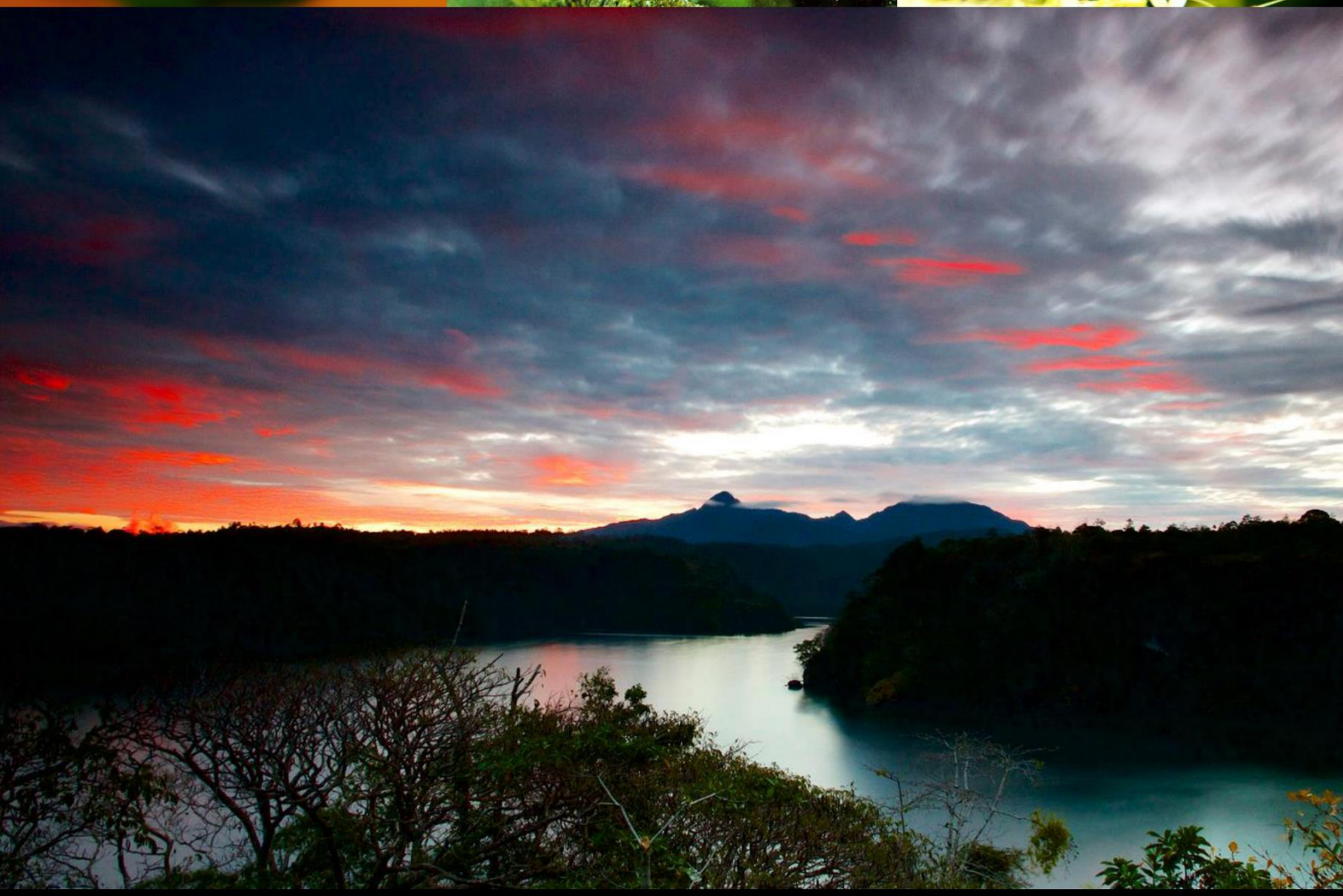
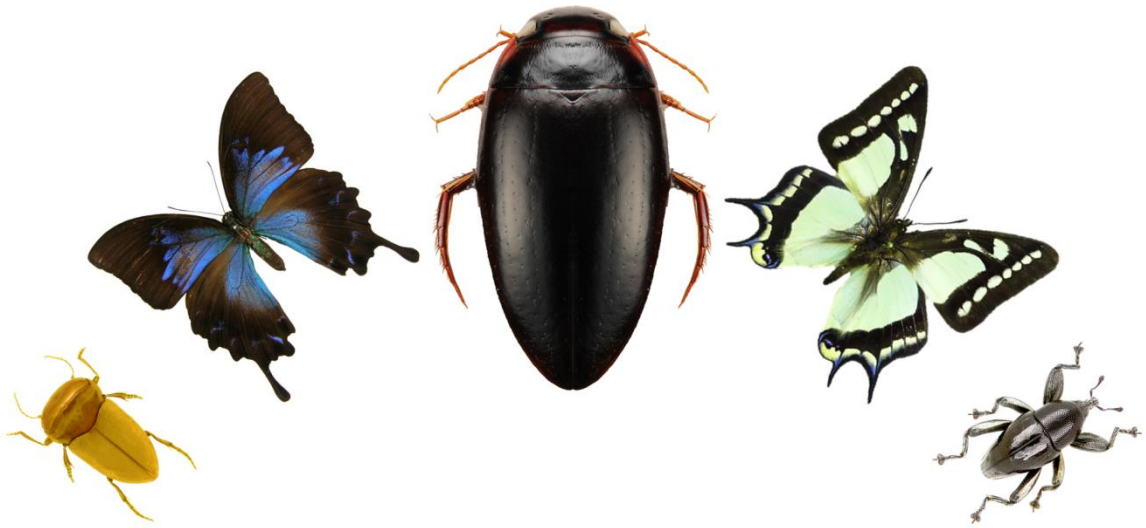


INSECTS

AS A MODEL TO PUZZLE OUT MECHANISMS OF LINEAGE DIVERSIFICATION IN THE INDOMALAYAN / AUSTRALASIAN ARCHIPELAGO

EMMANUEL F.A. TOUSSAINT



INSECTS AS A MODEL TO PUZZLE OUT MECHANISMS OF LINEAGE DIVERSIFICATION IN THE INDOMALAYAN / AUSTRALASIAN ARCHIPELAGO

DISSERTATION PRESENTED TO OBTAIN THE DEGREE OF DOCTOR AT THE FACULTY OF BIOLOGY
OF THE LUDWIG-MAXIMILIAN UNIVERSITY OF MUNICH

DISSERTATION ZUR ERLANGUNG DES DOKTORGRADES DER FAKULTÄT FÜR BIOLOGIE
DER LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

DISSERTATION PRESENTÉE EN VUE DE L'OBTENTION DU GRADE DE DOCTEUR DE LA FACULTÉ DE BIOLOGIE
DE L'UNIVERSITÉ LUDWIG MAXIMILIAN DE MUNICH

VORGELEGT VON / PRESENTED BY / PRESENTÉE PAR

EMMANUEL F.A. TOUSSAINT

GUTACHTER / REFEREES / RAPPORTEURS

Prof. Dr. Gerhard Haszprunar
Prof. Dr. Dirk Metzler
Prof. Dr. Susanne Renner
PD Dr. Michael Schrödl
Prof. Dr. Herwig Stibor
Prof. Dr. Gert Wörheide

TAG DER MÜNDLICHEN PRÜFUNG / DATE OF DEFENSE / DATE DE SOUTENANCE

12th December 2014

staatliche
naturwissenschaftliche
sammlungen bayerns

To my grandparents Paulette and Jean, who forged my passion for nature since my tender age and remain even today a source of inspiration…

Für meine Großeltern Paulette und Jean, die meine Leidenschaft für die Natur seit meinem jungen Alter geformt haben und auch heute eine Quelle der Inspiration sind...

A mes grands-parents Paulette et Jean, qui ont forgé ma passion pour la nature depuis l'âge tendre et qui demeurent une source d'inspiration encore aujourd'hui…

Declaration of Originality

I, Emmanuel François Adrien Toussaint, declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institute of tertiary education. To the best of my knowledge and belief, this Ph.D. dissertation contains no material previously published or written by another person except where due references are made.

Munich,

_____

Emmanuel F.A. Toussaint

Eidesstattliche Erklärung / Erklärung

Ich, Emmanuel François Adrien Toussaint, versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt ist.

Hiermit erkläre ich, dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist und dass ich mich anderweitig einer Doktorprüfung ohne Erfolg nicht unterzogen habe.

München, den

_____

Emmanuel F.A. Toussaint

Contents

List of publications and declaration of contribution 1

Publications included in the core of the dissertation 1
Summary of the Ph.D. candidate's contributions 2

Summaries 5

English abstract 3
German abstract (Zusammenfassung) 4
French abstract (Résumé) 5

GENERAL INTRODUCTION

Chapter 1. On the study of speciation 7

1.1 General background 8
1.2 Species discovery in the era of molecules 10
1.3 Speciation mechanisms and biogeography 15

Chapter 2. Phylogenetics as a tool to unveil patterns and processes of biodiversity 21

2.1 From characters to trees: A methodological overview 22
2.2 Enter clocks: a guide to dating trees 29
2.3 How trees help unfolding biogeographic and diversification patterns 34

Chapter 3. The Australasian archipelago 39

3.1 Glide over an intricate geological puzzle 40
3.2 From murky depths to sky islands: Diving beetles as an evolutionary model 43
3.3 On the study of tropical butterflies to unfold biogeographic scenarios 45

Chapter 4. Objectives of the thesis 47

PART 1: ON THE DISCOVERY OF NEW SPECIES IN THE ARCHIPELAGO

Chapter 5. A plea for an integrative taxonomy 50

5.1 Paper I - Suggestions to accelerate biodiversity assessments 51

Chapter 6. Integrative taxonomy at work 60

6.1 Paper II – Molecular species delimitation of charismatic tropical butterflies 61
6.2 Paper III- A new species of diving beetle from Timor 108
6.3 Paper IV - A new species of diving beetle from Biak Island 119
6.4 Paper V - A new genus of diving beetle from Australia 126

PART 2: ISLAND LINEAGE DIVERSIFICATION

Chapter 7. Biogeography on a continental-sized island: Australia 134

7.1 Paper VI - Diversification of Australian diving beetles in the Quaternary 135
7.2 Paper VII - Diversification of Australasian diving beetles in the Cenozoic 152

Chapter 8. Biogeography on a recent and geologically puzzling island: New Guinea ... 175

8.1 Paper VIII – Diversification of diving beetles in New Guinean highlands 176
8.2 Paper IX - Diversification of diving beetles during the New Guinean orogeny 196

PART 3: BIOGEOGRAPHIC INSIGHTS FOR THE ARCHIPELAGO

Chapter 9. On the role of biogeographic barriers 207

9.1 Paper X – Australasian weevil biogeography and the role of Wallace's line 208

Chapter 10. Unfolding historical biogeography of widespread clades 217

10.1 Paper XI - Fine-scale biogeography of widespread swallowtails 218

OUTCOMES OF THE THESIS 243

A truly integrative taxonomy within sight 244
Developments on insect diversification dynamics 246
Biogeography on the right tracks 249
Conclusion 250

References 251

Picture Credits 264

Acknowledgements 265

Curriculum Vitae 267

List of Publications and Declaration of Contribution

Publications included in the core of the dissertation

Paper I - Balke M, Hendrich L., **Toussaint EFA**, Zhou X, von Rintelen T, de Bruyn M (2013) Suggestions for a molecular biodiversity assessment of South East Asian freshwater invertebrates. Lessons from the megadiverse beetles (Coleoptera). *Journal of Limnology*, 72(2):61-68.

Paper II - **Toussaint EFA**, Morinière J, Muller CJ, Kunte K, Turlin B, Hausmann A, Balke M. An array of molecular species delimitation methods sheds light on species boundaries in the charismatic *Polyura* Nawab butterflies. *In review in Molecular Phylogenetics and Evolution.*

Paper III - Balke M**, Toussaint EFA**, Hendrich L, Hájek J (2013) A new species of the Australian genus *Necterosoma* from Timor (Coleoptera: Dytiscidae: Hydroporini). *Acta Entomologica Musei Nationalis Pragae*, 53(1):65-74.

Paper IV - Balke M, Warikar E, **Toussaint EFA**, Hendrich L (2013) *Papuadessus baueri* sp.nov. from Biak Island, Papua (Coleoptera: Dytiscidae: Hydroporinae). *Spixiana*, 36(2):283-288.

Paper V - Hendrich L, **Toussaint EFA**, Balke M (2014) A new genus of Hydroporini from south-western Australia. 37(1):103-109.

Paper VI - Hawlitschek O, Hendrich L, Espeland M, **Toussaint EFA**, Genner MJ, Balke M (2012) Pleistocene climate change promoted rapid diversification of aquatic invertebrates in South-East Australia. *BMC Evolutionary Biology*, 12:142.

Paper VII - **Toussaint EFA**, Condamine FL, Hawlitschek O, Watts CHS, Porch N, Hendrich L, Balke M (2015) Unveiling the diversification dynamics of Australasian predaceous diving beetles in the Cenozoic. *Systematic Biology*, 64(1):3-24.

Paper VIII - **Toussaint EFA**, Sagata K, Surbakti S, Hendrich L, Balke M (2013) Australasian sky islands act as a diversity pump facilitating peripheral speciation and complex reversal from narrow endemic to widespread ecological supertramp. *Ecology and Evolution,* 3(4):1031-1049.

Paper IX - **Toussaint EFA**, Hall R, Monaghan M, Sagata K, Ibalim S, Shaverdo HV, Vogler AP, Pons J, Balke M (2014) The towering orogeny of New Guinea as a trigger for arthropod megadiversity. *Nature Communications,* 5:ncomms5001

Paper X - Tänzler R, **Toussaint EFA**, Suhardjono YR, Balke M, Riedel A (2014) Multiple transgressions of Wallace's Line explain diversity of flightless *Trigonopterus* weevils on Bali. *Proceedings of the Royal Society of London, Series B,* 281(1782):20132528.

Paper XI - Condamine FL, **Toussaint EFA**, Cotton A, Sperling FAH, Genson G, Kergoat GJ (2013) Fine-scale biogeographic and temporal diversification processes of peacock swallowtails (*Papilio* subgenus *Achillides*) in the Indo-Australian Archipelago. *Cladistics,* 29(1):88-111.

Summary of the Ph.D. candidate's contributions

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI
Design of the study	-	●	-	-	-	-	●	●	●	●	●
Molecular biology	-	●	●	-	-	-	●	●	●	-	●
Analyses	-	●	●	●	●	●	●	●	●	●	●
Manuscript drafting	●	●	●	●	●	●	●	●	●	●	●
Figure design	●	●	●	●	●	●	●	●	●	●	●

●: Part of the study in which the Ph.D. candidate has been significantly involved

Munich,

Prof. Dr. Gerhard Haszprunar

Emmanuel F.A. Toussaint

English abstract

In the context of a dramatic erosion of natural habitats, the study of the magnitude, evolution and function of biodiversity is paramount. Novel integrative approaches strongly relying on molecular data help unknotting the processes engendering biodiversity. Using molecular phylogenies, the fossil record and geological evidence, it is possible to improve our understanding of diversification processes. Besides, efficient mathematical models have been developed to test hypotheses of character evolution, biogeographic scenarios and forces fostering the rise and fall of clades.

Here, I apply such methods to explore both fine and large-scale mechanisms governing speciation and evolutionary patterns. I focus on an extraordinarily diverse yet highly threatened area, the Indomalayan / Australasian archipelago. Its center, the Wallacea, is a transitional zone between Oriental and Australian biotas. Large-scale tectonic movements and volcanism triggered the emergence, drifting and collision of landmasses throughout the Cenozoic, resulting in arguably the most complex geological setting on Earth. Yet, most of the diversity on the archipelago remains to be discovered and the study of the underlying evolutionary mechanisms has only just begun.

Using phylogenetic trees, I show that the mountain ranges, especially so in New Guinea act as a species pump. Though geologically young, they served as a motor for diversification also for neighboring areas. I show that altitudinal preferences as well as local endemicity are reversible traits. Second, I reveal that the Quaternary climate changes have fostered lineage diversification in Australia through isolation in glacial refugia. I demonstrate that these climatic changes also triggered a wave of extinction in a group of Australasian diving beetles, providing the first empirical evidence for a declining trajectory of diversity for an invertebrate clade. Third, I unveil at a broader geographic scale that notorious biogeographic barriers have had little impact on the evolution of insect clades. Finally, I coauthor integrative new species descriptions firmly believing that formal names assigned to species-group entities are the very currency of comparative biology and ample effort should be made to flank molecular phylogenetics with modern taxonomy.

German abstract (Zusammenfassung)

Der dramatische Rückgang natürlicher Lebensräume erfordert eine schnelle Erforschung von Evolution und Biodiversität. Innovative und integrative Ansätze, die großteils auf molekularen Daten basieren, helfen die Prozesse hinter der globalen Biodiversität besser denn je zu verstehen. Die Verwendung „molekularer Phylogenien“, von Fossilien und immer besseren geologischen Daten ermöglichen es, Diversifikationsprozesse in Zeit und Raum besser rekonstruieren zu können. Zudem wurden mathematische Modelle entwickelt, um Hypothesen von Merkmalsevolution, biogeographischen Szenarien und dem Entstehen und Verschwinden von phylogenetischen Linien zu testen.

Hier wende ich solche Methoden an, um klein- und großräumige Mechanismen der Artbildung und Evolution zu erforschen. Ich fokussiere auf eine sehr diverse, jedoch höchst gefährdete Region, den indomalayisch / australasiatischen Archipel. Dieser umfasst unter anderem Wallacea, eine Übergangszone zwischen orientalischen und australischen Lebensräumen. Ausgedehnte tektonische Bewegungen und Vulkanismus haben das Auf- und Abtauchen, das Driften und die Kollision von Landmassen während des Känozoikums ausgelöst. Dadurch gilt die Region als die geologisch komplexeste Formation der Erde. Die Erforschung der Diversität und der zu Grunde liegenden evolutionären Mechanismen des Archipels stehen auch rund 150 Jahre nach Wallace’ Forschungsreisen gerade erst am Anfang.

Unter der Verwendung von Stammbäumen zeige ich, dass Gebirge in der Region, besonders in Neu Guinea, trotz geologisch relativ jungen Alters eine treibende Kraft für Diversifikationsprozesse sind. Es zeigt sich u.a., dass präferierte Höhenstufen sowie lokaler Endemismus umkehrbare Merkmale darstellen. Des Weiteren zeige ich, dass klimatische Veränderungen im Quartär die Artbildung in Australien durch Isolation in Gletscherrefugien gefördert und eine Aussterbewelle in einer Gruppe australischer Schwimmkäfer ausgelöst haben. Letzteres gibt einen ersten empirischen Anhaltspunkt für einen Rückgang der Diversität innerhalb der Invertebraten. Schließlich konnte ich aufzeigen, dass bekannte geologische Barrieren nur einen kleinen Einfluss auf die Evolution von Insektenlinien hatten. Zuletzt habe ich bei integrativen neuen Artbeschreibungen mitgewirkt, da ich davon überzeugt bin, dass wir in der vergleichenden Biologie nur dann nachhaltig arbeiten können, wenn wir Arten beim Namen nennen, welche ja eine wissenschaftliche Hypothese darstellen.

French abstract (Résumé)

Dans le contexte d'une érosion dramatique des habitats naturels, l'étude de la richesse, de l'évolution mais aussi des fonctions de la biodiversité se révèle cruciale. De nouvelles approches basées sur des données moléculaires permettent de dévoiler les mécanismes engendrant la biodiversité. A l'aide de phylogénies moléculaires, du registre fossile et de reconstructions géologiques, il est possible d'améliorer notre compréhension des processus de diversification. Par ailleurs, des modèles mathématiques ont été développés de façon à tester des hypothèses portant sur l'évolution des caractères, la biogéographie historique ainsi que les facteurs favorisant la diversification et le déclin de certains clades.

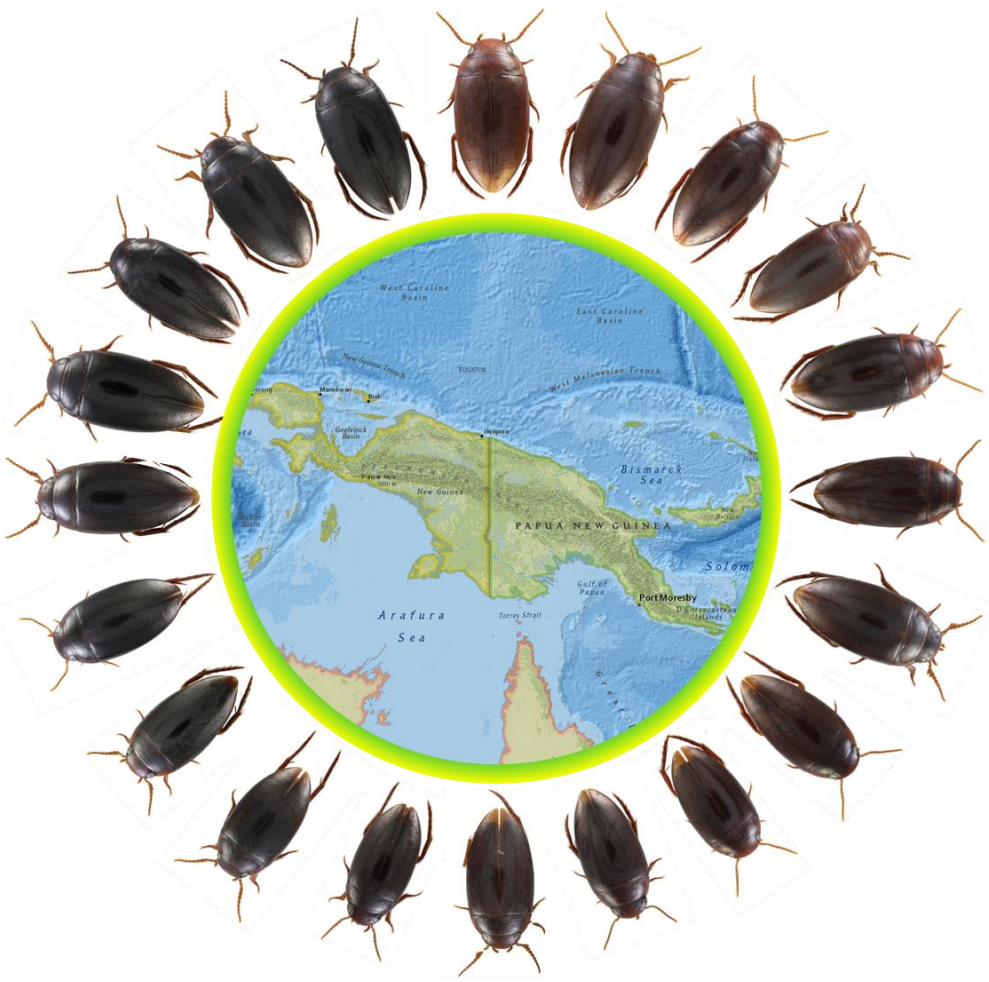
Au cours de mon doctorat, j'ai appliqué ces méthodes afin d'explorer les mécanismes gouvernant la spéciation et les patrons évolutifs à des échelles géographiques diverses. Je me suis intéressé en particulier à une région extraordinairement diverse bien que fortement menacée, l'archipel Indomalayen / Australasien. Son centre, la Wallacea, est une zone de transition entre les biotas Asiatique et Australien. Des mouvements tectoniques et un volcanisme de grande ampleur ont permis l'émergence, la dérive et la collision de masses continentales durant le Cénozoïque, engendrant l'un des arrangements géologiques les plus complexes sur Terre. Pourtant, la plupart de la diversité de l'archipel reste à découvrir, et l'étude des mécanismes évolutifs sous-jacents n'en est qu'à ses débuts.

A l'aide d'arbres phylogénétiques, je montre que les chaînes de montagnes, particulièrement en Nouvelle-Guinée, promeuvent la spéciation. Bien que récentes, elles servent de moteur pour la diversification également dans des régions périphériques. Je montre que les préférences en termes d'altitude ainsi que l'endémisme local sont des traits réversibles. Dans un second temps, je révèle l'implication des changements climatiques du Quaternaire dans la diversification de clades Australiens par isolation dans des refuges glaciaires. Je démontre que ces changements ont également engendré une vague d'extinction dans un groupe de coléoptères aquatiques Australasiens, constituant le premier exemple empirique d'un déclin de diversité chez les invertébrés. Je montre à une échelle plus large que les célèbres barrières biogéographiques de l'archipel ont eu un impact mineur sur l'évolution de certains clades d'insectes. Enfin, je participe à la description de nouvelles espèces à l'aide d'une approche intégrative, car les nouvelles approches de phylogénie moléculaire sont tributaires d'une taxonomie moderne.

Water beetle collecting in the heart of Bromo Tengger Semeru National Park tropical rainforest, Java.

GENERAL INTRODUCTION

Chapter 1. On the study of speciation



New Guinea shelters a tremendous diversity of Exocelina predaceous diving beetles

"*And thus, the forms of life throughout the universe become divided into groups subordinate to groups.*"

Charles Darwin, *On the Origin of Species*, 1859.

Chapter contents

1.1 General background 8
1.2 Species discovery in the era of molecules 10
1.3 Speciation mechanisms and biogeography 15

1.1 General background

Unfolding the mechanisms governing species formation and illuminating why lineages are unevenly distributed on Earth stand among the most enthralling and cardinal questions in evolutionary biology ever since Charles Darwin and Alfred Wallace first crucial observations (Darwin and Wallace 1858; Darwin 1859). Thanks to the advent of cutting-edge methods of evolutionary pattern inference, the discipline has made large strides forward. Yet, evolutionary biology still relies on an initial biodiversity assessment and on the description of (new) species as its basic currency. Recent studies have predicted a global species diversity of approximately 5±3 (Costello et al. 2013) to 8.7±1.3 million species (Mora et al. 2011). However, only a small fraction has already been catalogued (≈1.5 million) (Figure 1).

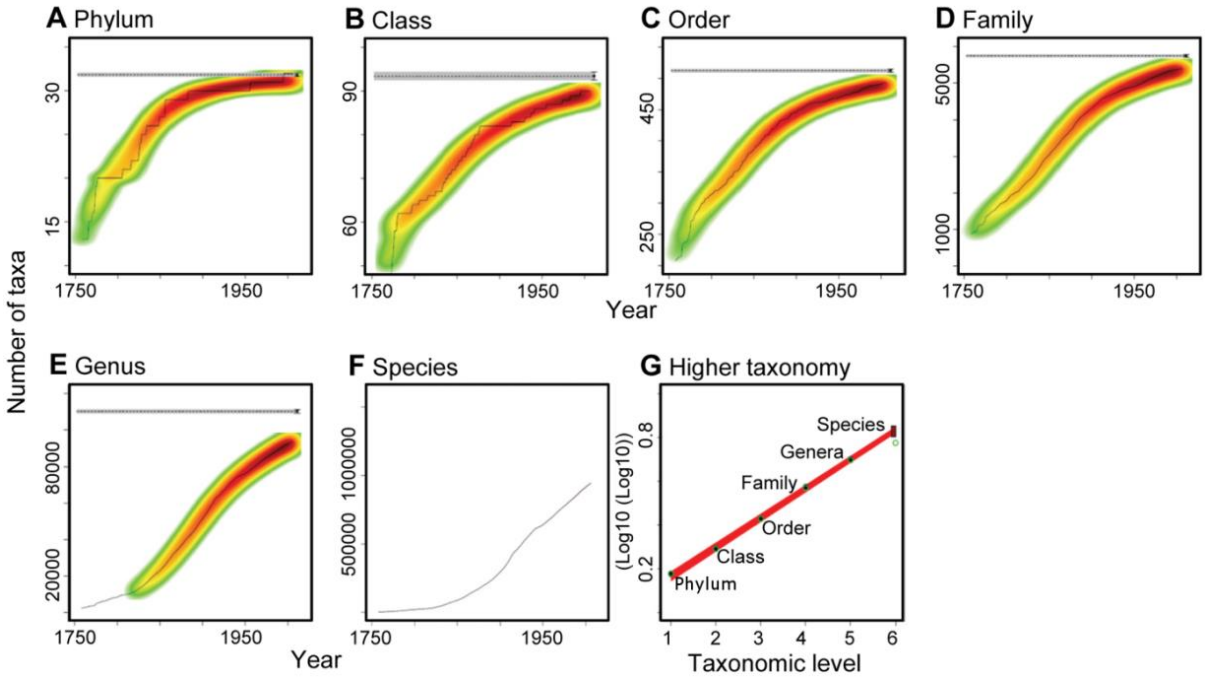


Figure 1. Graphics highlighting the number of new taxa discovered at different taxonomic ranks in the past 200 years and the respective trends for the future (Mora et al. 2011).

Paradoxically, whilst the exploration of biodiversity demands a major acceleration of species descriptions, this very same science has to face the wane of trained specialists and funding drying-up. Hence, novel approaches are needed to overcome this jeopardized enterprise and taxonomy has been shown to have great potential to be reinvigorated using an integrative approach bringing together traditional and modern molecular phylogenetic approaches (Riedel et al. 2013).

The science of discovering and naming new taxa is not recent. Since the origin of human society, we have always been intrigued by the living forms surrounding us. The most paramount achievement that initiated the modern taxonomy era was the publication of Carl von Linné's binomial nomenclature (von Linné 1758). The introduction to the community of this novel way to describe taxa and classify them has been a major revolution for science in general, and about 250 years later it remains a unique and universal reference to name and classify organisms. As one of the cornerstones of biology in general (de Queiroz 2007), species are compulsory to study fine to large-scale patterns of biodiversity ranging from community ecology to population genetics and from geographic gradients of biodiversity to uneven species-richness between clades. However, many different species definitions have been proposed since the early definition of Mayr (1942) implying that species are groups of interbreeding or potentially interbreeding populations reproductively isolated from other groups sharing the same characteristic. These novel definitions were based on different features such as the behavior, diagnosability, distribution, ecology, morphology or phenology of the taxon considered, which could be incompatible and thus contradictory (Figure 2; Hausdorf 2011).

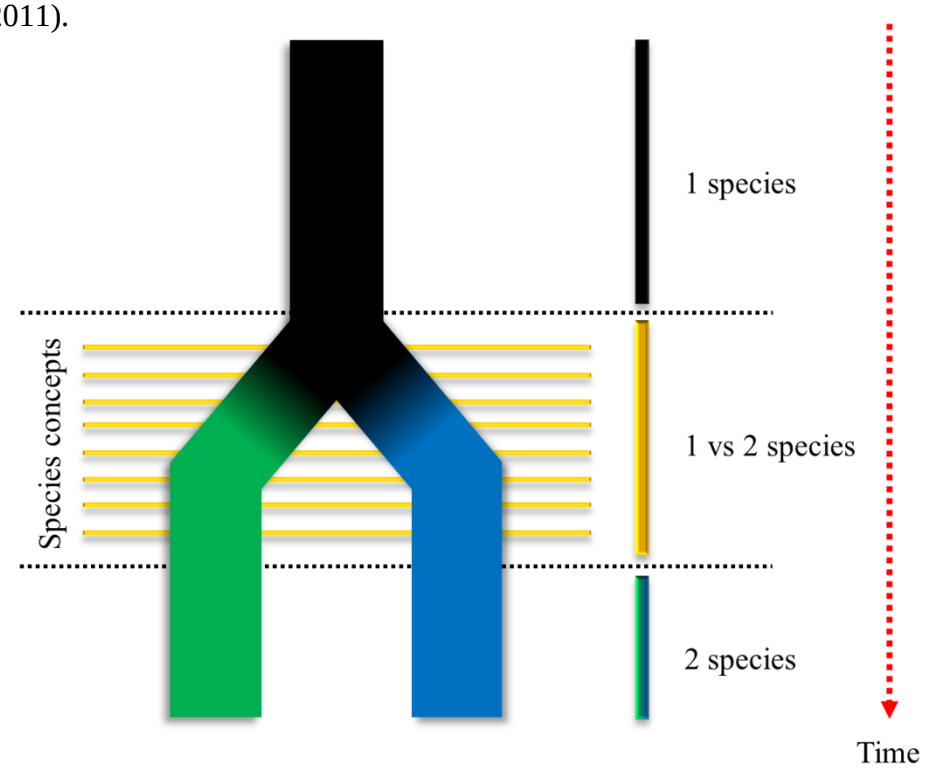


Figure 2. Diagram redrawn from de Queiroz (2007) representing the divergence from mother to daughter species according to the species concept considered. The yellow zone represents a period of time during which there is no clear separation but only acquisition of novel features such as a different ecology or the incompatibility to reproduce. These features might be the threshold to delineate species for some species concepts but not for others. Outside of this yellow area there is no conflict on the number of species.

Despite theoretical progress illustrated for instance by the Unified Species Concept developed by de Queiroz (2007) (see Hausdorf 2011 for a critical view of this concept), the species concept remains complex due to marked discrepancies in the way species are described across organisms (Coyne and Orr 2004). Species may be described following different criteria and definitions principally relying on ecology, geographic isolation, morphology or more recently DNA. For instance, most of the flowering plants offer straightforward morphological characters to establish species-level diagnosis, whereas some other organisms like prokaryotes, protists and viruses lack such clearly-defined features, rendering the task of species description more tedious (Butlin et al. 2009). Cryptic lineages, for which morphological characters might be subtle or not fixed yet, also constitute a challenge for traditional taxonomy (Avise 2000). There is therefore a need for a more integrative taxonomy combining different sources of characters in order to overcome the taxonomic impediment and initiate a new era for biodiversity assessment and description.

In this dissertation, I use an approximation of the biological species concept. Because in most cases it is unrealistic to test properly the feasibility of mating between taxa, I use different proxies to estimate species boundaries. Geographic distributions, genetic variation and morphological features are the three main proxies I used in the course of my thesis in order to delimit species. As a result, I assume that taxa presenting clear divergences in some of these proxies are different species under the biological species concept because the differences we observe are likely to be the result of differential lineage evolution.

1.2 Species discovery in the era of molecules

Only recently has it become feasible to combine multiple sources of data to investigate speciation processes in a truly integrative fashion and at larger scales (Padial et al. 2010; Riedel et al. 2013). Since the original plea for a DNA-based taxonomy (Tautz et al. 2003; Vogler and Monaghan 2007; Wiens 2007), major advances have been achieved to monitor biodiversity using molecules, offering a new perspective to speed-up biodiversity exploration by permitting to alleviate the shortcomings of taxonomy based on phenotypic features only (Jörger and Schrödl 2013). After a discrete emergence by the end of the 20th century (e.g. Cracraft 1983), molecular species identification and delineation methods have flourished at the beginning of the 21th century, offering tantalizing opportunities to accelerate species discovery and untangle cryptic speciation processes.

In 2003, DNA Barcoding was introduced as an approach to identify species and enhance biodiversity assessment using standard sequence data to calculate clusters that could potentially represent species (Hebert et al. 2003a,b). Using DNA barcoding, the identification of a specimen is done by algorithmic comparison of its barcode against a pre-existing database ideally comprising multiple specimens of multiple species which have been identified by a taxonomic authority (Hebert et al. 2003a,b). Although the initial aim of this new method was to provide quick and objective identifications and achieve the elaboration of a molecular library (Hebert et al. 2003a,b), a second goal emerged, to detect cryptic diversity and allow the delineation of species-level genetic clusters using DNA sequences (Hebert et al. 2004a). In their study, Hebert et al. (2004a) established DNA Barcoding as a way to discover species-level entities but also to question the essence of the species concept itself, because phenotypically homogeneous lineages started to be potential sources of an unknown diversity comprising ecological and/or reproductively isolated variants. The authors eventually concluded that the cryptic clusters they unveiled should be described as separate species since they presented ecological and reproductive particular characteristics, therefore raising DNA Barcoding not only to the stage of being a universal species identifier but also a potential engine to discover new species (Hebert et al. 2004a).

Roughly 10 years after DNA Barcoding was introduced to the scientific community, an astonishing accumulation of sequences has been achieved, and the method originally designed for the animal realm has been extended to fungi, plants and bacteria (see Hajibabaei et al. 2007 for a review). The success of this large project is undeniable for species identification purposes and in a broader view as the development of a library of life. In 2014, the database created to gather all the sequence data has now grown to approximately 3,000,000 barcodes representing more than 300,000 animal clusters. However, if DNA Barcoding has proved to be a valuable tool to identify species and reveal cryptic complexes, it did not accelerate the description of new species. As an illustration, the nine "barcoding species" of hesperid butterflies unveiled in Hebert et al. (2004a) have not been described by Hebert and colleagues but 6 years later by Brower (2010) no without emphasizing that "[…] the publication of these names does not imply the author's endorsement of using DNA barcode polymorphisms in lieu of more substantial diagnostic features as a general practice" and further down "My role here is simply that of a responsible citizen who, confronted by a sort of taxonomic oil slick, has taken it upon himself to clean it up. This nomenclatural remediation does not constitute an endorsement of what might be referred to as a barcode

species concept". In a time of taxonomic impediment, where classical taxonomy is unable to cope with the demand for a fast description of biodiversity, DNA Barcoding has certainly paved the way for a more integrative taxonomy, but one should keep in mind its major pitfalls such as sensitivity to introgression and incomplete lineage sorting (Monaghan et al. 2006), presence of pseudogenes (Song et al. 2008) or lineage idiosyncrasy (Hendrich et al. 2010) which may render it unsatisfactory when it comes to delineate species in a thorough fashion. These pitfalls are of particular importance for the "barcoding gap" which relies on the assumption that interspecific variation exceeds intraspecific variation to such an extent that under an a priori defined cutoff, it is possible to assign any individual to its databased species or to a new species-level genetic cluster (Hebert 2003a, 2004a,b). Empirical studies have proven that DNA Barcoding is sub-optimal to identify the delineation between intra and interspecific domains (e.g. Meyer and Paulay 2005; Wiemers and Fiedler 2007). Moreover, its potential for specimen identification is limited by the requirement of a complete database of living species identified beforehand and the correctness of the taxonomic status given to the species in the database. Despite clear advantages (Goldstein and DeSalle 2011), DNA Barcoding does not overcome the taxonomic impediment and there is a call for more thorough methodologies to identify species boundaries.

Methodologies based on molecular phylogenetic trees and the Coalescent theory (Kingman 1982), flourished in the past decade in response to a demand for integrative species delimitation procedures. In 2006, the first sophisticated methodology to delineate species using the Coalescent theory was introduced, the General Mixed Yule Coalescent (GMYC) (Pons et al. 2006). This method does not rely on a priori taxonomic information and is based on the principle that the branching pattern of a phylogenetic tree should present a threshold between inter and intraspecific levels, in other words between species and populations (Pons et al. 2006). The algorithm maximizes the likelihood of the GMYC model on a tree and the optimal solution yields a threshold before which branches correspond to diversification events and after which branches correspond to coalescent processes. The GMYC likely constitute a more reliable tool than DNA Barcoding to delineate species, because unlike the latter, it is based on proper phylogenetic trees based on models of evolution and possibly inferred using different genes than the unique COI. This method has become and remains very popular especially because of its good results in simulation and empirical tests (Fujisawa & Barraclough 2013; Talavera et al. 2013) despite presenting several issues such as a strong dependence to the correctness of the speciation model specified (Yule model), high sensitivity

to sampling incompleteness (Lohse 2009; see Papadopoulou et al. 2009 for a response), calculation on single-locus datasets only or use of an error-prone ultrametric tree. Several developments of this notorious method have been achieved, such as the possibility to have multiple thresholds (Monaghan et al. 2009) or account for gene tree uncertainty in a Bayesian framework (bGMYC, Reid and Carstens 2012). Many additional methods relying on the Coalescent theory (see Fujita et al. 2012 for a review) have been developed since the GMYC model, allowing these DNA-based species delimitation methods to be a growing fraction of the integrative taxonomy initiative.

Multiple models have been developed recently including discovery and validation methods. Discovery methods such as GMYC (see above), Gaussian clustering (Hausdorf and Hennig, 2010), STEM (Kubatko et al. 2009; Satler et al. 2013) or Structurama (Huelsenbeck et al. 2011), aim at dividing populations into clusters without any prior on the delineation of species. For instance, the Poisson Tree Processes model (PTP) is one of the most recent developments in species delimitation method, allowing to have a non-ultrametric tree unlike GMYC, therefore alleviating pitfalls linked to the calibration of the tree (Zhang et al. 2013). On the other hand, validation methods require a priori groupings of the samples included in the tree based on available information including DNA, geography, geology, morphology or ecology, and then test the accuracy of these assumptions in a probabilistic framework. Such methods like the Bayesian Phylogeography and Phylogenetics (BPP, Yang and Rannala 2010; Rannala and Yang 2013), Brownie (O'Meara 2010) or spedeSTEM (Ence and Carstens 2011), rely on multilocus sequence data but can deliver contradictory results and should be carried out in concert with discovery approaches in order to alleviate the potential biases of the priors used. Discovery and validation method combinations also known as chimeric approaches have also been developed (Leaché and Fujita 2010; Niemiller et al. 2012).

Species delimitation methods based on genetic information constitute a credible way to accelerate the discovery of new species and also permit to overcome the shortcomings of traditional species delimitation based on morphological characters, such as homoplasy (sharing of a similar character state not derived from a common ancestor) and cryptic speciation. However, the results yielded by DNA-based species delimitation methods should be cross-checked using multiple DNA-based methods instead of only one (Figure 3, Carstens et al. 2013) and several lines of evidences (Schlick-Steiner et al. 2010).

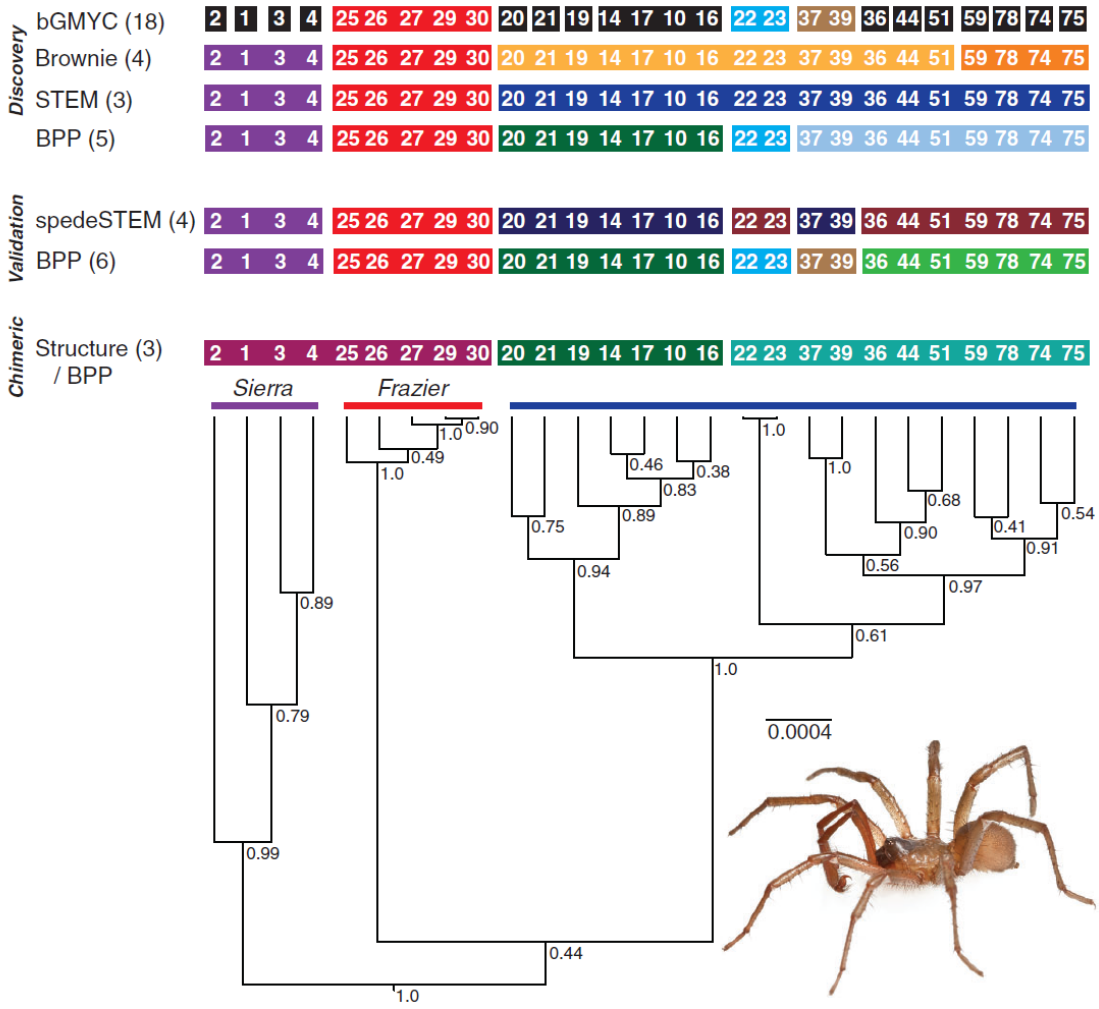


Figure 3. Incongruence in species delimitation results based on a dataset of six loci in a species complex of *Aliatypus* trapdoor spiders. Seven different methods (named at left) were used. Numbers in parentheses represent the number of lineages delimited (Salter et al. 2013).

Overall, new tools have been developed in the past decade to make of taxonomy a truly integrative science (Padial et al. 2010; Riedel et al. 2013). The species are pivotal units of biology and as such, are used in a wide array of studies in evolutionary biology and ecology. Even though the concept of species remained blurry for a long period of time, striking progress has been achieved towards a consensual view of species as independently evolving lineages (de Queiroz 2007). Species delineation has evolved to become an integrative approach combining DNA sequences, morphology but also geographic data (Raxworthy et al. 2007; Rissler and Apodaca 2007). If substantial progresses can be witnessed, much remains to be done using empirical data not only to accelerate inventorying Earth's biodiversity but also to bring new insights into a deeper understanding of speciation mechanisms.

1.3 Speciation mechanisms and biogeography

For most of the 20th century, speciation was understood as the mechanism by which geographically isolated populations would gradually accumulate genetic differences in a random fashion so that eventually they could not interbreed anymore (Hendry 2009). We know today that speciation is much more complex and can take on different forms. Selection (e.g. natural, ecological, competitive or sexual) in particular is assumed to have a leading role in its accomplishment by engendering reproductive isolation (Schluter 2009). Evolutionary mechanisms triggering the apparition of new species have been classified in four theoretical modes of speciation (Coyne and Orr 2004) which are influenced by several factors including for instance ecology, genetic drift, selection or polyploidy (Schluter 2001; Gavrilets 2003, Gavrilets and Vose 2007; Gavrilets et al. 2007).

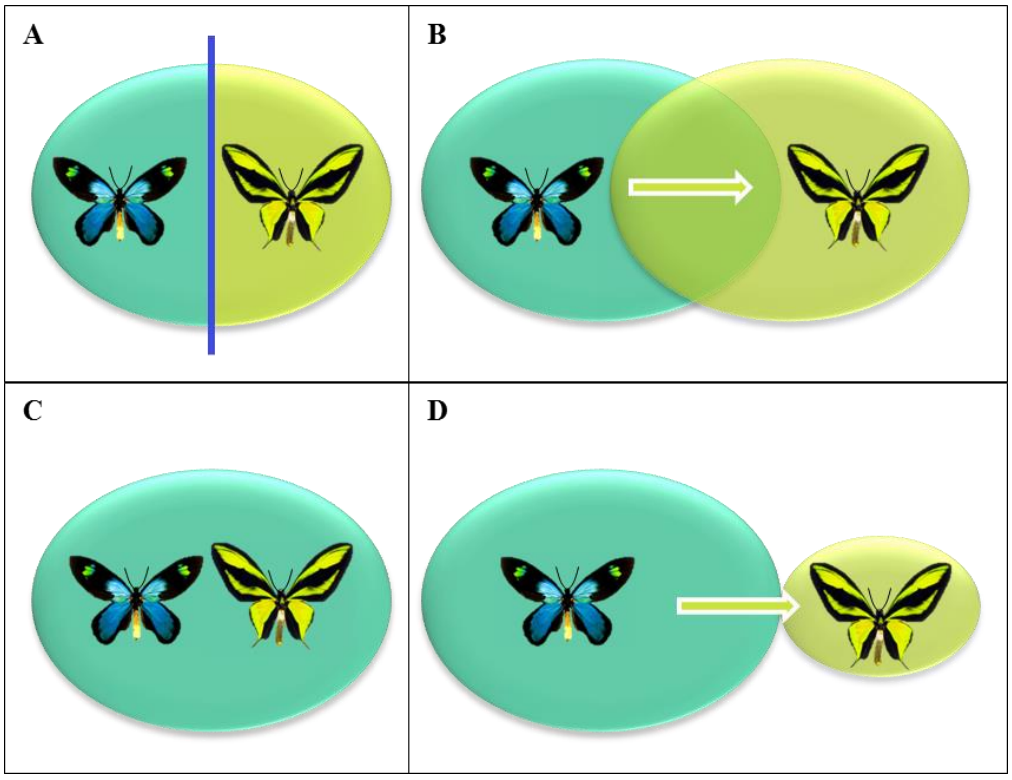


Figure 4. Schematization of the main modes of speciation. Blue and yellow butterflies illustrate distinct species. Blue and yellow circles indicate geographically or ecologically distinct areas. A: Allopatric speciation (the purple line represents a barrier fostering isolation); B: Parapatric speciation; C: Sympatric speciation; D: Peripatric speciation.

The process of speciation by isolation of two populations from an ancestral lineage which fix through their evolution apart some phenotypic and/or genotypic differences is called allopatric speciation (Figure 4). This mode of speciation, thought to be the most common in nature, can be easily illustrated by an ancestral population being progressively

divided by a mountain build-up in two population subsets. When the populations are separated geographically or "vicariant", then gene flow is prevented and evolutionary forces such as genetic drift, selection and mutation act independently in both subsets. After a certain amount of time, the two subsets will have fixed enough differences to become distinct lineages which in case of secondary reconnection would be unable to interbreed. Although the example of mountain build-up is convenient for illustration purposes, allopatric speciation can be fostered by any kind of barrier resulting in the isolation of populations. Possible barriers can be of geological (e.g. orogeny, marine incursion or tectonic separation of ranges) or climatic (e.g. glaciation or aridification) origin, even though both are often linked.

In a situation where a population resides in an area which overlaps with an adjacent one, the colonization of this latter by a subset of the population is referred to as parapatric speciation. In this case, the subset of the initial population can still be in contact and gene flow might happen at the edge of the respective ranges of the two populations. However, it is likely reduced and evolutionary forces might work differently, progressively leading to genetic divergences over time. The overlap of both population ranges can represent for instance an incomplete barrier as in the case of allopatric speciation, but still allowing exchange and interbreeding between the two populations.

Speciation can also be triggered by the colonization of a non-adjacent area or ecological niche by a subset of the population. This mode of speciation named peripatric speciation is similar and sometimes included in the allopatric speciation because the two populations are geographically separated and the gene flow is prevented, however the colonizing population is much smaller due to the fact that the adjacent area is remote and at the margin of the initial range. Examples of such mode of speciation can be found in polar and brown bears (Edwards et al. 2011) as well as in hermit crabs (Malay and Paulay 2010).

Sympatric speciation is a mode of speciation in which a new species arises from a population in a same range and without any barrier to gene flow. This process has retained the attention of the scientific community because of its apparent implausibility and aroused much controversy over the past century (Via 2001; Bolnick and Fitzpatrick 2007). Mayr in particular made a strong case against it in 1963, arguing that the model was unlikely in theory, referring to it as "the Learnean Hydra" because he predicted that investigations on this topic would not stop even after his plea. In the past decades, a growing line of simulation and

empirical evidences was gathered, suggesting that some ecological as well as genetic conditions were likely responsible for sympatric speciation. The role of selection and in particular disruptive selection (selection of extreme phenotypes over intermediate ones) was also underscored as a good candidate to trigger this mechanism in nature. Among the most convincing empirical examples of sympatric speciation are cichlid fishes of the Crater Lake Apoyo (Barluenga et al. 2006) or Howe island palms (Savolainen et al. 2006). In most cases, sympatric speciation is thought to be fostered by traits subject to disruptive selection and at the same time controlling non-random interbreeding, but alternative explanations exist relying on linkage disequilibrium of genes involved in mating behavior and adaptation (Dieckmann and Doebeli 1999; Gavrilets and Vose 2007; Gavrilets et al. 2007).

If the study of speciation mechanisms helps bringing new insights into our understanding of clade diversity, the study of the geographic origin of species in a paleogeological framework is paramount to appreciate the micro and macroevolutionary factors shaping species richness and patterns of distribution (Cox and Moore 2010). This discipline referred to as biogeography captures a tremendous number of topics focusing on species distribution patterns and processes. The roots of biogeography lie before the time of great explorers, with the French naturalist Georges Buffon who noticed in his "*Histoire Naturelle, générale et particulière"* (1749–1788) that faunas were different from a continent to another. He was particularly puzzled for instance to observe that what Peruvian natives called puma exhibited only little resemblances with the African lion, at a time where all animals were thought to be the creation of God and their distribution the result of Noah's Ark landing. He spent a good part of his life working and writing about these patterns of zoological geography, paving the way for what we know now as the "modern" biogeography.

Phytogeography played an important role in the development of biogeography as we know it today, with the works of Forster (1778) or von Humboldt (1805–1834) who questioned the distribution of plants in different regions of the World and at different altitudes. Forster in particular was the first to highlight a latitudinal gradient of diversity (Forster 1778). Between the beginning of the 19th century and the publication of Darwin's masterpiece "On the origin of species" (1859), biogeographers produced the first distribution maps of plants and animals. In 1858, Sclater published his view of the main biogeographic regions of the World based on the distribution of passerine songbirds, which would be the starting point of Wallace's drawings. With the theory of evolution by the means of natural

selection championed by Charles Darwin and Alfred Wallace shedding light on the origin of species (Darwin and Wallace 1858; Darwin 1859), biogeography was entering a new stage of development untied from reasoning constraints inherent to theological perspectives. In particular, biogeography entered a new era with the paramount achievements of Wallace in the Indomalayan / Australasian region (1869, 1876, 1880) which included reflections on the impact of climate shifts, now-vanished land bridges between islands or between islands and continents, extinction, dispersal or adaptive radiation among others factors shaping the zoogeographic patterns he observed.

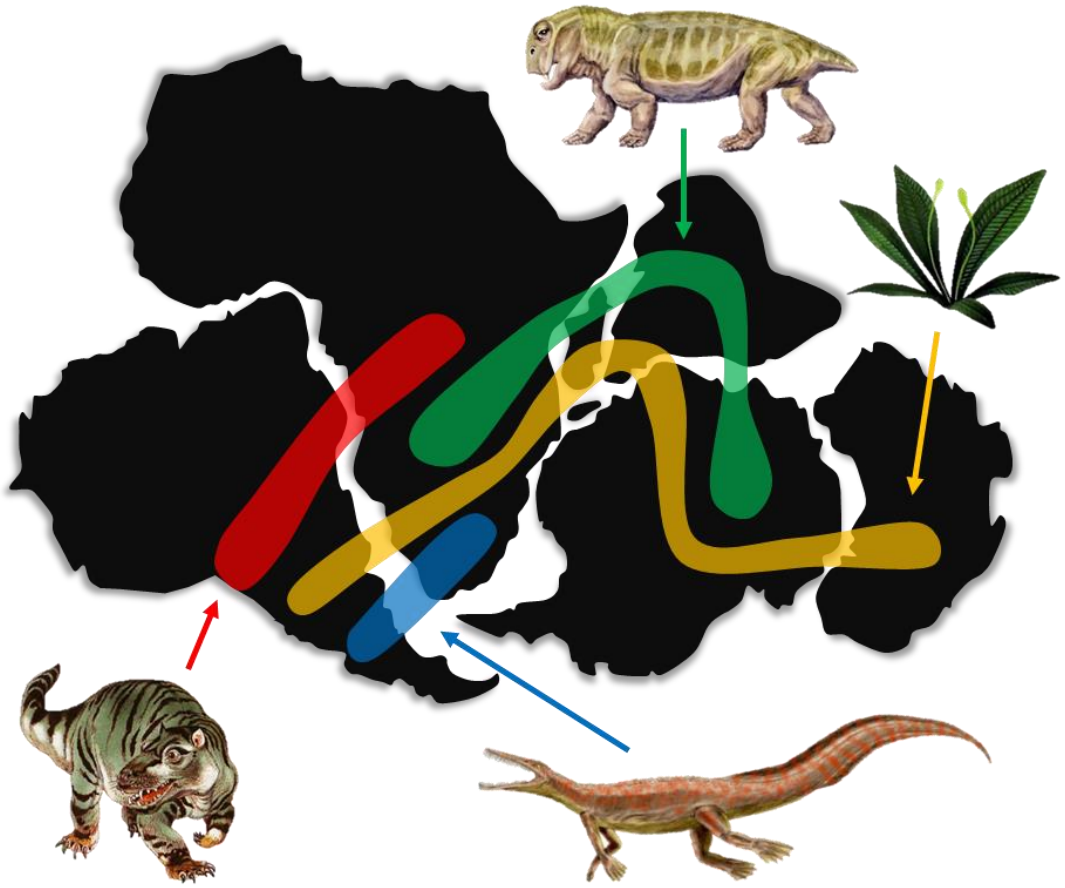


Figure 5. Fossil distribution of extinct lineages that occurred on different present-day continents and inspired Wegener's continental drift theory. In red: *Cynognathus* therapsids (mammals and extinct close relatives) from the Triassic (≈200–250 Ma); in green: *Lystrausaurus* therapsids from the Triassic; in yellow: *Glossopteris* seed ferns from the Permian (≈250–300 Ma); in blue: *Mesosaurus* aquatic reptiles from the Permian.

By the end of the 19th century and a large part of the 20th century (but see Croizat's panbiogeography hypotheses; 1958), the main concept was dispersalism, the idea that the distribution of species was the exclusive result of dispersal events, a theory derived from the work of the leading scientists and explorers who had made by that time most of the contributions to the theory of biogeography. Despite its early discovery by Wegener (1912),

the mechanism of continental drift or plate tectonics was only formally accepted much later in the second half of the 20th century through the publication of several robust lines of evidence. Following the processes unveiled by Wegener, the idea of distributions being shaped by landmass movements and geologic barriers (or vicariance) started to emerge, quickly becoming the favored concept to explain biogeographic patterns. Wegener's concept being accepted, paleontologist and biogeographers started to have a closer look at the fossil record and found fossils belonging to extant clades being far outside of the present-day distribution of these taxa (Figure 5). Therefore the fossil record could potentially inform biogeographers on past distribution of lineage ancestors extinct several million years ago (Ma), thus providing paramount information to understand global patterns of biodiversity (Lomolino et al. 2010). At long last, some astonishing paleogeological and paleogeographic maps highlighting the setting-up of landmasses and water bodies on Earth have been and continue to be designed for the past hundreds of millions of years (Myr) (e.g. Hall 2002, 2011, 2012; Metcalfe 2011). This development along with the advent of sophisticated molecular phylogenetic methods, allow biogeographers to work in a statistical and temporal framework to decipher the biogeographic scenarios that engendered the extant biodiversity observed today.

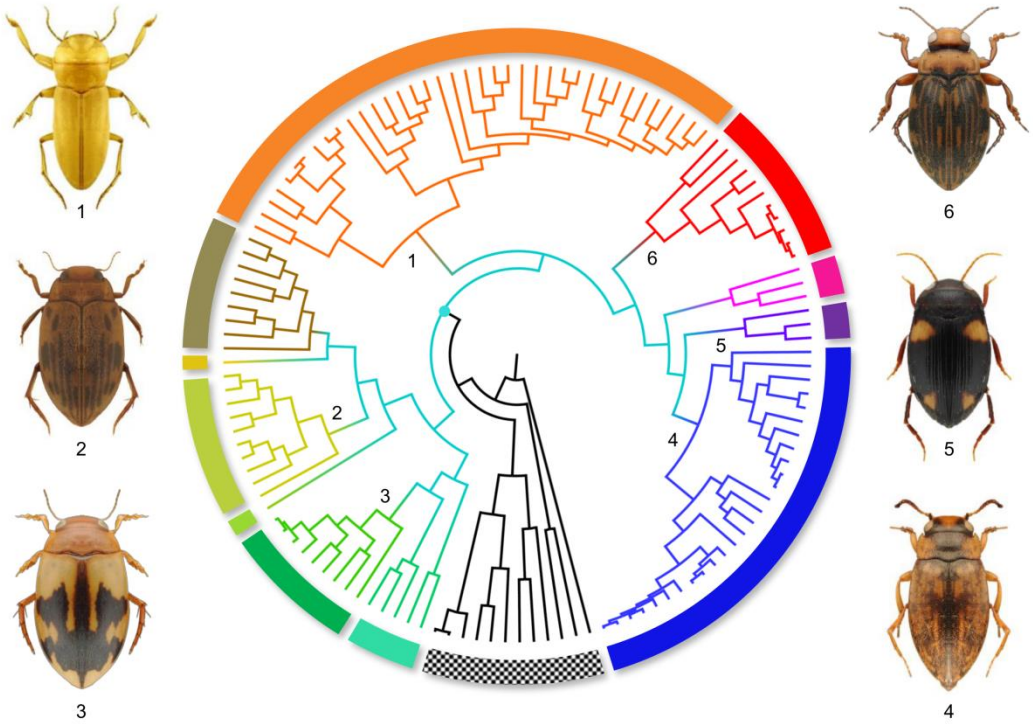
Among the numerous subsets of biogeography, one that has received an increasing attention since the middle of the 20th century is island biogeography. Although this field refers to different kinds of insular systems such as oceanic islands, mountain summits (sky islands), lakes or fragmented habitats, I will mostly focus on oceanic island biogeography in this part. Darwin and Wallace shed light on crucial evolutionary mechanisms and biogeographic patterns in the Galapagos and Indomalayan / Australasian archipelagoes respectively (Wallace 1876; Darwin 1909). Despite the lack of theoretical and geological background at their time, they proposed important hypotheses regarding the origin and evolution of biota on islands. Darwin the first postulated that the geographic isolation of populations by a barrier (allopatry) could lead to speciation through natural selection. He also suggested that species originating in one place might colonize new suitable habitats by dispersal. Approximately at the same time, Wallace was describing a demarcation line between Borneo and Sulawesi representing the crossroads of the Asian and Australian biotas, a barrier named after him and now acknowledged as one of the most notorious biogeographic barriers on Earth. The late discovery of plate tectonics reinvigorated the interest for island biogeography. Traditionally, three main types of islands are considered; drifting continental islands, islands along volcanic arcs and ephemeral hotspot islands (Cox and Moore 2010). As plate tectonics are a dynamic

mechanism, whilst new islands appear by subduction, fragmentation of continental landmasses or intense hotspot volcanism, some others drift and dock on other islands or continental landmasses, or simply disappear through subduction and/or erosion. In 1963, MacArthur and Wilson introduced the theory of island biogeography stipulating that species richness on islands is regulated by immigration and extinction, with immigration being higher when the island is close from potential sources of migrants (i.e other islands or mainland/continent), and extinction lower when the island is large. In other words, a small and remote island would be expected under this model to have a reduced species richness compared to a large island close from the continent and/or surrounding islands (MacArthur and Wilson 1963, 1967). Although this model based on simple predictions is very powerful, a key component is missing, the fact that colonizers might undergo speciation once settled on the island. As a result, species richness is not only the result of colonization and extinction but is also shaped by speciation in situ. Adaptive radiation, the fact for a lineage to radiate after the acquisition of a key innovation or the colonization of a new empty niche (Schluter 2000) has been proved to be an important mechanism to fuel lineage diversification on oceanic islands.

As described in the "Red Queen" (Van Valen 1973) and "Court Jester" (Barnosky 2001) models, biotic factors (e.g. competition or predation, see Antonelli and Sanmartín 2011) and abiotic factors (disruptions of climatic conditions, plate tectonics and associated geological dynamics) act alone or in concert to shape biodiversity patterns over different temporal and geographic scales (Benton, 2009). What we presently observe when we look at our planet is a snapshot of macro and microevolutionary processes at work. We can easily observe the unequivocal distribution of species richness around the globe as illustrated by altitudinal, latitudinal or longitudinal gradients of diversity (Gaston 2000). The fact that some clades comprise more species than others at a same taxonomic scale and that some groups have myriads of fossils when some others lack such a fossil record, is also remarkable. Thanks to the recent advent in molecular phylogenetics, paleoclimatology and paleogeology, but also the development of increasingly more powerful statistical models, we hold a chance to infer accurate patterns of biodiversity and understand their triggers.

GENERAL INTRODUCTION

Chapter 2. Phylogenetics as a tool to unveil patterns of biodiversity



Bayesian phylogenetic tree of Australasian Hydroporini diving beetles dated using fossil evidence

"*A fool sees not the same tree as the wise man sees.*"

William Blake, *The Marriage of Heaven and Hell*, 1793.

Chapter contents

2.1 From characters to trees: A methodological overview 22
2.2 Enter clocks: a guide to dating trees 29
2.3 How trees help unfolding biogeographic and diversification patterns 34

2.1 From characters to trees: A methodological overview

Inferring the genealogical relationships between organisms is what phylogenetics are all about. If nowadays evolutionary biologists mostly rely on molecular characters to reconstruct ancestor-descendant relationships between different taxa, any source of character is theoretically usable to accomplish this task. As a starting point, different heritable traits or characters (absence/presence of feathers, color of the wings, nucleotide positions, amino-acids, etc…) must be collected and arranged into distinct columns of a matrix where rows represent different taxa. A first step in phylogenetic inference is the alignment of the characters to ensure that what we compare is comparable, in other words that the character we are comparing between different taxa is the same (=homologous) because we do not want to compare pears and apples. Once matrices have been assembled and aligned, several methods and criterions can be used to infer the corresponding phylogenetic trees. It is important to note that phylogenetic trees are hypotheses based on a defined set of characters and are therefore not always the "true" phylogenetic tree depicting the evolutionary history of lineages.

Overall, two main groups of methods exist permitting to reconstruct phylogenetic relationships. The first one encapsulates the so-called pairwise distance methods which are not based on characters strictly speaking but on the fraction of differences between sets of characters. For instance, when using nucleotide sequences as an input, the first step of this type of analysis is to compute a matrix of distances between each possible pair of sequences (Swofford et al. 1996). For each pair of sequences, the fraction of positions which differ between the two sequences is a percentage called "p-distance". Nevertheless, some conceptual issues need to be considered when calculating p-distances. Indeed, in the course of the evolution of a taxon, the state of a given character can change several times through mutations. As a result, the sequence observed nowadays is only a snapshot of the evolutionary history of the considered taxon and therefore it is difficult to predict the number of times a given state has changed. Such shortcoming can be alleviated by the implementation of models of evolution in the calculation of the distances. These models of evolution make particular assumptions about the number of substitutions in the evolution of the sequence. Thus, only distances corrected with models of evolution should be used to reconstruct phylogenies. The selection of a model of evolution is possible through sound techniques mainly based on likelihood tests fitting different models to the data considered and then selecting the best-fitting model under different kinds of criterions (see Posada 2009 for more details). Once

corrected distances have been calculated under a proper model of evolution, two main types of distance methods can be applied. Clustering methods such as the UPGMA (Unweighted-Pair Group Method with Arithmetic means) gradually cluster sequences with the smallest distances until there is no more sequence to cluster. A major pitfall of these methods is the assumption that all the lineages are evolving at the same pace, a phenomenon referred to as "molecular clock". If the molecular clock assumption is met in some cases, in many other cases it is not, leading the calculation of the distances to be fallacious and the resulting trees prone to errors. Although new methods appeared to overcome this problem, much more powerful and explicit others have been develop to cope with the complexity of sequence data, and clustering algorithms are not used anymore in molecular phylogenetics. Methods relying on additive-distances have been developed to account for non-molecular-clock behavior of sequence data. In these methods such as the notorious Neighbor-Joining, sequences of more or less related taxa are allowed to have different rates of evolution. Unlike clustering methods, the starting point of these analyses is an unrooted star topology where all taxa are equally separated from a central node. From this tree, the two taxa with the smallest differences will be clustered together, and in a stepwise fashion all taxa are added progressively to reconstruct the tree. Although these methods are widely used because they are fast and potentially very efficient if the distance matrix is correct, more complex methods relying on characters and not distances have been proved to be more reliable and accurate in general because they do not reduce the phylogenetic signal between two sequences to a single fraction of divergence.

The second group of methods used to reconstruct phylogenetic trees comprises methods based on characters and not on a matrix of pairwise distance between taxa. These methods are based on an optimality criterion under which an algorithm tries to find an optimal tree. The first method I will introduce is the Parsimony method also called Maximum Parsimony (MP) who can easily be described as the search for a phylogenetic tree that will minimize the number of evolutionary changes required to explain the differences observed in a matrix of characters. The parsimony method seeks to reconstruct the phylogenetic tree that maximizes the number of character states inherited (synapomorphies) and minimizes the number of homoplasies. As a result, only the characters which present more than one state shared by at least two taxa are kept in the analysis because characters with only one state and characters for which only one taxon has a different state are uninformative to reconstruct a tree under the parsimony criterion. The method minimizes the number of changes by assigning character states to the nodes of a tree. To reconstruct the tree that counts the

smallest number of transformations needed, different techniques exist. Two techniques allow finding the shortest tree possible using different algorithms. First, an algorithm can compare all possible trees, but this method is highly time-consuming. For instance, a matrix containing 11 taxa can yield about 35 million possible trees (Swofford et al. 1996). Second, an algorithm can compare the score of a randomly generated tree with the score of a tree of three taxa on which the algorithm adds the rest of the taxa one by one. As long as the score of this gradually more complex tree is below the one of the randomly generated tree, the algorithm keeps adding new taxa. If the tree has a score higher than the randomly generated tree then the algorithm restarts from the beginning but will no longer consider all the trees that could be derived from the combination with a higher score. When a tree comprising all the taxa has been assembled and still has a lower score than the randomly generated tree, then it will be the new reference tree to which new combinations will be compared. This technique named "branch-and-bound" avoids considering all possible trees but remains considerably time-consuming and therefore some heuristic approaches have been developed to approximate the shortest tree without providing the insurance that the resulting tree is indeed the shortest one considering the data. Several algorithms have been created to perform those heuristic methods. Some add the taxa and retain the shortest tree after the addition of each taxon, when some others start from a random topology and swap some branches to stop when no shorter tree can be found after a certain amount of trials. All these heuristic methods are approximations and present a major pitfall, they can indicate a tree as the shortest whereas it is not just because they have found one of the shortest combinations possible which therefore will be difficult to outperform despite the fact that a shorter one actually exists. This problem which will be also discussed in the probabilistic methods below is known as local optimum. Although MP can be simple to understand and fast using heuristic approaches, it holds some major drawbacks. In particular, the impossibility to consider models of evolution is a problem which results in a failure to correct for multiple changes at a same site. When different taxa in a same tree evolve quicker than the rest of the taxa, they can accumulate such changes which have a higher probability to yield similar states than slow evolving taxa. In this situation, homoplasies can be interpreted as synapomorphies and artefactually gather the taxa together. This phenomenon called "long-branch attraction" results from the lack of models of evolution in MP and can also be encountered in other methods if simplistic models are specified.

Probabilistic methods comprising Maximum Likelihood (ML) and Bayesian Inference (BI) are also methods based on an optimality criterion. In these methods, the aim is to find the topology, branch lengths and models of evolution maximizing the likelihood of the data considered (nucleotide matrix for example). The ML method is defined as a search for the tree among all possibilities that maximizes the global likelihood (=maximum-likelihood tree). For a small amount of taxa, finding the optimal tree is possible, however, for multiple taxa and large amounts of data, this task can be extremely tedious and time-consuming. As a result, heuristic approaches have been developed to overcome this issue. The search usually starts with an initial tree which can be randomly generated or inferred using simpler methods. Additional trees are then generated based on this initial tree using tree rearrangement operations. These rearrangements imply swapping of branches using different algorithms (eg. Nearest neighbor interchange, sub-tree pruning and regrafting or tree-bisection and reconnection). The likelihoods of the newly generated trees are compared to the initial tree, and if one topology has a better score, it will be the starting point for the new step of the optimization. The procedure stops when no additional topology can improve the likelihood.

Figure 6. Sampling of beetles serving as an example to illustrate Bayes' theorem (From top to bottom: *Sandracottus rotundus*, *Laccophilus medialis*, *Exocelina munaso*, *Cybister dehaanii*).

When the ML method aims at finding the tree that optimizes the data, the BI method offers an alternative strategy. A simple example allows understanding it with ease. Let's assume that in a tropical pond there are diving beetles of different sizes (9 are small and 5 are large) and color patterns (8 have yellow markings and 6 are black) as in Figure 6.

If we assume that all the diving beetles have the same probability of getting caught, the probability to catch a large diving beetle will be 5/14. Now with a blindfold, what would be the probability to catch a diving beetle with yellow markings knowing that it is large? The straightforward answer is 3/5. Likewise, if we were not able to distinguish the size of a beetle but only that the beetle caught had yellow markings, what would be the probability to have caught a large one? The answer is 3/8. As a result, if we want to write these results as equations we would have:

$$P(\text{Yellow markings}|\text{Large}) = 3/5$$

and

$$P(\text{Large}|\text{Yellow markings}) = 3/8$$

These two probabilities are therefore quite different even though they might be confusing at first. The prior information that we have, markedly influences the results of the probability we are calculating. This is the basis of conditional probabilities, on which is based the theorem introduced by Thomas Bayes and used in BI:

$$P(H|D) = \frac{P(H)\times P(D|H)}{P(D)}$$

In our example, it would give:

$$P(\text{Yellow markings}|\text{Large}) = \frac{P(\text{Yellow markings})\times P(\text{Large}|\text{Yellow markings})}{P(\text{Large})}$$

$$P(\text{Yellow markings}|\text{Large}) = \frac{(8/14)\times(3/8)}{(5/14)} = 3/5$$

Bayes' theorem started to be applied to phylogenetic inference only recently (Felsenstein 1968; Huelsenbeck 2002). Its formulation is similar to the one used to calculate very basic conditional probabilities:

$$P(\tau i|X) = \frac{P(\tau i) \times P(X|\tau i)}{\sum_{j=i}^{Bs} P(X/\tau i) \times P(\tau i)}$$

$P(\tau i|X)$ is the posterior probability of a tree given the alignment X.
$P(\tau i)$ is the probability of the tree τi (often 1/Bs with Bs the total amount of trees).
$P(X|\tau i)$ is the probability to obtain the alignment X with the tree τi (maximum likelihood).
$\sum_{j=i}^{Bs} P(X/\tau i) \times P(\tau i)$ is the sum of all numerators over all possible hypotheses. Although it may seem very complex to understand, this term is a constant and is used as a scaling factor to ensure that posterior probabilities of the trees lie in an interval [0, 1].

In practice, the method requires an alignment and priors on the models of evolution applicable to the data, in order to generate a first tree and calculate its posterior probability. Theoretically, it would be compulsory to run the analysis long enough to generate all possible trees and calculate their posterior probabilities to find the tree with the highest posterior probability. However, this is most of the time impossible because the number of possible trees is a critical limitation for the analysis. As an example, the total amount of possible trees to be derived from a matrix of 50 taxa is larger than the number of atoms in the entire universe (Ronquist et al. 2009). Therefore, the posterior probability distribution of trees has to be estimated. The solution named Markov chain Monte Carlo sampling (MCMC) was introduced by Metropolis et al. (1953) and later extended by Hastings (1970). To illustrate the functioning of this Markov chain, a good schematic illustration would be a wide landscape (the entire space of parameters) with several mountain peaks of different heights (high posterior probability density regions of the parameter space; Figure 7).

The Markov chain starts from a random point in the entire landscape and makes a random move at each step of its progression in the landscape (a step=a generation). Each time the chain is moving randomly, the value of some parameters change, and a ratio of the posterior probability before and after the change is calculated to test for improvement. If the probability is improved (the chain would be climbing a mountain) then at the next generation, the chain will start from this new point. However, if the posterior probability is lower (the

chain is descending the mountain), the new state is accepted with a probability proportional to the height at which the chain is (the calculation takes also in count the asymmetry of the new distribution tested). The chain runs for several thousands to millions of generations to ensure that the space of parameters is thoroughly explored. Usually, the early phase of the run also called burn-in represent moves in low posterior probability regions of the parameter space and is therefore removed from the analysis afterwards.

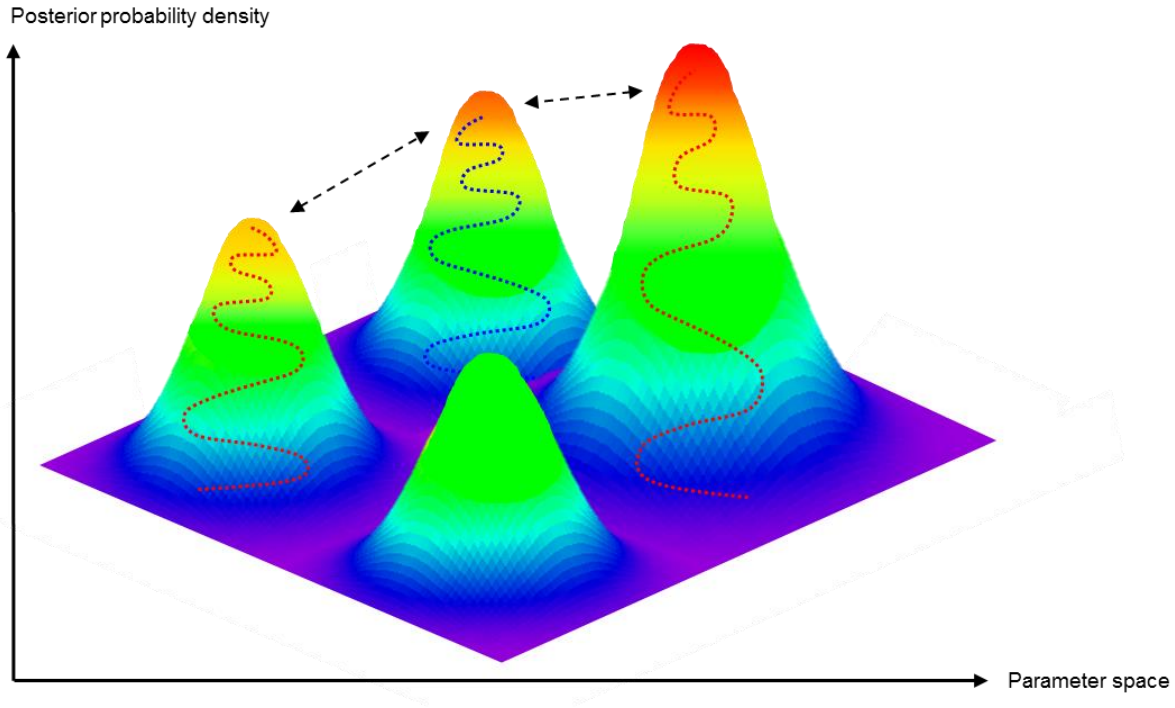


Figure 7. Schematic representation of an analysis in Bayesian inference. The multiple parameters of the analysis are represented in the x-axis even though this is a highly simplified vision of the multiple dimensional parameter space to explore. Thy y-axis illustrates the posterior probability density for the different regions of the parameter space. In purple the posterior probability is low and in red it is high. The three peaks on the left of the figure are local optima whereas the true optimum for the priors considered is shown on the right of the figure. Three Markov chains are illustrated; two hot in red and one cold in blue. The cold chain is stuck in a local optimum but has the possibility to swap states with one of the hot chains, one of which has already reached the optimal posterior probability density region of the entire parameter space (right peak).

Sometimes, a chain climbs a mountain and reaches a region of high posterior probability density, staying in this region because downhill moves are not accepted. Although this region might be the optimal posterior probability density region, it might also constitute an under optimal region to sample compared to the rest of the parameter space because there are some higher posterior probability density regions which were not sampled by the chain. Often multiple chains are used and different runs are conducted to avoid such local optimums also called tree islands (Geyer 1991). This procedure called Metropolis Coupling (Geyer

1991) implies the use of hot and cold chains. Several hot chains are exploring the parameter space faster than the cold chain because they evolve in deliberately flattened posterior probability density regions. However, hot and cold chains can swap states in order for the cold chain to accelerate its mixing over complex posterior distributions and avoid tree islands (Figure 7). Only the cold chain distribution is sampled, and all samples collected after the burn-in phase are then summarized to yield the posterior probability of each parameter. A consensus tree can be generated based on the trees collected during the stationary phase of the cold chain, and each node will have a posterior probability proportional to the number of time this node was recovered in the collected samples.

Whether molecular or morphological data should be used for phylogenetic inference is an old question which still has supporters of each side. I do not aim in this brief overview to encompass this topic especially since it relies as much on biological ground than it does on a philosophical one. It is however noteworthy to emphasize that even though molecular phylogenetics have taken over in the past decades with cheaper and more efficient sequencing methods, morphology remains a cornerstone of phylogenetics because we are not able to sequence the fossils permitting to date molecular phylogenetic trees in most of the cases. It is only based on morphological characters that we can assign with more or less certitude fossils to an extant clade. Recently, total evidence approaches in which morphological as well as molecular data are used in combination have helped to efficiently unfold the phylogenies of clades (eg. Prevosti 2010; Ronquist et al. 2012; Guil et al. 2013; Chen et al. 2014). Although morphological datasets are complex and time-consuming to generate, they provide a unique source of cross-validation for the hypotheses inferred using molecular matrices and should not be disregarded as new revolutionary models to test macroevolutionary processes relying on phylogenetic trees are being developed.

2.2 Enter the clocks: a guide to dating trees

As time brought more sophisticated techniques to infer phylogenetic trees, methods to estimate divergence times from these phylogenies underwent a parallel evolution. Empirical evidence highlighted the need to relax the initial hypothesis of molecular clock (Zuckerkandl and Pauling 1962). Indeed, many examples of clades have been proved to actually disrespect this assumption of lineages evolving at the same pace across a phylogenetic tree. As a result, a plea was made to take into account the heterogeneity in the

rate of evolution of different lineages (Thorne et al. 1998). Different models have been introduced to accommodate rate variation due to non-clock behavior (Figure 8). One of the first models introduced was the Bayesian autocorrelated relaxed clock (Thorne et al. 1998). In this method, it is assumed that closely related lineages have if not similar, close rates of evolution and that by extension, the most divergent lineages should have more different rates of evolution. As a result, each lineage has its own rate of evolution which is derived from its closest ancestor's rate. Therefore, the rates within the tree are all correlated, and the closest the lineages are, the more correlated the rates are. Although this method might perform well, it presents the major drawback that a prior needs to be fixed for the root to derive the other rates of the tree and that the degree of autocorrelation also requires a prior.

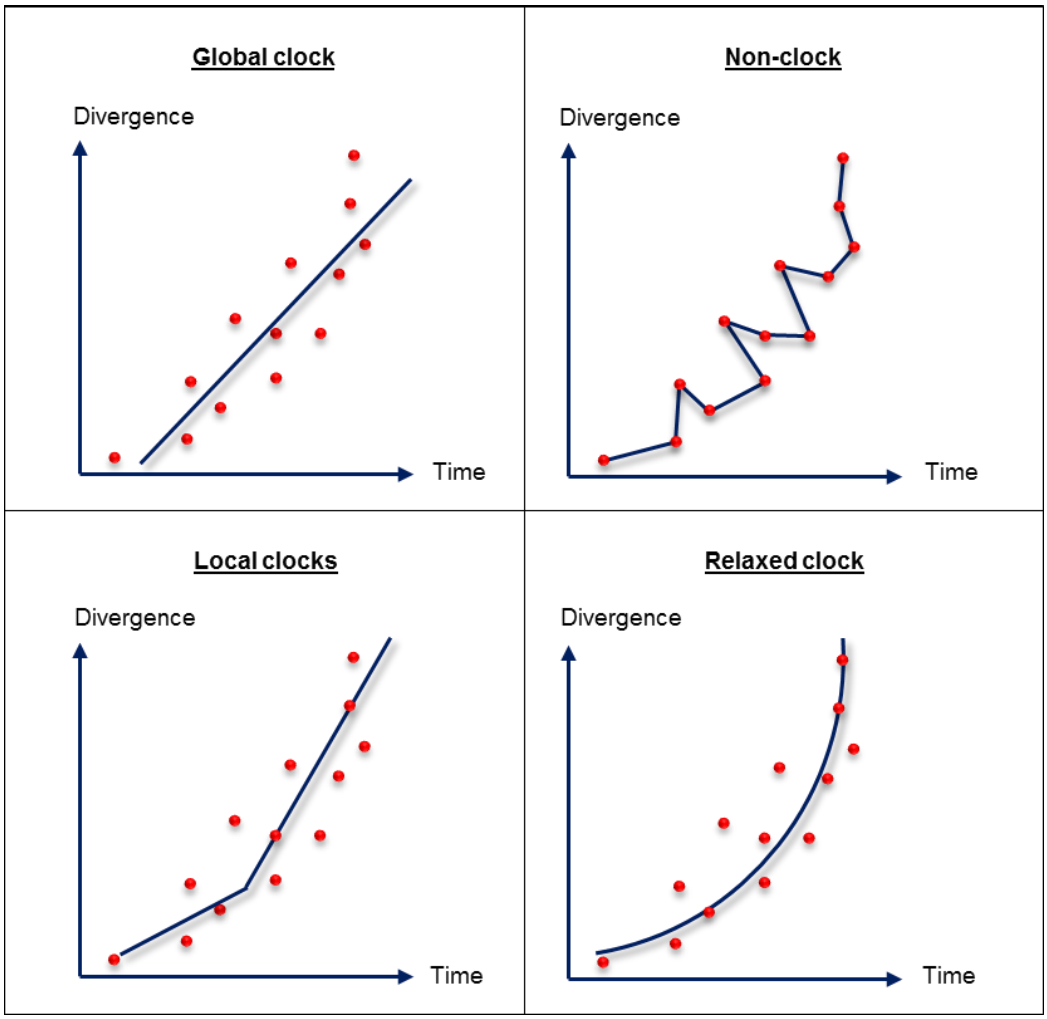


Figure 8. Schematic representation of different clock models applied to a random dataset following Pybus (2006) and Lemey and Posada (2009). Each point shows the divergence between two sequences. The different clock models are presented as black lines.

Local molecular clocks were also designed as a solution for non-clock behavior (Yoder and Yang 2000). Basically, different parameters of evolutionary rates are assigned to

subsets of a phylogenetic tree so that some collections of branches might have a different rate of evolution than the rest of the tree. Any number of local clocks can be specified, starting from one, assuming a clock behavior. However, the number of parameters linked to each local clock might be a computational burden and render complex the interpretation of time and rate distinct effects. Uncorrelated relaxed clocks are the most recently introduced method to acknowledge lineage rate heterogeneity (Drummond et al. 2006). In this method, rates can vary among lineages and across the tree without being correlated to each other. The rates of each branch are drawn from a rate distribution and therefore these clocks are an elegant alternative to autocorrelated relaxed clocks. This technique also allows using a collection of posterior topologies drawn from a MCMC analysis instead of a fixed phylogenetic tree as in the other methods. This later point is of particular interest when it comes to test hypotheses and provide credibility intervals for an evolutionary pattern because uncorrelated relaxed clocks permit the integration of phylogenetic uncertainty into the dating process.

These new methods have been proved to be very useful to provide age estimates in a robust statistical framework. However, the dating process based on clocks is only possible when we have an idea of the pace at which a given gene evolves. In some groups, a rate of evolution might be known and therefore used as a reference rate to calibrate the clock (eg. Brower 1994). However, even those calculated rates ready to be used have previously been estimated using some kind of calibration (Ho and Phillips 2009). Thus, there is a need for additional data from different sources to translate sequence divergences along the branches of phylogenetics trees into time. Broadly two main types of data have been used so far to calibrate clocks and provide age estimates; fossil and paleogeological/biogeographic data, even though the latter has received considerable criticism which are detailed below.

The fossil record contains a multitude of organisms preserved in different structures such as rocks or amber which are the ancestors of extant or extinct lineages and provide unique data on the extinctions, distributions and ages of clades. When fossils can be assigned to a given clade based on shared morphological features, they can provide a minimum age for this particular clade because the logical assumption is that the group must be at least as old as the fossil it comprises. As a result, if the geological formation (or amber) in which such fossil is encapsulated can be identified based on stratigraphic interpretation and then dated using radiometric data, it may constitute a reliable calibration point to calibrate the molecular clock (Benton and Donoghue 2007; Donoghue and Benton 2007, Benton et al. 2009).

Biogeographic data have also been advocated to calibrate molecular clocks and provide absolute age estimates. Indeed, by combining extant distributions of taxa and geological history of these distributional areas, one can derive maximum ages from lineage speciation by vicariance. Numerous examples are available in the literature of the use of such calibration points. Well-documented tectonic events implying vicariance have therefore been used to provide maximum ages on a node of taxa which are supposedly resulting from this particular event. Among these notorious biogeographic calibration points it is noteworthy to cite the Isthmus of Panama, New Caledonia or Sardinia which present different geological histories but rely on the same assumption that some of the lineages they hold are endemics, have none or low dispersal capacity and did not underwent extinction. For instance, New Caledonia has been suggested to have been entirely submerged 37 Ma (Grandcolas et al. 2008). Based on this hypothesis, one could suppose that a New Caledonian endemic found on the island today cannot possibly be older than the time of submergence. However, the taxon could have dispersed from New Caledonia to any kind of geographic refugia before the submergence and come back on the island after its re-emergence. Another scenario is that the taxon we observe today as an endemic might possibly have been a widespread taxon before running extinct in all its other ranges of distribution. As a result, its age would have nothing to do with the age of the island occupied by the remainder of this lineage. Finally, the endemic taxon observed nowadays might perfectly be the result of a recent dispersal event resulting in speciation by geographic isolation or ecological adaptation. In all cases, the biogeographic constraint used to calibrate the molecular clock would lead to fallacious ages. As a result, biogeographic calibration points are less extensively used than fossil constraints but they might provide interesting estimates for taxa with a scarce fossil record and not violating the numerous prerequisites of their application.

Placing a calibration point in a phylogeny is not a straightforward task. Until recently, it was only possible to fix the age of at least one node in the tree to match the estimated age of a fossil or a biogeographic constraint. This point calibration procedure however presents several major pitfalls such as the fallacious assumption that the constraint provides an approximate age for a given clade whereas really it only provides a minimum or a maximum age. The uncertainty on the age of the calibration point is also disregarded even though it represents a crucial parameter for the dating process. More recently, new calibration techniques allowed the placement of calibration points in the tree as minimum or maximum ages. These hard bounds allow the placement of a biogeographic/geological event or a fossil

on a node without restraining the node to have the same age as the constraint. This is a substantial improvement compared to the point calibration procedure, yet uncertainty regarding the age of the constraints and their placement remains problematic. Yang and Rannala (2006) and Drummond et al. (2006) independently introduced the concept of soft bounds, where the probability of the node age being outside of the pre-defined bound is not necessarily zero, but is modeled by a diminishing tail of probability beyond the bound. For instance, this method allows a node to have a younger age than the fossil placed on it. In order to permit this flexibility, parametric distributions are used as prior for the ages of nodes.

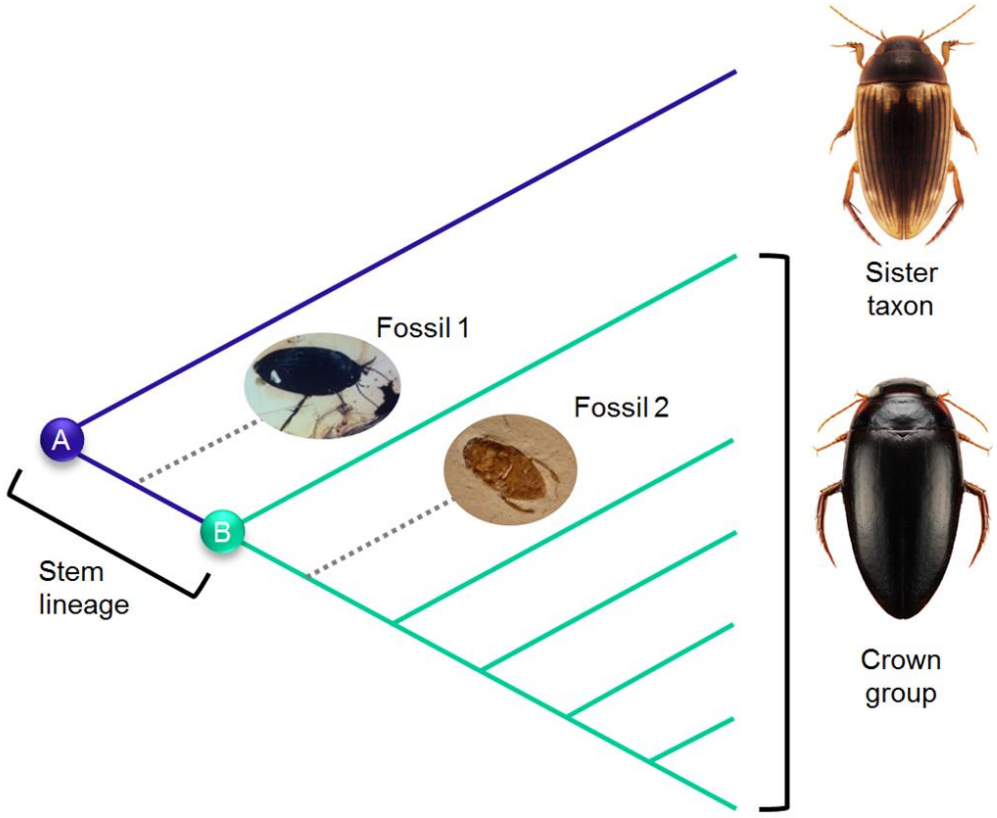


Figure 9. Representation of fossil use in molecular clock calibration. The clade rooted in A comprises all taxa including the two fossils and the crown-group. The clade rooted in B is the crown-group. The branch joining nodes A and B represents the stem lineage. Since the fossil 1 shares some synapomorphies with the crown-group, it can be used as a minimum bound for the node A. However, the fossil 2 shares all synapomorphies with the crown-group and also presents apomorphies which allow its inclusion into a particular clade of the crown-group. Therefore, the fossil 2 can be used as a minimum age for the node B. Illustrations: *Copelatus sociennus* for the sister taxon and *Exocelina* sp. for the crown-group.

In order to properly calibrate the molecular clock, different constraints can be placed either on a node from which all ancestors, extinct and extant lineages of a clade are derived, also called the crown-group, or on the branch that joins this crown-group and its sister-taxon, the stem-lineage (Figure 9). In the case of fossils, the placement of a constraint on the crown-

group requires that the fossil shares all the synapomorphies of at least one lineage of the crown-group. When such a diagnostic cannot be made or in the presence of a fossil presenting only a part of the synapomorphies of the crown-group, it cannot conservatively be placed on the latter but only on the stem-lineage (Ho and Phillips 2009). When fossil and biogeographic constraints have been correctly placed on the tree, several different hard or soft bounds can be used. In order to specify maximum or minimum soft bounds on nodes, a prior distribution should be chosen. Different prior distributions exist such as the normal, lognormal, uniform or exponential distributions which are the most commonly used in molecular dating. Except from the uniform distribution that requires no prior information in addition to the bound itself, all the other distributions cannot be specified without priors. All different distributions present alternative shapes that may render them more or less advisable depending on the type of fossil or biogeographic data considered. For instance, the shape of the lognormal distribution allows the maximum probability to be different from the bound itself. This might be relevant to account for the fact that the age of the clade is likely to be older than the oldest fossil discovered so far. However, all distributions except the uniform require the specification of different parameters a priori in order to give the distribution a particular shape. When there is no clear evidence supporting the selection of values for the different parameters, the uniform prior distribution (corresponding to a hard bound on the specified interval) might be advocated to avoid any error due to fallacious assumptions on the data used to calibrate the clock. Although cutting-edge methods have been developed to improve the power of dating based on fossil and biogeographic calibrations, one should keep in mind the different drawbacks of age estimate inference and be as thorough as possible in the search of information allowing a confident placement of the calibration points in the tree (Graur and Martin 2004; Heads 2011; Parham et al. 2012).

2.3 How trees help unfolding biogeographic and diversification patterns

With the advent of more sophisticated methods to infer and date phylogenetic inferences, new opportunities to investigate patterns and processes of biodiversity emerged. With them, models aiming at disentangling biogeographic scenarios and unfolding diversification dynamics of clades have experienced tremendous developments becoming one of the most paramount fields of evolutionary biology. Using dated phylogenetic trees, it is for instance possible to reconstruct ancestral character states such as morphological features, geographic areas or host-plant preferences.

Among all these new developments, the field of historical biogeography is probably the one that received the largest share of attention recently (Lomolino et al. 2010). If the use of phylogenetic trees to unfold geographic origins is as old as cladistics (Hennig 1966), the field has recently experienced an astonishing renewal with the advent of new methods and models where phylogenies play an essential role (Ree and Sanmartín 2009).

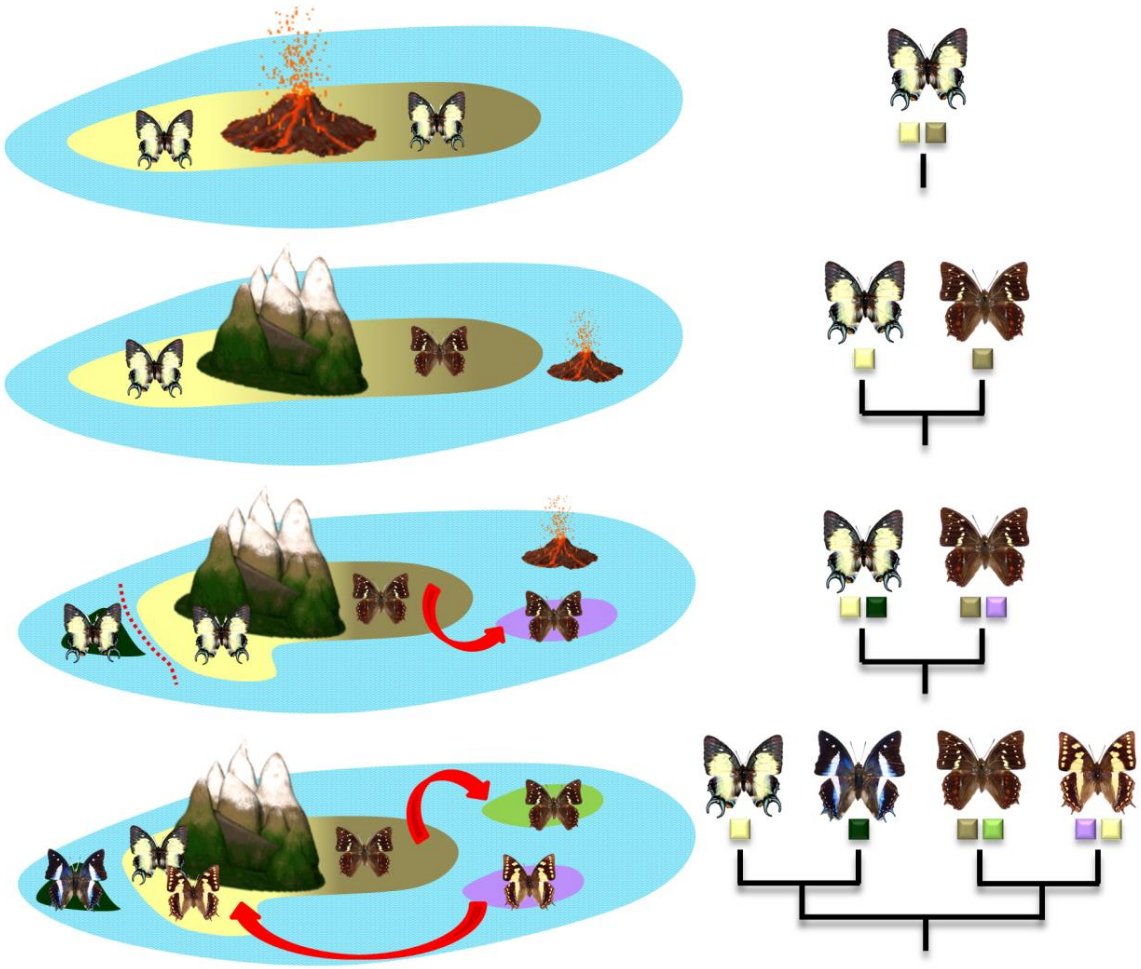
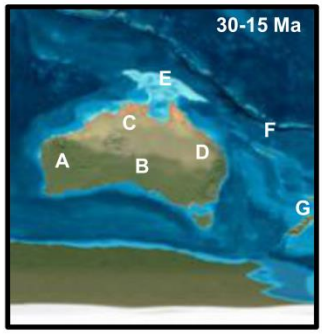


Figure 10. Biogeographic mechanisms at work. An initial taxon (yellow) present on an island is divided in two subsets by volcanism. Through geographic isolation, a vicariant species evolves (brown) whilst there is an intense volcanic activity offshore. This activity results in the formation of a new island where some populations of the brown taxon disperse actively resulting in range expansion. On the other side of the main island, tectonic activities foster the fragmentation of the yellow taxon distribution area allowing range expansion via vicariance. Both vicariant populations evolve in new taxa, on the fragment of the main island (black and blue) and on the island offshore (brown and yellow). The latter, via long-distance dispersal reaches the range of the yellow taxon whereas a part of the brown taxon colonizes a new offshore island east of the main island. The cladograms on the right illustrate the areas of distribution of each taxon during the different stages of evolution. Biogeographers have only access to the last cladogram that comprises extant taxa and aim at unveiling the origin of the entire clade using chronograms, models of vicariance/dispersal and paleotectonic and paleoclimatic data.

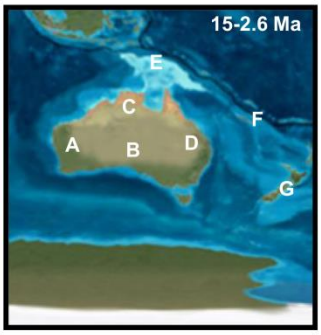
The discipline of historical biogeography aims at reconstructing the geographic origin of extant clades starting from phylogenetic relationships and present-day distribution of taxa. In an evolutionary perspective, geographic distributions of extant taxa are the result of dispersal and vicariant events on one hand, and the balance between speciation and extinction on the other. Unknotting the contribution of these events to the patterns of geographic distribution we observe today is the central goal of historical biogeography (Wiens and Donoghue 2004). Whilst plate tectonics and volcanism may trigger the apparition of aquatic or land barriers enabling vicariance and speciation by geographic isolation, dispersal might allow the colonization of new areas. Geological dynamics can also trigger connectivity between areas that were separated before, therefore allowing range expansion (Figure 10). For a long period of time, these processes where only deduced from the comparison of geographic area cladograms and phylogenies. In the past decades, new models have been developed to take advantage of the statistical framework provided by dated molecular phylogenies, in order to test alternative biogeographic scenarios and decipher mechanisms of range contraction and extension (Ronquist 1997; Ree et al. 2005; Ree and Smith 2008).

As for phylogenetic inference, parsimony and likelihood-based methods have been proposed to reconstruct character states at each node of a phylogenetic tree. The Dispersal Vicariance Analysis (DIVA, Ronquist 1997) is the one of the most notorious methods to investigate historical biogeography patterns using a parsimony-based algorithm. Based on a phylogenetic tree and distribution ranges of all extant taxa sampled in the tree, DIVA is searching for the most parsimonious biogeographic scenario to explain these extant distributions by minimizing dispersal and local extinction events. Phylogenetic uncertainty can be accounted using post-burnin tree samples from Bayesian phylogenetic inference analyses (Nylander et al. 2008). However, DIVA presents the major drawback of clearly favoring vicariance over dispersal therefore including a bias in the reconstruction of the biogeographic scenarios. Moreover, DIVA does not take into account branch lengths nor allow the inclusion of paleogeological and paleoclimatic data that could considerably improve both the realism and the accuracy of the model. More recently, the DEC model was proposed by Ree (2005) and Ree and Smith (2008). This model implemented in the software Lagrange (Ree and Smith 2008) is likelihood-based and allows reconstructing in a time-continuous framework a biogeographic scenario given a dated phylogenetic tree and dispersal rate matrices specified as a prior (Figure 11). Paleogeological, paleogeographic and paleoclimatic data but also taxa dispersal abilities are used to specify dispersal rates between areas at

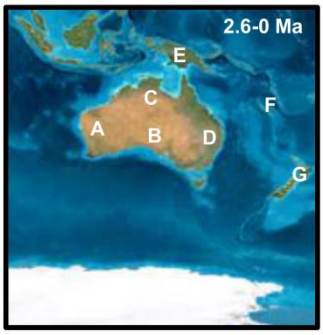
different moments of the evolution of the clade. Areas which are distant to each other at a certain point of time will have small rates of dispersal between each other, whereas connected areas will have greater dispersal rates (Figure 11). In the DEC model, dispersal, vicariant or local extinction events are allowed to happen with a probability proportional to the specified dispersal rates and topology branch lengths (Ree 2005; Ree and Smith 2008).



	A	B	C	D	E	F	G
A	-	0.75	0.75	0.5	0.01	0.01	0.01
B		-	0.75	0.75	0.01	0.01	0.01
C			-	0.75	0.01	0.01	0.01
D				-	0.01	0.01	0.01
E					-	0.01	0.01
F						-	0.01
G							-



	A	B	C	D	E	F	G
A	-	0.5	0.25	0.125	0.01	0.01	0.01
B		-	0.5	0.5	0.01	0.01	0.01
C			-	0.5	0.01	0.01	0.01
D				-	0.01	0.01	0.01
E					-	0.01	0.01
F						-	0.01
G							-



	A	B	C	D	E	F	G
A	-	0.25	0.125	0.01	0.01	0.01	0.01
B		-	0.25	0.5	0.01	0.01	0.01
C			-	0.5	0.5	0.01	0.01
D				-	0.125	0.01	0.01
E					-	0.125	0.01
F						-	0.01
G							-

Figure 11. Dispersal rate matrices for the Indomalayan / Australasian archipelago during the past 30 million years. Based on paleogeological and paleoclimatic reconstructions, the rates between pre-defined areas where the taxa occur are specified. Between 30 Ma and 15 Ma, most of the Indonesian archipelago is inexistent and there is supposedly no stable connection between the Australian and Indomalayan regions therefore dispersal rates are extremely reduced and only long-distance dispersal with a small rate are allowed. Between 15 Ma and 2.6 Ma, Australia is moving northwards and the tectonic activity in the archipelago leads to the emergence of islands whilst considerably bringing closer pre-existing landmasses. As a result the connectivity in the entire archipelago is enhanced and rates are greater. In the last time slice (2.5-0 Ma), most of the archipelago present-day landmasses exist and Quaternary climate changes trigger connectivity as sea-level fluctuates. New Guinea is fully emerged and its orogeny is massive allowing dispersal from Australia through it towards the rest of the region. Between 15 and the present Australia has become increasingly more arid and therefore the dispersal rates within the island have been reduced in the two last time slices.

Interestingly, as more complex and explicit biogeographic models are being developed, a strong focus has been made to establish a more integrative approach when it comes to understand the origin and evolution of biodiversity. As a result, the parallel field of diversification rate analysis has recently started to incorporate historical biogeography inference to investigate to which extent the biogeographic history of a clade can explain its diversification dynamics (Goldberg et al. 2011).

Diversification dynamics, the balance between speciation and extinction, is another field of evolutionary biology that greatly flourished along with molecular phylogenetics and dating. It arguably constitutes one of the most striking explosions of increasingly performing and complex models in any other field of macroevolution in the past 10 years (Paradis 2004; Rabosky 2006; Ricklefs 2007; Maddison et al. 2007; Harmon et al. 2008; Alfaro et al. 2009; Kembel et al. 2010; Goldberg et al. 2011; Morlon et al. 2011; Stadler 2011, 2013; Etienne et al. 2012; FitzJohn 2012; Condamine et al. 2013; see Pyron and Burbrink 2013; Morlon 2014 for recent reviews). Time-calibrated phylogenies provide an essential starting-point for these models to better reveal the tempo and mode of species diversification over geological times. When comparing reconstructed empirical phylogenies with phylogenies expected under certain models of diversification it is possible to unveil diversification dynamics of a clade. Starting from extremely simple models only implying speciation (pure birth models) towards much more complex models integrating extinction (birth-death models), a wide array of diversification models has been introduced to fit empirical data. Some models imply that all lineages have the same diversification rate and that the diversification is a constant process whilst more complex models allow relaxing these assumptions and therefore allow diversification rates to vary with time and among lineages (Morlon 2014). The latest models developed allow calculating the impact of biotic (e.g. competition, predation) and abiotic factors (e.g. climate, geography) on diversification rates (Goldberg et al. 2011; Condamine et al. 2013). The accumulation of lineages is also thought to be a potential trigger for diversification variation as a response to niche and geographic space filling for instance. As a result, models to test the impact of clade diversity on the diversification of the clade itself have also been developed (Etienne et al. 2012). Additional models accounting for character-dependent or clade-specific diversification have also been proposed. Overall, a drastically increasing set of models is made available to fit dated empirical phylogenetic trees in order to decipher the patterns and understand the processes of diversity dynamics through time (Pyron and Burbrink 2013; Morlon 2014).

GENERAL INTRODUCTION

Chapter 3. The Indomalayan / Australasian archipelago

Scattered tropical islands in the Indonesian archipelago

"From a look at a globe or a map of the Eastern hemisphere, we shall perceive between Asia and Australia a number of large and small islands forming a connected group distinct from those great masses of land, and having little connection with either of them. Situated upon the Equator, and bathed by the tepid water of the great tropical oceans, this region enjoys a climate more uniformly hot and moist than almost any other part of the globe, and teems with natural productions which are elsewhere unknown. The richest of fruits and the most precious of spices are Indigenous here. It produces the giant flowers of the Rafflesia, the great green-winged Ornithoptera (princes among the butterfly tribes), the man-like Orangutan, and the gorgeous Birds of Paradise."

Alfred Wallace, *The Malay Archipelago*, 1869.

Chapter contents

3.1 Glide over an intricate geological puzzle 40

3.2 From murky depths to sky islands: Diving beetles as an evolutionary model 43

3.3 On the study of tropical butterflies to unfold biogeographic scenarios 45

3.1 Glide over an intricate geological puzzle

The Indomalayan / Australasian archipelago is one of the most fascinating and diverse regions on the planet. This diversity is also highly threatened, as illustrated by the presence of no less than eleven biodiversity hotspots, but luckily still one out of only three of the world's last tropical wilderness areas (New Guinea) and a stunning assemblage of smaller more or less primary habitats distributed around the Equator (Mittermeier et al. 2004; Williams et al. 2011; Figure 12). The region has fascinated scientists ever since Wallace who made of it his very own laboratory to understand evolutionary processes (1860, 1863). With a highly complex geological history, it offers a good opportunity to study the origins and diversification of its biota through time and space and make predictions in the light of change induced by man (Hall 2002, 2011, 2012; Lohman et al. 2011; Metcalfe 2011).

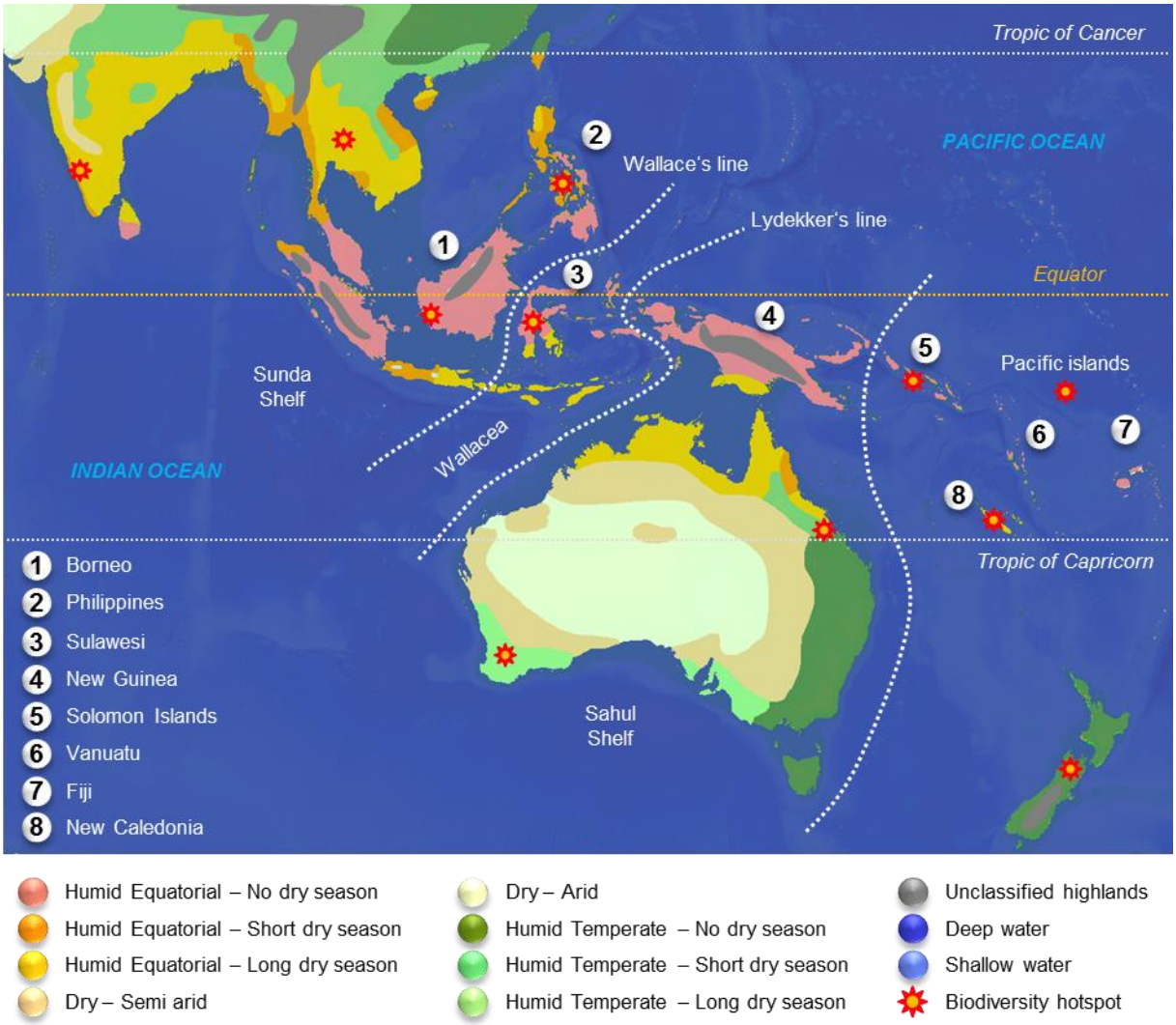


Figure 12. Map of the Indomalayan / Australasian archipelago featuring climatic zonation, bathymetry, major faunal boundaries and biodiversity hotspots.

The present-day geography of the archipelago is the result of complex movements of the Australian, Eurasian, Indian and Pacific plates and related geological events such as orogenies, volcanism and drift of terranes. Many areas are geologically young and some of complex composite origin such as Sulawesi and New Guinea (Hall 2002, 2011, 2012). The supercontinent Gondwana started to breakup in the Early Jurassic (≈ 180 Ma) with South America and Africa being separated from the remaining main landmasses (Antarctica, Australia, India). During the Early Cretaceous (≈ 100 Ma), India pulled away from the Antarctica-Australia block and initiated a northwards drift towards Asia. It is only much later that Australia completely detached from Antarctica, approximately during the Cenozoic (≈ 45 Ma) to move rapidly northwards and contribute to the paramount tectonic activity that yielded the Indomalayan / Australasian puzzling archipelago.

However, Gondwanean fragments started to dock on the Asian margin as soon as the Early Cretaceous roughly 110 Ma when a part of present-day Borneo docked on the Sunda margin (Hall 2012). A part of Java and Sulawesi present-day landmasses were also rather close from the Southeast Asian margin at that time, finally docking about 90 Ma. Simultaneously, the subduction beneath the Sunda arc temporarily ended allowing it to become an emergent continental region. Meanwhile, India and Australia continued to move northwards at different paces, with New Guinea and the Lesser Sunda Islands being attached to the latter. Whilst India started to drift faster in the late Cretaceous (75 Ma), Australia remained attached to Antarctica for a while. The collision of Gondwanean fragments with the Sunda margin permitted the extension of the margin as well as the formation of Borneo, Java and West Sulawesi between the Late Cretaceous and the Early Paleogene (90–60 Ma). Whilst India started to collide with Asia in the Early Paleogene (≈ 60–50 Ma), the rapid drift of Australia northwards provoked subduction at the Sunda margin (Figure 13). Arc material drifting westwards from the Pacific accreted on East Borneo, North Sulawesi and North New Guinea as soon as the mid-Paleogene (≈ 45 Ma). At the same time, West Sulawesi detached from the Sunda margin by rifting, resulting in the Makassar Strait separating it from Borneo. During the last 45 Myr, Australia continued to move northwards, with the Western part of the fragment getting closer from the Eastern part of the Sunda margin. The Sula Spur formed by the Lesser Sunda Islands attached to Western New Guinea collided with Western Sulawesi in the last 15 Myr. Meanwhile, highly complex arc material continued to drift westwards from the Pacific allowing the formation of the Philippines, North-East Borneo and North Sulawesi. Thus, present-day Sulawesi also called the anomalous island by Wallace (1880) is a complex

puzzle of Sunda margin material, drifting Pacific arcs and Gondwanean Sula Spur elements (Figure 13). As a result, Sulawesi is the geographic and geological core of the entire archipelago, a crucial junction between the Australian and Asian landmasses. During the past 10 Myr, a massive orogeny allowed the formation of present-day New Guinea with the accretion of arcs drifting southwards and the collision of the Pacific and Australian plates. At the same time, the Sula Spur reached the Sunda margin, engendering arc formation in the West (Bali, Lombok, Sumba, Timor) and yielding the present configuration of Moluccas between Sulawesi and New Guinea (Figure 13).

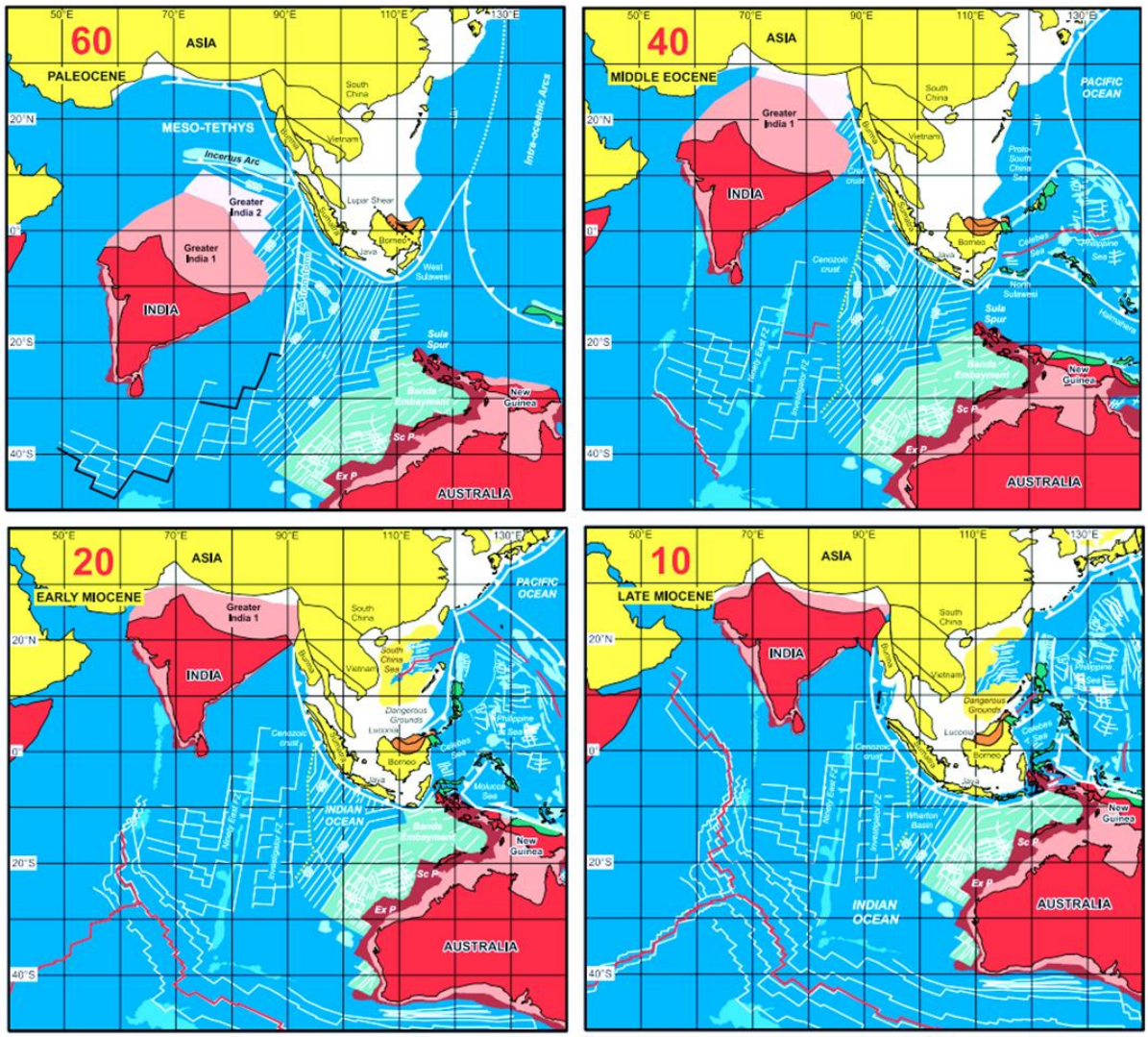


Figure 13. Paleogeographic reconstructions of the Indomalayan / Australasian archipelago during the Cenozoic era (Hall 2012). Color of landmasses: red is Gondwanean, yellow is Laurasian (North America and Eurasia) and green is arc, ophiolite or accreted material.

The role of abiotic factors linked to the region climatic and geological histories as potential major driving forces for biological diversification is not well understood and it is only recently that larger empirical studies started to unveil some of the underlying

mechanisms (Lohman et al. 2011). Nevertheless, some valuable insights have been brought by recent studies showing that Pleistocene glaciations had significant impacts on terrestrial and freshwater taxa dispersion across the Sunda Shelf and other archipelagos such as the Philippines, likely due to an enhanced connectivity of landmasses begotten by deep sea level fluctuations (de Bruyn et al. 2004, 2013; Azuma et al. 2006; de Bruyn and Mather 2007; Su et al. 2007; Klaus et al. 2013). In addition to the role of glaciations in the current pattern of distribution in South-East Asia, several studies support the paramount role of Wallacea and Borneo as cradles of diversity despite many examples of Wallace's line transgression (invertebrate studies: Balke et al. 2009; Müller et al. 2010; von Rintelen et al. 2011; Klaus et al. 2013). Although major progress has been made to understand the processes involved in the formation and diversification of the South-East Asian biota, there is a need for more comprehensive studies to investigate interactions of geology and past climate changes on the origins and evolution of the striking biodiversity indentured to the archipelago. It is especially true when focusing on mostly unresolved matters such as the origin of taxa on Sulawesi (Lohman et al. 2011; Stelbrink et al. 2012), the Moluccas, the Lesser Sunda Islands, New Guinea, or the evolution of biota prior and after the connection between Asia and Australasia.

3.2 From murky depths to sky islands: Diving beetles as an evolutionary model

There are about 4,300 species of Dytiscidae, the most species-rich water beetle family on Earth (Nilsson 2013). Diving beetles are well adapted to aquatic habitats and have a global distribution except for Antarctica. They are spread from Greenland to Tierra del Fuego (Balke 2005; Jäch and Balke 2008; Figure 14). Most of their diversity is concentrated around the tropics where high levels of endemism have been unveiled (Balke 2005; Jäch and Balke 2008). They spend most of their life as adults and larvae in a wide range of aquatic environments such as swamps, pools, lakes, rivers, streams but also bromeliad water tanks or underground aquifers (Balke 2005, 2008; Jäch and Balke 2008; Leys et al. 2003, 2008). Dytiscids also occur on a wide altitudinal gradient from a few meters above sea-level up to 5000m-high mountain summits in Peru and Tibet (Balke 2005). Although diving beetles are among the best known beetle groups, their higher-level phylogenetic relationships remain contentious to some degree despite recent larger scale analyses (Ribera et al. 2008; Miller and Bergsten 2014). However, because of their wide distribution, association with aquatic habitats and thus ease to collect in order to create comprehensively sampled datasets, they have become an interesting model to test evolutionary hypotheses.

Figure 14. From top to bottom and left to right: predaceous diving beetle larva also called water tiger consuming a mayfly larva; *Hydaticus pacificus*-group species (Dytiscinae, Hydaticini); group of *Cybister chinensis* (Dytiscinae, Cybistrini) devouring the carcass of a fish; *Platynectes chujoi* (Agabinae, Agabini).

Across the Indomalayan / Australasian archipelago, there are about 700 species of Dytiscidae although many of them are undescribed (Nilsson 2013). They occur in the most remote micro-archipelagos and their diversity might be unexpectedly high in islands such as New Guinea or the Solomons where only a few expeditions have been conducted to date. Recently, most studies on diving beetles of the region have employed an integrative approach combining morphological and molecular data, in some cases also ecological niche modelling. The integration of multiple kinds of data allows not only a sustainable assessment of the fauna, but also permits to decipher in a more thorough fashion the patterns and processes responsible for the observed diversity of these insects (Balke et al. 2007, 2009; Hendrich et al. 2010; Hawlitschek et al. 2011). During this PhD, I have continued to work to develop diving beetles as a proxy to study evolution across the Indomalayan / Australasian archipelago.

3.3 On the study of tropical butterflies to unfold biogeographic scenarios

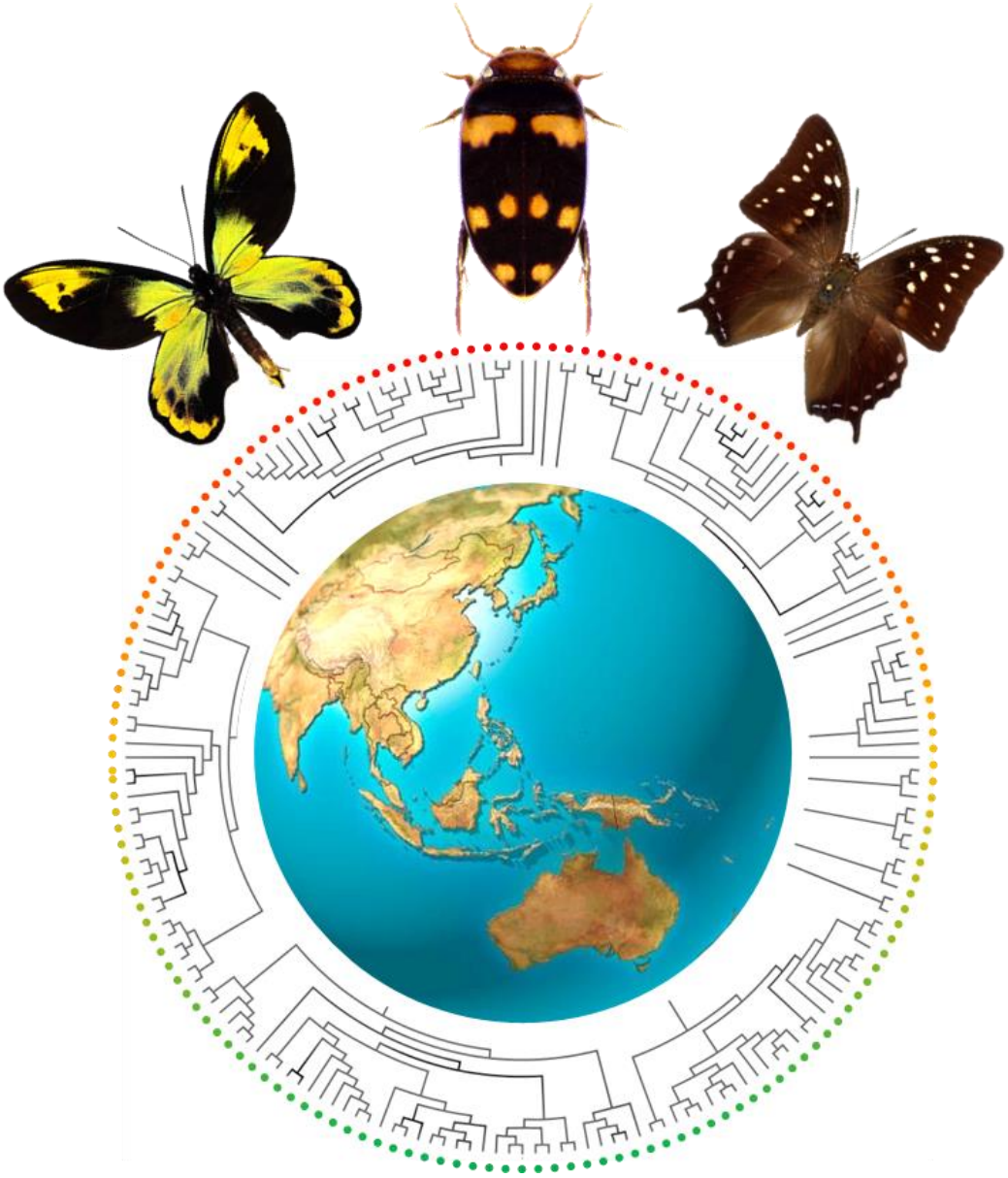
Among butterflies, two of the most charismatic families are the swallowtail and brush-footed butterflies (Papilionidae, 550 species and Nymphalidae, 6600 species). These remarkable insects are spread all over the globe except for Antarctica. The Indomalayan / Australasian archipelago holds without a doubt some of the most ornamental representatives of each family such as the birdwings (Papilionidae, *Ornithoptera* / *Trogonoptera* / *Troides*), peacock swallowtails (Papilionidae, *Papilio* subgenus *Achillides*), swordtails (Papilionidae, *Graphium*), Jezebels (Nymphalidae, *Delias*), Emperors and Nawabs (Nymphalidae, *Charaxes*) or Oakleafs (Nymphalidae, *Kallima*) (Figure 15).

Figure 15. From top to bottom and from left to right: *Trogonoptera brookiana* (Papilionidae, Troidini), *Papilio* (*Achillides*) *buddha* (Papilionidae, Papilionini), *Charaxes* (*Polyura*) *nepenthes* (Nymphalidae, Charaxini), *Ornithoptera priamus* (Papilionidae, Troidini).

Butterflies have been extensively used in studies relying on molecular phylogenies in the archipelago to untangle biogeographic scenarios and diversification processes. They are often excellent dispersers but heavily depend on their host plants and therefore provide a unique opportunity to study the impact of geological and especially climatic disruptions on patterns of distribution. Recent comprehensive phylogenies of South-East Asian butterflies helped to bring new insights into our understanding of lineage evolution in the region, already challenging some old paradigms. For instance, whilst a southwards colonization of the archipelago from the Palearctic was suggested for long as the most likely scenario to explain the present distribution of several groups (Kitching 1981; New 1999; Kondo et al. 2003), an "out-of-Australia" hypothesis was more recently put forward, illustrating cases of evolution from ancient Gondwanan stocks, and underpinning the complexity and variety of evolutionary histories among butterfly radiations of the region (de Jong 2004; Braby et al. 2005; Braby and Pierce 2007; Müller et al. 2013). It has been shown that the role of Wallacea and its famous biogeographic barriers to the east and west have been quite different among these insects, acting as putative impediment in some groups (Müller and Beheregaray 2010) whilst being of limited importance in others (Müller et al. 2010). These studies have therefore provided evidence that these biogeographic lines are permeable. As a result, butterfly clades can help to disentangle evolutionary patterns and processes. However, there is still a need for additional studies to shed light on the evolution of local endemism and the role of geological setting on diversification dynamics of butterflies.

GENERAL INTRODUCTION

Chapter 4. Objective of the thesis



Molecular phylogenies and reconstructions of past climate and geography help unknotting lineage evolutionary history

"*Nothing has such power to broaden the mind as the ability to investigate systematically and truly all that comes under thy observation in life.*"

Marcus Aurelius, *Meditations*, 161-180.

Discovering which factors govern lineage diversification through time is of paramount importance in evolutionary biology. Using integrative approaches combining a modern taxonomy coupled with DNA sequence data, molecular phylogenetics and model-based methods to infer population dynamics, historical biogeography and diversification dynamics, it is nowadays possible to unveil evolutionary mechanisms shaping biodiversity patterns. In this thesis, I aimed at unfolding evolutionary patterns of several groups of insects in the Indomalayan / Australasian archipelago in order to decipher their mechanisms of lineage diversification. In order to accomplish this objective, I studied different groups of beetles and butterflies of various taxonomic ranks, species richness and geographic distributions. I have chosen to divide the thesis in three main parts and six chapters.

The first part focuses mainly on species delimitation and description of new taxa using multiple lines of evidence with a strong focus on the utility of molecular data for this purpose. This part is divided in two chapters respectively introducing a molecular framework to study species boundaries and descriptions of new taxa to science using an integrative approach. The main objective is to underscore the interest of integrating DNA barcoding and molecular phylogenetics into species discovery and description to enhance biodiversity assessment and develop a global methodological framework to improve the consistency of biodiversity description in the future.

Although species description benefit from integrative approaches allowing the inclusion of additional characters such as nucleotide or amino acid positions contained in DNA strands, these molecular tools can also help to gain new insights to better understand how species originate and evolve. In Part 2, I study the mechanisms of speciation and clade diversification in island systems. The Australasian / Indomalayan archipelago holds a remarkable diversity of islands distributed from the Indomalayan peninsula towards the Pacific islands (e.g. Fiji, Vanuatu) and Australia. Within these myriads of islands, several are of particular interest because of their geological history. Some are of old continental origin such as Australia whereas some others are of much more complex origin resulting from puzzling geological assemblages as in the case of the Philippines or New Guinea for instance. I focus in a first chapter on the continental-sized island of Australia. This tremendous landmass is one of the largest Gondwanean fragment and its present position in the archipelago is the result of a slow drift from the South of the globe towards Asia triggered by

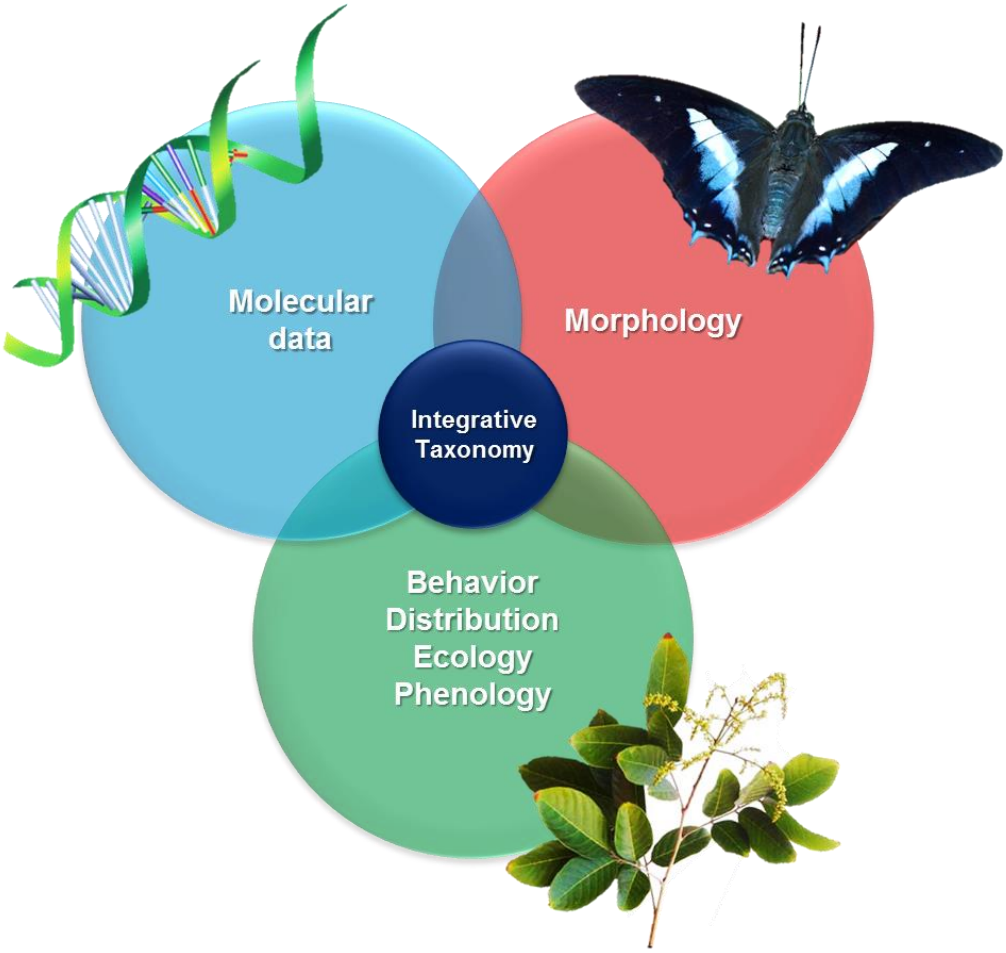
plate tectonics. If Australia has remained very stable during the past millions of years, its climate has dramatically changed through time, especially since the mid-Miocene (≈15 Ma). At that time, a global aridification still escalating nowadays started to deeply modify the ecosystems of the island. In order to decipher the impact of this climate change on the Australian biota, we study at two different periods of time the diversification dynamics of diving beetle clades. We use molecular phylogenetics, ecological niche modelling, historical biogeography, phylogeography and diversification rate analyses to unveil evolutionary patterns and understand the underlying mechanisms of clade evolution. In a second chapter, we apply the same set of cutting-edge analyses to the geologically very young and highly puzzling island of New Guinea to unravel the evolutionary history of two clades of diving beetles. We investigate the impact of recent glaciation cycles on the demographic history and associated cryptic speciation of a widespread lineage in the summits of New Guinea also referred to as sky islands. We also study the impact of the massive orogeny that built present-day New Guinea on the diversification of a young radiation resulting from an out-of-Australia dispersal event in the Miocene.

In a third part, I focus on biographic patterns and processes in the Australasian / Indomalayan archipelago. In a first chapter, we investigate the permeability of biogeographic barriers using wingless weevils spread on both parts of Wallace's line as a model. In a second chapter, we study the biogeographic history of a widespread clade of swallowtails indentured to the archipelago. In these two chapters, we combine molecular phylogenetics with fine-scale biogeographic reconstruction models to unfold global biogeographic patterns and suggest possible triggers for these patterns using paleogeological and paleoclimatic information.

Eventually, I underline all the outcomes of this Ph.D. in order to highlight several general insights allowing a better understanding of the origin and evolution of biodiversity not only in the Australasian / Indomalayan archipelago but also at a global scale. I also introduce future prospects to enhance our knowledge on the processes governing the wax and wane of lineages through the confluence of cutting-edge methodologies and the tremendous growth of molecular data availability.

PART 1: ON THE DISCOVERY OF NEW SPECIES IN THE ARCHIPELAGO

Chapter 5. A plea for an integrative taxonomy



Characters from different lines of evidence allow a more accurate and reliable estimation of species boundaries

"*The first step in wisdom is to know the things themselves; this notion consists in having a true idea of the objects; objects are distinguished and known by classifying them methodically and giving them appropriate names. Therefore, classification and name-giving will be the foundation of our science.*"

Carl von Linné, *Systema Naturae*, 1735.

Chapter contents

5.1 Paper I - Suggestions to accelerate biodiversity assessments 51

PAPER I

5.1 Suggestions for a molecular biodiversity assessment of South East Asian freshwater invertebrates. Lessons from the megadiverse beetles (Coleoptera)

Citation: Balke M, Hendrich L., Toussaint EFA, Zhou X, von Rintelen T, de Bruyn M (2013)Suggestions for a molecular biodiversity assessment of South East Asian freshwater invertebrates. Lessons from the megadiverse beetles (Coleoptera). *Journal of Limnology*, 72(2):61-68.

J. Limnol., 2013; 72(s2): 61-68
DOI: 10.4081/jlimnol.2013.s2.e4

Suggestions for a molecular biodiversity assessment of South East Asian freshwater invertebrates. Lessons from the megadiverse beetles (Coleoptera)

Michael BALKE,[1,2*] Lars HENDRICH,[1] Emmanuel F.A. TOUSSAINT,[1,2] Xin ZHOU,[3] Thomas von RINTELEN,[4] Mark DE BRUYN[5]

[1]Zoological State Collection, Münchhausenstrasse 21, 81247 Munich; [2]GeoBioCenter, Ludwig-Maximilians-University, Geschwister-Scholl-Platz 1, Munich, Germany; [3]China National GeneBank, BGI-Shenzhen, 518083 Shenzhen, China; [4]Museum für Naturkunde, Leibniz-Institut für Evolutions und Biodiversitätsforschung an der Humboldt-Universität zu Berlin, Invalidenstrasse 43, 10115 Berlin, Germany; [5]Molecular Ecology and Fisheries Genetics Laboratory, School of Biological Sciences, Environment Centre for Wales, University of Bangor, Deiniol Road, LL57 2UW Bangor, UK
*Corresponding author: Coleoptera-ZSM@zsm.mwn.de

ABSTRACT

The beginning of the 21th century was characterised by an unprecedented human-mediated loss of biodiversity, with an astonishing number of undescribed species disappearing from the Earth. To counter this major erosion of biodiversity, we need to describe and monitor what we want to preserve. Unfortunately, the velocity of deforestation, ecosystem degradation and the escalating threat to the last wilderness areas on the planet overwhelm traditional taxonomists in their bid to describe all of Earth's biodiversity. Based on empirical studies on weevils and diving beetles (Coleoptera), we show that biodiversity assessments based on cox1 *DNA sequence data deliver comparably accurate estimates of species diversity, even using a simple clustering method with a preset threshold. The method works best for large datasets, where lineage idiosyncratic errors such as species lumping or splitting compensate each other.* Cox1 *clusters cannot be translated into formal species* per se*, but can help taxonomists to accelerate their work. We suggest that large-scale sequencing campaigns for the Asian freshwater fauna will reveal patterns relevant for conservation priority setting, and enhance our understanding of macroevolutionary processes that have shaped current biodiversity in the region. Along with next generation sequencing approaches, we also suggest that our understanding of alpha taxonomy will benefit.*

Key words: South East Asia, macrozoobenthos, monitoring and inventories, DNA barcoding, next generation sequencing.

Received: January 2013. Accepted: June 2013.

INTRODUCTION

This paper was presented as a keynote at the *Freshwater Invertebrates of Southeast Asia: Biodiversity and Origin* (FISA) workshop (Maha Sarakham, Thailand, 2012) by M. Balke and has subsequently been modified to address issues raised by different speakers at the workshop.

With the dramatic increase of biodiversity erosion during the past decades, one of the greatest challenges in biodiversity research is biodiversity. How many species are there? How to best identify already described species? These are questions immediately relevant to paramount topics such as conservation of biodiversity, pest control and ultimately human welfare (Dincă *et al.*, 2011). However, with millions of undescribed species on Earth (Mora *et al.*, 2011), and most of the *ca*. 1,500,000 already described species (Wilson, 1988) rather difficult to identify, there is a need for reliable and fast tools to measure and describe this biodiversity (House of Lords, 2008). The common currency for biologists is the species entity, and for easiest communication and comparison of data, species require a formally assigned name based on one physically available voucher specimen, the holotype. In a series of publications, Hebert *et al.* (2003a, 2003b, 2004a, 2004b) advocated the use of a *DNA barcode*, a term coined to describe the use of a standard DNA fragment to build a truly global species database linking DNA sequence data with actual species names and thus morphology, ethology and *e.g.* species ecology. They suggested using a 658 bp fragment of the mitochondrial gene encoding for cytochrome *c* oxidase I (COI) (*cox1*). This marker can usually be amplified using universal primers and is informative at the population and species level (Hebert *et al.*, 2003a, 2003b, 2004a, 2004b), but this can also often be, but not always, for group species according to their higher classification (Hendrich *et al.*, 2010; Rach *et al.*, 2008). In the past decade, and in the wake of the Consortium for the Barcode of Life (CBOL) (http://barcoding.si.edu), more than 2 million specimens representing about 180,000 species have been sequenced and added to a growing central database [*Barcode Of Life Data systems* (BOLD)] (http://www.boldsystems.org). With this online library of DNA data, it is possible to use an identification interface to paste a query sequence and explore if there are close matches (Ratnasingham and Hebert, 2007). Species in this database typically have their own page, with se-

quence data and most notably voucher information, distribution maps and voucher photographs. While this system may not be perfect, it does illustrate very well how taxonomic data can be implemented and managed.

Many studies have shown that this DNA barcoding fragment can successfully be used to identify species (once a reference database has been created) (Janzen *et al.*, 2009; Dincă *et al.*, 2011), and potentially detect overlooked species even in well-studied faunas (Herbert 2004a, 2004b; Hausmann *et al.*, 2011). It was suggested that the genetic divergence among species is usually larger than within species, which should allow for species identification/delineation using simple preset divergence thresholds (Hebert *et al.*, 2003b), or using more elaborate methods employing statistical techniques to describe population genetic and speciation processes (Monaghan *et al.,* 2005, 2009; Pons *et al.*, 2006). However, several studies urge for caution (Meier *et al.*, 2006), as identification success might be low in certain cases, especially when moving from local to regional sampling levels where increasing numbers of closely related species are included (Bergsten *et al.*, 2012b; Hendrich *et al.*, 2010), or where interspecific DNA divergence levels between some species of one genus can be less than within other species of the same genus (Hendrich *et al.*, 2009). In young species, the genetic signal of *cox1* can be unstructured between species (Hendrich *et al.*, 2010), only allowing for delineation of a species-complex rather than individual species.

Taking these potential pitfalls into account (Moritz and Cicero, 2004; Will *et al.*, 2005), however, DNA sequence data can be a useful addition to the biologist's toolkit (Hebert, 2004a; Goldstein and DeSalle, 2011). An example of such an approach is the Barcoding Fauna Bavarica (Germany) project, which assembled data for more than 10,000 animal species, *all identified by experts*, and all vouchered (www.faunabavarica.de). This was possible because existing taxonomic knowledge was good, and experts for a range of taxa were relatively abundant. Even among the supposedly best-characterised species, overlooked sibling species were identified [*e.g.* documented by Bergsten *et al.* (2012a)], and taxonomic revision for the whole dataset has recently begun.

In most tropical regions, however, the situation is more challenging, mainly due to a lack of taxonomic expertise and a very high proportion of undescribed species (May, 2010; Riedel *et al.*, 2013a). This also became evident during the FISA workshop, for a range of taxa from rotifers to small Crustacea and aquatic insects. There is no doubt that a solid taxonomic framework is the most desirable foundation for any kind of biodiversity assessment. Proposed tools to accelerate taxonomy are manifold (Mallet and Willmott, 2003; Dayrat, 2005; Padial *et al.*, 2010; Riedel *et al.*, 2013a), and a review is not within the scope of this paper. Here, we rather focus on an alternative approach, and suggest the use of large scale DNA sequencing and objective clustering of data (Meier *et al.*, 2006) to provide first insights into large scale patterns of aquatic invertebrate diversity across South East Asia. Specifically, we suggest using standardised protocols based on the DNA barcoding method outlined above (Hebert *et al.*, 2003a), to create an objective, scientific basis to better understand freshwater diversity in Asia, to provide publicly accessible data for conservation and the setting of research priorities, and, for example, biogeographic analysis and community ecology. This is not to rival established taxonomic expertise and formal species description, but rather a supporting measure to these approaches (Riedel *et al.*, 2013b). We refer to our campaign as *molecular biodiversity assessment* (MBA), and this can be further developed into metabarcoding (see below).

METHODS

The DNA laboratory and analytical procedures required to obtain barcodes have been described in detail in previous open access publications (Hendrich *et al.*, 2010; Hawlitschek *et al.*, 2012; Tänzler *et al.*, 2012). Detailed information can also be found in our laboratory wiki at: http://zsm-entomology.de/wiki/The_Beetle_D_N_A_Lab. While museum specimens may deliver DNA adequate for barcoding purposes (Hebert *et al.*, 2013), the use of freshly collected specimens and/or specimens stored in 96% ethanol at -80°C is strongly advocated. Whenever possible, non-destructive DNA extractions (*e.g.* using the Qiagen™ column kit; Qiagen, Venlo, The Netherlands) should be performed using whole specimens for small taxa, and leg, thoracic, abdominal or antennae tissues for larger taxa. The preservation of specimen integrity is of great importance, since they will be retained in a public natural history collection as a physical voucher (Riedel *et al.*, 2013a). All vouchers are then digitised and the pictures uploaded to the database for further ease of identification and referencing. The DNA fragment we use is either the so-called barcoding fragment, the 5' end of cytochrome COI (Hebert *et al.*, 2003a) or the 3' end of *cox1* (for diving beetles, because we traditionally focused on that fragment; Hendrich *et al.*, 2010). Once a specimen has been processed through this pipeline, the corresponding sequence, picture and exhaustive information (classification, collector, locality) should be uploaded to complete the global database.

For our preliminary studies (Hendrich *et al.*, 2010; Tänzler *et al.*, 2012), we used the SpeciesIdentifier module of TaxonDNA software (Meier *et al.*, 2006). This module allows clustering sequence data at different thresholds, *e.g.* ranging 1-10% of sequence divergence using uncorrected *p*-distances (Fig. 1). SpeciesIdentifier accounts for threshold violations according to the *triangle inequity* (*i.e.* when the divergence between A-B and B-C is 3% or less, but A-C exceeds 3%, then A, B and C would still be

grouped into one 3% cluster by Taxon DNA) (Fig. 1). In our case, a maximum threshold of 10% was likely to capture any kind of lineage idiosyncrasy. It is very important to recall that such distances between species should not be seen as fixed values, as they depend on the lineage or even the area under consideration. Although a generalised cut-off level for the SE Asian freshwater beetles here is suggested to be *ca.* >3% (but see below), one should take into account that biodiversity assessments using molecular screening should always be interpreted within context. Hebert *et al.* (2003b), for example, show that in Cnidaria, *cox1* sequences between species tend to be extremely low (<1%), while in certain Crustacea, they can be very high (>20%), as we will discuss below for *Trigonopterus* weevils from New Guinea. This illustrates that the context should indeed be taken into account, but also that a combination of more than one technique (*e.g.* morphology+genetics) is highly important especially to test molecular methods for potential focal taxa and of course for proper taxonomic work.

SpeciesIdentifier recognises *a priori* delineated species from sequence names as long as the name follows the format *genus species*, *i.e. Rhantus suturalis*, or *Exocelina australiaone* MB1307. The output summarises the number of different species names in the dataset, the number of clusters found under the preset threshold (*e.g.* 1, 2, 3%), the number of clusters containing only one species name, and the number of *perfect clusters* (those that contain all individuals under one species name and only those individuals, *i.e.* monophyly). Therefore, we can calculate the number of split clusters (one species split into more than one cluster, *i.e.* paraphyly) and lumped clusters (more than one species name in a cluster). SpeciesIdentifier was also used for species richness estimation, with clusters taken as species surrogates. For any clustering threshold (*e.g.* at 1, 2, 3%), two values were reported.

The first of these values was the number of clusters found relative to the number of morphology-based species names in the dataset (*agreement* hereafter). This is illustrated in Fig. 2, left part (same height for columns means a perfect match); Fig. 3 shows the clustering performance at different preset thresholds. For example, a dataset with a hundred species names and a threshold clustering at 25% divergence would likely reveal only one cluster. Thus, our species richness estimation would amount to a meager 1% (*agreement*) of the true species, as delineated by morphology. Second, and more importantly, we report *taxonomic accuracy*, which was calculated as the number of perfect clusters (*i.e.* clusters containing all sequences of a morphology based species and only those sequences) relative to the number of species in the dataset. The number of perfect clusters can increase when the existing taxonomy is revised to accommodate cryptic or overlooked species. A 100% *accuracy* means that all clusters perfectly mirror putative species based on morphology (Fig. 2, right part – above circle means full congruence, middle circles are values in between, and lower circles mean full incongruence; Fig. 3). Higher numbers indicate species number overestimation/oversplitting, lower numbers indicate species number underestimation/lumping.

It is important to use species names, either formally identified described species, or operational expert taxonomic units (*e.g. Exocelina australiaone* MB1307) to be able to evaluate the performance of the method.

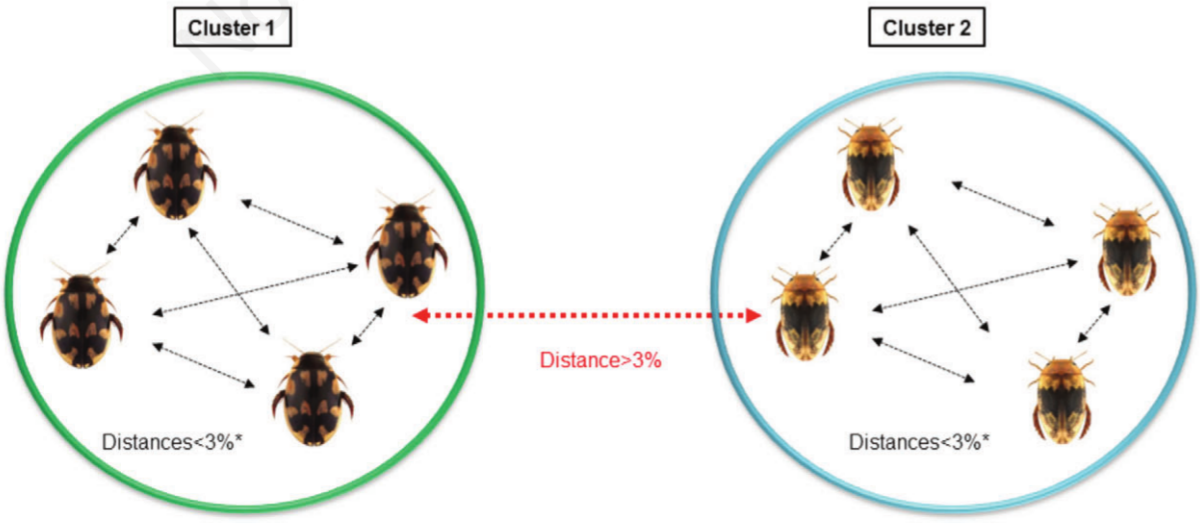


Fig. 1. Principle of clustering DNA sequence data. *The distance between specimens within a 3% cluster can exceed 3% because of the *triangle inequity*.

For the proposed molecular biodiversity assessment campaign, SpeciesIdentifier can and will have to use sequence names such as *Ephemeroptera sample MB7673* and so forth until further refinements are made based on taxonomists' feedback.

RESULTS AND DISCUSSION

Molecular biodiversity assessment

In preliminary studies we have asked: how well does the *number* of clusters agree with the *number* of species in the dataset (quantitative assessment) (Fig. 2)?

We also asked: how well do the *contents* of clusters agree with *contents* of species in the dataset? In other words: is there a method that detects species from our sequence matrix with 100% accuracy? (qualitative assessment) (Fig. 2).

For the entire Australian diving beetle (Coleoptera: Dytiscidae) fauna, with >270 sequenced species, we have empirically shown that the best results occur at cut-off thresholds between 2 and 3%. At 2-3%, we find a 100% correct estimation of the species number in a sample (quantity), however the quality was lower than expected, only *ca.* 80% of the clusters could be translated directly into formal species (Fig. 3) (Hendrich *et al.*, 2010). In a dataset of 279 species of New Guinean *Trigonopterus* weevils (Curculionidae) (Tänzler *et al.*, 2012), the quality of the clustering was 86% at a 3% cut-off, but the species number was 16% higher than that based on morphologically described species. Because of their high interspecific divergences, performance for *Trigonopterus* was best at 8% cut-off. The surprising result was that for both groups, utility of a standard threshold at *e.g.* 3% delivers error margins for taxonomic accuracy of <30% (diving beetles) (mean error for diving beetles individual genera was 21%) and <14% (*Trigonopterus* weevils), which were lower than in many studies using morphospecies sorting, where error could be up to 80% (Krell, 2004).

Clustering of *cox1* sequences can therefore provide a first insight into local species diversity (but of course not replace proper taxonomic revision). However, it has also

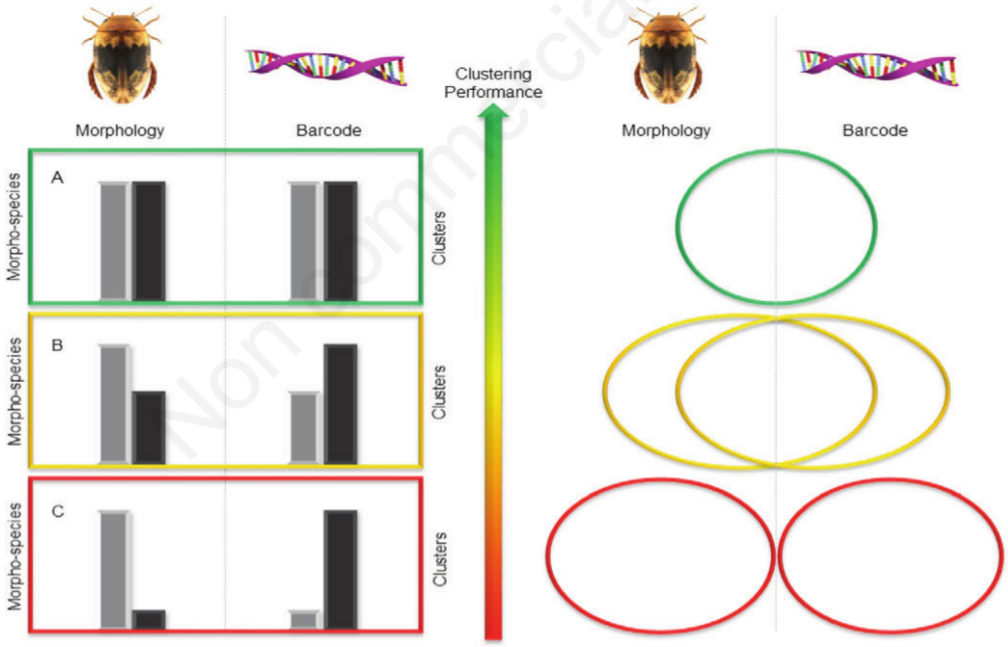


Fig. 2. Different cases of clustering performances and summary of the issue of lineage idiosyncrasy. A) The numbers of both morphospecies and genetic clusters are highly congruent. In such a case, the clustering performance is optimal, meaning that the agreement between morphologically identified species and genetic clusters based on the barcode fragment is excellent as represented by broadly overlapping circles. B) The number of morphospecies and genetic clusters are different, with two possibilities, *i.e.* the clustering based on barcodes recovers less putative species than the number of species based on morphology, or the opposite. In this case, morphology and barcoding are not in full agreement, and the clustering performance is then moderate. C) The number of species recovered by either morphology or barcoding is completely different, and the clustering performance is poor, as represented by non-overlapping circles between morphology-based and barcode-based identifications. It is noteworthy that the efficiency of clustering is highly dependent on lineage idiosyncrasy, and that despite this pitfall, the performance can be improved by tuning the threshold of clustering.

been shown that datasets should be large and ideally cover many lineages for such an MBA approach. Performance is highly lineage idiosyncratic, and errors for individual lineages can be significantly higher (Hendrich *et al.*, 2010). Studies on additional taxa are certainly needed. However, we feel confident that this approach can be widely applied to little-known communities to gain first insights into diversity patterns. Over time, and with improving taxonomic knowledge and databases, cluster-by-cluster sorting of lineages will be named, and their contents evaluated.

Molecular biodiversity assessment data for conservation, community ecology and biogeography

Haplotype clusters can be used to calculate similarity indices, such as the Sorenson index. Also, results for larger datasets were consistent between different analytical methods, *e.g.* traditional species delineation *vs* DNA entities, delineated by clustering or more sophisticated general mixed Yule coalescent (GMYC) analysis (Hendrich *et al.*, 2010; Tänzler *et al.*, 2012). Sequence data can moreover be used to calculate richness accumulation curves, using either haplotype or phylogenetic diversity. For regional samples of aquatic macroinvertebrates from Canada, a high correlation with morphologically delineated entities was found (Zhou *et al.*, 2009). Sequence data help to assign those lifestages to species names for which there is otherwise little hope to achieve reliable identification. This greatly aids ecological surveys and environmental impact assessments, especially when immature stages and adults occupy very different habitats, *e.g.* in the Trichoptera and Ephemeroptera, and it is crucial for conservation priority setting to understand where certain species actually breed (Ruiter *et al.*, 2013; Zhou *et al.*, 2007, 2010). Zhou *et al.* (2013) provide references on national biomonitoring programmes and possible sequencing-based approaches to aid these ecological surveys.

Limitations and pitfalls

DNA sequence data and MBA cannot automatically be translated into species entities, and does not provide a substitute for traditional taxonomy. Yet, the method can be a proxy for *species* sorting, and such *species* can be cross-checked with morphological or other evidence to accurately delineate species. This notwithstanding, problems associated with incompatible molecular *vs* morphological evidence are well established (Hendrich *et al.*, 2010; Hawlitschek *et al.*, 2012). Causes include, for example, very recent origins of species with associated incomplete lineage sorting and introgression (Monaghan *et al.*, 2006; Hawlitschek *et al.*, 2012), and this problem may occur particularly when a study covers species along with their closest relatives, *e.g.* densely sampled radiations or densely sampled faunas (Hendrich *et al.*, 2010; Bergsten *et al.*, 2012a). There is also not necessarily a straightforward approach when nuclear DNA markers are included (Skale *et al.*, 2012). However, these are not major problems as long as a study is carefully designed, taking these pitfalls into consideration. In doing so, a lot can be gained: MBA does not substitute carefully conducted taxonomic studies, it is really only a very first step. Nevertheless, MBA can be very useful as a framework for establishing further, integrative studies (Bergsten *et al.*, 2012a, 2012b).

A vision for a molecular biodiversity assessment campaign of Asian macrozoobenthos

Asian wetlands and streams are incredibly diverse biologically, yet we are only now beginning to comprehend the true magnitude of their species diversity, and the patterns of this biodiversity between sites. Concurrently, Asian wetlands, in particular, are under enormous anthropogenic pressure, and conservation measures to maintain their ecosystem services are of utmost importance.

The setting of conservation priorities is hampered by a lack of taxonomic knowledge, and the inability to place re-

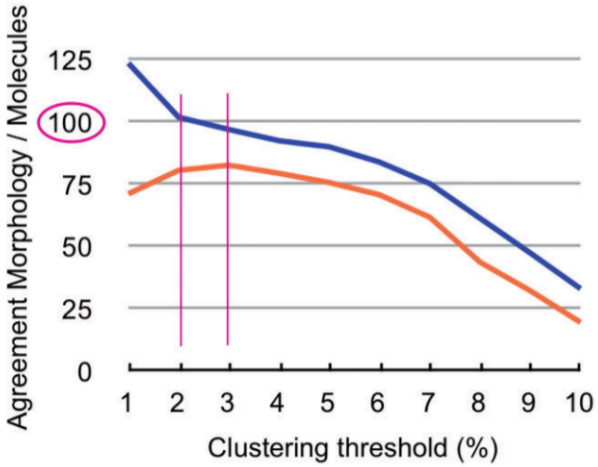


Fig. 3. Agreement and taxonomic accuracy of molecular clusters estimated at thresholds of 1-10% sequence divergence. Y-axis represents the relation of morphological entities *vs* molecular ones in % (*e.g.* if the *a priori* accepted morphological species number is 100, and we find 125 clusters using a 1% threshold, the value is 125%, meaning 25% overestimation of species number). Upper line is the agreement of species diversity estimation (percentage of clusters relative to the number of morphological species) while the lower line represents the taxonomic accuracy of clusters (number of clusters containing all sequences of a named species). At 100, there would be perfect agreement between both methods (modified from Hendrich *et al.*, 2010).

sults from local species inventories into a regional or global context. It has been shown that the full geographic range of a species and their genetic diversity has to be carefully considered (Bálint *et al.*, 2011), and the same is true for levels of endemism (species turnover or β-diversity).

We suggest the creation of a globally accessible database that has the power to strongly improve knowledge on Asian macrozoobenthos (and beyond), and begin to remedy the current problems associated with insufficient taxonomic knowledge.

The basic workflow is that of a typical DNA barcoding campaign, bringing together expert taxonomists, field biologists and molecular taxonomists. The major theme is sustainability and accountability, with all vouchers digitised and deposited in public natural history collections; examples can be seen at the *All Lepidoptera* website (http://www.lepbarcoding.org), or the already mentioned BOLD website (http://www. boldsystems.org). Pfrender *et al.* (2010) provide a review on project strategies that could also be adopted for work in SE Asia.

We envision a planning workshop at the next FISA meeting, to discuss an initial funding and sampling strategy. The obvious initial focus might be directed to lowland wetlands, which are under extreme threat. To gain first insights into diversity patterns, we suggest initial sampling at 200 localities across Southern and SE Asia west of Wallace line, covering the various biogeographical and ecological regions.

The DNA sequence data can serve as an immediate proxy for formally named species, to satisfy the requirement for a preliminary assessment of species diversity and diversity patterns using the MBA outlined above. Such a system would also support taxonomy, as data and vouchers will be freely available for experts, and monographic work should be strongly encouraged and financially supported, wherever possible. The preliminary data would certainly highlight areas of greatest potential species richness, which should incentivise taxonomists to prioritise efforts in these geographic areas of highest biodiversity.

A suggestion for taxon selection

We feel that a targeted approach is the key to success, and taxon selection should be based on discussions within the community to identify taxa where there is current taxonomic expertise to rapidly facilitate project outcomes, and also the selection of taxa where Asian projects can feed directly into existing global initiatives for added value. Examples are the international Barcode of Life's (iBOL) working group 1.7. Freshwater Bio-Surveillance which covers the groups commonly used for environmental impact assessments (EIAs) such as Ephemeroptera, Plecoptera, Trichptera (EPT) and Odonata, or the Trichoptera barcode of life (http://trichopterabol.org/). Others include different groups of Coleoptera, Crustacea and Mollusca where local (national) and international working groups already study Asian representative taxa. Groups should in general have been shown to add relevant information for environmental impact assessment, and should be comparably easy to handle and curate. The Barcoding Fauna Bavarica project (http://www.faunabavarica.de/) is an example where taxon selection was focused that way, funding and involving existing taxonomic expertise. After three years, more than 10,000 species have been collected, identified and sequenced. The setting for taxon identification is different in the widely understudied Asian fauna, but the basic framework is the same.

New technologies

There seems to exist a misconception as to the strength of DNA barcoding. It lies in the use of a standard marker that can link genotype to species. This requires a carefully well-finished and taxonomically revised database as basic knowledge of which haplotype belongs to which species, or which group of similar haplotypes represents which lineage of closely related young species (Hawlitschek *et al.,* 2012). Such a database forms the reference system for future mass sequencing technologies – no matter what percentage of a genome is being sequenced – so that matching the genetic data with taxonomy will only require a *cox1* database search in most cases.

Next generation sequencing

DNA barcoding can be coupled with next-generation sequencing (NGS) and this method is known as metabarcoding (Taberlet *et al*., 2012a, 2012b). The latest DNA sequencing technologies can provide incredible amounts of information economically and more rapidly than ever envisioned before. Next-generation sequencing makes it possible to sequence large amounts of specimens at the same time. The length of sequence reads is increasing as technologies advance, and this may allow efficient sequencing of bulk samples, an approach referred to as environmental sequencing or metabarcoding (Hajibabaei *et al.,* 2011; Thomsen *et al.,* 2012; Yu *et al.,* 2012).

Hajibabaei *et al.* (2011) suggested that this is feasible for bulk samples of aquatic insects. Thomsen *et al.* (2012) showed that it is possible to detect (large) target species from water samples alone. That means that NGS methods offer an opportunity to monitor rare and threatened animal species from DNA traces in their freshwater environments, and even potential for identifying new species. The new DNA-based method is effective even in locations where the animals are extremely rare. The method if further developed might also show some correlation between the amount of DNA in the environment and the density of individuals, meaning that the DNA detection method can even be used

to estimate population sizes. This is crucial in the monitoring of rare animals, where one often wants to know whether the population is large or small. The United Nations (UN) have agreed to halt the decline of biodiversity, but a prerequisite to do so is that we are capable of accurately documenting the status of threatened species. This new approach is a fundamental step forward making it cheaper and faster to monitor endangered species, and thus prioritise efforts to the benefit of biodiversity at a broad scale. Recent advances in NGS technologies enable such rapid quantification of these parameters directly from environmental water samples. Yu *et al.* (2012) show that NGS provides robust alpha and beta biodiversity assessments even in the absence of existing taxonomy, essentially taking the molecular biodiversity assessment methods suggested by Hendrich *et al.* (2010) and Tänzler *et al.* (2012) to the ultrasequencing level.

Finally, Zhou *et al.* (2013) describe an ultra-deep sequencing pipeline, *i.e.* NGS sequencing of environmental samples using a sequencing platform that produces extremely high amounts of sequence reads, which does not rely on pre-sequencing polymerase chain reaction (PCR) amplification. This means that the sample will go through a mitochondrial DNA enrichment step (centrifugation), and the DNA extraction step, followed by NGS sequencing. Such a pipeline, once optimised, would allow the simultaneous sequencing of mixtures of phylogenetically rather heterogenous samples, avoiding PCR primer mismatch problems. Also, without PCR amplification, there is some hope to be able to reveal relative abundance patterns in the sample.

CONCLUSIONS

Molecular biodiversity assessment might be a good alternative to tackle morphotype sorting, as it is more objective and relies on more standardised protocols especially in the context of larger-scale projects with diverse personnel.

Conducted carefully, MBA for the first time in history: i) helps to gather scientific data where there is little or no hope for other information; ii) enables meaningful assessment of diversity without guessing how many *morphospecies* there are; and iii) most importantly, it enables student/local participation in an objective framework.

New sequencing technologies will also enable us to conduct large surveys or monitoring for conservation purposes based on objective scientific data.

Understanding how ecosystems function, however, relies on knowledge of species and their biotic and abiotic interactions, so taxonomists and field biologists/ecologists will always play a key role. The approach outlined here should facilitate targeted taxonomic work, and free taxonomists from routine work that can certainly be standardised and automated more efficiently and effectively.

ACKNOWLEDGMENTS

Our sincere thanks to Dr. Maria K. Hołyńska and Prof. La-orsri Sanoamuang for organising the *Freshwater Invertebrates of Southeast Asia: Biodiversity and Origin* (FISA) workshop in Maha Sarakham, Thailand, 30th July to 4th August, 2012. This work was supported by DFG grant BA2152/11-1 to Dr. Michael Balke.

REFERENCES

Bálint M, Domisch S, Engelhardt CHM, Haase P, Lehrian S, Sauer J, Theissinger KT, Pauls SU, Nowak C, 2011. Cryptic biodiversity loss linked to global climate change. Nat. Clim. Chang. 1:313-318.

Bergsten J, Bilton DT, Fujisawa T, Elliott M, Monaghan M, Balke M, Hendrich L, Geijer J, Herrmann J, Foster G, Ribera I, Nilsson AN, Barraclough TG, Vogler AP, 2012a. The effect of geographical scale of sampling on DNA barcoding. Syst. Biol. 61:851-869.

Bergsten J, Brilmyer G, Crampton-Platt A, Nilsson AN, 2012b. Sympatry and colour variation disguised well-differentiated sister species: *suphrodytes* revised with integrative taxonomy including 5 kbp of housekeeping genes (Coleoptera: Dytiscidae). DNA Barcodes (Berlin) 1:1-18.

Dayrat B, 2005. Towards integrative taxonomy. Biol. J. Linn. Soc. 85:407-415.

Dincă V, Zakharov EV, Hebert PD, Vila R, 2011. Complete DNA barcode reference library for a country's butterfly fauna reveals high performance for temperate Europe. P. Roy. Soc. B-Biol. Sci. 278:347-355.

Goldstein PZ, DeSalle R, 2011. Integrating DNA barcode data and taxonomic practice: determination, discovery, and description. Bioessays 33:135-147.

Hajibabaei M, Shokralla S, Zhou X, Singer GAC, Baird DJ, 2011. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. PLoS ONE 6:e17497.

Hausmann A, Haszprunar G, Hebert PDN, 2011. DNA barcoding the geometrid fauna of Bavaria (Lepidoptera): successes, surprises, and questions. PloS ONE 6:e17134.

Hawlitschek O, Hendrich L, Espeland M, Toussaint EFA, Genner MJ, Balke M, 2012. Pleistocene climate change promoted rapid diversification of aquatic invertebrates in Southeast Australia. BMC Evol. Biol. 12:142.

Hebert PDN, Cywinska A, Ball SL, deWaard JR, 2003a. Biological identifications through DNA barcodes. P. Roy. Soc. Lond. B Bio. 270:313-321.

Hebert PDN, Penton EH, Burns J, Janzen DH, Hallwachs W, 2004a. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly, *Astraptes fulgerator*. P. Natl. Acad. Sci. USA 101:14812-14817.

Hebert PDN, Ratnasingham S, deWaard JR, 2003b. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. P. Roy. Soc. Lond. B Bio. 270(Suppl.1): S96-99.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM, 2004b. Identification of birds through DNA barcodes. PLoS Biol. 2:1657-1663.

Hebert PDN, Zakharov EV, Prosser SW, Sones JE, McKeown JT, Mantle B, La Salle J, 2013. A DNA 'Barcode Blitz': rapid digitization and sequencing of a natural history collection. PLoS ONE 8:e68535.

Hendrich L, Hawlitschek O, Balke M, 2009. A revision of Australasian *Neobidessodes Suphrodytes* gen.n. diving beetles - integrating morphology, cybertaxonomy, DNA taxonomy and phylogeny (Coleoptera: Dytiscidae, Bidessini). Zootaxa 2288:1-41.

Hendrich L, Pons J, Ribera I, Balke M, 2010. Mitochondrial cox1 sequence data reliably uncover patterns of insect diversity but suffer from high lineage-idiosyncratic error rates. PloS ONE 5:e14448.

House of Lords, 2008. Systematics and taxonomy: follow-up. House of Lords ed., London: 330 pp. Available from: http://www.publications.parliament.uk/pa/ld200708/ldselect/ldsctech/162/162.pdf

Janzen DH, Hallwachs W, Blandin P, Burns JM, Cadiou J-M, Chacon I, Dapkey T, Deans AR, Epstein ME, Espinoza B, Franclemont JG, Haber WA, Hajibabaei M, Hall JPW, Hebert PDN, Gauld ID, Harvey DJ, Hausmann A, Kitching IJ, Lafontaine D, Landry J-F, Lemaire C, Miller JY, Miller JS, Miller L, Miller SE, Montero J, Munroe E, Rab Green S, Ratnasingham S, Rawlins JE, Robbins RK, Rodriguez JJ, Rougerie R, Sharkey MJ, Smith MA, Solis MA, Sullivan JB, Thiaucourt P, Wahl DB, Weller SJ, Whitfield JB, Willmott KR, Wood DM, Woodley NE, Wilson JJ, 2009. Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. Mol. Ecol. Resour. 9:1-26.

Krell F-T, 2004. Parataxonomy vs. taxonomy in biodiversity studies - pitfalls and applicability of 'morphospecies' sorting. Biodivers. Conserv. 13:795-812.

Mallet J, Willmott K, 2003. Taxonomy: renaissance or tower of Babel? Trends Ecol. Evol. 18:57-59.

May RM, 2010. Tropical arthropod species, more or less? Science 329:41-42.

Meier R, Shiyang K, Vaidya G, Ng PKL, 2006. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. Syst. Biol. 55:715-728.

Monaghan MT, Balke M, Gregory TR, Vogler AP, 2005. DNA-based species delineation in tropical beetles using mitochondrial and nuclear markers. Philos. T. Roy. Soc. B 360: 1925-1933.

Monaghan MT, Balke M, Pons J, Vogler AP, 2006. Beyond barcodes. Complex DNA taxonomy of a South Pacific Island radiation. P. Roy. Soc. Lond. B Bio. 273:887-893.

Monaghan MT, Wild R, Elliot M, Fujisawa T, Balke M, Inward DJ, Lees DC, Ranaivosolo R, Eggleton P, Barraclough TG, Vogler AP, 2009. Accelerated species inventory on Madagascar using coalescent-based models of species delineation. Syst. Biol. 58:298-311.

Mora C, Tittensor DP, Adl S, Simpson AG, Worm B, 2011. How many species are there on Earth and in the ocean? PLoS Biol. 9:e1001127.

Moritz C, Cicero C, 2004. DNA barcoding: promise and pitfalls. PLoS Biol. 2:e354.

Padial JM, Miralles A, De la Riva I, Vences M, 2010. The integrative future of taxonomy. Front. Zool. 7:16.

Pfrender ME, Hawkins CP, Bagley M, Courtney GW, Creutzburg BR, Epler JH, Fend S, Schindel D, Ferrington LC Jr, Hartzell PL, Jackson S, Larsen DP, Lévesque A, Morse JC, Petersen MJ, Ruiter D, Whiting M, 2010. Assessing macroinvertebrate biodiversity in freshwater ecosystems: advances and challenges in DNA-based approaches. Q. Rev. Biol. 85:319-340.

Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, Kamoun S, Sumlin WD, Vogler AP, 2006. Sequence-based species delimitation 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Syst. Biol. 55:595-609.

Rach J, DeSalle R, Sarkar IN, Schierwater B, Hadrys H, 2008. Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. P. Roy. Soc. Lond. B Bio. 275:237-247.

Ratnasingham S, Hebert PDN, 2007. BOLD: the barcode of life data system. Mol. Ecol. Notes 7:355-364.

Riedel A, Sagata K, Suhardjono YR, Tänzler R, Balke M, 2013a. Integrative taxonomy on the fast track - towards more sustainability in biodiversity research. Front. Zool. 10:15.

Riedel A, Sagata K, Surbakti S, Tänzler R, Balke M, 2013b. One hundred new species of *Trigonopterus* weevils from New Guinea. Zookeys 280:1-150.

Ruiter DE, Boyle EE, Zhou X, 2013. DNA barcoding facilitates associations and diagnoses for Trichoptera larvae of the Churchill (Manitoba, Canada) area. BMC Ecol. 13:5.

Skale A, Tänzler R, Hendrich L, Balke M, 2012. High mitochondrial DNA sequence divergence in New Guinea *Carabdytes* stream beetles and the taxonomist's dilemma when other evidence is kind of subtle (and collecting localities are far far away). Zookeys 247:31-43.

Taberlet P, Coissac E, Hajibabaei MM, Rieseberg LH, 2012a. Environmental DNA. Mol. Ecol. 21:1789-1793.

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E, 2012b. Towards next generation biodiversity assessment using DNA metabarcoding. Mol. Ecol. 21:2045-2050.

Tänzler R, Sagata K, Surbakti S, Balke M, Riedel A, 2012. DNA barcoding for community ecology. How to tackle a hyperdiverse, mostly undescribed Melanesian fauna. PLoS ONE 7:e28832.

Thomsen PF, Kielgast J, Iversen LL, Wiuf C, Rasmussen M, Gilbert MT, Orlando L, Willerslev E, 2012. Monitoring endangered freshwater biodiversity using environmental DNA. Mol. Ecol. 2:2565-2573.

Will KW, Mishler BD, Wheeler QD, 2005. The perils of DNA barcoding and the need for integrative taxonomy. Syst. Biol. 54:844-851.

Wilson EO, 1988. Biodiversity. National Academies Press, Washington, DC: 538 pp.

Yu DW, Ji Y, Emerson BC, Wang X, Ye C, Yang C, Ding Z, 2012. Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. Methods Ecol. Evol. 3:613-623.

Zhou X, Adamowicz S, Jacobus LM, DeWalt RE, Hebert PDN, 2009. Towards a comprehensive barcode library for arctic life. Ephemeroptera, Plecoptera, and Trichoptera of Churchill, Manitoba, Canada. Front. Zool. 6:30.

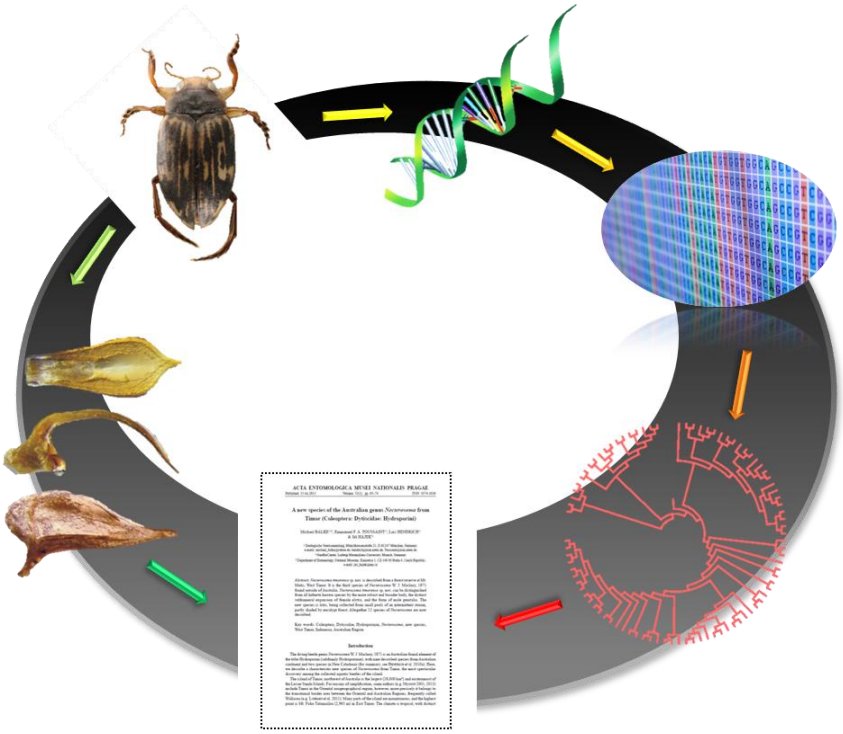
Zhou X, Jacobus LM, DeWalt RE, Adamowicz S, Hebert PDN, 2010. Ephemeroptera, Plecoptera, and Trichoptera fauna of Churchill (Manitoba, Canada): insights into biodiversity patterns from DNA barcoding. J. N. Am. Benthol. Soc. 29:814-837.

Zhou X, Kjer KM, Morse JC, 2007. Associating larvae and adults of Chinese Hydropsychidae (Insecta: Hydropsychidae) using DNA sequences. J. N. Am. Benthol. Soc. 26:718-741.

Zhou X, Li Y, Liu S, Yang Q, Su X, Zhou L, Tang M, Fu R, Li J, Huang Q, 2013. Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. Gigascience 2:4.

PART 1: ON THE DISCOVERY OF NEW SPECIES IN THE ARCHIPELAGO

Chapter 6. Integrative taxonomy at work



The combination of molecular and morphological characters is a powerful tool to describe new taxa

"*Taxonomy (the science of classification) is often undervalued as a glorified form of filing—with each species in its folder, like a stamp in its prescribed place in an album; but taxonomy is a fundamental and dynamic science, dedicated to exploring the causes of relationships and similarities among organisms. Classifications are theories about the basis of natural order, not dull catalogues compiled only to avoid chaos.*"

Stephen Gould, *Wonderful Life*, 1990.

Chapter contents

6.1 Paper II - Molecular species delimitation of charismatic tropical butterflies 61

6.2 Paper III - A new species of diving beetle from Timor 101

6.3 Paper IV - A new species of diving beetle from Biak Island 112

6.4 Paper V - A new genus of diving beetle from Australia 119

PAPER II

6.1 Molecular species delimitation of charismatic tropical butterflies

This manuscript is in a 2nd round of review in the journal Molecular Ecology

An array of molecular species delimitation methods casts light on species boundaries in the charismatic *Polyura* Nawab butterflies

Emmanuel F.A. Toussaint[1,*], Jérôme Morinière[1], Chris J. Müller[2], Krushnamegh Kunte[3], Bernard Turlin[4], Axel Hausmann[1,5] and Michael Balke[1, 5]

[1]*SNSB-Bavarian State Collection of Zoology, Münchhausenstraße 21, 81247 Munich, Germany.*

[2] *Australian Museum, 6 College Street, Sydney, NSW 2010, Australia.*

[3]*National Center for Biological Sciences, Tata Institute of Fundamental Research, Bengaluru 560065, India.*

[4]*14, Résidence du Nouveau Parc, 78570 Andrésy, France.*

[5]*GeoBioCenter, Ludwig-Maximilians University, Munich, Germany.*

[*] *Correspondence to be sent to: Emmanuel F.A. Toussaint (Toussaint@zsm.mwn.de), Tel: +49 898107155.*

Abstract

Tropical hotspots are astonishingly diverse, yet most of their species richness remains unknown. Because these hotspots are anthropogenically threatened, there have been recent pleas to accelerate species discovery using different approaches rather than only traditional taxonomy based on morphological characters. In particular, the use of increasingly more sophisticated methods of molecular species delimitation has received a growing attention. Whilst sequencing pipelines are making large strides towards cheaper and faster access to vast amounts of DNA sequence data, this trend constitutes a completion and putatively sometimes an alternative to morphology-based taxonomy. Cryptic species complexes for which fixed morphological synapomorphies are expected to be lacking may also be revealed and corroborated where traditional taxonomy might have its limits. In this study, we tested an array of species delimitation methods in the highly emblematic tropical *Polyura* Nawab butterflies found across twelve biodiversity hotspots of the Indomalayan / Australasian archipelago to investigate the discriminatory performance of these methods compared to merely morphology-based taxonomy. Based on a multimarker molecular phylogenetics framework, we describe two new species *Polyura paulettae* and *Polyura smilesi*, one synonym is proposed and six populations are raised to species status increasing diversity by more than 25%. Surprisingly, these species are not only island endemics but also independently evolving continental lineages. Species delimitation methods were mostly congruent and succeeded to cross-validate most extant morphological species. Our findings demonstrate the efficiency of such approaches on empirical data and pave the way for investigation of less well-known groups to unveil patterns of species richness and catalogue Earth's concealed, therefore unappreciated diversity.

Keywords

Australasian-Indomalayan archipelago; species delimitation; bGMYC; BPP; *Charaxes*; Nymphalidae; *Polyura*; Poisson Tree Processes.

Introduction

Habitat loss and climate disruptions threaten global species diversity (Thomas et al. 2004; Brooks et al. 2006). This is especially true in tropical regions that hold a majority of the 35 biodiversity hotspots found across the planet, representing regions of extreme yet highly threatened endemism (Mittermeier et al. 2004; Williams et al. 2011). Species disappear at an alarming pace whilst a growing body of evidence underpins the dramatic impact of biodiversity loss on ecosystem functioning (Wardle 2011; Cardinale et al. 2012; Hooper 2012). In this context, cataloguing Earth's biodiversity is urgent. Acceleration of species discovery and description is achievable using new techniques and modus-operandi. Recently, some pleas have been made to use molecular data instead or in addition to morphology and/or other lines of evidence to help discovering unknown diversity throughout the tree of life (e.g. Padial et al. 2010; Riedel et al. 2013a). Within this effervescence, the field of molecular species delimitation has taken a growing importance whilst opening a heated debate (e.g. Bauer et al. 2011; Fujita and Leaché 2011; Carstens et al. 2013). A wide array of new species delimitation methods have been developed in the past decade aiming at connecting molecular variation between organisms and taxonomy using models and thresholds of different nature and level of complexity (e.g. Hebert et al. 2003; O'Meara et al. 2006; Pons et al. 2006; Yang and Rannala 2010; Ence and Carstens 2011; Reid and Carstens 2012; Ratnasingham and Hebert 2013; Zhang et al. 2013). Although these new methods allow discriminating species-level molecular entities under a certain threshold, until recently only a few studies using such approaches led to taxonomic acts (but see Jörger and Schrödl 2013; Riedel *et al*. 2013b; Satler et al. 2013). To this regard, clades for which morphology-based taxonomy is relatively well-known might be of prime interest to empirically test the efficiency of these methods in order to pave the way for a more rapid assessment of diversity in focal clades.

Numerous studies have investigated cryptic diversity and species boundaries in Lepidoptera clades rendering the entire order a flagship for molecular biodiversity assessment studies (e.g. Hebert et al. 2004). Barcoding in particular has been widely used to describe new taxa using a combination of molecular clustering and other lines of evidence such as genitalia, host-plant preferences or caterpillar morphology (Burns et al. 2007, 2008, 2010; Hausmann et al. 2009; Chacón et al. 2012). However, studies using the most recent advances in the field of molecular species delimitation are scarce (but see Dincă et al. 2011; Le Ru et al. 2014). Yet, many lepidopteran groups count among the most taxonomically well-known groups of insects and therefore offer the opportunity to test the efficiency of such methods in an empirical framework. The tribe *Charaxini* (Lepidoptera, Nymphalidae) comprises the charismatic

Charaxes (Emperors and Rajahs), *Euxanthe* (Forest Queens) and *Polyura* (Nawabs) butterflies. Passion for this group among collectors and researchers has led to a thorough assessment of morphology-based alpha-taxonomy (e.g. Smiles 1982; Henning 1989; Turlin 2005- 2014). About 170 species have been described from the Afrotropical region with about 50 to 60 other species spread as far as Southeast Asia, Wallacea and Pacific Islands (Aduse-Poku et al. 2009; Müller et al., 2010). Molecular phylogenetic investigations of the group have revealed the affiliation of closely related clades within *Charaxes* despite a lack of morphological evidence (Aduse-Poku et al. 2009). Despite some taxonomic suggestions (Aduse-Poku et al. 2009), the systematics of *Charaxes* and its close relatives the genera *Euxanthe* and *Polyura* remain contentious and presently *Charaxes* is likely to represent a complex paraphyletic series. The Nawab butterflies (*Polyura*) are restricted to the Indomalayan / Australasian archipelago (Figure 1). This region encompasses 14 biodiversity hotspots (Mittermeier et al. 2004; Williams et al. 2011) and is the fruit of a highly complex geological history (Hall 2011, 2012), rendering it a natural laboratory to study processes of lineage diversification. *Polyura* contains 26 morphologically delineated species (*sensu* Smiles 1982) of fast-flying large butterflies exhibiting the typical patrolling, fighting and hill-topping behavior of the genus. As adults, *Polyura* feed on carrion and dung but also on rotten fruits and oozing sap. They are distributed from India to Fiji and from the Ryukyu archipelago to Southern Australia. Numerous endemic species occur on remote islands such as Christmas Island, Fiji, New Caledonia, the Solomons and Vanuatu. Since its description (Billberg, 1820), the genus *Polyura* has been surprisingly overlooked before receiving increasing attention in the past decades with a complete revision of the group (Smiles 1982) and some attempts to unravel phylogenetic relationships at regional scales (Wang *et al.*, 2003, 2004; Long *et al.*, 2006).). In his comprehensive revision of *Polyura*, Robert L. Smiles noted that characters from the genitalia, larval instars or venation were of little assistance to delineate species and therefore he based his taxonomic assessment on morphological features from wing undersides that he found very informative (Smiles 1982). As a result *Polyura* is a model of choice to investigate the performance of molecular species delimitation in a robust phylogenetic framework.

Here, we have generated a multi-marker DNA sequence matrix comprising more than 200 specimens of all extant species of the genus *Polyura* recognized by Smiles (1982) to study species boundaries and phylogenetic affinities among this group. We seek to (*i*) infer phylogenetic relationships between all sequenced specimens to investigate the monophyly of morphological species *sensu* Smiles (1982), (*ii*) delineate species boundaries using recent

methods of molecular species delimitation, and (*iii*) describe potential new species with respect to the results of species delimitation methods and the insights of geographical and morphological information derived from the literature

Materials and Methods

Taxon Sampling and Molecular Biology

We collected butterflies in the India and New Guinea (permit numbers are listed in the Acknowledgements), and used museum specimens to assemble a comprehensive taxonomic sampling of the genus *Polyura* comprising 205 specimens representing all described species but the most likely extinct Sulawesi endemic *P. inopinatus* (Appendix 1 // Maps of distribution). All specimens sequenced for this study are listed in Appendix 2. Total genomic DNA was extracted from legs and antennae tissues of dried specimens using the DNeasy kit (Qiagen, Hilden, Germany). Using standard PCR protocols (Wahlberg and Wheat 2008; Müller *et al.* 2010) we amplified and then sequenced the following gene fragments: *cytochrome oxidase subunit 1* (*CO1*, 471 bp), *NADH dehydrogenase subunit 5* (*ND5*, 417 bp), *ribosomal protein S5* (*RPS5*, 573 bp) and *Wingless* (*WGL*, 396 bp). All outgroup sequences were retrieved from Genbank except *Charaxes viola* which was sequenced for the purpose of this study. We specifically sampled representatives from most *Charaxes* species groups to test the monophyly of *Polyura* (Appendix 2). The DNA sequences were edited by eye in GENEIOUS R6 (Biomatters, http://www.geneious.com/), aligned using MUSCLE (Edgar 2004) and the reading frames checked under MESQUITE 2.75 (http://mesquiteproject.org). The different datasets used to infer phylogenetic relationships were generated under MESQUITE. All sequences were deposited in GenBank (accession Nos. XXX) and on a public data-set on BOLD (###) as well.

Molecular phylogenetics

We run preliminary analyses to reconstruct gene trees for the four markers in order to detect potential supported incongruences. Results indicated no conflict between mitochondrial gene trees and no supported incongruence between mitochondrial and nuclear gene trees. As a result we used Bayesian Inference (BI) and Maximum Likelihood (ML) to reconstruct phylogenetic relationships of all specimens sequenced using a concatenated dataset. The

partitions and corresponding optimal models of substitution were searched under PartitionFinder 1.1.1 (Lanfear et al. 2012) using the *greedy* algorithm, either the *mrbayes* or *raxml* set of models because MRBAYES 3.2.2 (Ronquist et al. 2012) and RAXML (Stamatakis, 2006) implement different sets of substitution models. The Akaike Information Criterion corrected (AICc) was used to compare the fit of the different models. The BI analyses were performed using MRBAYES 3.2.2 (Ronquist et al. 2012). Two simultaneous and independent runs consisting of eight Metropolis-coupled Markov chain Monte Carlo (MCMC, one cold and seven incrementally heated) running 80 million generations were used, with a tree sampling every 1000 generations to calculate posterior probabilities (PP). We used the partitions recovered in PartitionFinder, but instead of using the *a priori* substitution models recovered, we used reversible jump MCMC (rjMCMC) to sample the entire space of possible models (Huelsenbeck et al. 2004). In order to investigate the convergence of the runs we investigated the split frequencies and Effective Sample Size (ESS) of all the parameters, and plotted the log-likelihood of the samples against the number of generations in TRACER 1.5 (http://BEAST.bio.ed.ac.uk/Tracer). A value of ESS>200 was acknowledged as a good indicator of convergence. All the trees that predated the time needed to reach a log-likelihood plateau were discarded as burn-in, and the remaining samples were used to generate a 50% majority rule con*sensu*s tree. The ML analyses were conducted with the best partitioning scheme selected in PartitionFinder 1.1.1 (Lanfear et al. 2012) using RAXML (Stamatakis, 2006). We performed 1000 Bootstrap replicates (BS) to investigate the level of support at each node. A calculated PP≥0.95 or a BS≥70 was considered to indicate strong support for a given clade (Erixon et al. 2003; Felsenstein 2004).

Molecular species delimitation

Delimiting species boundaries in the absence of discriminating morphological features is a challenging task. It is often extremely difficult if not impossible to recognize the most suitable molecular species delimitation method to use on empirical data. As a result, the use of multiple methods has been recommended in order to avoid bias and to assess the consistency of delineated species across models (Astrin et al. 2012; Carstens et al. 2013; Satler et al. 2013). Because the genus *Polyura* has not been revised since the work of Smiles (1982), and because only few informative morphological characters exist, we investigated inter-specific relationships and species boundaries using multiple cutting-edge methods of molecular

species delimitation. First we used the Poisson Tree Processes (PTP) model (Zhang et al. 2013) to infer molecular clades based on our inferred molecular phylogeny. The PTP method estimates the mean expected number of substitutions per site between two branching events using the branch length information of a phylogeny and then implements two independent classes of Poisson processes (intra and inter-specific branching events) before clustering the phylogenetic tree according to the results. The analyses were conducted on the web server for PTP (available at http://species.h-its.org/ptp/) using the RAxML topology as advocated for this method (Zhang et al. 2013; Tang et al. 2014).

Second, we used bGMYC (Reid and Carstens 2012), a Bayesian implementation of the GMYC approach (Pons et al. 2006). The GMYC model searches in an ultrametric gene tree the threshold at which branching patterns represent coalescent events or speciation events (Pons et al. 2006). As a result, the phylogenetic uncertainty and the ultrametrization of the tree have a great impact on the calculation of this threshold. The bGMYC implementation allows alleviating such shortcomings by providing the mean to use posterior distributions of trees as an input instead of a single tree (Reid and Carstens 2012). We therefore conducted the bGMYC approach using CO1 and ND5 ultrametric gene trees inferred in the BEAST 1.8.0 (Drummond et al. 2012) without outgroups under a strict clock model and a *Speciation: Yule Process Tree Model*. The runs consisted of 10 million generations sampled every 1000 cycles. Convergence was assessed by ESS values. A conservative burn-in of 10% was performed after checking the log-likelihood curves in Tracer 1.5. As advocated (Reid and Carstens 2012), 100 trees sampled at intervals from the posterior distribution of trees using LOGCOMBINER 1.8.0 (Drummond et al. 2012) were used to perform the bGMYC analyses. Species delimitation analyses were conducted in R using the package '*bGMYC'*. The analyses consisted for each of the 100 trees selected of 250000 generations with a burnin of 25000 and a thinning parameter of 100.

Third, we used Bayesian species delimitation as implemented in Bayesian Phylogenetics and Phylogeography (BPP) 2.2 (Rannala and Yang 2003; Yang and Rannala 2010). This method accommodates the species phylogeny as well as lineage sorting due to ancestral polymorphism. A gamma prior $G\theta s(\alpha,\beta)$, with mean α/β, is used on the population size parameters (θs). The age of the root in the species tree (τ_0) is assigned the gamma prior $G\tau_0(\alpha,\beta)$, whereas the other divergence time parameters are assigned the Dirichlet prior (Yang and Rannala, 2010: equation 2). The morphological species recognized by Smiles (1982), geographic clades as well as species recovered by the PTP and bGMYC analyses (with

PP≥0.95) were used as putative species, yielding a total of 38 taxa to test (Figure 2). We used *BEAST 1.8.0 (Heled and Drummond 2010) to estimate the species tree using the four alignments and assigned each specimen to its corresponding putative species. Since some species were not successfully sequenced for all the genes, we generated artificial uninformative sequences only comprising ambiguities for the few specimens lacking, that we included afterwards in the alignment files. For the four different partitions, we specified an uncorrelated lognormal prior for the clock, a *Yule Process* model as *Species Tree Prior* and a *Piecewise constant Population Size Model*. The analysis consisted of 50 million generations with a sampling interval of 5000 and a conservative burnin of 25%. As advocated by Leaché and Fujita (2010), we conducted three different sets of analyses with different values of α and β allowing θs and τ_0 to account for (*i*) large ancestral population sizes and deep divergence between species using $G\theta$s(1,10) and $G\tau_0$(1,10), (*ii*) small ancestral population sizes and shallow divergence between species using $G\theta$s(2,2000) and $G\tau_0$(2,2000), and finally (*iii*) large ancestral population sizes and shallow divergence between species using $G\theta$s(1,10) and $G\tau_0$(2,2000). The analyses were performed with the following settings: *speciesdelimitation*=1, *algorithm*=0, *finetune e*=2, *usedata*=1 and *cleandata*=0. The reversible-jump MCMC analyses consisted of 50000 generations (sampling interval of 2) with 25000 samples being discarded as burn-in. Each analysis was run twice using different starting seeds to confirm consistency between runs.

Results

Phylogenetic Relationships

All information relative to the sequencing result and data quality is gathered in Table 1. Based on the full dataset comprising all specimens sequenced for the study, we recover the genus *Charaxes* as paraphyletic with strong support (PP=1.0/BS=91). A first clade C1 contains all *Charaxes* species from Africa (including the genus *Euxanthe*) and the Indomalayan / Australasian archipelago except *C. paphianus*. In a second clade C2, the genus *Polyura* is recovered as monophyletic with strong support (1.0/100) with *C. paphianus* as sister taxon (0.95/66) rendering *Charaxes* paraphyletic. Within *Polyura*, three main clades are recovered (Figure 2). The first clade C3 (1.0/99) contained all species of the *athamas* group with *P. schreiber sensu* Smiles (1982, MOTU 1 to 4) in a sister position to the rest of the other species. *P. schreiber*, *P. jalysus* (MOTU 6), *P. arja* (MOTU 14) and *P. hebe* (MOTU

15) are recovered as monophyletic with strong support except the last presenting lower support (0.64/49), whereas *P. agraria* (MOTU 7 to 10) and *P. athamas* (MOTU 11 to13) are found to be paraphyletic. In *P. schreiber sensu* Smiles (1982), specimens from the Philippines form a well-delineated clade sister to all other specimens distributed in the rest of the distributional range. *P. agraria sensu* Smiles (1982) is divided in three strongly supported monophyletic subclades representing different geographic distributions; (*i*) Malaysian peninsula, (*ii*) Sunda, the Lesser Sunda Islands and Sulawesi, and (*iii*) India. In the second subclade, specimens from Sulawesi on one hand and from Sunda and the Lesser Sunda Islands on the other hand are also clearly separated with strong support. *P. athamas sensu* Smiles (1982) is equally split in three subclades roughly matching the same geographic areas as in *P. agraria sensu* Smiles (1982).

The second clade C4 (0.54/61) contains all representatives of the *eudamippus*-group with *P. delphis* (MOTU 16) as sister to the rest of the species. In this clade all species are recovered as monophyletic with strong support. The main incongruence between BI and ML topologies is the placement of *P. posidonius* (MOTU 18) from Tibet which is recovered as sister to *P. narcaea* (MOTU 17) in ML and in a more derived position in BI (Figure 2). Specimens of *P. eudamippus weismanni* (MOTU 21) from Okinawa Island form a well delineated clade sister to the rest of the *P. eudamippus* specimens (MOTU 22).

The clade C5 (1.0/99) referred to below as the *pyrrhus* group *sensu lato* comprises *P. cognata* (MOTU 26), *P. dehanii* (MOTU 23), *P. epigenes* (MOTU 24 and 25) and all representatives of the *pyrrhus* group *sensu stricto* (MOTU 27 to 38). In BI two weakly supported subclades are recovered for this division whereas the ML topology only recovers a succession of increasingly derived clades. The first BI subclade contains *P. dehanii* (MOTU 23) and *P. epigenes* (MOTU 24 and 25) whereas in ML *P. epigenes* (MOTU 24 and 25) is found to be sister to the rest of the species of the subclade. All species recognized by Smiles (1982) are recovered as monophyletic with strong support except in the last clade C6 here referred to as the *pyrrhus* complex (MOTU 31 to 38). Within the latter, *P. pyrrhus* (MOTU 32) and *P. gilolensis* (MOTU 34) are recovered as monophyletic. *P. jupiter* (MOTU 31, 33 and 35) is recovered as polyphyletic in three subclades one of which comprising specimens from Solomons (1.0/100) is found sister to the rest of the *pyrrhus* complex and is clearly delineated from the rest of the other species. A second subclade from Seram is found sister to the *pyrrhus* complex except *P. pyrrhus* (MOTU 32) whereas the third one from New Britain, New Guinea and New Ireland is sister to *P. andrewsi* (MOTU 36), *P. sempronius* (MOTU 37)

and *P. galaxia* (MOTU 38).. Overall the phylogenetic reconstructions recover mostly congruent and highly supported clades at the inter- and intra-specific levels. The *BEAST cloudogram (Figure 3) from which was derived the input tree used to run the BPP analyses is broadly congruent with the topology presented in Figure 2 and does not showcase a signature of gene inconsistency in the latest clades. Deeper nodes on the other hand show a more contrasted congruence signal as should be expected with the combined use of mitochondrial and nuclear data.

Species delimitation

Across the entire tree, the number of molecular species-level clades varies depending on the methods used, ranging from 28 to 37 against 26 sampled species *sensu* Smiles (1982) (Figures 2, 4). Overall, the bGMYC analyses delineated the smallest numbers of species whereas BPP and PTP methods were splitting the tree in larger numbers of putative species (Figure 4). All BPP runs with the same parameters delivered identical results suggesting a good convergence and consistency of the results. Both analyses with a large ancestral population size resulted in the exact same results whereas the one based on a small ancestral population size delivered slightly different results. The only discrepancy between these two sets of analyses is localized in the *pyrrhus* complex. The analyses with a large ancestral population size recovered all the taxa tested as valid species except *P. galaxia* and *P. sempronius* that were lumped, whereas the analyses based on a small ancestral population size recovered the entire *pyrrhus* complex as one species also including *P. clitarchus*. Two paraphyletic series were found for which all different clades were not clearly delineated by the molecular species delimitation methods. First, we find in both BI and ML that *P. athamas* is paraphyletic due to the inclusion of *P. arja* and *P. hebe* in a derived position. Among the three clades recovered and tested, only the specimens from India are clearly delineated whereas the two remainder clades are not supported by the bGMYC analyses. Second, *P. jupiter* is recovered as polyphyletic within the *pyrrhus* complex with three distinct and well-delineated entities. Populations from the Solomon Islands form a distinct clade in the *pyrrhus* complex and populations from Seram also form a distinct clade.

Discussion

Consistency assessment across molecular species delimitation methods

In this study, we aimed at investigating species boundaries in the charismatic *Polyura* butterflies using molecular species delimitation and test the efficiency of such methods in comparison to morphologically recognized species. To do so, we inferred the first molecular phylogenetic hypothesis for the genus *Polyura*. Based on these phylogenetic results, we assessed the species-level diversity through the use of discovery methods to unfold putative species-level molecular clades and validation methods on candidate species trees to assess the consistency of the methods. Our phylogenetic trees disclosed several regions of interest where morphologically homogeneous species are scattered in different molecular clades of geographical relevance. Overall the different molecular species delimitation methods yielded consistent results. The GMYC approach proved to be much more conservative in a Bayesian framework than its original implementation (Pons et al. 2006) that was widely used and could engender oversplit results (see Talavera et al. 2012 for a discussion). Here, we find it difficult to meet the threshold of robustness (≥0.95) because of the use of randomly selected posterior probability trees with different branch lengths especially with the CO1 dataset. However, when looking at the maximum credibility clades, bGMYC delivered comparable results as the ones obtained with other methods and recovered most morphologically delineated species from Smiles (1982). The use of bGMYC with CO1 or ND5 yielded very similar results although clade support was generally higher with ND5. Among the only highly supported clades recovered by the method, three are surprisingly intra-specific clades which are discussed below. The BPP method based on a *BEAST species tree recovered the largest number of putative species with high support for each of them. Although BPP was the method that delineated the maximal number of species in our case it actually cross-validated most species from Smiles (1982). In particular it was the only method to recover *P. arja* with strong support, a species for which there are strong morphological characters described. The use of extremely loose priors for the input parameters of the models proved to give highly stable results except in the case of the *pyrrhus* complex. In this case, the model including small initial population sizes yielded a unique clade containing all species from Smiles (1982). Under the scenario of a large initial population size, all species were found as independent lineages. Globally, this method was highly congruent with the other ones and especially with PTP. The latter performed very well at delineating taxa from Smiles (1982) although it did not detect *P. arja*. Exception made of the *pyrrhus* complex, the four methods

gave identical results for 21 out of 31 MOTU (68%). If we do not consider the bGMYC results obtained with the CO1 dataset the congruence score climbs to 81%. The congruence score between PTP and BPP (when assuming a small initial population size) is even higher (94%), an unexpected similarity between a discovery and a validation method resting on extremely different priors and models. Finally, the congruence scores of the different models with the morphological species retained by Smiles (1982) vary from 54% (bGMYC ND5, BPP with small ancestral population size and PTP) up to 69% (BPP with large ancestral population size) demonstrating the need for a substantial part of the clades to be revised taxonomically under a generalized species concept (de Queiroz 2007).

Molecular species delimitation implications

Among the *athamas* group, specimens of *P. schreiber* from the Philippines (MOTU 1) are recovered in a well-supported clade which is delineated by all different methods of species delimitation. Interestingly the subspecies *praedictus* from Palawan is not recovered as being part of this clade. This pattern is supported by the geological affiliation of Palawan to the Sunda shelf whereas the Philippines have an independent and highly complex geological origin (Hall 2011, 2012). The remainder of the *P. schreiber* populations form a very widespread clade distributed from India to Borneo without disjunction, a case already highlighted in several butterfly species of the region (Wilson et al. 2013). Morphologically, Philippines specimens are extremely close to the ones from other regions except for a more slender wing shape. As a result we decide to raise all Philippine populations of *P. schreiber* in a new species for which the name *P. luzonica* stat. rev. is available. All populations from the remainder of the distributional range are kept under the name *P. schreiber*. We decide not to raise the subspecies from Palawan and India as separate entities because we dispose of only one specimen per taxon for which no nuclear gene has been sequenced. Future investigations will be needed to seek a potential demarcation between other populations of this widespread species. In a more derived position of the *athamas* group which is the most complex assemblage of lineages within *Polyura*, Thai and Myanmar populations of *P. agraria* (MOTU 8) are surprisingly found in a derived position in the tree without a clear connection with the Indian or Wallacean *P. agraria* populations (Figure 2). Specimens from Thailand and Myanmar are found in a well-supported clade that all species delimitation methods recover with moderate (bGMYC CO1) to strong support (bGMYC ND5 and BPP) for the ones

including robustness supports. These populations form a distinct geographic entity although specimens from this region do not exhibit morphological synapomorphies compared to the rest of the *P. agraria* and *P. athamas* populations. The same goes for the populations of *P. agraria* from the Wallacea (MOTU 9 and 10). Independently evolving lineages may not bear morphological differences because of a recent split as acknowledged in the generalized species concept (de Queiroz 2007). Following this concept, evidence from our results and the supported monophyly of these clades, we raise these three distinct populations to species level. The names *P. piepersianus* stat. rev. and *P. alphius* stat. rev. are available for the Wallacean species found in allopatry aside the Flores Sea respectively on Sulawesi (*P. piepersianus* stat. rev.) and on the Lesser Sunda Islands plus Java and Bali (*P. alphius* stat. rev.). Populations from Thailand and Myanmar form a distinct cryptic lineage and are described as a new species under the name *P. paulettae* nov. sp. whilst populations from India are kept under the name *P. agraria*. These four populations are monophyletic and likely represent recent independently evolving lineages presenting no clear morphological variation (Smiles 1982). Finally, Indian specimens of *P. athamas* (MOTU 11) are found in a well-delineated clade recovered in all analyses with different levels of support. This species is one of the most complex cases in *Polyura* because of its large distributional range and the morphological homogeneity of its populations. However, as for the populations of *P. agraria sensu* Smiles (1982) we decide to adopt the generalized species concept and raise Indian populations to species-level with the available name *P. bharata* stat. rev. We have gathered a relatively comprehensive geographical sampling for *P. athamas sensu* Smiles (1982) except for the easternmost populations in the Philippines and as a result we believe that the clear demarcation of Indian specimens from the rest of the representatives is not an artefact driven by sampling bias (Irwin 2002). One can observe the striking homogeneity between most of the members of the *athamas* group except *P. luzonica* and *P. schreiber* with two main wing patterns surprisingly found across the tree. For instance, *P. hebe* exhibits a wing pattern close to the one found in *P. jalysus* and *P. moori* with a hindwing upperside discal band much broader than in the other species often reaching the base of the wings (Appendix 1). In fact Smiles (1982) considered these three species as a monophyletic clade whereas our molecular phylogeny demonstrates that *P. hebe* is more closely associated with *P. arja* from which it is only slightly divergent genetically although striking morphological differences exist (Figure 2, Appendix 1). Therefore, it seems that the wing pattern observed in *P. hebe* is a convergence or a reversion if the ancestral form of this group was closer to the phenotype of *P. jalysus* and *P. moori* than the rest of the extant species. This kind of morphological homoplasy

exemplifies the need for careful assessment of species boundaries and relationships using other lines of evidence rather than only morphology. Within the other morphological group, only slight divergences have been put forward to delineate taxa including wing shapes, forewing band color variations and absence/presence of subapical spots in certain spaces of the forewing (Smiles 1982). These characters were not retained as diagnostic characters by Smiles (1982) in his morphological phylogeny and based on the study of large series of specimens in the ZSM and author collections from the entire distributional range of each species, it appears that these characters are highly variable and therefore render the delimitation of species tedious and error-prone. Interestingly, the conflict between morphological and molecular characters to delineate species was already recognized in the study of Aduse-Poku et al. (2010) investigating African *Charaxes* species relationships. In the context of recent diversification events where cryptic speciation can be unveiled, it might be difficult to distinguish between phenotypic variability and specific morphological divergences and therefore molecular characters might be more reliable to identify species-level taxa and diagnose them.

In the *eudamippus* group, all species delimitation methods recover the different species from Smiles (1982) as valid entities except for *P. eudamippus* in which all methods but bGMYC based on the CO1 dataset recover two independent lineages. The subspecies *P. eudamippus weismanni* (MOTU 21) found as a separate lineage presents the northernmost distributional range among *Polyura*, in the Ryukyu archipelago. This assemblage of islands is of recent tectonic origin with late connections to the continent before sea-level raised in the Pleistocene (Kimura 2000). In particular, land bridges may have existed between the archipelago and Taiwan until recently (Kimura 2000). As a result, this population could be either the vicariant of continental and Taiwanese populations or the result of dispersal at some point in the evolution of this group. In their study Long et al. (2006) suggested that Taiwanese *P. eudamippus* could be derived from continental populations via glacial land bridge colonization in the past thousands of years. However, the genetic demarcation we observe between *P. eudamippus weismanni* and the remainder of *P. eudamippus* is unlikely to be the result of such a short period of time and we hypothesize that the former is likely the result of a more ancient dispersal event out of China. This is supported by a greater morphological variation of this subspecies compared to the rest of *P. eudamippus* populations including the easternmost ones in Taiwan (Smiles 1982). Based on the results of our species delimitation methods, the strong phylogenetic signal, morphological features and these biogeographical

considerations we raise the populations of *P. eudamippus* from the Ryukyu archipelago as a valid species with the available name *P. weismanni* stat. rev.

Among the *pyrrhus* group *sensu lato*, most species are recovered by the species delimitation methods except for *P. epigenes* and within the *pyrrhus* complex (Figure 2). The investigation of species boundaries in *P. epigenes* revealed the potential existence of two separate lineages on different islands of the Solomons. The two subspecies *P. epigenes epigenes* and *P. epigenes monochromus* are found in the same clade although the populations are respectively restricted to the Southern and Northern part of the archipelago. The subspecies *P. epigenes bicolor* recently described from the island of Malaita (Turlin and Sato 1995) only a few kilometers away from Guadalcanal where *P. epigenes epigenes* occurs, is found to be paraphyletic with very low support in BI and as monophyletic with moderate support in ML. bGMYC consistently failed to recover *P. epigenes bicolor* as a valid species when BPP and PTP provided strong evidence for it. This subspecies has recently been described as a valid species (Müller and Tennent 1998) before being synonymized and downgraded to infraspecific level (Turlin 2001). In the original description (Turlin and Sato 1995) as well as in Turlin (2001), it was emphasized that despite presenting some strong morphological differences with populations from other islands this taxon should remain a subspecies of *P. epigenes*. We argue that in the light of our results a taxonomic reassessment of this lineage might be needed but in order to untangle this question additional taxon sampling and better gene coverage are compulsory. Therefore, and in a conservative manner we decide not to raise this lineage to species level. Within the *pyrrhus* complex, *P. jupiter sensu* Smiles (1982) is recovered as polyphyletic with three clades delineated by the phylogenetic analyses but with different levels of support by species molecular delimitation methods. Specimens from the Solomons (MOTU 31) are found in a very remote location compared to the remainder of the specimens and are recovered in all analyses of species delimitation except for BPP when ancestral population sizes are small. In the latter case this clade is found nested in one large species encompassing the *pyrrhus* complex and *P. clitarchus*. Specimens from the Solomons are an endemic with rather deviant morphology presenting a larger habitus than species from the pyrrhus complex recognized by Smiles (1982) and clear discriminatory morphological features on the upper- and underside of the wings. Although Solomons are close to New Guinea where some populations of *P. jupiter* (MOTU 35) occur, we find no direct relationship between these two taxa that could have been the result of a simple biogeographic event (vicariance during low sea level or active dispersal). Based on obvious morphological characters, geographic singularity as well as the

results of our species delimitation results we raise the populations of *P. jupiter sensu* Smiles (1982) from the Solomons to species rank under the available name *P. attila* stat. rev. We suspect that the subspecies *P. jupiter admiralitatis* which was not sampled in this study should be included within *P. attila* stat. rev. as it shares similar morphological features but since its distribution in the Admiralty archipelago overlaps the distribution of the New Guinean populations of *P. jupiter*, additional data is needed to confirm this taxonomic decision, . The placement of *P. attila* stat. rev. reveals a more complex evolutionary history for the group with possibly an early colonization of Pacific islands as illustrated by the branching of *P. sacco*, *P. caphontis* and *P. gamma* in the *pyrrhus* group *sensu lato*. The colonization of New Guinea and the Moluccas out of Pacific clades in a westward configuration would be in line with recent studies on the paleogeography of the region (Hall 2011, 2012) and the origin and timing of Melanesian clade diversification (Toussaint et al. 2014). However these hypotheses remain to be tested in a proper biogeographic framework. The remainder of the *pyrrhus* complex showcases a much more delicate pattern because most species delimitation methods disagree with morphology but also between each other. Based on strong morphological evidence (Smiles 1982; Turlin and Sato 1995), we argue that it would be unparsimonious to lump all extant species of this complex in one valid species. Moreover, our phylogenetic reconstructions clearly disclose a fine geographic structuration of populations in this group despite presenting moderate nodal support. Considering the validity of each extant species, two problems remain; (*i*) specimens of *P. jupiter sensu* Smiles (1982) from Seram (MOTU 33) are found in a distant clade whereas specimens from the New Guinean archipelago (MOTU 35) are found as sister to *P. gilolensis* (MOTU 34), and (*ii*) specimens from *P. galaxia* (MOTU 38) and *P. sempronius* (MOTU 37) *sensu* Smiles (1982) are found in a same clade without any structuration. Originally populations of *P. jupiter sensu* Smiles (1982) from Seram have been described as an aberration of *P. jupiter* with which they share a common morphology but that allows an easy separation from the sympatrically occurring *P. pyrrhus*. We recover a shallow genetic divergence between the two clades although both are found to be monophyletic. In order to keep *P. jupiter* monophyletic and reach a balanced decision, we describe the populations from Seram as a new species under the name *P. smilesi* sp. nov. and therefore leave the populations from the New Guinean archipelago under the name *P. jupiter*. The sister position of *P. pyrrhus* and *P. smilesi* sp. nov. which look quite alike in spite of the thicker anal vein of *P. pyrrhus* might indicate a case of sympatric speciation although additional taxon sampling and finer molecular techniques would be required to test this hypothesis thoroughly. In particular the relationships between *P.*

smilesi sp. nov. and specimens of *P. jupiter* from the two subspecies *keianus* and *watubela* respectively from Kei Island and Watubela Island east of Seram would be of interest. Finally, *P. galaxia* (MOTU 38) and *P. sempronius* (MOTU 37) *sensu* Smiles (1982) are consistently found as one species across the different methods used and the branching of the multiple specimens clearly indicates that these two taxa represent a unique species. The geographic distribution of both taxa also supports the view of a single widespread species ranging from Lombok to Lord Howe Island about 600km East of the Australian mainland, and encompassing most Lesser Sunda Islands and the entire coastal region of Australia from West to South-East (Figure 1). We therefore synonymize *P. galaxia* with *P. sempronius* in order to reflect our results. The case of the very peculiar Christmas Island endemic *P. andrewsi* is also of interest. This species could possibly belong to *P. sempronius* as the westernmost representative of this widespread species but additional data is also needed here. Interestingly this lineage might be of biogeographic relevance as Christmas Island is much closer to Java than it is to the Lesser Sunda Islands where its closest relative occurs, and no representative of the *pyrrhus* complex has yet reached the western part of Wallace's line in Bali or Java (Figure 1). Overall the *pyrrhus* complex is a geographically highly structured species complex with a distributional range encompassing Australia, the Moluccas, the New Guinean archipelago, the Solomons and Christmas Island. The shallow genetic divergence between members of this group would also be here in line with the late geological assemblage of Melanesia and Wallacea and the great dispersal ability of these insects.

Description of new species and taxonomic reassessment

The descriptions of two new species of *Polyura* are given following examples and recommendations retrieved from state-of.-the-art molecular taxonomy studies (e.g. Jörger and Schrödl 2013; Riedel *et al*. 2013b).

Family: Nymphalidae Rafinesque, 1815
Subfamily: Charaxinae Guenée, 1865
Genus: *Polyura* Billberg, 1820

Polyura paulettae Toussaint **sp. n.**
LSID : XXXXX

Species page: XXXX

Corresponding molecular operational taxonomic unit (MOTU): MOTU 8 (Figure 2).

Types: Holotype (ZSM - Bavarian State Collection of Zoology, Germany): Female from Wang Chin District, Phrae Province, Thailand, IV 1982, collected by local collectors, this is a dry-pinned specimen with voucher ET27 and red HOLOTYPE label (Figure 3). Paratypes (ZSM): three dry-pinned specimens collected by local collectors, with voucher # ET52 (male from the same locality as the holotype), ET61 (male from Shan States, Myanmar), ET94 (male from North Sagaing, Myanmar) and blue PARATYPE label.

Etymology: Named after the first author's grandmother Paule "Paulette" Toussaint, passionate butterfly admirer and collector.

Distribution: Currently known from Myanmar and Northern Thailand.

Diagnosis: A cryptic sister species of *P. agraria sensu* Smiles (1982) and found in the *athamas* group. Very similar to *P. agraria*, *P. alphius* and *P. piepersianus* with which the new species shares a more elongated wing shape than the rest of the *athamas* group. Among these four species *P. agraria* and *P. paulettae* are supposed to present a subapical spot in cell R4 but based on the careful study of multiple museum series of specimens it seems that this character is variable and thus not reliable for diagnosis as suspected by Smiles (1982). The characters suggested in Smiles (1982) and not retained as diagnostic characters were also variable in *P. alphius* and *P. piepersianus* therefore rendering the morphological identification of all four species difficult and error prone. Geographic localities on the other hand are helpful as these four species occur in allopatry (Appendix 1). Phylogenetic relationships also highlight a clear demarcation of these different lineages (Figures 2, 3).

Molecular diagnostic characters compared to the sister clade in the gene alignments (codon position): CO1: 90, C (3rd); 123, C (3rd); 276, A (3rd); 318, G (3rd); 405, A (3rd). ND5: 57, G (3rd); 64, T (1st); 147, G (3rd); 286, G (1st); 295, A (1st).

DNA sequences are EMBL accession numbers xxxxxxxx.

Polyura smilesi Toussaint **sp. n.**

Eriboea jupiter aberration *rectifascia* Talbot, 1920 (unavailable name, introduced at infrasubspecific level).

LSID : XXXXX.

Species page: XXXX.

Corresponding molecular operational taxonomic unit: MOTU 33 (Figure 2).

Types: Holotype (Indonesian Institute of Sciences LIPI, Division of Zoology, Cibinong, Indonesia): Female from Seram, Indonesia, X 1969, collected by local collectors, this is a dry-pinned specimen with voucher ET187 and red HOLOTYPE label (Figure 3). Paratypes (ZSM): three dry-pinned specimens collected by local collectors in the same locality as the holotype, with voucher # ET191 (male), ET192 (male), ET193 (female) and blue PARATYPE label.

Etymology: Named after Robert Leslie Smiles who realized the stunning revision of *Polyura* Nawab butterflies in 1982 and without whom this study would have never been possible.

Distribution: Endemic to Seram.

Diagnosis: Originally mentioned as ab. *rectifascia*, an aberration of *P. jupiter*, *P. smilesi* is morphologically very similar but with less grey-blue scaling on the hindwing. The molecular phylogeny recovers *P. smilesi* more closely related to *P. pyrrhus* than to *P. jupiter* even though a clear morphological demarcation renders false identifications theoretically impossible. As underlined by Smiles (1982), the black lines on the anal veins of the hindwing underside are exaggeratedly thick in *P. pyrrhus*, a unique character in *Polyura* which is not found in *P. jupiter* or *P. smilesi*.

Molecular diagnostic characters compared to the sister clade in the gene alignments (codon position): CO1: 145, A (1st); ND5: 345, T (3rd); 360, C (3rd).

We additionally propose the following taxonomic changes following our results with updated distributional ranges and habitus pictures presented in Appendix 1:

Polyura alphius Staudinger, 1886, **stat. rev.: species propria**

Polyura attila Grose-Smith, 1889, **stat. rev.: species propria**

Polyura bharata Felder & Felder, 1867, **stat. rev.: species propria**

Polyura luzonica Rothschild, 1899 (originally described in *Eulepis*), **stat. rev.: species propria**

Polyura piepersianus Martin, 1924, **stat. rev.: species propria**

Polyura weismanni Fritze, 1894, **stat. rev.: species propria**

Polyura sempronius Fabricius, 1793 (originally described in *Papilio*) = *Polyura galaxia* Butler, 1866, **syn. n.**

Conclusion

Molecular species delimitation methods offer a tantalizing opportunity to accelerate the discovery of biodiversity on our planet. This is especially true for cryptic species complexes that host a substantial fraction of this unknown species richness that cannot be unfolded with traditional morphology-based taxonomic approaches. Here, using molecular species delimitation techniques in addition to geographic and morphological data, we unveil new species in a group of tropical emblematic butterflies occurring in some of the most threatened regions of Earth. We argue that the proper use of molecular species delimitation methods might have at least two cardinal implications; (i) with an accelerated rate of species extinction and the notorious issue of traditional taxonomic description pace, these discovery methods represent an increasingly efficient tool that taxonomists should get a grip at in order to enhance the linkage with formal descriptions, and (ii) whilst anthropogenic erosion of habitats is greatly impacting the sustainability of known and unknown biodiversity, especially in the tropics, showcasing an unsuspected richness of flagship organisms can help capture the attention of conservation planners in order to preserve this ecological legacy for future generations.

Acknowledgements

This work was supported by German Science Foundation (DFG) grants BA2152/6-1, 7-1, 11-1 and 20-1. The Indian work was supported by a Ramanujan Fellowship to KK. The Indian states of Karnataka, Kerala, Nagaland and West Bengal are acknowledged for giving permission to conduct scientific research (Permit numbers: WL10-3781/2012, CWL/GEN/240/522, 3748(3)/WL/4R-1/13), The Department of Environment and Conservation, Papua New Guinea is acknowledged for giving permission to conduct scientific research in (Permit numbers: 013066, 013183, 013339),

References

Aduse-Poku K, Vingerhoedt E, Wahlberg N (2009) Out-of-Africa again: A phylogenetic hypothesis of the genus *Charaxes* (Lepidoptera: Nymphalidae) based on five gene regions. *Molecular Phylogenetics and Evolution*, **53**(2), 463-478.

Astrin JJ, Stüben PE, Misof B et al. (2012) Exploring diversity in cryptorhynchine weevils (Coleoptera) using distance-, character-and tree-based species delineation. *Molecular phylogenetics and evolution*, **63**(1), 1-14.

Bauer AM, Parham JF, Brown RM et al. (2011) Availability of new Bayesian-delimited gecko names and the importance of character-based species descriptions. *Proceedings of the Royal Society B: Biological Sciences*, **278**(1705), 490-492.

Billberg GJ (1820) *Enumeratio insectorum*. 138 pp. Holmiae.

Brooks TM, Mittermeier RA, da Fonsesca GAB et al. (2006) Global biodiversity conservation priorities. *Science*, **313**(5783), 58-61.

Burns JM, Janzen DH, Hajibabaei M, Hallwachs W, Hebert PD (2007) DNA barcodes of closely related (but morphologically and ecologically distinct) species of skipper butterflies (Hesperiidae) can differ by only one to three nucleotides. *Journal of the Lepidopterists Society*, **61**(3), 138-153.

Burns JM, Janzen DH, Hajibabaei M, Hallwachs W, Hebert PD (2008) DNA barcodes and cryptic species of skipper butterflies in the genus *Perichares* in Area de Conservacion Guanacaste, Costa Rica. *Proceedings of the National Academy of Sciences*, **105**(17), 6350-6355.

Burns JM, Janzen DH, Hallwachs W, Hajibabaei M, Hebert PD (2010) Genitalia, DNA barcodes, larval facies, and foodplants place the mimetic species *Neoxeniades molion* in *Rhinthon* (Hesperiidae: Hesperiinae). *Journal of the Lepidopterists Society*, **64**(2), 69-78.

Cardinale BJ, Duffy JE, Gonzalez A et al. (2012) Biodiversity loss and its impact on humanity. *Nature*, **486**(7401), 59-67.

Carstens BC, Pelletier TA, Reid NM, Satler JD (2013) How to fail at species delimitation. *Molecular Ecology*, **22**(17), 4369-4383.

Chacón IA, Montero-Ramírez J, Janzen DH et al. (2012) A new species of *Opsiphanes*. Doubleday, [1849] from Costa Rica (Nymphalidae: Morphinae: Brassolini), as revealed by its DNA barcodes and habitus. *Bulletin of the Allyn Museum*, **166**, 1-15.

Costello MJ, May RM, Stork NE (2013) Can we name Earth's species before they go extinct? *Science*, **339**(6118), 413-416.

De Queiroz K (2007) Species concepts and species delimitation. *Systematic Biology*, **56**(6), 879-886.

Dincă V, Lukhtanov VA, Talavera G, Vila R (2011) Unexpected layers of cryptic diversity in wood white *Leptidea* butterflies. *Nature Communications*, **2**, 324.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**(8), 1969-1973.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**(5), 1792-1797.

Ence DD, Carstens BC (2011) SpedeSTEM: a rapid and accurate method for species delimitation. *Molecular Ecology Resources*, **11**(3), 473-480.

Erixon P, Svennblad B, Britton T, Oxelman B (2003) Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Systematic Biology*, **52**(5), 665-673.

Felsenstein J (2004) *Inferring phylogenies*. Sinauer Associates, Sunderland, Massachusetts.

Fujita MK, Leaché AD (2011) A coalescent perspective on delimiting and naming species: a reply to Bauer et al. *Proceedings of the Royal Society B: Biological Sciences*, **278**(1705), 493-495.

Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C (2012) Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology and Evolution*, **27**(9), 480-488.

Hall R (2011) Australia–SE Asia collision: plate tectonics and crustal flow. *Geological Society, London, Special Publications*, **355**(1), 75-109.

Hall R (2012) Late Jurassic–Cenozoic reconstructions of the Indonesian region and the Indian Ocean. *Tectonophysics*, **570**, 1-41.

Hausmann A, Hebert P, Mitchell A et al. (2009) Revision of the Australian *Oenochroma vinaria* Guenée, 1858 species-complex (Lepidoptera, Geometridae, Oenochrominae): DNA barcoding reveals cryptic diversity and assesses status of type specimen without dissection. Zootaxa, **2239**, 1-21.

Hebert PD, Cywinska A, Ball SL (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **270**(1512), 313-321.

Hebert PD, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(41), 14812-14817.

Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, **27**(3), 570-580.

Henning S (1989) *The Charaxinae butterflies of Africa*. Johannesburg: Aloe Books.

Hooper DU, Adair EC, Cardinale BJ et al. (2012) A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature*, **486**(7401), 105-108.

Huelsenbeck JP, Larget B, Alfaro ME (2004) Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. *Molecular Biology and Evolution*, **21**(6), 1123-1133.

Irwin DE (2002) Phylogeographic breaks without geographic barriers to gene flow. *Evolution*, **56**:2383–2394.

Jörger KM, Schrödl M (2013) How to describe a cryptic species? Practical challenges of molecular taxonomy. *Frontiers in Zoology*, **10**(10.1186), 1742-9994.

Kimura M (2000) Paleogeography of the Ryukyu Islands. *Tropics*, **10**(1), 5-24.

Lanfear R, Calcott B, Ho SY, Guindon S (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, **29**(6), 1695-1701.

Leaché AD, Fujita MK. (2010) Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proceedings of the Royal Society B: Biological Sciences*, **277**(1697), 3071-3077.

Long Y, Wan H, Yan F et al. (2006) Glacial effects on sequence divergence of mitochondrial COII of *Polyura eudamippus* (Lepidoptera: Nymphalidae) in China. *Biochemical Genetics*, **44**(7-8), 359-375.

Mittermeier RA, Gil PR, Hoffman M et al. (2004) *Hotspots revisited: Earth's Biologically Richest and Most Threatened Ecoregions*. CEMEX, Mexico City, Mexico.

Mora C, Tittensor DP, Adl S, Simpson AG, Worm B (2011) How many species are there on Earth and in the ocean? *PLoS Biology*, **9**(8), e1001127.

Müller CJ, Tennent JW (1998) A new species of *Polyura* Billberg, 1820, from the Solomon Islands Lepidoptera Nymphalidae. *Lambillionea*, **98**, 591-594.

Müller CJ, Wahlberg N, Beheregaray LB (2010) 'After Africa': the evolutionary history and systematics of the genus *Charaxes* Ochsenheimer (Lepidoptera: Nymphalidae) in the Indo-Pacific region. *Biological Journal of the Linnean Society*, **100**(2), 457-481.

O'Meara BC, Ané C, Sanderson MJ, Wainwright PC (2006) Testing for different rates of continuous trait evolution using likelihood. *Evolution*, **60**(5), 922-933.

Padial JM, Miralles A, De la Riva I, Vences M (2010) Review: The integrative future of taxonomy. *Frontiers in Zoology*, **7**, 1-14.

Pons J. Barraclough TG, Gomez-Zurita J et al. (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, **55**, 595–609.

Rannala B, Yang Z (2003) Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* **164**, 1645-1656.

Ratnasingham S, Hebert PDN (2013) A DNA-based registry for all animal species: The Barcode Index Number (BIN) System. PLoS ONE, **8**(8), e66213.

Reid NM, Carstens BC (2012) Phylogenetic estimation error can decrease the accuracy of species delimitation: a Bayesian implementation of the general mixed Yule-coalescent model. *BMC Evolutionary Biology*, **12**(1), 196.

Riedel A, Sagata K, Suhardjono, YR, Tänzler R, Balke M (2013a) Integrative taxonomy on the fast track-towards more sustainability in biodiversity research. *Frontiers in Zoology*, **10**(1), 15.

Riedel A, Sagata K, Surbakti S, Tänzler R, Balke M (2013b) One hundred and one new species of *Trigonopterus* weevils from New Guinea. *ZooKeys*, **280**, 1.

Ronquist F, Klopfstein S, Vilhelmsen L et al. (2012) A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Systematic Biology*, **61**(6), 973-999.

Satler JD, Carstens BC, Hedin M (2013) Multilocus species delimitation in a complex of morphologically conserved trapdoor spiders (Mygalomorphae, Antrodiaetidae, *Aliatypus*). *Systematic biology*, **62**(6), 805-823.

Smiles RL (1982) The taxonomy and phylogeny of the genus *Polyura* Billberg (Lepidoptera: Nymphalidae). Bulletin of the British Museum (Natural History) (Entomology), **44**, 115-237.

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**(21), 2688-2690.

Talavera G, Dincă V, Vila R (2013) Factors affecting species delimitations with the GMYC model: insights from a butterfly survey. *Methods in Ecology and Evolution*, **4**(12), 1101-1110.

Tang CQ, Humphreys AM, Fontaneto D, Barraclough TG (2014) Effects of phylogenetic reconstruction method on the robustness of species delimitation using single-locus data. *Methods in Ecology and Evolution*, in press.

Thomas CD, Cameron A, Green RE et al. (2004) Extinction risk from climate change. *Nature*, **427**(6970), 145-14.

Toussaint EFA, Hall R, Monaghan MT et al. (2014) The towering orogeny of New Guinea as a trigger for arthropod megadiversity. *Nature Communications*, **5**, ncomms5001.

Turlin B (2001) Notes sur le genre *Polyura* Billberg dans les archipels du Pacifique (Lepidoptera, Nymphalidae, Charaxinae). *Lambillionea*, **101**(2): 243-247.

Turlin B (2005) *Nymphalidae X, Charaxes I* (Vol. 22), In: Bauer, E., Frankenbach, T. (Eds.), *Butterflies of the World*. Antiquariat Goecke & Evers, Germany.

Turlin B (2007a) *Nymphalidae XII, Charaxes II* (Vol. 25), In: Bauer, E., Frankenbach, T. (Eds.), *Butterflies of the World*. Antiquariat Goecke & Evers, Germany.

Turlin B (2007b) *Nymphalidae XIV, Charaxes III* (Vol. 28), In: Bauer, E., Frankenbach, T. (Eds.), *Butterflies of the World*. Antiquariat Goecke & Evers, Germany.

Turlin B (2009) *Nymphalidae XVII, Charaxes IV* (Vol. 32), In: Bauer, E., Frankenbach, T. (Eds.), *Butterflies of the World*. Antiquariat Goecke & Evers, Germany.

Turlin B (2011) *Nymphalidae XIX, Charaxes V* (Vol. 34), In: Bauer, E., Frankenbach, T. (Eds.), *Butterflies of the World*. Antiquariat Goecke & Evers, Germany.

Turlin B (2013) *Nymphalidae XXII, Charaxes VI* (Vol. 38), In: Bauer, E., Frankenbach, T. (Eds.), *Butterflies of the World*. Antiquariat Goecke & Evers, Germany.

Turlin B (2014) *Nymphalidae XXIV, Palla and Euxanthe* (Vol. 40), In: Bauer, E., Frankenbach, T. (Eds.), *Butterflies of the World*. Antiquariat Goecke & Evers, Germany.

Turlin B, Sato H (1995) Description of new subspecific taxa in the genus *Polyura* Billberg (Lepidoptera, Nymphalidae) from the Moluccas and Solomons Archipelagos. *Futao*, **20**, 20pp.

Wahlberg N, Wheat CW (2008) Genomic outposts serve the phylogenomic pioneers: designing novel nuclear markers for genomic DNA extractions of Lepidoptera. *Systematic Biology*, **57**(2), 231-242.

Wang R, Wan H, Long Y, Lei G, Li S (2004) Phylogenetic analysis of *Polyura* in China inferred from mitochondrial CO II sequences (Lepidoptera: Nymphalidae). *Acta Entomologica Sinica*, **47**(2), 243-247.

Wang RJ, Wan H, Lei GC (2003) DNA sequence divergence of mitochondrial cytochrome oxidase II of *Polyura narcaea* in China. *Acta Scientarum Naturalium Universitatis Pekinensis*, **39**(6), 775– 779.

Wardle DA, Bardgett RD, Callaway RM, Van der Putten WH (2011) Terrestrial ecosystem responses to species gains and losses. *Science*, **332**(6035), 1273-1277.

Williams KJ, Ford A, Rosauer DF et al. (2011) Forests of East Australia: the 35th biodiversity hotspot. In *Biodiversity Hotspots* (pp. 295-310). Springer Berlin Heidelberg.

Wilson JJ, Sing KW, Sofian-Azirun M (2013) Building a DNA Barcode Reference Library for the True Butterflies (Lepidoptera) of Peninsula Malaysia: What about the Subspecies? *PloS one*, **8**(11), e79969.

Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the USA*, **107**(20), 9264-9269.

Zhang J, Kapli P, Pavlidis P, Stamatakis A (2013) A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, **29**(22), 2869-2876.

Data Accessibility

DNA sequences will be submitted to EMBL.

Final DNA sequence assembly will be uploaded as online Supporting Information

Authors Contributions

E.F.A.T and M.B. designed the study; B.T, C.J.M. and E.F.A.T provided most of the samples with additional support from A.H. and K.K.; J.M. conducted the molecular work, E.F.A.T. carried out the analyses, manuscript writing and figure design, all authors commented on the manuscript. The authors declare no competing financial interests.

Tables

Table 1. Information relative to the molecular datasets

	Specimens	Length	Missing data	GC content	MOTU	Species
CO1	196 (95.6%)	471 bp	5.2%	27.9%	38 (100%)	33 (100%)
ND5	196 (95.6%)	417 bp	2.2%	20.3%	38 (100%)	33 (100%)
RPS5	125 (61.0%)	573 bp	5.3%	42.3%	32 (84.2%)	29 (87.9%)
Wingless	116 (56.6%)	396 bp	1.7%	24.8%	33 (86.8%)	30 (90.9%)
Full dataset	205	1857 bp	26.3 %	25.6 %	38 (100%)	33 (100%)

Figures

Figure 1. Distributional ranges of the different *Polyura* species groups in the Indomalayan-Australasian archipelago.

Geographic map of the Indomalayan-Australasian archipelago featuring the distribution of the three *Polyura* species groups spread on twelve out of fourteen hotspots of biodiversity found in the region. Stars do not indicate the exact location of biodiversity hotspots but a rough approximation of their center. Names and delineations of main biogeographic barriers and geological regions are provided. Habitus of six species are presented on the edge of their distribution (see Appendix 2). From left to right: *P. bharata* stat. rev., *P. dehanii*, *P. andrewsi*, *P. posidonius*, *P. luzonica* stat. rev., *P. weismanni* stat. rev., *P. epigenes*. Colored dots under the habitus of each species indicate the species group to which it belongs.

Figure 2. *Polyura* molecular phylogenetic relationships and species boundaries.

Bayesian molecular phylogeny of *CO1*, *ND5*, *Rps5* and *Wingless* gene fragments recovered under MrBayes. Posterior probabilities and bootstrap values from the RAxML analysis are presented for the most important nodes (asterisks indicate PP≥0.95 or BS≥70; - indicate that the node was not recovered in the RAxML topology). Double bar at the root indicates that grey branches have been reduced in length and are not proportional to the scale. Branches within the genera *Charaxes* and *Euxanthe* are shown in orange and branches for *Polyura* are respectively shown in blue, green and red for the *athamas*, *eudamippus* and *pyrrhus* group (*sensu lato*). Picture of habitus are presented for *C. fournierae* and all morphological species

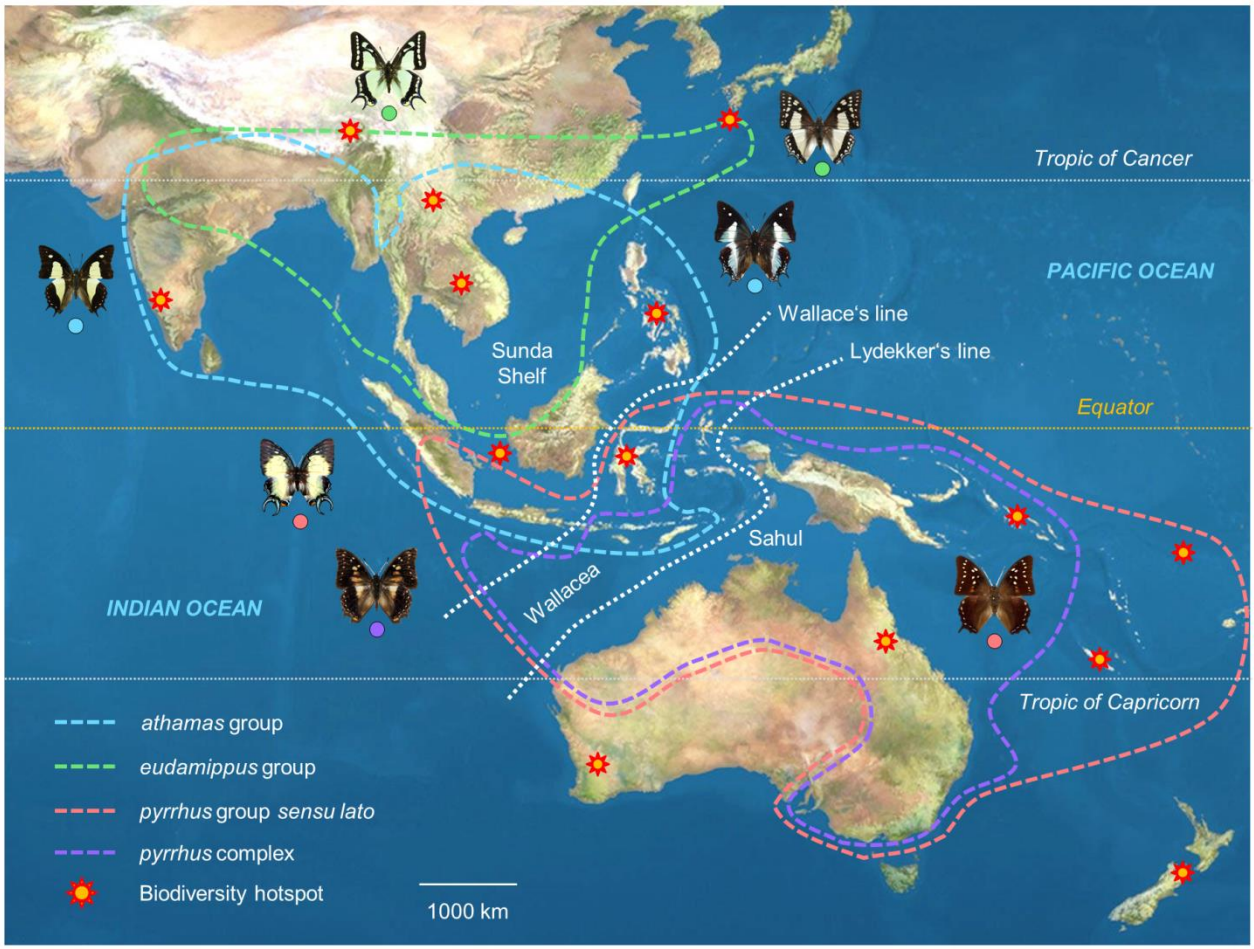
from Smiles (1982) as indicated by the delineation of the grey "Morphology" bars. Within the *pyrrhus* complex, the habitus presented from top to bottom are: *P. pyrrhus*, *P. jupiter*, *P. gilolensis*, *P. andrewsi*, *P. sempronius* and *P. galaxia sensu* Smiles(1982). Rectangles in the 4 other columns at the right present the results of the different species delimitation methods. Numbers on the right correspond to the 38 putative MOTU delineated using bGMYC and PTP and used in the *BEAST and BPP analyses.

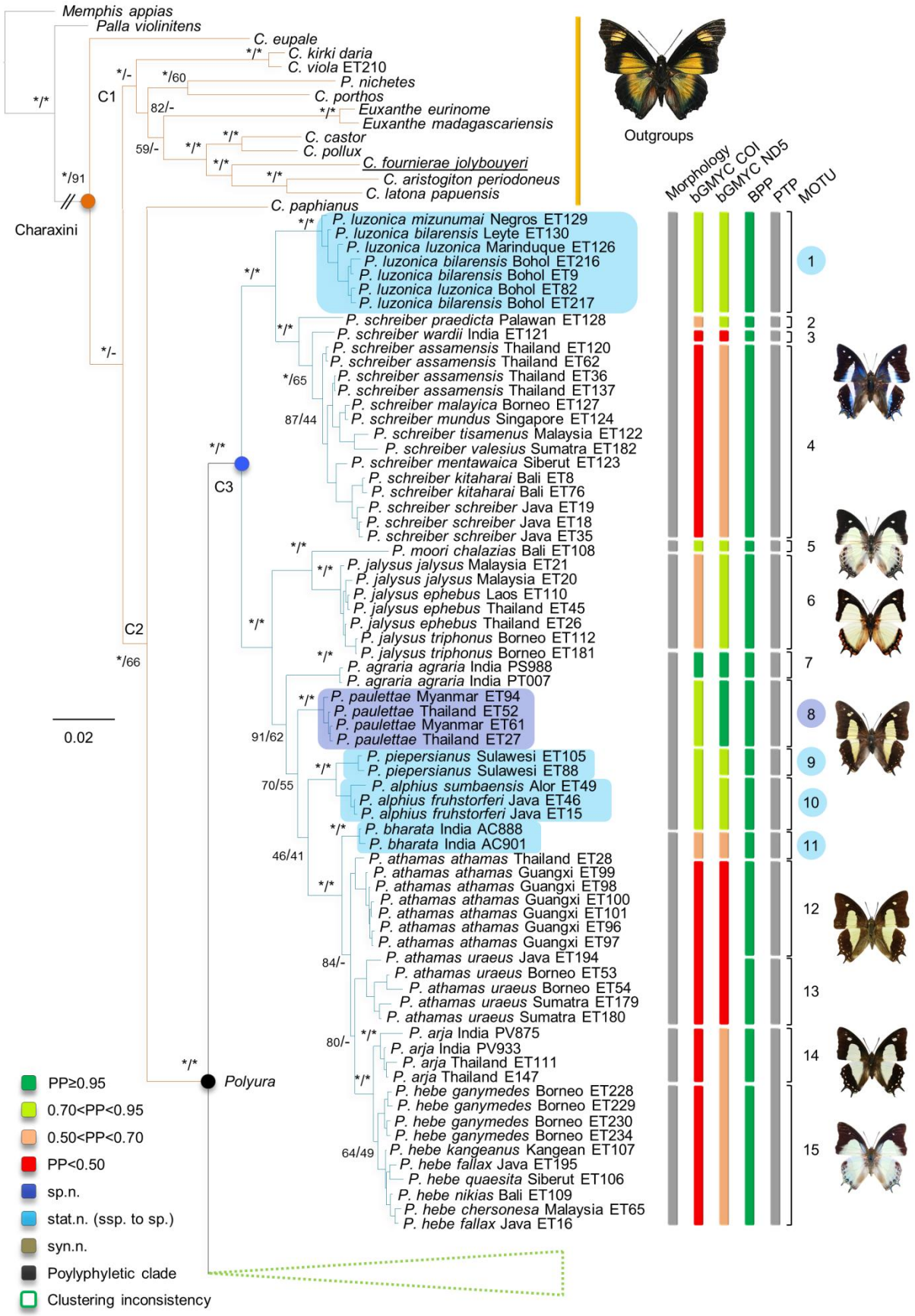
Figure 3. *BEAST cloudogram of candidate species trees.

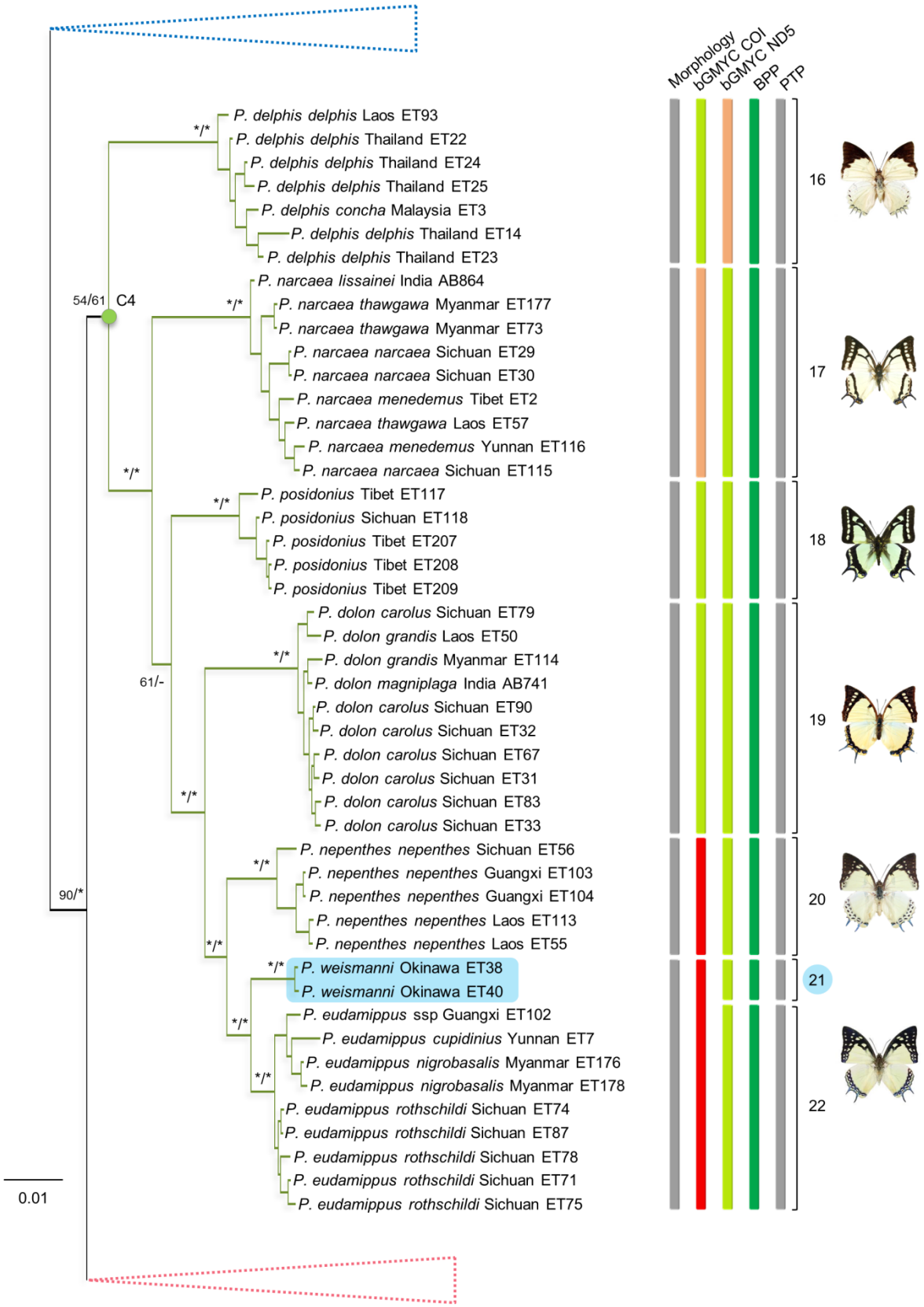
Left side: All post-burn-in posterior probability trees resulting from the *BEAST species tree analysis comprising the 38 putative species with corresponding numbers. Areas where the majority of topologies are in agreement are shown in dark and correspond to well-supported clades. A habitus for each species that was either described or raised to species-level in this study is presented on the right of the figure in the order of its MOTU number appearance in the cloudogram. From top to bottom: *P. paulettae* sp. nov. (MOTU 8), *P. alphius* stat. rev. (MOTU 10), *P. piepersianus* stat. rev. (MOTU 9), *P. bharata* stat. rev. (MOTU 11), *P. luzonica* stat. rev. (MOTU 1), *P. smilesi* sp. nov. (MOTU 33), *P. attila* stat. rev. (MOTU 31) and *P. weismanni* stat. rev. (MOTU 21).

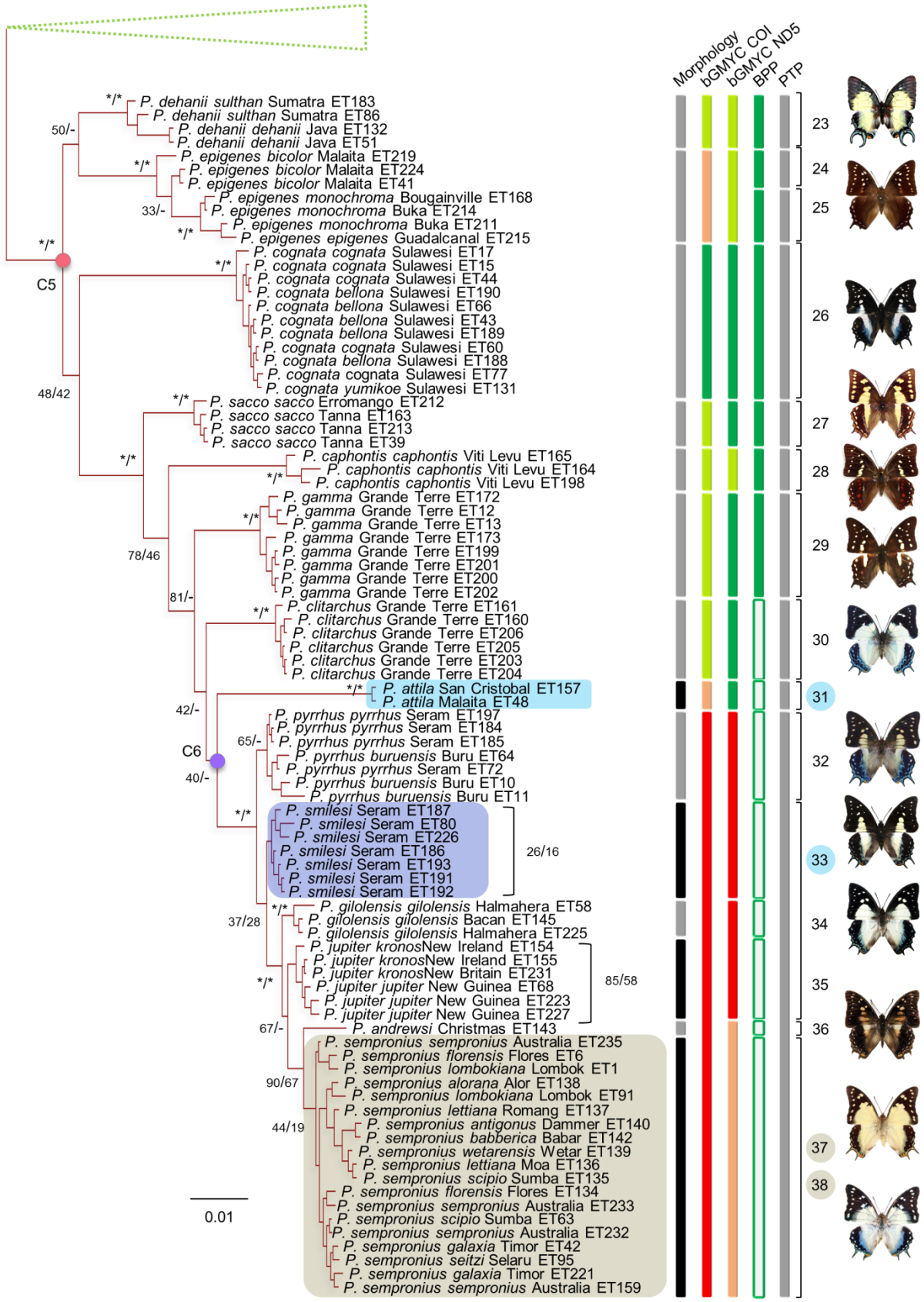
Figure 4. Performance of the different molecular species delimitation methods.

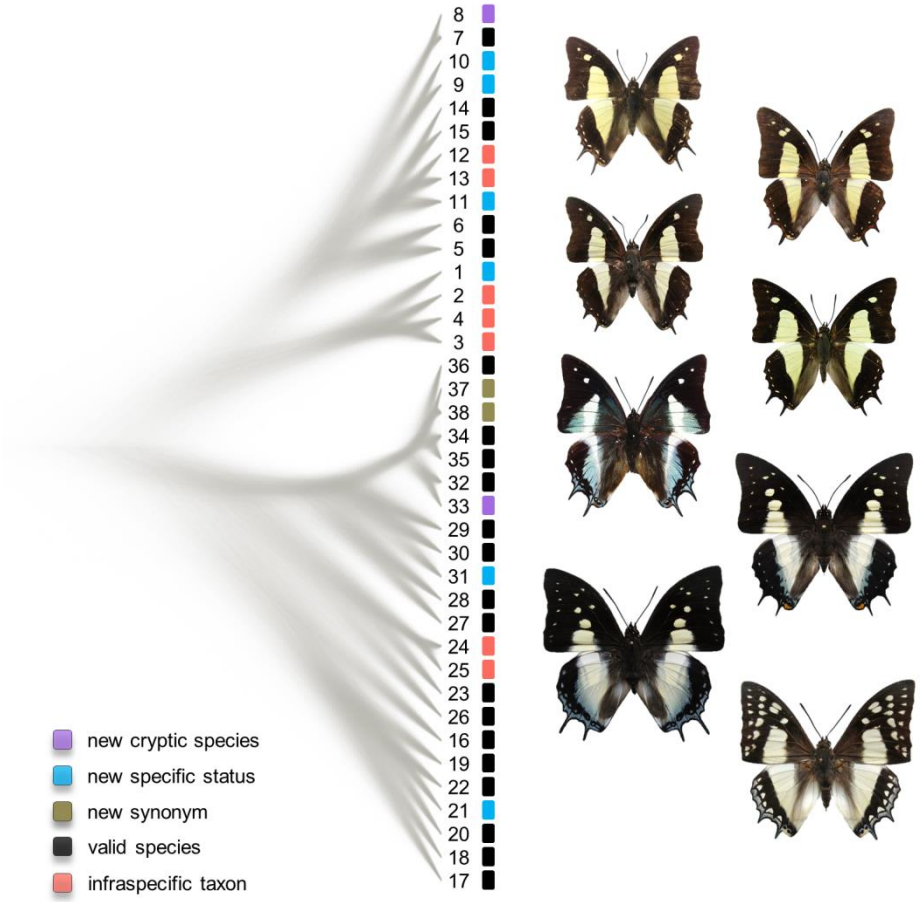
Histogram presenting the results of the different species delimitation methods used on the *Polyura* datasets. The following values are presented: proportion of MOTU recovered compared to the total amount of MOTU delineated (38 MOTU), proportion of morphospecies (26 species) from Smiles (1982) recovered, and proportion of valid species recognized in this study (33 species) recovered. BPP 1 summarizes the results of the models with large initial population size and BPP 2 with small initial population size.

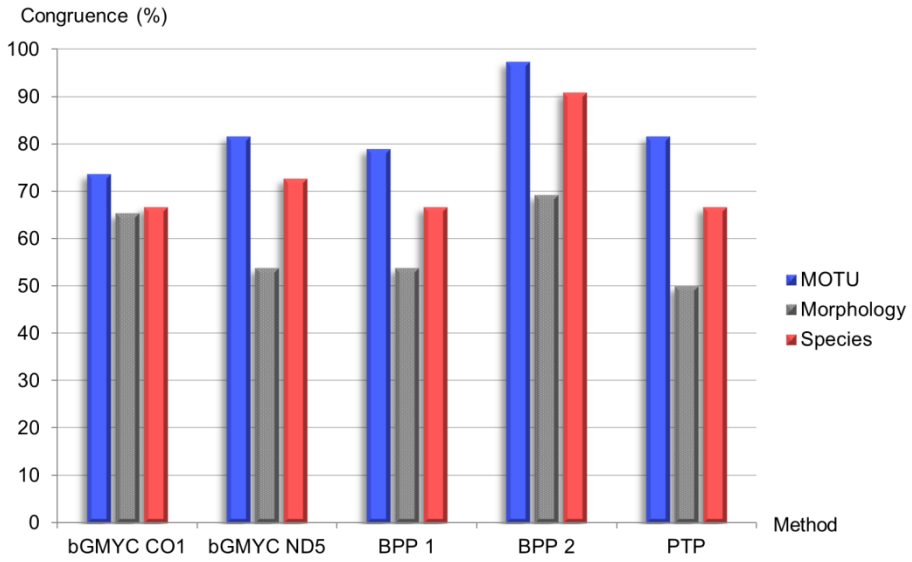

Tropic of Cancer
PACIFIC OCEAN
Wallace's line
Lydekker's line
Sunda
Shelf
Equator
Sahul
Wallacea
INDIAN OCEAN
Tropic of Capricorn
athamas group
eudamippus group
pyrrhus group sensu lato
pyrrhus complex
Biodiversity hotspot
1000 km


Memphis appias
Palla violinitens
C. eupale
C. kirki daria
C. viola ET210
P. nichetes
C. porthos
Euxanthe eurinome
Euxanthe madagascariensis
C. castor
C. pollux
C. fournierae jolybouyeri
C. aristogiton periodoneus
C. latona papuensis
C. paphianus
Outgroups
Charaxini
C1
C2
C3
Polyura
P. luzonica mizunumai Negros ET129
P. luzonica bilarensis Leyte ET130
P. luzonica luzonica Marinduque ET126
P. luzonica bilarensis Bohol ET216
P. luzonica bilarensis Bohol ET9
P. luzonica luzonica Bohol ET82
P. luzonica bilarensis Bohol ET217
P. schreiber praedicta Palawan ET128
P. schreiber wardii India ET121
P. schreiber assamensis Thailand ET120
P. schreiber assamensis Thailand ET62
P. schreiber assamensis Thailand ET36
P. schreiber assamensis Thailand ET137
P. schreiber malayica Borneo ET127
P. schreiber mundus Singapore ET124
P. schreiber tisamenus Malaysia ET122
P. schreiber valesius Sumatra ET182
P. schreiber mentawaica Siberut ET123
P. schreiber kitaharai Bali ET8
P. schreiber kitaharai Bali ET76
P. schreiber schreiber Java ET19
P. schreiber schreiber Java ET18
P. schreiber schreiber Java ET35
P. moori chalazias Bali ET108
P. jalysus jalysus Malaysia ET21
P. jalysus jalysus Malaysia ET20
P. jalysus ephebus Laos ET110
P. jalysus ephebus Thailand ET45
P. jalysus ephebus Thailand ET26
P. jalysus triphonus Borneo ET112
P. jalysus triphonus Borneo ET181
P. agraria agraria India PS988
P. agraria agraria India PT007
P. paulettae Myanmar ET94
P. paulettae Thailand ET52
P. paulettae Myanmar ET61
P. paulettae Thailand ET27
P. piepersianus Sulawesi ET105
P. piepersianus Sulawesi ET88
P. alphius sumbaensis Alor ET49
P. alphius fruhstorferi Java ET46
P. alphius fruhstorferi Java ET15
P. bharata India AC888
P. bharata India AC901
P. athamas athamas Thailand ET28
P. athamas athamas Guangxi ET99
P. athamas athamas Guangxi ET98
P. athamas athamas Guangxi ET100
P. athamas athamas Guangxi ET101
P. athamas athamas Guangxi ET96
P. athamas athamas Guangxi ET97
P. athamas uraeus Java ET194
P. athamas uraeus Borneo ET53
P. athamas uraeus Borneo ET54
P. athamas uraeus Sumatra ET179
P. athamas uraeus Sumatra ET180
P. arja India PV875
P. arja India PV933
P. arja Thailand ET111
P. arja Thailand E147
P. hebe ganymedes Borneo ET228
P. hebe ganymedes Borneo ET229
P. hebe ganymedes Borneo ET230
P. hebe ganymedes Borneo ET234
P. hebe kangeanus Kangean ET107
P. hebe fallax Java ET195
P. hebe quaesita Siberut ET106
P. hebe nikias Bali ET109
P. hebe chersonesa Malaysia ET65
P. hebe fallax Java ET16
Morphology
bGMYC COI
bGMYC ND5
BPP
PTP
MOTU
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
0.02
PP≥0.95
0.70<PP<0.95
0.50<PP<0.70
PP<0.50
sp.n.
stat.n. (ssp. to sp.)
syn.n.
Poylyphyletic clade
Clustering inconsistency


Morphology
bGMYC COI
bGMYC ND5
BPP
PTP
P. delphis delphis Laos ET93
P. delphis delphis Thailand ET22
P. delphis delphis Thailand ET24
P. delphis delphis Thailand ET25
P. delphis concha Malaysia ET3
P. delphis delphis Thailand ET14
P. delphis delphis Thailand ET23
P. narcaea lissainei India AB864
P. narcaea thawgawa Myanmar ET177
P. narcaea thawgawa Myanmar ET73
P. narcaea narcaea Sichuan ET29
P. narcaea narcaea Sichuan ET30
P. narcaea menedemus Tibet ET2
P. narcaea thawgawa Laos ET57
P. narcaea menedemus Yunnan ET116
P. narcaea narcaea Sichuan ET115
P. posidonius Tibet ET117
P. posidonius Sichuan ET118
P. posidonius Tibet ET207
P. posidonius Tibet ET208
P. posidonius Tibet ET209
P. dolon carolus Sichuan ET79
P. dolon grandis Laos ET50
P. dolon grandis Myanmar ET114
P. dolon magniplaga India AB741
P. dolon carolus Sichuan ET90
P. dolon carolus Sichuan ET32
P. dolon carolus Sichuan ET67
P. dolon carolus Sichuan ET31
P. dolon carolus Sichuan ET83
P. dolon carolus Sichuan ET33
P. nepenthes nepenthes Sichuan ET56
P. nepenthes nepenthes Guangxi ET103
P. nepenthes nepenthes Guangxi ET104
P. nepenthes nepenthes Laos ET113
P. nepenthes nepenthes Laos ET55
P. weismanni Okinawa ET38
P. weismanni Okinawa ET40
P. eudamippus ssp Guangxi ET102
P. eudamippus cupidinius Yunnan ET7
P. eudamippus nigrobasalis Myanmar ET176
P. eudamippus nigrobasalis Myanmar ET178
P. eudamippus rothschildi Sichuan ET74
P. eudamippus rothschildi Sichuan ET87
P. eudamippus rothschildi Sichuan ET78
P. eudamippus rothschildi Sichuan ET71
P. eudamippus rothschildi Sichuan ET75
54/61
C4
/
61/-
90/*
0.01
16
17
18
19
20
21
22


Morphology
bGMYC COI
bGMYC ND5
BPP
PTP
/ P. dehanii sulthan Sumatra ET183
P. dehanii sulthan Sumatra ET86
50/- P. dehanii dehanii Java ET132
P. dehanii dehanii Java ET51
P. epigenes bicolor Malaita ET219
/ P. epigenes bicolor Malaita ET224
P. epigenes bicolor Malaita ET41
P. epigenes monochroma Bougainville ET168
33/- P. epigenes monochroma Buka ET214
P. epigenes monochroma Buka ET211
/ P. epigenes epigenes Guadalcanal ET215
/ C5
P. cognata cognata Sulawesi ET17
/ P. cognata cognata Sulawesi ET15
P. cognata cognata Sulawesi ET44
P. cognata bellona Sulawesi ET190
P. cognata bellona Sulawesi ET66
P. cognata bellona Sulawesi ET43
P. cognata bellona Sulawesi ET189
P. cognata cognata Sulawesi ET60
P. cognata bellona Sulawesi ET188
P. cognata cognata Sulawesi ET77
P. cognata yumikoe Sulawesi ET131
48/42
/ P. sacco sacco Erromango ET212
P. sacco sacco Tanna ET163
P. sacco sacco Tanna ET213
P. sacco sacco Tanna ET39
/ P. caphontis caphontis Viti Levu ET165
/ P. caphontis caphontis Viti Levu ET164
P. caphontis caphontis Viti Levu ET198
P. gamma Grande Terre ET172
/ P. gamma Grande Terre ET12
P. gamma Grande Terre ET13
P. gamma Grande Terre ET173
78/46 P. gamma Grande Terre ET199
P. gamma Grande Terre ET201
P. gamma Grande Terre ET200
P. gamma Grande Terre ET202
81/- P. clitarchus Grande Terre ET161
/ P. clitarchus Grande Terre ET160
P. clitarchus Grande Terre ET206
P. clitarchus Grande Terre ET205
P. clitarchus Grande Terre ET203
P. clitarchus Grande Terre ET204
/ P. attila San Cristobal ET157
P. attila Malaita ET48
42/- P. pyrrhus pyrrhus Seram ET197
P. pyrrhus pyrrhus Seram ET184
65/- P. pyrrhus pyrrhus Seram ET185
C6 P. pyrrhus buruensis Buru ET64
P. pyrrhus pyrrhus Seram ET72
40/- P. pyrrhus buruensis Buru ET10
P. pyrrhus buruensis Buru ET11
/ P. smilesi Seram ET187
P. smilesi Seram ET80
P. smilesi Seram ET226
P. smilesi Seram ET186
P. smilesi Seram ET193
P. smilesi Seram ET191
P. smilesi Seram ET192
26/16
/ P. gilolensis gilolensis Halmahera ET58
37/28 P. gilolensis gilolensis Bacan ET145
P. gilolensis gilolensis Halmahera ET225
P. jupiter kronosNew Ireland ET154
P. jupiter kronosNew Ireland ET155
P. jupiter kronosNew Britain ET231
/ P. jupiter jupiter New Guinea ET68
P. jupiter jupiter New Guinea ET223
P. jupiter jupiter New Guinea ET227
85/58
67/- P. andrewsi Christmas ET143
P. sempronius sempronius Australia ET235
P. sempronius florensis Flores ET6
P. sempronius lombokiana Lombok ET1
90/67 P. sempronius alorana Alor ET138
P. sempronius lombokiana Lombok ET91
P. sempronius lettiana Romang ET137
44/19 P. sempronius antigonus Dammer ET140
P. sempronius babberica Babar ET142
P. sempronius wetarensis Wetar ET139
P. sempronius lettiana Moa ET136
P. sempronius scipio Sumba ET135
P. sempronius florensis Flores ET134
P. sempronius sempronius Australia ET233
P. sempronius scipio Sumba ET63
P. sempronius sempronius Australia ET232
P. sempronius galaxia Timor ET42
P. sempronius seitzi Selaru ET95
P. sempronius galaxia Timor ET221
P. sempronius sempronius Australia ET159
0.01
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38

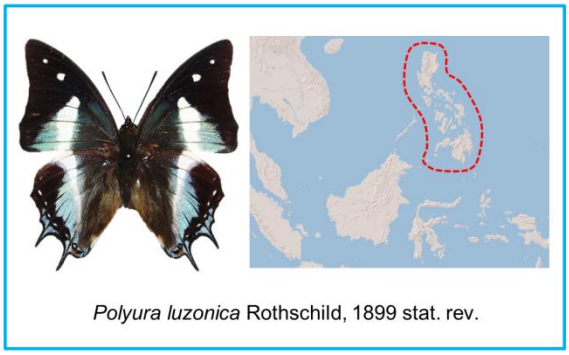

8
7
10
9
14
15
12
13
11
6
5
1
2
4
3
36
37
38
34
35
32
33
29
30
31
28
27
24
25
23
26
16
19
22
21
20
18
17
new cryptic species
new specific status
new synonym
valid species
infraspecific taxon


Congruence (%)
100
90
80
70
60
50
40
30
20
10
0
bGMYC CO1
bGMYC ND5
BPP 1
BPP 2
PTP
Method
MOTU
Morphology
Species

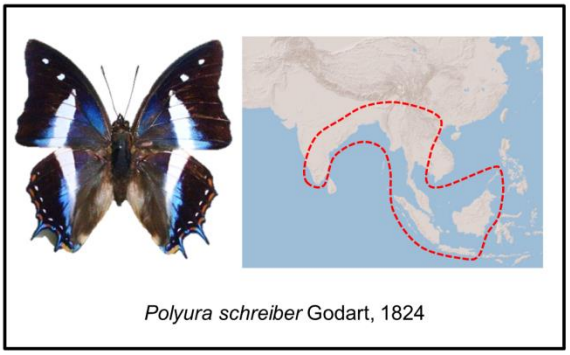
Appendices

Appendix 1. Distributional ranges of all extant *Polyura* species in the Indomalayan / Australasian archipelago.

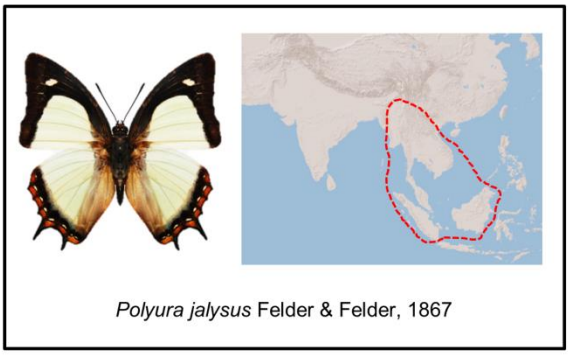
Appendix 2. List of taxa sequenced in this study



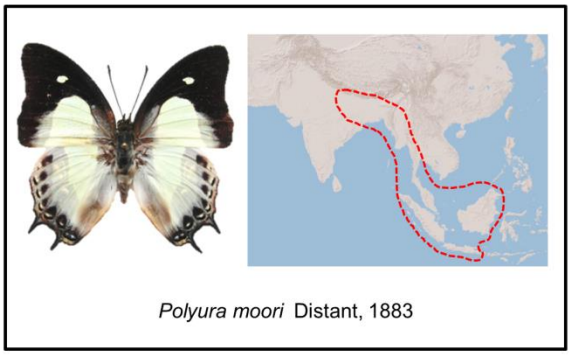
Polyura luzonica Rothschild, 1899 stat. rev.



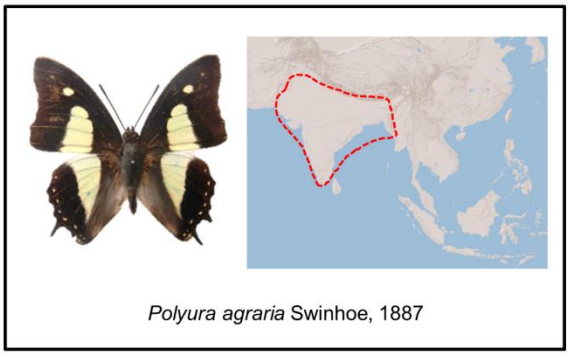
Polyura schreiber Godart, 1824



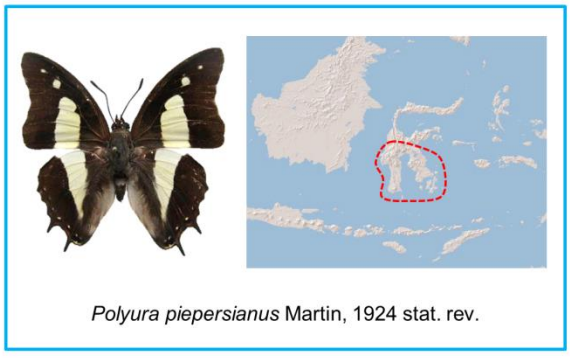
Polyura jalysus Felder & Felder, 1867



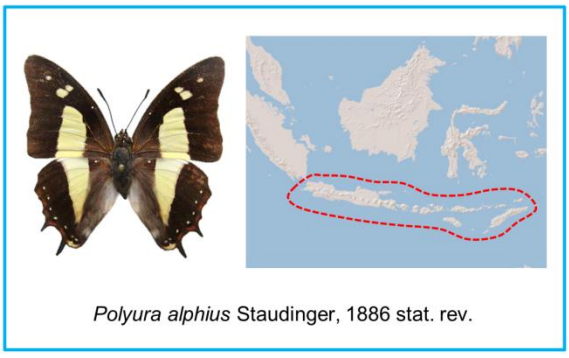
Polyura moori Distant, 1883



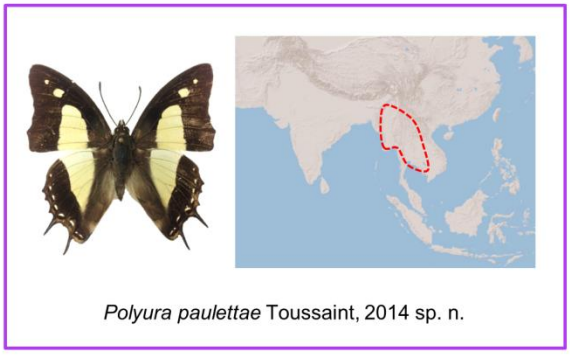
Polyura agraria Swinhoe, 1887



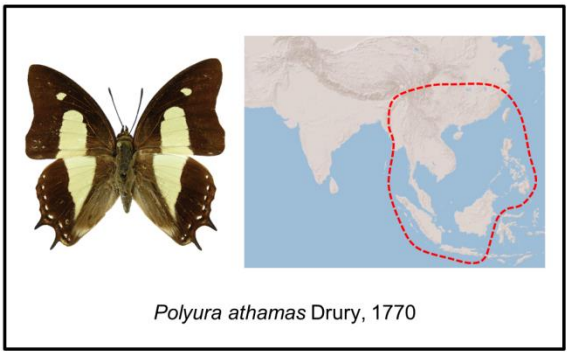
Polyura piepersianus Martin, 1924 stat. rev.



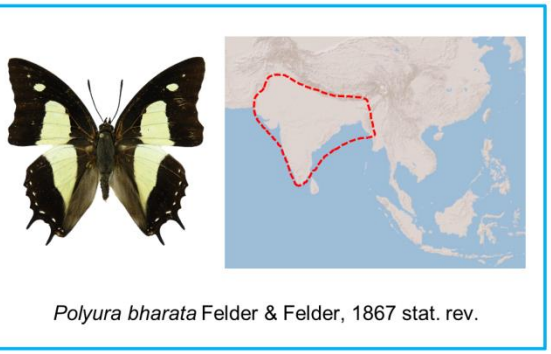
Polyura alphius Staudinger, 1886 stat. rev.



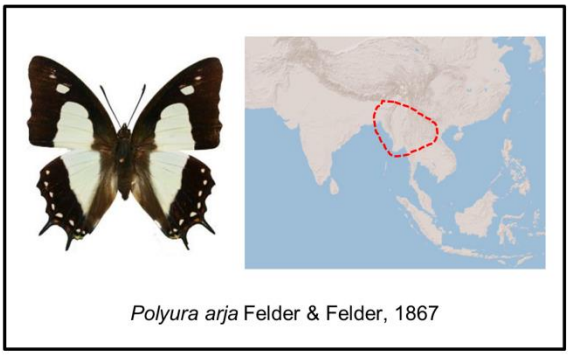
Polyura paulettae Toussaint, 2014 sp. n.



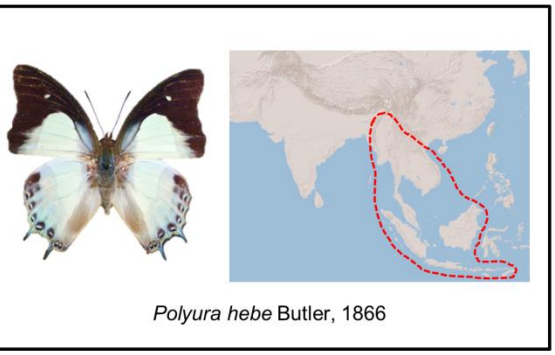
Polyura athamas Drury, 1770



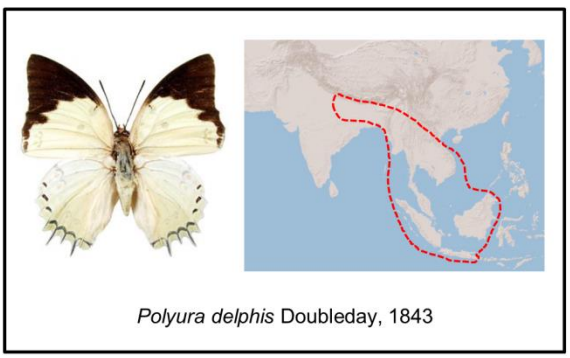
Polyura bharata Felder & Felder, 1867 stat. rev.



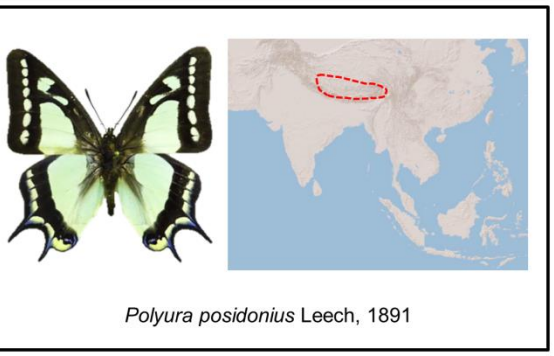
Polyura arja Felder & Felder, 1867



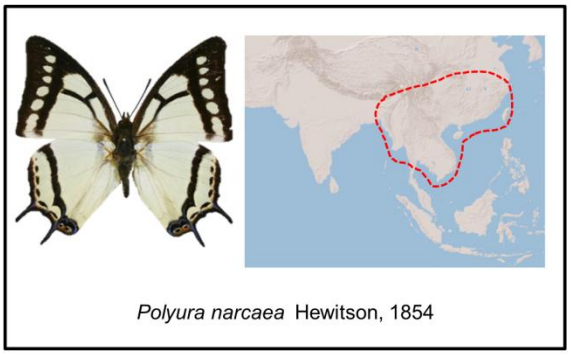
Polyura hebe Butler, 1866



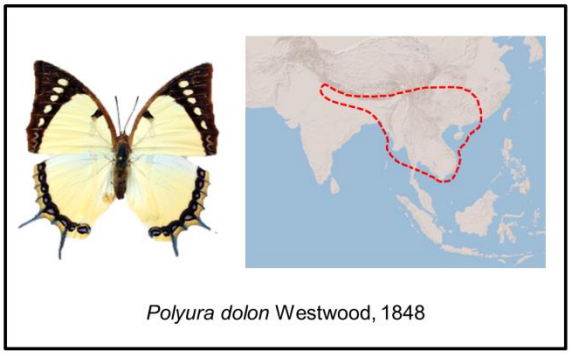
Polyura delphis Doubleday, 1843



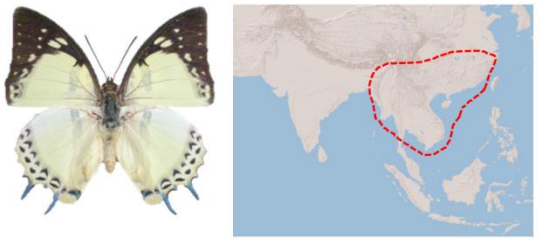
Polyura posidonius Leech, 1891



Polyura narcaea Hewitson, 1854



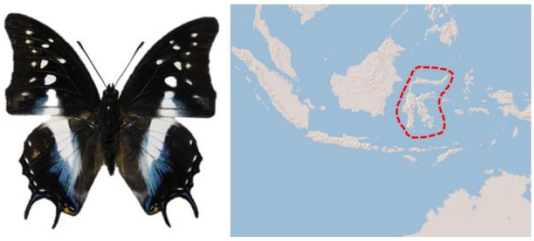
Polyura dolon Westwood, 1848



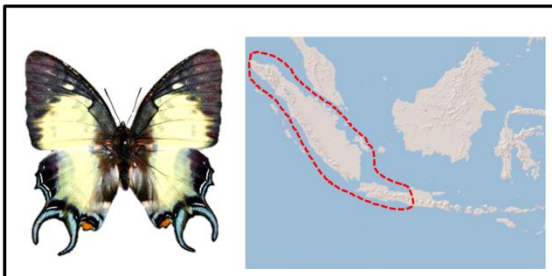
Polyura nepenthes Grose-Smith, 1883

Polyura eudamippus Doubleday, 1843

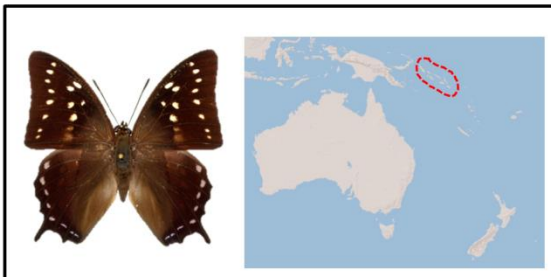
Polyura weismanni Fritze, 1894 stat. rev.



Polyura cognata Snellen van Vollenhoven, 1861



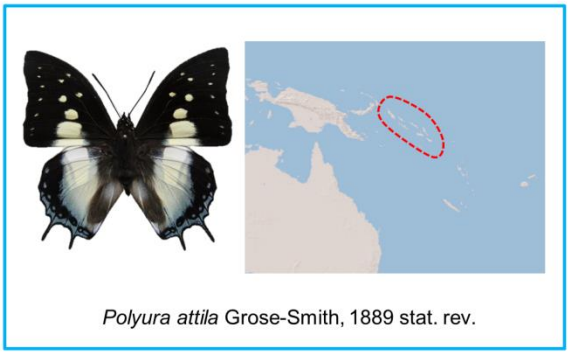
Polyura dehaani Westwood, 1850



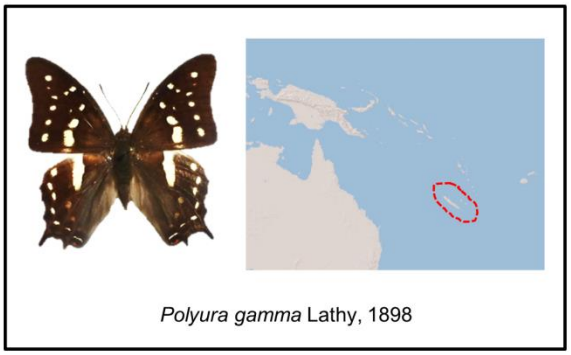
Polyura epigenes Godman & Salvin, 1888

Polyura sacco Smart, 1977

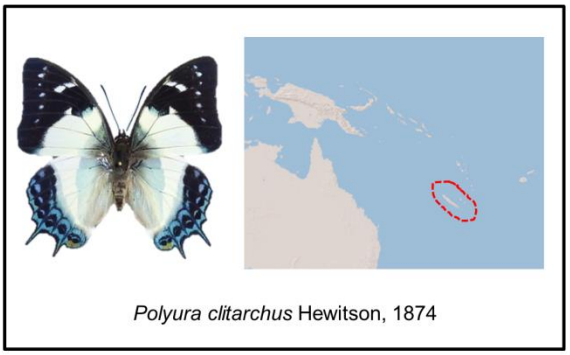
Polyura caphontis Hewitson, 1863



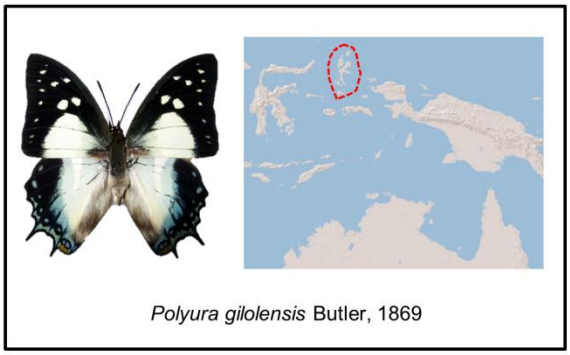
Polyura attila Grose-Smith, 1889 stat. rev.



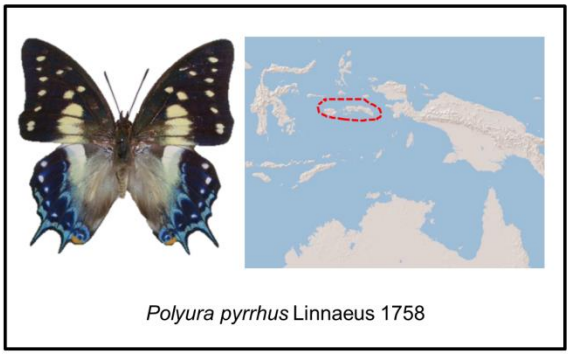
Polyura gamma Lathy, 1898



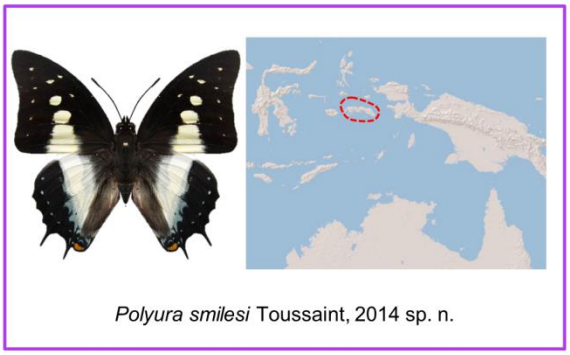
Polyura clitarchus Hewitson, 1874



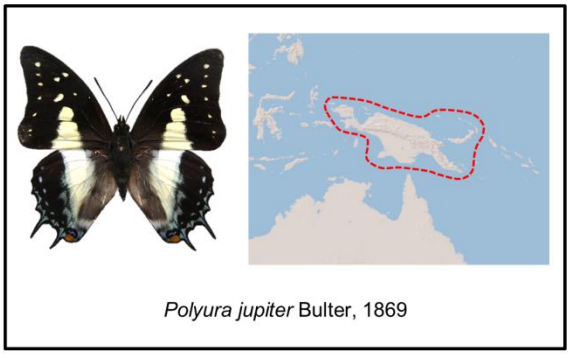
Polyura gilolensis Butler, 1869



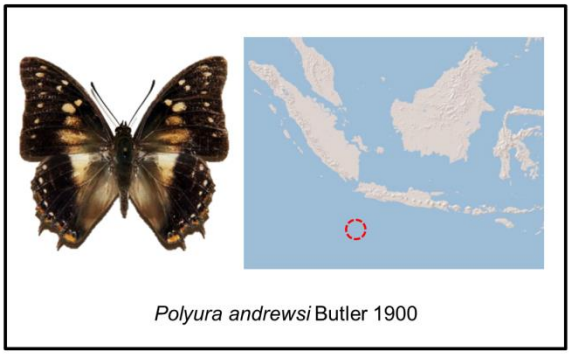
Polyura pyrrhus Linnaeus 1758



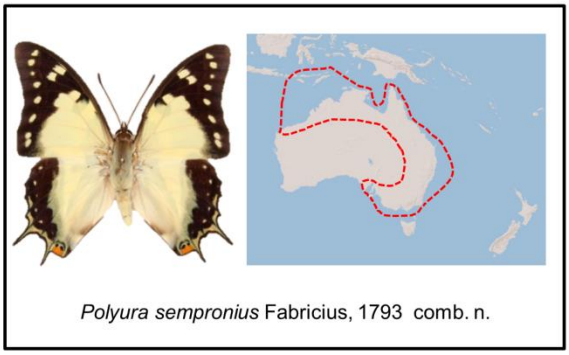
Polyura smilesi Toussaint, 2014 sp. n.



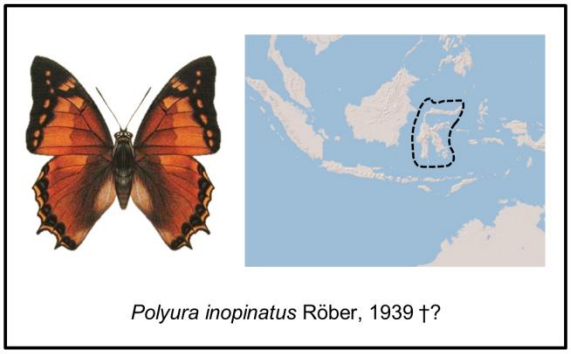
Polyura jupiter Bulter, 1869



Polyura andrewsi Butler 1900



Polyura sempronius Fabricius, 1793 comb. n.



Polyura inopinatus Röber, 1939 †?

Genus	Species	Subspecies	Author	Country	Locality	Coll.	Code
Memphis	*appias*	-	Hübner, 1825	Brazil	Atibai, Sao Paulo	UNI	NW127-6
Palla	*violinitens*	-	Crowley, 1890	Ghana	-	NSG	NW123-19
Charaxes	*aristogiton*	-	Felder, 1867	Thailand	Chiangmai	CJM	CJM-171-001
Charaxes	*castor*	-	Cramer, 1775	England	Stratford Pupae Farm	NSG	NW78-3
Charaxes	*eupale*	-	Drury, 1782	Uganda	Kibale National Park	NSG	NW164-3
Charaxes	*fournierae*	*jolybouyeri*	Vingerhoedt, 1998	Guinea	-	EVC	EV-0006
Charaxes	*kirki*	*daria*	Rothschild, 1903	Ethiopia	-	EVC	EV-0056
Charaxes	*latona*	*papuensis*	Butler, 1869	Papua New Guinea	Watut River, Morobe	CJM	CJM-194-001
Charaxes	*nichetes*	-	Grose-Smith, 1883	Zambia	North of Mwinilunga, Lesombo River	NSG	NW114-14
Charaxes	*paphianus*	-	Ward, 1871	Ghana	Bia	KAC	KAP108
Charaxes	*pollux*	-	Cramer, 1775	Ghana	Kakum National Park	KAC	KAP501
Charaxes	*porthos*	-	Grose-Smith, 1883	Uganda	Kibale National Park	NSG	NW118-11
Charaxes	*viola*	-	Butler, 1866	Central African Republic	-	ETC	ET210
Euxanthe	*eurinome*	-	Cramer, 1775	Ghana	Western Draw River	NSG	NW131-10
Euxanthe	*madagascariensis*	-	Lucas, 1843	Madagascar	-	EVC	EV-0066
Polyura	*agraria*	*agraria*	Swinhoe, 1887	India	Bengaluru Urban, Karnataka	NCBS	PS988
Polyura	*agraria*	*agraria*	Swinhoe, 1887	India	Bengaluru Urban, Karnataka	NCBS	PT007
Polyura	*alphius*	*fruhstorferi*	Röber, 1895	Indonesia	Java	ETC	ET15
Polyura	*alphius*	*fruhstorferi*	Röber, 1895	Indonesia	West Java	ZSM	ET46
Polyura	*alphius*	*sumbaensis*	Swinhoe, 1897	Indonesia	Alor, Nusa Teggara	ZSM	ET49
Polyura	*andrewsi*	-	Butler, 1900	Australia	Christmas Island	BTC	ET143
Polyura	*arja*	-	Felder & Felder, 1867	India	Rajabhatkhawa, Jalpaiguri, West Bengal	NCBS	PV875
Polyura	*arja*	-	Felder & Felder, 1867	India	Rajabhatkhawa, Jalpaiguri, West Bengal	NCBS	PV933
Polyura	*arja*	-	Felder & Felder, 1867	Thailand	Ban Mae Khanin, Neua, Wang Dong	BTC	ET111
Polyura	*arja*	-	Felder & Felder, 1867	Thailand	Chiang Mai	ZSM	ET47
Polyura	*athamas*	*athamas*	Drury, 1770	China	Guilin, Guangxi	ETC	ET100
Polyura	*athamas*	*athamas*	Drury, 1770	China	Guilin, Guangxi	ETC	ET101
Polyura	*athamas*	*athamas*	Drury, 1770	China	Guilin, Guangxi	ETC	ET96
Polyura	*athamas*	*athamas*	Drury, 1770	China	Guilin, Guangxi	ETC	ET97
Polyura	*athamas*	*athamas*	Drury, 1770	China	Guilin, Guangxi	ETC	ET98
Polyura	*athamas*	*athamas*	Drury, 1770	China	Guilin, Guangxi	ETC	ET99
Polyura	*athamas*	*athamas*	Drury, 1770	Thailand	Wang Chin, Phrae	ETC	ET28
Polyura	*athamas*	*uraeus*	Rothschild & Jordan, 1898	Indonesia	West Java	ETC	ET194
Polyura	*athamas*	*uraeus*	Rothschild & Jordan, 1898	Indonesia	Padang, West Sumatra	ETC	ET179
Polyura	*athamas*	*uraeus*	Rothschild & Jordan, 1898	Indonesia	Padang, West Sumatra	ETC	ET180
Polyura	*athamas*	*uraeus*	Rothschild & Jordan, 1898	Malaysia	Ranau, Kinabalu, Sabah, Northern Borneo	ZSM	ET53
Polyura	*athamas*	*uraeus*	Rothschild & Jordan, 1898	Malaysia	Ranau, Kinabalu, Sabah, Northern Borneo	ZSM	ET54
Polyura	*attila*	-	Grose-Smith, 1889	Solomon Islands	Maluu, North Malaita Island	ZSM	ET48
Polyura	*attila*	-	Grose-Smith, 1889	Solomon Islands	San Cristobal Island (Makira)	BTC	ET157
Polyura	*bharata*	-	Felder & Felder, 1867	India	Shendurney, Kollam District, Kerala	NCBS	AC888
Polyura	*bharata*	-	Felder & Felder, 1867	India	Thiruvananthapuram District, Neyyar, Kerala	NCBS	AC901
Polyura	*caphontis*	*caphontis*	Hewitson, 1863	Fiji	Biasevu, Viti Levu Island	BTC	ET165

Polyura	*caphontis*	*caphontis*	Hewitson, 1863	Fiji	Korotogo, Viti Levu Island	BTC	ET164
Polyura	*caphontis*	-	Hewitson, 1863	Fiji	Viti Levu	BTC	ET198
Polyura	*clitarchus*	-	Hewitson, 1874	France	Nouméa, New Caledonia	BTC	ET160
Polyura	*clitarchus*	-	Hewitson, 1874	France	Facola, New Caledonia	CMC	ET203
Polyura	*clitarchus*	-	Hewitson, 1874	France	Facola, New Caledonia	CMC	ET204
Polyura	*clitarchus*	-	Hewitson, 1874	France	Facola, New Caledonia	CMC	ET205
Polyura	*clitarchus*	-	Hewitson, 1874	France	Facola, New Caledonia	CMC	ET206
Polyura	*clitarchus*	-	Hewitson, 1874	France	Yahové, New Caledonia	BTC	ET161
Polyura	*cognata*	*bellona*	Tsukada, 1991	Indonesia	Bantimurung, Maros, Southern Sulawesi	ZSM	ET43
Polyura	*cognata*	*bellona*	Tsukada, 1991	Indonesia	Bantimurung, Maros, Southern Sulawesi	ZSM	ET66
Polyura	*cognata*	*bellona*	Tsukada, 1991	Indonesia	Bantimurung, Maros, Southern Sulawesi	ETC	ET188
Polyura	*cognata*	*bellona*	Tsukada, 1991	Indonesia	Bantimurung, Maros, Southern Sulawesi	ETC	ET189
Polyura	*cognata*	*bellona*	Tsukada, 1991	Indonesia	Bantimurung, Maros, Southern Sulawesi	ETC	ET190
Polyura	*cognata*	*cognata*	Snellen & Vollenhoven, 1861	Indonesia	Camba, Southern Sulawesi	ZSM	ET44
Polyura	*cognata*	*cognata*	Snellen & Vollenhoven, 1861	Indonesia	Camba, Southern Sulawesi	ZSM	ET60
Polyura	*cognata*	*cognata*	Snellen & Vollenhoven, 1861	Indonesia	Makki, Sulawesi	ETC	ET5
Polyura	*cognata*	*cognata*	Snellen & Vollenhoven, 1861	Indonesia	Papayato, Northern Sulawesi	ZSM	ET77
Polyura	*cognata*	*cognata*	Snellen & Vollenhoven, 1861	Indonesia	Southern Palopo, Sulawesi	ETC	ET17
Polyura	*cognata*	*yumikoe*	Nishimura, 1984	Indonesia	Peleng Island, Sulawesi	BTC	ET131
Polyura	*dehanii*	*dehanii*	Westwood, 1850	Indonesia	Gunung Halimun, West Java	BTC	ET132
Polyura	*dehanii*	*dehanii*	Westwood, 1850	Indonesia	Gunung Halimun, West Java	ZSM	ET51
Polyura	*dehanii*	*sulthan*	Hagen, 1896	Indonesia	Gunung Sanggul, West Sumatra	ZSM	ET86
Polyura	*dehanii*	*sulthan*	Hagen, 1896	Indonesia	Susuk, North Sumatra	ZSM	ET183
Polyura	*delphis*	*concha*	Snellen & Vollenhoven, 1861	Malaysia	Cameron Highlands	ETC	ET3
Polyura	*delphis*	*delphis*	Doubleday, 1843	Laos	Ban Na Hai, Lak Sao District	BTC	ET93
Polyura	*delphis*	*delphis*	Doubleday, 1843	Thailand	Wang Chin, Phrae	ETC	ET22
Polyura	*delphis*	*delphis*	Doubleday, 1843	Thailand	Wang Chin, Phrae	ETC	ET23
Polyura	*delphis*	*delphis*	Doubleday, 1843	Thailand	Wang Chin, Phrae	ETC	ET24
Polyura	*delphis*	*delphis*	Doubleday, 1843	Thailand	Wang Chin, Phrae	ETC	ET25
Polyura	*delphis*	*delphis*	Doubleday, 1843	Thailand	Wang Chin, Phrae	ETC	ET14
Polyura	*dolon*	*carolus*	Frushstorfer, 1904	China	Maipu, Jiulong County, Sichuan	ZSM	ET67
Polyura	*dolon*	*carolus*	Frushstorfer, 1904	China	Maipu, Jiulong County, Sichuan	ZSM	ET79
Polyura	*dolon*	*carolus*	Frushstorfer, 1904	China	Maipu, Jiulong County, Sichuan	ZSM	ET83
Polyura	*dolon*	*carolus*	Frushstorfer, 1904	China	Maipu, Jiulong County, Sichuan	ZSM	ET90
Polyura	*dolon*	*carolus*	Frushstorfer, 1904	China	Nanyuan (Sichuan) (900-1300m)	ETC	ET31
Polyura	*dolon*	*carolus*	Frushstorfer, 1904	China	Nanyuan (Sichuan) (900-1300m)	ETC	ET32
Polyura	*dolon*	*carolus*	Frushstorfer, 1904	China	Nanyuan (Sichuan) (900-1300m)	ETC	ET33
Polyura	*dolon*	*grandis*	Rothschild, 1899	Laos	Nam Phao post, Lak Sao District	ZSM	ET50
Polyura	*dolon*	*grandis*	Rothschild, 1899	Myanmar	Chudu Razi Hills, Karen State	BTC	ET114
Polyura	*dolon*	*magniplaga*	Frushstorfer, 1904	India	Dzulekie, Kohima District, Nagaland	NCBS	AB741
Polyura	*epigenes*	*bicolor*	Turlin & Sato, 1995	Solomon Islands	Auki, Malaita	CMC	ET219
Polyura	*epigenes*	*bicolor*	Turlin & Sato, 1995	Solomon Islands	Maluu, North Malaita Island	ZSM	ET41
Polyura	*epigenes*	*bicolor*	Turlin & Sato, 1995	Solomon Islands	New Mera Village, Malaita	CMC	ET224

Polyura	*epigenes*	*monochroma*	Niepelt, 1914	Papua New Guinea	Bougainville Island	BTC	ET168
Polyura	*epigenes*	*monochroma*	Niepelt, 1914	Papua New Guinea	Buka Island	CMC	ET211
Polyura	*epigenes*	*monochroma*	Niepelt, 1914	Papua New Guinea	Buka Island	CMC	ET214
Polyura	*epigenes*	*epigenes*	Godman & Salvin, 1888	Solomon Islands	Guadalcanal Island	CMC	ET215
Polyura	*eudamippus*	*cupidinius*	Frushstorfer, 1914	China	Manyan, Yunnan	ETC	ET7
Polyura	*eudamippus*	*nigrobasalis*	Lathy, 1898	Myanmar	Tarung Hka, North Sagaing	ETC	ET176
Polyura	*eudamippus*	*nigrobasalis*	Lathy, 1898	Myanmar	Tarung Hka, North Sagaing	ETC	ET178
Polyura	*eudamippus*	*rothschildi*	Leech, 1892	China	Maipu, Jiulong County, Sichuan	ZSM	ET71
Polyura	*eudamippus*	*rothschildi*	Leech, 1892	China	Maipu, Jiulong County, Sichuan	ZSM	ET74
Polyura	*eudamippus*	*rothschildi*	Leech, 1892	China	Maipu, Jiulong County, Sichuan	ZSM	ET75
Polyura	*eudamippus*	*rothschildi*	Leech, 1892	China	Maipu, Jiulong County, Sichuan	ZSM	ET78
Polyura	*eudamippus*	*rothschildi*	Leech, 1892	China	Maipu, Jiulong County, Sichuan	ZSM	ET87
Polyura	*eudamippus*	ssp	Doubleday, 1842	China	Guilin, Guangxi	ETC	ET102
Polyura	*gamma*	-	Lathy, 1898	France	La Foa, New Caledonia	ETC	ET13
Polyura	*gamma*	-	Lathy, 1898	France	La Foa, New Caledonia	ETC	ET12
Polyura	*gamma*	-	Lathy, 1898	France	Nouméa, New Caledonia	BTC	ET172
Polyura	*gamma*	-	Lathy, 1898	France	Facola, New Caledonia	CMC	ET199
Polyura	*gamma*	-	Lathy, 1898	France	Facola, New Caledonia	CMC	ET200
Polyura	*gamma*	-	Lathy, 1898	France	Facola, New Caledonia	CMC	ET201
Polyura	*gamma*	-	Lathy, 1898	France	Facola, New Caledonia	CMC	ET202
Polyura	*gamma*	-	Lathy, 1898	France	Yahoué, New Caledonia	BTC	ET173
Polyura	*gilolensis*	*gilolensis*	Butler, 1869	Indonesia	Kampung Makran, Labuha, Bacan Island	BTC	ET145
Polyura	*gilolensis*	*gilolensis*	Butler, 1869	Indonesia	North Halmahera Island, Moluccas	ZSM	ET58
Polyura	*gilolensis*	*gilolensis*	Butler, 1869	Indonesia	Halmahera	CMC	ET225
Polyura	*hebe*	*chersonesa*	Frushstorfer, 1898	Malaysia	Cameron Highlands	ZSM	ET65
Polyura	*hebe*	*fallax*	Röber, 1894	Indonesia	West Java	ETC	ET195
Polyura	*hebe*	*fallax*	Röber, 1894	Indonesia	Java	ETC	ET16
Polyura	*hebe*	*ganymedes*	Staudinger, 1886	Indonesia	Long Laai, North-East Kalimantan	CMC	ET228
Polyura	*hebe*	*ganymedes*	Staudinger, 1886	Indonesia	Long Laai, North-East Kalimantan	CMC	ET229
Polyura	*hebe*	*ganymedes*	Staudinger, 1886	Indonesia	Long Laai, North-East Kalimantan	CMC	ET230
Polyura	*hebe*	*ganymedes*	Staudinger, 1886	Indonesia	Long Laai, North-East Kalimantan	CMC	ET234
Polyura	*hebe*	*kangeana*	Frushstorfer, 1903	Indonesia	Kangean Islands, N.E. Java	BTC	ET107
Polyura	*hebe*	*nikias*	Frushstorfer, 1914	Indonesia	West Bali	BTC	ET109
Polyura	*hebe*	*quaesita*	Corbet, 1942	Indonesia	Siberut Island, Off Sulawesi	BTC	ET106
Polyura	*jalysus*	*ephebus*	Frushstorfer, 1914	Laos	Tad Xia Waterfall, Thabor district, North Laos	BTC	ET110
Polyura	*jalysus*	*ephebus*	Frushstorfer, 1914	Thailand	Samoeng Road, Hang-Dong, Chiang-Mai	ZSM	ET45
Polyura	*jalysus*	*ephebus*	Frushstorfer, 1914	Thailand	Wang Chin, Phrae	ETC	ET26
Polyura	*jalysus*	*jalysus*	Felder & Felder, 1867	Malaysia	Tapah Hills	ETC	ET20
Polyura	*jalysus*	*jalysus*	Felder & Felder, 1867	Malaysia	Tapah Hills	ETC	ET21
Polyura	*jalysus*	*triphonus*	Frushstorfer, 1914	Indonesia	Mount Bawang, West Kalimantan, Borneo	ETC	ET181
Polyura	*jalysus*	*triphonus*	Frushstorfer, 1914	Indonesia	West Kalimantan, Borneo	BTC	ET112
Polyura	*jupiter*	*jupiter*	Butler, 1869	Indonesia	Halilo, Tangma District, West Papua	BTC	ET68
Polyura	*jupiter*	*jupiter*	Butler, 1869	Papua New Guinea	Maprik, East Sepik Province	CMC	ET227

Polyura	*jupiter*	*jupiter*	Butler, 1869	Papua New Guinea	Maprik, East Sepik Province	CMC	ET223
Polyura	*jupiter*	*kronos*	Honrath, 1888	Papua New Guinea	Schleinitz Mts, New Ireland	BTC	ET155
Polyura	*jupiter*	*kronos*	Honrath, 1888	Papua New Guinea	Schleinitz Mts, New Ireland	BTC	ET154
Polyura	*jupiter*	*kronos*	Honrath, 1888	Papua New Guinea	Kimbe, West New Britain Province	CMC	ET231
Polyura	*luzonica*	*bilarensis*	Jumalon, 1975	Philippines	Bohol Island	ETC	ET9
Polyura	*luzonica*	*bilarensis*	Jumalon, 1975	Philippines	Bohol Island	ZSM	ET216
Polyura	*luzonica*	*bilarensis*	Jumalon, 1975	Philippines	Bohol Island	ZSM	ET217
Polyura	*luzonica*	*bilarensis*	Jumalon, 1975	Philippines	Mt St Bernard, Leyte island	BTC	ET130
Polyura	*luzonica*	*luzonica*	Rothschild, 1899	Philippines	Bohol Island	ZSM	ET82
Polyura	*luzonica*	*luzonica*	Rothschild, 1899	Philippines	Marinduque Island	BTC	ET126
Polyura	*luzonica*	*mizunumai*	Hanafusa & Sato, 1987	Philippines	Negros Island	BTC	ET129
Polyura	*moori*	*chalazias*	Frushstorfer, 1914	Indonesia	West Bali	BTC	ET108
Polyura	*narcaea*	*lissainei*	Tytler, 1914	India	Nagaland	NCBS	AB864
Polyura	*narcaea*	*menedemus*	Oberthür, 1891	China	Bai Han Chang, Yunnan	BTC	ET116
Polyura	*narcaea*	*menedemus*	Oberthür, 1891	China	Gamtog, Samda City-	ETC	ET2
Polyura	*narcaea*	*narcaea*	Hewitson, 1854	China	Maipu, Jiulong County, Sichuan	BTC	ET115
Polyura	*narcaea*	*narcaea*	Hewitson, 1854	China	Lushan, Sichuan (1000-1200m)	ETC	ET29
Polyura	*narcaea*	*narcaea*	Hewitson, 1854	China	Lushan, Sichuan (1000-1200m)	ETC	ET30
Polyura	*narcaea*	*thawgawa*	Tytler, 1940	Laos	Nam Phao post, Lak Sao District	ZSM	ET57
Polyura	*narcaea*	*thawgawa*	Tytler, 1940	Myanmar	Chud Razi Hills, Kawnglanghpu, Kachin State	ZSM	ET73
Polyura	*narcaea*	*thawgawa*	Tytler, 1940	Myanmar	Chud Razi Hills, Kawnglanghpu, Kachin State	ETC	ET177
Polyura	*nepenthes*	*nepenthes*	Grose-Smith, 1883	China	Guilin, Guangxi	ETC	ET103
Polyura	*nepenthes*	*nepenthes*	Grose-Smith, 1883	China	Guilin, Guangxi	ETC	ET104
Polyura	*nepenthes*	*nepenthes*	Grose-Smith, 1883	China	Maipu, Jiulong County, Sichuan	ZSM	ET56
Polyura	*nepenthes*	*nepenthes*	Grose-Smith, 1883	Laos	Luang Nam Tha, 1000m, North West Laos	BTC	ET113
Polyura	*nepenthes*	*nepenthes*	Grose-Smith, 1883	Laos	Nam Phao post, Lak Sao District	ZSM	ET55
Polyura	*paulettae*	-	Toussaint, 2014	Myanmar	Kalaw, Shan States	ZSM	ET61
Polyura	*paulettae*	-	Toussaint, 2014	Myanmar	Tarung Hka, North Sagaing	ZSM	ET94
Polyura	*paulettae*	-	Toussaint, 2014	Thailand	Wang Chin, Phrae	ZSM	ET27
Polyura	*paulettae*	-	Toussaint, 2014	Thailand	Wang Chin, Phrae	ZSM	ET52
Polyura	*piepersianus*	-	Martin, 1924	Indonesia	Bantimurung, Maros, Southern Sulawesi	ZSM	ET88
Polyura	*piepersianus*	-	Martin, 1924	Indonesia	Bantimurung, Maros, Southern Sulawesi	BTC	ET105
Polyura	*posidonius*	-	Leech, 1891	China	Batang, Ganzi Pref., 2740m, Sichuan	BTC	ET118
Polyura	*posidonius*	-	Leech, 1891	China	Chengdu, 2200m, Tibet	ZSM	ET207
Polyura	*posidonius*	-	Leech, 1891	China	Chengdu, 2200m, Tibet	ZSM	ET208
Polyura	*posidonius*	-	Leech, 1891	China	Chengdu, 2200m, Tibet	ZSM	ET209
Polyura	*posidonius*	-	Leech, 1891	China	Markam, 3500m, South East Tibet	BTC	ET117
Polyura	*pyrrhus*	*buruensis*	Rothschild & Jordan, 1898	Indonesia	Buru, Moluccas	BTC	ET64
Polyura	*pyrrhus*	*buruensis*	Rothschild & Jordan, 1898	Indonesia	Buru, Moluccas	ETC	ET10
Polyura	*pyrrhus*	*buruensis*	Rothschild & Jordan, 1898	Indonesia	Buru, Moluccas	ETC	ET11
Polyura	*pyrrhus*	*pyrrhus*	Linnaeus, 1758	Indonesia	Ceram Island	ETC	ET197
Polyura	*pyrrhus*	*pyrrhus*	Linnaeus, 1758	Indonesia	Ceram Island	ZSM	ET184
Polyura	*pyrrhus*	*pyrrhus*	Linnaeus, 1758	Indonesia	Ceram Island	ZSM	ET185

Polyura	*pyrrhus*	*pyrrhus*	Linnaeus, 1758	Indonesia	Saka Bridge, North Ceram Island	BTC	ET72
Polyura	*sacco*	*sacco*	Smart, 1977	Vanuatu	-	CMC	ET212
Polyura	*sacco*	*sacco*	Smart, 1977	Vanuatu	-	CMC	ET213
Polyura	*sacco*	*sacco*	Smart, 1977	Vanuatu	Lowanatum, Tanna Island, South New Hebrids	ZSM	ET39
Polyura	*sacco*	*sacco*	Smart, 1977	Vanuatu	Lowanatum, Tanna Island, South New Hebrids	BTC	ET163
Polyura	*schreiber*	*assamensis*	Rothschild, 1899	Thailand	Wang Chin, Phrae	ETC	ET36
Polyura	*schreiber*	*assamensis*	Rothschild, 1899	Thailand	Wang Chin, Phrae	ETC	ET37
Polyura	*schreiber*	*assamensis*	Rothschild, 1899	Thailand	Wang Chin, Phrae	BTC	ET120
Polyura	*schreiber*	*assamensis*	Rothschild, 1899	Thailand	Wang Chin, Phrae	ZSM	ET62
Polyura	*schreiber*	*kitaharai*	Hanafusa, 1987	Indonesia	Bali	ETC	ET8
Polyura	*schreiber*	*kitaharai*	Hanafusa, 1987	Indonesia	Singaraja, Bali	ZSM	ET76
Polyura	*schreiber*	*malayica*	Rothschild, 1899	Indonesia	Pontianak, West Kalimantan (Borneo)	BTC	ET127
Polyura	*schreiber*	*mentawaica*	Hanafusa, 1993	Indonesia	Siberut Island, Off Sulawesi	BTC	ET123
Polyura	*schreiber*	*mundus*	Tsukada, 1991	Indonesia	Singkep Island, South Singapore	BTC	ET124
Polyura	*schreiber*	*praedicta*	Schröder & Treadaway, 1980	Philippines	Palawan Island	BTC	ET128
Polyura	*schreiber*	*schreiber*	Godart, 1824	Indonesia	Java	ETC	ET35
Polyura	*schreiber*	*schreiber*	Godart, 1824	Indonesia	Java	ETC	ET18
Polyura	*schreiber*	*schreiber*	Godart, 1824	Indonesia	Java	ETC	ET19
Polyura	*schreiber*	*tisamenus*	Fruhstorfer, 1914	Malaysia	Cameron Highlands	BTC	ET122
Polyura	*schreiber*	*valesius*	Fruhstorfer, 1914	Indonesia	Padang, West Sumatra	ETC	ET182
Polyura	*schreiber*	*wardii*	Moore, 1896	India	Nilgiri Hills, Tanil Nadu (South India)	BTC	ET121
Polyura	*sempronius*	*alorana*	Rothschild, 1898	Indonesia	Alor Island, Nusa Tenggara	BTC	ET138
Polyura	*sempronius*	*antigonus*	Fruhstorfer, 1904	Indonesia	Dammer Island, Nusa Tenggara	BTC	ET140
Polyura	*sempronius*	*babberica*	Fruhstorfer, 1903	Indonesia	Babar Island, Nusa Tenggara	BTC	ET142
Polyura	*sempronius*	*florensis*	Tsukada, 1991	Indonesia	Flores Island, Nusa Tenggara	BTC	ET134
Polyura	*sempronius*	*florensis*	Tsukada, 1991	Indonesia	Flores Island, Nusa Tenggara	ETC	ET6
Polyura	*sempronius*	*galaxia*	Butler, 1866	Indonesia	Camplong, Timor	CMC	ET221
Polyura	*sempronius*	*galaxia*	Butler, 1866	Indonesia	West Timor Island	ZSM	ET42
Polyura	*sempronius*	*lettiana*	Rothschild, 1898	Indonesia	Moa Island, Nusa Tenggara	BTC	ET136
Polyura	*sempronius*	*lettiana*	Rothschild, 1898	Indonesia	Romang Island, Nusa Tenggara	BTC	ET137
Polyura	*sempronius*	*lombokiana*	Tsukada, 1991	Indonesia	Lombok	ZSM	ET91
Polyura	*sempronius*	*lombokiana*	Tsukada, 1991	Indonesia	Lombok	ETC	ET1
Polyura	*sempronius*	*scipio*	Rothschild, 1898	Indonesia	Sumba Island, Nusa Tenggara	ZSM	ET63
Polyura	*sempronius*	*scipio*	Rothschild, 1898	Indonesia	Sumba Island, Nusa Tenggara	BTC	ET135
Polyura	*sempronius*	*seitzi*	Rothschild, 1897	Indonesia	Selaru Island	ZSM	ET95
Polyura	*sempronius*	*wetarensis*	Fabricius, 1793	Indonesia	Wetar Island, Nusa Tenggara	BTC	ET139
Polyura	*sempronius*	*sempronius*	Fabricius, 1793	Australia	Grafton, New-South Wales	CMC	ET233
Polyura	*sempronius*	*sempronius*	Fabricius, 1793	Australia	Gunnedah, New-South Wales	CMC	ET232
Polyura	*sempronius*	*sempronius*	Fabricius, 1793	Australia	Gunnedah, New-South Wales	CMC	ET235
Polyura	*sempronius*	*sempronius*	Fabricius, 1793	Australia	Ningi, Queensland	BTC	ET159
Polyura	*smilesi*	-	Toussaint, 2014	Indonesia	Ceram Island	BTC	ET80
Polyura	*smilesi*	-	Toussaint, 2014	Indonesia	Ceram Island	ZSM	ET191
Polyura	*smilesi*	-	Toussaint, 2014	Indonesia	Ceram Island	ZSM	ET192

Polyura	*smilesi*	-	Toussaint, 2014	Indonesia	Ceram Island	ZSM	ET193
Polyura	*smilesi*	-	Toussaint, 2014	Indonesia	Ceram Island	ZSM	ET186
Polyura	*smilesi*	-	Toussaint, 2014	Indonesia	Ceram Island	LIPI	ET187
Polyura	*smilesi*	-	Toussaint, 2014	Indonesia	Ceram Island, South Moluccas	CMC	ET226
Polyura	*weismanni*	-	Fritze, 1894	Japan	Nago City, Higashie, Okinawa Island	BTC	ET38
Polyura	*weismanni*	-	Fritze, 1894	Japan	Ogimi Village, Okinawa Island	BTC	ET40

Notes: Coll., voucher locality; BTC, Bernard Turlin Research Collection; CMC, Chris Müller Research Collection; ETC, Emmanuel FA Toussaint Research Collection; EVC, Eric Vingerhoedt Research Collection; LIPI, Indonesian Institute of Sciences (Indonesia); NCBS, National Center for Biological Sciences (India); NSG, Nymphalidae Systematics Group Research Collection (Finland); UNI, State University of Campinas (Brazil); ZSM, Bavarian State Collection of Zoology (Germany).

PAPER III

6.2 A new species of diving beetle from Timor

Citation: Balke M, Toussaint EFA, Hendrich L, Hájek J (2013) A new species of the Australian genus *Necterosoma* from Timor (Coleoptera: Dytiscidae: Hydroporini). *Acta Entomologica Musei Nationalis Pragae*, 53(1):65-74.

ACTA ENTOMOLOGICA MUSEI NATIONALIS PRAGAE
Published 15.vii.2013 Volume 53(1), pp. 65–74 ISSN 0374-1036

A new species of the Australian genus *Necterosoma* from Timor (Coleoptera: Dytiscidae: Hydroporini)

Michael BALKE[1,2], Emmanuel F. A. TOUSSAINT[1], Lars HENDRICH[1] & Jiří HÁJEK[3]

[1] Zoologische Staatssammlung, Münchhausenstraße 21, D-81247 München, Germany; e-mails: michael_balke@yahoo.de, hendrich@zsm.mwn.de, Toussaint@zsm.mwn.de

[2] GeoBioCenter, Ludwig-Maximilians-University, Munich, Germany

[3] Department of Entomology, National Museum, Kunratice 1, CZ-148 00 Praha 4, Czech Republic; e-mail: jiri_hajek@nm.cz

Abstract. *Necterosoma timorense* sp. nov. is described from a forest reserve at Mt. Mutis, West Timor. It is the third species of *Necterosoma* W. J. Macleay, 1871 found outside of Australia. *Necterosoma timorense* sp. nov. can be distinguished from all hitherto known species by the more robust and broader body, the distinct subhumeral expansion of female elytra, and the form of male genitalia. The new species is lotic, being collected from small pools of an intermittent stream, partly shaded by eucalypt forest. Altogether 12 species of *Necterosoma* are now described.

Key words. Coleoptera, Dytiscidae, Hydroporinae, *Necterosoma*, new species, West Timor, Indonesia, Australian Region

Introduction

The diving beetle genus *Necterosoma* W. J. Macleay, 1871 is an Australian faunal element of the tribe Hydroporini (subfamily Hydroporinae), with nine described species from Australian continent and two species in New Caledonia (for summary, see Hendrich et al. 2010a). Here, we describe a characteristic new species of *Necterosoma* from Timor, the most spectacular discovery among the collected aquatic beetles of the island.

The island of Timor, northwest of Australia is the largest (28,000 km²) and easternmost of the Lesser Sunda Islands. For reasons of simplification, some authors (e.g. Nilsson 2001, 2013) include Timor in the Oriental zoogeographical region, however, more precisely it belongs to the transitional border area between the Oriental and Australian Regions, frequently called Wallacea (e.g. Lohman et al. 2011). Many parts of the island are mountainous, and the highest point is Mt. Foho Tatamailau (2,963 m) in East Timor. The climate is tropical, with distinct

rainy and dry seasons. The 12,000 hectares of the unique Mt. Mutis (2,427 m) mountain forest is dominated by homogenous stands of *Eucalyptus urophylla* S. T. Blake (Myrtaceae). The forested slopes of Mt. Mutis, the type locality of the new species, are a critical watershed for the island of Timor and play a strong role in the culture and economy of several villages located in and around Mt. Mutis (Lentz & Mallo 1998).

Material and methods

Collections. The specimens included in this study are deposited in the following institutional and private collections:

ANIC Australian National Insect Collection, CSIRO, Canberra, Australia;
HFCB Hans Fery collection, Berlin, Germany, property of NHMW;
LHCB Lars Hendrich collection, Berlin, Germany, property of NHMW;
MZBC LIPI Division of Zoology, Museum Zoologicum Bogoriense, Cibinong, Indonesia;
NMPC Národní muzeum, Prague, Czech Republic;
NHMW Naturhistorisches Museum Wien, Vienna, Austria;
SAMA South Australian Museum, Adelaide, Australia;
ZSM Zoologische Staatssammlung, Munich, Germany.

Morphological observations. The style of the descriptive notes of the new *Necterosoma* follows Watts (1978), Hendrich (2003) and Hendrich et al. (2010a). Photographs were taken with a Leica Photar 1:2 / 25 on bellows attached to a Nikon D700 camera; an image stack was produced with a custom built robotic macrorail. The principal setup is illustrated on our wiki: http://zsm-entomology.de/wiki/Digital_imaging_in_the_beetle_lab. The male genitalia were studied in dry condition; their photographs were taken with an Olympus camera DP 73 attached to an Olympus SZX16 stereomicroscope. Images at different focal planes were combined using the Helicon Focus 5.1.19 software and subsequently adapted with the Adobe Photoshop 9.0 software. The terminology to denote the orientation of the genitalia follows Miller & Nilsson (2003). Exact label data are cited for the material. Additional remarks are found in square brackets.

The following abbreviations are used in descriptions: TL – total length, length from front of head to apex of elytra; TL-h – total length minus head length, length of body from anterior margin of pronotum to apex of elytra; TW – maximum width of body measured at right angles to TL.

DNA extraction and amplification. DNA extractions were carried out on fresh material kept in 96% ethanol using the DNeasy Tissue Kit (Qiagen GmbH, Hilden, Germany). We sequenced a fragment of the cytochrome oxidase subunit 1 (CO1) using standard protocols (http://zsm-entomology.de/wiki/The_Beetle_D_N_A_Lab) for 91 specimens of *Necterosoma* along with three specimens of *Barretthydrus* Lea, 1927 as outgroups. The DNA strands obtained after sequencing were eye-corrected and aligned under Geneious R6 (available from http://www.geneious.com). Based on this matrix of nucleotides, the relationships within the genus *Necterosoma* were inferred in a Bayesian framework using MrBayes 3.2 (Ronquist et al. 2012). We used a partitioning scheme including a partition for each coding position and a GTR+Γ+I model for each partition. The analyses consisted of two independent runs of 4

Markov Chain Monte Carlo running 5 million generations and sampling every 500 cycles in order to calculate posteriori probabilities (PP). The convergence of the runs was assessed under Tracer 1.5 (Rambaut & Drummond 2007) using the shape of the log-likelihood curves and the Effective Sample Size values for each parameter. Once the generations prior to convergence were discarded, we used the remaining topologies to generate a 50% majority-rule consensus tree. In order to test the robustness of the resulting topology, we also performed phylogenetic inference using Parsimony under TNT 1.1 (Goloboff et al. 2008), with the *Tree Ratchet*, *Tree Fusing* and *Tree Drifting* algorithms (Goloboff 1999), and 1000 *Jacknife* replicates (JK).

The *Necterosoma Cox1* sequences were submitted to GenBank during earlier study (Hendrich et al. 2010b), new sequences were submitted under accession numbers HG003878 and HG003879.

Taxonomy

Necterosoma W. J. Macleay, 1871

Necterosoma W. J. Macleay, 1871: 124. Type species: *Necterosoma vittipenne* W. J. Macleay, 1871 (= *Hydroporus penicillatus* Clark, 1862) by subsequent designation of Guignot (1946).

Necterosoma: Sharp (1882): 412 (redescription); Branden (1885): 44 (catalogue); Zimmermann (1919): 145 (key); (1920): 63 (catalogue); Watts (1978): 90 (redescription); Zwick (1979): 179 (redescription); Pederzani (1995): 34 (key); Nilsson (2001): 174 (catalogue); Hendrich et al. (2010a): 154 (redescription, catalogue); Nilsson (2013): 160 (catalogue).

Diagnosis. Small to medium-sized beetles (4.0–5.4 mm), characterised by distinctly pentamerous protarsi, sides of elytra with small subapical spine, males with protibiae notched on inner side, often strongly so. Most species with distinct colour pattern on elytra, and raised elytral striae in three species. From *Sternopriscus* Sharp, 1880, another Australian genus the members of which have pentamerous protarsi, *Necterosoma* differs with 'normal' mesepimeron being in more or less vertical position, at pronounced angle with metanepisternum, and with medially contiguous metacoxal cavities with the interlaminary bridge totally concealed (cf. Zimmerman 1982).

Distribution. Eleven species were recently recognised in the genus: two species are endemic to New Caledonia; nine species occur in Australia (three species also reach Tasmania) (Watts 1978, 2002; Zwick 1979; Hendrich 2003; Hendrich et al. 2010a). Presence of the genus in New Guinea is possible.

Necterosoma timorense sp. nov.

(Figs 1–7)

Type locality. Indonesia, West Timor, Mount Mutis, 09°38′07.44″S, 124°12′48.00″E, ca. 1580 m alt.

Type material. Holotype: ♂: 'Indonesia: Timor, Mt. Mutis, creeks & streams, 1580m, 1.x.1992, 09 38.124S 124 12.800E, Balke (TIM04) [printed]', 'HOLOTYPE, *Necterosoma*, *timorense* sp. nov., Balke, Toussaint, Hendrich & Hajek des. 2013 [red label, printed]' (MZBC). Paratypes: 40 ♂♂ 39 ♀♀, same label data as the holotype (ANIC, HFCB, LHCB, MZBC, NHMW, SAMA, ZSM); 2 specimens with additional 'DNA MB 4488' and 'DNA MB 4489', respectively [green label indicating the specimen with voucher number was used for DNA extraction] (ZSM); 5 ♂♂ 3 ♀♀, 'INDONESIA, W Timor, Bali prov., Soe env., Desa Nenas, Mutis Mts, 20.-28.xi.2012, 1500-1600m, J. Horák leg. [printed]' (NMPC). All paratypes are provided with a red printed paratype label.

Additional material. 9 larvae from 'Indonesia: Timor, Mt. Mutis, creeks & streams, 1580m, 1.x.1992, 09 38.124S 124 12.800E, Balke (TIM04)'.

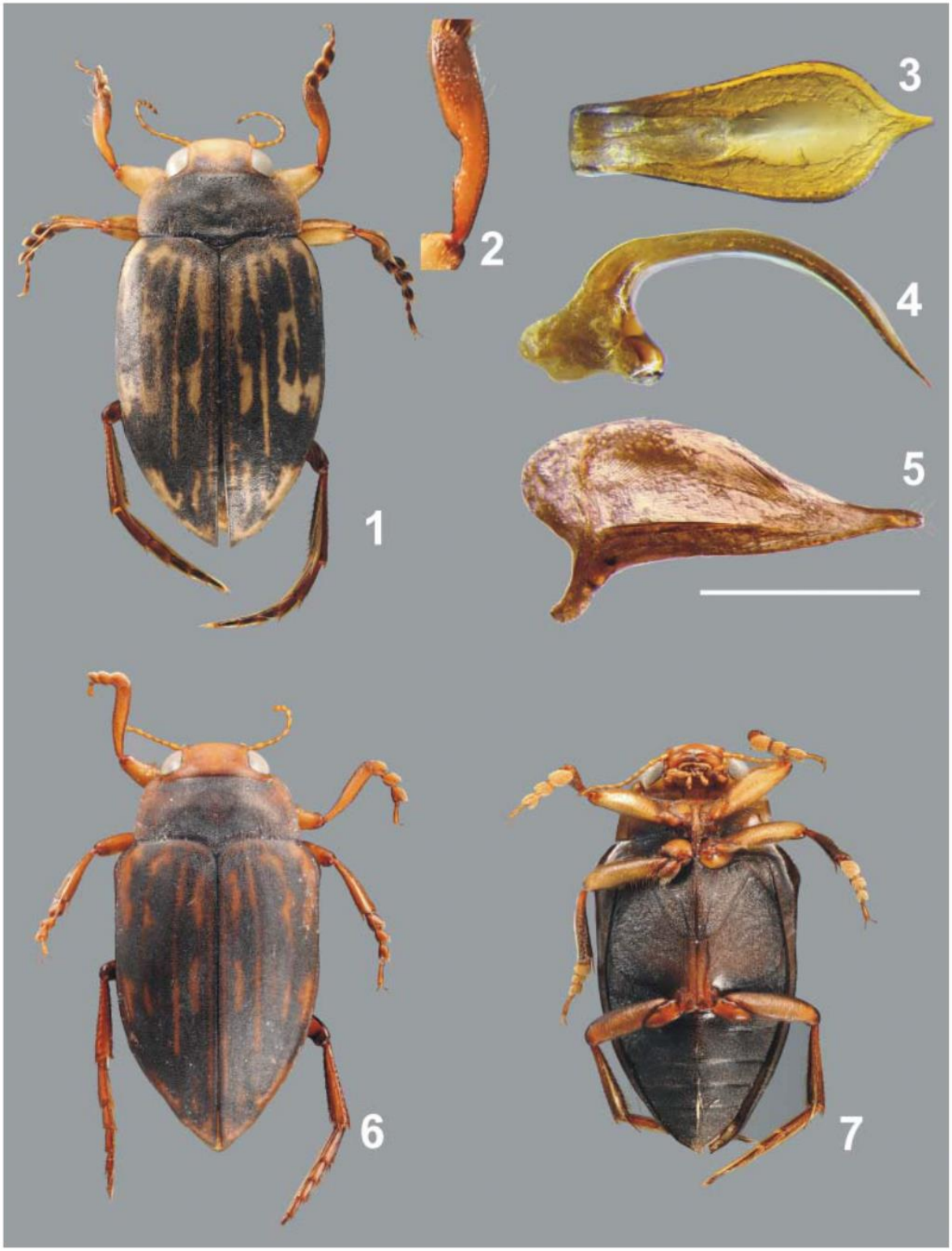
Description. ***Male holotype*.** Comparatively large to other species of genus, yellowish to dark brown species, with black longitudinal markings on elytra. Body elongate, dorsoventrally flattened, widest in middle of elytra. Beetle appearance rather shiny (Fig. 1).

Colouration. Head testaceous, darkened posteriorly; mouth appendages testaceous, terminal palpomere of maxillary palps piceous; antennae testaceous, antennomeres V–XI darkened apically. Pronotum dark brown with rufo-piceous lightenings along sides (broader anterolaterally). Elytra rufo-piceous with well separated black longitudinal stripes. Prosternum and metacoxal processes rufo-piceous, rest of ventral surface black. Legs testaceous; metafemora, tibiae apically and laterally, and tarsi darkened.

Sculpture (Figs 1, 3–4): Head microreticulated, reticulation composed of shallowly impressed polygonal meshes. Punctation simple, punctures irregularly dispersed, with diameter bigger than that of meshes; distance between punctures smaller than their diameter; punctures coarser and denser posteriorly. Head dorsally with two clypeal grooves. Fore margin of clypeus not bordered. Sides of pronotum rounded, weakly bordered; disc convex, medially with shallow longitudinal depression. Base of pronotum with several short longitudinal grooves, and shallow transverse depression along basal margin. Punctation coarse, especially on sides; punctures sometimes confluent, thus microreticulation poorly perceptible. Row of large setigerous punctures presents along anterior margin. Surface of pronotum and elytra covered with short recumbent golden setae. Elytra without carinae. Elytral punctation coarse, microreticulation absent; longitudinal puncture lines poorly perceptible. Subapical spine on lateral margin of elytra small but distinct. Prosternal process lanceolate, distinctly keeled; not narrowed compared to other species of genus. Metacoxal lines relatively close, subparallel anteriorly; metacoxal processes covered with recumbent setae. Punctation of metaventrite, metacoxae and abdominal ventrite I coarse, punctation of abdominal ventrites II–V progressively becoming finer and denser from ventrite II to apical ventrite; microreticulation absent. Abdominal ventrites medially with tuft of long golden setae, apical ventrite covered with recumbent setae; tip of ventrite slightly produced into a weak broad spine covered with setae. Protrochanters on anterior side, mesotrochanters and mesotibiae on posterior side with long golden setae; metatibia with very long natatorial setae on posterior side, metatarsi with natatorial setae on both sides. Pro- and mesotarsi moderately expanded, protarsomere II about 1.2× as wide as long. Protibia moderately expanded with large notch on inner edge near base (Fig. 2). Mesotibia expanded, quite strongly curved.

Genitalia. Median lobe of aedeagus broadly lanceolate in ventral view; its extreme tip truncate (Figs 3–4). Lateral lobes (parameres) with distinct longitudinal carina and fine long striolae on outer side; apically with setae (Fig. 5).

Female. Punctation of dorsal surface finer and denser than in male, thus beetle appearance matt. Microreticulation on pronotum absent. Body widest behind shoulders. Sides of pronotum converging anteriorly, nearly parallel-sided in basal two thirds; anterior angles acute, posterior angles rectangular. Sides of elytra with distinct subhumeral expansion (Figs 6–7). Puncture lines on elytra more distinct than in male. Pro- and mesotarsus only little expanded, tarsomere



Figs 1–7: *Necterosoma timorense* sp. nov. 1 – male habitus; 2 – male protibia; 3 – median lobe in ventral view; 4 – median lobe in lateral view; 5 – right lateral lobe (paramere); 6 – female habitus in dorsal view; 7 – female habitus in ventral view. Scale bar (Figs 3–5): 0.5 mm.

II about as wide as long. Pro- and mesotibia simple, their inner side nearly straight. Apical ventrite as in male but spine more robust and prominent, and setae longer.

Variability. Specimens of type series vary significantly in body colouration, especially in extension of rufo-piceous lightenings of anterolateral part of pronotum, and in colour on elytra, which varies from almost totally black to having black longitudinal stripes well separated.

Figs 8–10: Indonesia, West Timor, track to Mt. Mutis. Rest pools of an intermittent creek in mixed eucalypt forest (*Eucalyptus urophylla* S. T. Blake). Type locality of *Necterosoma timorense* sp. nov.

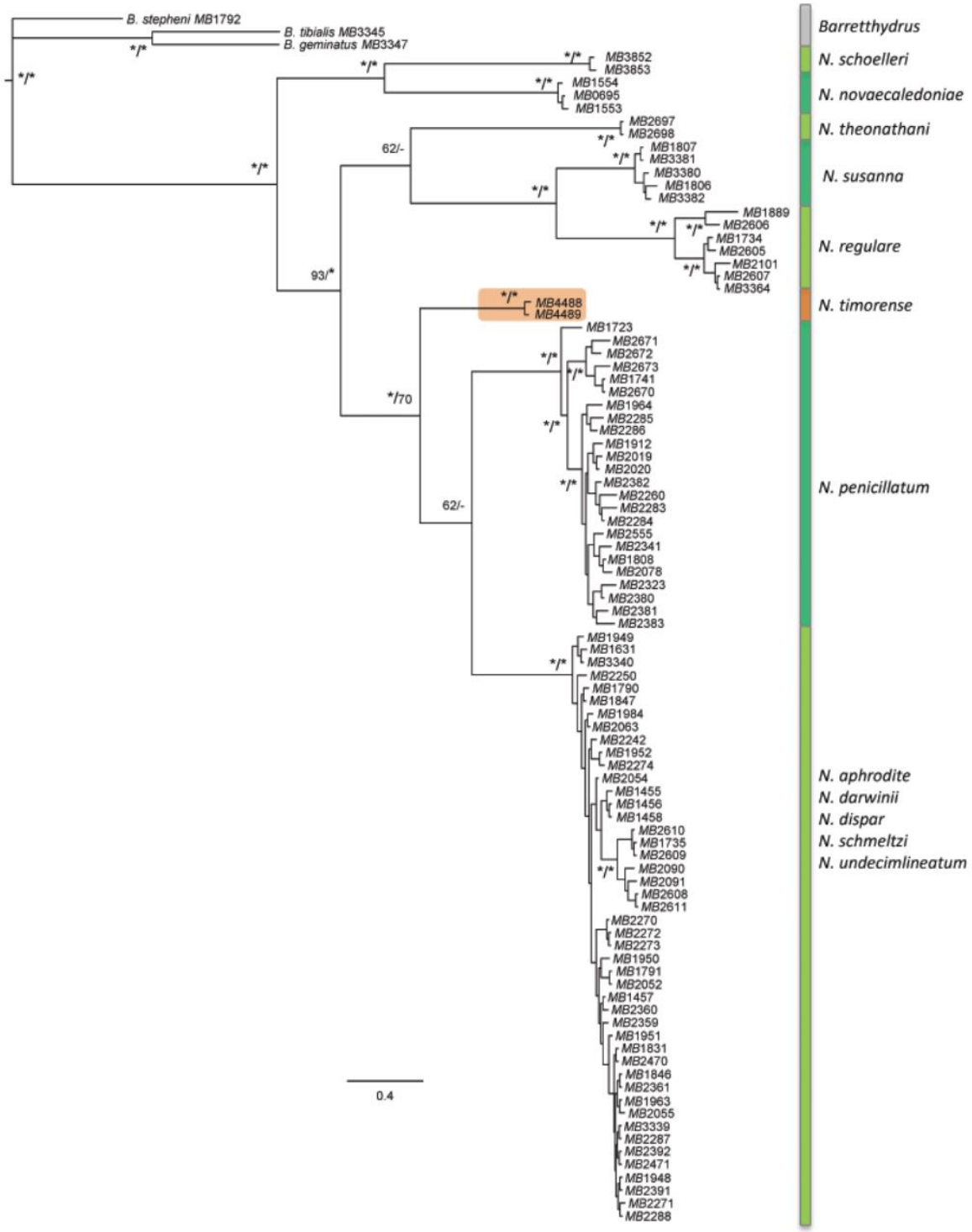


Fig. 11. Phylogenetic tree of the genus *Necterosoma* W. J. Macleay, 1871 under Bayesian Inference. The asterisks indicate a strong support in Bayesian Inference (posterior probability PP > 0.95) or in Parsimony (jackknife JK > 70). The nodes without value are poorly supported (PP < 70 or JK < 50).

Measurements. Males: TL 4.9–5.3 mm (holotype 5.3 mm); TL-h 4.5–4.9 mm (holotype 4.7 mm); TW 2.4–2.6 mm (holotype 2.6 mm). Females: TL 4.8–4.9 mm; TL-h 4.4–4.5 mm; TW 2.4–2.5 mm.

Differential diagnosis. In coloration, paler specimens of *N. timorense* sp. nov. are close to the Australian *N. penicillatum* (Clark, 1862), darker specimens resemble Australian *N. schmeltzi* Sharp, 1882 and *N. aphrodite* Watts, 1978. From all three species, *N. timorense* sp. nov. can be distinguished by a distinct subhumeral expansion of the female elytra (Figs 6–7), the shape of the notch on inner edge of the male protibia (Fig. 2), and the shape of the median lobe (Figs 3–4).

Etymology. Named after the island Timor, where the type material was collected; adjective in the nominative singular.

Collection circumstances. All specimens of *N. timorense* sp. nov. were collected in partly shaded rest pools of intermittent forest streams, at an altitude of 1,500–1,600 m. The beetles were found in shallow water, among roots, twigs and submerged leaf packs (Figs 8–10). *Necterosoma timorense* sp. nov. was found syntopically with the Gyrinidae *Dineutus regimbarti regimbarti* Régimbart, 1882 and *Macrogyrus obliquatus* Aubé, 1838, the Dytiscidae *Copelatus melanogrammus* Régimbart, 1883, *Hydaticus pacificus* Aubé, 1838, *Platynectes* sp., *Rhantus suturalis* (W. S. Macleay, 1825) and *Sandracottus chevrolati* Aubé, 1838, and some unidentified Hydrophilidae of the genera *Anacaena* Thomson, 1859, *Enochrus* Thomson, 1859, and *Helochares* Mulsant, 1844.

Distribution. Only known from the type locality on Mt. Mutis in West Timor (Indonesia) but probably more widespread on the island.

Discussion

Virtually nothing is known about the dytiscid fauna of Timor. Seven species were described from the island (Aubé 1838, Blanchard 1843, Régimbart 1899, Zimmermann 1923), additional four species were recorded by Régimbart (1899); together with the new *Necterosoma* and a record of *Rhantus suturalis* mentioned in the present work, only 12 species of Dytiscidae are known to occur in Timor. However occurrence of many other species is expected, including undescribed taxa from the genera *Laccophilus* Leach, 1815, *Neptosternus* Sharp, 1882 or *Platynectes* Régimbart, 1879 (M. Balke et al., unpublished data).

A preliminary analysis of the mitochondrial DNA CO1 fragment (Fig. 11) suggests that *Necterosoma timorense* sp. nov. is closely related to a clade of six Australian *Necterosoma*, including the most widespread, good flier *N. penicillatum*. We will compile a phylogenetic analysis based on additional genes in the future.

Acknowledgements

We are obliged to Jan Horák (Praha, Czech Republic) for placing specimens at our disposal. We are indebted to Hans Fery (Berlin, Germany) and one anonymous reviewer for helpful

comments on the manuscript. This study was partly supported by the Ministry of Culture of the Czech Republic (DKRVO 2013/12, National Museum, 0002327201).

References

AUBÉ C. 1838: Hydrocanthares et gyriniens. In: DEJEAN P. F.: *Species général des coléoptères de la collection de M. le Comte Dejean. Vol. 6.* Méquignon Père et Fils, Paris, xvi + 804 pp.

BLANCHARD C. É. 1843: Insectes. In: HOMBRON J. B. & JACQUINOT H. (eds.): Atlas d'histoire naturelle zoologie, par MM. Hombron et Jacquinot, chirurgiens de l'expédition. In: *Voyage au pôle Sud et dans l'Océanie sur les corvettes l'Astrolabe et la Zélée exécuté par ordre du roi pendant les années 1837-1838-1839-1840 sous le commandement de M. Dumont-D'Urville capitaine de vaisseau publié sous les auspices du department de la marine. Septième Livraison. Oiseaux pls 4, 5, 6; Poissons pl. 5; Insectes coléoptères pl. 4.* Gide, Paris.

BRANDEN C. VAN DEN 1885: Catalogue des coléoptères carnassiers aquatiques (Haliplidae, Amphizoidae, Pelobiidae et Dytiscidae). *Annales de la Société Entomologique de Belgique* **29**: 5–116.

GOLOBOFF P. A. 1999: Analyzing large data sets in reasonable times: solutions for composite optima. *Cladistics* **15**: 415–428.

GOLOBOFF P. A., FARRIS J. S. & NIXON K. C. 2008: TNT, a free program for phylogenetic analysis. *Cladistics* **24**: 774–786.

GUIGNOT F. 1946: Génotypes des Dytiscoidea et des Gyrinoidea. *Revue Française d'Entomologie* **13**: 112–118.

HENDRICH L. 2003: A new species of Necterosoma Sharp from the Kimberley region in north-western Australia (Coleoptera: Dytiscidae). *Entomologische Zeitschrift* **113**: 152–154, colour figures on pp 144–145.

HENDRICH L., BALKE M. & WEWALKA G. 2010a: Dytiscidae: Hydroporini (Coleoptera). Pp. 149–162. In: JÄCH M. A. & BALKE M. (eds.): Water Beetles of New Caledonia (part 1). *Monographs on Coleoptera* **3**: i–iv + 1–449.

HENDRICH L., PONS J., RIBERA I. & BALKE M. 2010b: Mitochondrial Cox1 sequence data reliably uncover patterns of insect diversity but suffer from high lineage-idiosyncratic error rates. *PLoS ONE* **5(12)**: e14448.

LENTZ C. & MALLO M. 1998: *Environmental management in Gunung Mutis: A case study from Nusa Tenggara, Indonesia.* World Wide Fund for Nature Nusa Tenggara, Cornell University, Nusa Tenggara Community Development Consortium, 17 pp.

LOHMAN D. J., DE BRUYN M., PAGE T., VON RINTELEN K., HALL R., NG P. K. L., SHIH H.-T., CARVALHO G. R. & VON RINTELEN T. 2011: Biogeography of the Indo-Australian Archipelago. *Annual Review of Ecology, Evolution and Systematics* **42**: 205–226.

MACLEAY W. J. 1871: Notes on a collection of insects from Gayndah. *Transactions of the Entomological Society of New South Wales* **2**: 79–205.

MILLER K. B. & NILSSON A. N. 2003: Homology and terminology: Communicating information about rotated structures in water beetles. *Latissimus* **17**: 1–4.

NILSSON A. N. 2001: Dytiscidae (Coleoptera). *World Catalogue of Insects* **3**: 1–395.

NILSSON A. N. 2013: *A World catalogue of the family Dytiscidae or the diving beetles (Coleoptera, Adephaga). Version 1.I.2013.* Distributed as a PDF file via Internet. Available from: http://www2.emg.umu.se/projects/biginst/andersn/

PEDERZANI F. 1995: Keys to the identification of the genera and subgenera of adult Dytiscidae (sensu lato) of the world (Coleoptera Dytiscidae). *Atti dell'Accademia Roveretana degli Agiati, Contributi della Classe di Scienze Matematiche, Fisiche e Naturali, a. 244 (1994), ser. VII* **4(B)**: 5–83.

RAMBAUT A. & DRUMMOND A. J. 2007: *Tracer v1.4.* Available at: http://beast.bio.ed.ac.uk/Tracer.

RÉGIMBART M. 1899: Révision des Dytiscidae de la région Indo-Sino-Malaise. *Annales de la Société Entomologique de France* **68**: 186–367.

RONQUIST F., TESLENKO M., VAN DER MARK P., AYRES D. L., DARLING A., HOHNA S., LARGET B., LIU L., SUCHARD M. A. & HUELSENBECK J. P. 2012: MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**: 539–542.

SHARP D. 1882: On aquatic carnivorous Coleoptera or Dytiscidae. *Scientific Transactions of the Royal Dublin Society, Series 2* **2**: 179–1003 + pls. 7–18.

WATTS C. H. S. 1978: A revision of the Australian Dytiscidae (Coleoptera). *Australian Journal of Zoology, Supplementary Series* **57**: 1–166.

WATTS C. H. S. 2002: Checklist and guides to the identification, to genus, of adults and larval Australian water beetles of the families Dytiscidae, Noteridae, Hygrobiidae, Haliplidae, Gyrinidae, Hydraenidae and the superfamily Hydrophiloidea (Insecta – Coleoptera). Cooperative Research Centre for Freshwater Ecology (Australia). *Identification and Ecology Guide* **43**: 1–110.

ZIMMERMAN J. R. 1982: The Deronectes of the southwestern United States, Mexico, and Guatemala (Coleoptera: Dytiscidae). *Coleopterists Bulletin* **36**: 412–438.

ZIMMERMANN A. 1919: Die Schwimmkäfer des Deutschen Entomologischen Museums in Berlin-Dahlem. *Archiv für Naturgeschichte (A)* **83(12)** (1917): 68–249.

ZIMMERMANN A. 1920: Pars 71: Dytiscidae, Haliplidae, Hygrobiidae, Amphizoidae. In: SCHENKLING S. (ed.): *Coleopterorum Catalogus, Volumen IV.* W. Junk, Berlin, 326 pp.

ZIMMERMANN A. 1923: Neue Schwimmkäfer. *Entomologische Blätter* **19**: 31–40.

ZWICK P. 1979: Notes on the genus Necterosoma (Col., Dytiscidae), with description of N. susanna sp. n. from Australia. *Aquatic Insects* **1**: 179–184.

6.3 A new species of diving beetle from Biak Island

Citation: Balke M, Warikar E, Toussaint EFA, Hendrich L (2013) *Papuadessus baueri* sp.nov. from Biak Island, Papua (Coleoptera: Dytiscidae: Hydroporinae). *Spixiana*, 36(2):283-288.

SPIXIANA	36	2	283–288	München, Dezember 2013	ISSN 0341–8391

Papuadessus baueri spec. nov. from Biak Island, Papua

(Coleoptera, Dytiscidae, Hydroporinae)

Michael Balke, Evi Warikar, Emmanuel F. A. Toussaint & Lars Hendrich

Balke, M., Warikar, E., Toussaint, E. F. A. & Hendrich, L. 2013. *Papuadessus baueri* spec. nov. from Biak Island, Papua (Coleoptera, Dytiscidae, Hydroporinae). Spixiana 36(2): 283–288.

Papuadessus baueri spec. nov. is decribed from Biak Island, Papua. The phylogenetic analysis of DNA sequence data suggested placement in that genus which otherwise contains *P. pakdjoko* Balke, 2001, a species widespread across mainland New Guinea. The new species seems to be endemic to Biak where it was collected from a limestone sinkhole. Important species characters (habitus, median lobe and paramere) are illustrated, and the habitat of *P. baueri* spec. nov. and its water beetle coenosis are briefly outlined.

Michael Balke (corresponding author), Lars Hendrich, Emmanuel F. A. Toussaint, Zoologische Staatssammlung München, Münchhausenstraße 21, 81247 München, Germany; email: michael_balke@yahoo.de

Evi Warikar, Department of Biology, Universitas Cendrawasih, Jayapura, Papua, Indonesia

Introduction

Papuadessus was described by Balke (2001) for the large and conspicuous New Guinea species *P. pakdjoko* (Fig. 1). Molecular phylogenetic investigations by Balke & Ribera (2004) and Ribera et al. (2008) established *Papuadessus* as a well delineated, isolated lineage within the Bidessini, but its closer relatives remain not well established. The species has been shown to be comparably widely distributed across New Guinea. Here, we add a second species to the genus. It appears morphologically rather divergent from *P. pakdjoko* and we decided to assign it to *Papuadessus* based on molecular phylogenetic inference, especially because the generic classification within the Bidessini using morphology is problematic (cf. Hendrich et al. 2009). We use the condensed descriptive format introduced by Riedel et al. (2013), integrating DNA sequencing, digital imaging and wiki publication of data.

Material and methods

The specimens included in this study are deposited in the following collections:

ANIC	Australian National Insect Collection, Canberra, Australia
CLH	Collection Lars Hendrich, Berlin, Germany; property of the NHMW
MZB	LIPI Division of Zoology, Museum Zoologicum Bogoriense, Cibinong, Indonesia
NMPC	Národní Museum, Prague, Czech Republic
NHMW	Naturhistorisches Museum Wien, Vienna, Austria
SAMA	South Australian Museum, Adelaide, South Australia, Australia
ZSM	Zoologische Staatssammlung München, Munich, Germany

Morphological observations. Photographs were taken with a Leica Photar 1:2/25 on bellows attached to a Nikon D3X camera, an image stack was produced with a custom built robotic macro-rail and combined with Helicon Focus software (www.heliconsoft.com). The principal setup is illustrated on our wiki: http://zsmentomology.de/wiki/Dig-

ital_imaging_in_the_beetle_lab. The male genitalia were studied in dry condition with a Leica M205C dissecting scope at 160×. Pencil sketches were produced with a drawing tube, scanned, and digitally inked using CorelDRAW 11.

DNA extraction and amplification. DNA extractions were carried out on fresh material kept in 96 % ethanol using the DNeasy Tissue Kit (Qiagen GmbH, Hilden, Germany). We sequenced fragments of the cytochrome oxidase subunit 1 (CO1, 611 bp) and ribosomal 16S (823 bp) using standard protocols (http://zsm-entomology.de/wiki/The_Beetle_D_N_A_Lab). The DNA strands obtained after sequencing were eye-corrected and aligned under Geneious R6 (available from http://www.geneious.com).

Phylogenetic analyses. The phylogenetic relationships were inferred in a Bayesian framework using MrBayes 3.2 (Ronquist et al. 2012). We used three partitioning schemes, namely P_1 with only one partition for both genes, P_2 with one partition for each gene, and P_3 including a partition for each coding position of the CO1 and one partition for the 16S. The substitution model for each partition was selected under jModelTest 0.1.1 (Posada 2008). The analyses consisted of two independent runs of 8 Markov Chain Monte Carlo running 20 million generations and sampling every 1000 cycles. The convergence of the runs was assessed under Tracer 1.5 (available at: http://BEAST.bio.ed.ac.uk/ Tracer) by checking the log-likelihood curves and the Effective Sample Size values for each parameter of the analyses. We applied a conservative burnin consisting of 25 % of the topologies sampled, and used the remaining ones to generate a 50 % majority-rule consensus tree based on the best partitioning scheme selected under Tracer 1.5 on the basis of Bayes Factors (B_F) calculated using 1000 Bootstrap replicates.

Results

Phylogeny

The final matrix comprised 1434 bps with no stop codons, and the GTR+Γ+I model was selected as the best-fitting for all partitions under jModelTest 0.1.1. All the Bayesian analyses carried out based on the molecular dataset converged well, and the partitioning scheme P_3 was selected under the B_F criterion. The topology resulting from this partitioning scheme is presented in Figure 7. Overall the phylogenetic tree is well to strongly supported, and two main clades were recovered. The first clade is strongly supported (PP = 1.0) and includes the following genera: *Allodessus* Guignot, 1953, *Gibbidessus* Watts, 1978, *Kakadudessus* Hendrich & Balke, 2009, *Neobidessodes* Hendrich & Balke, 2009, and *Uvarus* Guignot, 1939. The second clade contains the genus *Clypeodytes* Régimbart, 1894 along with the genera *Papuadessus* and *Hydroglyphus* Motschulsky, 1853 with a moderate support (PP = 0.7). Within this group, *P. pakdjoko* is found sister to the specimen from Biak Island with strong support (PP = 1.0). These two specimens are recovered in a sister position to *Hydroglyphus* with strong support (PP = 1.0).

Taxonomy

Genus *Papuadessus* Balke, 2001

Papuadessus Balke, 2001: 108; Balke & Ribera 2004: 125.

Type species. *Papuadessus pakdjoko* Balke, 2001, by original designation.

Online resource. SpeciesID page: http://species-id.net/wiki/Papuadessus

Papuadessus baueri spec. nov.

Figs 2–9

Type locality. Indonesia, Papua, Biak Island, road to Korim, 00°55.736'S 136°02.766'E

Type material. Holotype: ♂, Indonesia: "Indonesia/ Biak 7 BIA 1 Lake betw. Biak & Korem, 80 m, 13.7.1991 leg. Balke & Hendrich", "Holotype Papuadessus baueri sp. nov. Balke, Warikar, Toussaint & Hendrich des. 2013" [red printed label] (MZB). – Paratypes: 33 exs, same data as the holotype (ANIC, CLH, NHMW, NMPC, SAMA, ZSM); 12 exs, "Papua, Biak, road to Korim, sinkhole, 100 m, 24.x.2011, 00°55.736'S 136°02.766'E, Warikar, Surbakti & Balke leg. (PAP19) (MZB, ZSM)", 2 exs with DNA extraction numbers MB4486 and MB4487 (vouchers as well as DNA aliquots in ZSM). All paratypes are provided with red printed paratype labels.

Online resources. Genbank accession numbers: HG 327112 and HG327113. SpeciesID page: http://species-id.net/wiki/Papuadessus_baueri, where dorsal punctation and microreticulation can be examined based on high resolution images.

Etymology. To Jakob Bauer, volunteer in the Coleoptera section who visits us every week to mount specimens with a "thank you – your help is greatly appreciated"!

Description of the holotype

Sculpture and structure. Beetle with continuous body outline in dorsal view, narrowly oval and body rather narrowed towards apex. Head and pronotum with fine microreticulation, regularly and fine punctate, punctures smaller and weaker anteriorly and basally. Elytra with coarse microreticulation, regularly and coarsely punctured. Punctures on elytra not forming rows. Head without cervical line; pronotum and

Figs 1- 3. Habitus of *Papuadessus* species. **1**. *P. pakdjoko* Balke (length 3.2 mm). **2.** *P. baueri* spec. nov. (length 2.2 mm), with darker elytra. **3.** *P. baueri* spec. nov., with more extensive pale elytral bands.

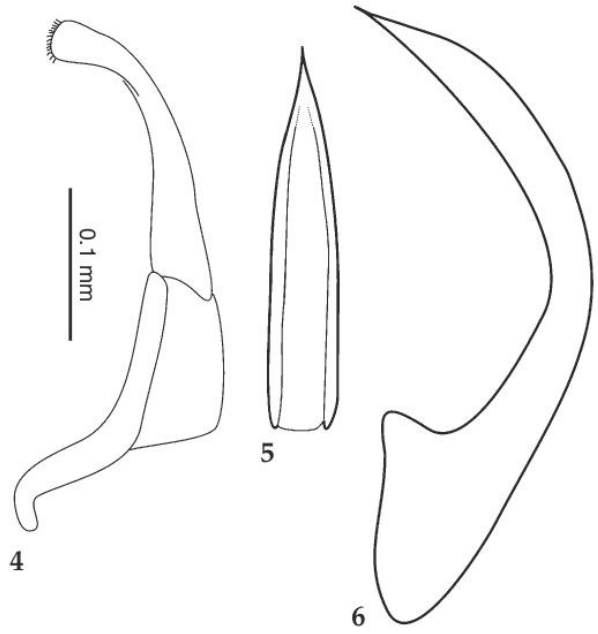
elytron with distinct basal plica; elytron without distinct sutural line (as in *Hydroglyphus* species) but apically with rather faint sutural impression. Epipleuron without basal pit or carina. Ventral side of elytron laterally with two distinct lamellae, as in *P. pakdjoko*, one caudad and one in a median position. Meta- and mesoventrites coarsely and densely punctured, punctures on all abdominal ventrites weaker and smaller.

Colour. Head dark brown, with yellow vertex and a lighter spot on each side of head above clypeus; pronotum yellow, with dark brown median patch at base of pronotum between the pronotal plicae; elytron dark brown with paler longitudinal markings discally, laterally and one distally (Fig. 3), ventrally yellow, head appendages and legs yellow.

Male. Pro- and mesotarsi not expanded. Median lobe of aedeagus in lateral view gently curved, produced into a very fine, acute tip, in ventral view also with pointed tip; parameres two-segmented (Figs 4–6).

Variability. The extend of the pale elytral markings varies, the surface can be mostly dark or the paler areas are more extended (Fig. 3). In the latter case, the general configuration of pale bands then agrees well with that of *P. pakdjoko* (Fig. 1).

Female. No sexual dimorphism observed.



Figs 4- 6. Aedeagus of *Papuadessus baueri* spec. nov. **4.** Paramere (lateral lobe) inner view. **5.** Median lobe in ventral view. **6.** Median lobe in lateral view.

Measurements. Total length of beetle 2.0–2.2 mm (holotype 2.1 mm), total length of beetle without head 1.8–2.0 mm (holotype 1.9 mm); width of beetle 0.9–1.0 mm (holotype 0.9 mm).

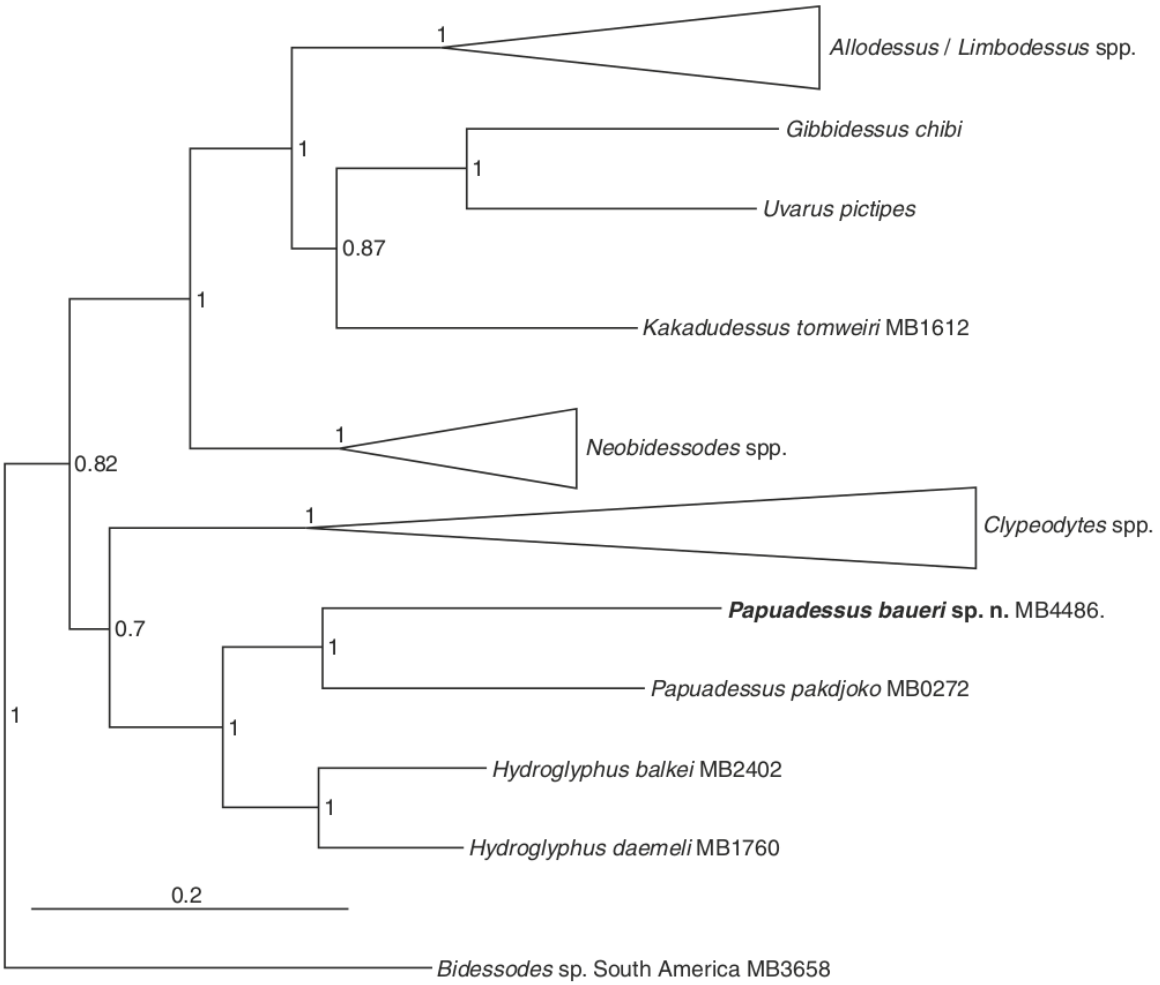


Fig. 7. Bayesian inference tree including Australasian Bidessini, node support are Bayesian posterior probabilities. Where applicable, our DNA extraction and voucher numbers stated after taxon name. Other sequences are from genbank, and for details on multiple species sequenced in the collapsed genera see Hendrich & Balke 2009.

Placement. Assigned to the genus *Papuadessus* based on the analysis of mitochondrial DNA sequence data (Fig. 7).

Collection circumstances. The beetles were collected from shallow water, where they were swimming around coarse limestone gravel, at the edge of a limestone sinkhole (doline), ca. 100 meters in diameter (Fig. 8). The species was collected in association with the following Dytiscidae: *Cybister sugillatus* Erichson, 1834 and *Laccophilus heidiae* Brancucci, 1983 (Balke et al. 1997).

Distribution. Only known from the type locality (Fig. 9).

Papuadessus pakdjoko Balke
Figs 1, 7, 9

Papuadessus pakdjoko Balke, 2001: 108; Balke & Ribera 2004: 125.

Online resources. Genbank accession numbers: 16S rRNA-AY368225; 3' cox1-AY368229. SpeciesID page: http://species-id.net/wiki/Papuadessus_pakdjoko

Notes. This species (Fig. 1) inhabits gravel banks of large lowland rivers and was also collected from smaller streams. It was described from West Papua, south of Nabire and later reported from Simbu Province (Crater Mountain) in Papua New Guinea (Balke & Ribera 2004). Here, we report additional localities:

4 exs, Papua New Guinea: Sandaun, Mianmin, 670 m, 20.x.2008, 4°53.292'S 141°34.118'E, Ibalim (PNG 191); 2 exs, Papua New Guinea: Sandaun, Mianmin area, >1000 m, 23.xii.2009, near 4°54.540'S 141°36.953'E, Ibalim & Pius (PNG232); 3 exs, Papua New Guinea: Sandaun, Mianmin area, >1000 m, 26.xii.2009, near 4°54.540'S 141°36.953'E, Ibalim & Pius (PNG233); 2 exs,

Fig. 8. Habitat of *Papuadessus baueri* spec. nov.

Papua New Guinea: Sandaun, Mianmin area, >600 m, 13.i.2010, 4°54.540'S 141°36.953'E, Ibalim & Pius (PNG 236) (all in MZB, ZSM).

Acknowledgements

We thank the staff of the Biology Department of the Cendrawasih University for help with the fieldwork, conducted during a training course on aquatic Coleoptera. Thanks are also due to Sentiko Ibalim and Pius (Binatang Research Center, Madang, PNG) for helping with the PNG fieldwork.

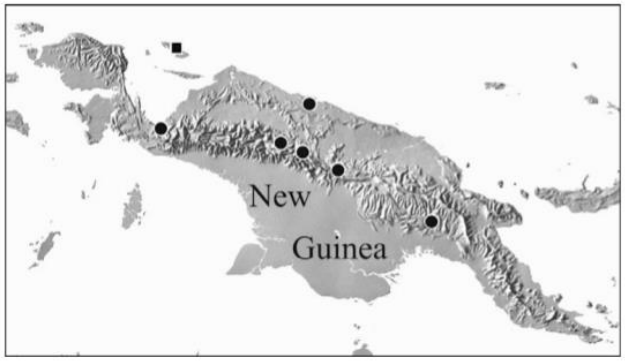


Fig. 9. Distribution of *Papuadessus* spp. (●: *P. pakdjoko*, ■: *P. baueri* spec. nov.).

References

Balke, M. 2001. *Papuadessus pakdjoko* – a new genus and species of rheobiont diving beetle from New Guinea of potential use for environmental impact assessments (Coleoptera, Dytiscidae). Hydrobiologia 464: 107–112.

– – , Larson, D. J. & Hendrich, L. 1997. A review of the New Guinea species of *Laccophilus* Leach, 1815 (Coleoptera: Dytiscidae) with notes on regional melanism. Tropical Zoology 10: 295–320.

– – & Ribera, I. 2004. Jumping across Wallace's line: *Allodessus* and *Limbodessus* revisited (Coleoptera: Dytiscidae, Bidessini) based on molecular phylogenetic and morphological data. Australian Journal of Entomology 43: 114–128.

Hendrich, L. & Balke, M. 2009. *Kakadudessus tomweiri*, a new genus and species of diving beetle from tropical northern Australia, based on molecular, phylogenetic and morphological data (Coleoptera, Dytiscidae, Bidessini). Zootaxa 2134: 49–59.

– – , Hawlitschek, O. & Balke, M. 2009. The epigean Australasian species of *Neobidessodes* gen. n. diving beetles – a revision integrating morphology, cybertaxonomy, DNA taxonomy and phylogeny (Coleoptera: Dytiscidae, Bidessini). Zootaxa 2288: 1–41.

Posada, D. 2008. jModelTest: Phylogenetic Model Averaging. Molecular Biology and Evolution 25(7): 1253–1256.

Ribera, I., Vogler, A. P. & Balke, M. 2008. Molecular phylogeny and diversification of diving beetles (Coleoptera: Dytiscidae). Cladistics 24(4): 563–590.

Riedel, A., Sagata, K., Suhardjono, Y. R., Tänzler, R. & Balke, M. 2013. Integrative taxonomy on the fast track – towards more sustainability in biodiversity research. Frontiers in Zoology 10: 15 (doi:10.1186/1742-9994-10-15).

Ronquist, F., Teslenko, M., van der Mark, P., Ayres D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A. & Huelsenbeck, J. P. 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. Systematic Biology 61(3): 539–542.

6.4 A new genus of diving beetle from Australia

Citation: Hendrich L, Toussaint EFA, Balke M (2014) A new genus of Hydroporini from south-western Australia. *Spixiana*, 37(1):103-109.

.

SPIXIANA	37	1	103–109	München, August 2014	ISSN 0341–8391

A new genus of Hydroporini from south-western Australia

(Coleoptera, Dytiscidae)

Lars Hendrich, Emmanuel F. A. Toussaint & Michael Balke

Hendrich, L., Toussaint, E. F. A. & Balke, M. 2014. A new genus of Hydroporini from south-western Australia (Coleoptera, Dytiscidae). Spixiana 37 (1): 103–109.

A molecular phylogenetic analysis of the four genera *Antiporus* Sharp, 1882, *Chostonectes* Sharp, 1882, *Megaporus* Brinck, 1943 and *Tiporus* Watts, 1985 of Australian Hydroporini shows that *Antiporus gottwaldi* Hendrich, 2001 forms a clade distant from the rest of the species of that genus. The Australian *Antiporus pennifoldae* Watts & Pinder, 2000 has not been studied genetically, however, based on several morphological characters it must also be included in the new genus *Brancuporus* Hendrich, Toussaint & Balke gen. nov. *Brancuporus gottwaldi* (Hendrich, 2001) comb. nov. and *Brancuporus pennifoldae* (Watts & Pinder, 2000) comb. nov. can be separated from *Antiporus* species by having 1) a distinctly asymmetric central lobe of the aedeagus, and 2) in having flanged elytra, at least in females of both species. They are restricted to the peatlands and seasonal swamps of the south-western corner of Southwest Australia.

Lars Hendrich (corresponding author), Emmanuel F. A. Toussaint & Michael Balke, Zoologische Staatssammlung München, Münchhausenstraße 21, 81247 München, Germany; e-mail: hendrich@zsm.mwn.de

Introduction

Australia houses a rich and diverse Hydroporini fauna with lots of endemic genera. The 147 known species (Nilsson 2001, 2013), inhabiting all kind of lentic and lotic aquatic habitats around Australia (e. g. Watts 1997, 2002; Hendrich 2003, 2008; Hendrich & Fery 2008; Hendrich & Watts 2009; Hawlitschek et al. 2011, 2012). Despite the fact that most Australian Hydroporini genera have been revised in recent years, and lots of species have been discovered, the situation on the generic level is quite stable (Nilsson 2013).

Surprisingly a comprehensive molecular phylogenetic analysis of an almost complete set of Australian Hydroporini (Toussaint et al. in press) has shown that a single species in the genus *Antiporus* Sharp, 1882 forms a clade distant from the rest of the genus and is thus assigned to a new generic name.

Material and methods

Taxon sampling and phylogenetic inference

We compiled the most complete molecular dataset of Australian dytiscid species to date (Hendrich et al. 2010) for the following genera: *Antiporus* Sharp, 1882, *Chostonectes* Sharp, 1882, *Megaporus* Brinck, 1943 and *Tiporus* Watts, 1985 to test the monophyly of each genus and investigate the relationships among *Antiporus* especially regarding *A. gottwaldi* and the morphological close *A. pennifoldae* (the specimens used in this study are listed in Table 1). In order to root the trees, the species *Carabhydrus niger* Watts, 1978 was selected. DNA was extracted from leg or thoracic tissues of freshly collected beetles stored in 96 % ethanol using the DNeasy kit from Qiagen (Hilden, Germany). We used standard protocols (http://zsm-entomology.de/wiki/The_Beetle_D_N_A_Lab) to amplify and sequence the Cytochrome b (CytB), Cytochrome oxidase subunit 1 (CO1), Histone 3 (H3) and Histone 4 (H4) (Table 2). Once both directions were sequenced, the sequences were eye-corrected and aligned using Geneious R6 (Biomatters, available from

http://www.geneious.com), and the reading frame of each gene was checked under Mesquite 2.75 (available from http://mesquiteproject.org). Under the same software, we concatenated the four genes to produce a combined dataset.

The phylogenetic inferences were completed using three different methods: Maximum Parsimony (MP), Maximum Likelihood (ML) and Bayesian Inference (BI). MP analyses were carried out under TNT 1.1 (Goloboff et al. 2008) with the Tree Ratchet, Tree Fusing and Tree Drifting algorithms (Goloboff 1999). 1000 Jackknife replicates (JK) were used to evaluate the robustness of the phylogenetic inference. For both ML and BI analyses, the best model of sequence evolution was selected using jModelTest 0.1.1 (Posada 2008). The ML analyses were realized on the combined dataset under RAxML (Stamatakis 2006) and we performed 1000 thorough Bootstrap replicates (BS) to investigate the level of support

Table 1. List of species of the subfamily Hydroporinae, tribe Hydroporini, used in this study. NSW = New South Wales, Gb = already in Genbank, number = newly submitted data to Genbank.

Species	Locality	co1	cytb	h3	h4
Antiporus bakewellii	Victoria, NSW	Gb	Gb	Gb	HG965722
Antiporus blakeii	NSW, South Australia, Tasmania	Gb	Gb	Gb	HG965737
Antiporus femoralis	NSW, South Australia, Tasmania	Gb	Gb	Gb	HG965742
Antiporus gilbertii	Western Australia	Gb	Gb	Gb	HG965747
Brancuporus gottwaldi	Western Australia	Gb	Gb	Gb	HG965732
Antiporus hollingsworthi	Western Australia	Gb	Gb	HG965673	HG965708
Antiporus interrogationis	New South Wales	Gb	Gb	Gb	HG965716
Antiporus jenniferae	Northern Territory	Gb	Gb	Gb	HG965712
Antiporus occidentalis	Western Australia	Gb	Gb	–	HG965707
Antiporus uncifer	New Zealand	HG965640	–	–	–
Antiporus wilsoni	Queensland	Gb	Gb	Gb	HG965724
Carabhydrus niger	Victoria	Gb	HG965672	HG965706	HG965750
Chostonectes johnsonii	New South Wales	Gb	HG965663	HG965696	HG965739
Chostonectes nebulosus	South Australia	Gb	HG965666	HG965699	HG965743
Chostonectes sharpi	Queensland	Gb	HG965664	HG965697	HG965740
Megaporus gardnerii	South Australia	Gb	HG965670	HG965703	HG965748
Megaporus hamatus	NSW, Victoria, South Australia	Gb	HG965653	HG965685	HG965725
Megaporus howittii	NSW, Victoria, South Australia	Gb	HG965649	HG965682	HG965718
Megaporus natvigi	Queensland	Gb	HG965661	HG965694	HG965736
Megaporus solidus	Western Australia	HG965630	HG965641	HG965674	HG965709
Megaporus wilsoni	South Australia	Gb	HG965648	HG965681	HG965717
Tiporus centralis	Northern Territory	Gb	HG965647	HG965680	HG965715
Tiporus collaris	Northern Territory	Gb	HG965657	HG965689	HG965729
Tiporus josepheni	Northern Territory	Gb	HG965665	HG965698	HG965741
Tiporus lachlani	Western Australia	Gb	HG965652	HG965684	HG965723
Tiporus tambreyi	Western Australia	Gb	HG965669	HG965702	HG965746
Tiporus undecimmaculatus	Northern Territory	Gb	HG965660	HG965693	HG965735

Table 2. Primers used to amplify regions of the cytochrome oxidase subunit 1 (CO1) and cytochrome B (CytB).

Locus	Primer	Sequence	Reference
CytB	CB3	GAGGAGCAACTGTAATTACTAA	Barraclough et al. 1999
	CB4	AAAAGAAA(AG)TATCATTCAGGTTGAAT	
CO1	Pat	CAACATTTATTTTGATTTTTTGG	Simon et al. 1994
	Jerry	TCCAATGCACTAATCTGCCATATTA	
H3	H3aF	ATGGCTCGTACCAAGCAGAC(AG)CGC	Colgan et al. 1998
	H3aR	ATATCCTT(AG)GGCAT(AG)AT(AG)GTGAC	
H4	H4F2s	TSCGIGAYAACATYCAGGGIATCAC	Pineau et al. 2005
	H42er	CKYTTIAGIGCRTAIACCACRTCCAT	

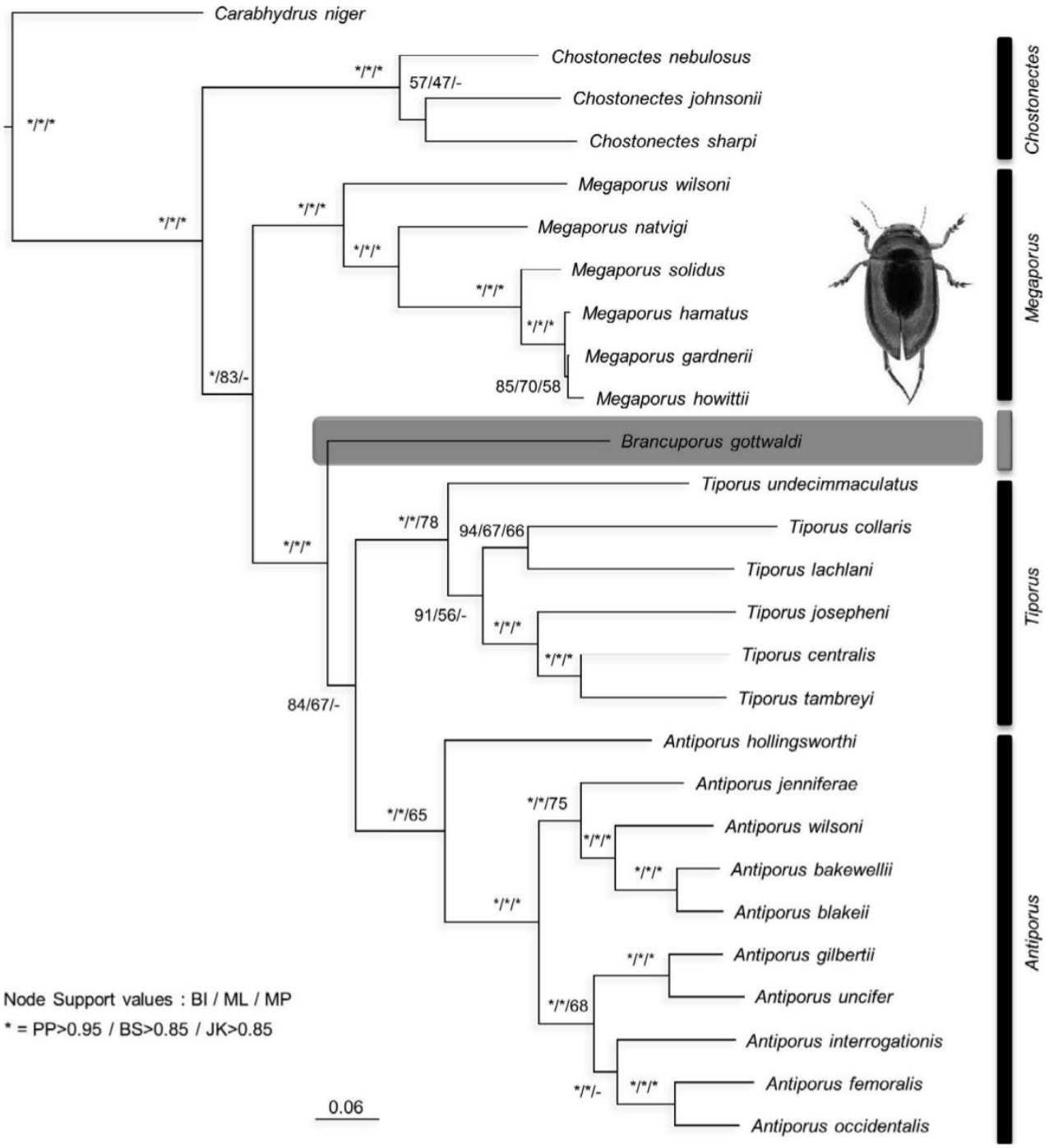
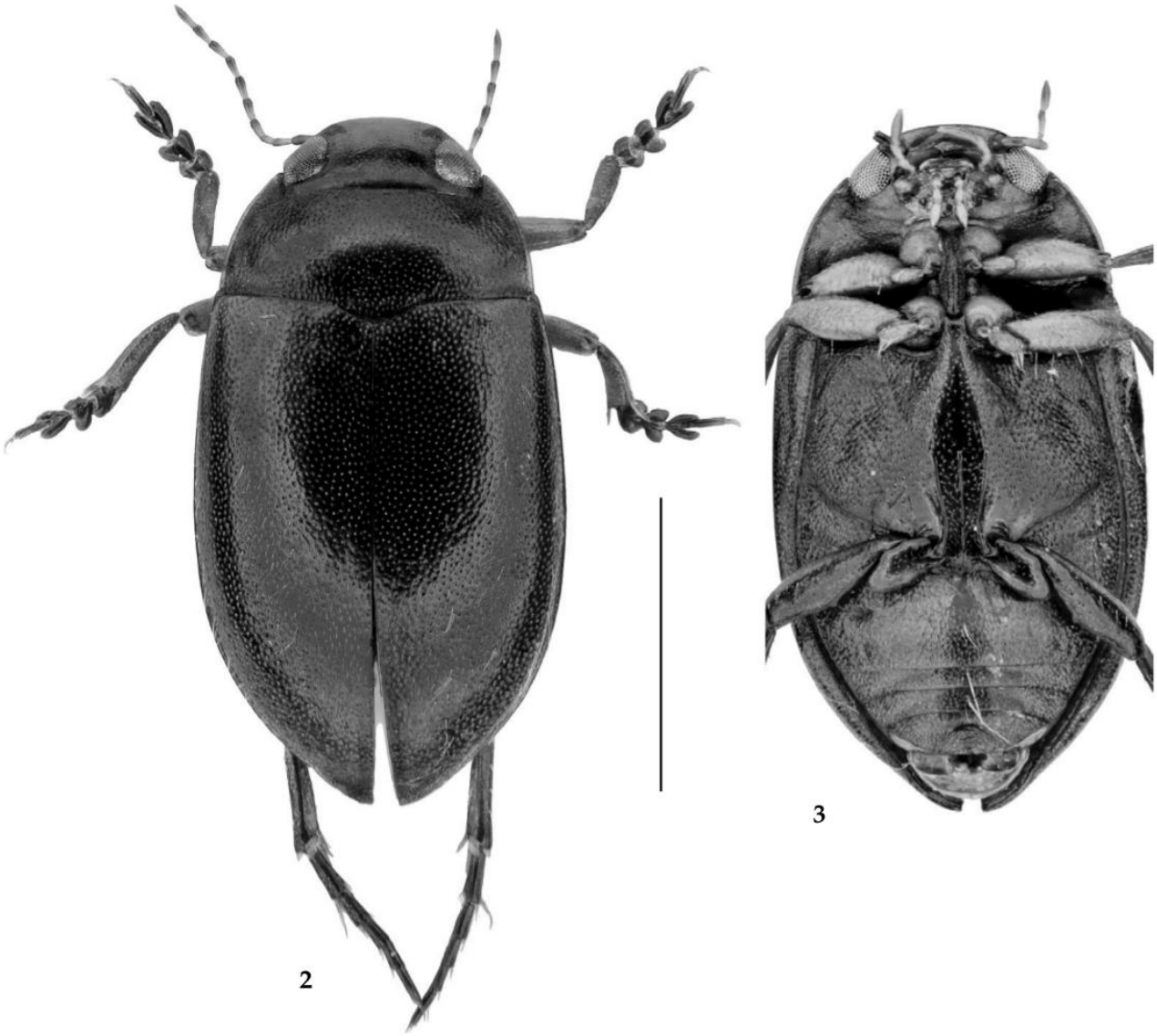


Fig. 1. MrBayes 50 % majority-rule consensus tree based on the molecular dataset. The support of each node recovered in the different analyses (BI, ML and MP) is indicated on the topology following the caption. A picture of the habitus of *B. gottwaldi* comb. nov. is provided.

at each node. Eventually, we also carried out BI analyses on the combined dataset under MrBayes 3.2 (Ronquist et al. 2012) with the following settings: the model of substitution set accordingly to the result of jModelTest, two runs of four Markov Chains Monte Carlo (MCMC, one cold and three incrementally heated) running for 8 million generations and sampling a topology every 1000 cycles. After checking the convergence of the runs under Tracer 1.5 (available at: http://beast.bio.ed.ac.uk/Tracer) and applying a conservative burn-in of 25 %, we used the command "sump" in MrBayes to calculate the posterior probabilities (PP) and produce a 50 % majority rule consensus tree. A PP≥0.95 and a BS or JK≥85 were recognized as indicating a strong support for a given node (Felsenstein 2004).

New sequences were submitted to Genbank, see Table 1.



Figs 2–3. Habitus of *Brancuporus gottwaldi* comb. nov. (scale bar = 1.5 mm). **2.** Dorsal side; **3.** ventral side.

Systematics

Brancuporus Hendrich, Toussaint & Balke gen. nov.

Figs 2–12

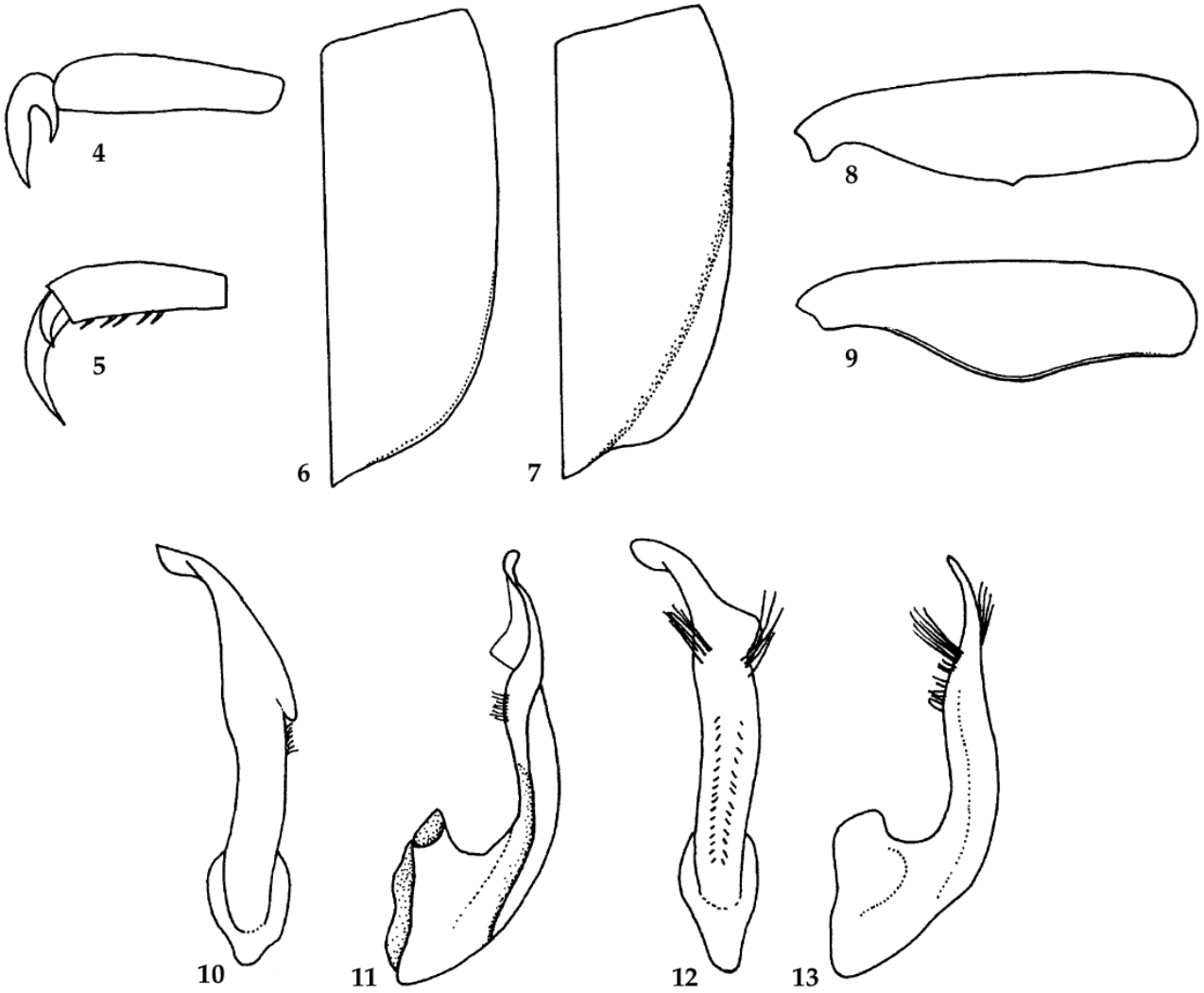
Type species. *Antiporus gottwaldi* Hendrich, 2001 by present designation.

Material studied. The material used for this study is listed in Watts & Pinder (2000) and Hendrich (2001).

Online resources. Registered in ZooBank under urn:lsid:zoobank.org:pub:91A7660B-51BC-4E24-878D-56DC61889A7F. Species page in wiki format under http://species-id.net/wiki/Brancuporus.

Diagnosis. *Brancuporus* Hendrich, Toussaint & Balke gen. nov. is assigned to the Hydroporini based on parameres of the aedeagus formed by one segment, and hind coxal process not in the same plane as the abdomen, but protruding like a step, in lateral view (Pederzani 1995). It is a genus of small sized elongate-oval, reddish-brown Hydroporini (3.00–3.35 mm), represented by two species restricted to south-western Australia.

The new genus is well separated from all other Hydroporini by the combination of the following characters: 1) Body elongate oval, reddish-brown, 2) Fourth tarsomere of protarsus scarcely visible; 3) Pronotum and elytron with narrow but well-marked lateral beading; 4) Elytra at least in female flanged; 5) Posterior part of epipleuron comparatively broad; 6) Humeral angle of elytron smoothly rounded; 7) Prosternal process blunt, sides weakly bowed, moderately ridged; 8) Distinctly asymmetric central lobe of aedeagus.



Figs 4. Lateral view of proclaw and apical tarsomere of *Brancuporus gottwaldi*; **5.** ditto *B. pennifoldae*; **6.** dorsal view of elytron of *Brancuporus gottwaldi*, female; **7.** ditto *B. pennifoldae*; **8.** ventral view of metafemur of *B. pennifoldae*; **9.** ditto *B. gottwaldi*; **10.** ventral view of median lobe of aedeagus of *B. gottwaldi*; **11.** ditto of lateral view; **12.** ventral view of median lobe of aedeagus of *B. pennifoldae*; **13.** ditto of lateral view (adapted from Hendrich 2001).

Brancuporus gen. nov. can be separated from *Antiporus* by having 1) a distinctly asymmetric central lobe of the aedeagus, and 2) in having flanged elytra, at least in females of both species.

Description

Measurements (N = 15). Total length of beetles 3.00–3.35 mm; length without head 2.75–2.95 mm; greatest width of beetles 1.70–1.85 mm.

Colour. Upper side comparably light; head reddish; pronotum ferrugineous anteriorly and laterally, dark posteriorly and medially; elytron dark brown, paler laterally (Fig. 2). Venter yellowish to brownish; pronotum, epipleuron, legs and abdominal ventrites yellowish to ferrugineous; metaventrite, metacoxal plates and processes brownish. Antennomeres yellowish and darkened anteriorly. Sculpture: dorsal surface, punctures dense, moderately sized; those on head weaker and sparser, a little smaller than eye facet. Pronotum and elytron with narrow but well marked lateral beading. Microreticulation on head and pronotum fine, moderately impressed, on elytron very fine and almost invisible. Ventral surface (Fig. 3) with punctures very dense, microreticulation similar to that on elytron. Prosternal process blunt, sides weakly bowed, moderately ridged. Metacoxal lines parallel in apical quarter, weakly diverging posteriorly, intralinear space flat, not depressed.

Male. Protarsi moderately expanded, single proclaw relatively stout, bent at right angles evenly curved with ventral basal spine (Fig. 4). Mesotibia normal, mesotarsi similar to protarsi except that the second and third tarsomere are a little shorter and two claws are present. Metafemur a little stouter than in female, with well marked beading in middle

at hind margin (Fig. 8). Apical third of elytron not flanged. Central lobe of aedeagus with asymmetric tip (ventral view, Figs 10–11).

Female. Protarsi weakly expanded, two claws. Mesotarsi moderately expanded, more so than protarsi. Metatibia simple. Elytron weakly flanged (Fig. 6).

Etymology. The name *Brancuporus* gen. nov. is derived from the name of our highly valued colleague, the late dytiscid specialist Dr Michel Brancucci (1950–2012), Basel, Switzerland. Its gender is masculine.

Systematic notes. The small size, relatively uniform reddish-brown colour and essentially simple metafemora (Hendrich 2001) suggest that *Brancuporus gottwaldi* comb. nov. is close to *Antiporus pennifoldae* (Watts & Pinder 2000), also described from Southwest Australia. Females of both species have flanged elytra [strongly flanged in *A. pennifoldae* (Fig. 7) and less flanged in *B. gottwaldi* (Fig. 6)]. Furthermore, the distinctly asymmetric central lobe of the aedeagus is a character shared by *B. gottwaldi* and *Antiporus pennifoldae* (Figs 12–13). Based on these morphological characters and despite the fact that there was no DNA of *A. pennifoldae* available for this study, *A. pennifoldae* is here transferred to *Brancuporus* gen. nov.

Distribution. The new genus is highly endemic to the most south-western part of Southwest Australia. Both species were collected in seasonal peatland swamps. The habitat and its water beetle coenosis are described in detail by Hendrich (2001).

Molecular systematics. Our phylogenetic analyses show (Fig. 1) that *Brancuporus* gen. nov. is not part of the *Antiporus* clade. It is rather part of a separate lineage sister to the *Tiporus* and *Antiporus* clades. The molecular data also show clearly that the species *Brancuporus gottwaldi* comb. nov. does not belong to any of the other known Australasian Hydroporini genera. Additionally, our data show that Australian *Brancuporus* gen. nov. does not create paraphyly among other Australasian genera (Toussaint et al. in press). This result is well supported in our analyses (Fig. 1). Essentially the same tree topology was recovered with different analytical approaches (maximum likelihood, parsimony and Bayesian probabilities as implemented in MrBayes, Fig. 1, node support values).

Acknowledgements

Dr Chris Watts is thanked for his steady support, generous assistance and hosting during the several visits of LH in Adelaide, South Australia. This work was supported by German Science Foundation (DFG) grants BA2152/6-1, 7-1 and 11-1. The Department of Conservation and Land Management (CALM) is acknowledged for giving permission to conduct scientific research in Nature Reserves in south-western Australia (permit numbers: SF 003017 and NE 002348). Katja Neven (ZSM, Munich, Germany) is thanked for providing the habitus photos.

References

Barraclough, T. G., Hogan, J. E. & Vogler, A. P. 1999. Testing whether ecological factors promote cladogenesis in a group of tiger beetles (Coleoptera: Cicindelidae). Proceedings of the Royal Society of London B 266: 1061–1067.

Colgan, D. J., McLauchlan, A., Wilson, G. D. F., Livingston, S. P., Edgecombe, G. D., Macaranas, J., Cassis, G. & Gray, M. R. 1998. Histone H3 and U2 snRNA DNA sequences and arthropod molecular evolution. Australian Journal of Zoology 46: 419–437.

Felsenstein, J. 2004. Inferring phylogenies. Sunderland, Massachusetts (Sinauer Associates).

Goloboff, P. A. 1999. Analyzing large data sets in reasonable times: solutions for composite optima. Cladistics 15: 415–428.

–– , Farris, J. S. & Nixon, K. C. 2008. TNT, a free program for phylogenetic analysis. Cladistics 24: 774–786.

Hawlitschek, O., Hendrich, L., Espeland, M., Toussaint, E. F. A., Genner, M. J. & Balke, M. 2012. Pleistocene climate change promoted rapid diversification of aquatic invertebrates in Southeast Australia. BMC Evolutionary Biology 12(1): 142.

–– , Porch, N., Hendrich, L. & Balke, M. 2011. Ecological niche modelling and nDNA sequencing support a new, morphologically cryptic beetle species unveiled by DNA barcoding. PloS ONE 6(2): e16662.

Hendrich, L. 2001. A new species of *Antiporus* Sharp, 1882 from peatland swamps of south-western Australia (Coleoptera: Dytiscidae). Linzer Biologische Beiträge 33(1): 299–308.

–– 2003. A new species of *Necterosoma* Sharp from the Kimberley region in north-western Australia (Coleoptera: Dytiscidae). Entomologische Zeitschrift 113: 152–154 + figs on pp. 144–145.

–– 2008. *Tiporus emmae* sp. n. from Northern Australia (Coleoptera: Dytiscidae, Hydroporinae). Linzer Biologische Beiträge 40(1): 311–320.

–– & Fery, H. 2008. *Paroster baylyi* sp. n., *P. ursulae* sp. n. (Col. Dytiscidae, Hydroporinae) and the water beetle diversity of pan-gnammas on isolated granite outcrops in the Mallee of south-western Australia. Zootaxa 1704: 27–41.

–– & Watts, C. H. S. 2009. Taxonomic revision of the Australian predaceous water beetle genus *Carabhydrus* Watts, 1978 (Coleoptera: Dytiscidae, Hydroporinae). Zootaxa 2048: 1–30.

–– , Pons, J., Ribera, I. & Balke, M. 2010. Mitochondrial cox1 sequence data reliably uncover patterns

of insect diversity but suffer from high lineage-idiosyncratic error rates. PloS ONE 5(12): e14448. doi:10.1371/journal.pone.00114448

Nilsson, A. N. 2001. Dytiscidae (Coleoptera). World Catalogue of Insects 3: 1–395.

– – 2013. A world catalogue of the family Dytiscidae or the diving beetles (Coleoptera, Adephaga). Version 1.I.2013. Distributed as a PDF file via Internet. Available from: http://www2.emg.umu.se/projects/biginst/andersn/

Pederzani, F. 1995. Keys to the identification of the genera and subgenera of adult Dytiscidae (sensu lato) of the world (Coleoptera Dytiscidae). Atti dell'Accademia Roveretana degli Agiati, Contributi della Classe di Scienze Matematiche, Fisiche e Naturali, a. 244 (1994), ser. VII, 4(B): 5–83.

Pineau, P., Henry, M., Suspène, R., Marchio, A., Dettai, A., Debruyne, R., Petit, T., Lécu, A., Moisson, P., Dejean, A., Wain-Hobson, S. & Vartanian, J.-P. 2005. A universal primer set for PCR amplification of nuclear histone H4 genes from all animal species. Molecular Biology and Evolution 22: 582–588.

Posada, D. 2008. jModelTest: Phylogenetic model averaging. Molecular Biology and Evolution 25: 1253–1256.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A. & Huelsenbeck, J. P. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61(3): 539–542.

Simon, C., Frati, F., Beckenbach, A. T., Crespi, B., Liu, H. & Fllok, P. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Annals of the American Entomological Society 87: 651–701.

Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688–2690.

Toussaint, E. F. A., Condamine, F. L., Hawlitschek, O., Watts, C. H. S., Hendrich, L. & Balke, M. in press. Unveiling the diversification dynamics of Australasian predaceous diving beetles in the Cenozoic. Systematic Biology.

Watts, C. H. S. 1997. Four new species of *Antiporus* Sharp (Coleoptera; Dytiscidae) from Australia, with notes on *A. femoralis* (Boheman) and *A. interrogationis* (Clark). Records of the South Australian Museum 30(1): 35–42.

– – 2002. Checklist and guides to the identification, to genus, of adults and larval Australian water beetles of the families Dytiscidae, Noteridae, Hygrobiidae, Haliplidae, Gyrinidae, Hydraenidae and the superfamily Hydrophiloidea (Insecta – Coleoptera). Cooperative Research Centre for Freshwater Ecology (Australia). Identification and Ecology Guide 43: 1–110.

– – & Pinder, A. 2000. Two new species of *Antiporus* from Western Australia (Coleoptera: Dytiscidae). Records of the South Australian Museum 33(1): 17–19.

PART 2: ISLAND LINEAGE DIVERSIFICATION

Chapter 7. Biogeography on a continental-sized island: Australia

View of the Southern part of the Australasian / Indomalayan archipelago from space

"*It is well known that the natural productions of Australia differ from those of Asia more than those of any of the four ancient quarters of the world differ from each other. Australia, in fact, stands alone […]*"

Alfred Wallace, *The Malay Archipelago*, 1869.

Chapter contents

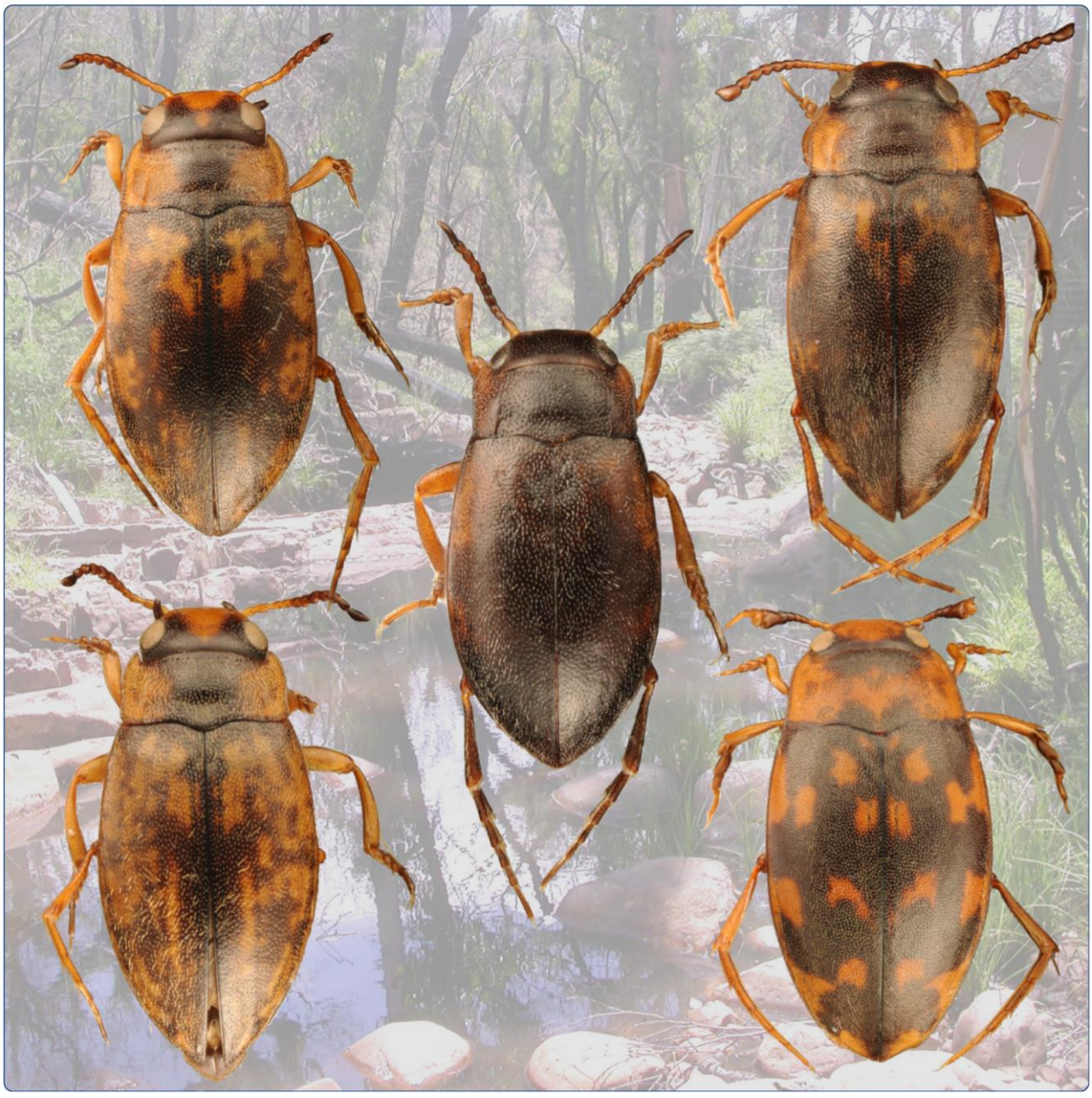
7.1 Paper VI - Diversification of Australian diving beetles in the Quaternary 135

7.2 Paper VII - Diversification of Australasian diving beetles in the Cenozoic 152

7.1 Pleistocene climate change promoted rapid diversification of aquatic invertebrates in South-East Australia

Citation: Hawlitschek O, Hendrich L, Espeland M, Toussaint EFA, Genner MJ, Balke M (2012) Pleistocene climate change promoted rapid diversification of aquatic invertebrates in South-East Australia. *BMC Evolutionary Biology*, 12:142.

BMC
Evolutionary Biology



Pleistocene climate change promoted rapid diversification of aquatic invertebrates in Southeast Australia

Hawlitschek *et al.*

BioMed Central

Hawlitschek *et al. BMC Evolutionary Biology* 2012, **12**:142
http://www.biomedcentral.com/1471-2148/12/142

Hawlitschek et al. BMC Evolutionary Biology 2012, 12:142
http://www.biomedcentral.com/1471-2148/12/142

BMC Evolutionary Biology

RESEARCH ARTICLE Open Access

Pleistocene climate change promoted rapid diversification of aquatic invertebrates in Southeast Australia

Oliver Hawlitschek[1*], Lars Hendrich[1], Marianne Espeland[2], Emmanuel FA Toussaint[1], Martin J Genner[3] and Michael Balke[1,4]

Abstract

Background: The Pleistocene Ice Ages were the most recent geohistorical event of major global impact, but their consequences for most parts of the Southern hemisphere remain poorly known. We investigate a radiation of ten species of *Sternopriscus*, the most species-rich genus of epigean Australian diving beetles. These species are distinct based on genital morphology but cannot be distinguished readily by mtDNA and nDNA because of genotype sharing caused by incomplete lineage sorting. Their genetic similarity suggests a Pleistocene origin.

Results: We use a dataset of 3858 bp of mitochondrial and nuclear DNA to reconstruct a phylogeny of *Sternopriscus* using gene and species trees. Diversification analyses support the finding of a recent rapid speciation event with estimated speciation rates of up to 2.40 species per MY, which is considerably higher than the proposed average rate of 0.16 species per MY for insects. Additionally, we use ecological niche modeling and analyze data on habitat preferences to test for niche divergence between species of the recent *Sternopriscus* radiation. These analyses show that the species can be characterized by a set of ecological variables referring to habitat, climate and altitude.

Conclusions: Our results suggest that the repeated isolation of populations in glacial refugia might have led to divergent ecological adaptations and the fixation of morphological traits supporting reproductive isolation and therefore may have promoted speciation. The recent *Sternopriscus* radiation fulfills many characteristics of a species flock and would be the first described example of an aquatic insect species flock. We argue that the species of this group may represent a stage in speciation past the species flock condition because of their mostly broad and often non-overlapping ranges and preferences for different habitat types.

Background

Global biodiversity is shaped by the processes of speciation and extinction, whose rates vary depending on region, environment, taxonomic group and geohistorical events [1-3]. Evidence for shifts in the rates of speciation and extinction have been inferred from the fossil record since early paleontology [4], and advances in molecular biology have greatly improved our capabilities to study these processes particularly for taxa with sparse or inconsistent fossil evidence [5,6].

The most recent geohistorical event of major global impact on biodiversity was the Pleistocene glaciations, or Ice Ages, which represent the largest expansion of cold climates since the Permian period 250 million years (MY) earlier. Until 10,000 years ago, temperatures repeatedly oscillated between warm and cold phases. The effects on the environment varied depending on geographical region, but were always accompanied by major biotic shifts. Boreal regions, particularly in the Northern hemisphere, were mostly glaciated and drove species into refugia [7]. In the tropics and subtropics, where glaciations were mostly restricted to high altitudes, a similar effect was attributed to the aridification of formerly humid forest habitats [8]. It has been a matter of discussion whether these cycles of environmental change

* Correspondence: oliver.hawlitschek@gmx.de
[1]Zoologische Staatssammlung, Münchhausenstr. 21, Munich 81247, Germany
Full list of author information is available at the end of the article

promoted speciation [9] or whether species responded solely by shifting their ranges toward ecologically suitable areas [10]. In Australia, glaciations occurred only at its highest elevations, but biota faced an ongoing process of aridification that was initiated in the Miocene *c.* 15 million years ago (MYA) when Australia drifted northward [11]. During the Ice Ages, the relatively rapid shifts between warm and wet *versus* cold and dry conditions had severe consequences particularly for the fauna [12,13]. Aquatic environments were strongly affected by oscillations between arid and humid conditions [14].

The genesis of the Australian arid zone promoted radiations in various organism groups, e.g., hypogean faunas in the ground waters underneath the spreading deserts, which most likely began with the onset of the aridification *c.* 15 MYA [14]. However, many rapid radiations of insects dating back only 2 MY or less have been described from all around the world. Coyne & Orr [15] proposed an average speciation rate of 0.16 species per MY, which is exceeded by an order of magnitude by the fastest known radiation [16-18]. Phylogenies of such young radiations based on mitochondrial gene trees are often poorly resolved, and species may appear para- or polyphyletic because of shared alleles with other species, which may be the result of incomplete lineage sorting or hybridization [19]. Species trees may cope with these problems: in a method based on a coalescent model and Bayesian inference, all gene trees are co-estimated and embedded in a single species tree whose tips represent species and not single samples [20,21].

Aside from morphological and molecular characters, ecological factors can be useful to distinguish and even delimit species. Many studies have shown that a variety of climate factors often have a profound effect on the distributions of species, and these factors can be combined to project potential distributions of species in an Ecological Niche Modeling (ENM) approach [22,23]. The predictive powers of this method have been demonstrated [24], and it has been successfully applied in species delimitations [22,25]. Naturally, the distinction of species based on differences in their responses to ecological factors is sensible only if there are actual response differences. Evidence of niche conservatism in closely related species, promoting allopatric speciation, is abundant [26]. However, in many examples of rapid radiations in limited geographic areas niche divergence appears to be the more common condition, and closely related species show different responses to ecological factors [2004, 27].

The focus of our study is on the genus *Sternopriscus* (Coleoptera: Dytiscidae: Hydroporini), which is the most species-rich epigean genus of Australian diving beetles and contains 28 species [27,28]. *Sternopriscus* species inhabit a wide variety of lentic and lotic habitats from sea level to high altitudes. 18 species are found in southeastern Australia, of which four species are endemic to Tasmania. The corresponding freshwater ecoregions according to Abell *et al.* [29] are Eastern Coastal Australia, Bass Strait Drainages, Southern Tasmania, and small parts of the Murray-Darling region. Unlike many other aquatic invertebrates, such as crustaceans and gastropods, most species of epigean aquatic beetles use flight to colonize new habitats. Therefore, the presence of suitable habitats most likely has a higher impact on aquatic beetle distribution than the drainage systems defining the biogeographic regions of Abell *et al.* [29]. Nevertheless, only 2 of these 18 species have a wider distribution over mainland Australia (*S. multimaculatus* and *S. clavatus*). 6 species, including some taxonomically and geographically isolated species, are endemic to peaty habitats in the southwest, in an area with cold and humid climate during winter, and 5 species are distributed over the tropical north, including one endemic species in the deep gorges of the Pilbara. None, or only one, species is shared by 2 or more of these areas of endemism. This distribution reflects the restriction of all but the widespread pioneer species *S. multimaculatus* to the more humid coastal areas of Australia. The high level of endemism in the southeast and southwest suggests that the arid barrier between these two regions is long-standing. Another strong pattern is the virtual absence of *S. tarsalis* group members from the north and southwest regions of the continent, whereas members of the *S. hansardii* group, with highly modified male antennae and median lobes, are more widespread [27,28].

Based on male morphological characters, the genus has been divided into 3 groups: the *S. hansardii* group (11 species), the *S. tarsalis* group (13 species), and 4 'phylogenetically isolated' species. The species in the *S. tarsalis* group have been assigned to 3 species complexes: the *S. tarsalis* complex (2 species), the *S. meadfootii* complex (5), and the *S. tasmanicus* complex (3). 3 species have not been assigned to any complex. The 10 species belonging to the *S. tarsalis*, *S. meadfootii* and *S. tasmanicus* complexes in the *S. tarsalis* group are genetically similar and centered in mesic southeastern Australia. Below, we refer to this group of species as the *S. tarsalis* radiation (STR). The STR is supposedly the result of recent diversification; some of these morphologically well-defined species occur in sympatry, and some in syntopy [27,28,30]. Previous genetic studies [30] suggest that species belonging to the STR are not easily delimited using mtDNA and nDNA.

In this study, we attempt to test the following hypotheses: (1) the delimitation of species in the STR, based on morphological characters, can be supported by genetic or ecological data; (2) the STR species originated in a rapid and recent diversification event, most likely in the

Pleistocene; and (3) the radiation of the STR was promoted by the Pleistocene climate oscillations. We use a molecular phylogeny with gene and species trees and diversification rate analyses to investigate how environmental change has affected speciation and extinction rates in the genus *Sternopriscus*. We then discuss which factors might have promoted lineage diversification in the STR and whether the molecular similarities are caused by hybridization or incomplete lineage sorting. Aside from the results of our molecular phylogeny, we use phylogeographic network analyses and ENM paired with empirical ecological data in an attempt to reveal how this diversification was promoted.

Methods

Sampling and laboratory procedures

Specimens were collected by sweeping aquatic dip nets and metal kitchen strainers in shallow water or operating black-light traps [27] and preserved in 96% ethanol. DNA was extracted non-destructively using Qiagen blood and tissue kits (Qiagen, Hilden). Primers are listed in Additional file 1: Table S1. New sequences were submitted to GenBank under accession numbers [EMBL: HE818935] to [EMBL:HE819178]; *cox1* data are [EMBL: FR732513] to [EMBL:FR733591]. The individual beetles from which we extracted and sequenced DNA each bear a green cardboard label that indicates our DNA extraction number (e.g., "DNA 1780 M. Balke"). This number links the DNA sample, the dry mounted voucher specimen and the GenBank entries.

Phylogenetic analyses

The aligned 3858 bp dataset contains three mitochondrial (16 S rRNA, cytochrome oxidase b (*cob*), and cytochrome *c* oxidase subunit I (*cox1*)) and four nuclear gene fragments (18 S rRNA, arginine kinase (ARK), histone 3 (h3), and histone 4 (h4)) for 54 specimens of 25 *Sternopriscus* species and 2 Hydroporini outgroups, *Barretthydrus stepheni* and *Carabhydrus niger*. Among the known species of *Sternopriscus*, only *S. mouchampsi* and *S. pilbaraensis* were not available for sequencing. *S. emmae* was excluded from the phylogenetic analyses because we only had DNA from museum specimens and only obtained a short *cox1* sequence. DNA alignment was performed in MUSCLE 3.7 [31]. We then used jModelTest 0.1.1 [32] to identify appropriate substitution models for each gene separately, assessing lnL, AIC and BIC results and giving preference to BIC. To evaluate different partition schemes, we performed a Bayes factor test with MrBayes 3.1 [33] and Tracer v1.5 [34]. The eleven schemes tested were mitochondrial *versus* nuclear, protein-coding *versus* ribosomal, and according to codon positions (1 + 2 *versus* 3 or one partition for each codon position). We used raxmlGUI 0.93 [35] for maximum likelihood analyses with 1000 fast bootstrap repeats. MrBayes 3.1 [33] was used for Bayesian analyses, with two runs and four chains with 30,000,000 generations (samplefreq = 1,000 and 25% burnin). Runs were checked for convergence and normal distribution in Tracer v1.5 [34]. We then used parsimony analysis to infer phylogenetic relations as implemented in the program TNT v1.1, which we also used to run 500 jackknife replications (removal 36%) to assess node stability [36] (hit the best tree 5 times, keep 10,000 trees in memory). Finally, we used coalescent-based species tree inference models in *BEAST v1.6.1 [21] for comparison with the results of the phylogenetic gene tree. *BEAST requires *a-priori* designation of species, which we performed based on morphological data [27,28]. We conducted two runs over 100,000,000 generations (sample freq = 1,000 and 20% burnin) and checked for convergence and normal distribution in Tracer v1.5 [34]. Additionally, as proposed in Pepper *et al.* [13], we repeated this analysis using simpler substitution models (HKY + G). All analyses in MUSCLE and MrBayes were run on the CIPRES Portal 2.2 [37]. Pairwise distances were calculated in MEGA 5.0 [38].

Lineage diversification and radiation

Analyses were conducted in R with the packages APE [39] and Laser [40]. Based on the phylogenetic tree created in MrBayes, we used the 'chronopl' function of APE to create an ultrametric tree in R and cropped all representatives but one of each species. We then constructed Lineage-Through-Time (LTT) plots [41] and calculated γ-statistics [42]. Because new species continue to be discovered in Australia and incomplete taxon sampling might influence γ-statistics, we conducted a Monte Carlo constant rates (mccr) test with 10,000 replicates, assuming 10% missing species. We then tested the fit of two rate-constant [41] and four rate-variable diversification models [43] to our dataset. Finally, we calculated *p*-values by simulating 10,000 trees with original numbers of present and missing species for a pure-birth scenario and for various birth-death rates (b = 0.5 and d = 0.0, 0.25, 0.5, 0.75 and 0.9). To be able to understand the effect of the near-tip radiation in the STR, we also tested γ for a tree in which this group was treated as a single taxon.

Because of a lack of reliable calibration points, we cannot rely on molecular clock analyses to estimate node ages in the *Sternopriscus* phylogeny. However, we attempt to approximate the age of the rapid radiation in the STR using the standard mutation rates of the *cox1* gene [44,45]. We apply the equation presented in Mendelson & Shaw [16] to estimate the relative speed of this radiation for comparison with other known rapid radiations in insects. For young and monophyletic radiations,

such as the STR, the equation is $\hat{r} = \ln N/t$, where $\hat{r}$ is the rate of diversification, N is the number of extant species, and t is the divergence time.

Phylogeographic structure analysis

We assembled a matrix of 710 bp of only *cox1* for 79 specimens of STR species to investigate the phylogeographic structure of this group. Additional sequences were obtained from Hendrich *et al.* [30]. The standard population genetic statistics Fu's Fs [46] and Tajima's D [47] were calculated, and mismatch distribution analyses to untangle demographic histories were performed using DnaSP 5.10 [48]. The multiple sequences were collapsed in haplotypes also using DnaSP 5.10. A minimum-spanning network was then inferred in Arlequin 3.5.1.3 [49] and used to create a minimum-spanning tree (MST) using Hapstar 0.5 [50]. The scalable vector graphics editor Inkscape 0.48 was further used to map geographic and taxonomic information on the MST.

Distinguishing incomplete lineage sorting from hybridization

We used an approach developed by Joly *et al.* [51], and employed in Joyce *et al.* [52] and Genner & Turner [53] to test whether the haplotype sharing between STR species was mainly the result of incomplete lineage sorting or influenced by hybridization. In this approach, mtDNA evolution is simulated using a species tree topology that assumes hybridization is absent. If low genetic distances between species pairs are due to incomplete lineage sorting, these similarly low genetic distances should be observed in the simulations. If low genetic distances between species pairs are due to hybridization, then significantly lower genetic distances should be present than observed in the simulations. First, we ran another *BEAST [21] analysis of a subset of the entire multilocus dataset containing only the STR species, using the HKY + G model for 11,000,000 generations (samplefreq = 1,000 and 10% burnin). Second, we used MrModeltest [54] to estimate the parameters of the substitution model for the *cox1* dataset from Hendrich *et al.* [30], which was previously used in the phylogeographic structure analysis. Third, we conducted a run of the JML software [55] using the same *cox1* dataset, the locus rate of *cox1* as yielded by *BEAST, a heredity scalar of 0.5, and the parameters yielded by MrModeltest.

Ecological niche modeling and analyses

In an attempt to detect possible divergence in response to climatic variables in their ranges, we created ecological niche models (ENMs) for the species of the STR. We excluded *S. montanus* and *S. williamsi* from the ENM analyses because of an insufficient number of localities. Our models were based on a total of 215 distribution points [27,28] (Additional file 2: Table S2) and unpublished data by L. Hendrich. With the exception of three records of *S. wehnckei*, all STR species occur in broad sympatry in southeastern Australia including Tasmania.

We preliminarily selected climate variables according to ecological requirements considered critical for the species. Bioclimatic variables [56] represent either annual means or maxima and minima in temperature and precipitation, or variables correlating temperature and precipitation, e.g., "mean temperature of wettest quarter" (BIO8). Such variables are useful for representing the seasonality of habitats [25]. After the preliminary selection, we used the ENMtools software [57] to calculate correlations between the selected climate layers in the area of interest. In our final selection, we removed layers until no two layers had correlation coefficients (r^2) higher than 0.75. ENMs for each species were created in Maxent 3.3.2 [58] (our procedure: Hawlitschek *et al.* [25]). Suitable background areas that were reachable by the species were defined by drawing minimum convex polygons around the species records, as suggested by Phillips *et al.* [59]. We conducted runs with 25% test percentage, 100 bootstrap repeats, jackknifing to measure variable importance and logistic output format. Model validation was performed by calculating the area under the curve (AUC) [60]. To compare ENMs of different *Sternopriscus* species, we measured niche overlap [57] in ENMtools. We also used ENMtools' niche identity test [61] with 500 repeats because the niche overlap values alone do not allow any statements whether the ENMs generated for the two species are identical or exhibit statistically significant differences. In each repeat of this test, pairwise comparisons of species distributions are conducted and their localities pooled, their identities are then randomized and two new random samples are extracted to generate a set of pseudoreplicates. The results are compared with the true calculated niche overlap (see above). The lower the true niche overlap is in comparison to the scores created by the pseudoreplicates of the pooled samples, the more significant the niche difference between the two compared species. Finally, we classified species by altitudinal and habitat preference and compared all data.

Results

Molecular phylogenetics

Bayes factor analyses favored separate partitioning of genes and codon positions (17 partitions in total). This was the most complex partition strategy tested. Substitution models applied were according to jModeltest: the GTR + I + G model (16 S rRNA, mitochondrial non-protein-coding), the GTR + G model (*cox1*, *cob*, mitochrondrial protein-coding), the HKY + I + G

model (18 S rRNA, nuclear non-protein-coding), and the HKY + G model (ARK, h3, h4, nuclear protein-coding). Bayesian, maximum likelihood, and maximum parsimony analyses revealed compatible topologies (Figure 1) that were largely congruent with the previously recognized classifications based on morphology. Here, we assign the four species previously supposed to be 'phylogenetically isolated' to either the *S. tarsalis* (*S. browni* and *S. wattsi*), or the *S. hansardii* (*S. eikei* and *S. marginatus*) group. Within the *S. tarsalis* group, all *S. tarsalis* complex species form a strongly supported clade (Figure 1).

The *BEAST species tree is largely congruent to the gene trees. The main difference is that in the gene trees,

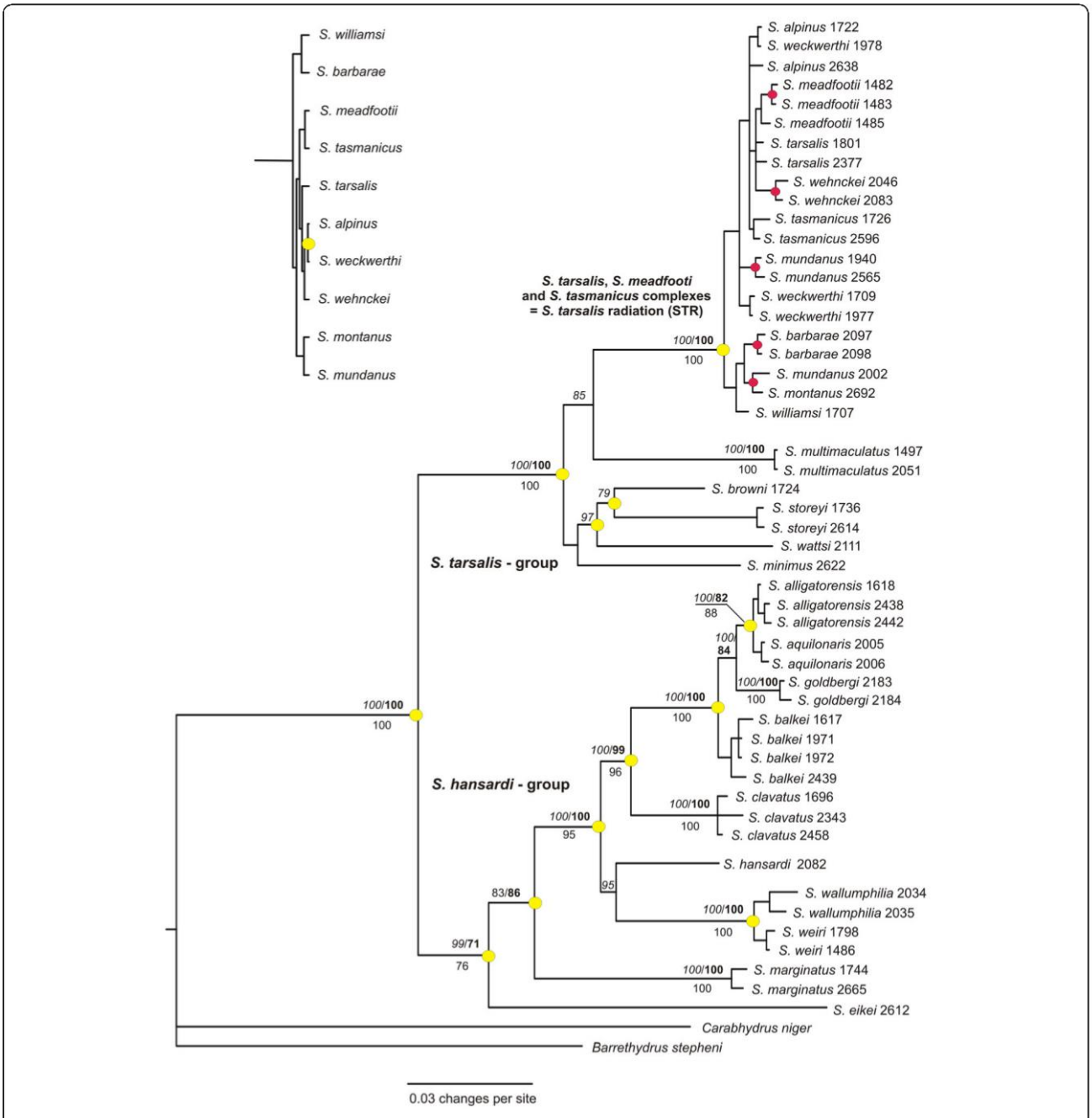


Figure 1 Phylogram of the genus *Sternopriscus*. The phylogram is based on a MrBayes tree with 7 gene loci and 3858 characters. Branch values are: MrBayes posterior probability (italic/above branch), RAxML bootstrap (bold/above branch), and TNT jackknife (below branch). Yellow circles mark nodes with *BEAST species tree posterior probabilities of 75 or more. Red circles mark nodes within the *S. tarsalis* radiation with PP, bootstrap and jackknife values of 75 or more (values not shown for layout reasons). Each tip represents one specimen. Specimen collection numbers are given after the species names. Upper left: *BEAST species tree fragment showing the *S. tarsalis* radiation specimens.

S. multimaculatus is the sister taxon to the STR, whereas in the *BEAST tree *S. minimus* is the sister taxon to the STR and *S. multimaculatus* is the sister taxon to all other members of the *S. tarsalis* group. Almost all species tree nodes within the STR are poorly supported. Notably, the analysis of the *BEAST run log file showed near-critically low posterior and prior effective sample sizes (< 120). This problem could neither be solved by repeating runs with higher sample frequencies nor with the application of simpler substitution models, as proposed in Pepper *et al.* [13], and indicates that the species tree results must be treated with caution.

The largest calculated *cox1 p*-distance between species in the STR was only 3.4% (*S. tarsalis*/*S. barbarae*), but interspecific distances may be as low as 0.3% (e.g., between *S. alpinus*, *S. mundanus* and *S. weckwerthi*, all belonging to different *S. tarsalis* complexes) or 0.2% (*S. alpinus/S. wehnckei*). Thus, no genetic distinction between the three complexes was possible because specimens often cluster with those belonging to other morphologically well-characterized species. This problem could not be solved by inspecting trees based on single or combined nuclear loci; the species *S. mundanus* and *S. weckwerthi* were polyphyletic in single-gene trees of *cob*, *cox1*, and ARK. The STR species shared identical haplotypes in all other nuclear genes studied.

Diversification analyses

Figure 2 shows the LTT plot for *Sternopriscus*. APE yielded a positive γ value of 3.22 (p = 0.0013*). According to the mccr test, the critical value is 1.73 (p = 0.9*10E-3**) and is therefore met by the true value of γ. The test in Laser yielded a Yule-2-rate model as significantly better than the next best model, which was a constant rate birth-death model. The level of significance was highest (p = 0.0073*) for equal rates of b (birth) and d (death) (both 0.5), but all tested combinations of b and d yielded significant test results. In the test run in which the *S. tarsalis*-group was treated as a single clade, γ was negative but not significant at a value of −0.01 (p = 0.4956). This means that for this dataset the null hypothesis that the diversification rates have not decreased over time cannot be rejected.

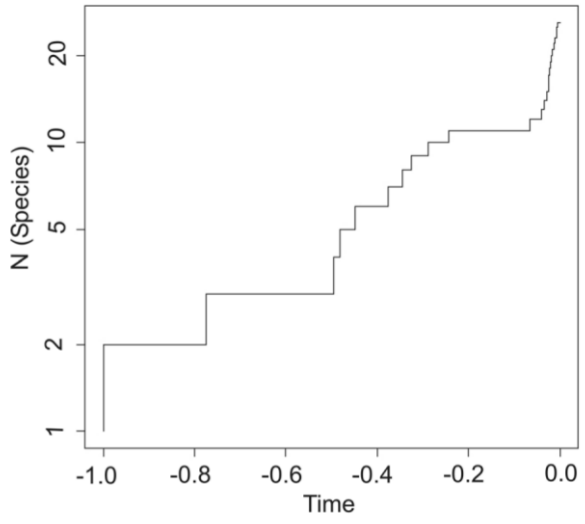


Figure 2 Lineage-through-time (LTT) plot for the genus *Sternopriscus*. Relative time (−1.0 is the time of the initial lineage split within the genus, 0.0 is the present) is given on the x-axis, number of species is given logarithmically on the y-axis.

The STR appears to have a thorough effect on the diversification analysis of the genus *Sternopriscus*. A high positive γ represents a rather unusual condition [6]. While many phylogenies are characterized by a decreasing rate of diversification (logistic growth or impact of extinctions [62]), a γ = 3.22 suggests a diversification rate that is highly increasing over time. This pattern is hard to explain in general. In the case of *Sternopriscus*, it appears appropriate to attribute this pattern to the recent speciation burst of the STR, which comprises 10 of 28 known species. This view is also supported by the test results that indicate a Yule-2-rate model as the most adequate, which fits to a sudden shift in diversification rates.

Papdopoulou *et al.* [44] suggested using substitution rates of 3.54% *cox1* divergence per MY which suggest an origin of the STR *c.* 0.96 MYA, and interspecific distances indicate divergence times as recent as 60,000 to 80,000 years ago. The slower substitution rate (2.3%) suggested by Brower [45] yields an approximate origin of the STR around 1.48 MYA and interspecific divergence times of 87,000 to 130,000 years ago (but see Papadopoulou *et al.* [44] for a discussion of these estimates). The equation by Mendelson & Shaw [16] was used to estimate speciation rates in the STR. Applying the proposed rate of Papdopoulou *et al.* [44], we estimate a speciation rate in the STR of 2.40 species per MY. Applying the proposed rate of Brower [45], we estimate a speciation rate in the STR of of 1.56 species per MY.

Phylogeographic structure

The matrix of 79 *cox1* sequences contained 69 polymorphic sites with a nucleotide diversity of π = 0.0121 and a haplotype diversity of H = 0.9815. We identified 61 distinct and mostly unique haplotypes within the STR with only 8 haplotypes comprising more than one sequence. Neither geographic nor taxonomic (Figure 3) mapping on the star-like MST yielded a comprehensive pattern. More precisely, no geographic structuring could be noticed based on the zoning of Australia, and the haplotypes of individuals of identical species were not systematically gathered in groups. Interestingly, the MST

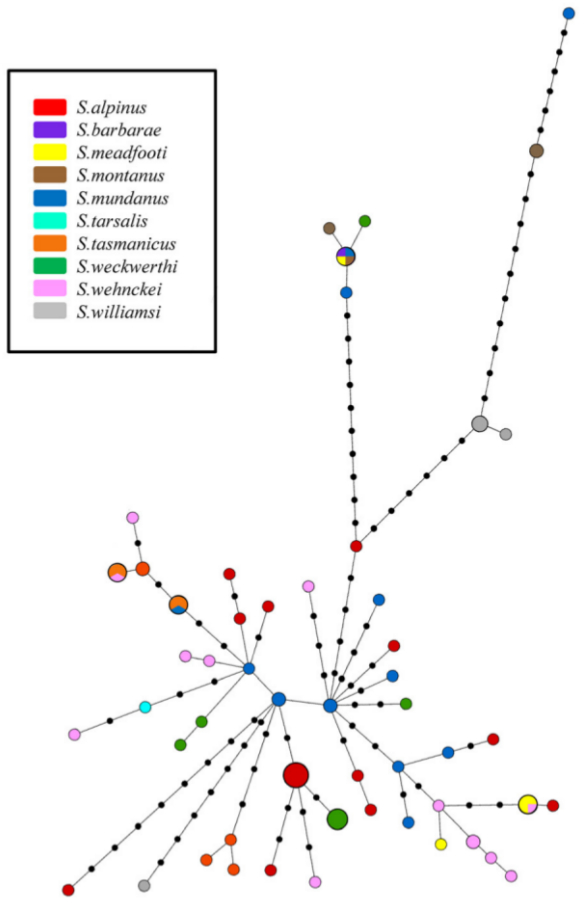


Figure 3 Minimum spanning tree of haplotypes of the *Sternopriscus tarsalis* radiation. The tree was created in Hapstar 0.5. Colors code the species determined according to morphology. Colored circles represent haplotypes, black dots represent mutational steps that are not represented by any haplotype.

appears to be composed of two central haplotypes of South Australian and Victorian *S. mundanus* from which the rest of the sequences appears to have derived. In addition, even if there is a lack of geographical or taxonomic structuration, one might notice that several haplotypes representing different species are separated from the central network by a deep break of multiple mutation steps. While Tajima's D value does not significantly support a scenario of demographic expansion (D = −1.27773, p-value = 0.06), Fu's Fs significantly support such a demographic history (Fs = −35.731, p-value = 0.01) (see Tajima [47] and Fu [46] regarding the interpretation of Tajima's and Fu's statistics). However, the mismatch distribution analyses yield a multimodal distribution of the pairwise genetic distances, which favors a scenario of demographic equilibrium for the STR even if unimodal distributions are recovered only for recent and fast expansions [63].

Incomplete lineage sorting vs. hybridization

*BEAST yielded a high relative locus rate of 2.332 for *cox1*, which was expected because many other markers included in our multilocus dataset, mainly nuclear markers, are known to evolve slower. The results of the JML run are given in Table 1. All species pairs exhibit genetic distances that are not significantly lower than expected. Thus, we cannot reject the hypothesis of incomplete lineage sorting in any cases.

Ecological niche modeling

Figure 4 summarizes all distribution points for all STR species and Figure 5 summarizes climate variables used for the creation of ENMs. The ENMs for the 8 STR species analyzed, supplemented with other ecological data,

Table 1 Results of the JML run

Distance obs./exp.	*S. alp.*	*S. bar.*	*S. mea.*	*S. mon.*	*S. mun.*	*S. tar.*	*S. tas.*	*S. wec.*	*S. weh.*	*S. wil.*
S. alp.		4.83	2.42	4.83	2.42	2.42	2.42	1.21#	1.21#	4.83
S. bar.	14.81		4.83#	2.42#	4.83#	4.83	4.83	4.83+	4.83	2.42
S. mea.	4.44	0		4.83#	2.42#	2.42	1.21	2.42+	2.42#	4.83
S. mon.	14.81	0	0		4.83#	4.83	4.83	4.83+	4.83	2.42
S. mun.	1.48	0	0	0		2.42	2.42#	2.42+	2.42#	4.83
S. tar.	5.92	23.70	8.89	23.70	4.44		2.42	2.42	2.42+	4.83
S. tas.	5.93	22.22	8.89	22.22	0	5.93		2.42	2.42#	4.83
S. wec.	0	1.48	1.48	1.48	1.48	5.93	4.44		1.21#	4.83
S. weh.	0	19.26	0	19.26	0	1.48	0	0		4.83
S. wil.	10.37	19.26	16.30	16.30	11.85	16.30	14.81	14.81	11.85	

Minimum genetic distance (*1,000), as estimated by JML, of STR species pairs. Lower left: observed minimum genetic distance. Upper right: expected minimum genetic distance (median). Species pairs in which the observed genetic distance is 0 due to the sharing of haplotypes are indicated by #. Species pairs in which the observed minimum genetic distance is higher than the expected distance are indicated by +. There is no case in which the probability that the minimum observed genetic distance is lower than expected is significant ($p \leq 0.05$).

Figure 4 Distribution of species of the *Sternopriscus tarsalis* radiation. Red dots represent specimen localities used for ecological niche modeling.

are given in Figure 6. AUC values for all models range from 0.981 to 0.997. Because all values are > 0.9, the ability to distinguish presence from random background points is considered "very good" for all models according to Swets [60]. We preliminarily selected the climate layers "maximum temperature of the warmest month" (BIO5), "minimum temperature of the coldest month" (BIO6), "mean temperature of the wettest quarter" (BIO8), "mean temperature of the driest quarter" (BIO9), "precipitation of the wettest month" (BIO13), "precipitation of the driest month" (BIO14), "precipitation of the warmest quarter" (BIO18) and "precipitation of the coldest quarter" (BIO19). In our final selection, we omitted BIO13 and BIO14 because of correlation coefficients with other variables of $r^2 > 0.75$. Thus, all models presented here are based on six climate variables. Jackknifing to measure the importance of variables showed that either "maximum temperature of the warmest month" (BIO5: *S. barbarae*, *S. weckwerthi*, *S. wehnckei*), "mean temperature of the wettest quarter" (BIO8: *S. alpinus*, *S. mundanus*), or "precipitation of the coldest quarter" (BIO19: *S. meadfootii*, *S. tarsalis*, *S. tasmanicus*) were the most important variables in creating ENMs. Niche overlap values (I and D) and identity test results are given in Table 2. The results of the identity test are highly significant (Bonferroni corrected) for I in all and for D in nearly all pairwise species comparisons. However, the null hypothesis of identity in the ENMs of two compared species can be rejected only if the true calculated niche overlap is below the 99.9% confidence interval of the values generated in the identity test. In a few cases, the true calculated niche overlap is above this interval, and the null hypothesis of niche identity cannot be rejected [61].

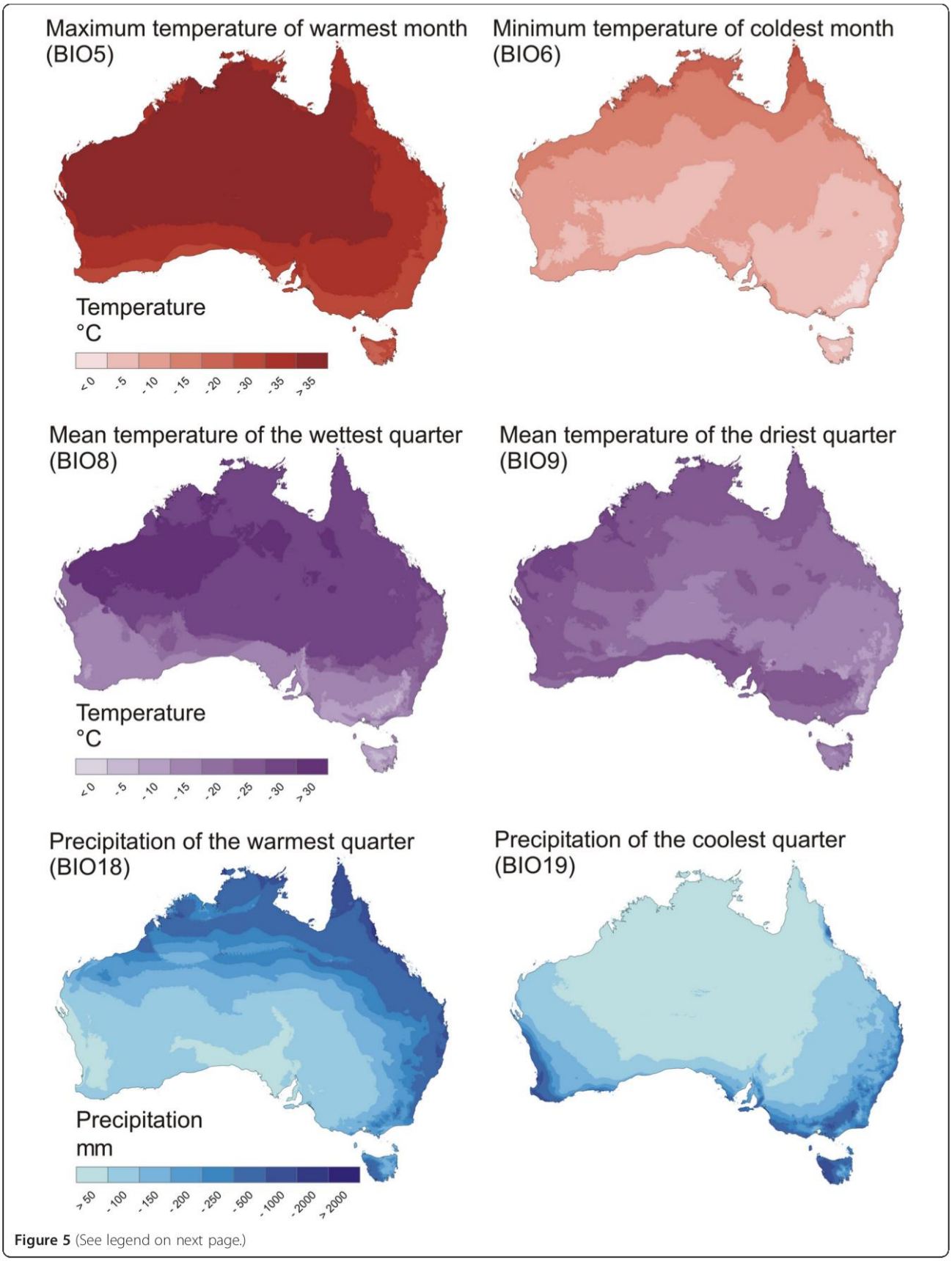


Figure 5 (See legend on next page.)

(See figure on previous page.)
Figure 5 Climate variables used for ENM creation. Variables were selected to represent the effects of temperature, precipitation and seasonality.

Ecological analyses

All species of the STR were compared for their preferences in altitude and habitat and for the most important climate factor in their ENM, which resulted from the jackknifing test in the ENM runs. Table 3 displays these three factors for all species coded by numbers for easy comparison. Only *S. tasmanicus* and *S. tarsalis* are identical in all three factors. *S. montanus* and *S. williamsi* might be identical to *S. alpinus* or *S. weckwerthi* depending on the most important climate factor, but no ENMs could be created. Within each of the three complexes in the *S. tarsalis* group, no two species are identical in all three factors.

Discussion

In the opening section of this article, we suggested three hypotheses: (1) species delimitation in the STR can be supported by genetic or ecological data; (2) the STR species originated in a rapid Pleistocene diversification event; and (3) Pleistocene climate oscillations promoted the radiation of the STR. In the following, we will discuss how our results support these hypotheses.

Our data shows that the molecular methods applied in our study do not serve to unambiguously distinguish and delimit the species of the STR. This is because of the widespread genotype sharing of mitochondrial genes and lack of diversification in nuclear genes between these species. However, the analysis of our ecological data shows that STR species appear to respond differently to ecological variables. Below, we initially discuss whether incomplete lineage sorting or hybridization may have caused the abundance of shared haplotypes in the STR. Then, we discuss the importance of the results of our ecological analyses in the context of the entire genus, and specifically for the STR.

Genotype sharings between species may be explained by incomplete lineage sorting, by hybridization, or a combination of both. Funk & Omland [19] also mention imperfect taxonomy, inadequate phylogenetic information and paralogs as causes for genotype sharing. However, the taxonomy of *Sternopriscus* based on morphological characters is well supported [27,28], and our multi-gene phylogeny is well supported by different analytical approaches. Paralogs can almost certainly be excluded because the patterns of species polyphyly are repeated by different mitochondrial and nuclear markers.

Hybridization, as a reason for genotype sharing in closely related species, has been proposed for various animal groups [64,65], including groups with strong sexual selection (e.g., mating calls [66]), and has been shown to contribute to speciation [64]. However, in the case of *Sternopriscus*, the results of our analyses, the diversity in genital morphology, and the absence of specimens identifiable as hybrids, do not support hybridization [67]. Incomplete lineage sorting, or the retention of ancestral polymorphism, is the more likely explanation for genotype sharing in the case of the STR. Incomplete lineage sorting has often been recognized as a problem in resolving phylogenies of young and closely related taxa [68]. This phenomenon affects nuclear loci more commonly than faster evolving mitochondrial loci, but mitochondrial genes can be equally affected, particularly in closely related taxa where hardly any diversification in nuclear genes is found [19]. Incomplete lineage sorting as an explanation for haplotype sharings in the STR supports the view that the STR is a recent radiation.

A comparison of our ecological findings concerning the STR species with data on other *Sternopriscus* species shows that the STR occupies ecological ranges similar to those of other related species. The currently known altitudinal distribution and ecology of all *Sternopriscus* species in Australia is shown in Additional file 3: Table S3, modified after Hendrich & Watts [27,28]. 10 species of the genus are rheophilic and inhabit rivers and streams that are mainly of intermittent character. 11 species are acidophilic and live in seasonal or permanent swamps, ponds and pools of different types of peatlands. 7 species appear to be more or less eurytopic and occur in various water bodies in open or forested country. The highest species diversity is in lowland or coastal areas and hilly or low mountain ranges from 0 to 500 m. Only 6 species were collected at 1000 m or above (*S. alpinus*, *S. meadfootii*, *S. montanus*, *S. mundanus*, *S. williamsi* and *S. weckwerthi*).

Within the STR, all species inhabit broadly overlapping areas in mesic southeast Australia, except for a few localities of *S. wehnckei* in the northeast (the Eastern Coastal Australia region and small parts of the Murray-Darling region of Abell *et al.* [29]. Many species also inhabit Tasmania, including two endemics (Bass Strait Drainages and Southern Tasmania). ENMs indicate niche diversification within this group of closely related and broadly sympatric species. Aside from the high levels of significance in the identity test, the degree of niche diversification is hard to measure. Therefore, we rely on the importance of the various climate variables used to characterize the species ENMs. The variables of highest importance are

Figure 6 Ecological Niche Models (ENMs) for species of the *Sternopriscus tarsalis* radiation. No ENMs were created for *S. montanus* and *S. williamsi* because of insufficient locality data. High Maxent values indicate high probabilities of occurrence of a species on a raster square (2.5 arc-minutes resolution). Maps include species name, taxonomic affinity, altitudinal range, habitat type and climate variable of highest importance in the ENM.

"maximum temperature of the warmest month" (BIO5), "mean temperature of the wettest quarter" (BIO8), or "precipitation of the coldest quarter" (BIO19). Figure 5 shows that all the species studied inhabit areas with relatively low maximum temperatures, with the lowest on Tasmania. The two species most characterized by this factor are the two Tasmanian endemics, *S. barbarae* and *S. weckwerthi*. A distinction between the two remaining factors is more

Table 2 Results of the niche identity test

Overlap D/I	*S. alp.*	*S. bar.*	*S. mea.*	*S. mun.*	*S. tar.*	*S. tas.*	*S. wec.*	*S. weh.*
S. alp.	0	0.674**	0.682**	0.676**	0.661*#	0.651**#	0.648**	0.571**
S. bar.	0.506**	0	0.733**#	0.569**	0.680**#	0.735**#	0.755**	0.582**
S. mea.	0.481**	0.589**#	0	0.691**	0.801**	0.847**#	0.606**	0.684**
S. mun.	0.474**	0.327**	0.496**	0	0.661**	0.602**	0.520**	0.637**
S. tar.	0.476	0.472	0.711**	0.456**	0	0.759**	0.548**	0.756**
S. tas.	0.433**	0.560	0.762**#	0.378**	0.648**	0	0.583**	0.627**
S. wec.	0.451**	0.642**	0.367**	0.241**	0.282**	0.331**	0	0.459**
S. weh.	0.374**	0.356**	0.523**	0.444**	0.639**	0.419**	0.177**	0

Niche overlap values (D and I), calculated with ENMtools, are given for species pairs and are mostly lower than the randomized overlap levels generated in the identity test at significant (*, $p \leq 0.05$, Bonferroni corrected) or highly significant (**, $p \leq 0.001$, Bonferroni corrected) level. This means that niches are more divergent than expected at random. In some cases, results are not significant, or significantly higher than the randomized overlap (indicated by #). In these cases, niches are not more divergent than expected by random. Note that results yielded by D and I do not accord in all cases.

Table 3 Taxonomic affinities and ecological preferences of species in the *Sternopriscus tarsalis* radiation

Species	Complex	Altitude	Habitat	Climate
S. alpinus	*2*	2	2	1
S. tasmanicus	*2*	0	1*	2
S. wehnckei	*2*	0	1	0
S. barbarae	*1*	0	0	0
S. meadfootii	*1*	1	1	2
S. montanus	*1*	2	2	?
S. mundanus	*1*	1 **	2	1
S. tarsalis	*0*	0	1	2
S. weckwerthi	*0*	2	2	0
S. williamsi	*1*	1	2	?

Complex: 0 = *S. tarsalis*, 1 = *S. meadfootii*, 2 = *S. tasmanicus*. Altitude: preferred altitude range, 0 = < 500 m, 1 = 500 – 1000 m, 2 = > 1000 m. Habitat: 0 = rheophilic, 1 = eurytopic, 2 = acidophilic. Climate: according to the dominating climate variables in the ENM, 0 = cool summers, 1 = cool winters, 2 = wet winters. *: Also occurs in habitats with moderate salinity. **: Actual altitudinal range is 200 – 1550 m.

difficult. Considering Figure 5, "mean temperature of the wettest quarter" is lowest in areas where winters (the wettest quarter in our region of interest) are cold, whereas "precipitation of the coldest quarter" is highest where winters are wet. Some species (e.g., the high-altitude *S. alpinus*) may be tolerant of winter temperatures that are too low for other species, whereas other species are more dependent on sufficient precipitation. Species that require the latter are eurytopic species that also inhabit ephemeral waters, such as ponds at the edge of rivers and creeks, which are only filled after heavy rainfall. The acidophilic species, which inhabit more permanent water bodies with dense vegetation, are often "cold winter" species.

The low divergences between haplotypes in the STR species suggest that these species originated in a recent and rapid radiation. Unfortunately, we could not rely on any calibration points to support our molecular clock approach. Instead, we attempted to estimate the origin of the STR based on standard *cox1* mutation rates [44,45]. We estimated an origin of *c.* 0.96 to 1.48 MYA, which leads to an estimated speciation rate of 2.40 or 1.56 species per MY. Genetic distance might indicate the age of the ancestral species, however divergence time estimates for the extant species should not be considered reliable beyond assumption of a comparably recent origin of the STR. This fact alone, however, suggests that the STR is an exceptional event for what is known of aquatic beetles. For other insect groups, little evidence exists for similarly fast diversification events. The fastest rate (4.17 species per MY) was estimated for a clade of 6 species of Hawaiian crickets over 0.43 MY [16]. However, in the same study, for a related clade comprising 11 species, the estimated rate was much lower at 1.26 species per MY over 1.9 MY. Additional estimates are available for *Galagete* moths in the Galapagos [17] of 0.8 species per MY (n = 12, t = 1.8 MY) and for Japanese *Ohomopertus* ground beetles [18] of 1.92 (n = 15, t = 1.4 MY) to 2.37 species per MY (n = 6, t = 0.76 MY). The average speciation rate in insects was proposed to be 0.16 species per MY [15]. This comparison shows that rapid radiation events, as exemplified in the STR, appear to be exceptional among insects and particularly in continental faunas because all other examples recorded were island radiations.

Species groups that originated from rapid radiation events have been detected in almost all organismic groups and habitats [69]. An overview of many recent and past events suggests three major promoters of rapid radiations: the appearance of a key innovation that allows the exploitation of previously unexploited resources or habitats [70], the availability of new resources [71], and the availability of new habitats, e.g., because of a rare colonization event or drastic environmental changes [72,73]. In the case of the STR, we find no evident key innovation distinguishing this group from other *Sternopriscus* species. We have no data concerning internal morphology or physiology. Additionally, our data show that the observation that STR species have ecological requirements similar to those of other *Sternopriscus* species does not indicate the presence of any key innovations. There is also no indication of any new resource that could be specifically exploited by the STR species. Therefore, we explore the possibility that drastic environmental changes during the Pleistocene climate oscillations mediated the radiation of the STR species.

During most of the Cenozoic, the climate of Australia was hot and humid and currently remains so in the northern rainforest areas [11]. Aridification began in the Miocene (*c.* 15 MYA) and gradually led to the disappearance of forests and to the spread of deserts over much of the present continent. Most of today's sand deserts, however, are geologically younger and appeared only after the final boost of aridification that accompanied the Ice Ages, particularly since the later Pleistocene (*c.* 0.9 MYA). The climate was subjected to large oscillations in temperature and rainfall, which drove many groups of organisms into refugia and also promoted speciation [12,13]. Our results also document a strong and abrupt increase in speciation in the genus *Sternopriscus* about 1 to 1.5 MYA, represented by the STR. This age estimate is congruent with the Pleistocene oscillations. Byrne *et al.* [12] present cases of organisms restricted to mesic habitats that were formerly most likely more widespread, but today occupy relictual areas with suitable climates. However, some of the young species of the STR occupy rather large areas in southwestern Australia. This distribution indicates good dispersal abilities, which are

necessary for organisms that inhabit habitats of relatively low persistence [74]. Ribera & Vogler [75] argue that for this reason, beetle species that inhabit lentic aquatic habitats often have better dispersal abilities than those inhabiting lotic habitats. However, it is possible that the STR species of lotic habitats only recently derived from an ancestor adapted to lentic habitats with good dispersal abilities that are maintained in the newly derived species.

Speciation in Pleistocene refugia was previously described for dytiscid beetles on the Iberian Peninsula [9]. During the Pleistocene climate oscillations, the ancestral species of the STR might have been forced into ongoing cycles of retreating into, and the re-expansion from, refugia. Under the recurrent, extremely unsuitable climate conditions, the isolation of small populations over many generations might have promoted speciation and the fixation of morphological traits. This scenario might also explain the lack of clear geographic or taxonomic structuring in the striking haplotypic diversity presented by the STR species. This diversity might be attributed to the cycles of expansion and retreat that repeatedly isolated haplotypes in various geographic locations before newly allowing the expansion and colonization of other areas.

The phenomenon of groups of young and closely related species within a defined distributional range is most familiar in ichthyology, in which it was termed "species flock". Among the most prominent species flocks are the cichlids of the African Great Lakes and other lake ecosystems around the world, the Sailfin Silversides of Sulawesi, and the Notothenioid Antarctic Ice Fishes (see review in Schön & Martens [76]). Schön & Martens [76] summarize the criteria for naming a group of species a species flock as "speciosity [= species-richness], monophyly and endemicity". Compared with the large fish species flocks, the STR is poor in species. Nevertheless, the number of species is "disproportionally high" [77] in relation to the surrounding areas, as no other region in Australia is inhabited by a comparable assemblage of closely related species. In the last decade, an increasing number of less species-rich radiations have been termed species flocks with as little as 3 or 4 species [76,78]. Most other species flocks inhabit lakes, islands or archipelagoes. These are areas more "narrowly circumscribed" [77] than the area of endemism of the STR, which can be broadly termed "the southeast Australian region". Most STR species have relatively large ranges that do not share a common limit and sometimes do not even overlap. Our results show that STR species often occupy different habitat types. Additionally, the clade is not strictly endemic to southeastern Australia, as shown by the northeastern records of *S. wehnckei*. Based on this criterion, other rapid radiations among insects [16,17] are much more adequate examples of species flocks.

Conclusions

Our results provide evidence that STR species are the result of an extremely recent, most likely Pleistocene, radiation. The STR species cannot be distinguished with the molecular methods used in this study, however, the species show clear divergences in their responses to ecological factors of habitat type and climate. We proposed a scenario in which the Pleistocene climate oscillations led to the repeated restriction and expansion of the ranges of the ancestral species of the STR, which may have promoted fixation of ecological adaptations and morphological traits in small and isolated populations restricted to refugia. This suggests that *Sternopriscus* is an example for the hypothesis that Pleistocene refugia promoted speciation.

Taking this scenario into account, the STR does not appear as an evolving or fully evolved species flock but as a radiation based on a species flock. While possibly confined to a narrowly circumscribed area during the Pleistocene, the STR species were able to break the boundaries of their refugia with the end of the Ice Ages and increase their ranges. Today, because the species are no longer confined to a common limited area, the term "species flock" may best fit a stage in speciation the STR has previously passed.

Additional files

Additional file 1: Table S1. Sequences of primers used for PCR and sequencing. Forward (F) and reverse (R) primers are given. Mitochrondrial gene loci: CO1 = cytochrome C oxidase 1, CytB = cytochrome B oxidase, 16 S = 16 S ribosomal RNA. Nuclear gene loci: H3 = histone 3, H4 = histone 4, ARK = arginine kinase,18 S = 18 S ribosomal RNA. I = inosine.

Additional file 2: Table S2. Localities of *Sternopriscus* species used in Ecological Niche Modeling. Coordinates are given in decimal degrees.

Additional file 3: Table S3. Ecological data on all *Sternopriscus* species. Data from Hendrich & Watts [27,28].

Abbreviations

ENM: Ecological niche modeling; MST: Minimum spanning tree; MY: Million years; MYA: Million years ago; STR: *Sternopriscus tarsalis* radiation.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

OH performed the laboratory work, the molecular genetic studies, the diversification analyses, the ecological niche modeling and analyses, and drafted the manuscript. LH collected the samples and ecological data and helped to draft the manuscript. ME coordinated the diversification analyses. EFAT conducted the phylogeographic analyses. MJG conducted the analysis of hybridization vs. incomplete lineage sorting. MB conceived the study, participated in its design and coordination, and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work and ongoing research on the Australian water beetle fauna was supported by grants from the German Research Foundation (DFG) to Lars Hendrich (HE5729/1-1) and Michael Balke (BA2152/7-1). We are grateful to the Department of Environment and Conservation in Western Australia for

giving us permission to conduct scientific research in Western Australia (Permit numbers: SF 003017 and NE 002348), and the Parks and Wildlife Commission of the Northern Territory for giving us permission to conduct scientific research in the Northern Territory (Permit Number: 23929 and RK-400/RK- 660). We are further grateful to the Department of Environment and Conservation in New South Wales (Scientific License No. S12040) for giving us permission to conduct scientific research in the National and State Parks. We thank the CIPRES portal for computing resources, Amanda Glaser and Jeannine Marquardt, Munich, for their assistance in the phylogenetic analyses, and Ulrich Schliewen, Munich, Jesús Gómez-Zurita, Barcelona, and the editor and two anonymous referees for greatly improving this manuscript with their comments.

Author details
[1]Zoologische Staatssammlung, Münchhausenstr. 21, Munich 81247, Germany. [2]Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA. [3]School of Biological Sciences, University of Bristol, Woodland Road, Bristol BS8 1UG, UK. [4]GeoBioCenter, Ludwig-Maximilians-Universität, Richard-Wagner-Str. 10, Munich 80333, Germany.

Received: 27 February 2012 Accepted: 30 July 2012
Published: 9 August 2012

References

1. Barraclough TG, Vogler AP, Harvey PH: **Revealing the factors that promote speciation.** *Phil Trans R Soc B* 1998, **353**:241–249.
2. Hillebrand H: **On the generality of the latitudinal diversity gradient.** *Am Nat* 2004, **163**:192–211.
3. Mittelbach GG, Schemske DW, Cornell HV, Allen AP, Brown JM, Bush MB, Harrison SP, Hurlbert AH, Knowlton N, Lessios HA, McCain CM, McCune AR, McDade LA, McPeek MA, Near TJ, Price TD, Ricklefs RE, Roy K, Sax DF, Schluter D, Sobel JM, Turelli M: **Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography.** *Ecol Lett* 2007, **10**:315–331.
4. Benton MJ: *When life nearly died: The greatest mass extinctions of all time.* London, UK: Thames & Hudson; 2005.
5. Ricklefs RE: **Estimating diversification rates from phylogenetic information.** *Trends Ecol Evol* 2007, **22**:601–610.
6. Quental TB, Marshall CR: **Diversity dynamics: molecular phylogenies need the fossil record.** *Trends Ecol Evol* 2010, **25**:434–441.
7. Hewitt GM: **Genetic consequences of climatic oscillations in the Quaternary.** *Phil Trans R Soc B* 2004, **359**:183–195.
8. Beheregaray LB: **Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere.** *Mol Ecol* 2008, **17**:3754–3774.
9. Ribera I, Vogler AP: **Speciation of Iberian diving beetles in Pleistocene refugia (Coleoptera, Dytiscidae).** *Mol Ecol* 2004, **13**:179–193.
10. Jansson R, Dynesius M: **The fate of clades in a world of recurrent climatic change: Milankovitch oscillations and evolution.** *Ann Rev Ecol Syst* 2002, **33**:741–777.
11. Frakes LA, McGowran B, Bowler JM: **Evolution of Australian Environments**. In *Fauna of Australia. General Articles.* 1Ath edition. Edited by Dyne GR, Walton DW. Canberra: Australian Government Publishing Service; 1987:1–17.
12. Byrne M, Yeates DK, Joseph L, Kearney M, Bowler J, Williams MAJ, Cooper S, Donnellan SC, Keogh JS, Leys R, Melville J, Murphy DJ, Porch N, Wyrwoll KH: **Birth of a biome: insights into the assembly and maintenance of the Australian arid zone biota.** *Mol Ecol* 2008, **17**:4398–4417.
13. Pepper M, Doughty P, Hutchinson MN, Keogh JS: **Ancient drainages divide cryptic species in Australia's arid zone: Morphological and multi-gene evidence for four new species of Beaked Geckos (*Rhynchoedura*).** *Mol Phylogenet Evol* 2011, **61**:810–822.
14. Leys R, Watts CHS, Cooper SJB, Humphreys WF: **Evolution of subterranean Diving Beetles (Coleoptera: Dytiscidae: Hydroporini, Bidessini) in the arid zone of Australia.** *Evolution* 2003, **57**:2819–2834.
15. Coyne JA, Orr HA: *Speciation*. Sunderland, MA: Sinauer Ass; 2004.
16. Mendelson TC, Shaw KL: **Sexual behavior: Rapid speciation in an arthropod.** *Nature* 2005, **433**:375–376.
17. Schmitz P, Cibois A, Landry B: **Molecular phylogeny and dating of an insular endemic moth radiation inferred from mitochondrial and nuclear genes: The genus *Galagete* (Lepidoptera: Autostichidae) of the Galapagos Islands.** *Mol Phylogenet Evol* 2007, **45**:180–192.
18. Sota T, Nagata N: **Diversification in a fluctuating island setting: rapid radiation of *Ohomopterus* ground beetles in the Japanese Islands.** *Phil Trans R Soc B* 2008, **363**:3377–3390.
19. Funk DJ, Omland KE: **Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA.** *Annu Rev Ecol Evol Syst* 2003, **34**:397–423.
20. Degnan JH, Rosenberg NA: **Discordance of species trees with their most likely gene trees.** *PLoS Genetics* 2006, **2**:762–768.
21. Heled J, Drummond AJ: **Bayesian inference of species trees from multilocus data.** *Mol Biol Evol* 2010, **27**:570–580.
22. Graham CH, Ron SR, Santos JC, Schneider CJ, Moritz C: **Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs.** *Evolution* 2004, **58**:1781–1793.
23. Soberón J, Peterson AT: **Interpretation of models of fundamental ecological niches and species' distributional areas.** *Biodiv Infor* 2005, **2**:1–10.
24. Guisan A, Zimmermann N: **Predictive habitat distribution models in ecology.** *Ecol Model* 2000, **135**:147–189.
25. Hawlitschek O, Porch N, Hendrich L, Balke M: **Ecological niche modelling and nDNA sequencing support a new, morphologically cryptic beetle species unveiled by DNA barcoding.** *PLoS One* 2011, **6**:e16662.
26. Kozak KH, Wiens JJ: **Does niche conservatism promote speciation? A case study in North American salamanders.** *Evolution* 2006, **60**:2604–2621.
27. Hendrich L, Watts CHS: **Taxonomic revision of the Australian genus *Sternopriscus* Sharp, 1882 (Coleoptera: Dytiscidae, Hydroporinae).** *Koleopterol Rdsch* 2004, **74**:75–142.
28. Hendrich L, Watts CHS: **Update of Australian *Sternopriscus* Sharp, 1882 with description of three new species (Coleoptera: Dytiscidae: Hydroporinae).** *Koleopterol Rdsch* 1882, **77**:49–59.
29. Abell R, Thieme M, Revenga C, Bryer M, Kottelat M, Bogutskaya N, Coad B, Mandrak N, Contreras-Balderas S, Bussing W, Stiassny MLJ, Skelton P, Allen GR, Unmack P, Naseka A, Ng R, Sindorf N, Robertson J, Armijo E, Higgins J, Heibel TJ, Wikramanayake E, Olson D, Lopez HL, Reis RE, Lundberg JG, Sabaj Perez MH, Petry P: **Freshwater ecoregions of the world: a new map of biogeographic units for freshwater biodiversity conservation.** *BioScience* 2008, **58**:403–414.
30. Hendrich L, Pons J, Ribera I, Balke M: **Mitochondrial Cox1 sequence data reliably uncover patterns of insect diversity but suffer from high lineage-idiosyncratic error rates.** *PLoS One* 2010, **5**:e14448.
31. Posada D: **jModelTest: Phylogenetic model averaging.** *Mol Biol Evol* 2008, **25**:1253–1256.
32. Huelsenbeck JP, Ronquist F: **MrBAYES: Bayesian inference of phylogenetic trees.** *Bioinformatics* 2001, **17**:754–755.
33. Rambaut A, Drummond AJ: *Tracer v1.4*. 2007. http://beast.bio.ed.ac.uk/Tracer.
34. Silvestro D, Michalak I: **raxmlGUI: a graphical front-end for RAxML.** *Org Divers Evol* 2012, in press.
35. Goloboff PA, Farris JS, Nixon KC: **TNT, a free program for phylogenetic analysis.** *Cladistics* 2008, **24**:774–786.
36. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32**:1792–1797.
37. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.** *Mol Biol Evol* 2011, **28**:2731–2739.
38. Paradis E, Claude J, Strimmer K: **APE: analyses of phylogenetics and evolution in R language.** *Bioinformatics* 2004, **20**:289–290.
39. Rabosky DL: **LASER, a maximum likelihood toolkit for detecting temporal shifts in diversification rates from molecular phylogenies.** *Evol Bioinform Online* 2006, **2**:247–250.
40. Nee S, Holmes EC, May RM, Harvey PH: **Extinction rates can be estimated from molecular phylogenies.** *Phil Trans R Soc B* 1994, **344**:77–82.
41. Pybus OG, Harvey PH: **Testing macroevolutionary models using incomplete molecular phylogenies.** *Proc R Soc B* 2000, **267**:2267–2272.
42. Rabosky DL: **Likelihood methods for detecting temporal shifts in diversification rates.** *Evolution* 2006, **60**:1152–1164.
43. Papadopoulou A, Anastasiou I, Vogler AP: **Revisiting the insect mitochondrial molecular clock: Mid-Aegean trench calibration.** *Mol Biol Evol* 2010, **27**:1659–1672.

44. Brower AVZ: **Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA Evolution.** *Proc Natl Acad Sci USA* 1994, **91:**6491–6495.
45. Fu YH: **Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection.** *Genetics* 1997, **147:**915–925.
46. Tajima F: **Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.** *Genetics* 1989, **123:**585–595.
47. Librado P, Rozas J: **DnaSP v5: A software for comprehensive analysis of DNA polymorphism data.** *Bioinformatics* 2009, **25:**1451–1452.
48. Excoffier L, Lischer HEL: **Arlequin suite ver 3.5: A new series of programs to performpopulation genetics analyses under Linux and Windows.** *Mol Ecol Res* 2010, **10:**564–567.
49. Teacher AGF, Griffiths DJ: **HapStar: Automated haplotype network layout and visualisation.** *Mol Ecol Res* 2011, **11:**151–153.
50. Joly S, McLenachan PA, Lockhart PJ: **A statistical approach for distinguishing hybridization and incomplete lineage sorting.** *Am Nat* 2009, **174:**54–70.
51. Joyce DA, Lunt DH, Genner MJ, Turner GF, Bills R, Seehausen O: **Repeated colonization and hybridization characterize the Lake Malawi cichlid fish species flock.** *Curr Biol* 2011, **21:**108–109.
52. Genner MJ, Turner GF: **Ancient hybridization and phenotypic novelty within Lake Malawi's cichlid fish radiation.** *Mol Biol Evol* 2012, **29:**195–206.
53. Nylander JAA: *MrModeltest v2*. Uppsala University: Program distributed by the author. Evolutionary Biology Centre; 2004.
54. Joly S: **JML: testing hybridization from species trees.** *Mol Ecol Res* 2012, **12:**179–184.
55. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A: **Very high resolution interpolated climate surfaces for global land areas.** *Int J Climatol* 2005, **2:**1965–1978.
56. Warren DL, Glor RE, Turelli M: **Environmental niche equivalency *versus* conservatism: quantitative approaches to niche evolution.** *Evolution* 2008, **62:**2868–2883.
57. Phillips SJ, Anderson RP, Schapire RE: **Maximum entropy modelling of species geographic distributions.** *Ecol Model* 2006, **190:**231–259.
58. Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S: **Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data.** *Ecol Appl* 2009, **19:**181–197.
59. Swets K: **Measuring the accuracy of diagnostic systems.** *Science* 1988, **240:**1285–1293.
60. Warren DL, Glor RE, Turelli M: **ENMTools: a toolbox for comparative studies of environmental niche models.** *Ecography* 2010, **33:**607–611.
61. Rabosky DL, Lovette IJ: **Explosive evolutionary radiations: decreasing speciation or increasing extinction through time?** *Evolution* 2008, **62:**1866–1875.
62. Rogers AR, Harpending H: **Population growth makes waves in the distribution of pairwise genetic differences.** *Mol Biol Evol* 1992, **9:**552–569.
63. Salzbuger W, Baric S, Sturmbauer C: **Speciation via introgressive hybridization in East African cichlids?** *Molec Ecol* 2002, **11:**619–625.
64. Grant PR, Grant BR, Petren K: **Hybridization in the recent past.** *Am Nat* 2005, **166:**56–67.
65. Vedenina V: **Variation in complex courtship traits across a hybrid zone between grasshopper species of the *Chorthippus albomarginatus* group.** *Biol J Linn Soc* 2011, **102:**275–291.
66. Usami T, Yokoyama J, Kubota K, Kawata M: **Genital lock-and-key system and premating isolation by mate preference in Carabid beetles (*Carabus* subgenus *Ohomopterus*).** *Biol J Linn Soc* 2006, **87:**145–154.
67. Takahashi K, Terai Y, Nishida M, Okada N: **Phylogenetic relationships and ancient incomplete lineage sorting among Cichlid fishes in Lake Tanganyika as revealed by analysis of the insertion of retroposons.** *Mol Biol Evol* 2001, **18:**2057–2066.
68. Schluter D: *The ecology of adaptive radiation*. Oxford, UK: Oxford University Press; 2000.
69. Nitecki MH: *Evolutionary innovations*. Chicago, IL: The University of Chicago Press; 2000.
70. Farrell BD, Mitter C: **Adaptive radiation in insects and plants: time and opportunity.** *Am Zool* 1994, **34:**57–69.
71. Simpson GG: *The major features of evolution*. New York: Columbia University Press; 1953.
72. Grant PR: *Ecology and evolution of Darwin's Finches*. Princeton, NJ: Princeton University Press; 1984.
73. Balke M, Ribera I, Hendrich L, Miller MA, Sagata K, Posman A, Vogler AP, Meier R: **New Guinea highland origin of a widespread arthropod supertramp.** *Proc R Soc B* 2009, **276:**2359–2367.
74. Ribera I, Vogler AP: **Habitat type as a determinant of species range sizes: the example of lotic-lentic differences in aquatic Coleoptera.** *Biol J Lin Soc* 2000, **71:**33–52.
75. Greenwood PH: **What is a species flock?** In *Evolution of fish species flocks*. Edited by Echelle AA, Kornfield I. Maine: Orono Press; 1994:13–19.
76. Schön I, Martens K: **Adaptive, pre-adaptive and non-adaptive components of radiations in ancient lakes: a review.** *Org Divers Evol* 2004, **4:**137–156.
77. Albrecht C, Trajanovski S, Kuhna K, Streita B, Wilke T: **Rapid evolution of an ancient lake species flock: Freshwater limpets (Gastropoda: Ancylidae) in the Balkan Lake Ohrid.** *Org Divers Evol* 2006, **6:**294–307.
78. Miller MA, Pfeiffer W, Schwartz T: **Creating the CIPRES Science Gateway for inference of large phylogenetic trees.** In *Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010*. New Orleans, Louisiana; 2010:1–8. http://www.ngbw.org/ee/index.php/portal/cite_us.

doi:10.1186/1471-2148-12-142
Cite this article as: Hawlitschek *et al.*: Pleistocene climate change promoted rapid diversification of aquatic invertebrates in Southeast Australia. *BMC Evolutionary Biology* 2012 **12**:142.

7.2 Unveiling the Diversification Dynamics of Australasian Predaceous Diving Beetles in the Cenozoic

Citation: Toussaint EFA, Condamine FL, Hawlitschek O, Watts CHS, Porch N, Hendrich L, Balke M (2015) Unveiling the diversification dynamics of Australasian predaceous diving beetles in the Cenozoic. *Systematic Biology*, 64(1):3-24.

Syst. Biol. 0(0):1–22, 2014

DOI:10.1093/sysbio/syu067

Unveiling the Diversification Dynamics of Australasian Predaceous Diving Beetles in the Cenozoic

EMMANUEL F.A. TOUSSAINT[1,*], FABIEN L. CONDAMINE[2], OLIVER HAWLITSCHEK[1], CHRIS H. WATTS[3], NICK PORCH[4], LARS HENDRICH[1], AND MICHAEL BALKE[1,5]

[1]*SNSB-Bavarian State Collection of Zoology, Münchhausenstraße 21, 81247 Munich, Germany;* [2]*CNRS, UMR 7641 Centre de Mathématiques Appliquées (Ecole Polytechnique), Route de Saclay, 91128 Palaiseau cedex, France;* [3]*South Australian Museum, Adelaide, South Australia, Australia;* [4]*Centre for Integrated Ecology & School of Life and Environmental Sciences, Deakin University, 221 Burwood Highway, Burwood, Victoria 3125, Australia; and* [5]*GeoBioCenter, Ludwig-Maximilians University, Munich, Germany.*

**Correspondence to be sent to: SNSB-Bavarian State Collection of Zoology, Münchhausenstraße 21, 81247 Munich, Germany; E-mail: Toussaint@zsm.mwn.de.*

Received 26 June 2013; reviews returned 18 June 2014; accepted 18 July 2014
Associate Editor: Brian Wiegmann

Abstract.—During the Cenozoic, Australia experienced major climatic shifts that have had dramatic ecological consequences for the modern biota. Mesic tropical ecosystems were progressively restricted to the coasts and replaced by arid-adapted floral and faunal communities. Whilst the role of aridification has been investigated in a wide range of terrestrial lineages, the response of freshwater clades remains poorly investigated. To gain insights into the diversification processes underlying a freshwater radiation, we studied the evolutionary history of the Australasian predaceous diving beetles of the tribe Hydroporini (147 described species). We used an integrative approach including the latest methods in phylogenetics, divergence time estimation, ancestral character state reconstruction, and likelihood-based methods of diversification rate estimation. Phylogenies and dating analyses were reconstructed with molecular data from seven genes (mitochondrial and nuclear) for 117 species (plus 12 outgroups). Robust and well-resolved phylogenies indicate a late Oligocene origin of Australasian Hydroporini. Biogeographic analyses suggest an origin in the East Coast region of Australia, and a dynamic biogeographic scenario implying dispersal events. The group successfully colonized the tropical coastal regions carved by a rampant desertification, and also colonized groundwater ecosystems in Central Australia. Diversification rate analyses suggest that the ongoing aridification of Australia initiated in the Miocene contributed to a major wave of extinctions since the late Pliocene probably attributable to an increasing aridity, range contractions and seasonally disruptions resulting from Quaternary climatic changes. When comparing subterranean and epigean genera, our results show that contrasting mechanisms drove their diversification and therefore current diversity pattern. The Australasian Hydroporini radiation reflects a combination of processes that promoted both diversification, resulting from new ecological opportunities driven by initial aridification, and a subsequent loss of mesic adapted diversity due to increasing aridity. [Australian aridification; diversification; Dytiscidae; Pleistocene extinction; freshwater biota; ground water organisms; Hydroporini.]

Unfolding macroevolutionary processes driving the assemblage of ecological communities across geological timescales is one of the most riveting challenges in biology (Ricklefs 2004). Modern molecular phylogenetic techniques allow the reconstruction of time-calibrated trees to unveil evolutionary radiations in taxonomic groups ranging from insects (Moreau et al. 2006; Hunt et al. 2007; Wiegmann et al. 2011) to tetrapods (Bininda-Emonds et al. 2007; Fritz and Rahbek 2012; Jetz et al. 2012), and from plants (Nagalingum et al. 2011; Soltis et al. 2011; Leslie et al. 2012) to microbial organisms (Morlon et al. 2012). In the last decade, starting from the simple lineages-through-time (LTT) plots (Ricklefs 2007) and the later development of more complex birth–death models to estimate speciation and extinction rates (Stadler 2013), macroevolutionary analyses of phylogenetic trees have made substantial progress to better reveal the tempo and mode of species diversification. A predominant pattern generally inferred in terrestrial radiations is characterized by an early rapid (or even explosive) speciation followed by declining diversification rates (Harmon et al. 2003; McPeek 2008; Phillimore and Price 2008; Rabosky and Lovette 2008; Gavrilets and Losos 2009; Morlon et al. 2012). This common pattern departs from the predictions of constant rate models and is often attributed to the theory of adaptive radiation (Glor 2010). By contrast, radiations of freshwater clades are less well-documented, yet they offer an opportunity to study evolutionary processes because of their ecological specialization, dependency on aquatic resources, often high species richness and relative ease to obtain good species-level coverage even on a continental scale. Very few studies have used macroevolutionary approaches to reveal diversification patterns in freshwater radiations (Day et al. 2013; Morvan et al. 2013). Freshwater clades studied so far exhibit a constant rate of diversification (Day et al. 2013; Morvan et al. 2013). Because too few empirical macroevolutionary case studies have been done to date, this result can be questioned and does not reflect a generality about the tempo and mode of freshwater diversification.

In this context, species-rich and globally distributed clades represent an ideal framework to advance our understanding of the processes governing the dynamics and assembly of biological diversity in freshwater ecosystems over spatio-temporal scales (Derryberry et al. 2011; Condamine et al. 2012; Drummond et al. 2012a). The ongoing development of methods to infer the tempo and mode of diversification at macroevolutionary scales has provided new opportunities to address questions regarding the underlying mechanisms of

geographic range evolution (Ree and Smith 2008; Ronquist and Sanmartín 2011) and among lineage variation in diversification rates (Rabosky 2006; Rabosky and Lovette 2008; FitzJohn et al. 2009; Goldberg et al. 2011; Morlon et al. 2011; Stadler 2011; Etienne et al. 2012). Although those methods are powerful tools, they highly rely on the quality of the taxon sampling (Heath et al. 2008). Ideally, about 80% of species should be included; otherwise diversification models can lead to inaccurate estimates of speciation and extinction rates (Cusimano and Renner 2010; Davis et al. 2013). This is particularly true when a decrease in diversification rates is inferred, a pattern that has been taken as evidence of adaptive radiation (e.g., Harmon et al. 2003; Glor 2010) or diversity-dependent diversification (Phillimore and Price 2008). Assembling such a taxon sampling for freshwater organisms of a species-rich clade would therefore offer a good opportunity to provide reliable estimates of diversification rates and gain insights into the processes shaping contemporary diversity (Day et al. 2013; Morvan et al. 2013).

Here, we study Australasian Hydroporini diving beetles (Coleoptera, Dytiscidae) also known as the "*Necterosoma* group" (Ribera et al. 2008), an endemic radiation of 11 genera distributed across the entire Australian continent and a few neighboring islands such as New Guinea, New Caledonia and Fiji. The 147 extant species occupy various types of lentic and lotic habitats such as well oxygenated rivers (*Barretthydrus*, *Carabhydrus*, *Sekaliporus*), protected embayments and rest pools of slow-flowing streams or standing water pools and saline lakes (*Chostonectes*, *Megaporus*, *Necterosoma*, *Sternopriscus*, *Tiporus*) but also swamps and peatlands (*Antiporus*, the new genus "*Brancuporus*" in description, *Sternopriscus*) (Watts 1997, 2002; Hendrich 2003, 2008; Hendrich and Watts 2009; Hawlitschek et al. 2011, 2012; Hendrich et al. 2014). Although some previous phylogenetic studies have investigated relationships and divergence times (Leys et al. 2003; Leijs et al. 2012), an accurate and comprehensive phylogenetic framework is still lacking to infer historical biogeography, diversification dynamics, and the processes governing this freshwater radiation.

This group is not only relevant to study a continental-scale Australasian radiation but also to investigate patterns and processes of diversification fostering radiations of two ecologically different lineages. The genera, *Paroster* (47 species) and *Sternopriscus* (29 species), are of particular interest because they represent the most species-rich hypogean and epigean Australian genera, respectively. *Paroster* contains species restricted to calcrete aquifers (underground water) of Western and Central Australia as well as epigean species distributed in mesic habitats of the southern part of the range (Watts and Humphrey 1999–2009; Watts et al. 2008; Hendrich and Fery 2008). Most authors aiming at deciphering the origin of hypogean taxa proposed that these lineages might have colonized underground ecosystems in response to climatic change (e.g., Leys et al. 2003; Faille et al. 2010). For Australian diving beetles, the dominant hypothesis has proposed that the *Paroster* radiation is the result of a groundwater colonization following the onset of Miocene aridification at *ca*. 15 million years ago (Ma) (Leys et al. 2003). At this time, epigean populations might have colonized subterranean aquifers to avoid increasing aridity. As a result, the genus harbors morphological features to specialized underground life including wing-loss, depigmentation and eye reduction. On the other hand, the genus *Sternopriscus*, whilst being characterized by elevated diversity, comprises species that do not show clear ecomorphological disparity (Hawlitschek et al. 2012). The genus is so morphologically homogeneous that many species are only revealed using genitalia and male secondary sexual characters. *Sternopriscus* is mostly distributed in Australian mesic areas such as southeastern, southwestern and northern coasts, and species inhabit a wide variety of lentic and lotic habitats from sea level to high altitudes. Such successful colonization is possible due to flight capacity. The high level of endemism in the southeast and southwest suggests that the arid barrier between these two regions is long-standing. Hypotheses on its origin supposed that the dynamics of biogeographical landscapes and the emergence of allo/parapatric barriers were the main drivers of this radiation (Hawlitschek et al. 2012). In particular, the role of complex climatic disruptions during the Pliocene and especially since the early Quaternary might have played an important role in shaping patterns of diversity and distribution in this radiation. The Quaternary climate changes constitute a period of deep modifications in both climatic regimes and ecological conditions in Australia. This period of time has been marked by global cooling and drying, alteration of rainfall seasonality, especially in Southern Australia, vegetation turnover and local emergence of new high altitude niches (Sniderman et al. 2007, 2009, 2013; Byrne et al. 2008). Rainfall seasonality disruptions which principally occurred in the Pleistocene might be one of the underlying mechanisms triggering diversification dynamics of these aquatic carnivorous beetles. Understanding and defining both radiation patterns is critical in order to understand the deterministic forces that shape the diversification and structure of freshwater community assemblies. Using these two clades that have diversified against a common biogeographic and ecological background, a multidisciplinary and integrative approach at the interface between macroevolution and ecological theory will help in the study of the processes of their radiation.

Here, we use a near-complete species-level coverage of the Australasian Hydroporini to: (i) provide for the first time a robust time-calibrated phylogenetic framework of the group combining nuclear and mitochondrial markers; (ii) investigate the potential effect of aridification on the radiation of the group using the latest methods to infer historical biogeography and diversification rates; (iii) contrast the diversification patterns and processes between the radiations of *Paroster* and *Sternopriscus*; and (iv) compare and discuss our results with previous studies regarding the evolution

of Australasian Hydroporini and also regarding the impact of past climate changes on the Australian fauna in general.

Materials and Methods

Taxon Sampling and Molecular Biology

We included 117 of the 147 described species (≈80%) and all genera of Australasian Hydroporini (Online Appendix 1 available on Dryad http://dx.doi.org/10.5061/dryad.c5g23/, Nilsson 2001, 2006; Watts et al. 2008; Watts and Humphrey 2006, 2009; Hendrich 2008; Hendrich and Fery 2008; Hendrich and Watts 2009; Hendrich et al. 2014). Outgroups were 12 species of Bidessini, Hyphydrini, Laccornini, and Vatellini (Appendix 1): the closest tribes to Australasian Hydroporini within Hydroporinae (Ribera et al. 2008). We included several genera for each tribe when possible to improve both phylogenetic resolution and branch length estimation. *Coptotomus* was selected to root the tree since Coptotominae has been found in sister position to the subfamily Hydroporinae (Ribera et al. 2008). Since the genus *Paroster* mainly comprises rare, sometimes monotypic and excessively difficult to sample hypogean species, almost all the sequences used in this study (CO1 and 16S) were recovered from GenBank. In addition to previously sequenced species and in order to improve the placement of the genus in the group, we sequenced mitochondrial and nuclear markers (see below) for four epigean species of *Paroster* diving beetles (Leys and Watts 2008).

Total genomic DNA was extracted from legs, thoracic and head tissues of specimens kept in 96% ethanol using the DNeasy kit (Qiagen, Hilden, Germany). Using standard PCR protocols (Appendix 2) we amplified and then sequenced the following genes: ribosomal 16S (16S, 769 base pairs [bp]), cytochrome oxidase subunit 1 (COI, 741 bp), cytochrome b (CytB, 390 bp), ribosomal 18S (18S, 607 bp), histone 3 (H3, 321 bp), histone 4 (H4, 203 bp), and arginine kinase (ARK, 636 bp). All sequences of the genus *Sternopriscus* were recovered from a recent publication (Hawlitschek et al. 2012.). The DNA sequences were eye corrected under Geneious R6 (Biomatters, http://www.geneious.com/), aligned using Muscle (Edgar 2004) and the reading frames checked under Mesquite 2.75 (http://mesquiteproject.org). The different datasets used to infer phylogenetic relationships were generated under Mesquite. All sequences were deposited in GenBank (accession Nos. HG965576-HG965750).

Phylogeny

We used Bayesian Inference (BI) and Maximum Likelihood (ML) to reconstruct phylogenetic relationships. For each partition (Table 1), the optimal model of substitution was selected under jModeltest 2.1.3 (Darriba et al. 2012) using the Bayesian Information Criterion (BIC). For BI analyses, we used MrBayes 3.2.1 (Ronquist et al. 2012) and partitioning schemes listed in Table 1. Two simultaneous and independent runs consisting of eight Metropolis-coupled Markov chain Monte Carlo (MCMC, one cold and seven incrementally heated) chains and 40 million generations were performed, with a tree sampling every 1000 generations to calculate posterior probabilities (PP). In order to investigate the convergence of the runs we investigated the split frequencies and Effective Sample Size (ESS) of all the parameters, and plotted the log-likelihood of the samples against the number of generations in Tracer 1.5 (http://BEAST.bio.ed.ac.uk/Tracer). A value of ESS>200 was acknowledged as a good indicator of convergence. All the trees that predated the time needed to reach a log-likelihood plateau were discarded as burn-in, and the remaining samples were used to generate a 50% majority rule consensus tree. The best partitioning scheme was selected according to Bayes Factors (B_F) based on average marginal likelihoods of dual runs estimated using Stepping-stone sampling (Xie et al. 2011). $2\times \ln(B_F)$ scores superior to 10 were considered good indicators of a significantly better partitioning scheme over another (Kass and Raftery 1995). The ML analyses were conducted with

Table 1. Partitioning schemes used for the phylogenetic reconstructions

Partitioning scheme	Details
P1 (NoPart) [1]	Unpartitioned dataset
P2 (ByType) [2]	One partition for the coding genes and one partition for the noncoding genes
P3 (ByGenome) [2]	One partition for the mitochondrial genes and one partition for the nuclear genes
P4 (ByThree) [3]	One partition for the mitochondrial coding genes, one partition for the nuclear coding genes and one partition for the noncoding genes
P5 (ByFour) [4]	One partition per codon position for the coding genes and one partition for the noncoding genes
P6 (BySeven) [7]	One partition per codon position for the mitochondrial genes, one partition per codon position for the nuclear genes and one partition for the noncoding genes
P7 (ByGene) [8]	One partition for each gene
P8 (ByEight) [8]	One partition per codon position for the mitochondrial genes, one partition per codon position for the nuclear genes and one partition for each noncoding gene
P9 (BySixteen) [16]	One partition per codon of each coding gene and one partition for the noncoding genes
P10 (ByMax) [17]	One partition per codon of each coding gene and one partition for each noncoding gene

Notes: The number of partition(s) is given in square brackets for each partitioning scheme.

the best partitioning scheme selected in BI using RAxML (Stamatakis 2006). We performed 1000 *Bootstrap* replicates (BS) to investigate the level of support at each node. A calculated PP >0.95 or a BS >70 was considered to indicate strong support for a given clade (Felsenstein 2004).

Divergence Time Estimation

In order to account for the difficulty of estimating divergence times with confidence, we chose to perform three independent sets of analyses using BEAST 1.7.4 (Drummond et al. 2012b). Prior to this, we tested whether or not the fragments contained in the sequence matrix evolve in a clockwise fashion, using PAUP* (Swofford 2003) to calculate the likelihood with and without enforcing a strict molecular clock. A likelihood ratio test (LRT) was carried out in the same software in order to compare both results, and since the molecular clock hypothesis was not statistically supported (P <0.0001), we used a relaxed clock allowing rate variation among lineages as implemented in BEAST.

First, we used the molecular matrix of the COI gene in combination with different divergence rates calculated for Coleoptera lineages (Balke et al. 2009; Papadopoulou et al. 2010; Andújar et al. 2012). Instead of running several independent analyses with the different rates of evolution calculated in these studies, we used an interval encompassing the different rate values and representative of the idiosyncratic variation of divergence rates within Coleoptera (for recent examples see Tänzler et al. 2014 and Toussaint et al. 2014). First, we used the substitution rate (r = 0.0195; r is the substitutions per site per million years per lineage, subs/s/Myr/l) of *Rhantus* diving beetles calculated by Balke et al. (2009) based on the age of Tahiti. Second, we included the rate of evolution (r = 0.0177 subs/s/Myr/l) of several darkling beetle genera (Tenebrionidae) calculated by Papadopoulou et al. (2010) using the biogeographic history of the Aegean archipelago. Finally, we used the rate of evolution (0.0145 subs/s/Myr/l) inferred from a dated phylogeny of the genus *Carabus* (Carabidae) based on multiple geological and fossil evidence (Andújar et al. 2012). The introduced interval (0.0145–0.0195 subs/s/Myr/l) was used to specify a uniform distribution on the *ucld.mean* (Lower = 0.0145, Upper = 0.0195) therefore taking into account the variation of divergence rates across lineages. The root of the tree was constrained with a Uniform distribution (Lower = 0, Upper = 150) so that the age could not be older than 150 Ma; approximately the age of the oldest dytiscid fossil ever discovered (Ponomarenko 1987). The *Substitution Model* was set accordingly to the jMODELTEST result for the COI dataset, and the *Tree Model* was set to a Yule and a birth–death model in different analyses, each analysis consisting of a 50 million generation run sampled every 1000 generations.

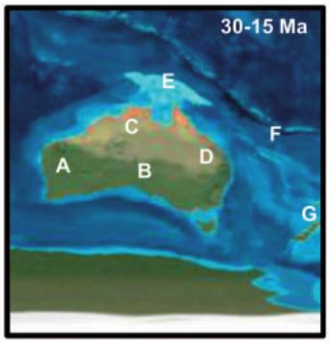
Second, we used the whole molecular dataset (seven genes) and the only unambiguous hydroporine fossil known to calibrate the tree: †*Calicovatellus petrodytes* Miller and Lubkin from the mid-Miocene (see Appendix 3 for more details). Since we included the two extant genera of the tribe Vatellini in the dataset (see Miller 2005 for a revision), but not all the species of the genera, and since the fossil is likely to be sister to the extant genera, we chose in a conservative way to place the calibration point on the stem of the tribe Vatellini based on the "apomorphy-based method" described in Sauquet et al. (2012). The choice of a prior distribution for calibration points is a critical step in dating inference (Ho and Phillips 2009). In order to account for possible biases related to the use of a single calibration point, we carried out different analyses with the Exponential, Lognormal, and Uniform distribution laws as prior for the stem of the tribe Vatellini. We therefore placed a minimum bound on this calibration point with the different priors, so that the 95% confidence interval ranged from 14.8 Ma, the age of the fossil (Woodburne et al. 1990; Miller and Lubkin 2001; towards 150.0 Ma. Parameters for the distribution laws were as follow: Exponential (*Mean* = 36.9, *Offset* = 13.865), Lognormal (Log_{Mean} = 1.44, Log_{Stdev} = 1.7691, *Offset* = 14.669), and *Uniform* (*Lower* = 14.8, *Upper* = 150.0). The tree root was constrained with the same prior as in the first calibration set, and likewise, both Yule and birth–death Tree Models were used in different analyses. The run settings were selected to be the same as the first set of calibration as well, with the best partitioning scheme recovered in phlogenetic analyses.

Third, we carried out a set of analyses based on the molecular matrix of the COI gene and the fossil data to calculate the rate of evolution within the radiation of Australasian Hydroporini. This set of analyses was performed as a cross-validation to check the applicability of the interval of substitution rates (0.0145–0.0195 subs/s/Myr/l) on our data. All settings were the same as the one used to calibrate the MRBAYES topology using the full dataset and the fossil of †*Calicovatellus petrodytes*. We also conducted different analyses using both Yule and birth–death Models.

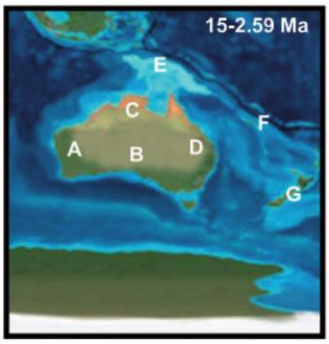
For all analyses, the best BI topology was used to perform dating analyses in order to optimize the search of optimal ages through a minimization of parameter space to explore. The convergence of the runs was investigated using ESS, a conservative burn-in of 25% applied after checking the log-likelihood curves and the different runs merged using LOGCOMBINER 1.7.4 (Drummond et al. 2012b). The maximum credibility tree, median ages and their 95% highest posterior density (HPD) were generated afterwards under TREEANNOTATOR 1.7.4 (Drummond et al. 2012b). The best analysis was selected based on BF estimates derived from marginal likelihoods of the runs using both the fossil and the CO1 dataset.

ANCESTRAL RANGE RECONSTRUCTION

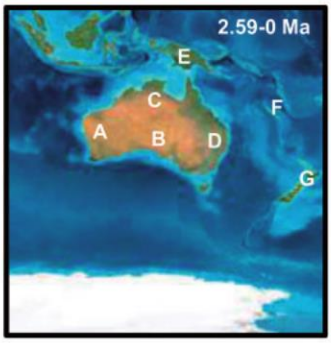
We used the likelihood model Dispersal–Extinction–Cladogenesis (DEC, Ree et al. 2005; Ree and Smith 2008) implemented in LAGRANGE (www.reelab.net/lagrange)



	A	B	C	D	E	F	G
A	-	0.75	0.75	0.5	0.01	0.01	0.01
B		-	0.75	0.75	0.01	0.01	0.01
C			-	0.75	0.01	0.01	0.01
D				-	0.01	0.01	0.01
E					-	0.01	0.01
F						-	0.01
G							-



	A	B	C	D	E	F	G
A	-	0.5	0.25	0.125	0.01	0.01	0.01
B		-	0.5	0.5	0.01	0.01	0.01
C			-	0.5	0.01	0.01	0.01
D				-	0.01	0.01	0.01
E					-	0.01	0.01
F						-	0.01
G							-



	A	B	C	D	E	F	G
A	-	0.25	0.125	0.01	0.01	0.01	0.01
B		-	0.25	0.5	0.01	0.01	0.01
C			-	0.5	0.5	0.01	0.01
D				-	0.125	0.01	0.01
E					-	0.125	0.01
F						-	0.01
G							-

FIGURE 1. Dispersal rate matrices used for biogeographic analyses. Paleogeographical model used for the stratified biogeographical analyses (transitional matrices of dispersal rates between areas) implemented in the Dispersal-Extinction-Cladogenesis model. Maps redrawn from Blakey (2008). A) Pilbara Province + Southwestern Province; B) Paleo Province + Central Australian Province; C) Northern Province + Queensland Province; D) Eastern Province and Tasmanian Province; E, New Guinea; F, New Caledonia + Fiji; G, New Zealand.

to infer ancestral ranges and colonization history of the Australasian members of the tribe Hydroporini. The analyses were carried out based on the different BEAST chronograms with outgroups removed.

The biogeographic regions used in this study were modified from Unmack (2001) to account for present distribution patterns of Australasian Hydroporini (Watts 1997, 2002; Hendrich 2003, 2008; Hendrich and Fery 2008; Hendrich and Watts 2009; Hawlitschek et al. 2011, 2012; and our field notes) (Fig. 1). The seven selected areas yielded a set of $2^7 = 128$ theoretically possible ranges from which we excluded those relying on biological implausibility such as widely disjoint areas. We computed a matrix in which each species was coded as present or absent for each area considered in the analyses (Appendix 1). The choice of temporal constraints and dispersal rates between the discrete distribution ranges was based on paleogeographic and paleoclimatic data (Drexel et al. 1995; Hall 2002, 2011; Hope et al. 2004; Miller 2005; Martin 2006; Byrne et al. 2008, 2011). In order to account for landmass movements and area heterogeneity through time, we defined three time slices spanning the past 30 Ma with the first ranging from 30 Ma to 15 Ma (corresponding to a tropical Australia), the second from 15 Ma to 2.59 Ma (representing the Australian aridification) and the last from 2.59 Ma to present (for the Quaternary climate change) (see Appendix 4 for rationales on the choice of time slices).

Following the principles described in Ree and Smith (2008), we constructed for each time slice a matrix of scaling factors representing dispersal rates between ranges (Fig. 1), and accounting for the geographic position of the areas and their connectivity (biogeographic barriers, ecological corridors). Dispersal rates values were set for each time slice based on paleo-reconstruction evidence (Voris 2000; Hall 2002, 2011; Hope et al. 2004), with long-distance dispersal events

(e.g., Southwestern Australia–New Zealand) authorized with a dispersal rate of 0.01 (in the last time slice, the dispersal rate is fixed to 0.125 instead of 0.01 to account for likely land-connections between New Guinea and Australia).

In order to improve the accuracy and feasibility of the analysis, we enforced all possible combinations of areas at the root and carried out further global likelihood comparisons to select the most likely ancestral area. A difference between potential combinations equal or >2 log-likelihood units was considered significant (Ree et al. 2005; Ree and Smith 2008).

Diversification Analyses

The tempo and mode of diversification were investigated using birth–death methods. We used all chronograms inferred with fossil or rate calibrations for the Australasian Hydroporini, *Paroster* and *Sternopriscus*. Diversification analyses were performed under the R 2.15 software implementing the *ape* ((Paradis et al., 2004)), *diversitree* ((FitzJohn, 2012)), *picante* ((Kembel et al., 2010)), and TreePar ((Stadler, 2011)) packages.

Time-dependent analyses.—First, we used the TreePar approach (Stadler 2011) with the "*bd.shifts.optim*" function allowing the estimation of discrete changes in speciation and extinction rates in possibly incompletely sampled phylogenies. It estimates the maximum likelihood speciation and extinction rates together with the shift times $t = (t_1,t_2,...,t_n)$ in a phylogeny (at the times t, the rates are allowed to change). TreePar analyses were run with: start = 0, end = crown age estimated by dating analyses, grid = 0.1 Myr, sampling fraction = 117/147, 38/47, 27/29 for, respectively, the Australasian Hydroporini, *Paroster*, and *Sternopriscus*, four possible shift times were tested, and posdiv = FALSE to allow the diversification rate to be negative (i.e., allows for periods of declining diversity).

Second, we used the approach of Morlon et al. (2011). Contrary to TreePar, this method has the advantage to take into account the heterogeneity of diversification rates across the tree such that clades may have their own speciation and extinction rates (and their own diversity dynamics), and to estimate continuous variations of rates over time (when discrete in TreePar). Like TreePar, the Morlon et al.'s approach is suitable to potentially infer a declining diversity pattern in which extinction can exceed speciation meaning that diversification rates can be negative (Morlon et al. 2011). We designed four models to be tested: (i) BCSTDCST, speciation and extinction rates are constant; (ii) BVARDCST, speciation rate is exponentially varying and extinction rate is constant; (iii) BCSTDVAR, speciation rate is constant and extinction rate is exponentially varying; and (iv) BVARDVAR, speciation and extinction rates are exponentially varying. The Australasian Hydroporini tree was analyzed as a whole using this approach and missing species were taken into account (stating f = 117/147). However, the heterogeneity of diversification rate can mask the phylogenetic signal of some clades for instance recently diversifying clades with short branch lengths vs. ancient clades that diversified early and then experienced slowdown of diversification (Morlon et al. 2011). To identify different evolutionary diversification among genera, we fitted the MEDUSA approach on a genus-level chronogram of which we informed the species richness of each genus in order to find whether rates varied across the tree (Alfaro et al. 2009). Based on the eventual presence of different rates per genus, we subsequently defined subtrees corresponding to each genus recovered by MEDUSA analyses. After isolating these genera and accounting for the missing species in each, we fitted the same diversification models.

Trait-dependent analyses

We assessed the impact of living in subterranean areas (calcrete aquifers) vs. living in epigean areas (surface habitats) by testing the hypothesis of higher diversification rates due to aridification of Australia in the Miocene and the extensive loss of mesic habitats. We used the Binary State Speciation Extinction model (BiSSE; Maddison et al. [2007]) implemented in the *diversitree* package (FitzJohn 2012). We built the likelihood function using "*make.bisse*," that is then optimized by maximum likelihood using "*find.mle*." Different models were run to test whether speciation, extinction, or transition rates were independent or constrained by the trait. Eight models were built with an increasing complexity starting from the simplest model with no difference in speciation, extinction and transition rates for all character states (three parameters) to the most complex model with speciation, extinction, and transition rates varying independently in each of the character states (six parameters). We also accounted for incomplete taxon sampling (FitzJohn et al. 2009). We estimated posterior density distribution with Bayesian MCMC analyses (10,000 steps) performed with the best-fitting models and the resulting speciation, extinction and dispersal rates.

Diversity-dependent analyses.—We tested the hypothesis that diversity is bounded or at equilibrium meaning that diversity has expanded rapidly in its early stage of diversification and saturates towards the present. We thus explored the effect of diversity on speciation and extinction rates. We used the method of Etienne et al. (2012) implemented in TreePar. The function "*bd.densdep.optim*" was used to fit this model with the following settings for the Australasian Hydroporini, *Paroster* and *Sternopriscus*: discrete = TRUE, the missing species were taken into account (rho = 117/147, 38/47, and 27/29, respectively), an initial carrying capacity (minK) set at 147, 47, and 29, respectively. The final carrying capacity (maxK) was fixed at 1.5× extant diversity of each clade.

RESULTS

Phylogenetic Relationships

The concatenated alignment of 16S, COI, CytB, 18S, H3, H4 and ARK gene fragments comprised 3668 bp for a total of about 37% of missing data in the final matrix and an average gene coverage of 2266 bp. Among the 1287 polymorphic sites, about 30% were contained in the nuclear markers and 70% in the mitochondrial ones. A table highlighting the gene coverage for each genus is available in Appendix 5, and the models of sequence evolution selected by jMODELTEST for each partition are listed in Appendix 6. All phylogenetic analyses using BI and ML converged well except the BI analysis for partitioning scheme *P8* (*ByEight*) that did not reach convergence even after 40 million generations and was discarded from further comparisons. In accordance with the B_F estimates (Table 2), the partitioning scheme *P6* (*BySeven*) was selected, and the resulting topology is presented in Figure 2 along with values of support for the ML analysis conducted with the partitioning scheme *P6*.

The subfamily Hydroporinae and its tribes were recovered monophyletic with strong support both in BI (PP = 1.0) and ML (BS = 100), except for lower support for Hyphydrini (PP = 1.0/BS = 52) and Australasian Hydroporini (PP = 1.0/BS = 84). Within the latter *Chostonectes* was paraphyletic in the BI analysis and monophyletic in the ML analysis due to the inclusion of *Megaporus* with low support in both analyses (PP = 78/BS = 27). A clade containing *Antiporus*, "*Brancuporus*," *Sekaliporus* and *Tiporus* is recovered as sister of *Chostonectes* + *Megaporus* with high support (PP = 1.0/BS = 100). This first large clade is recovered in a moderately to strongly supported sister position (PP = 1.0/BS = 64) to the rest of the genera nested in a second one (PP = 1.0/BS = 53). Within the latter, *Paroster* is the sister of ([*Necterosoma*] +[*Carabhydrus* + {*Barretthydrus* + *Sternopriscus*}]). The ML and BI analyses were highly congruent with almost 90% of the nodes recovered in both methods and except the possible paraphyly of *Chostonectes*, present no major divergences since the only conflicting nodes involve terminal taxa arrangements.

Divergence Time Estimates

The different types of calibration yielded highly similar results (Table 3), and the choice of distribution as a prior for the fossil calibration had a minor impact on the median ages inferred; the largest discrepancies were in the dating of certain outgroup tribes for which median ages are different even though the credibility intervals are broadly overlapping (Table 3). According to B_F calculations (Appendix 7), the best run under both the Yule Model and the Birth Death Model were based on the rate interval introduced in this study. The result of the analysis based on the rate interval and a birth–death model is presented in Figure 3.

The analysis based on the COI matrix and the fossil yielded an estimated mean rate of evolution for this gene ranging from 0.0175 subs/s/Myr/l (Yule model) to 0.0192 subs/s/Myr/l (birth–death model), and estimated ages highly similar with the ones inferred in the two other sets of calibration. Even though Australasian Hydroporini appear to have originated during the mid- to late Oligocene, the branching pattern and the divergence time estimates shown in Figure 3 suggest that most of the extant diversity in Australia is the result of a diversification that initiated during the late Miocene.

Ancestral Range Reconstruction

Figure 4 shows the result from the ancestral range reconstructions yielded by LAGRANGE with the same chronogram as in Figure 3. All analyses, based on the other chronograms gave identical results with significant support for the Eastern and Tasmanian Provinces (D) as the most likely ancestral area for the group (Table 4). Only the analysis comprising the chronogram based on a Lognormal prior and a Yule model supported D with a nonsignificant value.

Diversification Analyses

Based on the chronograms of all BEAST analyses, we reconstructed the corresponding lineages-though-time plots for all the Australasian Hydroporini and the genera *Paroster* and *Sternopriscus* separately (Fig. 5). Table 5

TABLE 2. BEST-fitting strategies of partitioning for the BI analyses with Bayes Factor (B_F) estimates

	Part.	ESS	SSML	*P1*	*P2*	*P3*	*P4*	*P5*	*P6*	*P7*	*P8*	*P9*	*P10*
P1	1	1935	−50847,80	–	*	*	*	*	*	*	NA	*	*
P2	2	1925	−50058,50	**	–	*	*	*	*	*	NA	*	*
P3	2	2601	−49639,95	**	**	–	*	*	*	*	NA	*	*
P4	3	2550	−49128,49	**	**	**	–	**	*	*	NA	*	*
P5	4	1895	−49019,00	**	**	**	*	–	*	*	NA	*	*
P6	7	2167	−47936,38	**	**	**	**	**	–	**	NA	**	**
P7	7	1242	−48697,14	**	**	**	**	**	*	–	NA	*	*
P8	8	118	NA	NA	NA	NA	NA	NA	NA	NA	–	NA	NA
P9	16	915	−48431,59	**	**	**	**	**	*	**	NA	–	*
P10	17	1442	−48143,71	**	**	**	**	**	**	*	NA	**	–

Notes: Part., number of partitions; ESS, Effective Sample Size; SSML, Stepping-Stone Marginal Likelihood; *, 2*ln(B_F)<1 **, 2*ln(B_F)>10, NA, not available.

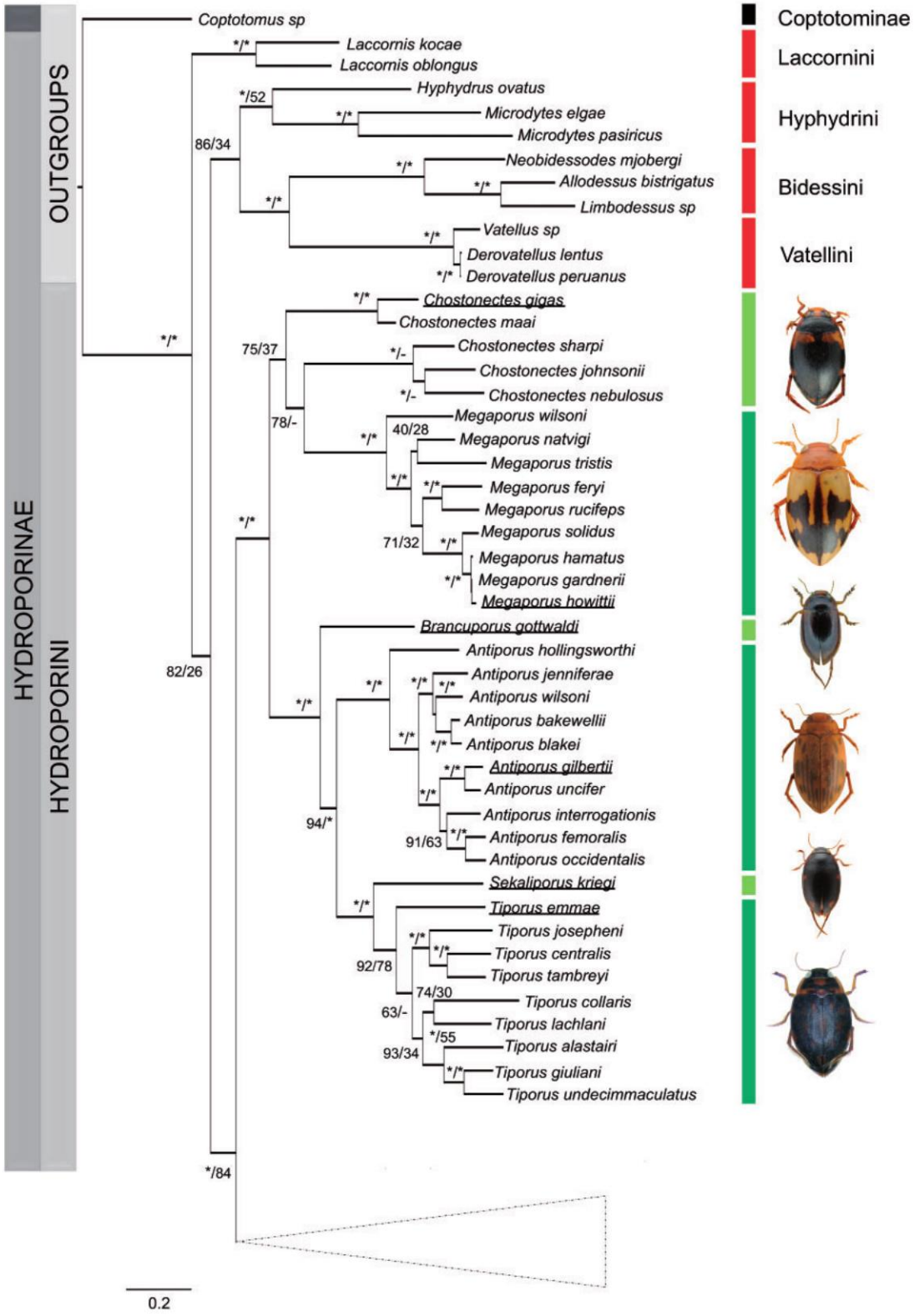


FIGURE 2. Phylogeny of Australasian Hydroporini carnivorous diving beetles. The tree is a 50% majority-rule consensus yielded by the MrBayes analysis based on 16S, 18S, COI, CytB, H3, H4, and ARK. The support of each node is indicated on the topology with respectively the posterior probability (PP) from the BI analysis on the left and the Bootstrap value (BS) from the ML analysis on the right. An asterisk indicates a PP = 1.0 or a BS = 100. The species for which a habitus is displayed have their name underlined.

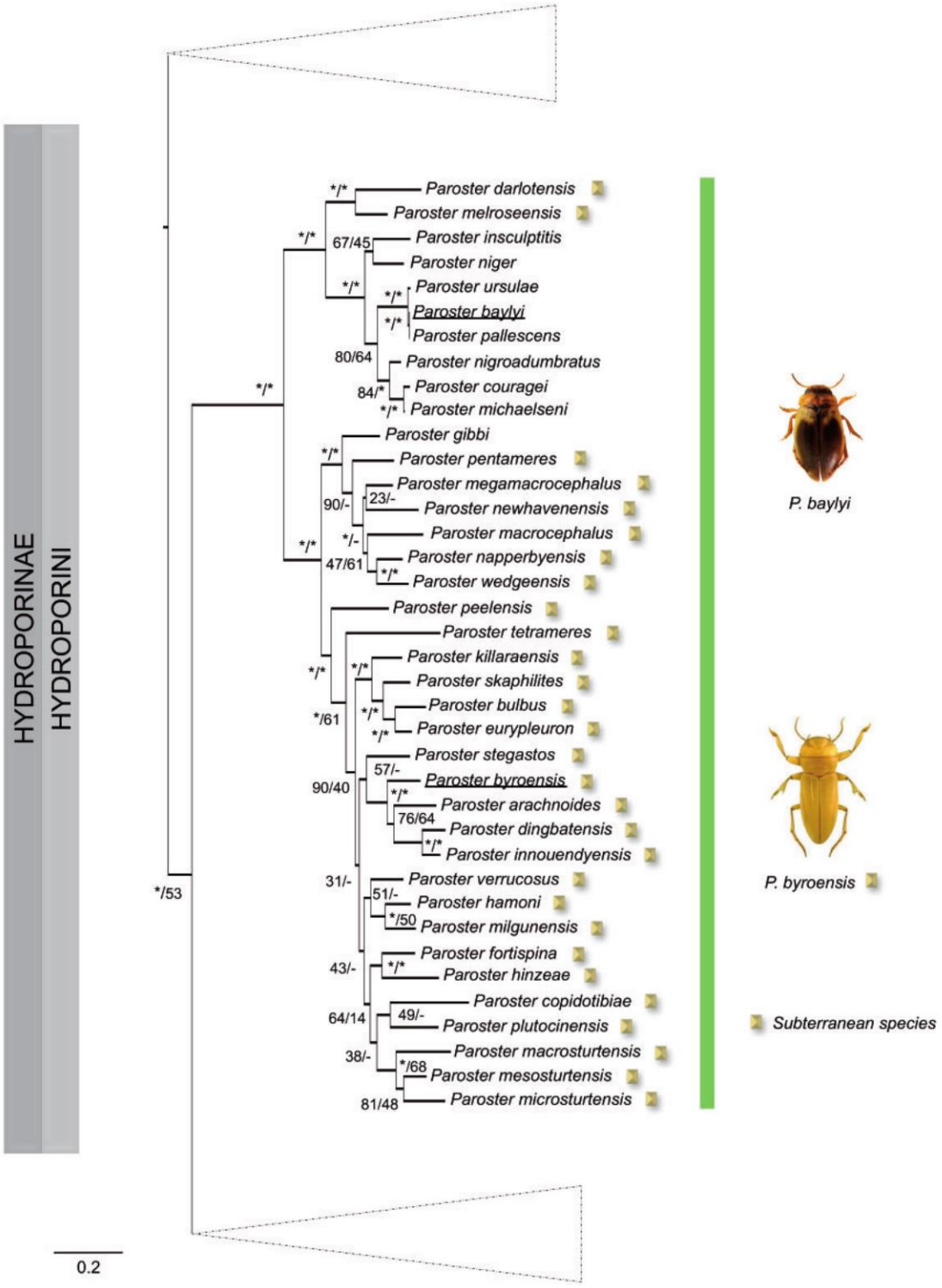


FIGURE 2. Continued

summarizes the main results of the different analyses conducted on the Australasian Hydroporini and both *Paroster* and *Sternopriscus* genera.

Time-dependent analyses.—For Australasian Hydroporini, TreePar analyses supported a diversification model with varying-rates compared to a constant-rate model (Appendix 8). Interestingly all chronograms (except one that has constant rates) show evidence of recent rate shifts. A rapid initial diversification rate is estimated before the first shift time, which generally occurred in the early Pliocene, and is always followed by a

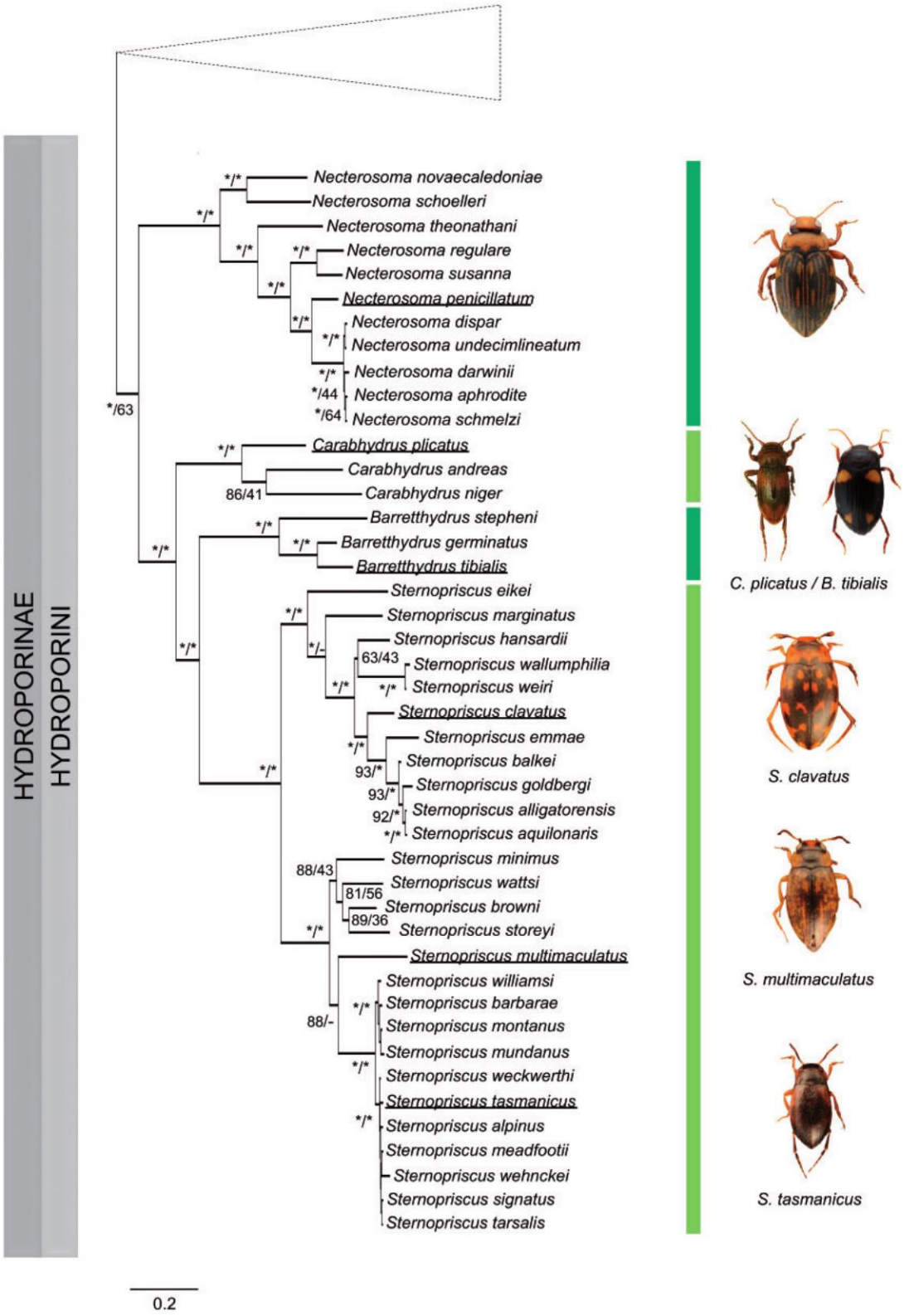


FIGURE 2. Continued

sharp decrease in diversification rate or a negative diversification rate. The second rate shift occurred in the middle-late Pleistocene. At this period, either a decrease in diversification rate or a negative diversification rate is also recovered. The negative diversification rate at present indicates a declining diversity pattern for the group (extinction rate exceeds speciation rate). The last shift time occurred in the late Pleistocene.

TABLE 3. Median ages and 95% credibility intervals for the different analyses

	Rate *Yule*	Rate *BD*	Fossil *Exp Yule*	Fossil *Exp BD*	Fossil *LogN Yule*	Fossil *LogN BD*	Fossil *Uni Yule*	Fossil *Uni BD*
Root	31.7 (24.3–39.7)	31.7 (24.4–40.0)	32.7 (15.9–68.4)	32.0 (15.8–65.7)	31.5 (18.9–51.7)	31.3 (18.4–57.2)	39.7 (19.4–100.8)	41.9 (19.4–110.2)
Hydroporinae	29.8 (23.6–36.8)	29.6 (23.4–36.9)	30.5 (15.3–63.5)	29.8 (15.4–61.2)	28.0 (17.8–44.2)	28.0 (17.3–49.8)	34.6 (17.6–86.5)	35.5 (18.8–98.7)
Laccornini	17.3 (9.9–25.6)	17.3 (9.7–25.5)	14.8 (5.1–28.8)	15.0 (5.1–29.6)	9.5 (1.6–25.8)	7.0 (2.2–20.0)	9.9 (1.3–30.2)	17.6 (2.5–57.4)
Hyphydrini	23.7 (17.6–30.0)	23.6 (17.5–30.0)	21.6 (11.8–40.0)	22.0 (11.4–41.3)	15.8 (5.2–27.5)	14.2 (5.2–26.7)	17.9 (6.9–44.6)	18.1 (6.3–48.7)
Bidessini	16.0 (11.1–21.4)	15.9 (10.7–21.2)	11.7 (5.7–21.5)	12.0 (5.6–22.7)	9.2 (4.3–14.8)	8.7 (4.0–15.3)	11.6 (4.0–29.3)	11.7 (4.1–28.6)
Vatellini	1.0 (0.4–1.7)	1.0 (0.5–1.7)	2.5 (0.9–4.9)	2.5 (0.8–5.1)	2.2 (0.5–7.9)	2.7 (0.6–9.3)	4.0 (0.7–11.6)	3.9 (0.5–12.8)
Australasian Hydroporini	27.6 (21.8–34.0)	27.5 (21.7–34.0)	27.1 (13.3–56.5)	26.4 (13.4–54.5)	24.1 (14.8–38.0)	23.7 (13.3–43.9)	29.4 (15.3–73.9)	32.5 (16.2–83.9)
Chostonectes+Megaporus	22.6 (17.4–28.4)	22.5 (17.4–28.3)	21.2 (9.9–44.3)	20.8 (9.9–43.0)	17.8 (9.7–28.9)	17.2 (9.1–30.7)	20.4 (8.4–51.2)	23.8 (7.7–61.1)
Megaporus	14.4 (10.3–18.7)	14.3 (10.4–18.7)	11.9 (5.1–25.1)	11.6 (5.2–24.5)	9.4 (4.6–10.3)	10.0 (4.7–16.8)	11.5 (4.4–29.5)	13.0 (4.7–36.4)
*Bra+Ant+Sek+Tip**	22.6 (17.4–28.4)	22.5 (17.4–28.1)	20.0 (9.2–41.6)	19.6 (9.4–40.1)	17.2 (10.1–27.5)	17.2 (8.7–30.6)	20.2 (9.6–52.2)	23.6 (10.6–61.1)
Antiporus	17.4 (12.8–22.7)	17.3 (12.6–22.4)	12.3 (5.3–25.8)	12.0 (5.5–25.0)	11.0 (6.3–17.8)	11.5 (5.7–19.7)	13.0 (6.5–33.9)	13.9 (5.3–36.8)
Sekaliporus+Tiporus	19.5 (14.8–24.7)	19.4 (14.6–24.6)	14.7 (6.8–30.9)	14.4 (6.9–30.2)	11.7 (6.7–19.6)	12.2 (5.9–21.9)	14.2 (6.1–35.8)	16.3 (7.0–44.7)
Tiporus	16.8 (12.5–21.5)	16.6 (12.4–21.3)	12.9 (5.9–27.2)	12.6 (5.8–26.3)	9.7 (5.5–16.2)	10.5 (4.6–18.5)	12.1 (5.2–30.3)	13.5 (5.7–37.9)
Paroster	19.8 (14.9–25.3)	19.7 (14.6–25.2)	14.9 (6.9–31.5)	14.5 (6.9–30.2)	16.4 (8.6–26.1)	15.7 (8.1–27.8)	20.0 (7.7–42.5)	20.7 (9.1–56.5)
Paroster N1	11.8 (8.2–15.7)	11.8 (8.1–15.6)	9.1 (4.5–16.8)	9.2 (4.5–17.7)	9.6 (4.3–17.7)	10.2 (4.5–18.6)	13.5 (4.7–35.2)	12.6 (4.5–37.2)
Paroster N2	13.2 (10.1–17.0)	13.1 (9.8–16.7)	11.4 (6.0–20.7)	11.6 (6.2–22.0)	13.2 (7.0–22.4)	12.8 (6.5–23.2)	16.6 (7.7–42.5)	16.1 (7.4–43.9)
Paroster N3	11.5 (7.9–15.3)	11.4 (7.7–15.4)	7.8 (3.8–14.4)	8.0 (3.8–15.2)	8.9 (3.5–16.7)	8.4 (3.2–17.4)	11.1 (4.0–28.5)	10.5 (3.4–30.0)
Paroster N4	12.5 (9.3–15.9)	12.4 (9.3–15.8)	10.5 (5.6–19.2)	10.7 (5.7–20.4)	11.6 (5.7–19.9)	11.0 (5.6–20.7)	14.6 (6.5–38.1)	14.3 (6.1–38.9)
Necterosoma	16.3 (11.9–21.1)	16.2 (11.7–20.9)	13.9 (5.9–29.5)	13.6 (5.8–28.3)	13.2 (5.1–23.3)	12.5 (5.4–22.1)	15.7 (5.5–39.9)	17.1 (5.3–44.1)
Carabhydrus	12.7 (8.0–17.7)	12.6 (8.1–17.6)	11.8 (4.7–25.2)	11.5 (4.6–24.1)	9.5 (3.2–16.7)	7.9 (2.6–17.7)	11.7 (4.4–30.1)	11.4 (2.5–34.2)
Barretthydrus	11.3 (7.3–15.9)	11.3 (7.3–15.8)	9.2 (3.5–19.7)	8.9 (3.3–18.8)	8.1 (3.6–15.0)	7.2 (1.5–13.9)	8.8 (3.0–22.7)	10.2 (2.3–29.3)
Sternopriscus	18.0 (13.7–22.8)	17.9 (13.4–22.6)	12.6 (5.8–26.6)	12.2 (5.7–25.5)	12.2 (7.1–19.8)	10.9 (5.7–20.2)	13.8 (6.2–34.8)	15.4 (7.2–39.4)
STR clade	1.2 (0.8–1.7)	1.2 (0.8–1.7)	1.7 (0.7–3.6)	1.6 (0.6–3.3)	3.4 (1.4–10.2)	2.6 (0.8–5.7)	3.9 (1.2–11.1)	3.8 (0.8–10.6)

Notes: The 95% confidence intervals are given in brackets; *BD, birth–death Tree Model; Exp, Exponential* distribution; *LogN, Lognormal* distribution; *Uni, Uniform* distribution; *Bra+Ant+Sek+Tip**, *Brancuporus+Antiporus+Sekaliporus+Tiporus*; STR clade, *Sternopriscus tarsalis* radiation (Hawlitschek et al. 2012).

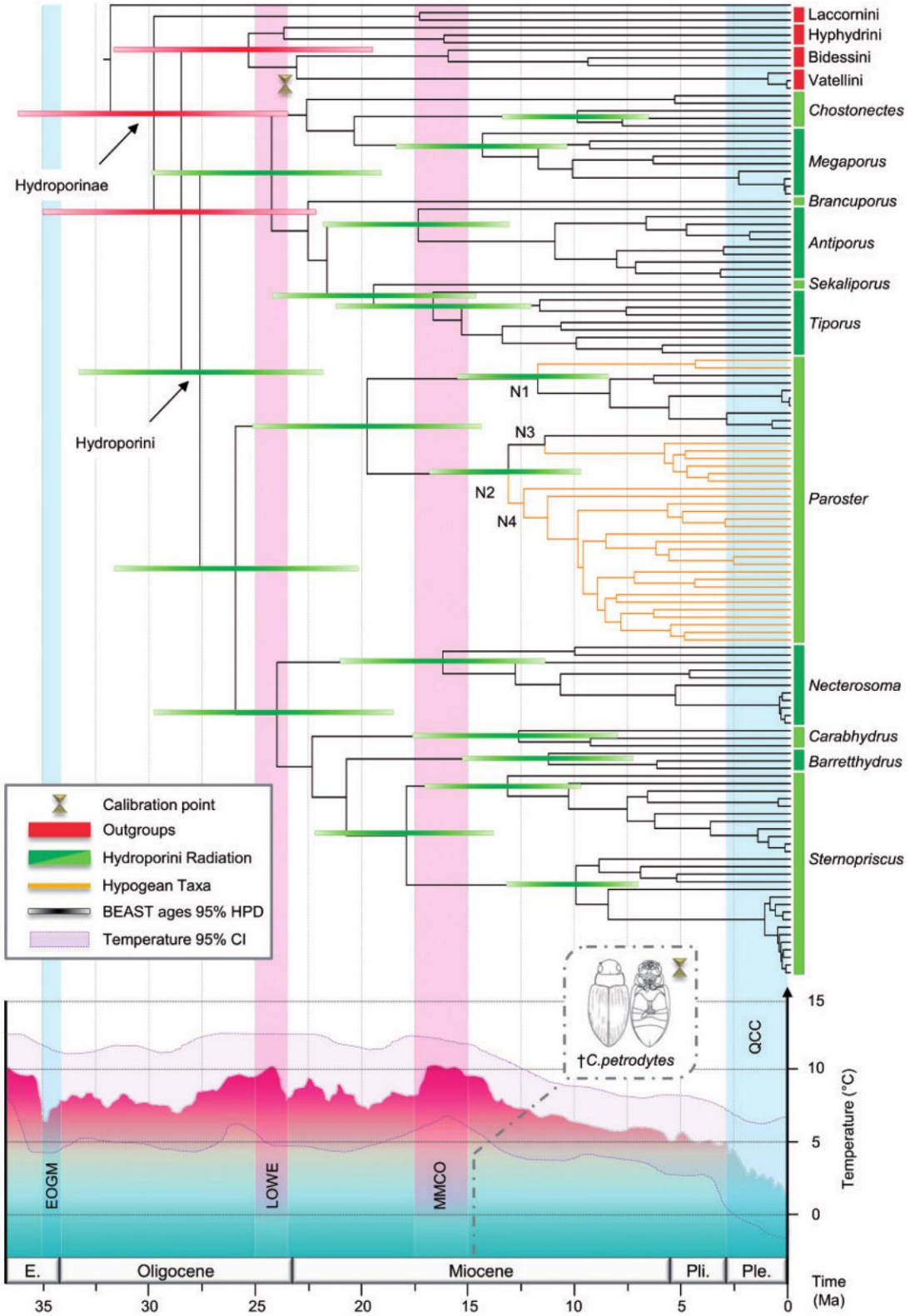


FIGURE 3. Bayesian molecular chronogram of Australian Hydroporini inferred under BEAST. Maximum clade credibility tree from the BEAST analysis. Illustrations on the right highlight the fossil used to calibrate the node indicated by a brownish hourglass. Four vertical bars represent the following major climatic events: Early Oligocene Glacial Maximum (EOGM), Late Oligocene Warming Event (LOWE), Mid-Miocene Climatic Optimum (MMCO) and QCC (Quaternary Climatic Change). A graphic showing the evolution of temperature during the last 37 Ma is presented at the bottom of the figure.

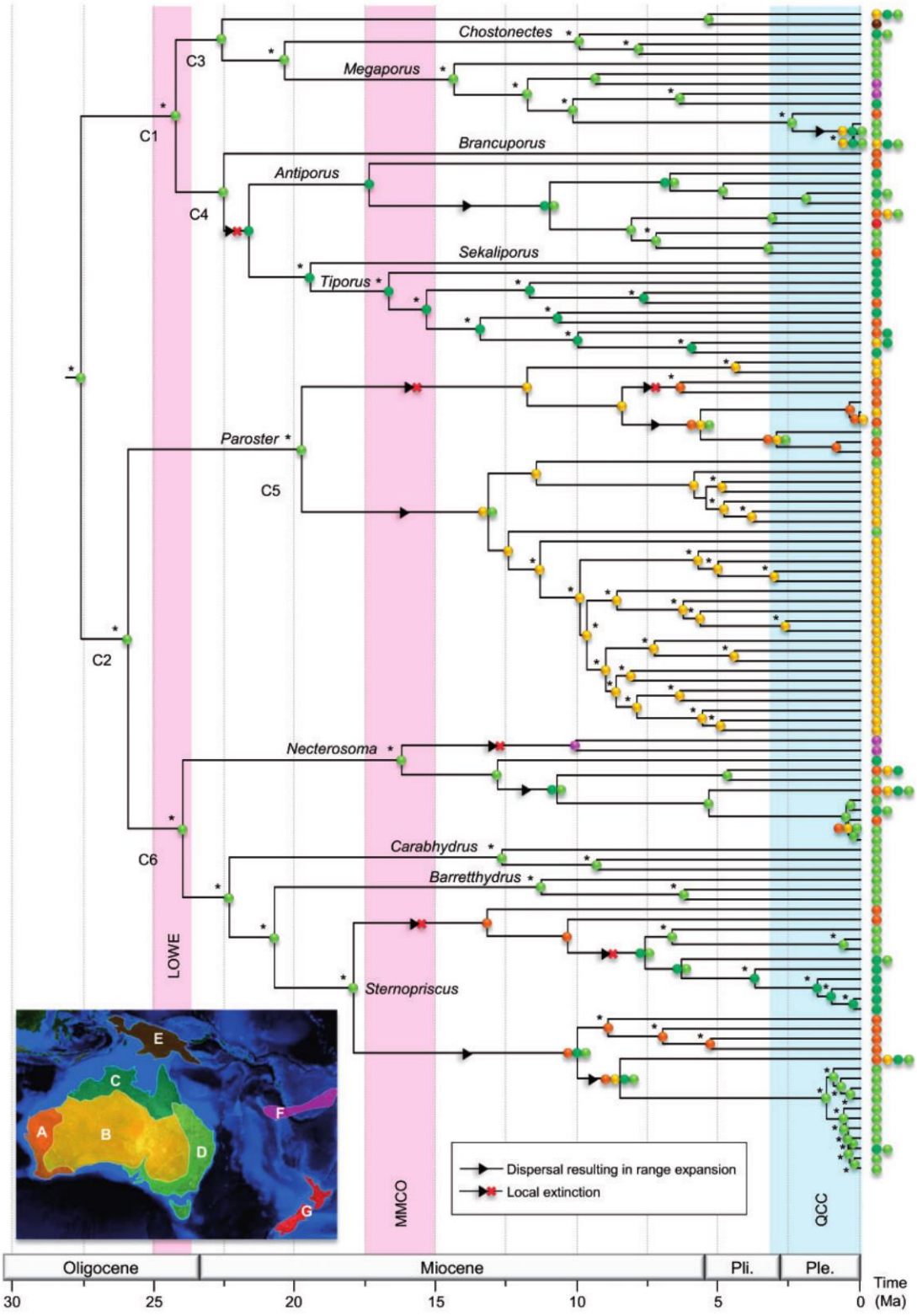


FIGURE 4. Historical biogeography of the Australasian Hydroporini. The bottom-left corner map represents the Australian region delimited in seven biogeographic regions. At the tips of the chronogram, the present-day distribution of the species (Appendix 1) is given. Colored pastilles at each node correspond to the most likely ancestral area recovered by the DEC model, and colored arrows on branches indicate dispersal events. Three vertical bars represent the following major climatic events: Late Oligocene Warming Event (LOWE), Mid-Miocene Climatic Optimum (MMCO) and QCC (Quaternary Climatic Change).

TABLE 4. Best ancestral area at the root using the different BEAST chronograms

Area	Rate *Yule*	Rate *BD*	Fossil *Exp Yule*	Fossil *Exp BD*	Fossil *LogN Yule*	Fossil *LogN BD*	Fossil *Uni Yule*	Fossil *Uni BD*
A	−252.2	−252.1	−253.8	−254.0	−251.9	−251.8	−256.2	−255.9
B	−248.0	−247.7	−249.8	−249.8	−247.3	−247.1	−250.0	−250.2
C	−248.8	−248.8	−248.9	−248.8	−248.9	−248.6	−249.3	−249.1
D	−245.5*	−245.2*	−246.0*	−245.6*	−245.5	−245.0*	−246.6*	−246.6*
E	−261.3	−261.1	−263.2	−263.0	−259.9	−260.1	−265.2	−264.9
F	−278.2	−275.9	−281.2	−281.3	−277.3	−276.0	−283.4	−283.3
G	−275.1	−274.9	−278.2	−279.0	−275.5	−274.4	−280.6	−281.1

Notes: *, significantly better likelihood than the second best ancestral area at the root

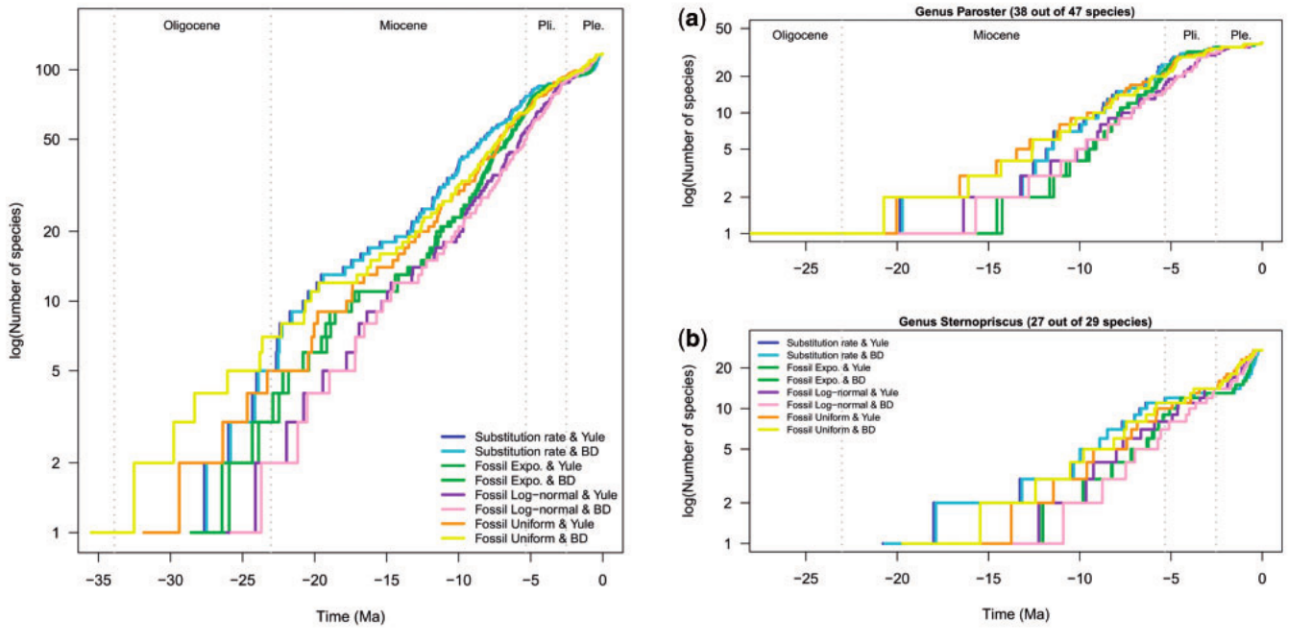


FIGURE 5. Lineages-through-time (LTT) plots for the Australasian Hydroporini and genera *Paroster* (a) and *Sternopriscus* (b). LTT plot on the displays the diversification pattern of all Australasian Hydroporini. The eight BEAST dating analyses are represented each with a different color. Time-scale is indicated spanning the full evolutionary history of the group. On the top, geological periods are indicated (Pli., Pliocene; Ple., Pleistocene).

For all chronograms, the MEDUSA analyses recovered a scenario with no rate change as the best model (Appendix 9), which means that no rate heterogeneity is evidenced among genera within Australasian Hydroporini. Consequently, we analyzed the whole tree with the approach of Morlon et al. (2011). Five out of eight chronograms (calibrated with a substitution rate or with an exponential distribution on the fossil age) were best explained by the BVARDVAR, in which both speciation and extinction rates vary through time (Appendix 10). Specifically, speciation rate increased slightly over time, and extinction rate did not increase until the Pleistocene when extinction exceeded speciation (Appendix 11). The value of the extinction rate at present is very high (around two species lost per lineage per million year, Appendix 10). As a result, a declining diversity scenario is revealed with a maximum of diversity culminating at about 500 species reached in the early Pleistocene (Appendix 12).

To compare diversification processes among the two ecologically-different clades, we analyzed the genera *Paroster* and *Sternopriscus* with both approaches. For *Paroster*, TreePar analyses indicated a scenario with one shift time (Appendix 8). Estimates of diversification rate showed an initial diversification phase with a high diversification rate until the first rate shift occurred in the early mid-Pliocene, and followed by a sharp decrease in diversification rate. With the Morlon et al. (2011)'s approach, the BVARDCST model is recovered for trees calibrated with the substitution rates or with an exponential distribution on the fossil age, and a BCSTDCST for the remaining trees (Appendices 10 and 13). We mostly inferred a damped-increase of species through time of lineages through time (Appendix 14).

For *Sternopriscus*, TreePar analyses provided various results depending on the calibration strategies (Appendix 8). Trees calibrated with the substitution rates or with an exponential distribution on the fossil age gave variable diversification rates with two shift times, whereas trees calibrated with lognormal and uniform distributions gave constant diversification rates. Trees showing shift in rates indicated that the

TABLE 5. Summary of diversification rate inference results

Type of birth–death	Models used	References	Settings	Australasian Hydroporini	*Paroster*	*Sternopriscus*
Time-dependence (rates vary as a function of time and clades)	TreePar (bd.shifts.optim)	Stadler (2011)	5 nested models testing from no rate shift up to 4 rate shifts	7 trees out of 8 show rate shifts in the Plio-Pleistocene (6 have declining diversity, i.e., negative *r*)	All 8 trees show shifts declining diversification rate since the Pliocene	4 trees have constant rates and 4 trees have declining diversification rates in the Pleistocene
	MEDUSA	Alfaro et al. (2009)	1 model estimating rate heterogeneity across clades	No rate shift detected	NA	NA
	Morlon et al.	Morlon et al. (2011)	4 nested models testing whether rates vary or not	5 trees with increase of extinction in the Pliocene, and 3 with constant rates	4 trees with declining speciation rate through time, and 4 with constant rates	6 trees with constant rates, and 2 trees with increase of extinction in the Pliocene
Trait-dependence (rates vary as a function of a character state for a trait)	BiSSE	Maddison et al. (2007); FitzJohn et al. (2009)	8 nested models testing which rates depend or not on the trait	No effect of the trait detected	No effect of the trait detected	NA
Diversity-dependence (rates vary as a function of the number of species)	TreePar (bd.densdep.optim)	Etienne et al. (2012)	1 model estimated the carrying capacity K	Not reached (mean $K = 220$)	Reached at 86% (mean $K = 55$)	Not reached (mean $K = 42$)

Notes: NA, not available.

diversification had an initial period with a medium rate until the first rate shift occurred in the early Pleistocene, and followed by an increase in diversification rate. The second shift time occurred in the middle Pleistocene and followed by a decrease in diversification rate. With the Morlon et al.'s (2011) approach, the BCSTDCST model is generally recovered but for the trees calibrated with substitution rates the BVARDVAR is supported (Appendices 10 and 15). We inferred either a steady accumulation or a recent declining diversity pattern of species through time (Appendix 16).

Trait-dependent analyses.—The BiSSE analyses of diversification suggested that there is no departure from the null model among the six diversification models (Appendix 17). Under these results, there is no difference in speciation, extinction and transition rates in relation to the ecological habitat (epigean vs. subterranean ecology) (Appendix 16). These results are consistent in all chronograms with analyses performed on the whole tree of Australasian Hydroporini and the tree for the genus *Paroster* as well (Appendix 16).

Diversity-dependence analyses.—TreePar analyses indicated that current Australasian Hydroporini diversity is not saturated as the ML estimate is obtained for the maximum carrying capacity at 220 species (Appendix 18). On the contrary, analyses for the genus *Paroster* indicated that the clade is near the equilibrium since the ML estimate gave a carrying capacity at 50 to 60 species, depending on the chronogram (47 species are presently known) (Appendix 18). Analyses of the genus *Sternopriscus* showed that the diversity is not saturated because the ML estimate is obtained for the highest value of the carrying capacity of 44 species (Appendix 18). However, the two rate-calibrated chronograms out of six reached a ML estimate for a carrying capacity equal to the extant number of species.

DISCUSSION

Deciphering biogeographical and diversification processes in the Australasian region has intrigued biologists since Wallace's pioneer works (1860, 1863) and still represents a source of great interest (Lohman et al. 2011 for a review, Carstensen et al. 2012, 2013; de Bruyn et al. 2012, 2013; Müller et al. 2012; Schweizer et al. 2012; Stelbrink et al. 2012; Condamine et al. 2013; Toussaint et al. 2013, 2014; Georges et al. 2014; Tänzler et al. 2014). Yet, the origin and evolution of endemic radiations on continental Australia is far less investigated despite offering a window towards a

better understanding of processes governing diversity dynamics in the region (McGuigan et al. 2000; Bell et al. 2007; Unmack 2010, 2012, 2013; Bowman et al. 2010; Fujita et al. 2010; Byrne et al. 2011; Kayaalp et al. 2013).

Australasian Hydroporini Phylogenetics

The phylogenetic relationships similarly inferred in BI and ML analyses for the Australian Hydroporini are highly congruent with the preliminary work on the family (Dytiscidae) carried out by Ribera et al. (2008). We recover the group monophyletic in all phylogenetic analyses with the highest support as assumed by Leys and Watts (2008) and Ribera et al. (2008). However, the branching pattern shown in Leys and Watts (2008) is incongruent with our inference. For instance, Leys and Watts (2008) recovered the clade (*Necterosoma* + *Paroster*) in a sister-clade to (*Antiporus* + *Brancuporus* + *Chostonectes* + *Megaporus* + *Sekaliporus* + *Tiporus*). This pattern is in contradiction with the one we inferred (Fig. 2), which is instead in agreement with the one recovered by (Ribera et al., 2008): the genus *Paroster* is the sister-clade to the large clade (*Necterosoma*,[*Barretthydrus*,{*Carabhydrus*,(*Sternopriscus*)}]).

Accuracy of Divergence Time Estimation

Dating phylogenies using fossils is a tantalizing concept, yet the methodology and data needed to properly estimate divergence times are paramount despite being sometimes overlooked (Graur and Martin 2004; Near and Sanderson 2004). One could acknowledge four principal sources of error that may lead to fallacious time estimates: (i) improper phylogenetic inference, (ii) misplacement of the fossil in the tree, (iii) wrong dating of the geological strata in which fossils are embedded, and (iv) methodological biases such as inappropriate prior probability distributions, models of sequence evolution and rate heterogeneity among lineages (Drummond et al. 2006; Graur and Martin 2004, Near and Sanderson 2004; Gandolfo et al. 2008; Brandley et al. 2011; Lukoschek et al. 2012).

We investigated some of these potential sources of errors using the most complete molecular dataset ever assembled for a group of Dytiscidae including seven gene fragments, around 80% of the described species, all the extant genera and four tribes representing a large part of the subfamily diversity (Ribera et al. 2008). Molecular dating was based on one of the best preserved specimens of the beetle fossil record that was not preserved in amber (Miller and Lubkin 2001), and we carried out a series of cross-validations to test the robustness of the age estimates using different settings, priors and calibrations (e.g., fossil-based or substitution rate-based dating, Yule model or birth–death process, different distribution laws for the calibrate node prior). Despite the well-known pitfalls of single calibration (Ho and Phillips 2009), our analyses show a close correspondence between the ages of all calibration schemes with overlapping confidence intervals and highly similar median ages whatever the priors or settings we used (Table 3).

All analyses conducted using the BEAST and multiple different priors recovered an Oligocene origin of the group (Fig. 5). In addition, the parallel analysis of cross-validation carried out with the COI dataset and the fossil resulted in rates of evolution congruent with the interval (0.0145–0.0195 subs/s/Myr/l) used to calibrate the tree. Overall, our divergence time estimates are in agreement with the results of Leys et al. (2003) who found a 21.5 Ma-origin for the Australian Hydroporini, an age moderately younger than the median ages we find in this study with different sets of calibration. The ages recovered in Leys et al. (2003, ≈17 Ma) and in Leijs et al. (2012, ≈14 Ma) for the *Paroster* radiation also appear in accordance with our results.

Origins and Biogeography of Australasian Hydroporini

Recent studies addressed the role of Cenozoic climate change on Australian invertebrate diversity and distributions (Sota et al. 2005; Cooper et al. 2011; Hugall and Stanisic 2011; Lucky 2011; Rix and Harvey 2012; Kayaalp et al. 2013), yet empirical studies of the freshwater fauna are either scarce (but see Schultz et al. 2009), focused on fine scales (Ponniah and Hughes 2004; Hawlitschek et al. 2012; Schwentner et al. 2012), or in many cases centered on groundwater-adapted organisms (Cooper et al. 2002, 2007, 2008; Leys et al. 2003; Leys and Watts 2008; Murphy et al. 2009, 2012; Guzik et al. 2012; Leijs et al. 2012). According to our biogeographic reconstructions, the common ancestor of the group most likely originated during the Oligocene in the mesic East coast of Australia whilst luxuriant tropical forests covered the whole region (Martin 2006; Byrne et al. 2008). During the onset of a more arid climate in the early to middle Miocene (Martin 2006; Byrne et al. 2008), two clades derived from the ancestor diversified in the eastern part of the continent and started to colonize the northern area whilst avoiding any westwards colonization (Fig. 4). As the arid zone expanded in central Australia during the middle Miocene around 14 Ma, the ancestors were isolated in mesic and monsoonal coastal ranges of the eastern and northern regions, where these conditions permitted their survival. During this period, a dispersal event occurred toward central Australia and was followed by the origin and diversification of the genus *Paroster* (clade C5, Fig. 4). This radiation likely represents a colonization of groundwater ecosystems which is potentially explained by a combination of modifications of the beetle's ecology via the availability of a new ecological niche, a drying Miocene climate, and a strong morphological adaptation to this new environment (Leys et al. 2003; Leijs et al. 2012).

During this period, the genus *Sternopriscus* (clade nested in C6, Fig. 4) colonized Western Australia (Pilbara) and Southwest Australia twice, a Mediterranean ecosystem with forests and woodlands. A reverse colonization towards the northern and eastern regions occurred during the late Miocene. It is also

during this last timeframe that the colonization of New Caledonian, New Zealand and Fiji archipelagos occurred by long-distance dispersals. Such oversea dispersals have already been highlighted in different diving beetle groups (Balke and Ribera 2004; Monaghan et al. 2006; Balke et al. 2007a, 2007b, 2009; Toussaint et al. 2013). Eventually, several recent dispersals resulting in multiple range expansion especially towards the western regions and the North shaped the extant distribution of the group in Australia. New Guinea and New Zealand therefore appear to have been colonized during the Pleistocene or an even more recent period of time (Balke 1995). The colonization of New Guinea from Eastern or Northern Australia was likely eased by lower sea levels during the Plio-Pleistocene (Voris 2000; Hope et al. 2004; Miller 2005). New Guinea and Australia were connected by a land bridge during this period, whereas today the shallow Torres Strait separates the Australian Cape York Peninsula from Southern New Guinea (Hall 2002, 2011).

Tempo and Mode of Species Diversification

Deciphering an Australasian radiation.—Diversity dynamics in continental Australia have been investigated in previous studies using diversification analyses (e.g., Harmon et al. 2003; Rabosky et al. 2007). However, most of these studies focused on terrestrial vertebrates with few studies investigating the diversification of invertebrates (but see Kayaalp et al. 2013). Here, we provide one of the first empirical studies for Australasian freshwater invertebrate diversity. We found that diversification rates did not remain constant through time (Table 5).

Diversification rates shifted from the early Pliocene (5.332–3.600 Ma) to the middle Pleistocene (0.781–0.126 Ma). Those shifts are associated with negative diversification rates in the last million years resulting in a decline of diversity dynamics of Australasian Hydroporini as recovered in the majority of the analyses (Table 5). This noteworthy pattern is the first empirical evidence for a declining diversity scenario for an invertebrate clade. Previous studies suggested that diversification rate shifted in the middle Miocene as a result of the progressive aridification that began around 15 Ma in Central Australia (Harmon et al. 2003; Rabosky et al. 2007). Although some Australian clades have suffered extinction during periods of environmental changes (Byrne et al. 2011; Sniderman et al. 2013), the pattern for the Australasian Hydroporini provides new insights that such extinction occurred in Australia. The ensuing question is: what can explain this declining diversity scenario? This pattern is attributable to a recent increase in the extinction rate, which exceeds the speciation rate (Table 5). Extinctions might have been fostered by the Quaternary climate change that contributed to increased aridity and perturbed rainfall seasonality in Australia (Sniderman et al. 2007, 2009; Byrne et al. 2008). As a result, freshwater ecosystems may have been quite strongly impacted by the on-going aridification resulting in fewer ecological niches and more geographic contractions.

Contrasting two ecologically different genera.—Variation in diversity dynamics among clades is a famous biological pattern ((Alfaro et al., 2009)). These differences are particularly well illustrated when we look at species richness between sister clades like angiosperms and gymnosperms, or birds and crocodilians. Differences in species richness are often attributable to biological traits or ecological preferences promoting or inhibiting diversification (Rabosky 2009; Wiens 2011). Here, we did not detect significant differences in diversification rates among Australasian Hydroporini genera with MEDUSA (Table 5), but it may be due to the hypothesis of constant-rate over time (Rabosky et al. 2007; Alfaro et al. 2009). Hence, we relaxed this assumption (Morlon et al. 2011) and compared the diversification patterns of the two richest genera that have different ecological features: *Paroster* (47 species), a hypogean-adapted clade that diversified in Central Australia and *Sternopriscus* (29 species), an epigean-adapted clade that diversified in Southeastern and Southwestern Australia (Fig. 4). By applying the same series of diversification analyses, striking differences in diversification processes between the two genera are revealed (Table 5).

Both genera originated in the early Miocene but currently have different species richness, which is explained by different evolutionary scenarios. The genus *Paroster* followed a diversity-dependent pattern characterized by high initial speciation rate that decreased over time. We also estimated a carrying capacity close to the extant species richness meaning that the genus is near equilibrium (Table 5). Given the extent of morphological changes the genus experienced during its evolution, these results are in agreement with the hypothesis of an adaptive radiation of the genus in groundwater ecosystems. Leys et al. (2003) proposed that the ancestor of the genus has colonized groundwater ecosystems as a result of Miocene climate change and has later evolved the morphological traits. The reverse hypothesis states that the clade first evolved the traits that allowed it to colonize the groundwater. Hence, the latter would have conferred a higher speciation rate or a lower extinction rate to this clade. When testing for this, we did not detect any difference in speciation or extinction rates when we applied the BiSSE method. This means that the trait "living in groundwater ecosystem" is not supported as a main driver of diversification (other biological traits may have more contributed). Thus our results support the climatic opportunity in the Miocene (Leys et al. 2003) that would have fostered the diversification of the group. On the contrary, the genus *Sternopriscus* followed a variable-rate diversification featured by recent shifts in diversification rates. The genus has not yet reached its carrying capacity and is still expanding with slowing diversification rate, which is in line with the "damped increase hypothesis"

(Table 5, Morlon et al. 2010; Cornell 2013). These slower speciation rates recovered by our analyses might be explained by recent evidence showing that southern Australia underwent a major alteration of rainfall seasonality in the Pleistocene, after 1.5 Ma. Sniderman et al. (2009) suggested that early Pleistocene Australia had a dominant summer rainfall system that supported (in upland areas) rainforest and wet sclerophyll forest that was very different to that which occurs today under an altered winter rainfall regime. We hypothesize that these deep macroecological disruptions of seasonality have played a major role in shaping slower speciation rates because many *Sternopriscus* diving beetles are restricted to seasonal aquatic habitats.

Including sufficient species-level sampling within clades is critical for diversification analyses. Our coverage within both examined genera (81% and 93% of described species for *Paroster* and *Sternopriscus*, respectively) falls above the threshold advocated by recent studies (Cusimano and Renner 2010; Davis et al. 2013). Furthermore, our analyses accounted for missing taxon sampling using taxonomic knowledge about the extant diversity of clades. The completeness of DNA matrices is another issue that may bias the inference of diversification patterns. In our study, the overall gene coverage within the group was sufficient for a good phylogenetic resolution and consistent Bayesian-relaxed-clock estimates. In the case of *Paroster*, however, the unknown extant diversity of subterranean species (Guzik et al. 2011) combined with a reduced nuclear dataset compared to other included genera may underestimate the diversification rates, though we believe this is unlikely to change the overall pattern obtained. The results recovered here highlight important differences in diversity dynamics for closely related genera that evolved in different freshwater ecosystems in Australia. Both ecosystems spurred different diversification rates but the groundwater ecosystem fostered an adaptive radiation with a diversity-dependent pattern whereas epigean ecosystems shaped a damped increase diversity pattern.

Conclusion

Using an integrative approach combining DNA sequences, distributional data, the fossil record, and latest methods in phylogenetics and likelihood-based approaches to diversification, we investigated the origin and evolution of Australasian Hydroporini diving beetles. We suggest that the aridification of Australia initiated in the Miocene might have played a cardinal role in shaping the extraordinary radiation of Australasian Hydroporini resulting in a striking diversity of species richness and types of colonized habitats, but also that later climatic adjustment, especially in seasonality of rainfall, contributed to a major wave of recent Pleistocene extinctions. Despite an astonishing adaptation to climate changes that occurred in the past million years, the group presents a boom-then-bust pattern of diversity dynamics, with a declining trajectory of diversity likely shaped by the on-going and increasing desertification that is occurring in Australia, which restricts the availability of suitable habitats and the likelihood of short-range dispersal events. As a result, this Australasian radiation has been influenced by climatic shifts and in particular the Quaternary climatic changes that likely opened new ecological opportunities, and it appears that whilst the climate continued to warm in the region, these beetles may have been the victims of the changes that once led to them thriving.

Supplementary Material

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.c5g23/ and TreeBASE data repository at http://purl.org/phylo/treebase/phylows/study/TB2:S16016.

Funding

This work, as well as ongoing research on the Australian predaceous water beetle fauna was supported by a grant from the German Science Foundation (DFG) grants BA2152/6-1, 7-1 and 11-1. FLC was funded by the CNRS and an ANR grant ECOEVOBIO-CHEX2011 awarded to his advisor, Hélène Morlon.

Acknowledgments

We want to thank the editor-in-chief Frank Anderson, associate editor Brian Wiegmann as well as two anonymous reviewers, Gael J. Kergoat and Ignacio Ribera for fruitful comments on previous versions of this manuscript. FLC wishes to thank Hélène Morlon for supervision and advice. The Department of Environment and Conservation in Western Australia is acknowledged for giving permission to conduct scientific research in Western Australia (Permit numbers: SF 003017 and NE 002348), and the Parks and Wildlife Commission of the Northern Territory is thanked for giving permission to conduct scientific research in the Northern Territory (Permit Number: 23929 and RK-400/ RK- 660). The Department of Environment and Conservation in New South Wales (Scientific License No. S12040), the Department of Sustainability and Environment in Victoria [Research Permit No. 10003840], and the Department of Primary Industries and Water on Tasmania [Permit No. FA 06487] are acknowledged for giving permission to conduct scientific research in National and State Parks.

References

Alfaro M.E., Santini F., Brock C., Alamillo H., Dornburg A., Rabosky D.L., Carnevale G., Harmon L.J. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proc. Natl. Acad. Sci. USA. 106(32):13410–13414.

Andújar C., Serrano J., Gómez-Zurita J. 2012. Winding up the molecular clock in the genus *Carabus* (Coleoptera: Carabidae): assessment of

methodological decisions on rate and node age estimation. BMC Evol. Biol. 12(1):40.
Balke M. 1995. The Hydroporini (Coleoptera: Dytiscidae: Hydroporinae) of New Guinea: systematics, distribution and origin of the fauna. Invertebr. Taxon. 9(5):1009–1019.
Balke M., Ribera I. 2004. Jumping across Wallace's line: *Allodessus* Guignot and *Limbodessus* Guignot revisited (Coleoptera: Dytiscidae, Bidessini) based on molecular-phylogenetic and morphological data. Aust. J. Entomol. 43(2):114–128.
Balke M., Pons J., Ribera I., Sagata K., Vogler A.P. 2007a. Infrequent and unidirectional colonization of hyperdiverse *Papuadytes* diving beetles in New Caledonia and New Guinea. Mol. Phylogenet. Evol. 42(2):505–516.
Balke M., Wewalka G., Alarie Y., Ribera I. 2007b. Molecular phylogeny of Pacific Island Colymbetinae: radiation of New Caledonian and Fijian species (Coleoptera, Dytiscidae). Zool. Scr. 36(2):173–200.
Balke M., Ribera I., Hendrich L., Miller M.A., Sagata K., Posman A., Vogler A.P., Meier R. 2009. New Guinea highland origin of a widespread arthropod supertramp. Proc. Roy. Soc. B. 276(1666):2359–2367.
Bell K.L., Moritz C., Moussalli A., Yeates D.K. 2007. Comparative phylogeography and speciation of dung beetles from the Australian Wet Tropics rainforest. Mol. Ecol. 16(23):4984–4998.
Bininda-Emonds O.R.P., Cardillo M., Jones K.E., MacPhee R.D.E., Beck R.M.D., Grenyer R., Price S.A., Vos R.A., Gittleman J.L., Purvis A. 2007. The delayed rise of present-day mammals. Nature. 446(7135):507–512.
Bowman D.M.J.S., Brown G.K., Braby M.F., Brown J.R., Cook L.G., Crisp M.D., Ford F., Haberle S., Hughes J., Isagi Y., Joseph L., McBride J., Nelson G., Ladiges P. Y. 2010. Biogeography of the Australian monsoon tropics. J. Biogeogr. 37(2):201–216.
Brandley M.C., Wang Y., Guo X., de Oca A.N.M., Fería-Ortíz M., Hikida T., Ota H. 2011. Accommodating heterogenous rates of evolution in molecular divergence dating methods: an example using intercontinental dispersal of *Plestiodon* (*Eumeces*) lizards. Syst. Biol. 60(1):3–15.
Byrne M., Yeates D.K., Joseph L., Kearney M., Bowler J., Williams M.A.J., Cooper S., Donnellan S.C., Keogh J.S., Leys R., Melville J., Murphy D.J., Porch N., Wyrwoll K.-H. 2008. Birth of a biome: insights into the assembly and maintenance of the Australian arid zone biota. Mol. Ecol. 17(20):4398–4417.
Byrne M., Steane D.A., Joseph L., Yeates D.K., Jordan G.J., Crayn D., Aplin K., Cantrill D.J., Cook L.G., Crisp M.D., Keogh J.S., Melville J., Moritz C., Porch N., Sniderman J.M.K., Sunnucks P., Weston P.H. 2011. Decline of a biome: evolution, contraction, fragmentation, extinction and invasion of the Australian mesic zone biota. J. Biogeogr. 38(9):1635–1656.
Carstensen D.W., Dalsgaard B., Svenning J.C., Rahbek C., Fjeldså J., Sutherland W.J., Olesen J.M. 2012. Biogeographical modules and island roles: a comparison of Wallacea and the West Indies. J. Biogeogr. 39(4):739–749.
Carstensen D.W., Dalsgaard B., Svenning J.C., Rahbek C., Fjeldså J., Sutherland W.J., Olesen J.M. 2013. The functional biogeography of species: biogeographical species roles of birds in Wallacea and the West Indies. Ecography 36(10):1097–1105.
Condamine F.L., Sperling F.A., Wahlberg N., Rasplus J.Y., Kergoat G.J. 2012. What causes latitudinal gradients in species diversity? Evolutionary processes and ecological constraints on swallowtail biodiversity. Ecol. Lett. 15(3):267–277.
Condamine F.L., Toussaint E.F.A., Cotton A.M., Genson G.S., Sperling F.A.H., Kergoat G.J. 2013. Fine-scale biogeographical and temporal diversification processes of peacock swallowtails (*Papilio* subgenus *Achillides*) in the Indo-Australian Archipelago. Cladistics. 29(1):88–111.
Cooper S.J.B., Hinze S., Leys R., Watts C.H.S., Humphreys W.F. 2002. Islands under the desert: molecular systematics and evolutionary origins of stygobitic water beetles (Coleoptera: Dytiscidae) from central Western Austalia. Invertebr. Syst. 16(4): 589–598.
Cooper S.J.B., Bradbury J.H., Saint K.M., Leys R., Austin A.D., Humphreys W.F. 2007. Subterranean archipelago in the Australian arid zone: mitochondrial DNA phylogeography of amphipods from central Western Australia. Mol. Ecol. 16(7):1533–1544.
Cooper S.J.B., Saint K.M., Taiti S., Austin A.D., Humphreys W.F. 2008. Subterranean archipelago: mitochondrial DNA phylogeography of stygobitic isopods (Oniscidea: *Haloniscus*) from the Yilgarn region of Western Australia. Invertebr. Syst. 22(2):195–203.
Cooper S.J.B., Harvey M.S., Saint K.M., Main B.Y. 2011. Deep phylogeographic structuring of populations of the trapdoor spider *Moggridgea* tingle (Migidae) from southwestern Australia: evidence for long-term refugia within refugia. Mol. Ecol. 20(15):3219–3236.
Cornell H.V. 2013. Is regional species diversity bounded or unbounded? Biol. Rev. 88(1):140–165.
Cusimano N., Renner S.S. 2010. Slowdowns in diversification rates from real phylogenies may not be real. Syst. Biol. 59(4):458–464.
Darriba D., Taboada G.L., Doallo R., Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9(8):772.
Davis M.P., Midford P.E., Maddison W. 2013. Exploring power and parameter estimation of the BiSSE method for analyzing species diversification. BMC Evol. Biol. 13(1):38.
Day J.J., Peart C.R. Brown, K.J. Friel, J.P., Bills R., Moritz T. 2013. Continental Diversification of an African Catfish Radiation (Mochokidae: Synodontis). Syst. Biol. 62(3):351–365.
de Bruyn M., von Rintelen T., von Rintelen K., Mather P.B., Carvalho G.R. 2012. Molecular biogeography and phylogeography of the freshwater fauna of the Indo-Australian Archipelago. In: Gower D.J., Johnson K., Richardson J., Rosen B., Rüber L., Williams S., editors. Biotic evolution and environmental change in Southeast Asia. Cambridge: Syst. Assoc. Vol., Cambridge University Press.
de Bruyn M., Rüber L., Nylinder S., Stelbrink B., Lovejoy N.R., Lavoué S., Tan H.H., Nugroho E., Wowor D., Ng P.K.L., Azizah M.N.S., Von Rintelen T., Hall R., Carvalho G.R. 2013. Paleo-drainage basin connectivity predicts evolutionary relationships across three Southeast Asian biodiversity hotspots. Syst. Biol. 62(3):398–410.
Derryberry E.P., Claramunt S., Derryberry G., Chesser R.T., Cracraft J., Aleixo A., Pérez-Emán J., Remsen J.V., Brumfield, R.T. 2011. Lineage diversification and morphological evolution in a large-scale continental radiation: the neotropical ovenbirds and woodcreepers (Aves: Furnariidae). Evolution 65(10):2973–2986.
Drexel J.F., Preiss W.V., Parker A.J. 1995. The Geology of South Australia: The Phanerozoic (Vol. 54). Mines and Energy, South Australia, Geological Survey of South Australia.
Drummond A.J., Ho S.Y., Phillips M.J., Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4(5):e88.
Drummond C.S., Eastwood R.J., Miotto S.T., Hughes C.E. 2012a. Multiple continental radiations and correlates of diversification in *Lupinus* (Leguminosae): testing for key innovation with incomplete taxon sampling. Syst. Biol. 61(3):443–460.
Drummond A.J., Suchard M.A., Xie D., Rambaut A. 2012b. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29(8):1969–1973.
Edgar R.C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5): 1792–1797.
Etienne R.S., Haegeman B., Stadler T., Aze T., Pearson P.N., Purvis A., Phillimore A.B. 2012. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. Proc. Roy. Soc. B. 279(1732):1300–1309.
Faille A., Ribera I., Deharveng L., Bourdeau C., Garnery L., Queinnec E., Deuve T. 2010. A molecular phylogeny shows the single origin of the Pyrenean subterranean Trechini ground beetles (Coleoptera: Carabidae). Mol. Phylogenet. Evol. 54(1):97–105.
Felsenstein J. 2004. Inferring phylogenies. Sinauer Associates, Sunderland, Massachusetts.
FitzJohn R.G., Maddison W.P., Otto S.P. 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. Syst. Biol. 58(6):595–611.
FitzJohn, R. G. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. Meth. Ecol. Evol. 3(6):1084–1092.
Fritz S.A., Rahbek C. 2012. Global patterns of amphibian phylogenetic diversity. J. Biogeogr. 39(8):1373–1382.
Fujita M.K., McGuire J.A., Donnellan S.C., Moritz C. 2010. Diversification and persistence at the arid–monsoonal interface: Australia-wide biogeography of the Bynoe's gecko (*Heteronotia binoei*; Gekkonidae). Evolution 64(8):2293–2314.

Gandolfo M.A., Nixon K.C., Crepet W.L. 2008. Selection of fossils for calibration of molecular dating models. Ann. Mo. Bot. Gard. 95(1):34–42.
Gavrilets S., Losos J.B. 2009. Adaptive radiation: contrasting theory with data. Science. 323(5915):732–737.
Georges A., Zhang X., Unmack P., Reid B. N., Le M., McCord W. P. 2014. Contemporary genetic structure of an endemic freshwater turtle reflects Miocene orogenesis of New Guinea. Biol. J. Linn. Soc. 111(1):192–208.
Goldberg E.E., Lancaster L.T., Ree R.H. 2011. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. Syst. Biol. 60(4):451–465.
Glor R.E. 2010. Phylogenetic insights on adaptive radiation. Annu. Rev. Ecol. Evol. Syst. 41:251–270.
Graur D., Martin W. 2004. Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. Trends Genet. 20(2):80–86.
Guzik M.T., Austin A.D., Cooper S.J.B., Harvey M.S., Humphreys W.F., Bradford T., Eberhard S.M., King R.A., Leys R., Muirhead K.A., Tomlinson M. 2011. Is the Australian subterranean fauna uniquely diverse? Invert. Syst. 24(5):407–418.
Guzik M.T., Adams M.A., Murphy N.P., Cooper S.J.B., Austin A.D. 2012. Desert Springs: Deep Phylogeographic Structure in an Ancient Endemic Crustacean (*Phreatomerus latipes*). PLoS One 7(7): e37642.
Hall R. 2002. Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific: computer-based reconstructions, model and animations. J. Asian Earth Sci. 20(4):353–431.
Hall R. 2011. Australia–SE Asia collision: plate tectonics and crustal flow. Geol. Soc. London Spec. Publ. 355(1):75–109.
Harmon L. J., Schulte II J.A., Larson A., Losos J.B. 2003. Tempo and mode of evolutionary radiation in iguanian lizards. Science. 301(5635):961–964.
Hawlitschek O., Porch N., Hendrich L., Balke M. 2011. Ecological niche modeling and nDNA sequencing support a new, morphologically cryptic beetle species unveiled by DNA barcoding. PLoS One. 6(2):e16662.
Hawlitschek O., Hendrich L., Espeland M., Toussaint E.F.A., Genner M.J., Balke M. 2012. Pleistocene climate change promoted rapid diversification of aquatic invertebrates in Southeast Australia. BMC Evol. Biol. 12(1):142.
Heath T.A., Zwickl D.J., Kim J., Hillis D.M. 2008. Taxon sampling affects inferences of macroevolutionary processes from phylogenetic trees. Syst. Biol. 57(1):160–166.
Hendrich L. 2003. A new species of *Necterosoma* SHARP from the Kimberley region in north-western Australia (Coleoptera: Dytiscidae). Entomologische Zeitschrift. 113:152–154, figures on pp. 144–145.
Hendrich L. 2008. *Tiporus emmae* sp.n. from Northern Australia (Coleoptera: Dytiscidae, Hydroporinae). Linzer biologische Beiträge. 40(1):311–320.
Hendrich L., Fery H. 2008. *Paroster baylyi* sp. n., *P. ursulae* sp. n. (Col. Dytiscidae, Hydroporinae) and the water beetle diversity of pan-gnammas on isolated granite outcrops in the Mallee of south-western Australia. Zootaxa. 1704:27–41.
Hendrich L., Watts C.H.S. 2009. Taxonomic revision of the Australian predaceous water beetle genus *Carabhydrus* WATTS, 1978 (Coleoptera: Dytiscidae, Hydroporinae). Zootaxa. 2048:1–30.
Hendrich L., Toussaint E.F.A., Balke M. 2014. A new genus of Hydroporini from south-western Australia (Coleoptera, Dytiscidae). *Spixiana*. 37(1):103–109.
Ho S.Y.W., Phillips M.J. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Syst. Biol. 58(3):367–380.
Hope G., Kershaw A.P., Kaars S.V.D., Xiangjun S., Liew P.M., Heusser L.E., Takahara H., McGlone M., Miyoshi M., Moss P.T. 2004. History of vegetation and habitat change in the Austral-Asian region. Quatern. Int. 118:103–126.
Hugall A.F., Stanisic J., 2011. Beyond the prolegomenon: a molecular phylogeny of the Australian camaenid land snail radiation. Zool. J. Linnean Soc. 161(3):531–572.
Hunt T., Bergsten J., Levkanicova Z., Papadopoulou A., John O.S., Wild R., Hammond P.M., Ahrens D., Balke M., Caterino M.S., Gómez-Zurita J., Ribera I., Barraclough T.G., Bocakova M., Bocak L., Vogler A.P. 2007. A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. Science. 318(5858): 1913–1916.
Jetz W., Thomas G.H., Joy J.B., Hartmann K., Mooers A.O. 2012. The global diversity of birds in space and time. Nature. 491(7424): 444–448.
Kass R., Raftery A. 1995. Bayes factors. J. Am. Statist. 90(430):773–795.
Kayaalp P., Schwarz M.P., Stevens M.I. 2013. Rapid diversification in Australia and two dispersals out of Australia in the globally distributed bee genus, *Hylaeus* (Colletidae: Hylaeinae). Mol. Phylogenet. Evol. 66(3):668–678.
Kembel S.W., Cowan P.D., Helmus M.R., Cornwell W.K., Morlon H, Ackerly D.D., Blomberg S.P., Webb C.O. 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics. 26(11): 1463–1464.
Leijs R., van Nes E.H., Watts C.H.S., Cooper S.J.B., Humphreys W.F., Hogendoorn K. 2012. Evolution of blind beetles in isolated aquifers: a test of alternative modes of speciation PLOS One 7(3):e34260.
Leslie A.B., Beaulieu J.M., Rai H.S., Crane P.R., Donoghue M.J., Mathews S. 2012. Hemisphere-scale differences in conifer evolutionary dynamics. Proc. Natl. Acad. Sci. USA. 109(40): 16217–16221.
Leys R., Watts C.H.S., Cooper S.J.B., Humphreys W.F. 2003. Evolution of subterranean diving beetles (Coleoptera: Dytiscidae: Hydroporini, Bidessini) in the arid zone of Australia. Evolution. 57(12):2819–2834.
Leys R., Watts C.H.S. 2008. Systematics and evolution of the Australian subterranean hydroporine diving beetles (Dytiscidae), with notes on *Carabhydrus*. Invertebr. Syst. 22(2):217–225.
Lohman D.J., de Bruyn M., Page T., von Rintelen K., Hall R., Ng P.K.L., Shih H-T., Carvalho G.R., von Rintelen T. 2011. Biogeography of the Indo-Australian Archipelago. Ann. Rev. Ecol. Evol. Syst. 42:205–226.
Lukoschek V., Keogh J.S., Avise J.C. 2012. Evaluating fossil calibrations for dating phylogenies in light of rates of molecular evolution: a comparison of three approaches. Syst. Biol. 61(1):22–43.
Lucky A. 2011. Molecular phylogeny and biogeography of the spider ants, genus *LeptoMamex* Mayr (Hymenoptera: Formicidae). Mol. Phylogenet. Evol. 59(2):281–292.
Maddison W.P., Midford P.E., Otto S.P. 2007. Estimating a binary character's effect on speciation and extinction. Syst. Biol. 56(5): 701–710.
Martin H.A. 2006. Cenozoic climatic change and the development of the arid vegetation in Australia. J. Arid Environ. 66(3): 533–563.
McGuigan K., Zhu D., Allen G.R., Moritz C. 2000. Phylogenetic relationships and historical biogeography of melanotaeniid fishes in Australia and New Guinea. Mar. Freshwater Res. 51(7): 713–723.
McPeek M.A. 2008. The ecological dynamics of clade diversification and community assembly. Am. Nat. 172(6):270–284.
Miller K.B., Lubkin S.H. 2001. *Calicovatellus petrodytes*, a new genus and species of primitive vatelline diving beetle (Coleoptera: Dytiscidae: Hydroporinae: Vatellini) from the Miocene Barstow Formation, southern California, USA. J. Paleontol. 75(4):890–894.
Miller K.B. 2005. Revision of the New World and south-east Asian Vatellini (Coleoptera: Dytiscidae: Hydroporinae) and phylogenetic analysis of the tribe. Zool. J. Linnean Soc.144(4):415–510.
Miller K.G., Kominz M.A., Browning J.V., Wright J.D., Mountain G.S., Katz M.E., Sugarman P.J., Cramer B.S., Christie-Blick N., Pekar S.F. 2005. The Phanerozoic record of global sea-level change. Science 310(5752):1293–1298.
Monaghan M.T., Balke M., Pons J., Vogler A.P. 2006. Beyond barcodes: complex DNA taxonomy of a South Pacific Island radiation. Proc. Roy. Soc. B. 273(1588):887–893.
Moreau C.S., Bell C.D., Vila R., Archibald S.B., Pierce N.E. 2006. Phylogeny of the ants: diversification in the age of angiosperms. Science. 312(5770):101–104.
Morlon H., Parsons T.L., Plotkin J.B. 2011. Reconciling molecular phylogenies with the fossil record. Proc. Natl. Acad. Sci. USA. 108(39):16327–16332.
Morlon H., Kemps B.D., Plotkin J.B., Brisson D. 2012. Explosive radiation of a bacterial species group. Evolution. 66(8):2577–2586.

Morvan C., Malard F., Paradis E., Lefébure T., Konecny-Dupré L., Douady C.J. 2013. Timetree of Aselloidea reveals species diversification dynamics in groundwater. Syst. Biol. 62(4):512–522.
Müller C.J., Matos-Maraví P.F., Beheregaray L.B. 2012. Delving into *Delias* Hübner (Lepidoptera: Pieridae): fine-scale biogeography, phylogenetics and systematics of the world's largest butterfly genus. J. Biogeogr. 40(5):881–893.
Murphy N.P., Adams M., Austin A.D. 2009. Independent colonization and extensive cryptic speciation of freshwater amphipods in the isolated groundwater springs of Australia's Great Artesian Basin. Mol. Ecol. 18(1):109–122.
Murphy N.P., Breed M.F., Guzik M.T., Cooper S.J.B., Austin A.D. 2012. Trapped in desert springs: phylogeography of Australian desert spring snails. J. Biogeogr. 39(9):1573–1582.
Nagalingum N.S., Marshall C.R., Quental T.B., Rai H.S., Little D.P., Matthews S. 2011. Recent synchronous radiation of a living fossil. Science. 334(6057):796–799.
Near T.J., Sanderson M.J. 2004. Assessing the quality of molecular divergence time estimates by fossil calibrations and fossil–based model selection. Phil. Trans. Roy. Soc. B. 359(1450):1477–1483.
Nilsson A.N. 2001. Dytiscidae (Coleoptera). World catalogue of insects. Vol. 3. Stenstrup (Danmark): Apollo Books. p. 395.
Nilsson A.N., Fery H. 2006. World Catalogue of Dytiscidae - corrections and additions, 3 (Coleoptera: Dytiscidae). Koleopterologische Rundschau. 76:55–74.
Papadopoulou A., Anastasiou I., Vogler A.P. 2010. Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. Mol. Biol. Evol. 27(7):1659–1672.
Paradis E., Claude J., Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 20(2):289–290.
Phillimore A.B., Price T.D. 2008. Density-dependent cladogenesis in birds. PLoS Biol. 6(3):e71.
Ponniah M., Hughes J.M. 2004. The evolution of Queensland spiny mountain crayfish of the genus Euastacus. I. Testing vicariance and dispersal with interspecific mitochondrial DNA. Evolution. 58(5):1071–1085.
Ponomarenko A.G. 1987. New Mesozoic water beetles (Insecta, Coleoptera) from Asia. Paleontol. Zh. 2:83–97.
Rabosky D.L. 2006. Likelihood methods for detecting temporal shifts in diversification rates. Evolution. 60(6):1152–1164.
Rabosky D.L., Donnellan S.C., Talaba A.L., Lovette I.J. 2007. Exceptional among-lineage variation in diversification rates during the radiation of Australia's most diverse vertebrate clade. Proc. Roy. Soc. B. 274(1628):2915–2923.
Rabosky D.L., Lovette I.J. 2008. Explosive evolutionary radiations: decreasing speciation or increasing extinction through time? Evolution. 62(8):1866–1875.
Rabosky D.L. 2009. Ecological limits and diversification rate: alternative paradigms to explain the variation in species richness among clades and regions. Ecol. Lett. 12(8):735–43.
Ree R.H., Moore B.R., Webb C.O., Donoghue M.J. 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. Evolution. 59(11):2299–2311.
Ree R.H., Smith S.A. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Syst. Biol. 57(1):4–14.
Ribera I., Vogler A.P., Balke M. 2008. Molecular phylogeny and diversification of diving beetles (Coleoptera: Dytiscidae). Cladistics. 24(4):563–590.
Ricklefs R.E. 2004. A comprehensive framework for global patterns in biodiversity. Ecol. Lett. 7(1):1–15.
Ricklefs R.E. 2007. Estimating diversification rates from phylogenetic information. Trends Ecol. Evol. 22(11): 601–610.
Rix M.G., Harvey M.S. 2012. Phylogeny and historical biogeography of ancient assassin spiders (Araneae: Archaeidae) in the Australian mesic zone: Evidence for Miocene speciation within Tertiary refugia. Mol. Phylogenet. Evol. 62(1):375–396.
Ronquist F., Sanmartín I. 2011. Phylogenetic methods in biogeography. Annu. Rev. Ecol. Evol. S. 42(1):441–464.
Ronquist F., Teslenko M., van der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A., Huelsenbeck J.P. 2012. MrBayes 3.2:Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. Syst. Biol. 61(3):539–542.
Sauquet H., Ho S.Y., Gandolfo M.A., Jordan G.J., Wilf P., Cantrill D.J., Bayly M.J., Bromham L., Brown G.K., Carpenter R.J., Lee D.M., Murphy D.J., Sniderman J.M.K., Udovicic F. 2012. Testing the impact of calibration on molecular divergence times using a fossil-rich group: the case of *Nothofagus* (Fagales). Syst. Biol. 61(2):289–313.
Schultz M.B., Smith S.A., Horwitz P., Richardson A.M.M., Crandall K.A., Austin C.M. 2009. Evolution underground: a molecular phylogenetic investigation of Australian burrowing freshwater crayfish (Decapoda: Parastacidae) with particular focus on *Engaeus* Erichson. Mol. Phylogenet. Evol. 50(3):580–598.
Schweizer M., Güntert M., Hertwig S.T. 2012. Phylogeny and biogeography of the parrot genus *Prioniturus* (Aves: Psittaciformes). J. Zool. Syst. Evol. Res. 50(2):145–156.
Schwentner M., Timms B.V., Richter S. 2012. Flying with the birds? Recent large-area dispersal of four Australian *Limnadopsis* species (Crustacea: Branchiopoda: Spinicaudata). Ecol. Evol. 2(7): 1605–1626.
Sniderman J.M.K., Pillans B., O'Sullivan P.B., Kershaw A.P. 2007. Climate and vegetation in southeastern Australia respond to Southern Hemisphere insolation forcing in the late Pliocene–early Pleistocene. Geology. 35(1):41–44.
Sniderman J.M.K., Porch N., Kershaw A.P. 2009. Quantitative reconstruction of Early Pleistocene climate in southeastern Australia and implications for atmospheric circulation. Quaternary Sci. Rev. 28(27):3185–3196.
Sniderman J.K., Jordan G.J., Cowling R.M. 2013. Fossil evidence for a hyperdiverse sclerophyll flora under a non–Mediterranean-type climate. Proc. Natl. Acad. Sci. USA. 110(9):3423–3428.
Soltis D.E., Smith S.A., Cellinese N., Wurdack K.J., Tank D.C., Brockington S.F, Refulio-Rodriguez N.F., Walker J.B., Moore M.J., Carlsward B.S., Bell C.D., Latvis M., Crawley S., Black C., Diouf D., Xi Z., Rushworth C.A., Gitzendanner M.A., Sytsma K.J., Qiu Y., Hilu K.W., Davis C.C., Sanderson M.J., Beaman R.S., Olmstead R.G., Judd W.S., Donoghue M.J., Soltis P.S. 2011. Angiosperm phylogeny: 17 genes, 640 taxa. Am. J. Bot. 98(4):704–730.
Sota T., Takami Y., Monteith G.B., Moore B.P. 2005. Phylogeny and character evolution of endemic Australian carabid beetles of the genus *Pamborus* based on mitochondrial and nuclear gene sequences. Mol. Phylogenet. Evol. 36(2):391–404.
Stadler T. 2011. Inferring speciation and extinction processes from extant species data. Proc. Natl. Acad. Sci. USA. 108(39): 16145–16146.
Stadler T. 2013. Recovering speciation and extinction dynamics based on phylogenies. J. Evol. Biol. 26(6):1203–1219.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 22(21):2688–2690.
Stelbrink B., Albrecht C., Hall R., von Rintelen T. 2012. The biogeography of Sulawesi revisited: is there evidence for a vicariant origin of taxa on Wallace's "anomalous island"? Evolution 66(7):2252–2271.
Swofford D.L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
Tänzler R., Toussaint E.F.A., Suhardjono Y.R., Balke M., Riedel A. 2014. Multiple transgressions of Wallace's Line explain diversity of flightless *Trigonopterus* weevils on Bali. Proc. Roy. Soc. B. 281(1782):20132528.
Toussaint E.F.A., Hall R., Monaghan M., Sagata K., Ibalim S., Shaverdo H.V., Vogler A.P., Pons J., Balke M. 2014. The towering orogeny of New Guinea as a trigger for arthropod megadiversity. Nat. Commun. 5:5001.
Toussaint E.F.A., Sagata K., Surbakti S., Hendrich L., Balke M. 2013. Australasian sky islands act as a diversity pump facilitating peripheral speciation and complex reversal from narrow endemic to widespread ecological supertramp. Ecol. Evol. 3(4) :1031–1049.
Unmack P.J. 2001. Biogeography of Australian freshwater fishes. J. Biogeogr. 28(9):1053–1089.
Unmack P.J., Dowling T.E. 2010. Biogeography of the genus *Craterocephalus* (Teleostei: Atherinidae) in Australia. Mol. Phylogenet. Evol. 55(3):968–984.

Unmack P.J., Bagley J.C., Adams M., Hammer M.P., Johnson J.B. 2012. Molecular phylogeny and phylogeography of the Australian freshwater fish genus *Galaxiella* with an emphasis on dwarf galaxias (*g. Pusilla*). PLoS One. 7(6):e38433.

Unmack P.J., Allen G.R., Johnson J.B. 2013. Phylogeny and biogeography of rainbowfishes (Melanotaeniidae) from Australia and New Guinea. Mol. Phylogenet. Evol. 67(1):15–27.

Voris H.K. 2000. Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. J. Biogeogr. 27(5): 1153–1167.

Wallace A.R. 1860. On the zoological geography of the Malay Archipelago. J. Proc. Linn. Soc. 4(16):172–184.

Wallace A.R. 1863. On the physical geography of the Malay Archipelago. J. Roy. Geog. Soc. 33:217–234.

Watts C.H.S. 1997. A new genus and species of Australian Dytiscidae (Coleoptera). Rec. South Australian Mus. 29(2):121–123.

Watts C.H.S. 2002. Checklist and guides to the identification, to genus, of adults and larval Australian water beetles of the families Dytiscidae, Noteridae, Hygrobiidae, Haliplidae, Gyrinidae, Hydraenidae and the superfamily Hydrophiloidea (Insecta – Coleoptera). Cooperative Research Centre for Freshwater Ecology (Australia). Identificat. Ecol. Guide. 43:1–110.

Watts C.H.S., Hancock P.J., Leys R. 2008. *Paroster peelensis* sp nov.: a new stygobitic water beetle from alluvial gravels in Northern New South Wales (Coleoptera: Dytiscidae). Aust. J. Entomol. 47(3):227–231.

Watts C.H.S., Humphreys W.F. 1999. Three new genera and five new species of Dytiscidae (Coleoptera) from underground waters in Australia.Rec. South Australian Mus. 32(2):121–142.

Watts C.H.S., Humphreys W.F. 2000. Six new species of *Nirridessus* and *Tjirtudessus* (Dytiscidae; Coleoptera) from underground waters in Australia Rec. South Australian Mus. 33:127–144.

Watts C.H.S., Humphreys W.F. 2001. A new genus and six new species of Dytiscidae (Coleoptera) from underground waters in the Yilgarn Palaeodrainage 15 system of Western Australia. Rec. South Australian Mus. 34(2):99–114.

Watts C.H.S., Humphreys W.F. 2003. Twenty-five new Dytiscidae (Coleoptera) of the genera *Tjirtudessus* Watts and Humphreys, *Nirripirti* Watts and Humphreys and *Bidessodes* Regimbart, from underground waters in Australia. Rec. South Australian Mus. 36(2):135–187.

Watts C.H.S., Humphreys W.F. 2004. Thirteen new Dytiscidae (Coleoptera) of the genera *Boongurrus* Larson, *Tjirtudessus* Watts and Humphreys and *Nirripirti* Watts and Humphreys, from underground waters in Australia. Trans. R. Soc. S. Aust. 128:99–129.

Watts C.H.S., Humphreys W.F. 2006. Twenty-six new Dytiscidae (Coleoptera) of the Genera *Limbodessus* Guignot and *Nirripirti* Watts and Humphreys, from underground waters in Australia. Trans. R. Soc. S. Aust. 130(1):123–185.

Watts C.H.S., Humphreys W.F. 2009. Fourteen new Dytiscidae (Coleoptera) of the genera *Limbodessus* Guignot and *Paroster* Sharp and *Exocelina* Broun from underground waters in Australia. Trans. R. Soc. S. Aust. 133(1):62–107.

Wiegmann B.M., Trautwein M.D., Winkler I.S., Barr N.B., Kim J.W., Lambkin C., Bertone M.A., Cassel B.K., Bayless K.M., Heimberg A.M., Wheeler B.M., Peterson K.J., Pape T., Sinclair B.J., Skevington J.H., Blagoderov V., Caravas J., Kutty N.S., Schmidt-Ott U., Kampmeier G.E., Thompson F.C., Grimaldi D.A., Beckenbach A.T., Courtney G.W., Friedrich M., Meier R., Yeates D.K. 2011. Episodic radiations in the fly tree of life. Proc. Natl. Acad. Sci. USA. 108(14):5690–5695.

Wiens J.J. 2011. The causes of species richness patterns across space, time, and clades and the role of "ecological limits". Q. Rev. Biol. 86(2):75–96.

Woodburne M., Tedford R., Swisher C. 1990. Lithostratigraphy, biostratigraphy, and geochronology of the Barstow Formation, Mojave Desert, Southern California. Geol. Soc. Am. Bull. 102(4): 459–477.

Xie W., Lewis P.O., Fan Y., Kuo L., Chen M-H. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. Syst. Biol. 60(2):150–160.

PART 2: ISLAND LINEAGE DIVERSIFICATION

Chapter 8. Biogeography on a recent and geologically puzzling island: New Guinea

Tavurvur active stratovolcano on the island of New Britain, Northern margin of New Guinea

"*The whole northern peninsula of New Guinea, as well as the islands of Wagion, Salwatty, and Balauta, are exceedingly rugged and mountainous. There is a continued succession of jagged and angular ranges of hills, and everywhere behind them, ridge beyond ridge stretch far away into the interior. Over the whole country spreads an unvarying forest, of a somewhat stunted appearance…*"

Alfred Wallace, *Notes of a Voyage to New Guinea*, 1859.

Chapter contents

8.1 Paper VIII – Diversification of diving beetles in New Guinean highlands 176

8.2 Paper IX - Diversification of diving beetles during the New Guinean orogeny 196

PAPER VIII

8.1 Australasian sky islands act as a diversity pump facilitating peripheral speciation and complex reversal from narrow endemic to widespread ecological supertramp

Citation: Toussaint EFA, Sagata K, Surbakti S, Hendrich L, Balke M (2013) Australasian sky islands act as a diversity pump facilitating peripheral speciation and complex reversal from narrow endemic to widespread ecological supertramp. *Ecology and Evolution.* 3(4):1031-1049.

Ecology and Evolution

Open Access

Australasian sky islands act as a diversity pump facilitating peripheral speciation and complex reversal from narrow endemic to widespread ecological supertramp

Emmanuel F. A. Toussaint[1], Katayo Sagata[2], Suriani Surbakti[3], Lars Hendrich[1] & Michael Balke[1,4]

[1]Zoological State Collection, Münchhausenstraße 21, Munich, 81247, Germany
[2]Papua New Guinea Institute for Biological research (PNG-IBR), Goroka, Papua New Guinea
[3]Jurusan Biology, FMIPA-Universitas Cendrawasih, Kampus Baru, Jayapura, Papua, Indonesia
[4]GeoBioCenter, Ludwig-Maximilians-University, Munich, Germany

Keywords
Australian region, diversity pump, highlands, New Guinea, New Zealand, peripheral speciation.

Correspondence
Emmanuel F.A. Toussaint, Zoological State Collection, Münchhausenstraße 21, Munich, 81247, Germany. Tel: +49 898107155; Fax: +49 898107300;
E-mail: Toussaint@zsm.mwn.de

Funding Information
This study was supported by German Science Foundation (DFG) grants BA2152/6-1, 7-1 and 11-1.

Received: 11 January 2013; Revised: 29 January 2013; Accepted: 1 February 2013

doi: 10.1002/ece3.517

Abstract

The Australasian archipelago is biologically extremely diverse as a result of a highly puzzling geological and biological evolution. Unveiling the underlying mechanisms has never been more attainable as molecular phylogenetic and geological methods improve, and has become a research priority considering increasing human-mediated loss of biodiversity. However, studies of finer scaled evolutionary patterns remain rare particularly for megadiverse Melanesian biota. While oceanic islands have received some attention in the region, likewise insular mountain blocks that serve as species pumps remain understudied, even though Australasia, for example, features some of the most spectacular tropical alpine habitats in the World. Here, we sequenced almost 2 kb of mitochondrial DNA from the widespread diving beetle *Rhantus suturalis* from across Australasia and the Indomalayan Archipelago, including remote New Guinean highlands. Based on expert taxonomy with a multigene phylogenetic backbone study, and combining molecular phylogenetics, phylogeography, divergence time estimation, and historical demography, we recover comparably low geographic signal, but complex phylogenetic relationships and population structure within *R. suturalis*. Four narrowly endemic New Guinea highland species are subordinated and two populations (New Guinea, New Zealand) seem to constitute cases of ongoing speciation. We reveal repeated colonization of remote mountain chains where haplotypes out of a core clade of very widespread haplotypes syntopically might occur with well-isolated ones. These results are corroborated by a Pleistocene origin approximately 2.4 Ma ago, followed by a sudden demographic expansion 600,000 years ago that may have been initiated through climatic adaptations. This study is a snapshot of the early stages of lineage diversification by peripatric speciation in Australasia, and supports New Guinea sky islands as cradles of evolution, in line with geological evidence suggesting very recent origin of high altitudes in the region.

Introduction

Deciphering the mechanisms of species formation is one of the most fascinating and challenging areas of evolutionary biology (Darwin and Wallace 1858; Darwin 1859; Mayr and Diamond 2001; Coyne and Orr 2004; Fitzpatrick et al. 2009; Santini et al. 2012). Many studies on biogeographic and ecological factors promoting speciation have helped to establish the separation by physical barriers ("vicariance") followed by genomic isolation as the null hypothesis that begets new species (Lynch 1989; Barraclough and Vogler 2000; Mayr and Diamond 2001; Johannesson 2010; Santini et al. 2012). However, although restriction of gene flow and allopatric speciation appear to be the most common mechanisms, different processes have also been documented through the years (White 1968; de Aguiar et al. 2009; Johannesson 2010). For instance, sympatric speciation, suggested already by

Darwin (1859) before being tested and supported for many metazoan taxa (e.g., Crow et al. 2010), implies that speciation events can occur within the same population through genetic polymorphism. Furthermore, special cases of allopatric speciation are found in parapatric and peripatric models, which invoke processes at the distributional periphery of an ancestral species, where individuals might enter a new habitat most likely facilitated by divergent ecological characteristics. In this case of peripheral speciation, also known as "budding speciation," either the geographic isolation, ecological factors, or a combination of the two leads to a cessation of gene flow, thus enabling the speciation process (Mayr 1982; Fitzpatrick and Turelli 2006).

Here, we focus on the Australasian archipelago and surrounding areas, a region that shelters a rich yet highly threatened biodiversity; seven biodiversity hotspots are situated in the Indomalayan-Australasian region (Mittermeier et al. 2004; conservation.org). This region, despite its highly complex geological history (e.g., Hall 2011; Metcalfe 2011), represents an ideal laboratory to study lineage diversification and speciation (Wallace 1860; Mayr and Diamond 2001; Condamine et al. 2013). Thousands of islands, many of them scattered across the Equator, varying in size from tiny patches to continental sized landmasses, ranging from young to geologically old and low-lying to high altitudes including snow-capped summits, harbor hyperdiverse biota and exceptional radiations. Particularly across the megadiverse Wallacea and Melanesia, most studies to date have investigated larger scale evolutionary patterns, whereas factors promoting speciation remain scarcely addressed, and the evolutionary processes involved are little known despite increased recent efforts (e.g., Von Rintelen et al. 2004; de Bruyn and Mather 2007; Joseph and Omland 2009; Craft et al. 2010; Deiner et al. 2011; Klaus et al. 2013). A unique feature in the Indomalayan-Australasian archipelago is its long chain of islands often with high mountains, usually surrounded by tropical lowland rain- or dry-forest. "Sky island" ecosystems are isolated patches surrounded by dramatically different lowland ecosystems (Heald 1967), in this case, isolated further from each other by ocean. Recently, such highland ecosystems have been shown to act as evolutionary cradles shaping a flourishing biota in diverse regions of the World (Hall 2005; Smith and Farrell 2005; Robin et al. 2010; Schultheis et al. 2012). In the Australasian region, sky islands are geologically young (<5 Mya; Cloos et al. 2005), yet highly diverse (e.g., Mittermeier et al. 2004). Numerous studies on Australian and New Zealand mountain ranges have investigated speciation patterns and indicate important vicariant effects from mountain uplift per se, and climate change as further promoter of species diversity (e.g., Trewick et al. 2000; McCulloch et al. 2010; Hawlitschek et al. 2012). The mountains of Indonesia and northern Australasia show especially striking altitudinal gradients, usually being surrounded by tropical lowland rain- or dry-forest. Within this biodiversity-rich assemblage of mountainous ecosystems, one of the largest and most remote highland regions is the central New Guinean cordillera, with vast expanses of tropical montane and subalpine habitat, and extensive areas above 3500-m altitude, including numerous summits above 4500 m. Their role as a diversity pump for the archipelago remains poorly studied and little appreciated, despite their vast geographic extent and extreme structuring (but see Mayr and Diamond 1976).

Here, we study pond-dwelling *Rhantus* diving beetles, often abundant in tropical montane and subalpine pond habitats across the Australasian region. There are at least 30 endemic species in the region including Oceania, mostly narrow endemics restricted to a single high valley, or few mountain tops. There is one striking exception, however: *Rhantus suturalis* MacLeay, 1825 (Coleoptera, Dytiscidae, Colymbetini) (Fig. 1), is a widespread ecological supertramp, ranging from the Azores islands to New Zealand, and inhabits diverse lentic habitats in mountainous or subalpine environments, for example, high altitude lakes (Fig. 2) and highland peat swamps (with pH around 4–5). It occurs in temperate lowland swamps, many anthropogenic habitats (freshly dug fish ponds, reservoirs, roadside ditches, cattle troughs), saline desert wetlands in North African deserts, and many more. Along with its closer relatives, it has never been found in tropical lowlands (Balke 1993, 2001; Balke et al. 2009). *Rhantus suturalis* is often an early colonizer of newly available habitats, hence referred to as a "supertramp species" (Balke et al. 2009). It is an ecological generalist, with high physiological tolerance, for example, in terms of salinity, temperature, and acidity. A comprehensive molecular phylogeny of *R. suturalis* from across its wide range revealed two

Figure 1. Habitus of *Rhantus suturalis* (Photo credit: Jan Hamrský)

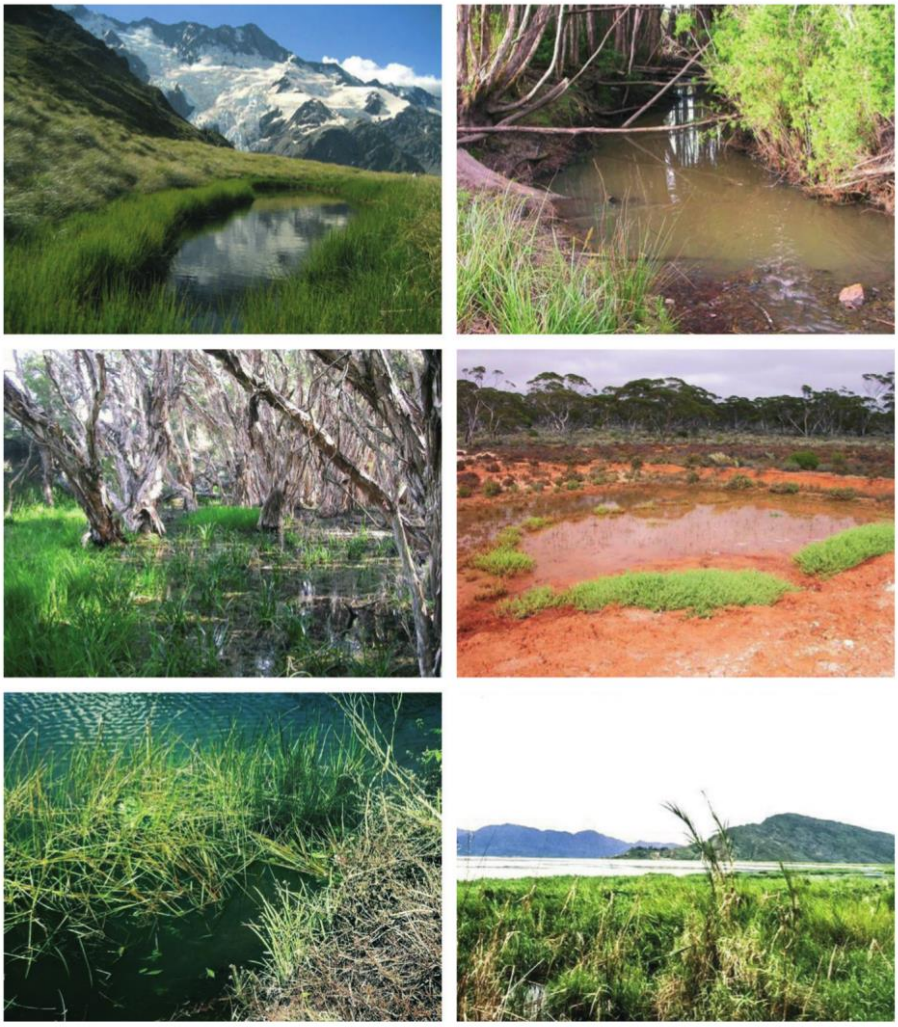


Figure 2. Habitats and habitus of the *Rhantus suturalis* southern clade in Southeast Asia and Australasia. Top left: montane peatland pond in Sealy Tarns (~1300 m) (New Zealand); top right: Welcome River in the North-West of Tasmania; center left: flooded paperbark swamp in Beeliar wetlands (Western Australia); centre right: temporary pool in the Mallee near Balladonia (Western Australia); bottom left: edge of a lake in Ranu Pani (East Java); bottom right: Lake Paniai (1700–3000 m) sheltering *R. ekari* in West Papua.

major clades – a northern one, from Portugal to Sumatra, and a southern one from adjacent Java eastward to New Caledonia (Fig. 3) (Balke et al. 2009). Both *R. suturalis* clades contain one or more narrow-endemic species previously described based on marked morphological divergence. This species paraphyly was supported by extensive mitochondrial and nuclear DNA sampling (>7000 bp). A recent origin of *R. suturalis*, *c.* 6.0–2.7 Ma ago, was suggested, possibly in New Guinea, followed by an ancestral colonization of the Malay Archipelago and a large part of the Australian region. The rise of a widespread generalist out of a clade of narrow endemics not only refutes the assumption of "specialisation as an evolutionary dead-end" (Mayr 1963; coined by Cope 1896 "the law of the unspecialized"), but offers an opportunity to study the early phases of lineage diversification across a wide species range, which is nevertheless constrained by climate and other ecological factors.

In this study, we use extensive sampling across the Indomalayan-Australasian region to (1) reconstruct phylogenetic relationships within the southern clade of the widespread *R. suturalis*, (2) investigate phylogeographic patterns using haplotype network inferences, (3) infer the historical demography and timing of divergence of this group in a paleoclimatic framework, in order to test the hypothesis of ongoing peripatric speciation in Australasia, particularly New Guinea, sky islands, and (4) examine whether those mountain chains act as a species pump in a "cradle of evolution" model, or as an ancient biotic pool, in a "museum" model.

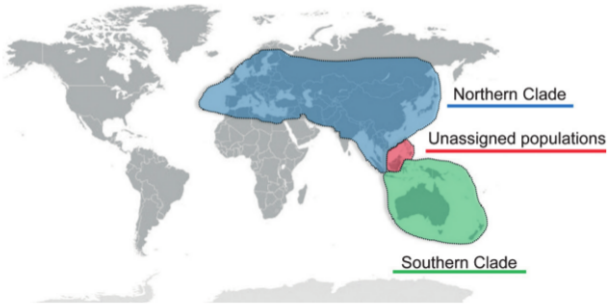


Figure 3. Distribution of *Rhantus suturalis*. Sequence data were not available for the area in red.

Materials and Methods

Taxon sampling and molecular biology

We sequenced 133 individuals of *R. suturalis* from 12 regions across the range of the southern clade (Table 1, Fig. 4) including the Sunda Islands, almost the entire New Guinean highland chain, Australia, New Zealand, and New Caledonia. *Rhantus suturalis* is known from old Philippine (Baguio) as well as Malaysian Borneo (Mt. Kinabalu) specimens (Balke 1993), but we did not manage to obtain fresh samples. We sampled most of the New Guinean endemic species (Fig. 5) subordinated within *R. suturalis*, that is, (1) *Rhantus dani*, Balke 2001 (usually shaded wetlands in an isolated montane depression, the Baliem Valley ~1700 m), (2) *Rhantus ekari*, Balke and Hendrich 1992 (swampy edge of a large montane lake, Lake Paniai ~1900 m) (Balke and Hendrich 1992), (3) *Rhantus riedeli*, Balke 2001 (same habitat but at Lake Anggi ~1900 m), and (4) *Rhantus supranubicus*, Balke 2001 (alpine peat swamp pools and edge of Lake Habbema as well as Mount Elit swampland, ~3300 m) (Balke 2001). *Rhantus kakapupu,* Balke 2001, described from old specimens collected across Lake Paniai where *R. ekari* occurs, was not found recently.

Specimens of closely related *Rhantus* species, that is, *R. bacchusi*, *R. elisabethae*, *Rhantus* new species 1 and 2, as well as *R. suturalis* from the northern clade were included as outgroups (Balke 2001; Balke et al. 2007, 2009).

Genomic DNA was extracted from legs or thoracic tissues using the DNeasy kit (Qiagen, Hilden, Germany). We sequenced 1095 bp from the mitochondrial cytochrome *c* oxidase subunit 1 (702 bp) and cytochrome b (393 bp) using the primers listed in Table 2 to conduct PCR reactions with standard protocols (http://zsm-entomology.de/wiki/The_Beetle_D_N_A_Lab). Both strands of the PCR products were then sequenced and sequences corrected and aligned using Geneious 5.6.5 (available from http://www.geneious.com) before being exported under Mesquite 2.75 (available from http://www.mesquiteproject.org) to check the reading frame and create three different datasets (CO1, CytB, and *Combined*). We sequenced fragments of the nuclear genes 18S and arginine kinase for several specimens of the southern clade, but no molecular variation in the alignment was identified (data not shown). This is in line with Balke et al. (2009) who used 18S rRNA, wingless, elongation factor 1 alpha (2 exons and 1 intron), and histone 3 and found little or no informative signal within the southern clade. Because of this lack of informative sites, we use fast evolving mitochondrial markers here. All the sequences used in this study are deposited in Genbank under the accession numbers KC604111 - KC604412.

Phylogeny

Different methods of phylogenetic inference were used for the *Combined* dataset to infer relationships of individuals within *R. suturalis*: (1) distance analyses using the Neighbor-Joining method implemented in Geneious 5.6.5 (Drummond et al. 2012) with 10000 bootstrap replicates and a HKY model of evolution (see below for a rationale on this setting); (2) Maximum Parsimony (MP) analyses using TNT 1.1 (Goloboff et al. 2008) with the Sectorial Searches, *Tree Ratchet, Tree Fusing* and *Tree Drifting* algorithms (Goloboff 1999), and 100 random additional sequences. A Symmetric Resampling with a probability fixed to 10 and 1000 replicates was performed as it allows avoiding uninformative characters, character weight, and transformation costs to affect the resampling unlike classic *Bootstrapping* and *Jacknifing*; (3) Maximum Likelihood (ML) analyses were performed with 1000 bootstrap replicates under RAxML (Stamatakis 2006) with different partitioning strategies: *NoPart* (no partitioning), *ByGene* (one partition for each gene), *ByCodon* (one partition for each codon position), and BySix (one partition for each codon position of each gene); and finally, (4) Bayesian Inference (BI) analyses were performed using the same strategies of partitioning, under MrBayes 3.1.2 (Ronquist and Huelsenbeck 2003). The datasets were analyzed with two independent runs consisting of eight Markov Chains Monte Carlo (MCMC, one cold and seven incrementally heated) sampling for 30 million generations. In order to compute support information, the trees were sampled every 1000 generations and each MCMC started from a random topology. The split-frequencies as long as the log-likelihood curves were investigated to provide a good estimate of the burn-in fraction. Once these samples were discarded, the remaining topologies were used to yield a 50% majority rule consensus tree. Best-fitting partitioning strategies for the ML and BI analyses were selected using Bayes Factors (BF; Kass and Raftery 1995) approximated under Tracer

Table 1. Collecting localities, with code and number of specimens.

Species	Country	Region	Code	Locality	Specimens
Rhantus bacchusi	Papua New Guinea	Eastern Highlands	PNGEHP	Aiyura	1
R. bacchusi	Papua New Guinea	Eastern Highlands	PNGEHP	Goroka	1
R. bacchusi	Papua New Guinea	Eastern Highlands	PNGEHP	Hogu	2
R. dani	Indonesia	Papua	INDPAP	Wamena	6
R. ekari	Indonesia	Papua	INDPAP	Enarotali	1
R. elisabethae	Papua New Guinea	Enga	PNGENG	Mt. Hagen Kumul Lodge	3
R. elisabethae	Papua New Guinea	Southern Highlands	PNGSHP	Mt. Giluwe Sopulkul	1
R. elisabethae	Papua New Guinea	Southern Highlands	PNGSHP	Tari	2
R. riedeli	Indonesia	Papua	INDPAP	Anggi	3
Rhantus sp.	Papua New Guinea	Central	PNGCEN	Myola	3
Rhantus sp.	Papua New Guinea	Morobe	PNGMOR	Huon	1
R. supranubicus	Indonesia	Papua	INDPAP	Lake Habbema	8
R. suturalis	Australia	New South Wales	AUSNSW	Bellingen	3
R. suturalis	Australia	New South Wales	AUSNSW	Braidwood	3
R. suturalis	Australia	New South Wales	AUSNSW	Casino	1
R. suturalis	Australia	New South Wales	AUSNSW	Delegate	9
R. suturalis	Australia	New South Wales	AUSNSW	Grafton	1
R. suturalis	Australia	New South Wales	AUSNSW	Kocsciousko NP	1
R. suturalis	Australia	New South Wales	AUSNSW	Nowra	3
R. suturalis	Australia	New South Wales	AUSNSW	Taraga	1
R. suturalis	Australia	New South Wales	AUSNSW	Wollongong	2
R. suturalis	Australia	Queensland	AUSQLD	Agnes	2
R. suturalis	Australia	Queensland	AUSQLD	Bundaberg	1
R. suturalis	Australia	Queensland	AUSQLD	Gladstone	3
R. suturalis	Australia	Southern Australia	AUSSA	Adelaide	1
R. suturalis	Australia	Southern Australia	AUSSA	Mt. Gambier	3
R. suturalis	Australia	Southern Australia	AUSSA	Meadows Creek	2
R. suturalis	Australia	Southern Australia	AUSSA	Penola	6
R. suturalis	Australia	Southern Australia	AUSSA	Robe	1
R. suturalis	Australia	Tasmania	AUSTAS	Geeveston	3
R. suturalis	Australia	Tasmania	AUSTAS	Togari	1
R. suturalis	Australia	Victoria	AUSVIC	Kyneton	2
R. suturalis	Australia	Victoria	AUSVIC	Tooborac	1
R. suturalis	Australia	Western Australia	AUSWA	Cataby	1
R. suturalis	Australia	Western Australia	AUSWA	Manjimup	1
R. suturalis	Australia	Western Australia	AUSWA	Manypeaks	1
R. suturalis	Australia	Western Australia	AUSWA	Northcliffe	1
R. suturalis	Australia	Western Australia	AUSWA	Pilbara	1
R. suturalis	Australia	Western Australia	AUSWA	Yanmah State For.	1
R. suturalis	Belarus	Minsk Oblast	BEL	Minsk	1
R. suturalis	Czech Republic	Liberec	CZE	Liberec	1
R. suturalis	France	New Caledonia	NEWCAL	Mt. Mou	1
R. suturalis	France	New Caledonia	NEWCAL	Poindimié	1
R. suturalis	France	New Caledonia	NEWCAL	Pouembout	1
R. suturalis	Indonesia	Flores	INDFLO	Mt. Ranaka/Ranamese Lake	6
R. suturalis	Indonesia	Java	INDJAVA	Dieng Plateau	6
R. suturalis	Indonesia	Lombok	INDLOM	Sumbalun Lawang	4
R. suturalis	Indonesia	Sulawesi	INDSUL	Malino	4
R. suturalis	Indonesia	West Sumatra	INDSUM	Danau di Atas	2
R. suturalis	Indonesia	Timor	INDTIM	Mt. Mutis	5
R. suturalis	Japan	Hokkaido	JAP	Tomakomai	1
R. suturalis	New Zealand	Auckland	NEWZEA	Auckland	1
R. suturalis	New Zealand	Nelson	NEWZEA	Canaan	3
R. suturalis	New Zealand	Southland	NEWZEA	Key Summit	1
R. suturalis	Papua New Guinea	Central	PNGCEN	Myola	2
R. suturalis	Papua New Guinea	Eastern Highlands	PNGEHP	Aiyura	3

Table 1. Continued.

Species	Country	Region	Code	Locality	Specimens
R. suturalis	Papua New Guinea	Enga	PNGENG	Wabag	1
R. suturalis	Papua New Guinea	Madang	PNGMAD	Mts. Finisterre	3
R. suturalis	Papua New Guinea	Sandaun	PNGSAN	Mianmin	4
R. suturalis	Papua New Guinea	Sandaun	PNGSAN	Telefomin	1
R. suturalis	Papua New Guinea	Southern Highlands	PNGSHP	Mt. Ambua	4
R. suturalis	Papua New Guinea	Western Highlands	PNGWHP	Giluwe	2
R. suturalis	Papua New Guinea	Western Highlands	PNGWHP	Mt. Hagen Town	2
R. suturalis	Papua New Guinea	Western Highlands	PNGWHP	Mt. Hagen Kumul Lodge	2

1.5 (Rambaut and Drummond 2007). The substitution models of evolution for each partition used in ML and BI analyses were selected under jModelTest 0.1.1 (Posada 2008), using the Bayesian information criterion (BIC) rather than the corrected Akaike information criterion (AICc) as advocated by Brown and Lemmon (2007). The HKY model was set for the NJ analysis as it is the closest one to the model selected for the *Combined* dataset implemented in Geneious 5.6.5.

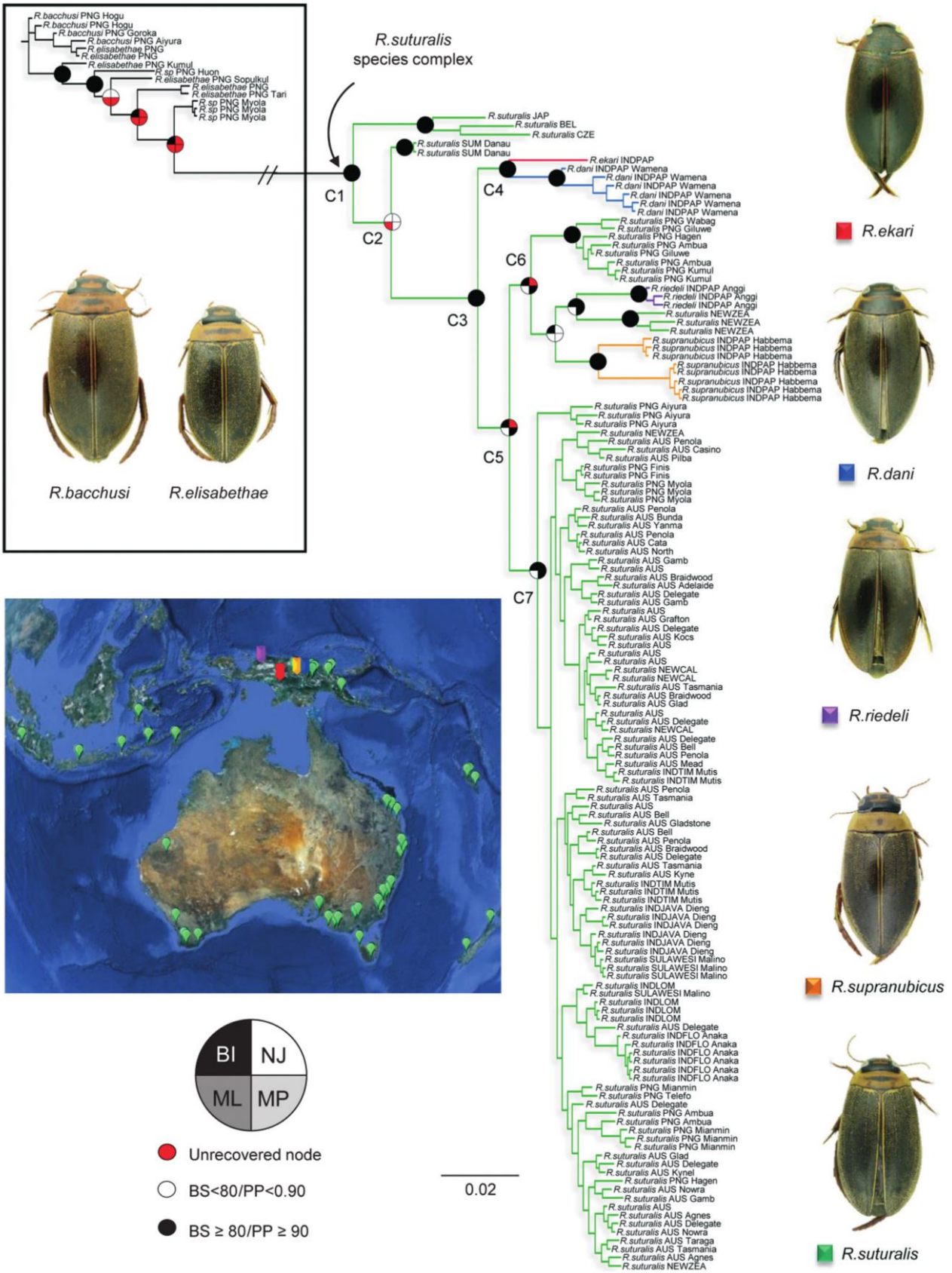
Phylogeography and historical demography

The phylogeographic pattern within the southern clade was analyzed for all the *Combined* dataset through haplotype network inferred from the 133 specimens of the southern clade in addition to 3 specimens of the northern clade included as outgroups. The sequences were collapsed into haplotypes under DnaSP 5.10 (Librado and Rozas 2009) and a network was inferred using Hapstar 0.7 (Teacher and Griffiths 2011) based on the connection lengths obtained in Arlequin 3.11 (Excoffier et al. 2005).

Historical demography was investigated with the southern clade specimens only using Tajima's *D*, Fu's *Fs*, and Harpending's raggedness index (Hri). Tajima's *D* (Tajima 1989) and Fu's *Fs* (Fu 1997) statistics were calculated using Arlequin 3.11 (Excoffier et al. 2005) with 10,000 permutations to assess whether the mitochondrial data shows evidence of deviation from the neutral theory of molecular evolution holding that stochastic processes such as molecular drift and mutation explain most of the genetic variation found in living organisms. In addition, these statistics can unveil demographic events such as population expansion (significant negative values) or contraction (significant positive values) (Tajima 1989; Fu 1997). The Harpending's raggedness index (Hri, Harpending 1994) based on mismatch distributions, was calculated using 1000 bootstrap replicates to investigate whether the population deviates from a sudden expansion model (Schneider and Excoffier 1999). A significant Hri ($P < 0.05$) indicates a poor fit to the model and therefore does not support a sudden demographic expansion (Harpending 1994).

The magnitude of historical demographic events was investigated using Bayesian Skyline Plots (BSP, Drummond et al. 2005) as well as Extended Bayesian Skyline Plots (EBSP, Heled and Drummond 2008) under BEAST 1.7.4 (Drummond et al. 2012). BSPs allow the inference of population historical demography in a Bayesian framework based on a coalescent model of evolution. The EBSPs are a slightly different method based on BSPs permitting the analyses of multiple loci separately (see Ho and Shapiro 2011 for a review). The.xml files were created with a partition for each gene and the respective models of evolution set according to the results obtained in jModelTest (Posada 2008). The applicability of a molecular clock was tested using PAUP* (Swofford 2003), and as the molecular clock hypothesis was not statistically supported ($P < 0.05$), we used a relaxed clock that allows rate variation among lineages. Therefore, *The Coalescent: Bayesian Skyline* and *Extended Bayesian Skyline* models were implemented with an estimated relaxed clock (uncorrelated lognormal) based on the rate of evolution calculated by Balke et al. (2009) regarding the evolution of the *R. suturalis* complex, including the 95% confidence interval ($r = 0.019$, 95% interval $I = 0.011$–0.028). The rate was set under a normal distribution with the following parameters: initial value = 0.0195, mean = 0.0195, and SD = 0.00435. Two

Figure 4. Phylogenetic relationships of the *Rhantus suturalis* species complex *Combined* dataset with the best-fitting strategy of partitioning under Bayesian Inference. Supports for each node are indicated according to the caption inserted in the figure (BI, Bayesian inference; NJ, neighbor-joining; MP, maximum parsimony; ML, maximum likelihood). A map highlighting collection localities is shown, in which the colors of the spots refer to the respective colored squares underneath the habitus of the different species (e.g., *R. riedeli* in purple). The major clades are labeled C1 to C7. Names of the species for which a habitus is displayed are specified under the pictures.



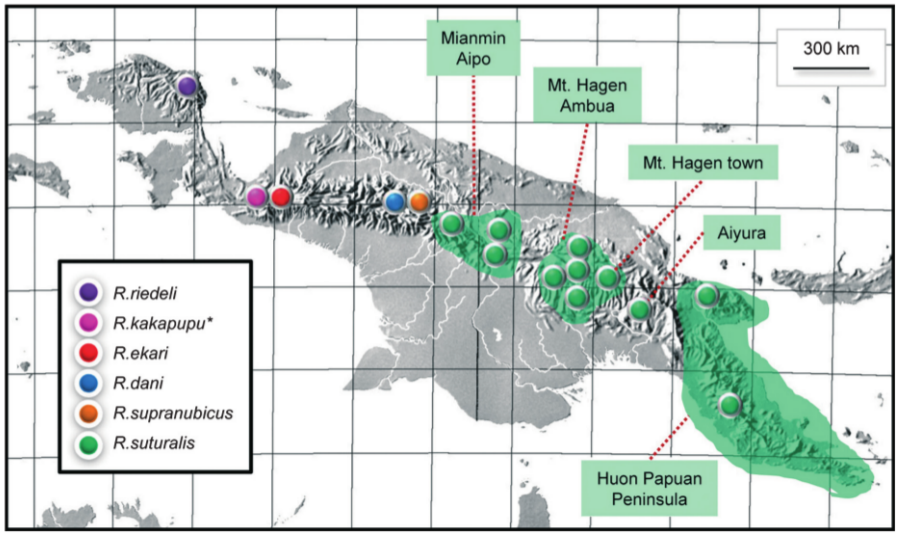


Figure 5. Distribution of New Guinean endemic species of the *Rhantus suturalis* complex. The different colors refer to the distribution of each species except *R. suturalis* for which they refer to the sampling localities (the distribution of *R. suturalis* in New Guinea is given by the green areas). The correspondences of the colors are shown in the legend at the bottom left corner of the figure. The asterisk indicates that *R. kakapupu* was not included in this study.

Table 2. Primers used to amplify regions of the cytochrome oxidase subunit 1 (CO1) and cytochrome B (CytB).

Locus	Primer	Primer sequence	Reference
CytB	CB3	GAG GAG CAA CTG TAA TTA CTA A	Barraclough et al. 1999
	CB4	AAA AGA AA(AG) TAT CAT TCA GGT TGA AT	Barraclough et al. 1999
Cox1	Pat	TCC AAT GCA CTA ATC TGC CAT ATT A	Simon et al. 1994
	Jerry	CAA CAT TTA TTT TGA TTT TTT GG	Simon et al. 1994

distinct runs of 50 million generations sampled every 1000 generations were performed for each model (BSP or EBSP). After discarding 10% of the samples as burn-in, the convergence of runs was assessed according to the ESS (Effective Sample Size) criterion and the plots were inferred under Tracer 1.5 (Rambaut and Drummond 2007) for the BSP and a graphic program for the EBSP.

Estimation of divergence times

As the fossil record is scarce for water beetles, and this study focuses on inter- as well as intra-specific levels, we chose to use the previously introduced evolutionary rate (Balke et al. 2009) with different models and parameters as advocated by previous studies (e.g., Ho and Phillips 2009) to infer diversification ages. A.xml file based on the *Combined* dataset was created with the following non-default settings and priors: the *Site Model* was chosen according to the models of evolution used in the phylogenetic analyses and the MCMC parameters were fixed to 30 million generations with sampling every 1000 generations and the first 25% discarded as burn-in. Divergence time analyses were carried out using BEAST 1.7.4 (Drummond et al. 2012) and were performed under both the *Coalescent: Constant Size* and *Speciation: Birth-Death models*. We used estimated relaxed clock rate (uncorrelated lognormal) with a normal distribution (initial value = 0.0195, mean = 0.0195, standard dev = 0.00435), and also a uniform distribution (initial value = 0.0195, upper = 0.028, lower = 0.011). The best topology obtained in BI for the *Combined* dataset was fixed as the reference topology for divergence time estimates by editing the.xml file manually. At the end of each analysis, a 50% majority rule consensus tree was created under TreeAnnotator 1.7.4. Likelihood scores and Bayes Factors (BF) were then calculated under Tracer 1.5 (Rambaut and Drummond 2007) to select the best analysis.

Results

Phylogenetic relationships

We obtained fragments of 524–702 bps length for CO1 and 281–393 bps for CytB for 133 specimens to produce

Table 3. Selection of the best-fitting models of sequence evolution under the corrected Akaike (AICc) and Bayesian (BIC) information criterions.

Dataset	AICc	BIC
Combined	TrN + I + G	TrN + I + G
Combined Position 1	K80 + I+G	TrN + I + G
Combined Position 2	JC	F81
Combined Position 3	HKY + G	TrN + G
Cytochrome oxidase 1	TrN + I + G	TrN + I + G
Cytochrome oxidase 1 Position 1	K80 + I+G	TrN + I + G
Cytochrome oxidase 1 Position 2	JC	F81
Cytochrome oxidase 1 Position 3	TrN + G	TrN + I + G
Cytochrome oxidase B	HKY + I + G	HKY + I + G
Cytochrome oxidase B Position 1	TrN + G	TrN + G
Cytochrome oxidase B Position 2	F81	F81
Cytochrome oxidase B Position 3	HKY + I + G	HKY + I + G

an alignment of 1905 bps without stop codons or frameshift mutations. The best-fitting evolutionary models for each partition are given in Table 3, and a phylogenetic hypothesis is shown in Figure 4, based on the best BI topology for the *Combined* dataset after selection under the BF criterion (Table 4) (for the best topologies recovered in NJ, MP, and ML, see Figs S1, S2, and S3 of the electronic supplementary materials, respectively).

Within the *R. suturalis* species complex, most of the internal nodes were well to strongly supported by bootstrap (BS $\geq$ 80) or posterior probability (PP $\geq$ 0.90) values. The monophyly of *R. suturalis* species complex labeled "C1" in Figure 4 was always retrieved (BS = 100/PP = 1.0). The next clade (C2) contains the southern clade and specimens from Sumatra, which are recovered as sister group of all remaining specimens. This clade was retrieved in all methods of inference except in the ML analysis in which they were the first branch of C1.

The next clade (C3) is the southern clade of *R. suturalis* of Balke et al. (2009), here always recovered with strong support (BS $\geq$ 98/PP = 1.0). *Rhantus ekari* and *R. dani* form clade C4 (BS $\geq$ 89/PP = 0.99) as the sister group of all the remaining specimens from the southern *R. suturalis* clade (C5). The next clade, C5, recovered in all analyses (BS $\geq$ 79/PP = 1.0) but in NJ, comprises the clades C6 (*R. riedeli*, *R. supranubicus*, and several specimens of *R. suturalis* from Papua New Guinea and New Zealand), and C7 (with all remaining specimens of *R. suturalis*, mainly from across Australia).

Strikingly, the Papua New Guinean specimens in C6 (all from Mt. Hagen-Ambua highlands region) were recovered in a well-supported, monophyletic, and genetically well-separated clade, the same is true for the New Zealand specimens recovered as the sister taxa of *R. riedeli* in clade C6. This means that there are two clades with specimens that are morphologically *R. suturalis* that group among narrow endemics, morphologically moderately to strongly divergent from *R. suturalis*, and these two clades are genetically isolated from the main clade of morphological *R. suturalis* specimens (C7).

Overall, the topology discloses a striking, partial lack of geographic structure. Exceptions are *R. ekari*, *R. dani*, *R. riedeli*, and *R. supranubicus* endemic to different West New Guinean highland regions and recovered as strongly supported monophyletic clades (BS $\geq$ 95/PP = 1.0), as well as the Papua New Guinea and New Zealand specimens in clade C6. Australian specimens are scattered in clade C7 without geographic signal, whereas specimens from Papua New Guinea (PNG) are found in small and scattered internal groups in C7, the ones from Eastern Highlands (Aiyura) and Huon-Papuan Peninsula being monophyletic (Fig. 5). Furthermore, Flores as well as Javanese, Lombok, New Caledonian, New Zealand, Sulawesi, or Timorese individuals were recovered as paraphyletic, or in poorly supported clusters.

Phylogeography and historical demography

All DNA matrices present striking haplotype diversities within the southern clade of *R. suturalis*, from 81% in the CytB to 93% in the *Combined* dataset, these results being supported by high nucleotide diversities (Table 5). Phylogeographic analyses based on the *Combined* dataset yielded a complex network with multiple haplotype series (Fig. 6). The species *R. dani*, *R. ekari*, *R. riedeli*, and *R. supranubicus* from the highlands of Papua New Guinea are well separated from the two central groups of

Table 4. Best-fitting strategies of partitioning for the BI and ML phylogenetic inferences with Bayes Factors (B_F) estimates, BI harmonic means, and ML optimization likelihoods.

			Bayes factors (B_F)			
Partitioning scheme	MrBayes harmonic mean	RAxML likelihood	NoPart	ByGene	ByCodon	BySix
Combined NoPart	−6317.20	−5875.28	–	0	0	0
Combined ByGene	−6390.39	−5862.60	>10	–	0	0
Combined ByCodon	−6197.00	−5526.02	>10	>10	–	0
Combined BySix	−6184.51	−5471.57	>10	>10	>10	–

Table 5. Genetic structure of each marker and results of demographic index calculations.

Dataset	*Cytochrome oxidase 1*	*Cytochrome B*	*Combined*
Length (bp)	702	393	1095
Mean number of pairwise differences	13.25 ± 6.00	6.37 ± 3.04	19.61 ± 8.73
Nucleotide diversity (Pi)	0.0219 ± 0.01	0.0241 ± 0.01	0.0225 ± 0.01
Number of haplotypes	107	106	122
Tajima's *D*	−1.201 NS (0.09)	−0.648 NS (0.30)	−1.068 NS (0.12)
Fu's *Fs*	−24.102 *** (0.0008)	−24.987 *** (0.0000)	−23.883 *** (0.0001)
Harpending's raggedness index	0.00320 NS (0.63)	0.00654 NS (0.96)	0.00174 NS (1.00)

NS indicates not significant values. *** indicates highly significant values.

haplotypes formed by numerous specimens of *R. suturalis* mainly from Australia, as well as from Flores, Java, Lombok, New Caledonia, New Zealand, Papua New Guinea, Sulawesi, and Timor. Among these central groups, there is no predominant haplotype and there is a close connection between haplotypes from the entire sampling area; however, no clear geographic structure is recovered, except for the specimens from Flores, which constitute a unique geographic cluster (Fig. 6). This pattern of geographic mixture is less well recovered within highland specimens from Papua New Guinea and specimens from New Zealand, that tend to represent distinct genetic entities similar to the subordinated, formally named Papuan species.

Eight individuals of *R. suturalis* from different high mountains in the Eastern Papua New Guinea highlands (Fig. 7) form a well-delineated clade. Strikingly, one specimen from the valley at the foot of Mt. Hagen, from Mount Hagen town, is in the main haplotype group in one of the star-bursts, and even more strikingly, so are two specimens collected from the high altitude Mt. Ambua locality, taken from the same pool as the specimens in the genetically well-delineated Mt.Hagen-Ambua clade (Fig. 7). These two Ambua specimens cluster with specimens from more western PNG-Papuan highlands localities (Min area in PNG west into Indonesian Papua in the Aipo area) (Fig. 5). Min area specimens are, however, also in a second, geographically proximate clade as well (Fig. 4). PNG specimens from Aiyura, Eastern highlands, form their own clade, as do the Huon Peninsula and Papuan Peninsula specimens. *R. suturalis* specimens from the northern clade are retrieved with a deep genetic divergence highlighted by very long connecting branches, and a geographic continuum, as the Sumatran specimens are more closely related to the southern clades than the Palearctic ones.

The non-significant negative values of Tajima's *D* obtained for all datasets suggest a demographic expansion statistically supported by highly significant negative values of Fu's *Fs*. The hypothesis of population expansion is supported by Harpending's raggedness index as well, as the values for all datasets are very low and non-significant (Table 5).

All the ESBP and BSP runs converged well according to the log-likelihood curves and ESS checked under Tracer 1.5 (Rambaut and Drummond 2007). The EBSP analysis based on the rate of Balke et al. (2009) highlights a scenario divided into a phase of constant population size, followed by a sudden demographic expansion, likely starting approximately 600,000 years ago during the Late Pleistocene (Fig. 8). The results of the BSP analysis highlight a demographic expansion as well (Fig. 8), with a later approximate age of 450,000 years ago.

Estimation of divergence times

The analysis based on the rate of Balke et al. (2009) optimized with a normal distribution and a *Birth-Death model* of speciation was selected under the BF, ESS, and likelihood criterions as the most likely, and the chronogram derived from this analysis is presented in Figure 9. The divergence time estimates obtained for the two best runs (*BDBalkeNorm* and *BDBalkeUni*, see Table 6 for abbreviations) were highly similar with a maximum divergence of less than 4% (~0.2 Ma) for the mean age of the root (Table 6). Our results show that the most recent common ancestor (MRCA) of the *R. suturalis* southern clade originated approximately 2.3 Ma ago (95% credibility interval: 1.2–3.6 Ma) during the Pliocene–Pleistocene transition. Interestingly, most of the intra-specific nodes within the southern clade radiation are young, with ages spanning a period of time from the Early Calabrian (~1.7 Ma) to the Tarantian (~100 kyr). The node C4 (*R. dani* and *R. ekari)* is dated to the Early Calabrian (~1.7 Ma), approximately the same age as the clade C6 (*R. riedeli*, *R. supranubicus*, the Papua New Guinean Mt. Hagen-Ambua clade, and one clade including some of the New Zealand specimens).

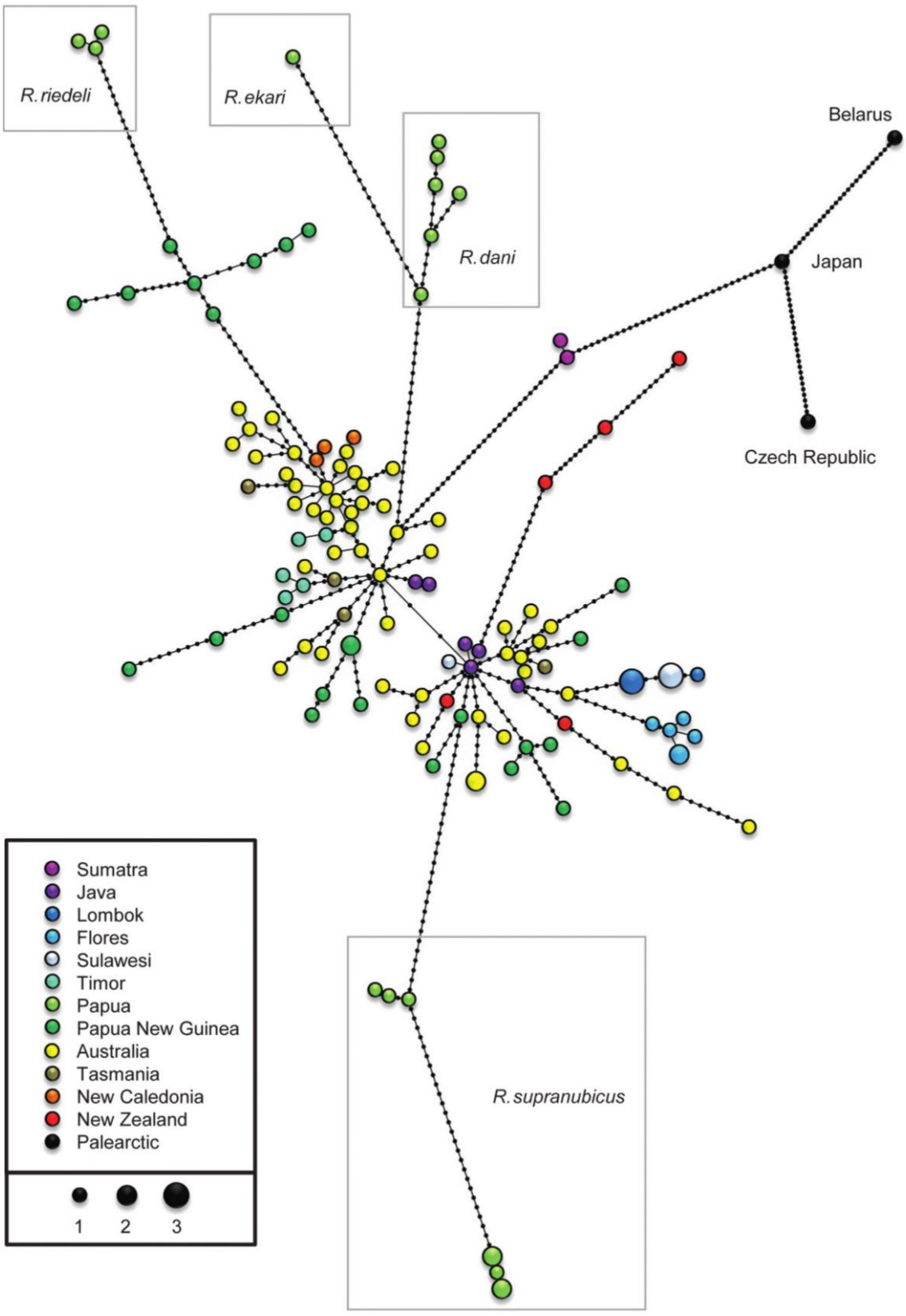


Figure 6. Network based on the *Combined* dataset. The locality and the number of specimen(s) are indicated according to the caption. The black dots indicate missing haplotypes.

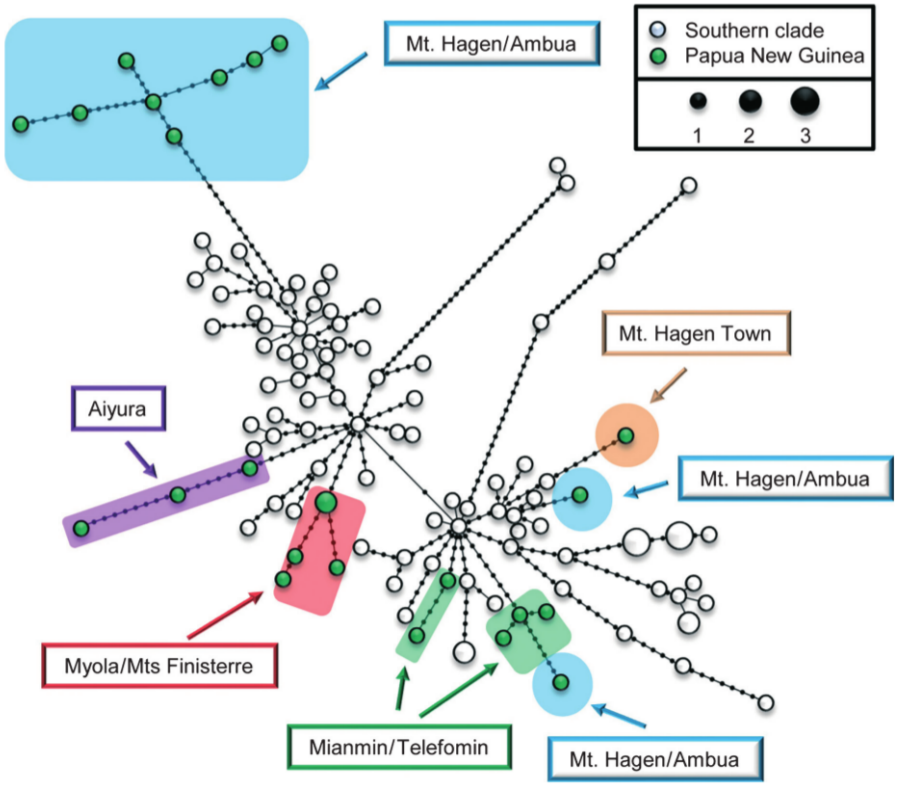


Figure 7. Simplified network based on the *Combined* dataset highlighting the New Guinean specimens. The locality and the number of specimen(s) are indicated according to the caption. The black dots indicate missing haplotypes.

Discussion

Phylogenetic relationships

We retrieve relationships congruent with the multi-gene mtDNA and nDNA study of Balke et al. (2009), with a robustly supported paraphyly of *R. suturalis*. The split between the northern and southern clades is also recovered here even though our placement of specimens from Sumatra is ambiguous, probably due to the scarce sampling for the northern clade. We find a lack of resolution for most of the *R. suturalis* radiation in the clade C5, except for *R. riedeli*, *R. supranubicus,* and for several individuals from New Guinea and New Zealand, which are grouped in C6 as the sister group of all remaining *R. suturalis* specimens from the entire archipelago. *Rhantus ekari* and *Rhantus dani* are recovered sister species, and they are indeed morphologically similar to each other as well as to *R. suturalis,* whereas the other New Guinea endemics *R. supranubicus* and *R. riedeli* are morphologically (male genital, color, for claws) more deviating. Even though Balke et al. (2009) proposed the inclusion of New Guinean highland specimens of *R. suturalis* among a clade comprising *R. supranubicus* and *R. riedeli*, the placement of specimens from southern New Zealand in this New Guinean clade was highly unexpected. Concerning the New Guinean specimens, the collection localities (remote alpine habitats) seem to indicate that these beetles belong to a well-differentiated population that may represent at least one new putative species. The New Zealand specimens, on the other hand, are recovered in a basal clade that is thought to be the ancestral clade of the *R. suturalis* as recovered by Balke et al. (2009). We suggest that isolated specimens from southern New Zealand are likely a relict population from a first colonization wave through the archipelago, and have most likely been evolving independently from the rest of the radiation for a long period of time. This population from mid-altitude lakes, well-separated from other New Zealand populations, might represent a new species, similar to the New Guinean specimens of the clade C6. More generally, the lack of resolution in C7 and the short branches within the different clades of the topology support a very recent radiation of the southern clade.

Phylogeographic network

The partial lack of geographic structure seen in the phylogenetic inference is recovered in the haplotype network (Fig. 6). The Papuan species, the Mt. Hagen-Ambua high-

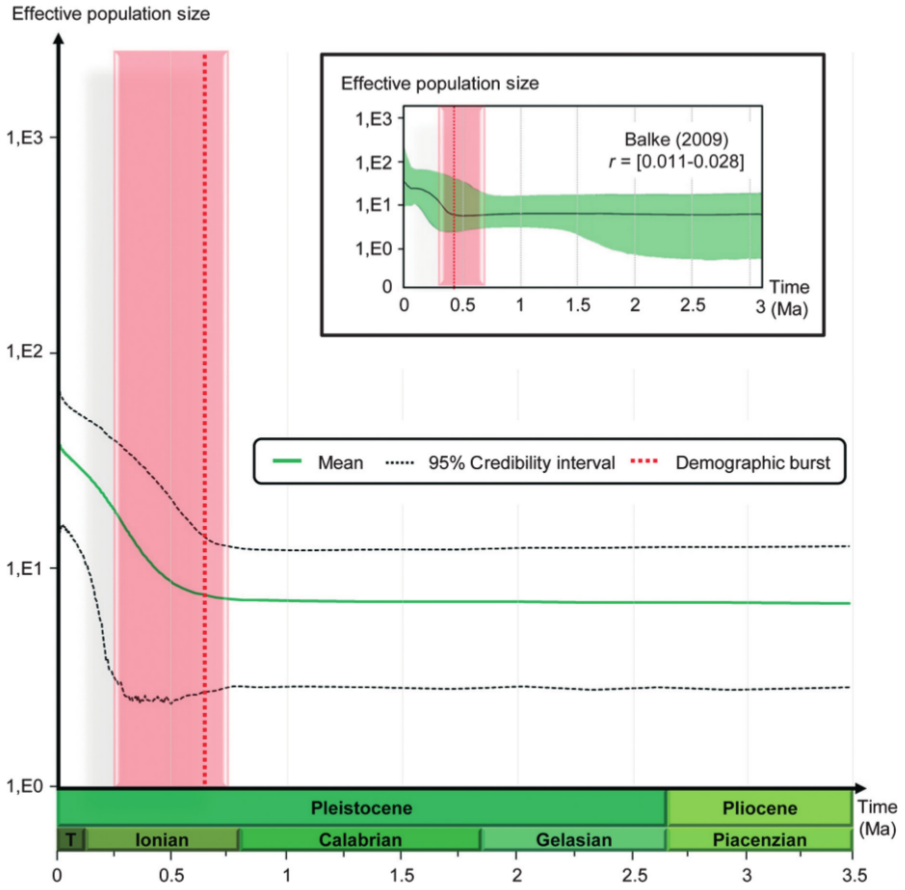


Figure 8. Extended Bayesian Skyline Plot based on the rate calculated by Balke et al. (2009). A 500-kyr timescale is shown at the bottom of the chronogram and spans a period of time from the late Pliocene to the present. Result of the Bayesian Skyline Plot is given in the right part of the figure. Demographic expansion and 95%HPD are shown according to the caption.

Table 6. Mean ages (in Ma) and 95% credibility intervals for the different analyses.

	Root	C1	C2	C3	C4	C5	C6	C7
1.*BDBalkeN*	**6.1 (3.0–9.9)**	**3.0 (1.5–4.7)**	**2.8 (1.4–4.4)**	**2.3 (1.2–3.6)**	**1.6 (0.6–2.8)**	**2.0 (1.0–3.1)**	**1.7 (0.8–2.7)**	**1.5 (0.8–2.4)**
2.*BDBalkeU*	5.9 (3.0–10.1)	2.9 (1.5–4.8)	2.7 (1.3–4.5)	2.2 (1.1–3.7)	1.6 (0.6–2.8)	1.9 (1.0–3.2)	1.6 (0.8–2.8)	1.5 (0.7–2.5)
5.*COALBalkeN*	8.8 (4.0–14.9)	3.8 (1.8–6.2)	3.5 (1.7–5.7)	2.8 (1.3–4.6)	1.9 (0.6–3.5)	2.4 (1.1–3.9)	2.0 (1.0–3.3)	1.7 (0.8–2.9)
6.*COALBalkeU*	9.9 (4.4–17.1)	4.1 (2.0–6.8)	3.7 (1.8–6.3)	3.0 (1.4–5.0)	2.1 (0.7–3.7)	2.5 (1.2–4.2)	2.1 (1.0–3.6)	1.8 (0.8–3.1)

BD, *Birth-Death* model; COAL, *Coalescent* model; N, Normal law of distribution with uncorrelated lognormal clock model; U, Uniform law of distribution with uncorrelated lognormal clock model.
The text of the best run based on BF, ESS, and likelihood criterions is bold.

land specimens from Papua New Guinea and some New Zealand specimens are separated from the core of other haplotypes by a large number of mutational steps. The numerous connections seem to indicate a star-like architecture even though there are some deviations. A star-like network suggests range expansion leading to the evolution of numerous closely related genotypes derived from a wide, central, and often ancestral haplotype (Avise 2000). Here, the lack of one central and widespread haplotype along with the general absence of geographic structuring in the main clades seem to suggest either (1) ongoing but moderate gene flow driven by dispersion within the wider area of distribution, therefore allowing the mixture of genotypes from different localities while avoiding complete homogenization, or (2) the recent cessation of gene flow across the Archipelago, thus leading to the isolation of populations that start to diverge genetically (e.g., Ribera et al. 2011). In the latter case, we suggest that the

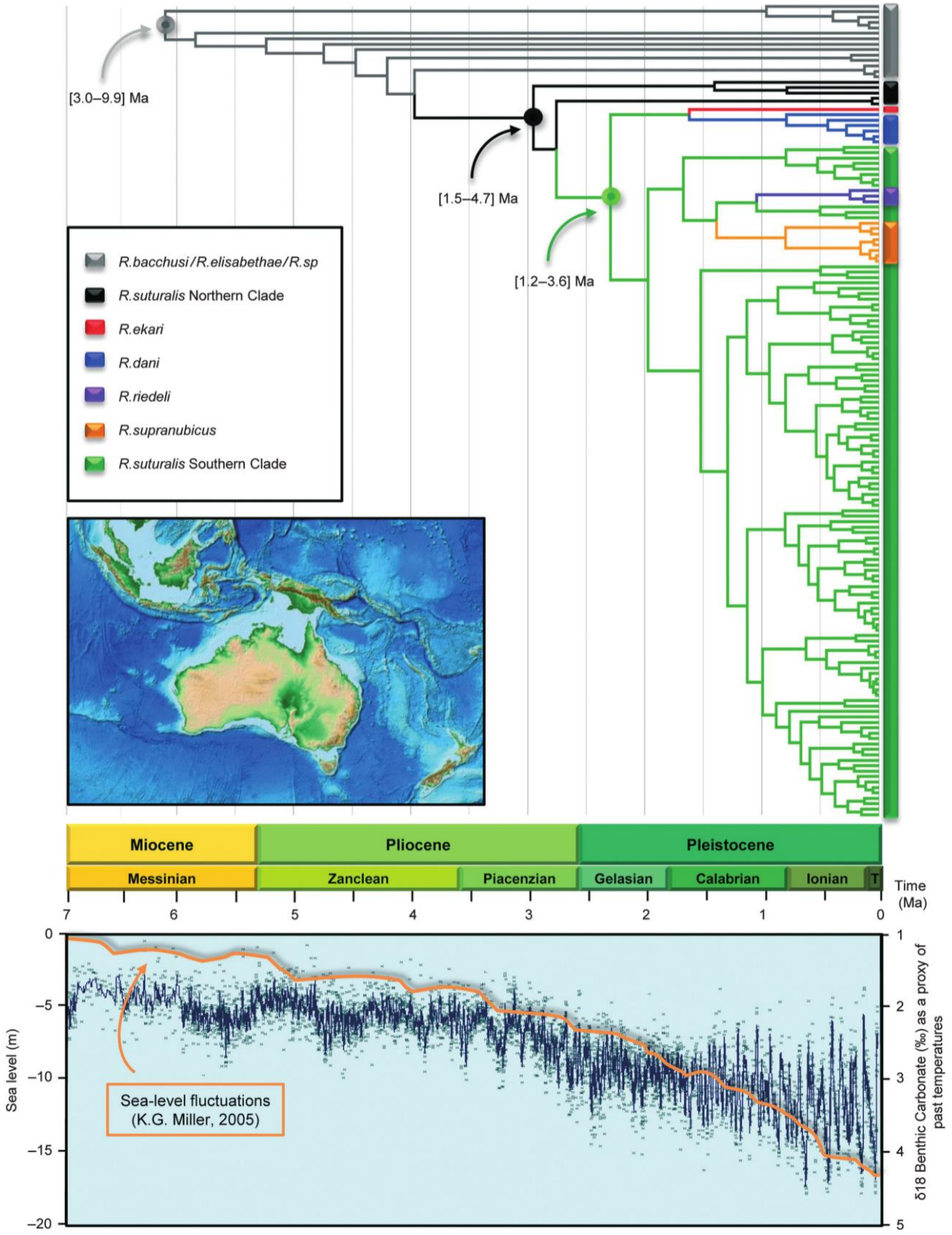


Figure 9. Maximum credibility tree with mean ages (Ma) from the BEAST analysis. A 1-Ma timescale is shown at the bottom of the chronogram and spans a period of time from the late Miocene to the present. The 95%HPD intervals of divergence times are shown between square brackets for the three major nodes of the chronogram. The vertical bands and pastilles at the nodes of different colors referring to the color of the clades highlight groups of interest for which the names are provided in the top left of the figure. A map with bathymetric information (light blue indicates shallow sea/dark blue indicates deep sea) is shown along with a graphic presenting the evolution of sea level and temperature during the last 7 Ma.

populations are not clustered in well-distinguished geographic groups because of the recency of the event. Even though there is a lack of clear geographic correlation, all specimens of some localities are closely related (e.g., Flores, New Caledonia, Timor, PNG Telefomin-Min area, PNG Aiyura area etc., Figs 6, 7), indicating the initiation of geographic structuring or colonization from the same or related sources. Interestingly, New Zealand specimens in a distinct clade in the phylogeny are not recovered close to New Guinean specimens, but are connected to a central and unique Javanese haplotype, most likely the result of incomplete lineage sorting. The deep divergence between these specimens and the central haplotype group, including the rest of the specimens from New Zealand, supports our hypothesis of an older isolation and restricted gene flow hinting toward ongoing speciation, a similar pattern as the one observed for the Papuan highland species. In agreement with the phylogenetic inference, the group of specimens from Papua recovered in C6 is separated from the central haplotype group and clusters with *R. riedeli*, while still exhibiting a deep genetic divergence between the two. In addition, most of the specimens from Papua New Guinea are restricted to small groups in the periphery of the phylogeographic networks, suggesting colonization or ongoing geographic isolation from the rest of the populations.

More importantly, it was astonishing to find syntopically occurring Mt. Ambua (Papua New Guinea) specimens in isolated clades C6 and C7 (Figs 6, 7). Specimens in C6 form a well-delineated clade in the network as well as in the phylogenetic trees. The Ambua specimens from the main clade C7 group with individuals from the mountain chain west of the Mt. Hagen-Ambua area, that is, the Telefomin-Mianmin-Aipo area (the wider Star Mountains, Fig. 7 map). The presence of two genetically very distinct populations of *R. suturalis* in the same locality suggests longer isolation of a population in the Mt. Hagen-Ambua area (Figs 6, 7), and recent secondary contact with dispersers out of the widespread clade C7. It is striking to note that the single specimen we obtained from the foot of Mountain Hagen, from the valley close to Mt. Hagen Town, also belongs to the large clade C7 and has no closer relatives, supported in both tree and network inference.

Pleistocene evolution in the *R. suturalis* complex

The divergence time estimates support an Early Pleistocene origin of the *R. suturalis* southern clade, while the branching pattern indicates that the radiation of the group started more recently, most likely in the Middle Pleistocene (~1 Ma). Furthermore, the sudden demographic expansion during the late Pleistocene, *c.* 600 kyr ago, corroborates this scenario of recent radiation during the Quaternary ice ages (2.4 Ma until present) (Hewitt 2000). By then, the high mountains of New Guinea and their (peat) swamps existed already, meaning that vast, highly structured highlands and associated habitats were available.

During the last decades, the impact of Pleistocene glaciations on tropical regions has been widely acknowledged, including global cooling, rising aridity, rainforest depletion, and ecosystem fragmentation associated with refugial budding especially in highlands (Hewitt 2000; Hope et al. 2004; Rull 2011). On the other hand, the dispersal or adaptation of taxa driven by habitat loss or alteration has been increasingly highlighted recently for several insect groups, assuming that the type of response to climate shifts lies on the timescale considered (e.g., Smith and Farrrell 2005; Winkler et al. 2009; Hawlitschek et al. 2012; Toussaint et al. 2012). Interestingly, our findings support the hypothesis that diversification and dispersal of *R.suturalis* started during the Plio-Pleistocene transition (~2.5 Ma), as advocated by Balke et al. (2009). Moreover, the placement of New Guinean species suggests a first transgression of Wallace's line during the early Pleistocene (~2.4 Ma) followed by a settlement in New Guinean highlands. Dispersal across the Australian region possibly out of New Guinea, toward New Zealand, New Caledonia, Sahul, and Wallacea, with secondary transgression of Wallace's line into Java occurred at the same time as the ongoing cooling of the region in the Pliocene (Figs 4, 9).

During this period, the sharp reduction in temperatures in New Guinean highlands may have led to altitudinal migration implying downward dispersion of most of the species, but also the adaptation of some populations to cooler climate, therefore promoting isolation by vicariance (Rull 2011). At this time, New Guinea had a similar relief to today, with extremely rugged highlands surrounded by

lowland tropical rainforests on either side, and often interspersed with chains of lowland forests in-between. Therefore, and even though forest expanse was declining, different geographic localities would have had significantly different microclimates during glacial maxima (Hewitt 2000; Hope et al. 2004), with the likely presence of multiple suitable refugia. Populations trapped in these sky islands then likely evolved in a similar way to the classic case of oceanic island isolation (Gillespie and Roderick 2002), while the downward migration of the New Guinean highland biota and its dispersion in Australasia might have promoted speciation in lowlands. As advocated by Verstappen (1997) and Hewitt (2000), the Pleistocene glaciations, even though driving global cooling, nevertheless constituted a succession of warmer and cooler periods known as the Milankovitch cycles. Therefore, highland populations might have been separated in times of cooling, promoting genetic isolation in sky islands, before being reconnected to other populations during warmer climatic phases.

Interestingly, our results underpin the different expected prospective stages of an early lineage diversification: (1) within the widespread morphologically delineated *R. suturalis*, there are different narrow-endemic species, which are morphologically quite distinct from *R. suturalis* (*R. riedeli*, *R. supranubicus*, with differences in male genitalia and male fore claws, the latter also displaying different coloration); (2) there are different narrow-endemic species with differences in male genitalia, but otherwise rather similar to *R. suturalis* (*R. dani*, *R. ekari,* and *R. kakapupu,* the latter not sequenced here); (3) there are genetically isolated groups morphologically, however, assigned to *R. suturalis* (e.g., the Mt. Hagen-Ambua clade, as well as an isolated New Zealand clade); and (4) within *R. supranubicus*, we find deep divergence between specimens from the same puddles, collected over a decade. The latter case, as well as the syntopical presence of the Mt. Hagen-Ambua clade with individuals from a distant clade, strongly support the idea that isolation of populations after dispersal can be comparably long, but secondary contact of populations might occur at any time.

Here, we suggest that this budding speciation might represent a good example of peripatric speciation, as the populations were not only geographically separated but also had to adapt to local ecological conditions in the periphery of the species distributional range. While this ongoing speciation process occurred in highlands, the Australian lowland populations obviously remained connected by strong gene flow as indicated by the low divergence levels of the haplotypes among the southern clades. Populations in Sunda and the Wallacean mountains, with the exception of Flores, were not monophyletic (Fig. 6), indicating incomplete lineage sorting after very recent arrival, or continuous gene flow across tropical lowland and oceanic barriers. The syntopic occurrence of beetles from distinct clades, as illustrated by New Guinea and New Zealand highland communities, might indicate that these parts of the *R. suturalis* species complex are at the end of an isolation stage. Interestingly, the exclusion of Sumatran specimens from the southern clade confirms the hypothesis of Balke et al. (2009), suggesting that the northern and southern clades are now separated by only a few 100 km between Sumatra and Java.

There might be another interpretation of these macroevolutionary patterns. The northern and southern clades of *R. suturalis* could represent two morphologically highly similar yet different species. We did not consider assigning two species names because we lack samples from southern Sumatra and mainland Southeast Asia and the Philippines, which we suggest might help to better understand species limits in this complex. If there were indeed two distinct species, and depending on the topology of the undersampled northern group, the supertramp trait might either still be ancestral in the *R. suturalis* complex, or has originated twice. In the southern group, this would be the origin of a supertramp (*R. suturalis*, which has its type locality in Java) out of the New Guinea clade of narrow-endemic species. Colonization of the Australasian region would have led to the formation of a paraphyletic series of narrow-endemic species, and then origin of the widespread New Guinea-Australian region-Wallacean supertramp. Peripheral speciation would, in that case, be ongoing in the rather deviating New Guinea and possibly Flores clades within *R. suturalis*.

Conclusion

R. suturalis is a morphologically rather uniform, very widespread, dispersive species and ecological generalist, occurring from saline ponds in oasis up to high altitude peat swamps, although being absent from tropical lowlands. Populations of *R. suturalis* across the Australasian-Indomalayan region are well connected by ongoing dispersal while peripheral speciation processes occur in highland ecosystems. Mountains of New Zealand and much more, so in the extremely rugged vast highlands of New Guinea, apparently harbor different stages of speciation. There are four narrow endemics, morphologically and genetically distinct species emerging from within the widespread genealogy. Furthermore, there are several genetically more or less divergent, isolated clades that agree morphologically, however, with the widespread form. We suggest that reproductive isolation has been shaped by Quaternary glaciations that promoted peripheral budding especially in New Guinea sky island ecosystems. The general cooling since the Pliocene might have promoted the demographic expansion and wide disper-

sion observed in *R. suturalis*, out of a clade of narrow endemics. *Rhantus suturalis* illustrates the reversal or switch of endemic species to widespread generalists, and then toward narrow-endemism (presumably with loss of physiological tolerance) again. This is an element of the taxon cycle (Wilson 1959, 1961), which predicts that higher physiological tolerance might occur at some stage in lineage evolution and promote colonization of new areas, and ultimately diversify into more specialized habitats. The high mountains of New Guinea act as a diversity pump for the region and therefore represent an evolutionary cradle of diversity, certainly deserving further investigation based on more extensive taxon and character sampling.

Acknowledgments

We thank Mark de Bruyn, Oliver Hawlitschek, Ignacio Ribera, Thomas von Rintelen, and two anonymous reviewers for the fruitful comments that considerably helped to improve this article. We also thank Jan Hamrský for the stunning photography of habitus illustrating this study. This study was supported by German Science Foundation (DFG) grants BA2152/6-1, 7-1 and 11-1.

Authors Contributions

Conceived and designed the experiments: EFAT MB. Collected the samples: KS LH MB SS. Analyzed the data: EFAT. Designed the figures: EFAT. Wrote the paper: EFAT MB. Contributed substantially to the modifications of manuscript drafts: LH.

Conflict of Interest

None declared.

References

de Aguiar, M. A. M., M. Baranger, E. M. Baptestini, L. Kaufman, and Y. Bar-Yam. 2009. Global patterns of speciation and diversity. Nature, 460:384–387.

Avise, J. C. 2000 Phylogeography. The history and formation of species. Harvard University Press, Cambridge, MA.

Balke, M. 1993. Taxonomische Revision der pazifischen, australischen und indonesischen Arten der Gattung *Rhantus* Dejean, 1833 (Coleoptera: Dytiscidae). Koleopterol. Rundsch., 63:39–84.

Balke, M.. 2001. Biogeography and classification of New Guinea Colymbetini (Coleoptera: Dytiscidae). Invertebr. Taxon., 15:259–275.

Balke, M., and L. Hendrich. 1992. Ein neuer *Rhantus* Dejean aus West-Neuguinea (Coleoptera: Dytiscidae). Entomol. Z. 102:37–39.

Balke, M., G. Wewalka, Y. Alarie, and I. Ribera. 2007. Molecular phylogeny of Pacific island Colymbetinae: radiation of New Caledonian and Fijian species (Coleoptera, Dytiscidae). Zool. Scr. 36:173–200.

Balke, M., I. Ribera, M. Miller, L. Hendrich, K. Sagata, A. Posman, et al. 2009. New Guinea highland origin of a widespread arthropod supertramp. Proc. R. Soc. B Bio. Sci. 276:2359–2367.

Barraclough, T. G., and A. P. Vogler. 2000. Detecting the geographical pattern of speciation from species level phylogenies. Am. Nat. 155:419–434.

Barraclough, T. G., J. E. Hogan, and A. P. Vogler. 1999. Testing whether ecological factors promote cladogenesis in a group of tiger beetles (Coleoptera: Cicindelidae). Proc. R. Soc. B Bio. Sci. 266:1061–1067.

Brown, J. M., and A. R. Lemmon. 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. Syst. Biol., 56:643–655.

de Bruyn, M., and P. B. Mather. 2007. Molecular signatures of Pleistocene sea-level changes that affected connectivity among freshwater shrimp in Indo-Australian waters. Mol. Ecol., 16:4295–4307.

Cloos, M., B. Sapiie, A. Q. van Ufford, R. J. Weiland, P. Q. Warren, and T. P. McMahon. 2005. Collisonal delamination in New Guinea: the geotectonics of subducting slab breakoff. Geol. S. Am. S. Paper, 400:1–51.

Condamine, F. L., E. F. A. Toussaint, A. M. Cotton, G. S. Genson, F. A. H. Sperling, and G. J. Kergoat. 2013. Fine-scale biogeographical and temporal diversification processes of peacock swallowtails (*Papilio* subgenus *Achillides*) in the Indo-Australian Archipelago. Cladistics, 29:88–111.

Cope, E. D. 1896. The primary factors of organic evolution. Open Court Publishing Company, Chicago.

Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer Associates, Inc., Sunderland, MA.

Craft, K. J., S. U. Pauls, K. Darrow, S. E. Miller, P. D. N. Hebert, L. E. Helgen, et al. 2010. Population genetics of ecological communities with DNA barcodes: an example from New Guinea Lepidoptera. Proc. Natl. Acad. Sci.USA, 107:5041–6.

Crow, K. D., H. Munehara, and G. Bernardi. 2010. Sympatric speciation in a genus of marine reef fishes. Mol. Ecol., 19:2089–2105.

Darwin, C. 1859. The Origin of Species by Means of Natural Selection. Penguin Classics, Harmondsworth, UK.

Darwin, C., and A. Wallace. 1858. On the tendency of species to form varieties, and on the perpetuation of varieties and species by natural means of selection. J. Linn. Soc. London, Zool. 3:45–62.

Deiner, K., A. R. Lemmon, A. L. Mack, R. C. Fleisher, and J. P. Dumbacher. 2011. A passerine bird's evolution corroborates the geological history of the island of New Guinea. PLoS ONE, 6:e19479.

Drummond, A. J., A. Rambaut, B. Shapiro, and O. Pybus. 2005. Bayesian coalescent inference of past population

dynamics from molecular sequences. Mol. Biol. Evol. 22:1185–1192.
Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29:1969–1973.
Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evol. Bioinform. Online 1:47–50.
Fitzpatrick, B. M., and M. Turelli. 2006. The geography of mammalian speciation: mixed signals from phylogenies and range maps. Evolution 60:601–615.
Fitzpatrick, B. M., J. A. Fordyce, and S. Gavrilets. 2009 Pattern, process and geographic modes of speciation. J. Evol. Biol., 22:2342–2347.
Fu, Y. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915–925.
Gillespie, R. G., and G. K. Roderick. 2002. Arthropods on islands: colonization, speciation, and conservation. Ann. Rev. Entomol. 47:595–632.
Goloboff, P. A.. 1999. Analyzing large data sets in reasonable times: solutions for composite optima. Cladistics 15:415–428.
Goloboff, P. A., J. S. Farris, and K. C. Nixon. 2008. *TNT*, a free program for *phylogenetic* analysis. Cladistics 24:774–786.
Hall, J. P.. 2005. Montane speciation patterns in Ithomiola butterflies (Lepidoptera: Rhiodinidae). Proc. Roy. Soc. B Biol. Sci., 272:2457–2466.
Hall, R. 2011. Australia-SE Asia collision: plate tectonics and crustal flow. Pp. 75–109 *in* R. Hall, M. A. Cottam and M. E. J. Wilson, eds. The SE Asian Gateway: History and Tectonics of Australia-Asia Collision. Geol. Soc. Lond., London.
Harpending, H. 1994. Signature of ancient population growth in a low resolution mitochondrial DNA mismatch distribution. Human Biol., 66:591–600.
Hawlitschek, O., L. Hendrich, M. Espeland, E. F. A. Toussaint, M. J. Genner, and M. Balke. 2012. Pleistocene climate change promoted rapid diversification of aquatic invertebrates in Southeast Australia. BMC Evol. Biol. 12:142.
Heald, W. 1967. Sky Island. Van Nostrand, Princeton, NJ.
Heled, J., and A. J. Drummond. 2008. Bayesian inference of population size history from multiple loci. BMC Evol. Biol. 8:289.
Hewitt, G. 2000. The genetic legacy of the quaternary ice ages. Nature, 405:907–913.
Ho, S. Y. W., and M. J. Phillips. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Syst. Biol. 58:367–380.
Ho, S. Y. W., and B. Shapiro. 2011. Skyline-plot methods for estimating demographic history from nucleotide sequences. Mol. Ecol. Resour. 11:423–434.
Hope, G., A. P. Kershaw, S. van der Kaars, S. Xiangjun, P. Liew, L. E. Heusser, et al. 2004. History of vegetation and habitat change in the Austral-Asian region. Quaternary International, 118-119 10:3–126.
Johannesson, K. 2010. Are we analyzing speciation without prejudice? Ann. NY. Acad. Sci. 1206:143–149.
Joseph, L., and K. E. Omland. 2009. ROWLEY REVIEW Phylogeography: its development and impact in Australo-Papuan ornithology with special reference to paraphyly in Australian birds. Emu 109:1–23.
Kass, R., and A. Raftery. 1995. Bayes factors and model uncertainty. J. Amer. Statist. Assoc. 90:773–795.
Klaus, S., S. Selvandran, J. W. Goh, D. Wowor, D. Brandis, P. Koller, et al. 2013. Out of Borneo : Neogene diversification of Sundaic freshwater crabs (Crustacea: Brachyura: Gecacinucidae: *Parathelphusa*). J. Biogeogr. 40:63–74.
Librado, P., and J. Rozas. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452.
Lynch, J. D. 1989. The gauge of speciation: on the frequencies of modes of speciation. Pp. 527–553 *in* D. Otte and J. A. Endler, eds. Speciation and Its Consequences . Sinauer Associates, Sunderland, Mass.
Mayr, E. 1963. Animal Species and Evolution. Harvard University Press, Cambridge.
Mayr, E. 1982. The growth of biological thought: diversity, evolution, and inheritance. Belknap Press, Cambridge, MA.
Mayr, E., and J. M. Diamond. 1976. Birds on islands in the sky: origin of the montane avifauna of northern Melanesia. Proc. Nat. Acad. Sci. USA 73:1765–1769.
Mayr, E., and J. M. Diamond. 2001. The Birds of Northern Melanesia: Speciation, Ecology and Biogeography. Oxford Univ. Press, New York.
McCulloch, G. A., G. P. Wallis, and J. M. Waters. 2010. Onset of glaciation drove simultaneous vicariant isolation of alpine insects in New Zealand. Evolution 64:2033–2043.
Metcalfe, I.. 2011. Tectonic framework and Phanerozoic evolution of Sundaland. Gondwana Res. 19:3–21.
Miller, K. G., M. A. Kominz, J. V. Browning, J. D. Wright, G. S. Mountain, M. E. Katz, et al. 2005. The Phanerozoic record of global sea-level change. Science 310:1293–1298.
Mittermeier, R. A., P. R. Gil, M. Hoffmann, J. Pilgrim, T. Brooks, C. G. Mittermeier, et al. 2004. Hotspots revisited: Earth's Biologically Richest and Most Threatened Ecoregions. CEMEX, Mexico City, Mexico.
Posada, D. 2008. JModelTest: Phylogenetic Model Averaging. Mol. Biol. Evol. 25:1253–1256.
Rambaut, A., and A. J. Drummond. 2007. Tracer v1.4, Available at: http://beast.bio.ed.ac.uk/Tracer.
Ribera, I., A. Castro, J. A. Díaz-Pazos, J. Garrido, A. Izquierdo, M. A. Jäch, et al. 2011. The geography of speciation in narrow range endemics of the "*Haenydra*" lineage (Coleoptera, Hydraenidae, *Hydraena*). J. Biogeogr. 38:502–516.
Robin, V. V., A. Sinha, and U. Ramakrishnan. 2010. Ancient geographical gaps and paleo-climate shape the phylogeography of an endemic bird in the sky islands of southern India. PLoS ONE 5:e13321.

Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.

Rull, V. 2011. Neotropical biodiversity: Timing and potential drivers. Trends Ecol. Evol. 26:508–513.

Santini, F., M. P. Miglietta, and A. Faucci. 2012. Speciation: where are we now? An introduction to a special issue on speciation. Evol. Biol. 39:141–147.

Schneider, S., and L. Excoffier. 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates very among sites: application to human mitochondrial DNA. Genetics, 152:1079–1089.

Schultheis, A. S., J. Y. Booth, L. R. Perlmutte, J. E. Bond, and A. L. Sheldon. 2012. Phylogeography and species biogeography of montane great basin stoneflies. Mol. Ecol. 21:3325–3340.

Simon, C., F. Frati, A. T. Beckenbach, B. Crespi, H. Liu, and P. Flook. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Am. 87:651–701.

Smith, C. I., and B. D. Farrell. 2005. Phylogeography of the longhorn cactus beetle *Moneilema appressum* LeConte (Coleoptera: Cerambycidae): was the differentiation of the Madrean sky islands driven by Pleistocene climate changes? Mol. Ecol. 14:3049–3065.

Stamatakis, A.. 2006. raxml-vi-hpc: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

Swofford, D. L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595.

Teacher, A. G. F., and D. J. Griffiths. 2011. HapStar: automated haplotype network layout and visualization. Mol. Ecol. Resour. 11:151–153.

Toussaint, E. F. A., F. L. Condamine, G. J. Kergoat, C. Capdevielle-Dulac, J. Barbut, J. Silvain, et al. 2012. Palaeoenvironmental shifts drove the adaptive radiation of a noctuid stemborer tribe (Lepidoptera, Noctuidae, Apameini) in the Miocene. PLoS ONE 7:e41377.

Trewick, S. A., G. P. Wallis, and M. Morgan-Richards. 2000. Phylogeographical pattern correlates with Pliocene mountain building in the alpine scree weta (Orthoptera, Anostostomatidae). Mol. Ecol. 9:657–666.

Verstappen, H. T. 1997. The effect of climatic change on Southeast Asian geomorphology. J. Quaternary Sci. Rev. 12:413–418.

Von Rintelen, T., A. B. Wilson, A. Meyer, and M. Glaubrecht. 2004. Escalation and trophic specialization drive adaptive radiation of freshwater gastropods in ancient lakes on Sulawesi, Indonesia. Proc. Roy. Soc. B Biol. Sci. 271:2541–2549.

Wallace, A. R. 1860. On the Zoological Geography of the Malay Archipelago. Journal of the Proceedings of the Linnean Society of London, Zoology, 4:172–184.

White, M. J. D. 1968. Models of speciation. Science, 159:1065–1070.

Wilson, E. O. 1959. Adaptive shift and dispersal in a tropical ant fauna. Evolution, 13:122–144.

Wilson, E. O. 1961. The nature of the taxon cycle in the Melanesian ant fauna. American Naturalist, 95:169–193.

Winkler, I. S., C. Mitter, and S. J. Scheffer. 2009. Repeated climate-linked host shifts have promoted diversification in a temperate clade of leaf-mining flies. Proc. Natl Acad. Sci. USA 106:18103–18108.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Best neighbour joining topology.
Figure S2. Best maximum parsimony topology.
Figure S3. Best maximum likelihood topology.

8.2 The towering orogeny of New Guinea as a trigger of arthropod megadiversity

Citation: Toussaint EFA, Hall R, Monaghan MT, Sagata K, Ibalim S, Shaverdo HV, Vogler AP, Pons J, Balke M (2014) The towering orogeny of New Guinea as a trigger of arthropod megadiversity. *Nature Communication*, 5:ncomms5001.

nature COMMUNICATIONS

ARTICLE

Received 7 Nov 2013 | Accepted 29 Apr 2014 | Published 29 May 2014

DOI: 10.1038/ncomms5001

The towering orogeny of New Guinea as a trigger for arthropod megadiversity

Emmanuel F.A. Toussaint[1], Robert Hall[2], Michael T. Monaghan[3], Katayo Sagata[4], Sentiko Ibalim[5], Helena V. Shaverdo[6], Alfried P. Vogler[7,8], Joan Pons[9] & Michael Balke[1,10]

Early studies on Melanesian mountain systems provided insights for fundamental evolutionary and ecological concepts. These island-like systems are thought to provide opportunities in the form of newly formed, competition-free niches. Here we show that a hyperdiverse radiation of freshwater arthropods originated in the emerging central New Guinea orogen, out of Australia, about 10 million years ago. Further diversification was mainly allopatric, with repeated more recent colonization of lowlands as they emerged in the form of colliding oceanic island arcs, continental fragments and the Papuan Peninsula, as well as recolonization of the central orogen. We unveil a constant and ongoing process of lineage accumulation while the carrying capacity of the island is about to be reached, suggesting that lineage diversification speed now exceeds that of landmass/new ecological opportunity formation. Therefore, the central orogeny of New Guinea acts as a motor of diversification for the entire region.

[1] Zoologische Staatssammlung München, Münchhausenstraße 21, D-81247 Munich, Germany. [2] Southeast Asia Research Group, Royal Holloway University of London, Egham, Surrey TW20 0EX, UK. [3] Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Müggelseedamm 301, 12587 Berlin, Germany. [4] Papua New Guinea Institute for Biological Research, 441 Goroka, Papua New Guinea. [5] New Guinea Binatung Research Centre, 511 Madang, Papua New Guinea. [6] Naturhistorisches Museum, Burgring 7, A-1010 Vienna, Austria. [7] Department of Life Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK. [8] Division of Biology, Imperial College London, Silwood Park Campus, Ascot SL5 7PY, UK. [9] IMEDEA (CSIC-UIB), Mediterranean Institute for Advanced Studies, c/Miquel Marquès 21, 07190 Esporles, Spain. [10] GeoBioCenter, Ludwig-Maximilians-University, 80333 Munich, Germany. Correspondence and requests for materials should be addressed to E.F.A.T. (email: Toussaint@zsm.mwn.de).

Tropical mountains are highly diverse natural laboratories that provide an ideal setting for the study of macro-evolutionary and ecological interactions during lineage diversification[1,2]. The tropical island of New Guinea in particular is exceedingly diverse[3], and naturalists ever since Wallace[4] have used the island as a natural laboratory to study the evolution of species and communities[5–7].

The second largest island on Earth, New Guinea (>800,000 km^2), is situated in the tropics but contains an extraordinary diversity of climate zones and landforms. These range from lowland plains with a seasonal climate to alpine zones with permanent ice on high peaks (>4,700 m above sea level). The island contains many mountain ranges and some of the world's last remaining tropical wilderness. Remarkably, most of this landmass was formed in the past 5 million years (Myr), resulting from major mountain building and what is arguably the world's most complex geotectonic history[8], further modified by extensive volcanism and glaciations[9]. Before 5 million years ago (Ma), and for much of the Cenozoic, the New Guinea region is thought to have been composed of many islands of varying geological origin (Fig. 1). This archipelago structure played an important role in the local radiation of rainbow fishes[10] and in the global evolution of a major group of songbirds[11], both *c.* 50–20 Ma. Speciation events resulting from island arc collisions and orogenies have been identified as key factors explaining the high biodiversity in Melanesia and in New Guinea in particular (general[3], Aves[6], Chelonii[12], Odonata[13], Hemiptera[14], Heteroptera[15]).

Increasingly detailed paleotectonic data and more extensive, species-level molecular phylogenies provide the means to test different spatio-temporal and ecological scenarios for the rise and diversification of New Guinea biodiversity. The general approach has been used for studies of birds[10,16–18], fishes[10] and mammals[19–23]. Results suggest that species-level diversification within New Guinea has been recent (<5 Ma), corroborating geological evidence that dates substantial landmass formation to <10 Ma[24,25]. Lowland vertebrate taxa tend to exhibit north–south divergence on either side of the predominant east–west cordillera[10,16,26]. Montane taxa more commonly exhibit east–west splits, thought to result from more local allopatric speciation among drainages along the cordillera[27].

Although New Guinea remains a region where comparatively few geological studies have been carried out (due to its difficult terrain, remote location and climate), plate tectonic models for the development of New Guinea have been proposed recently based on the evidence available. A model proposed by van Ufford and Cloos[28] suggested that an underthrusting of the Australian continent beneath an Inner Melanesian arc resulted in an orogeny restricted to eastern New Guinea *c.* 35–30 Ma. A later orogeny (from 15 Ma in the west to 3 Ma in the east) then gave rise to the Central Range at its present elevation *c.* 5 Ma. Biogeographically, this model[28] would suggest the Papuan Peninsula in eastern New Guinea to be an area of early lineage diversification[13], with successive lineages arising along the central range out of the Papuan Peninsula. In contrast, Hall[29], and Hill and Hall[30] proposed that convergence between the Pacific and Australian plates *c.* 5 Ma formed a fold-and-thrust belt and led to rapid rise of the central ranges that continues today. This model implies that New Guinea was largely submerged until the Early Pliocene and that formation of the present large emergent area occurred in the last 5 Ma[24]. Under this model, the expected biogeographic pattern would be rather divergent from the van Ufford and Cloos model[28] with an expected early diversification on the fold and thrust belts corresponding to present-day central orogen, followed by colonization of surrounding areas such as the oceanic terranes drifting from the North, the Bird's Head and the Papuan Peninsula.

Despite their overwhelming contribution to animal diversity, studies of invertebrate diversification on New Guinea remain in their infancy (but seerefs 7,31). In contrast to vertebrate diversification in the recent past on the landmass, invertebrate speciation on ancient island arcs is thought to be important in shaping extant species richness and distribution patterns[13,14] even though alternative hypotheses implying Quaternary diversification on the island exist[7]. Nonetheless, the macro-evolutionary and macroecological drivers of their diversification

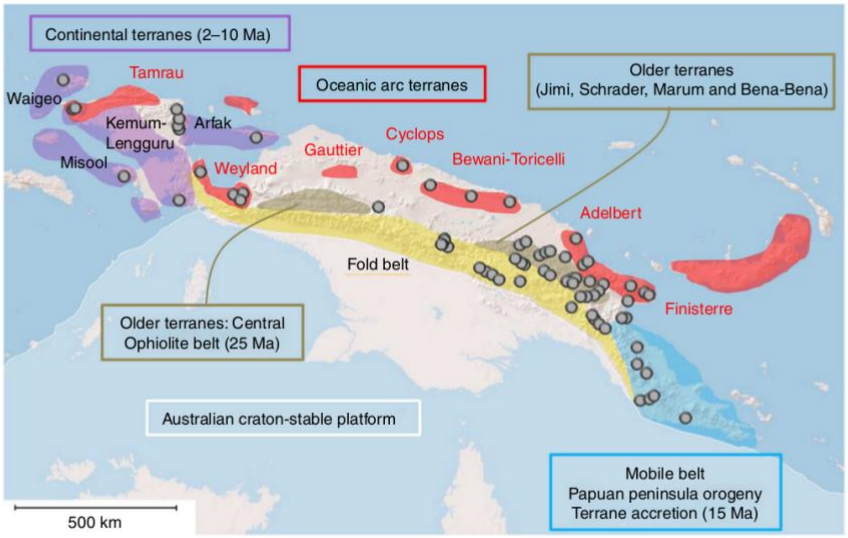


Figure 1 | Map of present-day New Guinea with major tectonic features and collecting localities for sequenced *Exocelina* specimens. Filled circles indicate localities from which at least one specimen of *Exocelina* was collected. Purple shading indicates continental terranes, red indicates oceanic arc terranes, brown indicates old terranes, and bright blue and yellow areas indicate the mobile and fold belts, respectively. The wide white area represents the Australian craton-stable platform.

remain little understood, and species-level patterns have yet to be tested in a phylogenetic framework.

Here we combine dense sampling of a single lineage of predaceous diving beetles (Dytiscidae, *Exocelina*) with molecular phylogenetics and sophisticated biogeographical and diversification analyses to test alternative geological scenarios and examine how extensive environmental change and very recent formation of most New Guinea terrestrial habitats have interacted to promote lineage diversification. We reconstruct a dated species-level phylogeny to analyse ancestral states and rates of lineage diversification and use this to test alternative hypotheses of (i) older island arc evolution[14,15] and (ii) early mountain building in the present Papuan Peninsula[13] of eastern New Guinea versus (iii) evolution and diversification in a more recent central orogen[29,30] with subsequent colonization of other areas in the New Guinea region. Our findings indicate that lineage diversification was comparatively recent, driven by the formation and subsequent modification of the central highlands, and that all other areas of New Guinea were colonized out of that region. Rather than being unidirectional, diversification into lower altitudes was reversed several times and included secondary highland radiation. We conclude that lineage diversification is a recent, complex, and dynamic process, in our case not closely linked to terrane movement as has been largely assumed.

Results

Molecular phylogenetics. The final concatenated alignment comprised 4,299 bps sequenced for 94 New Guinean *Exocelina* diving beetle species: 1,593 bps of mitochondrial *cytochrome c oxidase I* (*cox1*), *cytochrome c oxidase II* (*cox2*) and *cytochrome b* (*cob*) gene fragments, in addition to 2,706 bps of the nuclear *histone 3* (*H3*), *histone 4* (*H4*), *18S rRNA* (*18S*), *Carbomoylphosphate synthase* (*CAD*) and *Alpha-Spectrin* (*Asp*) gene fragments. The Bayesian phylogenetic analysis based on the combined data set and the best-fitting partitioning scheme (Tables 1 and 2) yielded a highly resolved and strongly supported tree highly compatible with the tree generated in the RAxML analysis (Supplementary Figs 1 and 2).

We recover a succession of branching events at the deepest level in the tree and six major clades in a more derived position (Fig. 2), for example, clade 6, corresponding to the *Exocelina ekari* group[32]. Sequence divergence was low, for example, in *cox1* the smallest interspecific uncorrected *P* distance was only 1.7% (s.d., 1.48) and the mean interspecific distance was 5.82%. Overall, interspecific divergence ranged from 0–9.2%. This was reflected in a shallow branching pattern at the tips as well as species clustering, which even at 1% threshold only recovered 69 clusters instead of the expected 94 due to lumping of taxonomic species. These results pinpoint that several species diverge by <1%: the parapatric *E. atowaso* + *E. astrophallus* by (0.67%, despite divergent male genital morphology), the allopatric *E. wondiwoiensis* + *E. irianensis* (0.13%, male genital similar) and the sympatric (not yet found syntopic) *E. weylandensis* + *E. soppi* (0.14%, male genital moderately divergent)[32].

Divergence times and diversification. The chronogram derived from the relaxed molecular clock analyses of *cox1* suggested a colonization of New Guinea occurring during the late Miocene (median age 8.2 Ma, 95% HPD 6.2–10.7 Ma; Fig. 2 and Supplementary Fig. 3) with most of the branching events occurring in the Plio-Pleistocene (5.3–0 Ma). The *TreePar* diversity-dependent analyses carried out on the chronogram support a birth–death model with no diversification rate shift during the evolution of the radiation. This model corresponds to a pattern of constant accumulation of lineages through time with a rather high rate of diversification ($r = 0.3950$) (Table 3). The density-dependent analyses indicate that the maximum carrying capacity K is about to be reached (150/169) and that therefore the radiation will soon attain equilibrium. The lineage-through-time

Table 1 | Partitioning schemes for the Bayesian inference analyses.

Partition	Composition
P1[1]	All genes combined
P2[2]	Mitochondrial + nuclear genes
P3[2]	Coding versus non-coding genes
P4[4]	18S + 1st cp. of coding genes + 2nd cp. of coding genes + 3rd cp. of coding genes
P5[7]	18S + 1st cp. of mitochondrial coding genes + 2nd cp. of mitochondrial coding genes + 3rd cp. of mitochondrial coding genes + 1st cp. of nuclear coding genes + 2nd cp. of nuclear coding genes + 3rd cp. of nuclear coding genes
P6[8]	One partition per gene
P7[18]	PartitionFinder scheme
P8[22]	18S + 1st cp. of each coding gene + 2nd cp. of each coding gene + 3rd cp. of each coding gene

cp., codon position.
Note: the number of partitions in each partitioning scheme is given in square bracketscp., codon position.

Table 2 | Best-fit partitioning strategies for the BI analyses with Bayes factors (BF) estimates.

	Part.	ESS	SSML	P1	P2	P3	P4	P5	P6	P7	P8
P1	1	5,919	−30,199.6	—	*	*	*	*	*	*	*
P2	2	2,136	−29,705.6	**	—	*	*	*	*	*	*
P3	2	1,598	−29,217.9	**	**	—	*	*	*	*	*
P4	4	4,433	−28,762.5	**	**	**	—	*	**	*	*
P5	7	2,013	−28,239.2	**	**	**	**	—	**	*	*
P6	8	270	−29,156.9	**	**	**	*	*	—	*	*
P7	18	1,019	−27,834.9	**	**	**	**	**	**	—	**
P8	22	357	−27,927.8	**	**	**	**	**	**	*	—

Part., number of partitions; ESS, effective sample size; SSML, stepping-stone marginal likelihood; *, $2 \times \ln(BF) < 1$; **, $2 \times \ln(BF) > 10$.

(LTT) plots inferred using either the BEAST incomplete phylogeny or 1,000 simulated phylogenies accounting for missing taxon sampling highlight the pattern of constant accumulation of *Exocelina* lineages during the evolution of the radiation (Fig. 3).

Ancestral state reconstructions. For the altitudinal species distribution, the first three nodes of the tree backbone had strong support (posterior probability (PP) ≥ 0.95) for the '≥ 500 m' state, followed by '$\geq 1{,}500$ m' at the 4th node. The earliest lineages in the evolution of the radiation, including clade 1, have independently colonized lowland and highland parts of the island from an initial montane zone during the past 8 Myr. Clades 2 and 3 contain species from the uppermost distributional limits, 2,000–2,800 m. Lower altitude colonization is a more recent event, independently occurring in clade 5 and with significant diversification in clade 6.

The results of the ancestral character state optimization for the horizontal origin trait recovered a central orogen origin with strong support (PP = 1.0). This character state also was recovered for most of the internal nodes with strong support (PP $\geq$ 0.95). Multiple shifts from central orogen to north coast range terranes (NCR) from the early Pleistocene were recovered with strong support (PP = 1.0), although the ancient lineage *E.* undescribed species MB1530 now occurs in the NCR. The colonization of the Bird's Head terranes (BHT) had a comparable timing in clade 6D, with subsequent radiation; one species (*E.* undescribed species MB1269, clade 6A) colonized the BHT earlier. Topological patterns congruent with speciation along an island arc were only recovered for the three NCR species in clade 1 (Fig. 2), with one species each occurring in the Cyclops, Bewani and Adelbert Mountains (Fig. 1). Their diversification is here inferred from the Pleistocene. The supposedly very old Papuan Peninsular orogen was colonized several times from western parts of the central highlands (Fig. 2).

Competitive exclusion. On the basis of our preliminary co-occurrence matrix, we found that up to five species were sampled from the same habitat (puddle or stream segment). Relative abundances appeared to differ although precise quantification was outside the scope of our collecting regime. In general, syntopic species were not closely related. Only In the *E. ekari* clade 6E did we find several closely related species syntopically, for example, *E. weylandensis*, *E. soppi*, *E. kakapupu* and *E. utowaensis* (Supplementary Fig. 4). Local diversity appears mostly to result from repeated colonization by different lineages. An example is the syntopic occurrence of highland species from the older clade 2 with species of clade 6C that have recently colonized the highlands from lower elevations.

Discussion

New Guinea is situated at the convergent boundary of the Australian and Pacific plates. Present-day New Guinea is a geologically young landmass of heterogeneous origin, composed of many terranes[33], including obducted ophiolites, accreted oceanic island arcs, continental slivers and the Australian continental margin[8,24]. In general terms, southern New Guinea and several sections of the western New Guinea (for example, the Bird's Head peninsula) are, or have been, formed from parts of the Australian plate (Fig. 1). The spine of New Guinea is a 1,300-km-long and up to 150-km-wide central highland chain. It includes a major fold-and-thrust belt in the central range that represents the deformed passive margin of the Australian continent, to the north of which are ophiolite belts (oceanic or arc lithosphere displaced during island arc–continent collision) and accreted island arc terranes. The ophiolites and arc terranes have been described as part of a mobile belt[30,34] in which there was significant deformation during the Neogene since $\sim$25 Ma.

Early studies interpreted New Guinea in terms of terranes[33,35], which are fault-bounded crustal fragments each with its own geological character, and suggested that at least 32 terranes had been added to the Australian margin in a series of collisions linked to subduction during the Cenozoic. A tentative plate tectonic model based on this terrane concept was outlined by Struckmeyer *et al.*[36] Later models[28–30,34] viewed New Guinea development in a more comprehensive plate tectonic framework, although there remain considerable differences between different models in terms of timing, events and inferred plate reconstructions.

van Ufford and Cloos[28] proposed a model which included two intra-oceanic arc systems north of New Guinea subducting discontinuously to the north and south from the Eocene to Middle Miocene with numerous small subduction zones developed since. In their model there is no role for the Philippine Sea plate and they interpreted several small plates between the Pacific and Australian plates. They suggested that at 35–30 Ma underthrusting of the Australian continent beneath their Inner Melanesian arc resulted in a peninsular orogeny. This orogeny was restricted to eastern New Guinea and initiated uplift and emergence of the Papuan Peninsula. Much later, the Central Range orogeny commenced 15 Ma in the west and up to 3 Ma later in the east, and was proposed to result from contractional thickening of passive margin strata and underthrusting of Australian continental basement. According to this model deformation began at the distal northern edge of the Australian passive margin where sediment cover was removed from oceanic or transitional crust, and collisional orogenesis involving continental crystalline basement began at 8 Ma. van Ufford and Cloos[28] observed a dramatic change in the coarseness and extent of clastic sediments, which they attribute to the rise of the Central Range to its present elevation at about 5 Ma.

Hall[29], and Hill and Hall[30] proposed a very different model. They suggested there was subduction during the Paleogene beneath a Philippines–Halmahera–Caroline intra-oceanic arc system after the Australian plate began to move rapidly northwards in the Eocene. Collision with the Philippines–

Figure 2 | Dated Bayesian phylogeny of the New Guinean *Exocelina* radiation and paleotectonic evolution of the New Guinean archipelago. A time-scale is indicated spanning the full evolutionary history of the group. Asterisks above the nodes indicate strong support (PP $\geq$ 0.95) for the reconstruction of the ancestral altitude state. Pastilles code for vertical distribution according to legend; purple stars = distribution on continental Bird's Head terranes; red diamonds = north coast terranes; blue triangles = Papuan Peninsula. Coloured branches indicate the reconstruction of the ancestral horizontal distribution state (black branches = central orogen). Panels **a**–**c** on the left side show maps of the distribution of land and sea at respectively 10 Ma, 5 Ma and at present after ref. 21 (green, land; dark blue, deep sea; lighter blue, shallow sea; red white brick, calcareous plateaus possibly exposed at times; orange, highland; grey, high altitude above 2,800 m). All maps are at the same scale indicated in **a**. Below is a schematic summary in a South-to-North orientation of major tectonic processes from proto- to present day New Guinea. Panels **d**–**f** highlight the orogenic dynamics that took place during the evolution of the *Exocelina* radiation respectively between 12–8 Ma, 8–4 Ma and 4–0 Ma. Far right in purple: drifting and colliding continental Bird's Head terranes; front in red: north coast terranes; central orogen with altitudinal zones colour codes as on the tree, Papuan Peninsula in blue at the far left. Numbers and letters in the tree refer to the major clades and subclades of the radiation.

Caroline arc commenced in the Late Oligocene at ~25 Ma. This caused a change to sinistral oblique convergence with the Philippine Sea and Pacific–Caroline Plates, which resulted in Paleogene arc fragments being displaced westwards along the New Guinea margin and sliced into many smaller terranes. The New Guinea Mobile Belt thus represents the complex Neogene

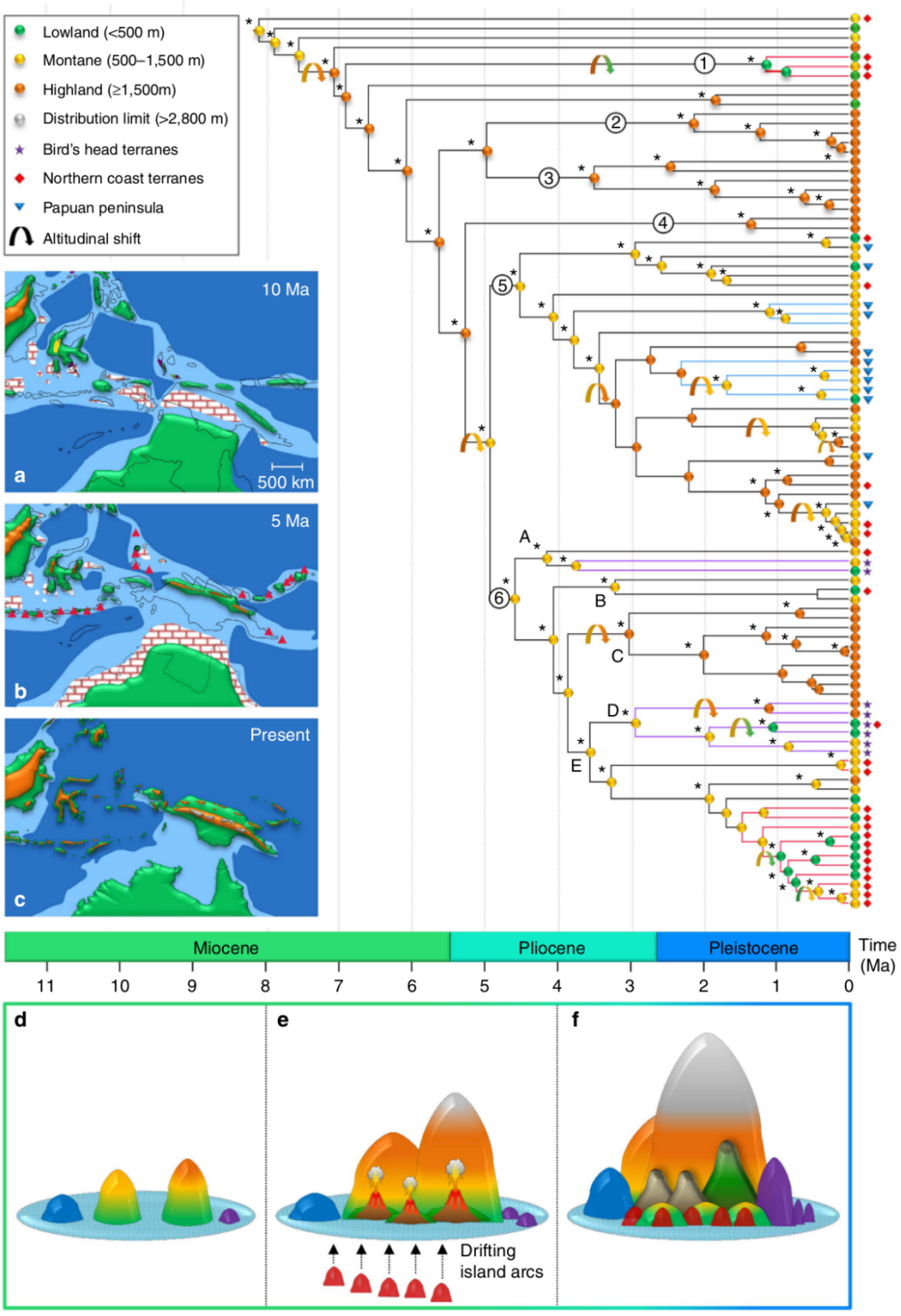


Table 3 | Results from the *TreePar* analyses conducted on the BEAST chronogram.

Model	Pm	− logL	*P*	r1	τ1	st1	r2	τ2	st2	r3	τ3	st3	r4	τ4	st4	r5	τ5
BD constant	**2**	**103.7828**	**—**	**0.3950**	**0.4904**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**
BD 1 shift	5	101.9753	0.3062	0.5861	0.2858	1.2000	0.4652	0.0000	NA	NA	NA	NA	NA	NA	NA	NA	NA
BD 2 shifts	8	99.3799	0.1583	0.4486	0.3795	1.1000	0.6185	0.6954	1.2000	0.4803	0.0000	NA	NA	NA	NA	NA	NA
BD 3 shifts	11	98.3047	0.5418	0.3268	0.6018	1.1000	0.3996	0.8472	1.2000	0.2856	0.0980	1.6000	0.4734	0.3432	NA	NA	NA
BD 4 shifts	14	97.3444	0.5089	0.3037	0.6410	1.1000	0.3960	0.8480	1.2000	0.3084	0.1730	1.6000	0.0001	0.9999	2.9000	0.4112	0.7572

Pm, number of parameters in the model; − LogL, the log-likelihood of the model; *P*, *P*-value of the likelihood ratio test between the incrementally more complex models (if $P<0.05$ the model is supported); r1, diversification rate at present; τ1, turnover rate at present; st1, most recent shift time. Other diversification and turnover rates, as well as shift times, going deeper in the past are denoted with numbers (for example, r2, τ2 and st2). The best-fit model is underlined in bold.

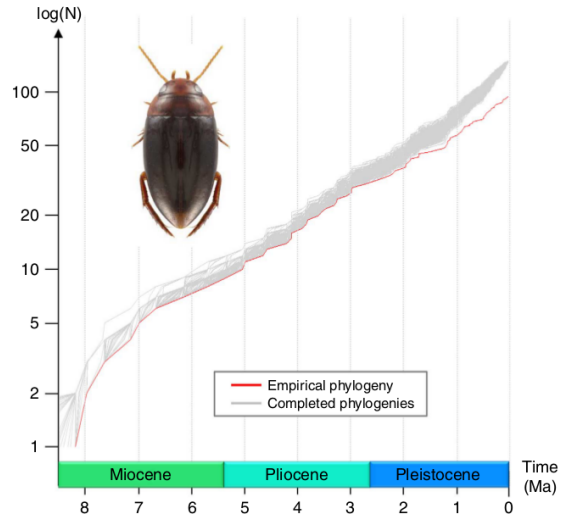


Figure 3 | LTT plots of the New Guinean *Exocelina* radiation. A time-scale is indicated spanning the full evolutionary history of the group. The vertical axis shows the number of species in logarithmic scale. The LTT plots of the New Guinean *Exocelina* (*E. unipo* is presented) have been generated using the BEAST incomplete phylogeny (red line) and 1,000 simulated complete phylogenies (grey lines) to illustrate the constant diversification of the radiation fostered by the orogeny of the island.

strike-slip zone to the north of the Australian passive margin with over 1,000 km sinistral displacement that has alternated between extension and compression. It is in this region that there were possibly small emergent areas during the Miocene. A change in plate motions at about 5 Ma initiated convergence between the Pacific and Australian plates and caused the formation of the fold and thrust belt and rapid rise of the central ranges that continues today. This model implies that New Guinea was largely submerged, except possibly for parts of the New Guinea Mobile Belt, until the Early Pliocene and that the formation of the present large emergent area occurred in the last 5 Ma[24], as mentioned above.

In the space of this paper it is not possible to discuss the reasoning and conclusions of these different geological models for New Guinea except to agree that '…the age, number and plate kinematics of the events that formed the island are vigorously argued'[28]. Nonetheless, despite significant differences in interpretation of events and plate tectonic reconstruction, the two models[28–30] have some features in common as far as the distribution of land is concerned. From the Oligocene until the Late Miocene there was likely to have been only small areas of land at the Pacific margin north of New Guinea. Disconnected islands may have formed where there were volcanic arcs, at local uplifts near to plate boundaries and in the Mobile Belt. There was shelf carbonate deposition along much of northern Australian margin in what are now the New Guinea ranges. From the Late Miocene there is evidence of increased siliciclastic sediment input, indicating erosion of land areas, and both models infer that massive and rapid uplift from 5 Ma formed the highest summits (3,000–5,000 m), accompanied by steady widening of the central ranges. Since 5 Ma, strike-slip movements in the northern part of New Guinea have also contributed to the current topography. Although precise timing is uncertain, it is clear that many parts of northern New Guinea, from the Bird's Head to New Britain, emerged above sea level in this period[24,25,28,30]. The major differences between the two models for biogeographers are that the first model[28] would suggest the Papuan Peninsula as an area of early lineage diversification and a successive phylogenetic tree branching pattern along the central range in an east–west direction, whereas the alternative[29,30] implies much younger colonization and diversification, centred on the Central Range.

Our results, based on the most extensive molecular phylogenetic and biogeographic species-level analysis for any New Guinean taxon to date, suggest a recent origin of New Guinean *Exocelina* during the late Miocene (~8.2 Ma) and a constant process of lineage diversification. The ongoing diversification suggests species have continuously colonized the plentiful, newly formed, biologically empty habitat. However, as highlighted by the calculated carrying capacity of the radiation, the diversification process might experience a slowdown in the next million years, corresponding to a saturation of ecological niches and habitat on the island. This pattern would imply that the rate of diversification will eventually exceed the pace of ecological opportunity formation fostered by the ongoing orogeny of the island.

The ancestral character state reconstructions revealed two important ancestral traits: an early occurrence at montane elevation, and an origin in the central orogen. In a geologically dynamic landscape like New Guinea, it is however difficult to infer ancestral altitudinal preference, because we have data only from present-day distributions. Ancestral New Guinea *Exocelina* might have colonized lower altitudes, diversified and undergone passive uplift to the high altitudes. They may also have colonized higher altitudes and migrated into lower areas as compensation for uplift into zones with colder temperatures. We suggest a complex mixture of both scenarios, as we uncovered several altitudinal shifts in our analysis. Ancestral Australian *Exocelina* are lowland species[37], which suggests lowlands may be the ancestral New Guinea habitat.

During mountain uplift along the northern margin of the Australian plate, an initial colonization out of Australia[37] would have reached emerging islands that might have had some elevation already. At that stage, the emerging orogen would have been an insular setting supporting lineage diversification, analogous to the initial branching found in rainbow fishes[10]. Strong landscape changes during the uplift as well as significant

amounts of volcanism during the late Neogene followed by Quaternary climate change[9] likely shifted environmental conditions. These changes would have further promoted the isolation of populations and fuelled lineage diversification[7,12,14]. The extremely structured central highland chain itself, with ongoing formation of rich aquatic resources in particular during the formation of extensive foothill chains, provided the setting for random colonization of new areas followed by isolation and speciation in remote valleys or mountain blocks[38]. Interestingly, New Guinean *Exocelina* are almost exclusively associated to running waters, and there is evidence that species in such habitats are weak dispersers. This trait has been suggested to enhance allopatric speciation and micro-endemism in diving beetles[39], and could therefore represent one of the underlying mechanisms fostering this astonishing radiation. More generally, tropical species are thought to be adapted to rather narrow climatic conditions because they do not experience seasons[40]. Colonization of new vertical bands might therefore be rare enough to support isolation.

Others have proposed that the present-day diversity of the New Guinea fauna is the product of ancient geological processes and landmass collisions. Instead, we found that this was not the case. The terranes of the oceanic South Caroline arc that formed much of the NCR, and the continental terranes that formed the Bird's Head and satellite islands, have been colonized from the central orogen after these different landmasses attained some proximity or had docked already. Previous hypotheses assigned a central role to these terranes in the diversification of diverse arthropod lineages on drifting island arcs. The biota was proposed to have reached island arcs while adrift in the Pacific, with subsequent diversification along the island arc, on islands remote from each other and from the rest of the emerging New Guinea[15] (also refer to the 'Discussion' in ref. 12). Closely related species would have reached closer geographic proximity only after terrane collision[15]. Analytically, this would predict (1) a greater age of the species on fully insular arcs far from the rising orogen, 40–10 Ma, and (2) a tree topology with no closely related species in other parts on present day New Guinea. It also implies that islands arcs did indeed provide terrestrial ecosystems over a long period of time, which is far from being unambiguously proven (ref. 24, page 116).

Our results provide strong evidence for an alternative and more complex scenario, namely that recent environmental change in the central highlands has been the primary driver of diversification in New Guinea. The evidence from tree topology, ancestral range reconstruction and lineage ages do not support the hypothesis of an older, island-arc evolution on the NCR and Bird's Head terranes[14,15]. Instead, the data indicate repeated colonization in the recent past. Present-day alpha-diversity thus appears more related to colonization events than to local diversification in a given terrane. The pattern expected from an older, island-arc scenario was seen only in clade 1, where single species occurred in each of the Cyclops, Bewani and Adelbert Mountains. While the clade was comparatively old, the species in each mountain range appear to have arisen in the last 2 Myr. This suggests recent allopatric speciation along the north coast as opposed to on ancient islands adrift. The relative isolation of the Bird's Head[12] was also suggested by our findings. There have been only two colonization of *Exocelina*, both of which appear to be recent (Fig. 1). This does not lend support to hypotheses of evolution on terranes adrift. The lack of colonization of the Bird's Head may be attributed to a lack of suitable stream habitats in the karstic geology of the Lengguru area that connects the Bird's Head with mainland New Guinea. The sister species to all other New Guinea *Exocelina* occurred in the Adelbert Mountains of the NCR, but its origin was the central orogen according to the ancestral range reconstructions. Related species occur along the central orogen as well. The hypotheses of early mountain building in eastern New Guinea were also not supported by our biological data. *Exocelina* diving beetles have colonized the Papuan Peninsula up to the Herzog Mountains out of the central orogen at least six times in comparatively recent time.

Distribution at high altitude in clades 2 and 3 was usually allopatric, related species occurring on different mountains. Thus, *Exocelina* do not exhibit a pattern described by Diamond[6], who identified (ecological) montane speciation in the absence of ecotones as a source of New Guinea bird diversity, where sister species occupy sharply delineated altitudinal zones on the same mountain. Allopatry appears to be the main mechanism in *Exocelina,* in general along more or less the same altitudinal zone as far as can be seen from our sampling. This differs from speciation patterns observed in the few other studies of running water organisms, where there is evidence that speciation may follow the river course longitudinally in peripatric speciation. Spatial patterns of speciation in running waters have interested ecologists since Illies[41], who suggested that warm-adapted lineages of aquatic insects arose from cold-adapted ones, with evolution within river systems progressing downstream. Statzner and Dolédec[42] found empirical evidence for this based on the distribution of ecological traits and phylogenetic relationships among French *Hydropsyche* (Trichoptera) species. They suggested a headwater ancestor with primarily downstream evolution of the clade. In contrast, Malagasy mayflies (Ephemeroptera) appear to have diversified from lowland ancestors to colder and faster-flowing upstream sections[43]. The syntopic or near-syntopic presence of closely related *Exocelina* species in the Weyland area (Fig. 2, clade 6E; Supplementary Fig. 4) suggests that peripatric speciation mechanisms may also contribute to the observed species richness, but the consistency of the pattern remains to be tested using denser longitudinal sampling.

There are no detailed ecological studies of *Exocelina* species in New Guinea. We observed that the relatively recent clade 6 includes many lowland species adapted to peripheral habitats along fast-flowing streams. Species in clades 2 and 3 live in habitats with slower flows that are similar to the habitat of the older lineages of *Exocelina* that occur in Australia. This suggests diversification into new, more extreme habitats, with the result that more emerging habitats are being utilized. Here, too, denser sampling is needed to further study mechanisms at work, that is, possible niche segregation and different abundances among sites.

In summary, our extensive biological data set implicates recent diversification and repeated colonization of sites by distantly related lineages as the primary drivers of the diversity patterns found in New Guinea *Exocelina*. Despite the clear biological results, a number of questions remain. In terms of geology, the origin of the Weyland and Wandammen regions remain unresolved. The extent and configuration of land available for colonization is still uncertain, but the general setting summarized above provides the framework for investigations of the biological evolution of New Guinea. Data from biologists and a large selection of organisms could potentially inform geologists about land configurations in the past, supporting a truly integrative science.

Methods

Taxon sampling and phylogenetic analyses. Using standard protocols (Supplementary Table 1) we obtained sequences from 94 in-group species from across New Guinea and representing all known morphological groups (Supplementary Table 2)[32]. Three species of *Exocelina* from Australia and New Caledonia were included as close outgroups as well as two representatives of the

subfamily (Copelatinae); *Copelatus irregularis*, *Lacconectus peguensis* and *Thermonectus sp* (Dytiscinae) in order to root the tree. We sequenced fragments of mitochondrial *cox1* (735 bp used in our alignment), *cox2* (552 bp) and *cob* (306 bp), as well as nuclear *H3* (318 bp), *H4* (198 bp), *18S* (570 bp), *CAD* (828 bp) and *Asp* (792 bp). Sequences were edited using Geneious R6 (Biomatters, http://www.geneious.com), aligned with Muscle[44] and codon positions were determined using Mesquite 2.75 (http://mesquiteproject.org).

We used Bayesian inference (BI) and maximum likelihood (ML) on the concatenated data set containing one specimen per species and seven different partitioning strategies (Table 1) to account for expected differences in sequence evolution in different genes. The best model for each partition was selected under jModelTest 2.1.3 (ref. 45), using the corrected Akaike information criterion (AICc) (Supplementary Table 3). An additional scheme was tested based on the partitions selected in PartitionFinder 1.1.1 (ref. 46) under the AICc (Table 1). The BI analyses were run in MrBayes 3.2 (ref. 47) with two independent runs consisting of eight Metropolis-coupled Markov Chains Monte Carlo (MCMC, one cold and seven incrementally heated) running for 50 million generations and sampling every 1,000 cycles. The split-frequencies and log-likelihood curves were investigated before applying a conservative burn-in of 25%, and the remaining topologies were used to generate a 50% majority rule consensus tree. The best strategy of partitioning was selected afterwards based on Bayes factors (BF)[48] and effective sample size (ESS) criteria approximated under Tracer 1.5 (ref. 49). BF tests were based on marginal likelihoods calculated using stepping-stone sampling to account for harmonic mean unsuitability to deliver unbiased estimates[50,51]. BF values superior to 10 were considered good indicators of a significantly better partitioning scheme over another, and ESS values greater than 500 indicative of a good convergence of the runs[52]. The ML analyses were carried out with the best model found in BI under RAxML[53]. We performed 5,000 'thorough bootstrap' replicates (BS) to investigate the level of support for each node. Calculated PP values $\geq$0.95 and BS values $\geq$70 were considered to indicate strongly supported nodes[54,55].

Estimation of divergence times. In order to calibrate the topology and since the fossil record is scarce for diving beetles, we used the information of three recent publications on Coleoptera in which a divergence rate for the same *cytochrome c oxidase subunit 1* fragment used in this study was calculated, that is on dytiscid beetles[56] (mean rate = 0.0195 substitutions per site per million years per lineage, subs/s/Myr/l), on tenebrionid beetles[57] (mean rate = 0.0177 subs/s/Myr/l) and on carabid beetles[58] (mean rate = 0.0145 subs/s/Myr/l). We set the *ucld.mean* with a normal distribution encompassing the three mean rates recovered in these studies (0.0145–0.0195 subs/s/Myr/l). To test the putative clockwise evolution of the matrix, we used PAUP*[59] and calculated likelihood with and without a strict molecular clock. The likelihood ratio test conducted under the same software resulted in a P value $<$0.001, implying that the strict molecular clock assumption was not statistically supported. Therefore, we used a relaxed clock model allowing rate variation among lineages. We carried out the analyses using the BEAST 1.7.4 (ref. 60) with the following non-default settings and priors: the *Site Model* was chosen according to the models of evolution used in the phylogenetic analyses, the *Molecular Clock Model* was set to an estimated *Relaxed Clock: Uncorrelated Lognormal*, the *Tree Model* to a *Speciation: Birth Death Process* and the MCMC parameters were fixed to 30 million generations with sampling every 1,000 generations and the first 10% discarded as burn-in. The best topology recovered from the BI phylogenetic reconstructions was fixed by manually editing the .xml file generated in BEAUTI 1.7.4 (ref. 60).

Ancestral state reconstructions. The ancestral state reconstructions were performed using the *'Traits'* and *'Sates'* options in the BEAST 1.7.4 (ref. 60) on the chronogram from which we pruned all outgroups. For the reconstruction of the 'Altitude' character, each species was assigned to one of the following categories: 'Lower altitude $<$500 m', which would capture species in lowland-foothill forest across the island, warmer climate and often faster-flowing streams on steep foothill slopes; 'Montane 500–1,500 m' as the lower montane zone leading towards the actual highlands, often fast flowing; and 'Highland $\geq$1,500 m', or mid montane zone, which contains the large intra-montane depressions of New Guinea up to the montane cloud forests. These bands are following altitudinal zonation and to some degree generalized due to regional differences in zonation[61,62]. For the 'Horizontal distribution origin' character, we use a generalized scheme, according to the two major island arcs or terrane systems that have formed distinct, clear cut geographic features of the island: the Bird's Head Including the Satellite islands in the west and the NCR. Finally, the central mountain range as well as the inferred older terranes closely attached to it in the north, and which drop into lowlands, marking a comparably clear-cut transition towards the NCR. *E. brahminensis* and *E. astrophallus* occur in NCR as well as the basin between central range and NCR, and were assigned to NCR. *E.* cf. *brahminensis* was reported from the Herzog Mountains[32], northern Papuan Peninsula, but we had no sequence data from that locality and decided to code our NCR specimens as NCR only. All analyses were run until the ESS of each parameter reached 500 and 25% of the samples were discarded as burn-in before generating the maximum credibility tree in TREEANNOTATOR 1.7.4 (ref. 60).

Diversification analyses. We assessed the diversification pattern of the *Exocelina* radiation using the *ape*[63], *TreePar*[64] and *TreeSim*[65] packages for R and the BEAST chronogram from which we pruned all outgroups. We used the approach developed by Stadler[64] to estimate putative shifts in speciation and extinction rates in possibly incomplete phylogenies. We used the function *'bd.shifts.optim'* to estimate the ML speciation and extinction rates along with the possible shift times in the *Exocelina* radiation. As an input, the function requires the number of sampled taxa in the phylogeny as well as an estimation of the current species richness in the clade. The analyses were therefore run with the following settings: the taxon sampling was set to 94 species and the extant diversity to 150 species, start = 0 and end = 8 and grid = 0.1 Myr for a fine-scale estimation of rate shifts. The best-fitting model was selected on the basis of likelihood ratio tests. We also used the method of Etienne *et al.*[66] implemented in *TreePar* to test for a potential saturation of *Exocelina* diversity on the island. Hence, we explored the effect of diversity on speciation and extinction rates. The function *'bd.densdep.optim'* was used to fit this model with the following settings: discrete = TRUE, missing species acknowledged using ρ = 94/150, the initial carrying capacity fixed to the extant diversity minK = 150 and the maximum carrying capacity tested fixed as a default parameter to maxK = 1.5 $\times$ minK = 225. Finally, we constructed LTT plots to visualize diversification rate over time using *ape*[63] for the 94 species included in this study and *TreeSim*[65] to draw the LTT of 1,000 simulated topologies accounting for missing taxa under a constant rate model[67].

Bionomics. Based on the sequenced specimens, as well as $>$3,000 additional specimens currently under taxonomic study, we compiled a preliminary survey of species composition per collecting locality. Most of our collecting localities only contain samples from one puddle or several small waterholes along one stream segment less than 200 m long. These data were used to evaluate the extent to which sister species or close relatives co-occur.

References

1. Mayr, E. & Diamond, J. M. Birds on islands in the sky: Origin of the montane avifauna of Northern Melanesia. *Proc. Natl Acad. Sci. USA* **73,** 1765–1769 (1976).
2. Knox, E. B. & Palmer, J. D. Chloroplast DNA variation and the recent radiation of the giant senecios (Asteraceae) on the tall mountains of eastern Africa. *Proc. Natl Acad. Sci. USA* **92,** 10349–10353 (1995).
3. Gressitt, J. L. *Biogeography and Ecology of New Guinea* Vol. 1 & 2 (Dr. W. Junk publishers, 1982).
4. Wallace, A. R. On the zoological geography of the Malay archipelago. *J. Proc. Linn. Soc.* **4,** 172–184 (1860).
5. Wilson, E. O. The nature of the taxon cycle in the Melanesian ant fauna. *Am. Nat.* **95,** 169–193 (1961).
6. Diamond, J. M. *Avifauna of the eastern highlands of New Guinea*, 438 (Nuttall Ornithological Club, 1972).
7. Gressitt, J. L. Systematics and biogeography of the longicorn beetle tribe Tmesisternini. *Pac. Insects Monogr.* **41,** 1–263 (1984).
8. Baldwin, S. L., Fitzgerald, P. G. & Webb, L. E. Tectonics of the New Guinea Region. *Annu. Rev. Earth Planet. Sci.* **40,** 495–520 (2012).
9. Hope, G. S. *The Equatorial Glaciers of New Guinea, Results of the 1971–1973 Australian Universities' Expeditions to Irian Jaya, Survey, Glaciology, Meteorology, Biology and Palaeoenvironments* (Balkema, 1976).
10. Unmack, P. J., Allen, G. R. & Johnson, J. B. (2013 Phylogeny and biogeography of rainbowfishes (Melanotaeniidae) from Australia and New Guinea. *Mol. Phylogenet. Evol.* **67,** 15–27 (2013).
11. Jønsson, K. A., Fabre, P.-H., Ricklefs, R. E. & Fjeldså, J. Major global radiation of corvoid birds originated in the proto-Papuan archipelago. *Proc. Natl Acad. Sci. USA* **108,** 2328–2333 (2011).
12. Georges, A. *et al.* Contemporary genetic structure of an endemic freshwater turtle reflects Miocene orogenesis of New Guinea. *Biol. J. Linn. Soc.* **111,** 192–208 (2014).
13. Kalkman, V. J., Dijkstra, K. B., Dow, R. A., Stokvis, F. R. & van Tol, J. *Out of Australia: the Argiolestidae reveal the Melanesian Arc System and East Papua Composite Terrane as possible ancient dispersal routes to the Indo-Australian archipelago (Odonata, Argiolestidae).* PhD thesis (Natural Biodiversity Center, Leiden (2013).
14. De Boer, A. J. & Duffels, J. P. Historical biogeography of the cicadas of Wallacea, New Guinea and the West Pacific, a geotectonic explanation. *Palaeogeogr. Palaeoclimatol. Palaeoecol* **124,** 153–177 (1996).
15. Polhemus, D. A. Island arcs, and their influence on Indo-Pacific biogeography. in *The origin and evolution of Pacific island biotas, New Guinea to eastern Polynesia: patterns and processes* 51–66 (SPB Academic Publishing, 1996).
16. Deiner, K., Lemmon, A. R., Mack, A. L., Fleischer, R. C. & Dumbacher, J. P. A Passerine Bird's evolution corroborates the geologic history of the island of New Guinea. *PloS ONE* **6,** e19479 (2011).

17. Norman, J. A., Rheindt, F. E., Rowe, D. L. & Christidis, L. Speciation dynamics in the Australo-Papuan Meliphaga honeyeaters. *Mol. Phylogenet. Evol.* **42,** 80–91 (2007).
18. Zwiers, P. B., Borgia, G. & Fleischer, R. C. Plumage based classification of the bowerbird genus Sericulus evaluated using a multi-gene, multi-genome analysis. *Mol. Phylogenet. Evol.* **46,** 923–931 (2008).
19. Macqueen, P., Goldizen, A. W., Austin, J. J. & Seddon, J. M. Phylogeography of the pademelons (Marsupialia: Macropodidae: Thylogale) in New Guinea reflects both geological and climatic events during the Plio-Pleistocene. *J. Biogeogr.* **38,** 1732–1747 (2011).
20. Meredith, R. W., Mendoza, M., Roberts, K. K., Westerman, M. & Springer, M. S. A phylogeny and timescale for the evolution of Pseudocheiridae (Marsupialia: Diprotodontia) in Australia and New Guinea. *J. Mamm. Evol.* **17,** 75–99 (2010).
21. Rowe, K. C., Reno, M. L., Richmond, D. M., Adkins, R. M. & Steppan, S. J. Pliocene colonization and adaptive radiations in Australia and New Guinea (Sahul): multilocus systematics of the old endemic rodents (Muroidea, Murinae). *Mol. Phylogenet. Evol.* **47,** 84–101 (2008).
22. Rowe, K. C., Aplin, K. P., Baverstock, P. R. & Moritz, C. Recent and rapid speciation with limited morphological disparity in the genus Rattus. *Syst. Biol.* **60,** 188–203 (2011).
23. Malekian, M., Cooper, S. J. B., Norman, J., Christidis, L. & Carthew, S. M. Molecular systematics and evolutionary origins of the genus Petaurus (Marsupialia, Petauridae) in Australia and New Guinea. *Mol. Phylogenet. Evol.* **54,** 122–135 (2010).
24. Hall, R. in *Biogeography and Geological Evolution of SE Asia*. (eds Hall, R. & Holloway, J. D.) 99–131 (Backhuys Publishers, 1998).
25. Cloos, M. *Collisional Delamination in New Guinea: The Geotectonics of Subducting Slab Breakoff* Vol. 400 (ed. Cloos, M.) (Geological Society of America, 2005).
26. Dumbacher, J. P. & Fleischer, R. C. Phylogenetic evidence for color-pattern convergence in toxic pitohuis: Müllerian mimicry in birds? *Proc. Roy. Soc. B* **268,** 1971–1976 (2001).
27. Joseph, L., Slikas, B., Alpers, D. & Schodde, R. Molecular systematics and phylogeography of New Guinean logrunners (Orthonychidae). *Emu* **101,** 273–280 (2001).
28. van Ufford, A. Q. & Cloos, M. Cenozoic tectonics of New Guinea. *Am. Assoc. Petr. Geol. B* **89,** 119–140 (2005).
29. Hall, R. Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific: computer-based reconstructions, model and animations. *J. Asian Earth Sci.* **20,** 353–434 (2002).
30. Hill, K. C. & Hall, R. Mesozoic-Cenozoic evolution of Australia's New Guinea margin in a west Pacific context. in *Evolution and Dynamics of the Australian Plate*. (eds Hillis, R. R. & Müller, R. D.) 265–289 (Geological Society of Australia Special Publication 22 and Geological Society of America Special Papers, 2003).
31. Beebe, N. W. & Cooper, R. D. Distribution and evolution of the Anopheles punctulatus group (Diptera: Culicidae) in Australia and Papua New Guinea. *Int. J. Parasitol.* **32,** 563–574 (2002).
32. Shaverdo, H. V., Surbakti, S., Hendrich, L. & Balke, M. Introduction of the Exocelina ekari-group with descriptions of 22 new species from New Guinea (Coleoptera, Dytiscidae, Copelatinae). *ZooKeys* **250,** 1–76 (2012).
33. Pigram, C. J. & Davies, H. L. Terranes and the accretion history of the New Guinea orogen. *BMR J. Aust. Geol. Geop.* **10,** 193–212 (1987).
34. Hill, K. C. & Raza, A. Arc-continent collision in Papua Guinea, Constraints from fission track thermochronology. *Tectonics* **18,** 950–966 (1999).
35. Pigram, C. J. & Symonds, P. A. A review of the timing of the major tectonic events in the New Guinea Orogen. *J. South Asian Earth* **6,** 307–318 (1991).
36. Struckmeyer, H. I. M., Yeung, M. & Pigram, C. J. in *Proc. 2nd PNG Petrol. Conv*. Petroleum Exploration and Development in Papua New Guinea 261–290 Port Moresby, Papua New Guinea, 1993.
37. Balke, M., Pons, J., Ribera, I., Sagata, K. & Vogler, A. P. Infrequent and unidirectional colonization of hyperdiverse Papuadytes diving beetles in New Caledonia and New Guinea. *Mol. Phylogenet. Evol.* **42,** 505–516 (2007).
38. Toussaint, E. F. A., Sagata, K., Surbakti, S., Hendrich, L. & Balke, M. Australasian sky islands act as a diversity pump facilitating peripheral speciation and complex reversal from narrow endemic to widespread ecological supertramp. *Ecol. Evol.* **3,** 1031–1049 (2013).
39. Ribera, I., Barraclough, T. G. & Vogler, A. P. The effect of habitat type on speciation rates and range movements in aquatic beetles: inferences from species-level phylogenies. *Mol. Ecol.* **10,** 721–735 (2001).
40. Janzen, D. H. Why mountain passes are higher in the tropics. *Am. Nat.* **101,** 233–249 (1967).
41. Illies, J. Versuch einer allgemeinen biozönotischen Gliederung der Fließgewässer. *Int. Rev. Hydrobiol.* **46,** 205–213 (1961).
42. Statzner, B. & Dolédec, S. Phylogenetic, spatial, and species-trait patterns across environmental gradients: the case of Hydropsyche (Trichoptera) along the Loire River. *Int. Rev. Hydrobiol.* **96,** 121–140 (2011).
43. Vuataz, L., Sartori, M., Gattolliat, J.-L. & Monaghan, M. T. Endemism and diversification in freshwater insects of Madagascar revealed by coalescent and phylogenetic analysis of museum and field collections. *Mol. Phylogenet. Evol.* **66,** 979–991 (2013).
44. Edgar, R. C. MUSCLE, Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).
45. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2, more models, new heuristics and parallel computing. *Nat. Methods* **9,** 772 (2012).
46. Lanfear, R., Calcott, B., Ho, S. Y. & Guindon, S. PartitionFinder, combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* **29,** 1695–1701 (2012).
47. Ronquist, F. *et al.* MrBayes 3.2, Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61,** 539–542 (2012).
48. Kass, R. & Raftery, A. Bayes factors. *J. Am. Statist.* **90,** 773–795 (1995).
49. Rambaut, A. & Drummond, A. J. Tracer, version 1.4 (2007).
50. Baele, G. *et al.* Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol. Biol. Evol.* **29,** 2157–2167 (2012).
51. Xie, W., Lewis, P. O., Fan, Y., Kuo, L. & Chen, M.-H. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Syst. Biol.* **60,** 150–160 (2011).
52. Brown, J. M. & Lemmon, A. R. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. *Syst. Biol.* **56,** 643–655 (2007).
53. Stamatakis, A. RAxML-VI-HP: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22,** 2688–2690 (2006).
54. Erixon, P., Svennblad, B., Britton, T. & Oxelman, B. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Syst. Biol.* **52,** 665–673 (2003).
55. Felsenstein, J. *Inferring phylogenies* (Sinauer Associates, 2004).
56. Balke, M. *et al.* New Guinea highland origin of a widespread arthropod supertramp. *Proc. R.. Soc. B* **276,** 2359–2367 (2009).
57. Papadopoulou, A., Anastasiou, I. & Vogler, A. P. Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. *Mol. Biol. Evol.* **27,** 1659–1672 (2010).
58. Andújar, C., Serrano, J. & Gómez-Zurita, J. Winding up the molecular clock in the genus Carabus (Coleoptera: Carabidae): assessment of methodological decisions on rate and node age estimation. *BMC Evol. Biol.* **12,** 40 (2012).
59. Swofford, D. L. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4 (Sinauer Associates, 2003).
60. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29,** 1969–1973 (2012).
61. Gressitt, J. L. & Nadkarni, N. *Guide to Mt Kaindi: Background to Montane New Guinea Ecology (No. 5)* (Wau Ecology Institute, 1978).
62. Johns, R. J. Plant zonation. in *Biogeography and Ecology of New Guinea* 309–330 (Springer, 1982).
63. Paradis, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20,** 289–290 (2004).
64. Stadler, T. Mammalian phylogeny reveals recent diversification rate shifts. *Proc. Natl Acad. Sci. USA* **108,** 6187–6192 (2011).
65. Stadler, T. Simulating trees on a fixed number of extant species. *Syst. Biol.* **60,** 676–684 (2011).
66. Etienne, R. S. *et al.* Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proc. Roy. Soc. B* **279,** 1300–1309 (2012).
67. Cusimano, N., Stadler, T. & Renner, S. A new method for handling missing species in diversification analysis applicable to randomly or non-randomly sampled phylogenies. *Syst. Biol.* **61,** 785–792 (2012).

Acknowledgements

We thank Alan Stewart (Plymouth) and Vojtech Novotny and his entire team at the PNG Binatang Research Center (Madang), and especially thank Aloysius Posman and Andrew Kinibel for their incredible efforts in the field. We also thank Alexander Riedel, Andrew Mack and Debra Wright (former WCS PNG Program in Goroka) and their entire team for much appreciated friendship and logistic help and the PNG Department of Environment and Conservation (DEC) for a series of export and research permissions. We are grateful to Suriani Surbakti and Evi Warikar (UNCEN, Jayapura, Papua) as well as Shita Prativi and the late Kris Tindige (Manokwari, Papua Barat) for logistic help and information. Funding came from Deutsche Forschungsgemeinschft DFG (BA2152/3-1; 4-1; 7-1; 11-1), the UK DEFRA Darwin Initiative (14054) and the FWF (Fonds zur Förderung der wissenschaftlichen Forschung—the Austrian Science Fund, project P 24312-B17 to H.V.S.).

Authors contributions

E.F.A.T. and M.B. designed the study; K.S., M.B. and S.I. collected the specimens; M.B. and H.V.S. performed the identification and compiled ecological data; E.F.A.T. carried out the molecular work, analyses and figure design; E.F.A.T, R.H. and M.B. drafted the manuscript with revisions of all the authors.

Additional information

Accession Codes: Sequences have been deposited in EMBL Protein Knowledgebase database under accession numbers HG974147 to HG974233 and HG973548 to HG973881.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

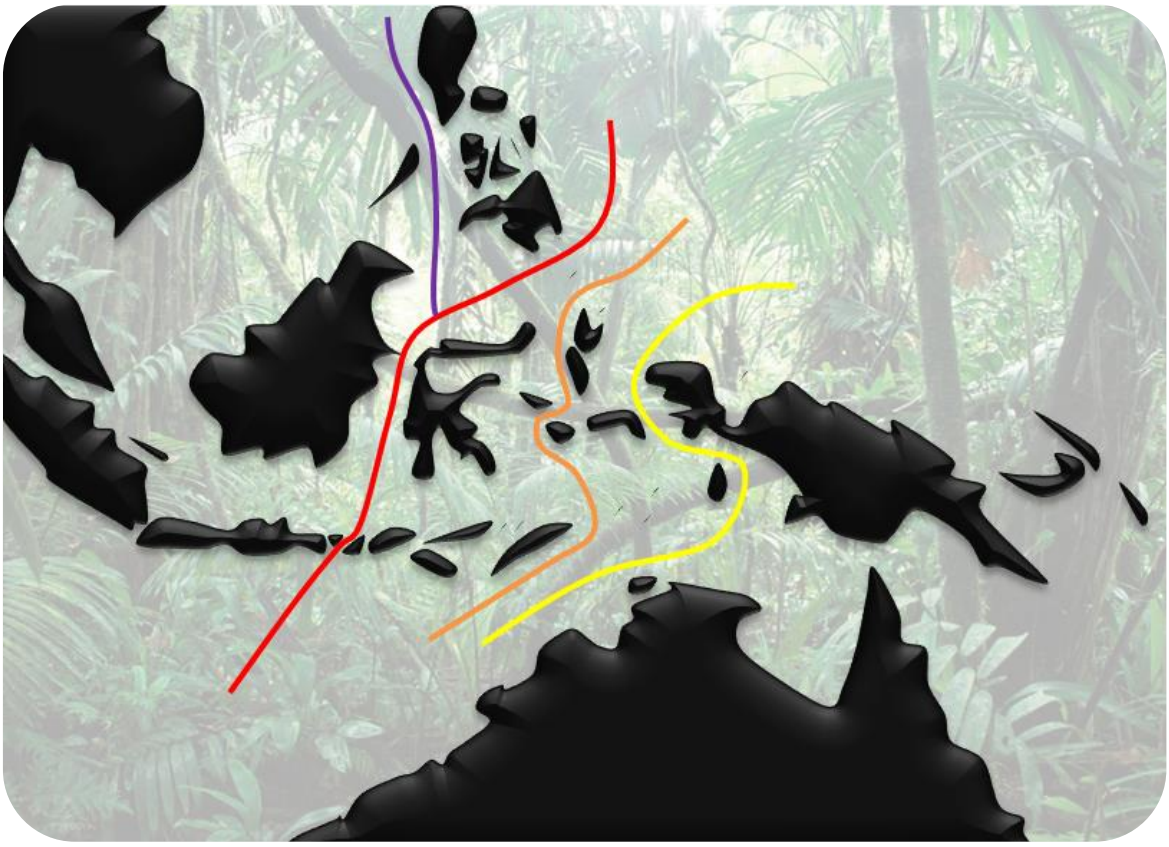
Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Toussaint, E. F. A. *et al.* The towering orogeny of New Guinea as a trigger for arthropod megadiversity. *Nat. Commun.* 5:4001 doi: 10.1038/ncomms5001 (2014).

PART 3: BIOGEOGRAPHIC INSIGHTS FOR THE ARCHIPELAGO

Chapter 9. On the role of biogeographic barriers



The Australasian / Indomalayan archipelago with Wallace's (red), Huxley's (purple), Weber's (orange) and Lydekker's (yellow) respective lines demarcating Asian and Australian biotas.

"*In this Archipelago there are two distinct faunas rigidly circumscribed, which differ as much as those of South America and Africa, and more than those of Europe and North America: yet there is nothing on the map or on the face of the islands to mark their limits. The boundary line often passes between islands closer than others in the same group. I believe the western part to be a separated portion of continental Asia, the eastern the fragmentary prolongation of a former Pacific continent."*

Alfred Wallace, *Letter to Henry Walter Bates*, 1858

Chapter contents

9.1 Paper X – Australasian weevil biogeography and the role of Wallace's line 208

PAPER X

9.1 Multiple transgressions of Wallace's Line explain diversity of flightless *Trigonopterus* weevils on Bali

Citation: Tänzler R, Toussaint EFA, Suhardjono YR, Balke M, Riedel A (2014) Multiple transgressions of Wallace's Line explain diversity of flightless *Trigonopterus* weevils on Bali. *Proceedings of the Royal Society of London, Series B*, 281:20132528.

PROCEEDINGS OF THE ROYAL SOCIETY B

rspb.royalsocietypublishing.org

Research

Cite this article: Tänzler R, Toussaint EFA, Suhardjono YR, Balke M, Riedel A. 2014 Multiple transgressions of Wallace's Line explain diversity of flightless *Trigonopterus* weevils on Bali. *Proc. R. Soc. B* **281**: 20132528. http://dx.doi.org/10.1098/rspb.2013.2528

Received: 26 September 2013
Accepted: 18 February 2014

Subject Areas:
evolution

Keywords:
Indo-Australian archipelago, phylogeography, biogeography, endemism, Bayesian relaxed clock

Author for correspondence:
Alexander Riedel
e-mail: riedel@smnk.de

Electronic supplementary material is available at http://dx.doi.org/10.1098/rspb.2013.2528 or via http://rspb.royalsocietypublishing.org.

Royal Society Publishing

Multiple transgressions of Wallace's Line explain diversity of flightless *Trigonopterus* weevils on Bali

Rene Tänzler[1], Emmanuel F. A. Toussaint[1], Yayuk R. Suhardjono[2], Michael Balke[1,3] and Alexander Riedel[4]

[1]SNSB-Zoological State Collection, Münchhausenstrasse 21, Munich 81247, Germany
[2]Division of Zoology, Cibinong Science Center, LIPI, Jalan Raya Jakarta - Bogor, Cibinong, Indonesia
[3]GeoBioCenter, Ludwig-Maximilians-University, Munich, Germany
[4]Museum of Natural History Karlsruhe (SMNK), Erbprinzenstrasse 13, Karlsruhe 76133, Germany

The fauna of Bali, situated immediately west of Wallace's Line, is supposedly of recent Javanese origin and characterized by low levels of endemicity. In flightless *Trigonopterus* weevils, however, we find 100% endemism for the eight species here reported for Bali. Phylogeographic analyses show extensive *in situ* differentiation, including a local radiation of five species. A comprehensive molecular phylogeny and ancestral area reconstruction of Indo-Malayan–Melanesian species reveals a complex colonization pattern, where the three Balinese lineages all arrived from the East, i.e. all of them transgressed Wallace's Line. Although East Java possesses a rich fauna of *Trigonopterus*, no exchange can be observed with Bali. We assert that the biogeographic picture of Bali has been dominated by the influx of mobile organisms from Java, but different relationships may be discovered when flightless invertebrates are studied. Our results highlight the importance of in-depth analyses of spatial patterns of biodiversity.

1. Introduction

The Indonesian island of Bali lies in a critical biogeographic position—on the edge of the Sunda shelf just west of the Lombok strait that demarcates Wallace's Line [1–4]. Yet, the numerous studies of the Indo-Australian fauna usually focus on the larger islands of Borneo and Sulawesi [5–8] and data on the fauna of Bali remain surprisingly scarce and scattered. Bali is essentially an extension of East Java, at the closest point only 2 km apart and repeatedly connected during lower sea levels [9]. This is reflected in Bali's presumably low degree of local endemism and fauna shared with Java.

Work in the first half of the last century focused mainly on lists of vertebrate species summarized by Rensch [10]: he concludes that there is only a single bird endemic to Bali, the now critically endangered Bali myna (*Leucopsar rothschildi* Stresemann). The majority of the remaining Balinese bird species also occur in Java, while only a very few are shared with islands to the East [10,11]. All native frog species except for one are shared with Java [12]. Large Asian mammals such as the banteng (*Bos javanicus* d'Alton) and the tiger (*Panthera tigris* Linnaeus) reached their easternmost area of distribution in Bali although the latter is now extinct on Bali [13]. The notion that the Balinese fauna is derived from Java relatively recently is confirmed by some rather mobile groups of invertebrates [14,15]. There are 14 species of land snails endemic to Bali and the neighbouring island of Nusa Penida [16], but the geographical provenance of their ancestors remains unknown.

There are no modern inventories of Balinese fauna, and molecular phylogenetic methods have never been applied to investigate macro-evolutionary processes explaining faunal origins or phylogeographic patterns within this island. Here, we conducted a comprehensive inventory of an Indo-Malayan–Australasian group of flightless weevils: *Trigonopterus* Fauvel is an ideal group

for studying the complex biogeographic history of the Indo-Australian archipelago. This genus has a marked tendency towards local species endemism, but despite the inability to fly, *Trigonopterus* has a wide range, from east Sumatra across Melanesia to the Samoan islands. It is hyperdiverse in New Guinea, with more than 300 species recorded [17,18]. Although only a single species has been described from Sulawesi to date, we have more than 100 new species awaiting formal description (A. Riedel 2014, unpublished data). Diversity decreases to the West, but is still substantial with more than 50 species recently discovered in Borneo, Sumatra and Java [19]. Species are confined to wet primary forests where they can be collected by sifting the leaf-litter. Many such habitats have been degraded or were converted to agricultural use as a consequence of the human population explosion on Java and Bali [20]. However, many lowland areas of East Java, Bali and the Lesser Sunda Islands currently support a seasonal type of monsoon forest, according to our experience not a suitable habitat for *Trigonopterus*. Therefore, these weevils are confined to remnants of wet primary forests, typically on mountainsides.

Here, we use molecular phylogenetic data coupled with ancestral area inference to show that *Trigonopterus* weevils have repeatedly colonized Bali from the East, thereby transgressing Wallace's Line. Surprisingly, there are no closer relationships with the fauna of nearby East Java, underpinning the need for comprehensively sampled, phylogeny-driven studies to better understand the region's faunal evolution.

2. Material and methods

(a) Taxon sampling and DNA sequencing

Species for our analysis were selected from a preliminary phylogenetic reconstruction which we performed to identify major clades. This initial analysis contained 138 *Trigonopterus* species with an alignment of 4646 bps consisting of fragments from CO1, 16S rRNA, arginine kinase, CAD, elongation factor 1α, enolase and histone 4. It included all 82 species from Sumatra, Java and the Lesser Sunda Islands found during a total of 212 days of fieldwork (17 days in Sumatra, 120 days in Java, 28 days in Bali, 27 days in Lombok, 12 days in Sumbawa and 8 days in Flores) covering 72 localities and resulting in 354 litter samples (electronic supplementary material, S6). A total of 3812 *Trigonopterus* specimens were available, and a full taxonomic treatment of this material is currently in preparation. All major areas of suitable habitat of the Sunda Arc were sampled. East Java, Bali and Lombok were sampled most intensely; we did not retrieve additional species upon repeated visits to the same localities. Areas of Sumatra, West Java and Flores may harbour additional species, but we are confident that all major clades from these areas have been discovered because additional visits did not reveal new lineages. A single clade containing most Balinese species together with a Lombok species (292) was well supported. Most species from Java and Sumatra, including species 299 and 348, are monophyletic with equally strong support. Eight species of *Trigonopterus* subgenus *Mimidotasia* from Java and Sumatra comprise an early diverging lineage that was omitted from the subsequent analysis because the group is missing from Bali and the Lesser Sunda Islands. The entire clade containing species 328 from Bali and species 317 from East Java was transferred to the subsequent analysis.

Our present dataset contains all Balinese species and their respective sister clades (electronic supplementary material, S3) along with a representative selection of the remaining fauna of the Sunda Arc comprising 40 species. Furthermore, the same number of species was added, representing all major lineages from Borneo, and the hyperdiverse islands of Sulawesi and New Guinea. Four cryptorhynchine species from Australia, New Guinea and Java were included as outgroup representatives (*Critomerus iliacus* (Pascoe); *Microporopterus* cf. *setosus* Voss; *Ouporopterus squamiventris* Lea; *Miocalles* sp.). Some of the *Trigonopterus* species from New Guinea possess a valid name [21], while others are currently being revised and described. Undescribed species are referred to by unique species numbers that will be given in future taxonomic treatments. All the species were monophyletic in a phylogeny using CO1 data of multiple specimens per species, and also well delineated by male genital characters.

DNA was extracted non-destructively using the DNeasy and NucleoSpin 96 Tissue kits (Qiagen, Hilden; Macherey-Nagel, Düren, Germany). For PCR amplification (electronic supplementary material, S1), we used standard protocols (http://zsm-entomology.de/wiki/The_Beetle_D_N_A_Lab). Sequences were edited using SEQUENCHER v. 4.10.1 (GeneCodes Corp., Ann Arbor, MI, USA).

(b) Alignment and data matrices

Twelve fragments representing nine genes were sequenced (electronic supplementary material, S2 and S7). Alignments were performed with MUSCLE [22] and reading frames checked in MESQUITE v. 2.75 (http://mesquiteproject.org). Alignment length was 6800 bps (two assembled fragments of CO1 (1416 bps), 16S (579 bps), 18S (584 bp), 28S (534 bps), arginine kinase (720 bps), CAD1 (462 bp), CAD2 (594 bps), CAD3 (663 bps), elongation factor 1α (372 bps), enolase (663 bps) and histone 4 (213 bps)).

(c) Phylogenetic inferences

We used maximum likelihood (ML) as well as Bayesian inference (BI) to reconstruct the relationships among *Trigonopterus* species. ML analyses were performed in RAxML [23] with 1000 thorough bootstrap replicates using five different partitioning strategies: no partitioning, one partition for each gene, one partition for each genome (mitochondrial versus nuclear), one partition for each type (coding versus non-coding genes) and one partition for each codon position (for non-coding genes, one partition for each). The same strategies were used for BI analyses carried out in MRBAYES v. 3.2 [24] (electronic supplementary material, S4). We sampled 30 million generations of two independent runs consisting of eight Markov chain Monte Carlo (MCMC) sampling every 1000th generation. A burn-in of 5000 trees was chosen after investigation of split-frequencies and log-likelihood curves in TRACER v. 1.5 [25]. A 50% majority rule consensus tree was constructed afterwards based on the remaining trees. The best-fitting partitioning strategy for BI was selected using Bayes factors [26] tests based on marginal likelihoods estimated through stepping-stone sampling [27]. The most appropriate substitution model for each partition was selected using the Bayesian information criterion as implemented in jMODELTEST v. 2.1.3 [28].

(d) Dating and ancestral area reconstruction

Divergence times were estimated with the Bayesian relaxed clock method implemented in BEAST v. 1.8.0 [29]. The only time-calibrated tree of Curculionidae available [30] did not recover Cryptorhynchinae as monophyletic, perhaps because more than 30% of the data was missing. The scant fossil record of this subfamily does not offer a taxon to which *Trigonopterus* could be safely attributed. As a result, we were not able to use a secondary calibration for the *Trigonopterus* radiation. In a first calibration, several substitution rates of Coleoptera have been used (calculated for the COI marker using multiple fossils and geological evidences [31–33]—see [34] for a rationale on the use of this interval). The early diversification of *Trigonopterus* would have taken place more than 60 Ma which appears significantly too old. The age of

flightless *Galapaganus* weevils older than the islands they inhabit was attributed to earlier, sunken islands [35]. In the case of *Trigonopterus*, no such land areas can be expected based on geological reconstructions of the New Guinea area more than 60 Ma [36,37]. The high interspecific divergences (mean 20% for CO1) previously reported for *Trigonopterus* [17] from supposedly young geological terranes might indicate accelerated molecular evolution, which has recently been linked to flightlessness [38]. Loss of flight in beetles not only promotes speciation, but also those flightless species retain a higher genetic differentiation on population level and show deeper genetic branching than flying species [39]. A life history that requires little movement is an equally important factor [40]. In *Trigonopterus*, both factors are given: these weevils, as well as all the other members of the subtribe Tylodina are fully wingless, and their habitat, the leaf-litter of humid forests, is a highly stable and relatively uniform resource. Groups of groundwater- and cave-dwelling Crustacea are also known for markedly accelerated evolutionary rates related to their fragmented populations and the frequent occurrence of bottlenecks [41].

Therefore, in order to obtain divergence time estimates, we used a geological calibration. We constrained the root of the tree not to be older than 30 Myr as a conservative estimate, because the early lineages in our phylogeny (figure 1) were all Papuan *Trigonopterus*. Present-day New Guinea has a highly complex orogenic history, but the most recent geological reconstructions of the region [36,37] suggest that at most small and low-lying islands were emergent before 30 Ma. If some land did exist before 30 Ma, it was lacking the horizontal and vertical dimension required to facilitate lineage diversification; habitats fully explaining the observed diversification patterns are more recent and likely of Miocenic age ([34]; R. Hall 2013, personal communication). This was reflected in empirical studies (birds [42], rainbow fishes [43], diving beetles [34]) which estimate the onset of Papuan lineage diversification around 30 Ma or more recent. Thus, our calibration is likely to be conservative and may yield slightly overestimated ages. Using more recent root calibration dates (20 Ma and 10 Ma) had no impact on the biogeographic scenario inferred for Bali. The analyses were performed under a *Speciation: Birth–Death Incomplete Sampling* [44] using an estimated relaxed clock rate (uncorrelated lognormal) because the hypothesis of a strict molecular clock was tested and rejected (p-value < 0.001) in PAUP* [45]. The MCMC parameters were fixed to 30 million generations with sampling every 1000th generation and discarding 5000 trees as burn-in. In order to reduce the computational time and the parameter space to explore, we fixed the best BI topology from which we removed all outgroups by manually editing the .xml file created in BEAUTi v. 1.8.0 [29]. A 50% majority rule consensus tree was created in TreeAnnotator v. 1.8.0.

Ancestral areas were inferred using the dispersal–extinction–cladogenesis (DEC) model in Lagrange [46,47] based on our BEAST topology. We defined seven areas: Bali, Kalimantan, Flores + Lombok + Sumbawa, Java, New Guinea, the Philippines and Sulawesi. No species of *Trigonopterus* occurred in more than one area. Palaeogeographic changes through time [36,37,48] were accommodated by two time slices encompassing the past 30 Myr. Rates of dispersal were based on distances between areas and geographical barriers (see the electronic supplementary material, S5). The maximum number of possible regions for each node was limited to three. We enforced all possible combinations of areas at the root and conducted likelihood comparisons to select the most likely ancestral area. A difference between potential combinations equal or greater than 2 log-likelihood units was considered significant [46,47].

(e) Phylogeography

We analysed the phylogeographic pattern for the eight Balinese species by producing a haplotype network derived from 70 specimens, using the 5′ CO1, 16S and CAD datasets. The sequences were collapsed into haplotypes using DnaSP v. 5.10 software [49], and networks were inferred with Hapstar v. 0.7 [50] based on connection lengths calculated in Arlequin v. 3.11 [51].

3. Results

(a) Molecular phylogenetics

Phylogenetic inference using ML and BI recovered highly congruent topologies for the species of the Sunda Arc (figure 1); some differences exist at the backbone formed by New Guinea species. For BI analyses, the high ESS (effective sample size) values indicated convergence for all runs. Bayes factor analyses suggested that the best-fit partitioning strategy was the one comprising one partition for each genome (electronic supplementary material, S4).

(b) Faunal evolution and biogeography

Balinese *Trigonopterus* are not monophyletic but belong to three separate lineages, each with its closest relatives outside of Bali. One clade comprises five species (sp. 285, sp. 334, sp. 289, sp. 340 and sp. 286), with its sister species found in Sumbawa (sp. 287); the second clade comprises the sibling species (sp. 280 + sp. 327) with its sister species in Lombok (sp. 282); the third clade represented by a single species (sp. 328) whose sister species is from Sumbawa (sp. 326).

(c) Dating and ancestral area reconstruction

High ESS values indicated that all dating analyses reached convergence. *Trigonopterus* had a median age of 22.59 Ma (95% HPD 18.04–29.16 Ma). For Balinese clades 1, 2 and 3, we estimate the following ages: 3.33 Ma (95% HPD 2.16–5.09), 2.24 Ma (95% HPD 1.69–3.19) and 1.15 Ma (95% HPD 0.73–1.86).

The AAR (figure 1) suggests that the early evolution of *Trigonopterus* was restricted to New Guinea until the Late Miocene. The New Guinea character state for the root was significantly recovered ($\ln = -127.3$ against the second best root character state Sulawesi with $\ln = -137.9$). Our results highlight a dispersal event towards Sulawesi at this period followed by the colonization out of this area of surrounding islands by the end of the Miocene and throughout the Neogene. Few lineages of basal clades reach Sulawesi, respectively, Sumbawa, another one the Philippines and Borneo. The Philippines might have served as a stepping stone for the colonization of Borneo from New Guinea as illustrated in the reconstruction (figure 1). All other species found on Borneo, Sumatra, Java and the Lesser Sunda Islands belong to one clade which also has a few Sulawesi endemics; the clade is completely absent from New Guinea. Evidence that the three clades of Balinese species have reached Bali coming from the Lesser Sunda Islands was found to be significant.

(d) Phylogeography

The CO1-based network was the most informative one and fully compatible with the slightly less resolved 16S-based network; the one based on CAD fragment 3 was hardly resolved as that marker was not informative at this hierarchical level. All Balinese species of *Trigonopterus* weevils were genetically distinctive (1.7–24.3% smallest interspecific CO1 p-distance,

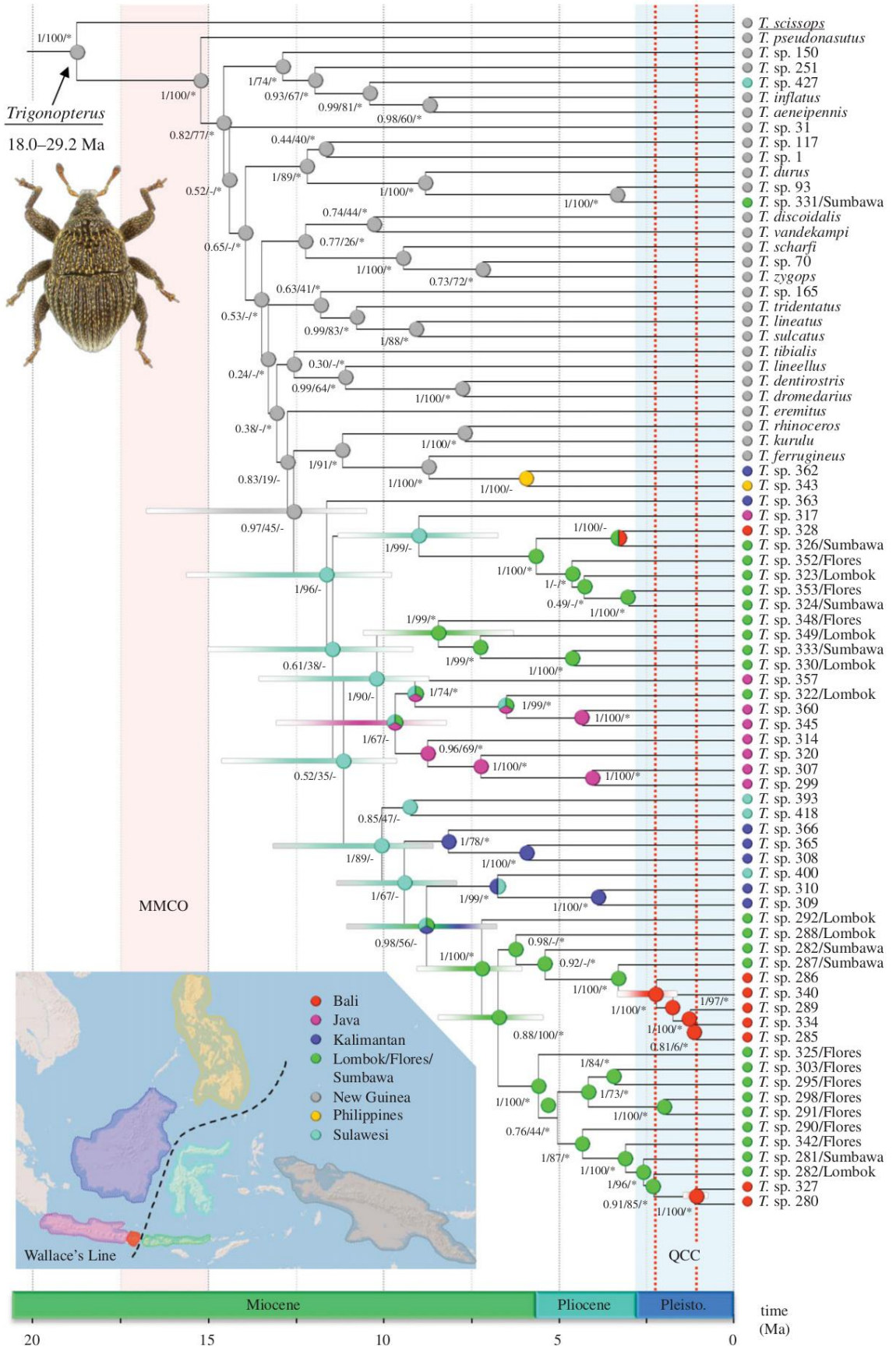


Figure 1. Bayesian phylogenetic relationships and historical biogeography of Australasian flightless *Trigonopterus* weevils. Values at each node (a/b/c) are (a) posterior probability of BI analysis, (b) Bootstrap support value of ML analysis (a hyphen indicates that this node is not found in the ML-based topology) and (c) relative probability of splits. Values above 95% are indicated by an asterisk, values below by a hyphen. A 2.5-Myr timescale is provided at the bottom of the chronogram spanning the epochs since 20 Ma. Horizontal bars indicate the 95% credibility interval of the divergence times. The bottom-left corner map represents the Australasian region along with the biogeographic regions used in the DEC analysis. Present-day distribution of the species is given at the tips of the topology. Coloured pastilles at each node correspond to the most likely ancestral area recovered by the DEC model. The mid-Miocene climatic optimum (MMCO) and quaternary climatic change (QCC) are illustrated with vertical coloured bars. The red vertical bars indicate independent colonization of Bali.

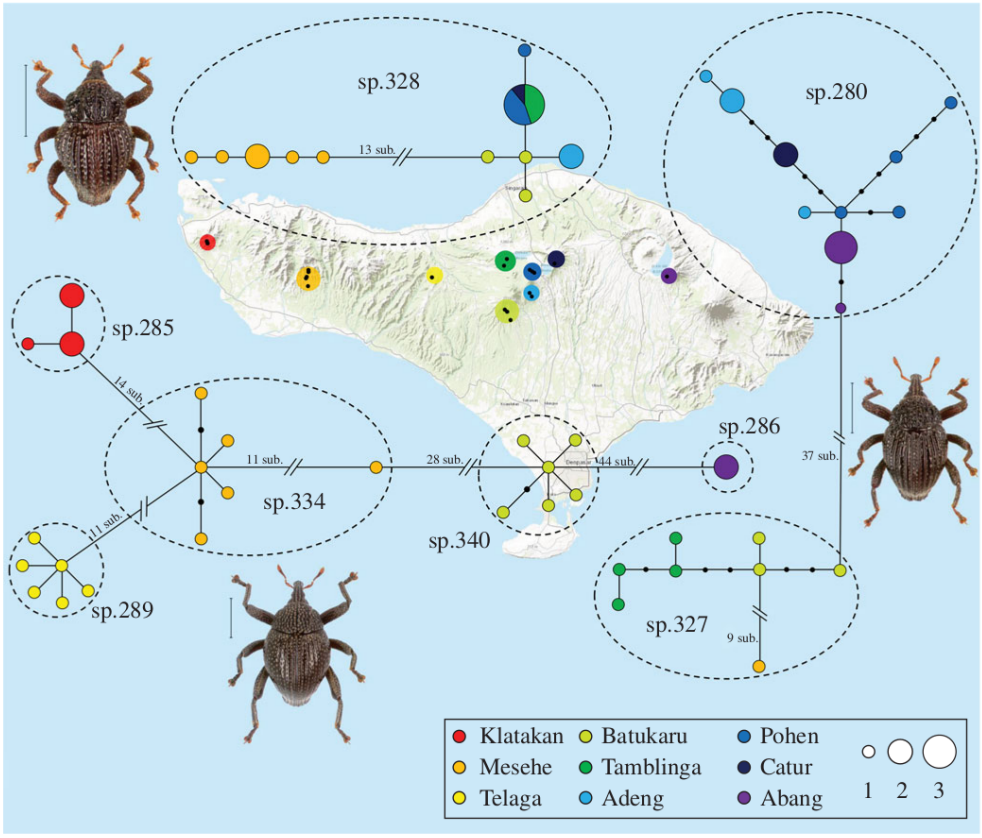


Figure 2. Haplotype networks based on the CO1 dataset of 71 *Trigonopterus* specimens from Bali. Substitutions (sub.) are marked by black dots between the haplotypes. More than two substitutions are stated as numbers above the respective branches. For colour code of localities, see inset. Scale bar of specimens = 1 mm.

only 0.1–1.1% intraspecific mean *p*-distance; figure 2). Most species are micro-endemics, i.e. to date only retrieved from a single locality each, with all haplotypes of a given species unique to a single locality. The clade containing five species shows a marked geographical east–west sequence, with, for example, the easternmost species (286) is more closely related to the central species (340) than to the western ones. Similarly, species 285, 334 and 289 are more closely related to each other than to the central (340) and eastern (286) species. Interestingly, individuals of species 289, 334 and 340 all had different haplotypes.

The other putative species occur in multiple localities and exhibit a more complex haplotype structure especially in the central region of the island where populations are not clearly delineated geographically.

Species 280 and 327 each occur in more than one locality and exhibit a marked haplotype structure. Species 280 is restricted to the east and eastern centre of Bali and is immediately replaced by species 327 from the western centre of Bali towards the West; the species are allopatric, but nearest localities are merely 6 km separated from each other.

Species 328, in its own clade, has a clear east/west separation of its populations but the eastern part of its distribution presents a less structured pattern as highlighted by a shared haplotype found in different localities. This species was found in all central localities. A high number (13) of CO1 substitutions distinguish specimens from the central localities and those from western Mt. Mesehe, but no other evidence for the presence of cryptic species could be found based on external and genital morphology.

4. Discussion

Here, we empirically show that comprehensive phylogenetic studies can reveal complex evolutionary histories of organisms in the geologically equally complex Indo-Australian archipelago. Using a densely sampled molecular phylogeny, we shed light on the origin of Bali's indigenous fauna, especially the origins of the little-known endemics which may not be as sparse as hitherto believed. Counterintuitively, the flightless weevils we studied arrived from east of Bali and east of Wallace's line rather than from the much closer East Java.

Based on present-day geography, a sea-level lowering of 50 m would result in a land connection between Bali and East Java [9]. Such conditions have likely prevailed at times during the Pleistocene allowing the influx of mobile terrestrial organisms from Java [52]. However, Bali remained insular throughout more than 50% of its Pleistocene history [9]. More importantly, during periods of connectivity, the lowlands of East Java and Bali were most likely dominated by savannah vegetation [53], an unsuitable habitat for *Trigonopterus* weevils, as well as most other forest-adapted taxa. Even at times when Bali and Java formed one landmass, the fauna of ever-wet rainforests was confined to the upper elevations of mountains, just as it is today [54]. Thus, an insular evolution persisted for forest species at all times.

Flightless, edaphic weevils fall into the category of less dispersive taxa [55]. Their chances to be lifted up by strong winds should be negligible. Ocean currents as a means of dispersal, maybe as part of or contained in flotsam, appear as a more plausible cause of dispersal. In this particular case, the general

scenario does not look favourable, either: the Indonesian throughflow [56] passing from north to south would form a barrier carrying organisms into the Indian Ocean instead of helping them to cross the straits from east to west.

Under these circumstances, colonization of Bali from the Lesser Sunda Islands by a group of flightless weevils appears to be unusual. However, this happened at least three times independently, and the conspicuous absence of any colonization event from the West, notwithstanding the rich presence of *Trigonopterus* in Java, is more than unexpected and highlights the need for comprehensively sampled phylogenetic analyses if we are to unravel the complexities of faunal evolution of a given area. While we cannot rule out the possible discovery of East Javanese sister species of one Balinese species, this appears highly unlikely for all three lineages.

Within Bali, on the contrary, the observed population-level pattern reflects the expected clear geographical structure for flightless organisms between different tropical mountain ranges [57–59]. This confirms the very limited dispersal abilities of *Trigonopterus* weevils and questions chance as an explanation for the multiple transgression of Wallace's Line by *Trigonopterus*. Once other taxa of poor dispersers are examined with appropriate methods, a general pattern may emerge that paints a more accurate picture of the early zoogeographic history of Bali, when land connections or ocean currents might have been very different from how we see them today. Taxa worth a comprehensive study may be some genera of snails (i.e. *Asperitas* Gray, *Sasakina* Rensch [60]) and weevils of Celeuthetini (i.e. *Syntrophus* Marshall [61]) that have distributions suggesting close relationships between Bali and islands to the East instead of Java, but no phylogenetic data are available for these yet.

Our study of Balinese *Trigonopterus* provides the first robust phylogeny and state-of-the-art biogeographic analysis for any Balinese taxon. The early evolution of *Trigonopterus* apparently took place in the area of present-day New Guinea, most likely in an archipelagic setting, before the formation of the main New Guinean landmass [36,37]. The Sunda-arc and the Sunda shelf were colonized from the East by rather derived lineages (figure 1). Thus, the possible centre of origin of *Trigonopterus* apparently coincides with its centre of diversity [62] in New Guinea/Australia. *Trigonopterus* was probably among the early groups to diversify on the proto-Papuan arc which was formed *ca* 30 Ma [36,63], a pattern also shown by some songbirds [42] and rainbow fish [43]. *Delias*, a diverse genus of butterflies, apparently also first diversified in the area of New Guinea and Wallacea, mainly during the Miocene [15,64].

The observed unexpected distribution patterns of *Trigonopterus* stress the importance of fine-grained and comprehensively sampled surveys in this biogeographic highly complex region. Relatively recent, largely Pleistocene processes of faunal exchange generated distribution patterns that include Bali along with Java as parts of the Sunda shelf contributing to what we perceive as Wallace's Line today [4]. However, the islands along the Sunda Arc from Sumatra to Flores are geologically heterogeneous and most likely emerged from the sea at very different times. Even Java is composed of a number of geologically distinct units [65]. The detailed study of less dispersive taxa undergoing endemic radiations during the Caenozoic will allow us to gain new insights into the development of a seemingly uniform chain of islands. Our present work highlights macro-evolutionary processes governing the biota of Bali and paves the way for future investigation of this frequently studied but still not fully understood area of the Indo-Australian Archipelago using molecular tools.

Acknowledgements. We thank LIPI (Indonesian Institute of Sciences), RISTEK (Ministry of State for Research and Technology, Indonesia) and the Indonesian Department of Forestry for providing relevant permits. Voucher specimens will be deposited in LIPI Division of Zoology Collection, formerly the Museum Zoologicum Bogoriense (Cibinong, Indonesia) and the Museum of Natural History Karlsruhe (Germany). We thank Dr Thomas von Rintelen (Berlin) and two anonymous reviewers for valuable comments. Last, but not least, the fieldwork in Bali and the other islands visited would not have been possible without the generous hospitality and help of many local people, and we thank all of them very warmly.

Data accessibility. DNA sequences: European Nucleotide Archive (accession nos. FN429126 - FN429350, HE613858–613921; 615156–616164). Phylogenetic data: TreeBASE accession no. 15388. Final DNA sequence assembly uploaded as the electronic supplementary material.

Funding statement. This work was financially supported by the German Research Foundation, DFG (RI 1817/1-1, 3-1, 3-3).

References

1. Wallace AR. 1860 On the zoological geography of the Malay Archipelago. *Zool. J. Linn. Soc.* **4**, 172–184. (doi:10.1111/j.1096-3642.1860.tb00090.x)
2. Huxley TH. 1868 On the classification and distribution of the Alectoromorphae and Heteromorphae. *Proc. Zool. Soc. Lond.* **1868**, 294–319.
3. Mayr E. 1944 Wallace's Line in the light of recent zoogeographic studies. *Q. Rev. Biol.* **19**, 1–14. (doi:10.1086/394684)
4. Lohman DJ, de Bruyn M, Page T, von Rintelen K, Hall R, Ng PKL, Shi H-T, Carvalho GP, von Rintelen T. 2011 Biogeography of the Indo-Australian archipelago. *Annu. Rev. Ecol. Evol. Syst.* **42**, 205–226. (doi:10.1146/annurev-ecolsys-102710-145001)
5. Whitmore TC (ed.) 1981 *Wallace's Line and plate tectonics*. Oxford, UK: Clarendon Press.
6. Hall R, Holloway JD (eds). 1998 *Biogeography and geological evolution of SE Asia*. Leiden, The Netherlands: Backhuys.
7. Stelbrink B, Albrecht C, Hall R, von Rintelen R. 2012 The biogeography of Sulawesi revisited: is there evidence for a vicariant origin of taxa on Wallace's 'anomalous island'? *Evolution* **66**, 2252–2271. (doi:10.1111/j.1558-5646.2012.01588.x)
8. Klaus S *et al.* 2013 Out of Borneo: Neogene diversification of Sundaic freshwater crabs (Crustacea: Brachyura: Gecarcinucidae: *Parathelphusa*). *J. Biogeogr.* **40**, 63–74. (doi:10.1111/j.1365-2699.2012.02771.x)
9. Voris HK. 2000 Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *J. Biogeogr.* **27**, 1153–1167. (doi:10.1046/j.1365-2699.2000.00489.x)
10. Rensch B. 1936 *Die Geschichte des Sundabogens: Eine tiergeographische Untersuchung*. Berlin, Germany: Gebr. Borntraeger.
11. MacKinnon JR, Phillipps K. 1993 *A field guide to the birds of Borneo, Sumatra, Java, and Bali, the Greater Sunda Islands*. Oxford, UK: Oxford University Press.
12. McKay JL. 2006 *A field guide to the amphibians and reptiles of Bali*. Malabar, FL: Krieger Publishing Company.
13. Whitten T, Soeriaatmadja RE, Afiff SA. 1996 *The ecology of Java and Bali*. The Ecology of Indonesia Series 2. Hong Kong, China: Periplus Editions.

14. Hendrich L, Balke M. 1995 Die Schwimmkäfer der Sunda-Insel Bali: Faunistik, Taxonomie, Ökologie, Besiedlungsgeschichte und Beschreibung von vier neuen Arten (Insecta: Coleoptera: Dytiscidae). *Faunist. Abh. Staatl. Mus. Tierk. Dresden* **20**, 29–56.
15. Braby MF, Pierce NE. 2007 Systematics, biogeography and diversification of the Indo-Australian genus *Delias* Hübner (Lepidoptera: Pieridae): phylogenetic evidence supports an 'out-of-Australia' origin. *Syst. Entomol.* **32**, 2–25. (doi:10.1111/j.1365-3113.2006.00349.x)
16. Vermeulen JJ, Whitten AJ. 1998 *Fauna Malesiana: guide to the land snails of Bali*. Leiden, The Netherlands: Backhuys Publishers.
17. Riedel A, Daawia D, Balke M. 2010 Deep *cox1* divergence and hyperdiversity of *Trigonopterus* weevils in a New Guinea mountain range (Coleoptera, Curculionidae). *Zool. Scripta* **39**, 63–74. (doi:10.1111/j.1463-6409.2009.00404.x)
18. Tänzler R, Sagata K, Surbakti S, Balke M, Riedel A. 2012 DNA barcoding for community ecology: how to tackle a hyperdiverse, mostly undescribed Melanesian fauna. *PLoS ONE* **7**, e28832. (doi:10.1371/journal.pone.0028832)
19. Riedel A, Suhardjono YR, Rahmadi C, Tänzler R. Submitted. Monograph of the *Trigonopterus* weevils from Sundaland and the Lesser Sunda Islands. *ZooKeys*.
20. Sodhi NS, Koh LP, Brook BW, Ng PKL. 2004 Southeast Asian biodiversity: an impending disaster. *Trends Ecol. Evol.* **19**, 654–660. (doi:10.1016/j.tree.2004.09.006)
21. Riedel A, Sagata K, Surbakti S, Tänzler R, Balke M. 2013 One hundred and one new species of *Trigonopterus* weevils from New Guinea. *ZooKeys* **280**, 1–150. (doi:10.3897/zookeys.280.3906)
22. Edgar RC. 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797. (doi:10.1093/nar/gkh340)
23. Stamatakis A. 2006 RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690. (doi:10.1093/bioinformatics/btl446)
24. Ronquist F *et al*. 2012 MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542. (doi:10.1093/sysbio/sys029)
25. Rambaut A, Drummond AJ. 2007 Tracer v. 1.4. See http://beast.bio.ed.ac.uk/Tracer.
26. Kass R, Raftery A. 1995 Bayes factors and model uncertainty. *J. Am. Stat. Assoc.* **90**, 773–795. (doi:10.1080/01621459.1995.10476572)
27. Xie W, Lewis PO, Fan Y, Kuo L, Chen M-H. 2011 Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Syst. Biol.* **60**, 150–160. (doi:10.1093/sysbio/syq085)
28. Darriba D, Taboada GL, Doallo R, Posada D. 2012 jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9**, 772. (doi:10.1038/nmeth.2109)
29. Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012 Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973. (doi:10.1093/molbev/mss075)
30. McKenna DD, Sequeira AS, Marvaldi AE, Farrell BD. 2009 Temporal lags and overlap in the diversification of weevils and flowering plants. *Proc. Natl Acad. Sci. USA* **106**, 7083–7088. (doi:10.1073/pnas.0810618106)
31. Balke M, Ribera I, Hendrich L, Miller M, Sagata K, Posman A, Vogler AP, Meier R. 2009 New Guinea highland origin of a widespread arthropod supertramp. *Proc. R. Soc. B* **276**, 2359–2367. (doi:10.1098/rspb.2009.0015)
32. Papadopoulou A, Anastasiou I, Vogler AP. 2010 Revisiting the insect mitochondrial molecular clock: the mid-Aegean Trench calibration. *Mol. Biol. Evol.* **27**, 1659–1672. (doi:10.1093/molbev/msq051)
33. Andújar C, Serrano J, Gómez-Zurita J. 2012 Winding up the molecular clock in the genus *Carabus* (Coleoptera: Carabidae): assessment of methodological decisions on rate and node age estimation. *BMC Evol. Biol.* **12**, 40. (doi:10.1186/1471-2148-12-40)
34. Toussaint EFA, Hall R, Monaghan MT, Sagata K, Ibalim S, Shaverdo HV, Vogler AP, Pons J, Balke M. Submitted. The towering oceanic orogeny of New Guinea as a trigger of arthropod megadiversity. *Nat. commun.*
35. Sequeira AS, Lanteri AA, Scataglini MA, Confalonieri VA, Farrell BD. 2000 Are flightless *Galapaganus* weevils older than the Galapagos Islands they inhabit? *Heredity* **85**, 20–29. (doi:10.1046/j.1365-2540.2000.00690.x)
36. Hall R. 2009 Southeast Asia's changing palaeogeography. *Blumea* **54**, 148–161. (doi:10.3767/000651909X475941)
37. Hall R. 2012 Sundaland and Wallacea: geology, plate tectonics and palaeogeography. In *Biotic evolution and environmental change in southeast Asia* (eds DJ Gower, KG Johnson, JE Richardson, BR Rosen, L Rüber, ST Williams), pp. 32–78. Cambridge, UK: Cambridge University Press.
38. Mitterboeck TF, Adamowicz SJ. 2013 Flight loss linked to faster molecular evolution in insects. *Proc. R. Soc. B* **280**, 20131128. (doi:10.1098/rspb.2013.1128)
39. Ikeda H, Nishikawa M, Sota T. 2012 Loss of flight promotes beetle diversification. *Nat. Commun.* **3**, 648. (doi:10.1038/ncomms1659)
40. Vogler AP, Timmermans MJ. 2012 Speciation: don't fly and diversify? *Curr. Biol.* **22**, R284–R286. (doi:10.1016/j.cub.2012.03.015)
41. Bauzà-Ribot MM, Juan C, Nardi F, Oromí P, Pons J, Jaume D. 2012 Mitogenomic phylogenetic analysis supports continental-scale vicariance in subterranean thalassoid crustaceans. *Curr. Biol.* **22**, 2069–2074. (doi:10.1016/j.cub.2012.09.012)
42. Jønsson KA, Fabre P-H, Ricklefs RE, Fjeldså J. 2011 Major global radiation of corvoid birds originated in the proto-Papuan archipelago. *Proc. Natl Acad. Sci. USA* **108**, 2328–2333. (doi:10.1073/pnas.1018956108)
43. Unmack PJ, Allen GR, Johnson JB. 2013 Phylogeny and biogeography of rainbowfishes (Melanotaeniidae) from Australia and New Guinea. *Mol. Phylogenet. Evol.* **67**, 15–27. (doi:10.1016/j.ympev.2012.12.019)
44. Stadler T. 2009 On incomplete sampling under birth–death models and connections to the sampling-based coalescent. *J. Theor. Biol.* **261**, 58–66. (doi:10.1016/j.jtbi.2009.07.018)
45. Swofford DL. 2002 *PAUP*. Phylogenetic analysis using parsimony* (**and other methods*), *v. 4.0b10*. Sunderland, MA: Sinauer Associates.
46. Ree RH, Moore BR, Webb CO, Donoghue MJ. 2005 A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution* **59**, 2299–2311. (doi:10.1111/j.0014-3820.2005.tb00940.x)
47. Ree RH, Smith SA. 2008 Maximum-likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Syst. Biol.* **57**, 4–14. (doi:10.1080/10635150701883881)
48. Metcalfe I. 2005 Asia: south-east. In *Encyclopedia of geology*, vol. 1 (eds RC Selley, LRM Cocks, IR Plimer), pp. 169–198. Oxford, UK: Elsevier.
49. Librado P, Rozas J. 2009 DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452. (doi:10.1093/bioinformatics/btp187)
50. Teacher AGF, Griffiths DJ. 2011 HapStar: automated haplotype network layout and visualization. *Mol. Ecol. Resour.* **11**, 151–153. (doi:10.1111/j.1755-0998.2010.02890.x)
51. Excoffier L, Laval G, Schneider S. 2005 Arlequin v. 3.0: an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* **1**, 47–50.
52. Miller KG *et al*. 2005 The Phanerozoic record of global sea-level change. *Science* **310**, 1293–1298. (doi:10.1126/science.1116412)
53. Bird MI, Taylor D, Hunt C. 2005 Palaeoenvironments of insular Southeast Asia during the last glacial period: a savanna corridor in Sundaland? *Q. Sci. Rev.* **24**, 2228–2242. (doi:10.1016/j.quascirev.2005.04.004)
54. Flenley JR. 1998 Tropical forests under the climates of the last 30,000 years. *Clim. change* **39**, 177–197. (doi:10.1023/A:1005367822750)
55. Gillespie RG, Baldwin BG, Waters JM, Fraser CI, Nikula R, Roderick GK. 2012 Long-distance dispersal: a framework for hypothesis testing. *Trends Ecol. Evol.* **27**, 47–56. (doi:10.1016/j.tree.2011.08.009)
56. Gordon AL, Fine RA. 1996 Pathways of water between the Pacific and Indian oceans in the Indonesian seas. *Nature* **379**, 146–149. (doi:10.1038/379146a0)
57. Hewitt GM. 2004 The structure of biodiversity–insights from molecular phylogeography. *Front. Zool.* **1**, 1–16. (doi:10.1186/1742-9994-1-4)
58. Chatzimanolis S, Caterino MS. 2008 Phylogeography of the darkling beetle *Coelus ciliatus* in California. *Ann. Entomol. Soc. Am.* **101**, 939–949. (doi:10.1603/0013-8746(2008)101[939:POTDBC]2.0.CO;2)

59. Kajtoch L. 2011 Conservation genetics of xerothermic beetles in Europe: the case of *Centricnemus leucogrammus*. *J. Insect. Conserv.* **15**, 787–797. (doi:10.1007/s10841-011-9377-8)
60. Hausdorf B. 1995 A preliminary phylogenetic and biogeographic analysis of the Dyakiidae (Gastropoda: Stylommatophora) and a biogeographic analysis of other Sundaland taxa. *Cladistics* **11**, 359–376. (doi:10.1111/j.1096-0031.1995.tb00095.x)
61. Marshall GAK. 1956 *The otiorrhynchine Curculionidae of the tribe Celeuthetini* (*Col.*). London, UK: Trustees of the British Museum.
62. Briggs JC. 2000 Centrifugal speciation and centres of origin. *J. Biogeogr.* **27**, 1183–1188. (doi:10.1046/j.1365-2699.2000.00459.x)
63. Hall R. 2011 Australia–SE Asia collision: plate tectonics and crustal flow. *Geol. Soc. Lond. Spec. Publ.* **355**, 75–109. (doi:10.1144/SP355.5)
64. Müller CJ, Matos-Maraví PF, Beheregaray LB. 2013 Delving into *Delias* Hübner (Lepidoptera: Pieridae): fine-scale biogeography, phylogenetics and systematics of the world's largest butterfly genus. *J. Biogeogr.* **40**, 881–893. (doi:10.1111/jbi.12040)
65. Pacey A, Macpherson CG, McCaffrey KJ. 2013 Linear volcanic segments in the central Sunda Arc, Indonesia, identified using Hough transform analysis: implications for arc lithosphere control upon volcano distribution. *Earth Planet. Sci. Lett.* **369–370**, 24–33. (doi:10.1016/j.epsl.2013.02.040)

PART 3: BIOGEOGRAPHIC INSIGHTS FOR THE ARCHIPELAGO

Chapter 10. Unfolding historical biogeography of widespread clades

Papilio (Achillides) ulysses, one of the most widespread species of peacock swallowtails distributed across the Moluccas, New Guinea, Solomon Islands and Queensland.

"*The Geographical distribution of Insects in the Archipelago is certainly far less strongly marked than that of Birds and Mammals, but I think that it may be in a great measure imputed to the much greater liability of insects to accidental dispersion.*"

Alfred Wallace, *Letter to Francis Polkinghorne Pascoe*, 1860.

Chapter contents

10.1 Paper XI - Fine-scale biogeography of widespread swallowtails 218

PAPER XI

10.1 Fine-scale biogeographic and temporal diversification processes of peacock swallowtails (*Papilio* subgenus *Achillides*) in the Indo-Australian Archipelago

Citation: Condamine FL, Toussaint EFA, Cotton A, Sperling FAH, Genson G, Kergoat GJ (2013) Fine-scale biogeographic and temporal diversification processes of peacock swallowtails (*Papilio* subgenus *Achillides*) in the Indo-Australian Archipelago. *Cladistics*, 29:88-111.

WILEY-BLACKWELL

Cladistics 29 (2013) 88–111

Cladistics

10.1111/j.1096-0031.2012.00412.x

Fine-scale biogeographical and temporal diversification processes of peacock swallowtails (*Papilio* subgenus *Achillides*) in the Indo-Australian Archipelago

Fabien L. Condamine[a,*], Emmanuel F.A. Toussaint[a], Adam M. Cotton[b], Gwenaelle S. Genson[a], Felix A.H. Sperling[c] and Gael J. Kergoat[a]

[a]*INRA, UMR 1062 Centre de Biologie pour la Gestion des Populations (INRA/IRD/CIRAD/Montpellier SupAgro), Campus International de Baillarguet, CS30016, 34988 Montferrier-sur-Lez, France;* [b]*86/2 Moo 5, Ban Hua Tung, Tambon Nong Kwai, Amphoe Hang Dong, Chiang Mai 50230 Thailand;* [c]*Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada*

Accepted 7 May 2012

Abstract

Explanations for the high species diversity of the Indo-Australian Archipelago are often challenged by the region's complex climatic and geological histories. Here, we investigated the evolutionary history of swallowtail butterflies of the *Papilio* subgenus *Achillides*, comprising up to 25 recognized species and about 100 subspecies distributed across the Indo-Australian Archipelago. To estimate the relative contributions of factors influencing their biodiversity, we used DNA sequences to infer the phylogeny and species limits of 22 species including most of their subspecies. We recovered a highly resolved and well-supported phylogeny for the subgenus, and clarified some taxonomic ambiguities at the species level. The corresponding DNA-based species phylogeny was then employed to reconstruct their historical biogeography using relaxed-clock and parametric-based analyses. Molecular dating and biogeographical analyses showed that *Achillides* originated around 19 Ma in Sunda + Wallacea. Biogeographical reconstructions indicated that geological vicariance shaped the early evolutionary history of *Achillides* whereas dispersal influenced late diversification. Birth–death likelihood analyses allowed exploration of their tempo and mode of diversification. We detected several shifts in diversification rates that are attributable to past climate-induced biogeographical events. By assessing both regional and fine-scale biodiversity patterns, this study brings new findings to a biogeographical understanding of the Indo-Australian Archipelago.

The Indo-Australian Archipelago encompasses a large proportion of the Earth's biodiversity, comprising 11 biodiversity hotspots and one of the world's three remaining tropical wilderness areas (Myers et al., 2000; Mittermeier et al., 2004). Biologists and naturalists have long been fascinated by the exceptional diversity of organisms in this region (e.g. birds of paradise) and its island systems have served as natural laboratories to study the evolutionary dynamics of island colonizations and radiations. The biogeography of the Indo-Australian Archipelago has been of great interest to evolutionary biologists since Wallace (1860, 1863), who first noted a major faunal discontinuity that is today well known as Wallace's line (Lohman et al., 2011). Surprisingly, the mechanisms responsible for the region's diversification remain poorly understood (Woodruff and Turner, 2009; Klaus et al., 2010; Müller and Beheregaray, 2010; Müller et al., 2010; for a review see Lohman et al., 2011). Understanding how organisms have diversified through time in the Indo-Australian Archipelago remains a significant challenge because numerous interdependent factors have probably been involved in shaping its present-day pattern of biodiversity.

Perhaps more than any other factor, geological history has been a major driving force of biogeographical diversification (Wiens and Donoghue, 2004;

*Corresponding author:
E-mail address: fabien.condamine@gmail.com

Lomolino et al., 2010). It is difficult to assess the effects of plate tectonics on biogeographical patterns because the origin, spread and radiation of many taxa took place when the Earth's surface had a different configuration (Ree and Sanmartín, 2009). The present-day geography of the Indo-Australian Archipelago is the result of complex geological events created by the convergence of the Australian, Eurasian and Indian plates (Hall, 2002; Metcalfe, 2005). India and Australia separated from Antarctica in the Early Cretaceous and Late Cretaceous, respectively (Scotese, 2001; Sanmartín and Ronquist, 2004). The two plates gradually drifted northward toward Eurasia (Scotese, 2001). Subsequently, there were three important periods in the geological development of the Indo-Australian Archipelago. First, the collision of India into Eurasia modified plate boundaries and initiated the formation of the Himalayas and the Tibetan Plateau in the Eocene about 45 Ma (Scotese, 2001). Consequently, mountain formation resulting from the collision led to major changes in habitats, climate and drainage systems, and promoted dispersal from Gondwana via India into Southeast Asia (e.g. Klaus et al., 2010). Second, the continued indentation of Asia by India further modified Sundaland, and the geographical changes of the Malay Peninsula began in the late Oligocene around 25 Ma (Hall, 1998, 2002; Metcalfe, 2005). Meanwhile, the Australian plate collided with Southeast Asia, resulting in major changes in the configuration of plate boundaries (Hall, 1998, 2002; Metcalfe, 2005). The progressive arrival of the Australian plate also provided possible pathways for dispersal between Asia and Australia, as well as forming new barriers (Hall, 1998; Lohman et al., 2011). Moreover, this collision favoured the creation of numerous islands in the region and New Guinea formed as a result of the collision of the Australian plate with the Pacific plate (Hall, 2002; Metcalfe, 2005). Third, the Sulawesi archipelago began to form following the movements of the tectonic plates and the collision of different terranes between 24 and 13 Ma (Wilson and Moss, 1999). Some of its biota may have vicariant origins from Sundaland in the west or New Guinea/Australia to the east (Wilson and Moss, 1999). The Philippines archipelago, whose complexity is well recognized (Honza and Fujioka, 2004), formed by accumulation of terranes and creation of oceanic islands during the middle Miocene approximately 15 Ma, and was completed 5 Ma. New Caledonia was part of the eastern margin of Gondwana until the Cretaceous (Cluzel et al., 2001; Pelletier, 2006) and split from Australia around 70–65 Ma (Grandcolas et al., 2008) but remained connected to New Zealand by a remnant Gondwanan fragment called Zealandia. While drifting eastwards, New Caledonia was completely immersed (Cluzel et al., 2001; Pelletier, 2006). Near 34–37 Ma, New Caledonia engaged the subduction zone initiated by the Australian and Pacific plates, which started the obduction of the first over the second, resulting in the formation of the current archipelago (Cluzel et al., 2001; Pelletier, 2006; Grandcolas et al., 2008).

These major tectonic events and geological changes were concomitant with a burst in the diversification of many groups of organisms (e.g. Bininda-Emonds et al., 2007; Hunt et al., 2007; Wahlberg et al., 2009), suggesting that geological history probably had a profound impact on the biotic diversification of the Indo-Australian Archipelago (Lohman et al., 2011). Nevertheless, studies documenting how the historical geology of the Indo-Australian Archipelago may have shaped the evolutionary history of a variety of taxa are scarce or have generally focused on particular areas (Yagi et al., 1999; Woodruff and Turner, 2009) with few studies recognizing the implications of Wallacea itself (Müller et al., 2010). Even if the region's highly complex biogeographical pattern can be accounted for by plate tectonic activity, other factors could have contributed to the diversification of biotas. Biogeographical consequences of these factors are manifold, but most stem from the effects of plate movements (Lomolino et al., 2010). Global circulation patterns of ocean and wind currents are not only controlled by equatorial-to-polar gradients of solar radiation and temperature, but also by the relative proportion of landmasses and water (Hall, 1998; de Queiroz, 2005). Oceanic currents and regional climates have also varied through time (Zachos et al., 2001; Miller et al., 2005) due to numerous past geological events (Hall, 2002). Although such changes have probably affected biotic diversification and distribution, identifying the involved processes remains an open question (Currie et al., 2004) particularly in the Indo-Australian Archipelago (Lohman et al., 2011). Pleistocene glaciations and linked sea-level fluctuations are often invoked as an important driving factor in recent speciation or extinction events (Voris, 2000; Woodruff and Turner, 2009; Lomolino et al., 2010, Lohman et al., 2011). Finally the geographical distributions of taxa can also be influenced by biotic factors such as interactions between species. It is especially true for phytophagous organisms (Winkler et al., 2009). When herbivores are highly specific, their distributions must depend in part on the availability of appropriate host plants, and consequently the geographical range of many specific phytophagous organisms correspond to those of their plant hosts (Becerra and Venable, 1999).

Swallowtail butterflies (Papilionidae) represent an ideal group of organisms to address biogeographical questions in the Indo-Australian Archipelago (Wallace, 1865). They exhibit high species richness especially in Sundaland and Wallacea, as well as a high level of endemism in all the corresponding biodiversity hotspots (e.g. Vane-Wright and de Jong, 2003). Despite being a fascinating group, since the pioneering work of Wallace

(1865) few studies have focused on how swallowtails originated and diversified in the Indo-Australian Archipelago (but see Zeuner, 1943; Yagi et al., 1999). The genus *Papilio*, in particular, is well represented in the region with about 100 species (Häuser et al., 2005) and corresponds to a highly complex and diverse assemblage. Although the genus *Papilio* has been the subject of many molecular phylogenetic studies (e.g. Caterino and Sperling, 1999; Reed and Sperling, 1999; Zakharov et al., 2004), these studies have never included comprehensive sampling of species occurring in the Indo-Australian Archipelago (Yagi et al., 2006). To better understand the fine-scale biogeographical patterns and processes of diversification of these subgenera, a comprehensive phylogenetic study is required.

Commonly named the peacock swallowtails, *Papilio* subgenus *Achillides* is particularly suitable to exploring diversification patterns because, first, it constitutes one of the most diverse clades of swallowtails in the Indo-Australian Archipelago. The subgenus *Achillides* is recovered as sister to an assemblage of the subgenus *Menelaides* (Southeast Asia) and part of subgenus *Princeps* (Africa and Southeast Asia) (Condamine et al., 2012). *Achillides* butterflies are widely distributed in this region, extending from Pakistan through Indo-China to Japan, and from the Malay Peninsula to Queensland (Australia) and New Caledonia (Fig. 1; see the Appendix S1). Nearly all species inhabit tropical or lowland rainforests, but some of them can be found in temperate areas or at high altitudes (e.g. *P. bianor*, *P. maackii*). In addition, many *Achillides* species have limited ranges, being restricted to long-isolated islands (e.g. *P. montrouzieri* in New Caledonia) or archipelagos (e.g. *P. neumoegeni* in Sumba, *P. pericles* in Timor) (see Fig. 1 for details).

Second, *Achillides* are popular among collectors, naturalists and researchers and so their taxonomy has been well documented (e.g. Harada, 1992; Izumi, 1993; Shimogori, 1997; Yoshimoto, 1998). Nevertheless, phylogenetic relationships within the subgenus remain unknown, and the number of species is variable among authors, ranging from 21 to 25 (see Shimogori, 1997; Bauer and Frankenbach, 1998; Häuser et al., 2005 for competing classifications). Traditionally *Achillides* has been split into four species groups, namely the *paris*, *palinurus*, *peranthus* and *ulysses* groups (Jordan, 1909; Munroe, 1961; Hancock, 1983). Although numerous taxonomic studies have focused on peacock swallowtails, the status of several species is still ambiguous, as is the status of more than 100 subspecies (Shimogori, 1997; Bauer and Frankenbach, 1998). For instance, *P. ulysses* (commonly known as the blue emperor) is mainly distributed in the Papua New Guinean region and contains at least 25 subspecies, most of which are isolated island endemics.

Third, *Achillides* species present a notable level of host plant specialization as they are known to feed exclusively on Rutaceae (Igarashi, 1984; Igarashi and Fukuda, 2000), a group of plants that is well distributed throughout the Indo-Australian Archipelago (Pfeil and Crisp, 2008). The distribution of phytophagous insects is dependent on those of their host plants (Becerra and Venable, 1999; Lomolino et al., 2010). Because any *Achillides* species is able to feed on a large array of Rutaceae (see Appendix S2), this suggests that they colonized an area previously occupied by Rutaceae species long before *Achillides* (see Pfeil and Crisp, 2008; Salvo et al., 2010 for an estimate of the origin of Rutaceae; Condamine et al., 2012). Thus, these kinds of biotic interactions arguably played a lesser role in the diversification of *Achillides* swallowtails.

Together, the above data suggest that the distribution pattern of *Achillides* diversity represents an excellent candidate for investigating biogeographical history and the causes that have promoted their diversification in the Indo-Australian Archipelago. Here, we aim to: (i) use dense taxon sampling to infer a robust phylogenetic framework for *Achillides* swallowtails; (ii) delimit several ambiguous taxa to re-examine Wallace's (1865) putative species clusters and their geographical distributions; (iii) investigate the historical biogeography of *Achillides* using a molecular relaxed-clock approach; and (*iv*) test the effect of palaeoclimate on diversification rates using birth–death likelihood analyses.

Material and methods

Taxon sampling

A total of 133 *Achillides* specimens were included in our study. Our sampling covers all species recognized by Shimogori (1997), Bauer and Frankenbach (1998) and Häuser et al. (2005) except three rare and endangered species (*Papilio buddha*, *P. chikae* and *P. elephenor*). Overall, this represents up to 22 described species recognized by different authors (Shimogori, 1997; Bauer and Frankenbach, 1998; Häuser et al., 2005). In addition, we attempted to include most of the described subspecies by covering 64% of the 106 known *Achillides* subspecies (Shimogori, 1997). The specimens were directly sampled in the field or were provided by collections of various institutes. One of the co-authors (A.M.C.) conducted taxonomic identifications of swallowtails, and all corresponding voucher materials were deposited in various research institutes (see Appendix S3). Permits were obtained by two co-authors (F.L.C. and A.M.C.) for all the species classified under the IUCN Red List of Threatened Species or in the CITES list. Because the outgroup selection is a crucial step in phylogenetics (Felsenstein, 2004), we relied on the most comprehensive and recent swallowtail phylogeny (Zakharov et al., 2004; Condamine et al., 2012) to

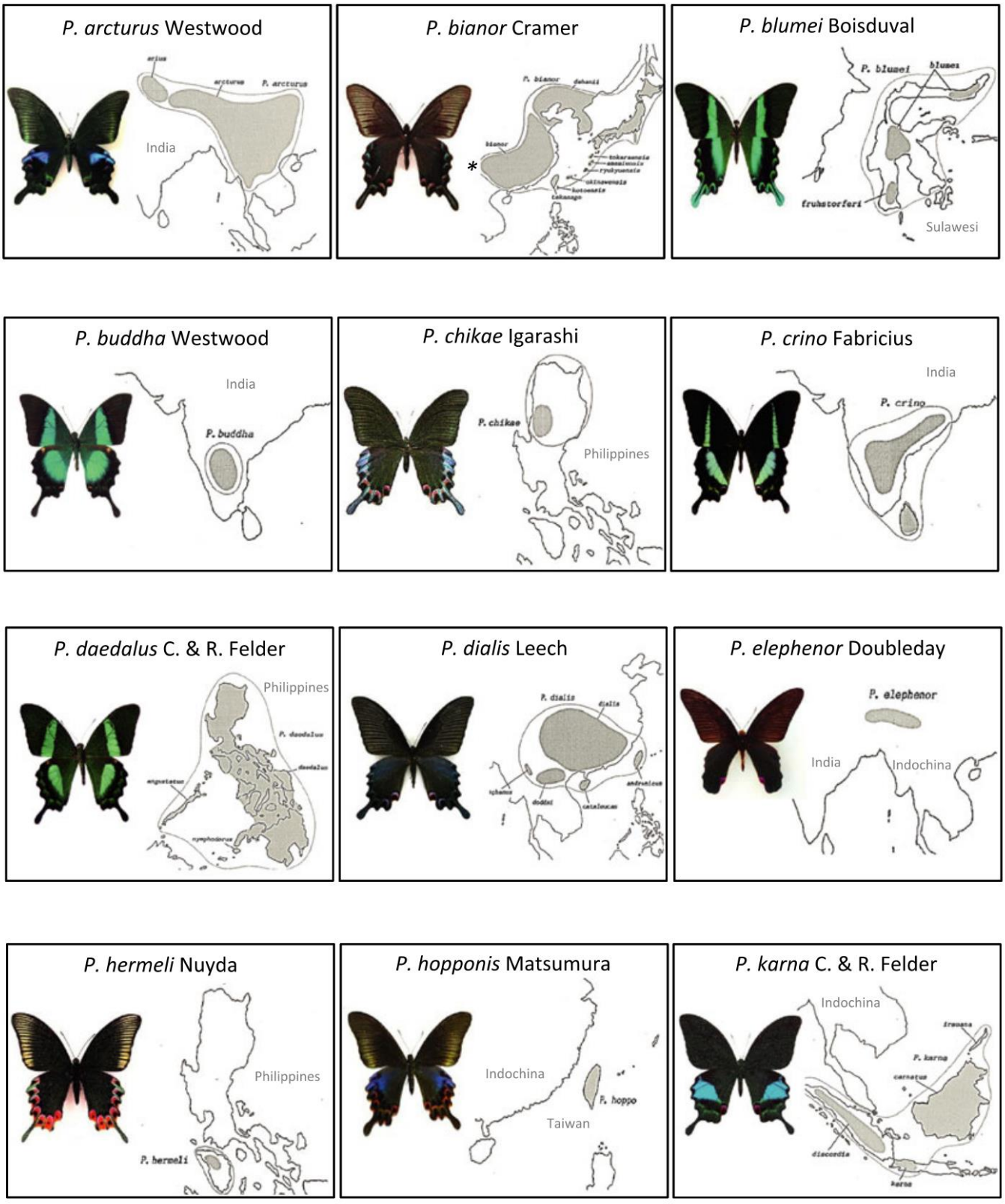


Fig. 1. Habitus of all species of the *Papilio* subgenus *Achillides* and their respective distributions (after Shimogori, 1997). For each species, the distribution of subspecies is also indicated. The asterisk indicates that the distribution of *P. bianor* now extends into the west through south-west China (*P. bianor gladiator*) to India (*P. bianor ganesa*) and Pakistan (*P. bianor polyctor*).

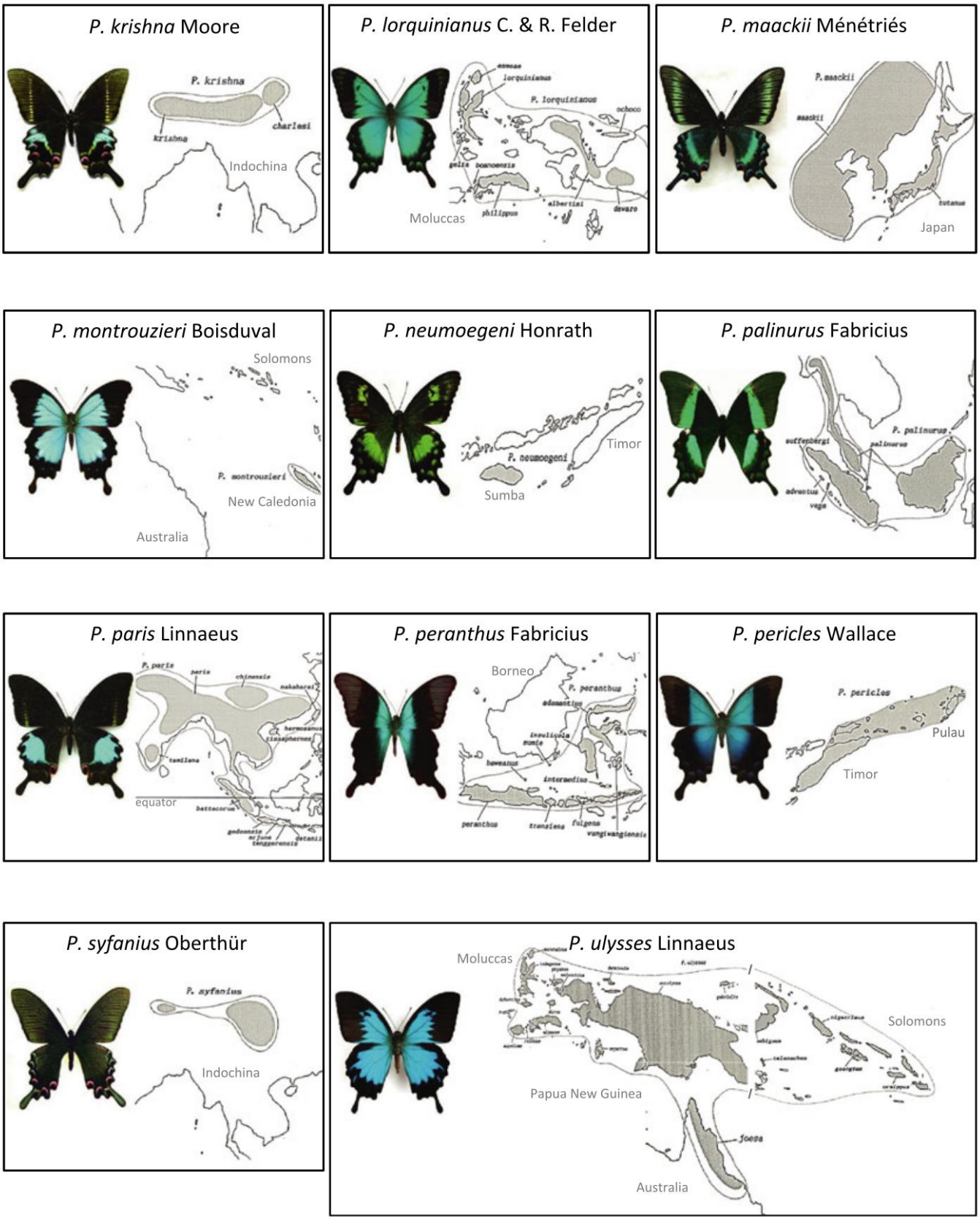


Fig. 1. (*Continued*).

choose four outgroups. These species (*Papilio demoleus*, *P. helenus*, *P. machaon* and *P. polytes*) were included as they represent subgenera that are close to the subgenus *Achillides* (Zakharov et al., 2004). Newly sampled taxa, sources of material and GenBank accession numbers for all the materials are given in Appendix S3 (new accession numbers are JQ982037–JQ982467).

Gene choice, molecular techniques and alignment

We followed previous work on swallowtail phylogeny that used DNA sequences (Caterino and Sperling, 1999; Reed and Sperling, 1999; Zakharov et al., 2004; Yagi et al., 2006) to expand the amount of analysed data available for global analyses. Here we used about 2.3 kb of three mitochondrial genes, namely cytochrome oxidase I (COI), ribosomal 16S RNA (16S) and NADH dehydrogenase 5 (ND5). We also used about 1.0 kb of the nuclear protein-coding gene elongation factor 1 alpha (EF-1α). The phylogenetic utility of these genes for Lepidoptera has been widely demonstrated (e.g. Cho et al., 1995; Zakharov et al., 2004; Simonsen et al., 2011), and a substantial dataset of lepidopteran sequences already exists for these genes. For newly sequenced species, total genomic DNA was extracted using the Qiagen® DNeasy tissue kit (Qiagen, Venlo, Netherlands). Polymerase chain reactions (PCR) were performed using the following programme: an initial 2 min denaturation at 94 °C, followed by 35 cycles of 1 min at 94 °C, 1 min at 45–52 °C (depending on primer combinations) and 1 min at 72 °C and then 7 min of final extension at 72 °C. Nucleotide sequences of the primers have been previously described and summarized (Simonsen et al., 2011). Positive PCR products were sent to Macrogen (Seoul, South Korea) and fragments were sequenced in both directions. Sequences were assembled into contiguous arrays using Geneious Pro 5.1 (available at: http://www.geneious.com/). The sequences were aligned with those retrieved from a previous study (Zakharov et al., 2004) using ClustalX 2.0 (Larkin et al., 2007) with default settings. Reading frames of COI, ND5 and EF-1α genes were checked under Mesquite 2.75 (available at: http://www.mesquiteproject.org).

Phylogenetic analyses

To estimate species relationships among *Achillides*, phylogenetic analyses were conducted using maximum parsimony (MP) and Bayesian inference (BI). Maximum parsimony analyses were carried out using TNT 1.1 (Goloboff et al., 2008). Initial heuristic searches were carried out using the Tree Bisection Reconnection (TBR) algorithm within the *Traditional Search* option in TNT, with random starting trees, 100 random-addition replicates and a *MaxTrees* value of 1000. More thorough analyses were further conducted using Random Sectorial Searches and Consensus-based Sectorial Searches (Goloboff, 1999), with the options for *Tree Ratchet*, *Tree Drifting* and *Tree Fusing* selected (Goloboff, 1999) within the *New Technology Search* option in TNT, with 100 random-addition replicates and with a *MaxTrees* value of 1000. For each analysis, 1000 non-parametric bootstrap replications were performed (standard sample with replacement).

Phylogenetic analyses were also performed with BI, which permits models of sequence evolution to be implemented (Holder and Lewis, 2003; Felsenstein, 2004; Nylander et al., 2004; Edwards, 2009). The dataset was combined and partitioned into five partitioning strategies (PS): (i) two partitions (one partition for the mitochondrial genes and one partition for the nuclear genes), (ii) four partitions (one partition per gene); (iii) four partitions (one partition per coding position plus one partition for 16S); (iv) seven partitions (one partition per coding position for the mitochondrial coding genes, one partition per coding position for the nuclear coding gene, one partition for 16S); and (v) ten partitions (one partition per coding position for each coding gene, one partition for the 16S). For each gene, the best-fit model of sequence evolution was selected with jModelTest (Posada, 2008) using both the corrected Akaike information criterion (AICc) and the Bayesian information criterion (BIC) (see Brown and Lemmon, 2007, for a discussion on the rationale for this setting).

Bayesian analyses were carried out with MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003). The BI analyses settings were: two independent runs with eight Markov chain Monte Carlo (MCMC) procedures (one cold and seven incrementally heated) that ran for 10^7 generations, sampling the trees every 100th cycle and each MCMC started from a random tree. A conservative burn-in of 25% was applied after checking for stability on the log-likelihood curves and the split-frequencies of the runs. All sample trees prior to reaching these generations were discarded and remaining trees were used to generate a 50% majority rule consensus tree.

Bootstrap values (BV) in MP and posterior probabilities (PP) in BI estimated node supports for the trees. Support was considered to be strong if BV ≥ 70% in MP (Hillis and Bull, 1993; Felsenstein, 2004) and PP ≥ 0.95 in BI (Erixon et al., 2003; Felsenstein, 2004). Selection of the best-fit partitioning strategy in BI analyses was performed using Bayes Factors (BF) (Kass and Raftery, 1995; Brown and Lemmon, 2007). Estimates of harmonic mean of the likelihood values, obtained with the *sump* command in MrBayes, are used to approximate the BF. We considered that BF values > 10 significantly favour one model over another (Kass and Raftery, 1995).

Species delimitation procedure

The main objective of the species delimitation procedure is to consistently determine putative molecular species clusters by classifying the observed branching time intervals defined by the nodes in a clock-constrained phylogeny as either being the result of interspecific (diversification) or intraspecific (coalescent) processes of lineage branching (Pons et al., 2006). Although this approach can be used to clarify taxonomic ambiguities, here we focus on using this method to delimit species clusters to better assess the geographical and temporal diversification pattern and processes in the evolution of *Achillides*.

For comparison purposes, these analyses were performed with the combined dataset (i.e. the phylogeny obtained with all the genes and the best-fit partitioning strategy), with an extended COI-barcode marker only (Hebert et al., 2003) and with the EF-1α nuclear gene only. From a specific topology, we inferred an arbitrary ultrametric tree with the mean path length algorithm using PATHd8 (Britton et al., 2007), we then used the General Mixed Yule Coalescent (GMYC) to perform analysis of species delimitation (Pons et al., 2006). The method is implemented in script codes running under the R 2.13 package using functions from the *ape* library (Paradis et al., 2004) and the *splits* library (available at http://www.r-forge.r-project.org/projects/splits/). First, we visualized the branching time data by plotting the log of the number of lineages through time. A transition in branching rates should be visible as a sudden increase in slope of the plot towards the present. Second, we ran the threshold version of the GMYC model on the calibrated ultrametric tree of *Achillides*. Third, we compared the likelihood with that obtained assuming a single branching process for the tree. For the GMYC model, assuming all species have the same parameter values, the threshold model has five parameters (λ1, λ2, p1, p2 and T), whereas the null model has two (λ1 and p1); hence, there are three degrees of freedom for the comparison.

The delimitation of molecular species was used to build a "reduced dataset". This new dataset comprised a single sequence for each recovered species cluster. A sequence for a species cluster was obtained from the consensus of all the sequences constituting a species cluster using Geneious 5.1 and default settings. This reduced dataset was used to perform MP and BI analyses with the corresponding aforementioned settings. This topology was used for estimates of divergence times and analyses of geographical range evolution.

Bayesian estimates of divergence times

The applicability of a molecular clock was preliminarily investigated for each node using PATHd8 (Britton et al., 2007). As the hypothesis of a molecular clock was not statistically supported for our dataset, methods of dating that account for rate variation across lineages were used (Thorne et al., 1998; Sanderson, 2002; Drummond et al., 2006). Here we implement Bayesian relaxed-clock (BRC) analyses, which use MCMC procedures to approximate the posterior distribution of rates and divergence times and simultaneously infer their credibility intervals. These molecular relaxed-clock methods have also introduced flexible techniques for incorporating calibrations leading to a discussion about approaches to calibrating estimates of divergence times (Yang and Rannala, 2006; Hug and Roger, 2007; Ho and Phillips, 2009). In our study, the BRC analyses were carried out with the program BEAST 1.6.2 (Drummond and Rambaut, 2007), which uses the log-normal model of Thorne et al. (1998), which is uncorrelated in BEAST. The xml-file for the BEAST analysis was created under the interface BEAUti (included in the BEAST package) with the following non-default settings and priors: the *Site Model* was set based on the model used in the original Bayesian analyses, the *Clock Model* was set to a relaxed clock with uncorrelated rates, the *Tree Model* was set to a Yule process of speciation, and the *MCMC parameters* were fixed to 2.5×10^7 or 5×10^7 generations with sampling every 1000 generations and the first 25% discarded as burn-in. The remaining parameters were left as default. Several independent analyses have been performed to check the likelihood scores under the LogCombiner and Tracer programs (included in the BEAST package) using the effective sample size criterion (Drummond et al., 2006).

In BRC analyses, a fundamental step is the choice of calibration points to adjust the molecular clock (Ho and Phillips, 2009). The selection and parameterization of calibration points is sensitive especially when considering further biogeographical analyses in the study (problem of circularity). Geological calibration, in particular, represents maximum ages for nodes (Ho and Phillips, 2009), but using this calibration is very dangerous because taxa might well have begun to diverge before the geological separation and also because geological reconstructions are not immutable (Heads, 2005; Sanmartín et al., 2008). For swallowtails, the fossil record is scarce and the three known unambiguous papilionid fossils do not belong to the genus *Papilio* (Condamine et al., 2012). We chose to use previous results on divergence times of the entire family Papilionidae based on the most comprehensive taxon sampling and molecular dataset, including all the available fossils (Condamine et al., 2012). The age and divergence times of Papilionidae were inferred using fossils as minimum age constraints bounded with maximum age corresponding to the origin of angiosperms (estimated from molecular dating analyses, *sensu* Bell et al., 2010). Thus, we followed the recommendations of several authors

(e.g. Yang and Rannala, 2006; Hug and Roger, 2007; Ho and Phillips, 2009) who advocated the use of a uniform distribution or normal distribution for secondary calibrations. As we recovered a similar phylogenetic pattern, we choose to use previous age estimates for the nodes that correspond to similar common ancestors. Five calibration points were retrieved and were set to uniform prior. The root was constrained to be within 19.17–28.61 Ma, the node between *P. demoleus* and *Achillides* was set to 16.74–25.24 Ma, that of *P. demoleus* and *P. helenus* constrained to 15.87–24.02 Ma, that of *P. helenus* and *P. polytes* set to 11.6–17.88 Ma, and finally the crown of *Achillides* was constrained to be within 13.09–21.85 Ma (for details see Condamine et al., 2012).

Historical biogeography

Ancestral areas and geographical speciation scenarios for *Achillides* were inferred using the dispersal–extinction–cladogenesis (DEC) model (Ree et al., 2005; Ree and Smith, 2008) of range evolution under the program Lagrange (available at: http://www.code.google.com/p/lagrange/). We chose to use this method because the DEC model describes ancestor–descendant transitions between geographical ranges by anagenetic and cladogenetic processes such as range expansion (dispersal), range contraction (local extinction) and range subdivision/inheritance (vicariance). When the ancestral range includes two or more areas, DEC requires that between-area vicariance events separate a single area from the remainder of the ancestral range. Furthermore, unlike DIVA, DEC allows speciation within one area of a wide ancestral range, which results in one descendant occupying the area of speciation whereas the other inherits the entire range (peripatric speciation). The method specifies the likelihood of species-range data arrayed at the tips of a phylogenetic tree as a function of rates of dispersal and local extinction (Ree et al., 2005; Ree and Smith, 2008), hereafter called the ML-DEC method.

The distribution of *Achillides* extends across the Oriental and Australasian regions. We divided these two biomes into smaller biogeographical identities to get more resolution in the inference of the ancestral area for the root. For that we defined our areas using paleogeography arguments with tectonic reconstructions (e.g. Scotese, 2001; Hall, 2002; Metcalfe, 2005; Pelletier, 2006), using biodiversity hotspot information (Myers et al., 2000; Mittermeier et al., 2004), and the present-day distribution of *Achillides* swallowtails (Shimogori, 1997). The model comprised 11 component areas delimited and listed in Appendix S1 (species richness is indicated for each area).

Species ranges were defined by presence–absence data by excluding marginal distribution or human introduction (Hines, 2008; Nylander et al., 2008). The 11 areas yield a set of $2^{11} = 2048$ theoretically possible geographical ranges (area subsets), many of which were excluded from consideration based on the biological implausibility of their spatial configurations (e.g. wide disjunction between Japan and New Caledonia). The goal is to define a palaeogeological model taking into account the geological history of the Earth with biologically plausible ranges (i.e. adjacent areas). We also discarded ranges larger than three areas in size that were not subsets of observed species ranges. The motivation for this step was to create the best-fit biogeographical model in relation to the known geological history of Earth. One benefit of this approach is to reduce the dimensions of the model's transition matrix, thus increasing its computational feasibility.

Following the principles described by Ree and Smith (2008), we constructed temporal constraints on rates of dispersal between areas based on palaeogeographical reconstructions of area position through time (e.g. Scotese, 2001; Hall, 2002; Metcalfe, 2005; Pelletier, 2006). These constraints were implemented as a series of four time slices. All time slices together spanned the past 20 Ma, with each slice being 5 Ma in duration. For each time slice, we constructed a matrix of scaling factors (between zero and 0.5) for the dispersal rate between areas according to their geographical position, interpreting greater distances and/or the extent of geographical barriers (e.g. sea straits, mountain chains) as being inversely proportional to the expected rate.

An obvious advantage of the ML-DEC method in biogeography is the statistical framework permitting local optimization for the root (Ree and Smith, 2008). We performed specific tests in which the root was constrained to be either a single area, or the union of two and three areas. A 2-log likelihood unit's threshold was used to choose which area was supported (Ree et al., 2005; Ree and Smith, 2008). This likelihood framework also allows specific geographical scenarios along the chronogram.

Diversification rates

To investigate the tempo and mode of species diversification rates over time, we followed a step-by-step procedure under the R 2.13 package with the *ape* (Paradis et al., 2004) and *laser* libraries (Rabosky, 2006b). Accumulation of lineages through time was graphically visualized using lineages-through-time (LTT) plots for all species. Second, we estimated the overall diversification rate (speciation minus extinction) under a simple birth–death model (Magallón and Sanderson, 2001), with net diversification rates being calculated for three relative extinction rates ($\varepsilon = 0/0.5/0.9$). We then tested the null hypothesis that per-lineage speciation and extinction rates have remained constant through time with the γ-statistics

under the pure birth (Yule) process (constant speciation rate; Pybus and Harvey, 2000). Fourth, we tested the hypothesis of rate shifts using a likelihood-based method to compare models with a constant diversification rate with those with one or more rate shifts (Rabosky, 2006a). Single-rate and two-rate models were compared by AIC and likelihood ratio tests (Rabosky, 2006a). For all analyses the tests were also conducted using simulated incomplete phylogenies with a Monte Carlo constant-rates test to account for the effects of taxon sampling on our conclusions, namely the lack of three species (*P. buddha*, *P. chikae* and *P. elephenor*). Finally, we tested the hypothesis of rate shifts occurring at major decreases or increases on the LTT plot by using a likelihood-based method to compare models with a constant diversification rate with those with one or more rate shifts (Rabosky, 2006a; Rabosky and Lovette, 2008). We did not include extinction rate in this model because estimates under a constant rate birth–death model tended toward 0, and it is difficult to obtain meaningful estimates of separate extinction and speciation parameters under discrete-shift models similar to those considered here (confidence intervals on extinction rates are very large; Rabosky, 2006a; Rabosky and Lovette, 2008).

Table 1
Selection of substitution model of sequence evolution using both the Bayesian information criterion (BIC) and the corrected Akaike information criterion (AICc)

Partition	BIC	AICc
All combined genes	GTR + G	GTR + G
16S rRNA	HKY + G	HKY + G
CO1	GTR + G + I	GTR + G + I
ND5	TrN + G	TrN + G
EF-1α	TrNef + G	TrNef + G
All coding genes (1st position)	TrN + G	GTR + G
All coding genes (2nd position)	HKY + G	TPM1uf + G
All coding genes (3rd position)	TPM1uf + G	GTR + G
Mitochondrial coding genes (1st position)	TrN + G	TrN + G
Mitochondrial coding genes (2nd position)	HKY + G	HKY + G
Mitochondrial coding genes (3rd position)	TPM1uf + G	TPM1uf + G
CO1 (1st position)	TrN + G	TrN + G
CO1 (2nd position)	HKY + G	F81 + G
CO1 (3rd position)	GTR + G	TPM1uf + G
ND5 (1st position)	HKY + G	HKY + G
ND5 (2nd position)	F81	JC
ND5 (3rd position)	TrN + G	TrN + G
EF1a (1st position)	F81	F81
EF1a (2nd position)	JC	JC
EF1a (3rd position)	HKY + G	K80 + G

Results

Phylogenetic relationships

The molecular matrix comprises 3259 nucleotides for 137 individuals (133 *Achillides* and four outgroups; see Appendix S3). We sequenced 67 subspecies among the 106 that are recognized in the most comprehensive work of Shimogori (1997). Under MP, Traditional Search analyses using the TBR algorithm in TNT yield 60 equally parsimonious trees (length = 2227). Only 13 equally parsimonious but not shorter trees (length = 2227) were recovered using the New Technology Search and the various algorithms in TNT. Both MP analyses recovered a similar phylogenetic pattern. The selection of the substitution model of sequence evolution is described for each partition in Table 1. In the BI analyses, the partitioning strategy with ten partitions was selected as the best-fit strategy through the BF comparisons (Table 2). The different PS yield similar topologies that differ only by the terminal branching of some taxa, and the low average split frequencies values revealed good convergence between the two runs (Table 2).

Combined and partitioned BI and MP analyses found similar phylogenetic relationships (Fig. 2). Overall, for MP and BI analyses, nodes have moderate to strong support values (BV and PP). Most importantly, almost all internal nodes and all the species nodes are supported by BV > 70% and PP = 1.0 (low supports are only recovered at the intraspecific level, Fig. 2). The subgenus *Achillides* is recovered as monophyletic with strong support (BV > 70% and PP = 1.0). Within the subgenus, we recovered a phylogenetic pattern with four clades labelled C1–C4 in Fig. 2. First, a two-species clade (*P. montrouzieri* and *P. ulysses*) was found in a sister position to the remaining *Achillides* species with high support (BV > 70% and PP = 1.0; C1 in Fig. 2). The species *P. crino*, labelled C2, was recovered in a sister position to a large clade (comprising C3 and C4) in all methods used, as opposed to its traditional placement within the *palinurus* group, although support for that position was moderate to strong (BV = 63% and PP = 1.0). The third and fourth clades were always recovered as sister groups in all MP and BI analyses (BV = 63% and PP = 0.92). Clade C3 (BV = 74% and PP = 1.0) comprises *P. blumei*, *P. daedalus*, *P. lorquinianus*, *P. neumoegeni*, *P. palinurus*, *P. peranthus* and *P. pericles*. Clade C4 (BV = 90% and PP = 1) encompasses the remaining species (*P. arcturus*, *P. bianor*, *P. dehaanii*, *P. dialis*, *P. hermeli*, *P. hopponis*, *P. karna*, *P. krishna*, *P. maackii*, *P. paris* and *P. syfanius*). Only two branching differences were recovered between MP and BI analyses (indicated by NA on Fig. 2). In MP, *P. krishna* is sister to *P. maackii* + *P. syfanius* (BV < 50%), whereas it is sister to *P. arcturus* + *P. hermeli* in

Table 2
Scores obtained by Bayesian Inference with different partitioning strategies and tests on the monophyly of some species

	Harmonic mean (BIC)	Harmonic mean (AICc)	No. of parameters	BF
Partitioning strategies (PS)				
PS 1 (two partitions)	17 457.15	17 457.15	25	**
PS 2 (four partitions)	17 104.00	17 104.00	49	**
PS 3 (four partitions)	16 847.22	16 841.34	49	**
PS 4 (seven partitions)	16 781.86	16 928.30	85	Best
PS 5 (ten partitions)	16 782.89	17 113.78	121	n.s.
Tests on the monophyly of some species				
P. paris monophyletic	16 788.24	16 941.82	85	*
P. maackii and *P. syfanius* monophyletic	16 804.88	16 954.13	85	**

The first part of the table comprises the results of the different Bayesian analyses using several partitioning strategies (PS 1–5). Harmonic means and number of parameters are used to estimate Bayes Factors (BF) to select the best-fit partitioning strategy for the dataset. The second part of the table corresponds to tests on the monophyly of some species recovered as paraphyletic with the best-fit PS. We constrained *P. paris* and *P. maackii* + *P. syfanius* to be monophyletic. As for the first part, the same scores are summarized for the three constrained analyses. To estimate BF for each analysis, we used the following formula: $BF = 2 \times (\ln L1 - \ln L0) + (P1 - P0) \times \ln(0.01)$ (Kass and Raftery, 1995).

n.s., non-significant; *, significant; **, very significant.

BI (PP = 0.47). *Papilio hermeli* is sister to the *bianor* group in MP (BV < 50%), but sister to *P. dialis* in BI (PP = 0.84).

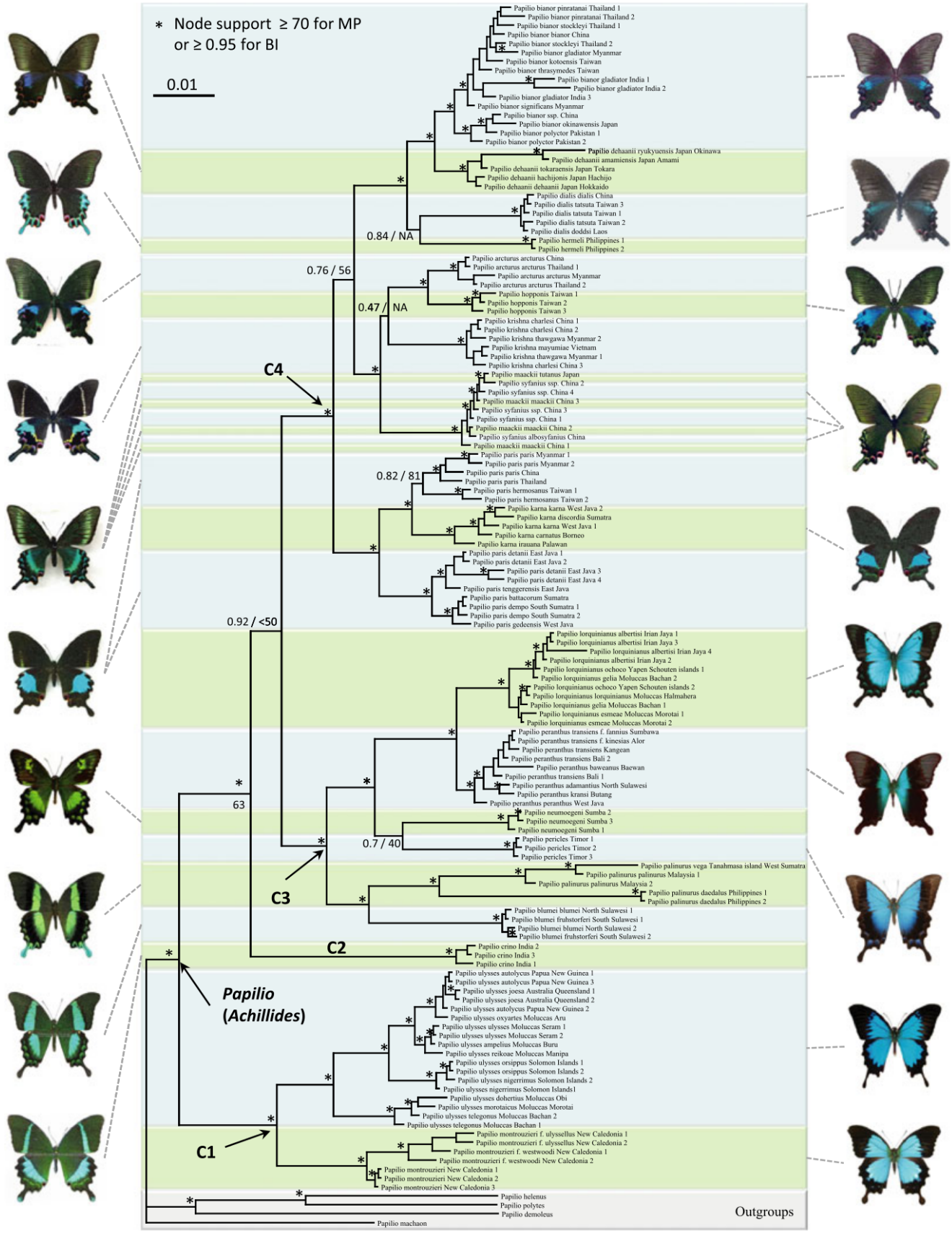
In addition, three species were recovered as paraphyletic in single gene analyses (especially with the nuclear gene EF-1α) and combined partitioned analyses as well, namely *P. paris*, *P. maackii* and *P. syfanius* (Fig. 2). *Papilio paris* is split into two strongly supported clades (BV > 70% and PP = 1.0), separated by a clade that encompasses all sampled individuals of *P. karna*. These two groups are geographically well structured: the first occurs in the Asian mainland whereas the second is distributed in Sundaland. The other case of paraphyly corresponds to the species *P. maackii* and *P. syfanius*, which are mixed in a single well-supported clade in all phylogenetic analyses (BV > 70% and PP = 1). Supplementary analyses were performed for each case of species paraphyly (Table 2): a first analysis was conducted with all sampled individuals of *P. paris* constrained to monophyly, and in a second analysis, members of *P. maackii* and *P. syfanius* were constrained to reciprocal monophyly. For each constrained analysis, BFs were assessed through the estimation of harmonic means. These analyses reveal that the paraphylies are confirmed because the unconstrained topology is significantly better than the constrained topologies enforcing the monophyly of *P. paris* or *P. maackii* + *P. syfanius* (Table 2).

Species delimitation

The existence of distinct phylogenetic lineages was corroborated by the analysis of the branching rate pattern. An LTT plot based on the ultrametric tree demonstrated a sudden increase in branching rate towards the present, corresponding to the switch from interspecific to intraspecific branching events (Fig. 3). To determine the position of the switch, the method of Pons et al. (2006) was applied to the ultrametric tree. The GMYC model was preferred over the null model of uniform branching rates (logL = 670.70, compared with the null model logL = 662.60; $2\Delta L = 16.196$, χ^2 test, d.f. = 3, $P < 0.00103$). The model fitted the switch in the branching pattern occurring at −0.157053 (i.e. T of the ML solution, the time separating the ingroup root from the present was arbitrarily assigned to 1), leading to an estimate of 26 putative species (Fig. 3). Overall, we recovered four putative species in clade C1, eight in clade C3 and 14 in clade C4 (Fig. 3). Note that four species (*P. bianor*, *P. dehaanii*, *P. paris* and *P. ulysses*) appear to correspond to several putative species (two, two, three and three, respectively) following this approach (therefore they are labelled with distinct numbers in Fig. 3) and two species are combined into one putative species [*P. maackii* and *P. syfanius*, hereafter *P. maackii sensu lato* (s.l.)]. The confidence interval for the threshold gave a range of 18–31 putative species with likelihood scores ranging from 668.85 to 668.89 respectively (i.e. estimates falling within 2 log-likelihood units of the ML solution).

By comparison, this approach was also applied to the topology obtained with the COI gene alone (the extended-barcode marker; see Appendix S4). The GMYC model was also preferred over the null model (logL = 313.44, compared with logL = 320.88; $2\Delta L = 14.878$, χ^2 test, d.f. = 3, $P < 0.0019$). By contrast, the

Fig. 2. Phylogenetic relationships of the *Papilio* subgenus *Achillides* obtained with the best-fit partitioning strategy (seven partitions) under Bayesian inference. Posterior probabilities (PP) and maximum parsimony bootstrap values (BV) are shown by nodes (an asterisk indicates PP ≥ 0.95 or BV ≥ 70%). NA indicates that MP analyses did not recover this topology. When no values are shown (especially at the intraspecific level) the BV and PP were ≤ 70% and 0.95, respectively. For deeper nodes, values are indicated. At the tips, the species and subspecies name is given, plus locality. A blue or green horizontal band delimits each species. *Achillides* and four major clades (C1–C4) are labelled.


* Node support ≥ 70 for MP
or ≥ 0.95 for BI
0.01
Papilio bianor pinratanai Thailand 1
Papilio bianor pinratanai Thailand 2
Papilio bianor stockleyi Thailand 1
Papilio bianor bianor China
Papilio bianor stockleyi Thailand 2
Papilio bianor gladiator Myanmar
Papilio bianor kotoensis Taiwan
Papilio bianor thrasymedes Taiwan
Papilio bianor gladiator India 1
Papilio bianor gladiator India 2
Papilio bianor gladiator India 3
Papilio bianor significans Myanmar
Papilio bianor ssp. China
Papilio bianor okinawensis Japan
Papilio bianor polyctor Pakistan 1
Papilio bianor polyctor Pakistan 2
Papilio dehaanii ryukyuensis Japan Okinawa
Papilio dehaanii amamiensis Japan Amami
Papilio dehaanii tokaraensis Japan Tokara
Papilio dehaanii hachijonis Japan Hachijo
Papilio dehaanii dehaanii Japan Hokkaido
Papilio dialis dialis China
Papilio dialis tatsuta Taiwan 3
Papilio dialis tatsuta Taiwan 1
Papilio dialis tatsuta Taiwan 2
Papilio dialis doddsi Laos
Papilio hermeli Philippines 1
Papilio hermeli Philippines 2
0.84 / NA
Papilio arcturus arcturus China
Papilio arcturus arcturus Thailand 1
Papilio arcturus arcturus Myanmar
Papilio arcturus arcturus Thailand 2
0.76 / 56
Papilio hopponis Taiwan 1
Papilio hopponis Taiwan 2
Papilio hopponis Taiwan 3
0.47 / NA
Papilio krishna charlesi China 1
Papilio krishna charlesi China 2
Papilio krishna thawgawa Myanmar 2
Papilio krishna mayumiae Vietnam
Papilio krishna thawgawa Myanmar 1
Papilio krishna charlesi China 3
C4
Papilio maackii tutanus Japan
Papilio syfanius ssp. China 2
Papilio syfanius ssp. China 4
Papilio maackii maackii China 3
Papilio syfanius ssp. China 3
Papilio syfanius ssp. China 1
Papilio maackii maackii China 2
Papilio syfanius albosyfanius China
Papilio maackii maackii China 1
Papilio paris paris Myanmar 1
Papilio paris paris Myanmar 2
Papilio paris paris China
Papilio paris paris Thailand
0.82 / 81
Papilio paris hermosanus Taiwan 1
Papilio paris hermosanus Taiwan 2
Papilio karna karna West Java 2
Papilio karna discordia Sumatra
Papilio karna karna West Java 1
Papilio karna carnatus Borneo
Papilio karna irauana Palawan
Papilio paris detanii East Java 1
Papilio paris detanii East Java 2
Papilio paris detanii East Java 3
Papilio paris detanii East Java 4
Papilio paris tenggerensis East Java
Papilio paris battacorum Sumatra
Papilio paris dempo South Sumatra 1
Papilio paris dempo South Sumatra 2
Papilio paris gedeensis West Java
0.92 / <50
Papilio lorquinianus albertisi Irian Jaya 1
Papilio lorquinianus albertisi Irian Jaya 3
Papilio lorquinianus albertisi Irian Jaya 4
Papilio lorquinianus albertisi Irian Jaya 2
Papilio lorquinianus ochoco Yapen Schouten islands 1
Papilio lorquinianus gelia Moluccas Bachan 2
Papilio lorquinianus ochoco Yapen Schouten islands 2
Papilio lorquinianus lorquinianus Moluccas Halmahera
Papilio lorquinianus gelia Moluccas Bachan 1
Papilio lorquinianus esmeae Moluccas Morotai 1
Papilio lorquinianus esmeae Moluccas Morotai 2
Papilio peranthus transiens f. fannius Sumbawa
Papilio peranthus transiens f. kinesias Alor
Papilio peranthus transiens Kangean
Papilio peranthus transiens Bali 2
Papilio peranthus baweanus Baewan
Papilio peranthus transiens Bali 1
Papilio peranthus adamantius North Sulawesi
Papilio peranthus kransi Butang
Papilio peranthus peranthus West Java
63
Papilio neumoegeni Sumba 2
Papilio neumoegeni Sumba 3
Papilio neumoegeni Sumba 1
0.7 / 40
Papilio pericles Timor 1
Papilio pericles Timor 2
Papilio pericles Timor 3
Papilio palinurus vega Tanahmasa island West Sumatra
Papilio palinurus palinurus Malaysia 1
Papilio palinurus palinurus Malaysia 2
Papilio palinurus daedalus Philippines 1
Papilio palinurus daedalus Philippines 2
C3
Papilio blumei blumei North Sulawesi 1
Papilio blumei fruhstorferi South Sulawesi 1
Papilio blumei blumei North Sulawesi 2
Papilio blumei fruhstorferi South Sulawesi 2
C2
Papilio crino India 2
Papilio crino India 3
Papilio crino India 1
Papilio ulysses autolycus Papua New Guinea 1
Papilio ulysses autolycus Papua New Guinea 3
Papilio ulysses joesa Australia Queensland 1
Papilio ulysses joesa Australia Queensland 2
Papilio ulysses autolycus Papua New Guinea 2
Papilio ulysses oxyartes Moluccas Aru
Papilio ulysses ulysses Moluccas Seram 1
Papilio ulysses ulysses Moluccas Seram 2
Papilio ulysses ampelius Moluccas Buru
Papilio ulysses reikoae Moluccas Manipa
Papilio ulysses orsippus Solomon Islands 1
Papilio ulysses orsippus Solomon Islands 2
Papilio ulysses nigerrimus Solomon Islands 2
Papilio ulysses nigerrimus Solomon Islands1
Papilio ulysses dohertius Moluccas Obi
Papilio ulysses morotaicus Moluccas Morotai
Papilio ulysses telegonus Moluccas Bachan 2
Papilio ulysses telegonus Moluccas Bachan 1
Papilio (Achillides)
C1
Papilio montrouzieri f. ulyssellus New Caledonia 1
Papilio montrouzieri f. ulyssellus New Caledonia 2
Papilio montrouzieri f. westwoodi New Caledonia 1
Papilio montrouzieri f. westwoodi New Caledonia 2
Papilio montrouzieri New Caledonia 1
Papilio montrouzieri New Caledonia 2
Papilio montrouzieri New Caledonia 3
Papilio helenus
Papilio polytes
Papilio demoleus
Papilio machaon
Outgroups

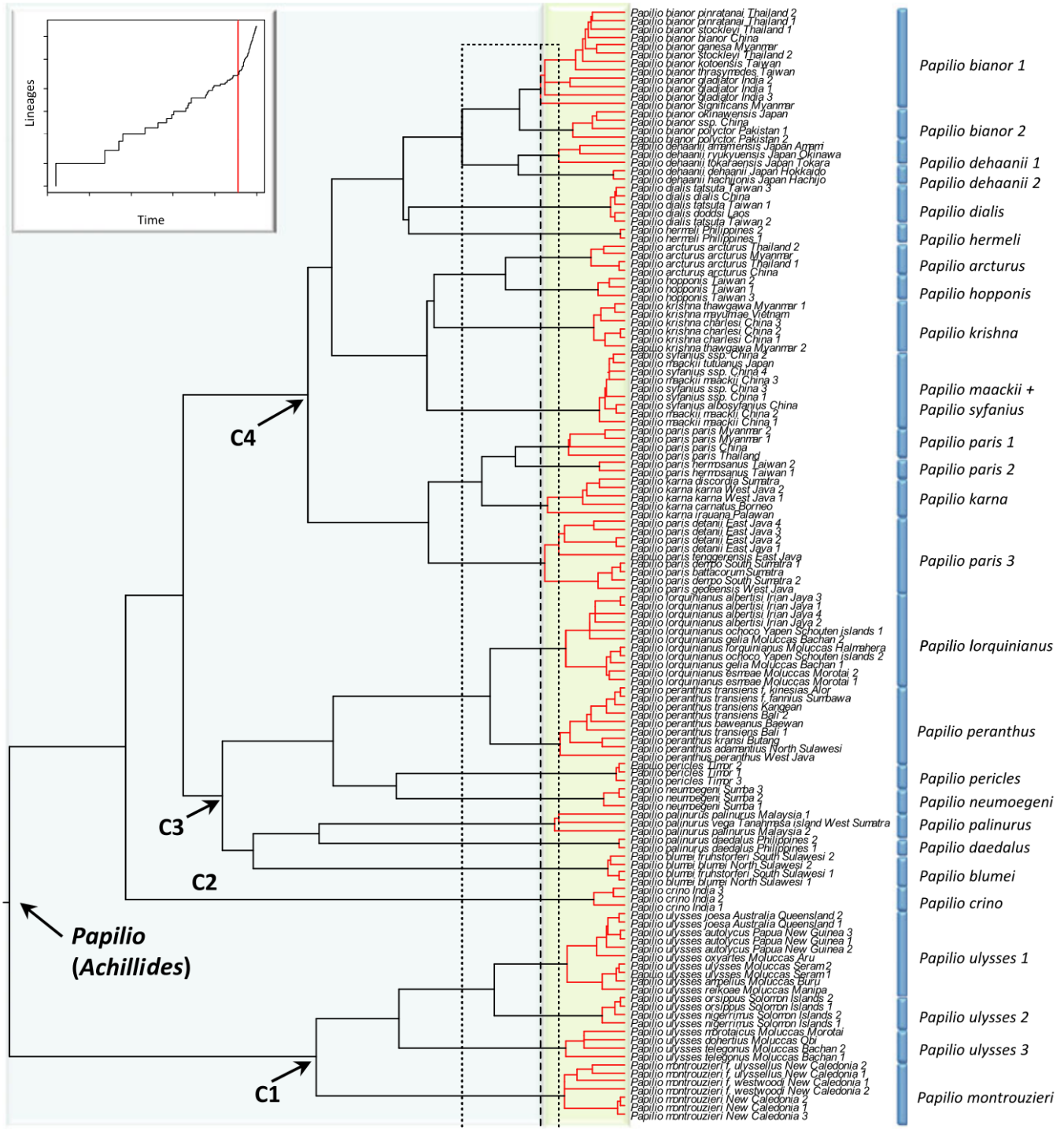


Fig. 3. Ultrametric tree of the *Papilio* subgenus *Achillides* obtained with PATHd8, showing clusters of specimens recognized as putative species. Genetic clusters recognized as a putative species are highlighted in red and separated by longer black branches. Blue and green background shadings represent the threshold (vertical black dashed line) between putative interspecific and intraspecific branching, respectively. The top corner graph shows the lineages-through-time plot based on the ultrametric tree. The sudden increase in branching rate, indicated by a red line, corresponds to the shift from interspecific to intraspecific lineage branching. The vertical bars group all sequences within each significant cluster, labelled by a temporary species name. The four major clades are labelled (C1–C4).

analysis recovered 33 putative species, with a confidence interval ranging from 19 to 44 estimated species (likelihood scores of 315.20 and 315.24, respectively). For details in the difference of species delimitation results between the multi-markers approach and the barcode approach, see Appendix S4. We conservatively chose to keep the number of species obtained with the multiple marker approach for the set of taxa used for the biogeographical analyses.

Estimates of divergence times and historical biogeography

The results of divergence time analysis under a BRC method are represented in Fig. 4. The complete list of median ages and their 95% highest posterior density (HPD) for each node are given in Table 3. Using Tracer, we observed through effective sample size (threshold fixed to 200) that all runs of the two analyses (2.5×10^7 and 5×10^7 generations) provide similar ages for *Achil-*

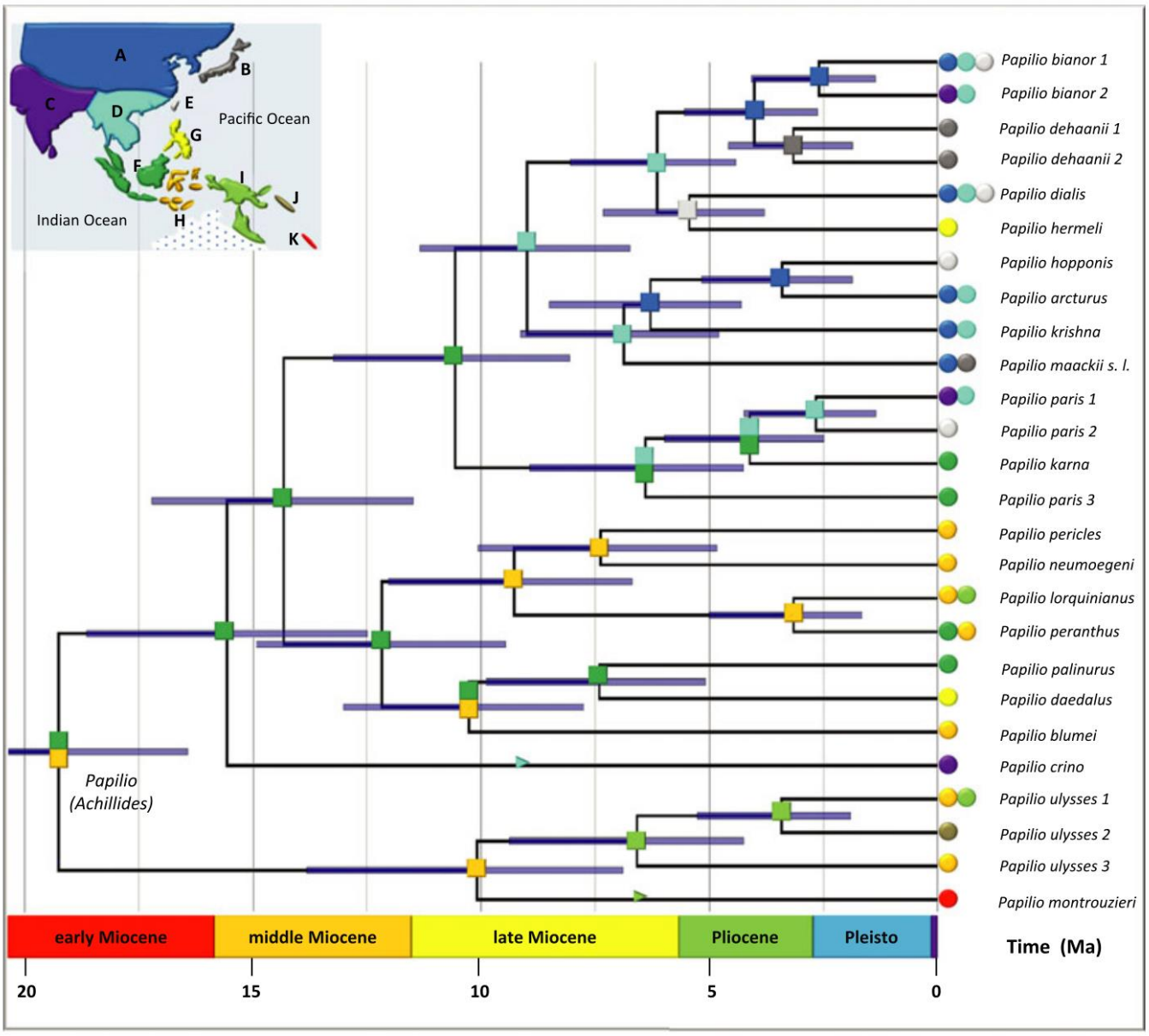


Fig. 4. Historical biogeography of the *Papilio* subgenus *Achillides*. The Bayesian chronogram shows phylogenetic relationships of *Achillides* (outgroups removed), with ages given as Ma and the 95% highest posterior density shown for each node. A 5-Myr timescale is placed at the bottom of the chronogram spanning the epochs since 20 Ma. The top left corner map represents the Indo-Australian Archipelago delimited into 11 areas. For each node, the biogeographical reconstruction is indicated by a coloured square for the ML-DEC (representing the most likely ancestral area). Colours of squares correspond to the coloured area on the map. Coloured triangles on branches indicate dispersal (optimizations of the ML-DEC analyses) from the ancestral region to the present-day region. The present-day distribution of each species is informed at the tips by coloured circles referring to the coloured areas on each map.

Table 3
Combined results of phylogeny, estimates of divergence times and biogeography of *Papilio* subgenus *Achillides*

		Phylogeny		BRC chronogram		ML-DEC reconstructions		
Taxonomy	Nodes	BV	PP	Age	95% CI	Range(s)	ln L	RP
Root*	1	NA	NA	22.86	19.17–26.8	NA	NA	NA
Papilio demoleus–Papilio (*Achillides*)*	2	100	NA	21.27	18.14–24.45	NA	NA	NA
*Papilio demoleus–Papilio helenus**	3	100	0.97	18.74	15.9–21.7	NA	NA	NA
*Papilio polytes–Papilio helenus**	4	100	1	13.12	11.6–15.87	NA	NA	NA
Papilio (*Achillides*)*	5	90	1	19.25	16.42–21.85	FH	89.54	Best
Clade 1	6	100	1	10.09	6.89–13.81	H	90.16	0.537
Papilio ulysses 1–*Papilio ulysses* 3	7	99	1	6.59	4.25–9.38	HI	90.74	0.302
Papilio ulysses 1–*Papilio ulysses* 2	8	99	1	3.42	1.91–5.26	I	90.79	0.285
Clade 2 (*P. crino*)	9	NA	1	15.57	12.49–18.64	F	90.16	0.537
Clade 3–Clade 4	10	61	1	14.32	11.48–17.21	F	90.49	0.387
Clade 3	11	75	0.8	12.17	9.47–14.94	F	90.17	0.532
Papilio blumei–Papilio daedalus	12	87	1	10.27	7.75–13	FH	90.92	0.252
Papilio daedalus–Papili palinurus	13	99	1	7.41	5.08–9.89	F	90.74	0.299
Papilio pericles–Papilio lorquinianus	14	89	1	9.28	6.69–12.04	H	90.92	0.252
Papilio pericles–Papilio neumoegeni	15	6	1	7.38	4.84–10.07	H	90.11	0.563
Papilio peranthus–Papilio lorquinianus	16	100	1	3.16	1.66–5.01	H	90.11	0.563
Clade 4	17	90	1	10.57	8.05–13.24	F	90.17	0.532
Papilio paris 1–*Papilio paris* 3	18	95	1	6.4	4.25–8.94	DF	90.26	0.487
Papilio karna–Papilio paris 2	19	85	1	4.12	2.5–5.99	DF	90.37	0.433
Papilio paris 1–*Papilio paris* 2	20	43	0.96	2.67	1.36–4.24	D	90.34	0.448
Papilio maackii s.l. –Papilio bianor 1	21	60	0.9	8.99	6.74–11.34	D	90.26	0.487
Papilio maackii s.l. –Papilio arcturus	22	NA	1	6.87	4.8–9.14	D	90.87	0.263
Papilio krishna–Papilio arcturus	23	72	0.96	6.29	4.31–8.5	A	91.54	0.135
Papilio hopponis–Papilio arcturus	24	98	1	3.41	1.87–5.17	D	91.13	0.204
Papilio hermeli–Papilio bianor 1	25	99	1	6.15	4.42–8.04	D	90.87	0.263
Papilio dialis–Papilio hermeli	26	8	0.94	5.44	3.79–7.33	DE	91.38	0.159
Papilio bianor 1–*Papilio dehaanii* 1	27	77	1	4.02	2.62–5.55	A	91.38	0.159
Papilio bianor 1–*Papilio bianor* 2	28	52	0.54	2.6	1.37–4.08	AB	90.88	0.262
Papilio dehaanii 1–*Papilio dehaanii* 2	29	4	0.99	3.16	1.87–4.58	B	90.88	0.262

The table refers to Fig. 4, organized by node number. For each node, we indicated (i) the phylogenetic support with the bootstrap values (BV) for maximum parsimony and posterior probability (PP) obtained in Bayesian analysis, (ii) the estimated median age with the 95% highest posterior density for the chronogram obtained using hard bounds under the Bayesian relaxed-clock method, and (iii) the ancestral range reconstruction obtained by ML-DEC analyses with the likelihood score (ln L) and the relative probability (RP). For geographical analysis, only the first split at a node is given and can be non-significant. Asterisks indicated where a constraint was used to calibrate the molecular clock. NA, not available because of the rooting process or because outgroups were removed for geographical range evolution analyses.

lides. Overall, an early Miocene origin of the subgenus *Achillides* was recovered (*ca*. 19.25 Ma, HPD 16.4–21.8). Clade C1 appeared at 10 Ma (HPD 6.9–13.8). Clade 2 (i.e. the divergence of *P. crino*) appeared at 15.6 (HPD 12.5–18.6). The split between clade C3 and C4 originated around 14.3 Ma (HPD 11.5–17.2). Finally, clade C3 and clade C4 appeared at 12.2 Ma (HPD 9.5–14.9 Ma) and 10.6 Ma (HPD 8.1–13.2 Ma), respectively. For more details on each node (date and 95% HPD), see Table 3.

The results of ML-DEC analyses of geographical range evolution are presented in Fig. 4. Ancestral areas and likelihood scores (L) with their relative probabilities (RP) are given for each node in Table 3. For the root, local optimizations under the ML-DEC method (with either two or three *maxareas*) recovered an optimal range area for the Sundaic region ($L = -91$ for two *maxareas*), which was nonetheless not significantly different from those of the Wallacea region ($L = -91.05$ for two *maxareas*). Other ranges were not statistically supported (more than 2 log-likelihood units of difference) by alternative local optimizations (e.g. for the Indo-Chinese region, $L = -94.71$ for two *maxareas*). Together, this evidence indicates that the ancestral area for the common ancestor of *Achillides* is more likely to be in the Sundaic and/or Wallacea regions. To improve the precision of this inference, supplementary ML-DEC analyses were conducted with the root enforced to be a combination of two or three areas. The combination of the Sundaland and Wallacea regions was significantly supported over the remaining combinations ($L = -89.54$ for two *maxareas*) even when three areas were constrained at the root (best score for Sundaland + Wallacea + Sahul, $L = -92.26$).

The results of both analyses of ancestral area evolution unambiguously support a dynamic dispersal pattern (rate of dispersal = 0.2481 for the optimization under the best ML-DEC model) and few vicariance events were reconstructed. Note that the DEC analyses, with-

out constraints on the root, recovered the same ancestral areas for the remaining nodes.

Diversification rates

Temporal changes in diversification rates are presented using an LTT plot in Appendix S5. Under a birth–death model and using three extinction rate values (ε = 0/0.5/0.9), net diversification rates were estimated at 0.135/0.122/0.064 per Ma for our dataset and at 0.137/0.124/0.065 per Ma taking into account the three missing species. The one-tailed γ-test rejected the null hypothesis of rate constancy (critical-γ = −2.7212, P = 0.0032). The likelihood-based method of diversification analysis provided strong support for a rate-variable model over the constant-rate model (AIC_{RC} = 37.84, AIC_{RV} = 27.99). Following Rabosky (2006a), we have confidence that the rate-variable model best approximates the data because the AIC difference between rate-constant and rate-variable is >4 (in our case ΔAIC_{RC} = 9.85). Both γ-statistic and likelihood analyses confirmed that diversification rates have varied through time. Among the rate-variable models (DDL, DDX, Yule-2-rate, Yule-3-rate and Yule-4-rate), we found that the best-fit model is the Yule-3-rate model. This result thus suggests that two rate shifts have occurred during the evolutionary history of *Achillides*.

When exploring the tempo of diversification and especially the impact of climate events that occurred during the diversification of the group, non-significant effects were found between the origin of *Achillides* and 13 Ma (P = 0.915, LR = 0.011) corresponding to the early divergence among the four clades (Appendix S5). However, significant effects on diversification rates were found at 10.3 Ma (P = 0.024, LR = 7.423; between 15 and 9 Ma), at 7.4 Ma (P = 0.014, LR = 8.572; between 9 and 6 Ma) and at 3.5 Ma (P = 0.017, LR = 11.899; between 5 and 2.5 Ma). Diversification rates were inferred to be higher after these events (under the ML-based method fitting pure birth model to branching times).

Discussion

Phylogenetic relationships and species delimitation within Achillides

The results of the phylogenetic analyses are well supported and consistent regardless of the inference method used (BI or MP), as has commonly been found in other studies (Rindal and Brower, 2011). This strong phylogenetic signal is probably attributable to dense taxonomic sampling and to the size and phylogenetic usefulness of the selected molecular markers (Sperling, 2003). Our results clearly supported the monophyly of *Achillides*, in agreement with previous studies (Zakharov et al., 2004; Yagi et al., 2006). Compared with the work of Yagi et al. (2006), which constitutes the first molecular phylogenetic study of *Achillides*, we found a similar topology even though their study was only based on the ND5 molecular marker and comprised either 11 or 14 species, depending on competing classifications (Shimogori, 1997; Bauer and Frankenbach, 1998; Häuser et al., 2005). The phylogenetic hypotheses recovered in the two studies are in general agreement, with the species belonging to *lorquinianus* and *palinurus* groups appearing as basal to the remaining species sampled by Yagi et al. (2006). Within our clade 4, two major subclades are also comparable in both studies, including *P. bianor*, *P. dehaanii*, *P. dialis*, *P. hermeli* and *P. polyctor* (Yagi et al.'s clade 1), which is sister to a clade comprising *P. arcturus*, *P. hopponis*, *P. karna*, *P. krishna*, *P. maackii*, *P. paris* and *P. syfanius* (Yagi et al.'s clade 2) (Fig. 2).

Apart from these similarities, our *Achillides* species tree presents significant new insights. In particular, clade 1, containing two lineages (*P. montrouzieri* and *P. ulysses*), was recovered in a sister position to all the remaining species. Interestingly, this placement has major implications with regard to the biogeographical pattern (see below). All species are clearly assigned to four well-resolved clades that exhibit distinctive geographical distributions, with the exception of *P. crino*. The latter species is found in an enigmatic and moderately supported position as sister to the C3–C4 clade (Fig. 2), in contrast to its traditional placement within the *palinurus* group (Jordan, 1909). We anticipate that the addition of *P. buddha*, also traditionally placed in the *palinurus* group and occurring in India as well, may bring more resolution on the issue of the phylogenetic position of *P. crino*.

Biogeographic history of Achillides*: roles of Wallacean geological history and palaeoclimate*

Molecular dating analyses suggest the subgenus *Achillides* diverged from its sister group (subgenera *Menelaides* and *Princeps*) around 21 Ma (HPD 18.1–24.5) and diversified around 19.2 Ma (HPD 16.4–21.8) in the early Miocene. This age estimate is also consistent with *Achillides* belonging to a relatively derived group within *Papilio*, a genus that mostly diversified during the middle Cenozoic (Zakharov et al., 2004; Condamine et al., 2012). In other groups of Lepidoptera, such as nymphalid and pierid butterflies, the origin and diversification of major lineages also mostly happened during the mid Cenozoic (e.g. Braby et al., 2006; Peña and Wahlberg, 2008; Wahlberg et al., 2009).

Broadly, the distribution of *Achillides* taxa can be grouped by clade, with clade 1 being endemic to Wallacea and the Australasian region, clade 2 (*P. crino*)

endemic to India, clade 3 ranging across Sundaland and Wallacea, and clade 4 being found in Sundaland and the Indo-Chinese region (Fig. 4). Speciation events are notably older for Sundaic and Wallacean taxa in comparison with species from remaining regions, and long branches imply extended isolation of Sundaic and Wallacean taxa or extinction events. Interestingly, such a pattern has been documented for other groups of butterflies co-occurring in the region (Müller and Beheregaray, 2010; Müller et al., 2010). Yet the temporal importance of Wallacea for the evolution of butterflies has remained enigmatic. Vane-Wright and de Jong (2003) showed that some endemic Sulawesi butterflies represented ancient lineages, a finding supported in other groups (de Boer, 1995).

Geological events as the motor of early geographical diversification

From a biogeographical viewpoint, our molecular dating indicates that the diversification of *Achillides* took place during major tectonic events (Hall, 2002; Metcalfe, 2005). When *Achillides* arose at 20 Ma, Southeast Asia underwent a period of intense tectonic activity with the formation of numerous biogeographical barriers (islands, mountain ranges and deep-water basins; Sanmartín and Ronquist, 2004). Our biogeographical analyses inferred that Sundaland + Wallacea constituted the most likely ancestral area of origin of the subgenus (Figs 4 and 5A), in agreement with a biogeographical analysis of Papilionidae that indicated that the subgenera *Achillides*, *Menelaides* and *Princeps* appeared in Southeast Asia (Condamine et al., 2012). According to several authors (e.g. Hall, 1998, 2002; Metcalfe, 2005), Sundaland appeared as a large peninsula, and geological evidence indicated that Wallacea was partially present 20 Ma (Fig. 5a; Hall, 2002). Shallow seas and palaeo-islands could have connected the two biogeographical entities, a scenario favouring allopatric diversification (Müller et al., 2010).

From this ancestral area, the common ancestor of *Achillides* evolved into three lineages (clade 1, clade 2 and clade 3 + clade 4). The diversification of the predominantly Wallacean clade (clade 1) and of the remaining *Achillides* in Sundaland (Figs 4 and 5b) is associated with a primarily vicariant event probably attributable to geological processes and not to climatic events (Appendix S5). A possible explanation of this vicariant event is the geological rearrangement of Wallacea, and especially for Sulawesi Island, which has changed substantially over the last 20 Myr (Hall, 2002; Metcalfe, 2005). Sulawesi has a unique geological history, and the islands that comprise it probably represent terranes that were sliced from the northern margin of New Guinea at different periods until its formation was completed 5 Ma (Wilson and Moss, 1999; Lohman et al., 2011). These events probably favoured the role of geology in the allopatric diversification of *Achillides*: the first lineage became isolated from the second with the opening of the Makassar Strait (comparable to Wallace's line) between Sundaland and Wallacea (Fig. 5b; Hall, 1998, 2002). The Makassar Strait and much of East Borneo and West Sulawesi was a wide, and locally deep, marine region forming a substantial barrier between Sundaland and Sulawesi (Fig. 5b; Hall, 2009). Such a barrier has also been evidenced in the distribution of the *Cethosia* (Lepidoptera, Nymphalidae) clades (Müller and Beheregaray, 2010). The initial genesis of Wallacea coincides temporally and closely with the earliest splitting in the two aforementioned *Achillides* clades with basal nodes at *ca*. 10 and 15 Ma, respectively. Within clade 3, the common ancestor extended its geographical range from Sundaland to Wallacea and two lineages separated from the common ancestor around 12.2 Ma (Fig. 5c). These two lineages arose with the anticlockwise rotation of Borneo, which provoked intense volcanism in the western part of the Wallacean region, probably facilitating dispersal to Wallacea through the formation of oceanic islands (Hall, 1998, 2002; Lohman et al., 2011).

Within the last 10 Myr, clade 1, which was first established in Wallacea (Fig. 4), extended its distribution southward to the Sahul and diversified (Fig. 5d). As the Australian plate drifted northward to collide with the Eurasian plate, new land bridges were created between Wallacea and Papua New Guinea through Halmahera and the Bird's Head (Wilson and Moss, 1999; Hall, 2002; Metcalfe, 2005). This pattern favours the role of dispersal events to explain the present distribution of clade 1. Trans-oceanic dispersal of the common ancestor can probably be accounted for by their flying abilities, as they have been able to colonize almost all tropical islands or archipelagos in the Australasian region (Fig. 1; Wallace, 1865). Indeed the four species of the *ulysses* group have colonized all Papua New Guinean islands, the Solomon Islands, and even New Caledonia. The colonization of New Caledonia may represent long-distance dispersal of the common ancestor of clade 1 from surrounding regions. The dispersal to New Caledonia would have been realized either from the Solomon Islands via the Vanuatu archipelago, implying extinction of an ancestral population in these islands, or directly from Australia/Papua New Guinea across intervening islands. Biogeographical analyses recovered an ancestral dispersal from Sahul ($L = -90.74$) that was more likely than from the Solomon Islands ($L = -93.94$).

Other notable events include the separation of the lineage that leads to the Indo-Chinese *P. dialis* and the Philippine *P. hermeli* in the late Miocene around 5.4 Ma, which cannot be attributed to climatic or dispersal events. Phylogenetic and dating results imply

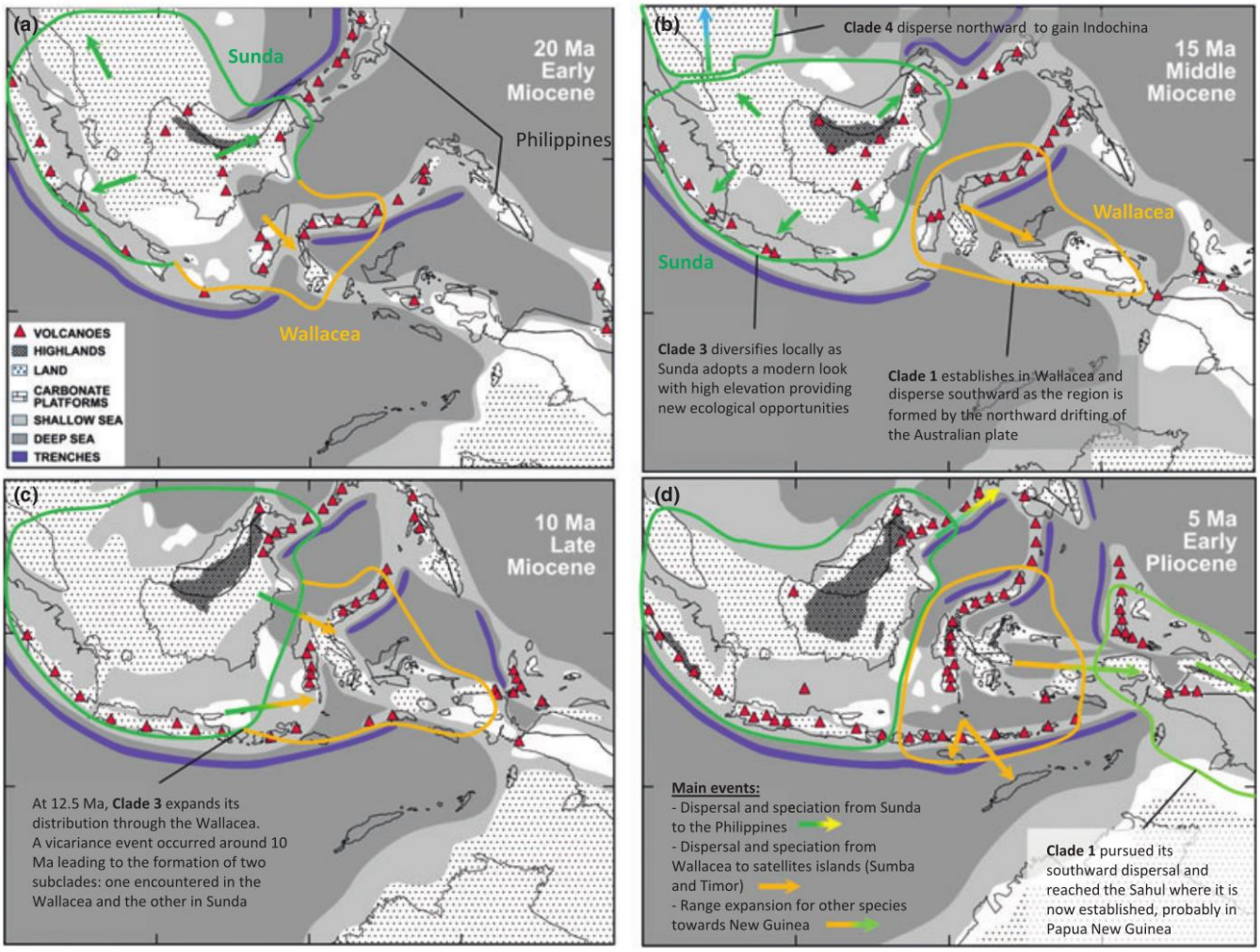


Fig. 5. Biogeographical scenarios for clades 1 and 3 of the *Papilio* subgenus *Achillides* illustrated by maps. (a) Origin and putative ancestral distribution of *Achillides*. *Achillides* appeared in the Sundaic–Wallacean region around 20 Ma. (b) The initial geographical divergence among the three main clades of *Achillides* near 15 Ma. (c) Range expansion of clade 3 followed by a vicariance event among the Sundaic lineage and Wallacean lineage around 10 Ma. (d) Local diversifications of clades 1 and 2 by dispersal events to the Philippines, Sumba and Timor and to Papua New Guinea near 5 Ma. Successive dispersal and range expansions are labelled with coloured arrows. Palaeogeographic maps were redrawn from Hall (1998, 2002).

past geological connections between the two regions. A connection between Taiwan and the Philippines (Luzon) through an arc–trench system has been proposed (Honza and Fujioka, 2004). It has been generally accepted that the formation of Taiwan resulted from the collision of the Luzon Arc with the Eurasian continental margin or with the former Ryukyu subduction zone (Honza and Fujioka, 2004). The Luzon Arc formed during the early to middle Miocene and probably connected Taiwan with Luzon from the late Miocene to late Pliocene (*ca.* 10–2 Ma, Honza and Fujioka, 2004), which facilitated the spread of organisms. As Taiwan was connected to southern China at various stages during the Pliocene and Pleistocene, based on bathymetric studies (Honza and Fujioka, 2004), it is plausible that the ancestor of *P. dialis* and *P. hermeli* was spread from Taiwan to the Philippines by this arc connecting Luzon and Taiwan, and that a vicariance event separated the two current lineages. This scenario is corroborated by the presence of *P. dialis* in Taiwan, as well as in Indo-China (Fig. 1). A similar palaeo-connection is likely to have resulted in the distributions of the closely related *Papilio* swallowtail species *P. xuthus* (Taiwan, Japan and China) and *P. benguetanus* (northern Luzon), as well as *Cethosia* nymphalids (Müller and Beheregaray, 2010).

Climate change as a major driving force in late diversification

Although it is difficult to link significant effects of past climate changes to the diversification rates of organisms (Currie et al., 2004; Winkler et al., 2009), in this study we uncovered evidence for three shifts in diversification

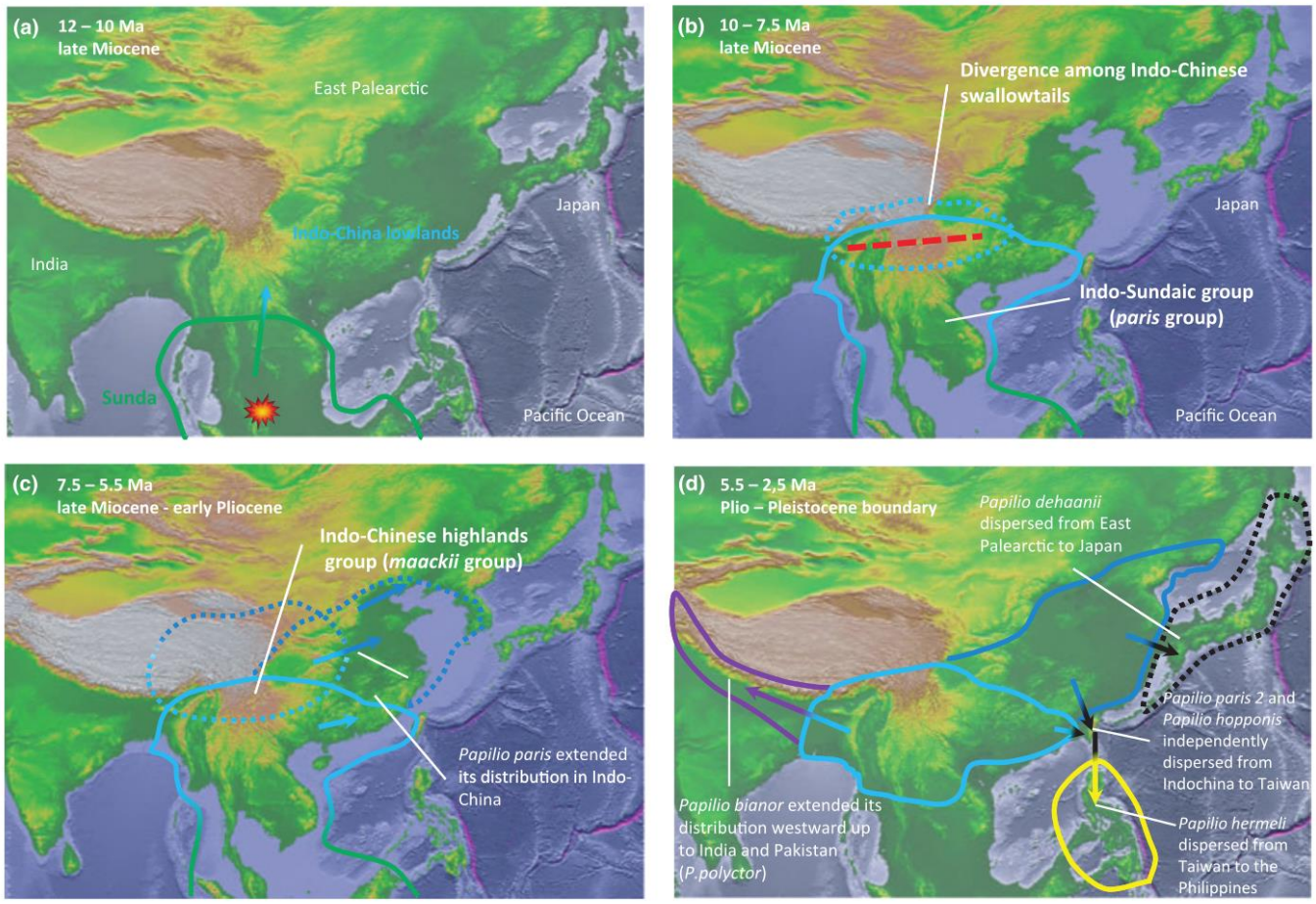


Fig. 6. Biogeographic scenarios for clade 4 of the *Papilio* subgenus *Achillides* illustrated by maps of the Asian mainland. As the palaeogeographical situation at 10 Ma was very similar to that of today, the same map is used, except for the sea-level decreases for maps A and C. (a) Origin of clade 4 and initial diversification in Sunda when sea level was low between 12 and 10 Ma. (b) Northward dispersal and early divergence among Indo-Chinese and Sundaic species. The diversification of clade 4 continued and three lineages arose: a Sundaic group, an Indo-Chinese group and a Chinese group between 10 and 7.5 Ma. (c) The Chinese group extended its distribution northward, whereas the Indo-Chinese lineage expanded westward to India (parapatric speciation). (d) Late range expansions and dispersals to islands when sea level dropped. During the Plio-Pleistocene glaciations, several island dispersals are identified to Japan (*P. dehaanii* 1 and 2), to Taiwan (*P. hopponis* and *P. paris* 2 from clade 3) and to the Philippines (*P. hermeli*). Lines indicate putative ancestral areas of distribution of clades (dashed lines for subclades), while arrows identify dispersals.

rates that may be correlated with specific climatic events. These results are of interest for: (i) understanding diversification processes, especially in the Indo-Australian Archipelago where geological vicariance is often postulated; and (ii) understanding the effects of global climate change as studies showing such climate effects are scarce (Winkler et al., 2009; Condamine et al., 2012). According to the LTT plot and birth–death likelihood tests, we found that diversification rates of *Achillides* shifted at 10.3 and at 7.4 Ma in the late Miocene, and also at 3.5 Ma at the Pliocene–Pleistocene boundary (Figs 6 and 7).

We recovered higher diversification rates just after these three periods and interpreted them as a response to global climate dynamics (Zachos et al., 2001), rather than to geological events, given that the area was tectonically stable during these periods (Wilson and Moss, 1999; Hall, 2002; Metcalfe, 2005). In the Cenozoic, Earth's climate underwent significant and complex evolution, including gradual trends of warming and cooling (Zachos et al., 2001). The most prominent response to climate change in the past has probably been sea-level fluctuations (Voris, 2000; Miller et al., 2005), which would have seen the islands of the Indo-Australian Archipelago repeatedly connected and isolated. This probably led to dispersal along the Sundaic shelf during periods of low sea level, and promoted subsequent diversification through isolation during periods of higher sea level (Figs 6 and 7; Woodruff and Turner, 2009).

Around 10.3 Ma, the shift in diversification rate is attributed to a cooling event (Fig. 6). It is well known

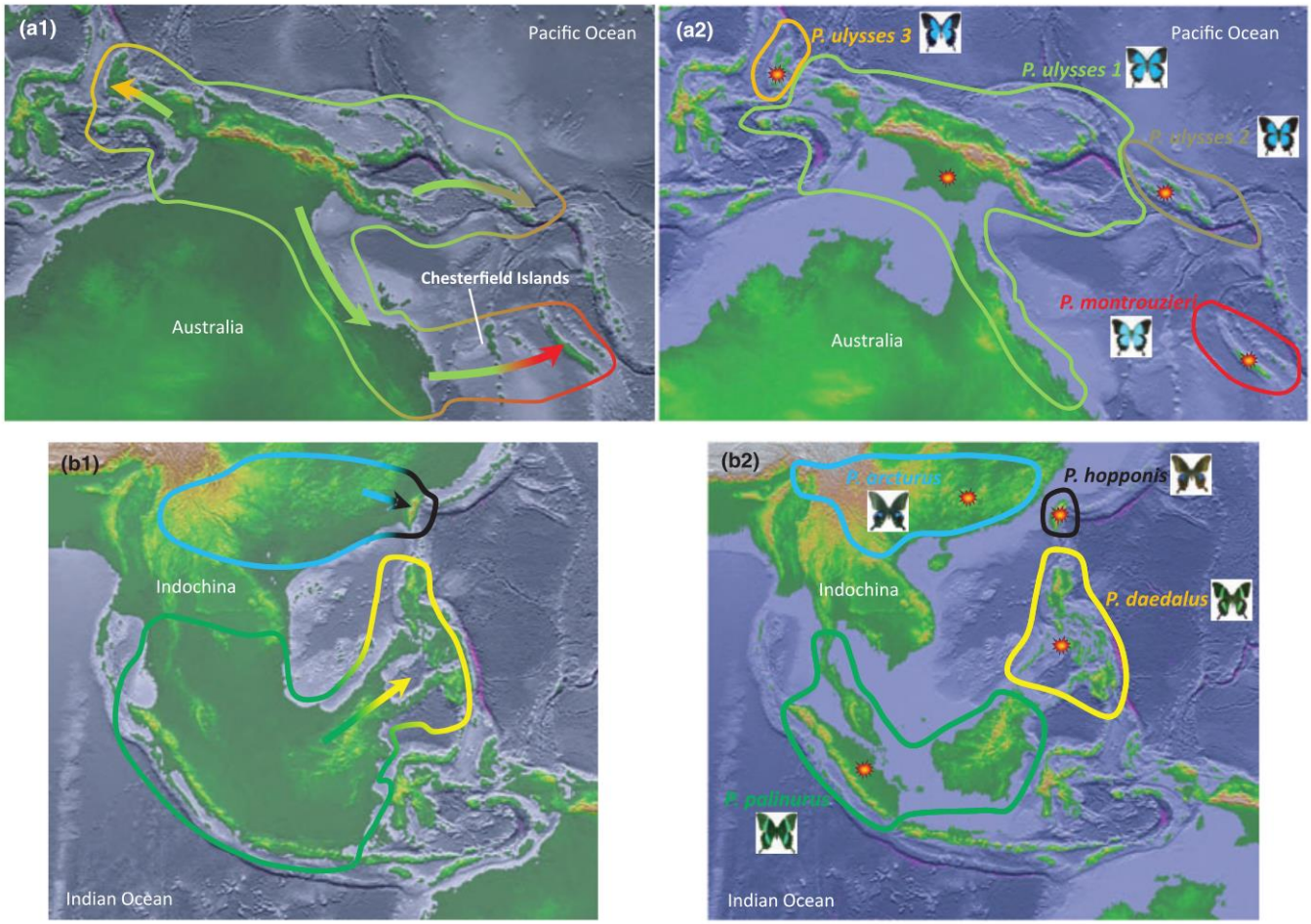


Fig. 7. Schematic maps showing the effects on diversification of the *Papilio* subgenus *Achillides* during the Plio-Pleistocene glaciations. (a) The maps represent the diversification of clade 1 (*ulysses* group). Ten million years ago, a period when sea level decreased due to a cooling event (a1), the common ancestor was probably distributed throughout the Sahulian region. Range expansion has been facilitated because land bridges existed and distribution areas were thus large. When sea level increased (a2), the populations became isolated and evolved by allopatry. (b) The maps represent the diversification of two pairs of sister species (*P. arcturus*/*P. hopponis* and *P. daedalus*/*P. palinurus*) in Indochina and Sunda. Similarly to the *ulysses* group, when sea level decreased during the cooling period (b1), dispersals were facilitated and distribution areas were large. When sea level increased (b2), populations became isolated and evolved by vicariance.

that cooling periods favoured a decrease of global sea level by accumulating water on ice-sheets at the poles, and the inverse is true during warmer times (Zachos et al., 2001; Miller et al., 2005). This fall in sea level probably facilitated range expansion among areas normally isolated, especially in the Indo-Australian Archipelago, and then favoured vicariance events (Hall, 2002; Woodruff and Turner, 2009). This climate change corresponds to the appearance of clade 1 and 3 (Fig. 6). Clade 1 diversified early from the common ancestor by geological vicariance (Fig. 4), but climate change also appeared to influence the diversification within clade 1, which involved a split between *P. montrouzieri* and the three putative *P. ulysses* species. Above, we assumed that *P. montrouzieri* appeared by long-distance dispersal from Sahul, and these results suggest that such an exceptional event (more than 1000 km of oceanic barrier) was probably facilitated by sea level decrease during the cooling event (Fig. 7a). North-eastern Australia and New Caledonia are surrounded by shallow seas and by lagoons with coral reefs. When sea level dropped, these extended areas of lagoons probably became land and previously emerged archipelagos, such as the Chesterfield Islands, have been recognized as occurring between Australia and New Caledonia on the Lord Howe ridge (Cluzel et al., 2001; Pelletier, 2006). Thus, the distance between the two biogeographical areas was reduced and dispersal could have been easier for flying insects along stepping-stones (Fig. 7a). We thus suggest that both geological and climatic events have contributed to the diversification of clade 1, which demonstrates that combined analyses are suitable for

understanding fine-grained biogeographical processes. Conversely, for clade 3, climate seems to be an important factor for its diversification. The split between clade 3 and clade 4 is probably not due to geological vicariance (Fig. 4). Climatic influence has been postulated for mammal diversifications in the Malay Peninsula (Woodruff and Turner, 2009), and it has been proposed that the peninsula was partially isolated several times by sea level rise (Lohman et al., 2011). This assumption is of notable importance because the divergence between clade 3 and clade 4 could have been the result of a climatic vicariance (Fig. 5b). Clade 3 occupied the southern part of Sundaland (Greater Sundas) whereas clade 4 inhabited the northern part (Thailand) (Figs 5b and 7a). Species of clade 4 have probably been able to adapt to more subtropical or temperate environments and further colonized the most northern areas towards China and Japan (Fig. 6). Finally in Wallacea, the cooling event provoked a decrease in sea level that has permitted dispersal between western Sulawesi and Sundaland. The ancestor of the Sundaic lineage, within clade 3, probably extended its range to Sulawesi and evolved into *P. blumei*. However, the Makassar Strait is a deep-water trench, so Borneo and Sulawesi cannot have been linked when sea level dropped (Wilson and Moss, 1999), corroborating that this biogeographical event is the result of an inter-island dispersal combined with climatic vicariance event.

Near 7.4 Ma, a sudden cooling event is recorded (Zachos et al., 2001) and corresponds to a new shift in diversification rates within *Achillides* (Appendix S5). The sister relationships of *P. neumoegeni* and *P. pericles*, in the Wallacean lineage of clade 3, may be attributed to this climatic event. We hypothesize that the Wallacean common ancestor dispersed from Sulawesi to the Lesser Sundas when sea level dropped. This ancestor then reached the southernmost islands, which have always been isolated, and when sea level increased, the populations evolved into two endemic species on the Lesser Sundas: *P. neumoegeni* on Sumba and *P. pericles* in Timor. Other species also resulted from the sea level decrease: *P. palinurus* extended its distribution from the Greater Sundas to the Philippines through temporal connections with Palawan, and evolved into *P. daedalus* (Fig. 7b). This scenario is corroborated by our divergence time estimates, which are congruent with the formation of the Philippines (Honza and Fujioka, 2004).

For *Achillides*, the Pliocene–Pleistocene boundary marked a period of peak diversification with five synchronous speciation events around 3.5 Ma affecting all biogeographical regions (Fig. 4; Appendix S5). It suggests that the impact of this late climatic phenomenon was widespread in the Indo-Australian Archipelago, in agreement with Lohman et al. (2011). The early Pliocene (5.5 Ma) is marked by a subtle warming event until 3.2 Ma, when benthic $\delta^{18}O$ increased again, reflecting the onset of northern hemisphere glaciations (Appendix S5; Zachos et al., 2001). Continuous sea level fluctuations have permitted the common ancestor of *P. ulysses 1–2* to disperse to the Solomon Islands and give rise to *P. ulysses 2* (Fig. 7a), as well as the dispersal of *P. arcturus* in Taiwan to evolve into *P. hopponis* (Fig. 7b). Moreover, the Pliocene–Pleistocene sea level rises and drops also appear to have shaped the subgenus at the subspecific level. The various subspecies of *P. lorquinianus*, *P. paris*, *P. peranthus* and *P. ulysses* (Fig. 1) in Sundaland and Wallacea probably diverged from an ancestor that spanned the entire island group when they were connected during a period of low sea level (Fig. 7). Woodruff and Turner (2009) showed that many rapid sea level rises in the last 5 Myr have probably resulted in faunal compression and expansion, corroborated here by the extant distribution pattern of *Achillides*, being mainly shaped by climatic vicariance.

Conclusions and insights for the biogeography of the Indo-Australian Archipelago

The historical biogeography of the genus *Papilio* represents a significant challenge. *Papilio* is a highly complex and diverse taxonomic unit with at least 200 species and many more subspecies (Häuser et al., 2005). Biogeographical analysis of the subgenus *Achillides* has provided insight into the history of the region and the processes that have shaped the evolution of these swallowtails. We showed that both geological and climatic events have probably influenced the geographical diversification of a species-rich insect group. This study has also produced valuable results as it is the first time that several significant effects of climate changes have been proposed for butterflies of the Indo-Australian Archipelago. The sea level fluctuations have probably been a major factor in their late geographical diversification. However, we remain cautious about our results and interpretations, especially in a region such as the Indo-Australian Archipelago. New Caledonia serves to remind us of fundamental geological and biogeographical problems in the region (Hall, 1998). Thus, investigation of the biogeographical history of organisms in this region requires good knowledge of palaeogeographical and palaeoclimatic history.

Beyond purely process-oriented scientific purposes, another goal of biodiversity surveys is to obtain a significant, broad representation of evolutionary history for extant organisms (Sechrest et al., 2002). Phylogenetics and a model-based historical biogeographical approach are necessary to assess the amount of evolutionary novelty represented in different lineages. Our study has thus revealed that biodiversity hotspots are not solely areas of high specific richness and high levels of ende-

mism, but are also important reservoirs of a unique and threatened evolutionary history (Sechrest et al., 2002). This particularly applies to the Wallacea hotspot in which the *Achillides* originated and further diversified. Within this framework, conservation efforts may be directed to ameliorate the current anthropogenic threats confronting these unique swallowtails while preserving a representation of their evolutionary history (Sechrest et al., 2002; Mittermeier et al., 2004).

Acknowledgements

We thank all the collectors and providers of numerous specimens such as Laurent Soldati, Hervé Jourdan, Kotaro Saito, Yutaka Inayoshi and Kiyoshi Miura. We thank Anne-Laure Clamens for technical assistance with the molecular work, and Frédérique Serqueira from the SFR 119 ''Montpellier Environnement Biodiversité'' for help in the sequencing of some specimens. We are also grateful to Alexandre Dehne Garcia for help with computer analysis, as well as Marie Pagès for advice on species delimitation analyses. We thank the Environment Management of Province Sud (DENV) in New Caledonia for providing us with collecting permits and field assistance. Partial funding for this study was obtained through the programme ''ANR Biodiversité'' of the French National Agency for Research (Project ANR BIONEOCAL) and by a grant from the graduate school of Montpellier II University (SIBAGHE). F.A.H.S. acknowledges funding from an NSERC Discovery Grant. F.L.C. conceived the study, participated in its design and coordination, collected samples, analysed data and drafted and revised the manuscript. E.F.A.T. carried out the molecular genetic studies, participated in the sequence analysis and revised the manuscript. A.M.C. collected and identified swallowtail specimens and revised the manuscript. G.G. carried out the molecular genetic studies and sequencing. F.A.H.S. provided additional samples and revised the manuscript. G.J.K. funded the study, analysed the data and revised the manuscript. All authors read and approved the final manuscript.

References

Bauer, E., Frankenbach, T., 1998. Butterflies of the World. Pt.1. Papilionidae I: *Papilio* subgenus *Achillides*, *Bhutanitis*, *Teinopalpus*. Goecke und Evers, Keltern.

Becerra, J.X., Venable, D.L., 1999. Macroevolution of insect-plant associations: the relevance of host biogeography to host affiliation. Proc. Natl. Acad. Sci. USA 96, 12625–12631.

Bell, C.D., Soltis, D.E., Soltis, P.S., 2010. The age and diversification of the angiosperms re-revisted. Am. J. Bot. 97, 1296–1303.

Bininda-Emonds, O.R.P., Cardillo, M., Jones, K.E., MacPhee, R.D.E., Beck, R.M.D., Grenyer, R., Price, S.A., Vos, R.A., Gittleman, J.L., Purvis, A., 2007. The delayed rise of present-day mammals. Nature 446, 507–512.

de Boer, A.J., 1995. Islands and cicadas adrift in the West-Pacific, biogeographic patterns related to plate tectonics. Tijd. v. Ent. 138, 169–244.

Braby, M.F., Vila, R., Pierce, N.E., 2006. Molecular phylogeny and systematics of the Pieridae (Lepidoptera: Papilionoidea): higher classification and biogeography. Zool. J. Linn. Soc. 147, 239–275.

Britton, T., Anderson, C.L., Jacquet, D., Lundqvist, S., Bremer, K., 2007. Estimating divergence times in large phylogenetic trees. Syst. Biol. 56, 741–752.

Brown, J.M., Lemmon, A.R., 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. Syst. Biol. 56, 643–655.

Caterino, M.S., Sperling, F.A.H., 1999. *Papilio* phylogeny based on mitochondrial cytochrome oxidase I and II genes. Mol. Phylogenet. Evol. 11, 122–137.

Cho, S., Mitchell, A., Regier, J.C., Mitter, C., Poole, R.W., Friedlander, T.P., Zhao, S., 1995. A highly conserved gene for low level phylogenetics: elongation factor-1 alpha recovers morphology-based tree for heliothine moths. Mol. Biol. Evol. 12, 650–656.

Cluzel, D., Aitchison, J.C., Picard, C., 2001. Tectonic accretion and underplating of mafic terranes in the Late Eocene intraoceanic fore-arc of New Caledonia (Southwest Pacific): geodynamic implications. Tectonophysics 340, 23–59.

Condamine, F.L., Sperling, F.A.H., Wahlberg, N., Rasplus, J.-Y., Kergoat, G.J., 2012. What causes latitudinal gradients in species diversity? Evolutionary processes and ecological constraints on swallowtail biodiversity. Ecol. Lett. 15, 267–277.

Currie, D.J., Mittelbach, G.G., Cornell, H.V., Field, R., Guégan, J.-F., Hawkins, B.A., Kaufman, D.M., Kerr, J.T., Oberdorff, T., O''Brien, E., Turner, J.R.G., 2004. Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. Ecol. Lett. 7, 1121–1134.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, e88.

Edwards, S.V., 2009. Is a new and general theory of molecular systematics emerging? Evolution 63, 1–19.

Erixon, P., Svennblad, B., Britton, T., Oxelman, B., 2003. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. Syst. Biol. 52, 665–673.

Felsenstein, J., 2004. Inferring Phylogenies. Sinauer, Sunderland, MA.

Goloboff, P.A., 1999. Analyzing large data sets in reasonable times: solutions for composite optima. Cladistics 15, 415–428.

Goloboff, P.A., Farris, J.S., Nixon, K.C., 2008. TNT, a free program for phylogenetic analysis. Cladistics 24, 774–786.

Grandcolas, P., Murienne, J., Robillard, T., Desutter-Grandcolas, L., Jourdan, H., Guilbert, E., Deharveng, L., 2008. New Caledonia: a very old Darwinian island? Philos. Trans. R. Soc. B 363, 3309–3317.

Hall, R., 1998. The plate tectonics of Cenozoic SE Asia and the distribution of land and sea. In: Hall, R., Holloway, J.D. (Eds.), Biogeography and Geological Evolution of SE Asia. Backhuys Publishers, Leiden, pp. 99–131.

Hall, R., 2002. Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific: computer-based reconstructions, model and animations. J. Asian Earth Sci. 20, 353–431.

Hancock, D.L., 1983. Classification of the Papilionidae (Lepidoptera): a phylogenetic approach. Smithersia 2, 1–48.

Harada, M., 1992. The status of *Achillides bianor* (Cramer) in Sichuan, China. Butterflies 2, 19–25.

Häuser, C.L., de Jong, R., Lamas, G., Robbins, R.K., Smith, C., Vane-Wright, R.I., 2005. Papilionidae revised GloBIS/GART

species checklist (2nd draft). Available at: http://www.insects-online.de/frames/papilio.htm [accessed January 2012].
Heads, M., 2005. Dating nodes on molecular phylogenies: a critique of molecular biogeography. Cladistics 21, 62–78.
Hebert, P.D.N., Cywinska, A., Ball, S.L., DeWaard, J.R., 2003. Biological identifications through DNA barcodes. Proc. Biol. Sci. 270, 313–321.
Hillis, D.M., Bull, J.J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42, 182–192.
Hines, H.M., 2008. Historical biogeography, divergence times, and diversification patterns of bumble bees (Hymenoptera: Apidae: *Bombus*). Syst. Biol. 57, 58–75.
Ho, S.Y.W., Phillips, M.J., 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Syst. Biol. 58, 367–380.
Holder, M., Lewis, P.O., 2003. Phylogeny estimation: traditional and Bayesian approaches. Nat. Rev. Genet. 4, 275–284.
Honza, E., Fujioka, K., 2004. Formation of arcs and backarc basins inferred from the tectonic evolution of Southeast Asia since the Late Cretaceous. Tectonophysics 384, 23–53.
Hug, L.A., Roger, A.J., 2007. The impact of fossils and taxon sampling on ancient molecular dating analyses. Mol. Biol. Evol. 24, 1889–1897.
Hunt, T., Bergsten, J., Levkanicova, Z., Papadopoulou, A., St. John, O., Wild, R., Hammond, P.M., Ahrens, D., Balke, M., Caterino, M.S., Gómez-Zurita, J., Ribera, I., Barraclough, T.G., Bocakova, M., Bocak, L., Vogler, A.P., 2007. A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. Science 318, 1913–1916.
Igarashi, S., 1984. The classification of the Papilionidae mainly based on the morphology of their immature stages. Trans. Lepido. Soc. Japan (Tyô To Ga) 34, 41–96.
Igarashi, S., Fukuda, H., 2000. The Life Histories of Asian Butterflies, Vol. 2. Tokai University Press, Tokyo.
Izumi, K., 1993. Geographic variation of wing shape in *Papilio bianor* (Insecta; Lepidoptera; Papilionidae). Bull. Biogeogr. Soc. Japan 48, 27–31.
Jordan, H.E.K., 1909. Fauna indoaustralica. In: Seitz, A. (Ed.), Macrolepidoptera of the World, Vol. 9. Kernen, Stuttgart, pp. 25–109.
Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.
Klaus, S., Schubart, C.D., Streit, B., Pfenninger, M., 2010. When Indian crabs were not yet Asian – biogeographic evidence for Eocene proximity of India and Southeast Asia. BMC Evol. Biol. 10, 287.
Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X, version 2.0. Bioinformatics 23, 2947–2948.
Lohman, D.J., de Bruyn, M., Page, T., von Rintelen, K., Hall, R., Ng, P.K.L., Shih, H.-T., Carvalho, G.R., von Rintelen, T., 2011. Biogeography of the Indo-Australian Archipelago. Ann. Rev. Ecol. Evol. Syst. 42, 205–226.
Lomolino, M.V., Riddle, B.R., Whittaker, R.J., Brown, J.H., 2010. Biogeography, 4th edn. Sinauer, Sunderland, MA.
Magallón, S., Sanderson, M.J., 2001. Absolute diversification rates in angiosperm clades. Evolution 55, 1762–1780.
Metcalfe, I., 2005. Asia: south-east. In: Selley, R.C., Cocks, L.R.M., Plimer, I.R. (Eds.), Encyclopedia of Geology, Vol. 1. Elsevier, Oxford, pp. 169–198.
Miller, K.G., Kominz, M.A., Browning, J.V., Wright, J.D., Mountain, G.S., Katz, M.E., Sugarman, P.J., Cramer, B.S., Christie-Blick, N., Pekar, S.F., 2005. The Phanerozoic record of global sea-level change. Science 310, 1293–1298.
Mittermeier, R.A., Robles Gil, P., Hoffmann, M., Pilgrim, J., Brooks, T., Mittermeier, C.G., Lamoreux, J.L., da Fonseca, G.A.B., 2004. Hotspots revisited: Earth's Biologically Richest and Most Threatened Ecoregions. CEMEX, Mexico City, Mexico.
Müller, C.J., Beheregaray, L.B., 2010. Palaeo island-affinities revisited – Biogeography and systematics of the Indo-Pacific genus *Cethosia* Fabricius (Lepidoptera: Nymphalidae). Mol. Phylogenet. Evol. 57, 314–326.
Müller, C.J., Wahlberg, N., Beheregaray, L.B., 2010. "After Africa": the evolutionary history and systematics of the genus *Charaxes* Ochsenheimer (Lepidoptera: Nymphalidae) in the Indo-Pacific region. Biol. J. Linn. Soc. 100, 457–481.
Munroe, E., 1961. The classification of the Papilionidae (Lepidoptera). Can. Entomol. Suppl. 17, 1–51.
Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Kent, J., 2000. Biodiversity hotspots for conservation priorities. Nature 403, 853–858.
Nylander, J.A., Ronquist, F., Huelsenbeck, J.P., Nieves-Aldrey, J.L., 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53, 47–67.
Nylander, J.A.A., Olsson, U., Alström, P., Sanmartín, I., 2008. Accounting for phylogenetic uncertainty in biogeography: a Bayesian approach to dispersal-vicariance analysis of the Trushes (Aves: *Turdus*). Syst. Biol. 57, 257–268.
Paradis, E., Claude, J., Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290.
Pelletier, B., 2006. Geology of the New Caledonia region and its implications for the study of the New Caledonian biodiversity. In: Payri, C.E., Richer de Forges, B. (Eds.), Compendium of Marine Species from New Caledonia Documents Scientifiques et Techniques, II4. Institut de Recherche pour le Développement, Nouméa, pp. 17–30.
Peña, C., Wahlberg, N., 2008. Prehistorical climate change increased diversification of a group of butterflies. Biol. Lett. 4, 247–278.
Pfeil, B.E., Crisp, M.D., 2008. The age and biogeography of *Citrus* and the orange subfamily (Rutaceae: Aurantioideae) in Australasia and New Caledonia. Am. J. Bot. 95, 1621–1631.
Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., Kamoun, S., Sumlin, W.D., Vogler, A.P., 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Syst. Biol. 55, 595–609.
Posada, D., 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–1256.
Pybus, O.G., Harvey, P.H., 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. Proc. Biol. Sci. 267, 2267–2272.
de Queiroz, A., 2005. The resurrection of oceanic dispersal in historical biogeography. Trends Ecol. Evol. 20, 68–73.
Rabosky, D.L., 2006a. Likelihood methods for detecting temporal shifts in diversification rates. Evolution 60, 1152–1164.
Rabosky, D.L., 2006b. LASER: a maximum likelihood toolkit for detecting temporal shifts in diversification rates from molecular phylogenies. Evol. Bioinform. Online 2, 257–260.
Rabosky, D.L., Lovette, I.J., 2008. Density-dependent diversification in North American wood warblers. Proc. R. Soc. B, 275, 2363–2371.
Ree, R.H., Sanmartín, I., 2009. Prospects and challenges for parametric models in historical biogeographical inference. J. Biogeogr. 36, 1211–1220.
Ree, R.H., Smith, S.A., 2008. Maximum-likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Syst. Biol. 57, 4–14.
Ree, R.H., Moore, B.R., Webb, C.O., Donoghue, M.J., 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. Evolution 59, 2299–2311.

Reed, R.D., Sperling, F.A.H., 1999. Interaction of process partitions in phylogenetic analysis: an example from the swallowtail butterfly genus *Papilio*. Mol. Biol. Evol. 16, 286–297.
Rindal, E., Brower, A.V.Z., 2011. Do model-based phylogenetic analyses outperform parsimony? A test with empirical data. Cladistics 27, 331–334.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.
Salvo, G., Ho, S.Y.W., Rosenbaum, G., Ree, R., Conti, E., 2010. Tracing the temporal and spatial origins of island endemics in the Mediterranean region: a case study from the citrus family (*Ruta* L., Rutaceae). Syst. Biol. 59, 705–722.
Sanderson, M.J., 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol. Biol. Evol. 19, 101–109.
Sanmartín, I., Ronquist, F., 2004. Southern hemisphere biogeography inferred by event-based models: plant versus animal patterns. Syst. Biol. 53, 216–243.
Sanmartin, I., van der Mark, P., Ronquist, F., 2008. Inferring dispersal: a Bayesian approach to phylogeny-based island biogeography, with special reference to the Canary Islands. J. Biogeogr. 35, 428–449.
Scotese, C.R., 2001. Atlas of Earth History, Vol. 1. PALEOMAP Project, Arlington, TX.
Sechrest, W., Brooks, T.M., da Fonseca, G.A.B., Konstant, W.R., Mittermeier, R.A., Purvis, A., Rylands, A.B., Gittleman, J.L., 2002. Hotspots and the conservation of evolutionary history. Proc. Natl. Acad. Sci. USA 99, 2067–2071.
Shimogori, Y., 1997. *Achillides* Butterflies (Papilionidae), Vol. 6. Endless Science, Tokyo.
Simonsen, T.J., Zakharov, E.V., Djernaes, M., Cotton, A.M., Vane-Wright, R.I., Sperling, F.A.H., 2011. Phylogenetics and divergence times of Papilioninae (Lepidoptera) with special reference to the enigmatic genera *Teinopalpus* and *Meandrusa*. Cladistics 27, 113–137.
Sperling, F.A.H., 2003. Butterfly molecular systematics: from species definitions to higher-level phylogenies. In: Boggs, C.L., Watt, W.B., Ehrlich, P.R. (Eds), Butterflies: Ecology and Evolution Taking Flight. Chicago University Press, Chicago, IL, pp. 431–458.
Thorne, J.L., Kishino, H., Painter, I.S., 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol. Biol. Evol. 15, 1647–1657.
Vane-Wright, R.I., de Jong, R., 2003. The butterflies of Sulawesi: annotated checklist for a critical island fauna. Zool. Verh. Leiden 343, 3–267.
Voris, H.K., 2000. Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. J. Biogeogr. 27, 1153–1167.
Wahlberg, N., Leneveu, J., Kodandaramaiah, U., Peña, C., Nylin, S., Freitas, A.V.L., Brower, A.V.Z., 2009. Nymphalid butterflies diversify following near demise at the Cretaceous/Tertiary boundary. Proc. Biol. Sci. 276, 4295–4302.
Wallace, A.R., 1860. On the zoological geography of the Malay Archipelago. J. Proc. Linn. Soc. Zool. 4, 172–184.
Wallace, A.R., 1863. On the physical geography of the Malay Archipelago. J. Roy. Geographic. Soc. London 33, 217–234.
Wallace, A.R., 1865. On the phenomena of variation and geographical distribution as illustrated by the Papilionidae of the Malayan region. Trans. Linn. Soc. London 25, 1–71.
Wiens, J.J., Donoghue, M.J., 2004. Historical biogeography, ecology and species richness. Trends Ecol. Evol. 19, 639–644.
Wilson, M.E., Moss, S.J., 1999. Cenozoic palaeogeographic evolution of Sulawesi and Borneo. Palaeogeogr. Palaeoclimatol. Palaeoecol. 145, 303–337.
Winkler, I.S., Mitter, C., Scheffer, S.J., 2009. Repeated climate-linked host shifts have promoted diversification in a temperate clade of leaf-mining flies. Proc. Natl. Acad. Sci. USA 106, 18103–18108.
Woodruff, D.S., Turner, L.M., 2009. The Indochinese–Sundaic zoogeographic transition: a description and analysis of terrestrial mammal species distributions. J. Biogeogr. 36, 803–821.
Yagi, T., Sasaki, G., Takebe, H., 1999. Phylogeny of Japanese papilionid butterflies inferred from nucleotide sequences of the mitochondrial ND5 gene. J. Mol. Evol. 48, 42–48.
Yagi, T., Sasaki, G., Omoto, K., 2006. Phylogeny, biogeography and wing pattern convergence of the subgenus *Achillides* (Lepidoptera, Papilionidae, *Papilio*) revealed by a mitochondrial DNA sequence analysis. Trans. Lepido. Soc. Japan 57, 137–147.
Yang, Z., Rannala, B., 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Mol. Biol. Evol. 23, 212–226.
Yoshimito, H., 1998. *Papilio bianor* and *Papilio dehaanii*, two distinct species. Butterflies 20, 45–49.
Zachos, J.C., Pagani, M., Sloan, L., Thomas, E., Billups, K., 2001. Trends, rhythms, and aberrations in global climate 65 Ma to Present. Science 292, 686–693.
Zakharov, E.V., Caterino, M.S., Sperling, F.A.H., 2004. Molecular phylogeny, historical biogeography, and divergence time estimates for swallowtail butterflies of the genus *Papilio* (Lepidoptera: Papilionidae). Syst. Biol. 53, 193–215.
Zeuner, F.E., 1943. Studies in the systematics of *Troides* Hübner (Lepidoptera Papilionidae) and its allies: distribution and phylogeny in relation to the geological history of the Australasian Archipelago. Trans. Zool. Soc. London 25, 107–184.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Map showing the region of study, distribution of the *Papilio* subgenus *Achillides* (black line) and delimited areas (red lines) used for biogeographical analyses. The region comprises 11 component areas: (A) East Palearctic (comprising central and northern China); (B) Japan archipelago; (C) India (encompassing Pakistan and Himalaya); (D) Indo-China (including non-peninsula Thailand, Laos, Cambodia, Vietnam, southern China and Myanmar); (E) Taiwan; (F) Sundaland (peninsular Thailand and Malaysia, Sumatra, Borneo and Java); (G) Philippines and Palawan; (H) Wallacea (including Lesser Sunda with all islands between Lombok and Tanimbar, including Timor and Wetar and the Moluccas with the islands of Ambon, Bacan, Buru, Halmahera, Morotai, Obi and Seram); (I) Irian Jaya, Papua New Guinea and Queensland (including Aru island); (J) Solomon Islands; and (K) New Caledonia and the Loyalty islands. Species richness is indicated by red dots as the number of species/subspecies that occur in each region. In the bottom-left corner, a map shows the delimitation of Wallacea, including various divisions of biotic affinities within the Indo-Australian Archipelago.

Appendix S2. Host plant preferences of the *Papilio* subgenus *Achillides*. These data are taken from Igarashi (1984) and Igarashi and Fukuda (2000). Pictures below show caterpillars of *P. dehaanii*, *P. paris* and *P. ulysses* on their host plant.

Appendix S3. Taxon sampling of *Papilio* subgenus *Achillides* used for this study. For each specimen, we specified the identification code (ID), the genus, species, author and date of description, subspecies if known, author and date of description, country, locality of capture and GenBank accession number for each molecular marker we used.

Appendix S4. DNA-based species delimitation with an extended-barcode marker (COI gene only).

Appendix S5. Temporal diversification of the *Papilio* subgenus *Achillides*. A lineages-through-time plot shows the number of lineages versus time (Ma). Superimposed palaeotemperature estimates (red line) are approximated by benthic $\delta^{18}O$ (Zachos et al., 2001). At the top, horizontal bars indicate whether the climate was warm (red) or cool (blue). Main climate changes for the last 20 Myr are indicated. Vertical teal bars indicate the significant effect of past climate on the diversification of *Achillides*. Light blue vertical bars, around 10.3, 7.4 and 3.9 Ma, identify three major shifts in diversification. At the bottom, the chronogram of *Achillides* is placed with a geological time scale from the Miocene to Holocene (H).

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

OUTCOMES OF THE THESIS

Rhantus suturalis, a widespread species of diving beetle used as a model for fine and large-scale studies of diversification in the Australasian / Indomalayan archipelago.

"*Hemmed in by the unfathomable depths of the unknown, let us be satisfied if it be vouchsafed to us to enlarge by a span the narrow domain of the known. Seekers, all of us, tormented by the desire for knowledge, let us move our lantern from point to point: with the particles explored we shall perhaps be able to piece together a fragment of the picture.*"

Jean-Henri Fabre, *Souvenirs Entomologiques 7th Edition*, 1900.

Contents

A truly integrative taxonomy within sight 244
Developments on insect diversification dynamics 246
Biogeography on the right tracks 249
Conclusion 250

A truly integrative taxonomy within sight

Although only a part of this dissertation is of pure taxonomic relevance, it should be underscored that the remainder (molecular phylogenies, biogeography and phylogeography, diversification analyses, etc…) heavily relies on formerly described species. In fact, most of the fields constituting our modern evolutionary biology need species to be given proper names. Interestingly, whilst most of these very same fields experience tremendous progress due to methodological and theoretical paramount advances, taxonomy is still awaiting for a new era to begin. Indeed, with a growing pressure on biodiversity, time has never been more pressing and solutions are needed to accelerate species descriptions. Recently, a new set of tools has been proposed to overcome the taxonomic impediment. Among these new methods used to detect species-level entities, molecular techniques have taken the lion's share because they are increasingly more efficient whilst being less costly.

In the first part of this thesis, I presented five papers. The first one is an opinion paper shedding light on the appealing advantages of DNA barcoding for a rapid assessment of Southeast Asian arthropod biodiversity. I already introduced in this dissertation the main advantages and pitfalls of DNA Barcoding regarding its multiple goals in biodiversity assessment, from accurately delineating existing taxa to discovering new ones. Although it seems clear now that this technique does not actually provide a way to accelerate the description of species, it may on the other hand speed-up its discovery by permitting a global screening of molecular datasets. Whilst applying extremely basic methods such as genetic clustering, DNA Barcoding provides excellent approximations of alpha diversity and as a result allows taxonomists to focus on particular regions of the Tree of life. With the advent of next-generation sequencing and the reduction of sequencing costs, this technique should strengthen and allow an increasingly more accurate assessment of biodiversity in a time of major threats of natural habitats.

Nevertheless, formal species descriptions still await a revolution to properly and in a timely fashion continue to catalogue life on Earth. Following this opinion paper, I included a manuscript that is currently in review and that illustrates a methodological pipeline to describe new taxa in an integrative fashion. In this article, I used cutting-edge molecular species delimitation techniques to investigate species boundaries in a group of tropical butterflies that has been recently revised taxonomically. I highlight that the use of one species delimitation

method only is perilous and that the use of multiple methods under different assumptions is crucial. On the other hand, using integrative approaches combining multiple methods of species delineation with other lines of evidence such as ranges of distribution or ecology, it is possible not only to have a satisfactory cross-validation of morphology-based taxonomy but also to uncover cryptic species. In this study, I recovered most of the extant species using these techniques but also eight cryptic species. Interestingly, among these eight species, six have already been recognized as distinct species at some point in the past and only two are given new names. Although these molecular techniques are still in the process of being developed and improved, there is a clear potential for a shift towards a faster way of delineating species using integrative molecular tools rather than the traditional morphological biodiversity assessment. Overall, integrating this new savoir-faire in the process of cataloguing life on Earth might be the most appealing track to develop a modern taxonomy.

Finally, I presented three taxonomic papers in which we used an integrative taxonomy to describe two new species to science and assign a new generic name to several species of Australian diving beetles. In the three cases, molecular data have been usefull to cross-validate morphological species delineation but also to help unfolding phylogenetic relationships. Diving beetles are known to present high degrees of cryptic speciation, especially in the Australasian / Indomalayan archipelago where speciation events linked to relatively recent geological and climatic disruptions are still at work (see Chapters 7 and 8). In the case of *Necterosoma timorense* and *Papuadessus baueri*, both species present clear morphological characters that supported the need for new descriptions. If the placement of *N. timorense* in the genus *Necterosoma* was eased by the very typical morphological characters of this genus, the assignment of *P. baueri* along with *P. pakdjoko* was motivated by molecular data only. It is also molecular evidence that suggested first the delineation of *Brancuporus* from its original genus *Antiporus*. These three studies illustrate several of the most important aspects linked to the integration of molecular data into taxonomic descriptions and revisions. First, it can help to cross-validate morphological data or on the contrary allow detecting cases of homoplasy or extreme phenotypic variability. Second, in the absence of any morphological evidence, it permits to uncover cryptic species in morphologically conserved taxa. Third, it allows the rapid phylogenetic placement of newly described taxa at a wider taxonomic scale whilst morphological phylogenies require tremendous amounts of time to be assembled when it is possible to collect enough data for this purpose.

The inclusion of additional data into traditional morphology-based taxonomy and the development of more efficient molecular delineation techniques are major advances to constitute a truly integrative taxonomy. On a more practical point of view, some pleas have been made to integrate crucial morphological data with DNA taxonomy in short descriptions of new taxa. Such turbo-taxonomy is a way to accelerate species discovery and description whilst providing sound scientific taxonomic acts in a timely manner, with high-resolution pictures of habitus and morphological key characters as well as short, intelligible and reproducible descriptions flanked with DNA information. Additionally, major advances have been brought to the scientific community regarding the description of cryptic species which are known to be a paramount part of the undescribed diversity on Earth. Eventually, with the advent of Wikispecies pages for instance, entire descriptions are reachable with a click and molecular data freely and easily retrievable from online repository. Overall, taxonomy is moving towards a truly different stage that will help to overcome the taxonomic impediment.

Developments on insect diversification dynamics

One of the aims of this thesis was to underpin the underlying mechanisms promoting speciation at diverse time and geographic scales. In Chapters 7 and 8, I presented new empirical evidences supporting divergent processes of lineage diversification in Australia and New Guinea. These two islands are of particular interest to compare and study the factors triggering the apparition of new species and the radiation of clades because of their extremely different geological histories. Australia, as an old fragment of the supercontinent Gondwana has remained geologically very stable throughout the Cenozoic whilst most of present-day New Guinea was formed during the past 10 million years. Yet, one can observe that the Australian fauna is far from being strikingly diverse whilst New Guinea holds some of the most remarkable groups of vertebrates and invertebrates such as the birds of paradise, birdwing butterflies or tree kangaroos among the most notorious.

In chapters 7.1 and 7.2, I investigated the evolutionary patterns of diving beetle clades in Australia. This island is extremely interesting because of its Gondwanean origin compared to most of the other islands in the Indomalayan / Australasian archipelago. Despite being of ancient origin, Australia presents a fauna that appears to be less rich than one could expect. If geological factors are unlikely to account for the observed species richness on this continental island, the effect of climate is of particular interest. Indeed, the separation of Australia from

Antarctica and its northwards migration have fostered deep climatic shifts on the island during the Cenozoic. As a result, Australia underwent a progressive aridification associated with the rarefaction of mesic habitats that may have triggered the wax and wane of clades especially aquatic ones. In Chapter 7.1, we reconstructed the phylogeny of the *Sternopriscus tarsalis* radiation (STR) and highlighted that Quaternary climate change was a likely trigger for species diversification. The evolution of the group has been shaped by disruptions in its habitat during the Plio-Pleistocene. At that time, an ongoing aridification initiated in the Miocene had already deeply modified the Australian ecosystems, changing tropical forest in more arid habitats and drying-out drainage systems. We suggested in Chapter 7.1 that glaciation cycles fostered speciation by isolation in glacial refugia even though the extent to which glaciations affected Southeastern Australia is not precisely known.

In Chapter 7.2, I suggested an alternative hypothesis to explain the diversification of the STR, implying that the ongoing aridification had a crucial effect on rainfall seasonality. Indeed, STR species are restricted to seasonal aquatic ecosystems. As highlighted in Chapter 7.2, early Pleistocene Australia had a dominant summer rainfall system that supported rainforest and wet sclerophyll forest that was very different to that which occurs today under an altered winter rainfall regime. From the mid Pleistocene, the disruptions in rainfall seasonality would have deeply changed the habitat of these beetles. As shown in Chapter 7.1, the STR species show clear divergences in the way they respond to climatic and ecological shifts. I argued that the Quaternary climate change in Australia fostered the specialization/adaptation of populations rather than isolated them in refugia and that these specializations allowed speciation to occur. I believe that this group is of particular interest because STR species are very recent and yet present overlapping distributions, therefore suggesting interesting speciation processes that could be deciphered using population genetics. It is possible that this radiation results from a complex mixture of adaptive radiation and both peripatric and sympatric speciation.

In chapters 8.1 and 8.2, I focused on diversification processes in New Guinea. This island is of recent origin even though some old terranes form its backbone. During the past 10 million years, drifting terranes and tectonic-driven orogeny have engendered one of the most dazzling geological assemblages in the region with summits up to 5000 meters covered by dense tropical forests stretched out to vast lowlands veined by rivers and torrents. In such a recent environment, it is striking to observe remarkable radiations such as tree kangaroos,

birds of paradise or birdwing butterflies. To investigate the mechanisms governing the rapid diversification of clades on this island, I used molecular phylogenies and reconstructed the evolutionary history of two radiations of diving beetles.

In chapter 8.1, I studied the mechanisms governing the evolution of a widespread species occurring from Portugal to New Zealand. *Rhantus suturalis* is actually a complex of cryptic species of recent origin comprising New Guinean highland endemics. I showed that glaciations acted in a different manner as for the STR by isolating populations in sky island refugia whilst the remainder of the populations were restricted to lowlands and connected by gene flow. As a result, the populations that successfully adapted to a cold climate in these refugia evolved independently from the main core population but also from each other yielding several mountain endemics. We suggested that after secondary contact when climate change allowed the core population to recolonize highlands, gene flow was prevented because of independent lineage evolution. Nowadays, both endemics and core populations can be found syntopically occurring in New Guinean highlands and therefore Quaternary climate changes in the New Guinean cordillera sky islands have triggered peripheral speciation in this group.

In chapter 8.2, I investigated the effect of the geological history of New Guinea on a highly diverse group of diving beetles. The genus *Exocelina* is widespread in Melanesia but in New Guinea its diversity is astonishing with about 150 known species and certainly much more to discover. Using paleogeological reconstructions and a dated molecular phylogeny, we were able to unfold the evolutionary history of the radiation on the island. About 8 million years ago, the ancestor of the New Guinean radiation colonized the island out of Australia. At that time New Guinea was beginning to emerge as a result of the Asian and Australian plate collision. During the next million years the island stretched vertically and horizontally through tectonics and collision of drifting island arcs. The availability of new niches fostered an extensive diversification of *Exocelina* diving beetles which are running-water associated and therefore prone to micro-endemism. Whilst the New Guinean orogeny continues, the pace at which these beetles diversify has not diminished. However, based on estimations of density-dependence effect on diversification rate, we suggest that the velocity of this diversification is likely to have exceeded the speed at which new habitats are created and therefore the radiation might undergo a slow-down in a close future.

Overall, the processes of lineage diversification in New Guinea are markedly different than the ones unveiled in Australia as expected from the different geological histories of these islands. Diversification patterns in Australia are mainly drove by climate shifts since the Miocene. The aridification engendered by tectonics has carved dramatically different habitats on the island forcing clades to migrate, adapt or run extinct. In particular the impact of Plio-Pleistocene glaciations is remarkable in the two clads investigated here. The disruptions of rainfall seasonality and the ongoing aridification during this period have considerably affected aquatic clades as illustrated by the pattern of declining-diversity highlighted in Hydroporini diving beetles. Across the Torres Strait, New Guinea presents different patterns of diversification. Climatic but also geological disruptions have triggered the evolution of clades on the island and regarding its young age, it seems clear that New Guinea acts as a species pump in the entire archipelago.

Biogeography on the right tracks

In Chapters 9 and 10, I focused on the impact of paleogeological and paleoclimatic factors on the biogeography of insect radiations in the Indomalayan / Australasian archipelago. In chapter 9, we investigated the role of Wallace's line on the phylogeography of *Trigonopterus* flightless weevils. This boundary is of particular interest because it is a notorious demarcation between the Asian and Australian biotas. To test its impact on biogeographical histories, flightless taxa constitute a much better model than vagile groups since this line represents a very narrow marine gap between the two regions. Indeed, it has been shown that during periods of sea-level fluctuations, the Sunda shelf was a large continental peninsula where land bridges existed between Sumatra, Java, Borneo and Bali. On the other hand, the deep Makassar Strait delineating Bali from Lombok and Borneo from Sulawesi prevented any land dispersal across Wallace's line and therefore groups of non-flying taxa are expected to comprise clades with more affinities between islands of the Sunda Shelf than between landmasses on either side of the barrier. In particular, Balinese taxa for which only a very scarce literature exists are thought to be of Javanese origin since the two islands are only a few kilometers away from each other. Strikingly, we found that Balinese *Trigonopterus* weevils are of Eastern origin therefore clearly highlighting several transgressions of Wallace's line. This result is even more striking given the fact that *Trigonopterus* beetles are microendemics as highlighted by our phylogeographic reconstructions and therefore do not represent a good candidate for dispersal. Although the dispersal mode involved in the

evolution of these weevils remains elusive (rafting, passive transport in bird stomach…), the biogeographical pattern highlighted in this study combined with other recent studies sheds light on the permeability of Wallace's line to insects.

In chapter 10, we investigated the biogeographical history of a widespread group of butterflies. The subgenus *Papilio* (*Achillides*) is spread in the entire archipelago and as such represents an ideal model to study the impact of past geological and climatic disruptions on distributional ranges. Using a comprehensive analytical toolbox, we recovered a potential cryptic diversity within this group but more importantly, we were able to link present-day distributions of the different species with fine-scale paleoreconstructions of the region. We revealed a Miocene origin of this group in the Sunda shelf / Wallacea with a complex mixture of vicariant and dispersal events responsible for the present-day distribution of these swallowtails in the Indomalayan / Australasian archipelago. The biogeographical pattern highlighted in this study has likely been fostered by climatic as well as geological factors. In particular sea-level fluctuations engendered by temperature oscillations throughout the Neogene and Quaternary have fostered the formation of island endemics therefore supporting the heart of the Indomalayan / Australasian archipelago has a cradle of diversity.

Conclusion

In this Ph.D. dissertation, I emphasized the importance of an emerging integrative taxonomy to accelerate species description. The discovery of new species is not only relevant for conservation planning and applied biodiversity research but also to study the origin of the astonishing diversity on Earth. Cardinal progresses in our understanding of lineage diversification can be achieved when looking at the evolution of clades in a phylogenetic framework. Here, I show that the overflowing biodiversity encountered in the Indomalayan / Australasian archipelago stems from multiple mechanisms of diversification often correlated with large-scale abiotic factors. Groundbreaking advances in molecular phylogenetic models and DNA sequencing allow the investigation of biodiversity patterns and processes at an increasingly finer-scale. In particular, the substantial development of next-generation sequencing in the past decade paves the way for genome-scale studies liable to provide an even deeper understanding of evolutionary processes governing biodiversity dynamics. Overall, the study of the origin of Earth's biodiversity is at a watershed, moving towards a highly integrative and blooming science that may assume a paramount role in the future.

References

Alfaro ME, Santini F, Brock C, Alamillo H, Dornburg A, Rabosky DL, Carnevale G, Harmon LJ (2009) Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences of the United States of America*, 106(32):13410–13414.

Antonelli A, Sanmartín I (2011) Why are there so many plant species in the Neotropics? *Taxon*, 60:403–414.

Avise JC (2000) *Phylogeography: the history and formation of species*. Harvard University Press.

Azuma N, Ogata K, Kikuchi T, Higashi S (2006) Phylogeography of Asian weaver ants, *Oecophylla smaragdina*. *Ecological Research*, 21(1):126–136.

Balke M (2005) 7.6. Dytiscidae Leach, 1915. Pp. 90–116 in: Beutel RG, Leschen RAB (eds.): *Handbook of zoology. A natural history of the phyla of the animal kingdom. Volume IV. Arthropoda: Insecta. Part 38. Coleoptera, Beetles. Volume 1: Morphology and systematics (Archostemata, Adephaga, Myxophaga, Polyphaga partim.).* Walter de Gruyter, Berlin, New York.

Balke M, Pons J, Ribera I, Sagata K, Vogler AP (2007) Infrequent and unidirectional colonization of hyperdiverse *Papuadytes* diving beetles in New Caledonia and New Guinea. *Molecular Phylogenetics and Evolution*, 42(2):505–516.

Balke M, Gómez–Zurita J, Ribera I, Viloria A, Zillikens A, Steiner J, Garcia M, Hendrich L, Vogler AP (2008) Ancient associations of aquatic beetles and tank bromeliads in the Neotropical forest canopy. *Proceedings of the National Academy of Sciences of the United States of America*, 105(17):6356–6361.

Balke M, Ribera I, Hendrich L, Miller MA, Sagata K, Posman A, Vogler AP, Meier R (2009) New Guinea highland origin of a widespread arthropod supertramp. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 276(1666):2359–2367.

Barluenga M, Stölting KN, Salzburger W, Muschick M, Meyer A (2006) Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature*, 439(7077):719–723.

Barnosky AD (2001) Distinguishing the effects of the Red Queen and Court Jester on Miocene mammal evolution in the northern Rocky Mountains. *Journal of Vertebrate Paleontology*, 21(1):172–185.

Benton MJ (2009) The Red Queen and the Court Jester: species diversity and the role of biotic and abiotic factors through time. *Science*, 323:728–732.

Benton MJ, Donoghue PC (2007) Paleontological evidence to date the tree of life. *Molecular biology and evolution*, 24(1):26–53.

Benton MJ, Donoghue PCJ, Asher RJ (2009) Calibrating and constraining molecular clocks. *The timetree of life*, 35–86.

Bolnick DI, Fitzpatrick BM (2007) Sympatric speciation: models and empirical evidence. *Annual Review of Ecology, Evolution, and Systematics*, 459–487.

Braby MF, Trueman JWH, Eastwood RG (2005) When and where did troidine butterflies (Lepidoptera: Papilionidae) evolve? Phylogenetic and biogeographic evidence suggests an origin in remnant Gondwana in the Late Cretaceous. *Invertebrate Systematics*, 19:113–143.

Braby MF, Pierce NE (2007) Systematics, biogeography and diversification of the Indo-Australian genus *Delias* Hübner (Lepidoptera: Pieridae): phylogenetic evidence supports an 'out-of-Australia'origin. *Systematic Entomology*, 32(1):2–25.

Brower AV (1994) Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 91(14):6491–6495.

Brower AV (2010) Alleviating the taxonomic impediment of DNA barcoding and setting a bad precedent: names for ten species of '*Astraptes fulgerator*' (Lepidoptera: Hesperiidae: Eudaminae) with DNA–based diagnoses. *Systematics and Biodiversity*, 8(4):485–491.

Buffon GL (1749–1788) *Histoire naturelle, générale et particulière*. Paris: Presses Universitaires de France.

Butlin R, Bridle J, Schluter D (2009) *Speciation and patterns of diversity*. Cambridge: Cambridge University Press.

Carstens BC, Pelletier TA, Reid NM, Satler JD (2013) How to fail at species delimitation. *Molecular ecology*, 22(17):4369–4383.

Chen J, Jiang LY, Qiao GX (2014) A total-evidence phylogenetic analysis of Hormaphidinae (Hemiptera: Aphididae), with comments on the evolution of galls. *Cladistics*, 30(1):26–66.

Condamine FL, Rolland J, Morlon H (2013) Macroevolutionary perspectives to environmental change. *Ecology letters*, 16:72–85.

Costello MJ, May RM, Stork NE (2013) Can we name Earth's species before they go extinct? *Science*, 339(6118):413–416.

Coyne JA, Orr HA (2004) *Speciation* (Vol. 37). Sunderland, MA: Sinauer Associates.

Cox CB, Moore PD (2010) *Biogeography: an ecological and evolutionary approach*. John Wiley & Sons.

Croizat L (1958) *Panbiogeography: The Old World. 2 v* (Vol. 2). L Croizat.

Cracraft J (1983) Species concepts and speciation analysis. In *Current ornithology* (pp. 159–187). Springer US.

Darwin C (1859) *On the Origin of Species.* John Murray, London.

Darwin C, Wallace A (1858) On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *Journal of the proceedings of the Linnean Society of London. Zoology*, 3(9):45–62.

Darwin C (1909) *The voyage of the beagle*. New York, NY: P.F. Collier.

de Bruyn M, Wilson JC, Mather PB (2004) Reconciling geography and genealogy: phylogeography of giant freshwater prawns from the Lake Carpentaria region. *Molecular Ecology*, 13:3515–26.

de Bruyn M, Rüber L, Nylinder S, Stelbrink B, Lovejoy NR, Lavoué S, Tan HH, Nugroho E, Wowor D, Ng PKL, Azizah MNS, Von Rintelen T, Hall R, Carvalho GR (2013) Paleo–drainage basin connectivity predicts evolutionary relationships across three southeast Asian biodiversity hotspots. *Systematic Biology*, 62(3):398–410.

de Bruyn M, Mather PB (2007) Molecular signatures of Pleistocene sea–level changes that affected connectivity among freshwater shrimp in Indo–Australian waters. *Molecular Ecology*, 16:4295–307.

de Jong R (2004) Phylogeny and biogeography of the genus *Taractrocera* Butler, 1870 (Lepidoptera: Hesperiidae), an example of Southeast Asian–Australian interchange. *Zoologische Mededelingen Leiden*, 78:383–415.

De Queiroz K (2007) Species concepts and species delimitation. *Systematic biology*, 56(6), 879–886.

Dieckmann U, Doebeli M (1999) On the origin of species by sympatric speciation. *Nature*, 400(6742):354–357.

Donoghue PC, Benton MJ (2007) Rocks and clocks: calibrating the Tree of Life using fossils and molecules. *Trends in Ecology & Evolution*, 22(8):424–431.

Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology*, 4(5):e88.

Edwards CJ, Suchard MA, Lemey P, Welch JJ, Barnes I, Fulton TL, Barnett R, O'Connell TC, Coxon P, Monaghan N, Valdiosera CE, Lorenzen ED, Willerslev E, Baryshnikov

GF, Rambaut A, Thomas MG, Bradley DG, Shapiro B (2011) Ancient hybridization and an Irish origin for the modern polar bear matriline. *Current Biology*, 21(15):1251–1258.

Ence DD, Carstens BC (2011) SpedeSTEM: a rapid and accurate method for species delimitation. *Molecular Ecology Resources*, 11(3):473–480.

Etienne RS, Haegeman B, Stadler T, Aze T, Pearson PN, Purvis A, Phillimore AB (2012) Diversity–dependence brings molecular phylogenies closer to agreement with the fossil record. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 279(1732):1300–1309.

Felsenstein J (1968) *Statistical inference and the estimation of phylogenies* (Doctoral dissertation, University of Chicago).

FitzJohn RG (2012) Diversitree: comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution*, 3:1084–1092.

Forster JR (1778) *Observations made during a voyage round the world*. University of Hawaii Press.

Fujisawa T, Barraclough TG (2013) Delimiting species using Single–Locus Data and the Generalized Mixed Yule Coalescent Approach: a revised method and evaluation on simulated data sets. *Systematic biology*, 62(5):707–724.

Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C (2012) Coalescent–based species delimitation in an integrative taxonomy. *Trends in ecology & evolution*, 27(9):480–488.

Gaston KJ (2000) Global patterns in biodiversity. *Nature*, 40:220–227.

Gavrilets S (2003) Perspective: models of speciation: what have we learned in 40 years? *Evolution*, 57(10):2197–2215.

Gavrilets S, Vose A (2007) Case studies and mathematical models of ecological speciation. 2. Palms on an oceanic island. *Molecular Ecology*, 16(14):2910–2921.

Gavrilets S, Vose A, Barluenga M, Salzburger W, Meyer A (2007) Case studies and mathematical models of ecological speciation. 1. Cichlids in a crater lake. *Molecular Ecology*, 16(14):2893–2909.

Geyer CJ (1991) Markov chain Monte Carlo maximum likelihood. pp 156–163 in: *Computing science and statistics: Proceedings of the 23rd Symposium on the Interface* (E. M. Keramidas, ed.). Interface Foundation, Fairfax Station.

Goldberg EE, Lancaster LT, Ree RH (2011) Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Systematic Biology*, 60(4):451–465.

Goldstein PZ, DeSalle R (2011) Integrating DNA barcode data and taxonomic practice: determination, discovery, and description. *Bioessays*, 33(2):135–147.

Grandcolas P, Murienne J, Robillard T, Desutter–Grandcolas L, Jourdan H, Guilbert E, Deharveng L (2008) New Caledonia: a very old Darwinian island? *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, 363(1508):3309–3317.

Graur D, Martin W (2004) Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. *Trends in Genetics*, 20(2):80–86.

Guil N, Machordom A, Guidetti R (2013) High level of phenotypic homoplasy amongst eutardigrades (Tardigrada) based on morphological and total evidence phylogenetic analyses. *Zoological Journal of the Linnean Society*, 169(1):1–26.

Hajibabaei M, Singer GA, Hebert PD, Hickey DA (2007) DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends in Genetics*, *23*(4), 167–172.7

Hall R (2002) Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific: computer–based reconstructions, model and animations. *Journal of Asian Earth Sciences*, 20(4):353–431.

Hall R (2011) Australia–SE Asia collision: plate tectonics and crustal flow. *Geological Society, London, Special Publications*, 355(1):75–109.

Hall R (2012) Late Jurassic–Cenozoic reconstructions of the Indonesian region and the Indian Ocean. *Tectonophysics*, 570:1–41.

Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W (2008) GEIGER: investigating evolutionary radiations. *Bioinformatics*, 24(1):129–131.

Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.

Hausdorf B (2011) Progress toward a general species concept. *Evolution*, 65(4):923–931.

Hausdorf B, Hennig C (2010) Species delimitation using dominant and codominant multilocus markers. *Systematic Biology*, 59(5):491–503.

Hawlitschek O, Porch N, Hendrich L, Balke M (2011) Ecological niche modeling and nDNA sequencing support a new, morphologically cryptic beetle species unveiled by DNA barcoding. *PLoS One*, 6(2):e16662.

Heads M (2011) Old taxa on young islands: a critique of the use of island age to date island–endemic clades and calibrate phylogenies. *Systematic Biology*, 60(2):204–218.

Hebert PD, Cywinska A, Ball SL (2003a) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1512):313–321.

Hebert PD, Ratnasingham S, de Waard JR (2003b) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270:S96–S99.

Hebert PD, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004a) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, 101(41):14812–14817.

Hebert PD, Stoeckle MY, Zemlak TS, Francis CM (2004b) Identification of birds through DNA barcodes. *PLoS Biology*, 2(10):e312.

Hendrich L, Pons J, Ribera I, Balke M (2010) Mitochondrial cox1 sequence data reliably uncover patterns of insect diversity but suffer from high lineage–idiosyncratic error rates. *PLoS One*, 5(12):e14448.

Hendry AP (2009) Evolutionary biology: speciation. *Nature*, 458(7235):162–164.

Hennig W (1966) *Phylogenetic Systematics*. Urbana: University of Illinois Press.

Ho SYW, Phillips MJ (2009) Accounting for Calibration Uncertainty in Phylogenetic Estimation of Evolutionary Divergence Times. *Systematic Biology*, 58(3):367–380.

Huelsenbeck JP, Larget B, Miller RE, Ronquist F (2002) Potential applications and pitfalls of Bayesian inference of phylogeny. *Systematic biology*, 51(5):673–688.

Huelsenbeck JP, Andolfatto P, Huelsenbeck ET (2011) Structurama: Bayesian inference of population structure. *Evolutionary bioinformatics online*, 7:55.

Jäch MA, Balke M (2008) Global diversity of water beetles (Coleoptera) in freshwater. *Hydrobiologia*, 595(1):419-442.

Jörger KM, Schrödl M (2013) How to describe a cryptic species? Practical challenges of molecular taxonomy. *Frontiers in Zoology*, 10:1742-9994.

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(11):1463–1464.

Kingman JFC (1982) The coalescent. *Stochastic processes and their applications*, 13(3):235–248.

Kitching RL (1981) The geography of the Australian Papilionoidea. *Ecological biogeography of Australia*, 977–1005.

Klaus S, Selvandran S, Goh J W, Wowor D, Brandis D, Koller P, Schubart CD, Streit B, Meier R, Ng PKL, Yeo D C (2013) Out of Borneo: Neogene diversification of Sundaic freshwater crabs (Crustacea: Brachyura: Gecarcinucidae: Parathelphusa). *Journal of Biogeography*, 40(1):63–74.

Kondo K, Shinkawa T, Matsuka H (2003) Molecular systematics of birdwing butterflies (Papilionidae) inferred from mitochondrial ND5 gene. *Journal of the Lepidopterists' Society*, 57:17–24.

Kubatko LS, Carstens BC, Knowles LL (2009) STEM: species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics*, 25(7):971–973.

Leaché AD, Fujita MK (2010) Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 277(1697):3071–3077.

Leys R, Watts CHS, Cooper SJB, Humphreys WF (2003) Evolution of subterranean diving beetles (Coleoptera: Dytiscidae: Hydroporini, Bidessini) in the arid zone of Australia. *Evolution*, 57(12):2819–2834.

Leys R, Watts CHS (2008) Systematics and evolution of the Australian subterranean hydroporine diving beetles (Dytiscidae), with notes on *Carabhydrus*. *Invertebrate Systematics*, 22(2):217–225.

Lohman DJ, de Bruyn M, Page T, von Rintelen K, Hall R, Ng PKL, Shih H–T, Carvalho GR, von Rintelen T (2011) Biogeography of the Indo–Australian Archipelago. *Annual Review of Ecology, Evolution, and Systematics*, 42:205–226.

Lohse K (2009) Can mtDNA barcodes be used to delimit species? A response to Pons et al. (2006). *Systematic Biology*, 58(4):439–442.

Lomolino MV, Brown JH, Sax DF (2010) Island biogeography theory. *The theory of island biogeography revisited*, 13–51.

MacArthur RH, Wilson EO (1963) An equilibrium theory of insular zoogeography. *Evolution*, 17(4):373–387.

MacArthur RH, Wilson EO (1967) *The Theory of Island Biogeography*. Princeton University Press, Princeton.

Maddison WP, Midford PE, Otto SP (2007) Estimating a binary character's effect on speciation and extinction. *Systematic Biology*, 56(5):701–710.

Malay MC MD, Paulay G (2010) Peripatric speciation drives diversification and distributional pattern of reef hermit crabs (Decapoda: Diogenidae: *Calcinus*). *Evolution*, 64(3):634–662.

Mayr E (1942) *Systematics and the origin of species, from the viewpoint of a zoologist*. Harvard University Press.

Mayr E (1963) *Animal species and evolution*. Belknap, Cambridge, MA.

Metcalfe I (2011) Tectonic framework and Phanerozoic evolution of Sundaland. *Gondwana Research,* 19(1):3-21.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (2004). Equation of state calculations by fast computing machines. *The journal of chemical physics,* 21(6):1087–1092.

Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology*, 3(12):e422.

Miller KB, Bergsten J (2014) The phylogeny and classification of predaceous diving beetles. In: *Ecology, Systematics, and the Natural History of Predaceous Diving Beetles (Coleoptera: Dytiscidae)*, pp. 49-172. Springer Netherlands.

Mittermeier RA, Robles Gil P, Hoffmann, Pilgrim J, Brooks T, Mittermeier CG, Lamoreux J, de Fonsescsa GAB, Seligmann PA, Ford H (2004) *Hotspots revisited: Earth's Biologically Richest and Most Threatened Ecoregions*. CEMEX, Mexico City, Mexico.

Monaghan MT, Balke M, Pons J, Vogler AP (2006) Beyond barcodes: complex DNA taxonomy of a South Pacific Island radiation. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 273(1588):887–893.

Monaghan MT, Wild R, Elliot M, Fujisawa T, Balke M, Inward DJ, Lees DC, Ranaivosolo R, Eggleton P, Barraclough TG, Vogler AP (2009) Accelerated species inventory on Madagascar using coalescent–based models of species delineation. *Systematic biology*, 58(3):298–311.

Mora C, Tittensor DP, Adl S, Simpson AG, Worm B (2011) How many species are there on Earth and in the ocean?. *PLoS Biology*, 9(8):e1001127.

Morlon H, Parsons TL, Plotkin JB (2011) Reconciling molecular phylogenies with the fossil record. *Proceedings of the National Academy of Sciences of the United States of America*, 108(39):16327–16332.

Morlon H (2014) Phylogenetic approaches for studying diversification. *Ecology letters*, 17(4):508–525.

Müller CJ, Wahlberg N, Beheregaray LB (2010) 'After Africa' – The Evolutionary History and Systematics of the genus *Charaxes* Ochsenheimer (Lepidoptera: Nymphalidae: Charaxinae) in the Indo–Pacific Region. *Biological Journal of the Linnean Society*, 100:457–481.

Müller CJ, Matos-Maraví PF, Beheregaray LB (2013) Delving into *Delias* Hübner (Lepidoptera: Pieridae): fine-scale biogeography, phylogenetics and systematics of the world's largest butterfly genus. *Journal Biogeography*, 40(5):881–893.

Müller CJ, Beheregaray LB (2010). Palaeo island–affinities revisited–Biogeography and systematics of the Indo–Pacific genus *Cethosia* Fabricius (Lepidoptera: Nymphalidae). *Molecular Phylogenetics and Evolution*, 57(1):314–326.

New TR (1999) *The evolution and characteristics of the Australian butterfly fauna*. In: Kitching RL, Jones RE, Scheermeyer E, Pierce NE, eds. Biology of Australian butterflies. Monographs of Australian Lepidoptera, Vol. 6. Melbourne: CSIRO Publishing, 33–52.

Niemiller ML, Near TJ, Fitzpatrick BM (2012) Delimiting species using multilocus data: diagnosing cryptic diversity in the southern cavefish, *Typhlichthys subterraneus* (Teleostei: Amblyopsidae). *Evolution*, 66(3):846–866.

Nilsson AN (2013) A world catalogue of the family Dytiscidae, or the diving beetles (Coleoptera, Adephaga). Version 1. I. 2013.

Nylander JA, Olsson U, Alström P, Sanmartín I (2008) Accounting for phylogenetic uncertainty in biogeography: a Bayesian approach to dispersal–vicariance analysis of the thrushes (Aves: *Turdus*). *Systematic Biology*, 57(2):257–268.

O'Meara BC (2010) New heuristic methods for joint species delimitation and species tree inference. *Systematic Biology*, 59(1), 59–73.

Padial JM, Miralles A, De la Riva I, Vences M (2010) Review: The integrative future of taxonomy. *Frontiers in Zoology*, 7:1–14.

Papadopoulou A, Monaghan MT, Barraclough TG, Vogler AP (2009) Sampling error does not invalidate the Yule–coalescent model for species delimitation. A response to Lohse (2009). *Systematic Biology*, 58(4):442–444.

Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20(2):289–290.

Parham JF, Donoghue PC, Bell CJ, Calway TD, Head JJ, Holroyd PA, Inoue JG, Irmis RB, Joyce WG, Ksepka DT, Patané JSL, Smith ND, Tarver JE, van Tuinen M, Yang Z, Angielczyk KD, Greenwood JM, Hipsley CA, Jacobs L, Makovicky PJ, Müller J, Smith

KT, Theodor JM, Warnock RCM, Benton MJ (2012) Best practices for justifying fossil calibrations. *Systematic Biology*, 61(2):346–359.

Pons J, Barraclough T, Gomez–Zurita J, Cardoso A, Duran DP, Hazell S, Kamoun S, Sumlin W, Vogler AP (2006) Sequence–based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, 55:595–609.

Posada D (2009) Selection of models of DNA evolution with jModelTest. In: *Bioinformatics for DNA sequence analysis* (pp. 93–112). Humana Press.

Prevosti FJ (2010) Phylogeny of the large extinct South American canids (Mammalia, Carnivora, Canidae) using a "total evidence" approach. *Cladistics*, 26(5):456–481.

Pybus OG (2006) Model selection and the molecular clock. *PLoS Biology*, 4(5):e151.

Pyron RA, Burbrink FT (2013) Phylogenetic estimates of speciation and extinction rates for testing ecological and evolutionary hypotheses. *Trends in ecology & evolution*, 28(12):729–736.

Rabosky DL (2006) Likelihood methods for detecting temporal shifts in diversification rates. *Evolution*, 60(6):1152–1164.

Rannala B, Yang Z (2013) Improved reversible jump algorithms for Bayesian species delimitation. *Genetics*, 194(1):245–253.

Raxworthy CJ, Ingram CM, Rabibisoa N, Pearson RG (2007) Applications of ecological niche modeling for species delimitation: a review and empirical evaluation using day geckos (*Phelsuma*) from Madagascar. *Systematic biology*, 56(6):907–923.

Ree RH, Moore BR, Webb CO, Donoghue MJ (2005) A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution*, 59:2299–2311.

Ree RH, Smith SA (2008) Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, 57(1):4–14.

Ree RH, Sanmartín I (2009) Prospects and challenges for parametric models in historical biogeographical inference. *Journal of Biogeography*, 36(7):1211–1220.

Reid NM, Carstens BC (2012) Phylogenetic estimation error can decrease the accuracy of species delimitation: a Bayesian implementation of the general mixed Yule–coalescent model. *BMC evolutionary biology*, 12(1):196.

Ribera I, Vogler AP, Balke M (2008) Molecular phylogeny and diversification of diving beetles (Coleoptera: Dytiscidae). *Cladistics*, 24(4):563–590.

Ricklefs RE (2007) Estimating diversification rates from phylogenetic information. *Trends in Ecology & Evolution*, 22(11):601–610.

Riedel A, Sagata K, Suhardjono YR, Tänzler R, Balke M (2013) Integrative taxonomy on the fast track–towards more sustainability in biodiversity research. *Frontiers in zoology*, 10(1):15.

Rissler LJ, Apodaca JJ (2007) Adding more ecology into species delimitation: ecological niche models and phylogeography help define cryptic species in the black salamander (*Aneides flavipunctatus*). *Systematic Biology*, 56(6):924–942.

Ronquist F (1997) Dispersal–vicariance analysis: a new approach to the quantification of historical biogeography. *Systematic Biology*, 46(1):195–203.

Ronquist F, van der Mark P, Huelsenbeck JP (2009) Bayesian phylogenetic analysis using MrBayes. In: *The phylogenetic handbook*.

Ronquist F, Klopfstein S, Vilhelmsen L, Schulmeister S, Murray DL, Rasnitsyn AP (2012) A total–evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Systematic Biology*, 61(6):973–999.

Satler JD, Carstens BC, Hedin M (2013) Multilocus species delimitation in a complex of morphologically conserved trapdoor spiders (Mygalomorphae, Antrodiaetidae, *Aliatypus*). *Systematic biology*, 62(6):805–823.

Savolainen V, Anstett MC, Lexer C, Hutton I, Clarkson JJ, Norup MV, Powell MP, Springate D, Salamin N, Baker WJ (2006) Sympatric speciation in palms on an oceanic island. *Nature*, 441(7090):210–213.

Schlick–Steiner BC, Steiner FM, Seifert B, Stauffer C, Christian E, Crozier RH (2010) Integrative taxonomy: a multisource approach to exploring biodiversity. *Annual review of Entomology*, 55:421–438.

Schluter D (2000) *The ecology of adaptive radiation*. Oxford University Press.

Schluter D (2001) Ecology and the origin of species. *Trends in Ecology & Evolution*, 16(7):372–380.

Schluter D (2009) Evidence for ecological speciation and its alternative. *Science*, 323(5915):737–741.

Sclater PL (1858) On the general geographical distribution of the members of the class Aves. *Journal of the Proceedings of the Linnean Society of London. Zoology*, 2(7):130–136.

Song H, Buhay JE, Whiting MF, Crandall KA (2008) Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences of the United States of America*, 105(36):13486–13491.

Stadler T (2011) Inferring speciation and extinction processes from extant species data. *Proceedings of the National Academy of Sciences of the United States of America*, 108(39):16145–16146.

Stadler T (2013) Recovering speciation and extinction dynamics based on phylogenies. *Journal of Evolutionary Biology*, 26:1203–1219.

Stelbrink B, Albrecht C, Hall R, von Rintelen T (2012) The biogeography of Sulawesi revisited: is there evidence for a vicariant origin of taxa on Wallace's "anomalous island"? *Evolution*, 66(7):2252–2271.

Su Y–C, Chang Y–H, Lee S–C, Tso I–M (2007) Phylogeography of the giant wood spider (*Nephila pilipes*, Araneae) from Asian–Australian regions. *Journal of Biogeography*, 34:177–91

Swofford DL, Olsen GJ, Waddell PJ, Hillis DM (1996) Phylogenetic inference. In: In *Molecular systematics* (2nd ed.) (1996), pp. 407-514 (eds Hillis DM, Moritz C, Mable BK).

Talavera G, Dincă V, Vila R (2013) Factors affecting species delimitations with the GMYC model: insights from a butterfly survey. *Methods in Ecology and Evolution*, 4(12):1101–1110.

Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. *Trends in Ecology & Evolution*, 18(2):70–74.

Thorne JL, Kishino H, Painter IS (1998) Estimating the rate of evolution of the rate of molecular evolution. *Molecular Biology and Evolution*, 15(12):1647–1657.

Van Valen, L. (1973). A new evolutionary law. *Evolutionary theory*, *1*, 1–30.

Via S (2001) Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology & Evolution*, 16(7):381–390.

Vogler AP, Monaghan MT (2007) Recent advances in DNA taxonomy. *Journal of Zoological Systematics and Evolutionary Research*, 45(1):1–10.

von Humboldt A (1805–1834) *Essai sur la géographie des plantes; accompagne d'un tableau physique des régions équinoxiales*. Levrault, Paris, France. 158 pp.

von Linné C (1758) *Systema Naturae*, 10th edn. Stockholm, Sweden: Laurentius Salvius

von Rintelen T, von Rintelen K, Glaubrecht M, Schubart CD, Herder F (2011) Aquatic biodiversity hotspots in Wallacea—the species flocks in ancient lakes of Sulawesi, Indonesia. In: *Biotic Evolution and Environmental Change in Southeast Asia*, ed. DJ Gower, KG Johnson, JE Richardson, BR Rosen, L Rüber, ST Williams. Cambridge: Cambridge Univ. Press.

Wallace AR (1860) On the Zoological Geography of the Malay Archipelago. *Journal of the proceedings of the Linnean Society*, 4:172–184.

Wallace AR (1863) On the physical geography of the Malay Archipelago. *Journal of the Royal Geographical Society of London*, 33:217–234.

Wallace AR (1869) *The Malay Archipelago: the land of the orang–utan and the bird of paradise; a narrative of travel, with studies of man and nature*. Courier Corporation.

Wallace AR (1876) *The Geographical Distribution of Animals: With a Study of the Relations of Living and Extinct Faunas as Elucidating the Past Changes of the Earth's Surface*. London: Macmillan.

Wallace AR (1880) *Island life: the phenomena and causes of insular faunas and floras*. London: Macmillan.

Wegener A (1912) Die entstehung der kontinente. *Geologische Rundschau*, 3(4):276–292.

Wiemers M, Fiedler K (2007) Does the DNA barcoding gap exist? A case study in blue butterflies (Lepidoptera: Lycaenidae). *Frontiers in Zoology*, 4(8):1–16.

Wiens JJ (2007) Species delimitation: new approaches for discovering diversity. *Systematic Biology*, 56(6):875–878.

Wiens JJ, Donoghue MJ (2004) Historical biogeography, ecology and species richness. *Trends in ecology & evolution*, 19(12):639–644.

Williams KJ, Ford A, Rosauer DF, De Silva N, Mittermeier R, Bruce C, Larsen FW, Margules C (2011) Forests of East Australia: the 35th biodiversity hotspot. In *Biodiversity Hotspots* (pp. 295–310). Springer Berlin Heidelberg.

Yang Z, Rannala B (2006) Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution*, 23(1):212–226.

Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, 107(20):9264–9269.

Yoder AD, Yang Z (2000) Estimation of primate speciation dates using local molecular clocks. *Molecular Biology and Evolution*, 17(7):1081–1090.

Zhang J, Kapli P, Pavlidis P, Stamatakis A (2013) A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, 29(22):2869–2876.

Zuckerkandl E, Pauling L (1962) Molecular disease, evolution and genetic heterogeneity. In: *Horizons in Biochemistry*, pp. 189-225.

Picture Credits

- Cover. jaiprox, zin.ru, LemonTeaYK, Stepehen Walford.
- Inside Cover. Chris Watts, Naturhistorisches Museum Wien; Thomas Neubauer, Michael Balke, The Insect Collector, Science/AAAS.
- Chapter 1 cover. National Geographic, Michael Balke.
- Figure 1. Mora et al. (2011)
- Figure 3. Satler et al. (2013)
- Figure 4. lepidopterran.blogspot.de.
- Figure 5. WillemSvdMerwe, Luciano Corbella, Encyclopedia Encarta, Nobu Tamura.
- Chapter 2 cover. Michael Balke.
- Figure 6. Udo Schmidt, Michael Balke, Warren Ishii.
- Figure 9. Grimaldi and Engel, Anders Leth Damgaard, Udo Schmidt, Michael balke.
- Figure 10. The Insect Collector, mushikiti.c.blog.so-net.ne.jp.
- Figure 11. Ronald Blakey.
- Chapter 3 cover. theclassytraveller.com.
- Figure 12. National Geographic.
- Figure 13. Hall (2012).
- Figure 14. Jan Hamrsky. homepage1.nifty.com, Museum Victoria Australia.
- Figure 15. Albert, Vengolis, 4eyes_tang, Lisa Cancade.
- Chapter 4 cover. durnfordi.blog62.fc2.com, Bill Frymire.
- Chapter 5 cover. team-consulting.com, vlupadya2000, stuartxchange.com.
- Chapter 6 cover. Balke et al. (2013), team-consulting.com, Kirk Hartley.
- Chapter 7 cover. NASA.
- Chapter 8 cover. Taro Taylor.
- Chapter 9 cover. http://larvalsubjects.wordpress.com.
- Chapter 10 cover. Bernard Dupont.
- Outcomes of the thesis cover. Jan Hamrsky

ACKNOWLEDGEMENTS

First and foremost I wish to thank my advisor Michael Balke who was daring enough to gamble on me despite my blatant fondness for hedonism. Not only I have never lacked of anything to conduct my research but you spoiled me with trips across the planet. What else could I possibly be asking for? Wine, cheese and barbecues naturally, which you provided like a horn of plenty along with Ditta and the girls. As you read these lines my Ph. D. ends, and I hope you have no regrets even though I busily participated to clear out your wine cellar. If it may reinvigorate you, know that I am genuinely persuaded our future collaboration will bring an even larger heap of geese, ducks, lambs, pheasants, well basically any roastable animal, but also tartiflettes, fondues and wine for us to relish… Thanks for everything man.

Second, I would like to warmly thank Lars Hendrich who was my co-advisor and helped me to "overcome" my horrifying taxonomic mediocrity even though I still cannot say that I am a specialist (I did start to pin butterflies though…). Thanks for that, your availability, boundless knowledge of the (Australasian) diving beetles and daily joviality.

My acknowledgements also go to Gerhard Haszprunar and the entire team of the ZSM. Thank you for running so efficiently the ZSM and helping students to achieve their ambitions.

I would also like to thank my co-authors and collaborators. I feel particularly indebted to Mark de Bruyn, Robert Hall, Axel Hausmann, Oliver Hawlitschek, Bruno Le Ru, Michael Monaghan, Chris Müller, Joan Pons, Nick Porch, Ignacio Ribera, Alexander Riedel, Helena Shaverdo, Andrew Short, Jean-Francois Silvain, Bernard Turlin and Thomas von Rintelen.

A very special acknowledgement goes to Fabien Condamine and Gael Kergoat. Both of you are not only outstanding colleagues but friends and I look forward to seeing you in Montpellier to (work and) enjoy some beach-volley/barbecue/beer afternoons. Santé les gars!

A warm thank you to the beetle office crew. Arnaud, I am proud of the French cliché we incarnated these past few years, eating saucisson and drinking wine whilst admiring the dazzling tropical sunsets of Bavaria through the office window. I hope the tradition will keep up... I am glad we succeeded to initiate a project on your creepy blind beetles because it will allow us to catch up from time to time. Jeromimo/Big C, I keep dantesque memories of some

great parties (Let's move yeah // PEDO) and I hope we will find time to catch up either in Munich, Paris or elsewhere in the future. Last but not the least, Rene my everyday wingman in the office. I hope you will never forget all these valuable lessons of French. If only you were slightly better at badminton I could have avoided gaining 15 pounds in 3 years. If only… Thanks for everything man and continue to send me the doodles for soccer who knows I might come along one day you won't expect it. Thanks guys for all the good times and I definitely look forward to drinking some Maß with you in a close future…

A special dedication to the ZSM soccer team, a little oasis to chill out at noon, and its members who were not previously mentioned: Bastian "Basti" Brenzinger, Alexander "Dunzi" Dunz, Peter "Pete" Kohnert and Vinicius "Vini" Padula.

I want to thanks all my friends from Munich and surroundings; Vio, Tom, Seb, Oliv, Fab, Tim, Sophie, etc… who have been as supportive as discombobulated when they first heard the topic of my research. Thanks also to my mates in Paris and elsewhere.

I would like to thank all the fantastic people I had the chance to come across during my journey in Sumatra, Java and Bali. I would particularly like to thank Ari Jamsari and all the participants of the IGN-TTRC project in Andalas and Brawijaya Universities. And of course Widodo the fearless driver with whom I had really great times drinking Bintang packs at night, eating (way too) spicy crab and smoking soaked cigarettes under the heavy rain.

Finally, I wish to thank my family in French for its support and understanding.

Merci pour votre soutien inconditionnel au cours de cette enrichissante escapade Bavaroise. Avec cette page qui se tourne, je souhaite tous vous remercier car au fond, qu'il s'agisse d'une oie rôtie à la table du célèbre 40 rue Frédéric Chopin, de plateaux d'huîtres au salon de la rue de Châtenay, de côtes de bœuf au feu de bois face aux Alpes, de poêlées de chanterelles grises à La Noë, d'une belle mousse Belge et souris d'agneau braisée à La Source, de jarrets grillés au Manoir, de salades estivales au Chemin du Châtaignier, ou tout simplement d'un café au coin d'une table, vous avez su me donner du cœur à l'ouvrage.

Merci aussi à ma plus fervente supportrice qui me suit bon gré mal gré dans mes périples dont la durée inconnue à une saveur douce-amère de découvertes et sacrifices…

Emmanuel FA Toussaint

E-Mail: toussaint@ku.edu

Curriculum Vitae

Current appointment

2015-2017: **Research Fellow**, University of Kansas, Department of Ecology and Evolutionary Biology, Division of Entomology, Biodiversity Institute, 1501 Crestline Drive, Suite 140, Lawrence, KS 66045-2811, USA.

Education

2011-2014: **Ph.D.**, Ludwig-Maximilians University of Munich, Professor-Huber-Platz 2, 80539 Munich, Germany / Bavarian State Collection of Zoology (ZSM), 21 Münchhausenstraße, 81247 Munich, Germany. *Summa cum laude.*

> Thesis: *Insects as a model to puzzle out mechanisms of lineage diversification in the Indomalayan / Australasian archipelago*.

2009-2011: **M.Sc.** Biology, Faculty of Sciences of Orsay, Paris-Sud XI University, 15 Rue Georges Clemenceau, 91400 Orsay, France.

> Thesis: *Origin and diversity of noctuid stemborers from the Afrotropical subtribe Sesamiina*.

Teaching and student mentoring

2011-2014: Jérôme Morinière (M.Sc., ZSM, Munich, Germany): Molecular phylogeny, diversification and historical biogeography of Colymbetinae diving beetles.

2012: Indonesian-German Network for Teaching, Training and Research Collaborations (IGN-TTRC), Module II: DNA Barcoding:

- Andalas University (18-29 June 2012, Padang, West Sumatra, Indonesia)
- Brawijaya University (2-13 July 2012, Malang, East Java, Indonesia)

2012: Daniel Frank (M.Sc. Student, ZSM, Munich, Germany): Molecular biodiversity assessment of diving beetles across the Indomalayan-Australasian archipelago.

Funded research programs

2014-2015: DFG Grant BA2152/20-1 (PI, Balke M, co-PI, **Toussaint EFA**): Old World tropical biogeography and the evolution of *Charaxes* butterflies. 32,000€.

2014-2015: DFG Grant BA2152/19-1 (PI, Balke M): New Guinea biogeography across the ophiolite and thrust-and-fold belt. 24,000€.

2011-2014: DFG Grant BA2152/11-1 (PI, Balke M): Evolution and biogeography in the Indomalayan-Melanesian-Oceanian region. 220,000€.

Publications in peer-reviewed journals

14. Bilton DT, **Toussaint EFA**, Turner C, Balke M (2015) *Capelatus prykei* gen. n., sp. n. (Coleoptera: Dytiscidae: Copelatinae) - a phylogenetically isolated diving beetle from the Western Cape of South Africa. *Systematic Entomology, in press.*

13. **Toussaint EFA**, Condamine FL, Hawlitschek O, Watts CHS, Porch N, Hendrich L, Balke M (2015) Unveiling the diversification dynamics of Australasian predaceous diving beetles in the Cenozoic. *Systematic Biology*, 64(1):3-24.

12. Le Ru BP, Capdevielle-Dulac C, **Toussaint EFA**, Conlong D, Van den Berg J, Pallangyo B, Ong'amo G, Chipabika G, Molo R, Overholt WA, Cuda JP, Kergoat GJ (2014) Molecular phylogenetics and systematics of *Acrapex* stem borers (Lepidoptera, Noctuidae, Apameini). *Invertebrate Systematics, 28(5):451-475.*

11. Hendrich L, **Toussaint EFA**, Balke M (2014) *Brancuporus* nov. gen. a new genus of hydroporine diving beetle (Coleoptera: Dytiscidae: Hydroporini) from Southwestern Australia. *Spixiana*, 37(1):103-109.

10. **Toussaint EFA**, Hall R, Monaghan M, Sagata K, Ibalim S, Shaverdo HV, Vogler AP, Pons J, Balke M (2014) The towering orogeny of New Guinea as a trigger for arthropod megadiversity. *Nature Communications*, 5:ncomms5001.

9. Tänzler R, **Toussaint EFA**, Suhardjono YR, Balke M, Riedel A (2014) Multiple transgressions of Wallace's Line explain diversity of flightless *Trigonopterus* weevils on Bali. *Proceedings of the Royal Society B, 281(1782):20132528*.

8. Balke M, Warikar E, **Toussaint EFA**, Hendrich L (2013) *Papuadessus baueri* sp.nov. from Biak Island, Papua (Coleoptera: Dytiscidae: Hydroporinae). *Spixiana*, 36(2):283-288.

7. Balke M, Hendrich L., **Toussaint EFA**, Zhou X, von Rintelen T, de Bruyn M (2013) Suggestions for a molecular biodiversity assessment of South East Asian freshwater invertebrates. Lessons from the megadiverse beetles (Coleoptera). *Journal of Limnology*, 72(2):61-68.

6. Balke M, Schmidt S, Hausmann A, **Toussaint EFA**, Bergsten J, Buffington M, Häuser C, Kroupa A, Riedel A, Polaszek A, Ubaidillah R, Krogmann L, Zwick L, Fikáček M, Hájek J, Michat M, Dietrich C, LaSalle J, Mantle B, Ng P (2013) Biodiversity into your hands – A call for a virtual global natural history 'metacollection'. *Frontiers in Zoology*, 10:55.

5. Balke M, **Toussaint EFA**, Hendrich L, Hájek J (2013) A new species of the Australian genus *Necterosoma* from Timor (Coleoptera: Dytiscidae: Hydroporini). *Acta Entomologica Musei Nationalis Pragae*, 53(1):65-74.

4. **Toussaint EFA**, Sagata K, Surbakti S, Hendrich L, Balke M (2013) Australasian sky islands act as a diversity pump facilitating peripheral speciation and complex reversal from narrow endemic to widespread ecological supertramp. *Ecology and Evolution.* 3(4):1031-1049.

3. Condamine FL, **Toussaint EFA**, Cotton A, Sperling FAH, Genson G, Kergoat GJ (2013) Fine-scale biogeographic and temporal diversification processes of peacock swallowtails (*Papilio* subgenus *Achillides*) in the Indo-Australian Archipelago. *Cladistics*, 29(1):88-111.

2. Hawlitschek O, Hendrich L, Espeland M, **Toussaint EFA**, Genner MJ, Balke M (2012) Pleistocene climate change promoted rapid diversification of aquatic invertebrates in South-East Australia. *BMC Evolutionary Biology*, 12:142.

1. **Toussaint EFA**, Condamine FL, Kergoat GJ, Capdevielle-Dulac C, Barbut J, Silvain J-F, Le Ru BP (2012) Palaeoenvironmental shifts drove the adaptive radiation of a noctuid stemborer tribe (Lepidoptera, Noctuidae, Apameini) in the Miocene. *PLoS ONE,* 7(7):e41377.

Publications submitted / under review

Toussaint EFA, Morinière J, Turlin B, Kunte K, Müller CJ, Balke M. An array of molecular species delimitation methods sheds light on species boundaries in the charismatic *Polyura* Nawab butterflies. Under review *(Molecular Phylogenetics and Evolution).*

Kergoat GJ, **Toussaint EFA**, Capdevielle-Dulac C, Clamens A-L, Ong'amo G, Conlong D, Van den Berg J, Cugala, D, Pallangyo B, Mubenga O, Chipabika G, Ndemah R, Sezonlin M, Bani G, Molo R, Catalayud P-A, Kaiser L, Silvain J-F, Le Ru BP. Morphological, molecular and biological evidence reveal six new species related to the Western corn borer *Sesamia nonagrioides* (Lefèbvre) (Lepidoptera, Noctuidae, Sesamiina). Under review *(Zoological Journal of the Linnean Society).*

Condamine FL, **Toussaint EFA**, Clamens A-L, Genson G, Sperling FAH, Kergoat GJ. Unfolding the evolution of birdwing butterflies 150 years after Alfred Russell Wallace. Under review *(Proceedings of the Royal Society B).*

Publications in preparation († equally contributing authors)

Toussaint EFA, Hendrich L, Shaverdo HV, Balke M. Mosaic patterns of post-colonization diversification dynamics in the Melanesian archipelago. *In preparation for The American Naturalist.*

Morinière J, **Toussaint EFA**, Bergsten J, Michat M, Hendrich L, Ribera I, Balke M. Not so tropical after all: Inverted latitudinal gradient of diversity in aquatic arthropods. *In preparation for Evolution*.

Toussaint EFA, Hendrich L, Lamm A, Short AEZ, Balke M. Long-distance dispersal events rather than Gondwanean splitting fostered the evolution of tropical freshwater beetles. *In preparation for Ecography.*

Hawlitschek O, **Toussaint EFA**, Gehring PS, Ratsoavina F, Cole N, Vences M, Glaw F. Overseas dispersal of a small gecko and the issue of geological calibration points. *In preparation for Journal of Biogeography*.

Toussaint EFA†, Morinière J, Jia F, Xu S, Beutel RG, Michat M, Ribera I, Balke M†. Molecular phylogenetics of Aspidytidae cliff beetles and description of the Chinese endemic *Sinoaspidytes* nov. gen.. *In preparation for Zoologica Scripta.*

Faille A†, Tänzler R, **Toussaint EFA†**. Shedding light on karstic structuration in populations of the micro-endemic cave beetle *Aphaenops cerberus* from southern France. *In preparation for Journal of Heredity.*

Toussaint EFA, Morinière J, Müller CJ, Wahlberg N, Balke M. Unfolding diversification processes of Emperor butterflies in the Old World Tropics. *In preparation for Proceedings of the National Academy of Sciences of the USA.*

Toussaint EFA, Morinière J, Müller CJ, Turlin B, Balke M. Molecular species delineations support the Nawab brush-footed butterfly *Polyura epigenes* as a widespread endemic from the Solomon Island archipelago. *In preparation for ZooKeys.*

Toussaint EFA, Morinière J, Dias FMS, Wahlberg N, Balke M, Vila R. Evolution of Anaeini brush-footed butterflies in the Neotropics. *In preparation for Journal of Biogeography*.

Expert reports and popular publications

1. Hendrich L, Hawlitschek O, **Toussaint EFA**, Tänzler R, Balke M (2014) Wiederfund des Schwimmkäfers *Graphoderus austriacus* (Sturm, 1834) in Bayern sowie weitere aktuelle Vorkommen von *Graphoderus bilineatus* (De Geer, 1774) im Gebiet des Starnberger Sees (Coleoptera: Dytiscidae, Noteridae, Hydrophilidae & Hydraenidae). *Nachrichtenblatt der bayerischen Entomologen*, 63(1/2) : 19-28.

Communications

11. **Toussaint EFA**. Les insectes d'un clos-masure au cœur de la biodiversité. Journées Européennes du Patrimoine 2014 au Manoir du Catel (20-21 September 2014, Ecretteville-lès-Baons, France).

10. **Toussaint EFA**, Morinière J, Müller CJ, Kunte K, Turlin B, Hausmann A, Balke M. Cryptic diversification of ornamental butterflies across tropical biodiversity hotspots. 7th International Conference on the Biology of Butterflies (11-14 August 2014, Turku, Finland). (Poster).

9. **Toussaint EFA**, Hall R, Monaghan M, Sagata K, Ibalim S, Shaverdo HV, Vogler AP, Pons J, Balke M. The towering orogeny of New Guinea as a trigger for arthropod megadiversity. 10th European Congress of Entomology (3-8 August 2014, York, England).

8. **Toussaint EFA**, Condamine FL, Hawlitschek O, Watts CHS, Porch N, Hendrich L, Balke M. Unveiling the diversification dynamics of Australasian predaceous diving beetles in the Cenozoic. 33rd Meeting of the Willi Hennig Society (6-10 July 2014, Trento, Italy).

7. **Toussaint EFA**, Condamine FL, Hawlitschek O, Watts CHS, Hendrich L, Balke M. Unveiling the diversification dynamics of Australasian predaceous diving beetles in the Cenozoic. International Biogeography Society Early Career Conference (7-10 January 2014, Canberra, Australia). *Recipient of a Travel Award.*

6. **Toussaint EFA**, Hall R, Monaghan M, Sagata K, Ibalim S, Shaverdo HV, Vogler AP, Pons J, Balke M. Deciphering the origins of massive radiations in the New Guinean archipelago. Entomology 2013 – The Entomological Society of America 61st Annual Meeting (10-13 November 2013, Austin, USA). *Invited talk - Recipient of a Travel Award.*

5. **Toussaint EFA**, Condamine FL, Hawlitschek O, Watts CHS, Hendrich L, Balke M. Evolutionary dynamics of beetles during the Australian drying-out. Entomology 2013 – The Entomological Society of America 61st Annual Meeting (10-13 November 2013, Austin, USA). *Invited talk - Recipient of a Travel Award.*

4. **Toussaint EFA**, Balke M. Massive orogeny as a biodiversity pump for the emerging Papuan giant. 32nd Meeting of the Willi Hennig Society (3-7 August 2013, Rostock, Germany). *Invited talk.*

3. **Toussaint EFA**, Shaverdo HV, Hendrich L, Balke M. Deciphering the origins of massive radiations in the Indo-Australian archipelago - The New Guinean archipelago as a major diversity pump in Australasia. 2nd Southeast Asian Gateway Evolution Meeting (11-17 March 2013, Berlin, Germany).

2. **Toussaint EFA**, Kergoat GJ, Capdevielle-Dulac C, Barbut J, Silvain J-F, Le Ru BP. Using a phylogenetic integrative approach to infer evolutionary patterns: Evolution, biodiversity and ecology of tropical African stem borer species of the subtribe Sesamiina (Lep. Noctuidae, Apameini). The 4th Printemps de Baillarguet (26-27 May 2011, Montpellier, France).

1. **Toussaint EFA**, Kergoat GJ, Capdevielle-Dulac C, Barbut J, Silvain J-F, Le Ru BP. Using a phylogenetic integrative approach to infer evolutionary patterns: Evolution, biodiversity and ecology of tropical African stem borer species of the subtribe Sesamiina (Lep. Noctuidae, Apameini). LEGSeriales 2011 (30-31 May 2011, Cernay-la-Ville, France).

Skills

Languages: French (Native speaker), English (Fluent), Spanish (Academic), German (Notions).

Fieldwork: Australia, Bali, Corsica, France, Germany, Italy, Java, Sicily, Slovenia, Sumatra, Tasmania.

Molecular Biology: Proficient in all protocols from sample to sequence data.

Bioinformatics:

- Sequence editing (Geneious)
- Phylogenetic inference (MrBayes, RAxML, TNT)
- Molecular species delimitation (BPP, GMYC/bGMYC, PTP/bPTP)
- Phylogeographic analyses (Arlequin, DnaSP, Network, SplitsTree)
- Divergence time estimation (BEAST)
- Historical Biogeography inference (BioGeoBEARS incl. BayArea, DEC, DIVA)
- Diversification rate analyses (*BAMM*, *diversitree*, *geiger*, *laser*, *TreePar* R packages)

Society memberships

International Biogeography Society
Society of Systematic Biologists
Willi Hennig Society

Referee Service

PLOS ONE
Molecular Phylogenetics and Evolution

Suggested references

Dr. Andrew EZ Short (Post-doctoral advisor), University of Kansas, Department of Ecology and Evolutionary Biology, Division of Entomology, Biodiversity Institute, 1501 Crestline Drive, Suite 140, Lawrence, KS 66045-2811, USA, Tel. +1 785 864 2323, E-Mail: aezshort@ku.edu

Dr. Michael Balke (Ph.D. advisor), Bavarian State Collection of Zoology (ZSM), 21 Münchhausenstraße, 81247 Munich, Germany, Tel. +49 (0)89 81 07 156, E-Mail: Coleoptera-ZSM@zsm.mwn.de

Dr. Gael J Kergoat (M.Sc. advisor), UMR 1062 CBGP-INR, Campus de Baillarguet, Av. Campus 34988 Montferrier-sur-Lèz, France, Tel. +33 (0)4 99 62 33 71, E-Mail: kergoat@supagro.inra.fr

Dr. Axel Hausmann (Collaborator), Bavarian State Collection of Zoology (ZSM), 21 Münchhausenstraße, 81247 Munich, Germany, Tel. +49 (0)89 81 07 149, E-Mail: Axel.Hausmann@zsm.mwn.de

Front page picture from left to right and top to bottom: Caterpillar of the genus *Polyura*, group of predaceous diving beetles of the genus *Eretes* devouring a fish, peacock swallowtail butterfly of the subgenus *Papilio* (*Achillides*) foraging, South-East New Guinean sunset in Oro Province.

INSECTS

AS A MODEL TO PUZZLE OUT MECHANISMS OF LINEAGE DIVERSIFICATION IN THE INDOMALAYAN / AUSTRALASIAN ARCHIPELAGO

Emmanuel F.A. Toussaint is an evolutionary biologist studying the origin and evolution of the stunning biodiversity living on Earth. Within this remarkable diversity of organisms, only a part has already been discovered. It is especially true for insects, in which an outstanding number of new species remains unknown. New advances in disciplines such as molecular phylogenetics, historical biogeography, ancestral character state reconstruction, ecological niche modeling or diversification analyses, coupled with the development of faster and cheaper ways to access increasingly larger amounts of DNA sequence data, allow to untangle the evolutionary processes that may have led to the setting up of this biodiversity. Interestingly, rock and amber fossils as well as paleoclimatic and paleogeological reconstructions help us to underpin which abiotic or biotic factors can explain those evolutionary histories and can therefore provide cardinal advances in our global understanding of the mechanisms governing biodiversity assemblages. In this dissertation are presented novel discoveries regarding the evolution of the Indomalayan / Australasian entomofauna, a poorly known biodiversity inhabiting one of the most threatened regions on Earth due to human activities.