{i \in \mathbf{B}_{\bar{T}}} L_i z_{i+}^2}.$$

The numerator is the same as that in  $\bar{C}_{2N}(\bar{T})$ , revealing that the similarity index  $\bar{U}_{2N}(\bar{T})$  also quantifies the node-by-node resemblance among the  $N$  abundance sets  $\{z_{ik}; i \in \mathcal{B}_T\}$ ,  $k = 1, 2, \dots, N$ ; but here the denominator (for the purpose of normalization) is different and takes a regional perspective.

### 3. A class of phylogenetic homogeneity measures

$$\bar{S}_{qN}(T) = \frac{1 / {}^q\bar{D}_\beta(T) - 1 / N}{1 - 1 / N}. \quad (12b)$$

This measure is linear in the proportion of regional phylogenetic diversity contained in a typical assemblage.

- (3a) For  $q=0$ , it reduces to the “phylo-Jaccard” measure  $\bar{U}_{0N}(T)$ , i.e.,  $\bar{S}_{0N}(T) = \bar{U}_{0N}(T)$ .
- (3b) For  $q=1$ , this measure does not reduce to the “phylo-Horn” overlap measure.
- (3c) For  $q=2$ , this measure is identical to  $\bar{C}_{2N}(T)$ , the “phylo-Morisita-Horn” similarity measure, i.e.,  $\bar{S}_{2N}(T) = \bar{C}_{2N}(T)$ .

### 4. A class of measures of the complement of “phylogenetic turnover rate”:

$$\bar{V}_{qN}(T) = \frac{N - {}^q\bar{D}_\beta(T)}{N - 1} = 1 - \frac{{}^q\bar{D}_\beta(T) - 1}{N - 1}. \quad (12c)$$

This measure is linear in the phylogenetic beta diversity and the corresponding differentiation measure  $[{}^q\bar{D}_\beta(T) - 1] / (N - 1)$  quantifies the relative branch turnover rate per assemblage.

- (4a) For  $q=0$ , the measure  $\bar{V}_{0N}(T)$  is identical to the “phylo-Sørensen” measure, i.e.,  $\bar{V}_{0N}(T) = \bar{C}_{0N}(T)$ .
- (4b) For  $q=1$ , this measure does not reduce to the “phylo-Horn” overlap measure.
- (4c) For  $q=2$ , this measure is identical to  $\bar{U}_{2N}(T)$ , the “phylo-regional-overlap” measure. That is,  $\bar{V}_{2N}(T) = \bar{U}_{2N}(T)$ .

As with the phylogenetic diversity measures, all the above similarity or differentiation measures are functions of two parameters: the sensitivity parameter  $q$  and the time perspective  $T$ . Thus, for each measure, we suggest using the two types of profiles described in section “[Phylogenetic Hill numbers and related measures](#)” for the two major similarity measures  $\bar{C}_{qN}(T)$  and  $\bar{U}_{qN}(T)$  (or their complements) to convey complete information about the similarity or differentiation of a set of assemblages. An example showing the two types of profiles is given in section “[An example](#)”.

The lineage excess  ${}^q\bar{D}_\gamma(T) - {}^q\bar{D}_\alpha(T)$  and the phylogenetic diversity excess  ${}^qPD_\gamma(T) - {}^qPD_\alpha(T)$  can be interpreted as the effective number of regional lineages

(or regional phylogenetic diversity) not contained in a typical local assemblage. However, they cannot be directly applied to compare the similarity or differentiation across multiple regions because both depend not only on the number of assemblages, but also on their corresponding alpha diversity. Following Chao et al. (2012) and Chiu et al. (2014, their Appendix D) proved that we can eliminate these dependencies by using an appropriate normalization. After proper normalizations, the two measures lead to the same four classes of normalized similarity and differentiation measures as those obtained from the phylogenetic beta diversity. This is another advantage of using the framework of phylogenetic Hill numbers. That is, a consensus can be achieved on phylogenetic similarity and differentiation measures, including  $N$ -assemblage phylogenetic generalizations of the classic Jaccard, Sørensen, Horn and Morisita-Horn measures, regardless of whether one prefers multiplicative or additive decompositions.

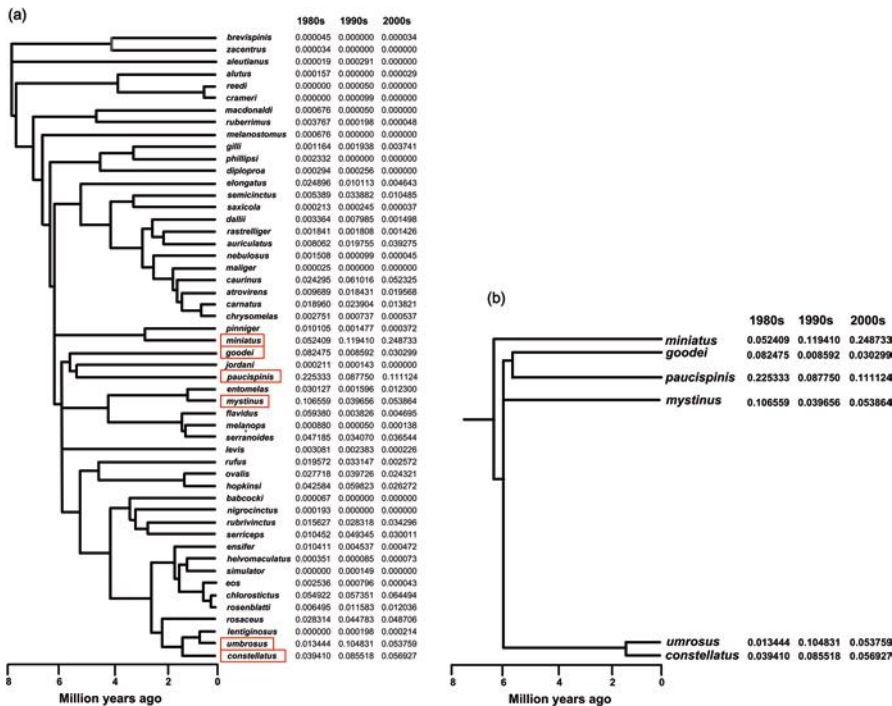
## An Example

We apply the phylogenetic diversity measures and similarity (or differentiation) measures considered in this chapter to a real conservation biology case discussed by Pavoine et al. (2009), a heavily-fished assemblage of 52 rockfish species of the genus *Sebastes* collected for 20 years over three decades (1980–1986, 1993–1994, 1996, 1998–2007) from the Southern California Bight, USA. The phylogenetic tree for these 52 species was obtained from Hyde and Vetter (2007); see Fig. 3a. The age of the root for these species is around 7.9 million years (Myr).

We separate the data into three decades: 1980s, 1990s and 2000s, which will be referred to as Assemblages (and Decades) I, II and III respectively. Within each decade's assemblage, species abundances are pooled. The species relative abundances for the three assemblages are shown in Fig. 3a. There were 48, 44 and 39 species in Decades I, II and III, respectively. (Note that each data point here is a mean of many years' observations.) A sub-tree containing only the six dominant species (those with relative abundance  $>8\%$  in at least one assemblage) is shown in Fig. 3b. All six species are shared in the three assemblages and four of them have been in isolated lineages for 6 Myr.

As suggested in section “Phylogenetic Hill numbers and related measures”, we present for each assemblage two types of profiles. In Fig. 4a, we plot the measure  ${}^q\bar{D}(T)$  as a function of order  $q$ ,  $0 \leq q \leq 3$ , for two selected values of temporal perspectives:  $T=0$  (phylogeny is ignored) and  $T=7.9$  Myr (whole phylogenetic tree in Fig. 3a is considered). In Fig. 4b, we plot  ${}^q\bar{D}(T)$  and  ${}^qPD(T)$  as functions of  $T$  separately for  $q=0, 1$ , and  $2$  for  $0 \leq T \leq 10$ .

Based on our phylogenetic diversity measures, all profiles in Fig. 4 reveal that the diversity in the most recent decade (Decade III) is the lowest among the three decades in the rockfish assemblage. This implies an appreciable loss of species (as shown in the first type of profile for  $T=0$ ), loss of lineages (as shown in the second type of profile based on the measure  ${}^q\bar{D}(T)$ ), and loss of evolutionary history (as

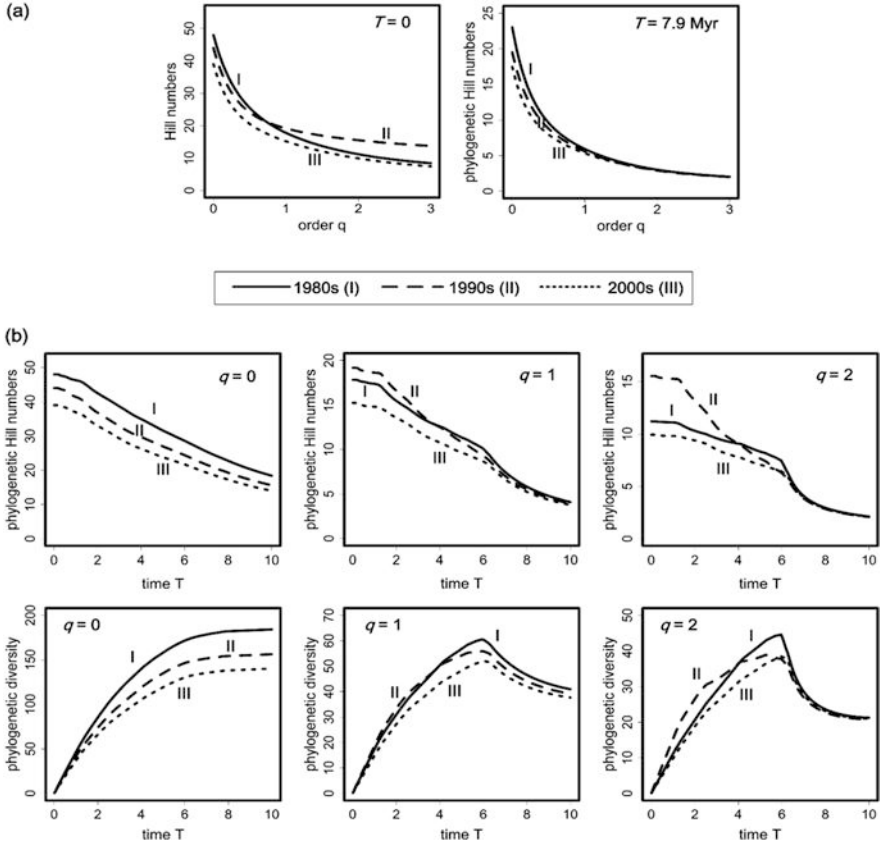


**Fig. 3** (a) The phylogenetic tree of 52 rockfish species of the genus *Sebastes* (Hyde and Vetter 2007) and the species relative abundances in three assemblages: 1980s (Decade I), 1990s (Decade II) and 2000s (Decade III). The age of the root is  $T=7.9$  Myr. (b) A sub-tree contains only the dominant species (those with relative abundance  $>8\%$  in at least one assemblage), and these species are marked in figure (a). All six species are shared by the three assemblages and four of them diverged around 6 Myr ago (i.e., they have been in isolated lineages for 6 Myr) (See Pavoine et al. (2009) for details)

shown in the second type of profiles based on the measure  ${}^qPD(T)$  over the three decades.

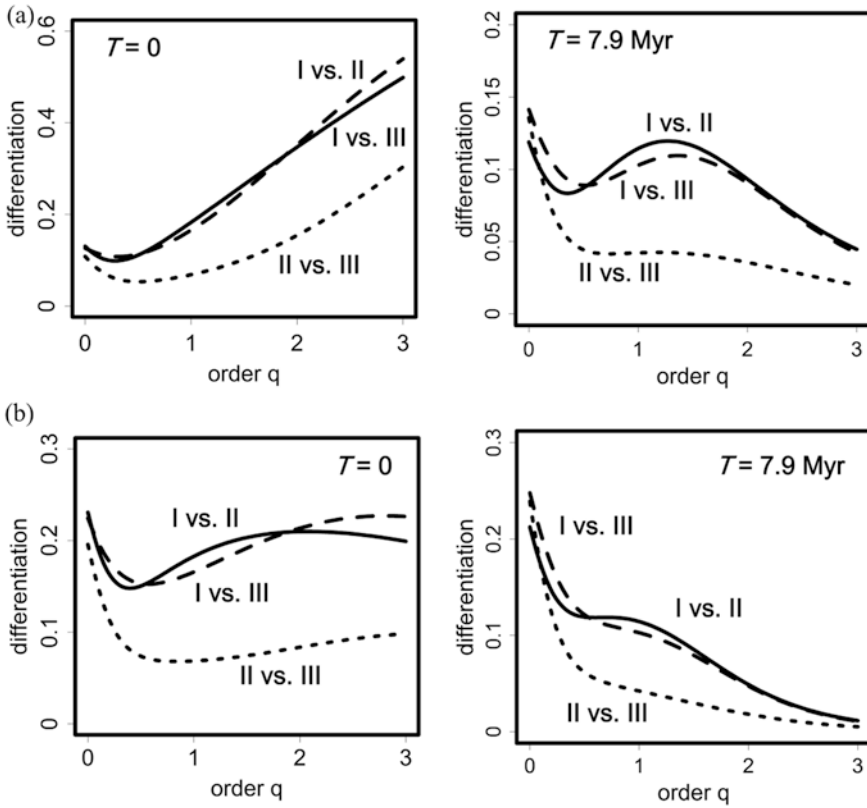
When species/lineage abundances are discounted ( $q=0$  in the left panels of Fig. 4b), both lineage richness (based on the measure  ${}^0\bar{D}(T)$ ) and total branch lengths (based on the measure  ${}^0PD(T)$ , i.e., Faith's PD) exhibit the expected ordering: Decade I > Decade II > Decade III. When species/lineage abundances are counted (i.e.  $q=1$  and 2 in Fig. 4b), the profiles for Decades I and II cross because the assemblage of Decade II has more even abundant species than that of Decade I (see the first type of profiles for  $T=0$  and Fig. 3a, b). Note that if the time-depth is greater than 6 Myr (including the age of the root), then all the abundance-sensitive phylogenetic measures for the three assemblages are very close because most of the dominant species began to diverge around 6 Myr (Fig. 3b). This also explains the closeness of the three profiles in the first type of profile for  $T=7.9$  Myr (the right panel in Fig. 4a).





**Fig. 4** (a) The first type of diversity profile plots  ${}^q\bar{D}(T)$  as a function of order  $q$ ,  $0 \leq q \leq 3$ , for two selected values of temporal perspectives:  $T=0$  (non-phylogenetic case) and  $T=7.9$  Myr (the age of the root of the phylogenetic tree in Fig. 3a). (b) The second type of diversity profile plots  ${}^q\bar{D}(T)$  (phylogenetic Hill number) and  ${}^qPD(T)$  (phylogenetic diversity) as functions of  $T$ ,  $0 \leq T \leq 10$ , separately for  $q=0$ , 1 and 2

To illustrate the phylogenetic differentiation among assemblages, we focus on measuring the phylogenetic differentiation between any two decades for three pairs (i.e. Decades I vs. II, Decades I vs. III and Decades II vs. III). To see how the phylogenetic differentiation measures vary with the time perspective  $q$  and with the order  $T$ , we show two types of profiles for each of the two differentiation measures  $1 - \bar{C}_{qN}(T)$  and  $1 - \bar{U}_{qN}(T)$  in Figs. 5 and 6. In Fig. 5a, we present the first type of profile that plots the measure  $1 - \bar{C}_{qN}(T)$  as a function of  $q$  where  $q$  is in the range  $[0, 3]$  for two time perspectives:  $T=0$  (non-phylogenetic case) and  $T=7.9$  Myr (the age of the root node). In Fig. 5b, the same type of differentiation profile is shown for the other measure  $1 - \bar{U}_{qN}(T)$ . Then in Fig. 6a, b, we present the second type of



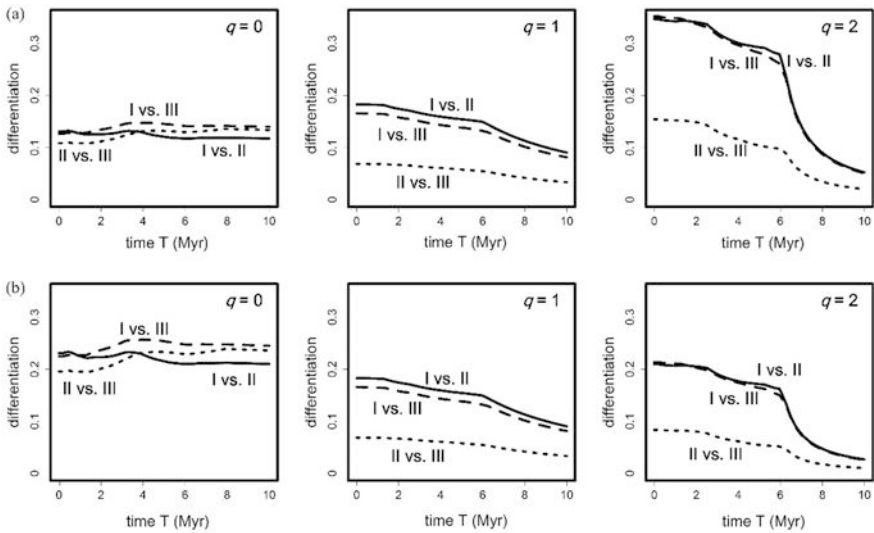
**Fig. 5** (a) Differentiation profiles of the measure  $1 - \bar{C}_{qN}(T)$  and (b) of the measure  $1 - \bar{U}_{qN}(T)$  as a function of order  $q$ ,  $0 \leq q \leq 3$ , for two specific time perspectives:  $T=0$  (left panels, corresponding to non-phylogenetic differentiation profiles), and  $T=7.9$  Myr (right panels, corresponding to the profiles for the age of the root node of the pooled phylogenetic tree in Fig. 3a) for three pairs of assemblages (I vs. II, I vs. III, and II vs. III)

profile that shows the two measures as a function of temporal perspective  $T$ ,  $0 \leq T \leq 10$ , for  $q=0, 1$  and  $2$  separately.

Based on the two phylogenetic differentiation measures, all profiles in Figs. 5 and 6 show consistent patterns. When species/lineages abundances are discounted ( $q=0$ ), the differences among the differentiation measures of the three pairs of assemblages are not appreciable, as shown in the two left panels in Fig. 6 and in the initial point in each of profiles in Fig. 5. When species/lineages abundances are counted ( $q > 0$ ), the compositional differentiation between Decades I vs. II is generally close to that between Decades I vs. III, and the differentiation between two recent decades (Decades II vs. III) is much lower than any of the other two pairs. This implies that the composition of species/lineage abundances has changed after 1990. Examining the relative abundances for those dominant species listed in

Fig. 3b, we see that the most abundant species *S. paucispinis* (23 %) in Decade I became less abundant in both Decade II (9 %) and Decade III (11 %); the second most abundant species *S. mystinus* (11 %) in Decade I became quite rare in both Decade II (4 %) and Decade III (5 %). Also, the species *S. miniatus* in Decade I was rare, but it became the most dominant species in both Decade II (12 %) and Decade III (25 %). These compositional changes for dominant species help explain the above findings.

As the time perspective  $T$  becomes large, more dominant shared lineages are added to the two assemblages, implying the differentiation between any two assemblages should exhibit a non-increasing trend as  $T$  is increased. Our two differentiation measures for  $q > 0$  in Fig. 6 show the expected decreasing trend, and the decline rates differ for  $q=1$  and  $q=2$ . Based on Fig. 3b, we see that most of the dominant and isolated species began to diverge around 6 Myr ago. Thus, the two differentiation profiles for  $q=1$  and 2 start to decrease sharply around 6 Myr especially for order  $q=2$ . Since the node abundances near roots (where the differentiation values are near zero) are relatively high and dominant in the whole tree, all values of the phylogenetic differentiation measures for  $T=7.9$  Myr (the first type of profile for  $T=7.9$  Myr in the right panel of Fig. 5) are substantially lower than their corresponding non-phylogenetic differentiation measure by comparing two figures ( $T=0$  and  $T=7.9$  Myr) in each row of Fig. 5. The two types of profiles (in Fig. 5a, b, and 6a, b) demonstrate that the two differentiation measures  $1-\bar{C}_{qN}(T)$  and  $1-\bar{U}_{qN}(T)$  can incorporate the differences in both tree structure and lineage abundances.



**Fig. 6** (a) Differentiation profiles of the measure  $1-\bar{C}_{qN}(T)$  and (b) of the measure  $1-\bar{U}_{qN}(T)$ , as a function of the time perspective (or time-depth)  $T$ ,  $0 \leq T \leq 10$ , for  $q=0$  (left panel),  $q=1$  (middle panel), and  $q=2$  (right panel) for three pairs of assemblages. All measures are computed for the interval  $[-T, 0]$ , where  $T$  varies from 0 to 10

In summary, our phylogenetic diversity measures have shown an appreciable loss of species, lineage and evolutionary history in rockfish assemblage over time due to fishing pressure, and our phylogenetic differentiation measures show a pronounced change of species/lineages composition after 1990.

## Conclusion

1. To quantify phylogenetic diversity of an assemblage, we suggest using two measures: (i) the phylogenetic Hill number  ${}^q\bar{D}(T)$  (Eqs. 4a and 4b) which measures the “the effective number of equally abundant and equally distinct lineages all with branch lengths  $T$ ”, and (ii) the phylogenetic or branch diversity  ${}^qPD(T)$  (Eqs. 5a and 5b) which measures the “effective total lineage-length”, i.e., the total evolutionary history on an assemblage since time  $T$ . These two measures depend explicitly on two parameters, the abundance sensitivity parameter  $q$  and the time perspective (or time-depth) parameter  $T$ .
2. Two types of diversity profiles are recommended for considering species/branch abundances and phylogenetic information: (i) The first type of diversity profile is obtained by plotting  ${}^qPD(T)$  or  ${}^q\bar{D}(T)$  as a function of order  $q$ , for some selected values of temporal perspective  $T$  including  $T=0$  (i.e., the non-phylogenetic profile based on the ordinary Hill numbers), and  $T$  = the age of the most basal node. See the upper panels of Fig. 4 for an example. It would be also informative to include  $T$  = the age of the divergence between the group under study and the rest of the tree. (ii) The second type of diversity profile is obtained by plotting  ${}^qPD(T)$  and  ${}^q\bar{D}(T)$  as functions of  $T$  separately for  $q=0, 1$ , and  $2$ ; see the middle and lower panels of Fig. 4 for an example. The second type of profile shows the effect of time-depth or evolution change on our diversity measures.
3. When there are multiple assemblages, the phylogenetic gamma Hill number is the effective number of equally abundant and equally distinct lineages in the pooled assemblage; the phylogenetic alpha Hill number is the effective number of equally abundant and equally distinct lineages per assemblage. Thus the phylogenetic beta Hill number, as the ratio of gamma and alpha, is interpreted as “the number of phylogenetically completely distinct assemblages”. In this case, alpha and beta are unrelated (or independent). The difference of phylogenetic gamma and alpha Hill numbers is lineage excess, which is dependent on both alpha and gamma. The phylogenetic beta Hill number and lineage excess lead to the same classes of similarity and differentiation measures, listed in section “[Normalized phylogenetic similarity measures](#)”. See Table 1 for the two major classes of phylogenetic overlap measures,  $\bar{C}_{qN}(T)$  from a local perspective and  $\bar{U}_{qN}(T)$  from a regional perspective.
4. To assess the phylogenetic resemblance or differentiation among assemblages, two types of similarity or differentiation profiles as those in Point 2 are suggested for the two major classes of measures,  $\bar{C}_{qN}(T)$  and  $\bar{U}_{qN}(T)$  (Table 1); see Figs. 5 and 6 for examples.

**Acknowledgements** The authors thank Roseli Pellens, Samuel Scheiner and an anonymous reviewer for helpful comments and suggestions. This work was supported by the Taiwan National Science Council under Contracts 100-2118-M007-006 and 103-2628-M007-007 for AC. CC is supported by a post-doctoral fellowship, National Tsing Hua University, Taiwan. LJ was supported by a donation from John V. Moore to the Population Biology Foundation. All numerical results presented in section “An example” of this chapter were obtained by expanding the R scripts provided in Pavoine et al. (2009, their Appendix S1) to include Hill numbers and our phylogenetic measures.

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Allen B, Kon M, Bar-Yam Y (2009) A new phylogenetic diversity measure generalizing the Shannon index and its application to phyllostomid bats. *Am Nat* 174(2):236–243
- Bryant JA, Lamanna C, Morlon H, Kerkhoff AJ, Enquist BJ, Green JL (2008) Microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. *Proc Natl Acad Sci U S A* 105:11505–11511
- Cavender-Bares J, Ackerly DD, Kozak KH (2012) Integrating ecology and phylogenetics: the footprint of history in modern-day communities. *Ecology* 93(sp8):S1–S3
- Cavender-Bares J, Kozak KH, Fine PV, Kembel SW (2009) The merging of community ecology and phylogenetic biology. *Ecol Lett* 12(7):693–715
- Chao A (2005) Species estimation and applications. In: Kotz S, Balakrishnan N, Read CB, Vidakovic B (eds) *Encyclopedia of statistical sciences*, vol 12, 2nd edn. Wiley, New York, pp 7907–7916
- Chao A, Jost L, Chiang SC, Jiang Y-H, Chazdon RL (2008) A two-stage probabilistic approach to multiple-community similarity indices. *Biometrics* 64(4):1178–1186
- Chao A, Chiu C-H, Jost L (2010) Phylogenetic diversity measures based on Hill numbers. *Phil Trans R Soc London B* 365:3599–3609
- Chao A, Chiu C-H, Hsieh TC (2012) Proposing a resolution to debates on diversity partitioning. *Ecology* 93(9):2037–2051
- Chao A, Chiu C-H, Hsieh TC, Davis T, Nipperess DA, Faith DP (2015) Rarefaction and extrapolation of phylogenetic diversity. *Methods Ecol Evol* 6:380–388
- Chiu C-H, Jost L, Chao A (2014) Phylogenetic beta diversity, similarity, and differentiation measures based on Hill numbers. *Ecol Monogr* 84(1):21–44
- Colwell RK, Coddington JA (1994) Estimating terrestrial biodiversity through extrapolation. *Philos Trans R Soc Lond B* 345:101–118
- Crozier R (1997) Preserving the information content of species: genetic diversity, phylogeny, and conservation worth. *Annu Rev Ecol Syst* 28:243–268
- Daróczy Z (1970) Generalized information functions. *Inf Control* 16(1):36–51
- de Bello F, Lavorel S, Díaz S, Harrington R, Cornelissen JH, Bardgett RD, Berg MP, Cipriotti P, Feld CK, Hering D (2010) Towards an assessment of multiple ecosystem processes and services via functional traits. *Biol Cons* 19(10):2873–2893
- Ellison AM (2010) Partitioning diversity. *Ecology* 91:1962–1963
- Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biol Cons* 61(1):1–10
- Faith DP (2013) Biodiversity and evolutionary history: useful extensions of the PD phylogenetic diversity assessment framework. *Ann NY Acad Sci* 1289(1):69–89
- Faith DP, Richards ZT (2012) Climate change impacts on the tree of life: changes in phylogenetic diversity illustrated for acropora corals. *Biology* 1(3):906–932

- Faith DP, Lozupone CA, Nipperess D, Knight R (2009) The cladistic basis for the phylogenetic diversity (PD) measure links evolutionary features to environmental gradients and supports broad applications of microbial ecology's "phylogenetic beta diversity" framework. *Int J Mol Sci* 10:4723–4741
- Ferrier S, Manion G, Elith J, Richardson K (2007) Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Divers Distrib* 13(3):252–264
- Gotelli NJ, Colwell RK (2011) Estimating species richness. In: Magurran AE, McGill BJ (eds) *Biological diversity: frontiers in measurement and assessment*. Oxford University Press, Oxford, pp 39–54
- Hannah L, Kay JA (1977) *Concentration in the modern industry: theory, measurement, and the U.K. experience*. MacMillan, London
- Havrdá J, Charvát F (1967) Quantification method of classification processes: concept of structural  $\alpha$ -entropy. *Kybernetika* 3(1):30–35
- Helmus MR, Bland TJ, Williams CK, Ives AR (2007) Phylogenetic measures of biodiversity. *Am Nat* 169(3):68–83
- Hill M (1973) Diversity and evenness: a unifying notation and its consequences. *Ecology* 54(2):427–432
- Horn HS (1966) Measurement of "overlap" in comparative ecological studies. *Am Nat* 100:419–424
- Hubbell SP (2001) *A unified theory of biodiversity and biogeography*. Princeton University Press, Princeton
- Hurlbert SH (1971) The nonconcept of species diversity: a critique and alternative parameters. *Ecology* 52(4):577–586
- Hyde JR, Vetter RD (2007) The origin, evolution, and diversification of rockfishes of the genus *Sebastes* (Cuvier). *Mol Phylogenet Evol* 44:790–811
- Ives AR, Helmus MR (2010) Phylogenetic Metrics of Community Similarity. *Am Nat* 176:128–142
- Ives AR, Helmus MR (2011) Generalized linear mixed models for phylogenetic analyses of community structure. *Ecol Monogr* 81(3):511–525
- Jost L (2006) Entropy and diversity. *Oikos* 113(2):363–375
- Jost L (2007) Partitioning diversity into independent alpha and beta components. *Ecology* 88:2427–2439
- Jost L (2009) Mismeasuring biological diversity: response to Hoffman and Hoffman (2008). *Ecol Econ* 68:925–927
- Jost L (2010) Independence of alpha and beta diversities. *Ecology* 91:1969–1974
- Jost L, DeVries P, Walla T, Greeney H, Chao A, Ricotta C (2010) Partitioning diversity for conservation analyses. *Divers Distrib* 16(1):65–76
- Keylock C (2005) Simpson diversity and the Shannon-Wiener index as special cases of a generalized entropy. *Oikos* 109:203–207
- Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. *Genetics* 49(4):725–738
- Lozupone C, Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* 71(12):8228–8235
- MacArthur RH (1965) Patterns of species diversity. *Biol Rev* 40(4):510–533
- MacArthur RH (1972) *Geographical ecology*. Harper & Row, New York
- MacArthur RH, Wilson EO (1967) *The theory of island biogeography*. Princeton University Press, Princeton
- Magurran AE (2004) *Measuring biological diversity*. Blackwell, Oxford
- Magurran AE, McGill BJ (eds) (2011) *Biological diversity: frontiers in measurement and assessment*. Oxford University Press, Oxford
- McPeck MA, Miller TE (1996) Evolutionary biology and community ecology. *Ecology* 77(5):1319–1320

- Morisita M (1959) Measuring of interspecific association and similarity between communities. *Memoires of the Faculty of Science, Kyushu University. Ser E (Biol)* 3:65–80
- Mouchet MA, Moullot D (2011) Decomposing phylogenetic entropy into  $\alpha$ ,  $\beta$  and  $\gamma$  components. *Biol Lett* 7(2):205–209
- Nipperess DA, Matsen FA IV (2013) The mean and variance of phylogenetic diversity under rarefaction. *Methods Ecol Evol* 4:566–572
- Patil GP, Taillie C (1979) A study on diversity profiles and orderings for a bird community in the vicinity of Colstrip, Montana. In: Patil GP, Rosenzweig M (eds) *Contemporary quantitative ecology and related econometrics*. International Co-operative Publishing House, Fairland, pp 23–48
- Patil GP, Taillie C (1982) Diversity as a concept and its measurement. *J Am Stat Assoc* 77:548–561
- Pavoine S, Love MS, Bonsall MB (2009) Hierarchical partitioning of evolutionary and ecological patterns in the organization of phylogenetically structured species assemblages: application to rockfish (genus: *Sebastes*) in the Southern California Bight. *Ecol Lett* 12(9):898–908
- Pavoine S, Baguette M, Bonsall MB (2010) Decomposition of trait diversity among the nodes of a phylogenetic tree. *Ecol Monogr* 80(3):485–507
- Peet RK (1974) The measurement of species diversity. *Annu Rev Ecol Syst* 5:285–307
- Petchey OL, Gaston KJ (2002) Functional diversity (FD), species richness and community composition. *Ecol Lett* 5(3):402–411
- Pielou EC (1975) *Ecological diversity*. Wiley, New York
- Purvis A, Hector A (2000) Getting the measure of biodiversity. *Nature* 405(6783):212–219
- Rao CR (1982) Diversity and dissimilarity coefficients: a unified approach. *Theor Popul Biol* 21(1):24–43
- Rényi A (1961) On measures of entropy and information. In: Neyman J (ed) *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, vol 1. University of California Press, Berkeley, pp 547–561
- Ricotta C, Szeidl L (2009) Diversity partitioning of Rao's quadratic entropy. *Theor Popul Biol* 76(4):299–302
- Routledge R (1979) Diversity indices: which ones are admissible? *J Theor Biol* 76(4):503–515
- Tilman D (2001) Functional diversity. In: Levin AS (ed) *Encyclopedia of biodiversity*, vol 3. Academic Press, San Diego, pp 109–120
- Tsallis C (1988) Possible generalization of Boltzmann-Gibbs statistics. *J Stat Phys* 52(1):479–487
- Vellend M, Cornwell WK, Magnuson-Ford K, Mooers A (2011) Measuring phylogenetic biodiversity. In: Magurran AE, McGill BJ (eds) *Biological diversity: frontiers in measurement and assessment*. Oxford University Press, Oxford, pp 194–207
- Warwick R, Clarke K (1995) New 'biodiversity' measures reveal a decrease in taxonomic distinctness with increasing stress. *Mar Ecol Prog Ser* 129(1):301–305
- Webb CO (2000) Exploring the phylogenetic structure of ecological communities: an example for rain forest trees. *Am Nat* 156:145–155
- Webb CO, Ackerly DD, McPeck MA, Donoghue MJ (2002) Phylogenies and community ecology. *Annu Rev Ecol Syst* 33:475–505
- Webb CO, Losos JB, Agrawal AA (2006) Integrating phylogenies into community ecology. *Ecology* 87(sp7):S1–S2
- Weiher E (2011) A primer of trait and functional diversity. In: Magurran AE, McGill BJ (eds) *Biological diversity: frontiers in measurement and assessment*. Oxford University Press, Oxford, pp 175–193
- Whittaker RH (1972) Evolution and measurement of species diversity. *Taxon* 12:213–251

# Split Diversity: Measuring and Optimizing Biodiversity Using Phylogenetic Split Networks

Olga Chernomor, Steffen Klaere, Arndt von Haeseler, and Bui Quang Minh

**Abstract** About 20 years ago the concepts of phylogenetic diversity and phylogenetic split networks were separately introduced in conservation biology and evolutionary biology, respectively. While it has been widely recognized that biodiversity assessment should better take into account the phylogenetic tree of life, it has also been widely acknowledged that phylogenetic networks are more appropriate for phylogenetic analysis in the presence of hybridization, horizontal gene transfer, or contradicting trees among genomic loci. Here, we aim to combine phylogenetic diversity and networks into one concept, split diversity (SD), which properly measures biodiversity for conflicting phylogenetic signals. Moreover, we reformulate well-known conservation questions under the SD framework and present computational methods to solve these, in general, computationally intractable questions. Notably, integer programming, a technique widely used to solve many real-life problems, serves as a general and efficient strategy that delivers optimal solutions to many biodiversity optimization problems. We finally discuss future directions for the new concept.

**Keywords** Biodiversity optimization • Phylogenetic diversity • Phylogenetic split networks • Split diversity • Integer programming

---

O. Chernomor • A. von Haeseler

Center for Integrative Bioinformatics Vienna, Max F. Perutz Laboratories,  
University of Vienna, Medical University of Vienna, Vienna, Austria

Bioinformatics and Computational Biology, Faculty of Computer Science,  
University of Vienna, Vienna, Austria

e-mail: [olga.chernomor@univie.ac.at](mailto:olga.chernomor@univie.ac.at); [arndt.von.haeseler@univie.ac.at](mailto:arndt.von.haeseler@univie.ac.at)

S. Klaere

Department of Statistics, School of Biological Sciences,  
University of Auckland, Auckland, New Zealand

e-mail: [s.klaere@auckland.ac.nz](mailto:s.klaere@auckland.ac.nz)

B.Q. Minh (✉)

Center for Integrative Bioinformatics Vienna, Max F. Perutz Laboratories,  
University of Vienna, Medical University of Vienna, Vienna, Austria

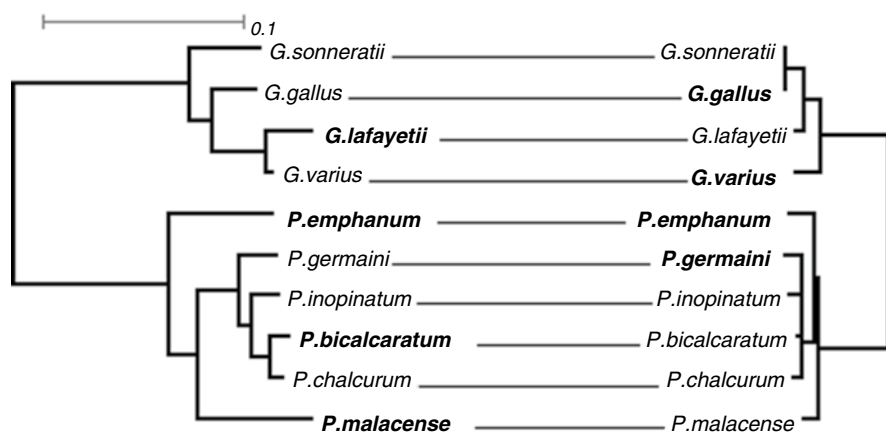
e-mail: [minh.bui@univie.ac.at](mailto:minh.bui@univie.ac.at)



## Introduction

The previous book chapters show that in the presence of phylogenetic information it is more appropriate to assess biodiversity based on phylogenetic trees than on the concept of species richness (see also May 1990; Vane-Wright et al. 1991). *Phylogenetic diversity* (PD; Faith 1992) is a popular measure of the amount of evolutionary history encompassed by the species under consideration. Given a phylogenetic tree for a set of taxa, PD of a taxon subset is defined as the sum of the branch lengths of the minimal subtree connecting those taxa. The definition of PD per se requires “a reliable estimate of phylogenetic relationships among the taxa” (Faith 1992). However, such a reliable estimate is sometimes hard to obtain due to, for example, model misspecification (Jermiin et al. 2008) or even intrinsically non-treelike evolutionary patterns. More recently, phylogenomic studies often revealed conflicting phylogenetic signals among genomic loci, adding the complication how to compute PD from multiple trees.

Figure 1 illustrates the problem. Here, phylogenetic trees are reconstructed for ten pheasant species from the mitochondrial cytochrome *b* gene (CYB) and the intron 3 of the dimerization cofactor of hepatocyte nuclear factor 1 (DCoH3) (data from Kimball and Braun 2008). The two resulting trees, denoted by  $T_{CYB}$  and  $T_{DCoH3}$ , clearly separate the two genera *Gallus* (junglefowl) and *Polyplectron* (peacock-pheasant). However, they strongly contradict within the *Gallus* clade. For example, *G. sonneratii* (grey junglefowl) and *G. varius* (green junglefowl) are the basal *Gallus* species in  $T_{CYB}$  and  $T_{DCoH3}$ , respectively. The trees also disagree on the phylogenetic positions of *P. emphanum* (Palawan peacock-pheasant) and *P. malacense* (Malayan peacock-pheasant). Moreover, edge lengths of the trees represented by



**Fig. 1** Maximum likelihood phylogenetic trees inferred with IQ-TREE (Minh et al. 2013) from the mitochondrial CYB and the nuclear intron DCoH3 for four *Gallus* (junglefowl) and six *Polyplectron* (peacock-pheasant) species. The *scalebar* represents the expected number of nucleotide substitutions per site. Highlighted in *boldface* are the four species maximizing phylogenetic diversity

the expected numbers of substitutions per site substantially differ between the trees. This particular example reflects the fact that the evolutionary relationships among these birds are still controversial and more data is needed to elucidate the galliform tree of life (e.g., Wang et al. 2013).

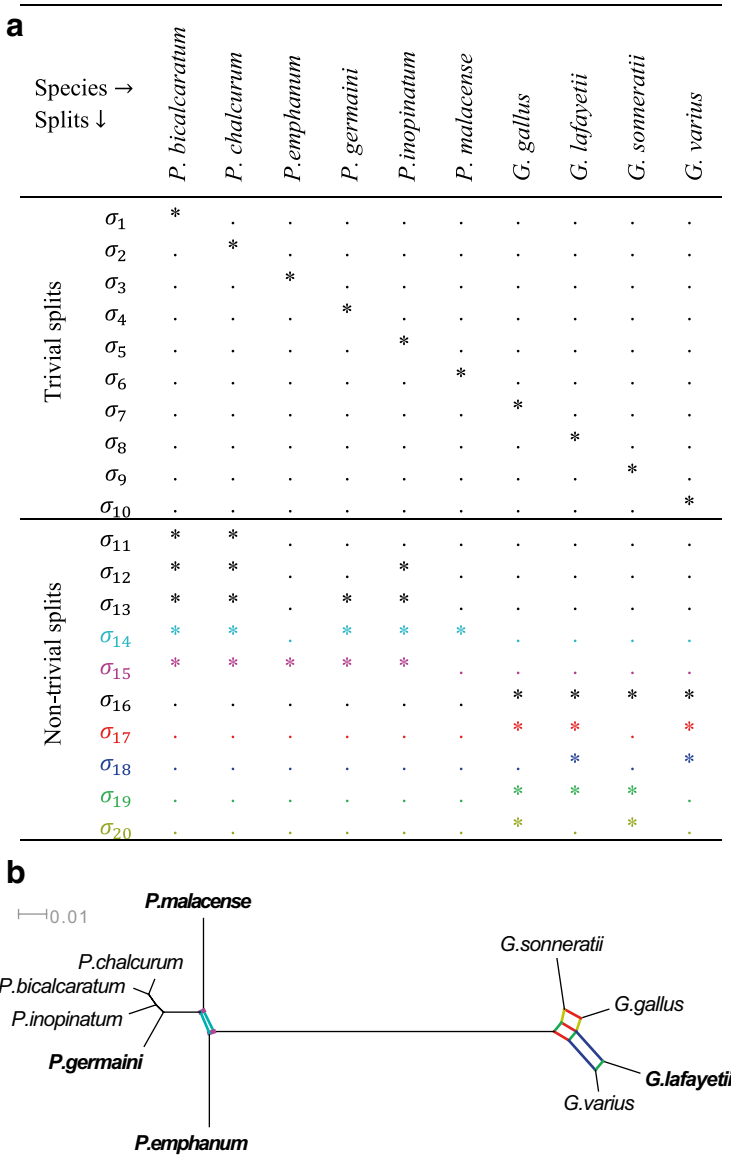
If one is interested in selecting four species maximizing PD, then one indeed ends up with two different sets of species (highlighted in bold-face, Fig. 1) and only *P. emphanum* occurs in both subsets.

To resolve this issue, we introduced the concept of *Split Diversity* (SD), which generalizes PD by combining information from multiple trees (Minh et al. 2009). For example, SD of a taxon set can be defined as the average PD of the two trees. By maximizing SD one then simultaneously maximizes PDs over all trees, which captures conflicting phylogenetic signals between the trees. Moreover, computing SD this way is equivalent to computing “phylogenetic diversity” from the so-called *phylogenetic split networks* (Bandelt and Dress 1992a; Huson et al. 2010). SD has also been recently applied to prioritize populations for conservation (Volkman et al. 2014). In the following we formalize the concept of split networks and the measure of split diversity. Further, we reformulate well-known biodiversity optimization problems under the framework of SD, present algorithmic solutions and computational tools to these problems. Finally conclude the chapter with future perspectives.

## Phylogenetic Split Networks

Rooted phylogenetic trees as shown in Fig. 1 are well understood. Here, both trees show that the common ancestor of the taxa considered has the ancestors of the two genera as direct descendants. In general, interior nodes indicate ancestral taxa of the leaf nodes, and the edge lengths give an estimate of the amount of change observed between nodes. However, if one wishes to combine the information in both trees, it becomes difficult to identify clear ancestors. For example,  $T_{CYB}$  and  $T_{DCoH3}$  disagree whether *G. sonneratii* or *G. varius* is the basal Gallus species. In order to visualize these conflicts phylogenetic split networks have been devised.

We start by describing splits. A *split*, denoted by  $A|B$ , is defined as a bipartition of the taxon set  $X$  into two disjoint subsets  $A$  and  $B$ , indicating that there is an observable amount of divergence between the two subsets. Every edge in a tree generates a split. If one removes an edge, the tree decomposes into two subtrees, each of which connects a unique set of leaves.  $T_{CYB}$  has 17 splits (edges), while  $T_{DCoH3}$  has 15 splits (2 splits in  $T_{DCoH3}$  have zero length and are collapsed as they do not influence subsequent computations). Figure 2a shows the union set  $\Sigma$  of 20 distinct splits occurring in the pheasant trees (Fig. 1).  $T_{CYB}$  and  $T_{DCoH3}$  share the ten *trivial splits*  $\sigma_1, \sigma_2, \dots, \sigma_{10}$  corresponding to external edges of the trees. The trees also share two non-trivial splits  $\sigma_{13}$  and  $\sigma_{16}$ , where  $\sigma_{16}$  corresponds to the internal edges separating Gallus from Polyplectron species. The remaining splits are unique to each tree.



**Fig. 2** (a) Set of all splits extracted from the trees in Fig. 1. Each split  $\sigma$  is a bipartition  $A|B$ , where ‘\*’ and ‘.’ represent taxa in  $A$  and  $B$ , respectively. Conflicting splits are colored. (b) Visualization of this split set as a phylogenetic split network. Conflicting splits are colored accordingly and depicted by parallelograms. Here, split weights are assigned as the mean of the weight of the corresponding edges in the two trees. Highlighted in **boldface** are the four species maximizing split diversity

This split set is visualized in a phylogenetic split network (Fig. 2b). The major difference to trees is that the interior nodes of a split network cannot be regarded as representing ancestral taxa. Instead, the weight of a split  $A|B$  indicates the amount of difference between the taxon set  $A$  and  $B$ . A split is visualized by a single edge or a set of parallel edges. The former indicates that the split does not conflict any other splits, while the latter indicates at least one conflict. Therefore, two conflicting splits are visualized by a parallelogram. For example,  $\sigma_{14}$  (in cyan, Fig. 2) and  $\sigma_{15}$  (in pink) contradict each other on the placement of *P. emphanum* and *P. malacense*. This disagreement generates a narrow parallelogram at the basal Polyplectron.

If more than two splits are in disagreement, the split network will show multiple connected parallelograms. For example,  $\sigma_{17}$  (in red, Fig. 2) conflicts with  $\sigma_{19}$  (in green) and  $\sigma_{20}$  (in yellow).  $\sigma_{19}$  also contradicts  $\sigma_{18}$  (in blue). Therefore,  $\sigma_{17}, \sigma_{18}, \sigma_{19}$  and  $\sigma_{20}$  are visualized by three red, two blue, three green, and two yellow parallel edges, respectively. This generates three parallelograms within Gallus (Fig. 2b).

Not every split set can be visualized in two dimensions. For example, assuming that we had a third tree that places *G. gallus* at the basal Gallus lineage. This would introduce one split contradicting with both  $\sigma_{17}$  and  $\sigma_{19}$ . These triple-wise conflicting splits are depicted by a three dimensional parallelepiped. The resulting split network is not easily visualized anymore. However, for the following it suffices to directly work on the split set (Fig. 2a).

## The Measure of Split Diversity

Given a split set  $\Sigma$ , the SD of a taxon subset  $Y$  is defined as the sum of the weights  $\lambda$  of all splits separating taxa in  $Y$ . Here, a split  $A|B \in \Sigma$  separates  $Y$  if  $Y \cap A$  and  $Y \cap B$  are both non-empty. Thus, we get

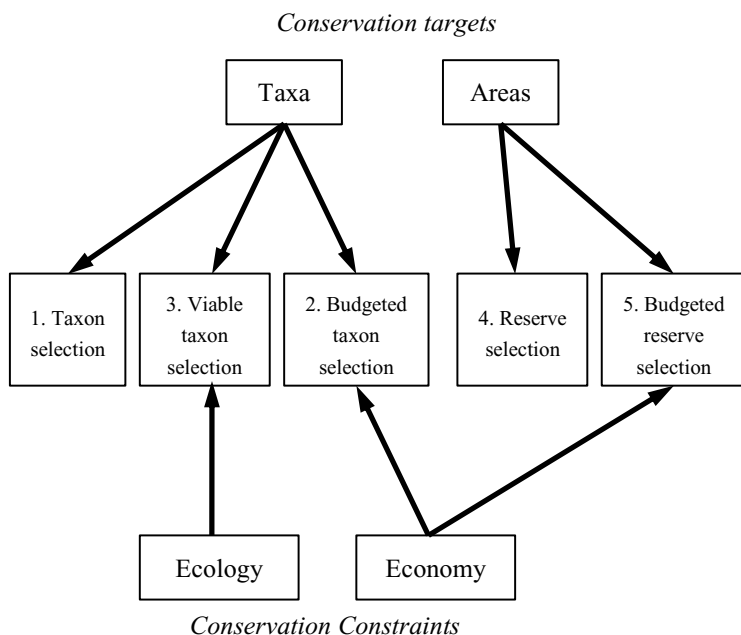
$$SD(Y) = \sum_{\sigma \in \Sigma: \sigma \text{ separates } Y} \lambda_{\sigma}$$

To illustrate, given  $\Sigma$  in Fig. 2, for  $Y = \{P. malacense, P. germaini, P. emphanum, G. lafayetii\}$  we have  $SD(Y) = \lambda_3 + \lambda_4 + \lambda_6 + \lambda_8 + \lambda_{13} + \lambda_{14} + \dots + \lambda_{19}$ , where  $\lambda_i = \lambda_{\sigma_i}$  is defined as the average of the corresponding branch lengths in  $T_{CYB}$  and  $T_{DCoH3}$ . Here, contradicting splits such as  $\sigma_{17}$  and  $\sigma_{19}$  are considered in the SD computation.

If the split set  $\Sigma$  corresponds to a tree (i.e. no conflicting splits exist in  $\Sigma$ ), then SD is equivalent to PD. The definition of SD therefore generalizes PD. For this reason we focus on SD for the remaining of the chapter.

## Biodiversity Optimization Problems

Conservation problems mainly fall into two categories: taxon selection and reserve selection (Fig. 3), where the conservation targets are either taxa or geographical areas, respectively. Under PD, the simplest taxon selection problem (Faith 1992) is



**Fig. 3** The “network” of biodiversity optimization problems

to identify a subset of  $k$  taxa that maximizes PD on a phylogenetic tree of  $n$  taxa ( $2 \leq k < n$ ). For reserve selection we define PD on a subset of areas as the PD of the union taxon set of the areas. The simplest reserve selection problem is analogously to identify a subset of  $k$  areas that maximizes PD over all subsets of  $k$  areas. In the following, we reformulate these problems using SD and further integrate economical and ecological constraints into the extensions.

## ***Taxon Selection Problems***

We start with the simplest taxon selection problem formally defined as:

### ***Problem 1 (Taxon Selection)***

Given a phylogenetic split set for  $n$  taxa, find a subset of  $k$  taxa that maximizes SD over all subsets of  $k$  taxa.

As an illustration, given the split set for ten pheasants (Fig. 2) we want to select four taxa maximizing SD. By doing so we yield an optimal subset (highlighted in

bold-face; Fig. 2b), which shares three taxa (*P. emphanum*, *P. malacense*, and *G. lafayetii*) with the CYB-based subset (left panel of Fig. 1) and only two taxa (*P. emphanum* and *P. germaini*) with DCoH3-based subset (right panel of Fig. 1). The SD approach therefore provides a “consensus” solution over the two independent PD analyses. Problem 1 is known to be NP-hard (Spillner et al. 2008), which means that to find an optimal set it may, in the worst case, necessary to compute the SD for the exponentially many subsets  $n$ .

Problem 1 implicitly assumes that each taxon requires the same amount of resources for conservation. If we knew the preservation costs for each taxon and were provided with a finite budget, then a more realistic scenario is to allocate this budget among the taxa so as to obtain the highest diversity. This process is known as conservation triage (Bottrill et al. 2008) and formally defined as:

**Problem 2 (Budgeted Taxon Selection)**

Given a split set and conservation costs for each taxon, find a subset of taxa whose total conservation costs do not exceed a predefined budget while maximizing SD.

Problem 1 and 2 ignore ecological relationships between taxa. In real life species interact with each other within a dependency network such as predator-prey relationships (Witting et al. 2000; van der Heide et al. 2005; Moulton et al. 2007). In general, a dependency network is, typically, an acyclic directed graph, where nodes in the graph represent taxa and edges represent dependencies between nodes. Figure 4 shows an artificial example of such a network for the pheasants. Here, *G. sonneratii* depends on *P. malacense* and *P. germaini*, depicted by two edges connecting *G. sonneratii* with these two taxa. We note that this is a purely fictional example, but it illustrates the major principles of including a dependency structure in conservation decisions.

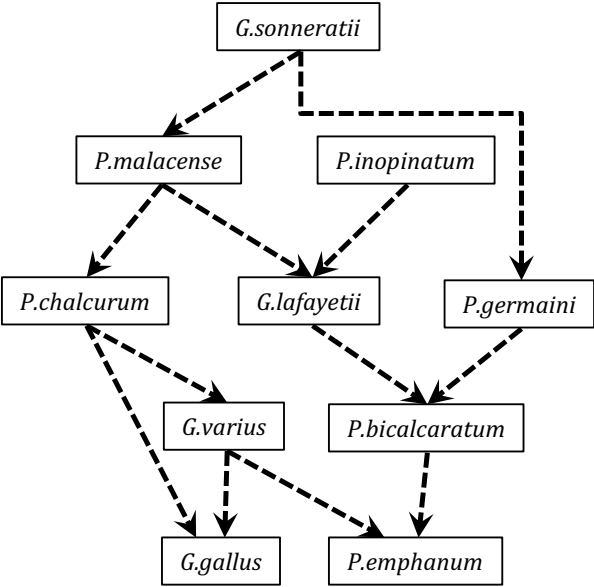
A taxon is called *viable in a subset* of taxa if this taxon does not depend on any other taxon, or if it does depend on some taxa, then at least one of them is also present in the subset. For example, *G. sonneratii* is viable in a subset if this subset also contains *P. malacense* or *P. germaini*. *P. emphanum* and *G. gallus* are viable in any (sub)set since they do not depend on any other species.

A subset is called *viable* if all its taxa are viable in this set. For example, {*P. emphanum*, *P. bicalcaratum*, *P. germaini*, *G. sonneratii*} is a viable subset, whereas {*P. emphanum*, *P. bicalcaratum*, *G. lafayetii*, *G. sonneratii*} is not viable.

We now formally define the viable taxon selection problem as

**Problem 3 (Viable Taxon Selection)**

Given a split set and a dependency network, find a viable subset of  $k$  taxa, which maximizes SD.



**Fig. 4** Artificial example of dependency network for the pheasant data set

**Reserve Selection Problems**

For reserve selection we define the SD of a subset of areas as the SD of the union set of taxa present in these areas. The reserve selection is formalized as:

**Problem 4 (Reserve Selection)**  
Given a split set for  $n$  taxa distributed in  $m$  areas, find a subset of  $k$  areas that maximizes SD over all subsets of  $k$  areas.

To illustrate the problem consider the geographical distribution of the ten pheasants (Table 1). The data were obtained from the global biodiversity information facility ([www.gbif.org](http://www.gbif.org); accessed on December 1st, 2013), where a country is listed as habitat only if there are at least three observations for the species. Table 1 shows that these pheasants occur in eight countries in South Asia. *G. gallus* and *P. bicalcaratum* occur in seven and two countries, respectively, whereas the remaining species are endemic to one country. Indonesia and Malaysia each host three species, Sri Lanka only one species, and the remaining five countries are home to two species each.

If one wants to select four countries with maximal diversity, then the decision heavily depends on the trees or network (Figs. 1 and 2b). Table 2 shows that using

**Table 1** Presence/absence of ten pheasants in eight countries obtained from the global biodiversity information facility (<http://www.gbif.org>)

Species	Bhutan, BT	Indonesia, ID	India, IN	Sri Lanka, LK	Philippines, PH	Malaysia, MY	Thailand, TH	Vietnam, VN
<i>P. bicalcaratum</i>	x						x	
<i>P. chalcuorum</i>		x						
<i>P. emphanum</i>					x			
<i>P. germaini</i>								x
<i>P. inopinatum</i>						x		
<i>P. malacense</i>						x		
<i>G. gallus</i>	x	x	x		x	x	x	x
<i>G. lafayetii</i>				x				
<i>G. sonneratii</i>			x					
<i>G. varius</i>		x						



**Table 2** Four countries maximizing PD on the CYB tree (first column), PD on the DCoH3 tree (second column), and SD on the split network (third column)

PD – CYB	PD – DCoH3	SD
<b>Malaysia</b> (3)	Indonesia (3)	<b>Malaysia</b> (3)
<b>Philippines</b> (2)	<b>Malaysia</b> (3)	<b>Philippines</b> (2)
Sri Lanka (1)	<b>Philippines</b> (2)	Indonesia (3)
India (2)	Vietnam (2)	India (2)

Highlighted in boldface are the countries present in all optimal sets. The number of species present in the country is given in brackets

the CYB and DCoH3 regions, the optimal sets only overlap in two countries: Malaysia and Philippines. If we now maximize SD instead, then the optimal set includes these two countries, the third one preferred by the PD-DCoH3 set (Indonesia), and the fourth one by the PD-CYB set (India). The union of the species sets for the selected areas contains seven species.

If budget data is available, then we have a budgeted reserve selection problem. Here, preserving these species in each country comes at a cost and we need to select those countries that maximize SD within an allocated budget.

**Problem 5 (Budgeted Reserve Selection)**

Given a split set for  $n$  taxa distributed on  $m$  areas and conservation costs for each area, find a subset of areas whose total conservation costs do not exceed a predefined budget while maximizing SD.

## Computational Methods in Conservation Planning

The algorithms to solve the aforementioned Problems 1–5 are those that are guaranteed to produce an optimal solution, often referred to as exact algorithms, and those that are not. The former includes algorithms that are based on integer programming and dynamic programming, whereas the latter comprise greedy algorithms, approximation algorithms and algorithms based on simulated annealing. We will start with greedy algorithms, as they are simple and probably most widely applied in conservation planning.

### Greedy Algorithms

Greedy algorithms are a simple and general heuristic strategy but, usually, do not guarantee optimal solutions. Kirkpatrick (1983) was probably the first to apply a greedy algorithm to find a solution to Problem 4, the simple reserve selection, but

under the species richness concept. His greedy algorithm coined “complementarity principle” first identifies the most species-rich area. In the second step, it finds the area, which “adds” the most numbers of new species to the firstly chosen area. This is repeated until  $k$  areas are obtained. Such a complementarity principle has been applied to maximize PD (Faith 1992) and also applied elsewhere (e.g., Vane-Wright et al. 1991; Pressey et al. 1997). Recently, Bordewich and Semple (2008) have proven that the greedy algorithm applied to Problem 5 under PD will generate a solution that has at least ~63 % of the PD of the optimal solution, which is the best possible approximation ratio.

The only case, where a greedy algorithm delivers the optimal solution is the taxon selection (Problem 1) under PD on trees (Pardi and Goldman 2005; Steel 2005). An efficient implementation of such a greedy algorithm (Minh et al. 2006) finds a solution for trees with millions of taxa within seconds on a standard PC. Greedy algorithms have been further examined in conservation biology (Moulton et al. 2007; Bordewich et al. 2008).

Obviously greedy algorithms can be applied for Problems 1–5 to maximize SD. The general idea is to start with one target (either taxon or area) having the highest SD. We then choose the second target “adding” the most SD while still satisfying the constraints (budget or viability constraints). We repeat this step until no further target can be added (e.g., exceeding  $k$  targets for Problem 1, 3, and 4 or exceeding the budget for Problem 2 and 5). As an illustration the greedy algorithm is applied for Problem 4 to find four countries showing the highest pheasant SD for the split network (Fig. 2b) and known geographical distribution (Table 1) as follows. Malaysia is first selected as it contains the highest SD. Philippines, Indonesia, and India are selected in the next steps. In this particular example the greedy algorithm happens to obtain the optimal set of four countries (Table 2).

## ***Integer Programming***

Integer Programming (IP; Dantzig et al. 1954; Gomory 1958) is a widely used and powerful optimization technique to solve a variety of decision-making problems (Wolsey 1998; Jünger et al. 2010). IP methods maximize or minimize a linear objective function subject to linear constraints (equalities or inequalities) when one or more variables are restricted to be integers. Theoretically solving IP is NP-hard. However, thanks to powerful solvers like CPLEX (2012) and GUROBI (Gurobi Optimization Inc. 2013), problems with thousands of variables and constraints can be solved optimally within reasonable time (Jünger et al. 2010; and references therein).

The first application of IP in conservation problems goes back to (Cocks and Baird 1989), who solved the reserve selection (Problem 4) under species richness. Such IP formulations have been extended to more realistic scenarios (Underhill 1994; Church et al. 1996; Possingham et al. 2000), to maximize PD (Rodrigues and Gaston 2002; Rodrigues et al. 2005), and to maximize SD (Minh et al. 2010).

Here, we show how to model biodiversity optimization problems 1–5 in IP parlance, which allows available IP software packages to solve the problem. We first introduce some notations and definitions and further exemplify IP formulations for Problems 1–5 using the pheasant data set.

### IP for Taxon Selection Problems

Given a set of  $n$  taxa, we encode a subset  $S$  by an  $n$ -element binary vector with entries of 0 and 1 indicating the absence and presence of the corresponding taxa in  $S$ . The elements of this vector are called taxon variables. For the pheasant data set there are ten taxon variables  $x_{PB}, x_{PC}, x_{PE}, x_{PG}, x_{PI}, x_{PM}, x_{GG}, x_{GL}, x_{GS}, x_{GV}$  (indices follow initials of species names). We say that a split  $\sigma = A|B$  is preserved in  $S$  if  $A$  and  $B$  each contain at least one taxon from  $S$ . For each split  $\sigma$  we introduce a binary split variable  $y_\sigma$ , where  $y_\sigma = 1$  if  $\sigma$  is preserved in  $S$  and 0 otherwise.

Each  $y_\sigma$  is fully identified from taxon variables by two split constraints as follows.  $\sigma_1$  is a trivial split that separates *P. bicalcaratum* from the remaining taxa.  $\sigma_1$  is preserved (i.e.  $y_1 = 1$ ) if *P. bicalcaratum* and at least another taxon are preserved (see Fig. 2a for the definition of the splits). This condition is expressed by two inequalities:

$$y_1 \leq x_{PB}, y_1 \leq x_{PC} + x_{PE} + x_{PG} + x_{PI} + x_{PM} + x_{GG} + x_{GL} + x_{GS} + x_{GV}$$

In fact, the second inequality always holds because  $k \geq 2$  and thus is ignored. Now consider the non-trivial split  $\sigma_{17}$ , which separates *G. gallus*, *G. lafayetii*, and *G. varius* from the remaining taxa.  $\sigma_{17}$  is preserved if at least one of *G. gallus*, *G. lafayetii*, and *G. varius* and one of the remaining taxa are preserved. Therefore,

$$y_{17} \leq x_{GG} + x_{GL} + x_{GV}, y_{17} \leq x_{PB} + x_{PC} + x_{PI} + x_{PG} + x_{PE} + x_{PM} + x_{GS}$$

The remaining split constraints are listed in Table 3.

Based on split variables one can rewrite SD of  $S$  as:

$$SD(S) = \sum_{\sigma} \lambda_{\sigma} y_{\sigma} \quad (1)$$

where  $\lambda_{\sigma}$  is the weight of split  $\sigma$ . This is the objective function that we want to maximize for all problems (1–5).

In the taxon selection Problem 1 the size of an optimal subset is constrained by a predefined number  $k$ , meaning that:

$$x_{PB} + x_{PC} + x_{PE} + x_{PG} + x_{PI} + x_{PM} + x_{GG} + x_{GL} + x_{GS} + x_{GV} \leq k \quad (2)$$

We also require that taxon and split variables are binary

$$x_i \in \{0,1\}, \forall \text{ taxon } i \quad (3)$$

$$y_\sigma \in \{0,1\}, \forall \text{ split } \sigma \quad (4)$$

### ***IP Formulation of Problem 1***

Maximize objective function (1), subject to subset size constraint (2), binary constraints (3, 4), split constraints (5) (see Table 3).

Suppose we are given a total budget  $B$ . Let  $c_i$  denote conservation costs for taxon  $i$ . We can then substitute constraint (2) by the budget constraint

$$\sum_i c_i x_i \leq B \quad (6)$$

Together with previous constraints we have the IP formulation of Problem 2 by:

### ***IP Formulation for Problem 2***

Maximize objective function (1), subject to budget constraint (6), binary constraints (3, 4), and split constraints (5) (Table 3).

We now model viability constraints that operate on taxon variables as follows. *G. sonneratii* depends on *P.malacense* and *P.germaini* (Fig. 4). Therefore, the viability constraint for *G. sonneratii* is simply

$$x_{PM} + x_{PG} \geq x_{GS}$$

This ensures that  $x_{GS}$  is 1 (i.e., *G. sonneratii* is selected for conservation) only if at least one of  $x_{PM}$  and  $x_{PG}$  is also 1. Viability constraints for all the other taxa are listed in Table 3. Now, the IP formulation for viable taxon selection can be obtained by simply including viability constraints to Problem 1:

### ***IP Formulation of Problem 3***

Maximize objective function (1), subject to subset size constraint (2), binary constraints (3, 4), split constraints (5), and viability constraints (7) (Table 3).

## **IP for Reserve Selection Problems**

For reserve selection we encode a subset  $W$  of  $m$  areas by a binary vector  $(z_1, z_2, \dots, z_m)$ , where  $z_r$  is 1 if area  $r$  is present in  $W$ , and 0 otherwise. We call  $z_r$  area variables. For the pheasant habitat (Table 1) we have eight area variables  $z_{ID}$ ,  $z_{LK}$ ,  $z_{BT}$ ,  $z_{IN}$ ,  $z_{PH}$ ,  $z_{MY}$ ,

$z_{TH}$ ,  $z_{VN}$  (indices follow two-letter country codes). We now redefine split constraints in terms of area variables instead of taxon variables as follows.

Split  $\sigma_{18}$ , which separates *G. lafayetii* and *G. varius* from the others, is preserved if (1) *G. lafayetii* or *G. varius* is preserved and (2) at least one of the remaining taxa is preserved. Because *G. lafayetii* or *G. varius* occur in Indonesia and Sri Lanka, condition 1 is equivalent to:

$$y_{18} \leq z_{ID} + z_{LK}$$

Similarly condition 2 is equivalent to:

$$y_{18} \leq z_{BT} + z_{ID} + z_{IN} + z_{PH} + z_{MY} + z_{TH} + z_{VN}$$

since the remaining taxa are found in all areas except Sri Lanka. Such area-split constraints for all other splits are listed in Table 4.

The subset size constraint has to be rewritten for countries:

$$z_{ID} + z_{LK} + z_{BT} + z_{IN} + z_{PH} + z_{MY} + z_{TH} + z_{VN} \leq k \quad (8)$$

We keep binary constraints for split variables and also include such for area variables

$$z_r \in \{0,1\} \forall \text{ area } r \quad (9)$$

Reserve selection problem is then formulated as follows:

#### ***IP Formulation of Problem 4***

Maximize objective function (1), subject to subset size constraint (8), binary constraints (4, 9), and area-split constraints (10) (Table 4).

For budgeted reserve selection we are given a total budget  $B$ . Let  $c_{ID}$ ,  $c_{LK}$ ,  $c_{BT}$ ,  $c_{IN}$ ,  $c_{PH}$ ,  $c_{MY}$ ,  $c_{TH}$ ,  $c_{VN}$  denote conservation costs for each country. Then a budget constraint for areas is

$$\sum_r c_r z_r \leq B \quad (11)$$

To obtain the IP formulation for Problem 5 we simply substitute subset size constraint (8) by the budget constraint (11).

#### ***IP Formulation of Problem 5***

Maximize objective function (1), subject to budgetary constraint (11), binary constraints (4, 9), and area-split constraints (10) (Table 4).

## ***Other Algorithms***

While greedy algorithms and IP are general strategies for all Problems 1–5, other algorithms have been applied to solve special cases. For example, simulated annealing algorithms (Possingham et al. 2000) were introduced to solve the reserve selection Problems 4 and 5 under species richness with an opportunity to minimize the connectivity between the areas such as the boundary lengths. Dynamic programming algorithms (DPA) have been applied to solve Problem 2 under PD (Pardi and Goldman 2007). DPA was further extended to maximize SD on circular split networks (Minh et al. 2009a, b). Other special types of split networks were exploited to solve Problem 1 (Spillner et al. 2008; Bordewich et al. 2009).

## **Computer Software**

Conservation planning software like *Marxan* (Ball et al. 2009) and *Zonation* (Moilanen et al. 2009) mainly focus on species richness. However, both programs can indirectly account for phylogenetic diversity (see also Silvano, Valdujo and Colli, chapter “[Priorities for Conservation of the Evolutionary History of Amphibians in the Cerrado](#)” and Arponen and Zupan, chapter “[Representing Hotspots of Evolutionary History in Systematic Conservation Planning for European Mammals](#)”). Only a few programs explicitly allow to compute phylogenetic diversity (Webb et al. 2008; Kembel et al. 2010). In the following we describe two programs relevant for the SD analysis.

### ***SplitsTree***

SplitsTree (Huson and Bryant 2006) is a user-friendly and leading software to reconstruct and visualize phylogenetic networks from multiple sequence alignments, distance matrices, or sets of trees. SplitsTree implements a wide range of split network inference methods such as split decomposition (Bandelt and Dress 1992b) and neighbor-net (Bryant and Moulton 2004). SplitsTree has a limited ability to compute PD and SD. It works for all major platforms including Windows, Mac OS X, and Unix. More information about SplitsTree is available at <http://www.splitsree.org>.

### ***PDA: Phylogenetic Diversity Analyzer***

PDA (Minh et al. 2009) is a software tool that computes and maximizes species richness, PD, and SD given a variety of user-defined constraints including budget, ecological, and geographical constraints. PDA can be used in conjunction with SplitsTree to work with SD. It solves all Problems 1–5 by greedy algorithms,

dynamic programming, and integer programming methods. Moreover, it supports weighted dependency networks for viable taxon selection and spatial reserve selection problems (Chernomor et al. 2015). Among other features is the computation of PD/SD endemism and complementarity (Faith et al. 2004). PDA is available as a command-line program for Windows, Mac OS X, and Unix as well as an online web service. More information about PDA is available at <http://www.cibiv.at/software/pda>.

## Conclusions and Perspectives

In this chapter we have presented the concept of split diversity, a generalization of PD to account for contradicting phylogenetic information in biodiversity optimization. We demonstrated the new concept with a small pheasant data set. We note that this example is not realistic because neither genera are vulnerable nor the selection of entire countries is reasonable. Moreover, genetic data for galliforms are available for more genera and genomic loci (Wang et al. 2013) and the methodology developed here is well applicable to this new data.

We then presented computational tools to perform the analysis under the SD framework. Both greedy algorithms and IP can be generally applied to solve the same conservation questions, where the former quickly computes a solution and the latter ensures optimal solutions. Moreover, IP works well for data set sizes usually encountered in real data. For example, we have recently applied IP to solve the viable taxon selection (Problem 3) for 242 marine species of Caribbean coral community and the budgeted reserve selection (Problem 5) for the Cape of South Africa with 735 plant genera (Chernomor et al. 2015). IP always returned optimal sets of taxa and areas within seconds to a few minutes.

SD can be extended to include species extinction risks as developed for PD (Weitzman 1992; Witting and Loeschcke 1995). Such a “probabilistic” PD approach (see chapters “[The Value of Phylogenetic Diversity](#)” and “[Reconsidering the Loss of Evolutionary History: How Does Non-random Extinction Prune the Tree-of-Life?](#)”) predicts future diversity given the fact that some species might become extinct in, say, 20 years. The problem, previously coined the Noah’s Ark Problem (NAP; Weitzman 1998), is then to maximize future PD given limited budgets. The same concept can be applied to SD as follows. One first computes “survival probabilities” for each split in split networks in the same fashion as for branches in phylogenetic trees. The future SD is then defined as the dot product of the split weights and split survival probabilities. This definition of future SD consistently generalizes that of future PD.

From a computational view point, solving the extended NAP under future SD is NP-hard as proven for PD (Hartmann and Steel 2006). Dynamic programming algorithms (DPA) optimally solve the NAP under future PD in a special scenario, where the species extinction probability becomes 0 if it is given enough resources (Pardi and Goldman 2007). For general scenarios Hickey et al. (2008) devised such a DPA that gives an approximation ratio of nearly 1 compared to the optimal solution. More recently, Billionnet (2013) presented an IP approach for the NAP that runs within a few minutes for simulated 4,000-taxon cases and provides near-optimal solutions, which are only 1.2 % away from the optimal solution. It will be interesting to investigate how such DPA and IP approaches can be adapted to solve the NAP under the more general SD framework.

**Acknowledgments** The authors thank Tung Lam Nguyen for developing PDA web service. The authors also thank Roseli Pellens and Philippe Grandcolas for inviting us to write this chapter and three anonymous reviewers for constructive comments. This work was supported by University of Vienna (Initiativkolleg I059-N) to O.C. and the Austrian Science Fund – FWF (I760-B17) to B.Q.M. and A.v.H.

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.



## Appendix

**Table 3** Objective function and constraints of taxon selection problems for the pheasant example

Maximize	$\lambda_1 y_1 + \dots + \lambda_{20} y_{20}$	(1)
Subject to		
Size constraint:	$x_{PB} + x_{PC} + x_{PE} + x_{PG} + x_{PI} + x_{PM} + x_{GG} + x_{GL} + x_{GS} + x_{GV} \leq k$	(2)
Binary constraints:	$x_i \in \{0,1\} \forall \text{ taxon } i$	(3)
	$y_\sigma \in \{0,1\} \forall \sigma = 1, \dots, 20$	(4)
Split constraints:	$y_i \leq x_i \forall \text{ taxon } i$	(5)
	$y_{11} \leq x_{PB} + x_{PC}$	
	$y_{11} \leq x_{PE} + x_{PG} + x_{PI} + x_{PM} + x_{GG} + x_{GL} + x_{GS} + x_{GV}$	
	$y_{12} \leq x_{PB} + x_{PC} + x_{PI}$	
	$y_{12} \leq x_{PE} + x_{PG} + x_{PM} + x_{GG} + x_{GL} + x_{GS} + x_{GV}$	
	$y_{13} \leq x_{PB} + x_{PC} + x_{PG} + x_{PI}$	
	$y_{13} \leq x_{PE} + x_{PM} + x_{GG} + x_{GL} + x_{GS} + x_{GV}$	
	$y_{14} \leq x_{PB} + x_{PC} + x_{PG} + x_{PI} + x_{PM}$	
	$y_{14} \leq x_{PE} + x_{GG} + x_{GL} + x_{GS} + x_{GV}$	
	$y_{15} \leq x_{PB} + x_{PC} + x_{PE} + x_{PG} + x_{PI}$	
	$y_{15} \leq x_{PM} + x_{GG} + x_{GL} + x_{GS} + x_{GV}$	
	$y_{16} \leq x_{PB} + x_{PC} + x_{PE} + x_{PG} + x_{PI} + x_{PM}$	
	$y_{16} \leq x_{GG} + x_{GL} + x_{GS} + x_{GV}$	
	$y_{17} \leq x_{PB} + x_{PC} + x_{PE} + x_{PG} + x_{PI} + x_{PM} + x_{GS}$	
	$y_{17} \leq x_{GG} + x_{GL} + x_{GV}$	
	$y_{18} \leq x_{PB} + x_{PC} + x_{PE} + x_{PG} + x_{PI} + x_{PM} + x_{GG} + x_{GS}$	
	$y_{18} \leq x_{GL} + x_{GV}$	
	$y_{19} \leq x_{PB} + x_{PC} + x_{PE} + x_{PG} + x_{PI} + x_{PM} + x_{GV}$	
	$y_{19} \leq x_{GG} + x_{GL} + x_{GS}$	
	$y_{20} \leq x_{PB} + x_{PC} + x_{PE} + x_{PG} + x_{PI} + x_{PM} + x_{GL} + x_{GV}$	
	$y_{20} \leq x_{GG} + x_{GS}$	

(continued)

**Table 3** (continued)

Budget constraint:	$c_{PB}x_{PB} + c_{PC}x_{PC} + \dots + c_{GV}x_{GV} \leq B$	(6)
Viability constraints:	$x_{PB} \leq x_{PE}$	(7)
	$x_{PC} \leq x_{GG} + x_{GV}$	
	$x_{PG} \leq x_{PB}$	
	$x_{PI} \leq x_{GL}$	
	$x_{PM} \leq x_{PC} + x_{GL}$	
	$x_{GL} \leq x_{PB}$	
	$x_{GS} \leq x_{PM} + x_{PG}$	
	$x_{GV} \leq x_{GG} + x_{PE}$	

**Table 4** Objective function and constraints of reserve selection problems for the pheasant example. Due to the fact that *G. gallus* is contained in all but one area there are many area-split constraints of the form  $y_\sigma \leq z_{BT} + z_{ID} + z_{IN} + z_{LK} + z_{PH} + z_{MY} + z_{TH} + z_{VN}$ . Such constraints are redundant since  $k \geq 2$ , and thus omitted

Maximize	$\lambda_1 y_1 + \dots + \lambda_{20} y_{20}$	(1)
Subject to		
Size constraint:	$z_{ID} + z_{LK} + z_{BT} + z_{IN} + z_{PH} + z_{MY} + z_{TH} + z_{VN} \leq k$	(8)
Binary constraints:	$z_r \in \{0,1\} \forall \text{ area } r$	(9)
	$y_\sigma \in \{0,1\} \forall \sigma = 1, \dots, 20$	(4)
Area-split constraints:	$y_1 \leq z_{BT} + z_{TH}$ $y_2 \leq z_{ID}$ $y_3 \leq z_{PH}$ $y_4 \leq z_{VN}$ $y_5 \leq z_{MY}$ $y_6 \leq z_{MY}$ $y_7 \leq z_{BT} + z_{ID} + z_{IN} + z_{PH} + z_{MY} + z_{TH} + z_{VN}$ $y_8 \leq z_{LK}$ $y_9 \leq z_{IN}$ $y_{10} \leq z_{ID}$ $y_{11} \leq z_{BT} + z_{ID} + z_{TH}$ $y_{12} \leq z_{BT} + z_{ID} + z_{MY} + z_{TH}$ $y_{13} \leq z_{BT} + z_{ID} + z_{MY} + z_{TH} + z_{VN}$ $y_{14} \leq z_{BT} + z_{ID} + z_{MY} + z_{TH} + z_{VN}$ $y_{15} \leq z_{BT} + z_{ID} + z_{PH} + z_{MY} + z_{TH} + z_{VN}$ $y_{16} \leq z_{BT} + z_{ID} + z_{PH} + z_{MY} + z_{TH} + z_{VN}$ $y_{17} \leq z_{BT} + z_{ID} + z_{IN} + z_{PH} + z_{MY} + z_{TH} + z_{VN}$ $y_{18} \leq z_{BT} + z_{ID} + z_{IN} + z_{PH} + z_{MY} + z_{TH} + z_{VN}$ $y_{18} \leq z_{ID} + z_{LK}$ $y_{19} \leq z_{BT} + z_{ID} + z_{PH} + z_{MY} + z_{TH} + z_{VN}$ $y_{20} \leq z_{BT} + z_{ID} + z_{LK} + z_{PH} + z_{MY} + z_{TH} + z_{VN}$ $y_{20} \leq z_{BT} + z_{ID} + z_{IN} + z_{PH} + z_{MY} + z_{TH} + z_{VN}$	(10)
Budget constraint:	$c_{ID} z_{ID} + c_{LK} z_{LK} + \dots + c_{VN} z_{VN} \leq B$	(11)

## References

- Ball IR, Possingham HP, Watts M (2009) Marxan and relatives: software for spatial conservation prioritisation. In: Moilanen A, Wilson KA, Possingham HP (eds) *Spatial conservation prioritisation: quantitative methods and computational tools*. Oxford University Press, New York, pp 185–195
- Bandelt HJ, Dress AWM (1992a) A canonical decomposition-theory for metrics on a finite-set. *Adv Math* 92:47–105
- Bandelt HJ, Dress AWM (1992b) Split decomposition: a new and useful approach to phylogenetic analysis of distance data. *Mol Phylogenet Evol* 1:242–252
- Billionnet A (2013) Solution of the generalized Noah's Ark problem. *Syst Biol* 62:147–156
- Bordewich M, Semple C (2008) Nature reserve selection problem: a tight approximation algorithm. *IEEE/ACM Trans Comput Biol Bioinform* 5:275–280
- Bordewich M, Rodrigo AG, Semple C (2008) Selecting taxa to save or sequence: desirable criteria and a greedy solution. *Syst Biol* 57:825–834
- Bordewich M, Semple C, Spillner A (2009) Optimizing phylogenetic diversity across two trees. *Appl Math Lett* 22:638–641
- Bottrill MC, Joseph LN, Carwardine J, Bode M, Cook C, Game ET, Grantham H, Kark S, Linke S, McDonald-Madden E, Pressey RL, Walker S, Wilson KA, Possingham HP (2008) Is conservation triage just smart decision making? *Trends Ecol Evol* 23:649–654
- Bryant D, Moulton V (2004) Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol* 21:255–265
- Chernomor O, Minh BQ, Forest F, Klaere S, Ingram T, Henzinger M, von Haeseler A (2015) Split diversity in constrained conservation prioritization using integer programming. *Methods Ecol Evol* 6:83–91
- Church RL, Stoms DM, Davis FW (1996) Reserve selection as a maximal covering location problem. *Biol Conserv* 76:105–112
- Cocks KD, Baird IA (1989) Using mathematical-programming to address the multiple reserve selection problem – an example from the Eyre Peninsula, South-Australia. *Biol Conserv* 49:113–130
- CPLEX (2012) IBM ILOG CPLEX optimizer
- Dantzig G, Fulkerson R, Johnson S (1954) Solution of a large-scale traveling-salesman problem. *J Oper Res Soc Am* 2:393–410
- Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61:1–10
- Faith DP, Reid CAM, Hunter J (2004) Integrating phylogenetic diversity, complementarity, and endemism for conservation assessment. *Conserv Biol* 18:255–261
- Gomory RE (1958) Outline of an algorithm for integer solutions to linear programs. *Bull Am Math Soc* 64:275–278
- Gurobi Optimization Inc (2013) Gurobi optimizer reference manual
- Hartmann K, Steel M (2006) Maximizing phylogenetic diversity in biodiversity conservation: greedy solutions to the Noah's Ark problem. *Syst Biol* 55:644–651
- Hickey G, Carmi P, Maheshwari A, Zeh N (2008) NAPX: a polynomial time approximation scheme for the Noah's Ark problem. *Algorithm Bioinforma Wabi* 5251:76–86
- Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23:254–267
- Huson DH, Rupp R, Scornavacca C (2010) *Phylogenetic networks: concepts, algorithms and applications*. Cambridge University Press, Cambridge
- Jermiin LS, Jayaswal V, Ababneh F, Robinson J (2008) Phylogenetic model evaluation. In: Keith (ed) *Bioinformatics: data, sequences analysis and evolution*. Humana Press, Totowa, pp 331–363
- Jünger M, Liebling TM, Naddef D, Nemhauser GL, Pulleyblank WR, Reinelt G, Rinaldi G, Wolsey LA (2010) 50 years of integer programming 1958–2008: from the early years to the state-of-the-art. Springer, Heidelberg

- Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, Blomberg SP, Webb CO (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26:1463–1464
- Kimball RT, Braun EL (2008) A multigene phylogeny of Galliformes supports a single origin of erectile ability in non-feathered facial traits. *J Avian Biol* 39:438–445
- Kirkpatrick JB (1983) An iterative method for establishing priorities for the selection of nature reserves: an example from Tasmania. *Biol Conserv* 25:127–134
- May RM (1990) Taxonomy as destiny. *Nature* 347:129–130
- Minh BQ, Klaere S, von Haeseler A (2006) Phylogenetic diversity within seconds. *Syst Biol* 55:769–773
- Minh BQ, Klaere S, von Haeseler A (2009a) Taxon selection under split diversity. *Syst Biol* 58:586–594
- Minh BQ, Pardi F, Klaere S, von Haeseler A (2009b) Budgeted phylogenetic diversity on circular split systems. *IEEE/ACM Trans Comput Biol Bioinform* 6:22–29
- Minh BQ, Klaere S, von Haeseler A (2010) SDA\*: a simple and unifying solution to recent bioinformatic challenges for conservation genetics. In: Pham SB, Hoang TH, McKay B, Hirota K (eds) *The second international conference on knowledge and systems engineering*. IEEE Computer Society, Hanoi, pp 33–37
- Minh BQ, Nguyen MAT, von Haeseler A (2013) Ultrafast approximation for phylogenetic bootstrapping. *Mol Biol Evol* 30:1188–1195
- Moilanen A, Kujala H, Leathwick JR (2009) The zonation framework and software for conservation prioritization. In: Moilanen A, Wilson KA, Possingham HP (eds) *Spatial conservation prioritization: quantitative methods and computational tools*. Oxford University Press, New York
- Moulton V, Semple C, Steel M (2007) Optimizing phylogenetic diversity under constraints. *J Theor Biol* 246:186–194
- Pardi F, Goldman N (2005) Species choice for comparative genomics: being greedy works. *PLoS Genet* 1:e71
- Pardi F, Goldman N (2007) Resource-aware taxon selection for maximizing phylogenetic diversity. *Syst Biol* 56:431–444
- Possingham HP, Ball IR, Andelman S (2000) Mathematical methods for identifying representative reserve networks. In: Ferson S, Burgman M (eds) *Quantitative methods for conservation biology*. Springer, New York, pp 291–305
- Pressey RL, Possingham HP, Day JR (1997) Effectiveness of alternative heuristic algorithms for identifying indicative minimum requirements for conservation reserves. *Biol Conserv* 80:207–219
- Rodrigues ASL, Gaston KJ (2002) Maximising phylogenetic diversity in the selection of networks of conservation areas. *Biol Conserv* 105:103–111
- Rodrigues ASL, Brooks TM, Gaston KJ (2005) Integrating phylogenetic diversity in the selection of priority areas for conservation: does it make a difference? In: Purvis A, Gittleman JL, Brooks T (eds) *Phylogeny and conservation*. Cambridge University Press, Cambridge, pp 101–119
- Spillner A, Nguyen BT, Moulton V (2008) Computing phylogenetic diversity for split systems. *IEEE/ACM Trans Comput Biol Bioinform* 5:235–244
- Steel M (2005) Phylogenetic diversity and the greedy algorithm. *Syst Biol* 54:527–529
- Underhill LG (1994) Optimal and suboptimal reserve selection algorithms. *Biol Conserv* 70:85–87
- van der Heide CM, van den Bergh JCJM, van Ierland EC (2005) Extending Weitzman's economic ranking of biodiversity protection: combining ecological and genetic considerations. *Ecol Econ* 55:218–223
- Vane-Wright RI, Humphries CJ, Williams PH (1991) What to protect – systematics and the agony of choice. *Biol Conserv* 55:235–254
- Volkman L, Martyn I, Moulton V, Spillner A, Mooers AO (2014) Prioritizing populations for conservation using phylogenetic networks. *PLoS One* 9:e88945

- Wang N, Kimball RT, Braun EL, Liang B, Zhang ZW (2013) Assessing phylogenetic relationships among Galliformes: a multigene phylogeny with expanded taxon sampling in Phasianidae. *PLoS One* 8:e64312
- Webb CO, Ackerly DD, Kembel SW (2008) Phylocom: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics* 24:2098–2100
- Weitzman ML (1992) On diversity. *Q J Econ* 107:363–405
- Weitzman ML (1998) The Noah's Ark problem. *Econometrica* 66:1279–1298
- Witting L, Loeschcke V (1995) The optimization of biodiversity conservation. *Biol Conserv* 71:205–207
- Witting L, Tomiuk J, Loeschcke V (2000) Modelling the optimal conservation of interacting species. *Ecol Model* 125:123–143
- Wolsey LA (1998) *Integer programming*. Wiley-Interscience, New York

# The Rarefaction of Phylogenetic Diversity: Formulation, Extension and Application

David A. Nipperess

**Abstract** Like other measures of diversity, Phylogenetic Diversity (PD) increases monotonically and asymptotically with increasing sample size. This relationship can be described by a rarefaction curve tracing the expected PD for a given number of accumulation units. Accumulation units represent individual organisms, collections of organisms (e.g. sites), or even species (or equivalent), giving individual-based, sample-based and species-based curves respectively. The formulation for the exact analytical solution for the rarefaction of PD is given in an expanded form to demonstrate congruence with the classic formulation for the rarefaction of species richness. Rarefaction is commonly applied as a standardisation for diversity values derived from differing numbers of sampling units. However, the solution can be simply extended to create measures of phylogenetic evenness, phylogenetic beta-diversity and phylogenetic dispersion, derived from individual-based, sample-based and species-based curves respectively. This extension, termed  $\Delta PD$ , is simply the initial slope of the rarefaction curve and is related to entropy measures such as PIE (Probability of Interspecific Encounter) and Gini-Simpson entropy. The application of rarefaction of PD to sample standardisation and measurement of phylogenetic evenness, phylogenetic beta-diversity and phylogenetic dispersion is demonstrated. Future prospects for PD rarefaction include the recognition of evolutionary hotspots (independent of species richness), the basis for ecological theory such as phylogeny-area relationships, and the prediction of unseen biodiversity.

**Keywords** Alpha diversity • Beta diversity • Evenness • Phylogeny • Sampling curves

---

D.A. Nipperess (✉)

Department of Biological Sciences, Macquarie University, North Ryde, NSW 2109, Australia  
e-mail: [david.nipperess@mq.edu.au](mailto:david.nipperess@mq.edu.au)

© The Author(s) 2016

R. Pellens, P. Grandcolas (eds.), *Biodiversity Conservation and Phylogenetic Systematics*, Topics in Biodiversity and Conservation 14,  
DOI 10.1007/978-3-319-22461-9\_10

197

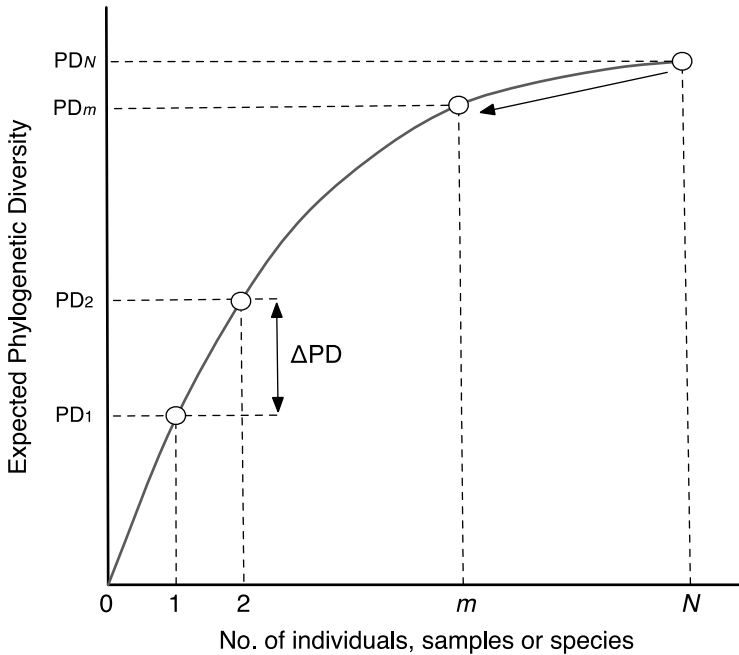
## Introduction

Phylogenetic Diversity (PD) is a simple, intuitive and effective measure of biodiversity. The PD of a set of taxa, represented as the tips of a phylogenetic tree, is the sum of the branch lengths connecting those taxa (Faith 1992). PD is a particularly flexible measure because it can be applied to any set of relationships among entities that can be reasonably portrayed as a tree. Thus, the tips do not, by necessity, need to represent species but could be higher taxa, Operational Taxonomic Units, Evolutionarily Significant Units, individual organisms or unique haplotypes. Further, the tree itself might not portray evolutionary relationships but instead be, for example, a cluster dendrogram portraying functional relationships among taxa (Petchey and Gaston 2002).

Since the original formulation by Faith (1992), PD has come to be not just a single measure equating to a phylogenetically weighted form of richness, but rather a general class of measures dealing with various aspects of alpha and beta-diversity (Faith 2013). The common feature of this class of measures is the summation of branch lengths rather than the counting of tips. By substituting branch segments (intervals between nodes on a phylogenetic tree) for species, and including a weighting for the length of that segment, it is possible to modify many of the classic measures of Species Diversity (SD) to a PD equivalent (Faith 2013). By this means, phylogenetically weighted measures of endemism (Faith et al. 2004; Rosauer et al. 2009), ecological resemblance (Ferrier et al. 2007; Nipperess et al. 2010), and entropy (Chao et al. 2010, and chapter “[Phylogenetic Diversity Measures and Their Decomposition: A Framework Based on Hill Numbers](#)”) have been developed, for example.

In its classic form, PD, like species richness, has the property of concavity (Lande 1996). That is, the addition of individuals or sets of individuals to a community can increase PD but never decrease it. Thus, just like species richness, PD increases monotonically with increasing sampling effort, creating a classic sampling curve that reaches an asymptote when all species (and branch segments) are represented (Fig. 1). Gotelli and Colwell (2001) recognise two general types of sampling curve, individuals-based and sample-based, that are distinguished by the units on the x-axis, representing either individual organisms or samples, respectively. Samples, in this context, are collections of individuals bounded in space and time, corresponding to the common ecological usage of the term. For PD, we can recognise a third type of sampling curve where the units on the x-axis are species or their equivalent (Fig. 1). Species, like samples, are also collections of individuals bounded, in this case, by some minimum degree of relatedness. Obviously, species-based sampling curves are meaningless when plotting species richness but have real value when plotting PD. For the purposes of generalisation, it is useful to be able to refer to these units (individuals, samples, species) with a single term. Chiarucci et al. (2008) used “accumulation units” to refer to individuals and samples. I extend this term to also include species as an additional unit of sampling effort in sampling curves. While these different units (individuals, samples, species) all measure





**Fig. 1** Sampling curve showing the relationship between Phylogenetic Diversity (PD) and sampling depth. The level of sampling is measured in accumulation units of individuals, samples (collections of individuals) or species as required.  $PD_N$  is the Phylogenetic Diversity of the full set of  $N$  accumulation units. Rarefaction is the process (indicated by *unidirectional arrow*) of randomly subsampling (rarefying) the pool of  $N$  accumulation units to a subset of size  $m$  and calculating the expected PD of that subset ( $PD_m$ ).  $\Delta PD$  is the expected gain in PD between the first and second accumulation unit, and can be used as a measure of phylogenetic evenness, beta-diversity or dispersion, depending on the nature of the unit of accumulation

sampling effort in some sense, they are not equivalent and sampling curves derived from them must be interpreted differently in each case.

Beside the units by which sampling effort is measured, Gotelli and Colwell (2001) distinguished between “accumulation curves” and “rarefaction curves”, based on the process by which the sampling curve is calculated. An accumulation curve plots a single ordering of individuals or samples (or species) against a cumulatively calculated concave diversity measure. The jagged shape of the resulting curve is highly dependent on the, often arbitrary, order of the accumulation units. To resolve this problem, rarefaction curves instead plot the *expected* value of the diversity measure against the corresponding number of accumulation units. Rarefaction can be achieved using an algorithmic procedure of repeated random sub-sampling of the full set of accumulation units and calculating the mean diversity (Gotelli and Colwell 2001). However, Hurlbert (1971) and Simberloff (1972) showed that expected diversity can be calculated using an exact analytical solution, obviating the need for computer-intensive repeated sub-sampling. Initially, this solution was for

individuals-based rarefaction curves, but it has since been shown that the same solution applies to sample-based rarefaction (Kobayashi 1974; Ugland et al. 2003; Mao et al. 2005; Chiarucci et al. 2008).

The original purpose of rarefaction was to allow the comparison of datasets with differing amounts of sampling effort (Sanders 1968). Assemblages can be compared “fairly” when rarefied to the same number of accumulation units (Gotelli and Colwell 2001). However, rarefaction has broader application than this single purpose. Depending of the unit of accumulation, the shape of the rarefaction curve provides information on ecological evenness (Olszewski 2004) and beta-diversity (Crist and Veech 2006). Rarefaction of species richness also forms the basis of estimators of species richness, including unseen species (Colwell and Coddington 1994). In the case of PD, species-based rarefaction curves also allow for a measure of phylogenetic dispersion (Webb et al. 2002), effectively the expected PD for some given number of species (Nipperess and Matsen 2013). A solution for the rarefaction of PD is therefore desirable as it will allow for these applications to be realised for phylogenetically explicit datasets.

Rarefaction of Phylogenetic Diversity, using an algorithmic solution of repeated sub-sampling, has now been done several times (see for example Lozupone and Knight 2008; Turnbaugh et al. 2009; Yu et al. 2012). However, an analytical solution for PD rarefaction, similar to that determined by Hurlbert (1971) for species richness, is preferable both because its results are exact (not dependent on the number of repeated subsamples) and substantially more computationally efficient. Nipperess and Matsen (2013) recently published just such a solution for both the mean and variance of PD under rarefaction. This solution is quite general, being applicable to rooted and unrooted trees, and even allowing partition of the tree into smaller components than the individual branch segments. As a result, the solution is given in a very generalised form and its relationship with classic rarefaction formula for species richness is not immediately clear.

In this chapter, I provide a detailed formulation for the exact analytical solution for expected (mean) Phylogenetic Diversity for a given amount of sampling effort. This formulation is for the specific but common case of a rooted phylogenetic tree where whole branch segments are selected under rarefaction. I use the same form of expression as used by Hurlbert (1971) to demonstrate the direct relationship between rarefaction of PD and rarefaction of species richness. I do not include a solution for variance of PD under rarefaction due to its complexity when given in this form and instead refer the reader to Nipperess and Matsen (2013). I extend this framework to show how the initial slope of the rarefaction curve ( $\Delta PD$ ) can be used as a flexible measure of phylogenetic evenness, phylogenetic beta-diversity or phylogenetic dispersion, depending on the unit of accumulation. I apply PD rarefaction and the derived  $\Delta PD$  measure to real ecological datasets to demonstrate its usefulness in addressing ecological questions. Finally, I discuss some future directions for the extension and application of PD rarefaction.

## Formulation

To begin, the classic rarefaction formula for species richness will be reviewed in order to demonstrate how it can be extended to the case of Phylogenetic Diversity. The expected species richness ( $S$ ) for a given amount of sampling is simply the sum of probabilities ( $p$ ) of each species occurring in a subset of  $m$  accumulation units (Eq. 1).

$$E[S]_m = \sum_i^S p_i \quad (1)$$

To solve Eq. 1, we need to determine the probability ( $p$ ) of each species being selected by a random draw of  $m$  accumulation units from the total set of  $N$  units. Regardless of whether the accumulation unit is an individual or a sample, this probability is a function of the frequency ( $n$ ) with which species  $i$  occurs across the set of  $N$  accumulation units (Chiarucci et al. 2008). Since  $N$  is a set of finite size, random draws from that set should be without replacement and thus  $p$  is defined by the hypergeometric distribution (Hurlbert 1971). Substituting into Eq. 1, the expected species richness is as follows (Eq. 2).

$$E[S]_m = \sum_i^S \left[ 1 - \frac{\binom{N-n_i}{m}}{\binom{N}{m}} \right] \quad (2)$$

The quantity within the square brackets in Eq. 2 corresponds to  $p$  in Eq. 1. Note that the expressions in curved brackets are binomial coefficients and not simple fractions, while the quantity subtracted from one within the square brackets is a fraction. The denominator in this fraction gives the number of distinct subsets of size  $m$  that can be drawn from the total set of  $N$  units. The numerator gives the number of distinct subsets of size  $m$  that do not contain species  $i$ . Equation 2 is the same as that originally proposed by Hurlbert (1971).

Phylogenetic Diversity is simply the sum of a set of branch lengths spanning a set of species (or, more generally, tips). So, for a set of  $S$  species, there is a corresponding set of  $T$  branch segments. Each branch segment ( $j$ ) has a length ( $L$ ) measured as sequence substitutions, millions of years, or some other biologically meaningful estimate of difference. Considering only rooted phylogenetic trees, PD is calculated as follows (Eq. 3).

$$PD = \sum_j^T L_j \quad (3)$$

In the original definition intended by Faith (1992), the PD of a subset of species is calculated by summing the branch lengths connecting that set of species to the root of the tree, even when the common ancestor of that subset is not the same as the root. In this definition, a subset containing a single species (or even a single individual) has a non-zero PD value, which in this case, would be the total path length from the tip to the root. This corresponds to the *rooted PD* value of Pardi and Goldman (2007). The alternative, called *unrooted PD* by Pardi and Goldman (2007), includes only the branch segments connecting a subset of species to their common ancestor, and thus a subset containing only a single species would have zero PD. The former definition, rooted PD, is adopted here because it allows for the straightforward formulation of a whole class of derived PD measures (Faith 2013), and because it is concordant with the original idea of PD acting as a surrogate for the feature diversity of a set (Faith 1992; Faith et al. 2009). Obviously, rooted PD requires a rooted phylogenetic tree, even if the choice of root is arbitrary (Nipperess and Matsen 2013).

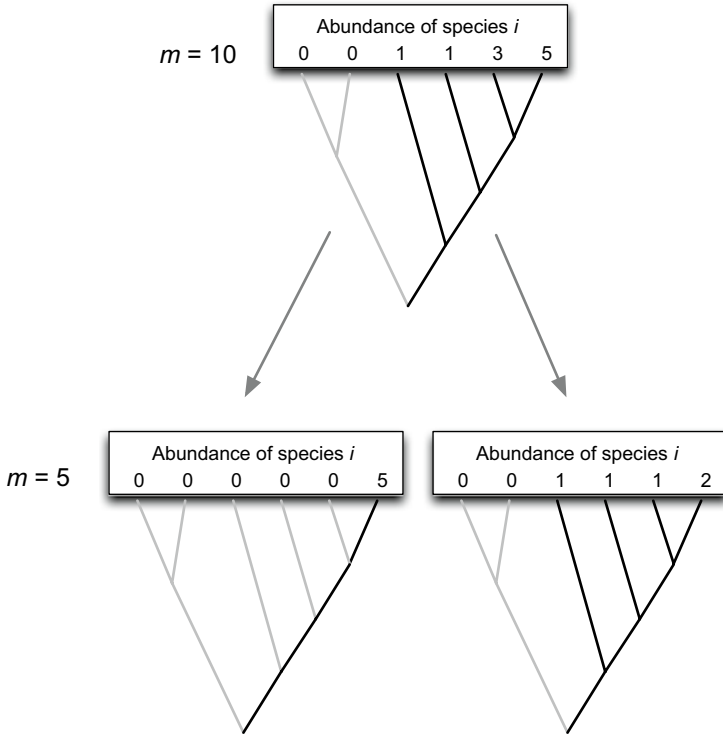
Given this definition, the rarefaction of PD involves finding the expected (average) sum of branch lengths (including the path to the root) for all possible distinct subsets of  $m$  accumulation units (Fig. 2). This is achieved by extending the classic rarefaction formula through a substitution of species for branch segments in a phylogenetic tree. Since PD is simply the sum of branch lengths, then the expected PD must also be the sum of branch lengths, each weighted by the probability ( $q$ ) of its occurrence in a subset of size  $m$  (O'Dwyer et al. 2012). So, for a rooted phylogenetic tree represented as a set of  $T$  branch segments, the expected PD is given as follows (Eq. 4).

$$E[PD]_m = \sum_j^T L_j \times {}_m q_j \quad (4)$$

The probability of each branch segment occurring in a subset is again a function of the frequency with which it occurs among accumulation units. The frequency of occurrence of a particular branch segment ( $o$ ) depends on the frequency of occurrence of species that are descendent from that branch segment. Let  $x$  be a binary value indicating whether species  $i$  is (1) or is not (0) a descendant of branch segment  $j$ . Multiplying  $x$  by  $n$  and summing across all species will give the total number of occurrences of branch segment  $j$  among  $N$  accumulation units (Eq. 5).

$$o_j = \sum_i^S (n_i \times x_{ij}) \quad (5)$$

Thus, by summing across branches instead of species, substituting branch occurrence for species occurrence, and including a branch length weighting, we are able to adapt the classic rarefaction formula for species richness for the purposes of calculating expected Phylogenetic Diversity (Eq. 6). Note this solution is equivalent to that of Nipperess and Matsen (2013) but is expressed in an expanded form for the specific case of calculating rooted PD. Equation 6 is very similar to the solution for



**Fig. 2** An illustration of the process of rarefying Phylogenetic Diversity (PD) by units of individuals. An initial sample of ten individuals ( $m=10$ ) distributed among four tips (species) is rarefied to a subset of five individuals ( $m=5$ ) by a process of random sampling without replacement. For the rarefied samples, 2 of the 252 possible subsets are shown. The expected PD under rarefaction is the average sum of branch lengths represented by each of these distinct subsets. The branch lengths summed to calculate PD are *black* while those not represented (and thus not summed) are *grey*. Note that the rooted definition of PD is used where the path length to the root is always included, even in the case where only a single tip is represented

*expected PD* of Faith (2013) but differs in that random draws are without replacement following the hypergeometric distribution.

$$E[PD]_m = \sum_j^T \left[ L_j \times \left( 1 - \frac{\binom{N-o_j}{m}}{\binom{N}{m}} \right) \right] \quad (6)$$

Finally, it is now possible to calculate the expected PD for a given number of *species*. A species, in this context, is simply a collection of individuals in much the same way as a sample is a collection of individuals, and the same equations apply.

Under these circumstances,  $o_j$  is equal to the sum of  $x_{ij}$  (over all species) as  $n_i$  will always equal 1, and  $N$  is equal to  $S$ . Substituting into Eq. 6 gives the following formula for rarefaction by species (Eq. 7).

$$E[PD]_m = \sum_j^T L_j \times \left[ 1 - \frac{\left( S - \sum x_{ij} \right)}{m} \right] \left( \frac{S}{m} \right) \quad (7)$$

## Extension

It has previously been recognised (Lande 1996; Olszewski 2004) that there is a relationship between individuals-based rarefaction curves and measures of evenness. Specifically, the initial slope of the individuals-based curve for species richness is equal to the PIE (Probability of Interspecific Encounter) index of Hurlbert (1971). The initial slope of the rarefaction curve is the difference between the expected species richness for two individuals ( $m=2$ ) and the expected species richness for one individual ( $m=1$ ), and is the probability that the second individual will be a different species from the first (Olszewski 2004). The PIE index is directly related to the Gini-Simpson index – the probability that two individuals selected at random will be different species. The difference between these two indices is in the form of random sampling – Gini-Simpson samples with replacement (thus assuming infinite population size) while PIE, just like rarefaction, samples without replacement. Following Olszewski (2004), PIE can be expressed as the following (Eq. 8) where  $E[S_1]$  and  $E[S_2]$  refer to the expected species richness of one and two randomly drawn individuals respectively. Note that  $E[S_1]$  always equals one in this case.

$$PIE = E[S_2] - E[S_1] \quad (8)$$

When considering a sample-based curve, it is clear that the initial slope is related to the beta-diversity of the set of samples from which the curve is calculated. In this case, the difference between  $E[S_1]$  and  $E[S_2]$  is the expected number of species in the second sample that are not found in the first. Thus, the PIE index can be used to measure beta-diversity if applied to sample-based rarefaction. This interpretation is directly related to the additive partitioning of species diversity into alpha and beta components where alpha-diversity is the mean (expected) richness of a single sample and beta-diversity is the gain in species richness from a single sample to a larger set of samples and can be read directly from a rarefaction curve (Crist and Veech 2006).

It follows that we can also define measures of phylogenetic evenness and phylogenetic beta-diversity using the initial slope of the PD rarefaction curve, where the units of accumulation are either individuals or samples respectively (Fig. 1). In either case, the initial slope is the expected gain in PD ( $\Delta PD$ ) when adding a second accumulation unit to the first. Further, because PD rarefaction curves can also meaningfully use species as accumulation units, we can extend this idea to include a measure of phylogenetic dispersion where the gain in PD is the expected branch length in the lineage (path from tip to root) of a second randomly selected species that is not shared with the first. Thus, we can define a general measure ( $\Delta PD$ ) for phylogenetic evenness, phylogenetic beta-diversity or phylogenetic dispersion, depending on the accumulation units chosen (Eq. 9, see also Fig. 1).  $\Delta PD$  is very similar to the  $\Delta PD_q$  measure of Faith (2013) although in that case, probabilities are not derived from the hypergeometric distribution. Further,  $\Delta PD_q$  is specifically applied to the problem of estimating loss of PD from extinction – a problem that is mathematically similar to rarefaction.

$$\Delta PD = E[PD_2] - E[PD_1] \quad (9)$$

If branch lengths are measured as millions of years between branching events, then  $\Delta PD$  is measured in units that make intuitive sense and allows for direct comparison across trees and systems. Alternatively, one could standardise the measure by dividing by its theoretical maximum.  $\Delta PD$  will be maximum when all individuals, species or samples represent wholly distinct lineages with no shared branch lengths. For an ultrametric tree, the lineage length (path from tip to root) is invariant across species and is equal to the depth of the tree. When rarefaction is by units of individuals or species,  $E[PD_1]$  is the lineage length. When rarefaction is by units of samples,  $E[PD_1]$  will equal the average PD of a sample and will be equal to  $\Delta PD$  in the extreme case where each sample shares no branch length with any other sample. Thus, whether referring to units of individuals, species or samples,  $E[PD_1]$  represents the theoretical maximum of  $\Delta PD$  and can be used to standardise the measure as follows.

$$\Delta PD_{standard} = \frac{\Delta PD}{\Delta PD_{max}} = \frac{E[PD_2] - E[PD_1]}{E[PD_1]} \quad (10)$$

## Application

The following is a demonstration of the application of PD rarefaction, and the derived  $\Delta PD$  statistics, to real ecological datasets. These applications are not intended to provide definitive answers to ecologically important questions but are, rather, simple demonstrations of how PD rarefaction can allow new analyses to be undertaken and, hopefully, new insights gained.

In all these applications, I have used published data on mammals. This is principally for convenience as mammals (Bininda-Emonds et al. 2007) and birds (Jetz et al. 2012) are the only major taxonomic groups for which comprehensive species-level supertrees are available. I have used an updated version of the mammal supertree of Bininda-Emonds et al. (2007) published as supplementary material by Fritz et al. (2009). In this supertree, all branch lengths are measured in units of time (millions of years between branching events), allowing for a straight-forward interpretation of PD as cumulative evolutionary history (Proches et al. 2006).

All analyses were conducted using the statistical software, *R* version 2.15.2 (R Core Team 2012). Phylogenetic information was processed using the *ape* package in *R* (Paradis et al. 2004). PD rarefaction analyses used the *phylodiv*, *phylocurve* and *phylorare* functions, written by the author and available from: <http://davidnipperess.blogspot.com.au>.

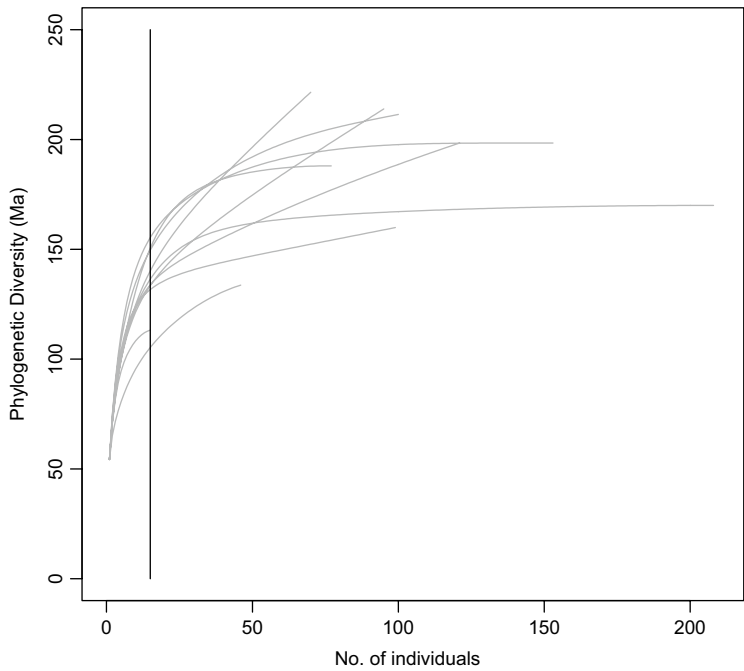
## ***Standardisation of Sampling***

The most commonly used application for rarefaction is standardisation to allow comparisons to be made between datasets with differing amounts of sampling effort. Standardisation can be achieved by rarefying all datasets back to a common (typically the minimum) number of accumulation units (Sanders 1968; Gotelli and Colwell 2001).

Law et al. (1998) surveyed bats in ten State Forests of the south-west slopes region of New South Wales, Australia. Survey methods were a combination of ultrasonic detectors, harp-traps, mist-nets and trip-lines. For the purposes of this demonstration, only data from the harp-traps will be used. A harp-trap is a rectangular frame, strung vertically with nylon line, placed so as to intercept the flight path of low-flying bats (Tidemann and Woodside 1978). A bat striking the nylon lines of the trap will tumble down into a collecting bag at the bottom.

Sampling effort among State Forests was variable with between 8 and 30 trap-nights. Comparison of bat diversity between State Forests is therefore confounded by variation in sampling effort, as can be seen when plotting separate PD rarefaction curves for each State Forest (Fig. 3). To correct for variation in trapping effort, expected PD for each State Forest was calculated for the common value of 15 individuals, which was the minimum number recovered from a State Forest (Fig. 3). While rarefying to eight trap-nights (samples) would also be an appropriate method of standardisation, data on the bat species caught per trap-night were not available in Law et al. (1998). Standardising for sample effort changed the rank order of the sites for Phylogenetic Diversity (Table 1). A test of the rank correlation between the standardised and non-standardised PD values was relatively high but non-significant (Spearman's correlation coefficient,  $\rho=0.57$ ,  $p=0.084$ ). Therefore, what one concludes about the relative bat diversity (and perhaps conservation importance) among these sites is dependent upon whether or not sampling effort is taken into account.





**Fig. 3** An example of standardisation of Phylogenetic Diversity (PD) by rarefaction. Data are abundances of bats caught in harp-traps in State Forests of the south-west slopes region of New South Wales, Australia. See Law et al. (1998) for a description of the data. Plotting separate individuals-based curves (grey lines) for each site shows considerable variation in sampling effort, with the raw value of PD being dependent on the number of trapped individuals. To allow for comparison between sites, PD is rarefied to an expected value for 15 individuals for all sites (indicated by black vertical line)

**Table 1** Comparison of diversity measures for bat assemblages for ten state forests of the south-west slopes region of New South Wales, Australia

State forest	Individuals (N)	Species richness (S)	Phylogenetic diversity ( $PD_N$ )	Standardised phylogenetic diversity ( $PD_{15}$ )
Bago	99	6	159	132
Maragle	208	8	170	136
Buccleuch	100	7	211	150
Bungongo	70	6	221	140
Woomargama	121	7	198	133
Carabost	153	8	198	155
Murragulldrie	95	6	214	133
Ellerslie	46	4	134	105
Tumblong	77	7	188	151
Minjary	15	3	113	113

Original data was taken from Law et al. (1998). Phylogenetic Diversity is measured in units of millions of years

*Phylogenetic Evenness*

The extension of PD rarefaction to  $\Delta$ PD allows for the measurement of phylogenetic evenness, which is essentially a measure of the distribution of individuals among branches in a phylogenetic tree (Webb and Pitman 2002). A phylogenetically even community is one where the most evolutionarily distinct species are also the most abundant. Because  $\Delta$ PD will increase with both increasing phylogenetic evenness and phylogenetic diversity, it is more correctly a measure of entropy (Jost 2006), directly comparable to the PIE and Gini-Simpson indices. It has a particularly close relationship with the quadratic entropy measure of Rao (1982). Rao’s quadratic entropy measures the average distance between individuals in an assemblage. When that distance is measured as patristic distance (path length on a phylogenetic tree),  $\Delta$ PD will be approximately half of Rao’s quadratic entropy.  $\Delta$ PD is also similar in intent, but not in form, to the phylogenetic entropy index of Allen et al. (2009).

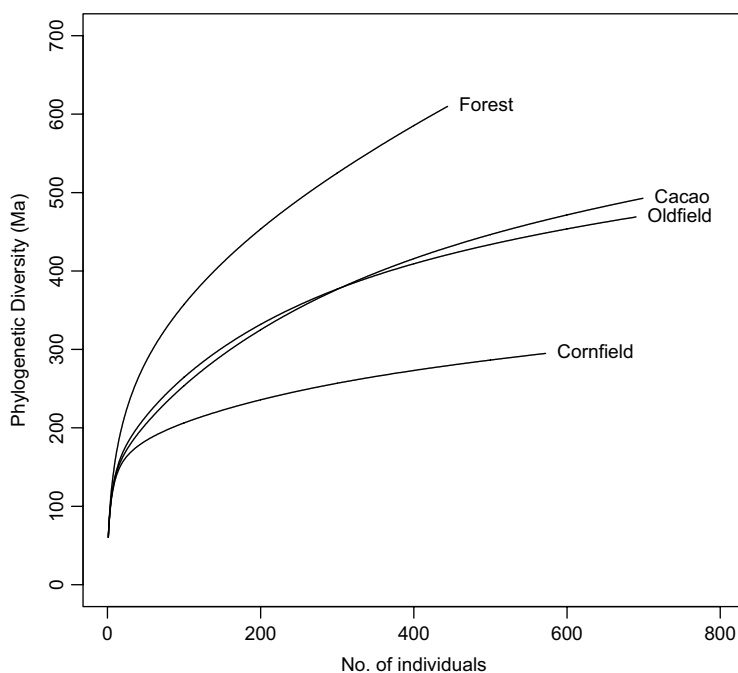
Low ecological evenness may be an indicator of disturbance where a small number of species are favoured. If those favoured species are also closely related, due to sharing a trait that allows exploitation of disturbance events, we can expect a reduction in phylogenetic evenness (Helmus et al. 2010). Medellin et al. (2000) surveyed the bat assemblages along a disturbance gradient in the Selva Lacandona, Chiapas, Mexico. The disturbance gradient consisted of four habitats, which, in order of disturbance, were cornfield, oldfield, cacao plantation and forest. Bats were sampled using mist nets and each habitat in the disturbance gradient was sampled using the same effort, thus making possible the comparison of habitats without the need for rarefaction. Medellin et al. (2000) found a trend of decreasing species richness and species evenness with increasing disturbance, and this trend is also reflected in the phylogenetic diversity and evenness of the assemblages (Table 2, Fig. 4).

The trend in phylogenetic evenness may simply be reflecting the abundance distribution among species. To determine the phylogenetic contribution to phylogenetic evenness,  $\Delta$ PD was divided by the PIE index (Table 2). Since PIE is the probability that the second randomly selected individual is a different species to the

**Table 2** Comparison of diversity measures for bat assemblages from four habitats along a disturbance gradient in the Selva Lacandona, Chiapas, Mexico

Habitat	Individuals	Species richness	PIE	Phylogenetic diversity	Phylogenetic evenness ( $\Delta$ PD)	Phylogenetic component ( $\Delta$ PD/PIE)
Cornfield	572	17	0.786	295	17.2	21.8
Oldfield	690	20	0.809	469	18.1	22.4
Cacao	699	21	0.851	493	18.2	21.3
Forest	444	27	0.884	609	20.4	23.0

Original data taken from Medellin et al. (2000). PIE refers to the Probability of Interspecific Encounter (Hurlbert 1971). Phylogenetic Diversity and phylogenetic evenness are measured in units of millions of years



**Fig. 4** Individuals-based PD rarefaction curves for bat assemblages from four habitats along a disturbance gradient in the Selva Lacandona, Chiapas, Mexico. (See Medellín et al. (2000) for a description of the data. Phylogenetic evenness ( $\Delta PD$ ) values are highest in the least disturbed habitat (Forest) and lowest in the most disturbed habitat (Cornfield)

first, we can divide  $\Delta PD$  by PIE to get the expected branch length of that species (conditional on the second individual being a different species). This value is related to phylogenetic dispersion ( $\Delta PD$  from a species-based rarefaction curve) but differs due to the conditional probability structure, and effectively measures the pure phylogenetic contribution to  $\Delta PD$  independent of the abundance distributions among species. We see, in this case, that the phylogenetic component generally decreases with increasing disturbance (Cacao being the exception), supporting the notion that disturbance favours more closely related species.

### *Phylogenetic Beta-Diversity*

Phylogenetic beta-diversity is effectively the turnover of branch lengths between samples in space and/or time. Like its species-level equivalent, phylogenetic beta-diversity can be measured on a pair-wise basis (Lozupone and Knight 2005; Bryant et al. 2008; Nipperess et al. 2010) or as a single value for a set of samples (Anderson et al. 2010). Rarefaction of PD provides a means for deriving a single value of

beta-diversity for a set of samples of any size via the  $\Delta$ PD measure, which is a phylogenetic analogue of the additive partitioning approach of Crist and Veech (2006).

Morton et al. (1994) compiled data on small mammal assemblages for 245 sites in arid Australia. I calculated beta-diversity for two regions from this dataset – Tanami desert and Uluru-Kata Tjuta National Park, Northern Territory. These regions had a similar number of sites (Table 3) covering a roughly similarly sized area but differed in the number of vegetation types. The Tanami sites were all spinifex grassland while the Uluru sites comprised a mix of spinifex grassland, acacia shrubland and woodland (Morton et al. 1994). It might be expected therefore that the Uluru sites will show higher beta-diversity due to the diversity of habitats represented. In addition to  $\Delta$ PD, I used the additive partitioning method to calculate species-level beta-diversity as the difference between total species richness of all sites in a region and the mean species richness of a single site (Lande 1996; Crist and Veech 2006).

Contrary to expectations, the Tanami desert sites showed greater species beta-diversity and phylogenetic beta-diversity despite the lack of variation in vegetation type (Table 3). This pattern is driven by the much higher site-level (alpha) species richness in Uluru-Kata Tjuta National Park (Table 3, Fig. 5) without a concomitant increase in overall (gamma) species richness, resulting in a high degree of species overlap. Given the overlap in species among Uluru sites, it appears that most small mammals are not specialised for particular vegetation types.

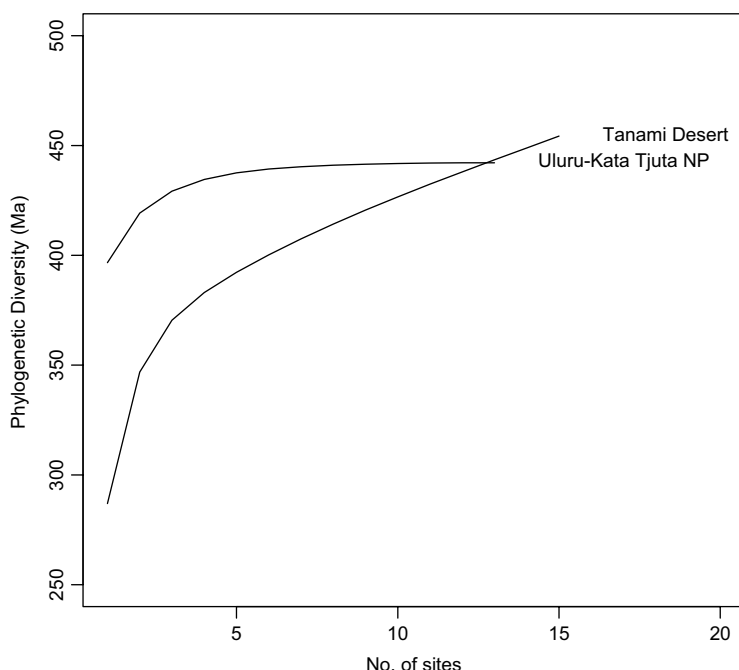
*Phylogenetic Dispersion*

Phylogenetic dispersion is a measure of the average phylogenetic distance among species (or tips) (Webb et al. 2002) and is in effect a measure of tree shape (Davies and Buckley 2012).  $\Delta$ PD provides a simple, intuitive measure of dispersion as the expected gain in PD of adding a second randomly selected species to the first. It can also be seen as a means of correcting for variation in species richness among samples, as it is well known that PD increases with species richness (Rodrigues and Gaston 2002).

**Table 3** Comparison of diversity measures for small mammal assemblages of sites in the Tanami Desert and Uluru-Kata Tjuta National Park, Northern Territory, Australia

Region	No. of sites	Species richness (alpha)	Species richness (gamma)	Species beta diversity (additive)	Phylogenetic beta diversity ( $\Delta$ PD)
Tanami	15	3.13	14	10.87	59.92
Uluru	13	6.54	13	6.46	22.54

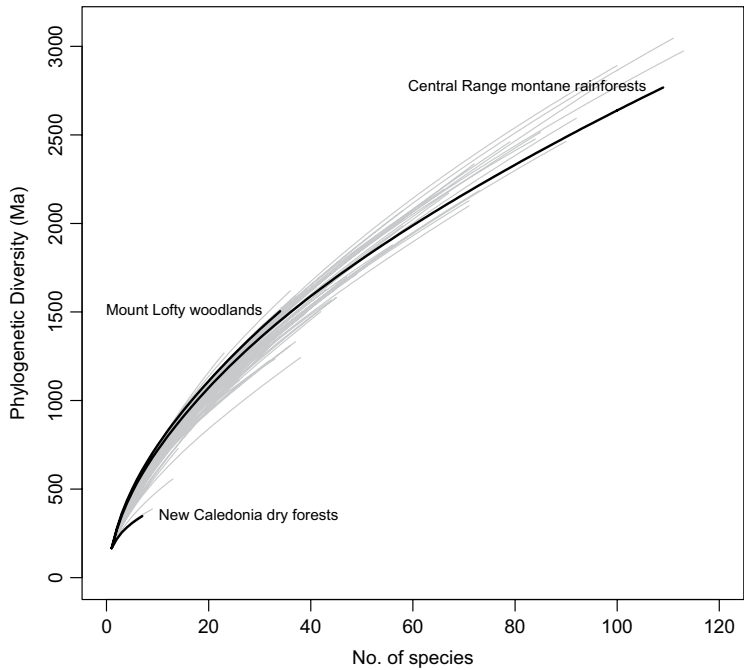
Species beta diversity is calculated as the difference between the total species richness of a region (gamma) and the mean site-level species richness (alpha)



**Fig. 5** Sample-based rarefaction curves for small mammal assemblages of sites in the Tanami Desert and Uluru-Kata Tjuta National Park, Northern Territory, Australia. See Morton et al. (1994) for a description of the data. Phylogenetic beta diversity ( $\Delta$ PD) is higher among the Tanami sites than the Uluru sites

I generated PD rarefaction curves and  $\Delta$ PD values for the mammal faunas of 71 of the 79 terrestrial ecoregions recognised by Olson et al. (2001) as constituting the Australasian biogeographic realm. Data were sourced from the wildfinder database (<http://worldwildlife.org/pages/wildfinder>) of the World Wildlife Fund. Eight ecoregions were excluded from the analysis because they had less than two species and thus a  $\Delta$ PD value could not be calculated.

The ecoregions show huge variation in species richness and, as expected, Phylogenetic Diversity is highly dependent on species richness (Fig. 6). Tropical ecoregions (such as the central range Montane rainforests, New Guinea) have high species richness and high Phylogenetic Diversity (Fig. 6, Table 4). When considering phylogenetic dispersion, however, other ecoregions show unusually high or low values given their species richness (Table 4). The ecoregion with the lowest  $\Delta$ PD is the New Caledonia dry forests. Because of its isolation, this fauna consists exclusively of bats and thus all the species are relatively closely related. The ecoregion with the highest  $\Delta$ PD was the Mount Lofty woodlands of South Australia, reflecting relatively high numbers of marsupial species compared to the more tropically distributed bats and rodents.



**Fig. 6** Species-based rarefaction curves for mammal assemblages of terrestrial ecoregions of the Australasian biogeographic realm Ecoregions are as defined by Olson et al. (2001). Data are sourced from the wildfinder database (<http://worldwildlife.org/pages/wildfinder>). Three ecoregions are highlighted, as having minimum (New Caledonia dry forests), maximum (Mount Lofty woodlands) or median (Central Range montane rainforests) values of phylogenetic dispersion ( $\Delta$ PD)

**Table 4** Comparison of diversity measures for mammal assemblages of selected ecoregions of the Australasian biogeographic realm

Ecoregion	Species richness	Phylogenetic diversity (Ma)	Phylogenetic dispersion ( $\Delta$ PD)
New Caledonia dry forests	7	347	51.3
Central range montane rainforests	109	2768	103.6
Mount Lofty woodlands	34	1504	110.7

### Future Directions

As demonstrated here, rarefaction of PD has a straightforward application in standardising PD across samples so that they can be compared directly. Further, depending on the accumulation unit, the rarefaction formula can be extended to the calculation of metrics of phylogenetic evenness, phylogenetic beta-diversity and phylogenetic dispersion. However, the application of the PD rarefaction formula

and its extension to other metrics is still very much in its infancy. Here I will outline some future directions for PD rarefaction.

Rarefaction by units of species allows for the comparison of locations while controlling for variation in species richness. This can easily be done by either rarefying all locations to a given number of species (Nipperess and Matsen 2013) or via  $\Delta$ PD as demonstrated here. This kind of correction has previously been done by including species richness as an explanatory variable in a statistical model and taking the residuals (Davies et al. 2008) or by comparison to a null model derived by repeated subsampling (Davies et al. 2007). The latter method is often used as a statistical test of phylogenetic dispersion (also known as phylogenetic structure) where random draws are taken from a species pool, representing a null community assembly process (Webb 2000). Such methods are no longer necessary as the exact relationship between species richness and PD is described by the rarefaction curve (Nipperess and Matsen 2013). Further, the exact analytical solution is computationally efficient, allowing for practical application to very large datasets.

By removing the effect of species richness, we can identify “evolutionary hotspots” with higher than expected phylogenetic diversity (Davies et al. 2008; Nipperess and Matsen 2013) on a regional or global scale. We can then use the standardised PD values (called relative PD by Davies et al. 2007) to explore the environmental, ecological and historical processes that lead to the observed patterns of high or low phylogenetic dispersion (Kooyman et al. 2013). Ultimately, we may be able to develop the theory to predict these patterns (Davies et al. 2007), in a similar vein to what has been done for species richness (Arrhenius 1921; MacArthur and Wilson 1963; Rosindell et al. 2011). For example, the relationship of species richness with area is well known but the phylogeny-area relationship has only recently begun to be explored (Morlon et al. 2011). Rarefaction curves have an obvious connection to species-area curves (Olszewski 2004) and thus the development of PD rarefaction may well improve understanding of the phylogeny-area relationship. In particular, species-based rarefaction of PD allows for the separation of species diversity effects from those purely explained by phylogeny.

It is possible to predict how much Phylogenetic Diversity is yet to be sampled from the observed rarefaction curve. Rarefaction is the basis of several species diversity estimators, which attempt to calculate total diversity (including unseen species) for a set of individuals or samples by effectively extending the curve beyond the observed sampling depth (Colwell and Coddington 1994). It follows that a useful extension of PD rarefaction would be a PD estimator that predicts unseen branch length, given the observed rate of accumulation of PD. It is important to note that PD rarefaction calculates the expected branch length gained by adding additional accumulation units but does not predict where on the tree these branches will come from. Similarly, a biodiversity estimator based on PD rarefaction may be able to predict the amount of PD not yet sampled but would not be able to predict where these unseen branches would be added to an existing tree. This would be, nevertheless, an exciting development.

It has recently been proposed that the standardisation of samples for species diversity should not be done by rarefaction to the same size (i.e. no. of individuals),

but rather by sample completeness (Alroy 2010; Jost 2010; Chao and Jost 2012). Completeness, when measured by a statistic known as coverage (Good 1953), is the proportion of individuals in a community that are represented by species in a sample from that community (Chao and Jost 2012). When samples differ in their coverage, they should be standardised to equal coverage before a “fair” comparison can be made. Much like expected species richness, the coverage of a sample can be estimated from the sample size and the distribution of individuals among the species in the sample (Chao and Jost 2012). Given that standardisation by sample completeness has been shown to yield a less biased comparison of species richness between communities (Chao and Jost 2012), it would be desirable to have a similar method of standardisation for PD. Since rarefaction of coverage is mathematically related to rarefaction of sample size, the recent work on estimating PD from sample size will no doubt form the basis from which estimated PD for sample coverage will be developed.

Finally, a general issue when considering any PD measure is uncertainty regarding the length of branches and the topology (branching pattern) of the tree. All PD measures (including those presented here) assume that the branch lengths and their arrangement in the tree are perfectly known. This is obviously an abstraction, although PD can be surprisingly robust to this source of variation (Swenson 2009). One solution to this dilemma is to calculate PD, including rarefied PD, for a large number of possible trees and report the mean and confidence limits. The output from a Bayesian phylogenetic analysis is a large number of trees, each with their own topology and corresponding branch lengths (see for example Jetz et al. 2012) and so lends itself well to this approach. However, when the possible trees number in the thousands and tens of thousands, this is obviously computationally intensive. An analytical solution, directly incorporating uncertainty into the calculation, would therefore be desirable. This is not an easy extension of the PD rarefaction solution because both variation in branch length and topology (affecting the probability of encountering internal branches) would need to be taken into account. It is worth remembering that phylogenetic relationships are not the only source of uncertainty when investigating real ecological communities – neither the abundance, nor even the presence (occupancy), of species are necessarily known with precision.

## Conclusion

The formulation for the rarefaction of Phylogenetic Diversity (PD) is given in expanded form to show its simplicity and its connection to the classic formula for the rarefaction of species richness (Hurlbert 1971; Simberloff 1972). The method is exact and efficient and should be preferred over the algorithmic (Monte Carlo) solution involving repeated random sub-sampling. Further, the extension to the calculation of  $\Delta$ PD provides a flexible and general framework for the measurement of biodiversity as phylogenetic evenness, phylogenetic beta-diversity or phylogenetic dispersion. The applications of PD rarefaction and  $\Delta$ PD presented here are



hopefully useful in improving understanding of the importance of rarefaction in ecology and in guiding future applications of the method. There are, I believe, exciting prospects for PD rarefaction in the future, including as a general method for standardising PD by removing variation with species richness, and for predicting unseen (i.e. un-sampled) PD. The recent availability of comprehensive phylogenies (Bininda-Emonds et al. 2007; Jetz et al. 2012) and rich data on species occurrences (Flemons et al. 2007), coupled with analytical advances such as PD rarefaction, allows us to better understand the distribution of Phylogenetic Diversity on the surface of the Earth and the processes giving rise to that distribution. This is valuable for its own sake but will also inform efforts to conserve as much of the Tree of Life as possible in the face of future extinctions (Rosauer and Mooers 2013).

**Acknowledgements** I would like to thank Frederick Matsen (Fred Hutchinson Cancer Research Center), Daniel Faith (Australian Museum), Peter Wilson (Macquarie University), Anne Chao (National Tsing Hua University) and Robert Colwell (University of Connecticut) for their input regarding the rarefaction of PD and what it all might mean. Daniel Miranda-Esquivel (Universidad Industrial de Santander) and Pedro Cardoso (University of Helsinki) provided helpful feedback on an earlier version of this chapter. This research was supported by the Australian Research Council (DP0665761 and DP1095200).

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Allen B, Kon M, Bar-Yam Y (2009) A new phylogenetic diversity measure generalizing the Shannon index and its application to phyllostomid bats. *Am Nat* 174:236–243
- Alroy J (2010) The shifting balance of diversity among major marine animal groups. *Science* 329:1191–1194
- Anderson MJ, Crist TO, Chase JM et al (2010) Navigating the multiple meanings of  $\beta$  diversity: a roadmap for the practicing ecologist. *Ecol Lett* 14:19–28
- Arrhenius O (1921) Species and area. *J Ecol* 9:95–99
- Bininda-Emonds ORP, Cardillo M, Jones KE et al (2007) The delayed rise of present-day mammals. *Nature* 446:507–512
- Bryant JA, Lamanna C, Morlon H et al (2008) Microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. *Proc Natl Acad Sci* 105:11505–11511
- Chao A, Jost L (2012) Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology* 93:2533–2547
- Chao A, Chiu CH, Jost L (2010) Phylogenetic diversity measures based on Hill numbers. *Phil Trans R Soc B* 365:3599–3609
- Chiarucci A, Bacaro G, Rocchini D, Fattorini L (2008) Discovering and rediscovering the sample-based rarefaction formula in the ecological literature. *Community Ecol* 9:121–123
- Colwell RK, Coddington J (1994) Estimating terrestrial biodiversity through extrapolation. *Phil Trans R Soc B* 345:101–118
- Crist TO, Veech JA (2006) Additive partitioning of rarefaction curves and species-area relationships: unifying alpha, beta and gamma-diversity with sample size and habitat area. *Ecol Lett* 9:923–932

- Davies TJ, Buckley LB (2012) Exploring the phylogenetic history of mammal species richness. *Glob Ecol Biogeogr* 21:1096–1105
- Davies RG, Orme CDL, Webster AJ et al (2007) Environmental predictors of global parrot (Aves: Psittaciformes) species richness and phylogenetic diversity. *Glob Ecol Biogeogr* 16:220–233
- Davies TJ, Fritz SA, Grenyer R et al (2008) Phylogenetic trees and the future of mammalian biodiversity. *Proc Natl Acad Sci* 105(Suppl 1):11556–11563
- Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61:1–10
- Faith DP (2013) Biodiversity and evolutionary history: useful extensions of the PD phylogenetic diversity assessment framework. *Ann N Y Acad Sci* 1289:69–89
- Faith DP, Reid CAM, Hunter J (2004) Integrating phylogenetic diversity, complementarity, and endemism for conservation assessment. *Conserv Biol* 18:255–261
- Faith DP, Lozupone CA, Nipperess DA, Knight R (2009) The cladistic basis for the phylogenetic diversity (PD) measure links evolutionary features to environmental gradients and supports broad applications of microbial ecology's 'Phylogenetic Beta Diversity' framework. *Int J Mol Sci* 10:4723–4741
- Ferrier S, Manion G, Elith J, Richardson K (2007) Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Divers Distrib* 13:252–264
- Flemons P, Guralnick R, Krieger J et al (2007) A web-based GIS tool for exploring the world's biodiversity: The Global Biodiversity Information Facility Mapping and Analysis Portal Application (GBIF-MAPA). *Ecol Informa* 2:49–60
- Fritz SA, Bininda-Emonds ORP, Purvis A (2009) Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics. *Ecol Lett* 12:538–549
- Good IJ (1953) The population frequencies of species and the estimation of population parameters. *Biometrika* 40:237–264
- Gotelli NJ, Colwell RK (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol Lett* 4:379–391
- Helmus MR, Keller W, Paterson MJ et al (2010) Communities contain closely related species during ecosystem disturbance. *Ecol Lett* 13:162–174
- Hurlbert S (1971) The nonconcept of species diversity: a critique and alternative parameters. *Ecology* 52:577–586
- Jetz W, Thomas GH, Joy JB et al (2012) The global diversity of birds in space and time. *Nature* 491:444–448
- Jost L (2006) Entropy and diversity. *Oikos* 113:363–375
- Jost L (2010) The relation between evenness and diversity. *Diversity* 2:207–232
- Kobayashi S (1974) The species-area relation I. A model for discrete sampling. *Res Popul Ecol* 15:223–237
- Kooyman RM, Rossetto M, Sauquet H, Laffan SW (2013) Landscape patterns in rainforest phylogenetic signal: isolated islands of refugia or structured continental distributions? *PLoS One* 8:e80685
- Lande R (1996) Statistics and partitioning of species diversity, and similarity among multiple communities. *Oikos* 76:5–13
- Law B, Anderson J, Chidel M (1998) A bat survey in State Forests on the south-west slopes region of New South Wales with suggestions of improvements for future surveys. *Aust Zool* 30:467–479
- Lozupone CA, Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* 71:8228–8235
- Lozupone CA, Knight R (2008) Species divergence and the measurement of microbial diversity. *FEMS Microbiol Rev* 32:557–578
- MacArthur RH, Wilson EO (1963) An equilibrium theory of insular zoogeography. *Evolution* 17:373–387
- Mao C, Colwell RK, Chang J (2005) Estimating the species accumulation curve using mixtures. *Biometrics* 61:433–441

- Medellin RA, Equihua M, Amin M (2000) Bat diversity and abundance as indicators of disturbance in neotropical rainforests. *Conserv Biol* 14:1666–1675
- Morlon H, Schwik DW, Bryant JA et al (2011) Spatial patterns of phylogenetic diversity. *Ecol Lett* 14:141–149
- Morton SR, Brown JH, Kelt DA, Reid JRW (1994) Comparisons of community structure among small mammals of North American and Australian deserts. *Aust J Zool* 42:501–525
- Nipperess DA, Matsen FA IV (2013) The mean and variance of phylogenetic diversity under rarefaction. *Methods Ecol Evol* 4:566–572
- Nipperess DA, Faith DP, Barton K (2010) Resemblance in phylogenetic diversity among ecological assemblages. *J Veg Sci* 21:809–820
- O'Dwyer JP, Kembel SW, Green JL (2012) Phylogenetic diversity theory sheds light on the structure of microbial communities. *PLoS Comput Biol* 8:e1002832
- Olson D, Dinerstein E, Wikramanayake E et al (2001) Terrestrial ecoregions of the world: a new map of life on Earth. *Bioscience* 51:933–938
- Olszewski TD (2004) A unified mathematical framework for the measurement of richness and evenness within and among multiple communities. *Oikos* 104:377–387
- Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290
- Pardi F, Goldman N (2007) Resource-aware taxon selection for maximizing phylogenetic diversity. *Syst Biol* 56:431–444
- Petchey OL, Gaston KJ (2002) Functional diversity(FD), species richness and community composition. *Ecol Lett* 5:402–411
- Proches S, Wilson J, Cowling R (2006) How much evolutionary history in a 10 x 10 plot? *Proc R Soc B* 273:1143–1148
- R Core Team (2012) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>
- Rao CR (1982) Diversity and dissimilarity coefficients: a unified approach. *Theor Popul Biol* 21:24–43
- Rodrigues ASL, Gaston KJ (2002) Maximising phylogenetic diversity in the selection of networks of conservation areas. *Biol Conserv* 105:103–111
- Rosauer DF, Mooers AO (2013) Nurturing the use of evolutionary diversity in nature conservation. *Trends Ecol Evol* 28:322–323
- Rosauer D, Laffan SW, Crisp MD et al (2009) Phylogenetic endemism: a new approach for identifying geographical concentrations of evolutionary history. *Mol Ecol* 18:4061–4072
- Rosindell J, Hubbell SP, Etienne RS (2011) The unified neutral theory of biodiversity and biogeography at age ten. *Trends Ecol Evol* 26:340–348
- Sanders HL (1968) Marine benthic diversity: a comparative study. *Am Nat* 102:243–282
- Simberloff D (1972) Properties of the rarefaction diversity measurement. *Am Nat* 106:414–418
- Swenson NG (2009) Phylogenetic resolution and quantifying the phylogenetic diversity and dispersion of communities. *PLoS One* 4:e4390
- Tidemann CR, Woodside DP (1978) A collapsible bat-trap and a comparison of results obtained with the trap and with mist-nets. *Wildl Res* 5:355–362
- Turnbaugh PJ, Hamady M, Yatsunenko T et al (2009) A core gut microbiome in obese and lean twins. *Nature* 457:480–484
- Ugland K, Gray JS, Ellingsen KE (2003) The species-accumulation curve and estimation of species richness. *J Anim Ecol* 72:888–897
- Webb CO (2000) Exploring the phylogenetic structure of ecological communities: an example for rain forest trees. *Am Nat* 156:145–155
- Webb CO, Pitman N (2002) Phylogenetic balance and ecological evenness. *Syst Biol* 51:898–907
- Webb CO, Ackerly DD, McPeck M, Donoghue MJ (2002) Phylogenies and community ecology. *Ann Rev Ecol Syst* 33:475–505
- Yu DW, Ji Y, Emerson BC, Wang X (2012) Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods Ecol Evol* 3:613–623

# Support in Area Prioritization Using Phylogenetic Information

Daniel Rafael Miranda-Esquivel

**Abstract** Human activities have accelerated the level of global biodiversity loss. As we cannot preserve all species and areas, we must prioritize what to protect. Therefore, one of the most urgent goals and crucial tasks in conservation biology is to prioritize areas. We could start by calculating ecological measures as richness or endemism, but they do not reflect the evolutionary diversity and distinctness of the species in a given area. The conservation of biodiversity must be linked to the understanding of the history of the taxa and the areas, and phylogeny give us the core for such understanding. In such phylogenetic context, evolutionary distinctiveness (ED) is a feasible way for defining a ranking of areas that takes into account the evolutionary history of each taxon that inhabits the area. As our knowledge of the distribution or the phylogeny might be incomplete, I introduce Jack-knife re-sampling in evolutionary distinctiveness prioritization analysis, as a way to evaluate the support of the ranking of the areas to modifications in the data used. In this way, some questions could be evaluated quantitatively as we could measure the confidence of the results, since deleting at random part of the information (phylogenies and/or distributions), would help to quantify the persistence of a given area in the ranking.

**Keywords** Phylogenetic conservation • Taxonomic distinctiveness • Jack-knife

## Conservation Planning

The biodiversity is at risk, therefore decisions must be made in order to tackle the biodiversity crisis. In the process of conservation planning, one or maybe the most important task is to evaluate the quality and importance of a given area. To fulfill this task there are many metrics, from species richness to endemism, but these two values do not consider the evolutionary uniqueness of a species (Purvis and Hector

---

D.R. Miranda-Esquivel (✉)

Escuela de Biología, Universidad Industrial de Santander, Santander, Colombia

e-mail: [dmiranda@uis.edu.co](mailto:dmiranda@uis.edu.co)

2000). Any useful metric must include the evolutionary value of the species (Rolland et al. 2012), where the most important and therefore the selected area is the one that harbors the highest biodiversity, but this does not mean the highest number of species but the highest number of unique species or evolutionary fronts.

There are many approaches in the context of phylogenetic diversity and conservation, from community ecology to taxon or area conservation. Given this broad spectrum, the questions are different and vary a lot. In the context of community ecology and phylogeny, the approach is to evaluate whether there is structure in the community given the phylogeny (Cavender-Bares et al. 2009), and therefore the null model approach is used to present the null hypothesis. The species by area matrix is shuffled (see: Gotelli and Graves 1996), or the species or area labels are shuffled. Here the “support” is closer to the traditional confidence limits and error evaluation.

To evaluate the diversity of an area using phylogenies as a general frame, two main perspectives could be used, evolutionary distinctiveness (ED) or phylogenetic diversity (PD). Evolutionary distinctiveness refers to species-specific measures developed to assign scores to the species and therefore the areas they inhabit (Vane-Wright et al. 1991). The measures are topology-based indices, calculated as “the sum of basic taxic weights,  $Q$ , and the sum of standardised taxic weights,  $W$ ” (Schweiger et al. 2008), and therefore are also known as Taxonomic distinctiveness indices. Phylogenetic diversity (PD) is a distance-based index using minimum spanning path of the subset in the tree (Faith 1992). Redding et al. (2008) identified some of the major differences between ED and PD. PD is effective only if all the species within the optimal subset are protected, otherwise other optimal subsets are possible; unlike ED, PD is not species-specific and thus does not offer priority species rankings, which are important to species conservation approaches as the IUCN Red List of Threatened Species. Furthermore, topologies are more stable than branch lengths. Increasing the number of characters or changing the set of characters seldom leads to entire shifts in the relationships among species, whereas branch lengths change considerably from one set of characters to another and permit only to state about the evolution of the data set that generated the topology and the branch lengths (Brown et al. 2010).

## Indexes Used

I present the general protocol to evaluate species or areas in a phylogenetic context in Fig. 1. The different indices for each species are calculated to obtain the species phylogenetic values, while the sum of the indices of all species in a given area produces the areal phylogenetic values.

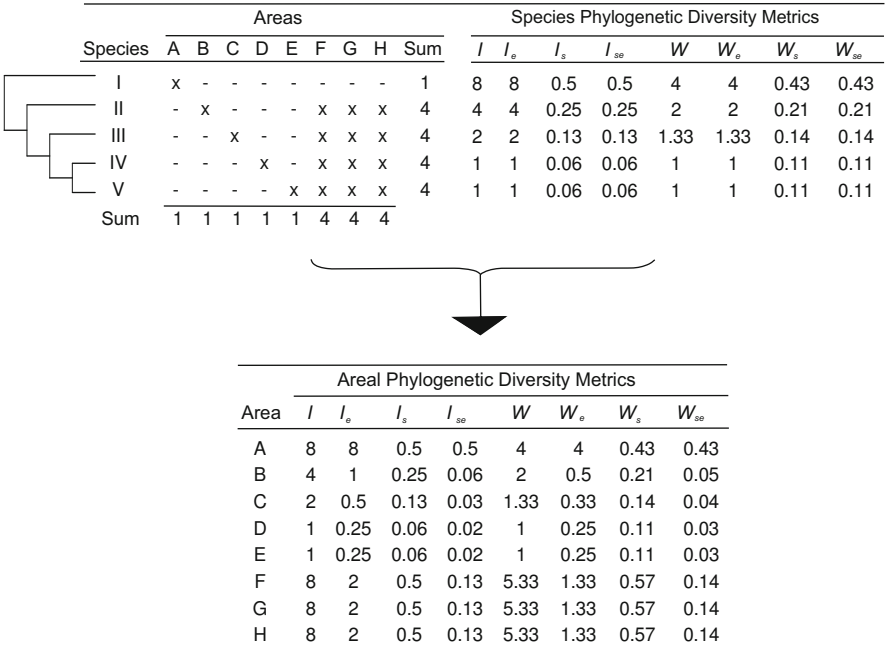
I used the traditional  $I$  &  $W$  indices created by Vane-Wright et al. (1991), along with the modifications introduced by Posadas et al. (2001) to consider endemism and widespread species ( $I_e/W_e$ ), the size of the topology ( $I_s/W_s$ ) or both variables at the same time ( $I_{es}/W_{es}$ ). The standardization of the indices  $I$  and  $W$  enables the

comparison of topologies with different number of species. In a topology with three species (I (II III)), distributed as taxon I in area A, II in B, and III in C. The taxon I and therefore the area A will have a value of 2.0 for indices  $I$  and  $W$ , while in a five species topology (I (II (III (IV V))))), the taxon I will have an index value of 8.0 for index  $I$  and 4.0 for index  $W$ , while the standardized  $I_s$  for this taxon and the area it inhabits will be 0.5 for both topologies.

If we consider the distributional pattern of the species, it could be endemic or widespread. We could apply the same index value to all areas where the species is present, but areas inhabited by widespread species will be selected, as we will sum the index values for each taxon, while an area inhabited only by an endemic taxon will be valued just for the single taxon it contains.

In a five taxa topology (Fig. 1), with four widespread species in the areas F, G, and H. If we use index  $I$  these three areas are as important as the area A, while using  $W$  index they are more important than the area A, as each area obtains the final index value because of the sum of all species inhabiting the area. Areas F, G and H are selected not because they are inhabited by unique species as area A but by widespread species. Using  $I_e/W_e$  or  $I_{es}/W_{es}$  the most important area is A, as it contains an evolutionary unique species, which is not found elsewhere.

Given the plethora of indices to choose, Winter et al. (2013) presented an important question: “We also call for a comprehensive guideline through the jungle of



**Fig. 1** An example for determining phylogenetic diversity metrics at species and area levels for a hypothetical topology with five species (four widespread), distributed in eight areas (Modified from Lehman (2006))

available phylogenetic diversity indices, with particular respect to the needs of conservationists – which index helps to protect what?”. Part of the answer to this question is given by the support to the decisions made, but in species or areas prioritization the literature does not present any kind of support measure (Whiting et al. 2000; Posadas et al. 2001; Pérez-Losada et al. 2002; López-Osorio and Miranda-Esquivel 2010; Prado et al. 2010), neither the most recent revisions cite any measure to evaluate the stability, confidence or support to the results (Schweiger et al. 2008; Vellend et al. 2011).

## Jack-Knife

In a jack-knife analysis, given a sample of observations and a parameter to evaluate, a subsample is made by eliminating a proportion of the original data and the parameter is calculated for the subsample. This procedure is repeated  $n$  times and summarized. Since the introduction of the jack-knife (Quenouille 1949), researchers have used it, to define limits of confidence in many sorts of analyses, from statistics (Efron 1979; Smith and van Belle 1984) and ecology (Crowley 1992) to phylogeny. It has been used not only as a measure of support (Lanyon 1987), but as a way to obtain the best solution for large data sets (Farris et al. 1996), to test competing hypotheses (Miller 2003), to generalize the performance of predictive models or for cross-validation to estimate the bias of a estimator. As the bootstrapping, it could be seen as “a measure of robustness of the estimator with regard to small changes in the data” (Holmes 2003).

I use this re-sampling approach to evaluate the support of the area ranking in the context of conservation and phylogeny. Therefore, some questions could be evaluated quantitatively.

## Jack-Knife in Conservation

The use of a meta-criterion to define an optimal parameter value has been used widely in phylogenetic analysis, i.e. the incongruence length difference test to define the ts/tv/gap costs (Wheeler 1995) or jack-knife frequencies to evaluate whether concavity parsimony outperforms linear parsimony (Goloboff et al. 2008).

In conservation biology, there must be a measure of the confidence and robustness of the results. A sensitivity analysis, deleting at random part of the information, helps to understand the support of the data as the persistence of a given area in the ranking. Therefore, jack-knife is the appropriate tool to explore the behavior of the results to perturbations in the data set (Holmes 2003).

In a conservation phylogenetic based analysis, there are three different items to evaluate, as we have three input parameters: the topology, the species in a given topology, and the distribution of a species.

The first question arises when we ask about the distributional pattern of the species: what if a locality (therefore all or some species in that area) is not included in the analysis? A species could not be included in a given locality for three reasons, because (1) it was never present there; (2) it is locally extinct; or (3) it was not sampled, although the species is present in the area. To evaluate such situation, the species can be deleted from a number areas to quantify the effect of missing information.

The second question arises when a species included in the phylogenetic analysis is not considered in the conservation analysis: what if a species is not included? A species not included in the analysis will affect the index value as this depends on the species included on the calculation. In this context, the presence of a species is deleted from all the areas it inhabits.

The third question arises when we do not include a given phylogeny: what if a phylogeny is not included? The whole topology might not be available for the conservation analysis. We could depend on a limited subset of phylogenies to the ranking of an specific area. Here, the topology, therefore the species and their distributions are deleted.

Given the three questions we can decide whether a phylogeny, a taxon or an area is deleted, with different probability values:

- j.topol is the probability to choose a topology ( $= p$ )
- j.tip is the probability to choose a species ( $= q$ )
- j.area is the probability to choose an area ( $= r$ )

In the first scenario, an area is deleted from the distribution of a species with a probability of  $p \times q \times r$  ( $0 < p, q, r < 1$ ), that is, the probability to select the topology and then select the species and then select the area. An area could be removed from the whole analysis, and this has to be run only the number of areas times, eliminating a single area each time. It would show the position of the area in the ranking of the areas and is equal to delete the area from the final results.

In the second scenario, a species is deleted from a single topology with a probability  $p \times q$  ( $0 < p, q < 1, r = 1.0$ ), therefore all areas inhabited by this species will not be included.

In the third scenario, the whole topology is not included in the analysis with a probability  $p$  ( $0 < p < 1, q = r = 1.0$ ), all the species and areas, belonging to that topology, will not be included in the analysis.

The first decision in the three scenarios, is made on the topology. As the number of topologies NOT included increases with the value of  $p$ , the absolute indices values would be small and inversely proportional to the value of  $p$ .

Those areas prioritized because of its position in a single or just a few topologies would change, the indices values would be lower, and the position of the area in the ranking might change. If an area is supported by all or most of the topologies, its position in the ranking must be stable, although the index value would be small in all the replicates, therefore the index values *per se* are meaningless, but the ranking is informative.



There is a fourth question, not considered here, related to the length of the branch. This question is valid in the context of Phylogenetic Diversity [PD] (Faith 1992), Genetic Diversity [GD] (Crozier 1992), or total lineage divergence (Scheiner 2012) [a metric similar to PD]. These methods require the precise estimation of the length, therefore the accuracy of the index value depends heavily on the length estimation.

Although Krajewski (1994) considers that the debate of the use and calculations of divergence in systematics and conservation are two topics, I consider that the same criticisms to the accuracy estimation of the length in systematics will have a profound impact in the decision made when the topology and its branch lengths are used in conservation. And as this quotation from Brown et al. (2010) states, “in any phylogenetic analysis, the biological plausibility of branch-length output must be carefully considered”. Therefore, we must be well aware of the methodological approach used to construct the phylogeny (Rannala et al. 2012).

Additionally, in some cases we must consider the sensitivity of PD value to intra-specific variation (Albert et al. 2012). Therefore, we must take into account the source of the tree (species vs. gene trees) [see for example Spinks and Shaffer (2009)].

## Optimal Scenario

Given a data set and  $n$  random perturbations on this data, if the index is robust, all (or most) perturbations would yield the same general ranking. Therefore, in the context of conservation in an optimal situation, we would prefer areas that:

1. *Have the same position in the ranking (original and re-sampled), no matter if we delete areas, species, or phylogenies*  
= same ranking or position, insensitive to changes in the item(s) deleted.
  2. if not, at least must be the same position in the ranking but considering just a subgroup (e.g. be first or second, or first to third).
  3. *Have the same position in the ranking (original and re-sampled), no matter the delete probability used (from 0.01 to 0.5).*  
= same ranking or position, insensitive to changes in the delete probability.
  4. or, have the same position for most of the probabilities used, but not counting extreme situations as a delete probability of 0.5.
- = not too sensitive to the probability values used.

In a real world, an scenario to meet the requirements of the first and third conditions is too strict and maybe impossible to fulfill. Therefore, my decision rules to select the best index and the best ranking are based in the second and fourth situations. The area must have the same position in the ranking considering just a subgroup, from the first to the third position in the ranking, no matter the type of item deleted, and for most of the probability values.

An alternative measure is to evaluate the behavior of an index and its success as the number of times that a replicate recovers part of the original ranking (e.g.

1st/2nd/3rd), but in any order. The researcher could consider only the first position in the ranking and evaluates the persistence of this area, or could consider the whole ordered ranking. These measures could be too strict and will be sensitive to the smallest perturbation to the data set, while the first to third position would be enough in terms of conservation planning.

Given any measure of success, the re-sampling approach in conservation have some possible applications as:

1. Which is the best index? that will answer also, what do we want to conserve/use to prioritize?

The best index would be defined as the most supported index, while the area used would be that found for most of the probabilities used.

2. How stable is the ranking (e.g. 1st/2nd/3rd position)?

This is a variation of the previous question, but focused in the ranking, as we prefer a supported ranking, we might evaluate the support for the original ranking.

## Proposed Protocol

Following the expected behavior in an optimal condition, first I evaluated the index. I considered the best index as the one that recovered most times the same original ranking -first to third areas-, as an ordered ranking. Then, using the selected index, I evaluated the best area, as the one found most often in the first place.

I tested six scenarios by modifying  $j.topol$  and  $j.tip$  values as follows:  $j.topol$  values of 0.50 and 0.32, and  $j.tip$  values of 1, 0.50 and 0.32. These values are just used to introduce the concept, but they are similar to strong, mild and relaxed tests. A value of 1 to delete a species means that all areas for that species will be deleted, while a value of 0.32 means that one out of three will be deleted. Smaller values as 0.01 are discarded, it would make no difference, as the perturbation to the data would be unimportant.

The effect of deleting areas is related to the number of areas inhabited. If the species is in an endemic area, the effect of deleting an area would be as deleting the whole species, while in a widespread species, the effect should be minimal with indices as  $I_e/W_e$  or  $I_{es}/W_{es}$ , but we can not define which is the best index as the four indices have similar properties. In all cases the probability of deleting areas was 1, therefore I tested the effect of the topology and species but not the effect of the distribution.

Number of Replicates

Hedges (1992) presented the number 1825 as the number of replicates needed to obtain an accuracy of  $\pm 1\%$  for a bootstrapping proportion of 95 %. Although the higher the number of replicates the higher the accuracy of the estimation of the bootstrap or jack-knife value, Pattengale et al. (2010) introduced a stopping criteria that yield lower figures as 500 replicates to get robust bootstrapping values for a 2500 taxa analysis. I randomized each scenario 10,000 times, that could be considered intuitively an appropriate number of replicates to estimate the jack-knife proportion for conservation purposes.

For these analyses, I used a modified version of the program *Richness* (Posadas et al. 2001) to randomize the data and to perform the index calculations [Jrich: available from <https://github.com/Dmirandae/jrich>], while the data analyses were performed using the software R (R Core Team 2013) and the figures were prepared using the library ggplot2 (Wickham 2009).

Empirical Examples

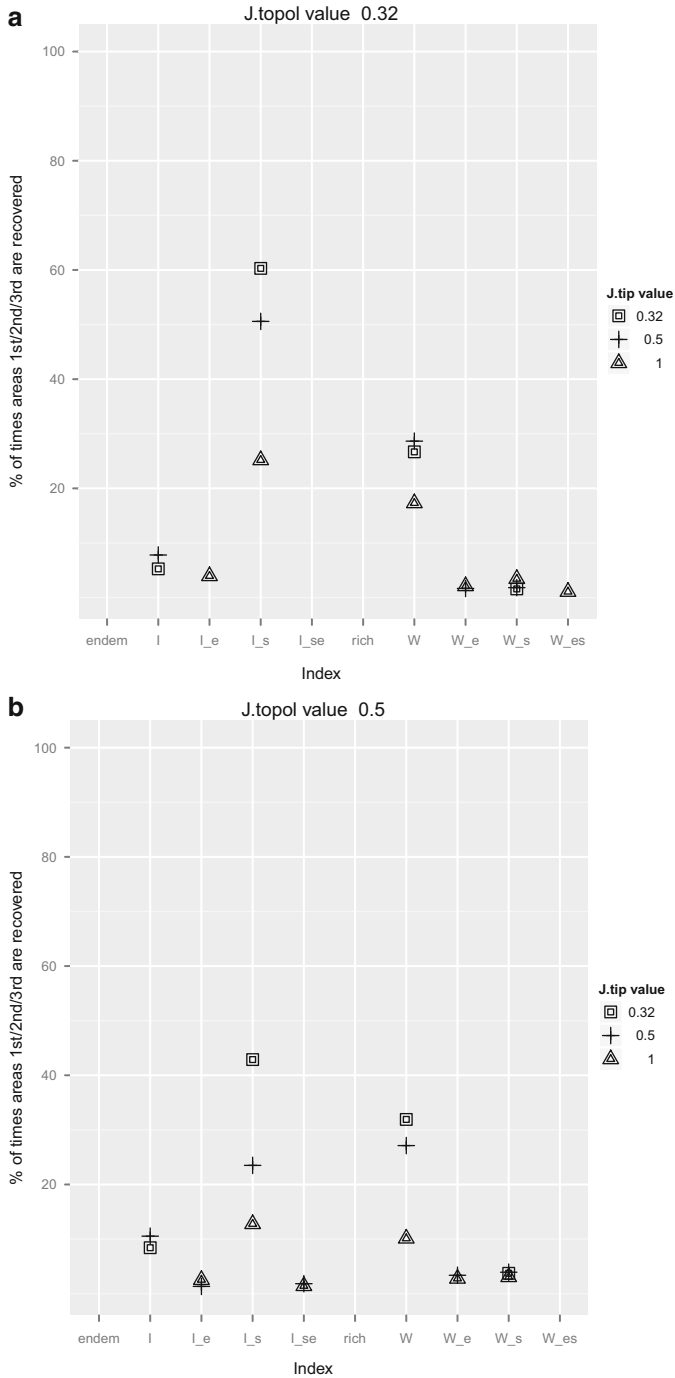
*First Case: The Original Ranking Does Not Mean Support*

Posadas et al. (2001) evaluated the conservation ranking in southern South America areas. They found that depending on the index used, the selected area changed, as the best area could be: Santiago (D), Ñuble (F), Valdivia (H), or the Malvinas islands (K). Also, for a single index, the values could be misleading, as the differences between the *W* index values are quite small, and the ranking could be an artifact rather than a real result (Table 1). I reanalyzed their dataset and found that the best index for this analysis is *I<sub>s</sub>* (Fig. 2) as this index that has the highest jack-knife frequency.

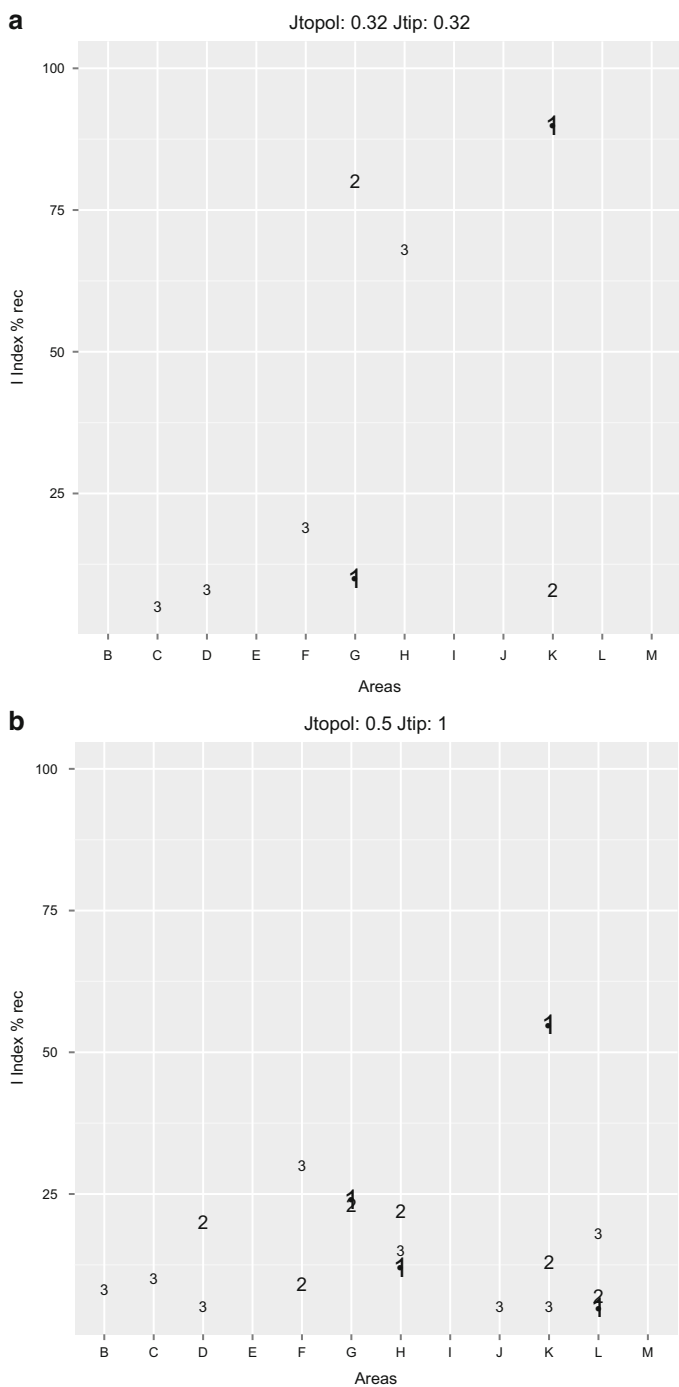
The most stable area using *I<sub>s</sub>* or raw *W* (the second best index), was the Malvinas islands, a candidate to be the best area (Fig. 3). The high uncertainty in the area chosen is eliminated when the support is included in the selection of the best area. Santiago has the highest number of species and harbors the highest number of endemic species, but it was not placed as the highest priority, while Malvinas island, the second most endemic area has the highest priority. The inferences based on the

**Table 1** First area in the ranking proposed by Posadas et al. (2001). For raw *W* the index values are 52.62/52.58/52.05. Labels follow Posadas et al. (2001)

	<i>I</i>	<i>W</i>
<i>raw</i>	Malvinas islands (K)	Valdivia (H) or Santiago (D) or Ñuble (F)
<i>e</i>	Malvinas islands (K)	Malvinas islands (K)
<i>s</i>	Santiago (D)	Ñuble (F)
<i>es</i>	Malvinas islands (K)	Ñuble (F)



**Fig. 2** For each index, the number of hits with a delete value of 0.32 or 0.5 for j.topol and three j.tip values of 0.32, 0.5, and 1 (in the last situation, the whole topology is deleted). Species endemism and richness (number of species) are included for comparative purposes (Data from Posadas et al. (2001)). (a) Number of hits with a j.topol value of 0.32 (b) Number of hits with a j.topol value of 0.5



**Fig. 3** Number of times an area is recovered as the first (number 1), second (number 2), or third area (number 3). These are the lowest (**a**) and the highest (**b**) delete probabilities used in this analysis. Acronyms for areas follow Posadas et al. (2001) and table 1. (**a**) Probabilities of  $j.topol=0.32$ ,  $j.tip=0.32$  (**b**) Probabilities of  $j.topol=0.5$ ,  $j.tip=1$

un-sampled data set might be misleading, while jack-knifing could help to decide which is the most supported solution.

## ***Second Case: The Support for the Original Ranking***

There are two main approaches to define amazonian areas of endemism, eight areas from Bates et al. (1998) and Da Silva et al. (2005) or 16 areas from Da Silva and Oren (1996). López-Osorio and Miranda-Esquivel (2010), used both ways to establish conservation priorities for Amazonia's areas of endemism.

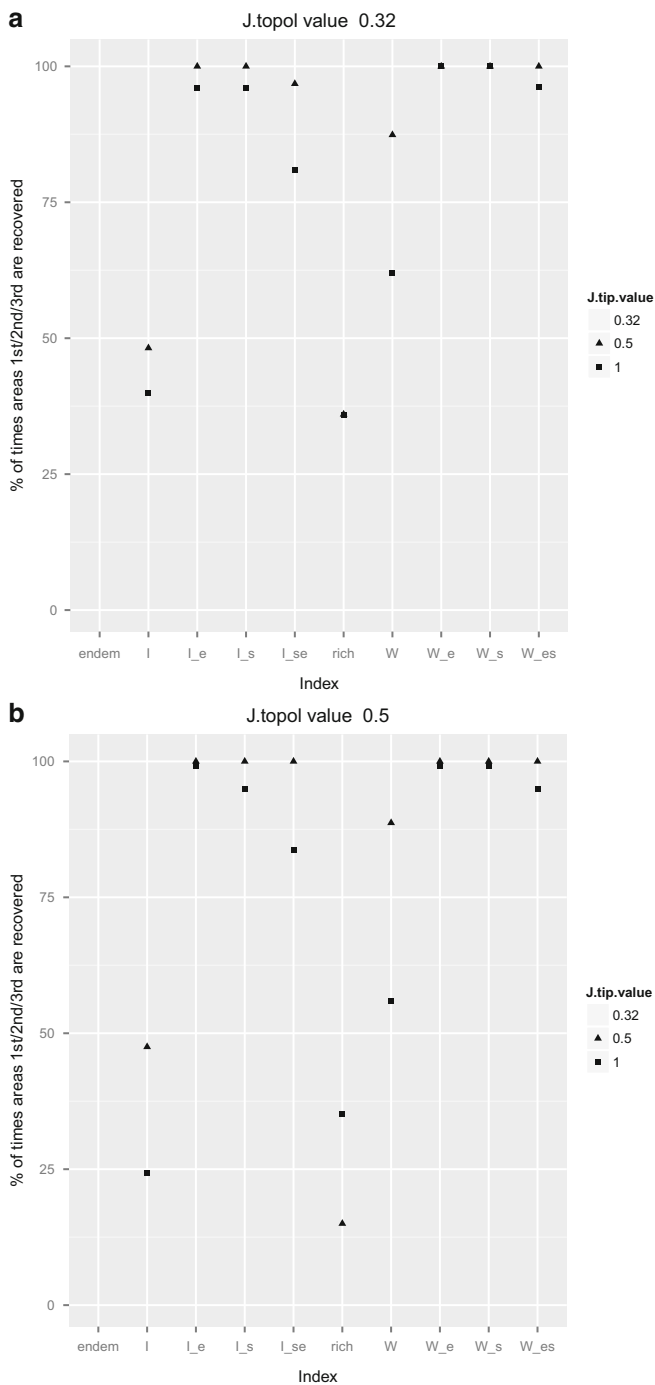
Using Bates et al. (1998) areas, they found that Guiana and Inambari are the first and second priority areas. Inambari is the richest area while Guiana presents the highest endemism value. Their inferences were based on  $W_{es}$ , on theoretical grounds as the index includes endemism and standardization (López-Osorio and Miranda-Esquivel 2010).

The reanalysis showed that the best index is either  $W_{es}$ ,  $W_e$  or  $W_s$  (Fig. 4). These three indices select Guiana as the first area and Inambari as the second area (Fig. 5), as stated in the original paper. In this example the re-sampling reinforces the original findings, giving a stronger support to the areas chosen as first and second in the ranking.

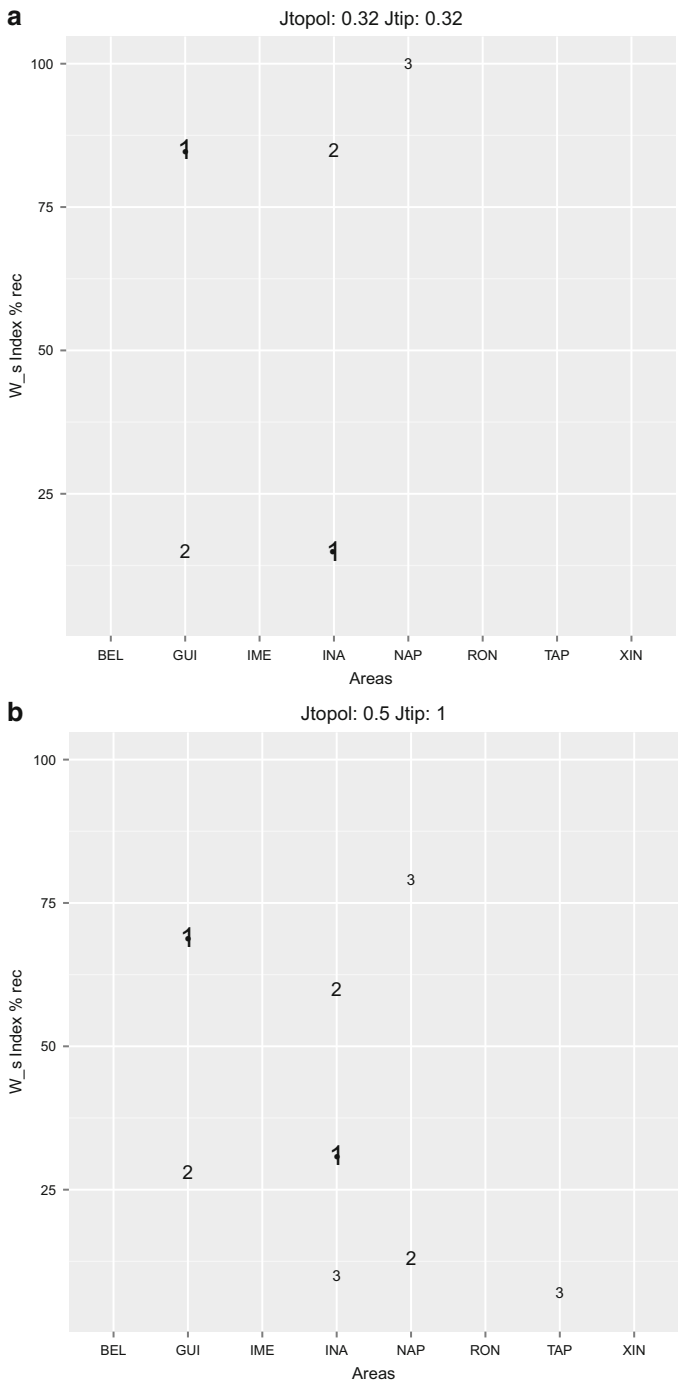
Using the areas from Da Silva and Oren (1996), López-Osorio and Miranda-Esquivel (2010) found that depending on the index, either Guiana2 or Rondonia could be the highest priority area, while the second area could be Guiana3, Inambari2 or even Rondonia or Guiana2. Therefore, the first question is, which is the best index for conservation in Amazonia? and given that index, which are the areas chosen as the first and second priority?

López-Osorio and Miranda-Esquivel (2010) found that most indices selected the same area Guiana2, which could be seen as there is no difference given the index. The reanalysis showed that in general  $I_s$  and  $W_s$  are more stable than any other index, and  $I_s$  behaves better than  $W_s$ . As the size of the topologies is different and some large topologies with more nodes may have more impact than smaller topologies, standard  $I$  and  $W$  indices are not stable (Fig. 6). The first area is Guiana2 in all indices used, while the second area varies: Rondonia, Guiana3 or Inambari2 (Fig. 7). These results are similar to those found by López-Osorio and Miranda-Esquivel (2010). Here the re-sampling helped to resolve the initial discrepancy as the highest priority is Guiana2 and not Rondonia, that could be a possible candidate. The second area could be any of the three initially considered, so the evidence is not misleading but inconclusive to define the second area, even after re-sampling the data.

These brief examples show that the confidence of the original ranking should be evaluated using re-sampling, as an un-sampled ranking analysis could be unstable when some information (phylogenies or species) is deleted. The results may render any output, from a different answer from the original ranking to a congruent answer with the original ranking. Only after the re-sampling analysis, the quality of the answer could be stated without hesitation. Even if we only calculate the support for

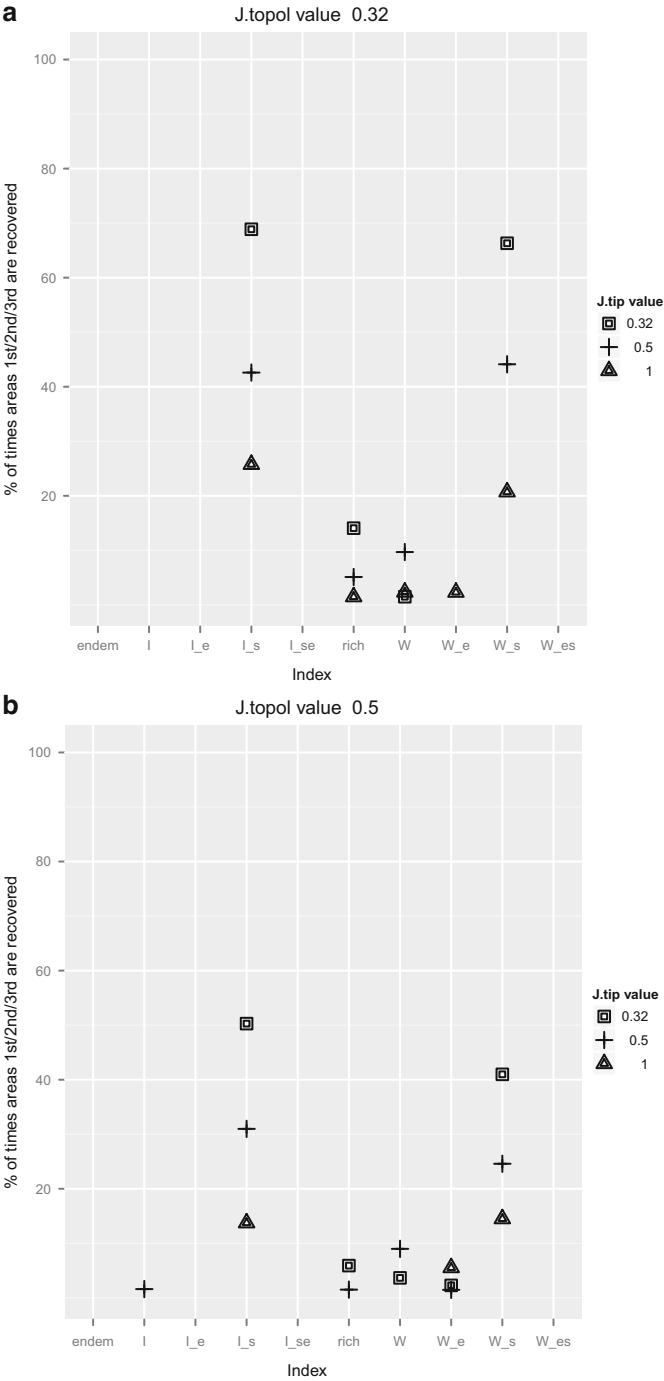


**Fig. 4** Number of hits with a delete value of 0.32 or 0.5 for j.topol and three j.tip values of 0.32, 0.5, and 1 (in the last situation, the whole topology is deleted) (Data from López-Osorio and Miranda-Esquivel (2010). Areas from Bates et al. (1998)). (a) Number of hits with a j.topol value of 0.32 (b) Number of hits with a j.topol value of 0.5

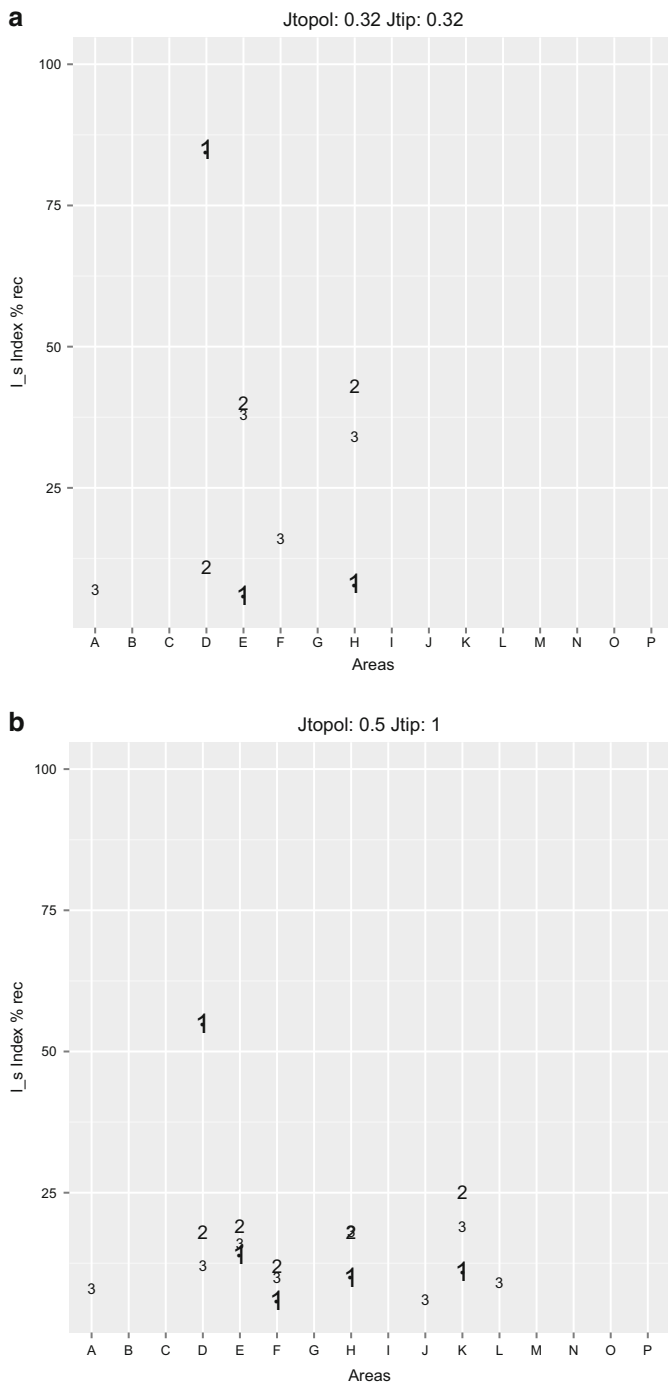


**Fig. 5** Number of times an Area is recovered as the first (number 1), second (number 2), or third area (number 3). These are the lowest **(a)** and the highest **(b)** delete probabilities used in this analysis. **(a)** Probabilities of  $j.topol=0.32$ ,  $j.tip=0.32$  **(b)** Probabilities of  $j.topol=0.5$ ,  $j.tip=1$





**Fig. 6** Number of hits with a delete value of 0.32 or 0.5 for j.topol and three j.tip values of 0.32, 0.5, and 1 (in the last situation, the whole topology is deleted) (Data from López-Osorio and Miranda-Esquivel (2010). Areas from Da Silva and Oren (1996)). (a) Number of hits with a j.topol value of 0.32 (b) Number of hits with a j.topol value of 0.5



**Fig. 7** Number of times an Area is recovered as the first (number 1), second (number 2), or third area (number 3). These are the lowest (**a**) and the highest (**b**) delete probabilities used in this analysis (Data from López-Orsio and Miranda-Esquivel (2010). Areas from Da Silva and Oren (1996)). (Acronyms *A* = Guiana3, *B* = Guiana2, *C* = Guiana1, *D* = Imeri2, *E* = Imeri1, *F* = Napo3, *G* = Napo2, *H* = Napo1, *I* = Inambari1, *J* = Inambari2, *K* = Inambari3, *L* = Inambari4, *M* = Rondonia, *N* = Tapajos, *O* = Xingu, *P* = Belen). (**a**) Probabilities of  $j.topol=0.32$ ,  $j.tip=0.32$  (**b**) Probabilities of  $j.topol=0.5$ ,  $j.tip=1$

a given ranking, the results after re-sampling would give a clue of the situation when the information is perturbed.

**Acknowledgements** I am grateful to Roseli Pellens for her kind invitation to participate in the book. Two anonymous referees helped to improve my perspective about Jack-knife and Conservation. I am indebted to División de Investigación y Extensión, Facultad de salud (project 5658), and Facultad de Ciencias, Universidad Industrial de Santander (project 5132) for their financial support.

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Albert CH, de Bello F, Boulangeat I, Pellet G, Lavorel S, Thuiller W (2012) On the importance of intraspecific variability for the quantification of functional diversity. *Oikos* 121(1):116–126
- Bates JM, Hackett SJ, Cracraft J (1998) Area-relationships in the neotropical lowlands: an hypothesis based on raw distributions of passerine birds. *J Biogeogr* 25(4):783–793
- Brown J, Hedtke S, Lemmon A, Lemmon E (2010) When trees grow too long: investigating the causes of highly inaccurate bayesian branch-length estimates. *Syst Biol* 59(2):145–161
- Cavender-Bares J, Kozak KH, Fine PVA, Kembel SW (2009) The merging of community ecology and phylogenetic biology. *Ecol Lett* 12(7):693–715
- Crowley P (1992) Resampling methods for computation-intensive data analysis in ecology and evolution. *Annu Rev Ecol Syst* 23:405–447
- Crozier R (1992) Genetic diversity and the agony of choice. *Biol Conserv* 61(1):11–15
- Da Silva JMC, Oren D (1996) Application of parsimony analysis of endemism in amazonian biogeography: an example with primates. *Biol J Linn Soc* 59(4):427–437
- Da Silva JMC, Rylands AB, Da Fonseca GAB (2005) The fate of the amazonian areas of endemism. *Conserv Biol* 19(3):689–694
- Efron B (1979) Bootstrap methods: another look at the jackknife. *Ann Stat* 7(1):1–26
- Faith D (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61(1):1–10
- Farris J, Albert V, Källersjö M, Lipscomb D, Kluge A (1996) Parsimony jackknifing outperforms neighbor-joining. *Cladistics* 12(2):99–124
- Goloboff PA, Carpenter JM, Arias JS, Miranda-Esquivel DR (2008) Weighting against homoplasy improves phylogenetic analysis of morphological data sets. *Cladistics* 24(5):758–773
- Gotelli N, Graves G (1996) Null models in ecology. Smithsonian Institution Press, Washington, DC
- Hedges SB (1992) The number of replications needed for accurate estimation of the bootstrap *P* value in phylogenetic studies. *Mol Biol Evol* 9(2):366–369
- Holmes S (2003) Bootstrapping phylogenetic trees: theory and methods. *Stat Sci* 18(2):241–255
- Krajewski C (1994) Phylogenetic measures of biodiversity: a comparison and critique. *Biol Conserv* 69(1):33–39
- Lanyon S (1987) Jackknifing and bootstrapping: important “new” statistical techniques for ornithologists. *Auk* 104(1):144–146
- Lehman SM (2006) Conservation biology of Malagasy strepsirrhines: a phylogenetic approach. *Am J Phys Anthropol* 130(2):238–253
- López-Osorio F, Miranda-Esquivel D (2010) A phylogenetic approach to conserving amazonian biodiversity. *Conserv Biol* 24(5):1359–1366

- Miller JA (2003) Assessing progress in systematics with continuous jackknife function analysis. *Syst Biol* 52(1):55–65
- Pattengale N, Alipour M, Bininda-Emonds O, Moret B, Stamatakis A (2010) How many bootstrap replicates are necessary? *J Comput Biol* 17(3):337–354
- Pérez-Losada M, Jara C, Bond-Buckup G, Crandall K (2002) Conservation phylogenetics of chilean freshwater crabs *Aegla* (anomura, aeglidae): assigning priorities for aquatic habitat protection. *Biol Conserv* 105(3):345–353
- Posadas P, Miranda-Esquivel DR, Crisci J (2001) Using phylogenetic diversity measures to set priorities in conservation: an example from southern South America. *Conserv Biol* 15(5):1325–1334
- Prado A, Hawkins J, Yesson C, Bárcenas R (2010) Multiple diversity measures to identify complementary conservation areas for the Baja California peninsular cacti. *Biol Conserv* 143(6):1510–1520
- Purvis A, Hector A (2000) Getting the measure of biodiversity. *Nature* 405(6783):212–219
- Quenouille M (1949) Approximate tests of correlation in time series. *R Stat Soc Ser B* 11:18–84
- R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <http://www.R-project.org>
- Rannala B, Zhu T, Yang Z (2012) Tail paradox, partial identifiability, and influential priors in bayesian branch length inference. *Mol Biol Evol* 29(1):325–335
- Redding DW, Hartmann K, Mimoto A, Bokal D, DeVos M, Mooers AØ (2008) Evolutionarily distinctive species often capture more phylogenetic diversity than expected. *J Theor Biol* 251(4):606–615
- Rolland J, Cadotte MW, Davies J, Devictor V, Lavergne S, Mouquet N, Pavoine S, Rodrigues A, Thuiller W, Turcati L, Winter M, Zupan L, Jabot F, Morlon H (2012) Using phylogenies in conservation: new perspectives. *Biol Lett* 8(5):692–694
- Schneider SM (2012) A metric of biodiversity that integrates abundance, phylogeny, and function. *Oikos* 121(8):1191–1202
- Schweiger O, Klotz S, Durka W, Kühn I (2008) A comparative test of phylogenetic diversity indices. *Oecologia* 157(3):485–495
- Smith E, van Belle G (1984) Nonparametric estimation of species richness. *Biometrics* 40:119–129
- Spinks P, Shaffer H (2009) Conflicting mitochondrial and nuclear phylogenies for the widely disjunct *Emys* (testudines: Emydidae) species complex, and what they tell us about biogeography and hybridization. *Syst Biol* 58(1):1–20
- Vane-Wright R, Humphries C, Williams P (1991) What to protect? – systematics and the agony of choice. *Biol Conserv* 55(3):235–254
- Vellend M, Cornwell W, Magnuson-Ford K, Mooers A (2011) Measuring phylogenetic biodiversity in Biological Diversity: Frontiers in Measurement and Assessment. eds A.E. Magurran B.J. McGill. Oxford University Press, New York, pp 194–207, Chap. 14
- Wheeler WC (1995) Sequence alignment, parameter sensitivity, and the phylogenetic analysis of molecular data. *Syst Biol* 44(3):321–331
- Whiting A, Lawler S, Horwitz P, Crandall K (2000) Biogeographic regionalization of Australia: assigning conservation priorities based on endemic freshwater crayfish phylogenetics. *Anim Conserv* 3(2):155–163
- Wickham H (2009) ggplot2: elegant graphics for data analysis. Springer, New York
- Winter M, Devictor V, Schweiger O (2013) Phylogenetic diversity and nature conservation: where are we? *Trends Ecol Evol* 28(4):199–204

# Assessing Hotspots of Evolutionary History with Data from Multiple Phylogenies: An Analysis of Endemic Clades from New Caledonia

Roseli Pellens, Antje Ahrends, Peter M. Hollingsworth,  
and Philippe Grandcolas

**Abstract** The great bulk of the present knowledge of the Tree of Life comes from many phylogenies, each with relatively few tips, but with lots of diversity concerning taxa and characters sampled and methods of analysis used. For several biodiversity hotspots this is the kind of data available and ready to be used to have a better understanding on the evolutionary patterns and to identify areas with remarkable evolutionary history. But relying on data coming from independent studies raises some methodological challenges of standardization, comparability and assessments of bias to make the best use of the currently available information. To bring light to this subject here we analyzed the distribution of phylogenetic diversity in New Caledonia, a biodiversity hotspot characterized by strong rates of regional and internal endemism. We used a dataset with 18 phylogenies distributed in 16 study sites, and based our analysis on the measure  $W_s$  sum. Our study comprises the analysis of (1) the role of the number of phylogenies on site' scores and a strategy of standardization of the dataset by the number of phylogenies; (2) the influence of species richness on site scores and the design of the measure  $W_s$  ranks to focus on the most divergent species of each phylogeny; (3) an assessment of the influence of individual phylogenies; (4) a resampling strategy using multiple phylogenies to verify the results' stability.

**Keywords**  $W_s$  • Resampling • Rarefaction • Meta-analysis

---

R. Pellens (✉) • P. Grandcolas  
Institut de Systématique, Evolution, Biodiversité, ISYEB – UMR 7205 CNRS MNHN  
UPMC EPHE, Muséum National d'Histoire Naturelle, Sorbonne Universités,  
45 rue Buffon, CP 50, 75005 Paris, France  
e-mail: [pellens@mnhn.fr](mailto:pellens@mnhn.fr); [pg@mnhn.fr](mailto:pg@mnhn.fr)

A. Ahrends • P.M. Hollingsworth  
Royal Botanic Garden Edinburgh, Edinburgh EH3 5LR, UK  
e-mail: [a.ahrends@rbge.ac.uk](mailto:a.ahrends@rbge.ac.uk); [P.Hollingsworth@rbge.ac.uk](mailto:P.Hollingsworth@rbge.ac.uk)

## Introduction

Opening wide the sampling window in biodiversity studies is a major goal today, and this leads to two major research challenges. On the one hand is the difficulty of dealing with big data, as, for example, those from entire genomes. Now that the main barriers to obtain enormous sequences seem to be broken, and that this kind of data is becoming easy and much cheaper to be obtained, the main constraint is the analysis of such huge datasets. On the other hand are the difficulties associated with synthesizing evidences produced by several independent studies where, by definition, the sampling protocols are not standardized. Both issues are at the core of the analysis of phylogenetic diversity for conservation and deserve more attention if we are to produce sound guidelines for conservation. However, in this chapter we will focus only on the last one.

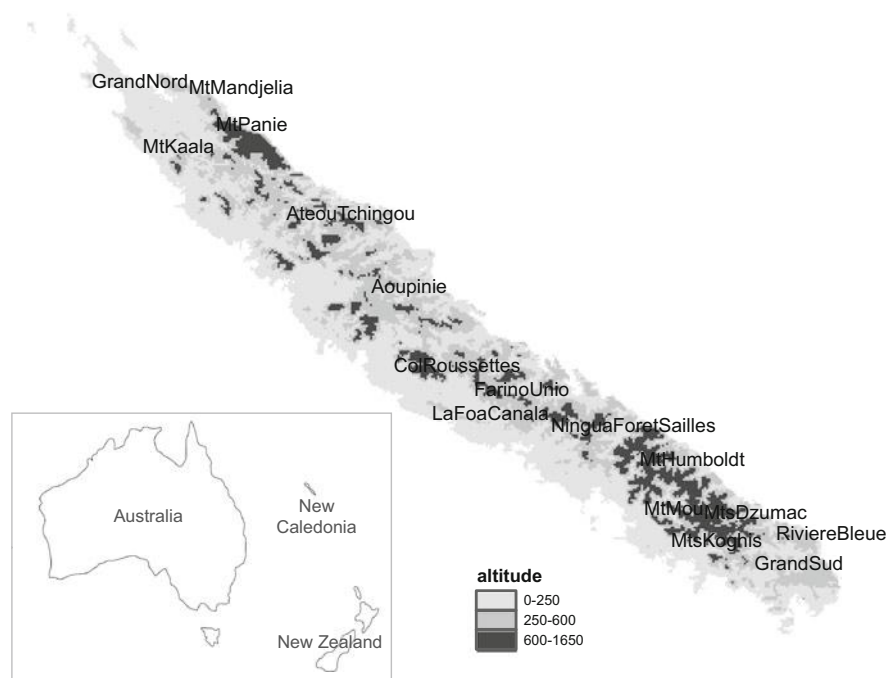
Our interest in this problem is due to the fact that the great bulk of present knowledge of the Tree of Life does not result from a comprehensive analysis with standardized samples of taxa and characters. Instead, the greatest part of published works comprises studies at the level of families or genera, with lots of diversity concerning taxon and characters sampling and methods of analysis. But the increased facility of molecular sequencing and phylogenetic analysis observed in the recent years has led to a substantial increase in available phylogenies. As a consequence, for some biodiversity hotspots, an important number of detailed phylogenetic studies for several distinct groups are now available. The data from these independent studies, associated with a greater accuracy and availability of species occurrence records, provide a rich material that can enhance biodiversity conservation decisions. This allows for detecting evolutionary patterns across a broader sample of the Tree of Life and, ultimately, for detecting hotspots of evolutionary history within these biodiversity Hotspots. Obviously, the higher the diversity of groups covered by the set of phylogenies the finer the picture of the Tree of Life in the region and the more reliable the contribution of phylogenetic information to the conservation planning (Rodrigues et al. 2005).

Although the possibility of integrating results from different phylogenies has been studied for a while (see Posadas et al. 2001, 2004; Faith et al. 2004; López-Orsorio and Miranda Esquivel 2010), we are only starting to explore the implications of different sampling effort and imperfect knowledge on studies of phylogenetic diversity for assessing areas for conservation (see Nipperess and Matsen 2013 and Nipperess, chapter “[The Rarefaction of Phylogenetic Diversity: Formulation, Extension and Application](#)” and Miranda-Esquivel, chapter “[Support in Area Prioritization Using Phylogenetic Information](#)”). In order to shed light to this problem here we propose some solutions when assessing hotspots for conservation within New Caledonia.

## ***Assessing Hotspots of Evolutionary Distinctiveness in New Caledonia***

New Caledonia is a Pacific Ocean island located some 1450 km east of Australia (Fig. 1). It is about 500 km long and 50 km wide and is classed as a globally significant biodiversity hotspot (Myers et al. 2000; Grandcolas et al. 2008; Kier et al. 2009). The island's biological diversity is threatened by activities associated with large-scale opencast nickel mining, an increased frequency of fires, and by ecological displacements caused by invasive species (Bouchet et al. 1995; Beauvais et al. 2006; Pascal et al. 2008; Pellens and Grandcolas 2010).

A key feature of the New Caledonian biota is its high level of endemism. The geographical isolation of the island and its ultramafic soils have all been proposed as factors promoting high levels of endemism. This endemism exists at the level of the island, but also at finer geographical scales, and within New Caledonia micro-endemism is common with many species restricted to individual mountains, mountain slopes, valleys, watercourses or edaphic 'islands' (e.g., Murienne et al. 2005; Sharma and Giribet 2009; Espeland and Johanson 2010b; Pillon et al. 2010; Nattier et al. 2012, 2013).



**Fig. 1** Localization of New Caledonia in the southern Pacific and the 16 study areas in New Caledonia's mainland

This high level of endemism has attracted considerable scientific interest in the evolutionary history of the island (see the review in Pellens and Grandcolas 2010) and phylogenies for several groups are now available. There have also been macro-analyses of the distribution of micro-endemic species (Wulff et al. 2013). However, to date there has been no systematic evaluation of the distribution of biodiversity in the context of evolutionary distinctiveness within the island, comparing multiple taxonomic groups over multiple geographical locations. In the current paper we tackle this topic with the aim to identify sites with high levels of phylogenetic diversity.

This type of study raises some methodological challenges. In an ideal world, one would design a sampling strategy involving equal sampling effort (or at least quantified sampling effort) at multiple sites for multiple sets of taxa, sampled for a common and comparable set of characters, and with the data analyzed in a common and comparable analytical framework. Unfortunately such a dataset does not exist presently for New Caledonia. Instead, we have taken an approach to make the best use of the currently available information by mining the literature for tree topologies, and then developing an analytical framework that copes with the shortcomings of the extant dataset.

Specifically we have pulled together the available data consisting of multiple phylogenies from different groups of organisms, of different levels of species richness, built from different character sets, different analytical methods and partially overlapping geographical locations. Our framework aims to standardize the contributions of these different datasets in a meta-analysis, and also to quantify the inevitably high-levels of uncertainty and variance in the range of possible conclusions that comes from dealing with (a) a complex biological system, and (b) imperfect data.

## Material and Methods

### *Data and Sampling*

We included all available phylogenetic studies up to 2010 that satisfied the three following conditions: (1) having a monophyletic group from New Caledonia with three species or more; (2) having extensive coverage of the geographic distribution of the group within New Caledonia's mainland; (3) having species represented in at least three out of the 16 selected geographical areas (see below). This resulted in 18 phylogenies encompassing both terrestrial and freshwater organisms (Table 1). The monophyletic clades in which New Caledonian species were found ranged from 3 to 59 species (mean 14.9, median 10.5) and in total these phylogenies included 269 species, all endemic to New Caledonia. They included organisms as diverse as insects, harvestmen, gastropods, vertebrates (geckos – Squamata), and vascular plants.



**Table 1** Overview of the 18 phylogenies used in this study, with complementary references of species distribution

Key		Family	Genus	Reference
1	Blattaria	Blattidae	<i>Angustonicus</i>	Murienne (2006)
2	Blattaria	Blattidae	<i>Lauraesilpha</i>	Murienne et al. (2008)
3	Heteroptera	Tingidae	<i>Cephalidiosus</i>	Murienne et al. (2009)
			<i>Nobarnus</i>	
4	Orthoptera	Eneopteridae	<i>Agnoteocus</i>	Desutter-Grandcolas and Robillard (2006)
5	Trichoptera	Hydrobiosidae	<i>Xanthochorema</i>	Espeland et al. (2008)
6	Trichoptera	Hydropsychidae	<i>various</i>	Espeland and Johanson (2010a)
7	Trichoptera	Ecnomidae	<i>Agmina</i>	Espeland and Johanson (2010b)
8	Coleoptera	Dytiscidae	<i>Rhantus</i>	Balke et al. (2007)
9	Opiliones	Troglosironidae	<i>Troglosiro</i>	Sharma and Giribet (2009)
10	Gastropoda	Hydrobiidae	<i>various</i>	Haase and Bouchet (1998)
11	Squamata	Scincidae	<i>Marmorosphax</i>	Sadlier et al. (2009)
12	Squamata	Scincidae	<i>various</i>	Sadlier et al. (2004)
13	Squamata	Diplodactylidae	<i>Dierokekko</i>	Bauer et al. (2006)
14	Squamata	Diplodactylidae	<i>Eurydactylodes</i>	Bauer et al. (2009)
15	Squamata	Diplodactylidae	<i>Rhacodactylus</i>	Good et al. (1997) and Bauer (1990)
16	Ericales	Sapotaceae	<i>Planchonella</i>	Swenson et al. (2007) and Munzinger and Swenson (2009)
17	Ericales	Sapotaceae	<i>various</i>	Munzinger and Swenson (2009), Swenson et al. (2008) and Swenson and Munzinger (2009, 2010a, b, c)
18	Ericales	Ebenaceae	<i>Diospyros</i>	Duangjai et al. (2009)

Key = the reference number that will be used in Tables 2 and 3 when referring to these studies

We considered 16 areas (sites) within New Caledonian mainland. This set of sites includes the very great majority of areas with remaining native forests and are distributed throughout the length of the island. These areas correspond to geographical entities with discrete boundaries, such as isolated mountains, or parts of large ridge systems or lowlands separated from the adjacent one by main valleys, rivers or lakes (Fig. 1). The basic condition for including an area in this analysis was the availability of at least five phylogenetic studies containing species represented at the site, and a minimum of ten species studied. Distributional data were collected from the original phylogenetic studies from the literature cited therein, and from the specialists working on the group in the region (Table 1). Species richness in this paper

refers to the number of species from the 18 phylogenies occurring in each area. A species was considered a microendemic if it was recorded in an area and nowhere else. In total, our data set consists of 523 records of occurrence of a given species from a phylogeny at one of these 16 sites.

### *Metric and Corrections for Bias*

We calculated evolutionary distinctiveness using the topology based metric, the  $W_s$  index from Posadas et al. (2001), which is derived from the Taxonomic Distinctness index conceived by Vane-Wright et al. (1991). We chose this metric for three reasons. (1) It assigns higher values to species with fewer and more distant relatives than to species with more and closer relatives, allowing for a better identification of areas with more phylogenetically divergent species (Redding et al. 2008). (2) It is designed for combining phylogenetic information from different cladograms, independently of the kind of characters (morphological, molecular, etc.) or reconstruction method, since it is a topology based metric. This way, we were able to integrate data from phylogenies of taxa as different as plants, reptiles, molluscs and arthropods to study the evolutionary distinctiveness of different areas in New Caledonia. (3) Each phylogeny contributes with the same amount of information, independently of its total species' number, as the  $W_s$  values for the species in any given phylogeny sum to one.

The traditional procedure is to sum  $W_s$  of all species present in each area and rank areas according to this sum (Posadas et al. 2001; Lehman 2006; McGoogan et al. 2007; López-Orsorio and Miranda Esquivel 2010). However, this practice often leads to strong correlations with species richness (see López-Orsorio and Miranda Esquivel 2010), having the possibility of masking important evolutionary divergence in sites with less species, or less phylogenies. Secondly, as  $W_s$  is bound between 0 and 1 for a given phylogeny, it is sensitive to the number of sampled species in each phylogeny. Although this will in part be driven by species richness, it is also simply affected by the scope of the study selected by the investigator (e.g. family level or genus level). Thus the wider the phylogenetic breadth of a study (the more species included), the lower the overall maximum value for any one species. Thirdly, in the absence of exhaustive location-based sampling, the data available on the evolutionary diversity of a given site will simply reflect the taxa that happen to have been sampled for individual research projects. If this bias is not corrected for, it will be hard to see the phylogenetic content, as the number of phylogenies and the number of species in each site might drive the result.

In order to address these shortcomings, we designed a method to highlight sites containing the most divergent taxa from each of the phylogenies. We firstly calculated  $W_s$  for each species in each phylogeny, and placed the species in order from the highest to the lowest  $W_s$  value. We then awarded "points" to the most divergent species in each phylogeny and compared the resulting scores among sites. As we were interested in the 'front-runners' from each phylogeny – we firstly took the top

three species, i.e. the most ‘basal’ species from each phylogeny, assigning them a score of 1 (for most basal) 0.67 (second place) and 0.33 (for third place). However, we latterly truncated this to scores for 1st and 2nd place (1 and 0.67) to emphasise the most divergent species. In the case of ties for the most divergent species, the total score of 1.67 was divided by the number of species that tied. Where there is a unique first place score, but ties for second place, the ‘second prize’ of 0.67 was ‘shared’ amongst the species which tied. The scores were then summed for all phylogenies at each site.

This method ensures that each phylogeny contributes a directly equal total score, and we are simply assessing in each case where the most divergent species are. The downside of using first and second ranked species, is that it discards information from all of the other species in each data set. To accommodate this, we also continue to report the (more conventional) sum of *Ws* values, also standardised by the number of phylogenies present at a given site.

## ***Resampling Analysis***

Our data set is constrained by the number of phylogenies that were available. To assess whether our findings are sensitive to the composition of the sample of phylogenies we have, we designed two tests. The first was through assessing the changes associated with the exclusion of a single phylogeny (single drops, a.k.a. Jackknifing). This is to see if the findings are being driven by a single influential phylogeny. Secondly, we undertook a resampling (or rarefaction) procedure, by defining subsets of 1, 2, 3... 15 phylogenies in a site and then calculating the mean and standard deviation of site’s scores with all possible combination of phylogenies with species occurring in it. This was to establish whether the results are stable with respect to the number of phylogenies we have available.

The R codes for these analyses are available from A.Ahrends@rbge.org.uk on request.

## **Results**

### ***The Role of the Number of Phylogenies on Site Scores***

In our dataset the number of phylogenies with species occurring at a site ranged between 5 and 16 (mean and median=11). So, the first point that we investigated was the role of the number of phylogenies in site’s scores. This showed that over 75 % of the site’s ranking with *Ws sum* was explained by the number of phylogenies with species in the site (Regression model:  $\text{Sum } Ws = -2.13 + 0.555 \text{ number of phylogenies}$ ;  $F=41.75$ ;  $DF=14$ ;  $p=0.000$ ;  $R^2=0.75$ ). With *Ws ranks* the influence of the number of phylogenies is a bit smaller but still important (Regression model:

$Ws\ ranks = -1.03 + 0.259 \text{ number of phylogenies}$ ;  $F=26.75$ ;  $DF=14$ ;  $p=0.000$ ;  $R^2=0.66$ ).

Based on it, we decided to standardize by dividing total *Ws sum* or total *Ws ranks* in the site by the number of phylogeny occurring in it. As expected, this came to a result where much less of the site's ranking is explained by the number the phylogenies with species occurring in the site, but the number of phylogenies still explains a substantial proportion of the variance (Regression model:  $Ws\ sum/\text{number of phylogenies} = 0.04 + 0.0105 \text{ number of phylogenies}$ ;  $F=8.9$ ;  $DF=14$ ;  $p=0.01$ ;  $R^2=0.39$ ; and  $Ws\ ranks/\text{number of phylogenies} = 0.082 + 0.0237 \text{ number of phylogenies}$ ;  $F=6.29$ ;  $DF=14$ ;  $p=0.025$ ;  $R^2=0.31$ ). In both cases, the standardized and non-standardized values are still correlated (Spearman  $r=0.9$ ,  $p<0.01$ ; and  $r=0.83$ ,  $p<0.01$  for *Ws sum* and *Ws ranks*, respectively). But ranking priorities change, putting in evidence the phylogenetic distinctiveness of some groups occurring in sites with less phylogeny (Figs. 2 and 3).

### ***The Influence of Species Richness on Site Scores***

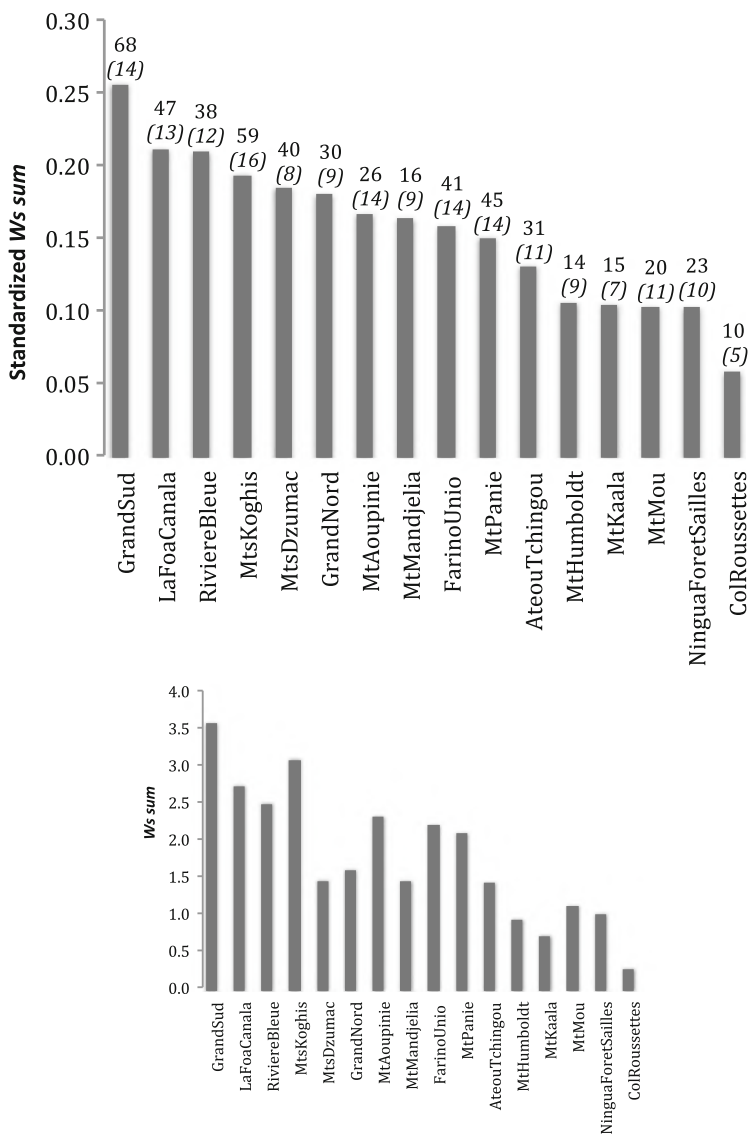
The number of species in the 16 sites varied between 10 and 68 (mean=33; median=31), and over 80 % of variation in the sum of *Ws* is explained by species richness. When *Ws sums* are standardized by the number of phylogenies, 70 % of the variation is still explained by species richness – sites with more species have greater chances of accumulating high *Ws sums* (Fig. 4a, b).

The analysis with *Ws ranks* shows that all sites had at least one top or second ranking species (1–14 per site, mean and median=7). The influence of species richness on *Ws ranks* is lower than *Ws sums* with just over 50 % of the variation in *Ws ranks* explained by species richness. When *Ws ranks* were standardized by the number of phylogenies, the influence of species richness became much lower (32 %), although still significant (Fig. 5a, b).

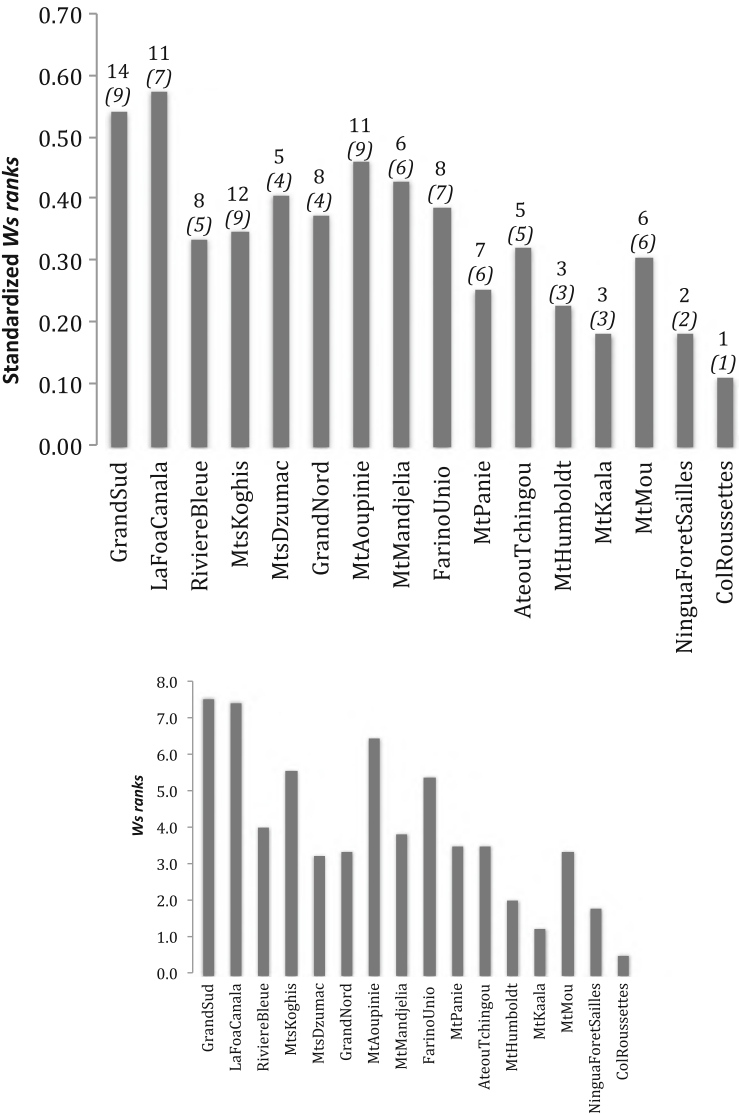
### ***Influence of Individual Phylogenies***

Tables 2 and 3 show the relative levels of evolutionary distinctiveness among sites when each of the 18 phylogenies is excluded from the analysis. It shows that some sites consistently have high levels of evolutionary distinctiveness, some have consistently lower levels, whereas some others show intermediate values and their ranking positions are more sensitive to the inclusion of any one phylogeny.

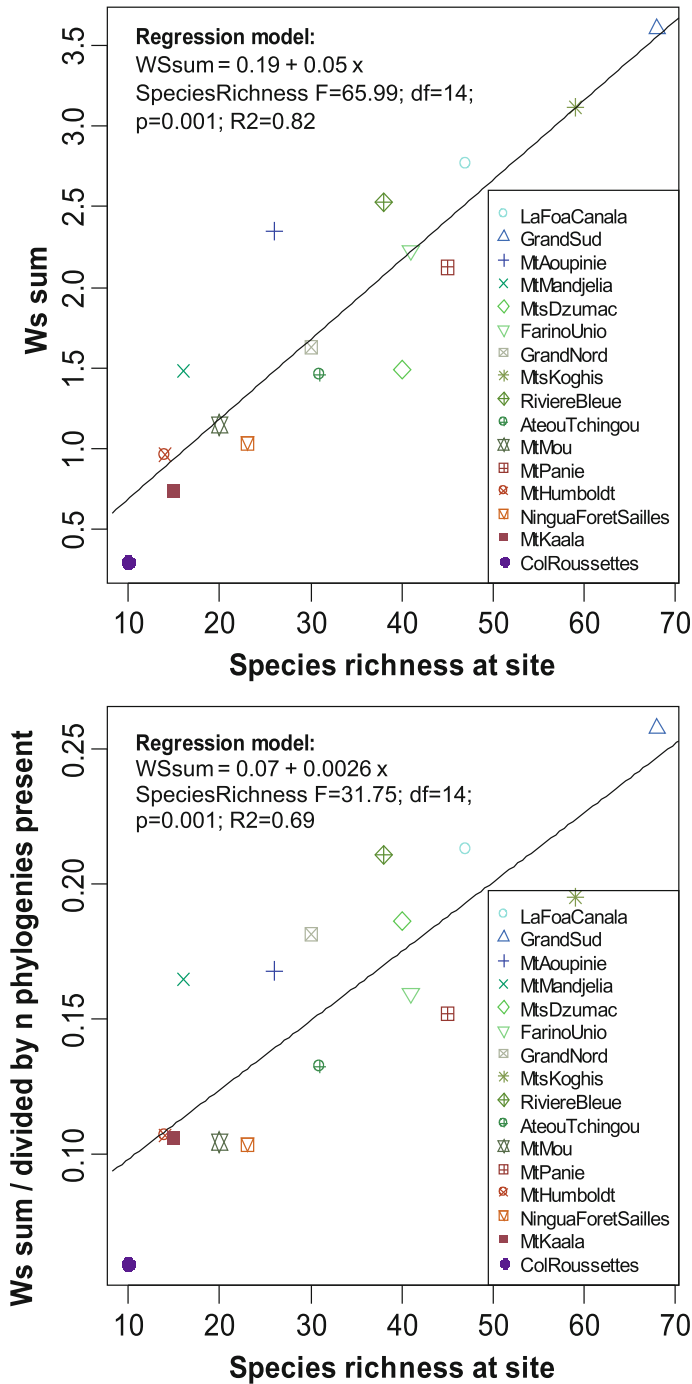
The sum of absolute difference in ranks when each of phylogeny was dropped synthesize this result (Figs. 6 and 7). It shows that several phylogenies contribute to site's ranking, refuting the hypothesis that site's ranking could be highly influenced by phylogenies with more species, or by a subset of phylogenies with more widespread species.



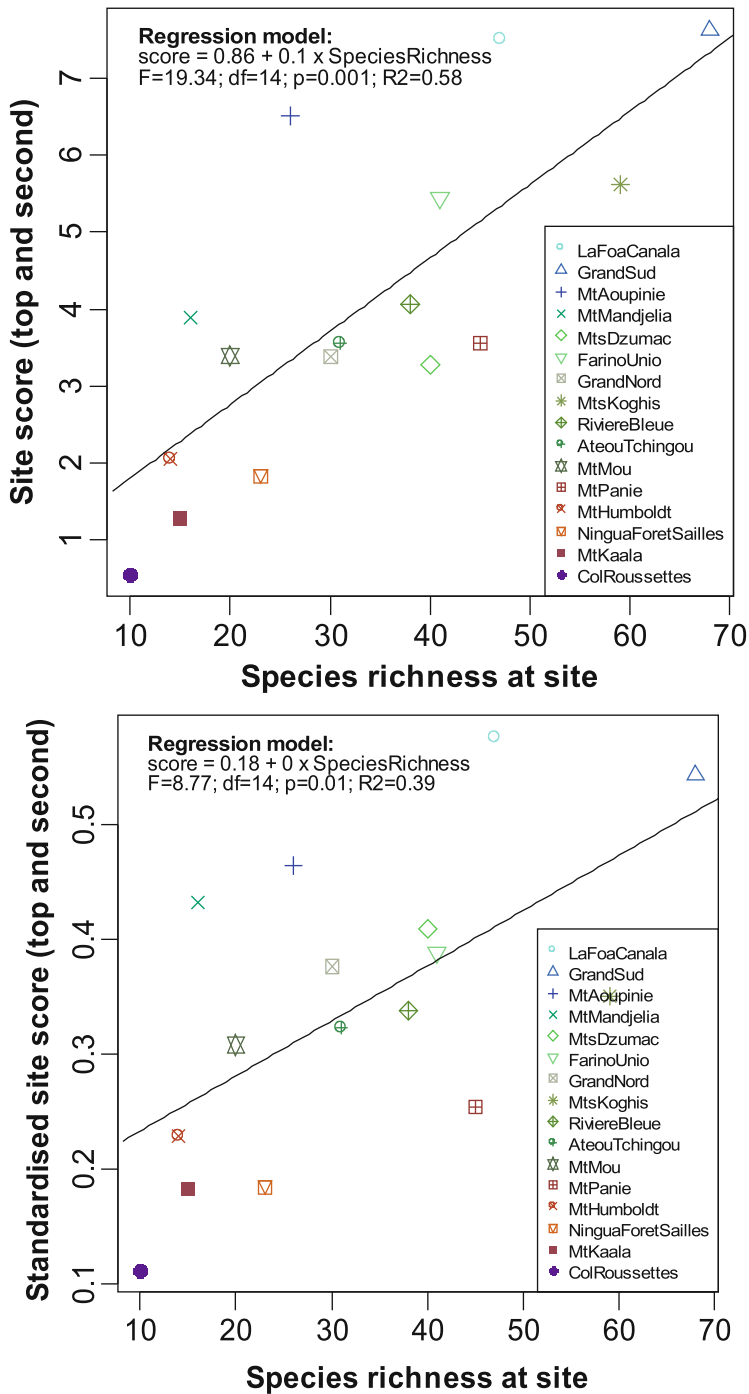
**Fig. 2** Summed Ws values. The main figure shows the values standardized (= divided) by the number of phylogenies present at the site. The numbers on top of each bar give the number of species and phylogenies (in *brackets and italics*) at each site. The small figure at the bottom shows the non-standardized values



**Fig. 3** Summed site scores for species on top and second ranks. The main figures shows the values standardized (= divided) by the number of phylogenies present at the site. The numbers on top of each bar give the number of scoring species and phylogenies (in *brackets and italics*) at each site. The small figure at the bottom shows the non-standardized values



**Fig. 4** Over 80 % of the variation in the sites' WS sums is explained by species richness (*upper figure part*). There is still a strong relationship between species richness and the WS sums when standardised by the number of phylogenies, suggesting that species rich sites also have more species with high WS values (*lower figure part*)



**Fig. 5** A little over 50 % of the variation in the sites' top and second top scores is explained by species richness (*upper figure part*). There is also still a dependency between species richness and the site scores when these are standardised by the number of phylogenies, suggesting that species rich sites also have more species with high WS values (*lower figure part*)



**Table 2** Site ranks (based *Ws sum*) if a given phylogeny is dropped

	Drop...																		Drop none
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
GrandSud	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
LaFoaCanala	3	2	2	3	2	4	3	5	2	2	3	2	2	3	2	3	2	2	2
RiviereBleue	2	3	4	2	3	2	2	2	3	3	2	3	3	4	6	2	3	3	3
MtsKoghis	4	4	6	4	4	5	5	3	5	4	4	6	4	2	3	4	4	4	4
MtsDzumac	5	5	3	5	8	8	6	4	4	5	5	5	5	5	4	5	5	6	5
GrandNord	6	6	5	6	5	3	4	6	6	6	7	4	11	6	5	7	7	5	6
MtAoupinie	7	9	7	7	6	6	7	8	7	7	8	7	7	7	8	6	8	7	7
MtMandjelia	8	7	8	8	7	7	9	10	8	9	6	8	6	8	7	8	6	9	8
FarinoUnio	9	8	10	9	10	9	8	7	9	8	9	9	8	9	9	9	9	8	9
MtPanie	10	10	9	10	9	10	10	9	10	10	10	10	9	10	10	10	10	10	10
AteouTchingou	11	11	11	11	11	11	11	11	11	11	11	11	10	11	11	11	11	11	11
MtHumboldt	12	12	13	12	12	12	15	13	15	15	12	13	14	12	12	12	14		12
MtKaala	13	14	14	13	13	15	15	12	14	12	12	13	12	13	13	15	15	12	13
MtMou	14	13	12	14	15	13	14	14	15	14	13	14	14	12	15	13	13	15	14
NinguaForetSailles	15	15	15	15	14	14	13	13	12	13	14	15	15	15	14	14	14	13	15
ColRoussettes	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16

**B**

	Drop...																		Drop none
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
GrandSud	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1
LaFoaCanala	3	3	2	3	3	4	3	5	3	3	3	3	3	3	3	3	3	3	3
RiviereBleue	4	4	4	4	5	3	4	3	4	4	4	4	4	5	7	4	4	4	4
MtsKoghis	2	2	3	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2
MtsDzumac	9	9	10	9	11	11	11	9	10	8	11	9	8	8	9	8	10	9	9
GrandNord	8	8	8	8	8	8	8	8	8	10	8	8	13	9	8	9	9	8	8
MtAoupinie	5	6	5	5	4	5	5	6	5	5	5	5	5	4	6	5	5	5	5
MtMandjelia	11	11	11	10	9	9	9	12	9	9	10	10	10	10	10	11	8	10	10
FarinoUnio	7	5	7	6	7	6	6	4	6	6	6	6	6	6	4	6	6	6	6
MtPanie	6	7	6	7	6	7	7	7	7	7	7	7	7	7	5	7	7	7	7
AteouTchingou	10	10	9	11	10	10	10	10	11	11	9	11	9	11	11	10	11	11	11
MtHumboldt	14	14	13	14	14	14	14	15	14	14	14	14	14	14	14	14	14	14	14
MtKaala	15	15	15	15	15	15	15	14	15	15	15	15	15	15	15	15	15	15	15
MtMou	12	12	12	12	12	12	12	13	12	12	12	12	11	12	13	12	12	12	12
NinguaForetSailles	13	13	14	13	13	13	13	11	13	13	13	13	12	13	12	13	13	13	13
ColRoussettes	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16

Values in red and bold are ‘real’ drops, i.e. the dropped phylogeny was indeed present at the site. A: ranks based on standardised values. B: ranks based on not standardised values. For a key to the phylogenies see Table 1

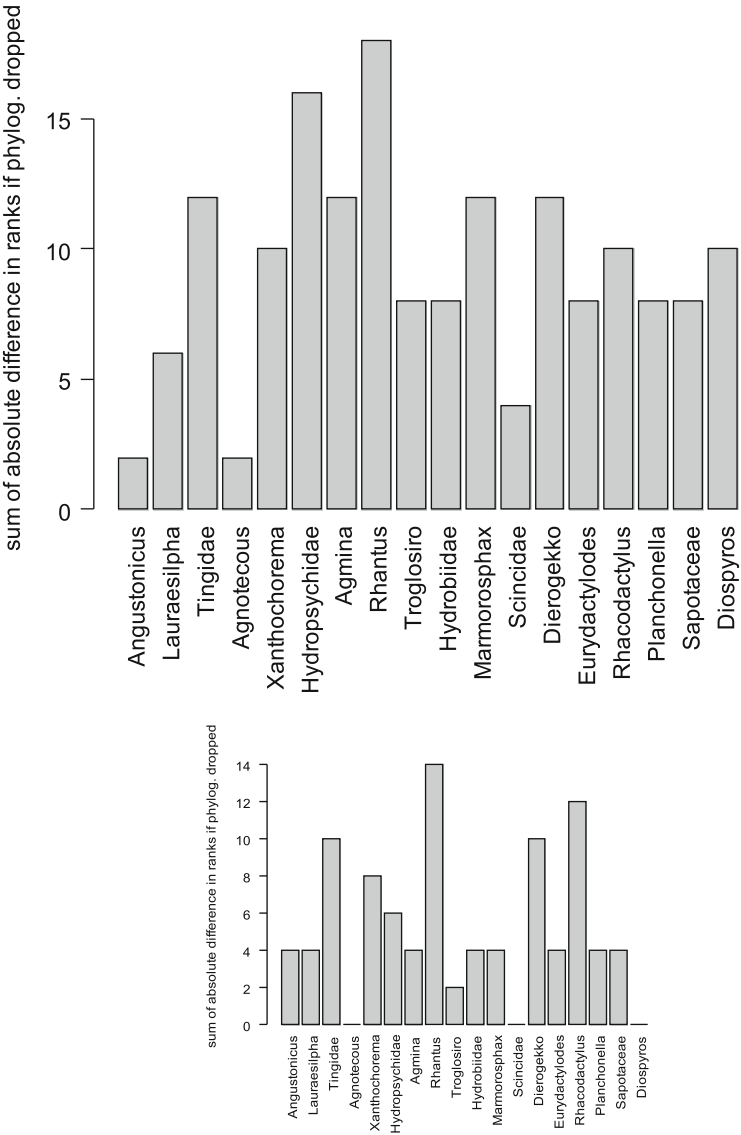
**Table 3** Site ranks (based on *Ws ranks*) if a given phylogeny is dropped

	Drop...																		Drop none
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
GrandSud	1	1	2	1	2	1	1	1	1	2	1	2	1	1	2	1	1	2	1
LaFoaCanala	2	2	1	2	1	3	2	2	2	1	2	1	2	2.5	1	2	2	1	2
RiviereBleue	6	6	5	6	7	6	6	6	6	6	6	6	6	12	12	6	11	6	6
MtsKoghis	4	5	4	4.5	5	5	4	5	4	5	4	5	4	4	5	4	4	4	4
MtsDzumac	11	12	10	12	12	12	11	9	11	10	12	12	11	8	10	11	10	11	12
GrandNord	10	10	9	11	10	9	8	8	10	12	11	11	14	11	9	10	12	10	11
MtAoupinie	3	3	3	3	3	2	3	4	3	3	3	3	3	2.5	3	3	3	3	3
MtMandjelia	12	7	11	7	6	7	7	10	7	7	7	7	7	6	6	12	6	7	7
FarinoUnio	5	4	6	4.5	4	4	5	3	5	4	5	4	5	5	4	5	5	5	5
MtPanie	7.5	11	12	11	8.5	10	13	7	8	8.5	8.5	8.5	8.5	9.5	7.5	7.5	7.5	8	8.5
AteouTchingou	7.5	8	7	8	8.5	8	9	11	12	8.5	8.5	8.5	8.5	9.5	7.5	7.5	7.5	12	8.5
MtHumboldt	13	13	13	13	14	13	12	15	13	13	13	13	12	13	13	13	13	13	13
MtKaala	15	15	14	15	15	15	15	14	15	15	15	15	15	14	15	15	15	15	15
MtMou	9	9	8	9	11	11	10	12	9	11	10	10	10	7	11	9	9	9	10
NinguaForetSailles	14	14	15	14	14	14	14	13	14	14	14	14	13	15	14	14	14	14	14
ColRoussettes	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16

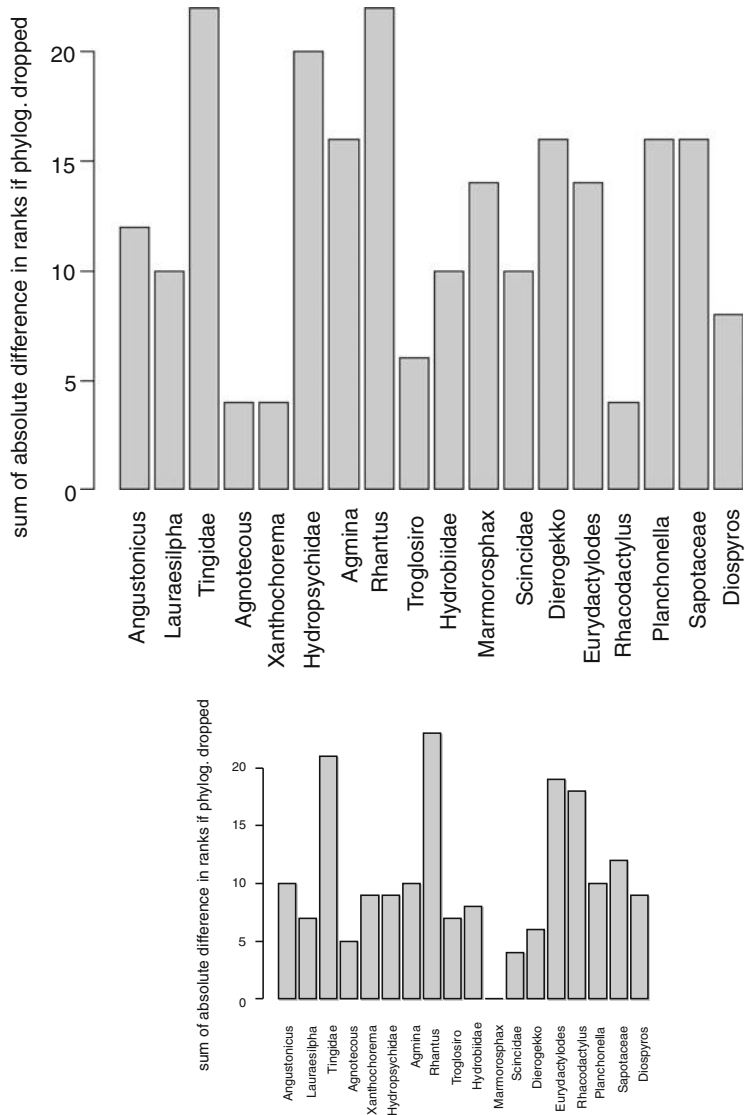
**B**

	Drop...																		Drop none
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
GrandSud	1	2	2	1	2	1	2	1	2	2	2	2	2	1	2	2	2	2	2
LaFoaCanala	2	1	1	2	1	2	1	2	1	1	1	1	1	2	1	1	1	1	1
RiviereBleue	8	8	7	9	9	8	8	8	9	8	8	8	8	12	11	8	12	9	9
MtsKoghis	7	9	10	8	10	10	7	9	8	9	7	11	7	8	8	7	7	8	8
MtsDzumac	4	6	3	6	5	9	9	3	3	6	10	6	5	5	5	3	4	7	5
GrandNord	6	7	6	7	7	5	5	6	7	10	6	5	13	7	7	5	9	5	7
MtAoupinie	3	4	4	3	3	3	3	7	4	3	3	3	4	3	3	4	5	3	3
MtMandjelia	5	3	5	4	4	4	4	5	5	4	4	4	3	4	4	6	3	4	4
FarinoUnio	9	5	11	5	6	7	6	4	6	5	5	7	6	6	6	10	6	6	6
MtPanie	13	13	13	12	12	12	15	10	12	12	12	12	11	11	12	12	11	12	12
AteouTchingou	10	11	9	10	8	6	10	11	11	7	11	9	9	9	9	9	8	11	10
MtHumboldt	12	12	12	13	13	11	11	15	13	13	13	13	12	14	13	13	13	13	13
MtKaala	15	15	14	15	15	15	14	14	15	15	16	15	14	13	15	14	14	14	15
MtMou	11	10	8	11	11	13	12	12	10	11	9	10	10	10	10	11	10	10	11
NinguaForetSailles	14	14	15	14	14	14	13	13	14	14	14	14	15	16	14	15	15	15	14
ColRoussettes	16	16	16	16	16	16	16	16	16	16	15	16	16	15	16	16	16	16	16

Values in red and bold are ‘real’ drops, i.e. the dropped phylogeny was indeed present at the site. A: ranks based on standardised values. B: ranks based on not standardised values. For a key to the phylogenies see caption Table 1



**Fig. 6** Sum over all sites of absolute differences in site ranks (based on *Ws sum*) if the phylogeny (x axis) is dropped. The main figures shows the values standardised (= divided) by the number of phylogenies present at the site when a phylogeny is dropped. The small figure at the bottom shows the nonstandardised values



**Fig. 7** Sum over all sites of absolute differences in site ranks (based on *Ws ranks*) if the phylogeny (x axis) is dropped. The main figure shows the values standardised (= divided) by the number of phylogenies present at the site when a phylogeny is dropped. The small figure at the bottom shows the non-standardised values

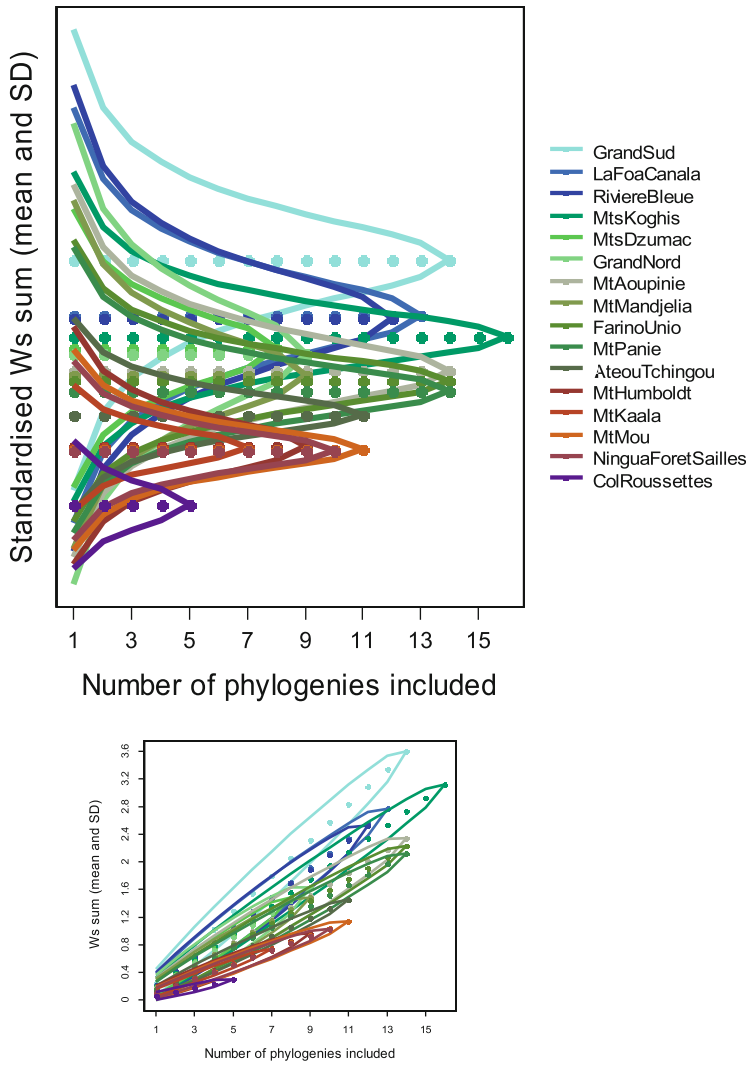
### ***Resampling Multiple Phylogenies: How Stable Are the Results?***

This resampling procedure is based on all possible combinations of phylogenies present at a site (from 1 to N, where N is the number of phylogenies with species in the site). Although there is (a) considerable overlap in the relative evolutionary divergence of sites in this resampling scheme and (b) the standard deviations are high, there are still some differences that emerge. For instance, when only 50 % of the phylogenies are used in the resampling ( $n=9$ ), the standard deviations of the top scoring sites do not overlap with those from the least phylogenetically diverse sites (Figs. 8 and 9). Thus with only nine phylogenies one can separate the four top scoring sites from the six least phylogenetically diverse ones when using *Ws sum*; and the two top sites and three bottom sites when *Ws ranks* are employed.

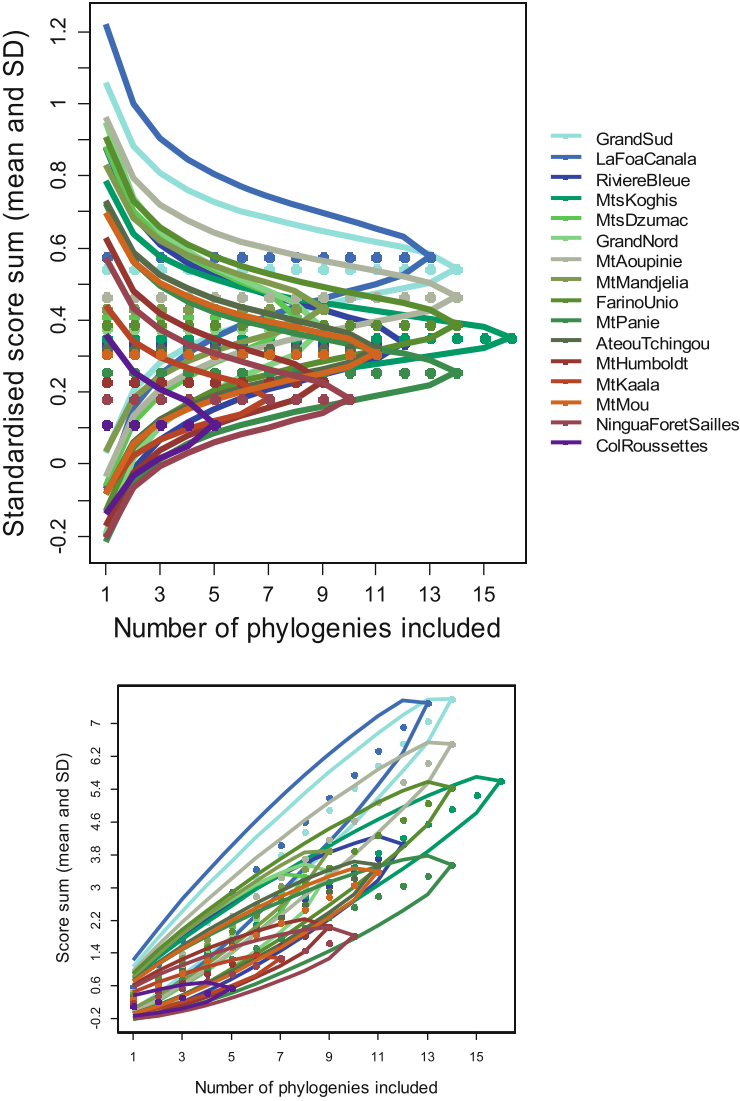
### ***Consideration of Individual Sites***

When the data set is evaluated using *Ws sums*, the site harbouring the greatest phylogenetic divergence is Grand Sud, and to a lesser degree La Foa Canala and Riviere Bleue. Grand Sud never drops in rank when individual phylogenies are dropped, and La Foa Canala and Riviere Bleue never below the 6th rank. The lower bound *Ws* (mean – SD) for Grand Sud still ranks 4th compared to the mean value of all others sites when all possible combinations of phylogenies are rarified to the smallest number of phylogenies present at any one site ( $n=5$ ) (the lower bound *Ws* for La Foa Canala and Riviere Bleue drop to ranks 9 and 10). The lowest scoring site is Col des Roussettes, and low phylogenetic diversity is also found in Ningua Foret Sailles, Mt Mou, Mt Kaala, and Mt Humboldt. These sites never move above the 12th rank when individual phylogenies are dropped, and their upper bound (mean + SD) ranks in the lowest two thirds compared to the mean value of all other sites when all possible combinations of phylogenies are rarified to the smallest number of phylogenies present at any one site.

When the same data set is evaluated using *Ws ranks* (summing the scores for the first and second most divergent species for each phylogeny), the sites harbouring the greatest phylogenetic divergence are also La Foa Canala and Grand Sud. These sites never drop below the 3rd rank when individual phylogenies are dropped, and their lower bound (mean – SD) still ranks in the upper half compared to the mean value of all others sites when all possible combinations of phylogenies are rarified to the smallest number of phylogenies present at any one site ( $n=5$ ). The lowest scoring sites are Col des Roussettes, Mt Kaala and Ningua Forest Sailles. These sites never move above the 13th rank when individual phylogenies are dropped, and their upper bound (mean + SD) ranks in the lowest quarter compared to the mean value of all other sites when all possible combinations of phylogenies are rarified to the smallest number of phylogenies present at any one site.



**Fig. 8** Mean and standard deviation for the sites' summed Ws values, resampled over all possible combinations of phylogenies present at the given site. The main figures shows the values standardised (= divided) by the number of phylogenies present at the site. The small figure at the bottom shows the non-standardised values



**Fig. 9** Mean and standard deviation for the sites’ summed site scores (*Ws ranks*), resampled over all possible combinations of phylogenies present at the given site. The main figures shows the values standardised (= divided) by the number of phylogenies present at the site. The small figure at the bottom shows the nonstandardised values



## Discussion

### *Methodological Considerations*

Even using metrics based on the  $W_s$ , there are several ways of evaluating evolutionary distinctiveness.  $W_s$  gives information on the total distribution of evolutionary divergence in the entire data set. An advantage of this index is that each phylogeny has its scores scaled between 0 and 1 and thus phylogenetic diversity can be represented by many species with small values (from phylogenies with many species), or few species with large values (from phylogenies with few species). However, this feature also introduces a limitation. If there is high beta-diversity (differentiation among sites) in each phylogeny (e.g. if each species only occurs at a single site), then small phylogenies have the potential to dominate the ranking of individual sites (as the most divergent species in small phylogenies have higher  $W_s$  values than the most divergent species in large phylogenies). In contrast, if there is low beta-diversity, then phylogenies with many species will have many species at individual sites, and thus will be able to ‘compete’ with the smaller phylogenies by having  $W_s$  totals that reflect the sum of several co-occurring species. In this latter case (low beta-diversity),  $W_s$  will be strongly correlated with overall species richness of a phylogeny.

Using the sum of 1st and 2nd ranks circumvents these problems. The power of this metric is that it gets at a simple question – where are the most divergent two species from each phylogeny, summed across sites and phylogenies. The downside is that, of course, it does not include information from species below the 1st and 2nd ranks. Thus it is purely targeted at examining the distribution of phylogenetically basal species, rather than the total sum of phylogenetic diversity. This needs to be borne in mind in its interpretation.

Another promising application of  $W_s$  ranks is in the detection of places of recent diversification. This can be achieved by focusing on the inverse of the most phylogenetic divergent species as used here, i.e., through awarding first and second prizes for the most and second most recent species of the phylogeny. Likewise, the methods of standardization and rarefaction can be very helpful for dealing with diverse sampling protocols and identifying the influence of different phylogenies to the ranking. Although evolutionary potential is a factor that requires genetic studies to be formally tackled (see Mace and Purvis 2008; the analysis of Grandcolas and Trewick in chapter “[What Is the Meaning of Extreme Phylogenetic Diversity? The Case of Phylogenetic Relict Species](#)”), the identification of sites that accumulate species with recent diversification is a first step to set out future study projects and monitoring strategies for testing this hypothesis. So, the possibility of identifying these sites should not be neglected.

Both of these metrics can then be adjusted to focus on micro-endemics, by using the measure  $W_{es}$  from Posadas et al. (2001) and the approach of 1st and 2nd ranks of  $W_s$  as developed here for the  $W_s$ .  $W_{es}$  is simply the  $W_s$  divided by the number of sites (or any measure of spatial distribution) the species occurs. The use of  $W_{es}$ ,



rather than  $W_s$  has the same issues with ‘sum’ versus 1st and 2nd ranks concepts as above. With  $W_s$  the  $W_s$  values are ‘diluted’ by being divided across each site that a species is recorded from and the main benefit is that sites will score more highly in proportion to the uniqueness of their species composition.

The resampling methods used here assure that ranking is not driven by a single or very small set of phylogenies, and the resampling with multiple drops indicates the tendency of sites remaining in similar ranking positions with the addition of phylogenies. To the best of our knowledge, this is the first time a set of phylogenetic studies are analysed this way (but see the proposition of Miranda-Esquivel, chapter “[Support in Area Prioritization Using Phylogenetic Information](#)”), and this seems to be a very promising way of integrating the problems of diversity of sampling effort.

### *Some Considerations About the Sites Prioritized*

The results of both analyses put in evidence that a few sites – Grand Sud, La Foa-Canala and Rivière Bleue are always ranking high. This clearly documents that these sites harbour remarkable species from a phylogenetic point of view. If ever these sites would be affected by disturbances, some more original evolutionary history would be lost in New Caledonia. How does it fit the conservation planning in New Caledonia? This planning is rather opportunistic, with the definition of small protected areas with very different status and varied protection level. Given the amazing level of micro-endemicity, every mountain or river harbours a conspicuous number of endemics so that any prioritization is difficult even among different protected areas. In every province, communication or action emphasis is often put on emblematical and large and supposedly virgin forested areas out of mining priorities, such as Massif du Panié in the North, or Rivière Bleue in the South. Our results do not adjust perfectly with this situation. The three high-ranking sites are not all emblematical and targetted areas and the protected areas concerned have different status. Grand Sud and Rivière Bleue areas are including natural reserves with high protection level but a large part of these areas are also situated outside the reserves, potentially putting at risk some populations of endemics. These risks are also increased because of the metalliferous soils derived from ultramafic rocks that are widespread in these southern areas and which are potentially places for nickel mining. La Foa-Canala area is another with less direct disturbances but with reserves with lower protection level. The reserve of Col d’Amieu is a place for forest logging and traditional seasonal bat hunting and is generally not targeted as an emblematical area.

Therefore, a recommendation based on our analysis of phylogenetic diversity should consider that conservation planning in New Caledonia is modified in two ways. The small natural parks in the South should become larger or connect with several new reserves, and the Reserve du Col d’Amieu should be carefully considered with improvement of the protection level.

## ***Future Perspectives***

In this work we focused in one method already adjusted to deal with prioritization of areas based on the evolutionary distinctiveness, the Ws (Posadas et al. 2001). The same procedure can be directly employed to any measure of evolutionary distinctiveness (ED), in which each species has a score related to its position in the phylogeny and the area ranks are assessed through the sum of the scores of the species occurring in it. So, it could be identically employed when using the EDGE or HEDGE measures, where ED is associated to threat status (Isaac et al. 2007; see also May-Collado et al. chapter “[Global Spatial Analyses of Phylogenetic Conservation Priorities for Aquatic Mammals](#)”), or in cases where ED is combined with geographical rarity, or with species abundance, as, for example, the AED from Cadotte and Davies (2010).

As shown by Faith et al. (2004) and Faith (chapter “[The PD Phylogenetic Diversity Framework: Linking Evolutionary History to Feature Diversity for Biodiversity Conservation](#)” this volume) PD could easily be used to assess site’s rank when using data from several phylogenies: in cases where phylogenies are based on different kinds of characters or method of analysis, PD can be employed on the simple basis of counting nodes. The great advantage is that PD (the sum of the minimum spanning path linking all the species in an area) is a group measure (see Hartman and Steel 2007) and takes in consideration the complementarity, which would result in avoiding redundancies. However, at the present state of knowledge the rarefaction as used here, or the standardization for number of phylogenies cannot be directly applied to group measures such as PD. As presented in the introduction of this chapter the rarefaction of PD is newly developed (Nipperess and Matsen 2013). Many solutions are designed in Nipperess’ (chapter “[The Rarefaction of Phylogenetic Diversity: Formulation, Extension and Application](#)”): the standardization of sampling effort; the calculation of phylogenetic evenness, phylogenetic beta diversity, and phylogenetic dispersion. So, an extension to the application of these solutions when using phylogenetic data from several phylogenies will complete this framework and provide more options about the measure to be employed.

Biodiversity conservation is a very complex issue, and conservation guidelines should take multiple variables in consideration. Ideally, the analysis should provide explicit information about the way each variable has been weighted and, as far as possible, a set of scenarios under different weights. In this perspective, complex frameworks for systematic conservation planning have been developed and are becoming to be employed more often. For example, the Zonation procedure (Moilanen 2007; Lehtomaki and Moilanen 2013) used by Arponen and Zupan (chapter “[Representing Hotspots of Evolutionary History in Systematic Conservation Planning for European Mammals](#)”), and the gap analysis (Ball and Possingham 2000) used in the study of Silvano et al. (chapter “[Priorities for Conservation of the Evolutionary History of Amphibians in the Cerrado](#)”). In these procedures, phylogenetic diversity is included as a weight along with other biological data like spe-

cies' distribution area, threat status, or some economic variables, such as the cost for conservation.

Although the results presented in this study highly stand by themselves, they can also be integrated in this kind of analysis as weights according to site's rank considering both Ws sums and Ws ranks amongst other variables. In this case, there is no doubt that the procedures conducted here will give a reliable picture of the phylogenetic distribution across this set of sites, and provide a better instrument to the conservation of the phylogenetic diversity.

To conclude, the analytical problems and need for the solutions outlined above will decrease as large-scale sequencing projects bring more directly comparable data together. However, until comprehensive and balanced sampling from common gene sets across taxa and sites are realized, the challenges of standardization, comparability and assessments of bias will remain relevant.

**Acknowledgements** This project was funded by the French Agency of Research, through the grant ANR Bioneocal to PG. We thank the authors of the phylogenies and in particular Ulf Swenson and Marianne Espeland, for giving access to their databases of species occurrences, and Hervé Jourdan (IRD Nouméa) for helping RP with the delimitation of the study areas. We also thank David Nipperess and Romain Nattier for their comments that helped to improve this manuscript.

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Balke M, Wewalka G, Alarie Y, Ribera I (2007) Molecular phylogeny of Pacific Island Colymbetinae: radiation of New Caledonian and Fijian species (Coleoptera, Dytiscidae). *Zool Scr* 36:173–200
- Ball I, Possingham HP (2000) MARXAN v1.8.2 – marine reserve design using spatially explicit annealing. University of Queensland, Brisbane
- Bauer AM (1990) Phylogenetic systematics and biogeography of the Carphodactylini (Reptilia: Gekkonidae). *Bonn Zool Monogr* 30:1–220
- Bauer AM, Jackman T, Sadlier RA, Whitaker AH (2006) A Revision of the *Bavayia validiclavis* group (Squamata: Gekkota: Diplodactylidae), a clade of New Caledonian Geckos exhibiting microendemism. *Proc Calif Acad Sci* 57:503–547
- Bauer AM, Jackman T, Sadlier RA, Whitaker AH (2009) In: Grandcolas P (ed) *Zoologia Neocaledonica 7, Systematics and biodiversity in New Caledonia*, 13–36 (Mémoires du Muséum National d'Histoire Naturelle, 198)
- Beauvais M-L, Coléno A, Jourdan H (eds) (2006) *Les espèces envahissantes dans l'archipel néo-calédonien – Un risque environnemental et économique majeur, vol 1. expertise collégiale*. IRD editions, Paris
- Bouchet P, Jaffré T, Veillon J-M (1995) Plant extinction in New Caledonia: protection of sclerophyll forest urgently needed. *Biodivers Conserv* 4:415–428

- Cadotte MW, Davies TJ (2010) Rarest of the rare: advances in combining evolutionary distinctiveness and scarcity to inform conservation at biogeographical scales. *Divers Distrib* 16(3):376–385
- Desutter-Grandcolas L, Robillard T (2006) Phylogenetic systematics and evolution of Agnateous in New Caledonia (Orthoptera: Grylloidea, Eneopteridae). *Syst Biol* 31:65–92
- Duangjai S, Samuel R, Munzinger J, Forest F, Wallnöfer B, Barfuss MH, Fischer G, Chase MW (2009) A multi-locus plastid phylogenetic analysis of the pantropical genus *Diospyros* (Ebenaceae), with an emphasis on the radiation and biogeographic origins of the New Caledonian endemic species. *Mol Phylogenet Evol* 52:602–620
- Espeland M, Johanson KA (2010a) The diversity and radiation of the largest monophyletic animal group on New Caledonia (Trichoptera: Ecnomidae: *Agmina*). *J Evol Biol* 23:2112–2122
- Espeland M, Johanson KA (2010b) The effect of environmental diversification on species diversification in New Caledonian caddisflies (Insecta: Trichoptera: Hydropsychidae). *J Biogeogr* 37:879–890
- Espeland M, Johanson KA, Hovmöller R (2008) Early *Xanthochorema* (Trichoptera, Insecta) radiations in New Caledonia originated on ultrabasic rocks. *Mol Phylogenet Evol* 48:904–917
- Faith DP, Reid CAM, Hunter (2004) Integrating phylogenetic diversity, complementarity, and endemism for conservation assessment. *Conserv Biol* 18(1):255–261
- Good DA, Bauer AM, Sadler RA (1997) Allozyme evidence for the phylogeny of the giant New Caledonian geckos (Squamata: Diplodactylidae: *Rhacodactylus*), with comments on the status of *R. leachianus henkeli*. *Aust J Zool* 45:317–330
- Grandcolas P, Murienné J, Robillard T, Desutter-Grandcolas L, Jourdan H, Guilbert E, Deharveng L (2008) New Caledonia: a very old Darwinian island? *Philos Trans R Soc Lond B* 363:3309–3317
- Haase M, Bouchet P (1998) Radiation of crenobiontic gastropods on an ancient continental island: the Hemistomia-clade in New Caledonia (Gastropoda: Hydrobiidae). *Hydrobiologia* 367:43–129
- Hartmann K, Steel MA (2007) Phylogenetic diversity: from combinatorics to ecology. In: Gascuel O, Steel MA (eds) *Reconstructing evolution: new mathematical and computational advances*. Oxford University Press, Oxford
- Isaac NJB, Turvey ST, Collen B et al (2007) Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PLoS One* 2: e296. doi:[10.1371/journal.pone.0000296](https://doi.org/10.1371/journal.pone.0000296)
- Kier G, Kreft H, Lee TM, Jetz W, Ibsch PL, Nowicki C, Mutke J (2009) A global assessment of endemism and species richness across island and mainland regions. *Proc Natl Acad Sci U S A* 106(23):9322–9327
- Lehman SM (2006) Conservation biology of Malagasy Strepsirrhines: a phylogenetic approach. *Am J Phys Anthropol* 130:238–253
- Lehtomäki J, Moilanen A (2013) Methods and workflow for spatial conservation prioritization using Zonation. *Environ Model Softw* 47:128–137. doi:[10.1016/j.envsoft.2013.05.001](https://doi.org/10.1016/j.envsoft.2013.05.001)
- López-Osorio F, Miranda Esquivel DR (2010) A phylogenetic approach to conserving Amazonian biodiversity. *Conserv Biol* 24(5):1359–1366
- Mace GM, Purvis A (2008) Evolutionary biology and practical conservation: bridging a widening gap. *Mol Ecol* 17(1):9–19
- McGoogan K, Kivell T, Hutchison M, Young H, Blanchard S, Keeth M, Lehman SM (2007) Phylogenetic diversity and the conservation biogeography of African primates. *J Biogeogr* 34(11):1962–1974
- Moilanen A (2007) Landscape Zonation, benefit functions and target-based planning: unifying reserve selection strategies. *Biol Conserv* 134(4):571–579. doi:[10.1016/j.biocon.2006.09.008](https://doi.org/10.1016/j.biocon.2006.09.008)
- Munzinger J, Swenson U (2009) Three new species of *Planchonella* (Sapotaceae) with a dichotomous and an online key to the genus in New Caledonia. *Adansonia* 31:175–189
- Murienné J (2006) Origine de la biodiversité en Nouvelle-Calédonie: analyse phylogénétique de l'endémisme chez les Insectes Dictyoptères, Université Pierre et Marie Curie – Paris 6

- Murienne J, Grandcolas P, Piulachs MD, Bellés X, D'Haese C, Legendre F, Pellens R, Guilbert E (2005) Evolution on a shaky piece of Gondwana: is local endemism recent in New Caledonia? *Cladistics* 21:2–7
- Murienne J, Pellens R, Budinoff RB, Wheeler W, Grandcolas P (2008) Phylogenetic analysis of the endemic New Caledonian cockroach *Lauraesilpha*. Testing competing hypothesis of diversification. *Cladistics* 24:802–812
- Murienne J, Guilbert E, Grandcolas P (2009) Species' diversity in the New Caledonian endemic genera *Cephalidiosus* and *Nobarnus* (Insecta: Heteroptera: Tingidae), an approach using phylogeny and species' distribution modelling. *Biol J Linn Soc* 97:177–184
- Myers N, Mittermeier RA, Mittermeier CG, Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* 403:853–858
- Nattier R, Grandcolas P, Elias M, Desutter-Grandcolas L, Jourdan H, Couloux A, Robillard T (2012) Secondary sympatry caused by range expansion informs on the dynamics of microendemism in a biodiversity hotspot. *PLoS ONE* 7(11):e48047
- Nattier R, Grandcolas P, Pellens R, Jourdan H, Couloux A, Poulain S, Robillard T (2013) Climate and soil type together explain the distribution of microendemic species in a biodiversity hotspot. *PLoS ONE* 8(12):e80811
- Nipperess DA, Matsen FA (2013) The mean and variance of phylogenetic diversity under rarefaction. *Methods Ecol Evol* 4(6):566–572
- Pascal M, Richer de Forges B, Le Guyader H, Simberloff D (2008) Mining and other threats to the New Caledonia biodiversity hotspot. *Conserv Biol* 22(2):498–499
- Pellens R, Grandcolas P (2010) Conservation and management of the biodiversity in a hotspot characterized by short range endemism and rarity: the challenge of New Caledonia. In: Rescigno V, Maletta S (eds) *Biodiversity hotspots*. Nova Publishers, New York, pp 139–151
- Pillon Y, Munzinger J, Amir H, Lebrun M (2010) Ultramafic soils and species sorting in the flora of New Caledonia. *J Ecol* 98:1108–1116. doi:[10.1111/j.1365-2745.2010.01689.x](https://doi.org/10.1111/j.1365-2745.2010.01689.x)
- Posadas P, Miranda Esquivel DR, Crisci JV (2001) Using phylogenetic diversity measures to set priorities in conservation: an example from Southern South America. *Conserv Biol* 15(5):1325–1334
- Posadas P, Miranda Esquivel DR, Crisci JV (2004) On words, tests, and applications: reply to Faith et al. *Conserv Biol* 18(1):262–266
- Redding DW, Hartmann K, Mimoto A, Bokal D, DeVosb M, Mooers AO (2008) Evolutionarily distinctive species often capture more phylogenetic diversity than expected. *J Theor Biol* 251:606–615
- Rodrigues ASL, Brooks T, Gaston KJ (2005) Integrating the phylogenetic diversity in the selection of priority areas for conservation: does it make a difference? In: Purvis A, Gittleman JL, Brooks T (eds) *Phylogeny and conservation*, *Conserv. Biol.* 8. Cambridge University Press, London, pp 101–119
- Sadlier RA, Smith SA, Bauer AM, Whitaker AH (2004) A new genus and species of live-bearing Scincid lizard (Reptilia: Scincidae) from New Caledonia. *J Herpetol* 38:320–330
- Sadlier RA, Smith SA, Bauer AM, Whitaker AH (2009) In: Grandcolas P (ed) *Zoologia Neocaledonica 7, Systematics and biodiversity in New Caledonia*, 247–263 (*Mémoires du Muséum National d'Histoire Naturelle*, 198)
- Sharma P, Giribet G (2009) A relict in New Caledonia: phylogenetic relationships of the family Troglodironidae (Opiliones: Cyphophthalmi). *Cladistics* 25:1–16
- Swenson U, Munzinger J (2009) Revision of *Pycnandra* subgenus *Pycnandra* (Sapotaceae), a genus endemic to New Caledonia. *Aust Syst Bot* 22:437–465
- Swenson U, Munzinger J (2010a) Revision of *Pycnandra* subgenus *Achradotypus* (Sapotaceae) with five new species from New Caledonia. *Aust Syst Bot* 23:185–216
- Swenson U, Munzinger J (2010b) Revision of *Pycnandra* subgenus *Sebertia* (Sapotaceae) and a generic key to the family in New Caledonia. *Adansonia* 32
- Swenson U, Munzinger J (2010c) Taxonomic revision of *Pycnandra* subgenus *Trouettia* (Sapotaceae) with six new species from New Caledonia. *Aust Syst Bot* 23:333–370

- Swenson U, Munzinger J, Bartish IV (2007) Molecular phylogeny of *Planchonella* (Sapotaceae) and eight new species from New Caledonia. *Taxon* 56:329–354
- Swenson U, Lowry PP II, Munzinger J, Rydin C, Bartish IV (2008) Phylogeny and generic limits in the *Niemeyera* complex of New Caledonian Sapotaceae: evidence of multiple origins of the anisomerous flower. *Mol Phylogenet Evol* 49:909–929
- Vane-Wright RI, Humphries CJ, Williams PH (1991) What to protect?-systematics and the agony of choice. *Biol Conserv* 55(3):235–254
- Wulff AS, Hollingsworth PM, Ahrends A, Jaffre T, Veillon JM, L'Huillier L, Fogliani B (2013) Conservation priorities in a biodiversity hotspot: analysis of narrow endemic plant species in New Caledonia. *PLoS One* 8(9):e73371

## **Part III**

# **Applications**

# Representing Hotspots of Evolutionary History in Systematic Conservation Planning for European Mammals

Anni Arponen and Laure Zupan

**Abstract** Systematic conservation planning deals with cost-effective allocation of conservation funds. There are diverse ways in which evolutionary history could be included in prioritization, but here we considered it at the local scale, valuing higher the locations where the local community has high phylogenetic diversity, while still aiming at maximizing overall species representation. We conducted the prioritization with the Zonation software for spatial conservation planning.

We prioritized areas for conservation in Europe using distribution data and phylogenies for 275 mammal species. We prioritized areas in Europe for conserving hotspots of evolutionary history. For comparison we made analyses with species occurrences alone. Analyses were done for the whole region and for each country separately. We explored the impacts of tree uncertainty, and analyzed how well existing protected areas performed with respect to Zonation priorities.

Our findings indicate that some hotspots of evolutionary history are missed by species-based prioritization, unless specifically accounted for. Uncertainty in spatial priorities caused by variation in phylogenetic tree structure was a minor concern for prioritization. Protected areas did not perform well when assessed against the Zonation priorities for species or for phylogenetic diversity, although highest national scale priorities had almost twice as much area protected as the overall average.

We emphasize that the chosen goals and analysis setups have strong impacts on spatial priorities and therefore care must be taken in defining them appropriately. But regardless of setups, the gap between the current conservation efforts and spatial prioritization outcomes is typically greater than the difference between including and excluding phylogenetic diversity. Therefore the focus should be on increasing the role of spatial analyses in practical conservation, but whenever

---

A. Arponen (✉)

Department of Biosciences, Faculty of Biological and Environmental Sciences, University of Helsinki, P.O. Box 65 (Viikinkaari 1), FI-00014 Helsinki, Finland  
e-mail: [anni.arponen@helsinki.fi](mailto:anni.arponen@helsinki.fi)

L. Zupan

Laboratoire d'Ecologie Alpine, UMR-CNRS 5553, Université J. Fourier, Grenoble I, 53, 38041 Grenoble cedex 9, France  
e-mail: [laure.zupan@gmail.com](mailto:laure.zupan@gmail.com)



feasible, also including evolutionary history in the analyses, because evolutionary history is not always well represented by targeting species for conservation.

**Keywords** Phylogenetic diversity • Quadratic entropy • Spatial prioritization • Zonation

## Introduction

**Systematic Conservation Planning** Protected areas around the world have typically been established in areas of low competing interests, which is not ideal from the perspective of biodiversity conservation (Pressey et al. 1993). Such biased allocation may even lead to existing protected areas performing worse than randomly chosen areas in representing diversity (Ferrier 2002). The realization that conservation would benefit from cost-effective practices led to the development of the field of Systematic conservation planning (Box 1, Margules and Pressey 2000; Margules and Sarkar 2007). More than 20 years of development have led to the integration of numerous aspects to the approach adding to its realism. In particular, the spatial prioritization for assessing the existing conservation areas and selecting new ones have become comprehensive and efficient, and nowadays they also provide more user-friendly graphical user interfaces, which has facilitated their broad use for practical conservation planning purposes (Ball et al. 2009; Moilanen et al. 2009).

**Evolutionary History in Conservation** Phylogenetic diversity or species originality are often mentioned as important for conservation (Rosauer and Mooers 2013; Winter et al. 2013), and the history of such discussion goes back already a few decades (Vane-Wright et al. 1991; Faith 1992). Evolutionary history is often quantified in community ecology for the purpose of understanding the diversity of current species distributions (Davies and Buckley 2011; Fritz and Rahbek 2012) or the potential functioning of ecosystems (Cadotte et al. 2012), whereas applications to conservation have remained limited.

### Box 1

The process of systematic conservation planning as described by Margules and Pressey (2000).

1. Compile data on the biodiversity of the planning region
2. Identify conservation goals for the planning region
3. Review existing conservation areas
4. Select additional conservation areas
5. Implement conservation actions
6. Maintain the required values of conservation areas

Numerous indices have been developed to measure the originality of species (Vane-Wright et al. 1991; Pavoine et al. 2005a; Isaac et al. 2007), or phylogenetic diversity (Faith 1992; Schweiger et al. 2008; Pavoine and Bonsall 2011; Faith chapter “[The PD Phylogenetic Diversity Framework: Linking Evolutionary History to Feature Diversity for Biodiversity Conservation](#)”). The former measures assign a value for each species based on their dissimilarity from other species, whereas the latter look at an assemblage of species as a whole.

Both types can be used in spatial conservation prioritization (Arponen 2012). Originality can be used for weighting species differently, whereas diversity indices can be used at different scales: either for measuring the diversity of all species across a network of protected areas, or for preferentially selecting areas with high local, alpha-level diversity of the community. Their use has been rare in published studies of spatial conservation prioritization. Arponen et al. (2005) used species weights based on species originality in conservation prioritization for plants in Finnish herb-rich forests. There are also some examples of considering assemblage-level phylogenetic diversity across a network of sites: The “Phylogenetic Diversity” of Faith (1992) has been used for conservation prioritization with birds (Rodrigues and Gaston 2002) and plants (Forest et al. 2007) in South Africa, as well as in a global analysis for mammals (Rodrigues et al. 2011). Instead of spatial prioritization of areas for protection, evolutionary history has been considered much more commonly in other kinds of conservation contexts (reviewed in Arponen 2012), such as creating priority lists of species for conservation. For example, Isaac et al. (2007) introduced the “Evolutionary distinctiveness” measure for species and used it in combination with extinction risk data to assign priorities for species in the EDGE program (see also May-Collado et al. chapter “[Global Spatial Analyses of Phylogenetic Conservation Priorities for Aquatic Mammals](#)”; Schnell and Safi chapter “[Metapopulation Capacity Meets Evolutionary Distinctness: Spatial Fragmentation Complements Phylogenetic Rarity in Prioritization](#)”).

To our knowledge, phylogenetic diversity has not been used at the scale of local communities in spatial conservation prioritization. The use of alpha-level phylogenetic diversity is based on the assumption that it would correlate with ecological processes better than species richness of the community (Forest et al. 2007), and therefore work as an indicator for functional diversity when species traits data are missing. This is based on the idea that phylogenetically distinct species are likely to be functionally different (Cadotte et al. 2008), although this assumption has also been challenged (Mouquet et al. 2012). For this purpose, phylogenetic diversity indices that account for species abundances (Chao et al. 2010; Chao et al. chapter “[Phylogenetic Diversity Measures and Their Decomposition: A Framework Based on Hill Numbers](#)”) might be more suitable than the ones that consider only presences and absences of species (such as Faith 1992): from the perspective of ecosystem function, viable populations and sparse individuals of a species should not be considered equally important.

**Case Study on European Mammals** Mammals are a fairly well known group of species regarding their ecology, distributions as well as phylogeny. Nevertheless, their phylogenies are not fully resolved, but contain polytomies. Resolving the

polytomies randomly results in variation among different trees, but having good spatial distribution data provides a good opportunity for investigating the influence of such uncertainty on spatial conservation prioritization. Mammals are also considered to be of high conservation interest due to their public appeal (Smith et al. 2012). They were the first focal taxon of the EDGE programme (Isaac et al. 2007), which was a pioneering endeavor to bring highly threatened and evolutionarily unique species to the limelight and to improve their conservation.

We conducted spatial prioritizations for European mammal conservation with the Zonation conservation planning software. We compared traditional, species based prioritization to one where alpha-level phylogenetic diversity, measured as the equivalent number of Rao's quadratic entropy, was allowed to influence site value through using its inverse as cost in the analyses. Because a continental scale analysis may not be politically feasible, we repeated both analyses at national scales, where Zonation performs identical prioritization but for each country separately. For mammals there is still some uncertainty related to the structure of the phylogeny. We acquired 100 different trees and ran Zonation analyses for each of them, comparing the similarities of outcomes to each other. We analyzed the trade-offs between species representation and the equivalent number of Rao's quadratic entropy in the solutions. Finally, we analyzed the performance of the current protected area network in representing hotspots of evolutionary history for mammals, as well as in representing species, both at the European and at national scales.

## Material and Methods

**European Mammal Distributions** We used data on the spatial distribution of European terrestrial mammals described in Maiorano et al. (2013). The primary data were extents of occurrence (EOOs) of the species occurring in Europe and Turkey obtained from the Global Mammal Assessment (<http://www.iucnredlist.org/initiatives/mammals>; accessed 15 August 2013 (IUCN 2012)). To refine EOOs and remove potential false presences, habitat requirements were used in an expert-based modelling approach. More specifically, for each species, habitat requirement was defined by experts (G. Amori, D. Russo and L. Boitani) and published literature (see Maiorano et al. 2013 for the full list of references) based on three environmental variables: land cover, elevation and distance to water. For each species, data collected were used to assign a suitability score (0, *unsuitable*; 1, *secondary habitat* and 2, *primary habitat*) to each of the 46 GlobCover land-use/land-cover classes. Elevation and distance to water were then combined to the habitat suitability score to refine the available EOOs and obtain current distribution with a cell size of 300 m resolution. The models were validated with help of field data (see Maiorano et al. 2013 for more details). From these 288 species we used 275 for which phylogenies were available.

As running the phylogenetic analyses and Zonation prioritization at 300 m resolution would have been too demanding for the equipment available at the time, we scaled up the species distributions following a regular grid of 10'. As a value for each 10' cell, we kept the percentage of 300 m cells considered as either 1 (primary habitat) or 2 (secondary habitat), and we refer to this value as “the proportion of suitable area” hereafter. For aesthetic reasons, all the maps presented hereafter have been projected using the Lambert conformal conic projection (UTM zone 34).

**Mammal Phylogenies** Phylogenetic data for mammals were based on the super-tree of Bininda-Emonds et al. (2007) updated by Fritz et al. (2009). We used 100 fully resolved phylogenetic trees, where polytomies were randomly resolved applying a birth-death model to simulate branch lengths (Kuhn et al. 2011).

**Protected Areas** We used the WDPA dataset on protected areas (UNEP 2010) categories I-IV (I: Strict nature reserve or wilderness area, II: National park, III: Natural monument or feature and IV: Habitat/Species management area) excluding the categories that are generally considered less beneficial for biodiversity conservation (categories V and VI), and areas where the category was either ‘not reported’ or ‘not applicable’. We used the proportions of area protected in each cell for our analyses of overlap of Zonation priorities with protected areas. WDPA data are polygons. As Zonation operates with raster data, we transformed the polygons into a raster, following the same grid as the species distribution data (10' cells regular grid). To do so, we overlapped the polygons on the grid and retained the proportion of area protected in each grid cell.

**Measuring Phylogenetic Diversity** To measure the phylogenetic diversity at each cell, we used the Rao's quadratic entropy (Rao 1982), an index of alpha-diversity, which is extended to account for the pair-wise dissimilarities of species:

$$QE = \sum_{i=1}^S \sum_{j=1}^S d_{ij} p_{ic} p_{jc}$$

$d_{ij}$  is derived from the ultrametric phylogenetic tree (Pavoine et al. 2005b) and corresponds to the phylogenetic dissimilarity between each pair of species  $i$  and  $j$ .  $p_i$  and  $p_j$  are the respective proportions of suitable habitat for the species  $i$  and  $j$  available in the 10' pixel  $c$ . It is now recognized in the literature that the values of most of diversity measures (like the Rao's quadratic entropy) do not behave intuitively because they do not satisfy the “replication principle” (Jost 2007; de Bello et al. 2010; Chao et al. 2010; Leinster and Cobbold 2012; Chao et al. chapter “Phylogenetic Diversity Measures and Their Decomposition: A Framework Based on Hill Numbers”). The replication principle (or “doubling property”) states that if we pool two equally diverse and equally large groups with no shared species, the total diversity should be two times the diversity of a single group (Chao et al. 2010; Chao et al. see their Fig. 2 in chapter “Phylogenetic Diversity Measures and Their Decomposition: A Framework Based on Hill Numbers”). To make the Rao's

quadratic entropy behave this way, we need to transform it into an equivalent number through a simple algebra step ( $1/(1-QE)$ , Jost 2007). The outcome is a raster layer with the value of QE in equivalent number for each of the 10' pixels with the same spatial extent and resolution as the mammal distribution data.

**The Zonation Approach** Zonation is a spatial prioritization software meant to be used as a decision support tool (Moilanen et al. 2009). While other approaches typically select a fraction of the landscape according to a pre-determined target, e.g. 10 % of species distributions, or maximize what is achieved with a pre-determined budget, Zonation instead ranks all cells in the entire landscape in the order of conservation value. A Zonation solution can be used to identify any best (or worst) fraction of the landscape.

The ranking is based on the evaluation of range size normalized richness of biodiversity features in each cell (Moilanen et al. 2005, 2011). In plain words, this means that features (e.g. species) with broad distributions contribute very little to the conservation value of a single cell, whereas narrowly distributed species substantially increase the conservation value of the cells they occupy. At every iteration (removal of one cell) Zonation recalculates the conservation value for the remaining cells based on the remaining feature distributions, which become smaller with each iteration. Thus, Zonation removes first cells with few, broadly distributed features, and during the ranking these features become rarer and rarer in the remaining landscape. As an outcome, the remaining highest priority fraction of the landscape will contain the cells with high species richness and narrow endemics.

Zonation provides two options as cell-removal rules that determine how the marginal value of a cell is calculated (Moilanen et al. 2005; Moilanen 2007). The additive benefit function approach allows for more flexible trade-offs to occur between features, because it considers cell value as the sum over benefit functions of representation of the features in the cell. This means that narrowly distributed species in species poor (or expensive) cells may be traded off against species rich cells. We chose to use the Core-area cell removal rule, which defines the cell value based on the most valuable occurrence over all species in the cell. This means that if a cell contains a large fraction of the range even for only one species, it will get high value, regardless of the species richness in the cell. This way the core areas of all species' ranges are retained in the highest priority fraction of the landscape. As species distribution data, we used the raster layers of proportion of suitable habitat per cell for each species, as described above in the section "European mammal distributions".

Even though Zonation does not consider phylogenetic data by default, it offers also options for accounting for evolutionary history in the prioritization. For example, species could be weighted based on their evolutionary distinctiveness either globally, or with different region-specific weights (Moilanen and Arponen 2011). Alternatively, locations can be weighted based on the phylogenetic diversity of the local community. In this case study we focus on the latter approach. Technically this happens through defining a "cost layer" as inversely proportional to the diversity. This way a cell with one-fifth of the phylogenetic diversity of another cell is

considered five times as costly to protect, lowering its position in the Zonation ranking. The cost layer can be scaled differently according to how much importance is given to phylogenetic diversity. The equivalent numbers of Rao's quadratic entropy values went from ca. 1 to 7, and the direct inverse was used in our "medium weighting" (that is, cost goes from 0.14 to 1), and this scale was halved ("low weights", 0.28–1) and doubled ("high weights", 0.07–1) to test for sensitivity to this parameter (see Fig. 1 for analysis setups).

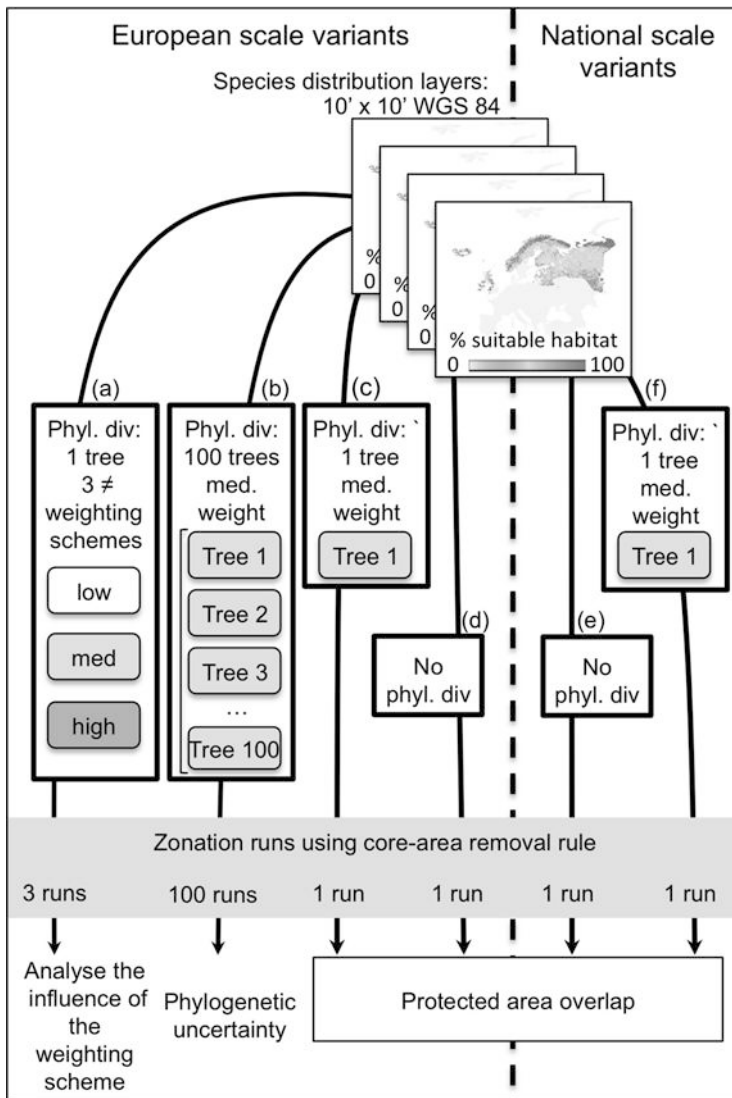
The latitudinal gradients in species richness and range sizes cause the spatial priorities in analyses at any scale to be concentrated in the more species rich lower latitude areas (Eklund et al. 2011; Moilanen et al. 2013). Even though cost-effective from the perspective of species conservation, focusing conservation efforts into these regions only would be very difficult for many reasons (see section "[Discussion and Conclusions](#)"). Therefore we also performed an analysis where countries were considered as independent administrative units, each aiming to conserve the diversity within their borders. This is implemented through the Administrative units analysis in Zonation (Moilanen and Arponen 2011). The analysis would allow for a compromise solution between purely European-scale and purely national-scale analyses, but for our analytical purposes, we chose the extreme cases only. A national-scale prioritization provides an interesting reference for comparison to protected areas. We did this for one tree only. Thus, we ended up with four main Zonation solutions to assess protected area performance regarding the representation of species and phylogenetic diversity at both European and national scales (Fig. 1).

## *Case Study Setup*

### **Results**

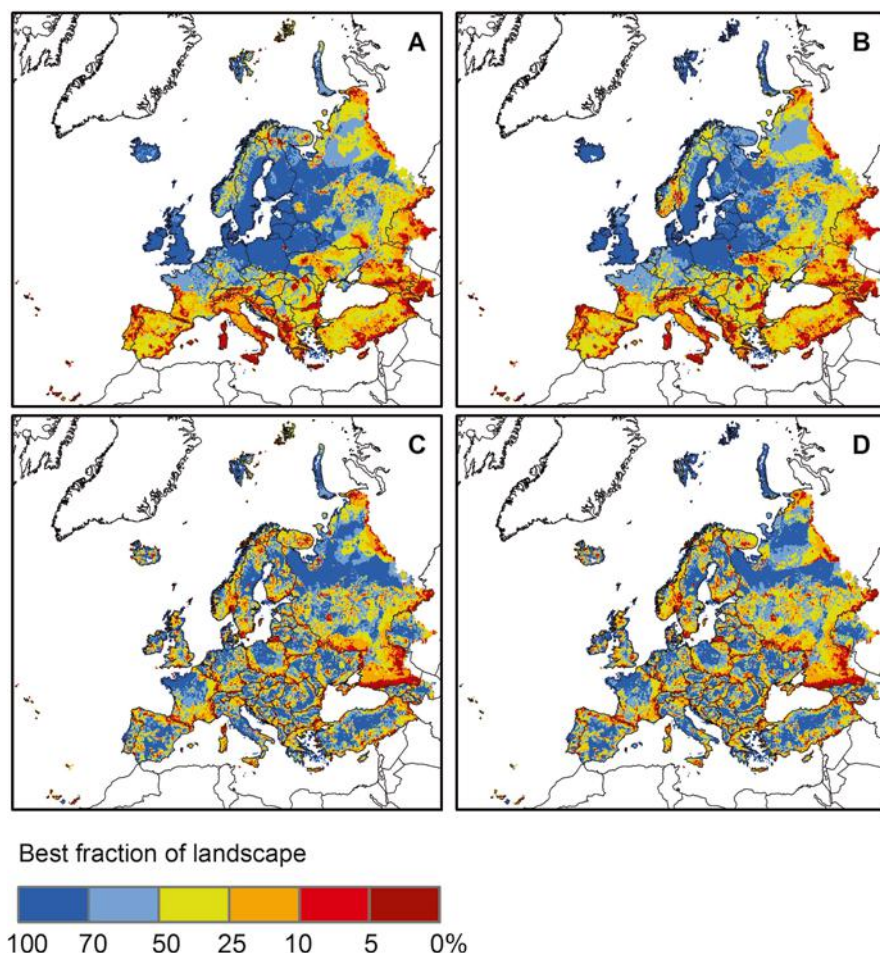
Spatial priorities in the European analyses were strongly concentrated around the southern parts as well as eastern border of the study region (Fig. 2a, b). Spatial priorities between the basic Core-area prioritization and the variants where phylogenetic diversity as the equivalent number of Rao's QE was included are extremely similar in some regions, but contain some rather dramatic differences in specific, especially northern parts of Europe (Fig. 2a, b). Spearman rank correlations between the rank values in the basic Core-area solution and the three weighting variants of the equivalent number of Rao's QE were 0.93, 0.91 and 0.89, for the low, medium and high weight scales, respectively.

We repeated the basic and phylogenetic diversity weighted analyses at the national scale, where Zonation performed the prioritization separately for each country (Fig. 2c, d). Here the priorities were forced to be evenly distributed among the countries, such that e.g. the best 10 % of the landscape consisted of the best 10 % in each country. Such priorities are much more scattered across Europe, and concentrated around country borders.



**Fig. 1** Diagram representing the flow of our analyses. The *left part* of the figure illustrates the analyses we ran at the European scale while at the right of the *dashed line* are the analyses conducted at the national level. In all analyses, the data input for the species were the proportion of suitable habitat (1 raster layer per species). We first **(a)** tested three different weightings for the phylogenetic diversity measured as the equivalent number of Rao's quadratic entropy (low, medium and high weighting, see main text) to assess whether this was influencing the prioritization results. In a second step **(b)** we ran Zonation 100 times using cost layers corresponding to the 100 different phylogenetic trees. We followed this procedure to evaluate the influence of the tree structure variation on the prioritization results. The remaining analyses were dedicated for the evaluation of the current protected areas network. We used only one cost-layer (corresponding to the equivalent number of Rao's quadratic entropy extracted from tree 1 and a medium weighting) to evaluate the protected area network at the **(c)** European scale and **(f)** the national scale. Finally we ran Zonation without any phylogenetic diversity data to assess the representation of species within the protected areas network at **(d)** European scale and **(e)** national scale. Abbreviations: *med* medium, *phyl. div* phylogenetic diversity measured as the equivalent number of Rao's quadratic entropy

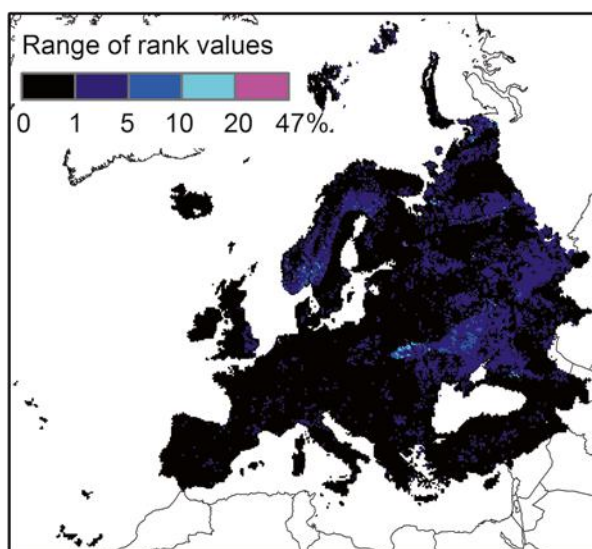




**Fig. 2** Zonation priority maps for mammals in Europe. The *red tones* represent the best 10 % of the solution and *blue tones* indicate the lowest 50 % of cells. (a) Is the basic, core-area Zonation solution for our data where conservation value in Zonation optimization is only based on species richness normalised by range size. (b) Is the Zonation solution where the conservation value of a cell is weighted with the medium variant of the phylogenetic diversity, i.e. the inverse of the equivalent number of Rao's quadratic entropy for the local community in each cell is used as cell costs. (c) Shows the national level basic Zonation priorities and (d) is the national analysis with phylogenetic diversity included

The three phylogenetic weightings give very similar results. The pair-wise Spearman rank correlations between these differently weighted Zonation analyses were very high (low-medium: 0.9965, low-high: 0.9916, and medium-high: 0.9987). Therefore, in the following analyses we used the medium weighting only, which corresponds to using the inverse of the equivalent number of Rao's QE directly as cell cost (see Fig. 1).



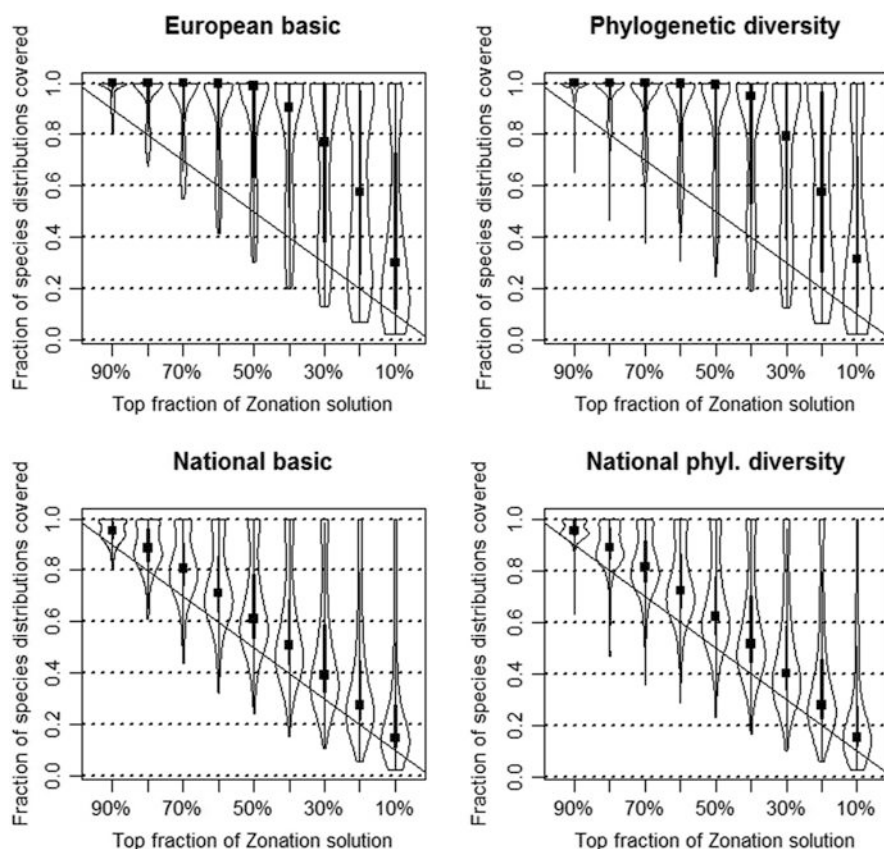


**Fig. 3** The range of variation in rank values among the 100 Zonation solutions done with different phylogenetic trees. Large majority of areas have very consistent rank values, with variation lower than 1 % (i.e. always placed within the same 1 % fraction in the Zonation ranking; *black* in the map). In some regions the variation is broader, but still keeping within the same 10 % fraction in Zonation (*medium blue*). Only very small regions have variation from 10 to 20 % (*light blue*), and only some sparse cells go through more dramatic changes in priority when different tree structures are considered, with variation up to 47 % (*pink cells*)

Similarly, pair-wise Spearman rank correlations for Zonation solutions done with the different 100 phylogenetic trees were also very high. The mean pair-wise correlation was 0.99985 and even the lowest pair-wise correlation was 0.99934. There were only a few regions across the study area where the rankings were not consistent (Fig. 3).

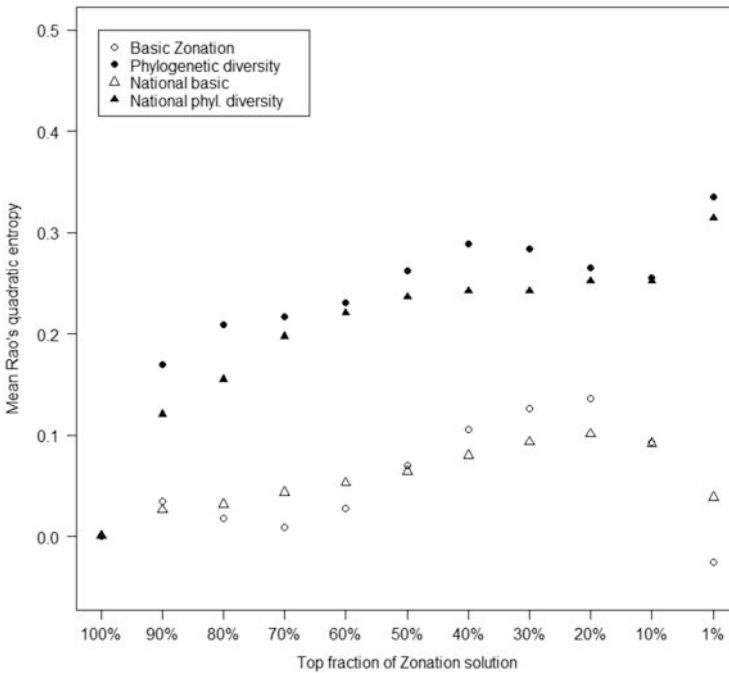
We also tested whether the uncertainty of tree structure was related to the position in Zonation rank, that is, whether there may have been more or less uncertainty associated with top ranking cells. Pair-wise Spearman rank correlations of uncertainty with each of the main Zonation variants gave weak, positive correlations of 0.10 for the basic solution, 0.12 for the phylogenetic diversity analysis, 0.08 for the basic national scale analysis, and 0.07 for the national scale phylogenetic diversity analysis. As the tree uncertainty seemed to play a very minor role in the prioritization outcome, in the following analyses we used one tree only (see Fig. 1).

To assess how well the Zonation priorities covered the different species' ranges, we plotted the proportions of species ranges retained in the landscape at different fractions of cell removal (Fig. 4). The median of representation is higher for the analysis with phylogenetic diversity than for the one without (Fig. 4, black squares). This may seem surprising, but is explained by the fact that Rao's QE correlates with species richness. Looking into the corresponding values for individual species



**Fig. 4** Proportions of species distributions retained (y-axis) in different top fractions of the landscape (x-axis). The *black squares* represent the median value across all species, which are surrounded by vertically plotted density distributions of all species' values around the median. For example, in the basic Zonation solution, the top 20 % of the landscape covers more than 55 % of the ranges for half of the species, but there are also (broadly distributed) species with only ca. 10 % of their ranges covered. A random selection at the continental scale would result in a 1:1 diagonal line for the medians (*solid line*)

(illustrated by the density distributions drawn around the medians in Fig. 4) reveals a very subtle trade-off: The basic core-area Zonation retains species representations more evenly, as it should by definition, whereas the phylogenetic diversity solution loses larger fractions of some species' ranges earlier on in the cell removal process (longer downward tails in the density distributions at lowest 50 % fractions). In other words, with the phylogenetic diversity weighting the protection of some species is traded off against protection of locations with higher phylogenetic diversity. But as this tradeoff is minor and most visible at poorest fractions of the landscape, it is unlikely to be of concern for practical conservation.



**Fig. 5** Mean phylogenetic diversity measured as the equivalent number of Rao's QE across cells at different top fractions of the landscape according to Zonation. The values have been standardized from an original mean QE of 5.06 across all cells. The different prioritizations converge at top fraction equal to 1, as that represents the mean value across all cells in the landscape. If cells were removed in random order, the points would form a flat line at this level

A major difference can be seen between the analyses at different spatial scales. When going from European to national priorities, the median representation values drop substantially, by even ca. 40 % (Fig. 4). Same pattern arises at the national scale from inclusion of phylogenetic diversity: again some species lose more of their ranges for the benefit of others that occur in locations with higher values of the equivalent number of Rao's QE.

To look at evolutionary history maintained by the Zonation solutions, we plotted the mean equivalent number of Rao's QE for the cells at different top fractions of the rankings (Fig. 5). We observed as an overall general trend that the mean QE is increasing as cells are removed from the landscape (from 100 to 1 % in the x-axis) for any selection procedure (with or without including phylogenetic diversity as selection criteria). This is caused by the positive correlation between QE and species richness. By default, Zonation values high species richness cells which will tend to be prioritized, and those cells are also more likely to have high QE values than species poor cells. The very highest priorities (top 1 % in Fig. 5) again diverge from this trend for the solutions that do not consider QE explicitly, because here Zonation tries to maintain a representation for as many species as possible, and thus the complementarity of species compositions overrides the importance of richness, and correlation with the QE weighted solutions disappear.

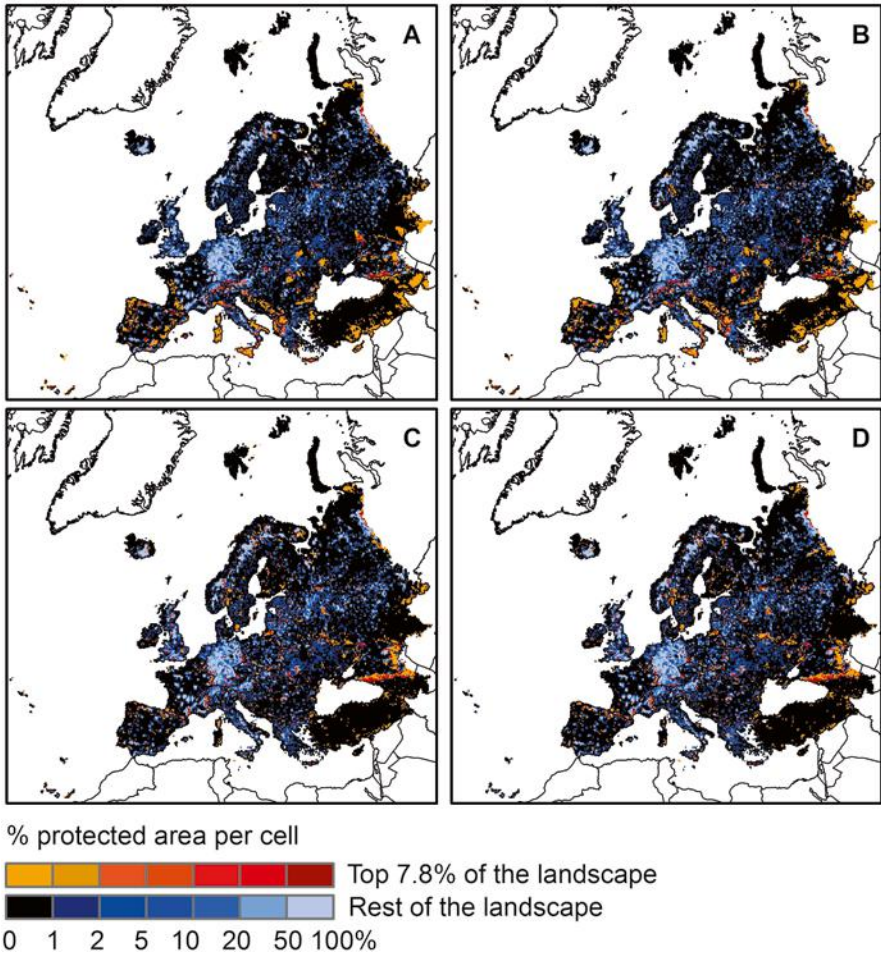
The mean equivalent number of Rao's QE retained by each fraction of the landscape is higher when phylogenetic diversity is accounted for (black-filled symbols are higher than empty symbols). This means that including phylogenetic diversity as a prioritization criterion improves the outcome of the Zonation solution from the perspective of evolutionary history. Our results also highlight that the scale at which the prioritization is conducted (European vs. National) does not appear to have a consistent impact on the mean QE retained in each fraction of the landscape (same colored symbols are close to each other for a given fraction). In other words, the choice whether to conduct a prioritization at the country level or at the continent level does not influence how much phylogenetic diversity is retained.

We overlaid the Zonation rankings with maps of existing protected areas to see how well the priorities and protected areas coincide. The protected areas we considered in our analyses (WDPA categories I–IV) cover a total of 7.8 % of the land area in the study region. We compared them to the same amount of land area prioritized by the Zonation variants (Fig. 6), i.e., 7.8 % top fraction of the Zonation solutions. A large majority of currently protected land is not considered of high priority by any of the Zonation variants (light blue areas), and conversely, much of the Zonation priorities are unprotected (yellow-orange tones). The best matching areas are shown in red, and are sparsely located across the study region without any clear spatial trends.

We plotted the mean proportions of cell area protected among the cells in different top fractions of Zonation solutions for each of the four main solutions (Fig. 7). For both of the European scale analyses the proportion protected did not seem to depend at all whether the cells were considered of high or low priority. Actually their pattern of distribution appeared near random. Instead, for the national scale analyses there was a consistent pattern of increasing protection with increasing rank in Zonation, for both the basic and phylogenetic diversity variants of Zonation. Topmost 1 % fractions had almost twice as much area under protection as compared with the mean across the whole study region.

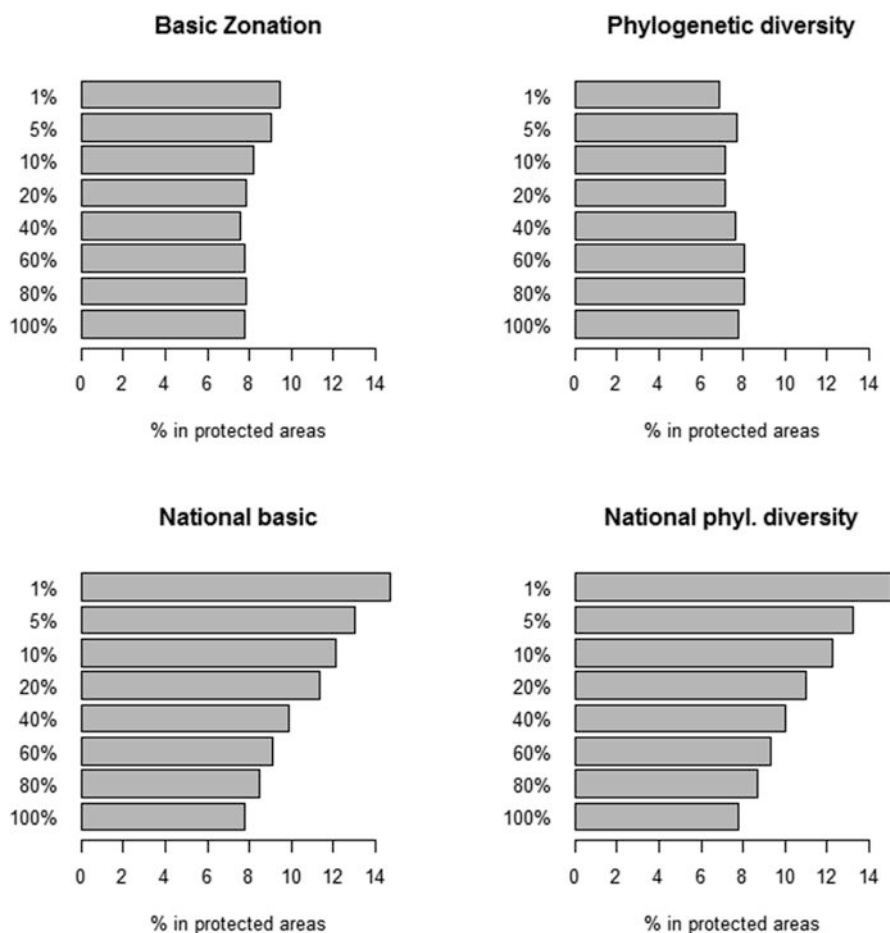
## Discussion and Conclusions

We prioritized areas for conservation of hotspots of mammalian evolutionary history with a spatial prioritization tool. Majority of high priority areas for species conservation are also of high priority for the conservation of evolutionary history, but there are some regions where substantial differences occur between the two different goals. This implies that targeting species alone does not necessarily succeed in protection of hotspots of evolutionary history. Past research has found mixed evidence of such surrogacy relationships between protecting species and phylogenetic diversity (Polasky et al. 2001; Rodrigues and Gaston 2002; Sechrest et al. 2002; Forest et al. 2007; Spathelf and Waite 2007; Rodrigues et al. 2011). Our findings show that it makes a difference in what regions such comparisons are made: we found little difference between priorities around the Mediterranean, but much more



**Fig. 6** Overlap of the area identified as priority for conservation with Zonation (7.8 % of the land area in the study region) with currently protected areas (WDPA categories I–IV). The two different color scales indicate the proportion of each cell under protection: *Light blue* cells have more than half of their area protected, *black* cells have less than 1 %. The tones from yellow to dark red indicate the same thing, but for the cells belonging to the Zonation 7.8 % top fractions. (a) Is the basic, core-area Zonation solution for our data where conservation value in Zonation optimization is only based on species richness normalised by range size. (b) Is the Zonation solution where the conservation value of a cell is weighted with the medium phylogenetic diversity, i.e. the inverse of the equivalent number of Rao's quadratic entropy for the local community in each cell is used as cell costs. (c) Shows the national level basic Zonation priorities and (d) is the national analysis with phylogenetic diversity included

e.g. in Scandinavia. Chazot et al. (chapter “[Patterns of Species, Phylogenetic and Mimicry Diversity of Clearwing Butterflies in the Neotropics](#)”) found that phylogenetic diversity and species richness are less correlated in areas of low species richness, which might explain some of the patterns in our results as well.



**Fig. 7** The mean proportion of area protected in the different top fractions of the Zonation rankings. For example, out of the best 1 % of the cells according to the continental scale phylogenetic diversity variant (with inverse of the equivalent number of Rao's QE as cell cost) approximately 7 % of area is under protection, whereas in both of the National Zonation variants more than twice as much of the top priority area is under protection. The 100 % bar indicates the overall mean of area protected across the whole study region, corresponding to 7.8 % of land area

Mammals are a group of species with broad distributions even at the European scale. Such patterns cause the priorities to be strongly concentrated around southern parts of the study area, where the diversity gradients peak. Whenever such a region is subdivided into smaller administrative units, species ranges will typically extend over multiple units. And whenever distributional ranges cross boundaries, selecting areas complementary to each other within a subunit of a larger area is likely to lead to selecting areas as far as possible from each other: the Northern border will host mostly different species from those along the southern border. This so-called edge artefact (Moilanen et al. 2013) is important to consider when discussing the relevance of spatial scales in priority setting.



A European scale prioritization is much more cost-efficient in covering species ranges, as compared with the national scale analyses that barely surpass a random selection (Fig. 4). National level prioritization is bound to be less cost-effective (Erasmus et al. 1999; Kark et al. 2009; Bladt et al. 2009) if all species, including the broadly distributed ones, must be conserved separately in each country. However, it would not be politically feasible to focus all conservation efforts to the European-scale hotspots either, because these cover disproportionate fractions of some countries, while leaving others virtually unprotected. Therefore, in reality a balanced compromise solution between the two extremes would be desirable, but such options are explored elsewhere (Moilanen and Arponen 2011; Moilanen et al. 2013).

Our results suggest that the amount of uncertainty related to the mammal phylogenies is not significant from the perspective of spatial prioritization of evolutionary hotspots. The differences between trees are minor and appear to occur in parts of the phylogeny with species that mainly occur in species rich communities, and thus the patterns of species distributions drive the prioritization and mask the impact of phylogenetic uncertainty. This is not to say that phylogenetic uncertainty in general would not matter in conservation prioritization. It may well be that for less well known taxa with higher uncertainty, or taxa with different phylogenetic structure and different kinds of patterns of spatial distributions of the species would show much higher variation in prioritization outcomes. The result could also be different for another conservation goal, e.g., if aiming at maximizing phylogenetic diversity across the study region (Rodrigues and Gaston 2002; Rodrigues et al. 2011) rather than considering it at the level of local community as we do here.

When including additional constraints to prioritization, such as a weighting based on phylogenetic diversity, some other aspect may have to be compromised and trade-offs sought, as priorities for different goals rarely perfectly coincide. In the case of the European mammals and alpha-level phylogenetic diversity, we found that the trade-offs were very reasonable, and indeed, negligible as compared with the losses incurred by restricting the prioritization to the national scale.

As expected, the mean phylogenetic diversity in cells prioritized by Zonation variants where phylogenetic diversity was included as a cost layer were higher than in those variants that did not include it. In relative terms the differences were not enormous (see Fig. 5), but one must consider that (1) Zonation can only work with the values that occur in the landscape, and in this case it had to select from a set of cells where the overall mean of phylogenetic diversity (measured as the equivalent number of Rao's quadratic entropy) was slightly larger than 5 and maximum was 7, and (2) the Core-area Zonation needs to retain core cells for all species, and cannot entirely give up on an "expensive" species – that is, a species occurring only in cells with very low phylogenetic diversity. Thus, the flexibility of the solutions strongly depends on the spatial patterns of species distributions and how they relate to the phylogenetic tree structure. For example, if species' range sizes are relatively small and overlap little, Zonation needs to retain a large number of cells to cover core distributions for all of them, and thus there is little flexibility in the solution even when variation in cell costs (or phylogenetic diversity) is high. If rare endemics happen to occur in cells with the highest phylogenetic diversity and all other species

have very broad distributions, then that leaves quite a lot of flexibility for ranking the rest of the landscape.

We found the current network of protected areas to perform rather poorly with respect to representation of areas perceived as high priorities by the European scale Zonation solutions. The proportions of different fractions from Zonation solutions were covered by protected areas roughly equally, close to the overall mean percentage of protected area in the study region. In other words, current protected areas appear equivalent to a random allocation of sites when compared with the European scale Zonation priorities. However, as discussed earlier, prioritization at the European scale may not be a reasonable comparison, as conservation planning in the real world mainly happens at more local scales. The comparison to the national scale prioritization was more positive, with highest Zonation priorities almost twice as likely to be protected as the mean across the region. Even though less than 15 % protection of the highest priorities is perhaps not an outcome to celebrate, it does indicate that at least according to some criteria, protected area allocation in Europe has not been fully opportunistic, and is not worse than random, as could be the case when the sites are biased towards areas of low economic interest (Ferrier 2002).

However, it is also important to remember, that even the best solution at national scale was only mildly better than a random selection (Fig. 4), making it another unreasonable baseline to compare against. It may well be, that the higher coincidence of protected areas with Zonation priorities is simply a consequence of countries preferably locating protected areas near borders, which coincides with spatial priorities due to the edge artefact mentioned above, rather than being a sign of cost-effective protected area planning. Such a pattern was found in the Americas in a previous study (Moilanen et al. 2013). Analyses at higher data resolutions that include other taxa and different aspects of diversity are required to make more realistic and useful assessments of protected areas, but our first attempt does provide some interesting insight into these questions.

Conservation of evolutionary history is generally acknowledged to be important, although the debate on the alternative justifications for it is still ongoing (Rosauer and Mooers 2013; Winter et al. 2013). The underlying reason for its conservation will influence the practical goals and conservation priorities. Our analysis identifies priority regions for conserving high alpha-level phylogenetic diversity for mammals. Such an approach is typically justified on the basis of representing higher functional diversity (see section “[Introduction](#)”), but due to the correlation of QE with species richness it may also be closer to a species-based solution than some alternative ways of considering evolutionary history in conservation prioritization. Therefore, our results should not be taken as proof of an existing surrogacy relationship of species and phylogenetic diversity-based prioritizations, especially as also with our approach there were some regions with clear differences to species-based prioritization. An important notion regarding Zonation, or any prioritization tool, is that it does not inherently “know” what is desirable in conservation. It can only answer the questions it is posed, and it is up to the user that the questions make sense (see also Moilanen 2008). For example, merely adjusting the strength of the weighting (cost layer) in the current analysis will shift priorities to some extent.



Similar prioritizations for different taxa are also quite likely to produce different outcomes. Conservation is always driven by value judgments (Vane-Wright and Coppock 2009), and there is even a risk of purposefully setting goals in a manner that produces desired spatial outcomes.

Since there necessarily are multiple potentially relevant objectives, a conservative, precautionary strategy would be to assess several of them and focus on areas where most priorities are in concordance, and consider as unimportant only the areas where no high priorities occur. However, in practice different types of conservation actions could be necessary to address the different objectives, and therefore the conflicts may be more apparent than real. For instance, regions with particularly low phylogenetic diversity may also be of conservation concern as they can represent areas of active diversification (Forest et al. 2007), but they might require different type of conservation from “museum” areas with relict species, as these areas and species in them might be threatened by very different processes.

Another open and closely related question is at what spatial scales should we operate when measuring and prioritizing evolutionary history? In our case the assumption was that phylogenetic diversity measured as the equivalent number of Rao’s quadratic entropy of the local community was the relevant unit, but especially when assessing the diversity across the study region, the delineation of the study region will have an impact on priorities as described above, but also through “pruning” of the phylogenetic tree: A specific region will cover parts of a full phylogeny, and regional scale prioritization with such a partial tree may prioritize areas different from a global prioritization with a full tree.

Considering the amount of literature on conservation of evolutionary history in general, it is surprising how rarely it is considered in systematic conservation planning applications. Phylogenetic data are increasing and the modern computational prioritization tools are better able to account for such data even at broad scales and for large numbers of species. These developments facilitate the inclusion of phylogenetic diversity into conservation planning. We hope that it will become a routine part of spatial conservation prioritization procedures, and that the message will also better reach the broader public through active communication.

**Acknowledgements** A.A. was funded by the Academy of Finland (grant #250126 to A.A.). LZ was funded by the ANR-BiodivERsA project CONNECT (ANR-11-EBID-002), as part of the ERA-Net BiodivERsA 2010 call. A part of the computations presented here were performed using the CIMENT infrastructure (<https://ciment.ujf-grenoble.fr>), which is supported by the Rhône-Alpes region (GRANT CPER07\_13 CIRA: <http://www.ci-ra.org>). We wish to thank L. Maiorano for the data compilation, W. Thuiller and T. Münckemüller for their programming support, as well as Mar Cabeza, 2 anonymous reviewers and Roseli Pellens for useful comments on the manuscript.

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Arponen A (2012) Prioritizing species for conservation planning. *Biodivers Conserv* 21:875–893. doi:[10.1007/s10531-012-0242-1](https://doi.org/10.1007/s10531-012-0242-1)
- Arponen A, Heikkinen RK, Thomas CD, Moilanen A (2005) The value of biodiversity in reserve selection: representation, species weighting, and benefit functions. *Conserv Biol* 19:2009–2014. doi:[10.1111/j.1523-1739.2005.00218.x](https://doi.org/10.1111/j.1523-1739.2005.00218.x)
- Ball IR, Possingham HP, Watts M (2009) Marxan and relatives: software for spatial conservation prioritisation. In: Moilanen A, Wilson KH, Possingham HP (eds) *Spatial conservation prioritisation: quantitative methods and computational tools*. Oxford University Press, Oxford, pp 185–195
- Bininda-Emonds ORP, Cardillo M, Jones KE et al (2007) The delayed rise of present-day mammals. *Nature* 446:507–512
- Bladt J, Strange N, Abildtrup J et al (2009) Conservation efficiency of geopolitical coordination in the EU. *J Nat Conserv* 17:72–86. doi:[10.1016/j.jnc.2008.12.003](https://doi.org/10.1016/j.jnc.2008.12.003)
- Cadotte MW, Cardinale BJ, Oakley TH (2008) Evolutionary history and the effect of biodiversity on plant productivity. *Proc Natl Acad Sci* 105:17012–17017. doi:[10.1073/pnas.0805962105](https://doi.org/10.1073/pnas.0805962105)
- Cadotte M, Dinnage R, Tilman GD (2012) Phylogenetic diversity promotes ecosystem stability. *Ecology* 93:223–233. doi:[10.1890/11-0426.1](https://doi.org/10.1890/11-0426.1)
- Chao A, Chiu C-H, Jost L (2010) Phylogenetic diversity measures based on Hill numbers. *Philos Trans R Soc Lond Ser B Biol Sci* 365:3599–3609
- Davies TJ, Buckley LB (2011) Phylogenetic diversity as a window into the evolutionary and biogeographic histories of present-day richness gradients for mammals. *Philos Trans R Soc Lond Ser B Biol Sci* 366:2414–2425
- De Bello F, Lavergne S, Meynard CN et al (2010) The partitioning of diversity: showing Theseus a way out of the labyrinth. *J Veg Sci* 21:992–1000. doi:[10.1111/j.1654-1103.2010.01195.x](https://doi.org/10.1111/j.1654-1103.2010.01195.x)
- Eklund J, Arponen A, Visconti P, Cabeza M (2011) Governance factors in the identification of global conservation priorities for mammals. *Philos Trans R Soc Lond B Biol Sci* 366:2661–2669. doi:[10.1098/rstb.2011.0114](https://doi.org/10.1098/rstb.2011.0114)
- Erasmus BFN, Freitag S, Gaston KJ et al (1999) Scale and conservation planning in the real world. *Proc R Soc Lond Ser B Biol Sci* 266:315–319
- Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61:1–10
- Ferrier S (2002) Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? *Syst Biol* 51:331–363. doi:[10.1080/10635150252899806](https://doi.org/10.1080/10635150252899806)
- Forest FF, Grenyer R, Rouget M et al (2007) Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature* 445:757–760. doi:[10.1038/nature05587](https://doi.org/10.1038/nature05587)
- Fritz SA, Rahbek C (2012) Global patterns of amphibian phylogenetic diversity. *J Biogeogr* 39:1373–1382. doi:[10.1111/j.1365-2699.2012.02757.x](https://doi.org/10.1111/j.1365-2699.2012.02757.x)
- Fritz SA, Bininda-Emonds ORP, Purvis A (2009) Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics. *Ecol Lett* 12:538–549
- Isaac NJB, Turvey ST, Collen B et al (2007) Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PLoS One* 2:e296. doi:[10.1371/journal.pone.0000296](https://doi.org/10.1371/journal.pone.0000296)
- IUCN (2012) The IUCN Red List of threatened species. Version 2012.2. <http://www.iucnredlist.org>
- Jost L (2007) Partitioning diversity into independent alpha and beta components. *Ecology* 88:2427–2439
- Kark S, Levin N, Grantham HS, Possingham HP (2009) Between-country collaboration and consideration of costs increase conservation planning efficiency in the Mediterranean Basin. *Proc Natl Acad Sci U S A* 106:15368–15373. doi:[10.1073/pnas.0901001106](https://doi.org/10.1073/pnas.0901001106)
- Kuhn TS, Mooers AØ, Thomas GH (2011) A simple polytomy resolver for dated phylogenies. *Methods Ecol Evol* 2:427–436. doi:[10.1111/j.2041-210X.2011.00103.x](https://doi.org/10.1111/j.2041-210X.2011.00103.x)
- Leinster T, Cobbold CA (2012) Measuring diversity: the importance of species similarity. *Ecology* 93:477–489

- Maiorano L, Amori G, Capula M et al (2013) Threats from climate change to terrestrial vertebrate hotspots in Europe. *PLoS One* 8:e74989. doi:[10.1371/journal.pone.0074989](https://doi.org/10.1371/journal.pone.0074989)
- Margules CR, Pressey RL (2000) Systematic conservation planning. *Nature* 405:243–253. doi:[10.1038/35012251](https://doi.org/10.1038/35012251)
- Margules C, Sarkar S (2007) Systematic conservation planning. Cambridge University Press, Cambridge, p 278
- Moilanen A (2007) Landscape Zonation, benefit functions and target-based planning: unifying reserve selection strategies. *Biol Conserv* 134:571–579. doi:[10.1016/j.biocon.2006.09.008](https://doi.org/10.1016/j.biocon.2006.09.008)
- Moilanen A (2008) Two paths to a suboptimal solution – once more about optimality in reserve selection. *Biol Conserv* 141:1919–1923. doi:[10.1016/j.biocon.2008.04.018](https://doi.org/10.1016/j.biocon.2008.04.018)
- Moilanen A, Arponen A (2011) Administrative regions in conservation: balancing local priorities with regional to global preferences in spatial planning. *Biol Conserv* 144:1719–1725. doi:[10.1016/j.biocon.2011.03.007](https://doi.org/10.1016/j.biocon.2011.03.007)
- Moilanen A, Franco AMA, Early RI et al (2005) Prioritising multiple use landscapes for conservation: methods for large multi species planning problems. *Proc R Soc Lond Ser B Biol Sci* 272:1885–1891. doi:[10.1098/rspb.2005.3164](https://doi.org/10.1098/rspb.2005.3164)
- Moilanen A, Kujala H, Leathwick JR (2009) The Zonation framework and software for conservation prioritization. In: Moilanen A, Wilson KH, Possingham HP (eds) *Spatial conservation prioritisation: quantitative methods and computational tools*. Oxford University Press, Oxford, pp 196–210
- Moilanen A, Leppänen J, Meller L et al (2011) Spatial conservation planning framework and software Zonation v. 3.0: user manual. University of Helsinki, Helsinki, p 278
- Moilanen A, Anderson BJ, Arponen A et al (2013) Edge artefacts and lost performance in national versus continental conservation priority areas. *Divers Distrib* 19:171–183. doi:[10.1111/ddi.12000](https://doi.org/10.1111/ddi.12000)
- Mouquet N, Devictor V, Meynard CN et al (2012) Ecophylogenetics: advances and perspectives. *Biol Rev Camb Philos Soc* 87:769–785. doi:[10.1111/j.1469-185X.2012.00224.x](https://doi.org/10.1111/j.1469-185X.2012.00224.x)
- Pavoine S, Bonsall MB (2011) Measuring biodiversity to explain community assembly: a unified approach. *Biol Rev Camb Philos Soc* 2:792–812. doi:[10.1111/j.1469-185X.2010.00171.x](https://doi.org/10.1111/j.1469-185X.2010.00171.x)
- Pavoine S, Ollier S, Dufour A-B, Crozier R (2005a) Is the originality of a species measurable? *Ecol Lett* 8:579–586. doi:[10.1111/j.1461-0248.2005.00752.x](https://doi.org/10.1111/j.1461-0248.2005.00752.x)
- Pavoine S, Ollier S, Pontier D (2005b) Measuring diversity from dissimilarities with Rao's quadratic entropy: are any dissimilarities suitable? *Theor Popul Biol* 67:231–239
- Polasky S, Csuti B, Vossler CA, Meyers SM (2001) A comparison of taxonomic distinctness versus richness as criteria for setting conservation priorities for North American birds. *Biol Conserv* 97:99–105
- Pressey RL, Humphries CJ, Margules CR et al (1993) Beyond opportunism – key principles for systematic reserve selection. *Trends Ecol Evol* 8:124–128
- Rao CR (1982) Diversity and dissimilarity coefficients: a unified approach. *Theor Popul Biol* 21:24–43
- Rodrigues ASL, Gaston KLJ (2002) Maximising phylogenetic diversity in the selection of networks of conservation areas. *Biol Conserv* 105:103–111
- Rodrigues ASL, Grenyer R, Baillie JEM et al (2011) Complete, accurate, mammalian phylogenies aid conservation planning, but not much. *Philos Trans R Soc Lond B Biol Sci* 366:2652–2660
- Rosauer DF, Mooers AO (2013) Nurturing the use of evolutionary diversity in nature conservation. *Trends Ecol Evol* 28:322–323. doi:[10.1016/j.tree.2013.01.014](https://doi.org/10.1016/j.tree.2013.01.014)
- Schweiger O, Klotz S, Durka W, Kühn I (2008) A comparative test of phylogenetic diversity indices. *Oecologia* 157:485–495. doi:[10.1007/s00442-008-1082-2](https://doi.org/10.1007/s00442-008-1082-2)
- Sechrest W, Brooks TM, da Fonseca GAB et al (2002) Hotspots and the conservation of evolutionary history. *Proc Natl Acad Sci U S A* 99:2067–2071. doi:[10.1073/pnas.251680798](https://doi.org/10.1073/pnas.251680798)
- Smith RJ, Veríssimo D, Isaac NJB, Jones KE (2012) Identifying Cinderella species: uncovering mammals with conservation flagship appeal. *Conserv Lett* 5:205–212. doi:[10.1111/j.1755-263X.2012.00229.x](https://doi.org/10.1111/j.1755-263X.2012.00229.x)

- Spathelf M, Waite T a (2007) Will hotspots conserve extra primate and carnivore evolutionary history? *Divers Distrib* 13:746–751. doi:[10.1111/j.1472-4642.2007.00386.x](https://doi.org/10.1111/j.1472-4642.2007.00386.x)
- UNEP I (2010) The World Database on Protected Areas (WDPA). UNEP-WCMC, Cambridge, [www.protectedplanet.net](http://www.protectedplanet.net)
- Vane-Wright RI, Coppock J (2009) Planetary awareness, worldviews and the conservation of biodiversity. In: Kellert SR, Speth JG (eds) *The coming transformation. Values to sustain human and natural communities*. Yale School of Forestry & Environmental Studies, New Haven, pp 353–382
- Vane-Wright RI, Humphries CJ, Williams PH (1991) What to protect? – systematics and the agony of choice. *Biol Conserv* 55:235–254
- Winter M, Devictor V, Schweiger O (2013) Phylogenetic diversity and nature conservation: where are we? *Trends Ecol Evol* 28:199–204. doi:[10.1016/j.tree.2012.10.015](https://doi.org/10.1016/j.tree.2012.10.015)

# Priorities for Conservation of the Evolutionary History of Amphibians in the Cerrado

Débora Leite Silvano, Paula Hanna Valdujo, and Guarino Rinaldi Colli

**Abstract** Population declines and species extinction can be abated through the establishment of effective conservation policies. Actions and policies towards biodiversity conservation must be well planned and priorities must be set. Besides the widely recognized principles of systematic conservation planning, it is also important to consider species attributes, such as their evolutionary distinctiveness (ED) and distribution pattern. In this study we did a gap analysis to evaluate protection status of anuran species endemic to the Brazilian Cerrado. We then selected priority areas for conservation in this biome based on a systematic conservation planning framework, also including species attributes as prioritization criteria. We found 65 gap species, for which less than 20 % of their conservation targets are met by the current network of protected areas, and 39 of them are not protected at all. Priority areas are located in the central portion of the Cerrado, and include river valleys and mountaintops. Mountains in southeastern and central Cerrado are especially rich in endemic and range-restricted species, resulting in higher priority values for these areas. Priority areas selected here are also the richest regions and have greater Total Evolutionary Distinctiveness than the rest of the biome, demonstrating their high potential for conserving evolutionary history of anuran lineages in the Cerrado. Despite their great importance for biodiversity, areas that have higher richness of endemic species are also those that suffered from more severe loss of habitat, which reinforces the urgency for effective actions towards species conservation.

---

D.L. Silvano (✉)

Laboratório de Zoologia – Universidade Católica de Brasília,  
Campus I – QS 07 Lote 01 EPCT, Águas Claras – CEP, 71966-700 Brasília, DF, Brazil  
e-mail: [deborasilvano@gmail.com](mailto:deborasilvano@gmail.com)

P.H. Valdujo

Laboratório de Ecologia da Paisagem WWF-Brasil,  
SHIS EQ QL 6/8, 71620-430 Brasília, DF, Brazil  
e-mail: [paulavaldujo@wwf.org.br](mailto:paulavaldujo@wwf.org.br)

G.R. Colli

Departamento de Zoologia, Universidade de Brasília,  
Campus Universitário Darcy Ribeiro, 70910-900 Brasília, DF, Brazil  
e-mail: [gcolli@unb.br](mailto:gcolli@unb.br)

**Keywords** Evolutionary distinctiveness • Systematic conservation planning • Gap analysis • Marxan • Brazil

## Introduction

Declines and extinctions of species often occur simply because many countries do not have an effective conservation policy. These declines are creating a demand for rapid and urgent strategies to maximize conservation efforts, especially in regions where there is little data on diversity, abundance and distribution of species, such as in Brazil (Young et al. 2001). Amphibians are perhaps the most threatened group of organisms at global scale (Wake and Vredenburg 2008; see Youssefou and Davies chapter “[Reconsidering the Loss of Evolutionary History: How Does Non-random Extinction Prune the Tree-of-Life?](#)”), with rapidly declining populations throughout the world (Stuart et al. 2004; Becker et al. 2007) and a significant concentration in the Neotropics (Becker and Loyola 2008). Brazil is the world leader in amphibian diversity. In spite of that, there is not yet a specific agenda for their conservation. There are some important initiatives undertaken by the government, such as lists of endangered species and the selection of priority areas for conservation (Silvano and Segalla 2005). However, these initiatives are quite general and often use subjective criteria.

Other initiatives are being conducted by the academic community, such as the Action Plan for Amphibian Conservation in Brazil (Verdade et al. 2012). Among the proposals outlined in this Action Plan for Amphibian Conservation, there is an indication of priority areas for their conservation (Verdade et al. 2012). To make this effective, it is recommended that they follow the same principles of systematic conservation planning (SCP) (Margules and Pressey 2000). SCP aims at a cost efficient protected areas network with the help of purposely built computer software that takes advantage of optimization algorithms. These criteria are essential to define the smallest set of areas necessary to achieve preset conservation goals (see Arponen and Zupan chapter “[Representing Hotspots of Evolutionary History in Systematic Conservation Planning for European Mammals](#)”). Since there are no resources neither enough time to conserve species one by one, we need to maximize the return on investment in conservation (Margules and Pressey 2000).

For conservation to be effective, in addition to the basic principles related to systematic conservation planning, it is necessary to consider certain attributes of the target species. Among these characteristics, we highlight Evolutionary Distinctiveness (ED) (Isaac et al. 2007) and their range size. The ED and range size should be considered independently for each species. The ED is a measure of species’ relative contributions to the total diversity in a phylogenetic tree (Isaac et al. 2007). In this framework more relictual species (i.e. those that belong to ancient

clades, with few species) should be prioritized by the unique evolutionary history they represent (Posadas et al. 2001). Similarly, species that have restricted distribution (e.g. endemic to Espinhaço range) require further attention over those widely distributed, since the species' range size is the most important predictor for the risk of extinction (Purvis et al. 2000a, b). This approach allows for preserving evolutionary history within a taxonomic group, providing more alternatives for responding to possible future environmental changes (Vazquez and Gittleman 1998; Avise 2005; Becker et al. 2010, and see Faith chapter "[The PD Phylogenetic Diversity Framework: Linking Evolutionary History to Feature Diversity for Biodiversity Conservation](#)").

Since half of the over 200 anuran species that occur in the Cerrado are endemic to this domain (Valdujo et al. 2012), it is critical that conservation strategies are outlined specifically to this region. Cerrado is one of 34 priority areas for conservation on the planet (Biodiversity Hotspots – Mittermeier et al. 2004), due to high levels of endemism of fauna and flora and the high rates of habitat destruction. However, few conservation actions are being carried out there. Currently, less than 2 % of the Cerrado range is under strict protection (CNUC 2010). This percentage is low for a region with high heterogeneity of vegetation and topography, and because the main threat to amphibian conservation in the Cerrado is the destruction of their habitats due to deforestation, expansion of agriculture, mining, fire and infrastructure development (Silvano and Segalla 2005). Therefore, strengthening and expanding the network of protected areas should be prioritized as an important conservation strategy, which could maximize the return on investment in conservation (Margules and Pressey 2000).

In spite of the recognized importance of including information on historical and evolutionary studies to define conservation priorities, in the Cerrado, just few and recent papers consider this information (e.g. Carvalho et al. 2010). The papers published over the last decade involving the prioritization of areas for anurans conservation in the Cerrado were based just on the species' extent of occurrence and richness, in a complementarity approach (e.g. Diniz-Filho et al. 2004, 2007, 2009). In order to contribute to enlarge this perspective, we conducted a gap analysis to check the conservation status of amphibian species endemic to the Cerrado and performed an exercise in prioritization of additional conservation areas needed for their protection. Information related to geographical distribution and evolutionary distinctiveness were considered in setting conservation goals for each species. Thus, we have prioritized the most relictual species, because they are phylogenetically rare, and the species of more restricted distribution, because restricted distribution ranges are associated with higher vulnerability to extinction in cases where habitat destruction pop up simultaneously in several points of the landscape. This study contributes to the proposed priority areas already published for the Cerrado through the inclusion of relevant evolutionary information and the use a more refined and complete database.

## Methods

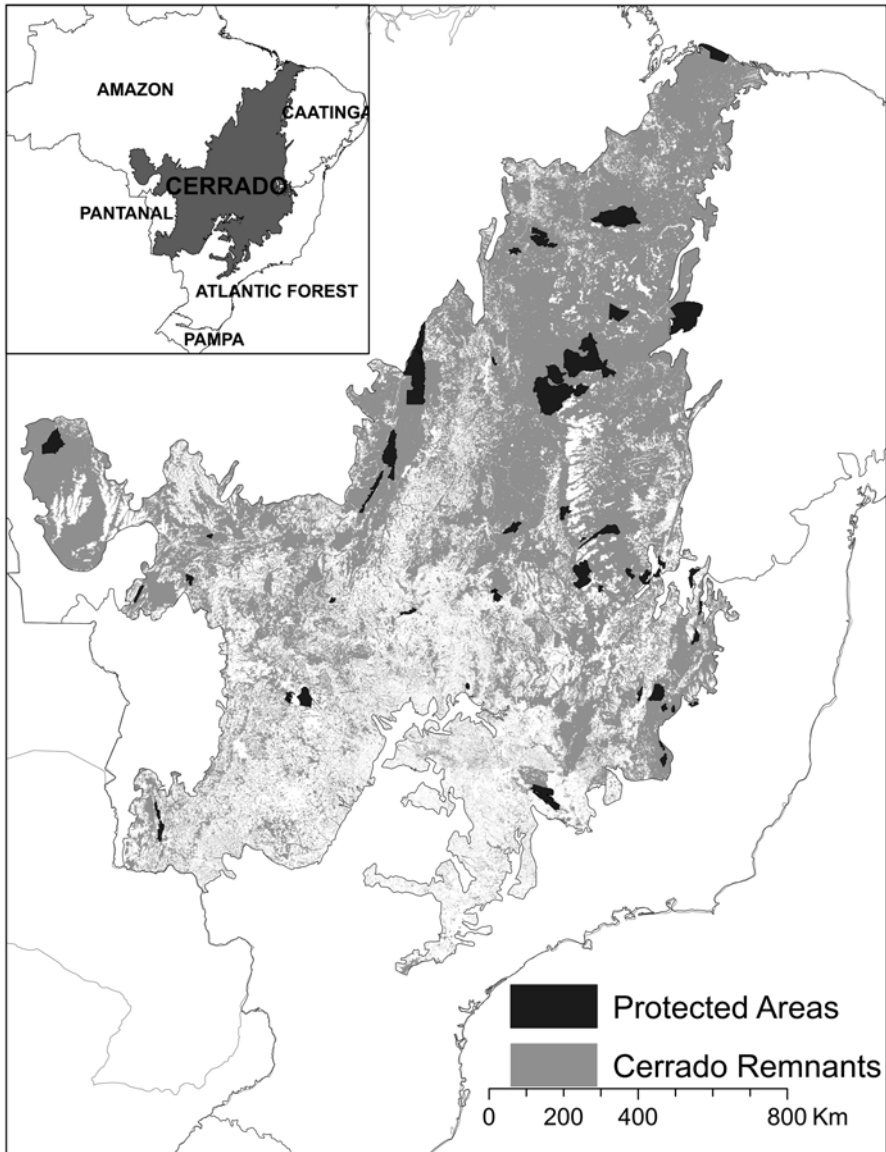
### *Study Area*

The Cerrado is located in central-eastern South America. It is covered by an heterogeneous mosaic of savannic and forest vegetation, including grasslands, shrublands and riverine forests, consisting of a gradient of altitude and vegetation density (Eiten 1972, 1982). Covering over 2.5 million km<sup>2</sup>, the Cerrado is renowned for its high species richness and endemism that places it as the planet's most diverse savannah. However, during the past 40 years their land has been converted mainly into crops and pastures, leading to an intense process of destruction and fragmentation of the vegetation (Klink and Machado 2005). Currently, the widest remnants of natural vegetation are mainly concentrated in the northern portion (Fig. 1). According to recent estimates, there are only 34 % of the original vegetation left and this is expected to disappear in 30 years if current rates of deforestation are maintained in the region, where traditional cultures are giving place to modern mechanized crops such as soybeans, cotton, corn, sorghum and sunflower (Machado et al. 2004). There is not a consensus about the delimitation of the Cerrado. However, since one of the main objectives of this study is to provide tools for decision-making related to conservation, we chose here to use the biome boundaries that are also adopted by the federal government's policies (IBGE 2004).

### *Data Used and Pre-processing*

**Planning Units** Planning units (PUs) are subdivisions of the study area into small spatially explicit units. Among many possible ways of obtaining PUs, we used a hydrosheds arrangement built from SRTM (Shuttle Radar Topography Mission; Hydrosheds, <http://hydrosheds.cr.usgs.gov/index.php>). This is the same database used by Brazilian government to set priority areas for conservation in Cerrado and Pantanal Biomes (MMA 2012, unpublished data). The use of sub basins as PUs has many advantages over other arrangements such as grids or hexagons: firstly, they have natural and biogeographically meaningful limits; secondly, they allow an hierarchical structure of basins within basins, which is very useful to switch scales and adjust data and results to different needs. To account for the complementarity principle of systematic conservation planning, strictly protected areas (IUCN categories I to IV) were included as PUs, using their actual boundaries regardless of the basin subdivision to design PUs. We only included protected areas wider than 350 km<sup>2</sup> to keep PUs sizes compatible with the scale of study and compatible to the official map of priority areas for conservation of the Cerrado, published by the Ministry of Environment. Twenty-three out of 108 protected areas were considered in the gap analysis, covering 50,640 out of 56,223 km<sup>2</sup> of IUCN categories I-IV protected areas





**Fig. 1** The Cerrado and its relation with other biomes (*inset*). Distribution of the Cerrado vegetation remnants (*gray*) and Protected Areas (PAs) greater than 350 km<sup>2</sup> (*black*)

in the Brazilian Cerrado. To define the area available within each PU, we overlaid the official map of extent of natural vegetation in the Cerrado in 2010 with PUs (data available from <http://siscom.ibama.gov.br/monitorabiomas/cerrado/index.htm>), and excluded any PU having no remnants of natural vegetation.

**Conservation Cost** The cost for each PU was obtained from WWF (Soares et al. 2012). The database was built by the Conservation Science Team based on potential future deforestation, using Land Change Modeler module of Idrisi Selva. Distance to roads, to cities, to infrastructure and to previously deforested areas were included as driver to changes in land cover from 2002 to 2010, and then applied to 2010 natural vegetation map to predict which areas are more likely to be deforested in the next 10 years.

**Focal Species** Eighty-two out of 209 amphibian species known to occur in the Cerrado (Valdujo et al. 2012) were selected as focal species. The criteria were based on endemism, range size (both obtained from Valdujo et al. 2012) and level of tolerance to anthropogenic alterations in habitat quality (two classes: tolerant and not-tolerant; species were classified based in our field experience, so that species commonly seen in disturbed areas were considered as tolerant). We used both endemism and extent of distribution as independent criteria because some species are endemic to the Cerrado but have a wide range within this biome, whereas some other species are range restricted (e.g. <60,000 km<sup>2</sup>) but occur in a transition zone between Cerrado and Atlantic Forest, and so they are not endemic to the Cerrado (Valdujo et al. 2012). Since we were prioritizing among natural areas within the Cerrado, widespread species do not add to the final solution, and neither do species that can tolerate habitat degradation.

**Species Distribution Models** We prepared geographic distribution maps for all 82 species, using distribution models constructed through the Maximum Entropy algorithm – MAXENT (Elith et al. 2006; Phillips and Dudik 2008). We included as predictors elevation and all 19 bioclimatic variables with a 10 arc-min spatial resolution provided by Worldclim (Hijmans et al. 2005). For each species we used the mean model of 20 runs and converted probabilistic models to binary models using the 10 percentile training presence logistic threshold. Distribution maps were lately validated by a group of experts during a workshop organized by the Ministry of Environment and WWF aiming to identify priority areas for biodiversity conservation in the Cerrado, in 2011, following the procedure recommended by Graham and Hijmans (2006). The distribution map for each species was superimposed onto the PUs' map in order to calculate how much of its distribution area is contained in each PU. All distribution maps were overlaid to obtain the richness surface of endemic species of amphibians in the Cerrado.

**Evolutionary History Prioritization** In some cases the outcome of area prioritization through SCP analyses fails to meet all targets. To ensure that at least the most important species meet their targets, it is possible to set a penalty factor (SPF) for each species that penalizes solutions more heavily when not achieving these targets. We assigned SPF based on both threat and phylogeny, using ED scores (Evolutionarily Distinctiveness) obtained from Isaac et al. (2012), ranging from 4669 to 17,903.

**Mapping Total Evolutionary Distinctiveness** We calculate the total ED of each PU by summing the value of all species occurring in it. As ED is highly correlated with richness, here we used a weighted value, obtained by dividing summed ED by richness in each PU.

**Table 1** Criteria for the definition of quantitative targets (percentage of range size already under legal protection), according to species range size

Species range size	Quantitative target
<60,000 km <sup>2</sup>	80 %
60,000–350,000 km <sup>2</sup>	50 %
>350,000 km <sup>2</sup>	20 %

**Table 2** Gap category, according to the percentage of quantitative target reached

Percentage of quantitative target reached	Gap category
<20 %	Gap species
20–90 %	Partial gap
>90 %	Covered

## Analysis

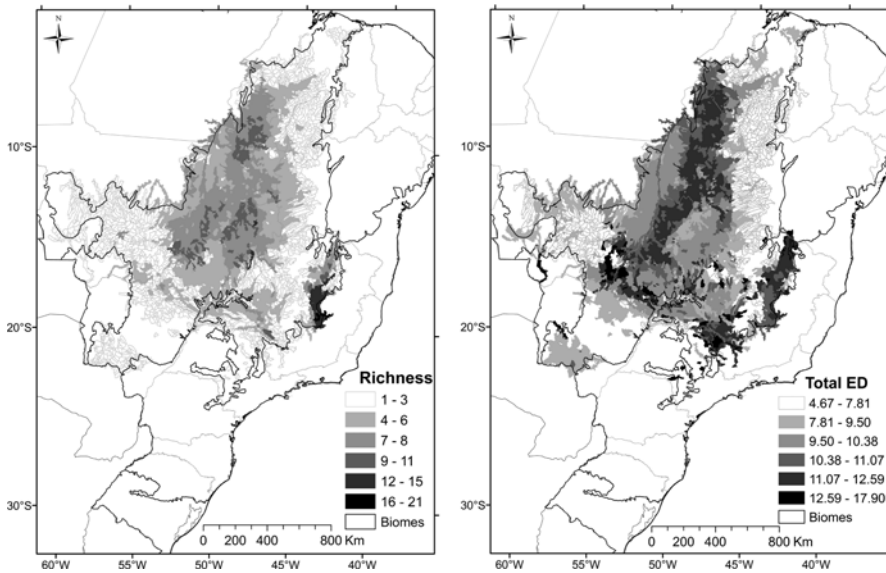
**Gap Analysis** To evaluate the conservation status of each of the focal species we performed a gap analysis (Rodrigues et al. 2003, 2004). This analysis consists of overlaying species distribution maps and protected areas to calculate how much of the quantitative target set for each species is already under legal protection. Spatial data for Brazilian protected areas were obtained from the Ministry of Environment website (<http://mapas.mma.gov.br/i3geo/mma/openlayers.htm?u3n6kqkh7ajn4igbe5jilhka56>). Targets were set to 20–80 % according to range size (Table 1). Those for which only up to 20 % of its conservation goal has been reached were considered “gap species”. The reaching from 20 to 90 % of the target were considered “partial gaps”, and above 90 % the species was considered “covered” (Rodrigues et al. 2003, 2004) (Table 2).

To select areas and define a conservation scenario for Cerrado amphibians we used the conservation planning software Marxan available online (<http://www.uq.edu.au/marxan/index.html>; Ball and Possingham 2000). Marxan uses a simulated annealing optimization algorithm for minimizing costs of achieving conservation targets. Planning units defined by protected areas were assigned to status 2, “reserved”. We set to 10,000 runs with 1 million iterations each run, temperature decreases = 10,000, and boundary modifier = 0.2. The identification of priorities for expanding the current network of protected areas was based on measures of “biological significance” (irreplaceability) of each PU within the study area.

Only to assist the identification of some areas within the basins we used geomorphological units denominations (IBGE 2011).

## Results

Species richness of amphibians endemic to the Cerrado varied between 0 and 21 species per PU (Fig. 2). Species are concentrated in the center of the biome, in its northwestern portion in the contact zone with the Amazon, and in the extreme



**Fig. 2** Species richness and total evolutionary distinctiveness of amphibian endemic to Cerrado per Planning Units (PUs)

southeastern region of the Espinhaço in the contact zone with the Atlantic Forest. The northeastern, southern and western Cerrado portions have low endemism, not exceeding four focal-species (Fig. 2). Total Evolutionary distinctiveness is also concentrated at the center, but with highest values at the Atlantic Forest contact zone (Espinhaço range), in the central western portion (Caiapônia plateau) and in some points at contact zone with Pantanal (Fig. 2).

Among the 82 species examined, over 80 % (66 species) have restricted distribution ranges (<6 million ha) and only 11 % (9 species) are widely distributed across the domain (>35 million ha). Sixty-five (79 %) have less than 20 % of its conservation target achieved being thus classified as gap species. Thirty-nine of these species are completely out of Protected Areas, all of them are restricted range species (<1.5 million ha) (Among the later, some more relictual species are also included, such as *Chiasmocleis mehelyi*, *Oreobates heterodactylus*, *O. remotus*, *Odontophrynus salvatori*, *Proceratophrys moratoi* e *P. cururu*). Only four species endemic to the Cerrado were considered covered (*Leptodactylus tapiti*, *Crossodactylus* sp., *Bokermannohyla ibitiguara* and *Phyllomedusa ayeaye*). All of these covered species have restricted ranges (<0.25 million ha) with most of their distribution in protected areas, and 13 species can be considered as partial gaps, presenting between 23 and 57 % of their conservation goal achieved (Table 3).

In the conservation prioritization analysis the “best solution” (lower cost and higher efficiency) offered by Marxan selected 742 PUs (18.6 % of the biome area)

**Table 3** Focal species evolutionary distinctiveness (ED), distribution area (million hectares), conservation goals, percentage of conservation goal achieved (area of distribution contained in protected areas), classified according to the percentage of goal achieved

Species	ED	Total area	Goal %	Goal area	Area in PAs	% achieved	Classification
<b>Aromobatidae</b>							
<i>Allobates brunneus</i>	5.1	0.202	80	0.162	0.000	0.0	Gap
<i>Allobates goianus</i>	5.1	0.160	80	0.128	0.000	0.0	Gap
<i>Allobates sp.</i>	5.1	0.284	80	0.228	0.000	0.0	Gap
<b>Bufonidae</b>							
<i>Melanophryniscus fulvoguttatus</i>	9.05	4.984	80	3.988	0.077	1.9	Gap
<i>Rhinella cerradensis</i>	4.67	35.147	10	3.515	0.699	19.9	Gap
<i>Rhinella ocellata</i>	4.67	98.528	10	9.853	3.953	40.1	Partial gap
<i>Rhinella scitula</i>	4.67	0.082	80	0.066	0.000	0.0	Gap
<i>Rhinella sp.</i>	4.67	0.118	80	0.095	0.000	0.0	Gap
<i>Rhinella veredas</i>	4.67	17.550	50	8.775	1.273	14.5	Gap
<b>Craugastoridae</b>							
<i>Barycholos ternetzi</i>	16.47	75.998	10	7.600	2.140	28.2	Partial gap
<i>Pristimantis dundeei</i>	11.38	0.952	80	0.762	0.042	5.5	Gap
<i>Oreobates crepitans</i>	11.38	0.238	80	0.190	0.000	0.0	Gap
<i>Oreobates heterodactylus</i>	15.68	0.339	80	0.271	0.000	0.0	Gap
<i>Oreobates remotus</i>	15.68	0.378	80	0.302	0.056	18.7	Gap
<b>Cycloramphidae</b>							
<i>Thoropa megatypanum</i>	13.5	5.359	80	4.287	0.280	6.5	Gap
<b>Dendrobatidae</b>							
<i>Ameerega berohoka</i>	5.44	0.430	80	0.344	0.000	0.0	Gap
<i>Ameerega braccata</i>	5.44	0.238	80	0.190	0.000	0.0	Gap
<i>Ameerega flavopicta</i>	5.44	27.314	50	13.657	0.336	2.5	Gap
<i>Ameerega picta</i>	5.44	0.082	80	0.066	0.000	0.0	Gap
<b>Hylidae</b>							
<i>Aplastodiscus sp.</i>	13.82	0.222	80	0.178	0.000	0.0	Gap
<i>Bokermannohyla alvarengai</i>	11.74	5.118	80	4.095	0.231	5.6	Gap
<i>Bokermannohyla ibitiiguara</i>	11.74	0.250	80	0.200	0.198	98.9	Covered
<i>Bokermannohyla izecksohni</i>	10.65	0.040	80	0.032	0.000	0.0	Gap
<i>Bokermannohyla nanuzae</i>	10.65	2.297	80	1.838	0.196	10.7	Gap
<i>Bokermannohyla pseudopseudis</i>	11.74	9.042	50	4.521	0.105	2.3	Gap
<i>Bokermannohyla ravida</i>	10.65	0.246	80	0.197	0.000	0.0	Gap

(continued)

**Table 3** (continued)

Species	ED	Total area	Goal %	Goal area	Area in PAs	% achieved	Classification
<i>Bokermannohyla saxicola</i>	11.74	5.098	80	4.079	0.266	6.5	Gap
<i>Bokermannohyla szimai</i>	10.65	2.482	80	1.985	0.198	10.0	Gap
<i>Dendropsophus anataliasiasi</i>	9.36	42.393	10	4.239	1.196	28.2	Partial gap
<i>Dendropsophus araguaya</i>	9.36	0.181	80	0.145	0.000	0.0	Gap
<i>Dendropsophus cerradensis</i>	9.36	0.036	80	0.029	0.000	0.0	Gap
<i>Dendropsophus cruzi</i>	9.36	108.799	10	10.880	3.599	33.1	Partial gap
<i>Dendropsophus jimi</i>	9.36	0.589	80	0.472	0.133	28.1	Partial gap
<i>Dendropsophus rhea</i>	9.36	0.225	80	0.180	0.000	0.0	Gap
<i>Dendropsophus tritaeniatus</i>	9.36	1.158	80	0.926	0.000	0.0	Gap
<i>Hypsiboas botumirim</i>	9.67	0.030	80	0.024	0.000	0.0	Gap
<i>Hypsiboas buriti</i>	9.67	0.565	80	0.452	0.000	0.0	Gap
<i>Hypsiboas cipoensis</i>	9.67	3.334	80	2.667	0.244	9.1	Gap
<i>Hypsiboas ericae</i>	9.67	0.149	80	0.119	0.000	0.0	Gap
<i>Hypsiboas goianus</i>	9.67	1.553	80	1.242	0.000	0.0	Gap
<i>Hypsiboas jaguariaivensis</i>	9.67	0.057	80	0.046	0.000	0.0	Gap
<i>Hypsiboas phaeopleura</i>	9.67	0.059	80	0.047	0.000	0.0	Gap
<i>Hypsiboas stenocephalus</i>	9.67	0.433	80	0.347	0.198	57.1	Partial gap
<i>Hypsiboas sp.</i>	11.19	22.337	50	11.169	0.971	8.7	Gap
<i>Lysapsus caraya</i>	12.6	0.129	80	0.104	0.000	0.0	Gap
<i>Phasmahyla jandaia</i>	12.19	0.169	80	0.135	0.032	23.4	Partial gap
<i>Phyllomedusa ayeaye</i>	10.77	0.257	80	0.206	0.198	96.2	Covered
<i>Phyllomedusa centralis</i>	10.77	0.195	80	0.156	0.000	0.0	Gap
<i>Phyllomedusa megacephala</i>	10.77	6.057	50	3.029	0.280	9.2	Gap
<i>Phyllomedusa oreades</i>	10.77	0.583	80	0.466	0.000	0.0	Gap
<i>Pseudis tocantins</i>	11.89	36.514	10	3.651	1.204	33.0	Partial gap
<i>Scinax cabralensis</i>	8.94	0.161	80	0.129	0.022	17.2	Gap
<i>Scinax canastrensis</i>	9.64	0.623	80	0.498	0.198	39.7	Partial gap
<i>Scinax centralis</i>	9.64	0.619	80	0.495	0.000	0.0	Gap
<i>Scinax constrictus</i>	8.94	88.952	10	8.895	3.562	40.0	Partial gap

(continued)

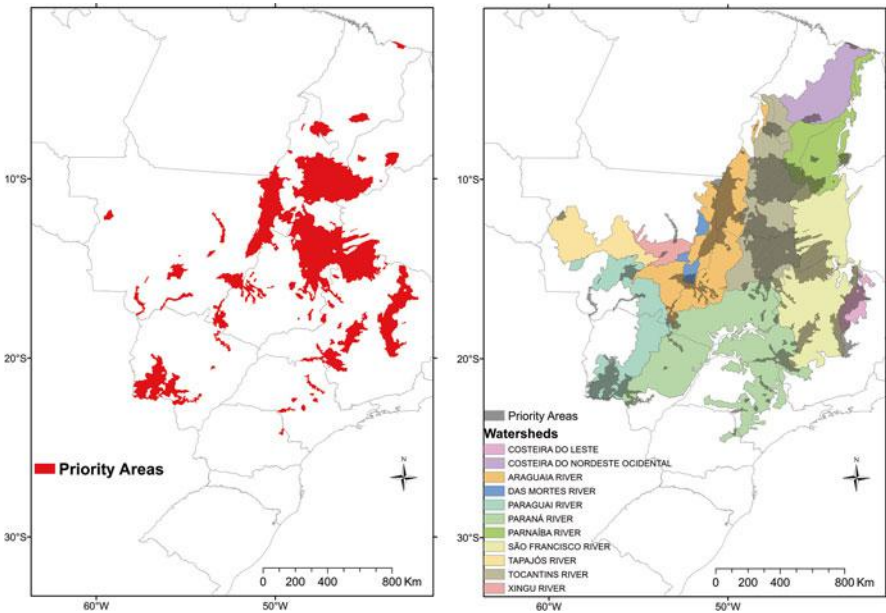
**Table 3** (continued)

Species	ED	Total area	Goal %	Goal area	Area in PAs	% achieved	Classification
<i>Scinax curicica</i>	8.94	1.787	80	1.430	0.196	13.7	Gap
<i>Scinax lutzorum</i>	8.94	0.039	80	0.032	0.000	0.0	Gap
<i>Scinax machadoi</i>	9.64	0.777	80	0.622	0.042	6.8	Gap
<i>Scinax maracaya</i>	8.94	0.622	80	0.498	0.198	39.7	Partial gap
<i>Scinax pinima</i>	8.94	0.100	80	0.080	0.000	0.0	Gap
<i>Scinax rogerioi</i>	8.94	0.200	80	0.160	0.000	0.0	Gap
<i>Scinax skaios</i>	9.64	0.886	80	0.709	0.000	0.0	Gap
<i>Scinax sp.</i>	9.64	0.563	80	0.451	0.011	2.4	Gap
<i>Scinax tigrinus</i>	8.94	0.332	80	0.265	0.000	0.0	Gap
<i>Trachycephalus mambaiensis</i>	10.23	0.180	80	0.144	0.000	0.0	Gap
<b>Hylodidae</b>							
<i>Crossodactylus bokermanni</i>	11.61	2.508	80	2.006	0.196	9.8	Gap
<i>Crossodactylus sp.</i>	13.05	0.243	80	0.194	0.198	101.7	Covered
<i>Crossodactylus trachystomus</i>	13.05	6.242	50	3.121	0.336	10.8	Gap
<i>Hylodes otavioi</i>	10.22	0.296	80	0.237	0.032	13.4	Gap
<b>Leptodactylidae</b>							
<i>Leptodactylus camaquara</i>	11.85	3.296	80	2.637	0.218	8.3	Gap
<i>Leptodactylus cunicularius</i>	11.85	11.644	50	5.822	0.478	8.2	Gap
<i>Leptodactylus pustulatus</i>	14.26	50.906	10	5.091	1.196	23.5	Partial gap
<i>Leptodactylus tapiti</i>	11.85	0.065	80	0.052	0.065	125.0	Covered
<i>Physalaemus deimaticus</i>	13.22	0.207	80	0.166	0.000	0.0	Gap
<i>Physalaemus evangelistai</i>	12.49	1.366	80	1.093	0.061	5.6	Gap
<i>Pleurodema fuscomaculatum</i>	12.74	0.339	80	0.271	0.000	0.0	Gap
<i>Pseudopaludicola mineira</i>	13.52	0.623	80	0.499	0.000	0.0	Gap
<b>Microhylidae</b>							
<i>Chiasmocleis mehelyi</i>	17.9	0.216	80	0.173	0.000	0.0	Gap
<b>Odontophrynidae</b>							
<i>Odontophrynus salvatori</i>	14.97	0.657	80	0.526	0.000	0.0	Gap
<i>Proceratophrys cururu</i>	14.13	0.511	80	0.409	0.055	13.4	Gap
<i>Proceratophrys goyana</i>	14.13	47.302	10	4.730	2.084	44.1	Partial gap
<i>Proceratophrys morato</i>	14.13	0.323	80	0.259	0.000	0.0	Gap



**Table 4** Category (frequency of selection) of the Planning Units (PU), number of PUs, area (million ha) and percentage of Cerrado corresponding to each category of PUs in the best solution of priority areas for Cerrado endemic species of anurans conservation

Category	Number of PUs	Area	% Cerrado
Very high conservation value (10,000)	153	10.49	4.39
High conservation value (7501–9999)	232	13.63	5.71
Intermediate conservation value (5001–7500)	167	10.26	4.30
Low conservation value (1–5000)	190	10.02	4.20
Protected	50	5.87	2.46
Total selected	792	50.27	21.06
Not selected	3760	188.43	78.94
Total	4552	238.7	100



**Fig. 3** Priority areas for the conservation of amphibians species endemic from Cerrado in 11 basins

as priorities. Among them, 153 PUs have a very high conservation value (selected in all 10,000 rounds) and 232 have high conservation value, matching 4.4 and 5.7 % of the biome area, respectively (Table 4). In contrast, 3760 PUs were not selected, representing 78.9 % of the Cerrado.

The selected areas, here termed as priorities for conservation of amphibian species endemic to the Cerrado, mostly occupy the central portion of the biome, following a northwest-southeast diagonal (Fig. 3). Some sparse areas can also be found at the contact with Pantanal biome. This set of areas is of fundamental importance



**Table 5** Richness and endemism of amphibians in major Cerrado basins. Cerrado endemic amphibians species richness, amphibians unique to each basin (Endemism) species richness, and geomorphological units, which concentrate priority areas

Basin	Richness	Endemism	Geomorphological units
São Francisco River	45	3	São Francisco baseline and tableland; Tocantins baseline; Espinhaço mountains
Paraná River	36	6	Canastra and Brazilian central uplands
Tocantins River	28	5	Tocantins and Araguaia rivers depressions and uplands; Brazilian central upland
Araguaia River	21	3	Araguaia, Tocantins and Pantanal rivers depressions
Costeira do Leste	21	1	Espinhaço mountains; Jequitinhonha and Pardo rivers uplands
Paraguay River	18	9	Paraguay and Guaporé rivers depressions and uplands; Guimarães upland
Mortes River	12	0	Tocantins and Araguaia rivers depressions
Parnaíba River	9	0	Meio Norte tableland and depressions
Xingu River	7	0	
Tapajós River	6	0	
Costeira do Nordeste Ocidental	4	0	Meio Norte tablelands

for achieving the conservation goals established. The prioritization analysis selected areas both in river valleys regions (below 400 m altitude), as well as elevated areas (above 1300 m). The selected areas include the depressions of the Araguaia, Tocantins and Paraguay rivers; the uplands in the São Francisco River, in western Bahia; the northern portion of the Central upland, and Canastra and Espinhaço uplands (Fig. 3, Table 5).

The priority areas are mainly concentrated in the Tocantins, Araguaia, São Francisco and Paraguay river basins and on the Costeira do Leste basin (Fig. 3). The São Francisco river basin has the largest number of frog species endemic to the Cerrado (45 species – Fig. 3, Table 5). Among them *Bokermannohyla ravida*, *Scinax cabralensis* and *S. pinima* occurs exclusively in this basin. The Paraná river basin is the second highest in richness, with 36 endemic amphibian species and is home to 6 species that occurs exclusively in this basin (*Bokermannohyla izecksohni*, *Dendropsophus cerradensis*, *D. rhea*, *Hypsiboas jaguariaivensis*, *Proceratophrys morato* and *Scinax centralis*). This basin is followed by the Tocantins river with 28 Cerrado endemic species and 5 species endemic to this basin (*Allobates* sp., *Hypsiboas ericae*, *H. phaeopleura*, *Leptodactylus tapiti* and *Trachycephalus mambaiensis*). The Paraguay river basin has the highest endemism with nine species that occur exclusively there (*Allobates bruneus*, *Ameerega braccata*, *A. picta*, *Chiasmocleis mehelyi*, *Oreobates heterodactylus*, *Phyllomedusa centralis*, *pleurodema fuscomaculatum*, *Oreobates crepitans* and *Rhinella scitula*). Another three species are endemic to the Araguaia river basin (*Dendropsophus araguaya*, *Lysapsus caraya* and *Scinax lutzorum*) and *Hypsiboas botumirim* is endemic to the Costeira do Leste basin.

## Discussion

Given the low number of protected areas and high species richness of amphibians with restricted range in the Cerrado, it was expected that most of the species were not adequately protected and that a large area of biome would be of high conservation value, as demonstrated by the results presented here. An aggravating fact is the greatest richness and total ED of endemic species associated with the central and southeastern regions of the biome. As shown in Fig. 1, these are the areas that suffered the greatest habitat destruction and where remnants are scarce. Forecasts of future habitat degradation also indicate that these areas will suffer further habitat loss if the current economic and political scenarios remain unchanged (see Silvano 2011).

The fact that 39 endemic and restricted range species of amphibians from the Cerrado are completely unprotected is alarming. Several studies have shown that limited range species are more prone to extinction (e.g. Purvis et al. 2000a, b; Cooper et al. 2008). This can happen simply because environmental change can affect all or most of their narrow distributions (Cooper et al. 2008). Most of these species are habitat specialists, and more susceptible to environmental changes (Hero et al. 2005). Moreover, many species occur in low abundance, and also have low reproductive success, and are subject to demographic stochasticity and inbreeding (O'Grady et al. 2006). Among these species are *Proceratophrys moratoi*, an example of threatened restricted range species, which occurs in small populations in extremely degraded grassland areas in the state of São Paulo (Carvalho-Jr et al. 2010; Rolim et al. 2010; Maffei et al. 2011).

More relictual species, such as *Chiasmocleis mehelyi*, *Oreobates heterodactylus*, and *Odontophrynus salvatori*, are completely unprotected and all of them are restricted range species. *Proceratophrys moratoi*, although currently detected within a protected area in São Paulo state, is also considered a gap-species because only a very small proportion of its limited range is actually protected. Others, like *Pristimantis dundeei* and *Oreobates crepitans* are restricted to the region of the cities of Cuiabá and Chapada dos Guimarães at Mato Grosso state. Recent studies indicate that these species are not closely related to others of the same genus, because of their low number of chromosomes and ecological characteristics (Siqueira et al. 2009), which makes them even more unique.

The areas of greatest conservation value for endemic amphibians species are concentrated in the central portion of the biome on a northwest-southeast diagonal, and represent 18.6 % of the Cerrado area. In recent studies, in order to define important areas for inclusion in an efficient network of protected areas for the conservation of all species of Cerrado frogs, 17 priority areas were defined, based on distribution maps (minimum convex polygons) for 131 species (Diniz-Filho et al. 2007, 2009). The results were very similar to those found in a previous study (Diniz-Filho et al. 2004), with the same purpose but using a shorter list of species, different algorithms and a grid of cells of different sizes. These results indicate priority regions for conservation of anurans distributed widely in the biome, but the most

important areas of concentration (“irreplaceable”) in the southeast part. Some of these areas are coincident with those found here and others are very different, as the northern portion of the biome, indicated by these studies as a priority and not selected here. The differences in results should be linked to the fact that (1) our study was based on a more complete database (see Valdujo et al. 2012), (2) we used modeled distribution maps based on topographic and climatic species requirements and (3) we included evolutionary characteristics.

The selection of areas along the elevation gradient, including both lowlands and river valleys as uplands and mountains, is related to the fact that endemic species have different habitat requirements (Valdujo et al. 2012). The São Francisco River basin has the highest species richness, certainly due to the high richness of endemic and restricted range species in the Espinhaço complex (see review in Leite et al. 2008). Other high elevation areas where endemic species have high richness are the Guimarães, Canasta and Central Brazil uplands (Valdujo et al. 2012).

The priority areas for achieving conservation goals established in this study seem to coincide with areas of high species richness and greater Total ED of amphibians in the Cerrado. According to our data, these areas incorporate most of the evolutionary history of Cerrado amphibians. The evolutionary history may be more important for maintaining ecosystem services than simply species richness (Cadotte et al. 2008). Conserving this diversity, we are also conserving the genotypic, phenotypic and functional diversity, giving more chances for ecosystems to respond appropriately to future changes (Cadotte and Davies 2010). In an assessment of the effects of climate change and habitat degradation on endemic amphibian species to the Cerrado, Silvano (2011) found that future scenarios are extremely unfavorable to the occurrence of these species. Thus, conservation strategies that consider the evolutionary diversity are mandatory tools for the future.

Since the resources available for conservation are limited and it is not possible to preserve the entire area due to conflicts with other social and especially economic interests, it is expected that the selection of these areas act as a starting point for decision makers. The areas considered here as priorities for the conservation of endemic Cerrado frogs should be investigated and appropriate plans for the conservation, management and control of these areas should be developed and implemented to ensure the existence of these species in the future.

**Acknowledgements** We thank Roseli Pellens for inviting us to write this chapter and for her review and comments. Anni Arponen, Annemarie Ohler and an anonymous reviewer for their comments and suggestions. Mario Barroso Ramos-Neto and Mariana da Silva Soares (Landscape Ecology Lab of WWF-Brazil) provided the planning units and cost layers for the Cerrado, built in a joint initiative with the Ministry of Environment. Hussam Zaher (MZUSP), Marcos Carvalho (ZUFMT), Rogério Bastos (ZUFG), Felipe Toledo (ZUEC), Célio Haddad (CFBH), Luciana Nascimento (MCN PUC-MG), Denise Rossa-Feres (DZSJRP), Franco Leandro de Souza (UFMS), José P. Pombal Junior (MNRJ), Ulisses Caramaschi (MNRJ) and Ulisses Galatti (MPEG) gently allowed us to examine museum specimens under their care. Mariana Gomes helped organizing the database. This study was funded FAPESP (fellowship 2007/51956-6 and grant 06/58011-4), CAPES, CNPq, FAP-DF, Conservation International (CP-FY 08/017 and CP-FY 08/018), Neotropical Grassland Conservancy, and Fundação O Boticário de Proteção à Natureza (0747\_20071 and 0798\_20082). Pequi –Pesquisa e Conservação do Cerrado provided logistical and institutional support for this project.

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Avice JC (2005) Phylogenetic units and currencies above and below the species level. In: Purvis A, Gittleman JL, Brooks TM (eds) *Phylogeny and conservation*. Cambridge University Press, Cambridge, pp 76–100
- Ball I, Possingham H (2000) MARXAN v1.8.2 – marine reserve design using spatially explicit annealing. University of Queensland, Brisbane
- Becker CG, Loyola RD (2008) Extinction risk assessments at the population and species level: implications for amphibian conservation. *Biodivers Conserv* 17:2297–2304
- Becker CG, Fonseca CR, Haddad CFB, Batista RF, Prado PI (2007) Habitat split and the global decline of amphibians. *Science* 318:1775–1777
- Becker CG, Loyola RD, Haddad CFB, Zamudio KR (2010) Integrating species life-history traits and patterns of deforestation in amphibian conservation planning. *Divers Distrib* 16:10–19
- Cadotte MW, Davies TJ (2010) Rarest of the rare: advances in combining evolutionary distinctiveness and scarcity to inform conservation at biogeographical scales. *Divers Distrib* 16:376–385
- Cadotte MW, Cardinale BJ, Oakley TH (2008) Evolutionary history and the effect of biodiversity on plant productivity. *Proc Natl Acad Sci* 105:17012–17017
- Carvalho RA, Cianciaruso MV, Trindade-Filho J, Sagnori MD, Loyola RD (2010) Drafting a blueprint for functional and phylogenetic diversity conservation in the Brazilian Cerrado. *Nat Conserv* 8(2):171–176
- Carvalho-Jr RR, Martinez RAM, Almeida SC, Ubaid FK, Maffei F, Jim J (2010) *Amphibia, Anura, Cycloramphidae, Proceratophrys moratoi* (Jim and Caramaschi 1980): distribution extension and new altitudinal record in state of Sao Paulo, Brazil. *Check List* 6:576–578
- CNUC (2010) Mapa das Unidades de Conservação. In: Cadastro Nacional de Unidades de Conservação. Ministério do Meio Ambiente, Brasília
- Cooper N, Bielby J, Thomas GH, Purvis A (2008) Macroecology and extinction risk correlates of frogs. *Glob Ecol Biogeogr* 17:211–221
- Diniz-Filho JAF, Bini LM, Vieira CM, Souza MC, Bastos RP, Brandão D, Oliveira LG (2004) Spatial patterns in species richness and priority areas for conservation of anurans in the Cerrado region, Central Brazil. *Amphibia-Reptilia* 25:63–75
- Diniz-Filho JAF, Bini LM, Pinto MP, Rangel TFLVB, Carvalho P, Vieira SL, Bastos RP (2007) Conservation biogeography of anurans in Brazilian Cerrado. *Biodivers Conserv* 16:997–1008
- Diniz-Filho JAF, Bini LM, Oliveira G, Barreto BS et al (2009) Macroecologia, biogeografia e áreas prioritárias para conservação no cerrado. *Oecologia Bras* 13:470–497
- Eiten G (1972) The Cerrado vegetation of Brazil. *Bot Rev* 38:201–341
- Eiten G (1982) Brazilian savannas. In: Huntley BJ, Walker BH (eds) *Ecology of tropical savannas*. Springer, Berlin, pp 25–47
- Elith J, Graham CH, Anderson RP et al (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129–151
- Graham CH, Hijmans RJ (2006) A comparison of methods for mapping species ranges and species richness. *Glob Ecol Biogeogr* 15:578–587
- Hero JM, Williams SE, Magnusson WE (2005) Ecological traits of declining amphibians in upland areas of eastern Australia. *J Zool* 267:221–232
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 25:1965–1978

- IBGE (2004) Mapa de Biomas do Brasil. In: Instituto Brasileiro de Geografia e Estatística, Brasília
- IBGE (2011) Mapa de compartimentos do relevo. In: Instituto Brasileiro de Geografia e Estatística, Brasília
- Isaac NJB, Turvey ST, Collen B, Waterman C, Baillie JEM (2007) Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PLoS One* 2:e296
- Isaac NJB, Redding DW, Meredith HM, Safi K (2012) Phylogenetically-informed priorities for amphibian conservation. *PLoS One* 7(8):e43912
- Klink CA, Machado RB (2005) Conservation of the Brazilian Cerrado. *Conserv Biol* 19:707–713
- Leite FSF, Juncá FA, Eterovick PC (2008) Status do conhecimento, endemismo e conservação de anfíbios anuros da Cadeia do Espinhaço, Brasil. *Megadiversidade* 4:182–200
- Machado RB, Ramos Neto MB, Pereira PGP et al (2004) Estimativas de perda da área do Cerrado brasileiro. In: Conservation International do Brasil, Brasília
- Maffei F, Ubaid FK, Jim J (2011) Discovery of the fifth population of a threatened and endemic toad of the Brazilian Cerrado, *Proceratophrys moratoii* (Anura, Cycloramphidae). *Herpetol Notes* 4:95–96
- Margules CR, Pressey RL (2000) Systematic conservation planning. *Nature* 405:243–253
- Mittermeier RA, Gil PR, Hoffman M et al (2004) Hotspots revisited: earth's biologically richest and most endangered terrestrial ecoregions. CEMEX and Agrupacion Sierra Madre, Mexico
- O'Grady JJ, Brook BW, Reed DH, Ballou JD, Tonkyn DW, Frankham R (2006) Realistic levels of inbreeding depression strongly affect extinction risk in wild populations. *Biol Conserv* 133:42–51
- Phillips SJ, Dudík M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31:161–175
- Posadas P, Esquivel DRM, Crisci JV (2001) Using phylogenetic diversity measures to set priorities in conservation: an example from southern South America. *Conserv Biol* 15:1325–1334
- Purvis A, Gittleman JL, Cowlishaw G, Mace GM (2000a) Predicting extinction risk in declining species. *Proc R Soc Lond Biol Sci* 267:1947–1952
- Purvis A, Jones KE, Mace GM (2000b) Extinction. *Bioessays* 22:1123–1133
- Rodrigues ASL, Andelman SJ, Bakarr MI et al (2003) Global gap analysis: towards a representative network of protected areas. In: Advances in applied biodiversity science 5. Conservation International, Washington, DC
- Rodrigues ASL, Akçakaya HR, Andelman SJ, Bakarr MI, Boitani L, Brooks TM, Chanson JS, Fishpool LDC, Fonseca GAB, Gaston KJ, Hoffmann M, Marquet PA, Pilgrim JD, Pressey RL, Schipper J, Sechrest W, Stuart SN, Underhill LG, Waller RW, Watts MEJ, Yan X (2004) Global gap analysis: priority regions for expanding the global protected-area network. *Bioscience* 54(12):1092–1100
- Rolim DC, Martinez RAM, Almeida SC, Ubaid FK, Maffei F, Jim J (2010) Amphibia, Anura, Cycloramphidae, *Proceratophrys moratoii* (Jim and Caramaschi, 1980): distribution extension and new altitudinal record in state of São Paulo, Brazil. *Check List* 6:576–578
- Silvano DL (2011) Distribuição e conservação de anfíbios no cerrado em cenários atuais e futuros. Thesis, Universidade de Brasília, Brasília
- Silvano DL, Segalla MV (2005) Conservation of Brazilian amphibians. *Conserv Biol* 19:653–658
- Siqueira S, Aguiar O Jr, Pansonato A, Giaretta AA, Strüssmann C, Martins I, Recco-Pimentel SM (2009) The karyotype of three Brazilian terrarana frogs (Amphibia, Anura) with evidence of a new *Barycholos* species. *Genet Mol Biol* 32:470–476
- Soares MS, Valdujo PH, Ramos-Neto MB, Oliveira MG, Rodrigues ST (2012) Building a probability surface of deforestation for Araguaia River basin. In: 8th International conference on ecological informatics, Brasília
- Stuart SN, Chanson JS, Cox NA, Young BE, Rodrigues ASL, Fischman DL, Waller RW (2004) Status and trends of amphibian declines and extinctions worldwide. *Science* 306:1783–1786

- Valdujo PH, Silvano DL, Colli GR, Martins M (2012) Anuran species composition and distribution patterns in Brazilian Cerrado, a neotropical hotspot. *S Am J Herpetol* 7(2):63–78
- Vazquez DP, Gittleman JL (1998) Biodiversity conservation: does phylogeny matter? *Curr Biol* 8:R379–R381
- Verdade VK, Valdujo PH, Carnaval AC et al (2012) A leap further: the Brazilian amphibian conservation action plan. *Alytes* 29(1–4):27–42
- Wake DB, Vredenburg VT (2008) Are we in the midst of the sixth mass extinction? A view from the world of amphibians. *Proc Natl Acad Sci* 105:11466–11473
- Young BE, Lips KR, Reaser JK et al (2001) Population declines and priorities for amphibian conservation in Latin America. *Conserv Biol* 15:1213–1223

# Global Spatial Analyses of Phylogenetic Conservation Priorities for Aquatic Mammals

Laura J. May-Collado, Carlos Zambrana-Torrelío, and Ingi Agnarsson

**Abstract** Several studies have shown how current climate change and human threats to aquatic environments are significantly impacting aquatic mammals worldwide. In response to these threats it is important to prioritize conservation efforts. A recent approach to evaluate conservation priorities is to combine information on species status from the International Union for Conservation of Nature (IUCN) Red List with information on the evolutionary history of the species from phylogenetic trees. This new approach provides a measure of biodiversity that complements estimates of species richness, adding evolutionary distinctiveness of species. Using near-complete species level phylogenies for the mammal groups with aquatic species (Carnivora, Cetacea, Sirenia) we calculated two measures (EDGE and HEDGE) of conservation priorities for 127 aquatic mammals under two scenarios of projected extinctions: a “pessimistic” approach, which represents a ‘worst case scenario’ for each species; and the “IUCN 50” a projected extinction risk over the next 50 years (Table 1). Then we analyzed the information to identify conservation priority areas (CPA) for aquatic mammals. We identified 22 CPAs distributed primarily along coastal waters in both northern and southern hemispheres. While thousands of marine protected areas (MPA) have been established in recent years, only 11.5 % of CPAs overlap with existing MPAs. Nevertheless, all phylogenetic CPAs identified in this study have also been proposed to be important by other independent studies using different prioritization criteria, highlighting the importance of focusing conservation efforts in these areas.

**Keywords** IUCN • Cetaceans • Pinnipeds • Manatees • EDGE • HEDGE

---

L.J. May-Collado (✉) • I. Agnarsson  
Department of Biology, University of Vermont,  
109 Carrigan Drive, Burlington, VT 05405, USA  
e-mail: [lmaycoll@uvm.edu](mailto:lmaycoll@uvm.edu); [Ingi.Agnarsson@uvm.edu](mailto:Ingi.Agnarsson@uvm.edu)

C. Zambrana-Torrelío  
EcoHealth Alliance, 460 West 34th Street – 17th floor, New York, NY 10001, USA  
e-mail: [zambrana@ecohealthalliance.org](mailto:zambrana@ecohealthalliance.org)

## Introduction

Current climate change and human threats to aquatic environments are significantly impacting aquatic mammals worldwide (Schipper et al. 2008; Davidson et al. 2011; Harnik et al. 2012; Harkonen et al. 2012). A recent study showed that 74 % of aquatic mammals experience high levels of human impact within their geographic range, with pollution and fisheries being the two most important threats (Davidson et al. 2011). Human overexploitation is proposed as the major cause of extinction of the Steller's sea cow (*Hydrodamalis gigas*), the tropical monk seal (*Monachus tropicalis*) (Hofman 1995) and more recently the Yangtze river dolphin or Baiji (*Lipotes vexillifer*) (Turvey et al. 2007). Also brought to the brink of extinction have been three additional aquatic mammals: the vaquita (*Phocoena sinus*) and the Hawaiian and Mediterranean monk seals (*Monachus schauinslandi* and *M. monachus*) whose populations have been reduced to fewer than 250 individuals (IUCN 2013.2). Some 28 % of aquatic mammals are threatened or near threatened under the International Union for Conservation of Nature (IUCN) risk classification and an additional 39 % are data deficient leaving only 33 % of aquatic mammal species at low risk. Furthermore, recent studies have suggested that even some of these species at relatively low risk should receive conservation attention due to their high evolutionary distinctiveness (May-Collado and Agnarsson 2011) and possible sudden changes in their risk of extinction due to rapidly changing environment (Davidson et al. 2011). Examples of such evolutionarily unique species are the data deficient Amazon River dolphin (*Inia geoffrensis*) and the walrus (*Odobenus rosmarus*). Both species are the only extant representatives of old lineages and live in habitats threatened by human activities and climate change, respectively. Additionally, based on the IUCN population trend information we estimate that most aquatic mammals are either decreasing (19.3 %) or unknown (64.3 %). In the light of these threats it is important to prioritize conservation efforts. In 2011, the Convention on Biological Diversity put in place a plan to protect 10 % of the world's marine and coastal ecological regions by 2020. Thus identifying geographic areas that could maximize these conservation goals is an urgent task (Davidson et al. 2011).

The International Union Conservation of Nature (IUCN) is the most influential conservation network in the world. Through its 'Red List' the IUCN has established conservation priorities prominently based on the imperilment levels of individual species. These categorizations are used by a number of organizations and government agencies to prioritize funding and conservation efforts. IUCN levels of imperilment result from the combination of several criteria including population size, evidence of population decline or recovery, distribution patterns and factors threatening species (<http://www.iucn.org>).

However, there are other criteria to establish conservation priorities including the use of 'umbrella species' also known as keystone or flagship species (Zacharias and Roff 2001), 'sentinel species' (Moore 2008), 'latent extinction risk' (Cardillo et al. 2006), regional and local habitat models (e.g., Praca et al. 2009; Azzellino et al. 2012), and identification of hot-spots of species richness (Davidson et al. 2011; Kaschner et al. 2011; Pompa et al. 2011). Recent approaches to identify conservation priorities combine information on species status from the IUCN with information



on the evolutionary history of the species from phylogenetic trees (Faith 1992, 2002, 2008; Faith et al. 2004; Redding and Mooers 2006; Schipper et al. 2008; Isaac et al. 2007; Mooers et al. 2008; Kuntner et al. 2009; Agnarsson et al. 2010; May-Collado and Agnarsson 2011). This new approach to conservation provides a measure of biodiversity that complements estimates of species richness, that is, of evolutionary distinctiveness of species. The fundamental argument is that the loss of evolutionarily unique species lacking close relatives represents a greater loss of biodiversity than the loss of species whose evolutionary history is, to a great extent, preserved in other closely related species (May-Collado and Agnarsson 2011).

Considering both evolutionary histories of lineages and species' threats can help the goal of maximizing biodiversity conservation. This approach of identifying areas protecting both threatened species and containing high phylogenetic diversity provides another tool for decision-making. Here we examine global patterns of aquatic mammal phylogenetic conservation priorities using four recently proposed metrics for 127 aquatic mammals. We identify Conservation Priority Areas (CPAs), estimate the degree to which such areas are contained within current Marine Protected Areas (MPA), and suggest areas where focusing future conservation effort might be particularly valuable.

## Material and Methods

We used the most detailed primary-data species-level phylogenies available for the three major mammalian groups containing aquatic species Cetacea (May-Collado and Agnarsson 2006; May-Collado et al. 2007), Carnivora (Agnarsson et al. 2010), and Afrotheria (Kuntner et al. 2009). The conservation status for 127 aquatic mammals was obtained from the IUCN Red List of Threatened Species database (2010.4–2013.2) and transformed to probability estimates of extinction risk using two of the methods discussed in Mooers et al. (2008) “pessimistic” and “IUCN50”. The “pessimistic” method is an arbitrary transformation that designates a sizable probability of extinction to every category. So, even for the ‘least concern’ species has a probability of 0.2, which is much higher than in the IUCN 50 scenario (see Table 1) (Mooers et al. 2008). The “IUCN50” is a projection of extinction risk over the next 50 years given current conservation status, proposed by the IUCN. This scenario assumes that species in the ‘least concern’ category are essentially ‘safe’, assigning to them low probability of extinction (Mooers et al. 2008) (Table 1). We selected these two transformation methods because they offer contrasting scenarios based on how they treat species that are currently thought to be at relatively low risk.

Using these transformation methods we calculated conservation priority measures using the TUATARA module version 1.01 (Maddison and Mooers 2007) in the evolutionary analysis package MESQUITE version 2.75 (Maddison and Maddison 2011). We used the conservation priority methods EDGE (Evolutionary Distinct, Globally Endangered) (<http://www.edgeofexistence.org>), which measures evolutionary distinctiveness (ED) weighted by current IUCN levels of extinction risk. EDGE scores are equivalent to a logarithmic transformation of the product of

**Table 1** Extinction probabilities for IUCN levels of imperilment transformed into extinction probabilities using “pessimistic” and IUCN 50 transformations, as proposed by Mooers et al. 2008

IUCN level of imperilment	Pessimistic	IUCN 50
Least concern <sup>a</sup>	0.2	0.0001
Data deficient <sup>b</sup>	0.3	0.005
Near threatened <sup>a</sup>	0.4	0.01
Vulnerable <sup>a</sup>	0.8	0.1
Endangered <sup>a</sup>	0.9	0.667
Critically endangered <sup>a</sup>	0.99	0.999

<sup>a</sup>Mooers et al. 2008  
<sup>b</sup>May-Collado and Agnarsson 2011

a species’ evolutionary distinctiveness and the probability it will go extinct (see Isaac et al. 2007). We also use the conservation priority method HEDGE (Heightened EDGE), which like EDGE measure evolutionary distinctiveness by IUCN levels of extinction but additionally considers how future extinction of species will affect the evolutionary distinctiveness of remaining species. In sum, HEDGE estimates the expected terminal branch length of the focal species in light of future extinction risk (Steel et al. 2007). Both conservation methods generate fixed probabilities of extinction as described in Table 1. For more information on how IUCN levels of imperilments are transformed into probabilities of extinction see Moores et al. (2008). Because there are not estimated probabilities of extinction available for data deficient species we arbitrarily assigned an extinction risk score in between the two lowest IUCN extinction categories: least concern and near threatened (Table 1). All four metrics were calculated using both ‘raw’ branch lengths (estimated by MrBayes) as they contain information on the unique evolutionary information of terminal taxa and ultrametrized trees. Furthermore, with the purpose of comparing this approach to identify conservation priority areas with other commonly used conservation prioritization criteria we calculated evolutionary distinctiveness (ED) (Isaac et al. 2007) also implemented in the TUATARA module and gathered information on species richness.

To identify conservation priority areas (CPAs) we used distribution range maps from the IUCN spatial database (2013) as a baseline to produce species richness, ED, EDGE and HEDGE maps, under both IUCN extinction probabilities transformation methods, pessimistic and IUCN50. The IUCN spatial database depict species’ range distribution as polygons based on the extent of occurrence, which is defined as the area contained within a minimum convex hull around species’ observations or records. This convex hull or polygon is further improved by including areas known to be suitable or by removing unsuitable or unoccupied areas based on expert knowledge.

For each species the distribution range was converted to a grid system with cells of 5’ × 5’ (approximately 10 × 10 km at the Equator line). This spatial resolution was selected for its practical compromise between intensive computing and a reasonable representation of geographic patterns. Traditionally, a one-degree cell (100 × 100 km) has been used in macroecological analyses, but there is no ecological reason behind

the resolution. More importantly, it has been shown that higher spatial resolutions distort the geographical patterns of species richness (Rahbek 2005; Graham and Hijmans 2006). In contrast, lower spatial resolutions minimize the overestimation of distribution ranges, in particular of those species with small range distributions. For example, Rondinini et al. (2011) used a resolution of 300 × 300 m for their estimations of global mammal species richness.

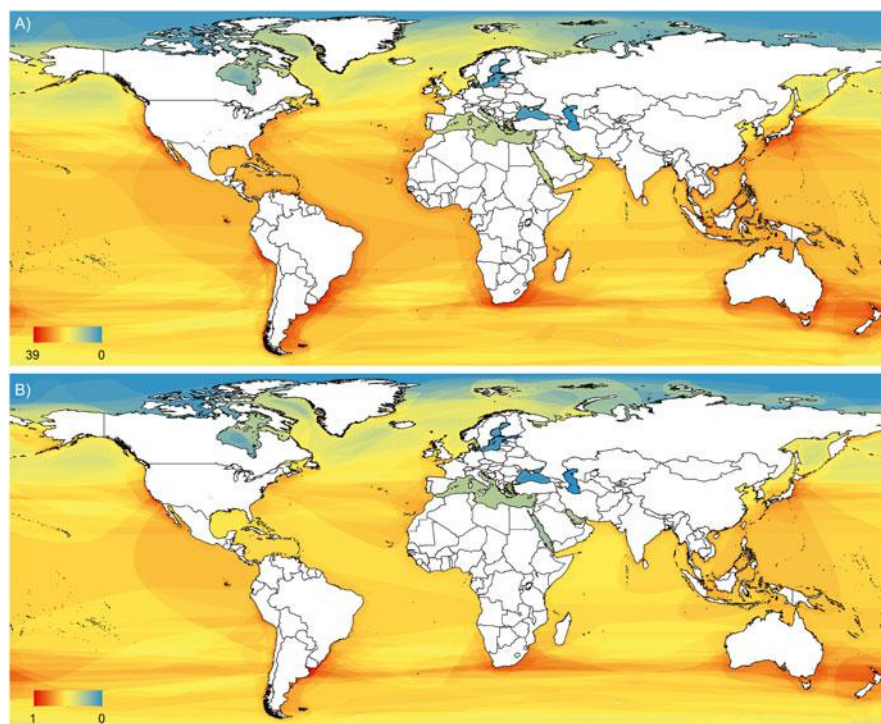
Species's presence in each 5' × 5' grid cell was assigned with a value of one. The same procedure was repeated to assign the estimated values of ED, EDGE and HEDGE to the grid cells of each species' occurrence, and maps were calculated by overlying individual grids. For example, the species richness map represents the sum of all presence grids. Under this spatial framework CPAs represent areas with the highest scores due to a high number of species regardless of ED, EDGE, and HEDGE scores or to a few species with high probability scores.

To understand how well these patterns of aquatic mammal conservation priorities are already included in existing MPAs, we used information from the World Database on Protected Areas website (<http://protectedplanet.net/>). We calculated the percentage of each species range within all designated MPAs of the world and to preserve areal relationships we first re-projected both the species ranges and the MPAs, to an equal area using Behrmann projection. We then iteratively intersected each marine species range with all the MPAs using the function *joinPolys* with the operator 'INT' (intersection) from the package PBS-Mapping (version 2.67). Species range percentage within all MPAs was calculated by contrasting the sum of their intersected areas with the total species range area. All analyses were performed in R (R Core Team 2013) and the final maps were created using ArcGIS v10.1.

## Results

Aquatic mammal species richness and the sum of species evolutionary uniqueness peaked in coastal waters of both northern and southern hemispheres. Both metrics showed high scores at the coasts of California and Japan in the northern hemisphere, and along the coast of Peru, Argentina, Uruguay, Southern Brazil, South Africa, Southern Australia, Tasmania, and New Zealand in the southern hemisphere (Fig. 1).

All methods identified the Hawaiian, Kurl, and Aleutian Islands, the coastal waters of northern California, Nouadhibou, Yangtze River, southern Brazil to Argentina, where both metrics had the highest cumulative scores in Rio de la Plata, and Southern Australia and Japan as CPAs (Figs. 2 and 3). Furthermore, the Mediterranean Sea was identified as other CPA under pessimistic EDGE and IUCN 50 HEDGE (Fig. 2a, d); South Africa, Patagonia, New Zealand, Tasmania, Bay of Bengal, Arabian Sea, Indonesia, and South China Sea under pessimistic HEDGE (Fig. 2b); and North Atlantic Ocean and Galapagos Islands, under IUCN50 EDGE (Fig. 2c). Figure 3 summarizes these four conservation priority metrics into a single map showing all CPAs. In each of them we highlight examples of top ranking phylogenetic conservation priority species (Table 2).



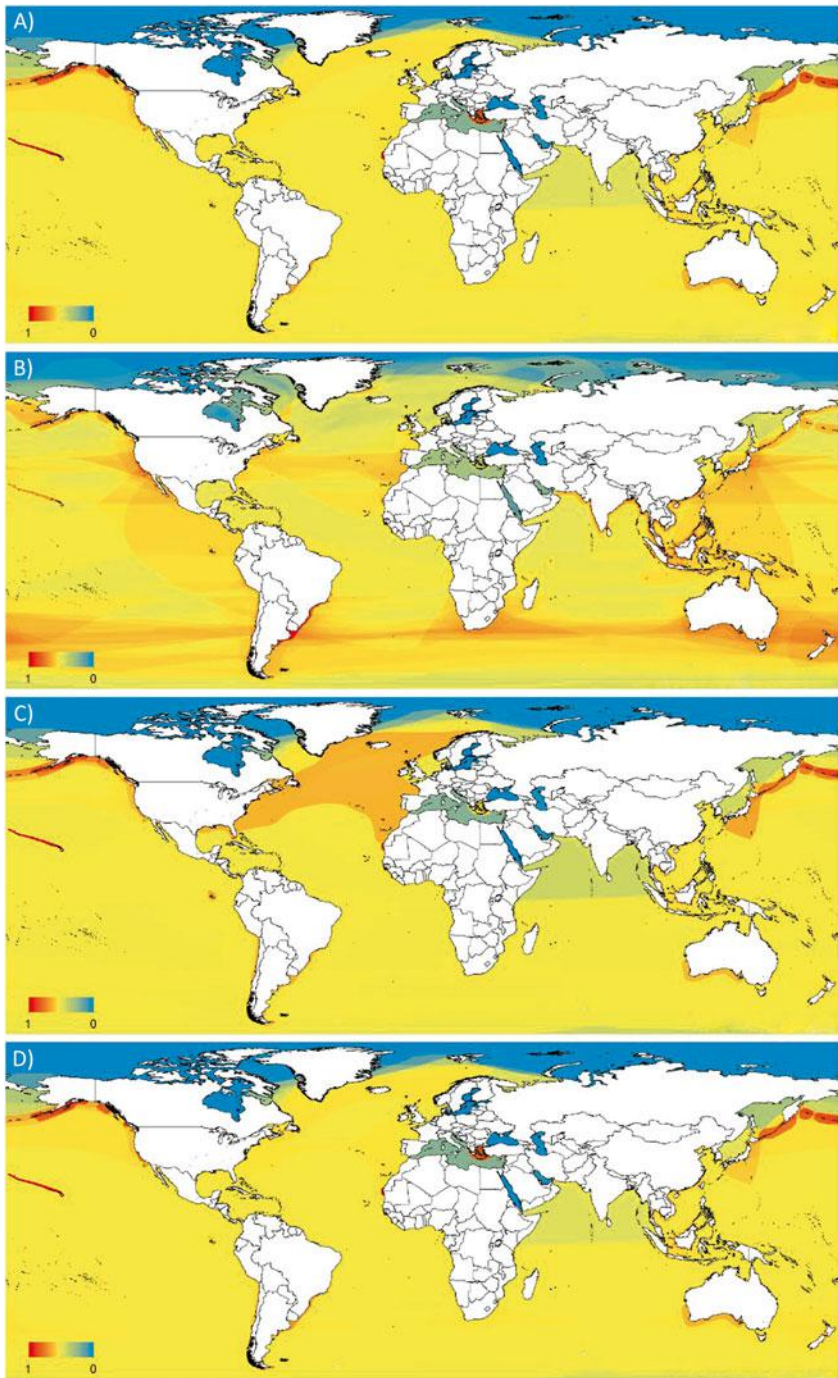
**Fig. 1** Global patterns of aquatic mammal (a) species richness and (b) evolutionary distinctiveness (ED) (red tones represent the highest scores, cold colors indicate the lowest ones)

When we analyzed the overlap with designated marine protected areas we found that although there are 847 of MPAs, most are very small (mean=201 km<sup>2</sup>; sd=424 km<sup>2</sup>) when compared to the average size of the CPAs (2,269,520 km<sup>2</sup>, sd=4,313,740 km<sup>2</sup>). Nearly 50 % of large MPAs (those MPAs with an area greater than 500 km<sup>2</sup>) do not overlap with any CPAs failing to protect important areas of aquatic mammal diversity as identified in this study (Fig. 4). Table 2 provides information on the % overlap of top ranking phylogenetic conservation priority species (see highlighted species in Fig. 3) distribution that is under any form of protected area. With the exception of the Galapagos sea lion and fur seal and the Hawaiian monk seal most top conservation priority species habitat is unprotected.

## Discussion

Here we provide the first spatial analysis of phylogenetic conservation priorities for aquatic mammals. We consider four methods that essentially reflect two possible scenarios differing in how extinction risk is evaluated for the lower threat categories and data deficient species: the pessimistic approach and the IUCN50 approach. The two approaches give dramatically different results, in many cases highlighting





**Fig. 2** Global patterns of conservation priorities using the conservation priority methods: (a) EDGE Pessimistic, (b) HEDGE Pessimistic, (c) EDGE IUCN 50 and (d) HEDGE IUCN 50 (red tones represent the highest scores, cold colors indicate the lowest ones)



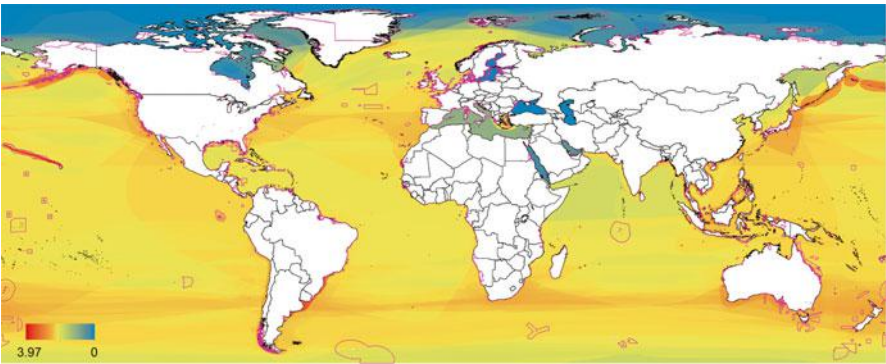
**Table 2** Examples of aquatic mammal species that ranked among the top species for one or all methods, their corresponding conservation priority areas (CPAs), population status, and overall overlap with marine protected areas (MPAs)

Species	IUCN value	Population trend	Conservation priority area	Overlap of species range and MPAs (%)
Cetaceans				
<i>Lipotes vexillifer</i>	Critically endangered	Unknown	Yangtze River	<1 %
Baiji				
<i>Pontoporia blainvillei</i>	Vulnerable	Unknown	South America	<1 %
Franciscana			Brazil to Argentina	
<i>Neophocaena asiaeorientalis</i> and <i>N. phocaenoides</i>	Vulnerable	Decreasing	Japan, East and South China Sea, Bay of Bengal, Arabian Sea	Not in database/2.35 % respectively
Finless porpoises				
<i>Phocoena sinus</i>	Critically endangered	Decreasing	Baja California	33.6 %
Vaquita				
<i>Cephalorhynchus hectori</i>	Endangered	Decreasing	New Zealand	<1 %
Hector's Dolphin				
<i>Eubalaena glacialis</i>	Endangered	Unknown	North Atlantic Ocean	<1 %
North Atlantic Right Whale				
<i>Eubalaena japonica</i>	Endangered	Unknown	Japan	<1 %
North Pacific Right Whale				
Pinnipeds				
<i>Monachus schauinslandi</i>	Critically endangered	Decreasing	Hawaii	69.1 %
Hawaiian Monk Seal				
<i>Monachus monachus</i>	Critically endangered	Decreasing	Mediterranean Sea, Madeira, Nouadhibou	4.54 %
Mediterranean Monk Seal				
<i>Enhydra lutris</i>	Endangered	Decreasing	Gulf of Alaska, California	6.9 %
Sea Otter				
<i>Lontra felina</i>	Endangered	Decreasing	South America	12.4 %
Marine Otter			Peru and Chile	
<i>Callorhinus ursinus</i>	Vulnerable	Decreasing	Kurl and Aleutian Islands, Gulf of Alaska, California	1.6 %
Northern Fur Seal				
<i>Eumetopias jubatus</i>	Near threatened	Increasing	Kurl and Aleutian Islands, Gulf of Alaska, California	3.8 %
Steller Sea Lion				

(continued)

**Table 2** (continued)

Species	IUCN value	Population trend	Conservation priority area	Overlap of species range and MPAs (%)
<i>Neophoca cinerea</i>	Endangered	Decreasing	Southern Australia	6.4 %
Australian Sea Lion				
<i>Zalophus wollebaeki</i>	Endangered	Decreasing	Galapagos Island	60.2 %
Galapagos Sea Lion				
<i>Odobenus rosmarus</i>	Data deficient	Unknown	Aleutian Islands	6.86 %
Walrus				
<i>Arctocephalus galapagoensis</i>	Endangered	Decreasing	Galapagos Island	99.5 %
Galapagos Fur Seal				
<i>Erignathus barbatus</i>	Least concern	Stable	Aleutian Islands	3.4 %
Bearded seal				



**Fig. 4** Overlap of conservation priority areas (CPAs) and global distribution of designated marine protected areas (MPAs) (in pink). Lower panels highlight those CPAs with low spatial overlap with MPAs

different CPAs. The pessimistic approach gives more weight to phylogenetic diversity, while the IUCN50 gives more weight to extinction risk of species (May-Collado and Agnarsson 2011). Given that both perspectives are valid, we used a combination of these two scenarios to identify Conservation Priority Areas, emerging as highly ranking under one or both of these approaches (see Fig. 3). These results provide a tool for conservation planning for aquatic mammals that supplements previous spatial studies using other prioritization criteria (Davidson et al. 2011; Kashner et al. 2011; Pompa et al. 2011), and may thus be useful in helping to guide future conservation effort.

Our results indicate that accumulative evolutionary distinctiveness and conservation priorities are in general concentrated in coast waters. This pattern could be an artifact of survey effort in coastal waters and in general reflect that aquatic mammal



survey coverage, whether coastal or oceanic, is very limited (Jewell et al. 2012). Less than 25 % of the ocean surface has been surveyed and only 6 % has been covered frequently enough to allow estimations of population trends (Kaschner et al. 2012). In addition, spatial coverage is also significantly biased towards some ocean basins. Karchner et al. (2012) reported that with the exception of Antarctic waters, survey coverage was biased toward the northern hemisphere, especially US and northern European waters, which may explain the consensus among methods identifying the Aleutian and Hawaiian Islands as CPAs. Nevertheless, despite this potential data bias most CPAs were found in the southern hemisphere, suggesting that phylogenetic conservation priority methods do not simply reflect sampling effort, but identify areas that contain aquatic mammal communities including both evolutionarily unique species and those at risk.

As we discussed in our methods, CPAs were the result from the cumulative values for each metric in each cell. Thus CPAs may reflect a large number of species varying in conservation priority values or possibly only a few species with high values. The later seems to be the case for Hawaii, Nouadhibou, Madeira Island, Yangtze river, and Southern Brazil-Argentina where highly evolutionary unique species are endemic to those areas (Table 2). Other CPAs such as California, Southern Australia and New Zealand include many species, but only some of which are evolutionarily unique species. These areas are part of ranges of several species with very broad distribution ranges such as the sperm whale, pygmy and dwarf sperm whales, and blue whale. Interestingly, previous studies that have used species richness to identify ‘hotspot’ of aquatic mammal diversity (Pompa et al. 2011) and a combination of levels of imperilment with intrinsic and extrinsic factors to identify high risk areas for aquatic mammals (Davidson et al. 2011) agree with the CPAs identified here. Davidson et al. (2011) identified five major global hotspots of marine mammal species at risk. Within these major hotspots several locations overlap with those found in this study: Aleutian Islands, Alaska, California, Galapagos, Patagonia, South Africa, Japan, Indonesia, South Australia, and New Zealand. Pompa et al. (2011) identified nine ‘hotspots’ based solely on species richness and 11 irreplaceable key conservation sites, based on the presence of endemic species, five of these sites Hawaiian and Galapagos Islands, Mediterranean Sea, and the Yangtze river network were also identified as CPAs. Finally, Kashner et al. (2011) using an environmental suitability model predicted highest marine mammal richness in New Zealand, Japan, Baja California, Galapagos, the Southeast Pacific and Southern Ocean, all congruent with our study.

Within the CPAs identified here we highlight the presence of several top ranking conservation priority species among those are the extant monk seals (see Table 2). The IUCN has estimated a 68 % reduction of Hawaiian monk seal abundance in the past 49 years, and projects an 86 % reduction in the next 15 years. The future for the Mediterranean monk seal seems bleak, current population estimates are about 350–450 individuals (IUCN 2013). The Cap Blanc population in Nouadhibou is probably the most threatened, with less than 220 individuals. This is the last population with colonial structure, so its loss would also lead to the loss of a peculiar behavior amongst monk seals (e.g., Gonzalez 2006; Martinez-Jauregui et al. 2012; Gonzalez

and Fernandez de Larrinoa 2013). Both species are facing fragmentation of their habitat that overlaps with a number of human activities, some of which affects them directly as bycatch in gillnets and bottom trawl nets particularly in the case of the Mediterranean monk seal (Gonzalez and Fernandez de Larrinoa 2013). In 1988 Hawaiian monk seal habitats were declared as 'critical areas' by the Endangered Species Act, limiting several federally authorized activities such as permits for fishing, coastal development, and a number of military activities. However, the designation of critical habitat offers limited protection, allowing a number of non-federal activities such as boating and jet-skiing, and tour operations that might have an indirect impact in their recovery. For the Mediterranean monk seals, surveyed protected marine areas might help mitigate interactions with fisheries (Rowwe 2007). Despite growing efforts in protecting the species, only 4.5 % of their habitat is currently protected. In areas where it is protected such as in Madeira Island, Portugal the creation of a natural reserve and change in fishing gear has halted monk seal decline and helped their recovery (Pires et al. 2008; Hale et al. 2011) but such protected area is not in place in Nouadhibou yet (Gonzalez and Fernandez de Larrinoa 2013).

For cetaceans we would like to highlight the river dolphin Baiji and the coastal Franciscana. The Baiji dolphin is likely extinct, no sightings of the species have been made in recent years. In 2005, its population size was estimated to be less than 100 individuals (Dudgeon 2005). A number of restoration efforts have been made including establishing of natural and seminatural reserves in the middle and lower parts of the Yangtze River, and breeding programs. However, the extent of the human impact on this species habitat may not allow its recovery with ~5 % of the world's total population living along the Yangtze River (Yang et al. 2006). In contrast with the Baiji, the Franciscana population size is considered 'healthy' by the IUCN. However, the extent of bycatch mortality by nearshore gillnets in southern Brazil results in thousands of individuals killed every year, this is a major reason for concern (Danilewicz et al. 2010; Prado et al. 2013), particularly when less than 1 % of the species habitat is under protection. Current MPAs within the species range are few, most are small, sparse, and outside the Rio del Plata area where conservation priority values peaked. Conservation priority species such as the Hector's dolphins in New Zealand and finless porpoises (see Figs. 3 and 4) may benefit from expansion of local MPAs, which currently protect only a small percent of their habitat. In contrast, conservation priority species with global distribution such as sperm, blue, sei and fin whales, may benefit for a multinational management approach at the species level combined with protected areas in key breeding and feeding grounds.

Our results offer a spatial phylogeny-based conservation prioritization for aquatic mammals that complements prior findings. Given urgent need to invest management and conservation efforts and the Convention on Biological Diversity plan to protect 10 % of the world's marine and coastal ecological regions by 2020, such analyses should be helpful tools in identifying important areas for consideration.

**Acknowledgements** We thank EcoHealth Alliance for providing us with the resources to produce the spatial analysis and to Lisa Ballance, Roseli Pellens, and an anonymous reviewer for their comments that greatly improve this manuscript.

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Agnarsson I, Kuntner M, May-Collado LJ (2010) Dogs, cats, and kin: a molecular species-level phylogeny of Carnivora. *Mol Phylogenet Evol* 54:726–745
- Azzellino A, Panigada S, Lanfredi C et al (2012) Predictive habitat models for managing marine areas: spatial and temporal distribution of marine mammals within the Pelagos Sanctuary (Northwestern Mediterranean sea). *Ocean Coast Manag* 67:63–74
- Cardillo M, Mace GM, Gittleman JL, Purvis A (2006) Latent extinction risk and the future battlegrounds of mammal conservation. *Proc Natl Acad Sci U S A* 103:4157–4161
- Danilewicz D, Moreno IB, Ott PH et al (2010) Abundance estimate for a threatened population of franciscana dolphins in southern coastal Brazil: uncertainties and management implications. *J Mar Biol Assoc UK* 90:1–9
- Davidson AD, Boyer AG, Kim H et al (2011) Drivers and hotspots of extinction risk in marine mammals. *Proc Natl Acad Sci U S A* 109:3395–3400
- Dudgeon D (2005) Last chance to see: ex situ conservation and the fate of the baiji. *Aquat Conserv Mar Freshw Ecosyst* 15:105–108
- Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserve* 61:1–10
- Faith DP (2002) Quantifying biodiversity: a phylogenetic perspective. *Conserv Biol* 16:248–252
- Faith DP (2008) Threatened species and the potential loss of phylogenetic diversity: conservation scenarios based on estimated extinction probabilities and phylogenetic risk analysis. *Conserv Biol* 22:1461–1470
- Faith DP, Reid CAM, Hunter J (2004) Integrating phylogenetic diversity, complementarity, and endemism for conservation assessment. *Conserv Biol* 18:255–261
- González LM (2006) The action plan for the recovery of the Mediterranean monk seal in the Eastern Atlantic. O.A. Parques Nacionales. Ministerio de Medio Ambiente, Madrid, pp 290
- González LM, Fernández de Larrinoa P (2013) Mediterranean monk seal *Monachus monachus* distribution and fisheries interactions in the Atlantic Sahara during the second half of the 20th century. *Mammalia* 77:41–49
- Graham CH, Hijmans RJ (2006) A comparison of methods for mapping species ranges and species richness. *Global Ecol Biogeogr* 15:578–587
- Hale R, Pires R, Santos P, Karamanlidis A (2011) Mediterranean monk seal (*Monachus monachus*): fishery interactions in the Archipelago of Madeira. *Aquat Mamm* 37:298–304
- Harkonen T, Harding KC, Wilson S et al (2012) Collapse of a marine mammal species driven by human impacts. *PLoS One* 7:e43130
- Harnik PG, Lotze HK, Anderson SC et al (2012) Extinctions in ancient and modern seas. *Trends Ecol Evol* 27:608–617
- Hofman RJ (1995) The changing focus of marine mammal conservation. *Trends Ecol Evol* 10:462–464
- Isaac NJB, Turvey ST, Collen B et al (2007) Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PLoS One* 2:e296
- Jewell R, Thomas L, Harris CM et al (2012) Global analysis of cetacean line-transect surveys: detecting trends in cetacean density. *Mar Ecol Prog Ser* 453:227–240
- Kashner K, Tittensor DP, Ready J et al (2011) Current and future patterns of global marine mammal biodiversity. *PLoS One* 6:e19653
- Kashner K, Quick NJ, Jewell R et al (2012) Global coverage of cetacean line-transect surveys: status quo, data gaps, and future challenges. *PLoS One* 7:e44075

- Kuntner M, May-Collado LJ, Agnarsson I (2009) Phylogeny and conservation priorities of afrotherian mammals (Afrotheria, Mammalia). *Zool Scr* 40:1–15
- Maddison WP, Maddison DR (2011) Mesquite: a modular system for evolutionary analysis. Version 2.75. Available via <http://mesquiteproject.org>
- Maddison WP, Mooers AØ (2007) Tuatara: conservation priority in a phylogenetic context, Version 1.0. Available: <http://mesquiteproject.org/packages/tuatara>
- Martínez-Jaúregui M, Tavecchia G, Cedenilla MA et al (2012) Population resilience of the Mediterranean monk seal *Monachus monachus* at Cabo Blanco peninsula. *Mar Ecol Prog Ser* 461:273–281
- May-Collado LJ, Agnarsson I (2006) Cytochrome b and Bayesian inference of whale phylogeny. *Mol Phylogenet Evol* 38:344–354
- May-Collado LJ, Agnarsson I (2011) Phylogenetic analysis of conservation priorities for aquatic mammals and their terrestrial relatives, with a comparison of methods. *PLoS One* 6(7):e22562
- May-Collado LJ, Agnarsson I, Wartzok D (2007) Phylogenetic review of tonal sound production in whales in relation to sociality. *BMC Evol Biol* 7:136
- Mooers A, Faith DP, Maddison WP (2008) Converting endangered species categories to probabilities of extinction for phylogenetic conservation prioritization. *PLoS ONE* 3:e3700
- Moore SE (2008) Marine mammals as ecosystem sentinels. *J Mammal* 89:534–540
- Pires R, Costa Neves N, Karamanlidis A (2008) The critically endangered Mediterranean monk seal *Monachus monachus* in the archipelago of Madeira: priorities for conservation. *Oryx* 42:278–285
- Pompa S, Ehrlich P, Ceballos G (2011) Global distribution and conservation of marine mammals. *Proc Natl Acad Sci U S A* 108:13600–13605
- Praca E, Gannier A, Das K, Laran S (2009) Modelling the habitat suitability of cetaceans: example of the sperm whale of the northwestern Mediterranean Sea. *Deep Sea Res I* 56:648–657
- Prado JHF, Secchi ER, Kinas PG (2013) Mark-recapture of the endangered franciscana dolphin (*Pontoporia blainvillei*) killed in gillnet fisheries to estimate past bycatch from time series of stranded carcasses in southern Brazil. *Ecol Indic* 32:35–41
- R Core Team (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
- Rahbek C (2005) The role of spatial scale and the perception of large-scale species-richness patterns. *Ecol Lett* 8:224–239
- Redding DW, Mooers AØ (2006) Incorporating evolutionary measures into conservation prioritization. *Conserv Biol* 20:1670–1678
- Rondinini C, Di Marco M, Chiozza F et al (2011) Global habitat suitability models of terrestrial mammals. *Philos Trans R Soc Lond B Biol Sci* 366:2633–2641
- Rowe SJ (2007) A review of methodologies for mitigating incidental catch of protected marine mammals. DOC research & development series 283. Science & Technical Publishing, New Zealand Department of Conservation, Wellington
- Schipper J, Chanson J, Chiozza F et al (2008) The status of the world's land and marine mammals: diversity, threat, and knowledge. *Science* 322:225–230
- Steel M, Mimoto A, Mooers AØ (2007) Hedging one's bets: quantifying a taxon's expected contribution to future phylogenetic diversity. *Evol Bioinform* 3:237–244
- The IUCN Red List of Threatened Species (2013). <http://www.iucnredlist.org>
- Turvey ST, Pitman RL, Taylor BL et al (2007) First human-caused extinction of a cetacean species? *Biol Lett* 3:537–540
- Yang G, Bruford MW, Wei F, Zhou K (2006) Conservation options for the Baiji: time for realism? *Conserv Biol* 20:620–622
- Zacharias MA, Roff JC (2001) Use of focal species in marine conservation and management: a review and critique. *Aquat Conserv Mar Freshw Ecosyst* 11:59–76

# Metapopulation Capacity Meets Evolutionary Distinctness: Spatial Fragmentation Complements Phylogenetic Rarity in Prioritization

Jessica K. Schnell and Kamran Safi

**Abstract** Many species have declined or already gone extinct due to the human activities across the world causing what is termed the current sixth mass extinction event. The biggest determinant of species survival is the availability of a network of suitable habitat, affecting population size and eventual extinction risk. Considering that modern technology allows us to efficiently quantify habitat loss, species distribution data can inform us of the required minimum connectivity of habitats. Evolutionary distinctiveness (ED) is already part of conservation schemes to prioritize rare traits and unique phylogenetic history. However, so far none of these prioritisations quantifies the spatial constraints of a species to estimate long-term persistence based on the fragmentation of the landscape. Metapopulation capacity ( $\lambda_M$ ) is one such measurement for quantifying fragmentation. Here we propose a combination of metapopulation capacity and phylogenetic distinctiveness to prioritize important specific habitat patches for evolutionary distinct species. We applied the new framework to prioritize island mammals and found Data Deficient and Least Concern species with a high combined value in ED and  $\lambda_M$ . Balancing between the extinction risks of solitary islands and the potential loss of unique evolutionary history of rare species on these islands can be a worthwhile exercise in prioritization schemes.

**Keywords** Habitat fragmentation • Metapopulation capacity • EDGE of existence • Conservation • Islands

---

J.K. Schnell (✉) • K. Safi

Department of Migration and Immuno-ecology, Max Planck Institute for Ornithology,  
Am Obstberg 1, 78315 Radolfzell, Germany

Department of Biology, University of Konstanz, 78457 Konstanz, Germany  
e-mail: [schnell@orn.mpg.de](mailto:schnell@orn.mpg.de)

© The Author(s) 2016

R. Pellens, P. Grandcolas (eds.), *Biodiversity Conservation and Phylogenetic Systematics*, Topics in Biodiversity and Conservation 14,  
DOI 10.1007/978-3-319-22461-9\_16

## Introduction

Conservation is an increasing necessity for the world (Pimm et al. 1995), and one that requires immediate action. Extinction occurs at a progressive rate, and we want to mitigate it before more species, known and unknown, are lost forever (Loehle and Eschenbach 2012). What is now recognised as the sixth mass extinction event is clearly attributable to anthropogenic action, mainly in the last few decades (Barnosky et al. 2011; Pereira et al. 2012). We will face great future challenges in preserving life on Earth, or at the least, in slowing down the rate of species loss. By setting priorities, as to which species or areas should receive the immediate attention, we can focus conservation efforts and resources in a bid to minimize the global biodiversity decline.

### *Evolutionary Distinctness*

The EDGE of Existence program is a conservation program guided by a straightforward combination of two characteristics, evolutionary distinctness (ED) and global endangerment (GE); simply put, it prioritizes for phylogenetic rarity/uniqueness, and threat status (Isaac et al. 2007; Collen et al. 2011). ED is a species-level prioritisation that weighs each species by its relative importance with regards to the unique evolutionary history it represents as a consequence of its specific phylogenetic history. The calculation of ED is essentially distributing the amount of shared ancestry from the root to tip of a phylogenetic tree by hierarchically distributing each branch's length equally to all of its subtending branches, thus accumulating evolutionary history up to the species level. This is calculated by taking the branch length and dividing by the number of species leading up to that branch, and then the ED of a species is the sum of these values for all branches from which the species is descended (Isaac et al. 2007).

For including global endangerment, the EDGE score adds the global IUCN assessment criteria by adding a quasi-probability of extinction associated with a doubling of extinction risk with increasing threat category (Isaac et al. 2012). However, the IUCN criteria include a wide, varied assortment of factors to determine the threat status of every species in the world. While some aspects of the criteria are standardized and quantified, others are somewhat equivocal terminology, ultimately based on expert opinion, particularly so when data is lacking (IUCN 2013).

### *Spatial Analysis*

The importance of habitat to animals cannot be overstated, particularly when their long-term survival is at stake. It is important to take advantage of high-resolution habitat data and furthermore, to analyse and quantify the available space (Kerr and

Ostrovsky 2003; Gillespie et al. 2008; Kearney and Porter 2009). By first focusing on the spatial aspects of a threat status, we may better assess what is often the main driver of species' extinction. Then conservation areas can target protection of those species with rare traits that are simultaneously habitat-limited.

With access to environmental data that fundamentally shapes species distributions, we now have the possibility to reveal what we need to prioritize through modelling (Moilanen et al. 2009). Major conservation tools often focus on protecting either particular species or specific areas. Good examples of species prioritisation schemes include the IUCN Red List and the phylogenetically informed EDGE of Existence concept (Isaac et al. 2007; IUCN 2013). In combination with spatial approaches, prioritization allows us to recognise the urgency and mitigate using what limited resources are available to conservationists. So, how to refine this focus to some criterion that is both highly quantifiable and universally important?

## *Metapopulation Capacity*

Gathering distribution estimates is difficult for rare or elusive species, and gathering population data more so, often because of the inaccessibility of their habitat which in turn biases ecological studies around the world (Martin et al. 2012). Population viability analysis can predict species trends, but such modelling also requires a certain level of life history data (Brook et al. 2000) that is unavailable for the full spectrum of species of concern. We have quality landscape data, but we want to know how this affects the species that reside in such landscapes.

Once such tactic is looking at metapopulation capacity ( $\lambda_M$ ), calculated from spatial input (i.e. patch areas and distances) of spatially explicit metapopulation models. We can consider metapopulation theory as a compromise between landscape ecology and species distribution modelling (Hanski 1998). The resulting value is the capacity of a landscape to support long-term species persistence (Hanski and Ovaskainen 2000).  $\lambda_M$  is one way of assessing risk for species living in fragmented landscapes, as a relative quantification of fragmentation. Schnell and co-workers (2013a) further developed a modification of  $\lambda_M$  for large-scale landscapes.

Species' habitats fragment over time, often due to human land use changes, and eventually the animals grow increasingly endangered. When isolated populations are too small and isolated, the metapopulation as a whole goes extinct. Therefore,  $\lambda_M$  can be useful in prioritising species conservation from a spatial standpoint (Hanski and Simberloff 1997; Hanski and Ovaskainen 2002; Schnell et al. 2013b). In the realm of conserving evolutionary history we can argue in much the same way, so combining the  $\lambda_M$  and ED could help us to prioritise and plan conservation areas in a spatially explicit manner, by factoring in the underlying processes of fragmentation, while balancing the objective of conserving evolutionary history.

We can even calculate  $\lambda_M$  at the patch level, allowing us to target specific areas within a species distribution for conservation prioritization (Ovaskainen and Hanski 2003). Since the spatial aspects would influence upon the evolutionary history of

animals, we study this by quantifying isolation and size of patches (or islands). Relatedly, metapopulation theory itself was founded on such spatial assumptions of island biogeography (MacArthur and Wilson 1967).

## *Island Biogeography*

Current global databases often lack the spatial and ecological granularity necessary to conduct such a large-scale analysis, without requiring great effort in obtaining and polishing the data. However, one way that we can at least test this proposed conservation prioritisation method is by examining islands, which we do here on mammals.

In this chapter, we use  $\lambda_M$  in combination with the current prioritization scheme of EDGE for two purposes. First, we investigate whether phylogenetic diversity correlates with characteristics of islands. We expect, based on the principles of island theory that predict lower immigration and emigration rates, that with increasing remoteness and decreasing size, species could accumulate evolutionary history. Second, we prioritise important islands containing an over proportional amount of evolutionary distinct species, indicating a potentially increased risk of living on small remote islands, requiring special attention. IUCN spatial data on species geographic ranges are typically somewhat general and broad, owing to the scope of species assessed. By incorporating more accurate, updated distribution data, we are vastly improving our collective understanding as to how threatened a particular species really is. We want to measure biodiversity value with readily available data and tools to identify conservation priority sites in a heavily fragmented landscape.

## **Methods and Materials**

Islands are an ideal system to examine, because they are spatially segregated, but are also of importance, as they are home to many potentially important species under threat (Steadman 1995). We assume islands are associated with a greater ED than mainland areas, since islands are more isolated and therefore should be more likely to accumulate ED than other landforms. We already know that island area correlates with phylogenetic structure (Cardillo et al. 2008), and we too found a correlation between island size and ED.

The next logical question then is how could we quantify the different islands, with respect to species and each island's overall community. We take the ED score of mammal species on islands, and then calculate the  $\lambda_M$  of every patch within a species' distribution to prioritise spatially among the island patches. Metapopulation theory suggests that a population made up of smaller populations with potential gene flow might better persist than otherwise expected when considering each population alone and individually. Thus, distributions made up of closer, larger islands would be better off because of the increased probability of dispersal and rescue effect.



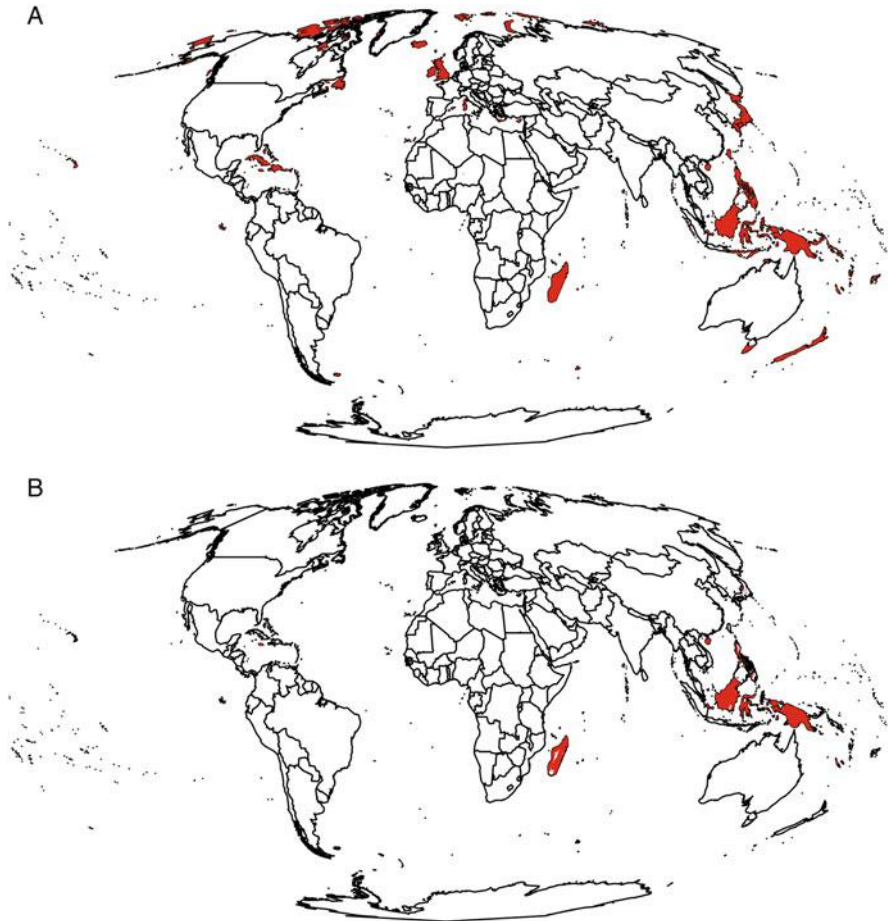
## ***Global Self-Consistent Hierarchical High-Resolution Shoreline Data***

We began with Global Self-consistent Hierarchical High-resolution Shorelines (GSHHS) data to identify island boundaries (Wessel and Smith 1996), before selecting out the qED (the position or quantile of the observed realised cumulative score) values from IUCN geographic ranges (see Safi et al. 2013). We considered islands closer to the mainland than 5 km as belonging to the mainland itself. Likewise, we clumped islands that had distances below 5 km on average to belong together and forming “connected” archipelagos. In order to assess the distances and identify archipelagos, we used the “raster” and “sp” packages in R (2.15.1). We first rasterised the GSHHS coast line with a resolution of 5 by 5 km, where a raster cell was considered landmass, if the grid cell lay on or touched a landmass. We then identified patches of connected raster cells using the queen’s case to decide on the connectedness of raster cells forming “clumps”. Following this procedure, we excluded all patches of connected landmass with an area equal to and larger than Greenland. Finally, we extracted from the original GSHHS vector data all those polygons that contained or touched the remaining grid cells, identifying islands, and archipelagos of the appropriate size and with the approximate required distances to each other and to the main lands. For all islands (and archipelagos), we overlay the IUCN geographic range data previously gridded to a resolution of  $25 \times 25$  km onto the island polygons of the GSHHS vector data to identify the species and the respective ED scores for each island (see Fig. 1a).

## ***Digital Distribution Maps of the IUCN Red List of Threatened Species***

We began with the datasets of terrestrial mammal species as defined by the IUCN Red List database (IUCN 2013). Then we focused on terrestrial mammal species living only on islands, and excluded all species that did not have distributions confined to islands only. We defined islands as landmasses smaller than Greenland ( $2,130,800 \text{ km}^2$ ), with New Guinea ( $785,753 \text{ km}^2$ ) as the largest island. IUCN’s terrestrial mammal spatial data had 1728 unique species identified as residing on an island. When we intersected this with the GSHHS shoreline data, which fulfilled our definition for island, there were 1501 species.

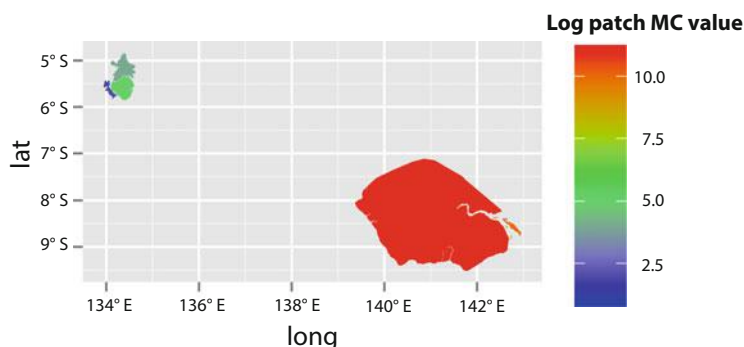
Finally, we restricted this to obligate islanders only, i.e. species not found on any continental mainland, and had 389 species with island-only distributions. We excluded those species with distributions that also encompassed continental mainland because we expected that they would not experience the same level of fragmentation threat as species with an island-confined existence. The mainland can be a potential population source that would not compare evenly in the calculations, particularly as our GSHHS data would not be able to define the species distribution extent on mainland.



**Fig. 1** (a) Map of GSHHS-defined islands, highlighting all those containing mammals for which we have ED scores. (b) The highest  $\lambda_M$  score (1.0) of IUCN-defined island mammals ranges, where endemics confined to one island are automatically assigned an  $\lambda_M$  score of 1.0. This indicates where the most valuable patches are within a species distribution, and consequently what would be most worth saving

### *Data Analysis*

After finding those islands where both GSHHS and IUCN datasets intersected, we calculated the relative  $\lambda_M$  of every patch within a species' distribution and scaled their values from 0 to 1.0, with the highest value indicating the island/patch that contributed most to the overall long-term persistence (see Fig. 2). We also designated any species with only one island/patch in their distribution automatically with a  $\lambda_M$  score of 1 (see Fig. 1b), because of its significant importance for that species. We then took these scores and for each, multiplied by the species' ED score. To



**Fig. 2** Example map showing how the relative log-scaled  $\lambda_M$  scores rank within a species' distribution. Here is the distribution of the Wallace's three-striped dasyure (*Myoictis wallacei*), which occurs in the Aru Islands (Indonesia), and in the southern lowlands on the island of New Guinea (Indonesia and Papua New Guinea) from Merauke in the west to Avera on the Aroa River in the east (Leary et al. 2008)

further give an average  $\lambda_M$ -ED score per island, we took the sum of species' scores and divided this by the number of island mammal species (in our dataset) residing on that island.

## Results

We found 40 Least Concern and Data Deficient species that possess a high combined score of  $\lambda_M$  and ED (see Table 1). In total, 42 of the island mammal species we assessed were listed by the IUCN as Data Deficient, 47 as Least Concern, with the remainder as threatened species. Those species already listed as threatened were potentially suffering from other threats (e.g. non-native species as predators/competitors). Focusing on those species that are Data Deficient or Least Concern and have higher  $\lambda_M$ -ED scores would be most beneficial, as their rarity indicate them to be at risk and a high  $\lambda_M$  value represents an important patch, and one that would pay off greatly to conserve.

The five islands with the highest average  $\lambda_M$ -ED scores, taken by adding all the scores and dividing by our (island-restricted mammals) species richness per island were Jamaica, Guadalcanal, Isle of Pines, Madagascar, and Nggela Sule (see Table 2, Fig. 3 for map). Interestingly, Madagascar held 39 of the highest  $\lambda_M$ -ED species, and ranked fourth in our  $\lambda_M$ -ED islands list.

We found that combining evolutionary distinctness with  $\lambda_M$  revealed species that may be of concern that were not otherwise noticed. Because quantifying fragmentation effects on species takes into account spatial configuration, this can help to improve threat status assessments. The EDGE programme has already sought to visualize regions in the world with the most rare species and moved to prioritize those particular species. This adds a spatial understanding of the species distribution to that prioritization.

**Table 1** Top 10 species in order of decreasing  $\lambda_M$ -ED score

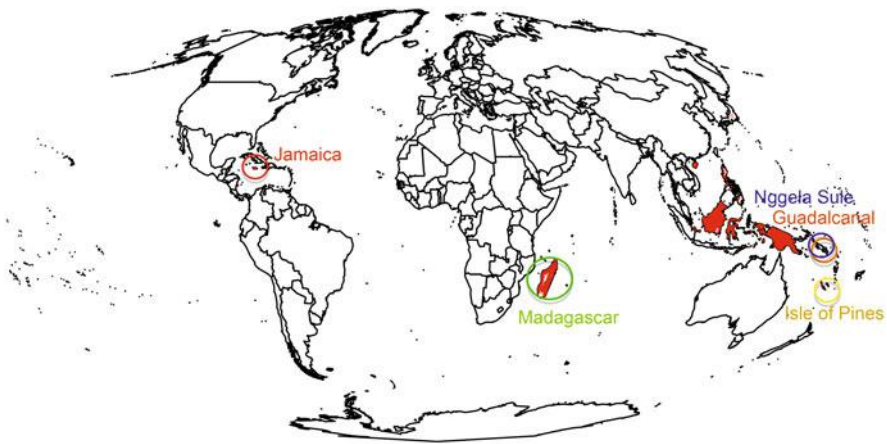
Island	Species	$\lambda_M$ *ED
Jamaica	<i>Ariteus flavescens</i>	0.93350
Madagascar	<i>Emballonura tiavato</i>	0.92111
Madagascar	<i>Avahi unicolor</i>	0.92111
Madagascar	<i>Microgale brevicaudata</i>	0.92111
Madagascar	<i>Eulemur rubriventer</i>	0.92111
Madagascar	<i>Microgale drouhardi</i>	0.92111
Madagascar	<i>Brachytarsomys villosa</i>	0.92111
Madagascar	<i>Gymnuromys roberti</i>	0.92111
Madagascar	<i>Pteropus rufus</i>	0.92111
Madagascar	<i>Avahi occidentalis</i>	0.92111

We consider those species to be of high concern to be a  $\lambda_M$  \*ED value above 0.8

**Table 2** Top 10 islands, in order of decreasing  $\lambda_M$ -ED score

Island	$\lambda_M$ *ED
Jamaica	0.93350
Guadalcanal	0.88409
Isle of Pines	0.85935
Madagascar	0.68196
Nggela Sule	0.57726
Bangka	0.52276
Biak and Supiori	0.50898
Dinagat	0.40916
Fergusson Island	0.38875
New Ireland	0.38470

We calculated this by dividing all species'  $\lambda_M$ -ED score by the number of resident island mammal species for which we had range data per island



**Fig. 3** Map highlighting the top five islands, coloured from warm to cool (i.e. red to blue), in decreasing  $\lambda_M$ -ED score (see also Table 2)

## Discussion

### *Summary*

We found Least Concern and Data Deficient island-restricted mammals that possess a high combined score of  $\lambda_M$  and ED. This method can be the start to finding species with a combination of phylogenetic rarity and long-term extinction risk due to island isolation. Further analyses are needed, as global prioritizations risk overgeneralizing among distinct animals, and yet suitable datasets, spatial and otherwise, are difficult to come by.

### *Island Studies*

Islands represent less than 5 % of the earth's land area, harbour 80 % of known species extinctions since 1500 (Ricketts et al. 2005), and make up 39 % of today's IUCN Critically Endangered species (TIB 2012). Endangered island species, such as those targeted and listed in the Threatened Island Biodiversity (TIB) database, are currently of major concern due to invasive species. However, we can still examine the effects of isolation and area from an island point of view. On a global scale, this method aims to show which islands or species are most important for conservation, based on the spatial properties of the islands and the phylogenetic rarity of the species themselves.

Islands are a natural laboratory for evolutionary specialization and adaptation, because such an environment greatly shapes the select set of species living there in such isolation (Losos and Ricklefs 2009). From a conservation perspective, islands are unique because with less spatial area to begin with, they can only support smaller populations to evolve on them (Diamond 1975; Frankham 1998). Furthermore, recolonisation, the process responsible for maintaining population size from a larger source population, decreases because of spatial isolation and size (MacArthur and Wilson 1963, 1967; Simberloff and Wilson 1970), and dispersal amongst islands can be far more limited than on terrestrial "islands". We expect that islands suffer more from stochastic extinction processes, in addition to anthropogenic effects such as introduced species, so they are on the whole in much greater need of immediate conservation action. In fact, islands have previously been the focus of research on prioritisation schemes for conservation planning (TIB 2012).

However, much complexity remains in studying islands. Most threatened species have small geographic distributions, and the distributions of island species are inevitably smaller than the distributions of continental species (Manne et al. 1999). Yet, some island populations can "show greater persistence than mainland populations of the same species, notwithstanding their smaller range sizes" (Channell and Lomolino 2000), perhaps reflecting the advantages of living in sheltered isolation. Another study found that island endemics are not relatively more threatened than

continental ones, considering their distribution size, “suggesting that evolutionary isolation is not the reason for their vulnerability” (Purvis et al. 2000). Perhaps unravelling isolation and evolutionary factors can lead to a greater understanding of the unique state that island animals seem to occupy.

Small distribution area and island endemism were the most important predictors of mammal extinction risk found through literature survey (Purvis et al. 2000). Because of such isolation, we would expect evolutionary history to reflect the spatial fragmentation. Moreover, there is a certain importance to the isolation of islands, given the limits of animal dispersal (Diamond 1974). For instance, the number of threatened endemic bird species has been found to correlate with deforestation on islands, and single-island endemics are considerably more at risk than more widespread species (Brooks et al. 1997), hence examining spatial aspects of islands is a sensible route.

Islands, particularly larger ones, are likely to contain multiple landscape types, and our islands borders, although defined at high resolution by GSHHS, can likely overestimate the amount of suitable habitat for a species. For instance, we found Madagascar ranked fourth in our list, but including additional information would scale down the habitat size from islands to the actual size of primary habitat. Then Madagascar might very well outrank all the other islands, due to unique species that possess ranges limited to parts of the island. With species records from GBIF and publicly available environmental layers, we could perhaps improve on this by creating approximate species distribution “maps” that we might be able to prune down the current IUCN extent of occurrence maps to a more realistically “fragmented” habitat extent. Calculating the  $\lambda_M$  of such maps would be an improved and more realistic estimate as to long-term species persistence.

It might be that island species have some adaptation for having historically small isolated populations, such that the little area available has shaped the species’ phylogeny (Cardillo et al. 2008). On the other hand, age of the islands (equivalently, patches) might have a significant influence on metapopulation persistence (Hastings 2010). It could also be that the most sensitive species were previously driven to extinction and modern day survivors have already been selected for (Manne et al. 1999). Human impact cannot be overestimated, because despite exceptional habitat loss on all terrestrial land types, “the human impact index” was considerably greater on islands (Kier et al. 2009). It is still a puzzle to be teased apart, how the interaction of intrinsic factors, e.g. innate biological susceptibility, and extrinsic factors, i.e. those mostly due to human impact, affect the outcome that ultimately leads to extinction (Bennett and Owens 1997).

Already there are numerous efforts underway to stave off the extinction of island species, such as the previously mentioned Threatened Island Biodiversity (TIB) database (<http://tib.islandconservation.org/>), whose primary focus is on eradicating threatening non-natives. The high levels of endemic richness already warrant special conservation protection (Kier et al. 2009). Species on continents can experience island effects, e.g. mountains or islands within lakes, which would still make island conservation studies, such as this, applicable to them.

## *Next Steps*

Several aspects of this analysis can be modified depending on the user's goals. For example, we took 5 km to be the minimum distance from continental mainland for an archipelago isolated enough to not experience a strong mainland source population. At one extreme, Davies et al. (2007) previously defined oceanic islands as those more than 200 km away from a continental shelf edge. Distance to mainland would understandably have different consequences on the species if (1) they have some portion of their metapopulation residing on the mainland, or (2) they are able to cross this water gap, albeit rarely. If this assessment was of larger sized islands or patches, one could implement a  $\lambda_M$  score per area (e.g. square kilometre).

It is worth mentioning that species richness does not play any role in this ranking. Species richness is an anthropogenic valuation scheme, and this method is unique in considering from the phylogenetic and spatial considerations of the animals themselves. However, something that could be accounted for is complementarity, as in the case where two islands contain the same sets of species. Many sophisticated spatial planning tools try to take this into account, one such being Zonation (Moilanen et al. 2005; Moilanen 2007).

It seems logical that species endemic to only one island require the most accurate distribution data, and most rigorous of assessments, because these cases have all their “eggs in one basket”. Incorporating movement functions would greatly improve the model's connectivity aspect, determining how fragmented such oceanic islands are. The availability of such data is increasing, fortunately, and ideally they will improve habitat utilization and connectivity estimates in the future. This method can go beyond islands, however.

We had excluded those species with distributions including continents because of how it would influence the biogeography dynamics. Facultative islanders (of which we found 1611 species), those species with distribution on both island and continent, made up a longer list that could be worthwhile for further study. This would be an interesting question to tackle, because it would be a step closer to quantifying mainland “value” for islands, how to go about quantifying its contribution. Nevertheless, looking at only islands made for a simpler study, and a further interesting one is then to shift our focus towards continents. It would be more broadly useful, and also computationally challenging, to do the same analysis for higher precision information of animal distributions on the continents. The  $\lambda_M$  has the potential to identify important areas for connectivity, so that we might better respond to extinction threats, and therefore might be a better way of prioritising specific areas for conservation. This index weighs those island “patches” which are most valuable to species with limited ranges and for species with unique phylogenies. Future schemes could consider different weightings and combinations of these two indices. More importantly, for islands a score is calculated by taking an average score over all species.

As for island species, we would like to compare our lists with the outcome of the EDGE zones papers. It would be interesting to see whether the islands important for



$\lambda_M$ -ED island species are similar to those we identified in the global EDGE analysis. We also need to discuss GE and how best to handle this additional information. We already know we can be so much more effective in conservation when a targeted approach is taken, particularly for critically endangered species (Brooke et al. 2008).

**Acknowledgements** We would like to thank Bart Kranstauber for discussions, and Laura J. May-Collado, Roseli Pellens, and an anonymous reviewer for helpful feedback.

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Barnosky AD, Matzke N, Tomiya S, Wogan GO, Swartz B, Quental TB et al (2011) Has the Earth's sixth mass extinction already arrived? *Nature* 471(7336):51–57
- Bennett PM, Owens IP (1997) Variation in extinction risk among birds: chance or evolutionary predisposition? *Proc R Soc Lond Ser B Biol Sci* 264(1380):401–408
- Brook BW, O'Grady JJ, Chapman AP, Burgman MA, Akçakaya HR, Frankham R (2000) Predictive accuracy of population viability analysis in conservation biology. *Nature* 404(6776):385–387
- Brooke MDL, Butchart SHM, Garnett ST, Crowley GM, Mantilla-Beniers NB, Stattersfield AJ (2008) Rates of movement of threatened bird species between IUCN Red List categories and toward extinction. *Conserv Biol* 22(2):417–427
- Brooks TM, Pimm SL, Collar NJ (1997) Deforestation predicts the number of threatened birds in insular Southeast Asia. *Conserv Biol* 11(2):382–394
- Cardillo M, Gittleman JL, Purvis A (2008) Global patterns in the phylogenetic structure of island mammal assemblages. *Proc R Soc B* 275(1642):1549–1556. doi:[10.1098/rspb.2008.0262](https://doi.org/10.1098/rspb.2008.0262)
- Channell R, Lomolino MV (2000) Dynamic biogeography and conservation of endangered species. *Nature* 403:84–86
- Collen B, Turvey ST, Waterman C, Meredith HM, Kuhn TS, Baillie JE, Isaac NJ (2011) Investing in evolutionary history: implementing a phylogenetic approach for mammal conservation. *Philos Trans R Soc B* 366(1578):2611–2622
- Davies RG, Orme CDL, Storch D, Olson VA, Thomas GH, Ross SG, ..., Gaston KJ (2007) Topography, energy and the global distribution of bird species richness. *Proc R Soc B Biol Sci* 274(1614):1189–1197
- Diamond JM (1974) Colonization of exploded volcanic islands by birds: the supertramp strategy. *Science* 184(4138):803–806
- Diamond JM (1975) The island dilemma: lessons of modern biogeographic studies for the design of natural reserves. *Biol Conserv* 7(2):129–146
- Frankham R (1998) Inbreeding and extinction: island populations. *Conserv Biol* 12(3):665–675
- Gillespie TW, Foody GM, Rocchini D, Giorgi AP, Saatchi S (2008) Measuring and modelling biodiversity from space. *Prog Phys Geogr* 32(2):203–221
- Hanski I (1998) Metapopulation dynamics. *Nature* 396(6706):41–49
- Hanski I, Ovaskainen O (2000) The metapopulation capacity of a fragmented landscape. *Nature* 404(6779):755–758
- Hanski I, Ovaskainen O (2002) Extinction debt at extinction threshold. *Conserv Biol* 16(3):666–673



- Hanski I, Simberloff D (1997) The metapopulation approach, its history, conceptual domain, and application to conservation. In: Hanski I, Simberloff D (eds) *Metapopulation biology: ecology, genetics, and evolution*. Academic, San Diego
- Hastings A (2010) Timescales, dynamics, and ecological understanding 1. *Ecology* 91(12):3471–3480
- Isaac NJB, Redding DW, Meredith HM, Safi K (2012) Phylogenetically-Informed Priorities for Amphibian Conservation. *PLoS ONE* 7(8), e43912. doi: [10.1371/journal.pone.0043912](https://doi.org/10.1371/journal.pone.0043912)
- Isaac NJ, Turvey ST, Collen B, Waterman C, Baillie JE (2007) Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PLoS One* 2(3):e296
- IUCN (2013) The IUCN Red List of threatened species. Version 2013.2. <http://www.iucnredlist.org>. Downloaded on 25 Nov 2013
- Kearney M, Porter W (2009) Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecol Lett* 12(4):334–350
- Kerr JT, Ostrovsky M (2003) From space to species: ecological applications for remote sensing. *Trends Ecol Evol* 18(6):299–305
- Kier G, Kreft H, Lee TM, Jetz W, Ibisch PL, Nowicki C, ..., Barthlott W (2009) A global assessment of endemism and species richness across island and mainland regions. *Proc Natl Acad Sci* 106(23):9322–9327
- Leary T, Seri L, Wright D, Hamilton S, Helgen K, Singadan R, Menzies J, Allison A, James R, Dickman C, Lunde D, Aplin K, Flannery T, Woolley P (2008) *Myioictis wallacei*. In: IUCN 2013. IUCN Red List of threatened species. Version 2013.2. [www.iucnredlist.org](http://www.iucnredlist.org). Downloaded on 30 May 2014
- Loehle C, Eschenbach W (2012) Historical bird and terrestrial mammal extinction rates and causes. *Divers Distrib* 18(1):84–91
- Losos JB, Ricklefs RE (2009) Adaptation and diversification on islands. *Nature* 457(7231):830–836
- MacArthur RH, Wilson EO (1963) An equilibrium theory of insular zoogeography. *Evolution* 17(4):373–387
- MacArthur RH, Wilson EO (1967) *The equilibrium theory of island biogeography*. Princeton University Press, Princeton
- Manne LL, Brooks TM, Pimm SL (1999) Relative risk of extinction of passerine birds on continents and islands. *Nature* 399:258–261
- Martin LJ, Blossey B, Ellis E (2012) Mapping where ecologists work: biases in the global distribution of terrestrial ecological observations. *Front Ecol Environ* 10(4):195–201
- Moilanen A (2007) Landscape zonation, benefit functions and target-based planning: unifying reserve selection strategies. *Biol Conserv* 134(4):571–579
- Moilanen A, Franco AM, Early RI, Fox R, Wintle B, Thomas CD (2005) Prioritizing multiple-use landscapes for conservation: methods for large multi-species planning problems. *Proc R Soc B Biol Sci* 272(1575):1885–1891
- Moilanen A, Wilson KA, Possingham HP (eds) (2009) *Spatial conservation prioritization: quantitative methods and computational tools*. Oxford University Press, Oxford
- Ovaskainen O, Hanski I (2003) How much does an individual habitat fragment contribute to metapopulation dynamics and persistence? *Theor Popul Biol* 64(4):481–495
- Pereira HM, Navarro LM, Martins IS (2012) Global biodiversity change: the bad, the good, and the unknown. *Ann Rev Environ Resour* 37:25–50
- Pimm SL, Russell GJ, Gittleman JL, Brooks TM (1995) The future of biodiversity. *Sci AAAS Wkly Pap Ed* 269(5222):347–349
- Purvis A, Gittleman JL, Cowlishaw G, Mace GM (2000) Predicting extinction risk in declining species. *Proc R Soc Lond B* 267:1947–1952
- Ricketts TH, Dinerstein E, Boucher T, Brooks TM, Butchart SH, Hoffmann M, ..., Wikramanayake E (2005) Pinpointing and preventing imminent extinctions. *Proc Natl Acad Sci U S A* 102(51):18497–18501

- Safi K, Armour-Marshall K, Baillie JE, Isaac NJ (2013) Global patterns of evolutionary distinct and globally endangered amphibians and mammals. *PLoS One* 8(5):e63582
- Schnell JK, Harris GM, Pimm SL, Russell GJ (2013a) Estimating extinction risk with metapopulation models of large-scale fragmentation. *Conserv Biol* 27(3):520–530
- Schnell JK, Harris GM, Pimm SL, Russell GJ (2013b) Quantitative analysis of forest fragmentation in the Atlantic forest reveals more threatened bird species than the current Red List. *PLoS One* 8(5):e65357
- Simberloff DS, Wilson EO (1970) Experimental zoogeography of islands. A two-year record of colonization. *Ecology* 51(5):934–937
- Steadman DW (1995) Prehistoric extinctions of Pacific island birds: biodiversity meets zooarchaeology. *Science* 267(5201):1123–1131
- TIB Partners (2012) Threatened island biodiversity database. Version 2012.1. <http://www.tib.islandconservation.org>. Downloaded on Oct 2013
- Wessel P, Smith WHF (1996) A global, self-consistent, hierarchical, high-resolution shoreline database. *J Geophys Res* 101(B4):8741–8743

# Patterns of Species, Phylogenetic and Mimicry Diversity of Clearwing Butterflies in the Neotropics

Nicolas Chazot, Keith R. Willmott, André V.L. Freitas, Donna Lisa de Silva, Roseli Pellens, and Marianne Elias

**Abstract** The Neotropical region comprises six of the major biodiversity hotspots of the planet, including the Andean foothills, which harbour the most diverse terrestrial ecosystems. It is also one of those most threatened by habitat destruction and climatic changes, which cause species extirpation and sometimes extinction, resulting in community disassembly and loss of interspecific interactions. The effects of community disassembly can be particularly strong in highly coevolved mutualistic species assemblages, such as Müllerian mimetic species. Conservation strategies should therefore aim at preserving not only evolutionary diversity, but also species interactions. Here we use mimetic ithomiine butterflies (Nymphalidae: Danainae, Ithomiini) as a model to identify areas of both evolutionary and ecological importance, and hence conservation significance. Ithomiine butterflies form a tribe of ca. 380 species that inhabit lowland and montane Neotropical forests. All species engage in Müllerian mimicry, and drive mimicry in other, distantly related, Lepidoptera. We analyse phylogenetic, distribution and mimicry data for three diverse ithomiine genera, *Napeogenes*, *Ithomia* and *Oleria*. We use different metrics to study geographical patterns of diversity. Patterns of species richness, phylogenetic diversity and mimicry diversity are highly congruent within genera but slightly different among genera. Mountainous regions contain the greatest taxonomic and mimetic diversity in ithomiines, with the Andean foothill region being

---

N. Chazot (✉) • D.L. de Silva • R. Pellens • M. Elias

Institut de Systématique, Evolution, Biodiversité, ISYEB – UMR 7205 CNRS MNHN

UPMC EPHE, Muséum National d'Histoire Naturelle, Sorbonne Universités,

45 rue Buffon, CP 50, 75005 Paris, France

e-mail: [chazotn@gmail.com](mailto:chazotn@gmail.com); [lisadesilva@yahoo.co.uk](mailto:lisadesilva@yahoo.co.uk); [pellens@mnhn.fr](mailto:pellens@mnhn.fr); [elias@mnhn.fr](mailto:elias@mnhn.fr)

K.R. Willmott

McGuire Center for Lepidoptera and Biodiversity, Florida Museum of Natural History,

University of Florida, Gainesville, FL 32611, USA

e-mail: [kwillmott@flmnh.ufl.edu](mailto:kwillmott@flmnh.ufl.edu)

A.V.L. Freitas

Departamento de Zoologia and Museu de História Natural, Instituto de Biologia,

Universidade Estadual de Campinas, Campinas, São Paulo, Brazil

e-mail: [baku@unicamp.br](mailto:baku@unicamp.br)

the area of highest diversity, but other regions, such as Central America and the upper Amazon, are also important. Finally, a measure of vulnerability related to mimicry indicates that mutualistic interactions are not distributed evenly across space and genera. We argue that mutualistic interactions should be taken into account in conservation strategies.

**Keywords** Ithomiini • Müllerian mimicry • Phylogenetic diversity • Amazonia • Andes

## Introduction

The Neotropical region extends from Mexico to northern Argentina, including the Amazon basin, the Andean cordillera and the Atlantic Forest. It is the most biologically diverse of the world's major biogeographic regions (Gaston and Hudson 1994; Myers et al. 2000; Hawkins et al. 2007). At least one million species of insects, 40,000 of plants, 3000 of fishes, 1294 of birds, 427 of mammals, 427 of amphibians, and 378 of reptiles inhabit the Amazonian basin (Da Silva et al. 2005a). It is also a region of high endemism, including 6 of the world's 25 biodiversity 'hotspots' (Myers et al. 2000).

Many areas of the Neotropics are under continual threat from deforestation. In 2004, ca. 2.7 million ha of forest were cleared in Amazonia alone (INPE). In Central America, only 1.7 % of the original dry forest remains, and most of this comprises small, isolated patches (Griscom and Ashton 2011). Similarly, the Atlantic forest has been reduced to 12 % of its original area, with astonishing rates of deforestation every year (SOS Mata Atlântica, INPE, ISA 1998; Ribeiro et al. 2009). In Amazonia, wood extraction, industrial logging, cattle pastures, banana plantations and more recently oil palm culture are the main causes of the ongoing deforestation. The Neotropics are also threatened by climatic changes, which are likely to be particularly serious in mountain habitats (e.g., Engler et al. 2009; Chen et al. 2011; Feeley et al. 2011). Habitat destruction and climatic changes may cause species extirpation, displacements and extinction (e.g., Loiselle et al. 2010), which may in turn result in community disassembly, with loss of interspecific interactions (Sheldon et al. 2011). The consequences of community disassembly can be particularly strong in highly coevolved mutualistic species assemblages, such as insect-pollinator networks, plant species engaged in facilitation, or Müllerian mimetic species (Chazot et al. 2014).

To preserve Neotropical biodiversity, given constraints of time and money, it is essential to identify priority areas for conservation (Williams et al. 1996). However, there are several problems in identifying such areas. Firstly, distribution data are not available for all taxa, so attention has focused on indicator taxa, which are expected to reliably indicate patterns of diversity in other, more poorly known groups (e.g., Howard et al. 1998; Lamoreux et al. 2006). Insects constitute at least 70 % of all terrestrial organisms (Samways 1994), and their outstanding evolutionary success

in virtually all terrestrial habitats potentially makes them one of the most valuable study groups for understanding the distribution and origins of biodiversity, and for developing efficient means to conserve that biodiversity (Brown 1997). However, because of the diversity of insects and taxonomic difficulties in many groups, some authors have suggested that conservation research be focused on a few suitable taxa, such as butterflies (New 1993; Brown and Freitas 2000; Bonebrake et al. 2010; Basset et al. 2013). Butterflies can be used to monitor ecosystem health (e.g. Warren et al. 2001), reveal broadly applicable patterns of diversity and endemism, and effectively communicate complex scientific ideas to the public and generate popular support for conservation (Sparrow et al. 1994; Boggs et al. 2003).

One of the best studied diverse groups of Neotropical butterflies is the tribe Ithomiini (Nymphalidae: Danainae), an exclusively Neotropical group which currently includes about 380 species placed in 47 genera (Lamas 2004; Willmott and Lamas 2007). Ithomiines occur from Mexico to Argentina and are largely restricted to moist forest habitats from sea level up to 3000 m (Beccaloni 1997a). Among the attributes that make the group a potentially useful indicator of conservation priorities for other taxa are its diversity, its broad range of occupied elevations, its abundance in the field and collections, the broad variation in range size among species, and its good level of taxonomic knowledge.

Having selected a potentially suitable indicator taxon, the next issue is to decide what surrogate measure of biodiversity will be used (Williams et al. 1996). Species richness is the most commonly used measure, but it may not represent important aspects of the structure and composition of natural communities. A species richness measure considers all species as equal, ignoring their functional or phylogenetic relationships (e.g., Safi et al. 2011). As an alternative, measures of phylogenetic diversity evaluate species in terms of the amount of unique evolutionary history they represent. The loss of species with no close relatives represents the extinction of an entire lineage, resulting in a greater loss to biodiversity than the loss of a species that shares most of its evolutionary history with another (Mace et al. 2003; Mooers et al. 2005; Maclaurin and Sterelny 2008). During the last two decades several metrics have been developed to assess the phylogenetic diversity of clades and to evaluate and compare communities for conservation based on the phylogenetic diversity of the species they harbour (e.g., Vane-Wright et al. 1991; Faith 1992). Despite the difficulty of defining the most adequate metric (see Redding et al. 2008; Schweiger et al. 2008), and the circumstances where phylogenetics can be useful for conservation (e.g., Rodrigues et al. 2005; Hartmann and Andre 2013), two points emerged from these studies. The first is that conservation strategies based on phylogenetic measures capture more evolutionary diversity than strategies ignoring phylogeny (e.g., Hartmann and Steel 2007; Redding et al. 2008), and the second is that extinctions are not random in the tree of life, but rather are phylogenetically and functionally clumped (Purvis et al. 2000; Yessoufou et al. 2012). In the last few years several phylogenies have become available for the Ithomiini as a whole (Brower et al. 2014; Willmott and Freitas 2006), and also for some speciose clades inside this tribe (Mallarino et al. 2005; Elias et al. 2009; de-Silva et al. 2015). This phylogenetic information thus allows us to consider phylogenetic diversity in an assessment of conservation priorities for Neotropical insects.

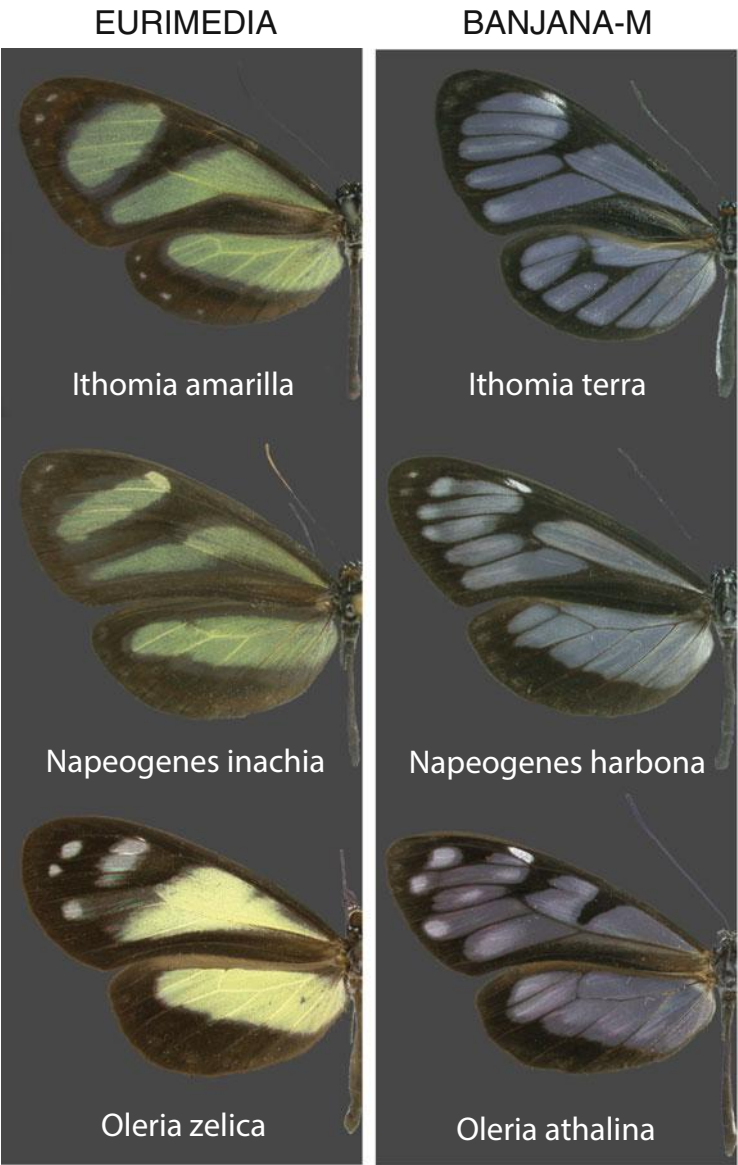
In addition to evolutionary diversity, ecological interactions represent another important component of biodiversity that is rarely addressed in conservation prioritization. While difficult to characterize for many insects, ecological interactions among ithomiines have received an unusual amount of attention since these butterflies illustrate some of the most outstanding examples of mutualistic, Müllerian mimicry (Müller 1879). Ithomiines are considered, together with the Heliconiini, the central models in many Neotropical Lepidoptera mimicry rings (Brown and Benson 1974; Beccaloni 1997a). Chemical compounds acquired by adult feeding (Brown 1984, 1985) make ithomiines unpalatable to their predators, which learn to avoid the aposematic wing patterns exhibited by ithomiines. The wing colour patterns of co-existing species are under strong selection for convergence, thereby reducing the individual cost of educating predators (Müller 1879; Mallet 1999), and the resulting ‘mimicry rings’ contain multiple co-mimetic species which interact mutualistically (Fig. 1). Co-mimetic ithomiines tend to share habitats (Chazot et al. 2014). They also tend to share hostplants (Willmott and Mallet 2004), fly at similar heights above the ground (Beccaloni 1997b; Elias et al. 2008) and fly in similar forest microhabitats (DeVries et al. 1999; Elias et al. 2008; Hill 2010). These tightly-knit webs of interactions may thus be particularly sensitive to community disassembly caused by habitat or climate change (Sheldon et al. 2011), with the potential for cascading co-extinctions due to the loss of a few species whose presence facilitates the existence of other species. Because mutualistic interactions are particularly easy to characterize in ithomiines (species that share the same wing pattern interact mutualistically, species with different wing patterns do not), this group provides a unique opportunity to assess the importance of mutualistic interactions from a conservation perspective.

Studies combining phylogenetic and ecological or functional data to characterize biodiversity patterns and their association with environmental gradients (Devictor et al. 2010; Flynn et al. 2011; Duarte et al. 2012), as well as to test conservation-focused hypotheses (Faith 2008; Safi et al. 2011), are likely to better represent biodiversity and hopefully to guide conservation in a more precise way. Here, we use distribution, phylogenetic and mimicry data for three diverse ithomiine genera, *Ithomia*, *Napeogenes* and *Oleria*, to identify areas of maximal species, phylogenetic and mimicry diversity for each of these genera. We also identify areas of maximum and minimum vulnerability in terms of proportion of potential loss of species due to disruption of mimicry rings. With these three independent replicates we then explore whether different metrics show peaks in the same areas, and whether the different taxa show similar spatial patterns of diversity.

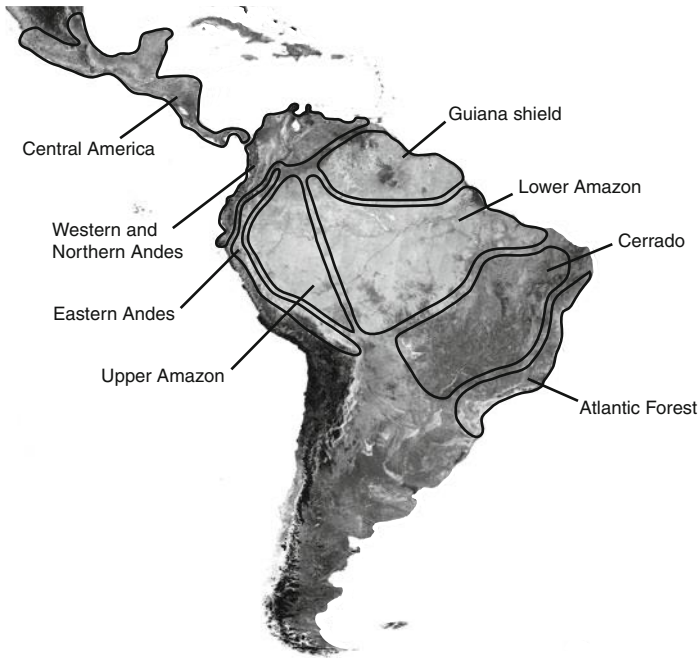
## Material and Methods

### *The Neotropics*

In this study we refer to the following specific areas within the Neotropical region (Fig. 2): (1) Central America, which extends from Panama to Mexico; (2) the western/northern Andes and (3) the eastern Andes, which usually exhibit distinct faunas;



**Fig. 1** Two examples of mimicry rings named eurimedia and banjana-m, each one shared by a species of the three genera under interest



**Fig. 2** Neotropical regions used in this study

(4) the upper Amazon along the eastern foothills of the Andes; (5) the lower Amazon; (6) the Guiana shield; (7) the Cerrado, which separates the Amazon basin and the Atlantic Forest and (8) the Atlantic Forest, which extends along the south-east and southern Brazilian coast and adjacent inland regions.

### ***Study Groups and Phylogenies***

In this chapter we focus on three ithomiine genera (Table 1) for which nearly comprehensive calibrated molecular phylogenies are available in the literature: *Ithomia* (24 out of 25 extant species, Mallarino et al. 2005; Jiggins et al. 2006; Elias et al. 2009); *Napeogenes* (24 out of 24 extant species, Elias et al. 2009) and *Oleria* (42 out of 49 extant species, de-Silva et al. 2015; de-Silva et al. 2010). We analyse the three genera independently to assess the congruence of geographical diversity patterns among genera. All the trees used are ultrametric, with branch lengths proportional to time.



## Mimicry Classification

Species were classified among 23 mimicry groups, following previous classification (e.g., Willmott and Mallet 2004; Jiggins et al. 2006; Elias et al. 2008, Table 1 and Fig. 1).

**Table 1** List of the species of *Ithomia*, *Napeogenes* and *Oleria*, with their mimicry patterns. (Species may harbour multiple geographical races with different mimicry patterns)

Genus	Species	Mimicry pattern
<i>Ithomia</i>	<i>adelinda</i>	confusa, mestra, praxilla
<i>Ithomia</i>	<i>agnosia</i>	agnosia, lerida
<i>Ithomia</i>	<i>amarilla</i>	eurimedia
<i>Ithomia</i>	<i>arduinna</i>	agnosia, eurimedia
<i>Ithomia</i>	<i>avella</i>	ticida-m, banjana-m, panthyle
<i>Ithomia</i>	<i>celemla</i>	hermlas, mamercus, parallelis
<i>Ithomia</i>	<i>cleora</i>	mantineus
<i>Ithomia</i>	<i>diasia</i>	amalda, lerida
<i>Ithomia</i>	<i>drymo</i>	lerida
<i>Ithomia</i>	<i>eleonora</i>	banjana-m
<i>Ithomia</i>	<i>ellara</i>	banjana-m
<i>Ithomia</i>	<i>heraldica</i>	mamercus
<i>Ithomia</i>	<i>hyala</i>	lerida
<i>Ithomia</i>	<i>hymettia</i>	agnosia, banjana-m
<i>Ithomia</i>	<i>iphianassa</i>	dilucida, hermlas, idae
<i>Ithomia</i>	<i>jucunda</i>	amalda, lerida
<i>Ithomia</i>	<i>lichyi</i>	agnosia, lerida
<i>Ithomia</i>	<i>patilla</i>	lerida
<i>Ithomia</i>	<i>praeithomia</i>	banjana-m
<i>Ithomia</i>	<i>pseudoagalla</i>	dilucida
<i>Ithomia</i>	<i>salapia</i>	agnosia, derama, eurimedia
<i>Ithomia</i>	<i>terra</i>	agnosia, banjana-m, lerida
<i>Ithomia</i>	<i>xenos</i>	dilucida
<i>Napeogenes</i>	<i>aethra</i>	hermlas
<i>Napeogenes</i>	<i>apulja</i>	mestra, ocna
<i>Napeogenes</i>	<i>benigna</i>	dilucida, panthyle
<i>Napeogenes</i>	<i>cranto</i>	dilucida, eurimedia
<i>Napeogenes</i>	<i>duessa</i>	duessa, eurimedia, mamercus
<i>Napeogenes</i>	<i>flossina</i>	panthyle
<i>Napeogenes</i>	<i>glycera</i>	mestra, ocna, praxilla

(continued)

**Table 1** (continued)

Genus	Species	Mimicry pattern
<i>Napeogenes</i>	<i>gracilis</i>	ozia
<i>Napeogenes</i>	<i>harbona</i>	banjana-m, susiana, derasa, mestra, unknown
<i>Napeogenes</i>	<i>inachia</i>	eurimedia, hemixanthe
<i>Napeogenes</i>	<i>larilla</i>	banjana-m, hewitsoni, panthyale, theudelinda
<i>Napeogenes</i>	<i>lycora</i>	ozia, praxilla
<i>Napeogenes</i>	<i>nsp1</i>	ocna
<i>Napeogenes</i>	<i>nsp2</i>	banjana-m, ocna
<i>Napeogenes</i>	<i>peridia</i>	dilucida, excelsa, hecalesina, hermiias
<i>Napeogenes</i>	<i>pharo</i>	confusa, derasa, eurimedia, ozia
<i>Napeogenes</i>	<i>rhezia</i>	eurimedia, hemixanthe, hermiias, mamercus, mothone
<i>Napeogenes</i>	<i>sodalis</i>	agnosia
<i>Napeogenes</i>	<i>stella</i>	dilucida, hermiias
<i>Napeogenes</i>	<i>sulphureophila</i>	ocna
<i>Napeogenes</i>	<i>sylphis</i>	agnosia, aureliana, egra, illinissa, lerida
<i>Napeogenes</i>	<i>tolosa</i>	dilucida, eurimedia, excelsa, mamercus
<i>Napeogenes</i>	<i>verticilla</i>	agnosia
<i>Napeogenes</i>	<i>zurippa</i>	hermiias, mamercus, orestes
<i>Oleria</i>	<i>aegle</i>	egra, lerida
<i>Oleria</i>	<i>agarista</i>	aureliana, lerida, sinilia
<i>Oleria</i>	<i>alexina</i>	agnosia
<i>Oleria</i>	<i>amalda</i>	amalda, lerida
<i>Oleria</i>	<i>antaxis</i>	egra, lerida, sinilia
<i>Oleria</i>	<i>aquata</i>	lerida
<i>Oleria</i>	<i>assimilis</i>	angosia, aureliana, lerida
<i>Oleria</i>	<i>astrea</i>	lerida
<i>Oleria</i>	<i>athalina</i>	banjana-m, susiana
<i>Oleria</i>	<i>attalia</i>	mestra, susiana
<i>Oleria</i>	<i>baizana</i>	banjana-m, hewitsoni
<i>Oleria</i>	<i>bioculata</i>	agnosia
<i>Oleria</i>	<i>boyeri</i>	agnosia
<i>Oleria</i>	<i>cyrene</i>	banjana-m, susiana
<i>Oleria</i>	<i>deronda</i>	banjana-m, thabena-f
<i>Oleria</i>	<i>derondina</i>	banjana-m, thabena-f, panthyale
<i>Oleria</i>	<i>didymaea</i>	agnosia, lerida

(continued)

**Table 1** (continued)

Genus	Species	Mimicry pattern
<i>Oleria</i>	<i>enania</i>	agnosia, aureliana, lerida
<i>Oleria</i>	<i>estella</i>	agnosia, quintina
<i>Oleria</i>	<i>fasciata</i>	banjana-m, susiana
<i>Oleria</i>	<i>flora</i>	egra, lerida
<i>Oleria</i>	<i>fumata</i>	banjana-m
<i>Oleria</i>	<i>gunilla</i>	agnosia, aureliana, illinissa, lerida, quintina, sinilia
<i>Oleria</i>	<i>ilerdina</i>	aureliana, illinissa, lerida
<i>Oleria</i>	<i>makrena</i>	agnosia, banjana-m
<i>Oleria</i>	<i>onega</i>	agnosia, aureliana, lerida, quintina
<i>Oleria</i>	<i>padilla</i>	agnosia, banjana-m
<i>Oleria</i>	<i>paula</i>	amalda, lerida
<i>Oleria</i>	<i>phenomoe</i>	agnosia
<i>Oleria</i>	<i>quadrata</i>	agnosia
<i>Oleria</i>	<i>quintina</i>	quintina
<i>Oleria</i>	<i>radina</i>	panthyale, banjana-m, unknown
<i>Oleria</i>	<i>rubescens</i>	lerida
<i>Oleria</i>	<i>santinezia</i>	agnosia, banjana-m
<i>Oleria</i>	<i>sexmaculata</i>	aureliana, lerida, sinilia
<i>Oleria</i>	<i>similigena</i>	egra, lerida
<i>Oleria</i>	<i>tigilla</i>	agnosia, lerida
<i>Oleria</i>	<i>tremona</i>	banjana-m
<i>Oleria</i>	<i>vicina</i>	agnosia
<i>Oleria</i>	<i>victorine</i>	agnosia, lerida
<i>Oleria</i>	<i>nsp</i>	agnosia
<i>Oleria</i>	<i>zelica</i>	eurimedia

## Species Distribution

To map species distributions we compiled a database of 5386 species-locality records. This database combines fieldwork data from the authors and records from more than 60 museums and private collections, with the most significant contributions (>200 records each) from the Natural History Museum, London (BMNH), the Museo de Historia Natural, Universidad Mayor de San Marcos, Lima (MUSM), the McGuire Center, Florida Museum of Natural History, Gainesville (FLMNH), the United States National Museum, Washington D.C. (USNM), the American Museum of Natural History, New York (AMNH), and sight records (Willmott & Hall, unpublished data). Despite ithomiines being one of the best collected Neotropical butterfly

groups, representation is still patchy at fine scales because many regions are yet to be sampled. We therefore used species distribution modelling to better represent the distribution of species and subspecies. We used ArcGIS 9 for most geoprocessing, with the World Cylindrical Equal Area projection, and DIVA-GIS version 7.5 (<http://www.diva-gis.org/>) for distribution modelling. Briefly, the procedure was as follows. First, we calculated the maximum nearest neighbour distance between any two points for each species, as an approximate measure of the extent of our knowledge of the distribution of that species. For two species with disjunct ranges (*Oleria aquata* and *Oleria victorine*) we calculated this distance separately for each population. Second, for each species we created a minimum convex polygon around its distribution points buffered at the distance calculated in step 1. Third, we used the BIOCLIM model in DIVA-GIS to estimate climatically suitable areas for each species on a 2.5 min grid, using two climatic variables, Annual Mean Temperature and Annual Precipitation. We converted the resulting distributions into presence-absence rasters with a value of 0 representing absence (predicted distributions with less than 5 % certainty, i.e. values of 0–50 in the DIVA-GIS output grid file), and 1 for presence (values of 50–500 in the DIVA-GIS output grid file). Fourth, we overlaid the DIVA-GIS model with the buffered minimum convex polygon, and calculated the intersection of these layers as the final predicted distribution for the species. In cases where the distribution was predicted to occur in areas without any record which were separated by a significant barrier (e.g., the Andes mountains) from areas with records, we cropped the distribution to remove those areas with no records. The resulting distribution was converted to a point shapefile (at quarter degree grid resolution) for ease of analysis. As a further step to model the distribution of subspecies, we used the Thiessen polygon (TP) tool in ArcGIS to divide the Neotropical region for each species into a series of contiguous polygons. Each polygon contains a single empirical distribution point, and everywhere within that polygon is nearer to that point than to any other point. We assumed that any modelled distribution point falling within a TP was most likely to be represented by the subspecies occurring at the source point for the TP. We thus overlaid the TP layer with our modelled point shapefiles and assigned each modelled point to a subspecies.

The resulting data were finally analysed by quarter degree grid cell. Distribution maps were overlaid to determine the species/subspecies composition and to calculate six measures of diversity listed below for each grid cell. The three measures of phylogenetic diversity were computed with the package Picante in R.

### ***Species, Mimicry and Phylogenetic Diversity***

We used several metrics to measure different aspects of ithomiine diversity in each grid cell, as outlined below:

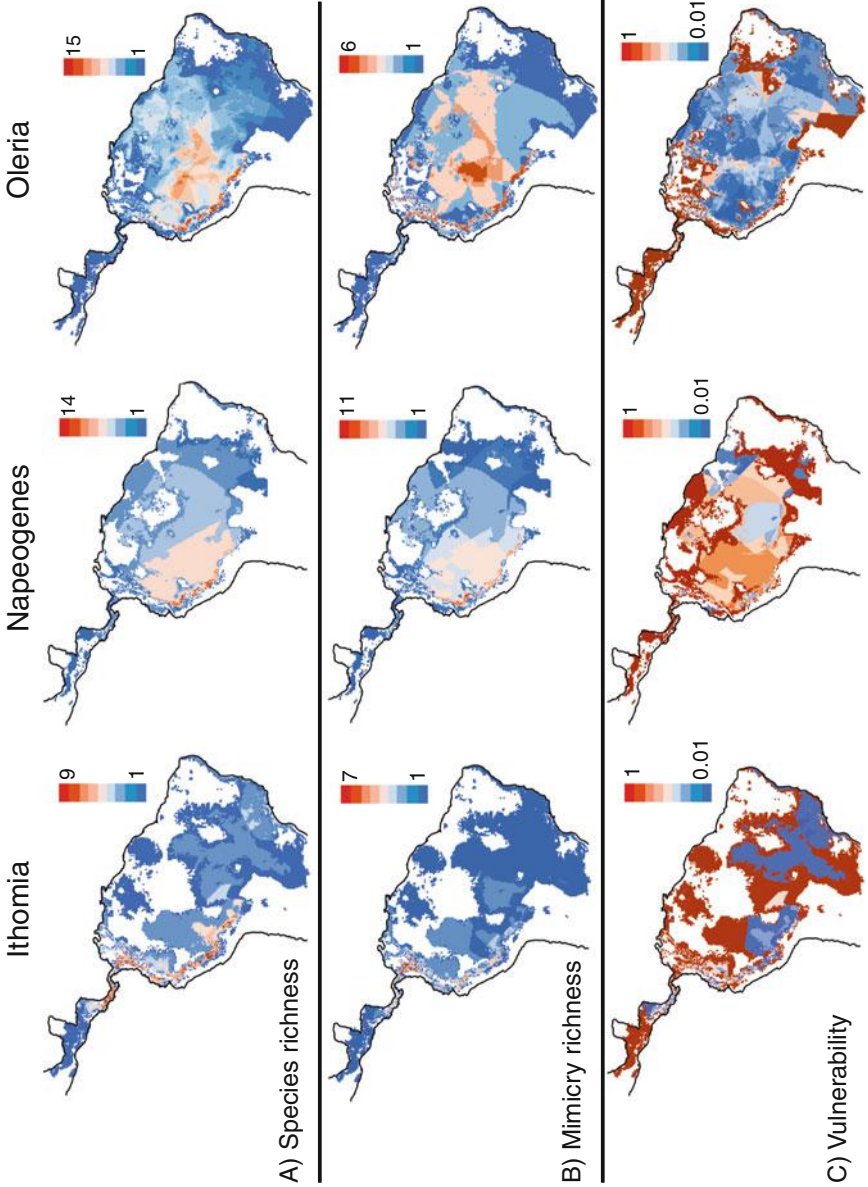
- *Species richness* is the most commonly used measure of diversity, and is computed as the number of species present in each grid cell.
- *Mimicry richness* corresponds to the number of mimicry patterns in each grid cell.

- *Mimicry vulnerability* is a relative vulnerability index based on the hypothesis that the smaller a mimicry ring, the more vulnerable it is (i.e., the more likely it is to lose species). Specifically, we define the vulnerability of mimicry ring  $i$  as  $1/n_i$ , with  $n_i$  the number of species in the mimicry ring. The total vulnerability of a grid cell is the sum of vulnerabilities of each of its constituent mimicry rings  $\sum 1/n_i$ . The relative vulnerability of a grid cell (i.e., scaled by its species richness), is  $(\sum 1/n_i)/\Sigma n_i$ . The smaller this index is, the more robust the community of the grid cell.
- *Faith's Phylogenetic Diversity* (PD, Faith 2002, 2008) is recognised as the most complex measure of phylogenetic diversity. It is a group measure calculated as the total branch length connecting the species present in each grid cell.
- *Equal-Splits* (ES, Redding et al. 2008) is a measure of evolutionary distinctiveness, and it reflects how evolutionarily isolated a species is. Unlike the other measures used, it is a species property measure obtained by dividing the evolutionary time represented by a branch equally among its daughter branches. So, species that diverge early in the tree have higher ES values because much of their evolutionary time is not shared with any other species. The ES of a grid cell is given by the sum of the ES of all species occurring in it.
- *Mean Phylogenetic Diversity* (MPD) is the mean phylogenetic distance between all pairs of species occurring in a grid cell. While PD and ES are expected to be influenced by species richness (see Rodrigues et al. 2005; Nipperess chapter “The Rarefaction of Phylogenetic Diversity: Formulation, Extension and Application”), MPD is totally decoupled from it. High values of MPD indicate the presence of pairs of distantly related species in the grid cell.

For all metrics, we present only results for species present in the phylogeny (i.e., we ignore the single and seven species missing from the phylogenies of *Ithomia* and *Oleria*, respectively). Including all extant species for the metrics that do not depend on the phylogeny does not affect our results and conclusions (results not shown).

## Results

*Ithomia* Species richness (Fig. 3a, left) is low across the lowlands (Guyana shield, lower Amazon, parts of the upper Amazon, Cerrado and Atlantic Forest), and peaks along the eastern and northern Andes and in Central America. The distribution of PD and ES (Fig. 3d, e, left) is very similar to that of species richness, all peaking along the Andes. MPD is also high along the Andes, but the highest values are observed in the upper Amazon and northern Andes, including many grid cells where species richness is very low (Fig. 3f, left). Central America exhibits a low MPD despite high species richness. Mimicry richness (Fig. 3b, left) is highest in the Andes and in the southern part of Central America, and, to a lesser extent, in the upper Amazon, with little diversity in the lower Amazon, Atlantic Forest and Cerrado. Vulnerability (Fig. 3c, left) is generally lowest in areas of high species richness (Andes and Central America), but also in areas where intermediate or low



**Fig. 3** Six measures of diversity of mimetic ithomiine butterflies in the Neotropics for the genera *Ithomia*, *Napeogenes* and *Oleria*: (a) species richness, (b) mimicry richness, (c) mimicry vulnerability, (d) Phylogenetic Diversity (PD), (e) Equal-splits (ES) and (f) Mean Phylogenetic Distance (MPD)

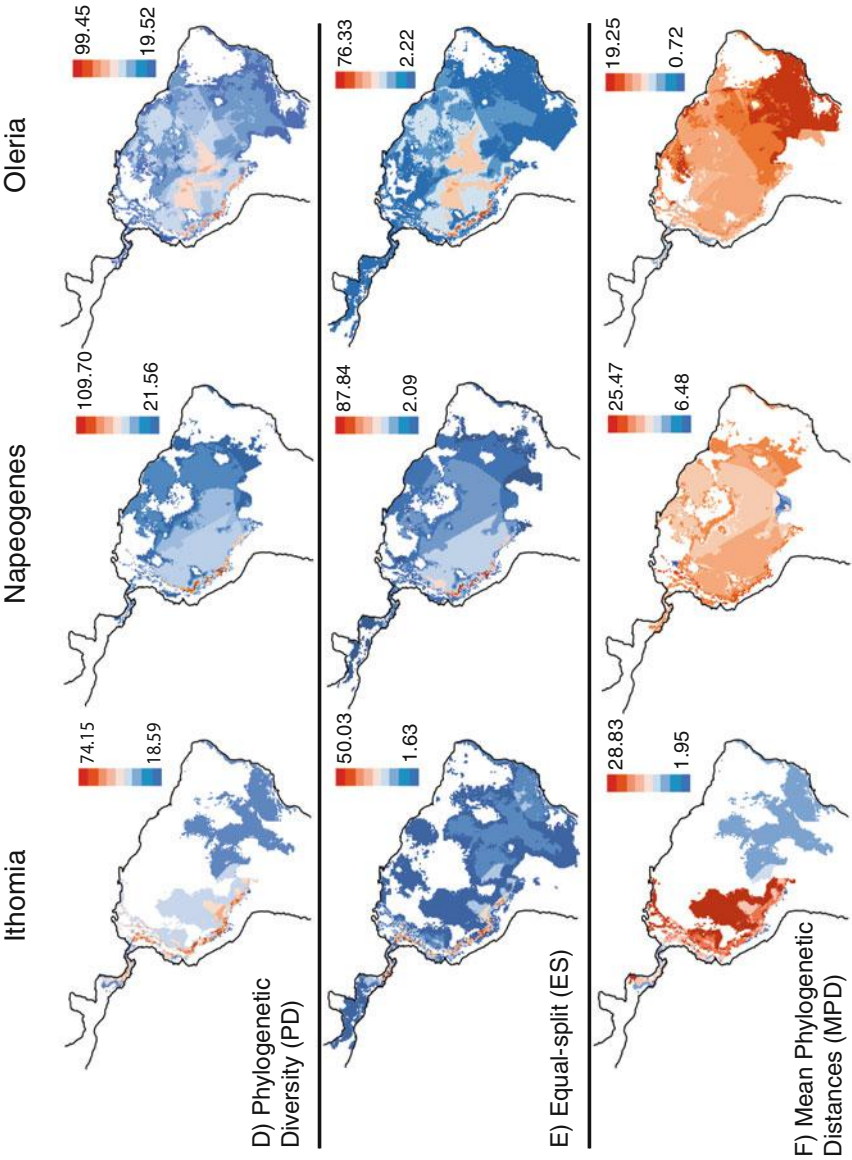


Fig. 3 (continued)

species richness is associated with low mimicry diversity (southern upper Amazon, Cerrado and Atlantic forest).

*Napeogenes* Species richness, mimicry richness, PD, and ES (Fig. 3a, b, d, e, middle) exhibit a nearly identical pattern. They peak along the eastern Andes, remain high in the upper Amazon, and decrease toward the east and south (Guiana shield, lower Amazon and Atlantic Forest). Northern Andes and Central America have low values for these metrics. MPD (Fig. 3f, middle) peaks all along the Andes, from south of Peru, to north Colombia, and exhibits intermediate values in the Cerrado and the edges of the lower Amazon. Vulnerability (Fig. 3c, middle) is generally high, with lowest values in the lower Amazon and in Peruvian eastern Andes. Vulnerability appears less related to species richness than for *Ithomia*.

*Oleria* Species richness, PD and ES (Fig. 3a, d, e, right) peak in the eastern Andes, followed by the upper Amazon and the western part of the lower Amazon. Central America and Atlantic Forest are low-diversity areas. Mimicry richness (Fig. 3b, right) peaks in the Andes and exhibits a second important peak in central Amazon while the Amazonian basin is mimetically rich. MPD increases from north-west (Central-America) toward south-east (Atlantic Forest) (Fig. 3c, right). Mimicry vulnerability (Fig. 3c, right) is lowest in the entire Amazonian basin, Cerrado and Atlantic Forest and increases at the edges of the generic distribution and in Central America.

## Discussion

Understanding global patterns of biodiversity distribution is at the core of macroecology (e.g., Gaston 2000; Gaston and Blackburn 2000; Crisp et al. 2009), and represents a basis for identifying regions that should be the focus for conservation (e.g., Myers et al. 2000). In our study we used six different measures to assess large-scale patterns of diversity for three butterfly genera in the Neotropics. These measures capture different attributes of biodiversity and their simultaneous use contributes to a better picture of how they are related. We found that the patterns of distribution of species richness, phylogenetic diversity and mimicry diversity remain relatively consistent across different ithomiine genera. However, sensitivity to extinction related to mutualistic interactions strongly varies across regions and shows incongruence across the groups studied here.

### *Hotspots of Species Richness and Phylogenetic Diversity in the Neotropics*

For the three genera studied here, the eastern part of the Andes is one of the regions with highest species richness and phylogenetic diversity (PD and ES) while the poorest regions are the lower-Amazon, the Cerrado and the Atlantic-forest. *Napeogenes* and *Oleria* show a relatively similar secondary peak of diversity in the



upper-Amazon. By contrast, *Ithomia* exhibits low diversity in the upper-Amazon but maximum species richness in Central-America. The latter pattern is due mainly to the diversification of a single clade in Central-America, which explains the relatively intermediate values of PD and ES (i.e., not maximum values). Interestingly, the Central-America diversity peak also corresponds to a mountainous region.

Within each genus, species richness, PD and ES show a very strong pattern of covariation. This is likely due to the fact that these indices are summed across all species in a grid cell and are therefore strongly influenced by species richness (see Rodrigues et al. 2005; Davies and Cadotte 2011). This may be particularly important for the three genera studied here because their phylogenetic trees are rather balanced, resulting in no major differences in phylogenetic diversity among species (see Rodrigues et al. 2005 for an analysis of PD in this respect). But, as shown in previous works, the congruence among different indices is not perfect throughout the spectrum of species richness distribution. Here, differences among measures are more obvious in areas with intermediate or intermediate to low species richness. Differences between species and phylogenetic diversity are likely to be common for relatively low species richness areas, because such areas could harbour distantly related species and/or phylogenetically distinctive species, resulting in high PD and ES values. For example, Arponen and Zupan (chapter “[Representing Hotspots of Evolutionary History in Systematic Conservation Planning for European Mammals](#)”) found major differences between phylogenetic diversity and species richness for mammals in areas of low diversity in the north of Europe.

MPD captures the average relatedness of the pairs of species in each grid cell, and high values indicate the presence of pairs of distant relatives in species assemblages. As a mean value it is independent of species richness, but its variance increases with low species richness. However, it provides useful information related to the diversification history of a clade. For example, the increase of MPD for *Oleria*, from northwest toward southeast, is explained by phylogenetically independent colonisation of these regions.

One of the first studies investigating the usefulness of ithomiines as biogeographic indicators suggested that they could be good surrogates of total butterfly diversity in lowland Neotropical forests (Beccaloni and Gaston 1995). Our results are mostly consistent with that suggestion, since peaks in richness in the eastern Andes and upper Amazon, as identified here, have also been reported in *Heliconius* butterflies (Rosser et al. 2012) and the genus *Adelpha* (Willmott 2003).

Studies on other taxa have also found a pattern of high diversity in the upper Amazon, based on various different measures. For example, using a dataset of 50 clades (López-Osorio and Miranda-Esquivel 2010), found that species richness and evolutionary distinctiveness of several groups of vertebrates and some groups of insects and plants are high in the southern upper Amazon. But, unlike the genera studied here, they also found very high diversity in the Guianas (see also Miranda-Esquivel chapter “[Support in Area Prioritization Using Phylogenetic Information](#)”). Similarly, Amori et al. 2013 noted that rodent diversity peaks in the upper Amazon, but also found diversity hotspots in the Guianas and Atlantic forest. Primates similarly show increasing diversity from east to west (Da Silva et al. 2005b), as well as birds (Haffer 1990), non-volant mammals (Costa et al. 2000) and plants (Ter Steege

et al. 2000). Many factors are likely to contribute to the general high species richness in the western Amazon and Andean foothills. Along the eastern Andes, high turnover in abiotic conditions, habitat types, vegetation and host-plants for phytophagous insects, in addition to topological complexity, may explain a high species turnover within a grid cell, therefore increasing diversity. All these factors are also potential drivers of speciation, which also contributes to increase diversity. The diversification histories across geographical areas also account for patterns of diversity of different organisms. In the case of ithomiine butterflies, previous studies found that *Napeogenes*, *Ithomia* and *Oleria* likely originated in the northern Andes and subsequently diversified throughout both the Andes and the rest of the Neotropics (Elias et al. 2009; de-Silva et al. 2015; de-Silva et al. 2010). Shifts of altitudinal range and colour pattern are also correlated (Chazot et al. 2014) and are involved in speciation events (Jiggins et al. 2006; Elias et al. 2009), and may likely have increased speciation rate in montane regions. In addition, hostplant diversity has been proposed to drive diversification in phytophagous insects (Janz et al. 2006) and particularly in ithomiines (Willmott and Freitas 2006), whose Solanaceae hostplants are most diverse in the Andes and the upper Amazon, and to a lesser extent in the Atlantic Forest (Knapp 2002; PBI Solanum Project; <http://www.nhm.ac.uk/research-curation/research/projects/solanaceaesource/>). Understanding the ecology and the diversification history of different groups of organisms may therefore lead to a better explanation of diversity patterns in the Neotropics and improve conservation strategies. For this reason, no single group of organism can be a good indicator of general patterns of diversity. Approaches that rely on a wide range of taxa (e.g., López-Osorio and Miranda-Esquivel 2010) are more powerful in this respect.

### ***Müllerian Mimicry: Patterns of Diversity and Community Vulnerability***

Müllerian mimicry affects both local and regional species assemblages, and mutualistic mimetic interactions have apparently led to adaptively structured assemblages (Elias et al. 2008; Chazot et al. 2014). In this study we tried to capture the importance of ecological interactions by using Müllerian mimicry as an example. We first measured mimicry pattern of richness. This measure is relatively correlated with both species richness and phylogenetic diversity and thus shows a consistent peak along the Andes across the three genera and in the upper-Amazon for *Napeogenes* and *Oleria*. *Napeogenes* appears to be the genus in which most mimicry patterns co-occur, and it is the most polymorphic genus studied here, with, for example, twice as many mimicry patterns as *Oleria*. In contrast to *Oleria* and *Ithomia*, some *Napeogenes* species have opaque wings with bright orange, yellow and black patterns. Interestingly, we found two main centres of similar mimicry diversity in *Oleria*: along the Andes and in the upper Amazon. *Oleria* is the most

species-rich genus but, perhaps surprisingly, the poorest in terms of mimicry diversity. However, within genera a positive correlation between local species richness and the local number of mimicry patterns across the generic range is expected, a pattern confirmed in our analyses for all genera. This result implies that, at a local scale, preserving high species diversity should preserve mimicry diversity as well.

Mimicry ring formation is the convergence of multiple co-occurring species toward a similar aposematic colour pattern. A species benefits from the presence of locally co-mimetic species. Thus there is high interdependency between co-mimetic species, such that extinction of one species may affect its co-mimics and induce cascading extinctions. Our vulnerability metric was based on the assumption that the more species share a mimicry pattern the less species are sensitive to extinction. We found very distinct patterns across genera: (1) for *Oleria*, vulnerability is lowest in the upper-Amazon, Cerrado and Atlantic Forest and sharply increases at the edges of these regions and in Central America; (2) vulnerability of *Napeogenes* communities is high everywhere, with the lowest values in central Amazonia and east of the lower Amazon; (3) for *Ithomia*, the lowest vulnerability is found in the southeastern Neotropics and in Central America. Overall, these results broadly reflect the extent to which each genus numerically dominates butterfly communities in each region. *Oleria* are abundant members of Amazon forest communities and tend to be co-mimetic, leading to low vulnerability in these regions. *Ithomia* are abundant in the southwestern Amazon and Atlantic forests and in Nicaragua-Costa Rica, where they show low vulnerability, while *Napeogenes*, most of whose species are rare everywhere, proves also to be vulnerable everywhere.

This analysis is a preliminary exploration of the possibility of using more sophisticated measures related to the ecology of Neotropical butterflies. As such, it suffers from several problems. Our dataset includes about 100 species out of ca. 380 ithomiine species, many of which are involved in the mimicry rings considered here but ignored in our vulnerability index. Similarly, many other taxonomic groups are members of ithomiine mimicry rings, particularly Heliconiini (Nymphalidae) and several diurnal moths. Finally, mimicry rings may also involve Batesian (non-poisonous) mimics such as some Pieridae and Nymphalidae (e.g., Beccaloni 1997a), which weakens the protection given by mimicry. Our metric may therefore misestimate (and most likely overestimate) vulnerability of some mimicry rings, particularly those found in *Ithomia* and *Napeogenes*, which include many species of other ithomiine genera. Nevertheless, despite not being optimal our vulnerability analysis draws attention to three aspects. Firstly, biotic interactions – mutualistic interactions in this case – are not homogeneously distributed across the Neotropics, and may strongly influence sensitivity to extinction of butterfly assemblages across space. Secondly, biotic interactions are not homogeneously distributed across taxa, meaning that the pattern of one clade is not necessarily similar to another one. And thirdly, because ithomiines numerically dominate many butterfly assemblages across the Neotropics, they probably condition to a certain extent the distribution of other species that interact with them in a positive or negative way.

## Conclusion

By studying large-scale diversity patterns of three butterfly genera we found a common pattern of high diversity along the eastern Andes, broadly similar to what has previously been found in other animal and plant groups. However, we also found some differences among the genera, which result from different evolutionary histories. For ithomiine butterflies, we argue that the east Andean slopes and foothills and the upper Amazon region should be the highest priority for conservation. The upper Amazon has already received attention and protected areas have been defined. Similar conservation plans should now focus on the Andean region. However, mountainous regions in Panama and Costa-Rica appear as a secondary diversity hotspot, not as rich but with highly distinct and endemic faunas that are also signifi-

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

cant for conservation. Moreover, conservation efforts should not only focus on diversity hotspots, but also on regions where diversification is high. Diversification rates are likely to be particularly high in mountain areas, where rapid turnover of environmental conditions and complex topography are drivers of speciation.

The present study clearly shows that a continental approach can assist in identifying conservation priorities in macroregions, based on regional phylogenetic diversity and vulnerability of regional (and local) networks (in this case, using mimicry rings as a proxy), as also proposed by Arponen and Zupan (chapter “[Representing Hotspots of Evolutionary History in Systematic Conservation Planning for European Mammals](#)”) for mammals in Europe. Although macroecological, regional patterns can appear imprecise and general in some cases, they are of extreme importance for identifying areas where local-scale studies should be conducted to better understand the value of the interaction networks, and their vulnerability to environmental disturbance. Furthermore, metrics which appear similar or to be surrogates of one another can be used to identify priorities among alternative sites. For example, given two sites with identical species richness, differences in phylogenetic and/or ecological indices could help to discriminate between them. In our study, mimicry diversity and vulnerability are clearly related to functional diversity. In the future, our results could be expanded with the addition of other important functional traits of these butterflies, such as body size, host plant use or flight height, helping to better understand the complex and megadiverse Neotropical communities.

**Acknowledgements** AVLF acknowledges support from the FAPESP (BIOTA-FAPESP 2011/50225-3), from the Brazilian Research Council-CNPq (fellowship 302585/2011-7, and “SISBIOTA-Brasil/CNPq grant 563332/2010-7), and from the collaborative grant “Dimensions US-Biota-São Paulo: Assembly and evolution of the Amazon biota and its environment: an inte-

grated approach”, US NSF, NASA, and FAPESP (2012/50260-6), and from the National Science Foundation (DEB-1256742). KRW acknowledges support from the Leverhulme Trust, the National Geographic Society, the Darwin Initiative, the National Science Foundation (DEB-0639861, DEB-0103746), and the Museo Ecuatoriano de Ciencias Naturales and Ministerio del Ambiente in Quito, for support in museum and field work. We also thank the many museum curators and individuals who permitted us to examine their collections and record distributional data, especially Philip Ackery and Jim Reynolds (BMNH), Gerardo Lamas (MUSM), Robert Robbins (USNM), and Eric Quinter (AMNH). Finally, KRW thanks Jim Mallet for his support of ithomiine research over many years, Jason Hall for many Ecuadorian distribution records, and the multiple students, volunteers and assistants, especially Fraser Simpson, who helped in databasing specimens. ME, LdS and NC acknowledge support from CNRS (ATIP funding). NC is funded by a doctoral grant from Doctoral School ED227, Ministère de la Recherche, at the Muséum National d'Histoire Naturelle (Paris, France). Finally we acknowledge reviewers Matjaž Kuntner and Claudia Elizabeth Moreno Ortega for their comments on the manuscript.

## References

- Amori G, Chiozza F, Patterson BD, Rondinini C, Schipper J, Luiselli L (2013) Species richness and distribution of Neotropical rodents, with conservation implications. *Mammalia* 77:1–19
- Basset Y, Eastwood R, Sam L, Lohman DJ, Novotny V, Treuer T, Miller SE, Weiblen GD, Pierce NE, Bunyavejchewin S, Sakchoowong W, Kongnoo P, Osorio-Arenas MA (2013) Cross-continental comparisons of butterfly assemblages in tropical rainforests: implications for biological monitoring. *Insect Conserv Divers* 6:223–233
- Beccaloni GW (1997a) Ecology, natural history and behaviour of ithomiine butterflies and their mimics in Ecuador (Lepidoptera: Nymphalidae: Ithomiinae). *Trop Lepidoptera* 8:103–124
- Beccaloni GW (1997b) Vertical stratification of ithomiine butterfly (Nymphalidae: Ithomiinae) mimicry complexes: the relationship between adult flight height and larval host-plant height. *Biol J Linn Soc* 62:313–341
- Beccaloni GW, Gaston KJ (1995) Predicting the species richness of neotropical forest butterflies – Ithomiinae (Lepidoptera, Nymphalidae) as indicators. *Biol Conserv* 71:77–86
- Boggs CL, Watt WB, Ehrlich PR (2003) Butterflies: ecology and evolution taking flight. The University of Chicago Press, Chicago
- Bonebrake TC, Ponisio LC, Boggs CL, Ehrlich PR (2010) More than just indicators: a review of tropical butterfly ecology and conservation. *Biol Conserv* 143:1831–1841
- Brower AVZ, Willmott KR, Silva-Brandão KL, Garzón-Orduña JJ, Freitas AVL (2014) Phylogenetic relationships of ithomiine butterflies (Lepidoptera: Nymphalidae: Danainae) as implied by combined morphological and molecular data. *Syst Biodivers* 12(2):133–147
- Brown KS (1984) Adult-obtained pyrrolizidine alkaloids defend ithomiine butterflies against a spider predator. *Nature* 309:707–709
- Brown KS (1985) Chemical ecology of dehydropyrrolizidine alkaloids in adult Ithomiinae (Lepidoptera, Nymphalidae). *Rev Bras Biol* 44:435–460
- Brown KS (1997) Diversity, disturbance, and sustainable use of Neotropical forests: insects as indicators for conservation monitoring. *J Insect Conserv* 1:25–42
- Brown KS, Benson WW (1974) Adaptive polymorphism associated with multiple Müllerian mimicry in *Heliconius numata* (Lepid. Nymph.). *Biotropica* 6:205–228
- Brown KS, Freitas AVL (2000) Atlantic forest butterflies: indicators for landscape conservation. *Biotropica* 32:934–956
- Chazot N, Willmott KR, Santacruz Endara PG, Toporov A, Hill RI, Jiggins CD, Elias M (2014) Mutualistic mimicry and filtering by altitude shape the structure of Andean butterfly Communities. *Am Nat* 183:26–39

- Chen IC, Hill JK, Shiu HJ, Holloway JD, Benedick S, Chey VK, Barlow HS, Thomas CD (2011) Asymmetric boundary shifts of tropical montane Lepidoptera over four decades of climate warming. *Glob Ecol Biogeogr* 20:34–45
- Costa LP, Leite YLR, da Fonseca GAB, da Fonseca MT (2000) Biogeography of South American forest mammals: endemism and diversity in the Atlantic Forest. *Biotropica* 32:872–881
- Crisp MD, Arroyo MTK, Cook LG, Gandolfo MA, Jordan GJ, McGlone MS, Weston PH, Westoby M, Wilf P, Linder HP (2009) Phylogenetic biome conservatism on a global scale. *Nature* 458:754–756
- Da Silva JMC, Rylands AB, Da Fonseca GAB (2005a) The fate of the Amazonian areas of endemism. *Conserv Biol* 19:689–694
- Da Silva JMC, Rylands AB, Júnior JSS, Gascon C, da Fonseca GAB (2005b) Primate diversity patterns and their conservation in Amazonia. In: *Phylogeny and conservation*. University Press Cambridge, Cambridge, pp 337–364
- Davies TJ, Cadotte MW (2011) Quantifying biodiversity: does it matter what we measure? In: Zachos FE, Habel JC (eds) *Biodiversity hotspots*. Springer, Berlin, pp 43–60
- De Silva DL, Elias M, Willmott K, Mallet J, Day JJ (2015) Diversification of clearwing butterflies with the rise of the Andes. *J Biogeogr* (in press)
- de-Silva DL, Day JJ, Elias M, Willmott K, Whinnett A, Mallet J (2010) Molecular phylogenetics of the neotropical butterfly subtribe Oleriina (Nymphalidae: Danainae: Ithomiini). *Mol Phylogenet Evol* 55:1032–1041
- Devicor V, Mouillot D, Meynard C, Jiguet F, Thuiller W, Mouquet N (2010) Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: the need for integrative conservation strategies in a changing world. *Ecol Lett* 13:1030–1040
- DeVries PJ, Lande R, Murray D (1999) Associations of co-mimetic ithomiine butterflies on small spatial and temporal scales in a neotropical rainforest. *Biol J Linn Soc* 67:73–85
- Duarte LDS, Prieto PV, Pillar VD (2012) Assessing spatial and environmental drivers of phylogenetic structure in Brazilian Araucaria forests. *Ecography* 35:952–960
- Elias M, Gompert Z, Jiggins C, Willmott K (2008) Mutualistic interactions drive ecological niche convergence in a diverse butterfly community. *Plos Biol* 6:2642–2649
- Elias M, Joron M, Willmott K, Silva-Brandao KL, Kaiser V, Arias CF, Pinerez LMG, Uribe S, Brower AVZ, Freitas AVL, Jiggins CD (2009) Out of the Andes: patterns of diversification in clearwing butterflies. *Mol Ecol* 18:1716–1729
- Engler R, Randin CF, Vittoz P, Czaka T, Beniston M, Zimmermann NE, Guisan A (2009) Predicting future distributions of mountain plants under climate change: does dispersal capacity matter? *Ecography* 32:34–45
- Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61:1–10
- Faith DP (2002) Quantifying biodiversity: a phylogenetic perspective. *Conserv Biol* 16:248–252
- Faith DP (2008) Threatened species and the potential loss of phylogenetic diversity: conservation scenarios based on estimated extinction probabilities and phylogenetic risk analysis. *Conserv Biol* 22:1461–1470
- Feeley KJ, Silman MR, Bush MB, Farfan W, Cabrera KG, Malhi Y, Meir P, Revilla NS, Quisipupanqui MNR, Saatchi S (2011) Upslope migration of Andean trees. *J Biogeogr* 38:783–791
- Flynn DFB, Mirotchnick N, Jain M, Palmer MI, Naeem S (2011) Functional and phylogenetic diversity as predictors of biodiversity-ecosystem-function relationships. *Ecology* 92:1573–1581
- Gaston KJ (2000) Global patterns in biodiversity. *Nature* 405:220–227
- Gaston KJ, Blackburn TM (2000) *Pattern and process in macroecology*. Blackwell Science, Oxford
- Gaston KJ, Hudson E (1994) Regional patterns of diversity and estimates of global insect species richness. *Biodivers Conserv* 3:493–500
- Griscom HP, Ashton MS (2011) Restoration of dry tropical forests in Central America: a review of pattern and process. *For Ecol Manage* 261:1564–1579
- Haffer J (1990) Avian species richness in tropical South-America. *Stud Neotropical Fauna Environ* 25:157–183

- Hartmann K, Andre J (2013) Should evolutionary history guide conservation? *Biodivers Conserv* 22:449–458
- Hartmann K, Steel MA (2007) Phylogenetic diversity: from combinatorics to ecology. In: Gascuel O, Steel MA (eds) *Reconstructing evolution: new mathematical and computational advances*. Oxford University Press, New York, pp 171–196
- Hawkins BA, Diniz-Filho JAF, Jaramillo CA, Soeller SA (2007) Climate, niche conservatism, and the global bird diversity gradient. *Am Nat* 170:S16–S27
- Hill RI (2010) Habitat segregation among mimetic ithomiine butterflies (Nymphalidae). *Evol Ecol* 24:273–285
- Howard PC, Viskanic P, Davenport TRB, Kigenyi FW, Baltzer M, Dickinson CJ, Lwanga JS, Matthews RA, Balmford A (1998) Complementarity and the use of indicator groups for reserve selection in Uganda. *Nature* 394:472–475
- Janz N, Nylin S, Wahlberg N (2006) Diversity begets diversity: host expansions and the diversification of plant-feeding insects. *BMC Evol Biol* 6:4
- Jiggins CD, Mallarino R, Willmott KR, Bermingham E (2006) The phylogenetic pattern of speciation and wing pattern change in neotropical Ithomia butterflies (Lepidoptera: Nymphalidae). *Evolution* 60:1454–1466
- Knapp S (2002) Assessing patterns of plant endemism in neotropical uplands. *Bot Rev* 68:22–37
- Lamas G (2004) Checklist: Part 4A. Hesperioidea – Papilionoidea. In: *Atlas of neotropical Lepidoptera*. Scientific Publisher, Gainesville
- Lamoreux JF, Morrison JC, Ricketts TH, Olson DM, Dinerstein E, McKnight MW, Shugart HH (2006) Global tests of biodiversity concordance and the importance of endemism. *Nature* 440:212–214
- Loiselle BA, Graham CH, Goerck JM, Ribeiro MC (2010) Assessing the impact of deforestation and climate change on the range size and environmental niche of bird species in the Atlantic forests, Brazil. *J Biogeogr* 37:1288–1301
- López-Osorio F, Miranda-Esquivel DR (2010) A phylogenetic approach to conserving Amazonian biodiversity. *Conserv Biol* 24:1359–1366
- Mace GM, Gittleman JL, Purvis A (2003) Preserving the tree of life. *Science* 300:1707–1709
- MacLaurin J, Sterelny K (2008) *What is biodiversity*. The University of Chicago Press, Chicago
- Mallarino R, Bermingham E, Willmott KR, Whinnett A, Jiggins CD (2005) Molecular systematics of the butterfly genus Ithomia (Lepidoptera: Ithomiinae): a composite phylogenetic hypothesis based on seven genes. *Mol Phylogenet Evol* 34:625–644
- Mallet J (1999) Causes and consequences of a lack of coevolution in Mullerian mimicry. *Evol Ecol* 13:777–806
- Mooers AO, Heard SB, Chrostowski E (2005) Evolutionary heritage as a metric for conservation. In: Purvis A, Gittleman JL, Brooks T (eds) *Phylogeny and conservation*, Cambridge University Press, pp 120–138
- Müller F (1879) Ituna and Thyridia; a remarkable case of mimicry in butterflies. *Trans Entomol Soc Lond*, xx–xxix
- Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* 403:853–858
- New TR (1993) Angels on a pin – dimensions of the crisis in invertebrate conservation. *Am Zool* 33:623–630
- Purvis A, Agapow PM, Gittleman JL, Mace GM (2000) Nonrandom extinction and the loss of evolutionary history. *Science* 288:328–330
- Redding DW, Hartmann K, Mirmoto A, Bokal D, Devos M, Mooers AO (2008) Evolutionarily distinctive species often capture more phylogenetic diversity than expected. *J Theor Biol* 251:606–615
- Ribeiro MC, Metzger JP, Martensen AC, Ponzoni FJ, Hirota MM (2009) The Brazilian Atlantic forest: how much is left, and how is the remaining forest distributed? Implications for conservation. *Biol Conserv* 142:1141–1153

- Rodrigues ASL, Brooks TM, Gaston KJ (2005) Integrating phylogenetic diversity in the selection of priority areas for conservation: does it make a difference? In: Purvis A, Gittleman JL, Brooks T (eds) *Phylogeny and conservation*, Cambridge University Press, pp 101–119
- Rosser N, Phillimore AB, Huertas B, Willmott KR, Mallet J (2012) Testing historical explanations for gradients in species richness in heliconiine butterflies of tropical America. *Biol J Linn Soc* 105:479–497
- Safi K, Cianciaruso MV, Loyola RD, Brito D, Armour-Marshall K, Diniz JAF (2011) Understanding global patterns of mammalian functional and phylogenetic diversity. *Phil Trans R Soc B-Biol Sci* 366:2536–2544
- Samways MJ (1994) *Insect conservation biology*. Chapman & Hall, London
- Schweiger O, Klotz S, Durka W, Kuhn I (2008) A comparative test of phylogenetic diversity indices. *Oecologia* 157:485–495
- Sheldon KS, Yang S, Tewksbury JJ (2011) Climate change and community disassembly: impacts of warming on tropical and temperate montane community structure. *Ecol Lett* 14:1191–1200
- SOS Mata Atlântica, INPE, ISA, eds (1998) *Atlas da Evolução dos Remanescentes Florestais e Ecossistemas Associados do Domínio da Mata Atlântica no Período 1990–1995*, Fundação SOS Mata Atlântica, Instituto Nacional de Pesquisas Espaciais e Instituto Socioambiental, São Paulo
- Sparrow HR, Sisk TD, Ehrlich PR, Murphy DD (1994) Techniques and guidelines for monitoring neotropical butterflies. *Conserv Biol* 8:800–809
- Ter Steege H, Sabatier D, Castellanos H, Van Andel T, Duivenvoorden J, De Oliveira AA, Ek R, Lilwah R, Maas P, Mori S (2000) An analysis of the floristic composition and diversity of Amazonian forests including those of the Guiana Shield. *J Trop Ecol* 16:801–828
- Vane-Wright RI, Humphries CJ, Williams PH (1991) What to protect – systematics and the agony of choice. *Biol Conserv* 55:235–254
- Warren MS, Hill JK, Thomas JA, Asher J, Fox R, Huntley B, Roy DB, Telfer MG, Jeffcoate S, Harding P, Jeffcoate G, Willis SG, Greatorex-Davies JN, Moss D, Thomas CD (2001) Rapid responses of British butterflies to opposing forces of climate and habitat change. *Nature* 414:65–69
- Williams P, Gibbons D, Margules C, Rebelo A, Humphries C, Pressey R (1996) A comparison of richness hotspots, rarity hotspots, and complementary areas for conserving diversity of British birds. *Conserv Biol* 10:155–174
- Willmott KR (2003) Cladistic analysis of the Neotropical butterfly genus *Adelpha* (Lepidoptera: Nymphalidae), with comments on the subtribal classification of Limenitidini. *Syst Entomol* 28:279–322
- Willmott KR, Freitas AVL (2006) Higher-level phylogeny of the Ithomiinae (Lepidoptera: Nymphalidae): classification, patterns of larval hostplant colonization and diversification. *Cladistics* 22:297–368
- Willmott KR, Lamas G (2007) A revision of *Pachacutia*, a new genus of rare Andean ithomiine butterflies (Nymphalidae: Ithomiinae), with the description of two new species. *Ann Entomol Soc Am* 100:449–469
- Willmott KR, Mallet J (2004) Correlations between adult mimicry and larval host plants in ithomiine butterflies. *Proc R Soc B-Biol Sci* 271:S266–S269
- Yessoufou K, Daru BH, Davies TJ (2012) Phylogenetic patterns of extinction risk in the eastern arc ecosystems, an African biodiversity hotspot. *PLoS One* 7:e47082



# Conservation of Phylogenetic Diversity in Madagascar's Largest Endemic Plant Family, Sarcolaenaceae

Anaëlle Soulebeau, Roseli Pellens, Porter P. Lowry II, Xavier Aubriot, Margaret E.K. Evans, and Thomas Haevermans

**Abstract** Madagascar is renowned for its impressive species richness and high level of endemism, which led to the island being recognized as one of the world's most important biodiversity hotspots. As in many other regions, Madagascar's biodiversity is highly threatened by unsustainable anthropogenic disturbance, leading to widespread habitat loss and degradation. Although the country has significantly expanded its network of protected areas (PAs), current protocols for identifying priority areas are based on traditional measures that could fail to ensure maximal

---

A. Soulebeau (✉) • T. Haevermans

Institut de Systématique, Evolution, Biodiversité, ISYEB – UMR 7205 CNRS MNHN  
UPMC EPHE, Muséum National d'Histoire Naturelle, Sorbonne Universités,  
57 rue Cuvier, CP 50, 75005 Paris, France  
e-mail: [asoulebeau@mnhn.fr](mailto:asoulebeau@mnhn.fr); [haever@mnhn.fr](mailto:haever@mnhn.fr)

R. Pellens

Institut de Systématique, Evolution, Biodiversité, ISYEB – UMR 7205 CNRS MNHN  
UPMC EPHE, Muséum National d'Histoire Naturelle, Sorbonne Universités,  
45 rue Buffon, CP 50, 75005 Paris, France  
e-mail: [pellens@mnhn.fr](mailto:pellens@mnhn.fr)

P.P. Lowry II

Institut de Systématique, Evolution, Biodiversité, ISYEB – UMR 7205 CNRS MNHN  
UPMC EPHE, Muséum National d'Histoire Naturelle, Sorbonne Universités,  
57 rue Cuvier, CP 50, 75005 Paris, France

Missouri Botanical Garden, PO Box 299, St. Louis, MO 63166-0299, USA  
e-mail: [pete.lowry@mobot.org](mailto:pete.lowry@mobot.org)

X. Aubriot

Institut de Systématique, Evolution, Biodiversité, ISYEB – UMR 7205 CNRS MNHN  
UPMC EPHE, Muséum National d'Histoire Naturelle, Sorbonne Universités,  
57 rue Cuvier, CP 50, 75005 Paris, France

Department of Life Sciences, Natural History Museum,  
Cromwell Road, London SW7 5BD, UK  
e-mail: [x.aubriot@gmail.com](mailto:x.aubriot@gmail.com)

inclusion of the country's biodiversity. In this study, we use Madagascar's largest endemic plant family, Sarcolaenaceae, as a model to identify areas with high diversity and to explore the potential conservation importance of these areas. Using phylogenetic information and species distribution data, we employ three metrics to study geographic patterns of diversity: species richness, Phylogenetic Diversity (PD) and Mean Phylogenetic Diversity (MPD). The distributions of species richness and PD show considerable spatial congruence, with the highest values found in a narrow localized region in the central-northern portion of the eastern humid forest. MPD is comparatively uniform spatially, suggesting that the balanced nature of the phylogenetic tree plays a role in the observed congruence between PD and species richness. The current network of PAs includes a large part of the family's biodiversity, and three PAs (Ankeniheny Zahamena Forest Corridor, the Bongolava Forest Corridor and the Itremo Massif) together contain almost 85 % of the PD. Our results suggest that PD could be a valuable source of complementary information for determining the contribution of Madagascar's existing network of PAs toward protecting the country's biodiversity and for identifying priority areas for the establishment of new parks and reserves.

**Keywords** Protected areas • Extinction • Endemism • Biodiversity • Species richness

## Introduction

Among the areas identified by biologists and conservationists as biodiversity hotspots (Myers et al. 2000; Myers 2003), Madagascar is one of the most important because of its exceptionally high levels of species diversity and endemism, along with an unprecedented rate of habitat loss due to anthropogenic activities, leading to species extinction (Goodman and Benstead 2005; Callmänder et al. 2011; Buerki et al. 2013). Less than 10 % of the original natural habitats present on the island before human colonization are still intact (Myers et al. 2000). Although the conservation of Madagascar's biodiversity is a high priority, the dearth of reliable information for identifying priority sites in need of protection complicates the establishment of a robust national conservation program and policy.

---

M.E.K. Evans

Institut de Systématique, Evolution, Biodiversité, ISYEB – UMR 7205 CNRS MNHN  
UPMC EPHE, Muséum National d'Histoire Naturelle, Sorbonne Universités,  
57 rue Cuvier, CP 50, 75005 Paris, France

Laboratory of Tree Ring Research, University of Arizona,  
Bryant Bannister Building, 1215 East Lowell St, Tucson, AZ 85721, USA  
e-mail: [mekevans@email.arizona.edu](mailto:mekevans@email.arizona.edu)

In Madagascar, as in many other regions of the world, species richness and the number of endemic species are the parameters most frequently used to define priorities for biodiversity conservation (Callmander et al. 2007; Kremen et al. 2008). However, as illustrated throughout this book, phylogenetic diversity is another important element that should be taken into consideration, for two main reasons. First, phylogenetic diversity takes into account not only the number of species or endemics in an area but also the evolutionary distinctiveness of those species, such that a site with a legume, an orchid and a fern would be considered to have higher phylogenetic diversity than another site with three species belonging to just one of these groups (Vane-Wright et al. 1991; Faith 1992). Second, measures of phylogenetic diversity are useful in conservation decision-making because extinctions are not random – in many groups where one species is vulnerable, several other related species will tend to be as well. The use of phylogenetic diversity as a criterion in conservation planning thus reduces the risk of losing entire groups or lineages (see Yessoufou and Davies, chapter “[Reconsidering the Loss of Evolutionary History: How Does Non-random Extinction Prune the Tree-of-Life?](#)”).

We might then ask to what extent does Madagascar's system of protected areas help protect key features of the biodiversity within a clade, including not only the number of species, but also phylogenetic and ecological diversity. Patterns in biodiversity distribution can vary considerably from one lineage to another, as shown by two recently published studies on the conservation of biodiversity in Madagascar. While Isambert et al. (2011) showed a striking difference in the spatial distribution of the number of endemic species and phylogenetic diversity of adepagan water beetles, Buerki et al. (2015) revealed a strong congruence between species richness and phylogenetic diversity in the plant family Fabaceae.

Here we use Sarcolaenaceae, the largest plant family endemic to Madagascar, as a case study to identify areas of high phylogenetic diversity and to assess whether the current network of protected areas provides adequate conservation of that diversity.

## ***Madagascar***

Madagascar, located in the Indian Ocean off the coast of southeastern Africa, is well known for its rich and highly endemic flora and fauna (Myers et al. 2000; Myers 2003; Goodman and Benstead 2005). This large continental island separated from mainland Africa ca. 165 Million Years Before Present (MYBP) as part of a block that also included Antarctica and India, subsequently becoming detached from the latter two by 80 MYBP (Schettino and Scotese 2005; Jöns et al. 2009). The resulting long isolation has played a key role in the development and maintenance of Madagascar's striking biota, which exhibits affinities with neighboring Africa, but is also home to groups thought to have reached Madagascar by long-distance dispersal, with their closest relatives occurring in more distant areas such as India, Sri Lanka, Southeast Asia, Australia, New Caledonia and America (Leroy 1978; Schatz

1995; Yoder and Nowak 2006; Warren et al. 2010; Gautier et al. 2012; Buerki et al. 2013; Torsvik et al. 2013).

The evolution of Madagascar's biota has also been driven by the tremendous diversity of environments found on the island, which is underscored by the fact that it has one of the world's highest rates of vertebrate beta-diversity (Holt et al. 2013). The landscape is characterized by a mountainous interior that extends the entire north-south length of the island (ca. 1600 km) resulting in an often sharp altitudinal gradient from the coasts to well over 1000 m in large areas, with many massifs reaching above 1500 m and several dozen peaks surpassing 2000 m, the highest being Maromokotro in the Tsaratanana massif (2876 m). The climate is characterized by a strong precipitation gradient from perennially humid areas on the mountain slopes in the northeastern part of the island, where rainfall may exceed 6000 mm in some years (Thorstrom et al. 1997 in Rakotoarisoa and Be 2004), to a subarid zone in the southwest, which receives less than 300 mm of rain per year and can go without precipitation for 10 months or more (Cornet 1974; Goodman and Benstead 2003). Madagascar's ecosystems in turn reflect the island's relief and climate, ranging from perhumid tropical and montane forests in the east to subhumid and dry formations in the center and west, and subarid ecosystems in the southwest, often with fairly sharp, well delimited boundaries between them. Compounding the spatial arrangement of these biomes is the fact that they are thought to be of different ages. For instance, the spiny subarid vegetation of the southwest is regarded as comparatively old (Paleogene, 23–66 MYBP) while the Sambirano humid forest is likely the youngest biome (Late Miocene, 8 MYBP), originating with the advent of a Monsoon regime in Asia following the uplift of the Himalayas (Wells 2003).

As mentioned above, two particularly striking features of the Malagasy flora are its remarkable species richness and its high level of endemism. A recent assessment indicated that some 11,220 described native vascular plant species are currently recognized, belonging to 1730 genera and 243 families, and that 82 % of these species are endemic (Callmander et al. 2011). Moreover, based on recent taxonomic revisions in a wide range of families (Madagascar Catalogue 2015), an additional ca. 2200 species have been described in the last few years or are awaiting description, nearly all of which will be found to be endemic, along with an estimated 600 more species still to be discovered, thus increasing the total number of native species to ca. 14,000 and the level of endemism to well over 85 % (P. Phillipson personal communication). Equally striking is the level of lineage diversification in Madagascar's flora. The 30 most species-rich families include almost 70 % of the total vascular plant flora as well as 30 % of the genera present on the island, and 38 families include 10 or more genera (Gautier et al. 2012). Moreover, more than 320 genera (19 %) and a total of 5 families are endemic to the island (Callmander et al. 2011; Buerki et al. 2013), Sarcolaenaceae being the largest of these endemic families.

Understanding the origin and diversification of lineages in Madagascar requires consideration of the interplay among the complex eco-geography and geological history of the island, the varying dispersal abilities of the members of the lineages present there, and Madagascar's proximity to potential source areas, in particular the African continent but also Asia and areas beyond. Many studies have been pub-

lished on the evolutionary and ecological processes that have shaped diversity in Madagascar's fauna, but very little attention has been given to its flora. One of the most notable features, still to be explored, is the spatial distribution of plant species richness on the island and the drivers underlying this distribution. Some have suggested that abiotic factors have played an important role, e.g. bio-climate, substrate type, elevation or paleo-precipitation (Yoder and Nowak 2006; Pearson and Raxworthy 2009; Agnarsson and Kuntner 2012; Buerki et al. 2013; Mercier and Wilmé 2013; Rakotoarinivo et al. 2013). Others have explored the role of potential key innovations in species diversification and niche expansion (Vary et al. 2011; Evans et al. 2014; Moore and Robertson 2014).

### ***Biodiversity Conservation in Madagascar***

As mentioned earlier, Madagascar is recognized as one of world's 'hottest' biodiversity hotspots (Myers et al. 2000; Myers 2003; Goodman and Benstead 2005) because its large, diverse and highly endemic biota is severely threatened by unsustainable practices such as shifting agriculture, uncontrolled burning and extensive charcoal production, all of which place intense pressure on the island's remaining natural areas. Over the last three decades a major effort has been made to expand and strengthen the system of protected areas, which now includes ca. 5.7–5.9 million hectares of terrestrial parks and reserves, many of which were established during the last 15 years (Kremen et al. 2008, <http://atlas.rebioma.net>). Despite these efforts, however, deforestation and habitat degradation have continued at an alarming rate as the human population has doubled in the last 25 years, reaching an estimated 22.4 million by mid-2014 (Population Reference Bureau 2015). More than three-quarters of the population lives below the poverty level (World Bank 2015) and almost all Malagasy are directly or indirectly dependent on the island's natural resources as a major source of food, shelter, fuel, and traditional medicine. Studies of Madagascar's forest cover using aerial photographs and Landsat images have estimated a decline in area of 40 % since the 1950s, with a rate of forest loss of 0.9 % per year between 1990 and 2000 (Harper et al. 2007). Organized, large-scale illegal exploitation of precious hardwoods and endangered plant and animal species increased dramatically during the last political crisis (roughly 2009–2014), adding to an already alarming situation (Schuurman and Lowry II 2009; Waeber 2009; Caramel 2015). Furthermore, because a high proportion of Madagascar's species have restricted geographic ranges, they are particularly vulnerable to changes in forest cover. For instance, in a study of 2243 species in 12 different taxonomic groups (including both plants and invertebrates), Allnutt et al. (2008) estimated that 9.2 % of them were driven to extinction between 1950 and 2000 due to forest loss, in addition to the 32.9 % thought to have gone extinct prior to 1950.

Beyond these alarming conclusions, it remains to be seen to what extent the present system of protected areas can effectively preserve what remains of Madagascar's unique biodiversity. Does the system include the full array of species, and do they have populations large enough to be viable over time?

## *Sarcolaenaceae as a Model Group*

Sarcolaenaceae are ideally suited for such a study of phylogenetic diversity because (1) the taxonomy and distribution of its members are particularly well understood and documented, (2) its genera vary in size and its species have a wide range of eco-geographic preferences, and (3) a well-resolved phylogeny is available based on a large, representative sample of species that includes members of all ten genera (Haevermans et al. [in prep](#)).

Sarcolaenaceae comprise 71 species of shrubs and trees belonging to 10 genera (Madagascar Catalogue [2015](#)), each of which has been the subject of a recent taxonomic revision (Hong-Wa [2009](#); Lowry II et al. [1999](#), [2000](#), [2002](#); Randrianasolo and Miller [1994](#), [1999](#); Schatz et al. [2000](#), [2001](#)), followed by the description of several newly discovered species (Lowry II and Rabehevitra [2006](#); Rabehevitra and Lowry II [2009](#); Lowry II et al. [2014](#)). Members of the family are found almost throughout the island, with the notable exception of the subarid southwest, and the distribution of each species has been carefully mapped using the locality information associated with herbarium collections (Ramanananjahary et al. [2010](#); Madagascar Catalogue [2015](#)). Based on the collections in the herbaria of Paris Museum and the Missouri Botanical Garden, we estimate that more than 2000 specimens are available for the family, with an average of 30 geographic occurrences per species and a total number of collections ranging from more than 300 for common, widespread species such as *Leptolaena pauciflora* Baker to just one or a few for species known from a single locality such as *Leptolaena masoalensis* G.E. Schatz & Lowry II, *Schizolaena capuronii* Lowry II et al. and *Schizolaena raymondii* Lowry II and Rabehevitra. The genera of Sarcolaenaceae vary considerably in size, from *Schizolaena* with 22 species, *Sarcolaena* with 8 described species (as well as 6 that remain to be described), to *Mediusella* and *Eremolaena*, which include just 2 and 3 species respectively (Madagascar Catalogue [2015](#)). A little more than half of the species in the family have a restricted geographic distribution, known from fewer than ten localities, and several genera are largely or entirely restricted to a particular climatic region, such as *Eremolaena*, *Leptolaena* *Rhodolaena* and *Schizolaena* which are found primarily or exclusively in humid areas, and *Mediusella* and *Xerochlamys*, which occur only in drier habitats.

The goal of this chapter is to identify areas with the highest levels of phylogenetic diversity of Sarcolaenaceae and to evaluate the degree to which that diversity is captured in existing protected areas. Toward that end, we first show how members of the ten genera are distributed and analyze the geographic distribution of three important diversity measures: species richness, Phylogenetic Diversity (PD) and Mean Phylogenetic Diversity (MPD). We then compare the distribution of these diversity statistics to Madagascar's system of protected areas and point out areas of greatest importance for conservation of phylogenetic diversity in Sarcolaenaceae.

## Material and Methods

### *Phylogenetic Data*

To estimate phylogenetic diversity we used the plastid and nuclear phylogeny of Sarcolaenaceae produced by Haevermans et al. ([in prep](#)). The taxon sampling includes 47 species belonging to the family, for a total of 91 Operational Taxonomic Units (OTUs). All ten genera were represented and 66 % of the species were sampled, thereby capturing most of the taxonomic and morphological/ecological diversity and covering the full geographic distribution of Sarcolaenaceae.

In addition to the 47 Sarcolaenaceae sampled, 6 species were selected from their sister group, Dipterocarpaceae, and 1 species from the next most-closely related family, Cistaceae (Dayanandan et al. 1999; Ducousso et al. 2004; Haevermans et al. [in prep](#)), all of which served as outgroup taxa. Sequence data were obtained from one nuclear (ITS) and three plastid (*rbcL*, *psbA-trnH*, and *psaA-ORF170*) markers (Haevermans et al. [in prep](#)). We performed a Bayesian dating analysis using BEAST v1.7.2 (Drummond et al. 2012) under the uncorrelated lognormal relaxed clock model with a Yule prior on speciation. Data were partitioned according to the number of DNA regions and we applied to each partition the GTR+ I+G substitution model, for reasons outlined by Huelsenbeck and Rannala (2004). An individual MCMC run was conducted for  $20 \times 10^6$  generations, with sampling every 1,000 iterations, thus generating 20,000 chronograms. We discarded the first 25 % of samples as burn-in. Mixing of the chains and their convergence were verified in Tracer 1.4 (Rambaut and Drummond 2007). Using Logcombiner, we merged the remaining 15,000 trees and produced a maximum clade credibility (MCC) chronogram using TreeAnnotator. We applied two temporal constraints to calibrate the tree, one at the split between Dipterocarpaceae and Sarcolaenaceae based on Wikström et al. (2001), and another for the age of the stem-group of the clade comprising *Leptolaena*, *Mediusella*, *Sarcolaena*, *Xerochlamys* and *Xyloolaena* based on the estimated age of a fossil pollen attributable to this group (Coetzee and Muller 1984).

In order to assess the cladogenesis process in Sarcolaenaceae, we measured the degree of imbalance of the Sarcolaenaceae consensus tree topology using the R package *apTreeshape* (Bortolussi et al. 2006), in conjunction with the R package *ape* (Paradis et al. 2004). The imbalance was estimated by calculating the Colless's index (Mooers and Heard 1997). We compared this experimental value against those obtained for 500 simulated trees built under the Equal Rate Markov (ERM) Yule model or the PDA (Proportional to Distinguishable Arrangement) model in which each tree is equally probable (Mooers and Heard 1997), using the function *colless.test()* implemented in R package *apTreeshape* (Bortolussi et al. 2006). We used the “less” and “greater” alternatives to test whether the tree is less unbalanced or more unbalanced than predicted by the null model.



## Measures and Analysis

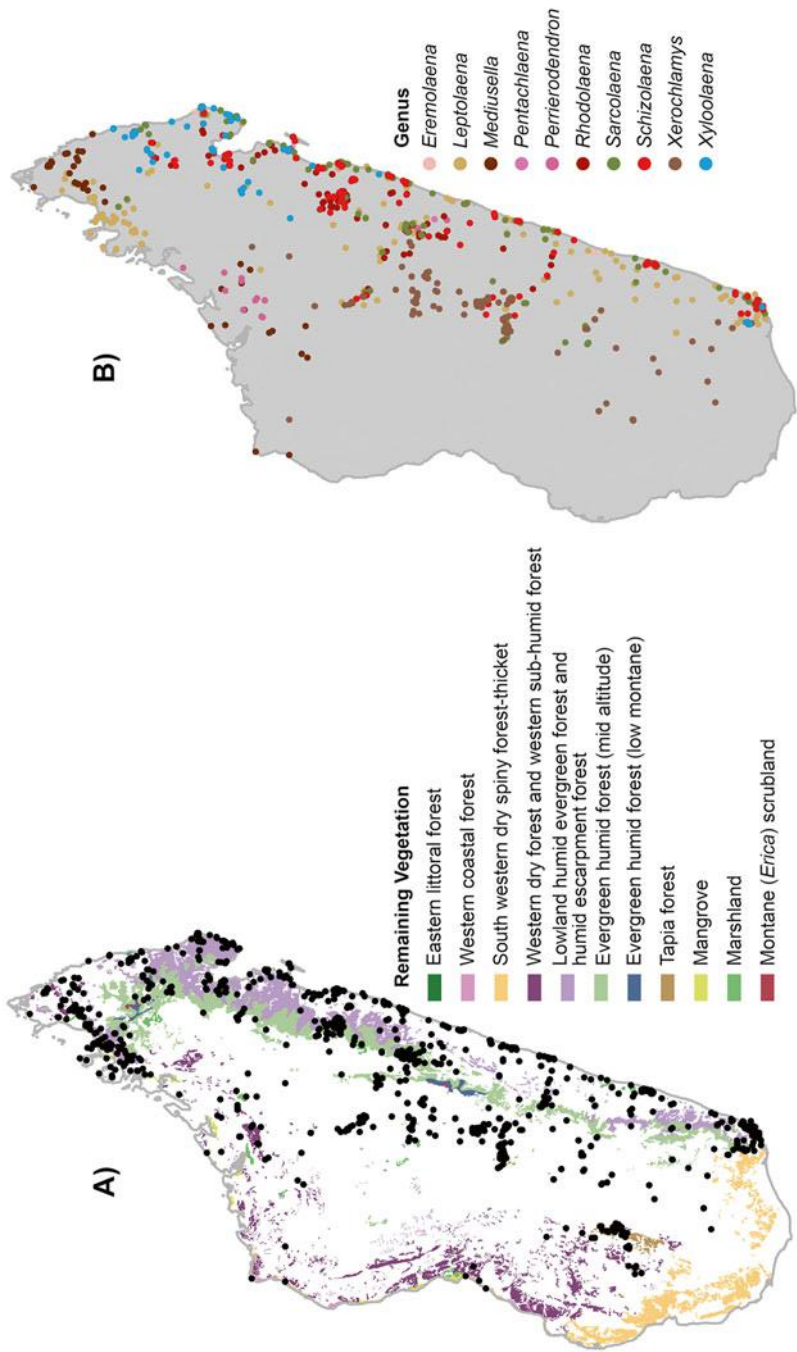
We estimated the area of each species' geographic distribution by creating a minimum convex polygon based on 2148 occurrence points, 1899 of them corresponding to species included in the phylogeny (using ArcMap version 10.2). Occurrences of species with less than three points (for which polygons cannot not be generated) were directly assigned to  $\frac{1}{4}$  degree grid cells (each covering  $30 \times 30$  min) overlaid on a map of Madagascar. Then, a global polygon for the entire family was produced by overlaying all the species polygons, with limits calculated to exclude the sea. Species richness (the number of recorded species) was then calculated in each grid cell, along with two measures of phylogenetic diversity, Faith's PD (Faith 1992) and Mean Phylogenetic Diversity (MPD). PD is a group measure of phylogenetic diversity given by the minimum spanning path along the tree linking all species occurring in a grid cell (see Faith chapter "[The PD Phylogenetic Diversity Framework: Linking Evolutionary History to Feature Diversity for Biodiversity Conservation](#)"). For cells with only one species, the PD value corresponds to the branch length from the tip to the root of the tree. MPD is the mean distance (i.e., mean branch length) between all pairs of species occurring in a given grid cell; this measure provides information on phylogenetic relatedness of the set of species occurring in that cell, controlling for species richness. These two measures were computed using the R package *picante* (Kembel et al. 2010). The distributions of the three measures (species richness, PD and MPD) were then overlaid on the polygon of Sarcolaenaceae occurrence and plotted on a map of Madagascar, which results in parts of the island not being represented in the overall polygon. The resulting maps were compared to the most recent map of protected areas (PA) in Madagascar (<http://atlas.rebioma.net/>). This enabled us to identify whether the cells containing the highest level of PD correspond to those occupied by PAs, and to determine which, if any, cells with high values of PD are located out of the current coverage of Madagascar's PA network.

## Results

Sarcolaenaceae species occur in a wide range of forest ecosystems in Madagascar, from remnant littoral forests scattered along the entire east coast to montane forests on the highest massifs and woodlands in the center, and from the north to the south of the island. By contrast, very few occurrences have been recorded in deciduous seasonally dry forest of the west and none at all from deciduous dry forests and scrubland in the south and southwest (Fig. 1a). By plotting the points for each of the ten genera (Fig. 1b), it can be seen that eastern littoral forest and low- and mid-altitude evergreen humid forests have the greatest diversity, with species in several genera.

The Colless's index obtained when estimating the balance of the Sarcolaenaceae phylogenetic tree is 93, with non-significant p-values for all tests, except when using a PDA model under the "less" alternative (p-value <0.05). These results sug-





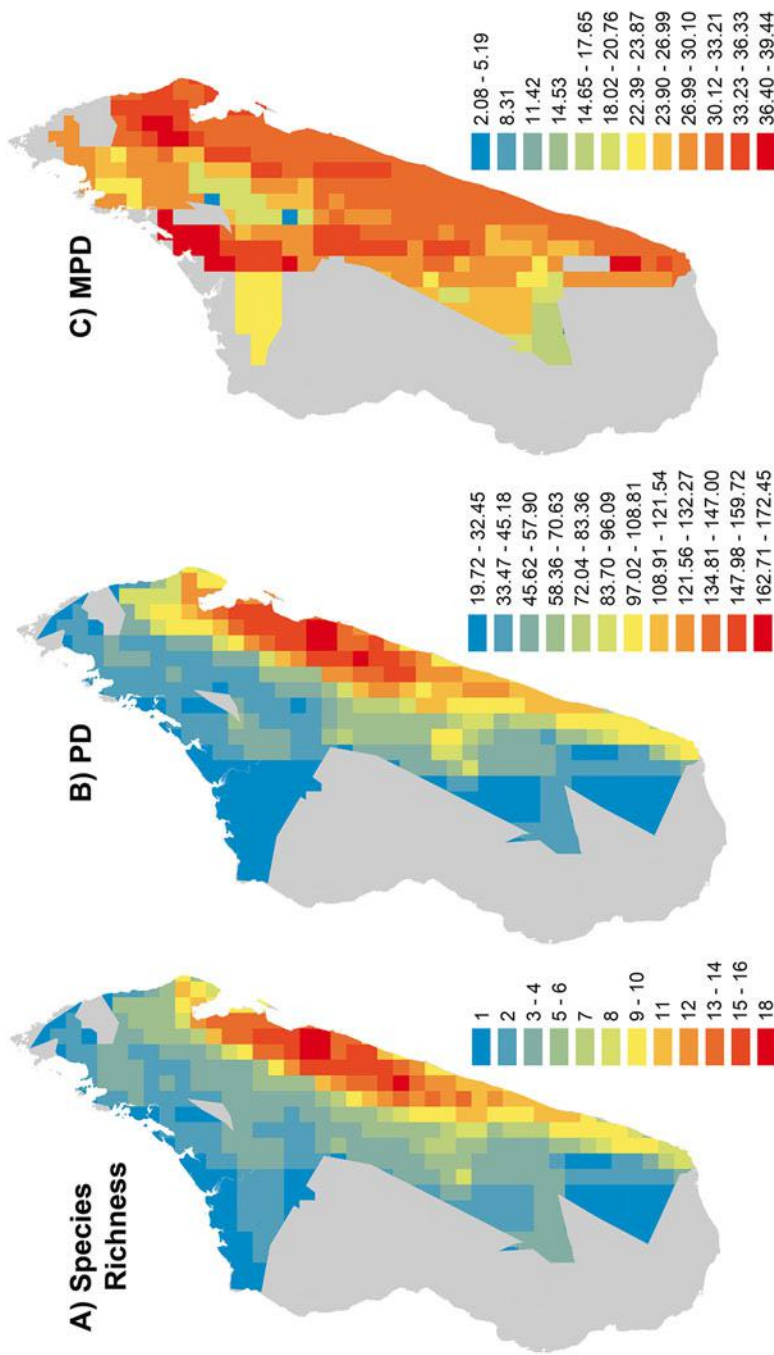
**Fig. 1** Points of occurrence of Sarcolaenaceae in Madagascar. (a) Points of occurrence of 69 species of Sarcolaenaceae in the remaining vegetation of Madagascar. The vegetation map layer is adapted from the vegetation map of Faramelala (1988, 1995), available on ArcGis 10.2. (b) Points of occurrence of 47 species used in this phylogenetic analysis according to the genera they belong. Colors of the same hue indicate close related genera

gest that the Sarcolaenaceae consensus tree is significantly more balanced than one built under a PDA model. The evolutionary history of Sarcolaenaceae thus seems to be characterized by processes that have operated evenly among clades.

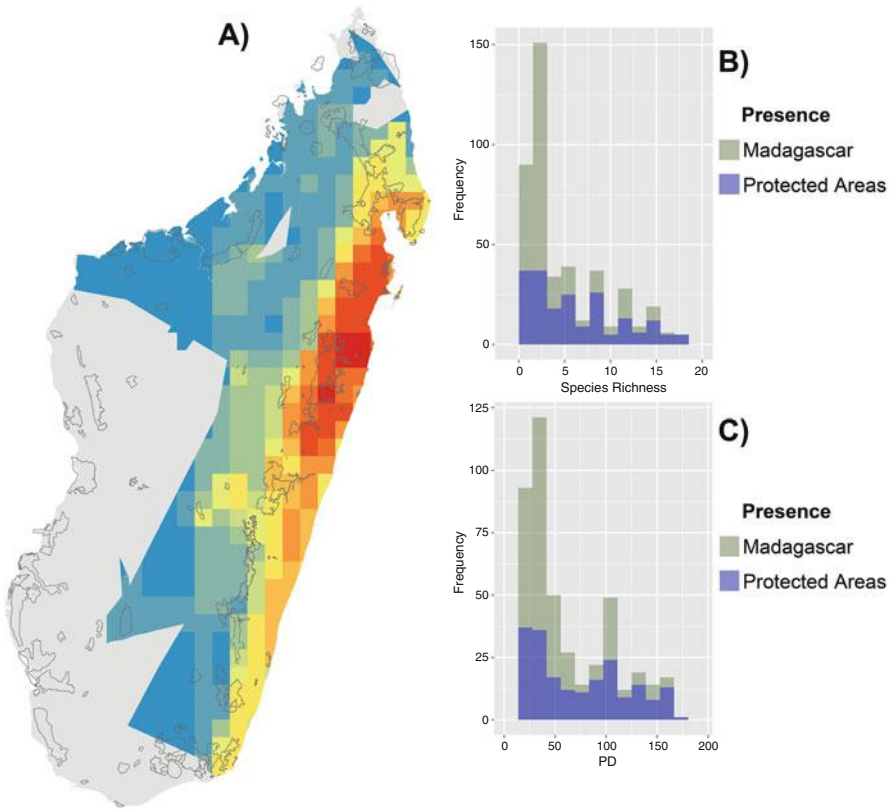
Species richness and Phylogenetic Diversity (PD) vary markedly, from high in eastern Madagascar to low in the west, and to a lesser degree along a north-south gradient (Fig. 2a, b). These two variables show a high degree of spatial congruence (Fig. 2a, b). The areas with the highest values of both species richness and PD are concentrated in the central-northern portion of the eastern escarpment, in regions with both low- and mid-altitude evergreen humid forest. By contrast, Mean Phylogenetic Diversity (MPD) varies much less than species richness and PD (Fig. 2c) and a comparison among the distributions of all three variables does not suggest that cells with higher PD values harbor sets of species that are more distantly related to one another than those found in cells with lower values of PD. The cells with the highest values of MPD, mostly located in the northwestern and northeastern parts of the island (Fig. 2c), have comparatively low values of PD, indicating the occurrence of a limited number of species that are evolutionarily distinct.

Forty-five percent of the cells occupied by Sarcolaenaceae contain at least part of a protected area (Fig. 3a) and the system of PAs is comparatively better represented in cells with higher values of species richness (Fig. 3b) and PD (Fig. 3c). By contrast, most of the cells lacking any PA correspond to those with the lowest values of species richness and PD. All lineages of Sarcolaenaceae and 97.6 % of the total PD are thus found in cells that contain PAs.

Figure 4 shows the PD values for each protected area (Fig. 4a), indicating that areas with the highest PD values in Sarcolaenaceae are concentrated in grid cells in the central-northern portion of the eastern escarpment, centered in the sites comprising the Ankeniheny Zahamena Forest Corridor (Fig. 4b) which contain 64 % of the Sarcolaenaceae's PD. These cells include eight out of ten lineages and all lineages deeply branched. Two lineages are not represented in the Ankeniheny Zahamena Forest Corridor: The genus *Xerochlamys*, which is found from the central region to the south, and the genus *Mediusella* which occurs in the extreme north and in the northeast (see Fig. 1b). Some other PAs exhibit a high level of heterogeneity in PD with parts that show high PD and others that display low PD values. It is the case for Midongy du Sud (Southeast Madagascar), Masoala (perhumid forest in Northeast Madagascar) and the Itremo massif (Central Highlands) (Fig. 4c–e). Interestingly, a few PAs are located in cells with low species richness and PD, but with high value of MPD, in particular Behara-Tranomaro in the southeast (Fig. 4f) and the Bongolava Forest Corridor in the northwest (Fig. 4g).



**Fig. 2** (a) Distribution of the species richness of Sarcolaenaceae in Madagascar; (b) patterns of distribution of the phylogenetic diversity (Faith's PD), and (c) Mean Phylogenetic Distance (MPD)

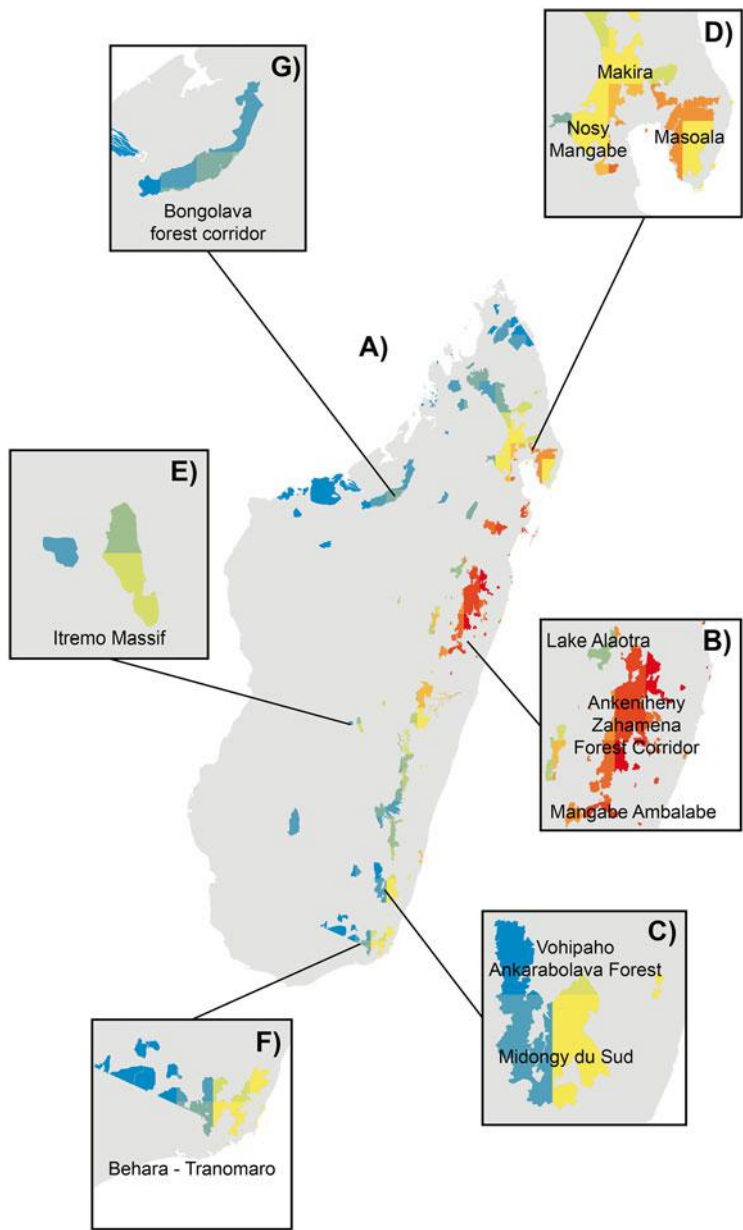


**Fig. 3** (a) PD of Sarcolaenaceae in the network of protected areas in Madagascar; (b) Frequency of species richness and PD (c) all over Madagascar, and in the cells including protected areas

## Discussion

### *Sarcolaenaceae as a Model Group for Conservation in Madagascar*

As mentioned above, Madagascar's biodiversity is among the most distinctive and highly endemic in the world, and the multiple threats it faces result in it being among the most threatened as well. There is thus a strong need to evaluate the effectiveness of conservation efforts, in particular the existing network of protected areas in Madagascar, with respect to their ability to ensure the survival of biodiversity, as measured not only by species richness but also in terms of phylogenetic diversity. In particular, PD is assumed to serve as a valuable tool for developing conservation policies, but to date very few studies have explored it for Madagascar (Sechrest et al. 2002; Magnuson-Ford et al. 2010; Isambert et al. 2011), only one of which



**Fig. 4** (a) PD of Sarcolaenaceae in the Malagasy system of protected areas; (b) Ankeniheny Zahamena Forest Corridor; (c) Midongy du Sud; (d) Mangabe/Masoala; (e) Itremo massif; (f) Behara-Tranomaro; (g) Bongolava Forest Corridor

involving plants (Buerki et al. 2015). In that respect, we focused on Sarcolaenaceae in order to help provide a better understanding of the potential value of PD for conserving Madagascar's biodiversity.

A robust analysis of PD requires a dated phylogeny based on dense taxon sampling, as well as reliable data on the distribution of each species. For Sarcolaenaceae, our sampling comprised nearly 70 % of the total species diversity, with good representation from each of the ten genera in the family. We used the most up-to-date and reliable distributional information, based on more than 2000 occurrence points from the collections kept in the herbaria of the Paris Museum and of the Missouri Botanical Garden examined for recent taxonomic revisions and data from the ca. 40–50 new collections made each year since. The result presented in Fig. 2 provides the first insights into the distribution of species richness and PD for Madagascar's largest endemic plant family, showing that both measures of diversity are highest in areas with humid forest and lowest in dry forests and subarid thickets.

### ***Measures of Biodiversity and Madagascar's Network of Protected Areas***

Our results show a high level of congruence between the distribution of species richness (Fig. 2a) and PD (Fig. 2b). Although not a rule, congruence between species richness and PD is often observed (see for example, Arponen and Zupan, chapter “[Representing Hotspots of Evolutionary History in Systematic Conservation Planning for European Mammals](#)” and Chazot et al. chapter “[Patterns of Species, Phylogenetic and Mimicry Diversity of Clearwing Butterflies in the Neotropics](#)”). This is primarily due to the fact that they both increase as more species are included (see Nipperess, chapter “[The Rarefaction of Phylogenetic Diversity: Formulation, Extension and Application](#)”). But tree shape and the structure of geographic distributions also contribute to variation in congruence between these two statistics. The more balanced a tree is, the more similar each species' contribution will be to overall PD. Likewise, the more species from different parts of the tree co-occur, the higher the congruence between species richness and PD (Rodrigues et al. 2005).

Sarcolaenaceae present a case where both of these factors are at play. The phylogenetic tree is balanced, as shown by Colless's index, yielding little variation among species in PD values. Moreover, the areas with the highest level of species richness contain species belonging to several genera (Fig. 1b) rather than many species in a single genus, as would be expected if overall diversity were the result of radiation of a single lineage within a given eco-geographic zone. Sarcolaenaceae thus present a situation very different from that observed in Malagasy adephagan water beetles by Isambert et al. (2011) but highly similar to that observed in Fabaceae by Buerki et al. (2015), where the distributions of PD and species richness are highly congruent.



In contrast, MPD is independent of species richness, and high values of MPD indicate the presence of distantly-related species co-occurring in a particular area. The balanced nature of the phylogenetic tree partly explains the low variation in Mean Phylogenetic Diversity (MPD) across Madagascar (Fig. 2c). This measure provides additional insight into the distribution of phylogenetic diversity, being of particular interest in those areas with low species richness and low values of PD. In this case, high values of MPD could indicate areas in which ecological convergence has occurred in separate lineages of Sarcolaenaceae. We note, however, the absence of areas that concurrently exhibit low species diversity and high values of PD, although some areas do have low species diversity and low PD but high MPD.

The most important areas for conserving PD in Sarcolaenaceae are concentrated in the central-northern portion of Madagascar's Eastern region, including and adjacent to the eastern edge of the Ankeniheny Zahamena Forest Corridor (Figs. 3 and 4b). However, as this area does not include any representatives of *Xerochlamys* and *Mediusella*, and because the distributions of these two genera do not overlap, the ideal strategy for protecting all lineages of Sarcolaenaceae and to maximize conservation of PD for this family, would be to include two additional protected areas: the Bongolava Forest Corridor in the northwest (Fig. 4g) and the Itremo Massif (Fig. 4e). Taken together, these three regions contain 84.9 % of the PD of Sarcolaenaceae. It is also of critical importance to consider preserving sites with high MPD values because (1) they harbor distantly related species that do not share the same evolutionary history, (2) might be impacted by different threats, and (3) require different conservation procedures. Buerki and colleagues (2015) recently suggested that the current distribution of MPD in endemic Malagasy legumes could be explained by a range of factors, such as the role of watersheds and dispersal corridors during past climatic changes, as well as by the evolutionary history of the group's most important dispersers, viz. extant and extinct lemurs. They conclude by advocating that a sound conservation plan should incorporate, in addition to the traditional biodiversity measures (species richness, PD and MPD), a detailed investigation of the biotic and abiotic factor that play (or have played) a role in the dynamics of each ecosystem.

The trends observed in the PD of Sarcolaenaceae differ significantly from those observed in Malagasy legumes by Buerki et al. (2015), where high values of species richness and PD are found in the subhumid highlands and lower values in humid eastern forests. However, the Bongolava Forest Corridor (Fig. 4g) and Midongy du Sud (Fig. 4c) are two sites where MPD values are high for Sarcolaenaceae that were regarded by Wilmé et al. (2006) and Buerki et al. (2015) as low- and high-elevation watersheds, respectively, and considered by them to represent potential refugia during the Quaternary climatic shifts. The list of important areas for conserving Sarcolaenaceae would thus also include the Bongolava Forest Corridor, Midongy du Sud, along with Makira and Masoala in the northeast and the eco-geographically diverse Behara-Tranomaro-Andohahela-Tsotongambarika area in the southeast, which spans a sharp ecotone from humid forest in the east to subarid tundra in the west (Fig. 4).

The recent expansion of Madagascar's network of protected areas has strengthened conservation in several areas that exhibit high levels of PD for Sarcolaenaceae, such as Makira, Pointe à Larrée, the Ankeniheny Zahamena and Fandriana Marolambo Forest Corridors, Ambalabe and Alan'Agalazaha (Fig. 4), all receiving legal protection within the last 5 years. Our results show that while Madagascar's present system of PAs was not designed to protect the phylogenetic diversity of Sarcolaenaceae, it nevertheless does a very good job of this, as indicated by the fact that 97.6 % of the total PD is included in cells that contain PAs. Furthermore, recent studies have shown that Sarcolaenaceae are part of a cohort of woody groups that are host to a diverse array of ectomycorrhizal fungi, which also includes members of two other endemic families, Asteropeiaceae and Sphaerosepalaceae, as well as the Malagasy species of the widespread tropical genus *Uapaca* (Phyllanthaceae), all of which are likewise endemic (Ducousso et al. 2008). Our results suggest that the overall distribution of Sarcolaenaceae (Fig. 1) might be constrained by aridity. As the presence of ectomycorrhizal fungi has been documented in the family, the spatial distribution of Sarcolaenaceae might also be limited by the dispersal ability of the associated fungi. This ecological interaction should therefore be taken into account when seeking to conserve the full diversity of this plant family. Members of these groups often co-occur and form an important component of the local vegetation, which suggests that habitat loss in areas rich in Sarcolaenaceae may also impact members of these other groups. The integration of information on the phylogenetic diversity of Sarcolaenaceae into conservation planning could thus also lead to species protection in these associated groups.

## Conclusion

As indicated earlier, the type of analysis presented here requires a dated phylogeny based on sampling that is representative of the study group, as well as reliable data on the distribution of each species. For Sarcolaenaceae, our sampling comprised ca. 70 % of the total species diversity, with good representation from each of the 10 genera in the family. We were also able to access reliable distributional information from recent taxonomic revisions augmented by ongoing identifications made of subsequently collected material. Our study has shown the potential value of determining the spatial distribution of species richness, PD and MPD for understanding whether the current network of protected areas provides adequate conservation of these important biodiversity values and for identifying gaps in the existing network that should be targeted for the establishment of new PAs. The study presented here suggests that it may be worthwhile to expand this approach to other endemic Malagasy clades that contain a sufficient number of well-delimited species and are present in a range of eco-geographic zones. By carefully selecting study groups it should be possible to cover regions of Madagascar in which Sarcolaenaceae are poorly represented or absent and thereby generate results from a set of lineages that are collectively representative of the Malagasy flora as a whole. It would then be



possible to deliver another valuable source of information – phylogenetic diversity – to the set of criteria used to assess the value and effectiveness of Madagascar's existing network of PAs and to identify priority areas for the establishment of new parks and reserves.

**Acknowledgments** Anaëlle Soulebeau is funded by a doctoral grant from Doctoral School ED227, Ministère de la Recherche, at the Muséum National d'Histoire Naturelle (Paris, France). The laboratory work was supported by the ANR-funded project EVORANGE (ANR-09-PEXT-011) and agreement No. 2005/67 between Genoscope and the MNHN under the project "Macrophylogeny of life" directed by Guillaume Lecointre. We thank Nicolas Chazot and Benoit Lefeuvre for their help with the R script and GIS analysis, and Philippe Grandcolas and Sven Buerki for comments that helped to improve this manuscript.

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Agnarsson I, Kuntner M (2012) The generation of a biodiversity hotspot: biogeography and phylogeography of the western Indian Ocean islands. In: Anamthawat-Jonsson K (ed) Current topics in phylogenetics and phylogeography of terrestrial and aquatic systems. InTech Publishers, Rijeka, pp 33–82
- Allnutt TF, Ferrier S, Manion G et al (2008) A method for quantifying biodiversity loss and its application to a 50-year record of deforestation across Madagascar. *Conserv Lett* 1(4):173–181
- Atlas numérique du Système des Aires Protégées de Madagascar (2015). <http://atlas.rebioma.net/>. Accessed 30 Jan 2014
- Bortolussi N, Durand E, Blum et al (2006) apTreeshape: statistical analysis of phylogenetic tree shape. *Bioinformatics* 22(3):363–364
- Buerki S, Devey DS, Callmender MW et al (2013) Spatio-temporal history of the endemic genera of Madagascar. *Bot J Linn Soc* 171(2):304–329
- Buerki S, Callmender MW, Bachman S et al (2015) Incorporating evolutionary history into conservation planning in biodiversity hotspots. *Philos Trans R Soc Lond B Biol Sci* 370(1662):20140014
- Callmender MW, Schatz GE, Lowry II PP et al (2007) Identification of priority areas for plant conservation in Madagascar using Red List criteria: rare and threatened Pandanaceae indicate sites in need of protection. *Oryx* 41(02):168–176
- Callmender MW, Phillipson PB, Schatz GE et al (2011) The endemic and non-endemic vascular flora of Madagascar updated. *Plant Ecol Evol* 144:121–125
- Caramel F (2015) Bolabola, le bois qui saigne. De Madagascar à la Chine, le scandaleux trafic du bois de rose. [http://www.lemonde.fr/planete/article/2015/01/24/bolabola-le-bois-qui-saigne\\_4562855\\_3244.html](http://www.lemonde.fr/planete/article/2015/01/24/bolabola-le-bois-qui-saigne_4562855_3244.html). Accessed 16 Feb 2015
- Coetzee JA, Muller J (1984) The phytogeographic significance of some extinct Gondwana pollen types from the Tertiary of the southwestern Cape (South Africa). *Ann Mo Bot Gard* 71:1088–1099
- Cornet A (1974) Essai de cartographie bioclimatique à Madagascar. Notice Explicative de l'ORSTOM 55:1–38

- Dayanandan S, Ashton PS, Williams SM et al (1999) Phylogeny of the tropical tree family Dipterocarpaceae based on nucleotide sequences of the chloroplast *rbcL* gene. *Am J Bot* 86(8):1182–1190
- Drummond AJ, Suchard MA, Xie D et al (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29(8):1969–1973
- Ducousso M, Béna G, Bourgeois C et al (2004) The last common ancestor of Sarcolaenaceae and Asian dipterocarp trees was ectomycorrhizal before the India-Madagascar separation, about 88 million years ago. *Mol Ecol* 13(1):231–236
- Ducousso M, Ramanankierana H, Duponnois R et al (2008) Mycorrhizal status of native trees and shrubs from eastern Madagascar littoral forests with special emphasis on one new ectomycorrhizal endemic family, the Asteropeiaceae. *New Phytol* 178(2):233–238
- Evans M, Aubriot X, Hearn D et al (2014) Insights on the evolution of plant succulence from a remarkable radiation in Madagascar (*Euphorbia*). *Syst Biol* 63(5):698–711. doi:[10.1093/sysbio/syu035](https://doi.org/10.1093/sysbio/syu035)
- Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61(1):1–10
- Faramalala MH (1988) Étude de la Végétation de Madagascar à l'aide des Données spatiales. Doctoral Thesis, Univ. Paul Sabatier de Toulouse, 167 p. + map at 1:1,000,000
- Faramalala MH (1995) Formations Végétales et Domaine Forestier National de Madagascar. Conservation International (et al.), 1 map
- Gautier L, Chatelain C, Callmander MW et al (2012) Richness, similarity and specificity of Madagascar flora compared with sub-Saharan Africa. *Plant Ecol Evol* 145(1):55–64
- Goodman SM, Benstead J (eds) (2003) The natural history of Madagascar. University of Chicago Press, Chicago
- Goodman SM, Benstead JP (2005) Updated estimates of biotic diversity and endemism for Madagascar. *Oryx* 39:73–77
- Haevermans T, Schatz GE, Applequist W et al Molecular phylogenetics of Sarcolaenaceae (Malvales), Madagascar's largest endemic plant family. (in preparation)
- Harper GJ, Steininger MK, Tucker CJ et al (2007) Fifty years of deforestation and forest fragmentation in Madagascar. *Environ Conserv* 34(04):325–333
- Holt BG, Lessard JP, Borregaard MK et al (2013) An update of Wallace's zoogeographic regions of the world. *Science* 339(6115):74–78
- Hong-Wa C (2009) Endemic families of Madagascar. XII. Resurrection and taxonomic revision of the genera *Mediusella* (Cavaco) Hutchinson and *Xerochlamys* Baker (Sarcolaenaceae). *Adansonia* 31(2):311–339
- Huelsenbeck JP, Rannala B (2004) Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Syst Biol* 53(6):904–913
- Isambert B, Bergsten J, Monaghan MT et al (2011) Endemism and evolutionary history in conflict over Madagascar's freshwater conservation priorities. *Biol Conserv* 144(6):1902–1909
- Jöns N, Emmel B, Schenk V et al (2009) From orogenesis to passive margin—the cooling history of the Bemarivo Belt (N Madagascar), a multi-thermochronometer approach. *Gondwana Res* 16(1):72–81
- Kembel SW, Cowan PD, Helmus MR et al (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26:1463–1464
- Kremen C, Cameron A, Moilanen A et al (2008) Aligning conservation priorities across taxa in Madagascar, a biodiversity hotspot, with high-resolution planning tools. *Science* 320:222–226
- Leroy JF (1978) Composition, origin, and affinities of the Madagascan vascular flora. *Ann Mo Bot Gard* 65:535–589
- Lowry II PP, Rabehevitra D (2006) Endemic families of Madagascar. IX. A new littoral forest species of *Schizolaena* (Sarcolaenaceae). *Adansonia* 3(28):149–153
- Lowry II PP, Schatz GE, Leroy JF et al (1999) Endemic families of Madagascar. III. A synoptic revision of *Schizolaena* (Sarcolaenaceae). *Adansonia* 3(21):183–212

- Lowry II PP, Haevermans T, Labat JN et al (2000) Endemic families of Madagascar. V. A synoptic revision of *Eremolaena*, *Pentachlaena* and *Perrierodendron* (Sarcolaenaceae). *Adansonia* 3(22):11–31
- Lowry II PP, Schatz GE, Wolf AE (2002) Endemic families of Madagascar. VIII. A synoptic revision of *Xyloolaena* Baill. (Sarcolaenaceae). *Adansonia* 3(24):7–19
- Lowry II PP, Nusbaumer N, Randrianasolo A et al (2014) Endemic families of Madagascar. XIII. New, restricted range species of *Eremolaena* Baill. and *Schizolaena* Thouars (Sarcolaenaceae). *Candollea* 69:183–193
- Madagascar Catalogue (2015) Catalogue of the vascular plants of Madagascar. Missouri Botanical Garden, St. Louis. <http://www.efloras.org/madagascar>. Accessed 30 Jan 2015
- Magnuson-Ford K, Mooers A, Paquette SR et al (2010) Comparing strategies to preserve evolutionary diversity. *J Theor Biol* 266(1):107–116
- Mercier JL, Wilmé L (2013) The Eco-Geo-Clim model: explaining Madagascar's endemism. *Madag Conserv Dev* 8(2):63–68
- Mooers AO, Heard SB (1997) Inferring evolutionary process from phylogenetic tree shape. *Q Rev Biol* 72(1):31–54
- Moore W, Robertson JA (2014) Explosive adaptive radiation and extreme phenotypic diversity within ant-nest beetles. *Curr Biol* 24(20):2435–2439
- Myers N (2003) Biodiversity hotspots revisited. *Bioscience* 53(10):916–917
- Myers N, Mittermeier RA, Mittermeier CG et al (2000) Biodiversity hotspots for conservation priorities. *Nature* 403:853–858
- Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20(2):289–290
- Pearson RG, Raxworthy CJ (2009) The evolution of local endemism in Madagascar: watershed versus climatic gradient hypotheses evaluated by null biogeographic models. *Evolution* 63(4):959–967
- Population Reference Bureau (2015). <http://www.prb.org/>. Accessed 30 Jan 2015
- Rabehevitra D, Lowry II PP (2009) Endemic families of Madagascar. XI. A new critically endangered species of *Schizolaena* (Sarcolaenaceae) from tapia woodland in South-Central Madagascar. *Adansonia* 3(31):149–155
- Rakotoarivivo M, Blach-Overgaard A, Baker WJ et al (2013) Palaeo-precipitation is a major determinant of palm species richness patterns across Madagascar: a tropical biodiversity hotspot. *Proc R Soc Lond B* 280(1757):20123048
- Rakotoarisoa JE, Be B (2004) Breeding and foraging behaviour and habitat characteristics of the scaly ground-roller *Geobiastes squamigerus* in Madagascar. *Ostrich-J Afr Ornithol* 75(1–2):25–31
- Ramananjahary RH, Frasier CL, Lowry II PP et al (2010) Madagascar's endemic plant families species guide – Guide Des Espèces de Plantes Appartenant Aux Familles Endémiques de Madagascar. Botanical Garden Press, Missouri
- Rambaut A, Drummond AJ (2007) Tracer ver. 1.4. <http://beast.bio.ed.ac.uk/Tracer>. Accessed 30 Sept 2013
- Randrianasolo A, Miller JS (1994) *Sarcolaena isaloensis*, a new species of Sarcolaenaceae from Isalo, south-central Madagascar. *Novon* 4:290–292
- Randrianasolo A, Miller JS (1999) Taxonomic revision of the genus *Sarcolaena* (Sarcolaenaceae). *Ann Mo Bot Gard* 86:702–722
- Rodrigues ASL, Brooks T, Gaston KJ (2005) Integrating the phylogenetic diversity in the selection of priority areas for conservation: does it make a difference? In: Purvis A, Gittleman JL, Brooks T (eds) *Phylogeny and conservation*, vol 8, Conservation biology. Cambridge University Press, London, pp 101–119
- Schatz GE (1995) Malagasy/Indo-Malesian phytogeographic connections. In: Lourenço WR (ed) *Biogéographie de Madagascar*. Orstom, Paris, pp 73–83
- Schatz GE, Lowry II PP, Wolf AE (2000) Endemic families of Madagascar. VI. A synoptic revision of *Rhodolaena* (Sarcolaenaceae). *Adansonia* 3(22):239–252

- Schatz GE, Lowry II PP, Wolf AE (2001) Endemic families of Madagascar. VII. A synoptic revision of *Leptolaena* Thouars sensu stricto (Sarcolaenaceae). *Adansonia* 3(23):171–189
- Schettino A, Scotese CR (2005) Apparent polar wander paths for the major continents (200 Ma to the present day): a palaeomagnetic reference frame for global plate tectonic reconstructions. *Geophys J Int* 163(2):727–759
- Schuurman D, Lowry II PP (2009) The Madagascar rosewood massacre. *Madag Conserv Dev* 4(2):98–102
- Sechrest W, Brooks TM, da Fonseca GA et al (2002) Hotspots and the conservation of evolutionary history. *PNAS* 99(4):2067–2071
- Torsvik TH, Amundsen H, Hartz EH et al (2013) A Precambrian microcontinent in the Indian Ocean. *Nat Geosci* 6(3):223–227
- Vane-Wright RI, Humphries CJ, Williams PH (1991) What to protect? – systematics and the agony of choice. *Biol Cons* 55(3):235–254
- Vary LB, Gillen DL, Randrianjanahary M et al (2011) Dioecy, monoecy, and their ecological correlates in the littoral forest of Madagascar. *Biotropica* 43(5):582–590
- Waeber PO (2009) Madagascar – ‘down the river without a paddle’ or ‘turning the corner’? *Madag Conserv Dev* 4(2):72–74
- Warren BH, Strasberg D, Bruggemann JH et al (2010) Why does the biota of the Madagascar region have such a strong Asiatic flavour? *Cladistics* 26(5):526–538
- Wells NA (2003) Some hypotheses on the Mesozoic and Cenozoic paleoenvironmental history of Madagascar. In: Goodman SM, Benstead JP (eds) *Natural history of Madagascar*. The University of Chicago Press, Chicago, pp 16–34
- Wikström N, Savolainen V, Chase MW (2001) Evolution of the angiosperms: calibrating the family tree. *Proc R Soc B* 268(1482):2211–2220
- Wilmé L, Goodman SM, Ganzhorn JU (2006) Biogeographic evolution of Madagascar’s microendemic biota. *Science* 312:1063–1065. doi:10.1126/science.1122806
- World Bank (2015). <http://data.worldbank.org/country/madagascar>. Accessed 15 Jan 2015
- Yoder AD, Nowak MD (2006) Has vicariance or dispersal been the predominant biogeographic force in Madagascar? Only time will tell. *Annu Rev Ecol Evol Syst* 37:405–431

# The Future of Phylogenetic Systematics in Conservation Biology: Linking Biodiversity and Society

Roseli Pellens, Daniel P. Faith, and Philippe Grandcolas

*Given the rate at which sequence data in the public domain are accumulating, with initiatives to sequence the entire biota ... on the horizon, it seems likely that within a decade or two, phylogenetic data will cease to be the limiting factor: It could even be that an organism's place in the Tree of Life often will be one of the few things we know about it. Mace et al. (2003)*

**Abstract** Phylogenetic diversity has become invaluable for conservation biology in the last decades, reflecting its link to option values and to evolutionary potential. We argue that its use will continue to grow rapidly in the next decades because of the transformation of systematics with new molecular techniques and especially metagenomics. In a near future, phylogenetic diversity typically will be the very first result at hand, and the great challenge of biodiversity sciences will be to preserve its link with natural history and the remainder of biological knowledge through species vouchers and names. The phylogeny availability and the very wide sampling allowed will facilitate obtaining detailed biodiversity information at local scale and considering the transition across scales – a fundamental need well highlighted in international conservation guidelines, and historically so difficult to achieve. All this suggests that phylogenetic diversity might be at the center of more explicit identification of conservation priorities and options. For concluding, we explore an emerging local-to-regional-to-global challenge: the possibility of defining “planetary boundaries” for biodiversity on the basis of phylogenetic diversity.

**Keywords** Species molecular characterization • Metagenomics • Knowledge databases • Option values • Planetary boundaries

---

R. Pellens (✉) • P. Grandcolas

Institut de Systématique, Evolution, Biodiversité, ISYEB – UMR 7205 CNRS MNHN  
UPMC EPHE, Muséum National d'Histoire Naturelle, Sorbonne Universités,  
45 rue Buffon, CP 50, 75005 Paris, France  
e-mail: [pellens@mnhn.fr](mailto:pellens@mnhn.fr); [pg@mnhn.fr](mailto:pg@mnhn.fr)

D.P. Faith

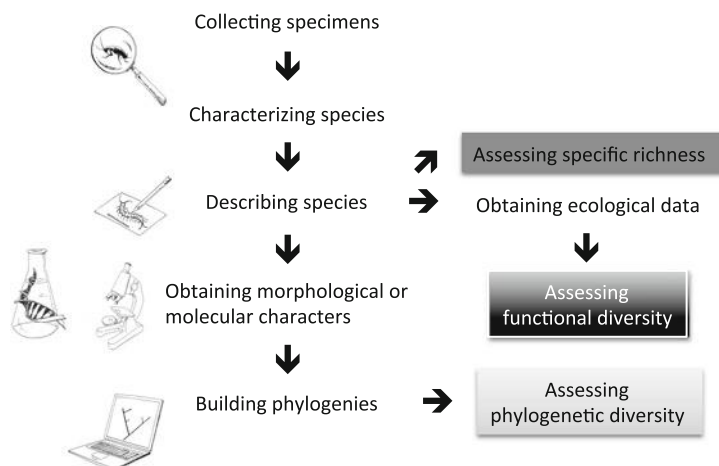
AMRI, The Australian Museum, Sydney, NSW 2010, Australia  
e-mail: [dan.fait@austmus.gov.au](mailto:dan.fait@austmus.gov.au)

Phylogenetic diversity is now a core part of conservation biology, reflecting its link to option values and to evolutionary potential. Further, there is good overlap with related issues in broader ecology. These include community ecology's interest in productivity (e.g. Cadotte et al. 2012), resilience (e.g. Pugliesi and Rapini 2015) and the functioning of ecosystems (e.g. Srivastava et al. 2012) and microbial ecology's use of PD as a cornerstone for exploring diversity patterns at multiple scales (Lozupone and Knight 2005, 2008; Faith et al. 2009). As the chapters in this book demonstrate, the development of new methods and their applications are very much tuned into human impacts and sustainability issues. Thus, red listings, drivers of extinction, and changes in spatial and temporal distribution of phylogenetic diversity are common elements of these studies. All this promotes the incorporation of phylogenetic diversity in the international conservation agenda.

These prospects are magnified by the remarkable facilities for obtaining entire or large parts of genomes or other molecular sequences of any kind of organisms, and by the sheer magnitude of biological (gene sequences, trait databases, species occurrences, red lists) and environmental data (climate layers for past, present and future interpolated to very fine spatial scales; land-use layers, spatial data indicating particular important risks such as fires, floods, and so on) now available in the public domain. All these allow for rapid estimation of the phylogenetic relationships for a large number of organisms in association with potential distribution and threats for species and lineages. In addition, under the stimulus of modern phylogenetic and molecular methods, systematics is going through a significant transformation that will certainly influence biodiversity conservation (Mace et al. 2003; Pons et al. 2006; Vogler and Monaghan 2006; Faith et al. 2010; Yahara et al. 2010). For closing this book, we will briefly describe this transformation of systematics and then discuss some impacts of these changes in biological conservation. We finish by exploring the possibility of defining “planetary boundaries” for biodiversity on the basis of phylogenetic diversity, and its important role in linking biodiversity into broader societal perspectives and needs.

## **In Phase with Modern Systematics and NGS Methods: The Tree First, Then the Species**

Conventionally, species are first characterized, then described with morphological or molecular data, and only then analyzed for building a phylogenetic tree (Fig. 1). As the entire operation demanded a long time and effort of specialists, the extent to which the later stage of the process – calculations of phylogenetic diversity – provided additional information “worth waiting for” was a recurrent and important question. Stopping at the first step and using species richness was accepted as a good *proxy* of biodiversity and sometimes justified, as when phylogenetic trees were expected to be balanced, or when the species with higher values of phylogenetic diversity were widespread, so not bringing important additional information (Rodrigues et al. 2005; Hartmann and André 2013). This rationale involved an

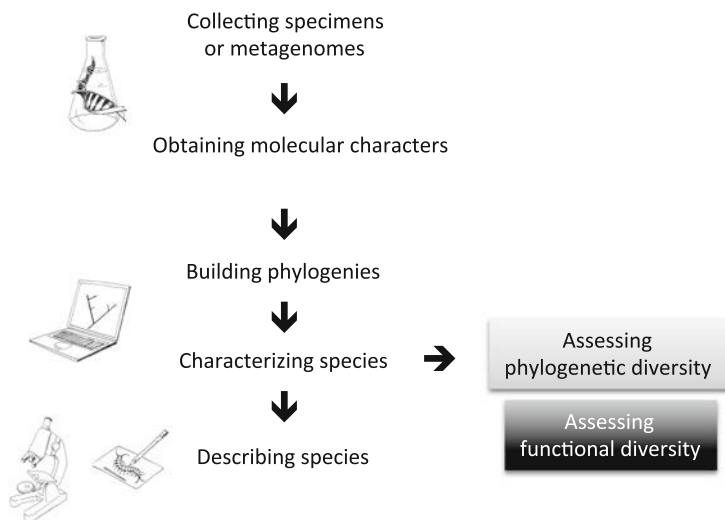


**Fig. 1** The traditional data processing in systematics, beginning with the sampling of specimens and the characterization and description of species. Specimens were then specifically sampled for phylogenetic characters, allowing to build phylogenetic trees and to compute phylogenetic diversity. In parallel, other biological knowledge was obtained separately. Note that in this framework, the number of species is obtained at the early step of species characterization (the so-called “morphospecies” may be obtained before description if necessary) before the phylogenetic analysis (Drawings by Agathe Haevermans)

unfortunate circular reasoning in that it required the knowledge about the phylogeny of the group to be able to discard it.

This process now has been turned upside down. In the new paradigm, systematics proceeds in an all-in-one operation i.e. the result of the data processing makes that the species position on the tree is part of species delimitation and characterization (Fig. 2). A global sample of characters (e.g. DNA)  $\times$  individuals can be partitioned into clusters – potentially species – through a tree-like guidance. The new rationale is simple: to define species, we need first to recognize and delimit different groups of individuals, by contrasting their characters (Goldstein et al. 2000; Pons et al. 2006; Vogler and Monaghan 2006; O’Meara 2010; Pante et al. 2015; but see DeSalle et al. 2005). This phylogenetic perspective is still certainly new for many, although it is inexorably implemented in the most recent molecular methods used for biodiversity exploration and characterization, such as molecular species delimitation or metagenomics.

Metagenomics recently went one step further by considering global amounts of DNA from environmental samples. In this approach there is no need to assemble the matrix ‘individuals  $\times$  characters’ that is already all in the test tube. This technique is also remarkable by capturing all DNA at the same time and carrying out a very wide sampling including microbes and all organisms usually ignored by traditional taxonomic screening (Tringe and Rubin 2005; Yahara et al. 2010). Combined with proteomics, it can even provide functional information at the same time, by obtaining both DNA species and protein synthesis. At the point we are now, systematics is therefore able to offer a comprehensive picture of diversity, linking species, their



**Fig. 2** The new and upcoming data processing in systematics, beginning with molecular characterization or even metagenomics, jointly allowing phylogenetic analysis and species characterization, and therefore computation of phylogenetic diversity. Note also that some assessments of phylogenetic diversity may proceed without the species characterization. The species description and name attribution is the last, but not the least, step to keep molecular data connected with biological knowledge. Note that in the same line, proteomics could provide – to some extent – the species functional characterization, with molecular analysis on its own (Drawings by Agathe Haevermans)

relationships and their characters to conservation biology (Funk et al. 2002; Wilson 2003; Faith et al. 2010; Lean and MacLaurin chapter “[The Value of Phylogenetic Diversity](#)”).

Obviously, this new framework enhanced by molecular biology and metagenomics will maintain biological significance and usefulness as long as molecular proxies will remain related to species concepts, taxon names and classifications linking to the wider biological knowledge (Mace 2004; German National Academy of Sciences Leopoldina 2014). The peril to invest only in an isolated and blind molecular database was already keenly emphasized by many taxonomists at the occasion of the rise of the barcoding initiative (e.g., Will et al. 2005). Building the network between names, biological knowledge and molecular data is from far the biggest challenge of present-day systematics and other sciences of diversity, much beyond the molecular technical tour de force (Grandcolas et al. 2013). We must keep in mind that this challenge takes place in a difficult moment when discovery rates of species new to science do not decline (Tancoigne and Dubois 2013) but in a context of rising rates of extinction (Régnier et al. 2015).



## Impacts on Biodiversity Conservation

The first impact of these methodological breakthroughs in biodiversity conservation is the growing availability of phylogenies with adequate taxon and character sampling at fine scale. As a consequence, it will increase not only the possibility of identifying taxa and areas whose conservation will maximize phylogenetic diversity (Forest et al. 2007; Buerki et al. 2015; Soulebeau et al. chapter “[Conservation of Phylogenetic Diversity in Madagascar’s Largest Endemic Plant Family, Sarcolaenaceae](#)”) or whose loss would contribute to major losses of our evolutionary heritage (Faith and Richards 2012; Faith 2015). It will also facilitate the transition across scales, a fundamental need well highlighted in international conservation guidelines (e.g. the Convention on Biological Diversity, “CBD”), and historically so difficult to achieve. For example, some targets can be established at a global scale based on a general phylogeny (for example, a phylogeny with samples of all genera or families), and a more detailed phylogeny with the regional diversification of the group (including for example a large sample of the species occurring in this region) will allow for establishing the areas to be protected for attainment of the broader target. This, associated with modern methods of Systematic Conservation Planning (Moilanen and Arponen 2011; Kukkala and Moilanen 2013; Faith chapter “[Using Phylogenetic Dissimilarities Among Sites for Biodiversity Assessments and Conservation](#)”) in which biological variables, including phylogenetic diversity can be considered along with costs, risks and return to investment, will certainly contribute to more explicit identification of conservation priorities and options (Pollock et al. 2015; Arponen and Zupan chapter “[Representing Hotspots of Evolutionary History in Systematic Conservation Planning for European Mammals](#)”; Silvano et al. chapter “[Priorities for Conservation of the Evolutionary History of Amphibians in the Cerrado](#)”). With these developments in mind, we will close this book by exploring an emerging local-to-regional-to-global challenge: the possibility of defining “planetary boundaries” for biodiversity on the basis of phylogenetic diversity.

## Phylogenetic Diversity as a Basis for Defining “Planetary Boundaries” for Biodiversity

The idea that we are approaching a state of shift in the planet’s environment, due to various human activities within the “Anthropocene” (Barnosky et al. 2012) is attracting attention in the scientific community. The definition and quantification of “planetary boundaries” is one approach to respond to this. “Planetary boundaries” (see Rockström et al. 2009; Steffen et al. 2015) refer to the idea of a “safe operating space” for humanity. The planetary boundaries framework considers processes relating to climate change, biodiversity loss, land-system change, biogeochemical flows, stratospheric ozone depletion, ocean acidification, freshwater use,

atmospheric aerosol loading, and chemical pollution. The rationale is that exceeding the identified boundaries means that thresholds and undesirable changes threaten human well-being.

There has been much debate about how to define a meaningful boundary related to biodiversity. The current rate of extinctions and the corresponding biodiversity crisis suggest a possible focus on global extinction rates. However, recent work has focused more on phylogenetic and functional diversity (Faith et al. 2010; Mace et al. 2014; Steffen et al. 2015). These aspects may have a good regional-to-global scope, and appealing links to current and future well-being. These two key aspects for a biodiversity boundary are now being investigated through a global change international program called “Future Earth”. The PD calculus may provide ways to describe boundaries related to phylogenetic diversity “tipping points” (Faith et al. 2010). Such phylogenetic tipping points correspond to the irreversible loss of deep branches of the tree of life, following successive losses over time of descendent taxa. The tipping points, and corresponding boundaries, then link naturally to concerns about the loss of evolutionary or evosystem services, including option values (unanticipated future benefits for humans) and evolutionary potential. Such option values of biodiversity typically reflect global-scale benefits for future generations, and so they are a natural consideration for planetary boundaries. At the same time, phylogenetic diversity has local importance (e.g. for resilience and delivery of evosystem services) and may be part of regional-scale planning. Early warnings with respect to a phylogenetic planetary boundary may focus on the changing status of Phylogenetic Key Biodiversity Areas – those places on the planet that are outstanding in their current contribution to retaining global phylogenetic diversity (Brooks et al. 2015; Faith chapter “[The Value of Phylogenetic Diversity](#)”).

The interest in Planetary Boundaries also reminds us that there are “boundaries” in the utility of phylogeny for conservation. The PD measure (Faith 1992; Faith chapter “[The Value of Phylogenetic Diversity](#)”) is useful, but does not tell us all we need to know about functional traits – one of the other possible foci for a biodiversity boundary. Functional traits, by their nature, are not always well accounted for by the PD assumption that shared ancestry explains shared features. This assumption could be especially hard to justify if these traits are defined too intrinsically and are therefore not heritable (Grandcolas et al. 2010; Weiher et al. 2011). Therefore, an alternative model assuming that shared *habitat* explains shared traits may be useful. Such companion models to phylogenetic diversity are in development (Faith chapter “[Using Phylogenetic Dissimilarities Among Sites for Biodiversity Assessments and Conservation](#)”). At the global scale, such approaches could provide, for multiple taxonomic groups, a running report card on risk of loss of functional trait diversity. This would nicely complement the emerging use of a PD report card to assess risks associated with resilience-loss, tipping points and planetary boundaries.

These issues highlight the broader need to integrate phylogenetic diversity – and its associated option values – into the broader perspectives on sustainability and multiple needs of society. This book demonstrates that effective development of the

measures – the toolbox – enables phylogenetic diversity to be “on the table” in these policy contexts.

**Acknowledgements** We thank Agathe Haevermans for kindly providing the drawings for the two figures of this chapter.

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Barnosky AD, Hadly EA, Bascompte J et al (2012) Approaching a state shift in Earth's biosphere. *Nature* 486(7401):52–58. doi:[10.1038/Nature11018](https://doi.org/10.1038/Nature11018)
- Brooks TM, Cuttelod A, Faith DP et al (2015) Why and how might genetic and phylogenetic diversity be reflected in the identification of key biodiversity areas? *Phil Trans R Soc Lond B* 370(1662). Unsp 20140019 doi:[10.1098/Rstb.2014.0019](https://doi.org/10.1098/Rstb.2014.0019)
- Buerki S, Callmander MW, Bachman S et al (2015) Incorporating evolutionary history into conservation planning in biodiversity hotspots. *Phil Trans R Soc B* 370(1662). Artn 20140014 doi:[10.1098/Rstb.2014.0014](https://doi.org/10.1098/Rstb.2014.0014)
- Cadotte MW, Dinnage R, Tilman D (2012) Phylogenetic diversity promotes ecosystem stability. *Ecology* 93(8):S223–S233
- DeSalle R, Egan MG, Siddall M (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Phil Trans R Soc Lond B* 360(1462):1905–1916
- Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61:1–10
- Faith DP (2015) Phylogenetic diversity and extinction: avoiding tipping points and worst-case losses from the tree of life. *Phil Trans R Soc Lond B* 370:20140011. doi:[10.1098/rstb.2014.0011](https://doi.org/10.1098/rstb.2014.0011)
- Faith DP, Lozupone C, Nipperess D, Knight R (2009) The cladistic basis for the phylogenetic diversity (PD) measure links evolutionary features to environmental gradients and supports broad applications of microbial ecology's “phylogenetic beta diversity” framework. *Int J Mol Sci* 10(11):4723–4741
- Faith DP, Magallón S, Hendry AP et al (2010) Ecosystem services: an evolutionary perspective on the links between biodiversity and human well-being. *Curr Opin Environ Sustain* 2:66–74
- Faith DP, Richards Z (2012) Climate change impacts on the tree of life: changes in phylogenetic diversity illustrated for *Acropora* corals. *Biology* 1:906–932
- Forest F, Grenyer R, Rouget M et al (2007) Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature* 455:757–760
- Funk VA, Sakai AK, Richardson K (2002) Biodiversity: the interface between systematics and conservation. *Syst Biol* 51(2):235–237
- German National Academy of Sciences Leopoldina (2014) Challenges and opportunities of integrative taxonomy for research and society – taxonomic research in the era of OMICS technologies. Deutsche Akademie der Naturforscher Leopoldina e.V. Nationale Akademie der Wissenschaften (German National Academy of Sciences), Halle, 54 pages
- Goldstein PZ, Desalle R, Amato G, Vogler AP (2000) Conservation genetics at the species boundary. *Conserv Biol* 14(1):120–131. doi:[10.1046/j.1523-1739.2000.98122.x](https://doi.org/10.1046/j.1523-1739.2000.98122.x)
- Grandcolas P, Nattier R, Legendre F, Pellens R (2010) Mapping extrinsic traits such as extinction risks or modelled bioclimatic niches on phylogenies: does it make sense at all? *Cladistics* 26:1–5
- Grandcolas P, Daubin V, Chave J et al (2013) Systématique, Phylogénie. In: Thiébaud S, Hadi H (eds) Prospective de l'Institut Ecologie Environnement du CNRS. *Compte-Rendu des Journées*

- des 24 et 25 Octobre 2012, Avignon. Les Cahiers Prospectives. Hors série. CNRS, Paris, pp 75–78 [<http://www.cnrs.fr/inee/outils/docs/prospectiveavignon2012.pdf>]
- Hartmann K, André J (2013) Should evolutionary history guide conservation? *Biodivers Conserv* 22(2):449–458. doi:[10.1007/S10531-012-0422-Z](https://doi.org/10.1007/S10531-012-0422-Z)
- Kukkala AS, Moilanen A (2013) Core concepts of spatial prioritisation in systematic conservation planning. *Biol Rev* 88(2):443–464. doi:[10.1111/Brv.12008](https://doi.org/10.1111/Brv.12008)
- Lozupone C, Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* 71(12):8228–8235
- Lozupone C, Knight R (2008) Species divergence and the measurement of microbial diversity. *FEMS Microbiol Rev* 32(4):557–578
- Mace GM (2004) The role of taxonomy in species conservation. *Phil Trans R Soc Lond B* 359:711–719
- Mace GM, Gittleman JL, Purvis A (2003) Preserving the tree of life. *Science* 300:1707–1709
- Mace GM, Reyers B, Alkemade R et al (2014) Approaches to defining a planetary boundary for biodiversity. *Glob Environ Chang* 28:289–297. doi:[10.1016/J.Gloenvcha.2014.07.009](https://doi.org/10.1016/J.Gloenvcha.2014.07.009)
- Moilanen A, Arponen A (2011) Setting conservation targets under budgetary constraints. *Biol Conserv* 144(1):650–653. doi:[10.1016/J.Biocon.2010.09.006](https://doi.org/10.1016/J.Biocon.2010.09.006)
- O'Meara BC (2010) New heuristic methods for joint species delimitation and species tree inference. *Syst Biol* 59:59–73
- Pante E, Schoelincx C, Puillandre N (2015) From integrative taxonomy to species description: one step beyond. *Syst Biol* 64(1):152–160. doi:[10.1093/Sysbio/Syu083](https://doi.org/10.1093/Sysbio/Syu083)
- Pollock LJ, Rosauer DF, Thornhill AH et al (2015) Phylogenetic diversity meets conservation policy: small areas are key to preserving eucalypt lineages. *Phil Trans R Soc B* 370(1662). Unsp 20140007 doi:[10.1098/Rstb.2014.0007](https://doi.org/10.1098/Rstb.2014.0007)
- Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, Kamoun S, Sumlin WD, Vogler AP (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst Biol* 55(4):595–609. doi:[10.1080/10635150600852011](https://doi.org/10.1080/10635150600852011)
- Pugliesi L, Rapini A (2015) Tropical refuges with exceptionally high phylogenetic diversity reveal contrasting phylogenetic structures. *Int J Biodivers*. <http://dx.doi.org/10.1155/2015/758019>
- Régner C, Achaz G, Lambert A, Cowie RH, Bouchet P, Fontaine B (2015) Mass extinction in poorly known taxa. *Proc Natl Acad Sci USA* 112(25):7761–7766. doi:[10.1073/pnas.1502350112](https://doi.org/10.1073/pnas.1502350112)
- Rockström J, Steffen W, Noone K, Persson A, Chapin FS, Lambin EF, Lenton TM, Scheffer M, Folke C, Schellnhuber HJ, Nykvist B, de Wit CA, Hughes T, van der Leeuw S, Rodhe H, Sorlin S, Snyder PK, Costanza R, Svedin U, Falkenmark M, Karlberg L, Corell RW, Fabry VJ, Hansen J, Walker B, Liverman D, Richardson K, Crutzen P, Foley JA (2009) A safe operating space for humanity. *Nature* 461(7263):472–475. doi:[10.1038/461472a](https://doi.org/10.1038/461472a)
- Rodrigues AS, Brooks TM, Gaston KJ (2005) Integrating phylogenetic diversity in the selection of priority areas for conservation: does it make a difference? In: Purvis A, Gittleman JL, Brooks T (eds) *Phylogeny and conservation*. Cambridge University Press, Cambridge, pp 101–119
- Srivastava DS, Cadotte MW, MacDonald AAM, Marushia RG, Mirotnick N (2012) Phylogenetic diversity and the functioning of ecosystems. *Ecol Lett* 15(7):637–648. doi:[10.1111/J.1461-0248.2012.01795.X](https://doi.org/10.1111/J.1461-0248.2012.01795.X)
- Steffen W, Richardson K, Rockström J et al (2015) Planetary boundaries: guiding human development on a changing planet. *Science* 347(6223):736. doi:[10.1126/science.1259855](https://doi.org/10.1126/science.1259855)
- Tancoigne E, Dubois A (2013) Taxonomy: no decline, but inertia. *Cladistics* 29(5):567–570. doi:[10.1111/cla.12019](https://doi.org/10.1111/cla.12019)
- Tringe SG, Rubin EM (2005) Metagenomics: DNA sequencing of environmental samples. *Nat Rev Genet* 6:805–814
- Vogler AP, Monaghan MT (2006) Recent advances in DNA taxonomy. *J Zool Syst Evol Res* 45(1):1–10. doi:[10.1111/j.1439-0469.2006.00384.x](https://doi.org/10.1111/j.1439-0469.2006.00384.x)
- Weier E, Freund D, Bunton T et al (2011) Advances, challenges and a developing synthesis of ecological community assembly theory. *Phil Trans R Soc B* 366:2403–2413

- Will KW, Mishler BD, Wheeler QD (2005) The perils of DNA barcoding and the need for integrative taxonomy. *Syst Biol* 54(5):844–851. doi:[10.1080/10635150500354878](https://doi.org/10.1080/10635150500354878)
- Wilson EO (2003) The encyclopedia of life. *Trends Ecol Evol* 18(2):77–80. doi:[10.1016/S0169-5347\(02\)00040-X](https://doi.org/10.1016/S0169-5347(02)00040-X)
- Yahara T, Donoghue M, Zardoya R et al (2010) Genetic diversity assessments in the century of genome science. *Curr Opin Environ Sustain* 2:43–49, <http://dx.doi.org/10.1016/j.cosust.2010.03.008>

# Index

## A

Accumulation curves, 199  
 Aesthetic, 29, 31, 33, 269  
 Alpha-diversity, 146, 148, 156, 157, 164, 204, 269  
 Alpha-level phylogenetic diversity, 267, 268, 280, 281  
 Amazonia, 45, 229, 334, 349  
 Amphibian, 9, 59, 61, 62, 187, 258, 287–302, 334, 379  
 Anagenesis, 40, 41, 120  
 Ancestry, 7, 43, 86, 120, 122, 124, 320, 380  
 Andean foothills, 348  
 Andes, 336, 338, 342, 343, 346–348, 350  
 Anuran, 9, 289, 298, 300  
 Archaea, 90, 91  
 Areas, 3, 4, 7, 9, 11, 25, 42, 44–53, 61, 64, 104, 106–108, 110, 120–123, 125, 127–130, 133–136, 155, 178, 180, 182, 183, 185–187, 192, 210, 213, 219–226, 228, 229, 231, 233, 240–242, 257–259, 266–275, 277–282, 288, 301, 306–310, 312–316, 320–323, 327–329, 334, 336, 342, 343, 346–348, 350, 356–360, 362, 364, 366–371

## B

Bacteria, 91, 92  
 Basal taxa, 103, 110  
 Benefit, 29, 32–34, 40, 67, 93, 111, 257, 266, 270, 276, 316  
 Beta-diversity, 8, 121, 157, 158, 160–161, 163, 164, 198–200, 204, 205, 209–212, 214, 256, 357

## Big five, 2

Biodiversity, 2–8, 11, 20–26, 28–35, 39–44, 48, 52, 53, 58, 59, 64, 82, 89, 99, 100, 120, 122–125, 129–131, 134–137, 142, 144, 145, 174, 175, 177–182, 184, 188, 198, 213, 214, 219, 220, 238–240, 258, 266, 267, 269, 270, 289, 292, 307, 320, 322, 327, 328, 334–336, 346, 356, 357, 359, 362, 366, 369, 370, 376, 377, 379, 380  
 decline, 2, 41, 46, 47, 59, 60, 63, 65, 67, 90, 168, 288, 306, 316, 320, 359, 378  
 Biological diversity, convention, 3, 23, 24, 29, 31, 306, 316, 379  
 Branch length, 5, 6, 27, 40, 49, 51, 68, 71–73, 83, 84, 108, 120, 143, 147, 149–151, 153–159, 162, 165, 169, 174, 177, 198, 201–203, 205, 206, 209, 213, 214, 220, 224, 269, 308, 320, 338, 343, 362  
 Brazil, 9, 288, 290, 291, 293, 299, 301, 309, 313, 316, 388  
 Brownian motion, 70–73  
 Budget, 8, 179, 182, 183, 185–187, 191, 192, 270

## C

Carnivora, 71, 307  
 CBD. *See* Convention on Biological Diversity (CBD)  
 Central America, 334, 336, 343, 346, 347, 349  
 Cerrado, 9, 187, 258, 289–294, 298–301, 338, 343, 346, 349, 379  
 Cetacea, 307  
 Cetaceans, 313, 316

**Character**

- archaic, 102, 103
- molecular, 6, 11, 26, 82–84, 86, 87, 100–102, 104, 126, 238, 242, 338, 377–379
- morphological, 6, 26, 27, 33, 34, 70, 72, 87, 153, 242, 293, 299, 361, 376
- original, 3
- primitive, 103

Cladogenesis, 40, 41, 84, 120, 361

Climate change, 2, 58, 62, 64, 65, 68, 301, 306, 336, 379

Community, 2, 7, 21, 26, 48, 92, 101, 109, 122, 126, 131, 132, 142, 148, 150, 188, 198, 208, 213, 214, 220, 266, 267, 270, 273, 278, 280, 282, 288, 322, 334, 336, 343, 348–349, 376, 379

Comparison of phylogenetic measures, 4

Competition, 104, 109

Complementarity, 44–52

Concept, 2–4, 7, 8, 21–23, 25, 39, 42, 48, 64, 101, 103, 120, 174, 175, 183, 188, 225, 257, 321, 378

Confidence, 89, 214, 222, 229

Congruence, 61, 338, 347, 356, 357, 364, 368

Consequentialism, 30, 32, 33, 41, 47, 110, 130, 131, 146, 238, 281, 320, 329, 334, 379

**Conservation**

- goal, 20, 21, 23, 28, 40, 41, 62, 66, 144, 266, 288, 289, 293–295, 299, 301, 306
- priority areas, 9, 10, 288, 290, 292, 334
- projects, 21
- target, 10, 71, 293
- triage, 179

Continent, 279, 329, 350

Convention on Biological Diversity (CBD), 3, 23, 306, 316, 379

Cost, 4, 32, 33, 67, 124, 179, 182, 185, 186, 222, 259, 266, 268, 270–273, 278–281, 288, 292–294, 301, 336, 379

Country, 9, 131, 180, 182, 186, 268, 271, 277, 280, 356

Criterion, 41, 102, 104, 106, 108, 125, 129, 132, 134, 222, 277, 321, 357

**D**

Data deficient, 10, 306, 308, 314, 325, 327

Delete, 223–225, 227, 228, 230–233

Dispersion, 8, 28, 34, 199, 200, 205, 209–214, 258

Dissimilarities, 27, 43, 44, 121–127, 129, 130, 132, 134–137, 269, 379, 380

**Diversity**

abundance-based, 8, 142–145, 148, 155, 158

alpha, 142, 146

beta (*see* Beta-diversity)

decomposition, 8, 141, 144, 145

gamma, 146, 148, 156, 157

species, 5

DNA, 83, 85, 90, 91, 101, 146, 361, 377

Doubling property, 143, 148, 269

**E**

Ecological diversity, 82, 89, 357, 361

ED (Environmental Diversity), 123–126, 134, 137

ED (Evolutionary Distinctness or Evolutionary Distinctiveness), 4–6, 8, 10, 39, 45, 48, 99, 219, 220, 239, 240, 242, 244, 256, 258, 267, 288, 289, 292, 294, 295, 306–308, 310, 312, 357

EDGE, 10, 48, 51, 52, 66, 67, 258, 267, 268, 307–309, 311, 312, 320–322, 325, 329

Edge of existence, 28, 66

Edge = split, 343, 344

Endemic, 2, 8–11, 89, 91, 103, 180, 221, 225, 226, 240, 289, 292–294, 298–301, 314, 328, 355–371

Endemism, 43–46, 50, 51, 108, 188, 198, 227, 229, 239, 240, 289, 290, 292, 294, 334, 335, 356, 358

Entropy, 142–149, 151, 158, 159, 161, 198, 208, 268–273, 278, 280, 282

Epistemic, 31

Equal splits (ES), 43, 44

Essential biodiversity variables, 2, 21

Ethic, 2, 24

Eukarya, 90

Europe, 9, 268, 271, 273, 281, 347, 350

Evenness, 8, 25, 26, 199, 200, 204, 205, 208, 209, 212, 214, 258

Evolutionary distinctness, 4–6, 8, 10, 39, 45, 48, 99, 220, 239, 240, 242, 244, 256, 258, 267, 288, 292, 320, 325, 347, 357

Evolutionary heritage, 4, 5, 66, 379

Evolutionary history, 3, 67–72, 266

Evolutionary model, 6, 44, 67, 70–72, 120

Evolutionary potential, 7, 28, 82, 83, 89, 93, 104, 105, 110, 256, 376, 380

Evolutionary time, 149, 343

Evolutionary uniqueness, 219, 309

Ecosystem services, 4, 28, 380

Extinction, 6, 59–72, 102, 109–110, 188, 288, 308, 320, 357

- risk, 7, 10, 41, 59, 61–69, 71, 72, 109, 110, 188, 267, 306–308, 310, 314, 320, 327, 328
- F**
- Feature diversity, 3, 5, 28, 34, 35, 39–53, 67, 70–72, 120, 122, 134, 202, 258, 267, 289, 362
- Fossil records, 59, 68, 83, 85, 101–103, 106, 107
- Function(al)
  - diversity, 4, 21, 26, 28, 33, 46, 49, 59, 144, 153, 267, 281, 301, 350, 380
  - traits, 26, 143, 350, 380
- G**
- Gap
  - analysis, 108, 258, 289, 290, 293
  - phylogenetic diversity, 45, 53, 123, 124
- Generalized entropy, 145, 147–149
- Geography, 67, 358
- Gini-Simpson entropy, 208
- Gini-Simpson index, 142–148, 204, 208
- Global, 4, 7, 10, 24, 28, 33, 42, 46, 51–53, 58, 61, 64, 70, 89, 91, 100, 102, 103, 105, 109–111, 137, 180, 181, 213, 239, 258, 267, 268, 270, 282, 288, 305–316, 320, 322, 323, 327, 330, 346, 362, 377, 379, 380
- Global Biodiversity Information Facility (GBIF), 328
- Glossary, 3, 5
- Greedy algorithm, 182, 183, 187, 188
- Group on Earth Observation Biodiversity Observation Network (GEO BON), 137
- Guadalcanal, 10, 325, 326
- H**
- Habitat, 2, 7, 10, 21, 32, 58, 62, 64, 67, 68, 84, 89–91, 109, 111, 122–124, 130, 136, 180, 185, 186, 208–210, 268–270, 289, 292, 300, 301, 306, 310, 316, 320, 322, 328, 329, 334–336, 348, 356, 359, 360, 370, 380
- Haplotypes, 40, 41, 104, 198
- Heightened EDGE (HEDGE), 10, 258, 308, 309, 311, 312
- Higher taxa, 42, 198
- Hill number, phylogenetic, 8, 44, 142–158, 164, 166, 169, 170
- Hotspots
  - of evolutionary history, 8, 9, 187, 237–259, 265–282, 288, 347, 350, 368, 379
  - of phylogenetic diversity, 8, 9, 237–259, 346–348
  - of species richness, 346–348
- Hydrosheds, 290
- I**
- Imperfect sampling, 238
- Integer programming (IP), 182–189
- Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES), 41
- International Union for Nature Conservation (IUCN), 2, 10, 40, 51, 58, 61, 62, 71, 220, 268, 290, 306–309, 311, 313–315, 320–325, 327, 328
- Intrinsic value, 29, 30, 41
- $I_s$ , 7, 93, 100–108, 220, 226, 229, 256, 273, 278
- Island biogeography, 60, 322
- Islands, 9–11, 60, 61, 86, 226, 239–241, 309, 313, 314, 322–325, 327–330, 356–360, 362, 364
- Isle of Pines, 10, 325, 326
- Isolation, 10, 86, 107, 211, 239, 322, 327, 328, 357
- Ithomia, 336, 338–341, 343, 344, 346–349
- Ithomiini, 335
- J**
- Jack-knife, 222, 226, 234
- Jamaica, 325, 326
- Justification, 6, 28–30, 33, 44, 67, 120, 126, 281
- K**
- Kakapo, 6, 7, 85, 86, 93
- Key biodiversity areas (KBA), 52, 53, 380
- Kōtare, 85–87
- L**
- Landscape transformation, 2
- Least concern, 10, 51, 307, 308, 314, 325, 327
- Lineage
  - common, 151
  - dominant, 151
- Living ancestor, 103, 104, 110



Local, 10, 22, 64, 83, 101, 107, 123, 156,  
159, 160, 162, 164, 169, 267, 270, 273,  
278, 280–282, 306, 316, 348–350, 370,  
379, 380

Long branches, 48, 50, 82–87, 89, 101, 102,  
108, 109, 135

## M

Macrobial, 42

Madagascar, 10, 11, 325, 326, 328, 356–360,  
362–370, 379

Mammals

aquatic, 10, 258, 267, 305–316

marine, 10, 315

Manatees

Marxan, 187, 293, 294

Mean base change, 153

Mean branch length, 153–155, 158, 159,  
162, 362

Mean phylogenetic diversity (MPD), 149,  
151, 280, 343–346, 360, 362, 364,  
365, 369, 370

Measure, 2, 20, 40, 58, 89, 104, 120, 141–170,  
174, 177, 198, 220, 256, 267, 288, 307,  
322, 335, 357, 380

Measurement, 21, 23–25, 27, 29, 34, 110,  
208, 214

Megadiverse, 9, 350

Meta-analysis, 240

Metagenomics, 101, 377, 378

Metaphor, 3, 110, 111

Metapopulation, 10, 267, 319–330

Metapopulation capacity, 10, 321–322

Microbial, 7, 42, 82, 91–93, 123, 126,  
137, 376

Microbiomes, 7, 82, 91

Mimicry, 10, 333–350

Missing link, 103, 110

Monitoring, 2, 7, 8, 137, 142, 256

Mullerian mimicry, 336, 348–349

Mutualistic interactions, 334, 336, 346, 349

## N

Napeogenes, 336, 339–341, 344, 346, 348, 349  
Neotropics, 10, 278, 288, 334, 336–338, 344,  
346–349, 368

Network, 8, 11, 28, 137, 173–192, 267, 268,  
272, 281, 288, 289, 293, 300, 306, 315,  
334, 350, 356, 357, 362, 366, 368, 370,  
371, 378

Network of protected areas, 267, 281, 289,  
293, 300, 356, 357, 366, 368, 370

New Caledonia, 8, 103, 108, 211, 212,  
238–242, 257, 357

Nggela Sule, 10, 325, 326

Niche, 32, 60, 62, 67, 104, 109, 359

Niche conservatism, phylogenetic, 109

Noah's Ark problem (NAP), 188, 189

Non-random extinction, 6, 57–73, 109,  
188, 357

NP-hard, 179, 183, 189

## O

*Oleria*, 336, 339–344, 346–349

Option value, 4, 30–32, 40–43, 120, 144,  
376, 380

Originality, 27, 266, 267

## P

Parallel edge, 177

Parameters, 142, 146, 148, 152, 153, 163, 169,  
222, 271

Pattern, 10, 28, 40, 41, 43, 60, 61, 63, 68, 105,  
122, 151, 210, 214, 221, 223, 276, 277,  
281, 312, 333–350

Pattern-process model, 41, 43

PD-dissimilarity, 121, 122, 125, 126,  
134, 137

Philosophical, 29, 30, 33

Phyla, 7, 91, 92

Phylogenetic distances, 43, 144, 146, 147,  
150, 210, 343

Phylogenetic diversity (PD)  
decomposition, 8, 155, 158  
mean, 8

Phylogenetic diversity analyzer (PDA),  
187–189, 361, 362

Phylogenetic overlap, 45, 46, 51, 169

Phylogenetic redundancy, 109

Phylogenetic tipping points, 52, 380

Phylogenetic tree, 27, 34, 42, 46, 71, 84, 88,  
120–122, 288, 320, 362

Phylogenetic, turn over rate, 132, 163

Phylo-Horn, 159, 162, 163

Phylo-Jaccard, 160, 162, 163

Phylo-Morisita-Horn, 159, 161, 163

Phylo-regional-overlap, 160–163

PhyloSor, 159

Phylo-Sørensen, 159, 160, 163

Pinnipeds, 313

Planetary boundaries, 4, 11, 376,  
379, 380

Planning units (PUs), 290, 294, 298

Porifera, 91

- Prioritization, 5, 9, 10, 50, 222, 257, 258,  
266–272, 274, 277, 280–282, 289,  
292, 294, 299, 308, 314, 316, 321,  
322, 325, 336
- Priority, 4, 8, 10, 238, 257, 267, 292, 298, 307,  
314, 350
- Probability of Interspecific Encounter (PIE),  
204, 208
- Process, 3, 41, 43, 59, 65, 67, 70, 71, 89, 94,  
102, 179, 199, 203, 213, 219, 266, 275,  
290, 327, 361, 376, 377
- Protected area, marine, 10, 307, 310, 313, 316
- Protected areas, 9–11, 125, 129, 130, 257,  
266, 267, 269, 271, 277, 278, 281,  
288–291, 293–295, 300, 307, 309, 310,  
316, 356, 357, 359, 360, 362, 366–370
- Prune, 6, 57–74, 109, 188, 288, 357
- Punctuated evolution, 72
- Q**
- Quadratic entropy, 25, 144, 146, 147, 151,  
158, 159, 161, 208, 266, 268–271, 273,  
278, 280, 282
- R**
- Rao, 25, 144, 146, 147, 151, 208, 268, 269,  
271–274, 276–280, 282
- Rao's quadratic entropy, 144, 146, 147, 151,  
208, 268, 269, 271–273, 278, 280, 282
- Rarefaction, 8, 143, 199, 200, 209, 213, 238,  
258, 343, 368
- Rarefaction curves, 8, 199, 200, 204–206, 209,  
211–213
- Raup, 34, 59, 60, 63
- Red list, 109, 376
- Regional, 42, 46, 124, 159–163, 169, 213,  
282, 306, 348, 350, 379, 380
- Relict  
climatic, 100–102, 110  
geographic, 105  
phylogenetic, 100  
species, 100
- Replication principle, 143–148, 153, 154,  
157, 269
- Resampling, 8, 9, 243, 253, 257
- Reserve, 180–182, 185–187, 257
- Rhynchocephalia, 87–89
- Richness, 2, 3, 9, 10, 23, 25, 48, 58–61, 73,  
125, 128–132, 134, 142–147, 150, 151,  
155, 161, 165, 174, 183, 187, 198,  
200–202, 204, 208, 210, 211, 213, 214,  
219, 227, 240, 242, 244, 247, 248, 256,  
267, 270, 271, 273, 274, 276, 278, 281,  
289, 290, 292, 299–301, 306, 308–310,  
315, 325, 328, 329, 335, 342–344, 346,  
347, 349, 350, 356–360, 362, 364–366,  
368–370, 376
- Robustness, 144, 222
- Rooted PD, 202
- S**
- Sampling curves, 198, 199
- Sampling effort, 2, 8, 198–200, 206, 207, 238,  
240, 257, 258, 315
- Sarcolaenaceae, 11, 357, 358, 360–370, 379
- Scale, 5, 9, 10, 20, 22–24, 28, 29, 33–35, 60,  
61, 66, 87, 93, 102, 105, 174, 213, 239,  
259, 267, 268, 271, 272, 274–277,  
279–282, 288, 290, 321, 322, 327, 328,  
342, 349, 359, 379, 380
- Sensitivity analysis, 222
- Shannon entropy, 142–148, 158
- Single edge, 177
- Sirenia, 305
- Sixth mass extinction, 2, 58, 320
- Spatial fragmentation, 10, 267, 319–330
- Spatial prioritization, 9, 266–268, 270,  
277, 280
- Spatial scale, 2, 9, 276, 279, 282, 376
- Speciation, 59–60
- Species richness, 25, 148, 207, 208, 210, 212,  
241, 244, 293, 294, 308, 329, 335,  
342–344, 346–350, 357–360, 362,  
364–366, 368–370, 376
- Specional, model of evolution, 6, 71, 72
- Sphenodon*, 87, 89
- Split decomposition, 187
- Split diversity (SD), 8, 173–192
- Split network, 8, 173–191
- SplitsTree, 187
- Sponges, 7, 91, 92
- Stability, 9, 58, 222
- Standardization, 220, 229, 256, 258, 259
- State of shift, 379
- Subsampling, 199, 213
- Support, 8, 9, 30, 135, 144, 220, 222, 225,  
226, 229, 234, 238, 257, 270, 282, 301,  
321, 327, 335, 347
- Surrogate, 20, 23, 24, 28, 124, 125, 135, 202,  
335, 347, 350
- Systematic conservation planning, 4, 9, 134,  
187, 258, 265–282, 288, 290, 347, 350,  
368, 379
- System of protected areas, 9–11, 357, 359,  
360, 367

**T**

- Taxa, 5, 7, 40, 45, 52, 59, 61, 66, 70, 82, 84, 91, 93, 102–104, 108, 110, 120, 152, 198, 221, 226, 238, 240, 242, 259, 280–282, 308, 334–336, 347–349, 361, 379, 380
- Taxonomic distinctness, 27, 242
- Threat categories, 310
- Threatened, 2, 32, 46, 51–53, 58, 61–63, 65, 66, 70, 71, 239, 268, 282, 288, 300, 306–308, 315, 322, 325, 327, 328, 334, 359, 366
- Threatened species, 2, 32, 46, 51, 52, 61, 66, 70, 71, 220, 307, 323–325, 327
- Tipping points, 4, 52, 380
- Topology, 5, 27, 67, 70, 103, 108, 214, 220–225, 227, 230, 232, 242, 361
- Traits, 6, 7, 21, 22, 26, 33, 34, 40, 41, 45, 59, 62, 63, 65–68, 70–73, 83, 86, 89, 93, 103, 109, 122, 143, 208, 267, 321, 350, 376, 380
- Tree
  - balance, 60
  - of life, 3, 5–7, 57–73, 90, 92, 103, 109, 120, 175, 215, 238, 335, 380
  - non ultrametric, 150, 153–155, 158–160, 162
  - phylogenetic, 82
  - ultrametric, 70, 147, 149, 150, 205, 269, 338

- Tuatara, 6, 7, 87, 89, 93
- Turnover of branch lengths, 209

**U**

- Uncertainty, 29, 32–35, 145, 214, 226, 240, 268, 274, 280
- UniFrac, 160, 162
- Unrooted PD, 202

**V**

- Value
  - instrumental, 29, 31, 33, 35
  - intrinsic, 29–30, 41
  - option, 4, 31, 32, 40–43, 120, 144, 376, 380
- Variables, 2, 21, 22, 24, 25, 121, 126, 183, 184, 186, 206, 213, 220, 258
- Variance, 26, 70–72, 83, 200, 240, 244, 347
- Vulnerability index, 343, 349

**W**

- Wes, 220, 221, 225, 229, 256, 257
- Widespread, 220, 221, 225, 244, 257, 292, 328, 360, 370, 376
- Ws, 220, 229, 242–245, 247, 249–256, 258, 259

**Z**

- Zonation, 9, 187, 258, 268–281, 329