Cumulative dissertation submitted in partial fulfillment of the requirements for the doctoral degree (Dr. rer. nat.)

Reconciling Molecular with Biological and Morphological Data Towards an Integrative Analysis of the Evolutionary Biology of Chafers

by

Jonas Eberle

from Giengen/Brz.

Bonn, April, 2017

Faculty of Mathematics and Natural Sciences, Rheinische Friedrich-Wilhelms-Universität, Bonn
prepared at the Zoological Research Museum A. Koenig, Bonn

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn.

Finanziert durch das DFG Projekt AH175/3.

Betreut von: Dr. Dirk Ahrens

Erstgutachter: Prof. Dr. Bernhard Misof

Zweitgutachter: Prof. Dr. Susanne Dobler

Fachnahes Kommissionsmitglied: Prof. Dr. Maximilian Weigend

Fachfremdes Kommissionsmitglied: PD Dr. Torsten Wappler

Tag der Promotion: 29. September 2017

Erscheinungsjahr: 2018

Disclaimer: The nomenclaturally relevant acts in this thesis have to be regarded as unpublished according to Article 8 of the *International Code of Zoological Nomenclature*, and will only become available by the referring publications.

The first part of knowledge is
getting the names right.

(Chinese proverb)

Summary

The research constituting this thesis acts at the interface of multiple disciplines related to biodiversity, combining data from morphology, genetics, geography, and ecology. Seven independent chapters (chapter II – chapter VIII) deal with diverse topics from a traditional taxonomic revision complemented by molecular phylogenetic analyses, over species delimitation and evolutionary studies of morphological diversification at different taxonomic scales, to population level landscape genetics integrating ecological niche modeling. Given the difficulties associated with a mega-diversity of taxonomically challenging species and yet rarely used high dimensional trait data in comparative analyses, innovative methods were employed and tested, often pushing the limits of currently available methods.

The subject of all studies were herbivore scarabs which represent one of the most diverse groups of living organisms. This property rendered them particularly suitable for the investigation of biodiversity-related questions. Three of the major known mechanisms for diversification of life were identified to have had potential impact: evolutionary key innovations, entry into new adaptive zones, and sexual selection. The former two of them were found in a single tribe of herbivore scarabs, the Sericini. The vast variability in their male genitalia, rendered a diversifying impact of sexual selection likely as well.

Accurate delimitation of species and handling of taxonomic implications provided a proper basis for evolutionary research. The thesis thus contributes to the understanding of the hyper-diversity of herbivore scarabs, which might be explained by a series of major diversifying events and mechanisms, and provides a first step to the conservation of this diversity in South Africa.

Chapter II Accurate delimitation of species is crucial for a stable taxonomy which is the foundation for evolutionary, ecological, and basically all biological studies. Several approaches towards an impartial and repeatable practice are available but yet have shortcomings. Problems may particularly arise in case of violation of underlying model assumptions, for instance in very recent speciations or cases of reduced gene flow. The latter is also observed in case of sex biased dispersal which is a common feature in many groups of organisms and may lead to over-estimations of actual species numbers. We evaluate the bias introduced by extreme female philopatry on a range of *de novo* (GMYC, PTP, ABGD, statistical parsimony, tr2) and validation (STACEY, iBPP) species delimitation approaches in the empirical test case of the scarab beetle genus *Pachypus*. Because female philopatry may particularly affect mitochondrial gene flow, we compared results from analyses of single loci, mitochondrial loci, nuclear loci, and combined data, as well as the performance of morphometric data as a second data source in a fully integrative Bayesian framework. Vast over-estimation of species numbers was observed in the analyses of combined and mitochondrial DNA data. The use of nuclear data resulted in more realistic estimations of species boundaries which were largely confirmed by morphometrics of linear measurements. Geometric morphometrics of body outlines resulted in stronger splitting. Our results suggest that nuclear DNA is better suited for species delimitation in many cases and reveal severe potential shortcomings of species delimitation solely relying on single mitochondrial loci. Particularly the integration of morphometric and molecular data yielded promising results. The impact on *cox1*-barcoding is briefly discussed.

Chapter III Defining species units can be challenging, especially during the earliest stages of speciation, when phylogenetic inference and delimitation methods may be compromised by incomplete lineage sorting (ILS) or secondary gene flow. Integrative approaches to taxonomy, which combine molecular and morphological evidence, have the potential to be valuable in such cases. In this study we investigated the South African scarab beetle genus Pleophylla using data collected from 110 individuals of eight putative morphospecies. The dataset included four molecular markers (*cox1*, *rrnL*, 28S, ITS1) and morphometric data based on male genital morphology. We applied a suite of molecular and morphological approaches to species delimitation,

and implemented a novel Bayesian approach in the software iBPP, which enables continuous morphological trait and molecular data to be combined.

Traditional morphology-based species assignments were supported quantitatively by morphometric analyses of the male genitalia (eigenshape analysis, CVA, LDA). While the ITS1-based delineation was also broadly congruent with the morphospecies, the *cox1* data resulted in over-splitting (GMYC modelling, haplotype networks, PTP, ABGD). In the most extreme case morphospecies shared identical haplotypes, which may be attributable to ILS based on statistical tests performed using the software JML. We found the strongest support for putative morphospecies based on phylogenetic evidence using the combined approach implemented in iBPP. However, support for putative species was sensitive to the use of alternative guide trees and alternative combinations of priors on the population size (θ) and root age (τ_0) parameters, especially when the analysis was based on molecular or morphological data alone.

We demonstrate that continuous morphological trait data can be extremely valuable in assessing competing hypotheses to species delimitation. In particular, we show that the inclusion of morphological data in an integrative Bayesian framework can improve the resolution of inferred species units. However, we also demonstrate that this approach is extremely sensitive to guide tree and prior parameter choice. These parameters should be chosen with caution – if possible – based on independent empirical evidence, or careful sensitivity analyses should be performed to assess the robustness of results. Young species provide exemplars for investigating the mechanisms of speciation and for assessing the performance of tools used to delimit species on the basis of molecular and/or morphological evidence.

Chapter IV The species of *Pleophylla* Erichson, 1847 are revised here and the phylogenetic relationships of all known species inferred based on external and genital morphology. For almost half of the species we were able to reconstruct gene trees for one nuclear marker (internal transcribed spacer 1) and one mitochondrial gene *(cytochrome oxidase subunit 1)*. Based on the morphology-based taxonomic revision of the type material, the following new synonymy is established: *Pleophylla navicularis* Burmeister, 1855 (= *P. flavicornis* Schaufuss, 1871, syn. nov.); *P. pilosa* Boheman, 1857 (= *P. opalina* Schaufuss, 1871, syn. nov.). Additionally, both *Pleophylla ferruginea* Burmeister, 1855 and *P. pilosa* Boheman, 1857 are removed from synonymy

with *Pleophylla fasciatipennis* Blanchard, 1850. Nineteen new species are described, mainly from South Africa (RSA): *Pleophylla burundiensis* sp. nov. (Burundi), *P. charlyi* sp. nov. (Tanzania), *P. congoensis* sp. nov. (Rwanda), *P. harrisoni* sp. nov. (RSA), *P. kruegeri* sp. nov. (RSA), *P. lizleri* sp. nov. (Tanzania), *P. mlilwaneensis* sp. nov. (Swaziland), *P. mpumalanga* sp. nov. (RSA), *P. murzini* sp. nov. (RSA), *P. nelshoogteensis* sp. nov. (RSA), *P. pseudopilosa* sp. nov. (RSA), *P. ruthae* sp. nov. (RSA), *P. settentrionalis* sp. nov. (Dem. Rep. of Congo), *P. silvatica* sp. nov. (RSA), *P. stalsi* sp. nov. (RSA), *P. taitaensis* sp. nov. (Kenya), *P. transkeiensis* sp. nov. (RSA), *P. wakkerstroomensis* sp. nov. (RSA), *P. warnockae* sp. nov. (RSA). The lectotypes of the following taxa are designated: *Pleophylla fasciatipennis* Blanchard, 1850, *P. ferruginea* Burmeister, 1855, *P. flavicornis* Schaufuss, 1871, *P. maculipennis* Boheman, 1857, *P. navicularis* Burmeister, 1855, *P. opalina* Schaufuss, 1871, *P. pilosa* Boheman, 1857, and *P. tongaatsana* Péringuey, 1904.

Chapter V Body shape reflects species' evolution and mediates its role in the environment as it integrates gene expression, life style, and structural morphology. Its comparative analysis may reveal insight on what shapes shape, being a useful approach when other evidence is lacking. Here we investigated evolutionary patterns of body shape in the highly diverse phytophagous chafers (Scarabaeidae: Pleurosticti), a polyphagous group utilizing different parts of angiosperms. Because the reasons of their successful diversification are largely unknown, we used a phylogenetic tree and multivariate analysis on twenty linear measurements of body morphology including all major Pleurosticti lineages to infer patterns of morphospace covariation and divergence. The chafer's different feeding types resulted to be not distinguishable in the described morphospace which was largely attributed to large occupancy of the morphospace of some feeding types and to multiple convergences of feeding behavior (particularly of anthophagy). Low correlation between molecular and morphological rates of evolution, including significant rate shifts for some lineages, indicated directed selection within feeding types. This is supported by morphospace divergence within feeding types and convergent evolution in Australian Melolonthinae. Traits driving morphospace divergence were extremities and traits linked with locomotion behavior, but also body size. Being highly adaptive for burrowing and locomotion, these traits showed major changes in the evolution of pleurostict scarabs. These activities also

affected another trait, the metacoxal length, which is highly influenced by key innovations of the metacoxa (extended mesal process, secondary closure) particularly in one lineage, the Sericini. Significant shape divergence between major lineages and a lack of strong differentiation among closely related lineages indicated that the question about the presence or absence of competition-derived directed selection needs to be addressed for different time scales. Striking divergence between some sister lineages at their origin revealed strong driven selection towards morphospace divergence, possibly linked with resource partitioning.

Chapter VI Left–right asymmetry is a frequently encountered phenomenon in the copulation organs of insects. While various causes have been proposed for genital asymmetry, we raise the question of whether asymmetry might facilitate, or even accelerate, morphological divergence of genitalia between species. We tested this hypothesis in the scarab chafer genus *Schizonycha*, which comprises species with symmetric as well as asymmetric male genitalia. Morphometric analyses were conducted in the context of their phylogeny, inferred from mitochondrial and nuclear ribosomal DNA sequence data (*cox1, rrnL*, and 28S) for a sample of 99 South African specimens, including 34 species and 5 outgroup taxa. Trees were reconstructed with maximum likelihood and Bayesian analysis. The extent of asymmetry and the variation of male copulation organs were analyzed with Generalized Procrustes analysis (GPA), by quantifying shape divergence of the parameres. We found a continuous transition in the degree of asymmetry among the investigated species. Ancestral state reconstruction revealed multiple origins and a high degree of evolutionary plasticity of paramere asymmetry in *Schizonycha*. However, no significant correlation between evolutionary rates of paramere shape divergence and the degree of paramere asymmetry was found, and so we conclude that asymmetric genitalia in *Schizonycha* do not increase the rate of genital shape divergence.

Chapter VII Megadiverse insect groups present special difficulties for biogeographers because poor classification, incomplete knowledge of taxonomy, and many undescribed species can introduce *a priori* sampling bias to any analysis. The historical biogeography of Sericini, a tribe of melolonthine scarabs comprising about 4,000 species, was investigated using the most comprehensive and time-calibrated molecular

phylogeny available today. Problems arising through nomenclatural confusion were overcome by extensive sampling (665 species) from all major lineages of the tribe. A West Gondwanan origin of Sericini (ca 112 Ma) was reconstructed using maximum parsimony, maximum-likelihood and model-based ancestral area estimation. Vicariance in the tribe's earliest history separated Neotropical and Old World Sericini, whereas subsequent lower Cretaceous biogeography of the tribe was characterized by repeated migrations out of Africa, resulting in the colonization of Eurasia and Madagascar. North America was colonized from Asia during the Cenozoic and a lineage of "Modern Sericini" reinvaded Africa. Diversification dynamics revealed three independent shifts to increased speciation rates: in African ant-adapted *Trochalus*, Oriental *Tetraserica*, and Asian and African Sericina. Southern Africa is proposed as both cradle and refuge of Sericini. This area has retained many old lineages that portray the evolution of the African Sericini fauna as a series of taxon pulses.

Chapter VIII Today, indigenous forests cover less than 0.6% of South Africa's land surface and are highly fragmented. Most forest relicts are very small and typically occur in fire-protected gorges along the eastern Great Escarpment. Yet, they hold a unique and valuable fauna with high endemism and ancient phylogenetic lineages, fostered by long term climatic stability and complex micro-climates. Despite numerous studies on southern African vegetation cover, the current state of knowledge about the natural extension of indigenous forests is rather fragmentary. We use an integrated approach of population-level phylogeography and climatic niche modeling of forest associated chafer species to assess connectivity and extent of forest habitats since the last glacial maximum. Current and past species distribution models ascertained potential fluctuations of forest distribution and supported a much wider potential current extension of forests based on climatic data. Considerable genetic admixture of mitochondrial and nuclear DNA among many populations and an increase of mean population mutation rate in Extended Bayesian Skyline Plots of all species indicated more extended or better connected forests in the recent past (< 5 kya). Genetic isolation of certain populations, as revealed by population differentiation statistics (G'_{ST}), as well as landscape connectivity statistics and models as well as scenarios of near future habitat conditions suggest considerable loss of habitat connectivity. Since major anthropogenic influence is likely, conservational actions need to be considered.

Contents

List of Figures

List of Tables

List of Electronic Supplements

Chapter I.

General Introduction

In times of rapidly declining species numbers (Butchart et al. 2010; Pereira et al. 2010; Newbold et al. 2016), one of the most stirring questions is about the drivers of biodiversity: which forces foster and sustain diversity of life? Since sustainable protection of biodiversity requires knowledge of the underlying forces (Cotterill 1995) that foster and maintain it, it is crucial to explore the underlying mechanisms that drive diversification of life and shape the global distribution patterns of organisms.

The biodiversity that we observe today is the result of long lasting processes of ancestral species' diversification and of extinctions. Highly diverse groups may thus result from exceptionally rapid diversifications or from generally low rates of extinction over evolutionary time, or both. With the rise of computer-based phylogenetic analyses and in particular with the utilization of genetic data for the quantification of relationships among species (Cavalli-Sforza and Edwards 1967), a new dimension of insights into major factors fostering diversification was possible. The knowledge of the phylogenetic relationships of biological species enables hypothesis testing of potentially diversifying traits. Such thoughts on comparative analyses among species were already mentioned by Lamarck (1809) and Darwin (1859) and dominate comparative methods in evolutionary biology until today (Garamszegi 2014). By the addition of chronological data to phylogenies, like it can be gained from the fossil record, more precise predictions about speciation rates and the timing of events may be inferred and correlated with historic geographical or environmental processes that potentially triggered diversification. Among the factors driving an extraordinary formation of species were identified evolutionary key innovations (Hunter 1998), sexual selection (Hosken and Stockley 2004), competition for resources (Hanski and Cambefort 1991; Giller and Doube 1994), and evolutionary arms races (Dawkins and Krebs 1979). More recently, exceptional diversity was also attributed to steady speciation along with low extinction rates and clade-specific innovations (Hunt et al. 2007; Condamine et al. 2016).

Diversification of biological species is linked to speciation. Understanding the process of diversification thus requires understanding the causes leading to the rise of a new species. For this reason, in addition to the employment of phylogenetic information, proper recognition of species entities is of major importance, because they are the basal entities for measuring biodiversity and any kind of diversifying forces. Delimiting species is, however, far from being trivial. A multitude of definitions for

species that were developed during the last decades (e.g., Simpson 1962; Valen 1976; Paterson 1985; Cracraft 1989; Templeton 1989; Mallet 1995; Mayr 1995) reflect the controversial discussion how to define a species. Also, the development of a wide variety of methods, aiming at more objective approaches to species delimitation than traditionally used, attributes to the fundamental complexity encountered in species recognition. This makes taxonomy, the science of delimiting, describing, and naming species, more and more a sophisticated science, being the base of all branches of biodiversity research.

Phytophagous insects constitute more than one quarter of all known animal species. They were proposed to have undergone huge radiations with the acquisition of phytophagy (Mitter et al. 1988). In particular beetles hold huge phytophagous lineages (Hunt et al. 2007; Beutel and Leschen 2016). It is likely that phytophagy is indeed correlated with this vast diversity (Farrell 1998), not only in beetles but in many phytophagous insect groups (Mitter et al. 1988). A popular explanation is specialization of the insect to specific plant species, potentially leading to diversification (Ehrlich and Raven 1964). Specifically, increased differentiation among populations by more patchily distributed host resources might cause increased speciation; however, increasing numbers of niches and the colonization of a new plant species by a specialized insect are more likely to be the "fuel in the engine of diversification" (Janz et al. 2006). In spite of that, alternative explanations are needed for species that feed polyphaguously on a variety of plants like it is the case for phytophagous scarab beetles of the family Scarabaeidae. For instance, *Popilia japonica* is known to feed on 435 plant species out of 95 families (Fleming 1972). The idea of an evolutionary arms race between the insect and its host plant is closely linked with an alternative hypothesis that might explain biological diversity: the entry of evolutionary lineages into new adaptive zones (ecological speciation) (Simpson 1953; Mitter et al. 1988). This might be the case when a species enters an island habitat with many ecological niches available, i.e., unoccupied or occupied by less competitive species. Many cases are described among which is also the famous example of Darwin's finches (Lowe 1936; Lack 1947). However, new adaptive zones might not only be spatially isolated areas but also ecologically isolated. Invasions into new environments like the air or utilization of a new food resource have similar effects. The rise of angiosperms and the accompanying effect on the soil strata is an instance for a huge food and brood

Figure I.1. A specimen of the forest-associated genus *Pleophylla* which is mainly distributed in South Africa but also found in the Eastern Arc mountain range. *Pleophylla* is the subject of three chapters of this thesis.

resource which became available for species capable of using it. Indeed, several phylogenetic lineages of herbivorous beetles are thought to have tremendously diversified with the rise of Angiosperms in the Lower Cretaceous (Mitter et al. 1988; Farrell 1998; Ahrens et al. 2014).

While phytophagy certainly facilitated some of the largest known radiations (Mitter et al. 1988; Farrell 1998; Janz et al. 2006), the unmatched absolute species richness of phytophagous beetle lineages surely also origins in key features that promoted the extraordinary diversity that is generally observed in beetles. Beetles comprise 300,000 – 450,000 described species, which is nearly one fifth of all known organisms (Grove and Stork 2000). The real number is estimated to be quadrupled in size, some authors even expect up to an order of magnitude more species than currently known (Grove and Stork 2000; Stork et al. 2015). Their diversity is often attributed to the elytra, the heavily sclerotized fore-wings of beetles that can – in most species – be locked

against each other and the beetles body to form a strong protective entity (Heberdey 1938). Elytra are supposed to be one of the most important key innovations that enabled beetles to invade the multitude of habitats they are found in (Lawrence and Newton, Jr. 1982; Beutel and Leschen 2016). Further proposed key innovations are likewise related to mechanical protection and the locomotor-system, enabling the penetration into crevices and hidden protective micro-habitats like under-bark (Beutel and Leschen 2016). This high adaptability to new (micro-) habitats and increased competitiveness is thus likely to also explain increased species numbers of phytophagous beetles compared to other, less diverse plant-feeding insects.

Vastly differing species numbers among phytophagous clades moreover suggest the operation of further diversifying mechanisms. A prominent example of a radiation that is likely independent from herbivory may for instance be seen in placental mammals, for which the development of the placenta, lactation and social bonds are seen as key innovations (Wilson et al. 1975; Heard and Hauser 1995). Also, the evolution of the oviposition rostrum in weevils (Curculionoidea) is seen as key innovation that boosted speciation (Oberprieler et al. 2007). The convergent occurrence of such a trait in multiple highly diverse groups may indicate its significance for diversification (Hunter 1998). Another factor continuously driving biodiversity also over long evolutionary time periods is sexual selection. Having the power to drive changes in mate recognition, it theoretically has enormous potential to foster speciation (Lande 1981; West-Eberhard 1983; Pomiankowski and Iwasa 1998; Panhuis et al. 2001). Although final evidence is lacking (Panhuis et al. 2001; Ritchie 2007), good indications are given by comparative analyses of sexual selection in speciose groups (Barraclough et al. 1995; Mendelson and Shaw 2005).

Phytophagy is considered to have developed about 10 times in beetles, including the case of "Phytophaga" (Farrell 1998; Marvaldi et al. 2002), one of the largest radiations on earth comprising leaf beetles, longhorns, and weevils (Chrysomelidae, Cerambycidae, and Curculionidae). Presumably, the vast availability of food and new brood resources triggered an excessive formation of species by reducing interspecific competition (Strong et al. 1984). Scarab beetles (Scarabaeidae), the second largest family of beetles (ca 27,000 species; Erichson 1847; Scholtz and Grebennikov 2005), include well known and charismatic beetle representatives like dung beetles (Scarabaeinae), cock chafers (Melolonthinae), and rose chafers (Cetoniinae). They belong to the su-

Figure I.2. *Schizonycha* comprises 370 described species with asymmetric and with symmetric male genitalia, making it highly suitable for the study of sexual selection. (Foto: D. Ahrens)

perfamily Scarabaeoidea (= Lamellicornia), placing them in near relationship with for instance stag beetles (Lucanidae), earth-boring dung beetles (Geotrupidae), and bess beetles (Passalidae). Approximately two third of scarab beetles are found in a monophyletic lineage of phytophagous species (Ritcher 1958; Balthasar 1963; Browne and Scholtz 1998; Scholtz and Grebennikov 2005). Herbivore scarabs utilize nearly all parts of angiosperm plants. They feed on leaves, flowers, and pollen as adults, and on living roots, soil humus or decaying wood in the larval stages (Scholtz and Chown 1995). In particular, the larvae are thus also encountered as crop pests (e.g. Jackson 2006).

The present studies focus on three groups within phytophagous scarab beetles, all providing interesting case studies related to biodiversity research. With ca 4,000

species, the tribe of Sericini (Fig. I.1) places among the largest groups of herbivore scarabs and is subject to most of the studies comprising this thesis. Despite their species richness and near global distribution (except Australia and circumpolar regions), their external morphology is of remarkably uniform appearance and is thus posing special difficulties for taxonomists. Consequently, historical systematic classifications are error-prone: on the one hand, 60% of Sericini genera are monotypic (Ahrens 2007a,b; Ahrens, unpublished data), on the other hand, huge paraphyletic collective groups have formed (Ahrens and Vogler 2008; Liu et al. 2015). The majority of species is found on the Asian subcontinent, exhibiting high levels of endemism (Ahrens 2004; Liu et al. 2015). However, the geographical origin of Sericini is thought to be in Africa, where its sister tribe, Ablaberini, exclusively occurs (Ahrens 2006).

The second group exhibiting characteristics that potentially drives increased speciation is the genus *Schizonycha* (Fig. I.2). It comprises 370 species with most of them (349) occurring in the Afro-tropical region (Pope 1960; Lacroix 2010). Like Sericini, they feed polyphagously on plants, while their larvae are soil dwellers and feed on humus or plant roots. Within *Schizonycha*, asymmetric and symmetric genitalia are found, which is most likely attributable to sexual selection (Simmons 2014). Providing a huge amount of well comparable species, the genus is predestined for the investigation of sexual selection as diversifying force.

A highly challenging case for species delimitation is found in the ancient chafer tribe Pachypodini which is exclusively distributed in the western Mediterranean (Sparacio 2008). The only genus within the tribe, *Pachypus* (Fig. I.3), is composed of five described species (Baraud 1985; Baraud 1992; Sparacio 2008; Guerlach et al. 2013). However, its homogeneous external morphology and barely distinctive genital characters pose difficulties to taxonomists, which is also reflected in a large number of synonymies (Olivier 1789; Fabricius 1792; Erichson 1840; Reitter 1898; Luigioni 1923; source: Schoolmeesters 2016) associated so far with the species *P. candidae*. Moreover, due to the complete lack of female elytra and hindwings and their almost entirely hypogean life style (Crovetti 1969; Arnone and Sparacio 1990), exaggerated mitochondrial genetic structure is to expect so that basic assumptions of available species delimitation methods might be violated. *Pachypus* thus represents an exemplar case study for the utilization of newly developed and powerful integrative species delimitation methods.

Figure I.3. Left: Female *Pachypus candidae*; right: male *Pachypus melonii*. Females of the genus *Pachypus* completely lack fore- and hindwings. Increased genetic divergence of mitochondrial DNA and the uniform external morphology among species makes the genus exceptionally challenging for species delimitation. (Foto: E. Bazzato)

The enormous diversity of herbivore scarabs and in particular of Sericini makes them an ideal group for the studies of systematics, taxonomy, biogeography, macroecology, and species formation that are covered in this thesis. The present study contributes to a better understanding of scarab evolutionary biology on several levels, from the delimitation of species in challenging cases, the description of new species as entities for subsequent studies, exploration of major diversifying factors in highly speciose groups of herbivore scarabs, to the conservation-relevant study of habitat fragmentation in a genus of South African chafers. Thereby, a major focus of this work lies in the combination and integration of multiple data sources, applying the latest methods developed in integrative analyses and phylogenetic comparative methods.

The practice of comparing and integrating data from multiple species led to major developments in evolutionary biology, triggered by the recognition of the fact that data from biological species are not statistically independent (Harvey and Pagel 1991). Statistical inferences on data of more than one species that is derived from morphology, ecology, ontology, behavior, or any other source, must therefore generally account for the phylogenetic relationship of the species. Otherwise, basic assumptions of most methods are violated and may lead to false conclusions: related species are more similar to each other due to their shared history and thus the phylogenetic and the environmental component of a trait should be separated (Lynch 1991). The ability to

quantify evolutionary distances between species by genetic data facilitated the development of phylogenetic comparative methods. Phylogenetic independent contrasts (PIC; Felsenstein 1985) and phylogenetic generalized least squares (PGLS; Grafen 1989) are presently the most commonly used methods. In combination with further refinements of trait evolution models (e.g., Hansen 1997), particularly PGLS provides a sophisticated framework to infer evidence of evolutionary association among traits, ancestral state estimation, and the assessment of mode and directionality of evolution (Garamszegi 2014).

Species delimitation is an example of a discipline that particularly benefits from integration of data. Being traditionally based on morphological examination of specimens, new approaches like the utilization of DNA-barcodes caused discussion (Wheeler 2005; Will et al. 2005). Integrative approaches intend to reconcile the different sources of data. Applying the principle of reciprocal illumination (Hennig 1966), strong evidence might be reached by consilience in Whewell's sense (Whewell 1847; Santos and Capellari 2009). However, disagreement of conclusions from different data is likely to occur and must be handled. Taxonomists may decide to delimit two species if there is evidence from at least one source of data (e.g., DNA or morphology) or only, if there is evidence from more than one or all data (e.g., DNA and morphology). Both approaches, also known as integration by cumulation and integration by congruence (Padial et al. 2010) are prone to errors, either by over- or under-estimation of diversity (Padial et al. 2010). One solution might be to solve disagreement of data by invoking evolutionary explanations for the disagreement (Schlick-Steiner et al. 2010). Some of the latest developments in the field of species delimitation fully integrate molecular with morphological information (Solís-Lemus et al. 2015). They combine the data in a Bayesian probabilistic framework and thus overcome the need for the decision between a cumulation and a congruence approach. Regarding the set of different approaches to species delimitation, the question on the underlying species concept is crucial but not as decisive as thought in the past. In this regard, it should be differentiated between species concepts that aim at defining the particular mechanism that led to isolation and species delimitation that aims to detect the ultimate outcome of speciation – genetic isolation on an evolutionary timescale (Queiroz 2005; De Queiroz 2007; Rannala 2015). This general species concept reconciles the multitude of species concepts by pointing out the central property shared by all species concepts of species

being independently evolving metapopulations (De Queiroz 2007). Such metapopulations might then be best defined by features that have negative fitness effects on other groups of individuals (Hausdorf and Hennig 2010).

However, discordances in data of different origin might be enlightening as well. So can directed evolutionary selection on a trait be inferred by the comparison of its morphological development with a neutrally evolving molecular locus (Polly 2004; Ricklefs 2006; Ahrens and Ribera 2009). Doing so might reveal cases of sexual selection or the entrance into a new adaptive zone, where functional traits are adapted to a new environment. Likewise, the evolutionary plasticity of a trait and its independent development in multiple groups might be inferred by its projection on a molecular phylogenetic hypothesis, and thus indicate its potential biological relevance. Such projections of the results from one type of data to another, being discordant or not, are a powerful tool to infer various questions. For instance, through the link by distribution data, species delimitations may be used to derive ecological niches and habitat availability. Mapping spatial data on molecular relationships might give insights into habitat connectivity and gene-flow, hence contributing to conservation-relevant problems.

The data for integrative analyses of biodiversity can be of various origins, including DNA, morphology, ecology, distribution, ontology, and behavior. Commonly used and particularly promising are molecular and morphological data. Geometric morphometrics provide impartial and quantifiable morphological trait data and are thus particularly suitable for the use in evolutionary biology and species delimitation. Shape variation of biological specimens is mathematically captured at homologizable points or from outline curves. The most common analytical methods are Generalized Procrustes analysis (Gower 1975; Rohlf and Slice 1990) and Eigenshape analysis (MacLeod and Rose 1993; MacLeod 1999), both providing quantitative measures of specimen divergence in a multidimensional morphospace. Morphometrics thus enable evolutionary biologists to investigate whole trait complexes or traits that are otherwise hard to measure. The challenge lies in the high dimensionality of the resulting data that pushes many available methods to their limits. The present work utilized such data for the first time for truly integrative species delimitation (Solís-Lemus et al. 2015), to infer the effects of phytophagy in chafers on their phenotype, and to

evaluate the effects of asymmetric genitalia to the diversification of a hyper-diverse genus of chafers.

Research questions

Basically, all studies related to biodiversity rely on an accurate delimitation of species which is also crucial for a stable taxonomy. One of the oldest and most controversial biological questions might thus be about the definition of a species and its proper delimitation (Darwin 1859). Chapter II to IV address species delimitation and taxonomy of two scarab genera, *Pachypus*(Fig. I.3) and *Pleophylla* (Fig. I.1), bringing together molecular methods and collection based research. The latest available methods for species delimitation are compared to other frequently used approaches, and their performance and pitfalls are thoroughly tested in extremely challenging empirical test cases. Such a new method is integrated Bayesian Phylogenetics and Phylogeography (iBPP; Solís-Lemus et al. 2015) which is not much tested yet but a very promising step towards the future of fully integrative species delimitation (only 16 citations, 6 of which actually applying it to species delimitation; `www.scholar.google.com`, accessed on August 13, 2016). Chapter IV implements these results (Chapter III) in a taxonomic revision of the South African chafer genus *Pleophylla* and provides spatial distribution data of the species.

After the thorough investigation of species limits in *Pachypus* and *Pleophylla*, chapter V and VI pose questions about the macroecology, evolution, and diversification of herbivore scarab beetles. Chapter V investigates the response of herbivore scarab body shape to the utilization of angiosperms at the subfamily level. Features of scarab morphology that might foster diversification are examined in a phylogenetic comparative framework. The findings are compared to the established model system of dung beetles which are known to have greatly diverged morphologically due to strong competition. The process of diversification by entering the new adaptive zone given through the rise of angiosperms is discussed in the light of potentially lacking competition among species. Using similar methods, chapter VI deals with potential diversification driven by sexual selection: the influence of asymmetry on male genital divergence is evaluated in the scarab genus *Schizonycha*. The evolutionary plasticity

of male genital asymmetry is investigated in the background of the first molecular phylogeny of the genus.

While comparative analyses of body-shape and genital morphology might give insights into evolutionary key innovations and sexual selection as diversifying factors of diversification in phytophagous scarabs, chapter VII attempts to reconstruct the historical biogeographic processes leading to the present-day distribution of Sericini chafers in the background of the most comprehensive phylogenetic hypothesis of the tribe based on molecular data. The integration of the results from molecular analyses with spatial distribution data aims to clarify of the origin of the tribe, major global scale migrations, and the timing of major diversifications leading to the mega-diversity observed today. The identification of ancient and evolutionary unique Sericini lineages facilitates the recognition of high priority conservation areas.

Pleophylla represents such an evolutionary unique lineage which predominantly occurs in South Africa. Building on the updated taxonomy of *Pleophylla* (chapter IV) chapter VIII examines the status of forest habitat connectivity in South Africa by the investigation of the species' climatic niche models and genetic interchange at the population level. Based on these current and past species distribution models of the apparently strictly forest associated genus it is investigated whether the current patchy distribution of forests in South Africa is the natural state or shaped by anthropogenic influence. With this final chapter of the thesis, species delimitation and museum collection based taxonomy is tied to the application of population level landscape genetics for actual use in conservation management with concrete conservation-related actions.

The research constituting this thesis acts at the interface of multiple disciplines, combining data from morphology, genetics, geography, and ecology. Seven independent chapters deal with diverse topics from a traditional taxonomic revision complemented by molecular phylogenetic analyses, over species delimitation and evolutionary studies of morphological diversification at different taxonomic scales, to population level landscape genetics integrating ecological niche modeling. Given the difficulties associated with a mega-diversity of taxonomically challenging species and yet rarely used high dimensional trait data in comparative analyses, innovative methods were employed, often pushing and extending the limits of currently available methods.

References

Ahrens, D. (2004). *Monographie der Sericini des Himalaya (Coleoptera, Scarabaeidae)*. Dissertation.de - Verlag im Internet GmbH, Berlin, p. 534.

— (2006). The phylogeny of Sericini and their position within the Scarabaeidae based on morphological characters (Coleoptera: Scarabaeidae). *Systematic Entomology* 31, 113–144. DOI: 10.1111/j.1365-3113.2005.00307.x.

— (2007a). Taxonomic changes and an updated catalogue for the Palaearctic Sericini (Coleoptera: Scarabaeidae: Melolonthinae). *Zootaxa* 1504, 1–51.

— (2007b). Type species designations of Afrotropical Ablaberini and Sericini genera (Coleoptera: Scarabaeidae: Melolonthinae). *Zootaxa* 1496, 53–62.

Ahrens, D. and I. Ribera (2009). Inferring speciation modes in a clade of Iberian chafers from rates of morphological evolution in different character systems. *BMC evolutionary biology* 9, 234. DOI: 10.1186/1471-2148-9-234.

Ahrens, D., J. Schwarzer, and A. P. Vogler (2014). The evolution of scarab beetles tracks the sequential rise of angiosperms and mammals. *Proceedings of the Royal Society B: Biological Sciences* 281, 20141470–20141470. DOI: 10.1098/rspb.2014.1470.

Ahrens, D. and A. P. Vogler (2008). Towards the phylogeny of chafers (Sericini): analysis of alignment-variable sequences and the evolution of segment numbers in the antennal club. *Molecular phylogenetics and evolution* 47, 783–98. DOI: 10.1016/j.ympev.2008.02.010.

Arnone, M. and I. Sparacio (1990). Il *Pachypus caesus* Erichson 1840: brevi note sulla biologia e la distribuzione in Sicilia (Coleoptera: Scarabaeoidea). *Naturalista Siciliano* 14, 63–71.

Balthasar, V. (1963). *Monographie der Scarabaeidae und Aphodiidae der palaearktischen und orientalischen Region. Coleoptera: Lamellicornia*. Band 1. Prag: Verlag der Tschechoslowakischen Akademie der Wissenschaften, p. 391.

Baraud, J. (1985). *Tropinota (Epicometis) villiersi* nouvelle espèce du Moyen-orient (Coleoptera, Scarabaeoidea, Cetoniidae). *Revue française d'entomologie* 6, 61–63.

Baraud, J. (1992). *Coléoptères Scarabaeoidea d'Europe – Faune de France*. Paris ;Lyon: Fédération française des Sociétés de Sciences naturelles, Paris, Lyon, p. 856.

Barraclough, T., P. Harvey, and S. Nee (1995). Sexual selection and taxonomic diversity in passerine birds. *Proceedings: Biological Sciences* 259, 211–215.

Beutel, R. G. and R. A. Leschen, eds. (2016). *Handbook of Zoology. Coleoptera, Beetles. Volume 1: Morphology and Systematics (Archostemata, Adephaga, Myxophaga, Polyphaga partim) 2nd edition*. Walter de Gruyter, Berlin/Boston, p. 701.

Browne, J. and C. Scholtz (1998). Evolution of the scarab hind wing articulation and wing base: a contribution toward the phylogeny of the Scarabaeidae (Scarabaeoidea: Coleoptera). *Systematic Entomology* 23, 307–326.

Butchart, S. H. M., M. Walpole, B. Collen, A. van Strien, J. P. W. Scharlemann, R. E. A. Almond, J. E. M. Baillie, B. Bomhard, C. Brown, J. Bruno, K. E. Carpenter, G. M. Carr, J. Chanson, A. M. Chenery, J. Csirke, N. C. Davidson, F. Dentener, M. Foster,

A. Galli, J. N. Galloway, P. Genovesi, R. D. Gregory, M. Hockings, V. Kapos, J.-F. Lamarque, F. Leverington, J. Loh, M. A. McGeoch, L. McRae, A. Minasyan, M. H. Morcillo, T. E. E. Oldfield, D. Pauly, S. Quader, C. Revenga, J. R. Sauer, B. Skolnik, D. Spear, D. Stanwell-Smith, S. N. Stuart, A. Symes, M. Tierney, T. D. Tyrrell, J.-C. Vie, and R. Watson (2010). Global Biodiversity: Indicators of Recent Declines. *Science* 328, 1164–1168. DOI: 10.1126/science.1187512.

Cavalli-Sforza, L. L. and A. W. F. Edwards (1967). Phylogenetic Analysis Models and Estimation Procedures. *American journal of human genetics* 19, 233–257. DOI: 10.2307/2406616.

Condamine, F. L., M. E. Clapham, and G. J. Kergoat (2016). Global patterns of insect diversification: towards a reconciliation of fossil and molecular evidence? *Scientific Reports* 6, 19208. DOI: 10.1038/srep19208.

Cotterill, F. P. D. (1995). Systematics, biological knowledge and environmental conservation. *Biodiversity and Conservation* 4, 183–205. DOI: 10.1007/BF00137784.

Cracraft, J. (1989). "Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation." In: *Speciation and its consequences*. Ed. by D. Otte and J. A. Endler. Sunderland, MA: Sinauer Associates, Sunderland, MA, pp. 28–59.

Crovetti, A. (1969). Contributo alla conoscenza dei Coleotteri Scarabeidi I. Il genere *Pachypus* ServiIle (Coleoptera, Scarabaeidae, Pachypodinae). *Bollettino di Zoologia Agraria e di Bachicoltura* 9, 133–188.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection*. London: John Murray, London, p. 502. DOI: 10.1016/S0262-4079(09)60380-8.

Dawkins, R. and J. R. Krebs (1979). Arms races between and within species. *Proceedings of the Royal Society of London. Series B* 205, 489–511. DOI: 10.1098/rspb.1979.0081.

De Queiroz, K. (2007). Species Concepts and Species Delimitation. *Systematic Biology* 56, 879–886. DOI: 10.1080/10635150701701083.

Ehrlich, P. and P. Raven (1964). Butterflies and plants: A study in coevolution. *Evolution* 18, 586–608.

Erichson, W. F. (1840). Die Pachypoden, eine kleine Gruppe aus der Familie der Melolonthen. Entomographien, Untersuchungen in dem Gebiete der Entomologie mit besonderer Benutzung der Königl. Sammlung in Berlin. *Morin* 1, 29–43.

Erichson, W. (1847). *Naturgeschichte der Insecten Deutschlands. Erste Abtheilung. Coleoptera*. 3, Lfrg. 4. Berlin: Nicolaische Buchhandlung, Berlin, pp. 552–986.

Fabricius, J. C. (1792). *Entomologia systematica emendata et aucta. Secundum Classes, Ordines, Genera, Species adjectis Synonimis, Locis, Observationibus, Descriptionibus*. Hafniae, p. 538.

Farrell, B. (1998). "Inordinate Fondness" explained: Why are there so many beetles? *Science* 281, 555–559.

Felsenstein, J. (1985). Phylogenies and the comparative method. *The American Naturalist* 125, 1–15.

Fleming, W. E. (1972). Biology of the japanese beetle. *Technichal Bulletin* 1449.

Garamszegi, L. Z. (2014). Modern phylogenetic comparative methods and their application in evolutionary biology. *Concepts and Practice. London, UK: Springer*.

Giller, P. S. and B. M. Doube (1994). Spatial and Temporal Co-Occurrence of Competitors in Southern African Dung Beetle Communities. *Journal of Animal Ecology* 63, 629–643.

Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika* 40, 33–51. DOI: 10.1007/BF02291478.

Grafen, A. (1989). The Phylogenetic Regression. *Philosophical Transactions of the Royal Society B: Biological Sciences* 326, 119–157. DOI: 10.1098/rstb.1989.0106.

Grove, S. J. and N. E. Stork (2000). An inordinate fondness for beetles. *Invertebrate Systematics* 14, 733–739. DOI: 10.1071/IT00023.

Guerlach, G., E. Bazzato, and D. Cillo (2013). Description d'une nouvelle espèce de *Pachypus* Dejean, 1821: *Pachypus sardiniensis* n.sp. *Lambillionea* 113, 73–76.

Hansen, T. F. (1997). Stabilizing Selection and the Comparative Analysis of Adaptation. *Evolution* 51, 1341–1351. DOI: 10.2307/2411186.

Hanski, I. and Y. Cambefort (1991). *Dung beetle ecology*. Princeton University Press.

Harvey, P. and M. Pagel (1991). *The comparative method in evolutionary biology*. Vol. 239. Oxford university press Oxford.

Hausdorf, B. and C. Hennig (2010). Species Delimitation Using Dominant and Codominant Multilocus Markers. *Systematic Biology* 59, 491–503. DOI: 10.1093/sysbio/syq039.

Heard, S. B. and D. L. Hauser (1995). Key innovations and their ecological mechanisms. *Historical Biology* 10, 151–173.

Heberdey, R. F. (1938). Beiträge zum Bau des Subelytralraumes und zur Atmung der Coleopteren. *Zeitschrift für Morphologie und Ökologie der Tiere* 33, 667–734.

Hennig, W. (1966). *Phylogenetic Systematics*. Urbana, Chicago, London: University of Illinois Press, Urbana, Chicago, London, p. 263.

Hosken, D. J. and P. Stockley (2004). Sexual selection and genital evolution. *Trends in ecology & evolution* 19, 87–93. DOI: 10.1016/j.tree.2003.11.012.

Hunt, T., J. Bergsten, Z. Levkanicova, A. Papadopoulou, O. S. John, R. Wild, P. M. Hammond, D. Ahrens, M. Balke, M. S. Caterino, J. Gomez-Zurita, I. Ribera, T. G. Barraclough, M. Bocakova, L. Bocak, A. P. Vogler, J. Gómez-Zurita, I. Ribera, T. G. Barraclough, M. Bocakova, L. Bocak, and A. P. Vogler (2007). A Comprehensive Phylogeny of Beetles Reveals the Evolutionary Origins of a Superradiation. *Science* 318, 1913–1916. DOI: 10.1126/science.1146954.

Hunter, J. P. (1998). Key innovations and the ecology of macroevolution. *Trends in ecology & evolution* 13, 31–6.

Jackson, T. A. (2006). Scarabs as Pests: a continuing problem. *Coleopterists Society Monograph* 5, 102–119.

Janz, N., S. Nylin, and N. Wahlberg (2006). Diversity begets diversity: host expansions and the diversification of plant-feeding insects. *BMC Evolutionary Biology* 6, 4. DOI: 10.1186/1471-2148-6-4.

Lack, D. (1947). *Darwin's finches*. CUP Archive, p. 208.

Lacroix, M. (2010). *Melolonthinae afrotropicaux (Scarabaeoidea, Melolonthidae): genera et catalogue commenté*. Paris: Lacroix, p. 277.

Lamarck, J. B. (1809). *Philosophie zoologique*. 1994 editi. Paris: Flammarion.

Lande, R. (1981). Models of speciation by sexual selection on polygenic traits. *Proceedings of the National Academy of Sciences of the United States of America* 78, 3721–3725. DOI: 10.1073/pnas.78.6.3721.

Lawrence, J. F. and A. F. Newton, Jr. (1982). Evolution and Classification of Beetles. *Annual Review of Ecology and Systematics* 13, 261–290.

Liu, W.-G., J. Eberle, M. Bai, X.-K. Yang, and D. Ahrens (2015). A phylogeny of Sericini with particular reference to Chinese species using mitochondrial and ribosomal DNA (Coleoptera: Scarabaeidae). *Organisms Diversity & Evolution* 15, 343–350. DOI: 10.1007/s13127-015-0204-z.

Lowe, P. R. (1936). The Finches of the Galapagos in relation to Darwin's Conception of Species. *Ibis* 78, 310–321.

Luigioni, P. (1923). Le specie et le varieta del gen. *Pachypus* Serv. in Italia. *Memorie della Società Entomologica Italiana* 2, 50–64.

Lynch, M. (1991). Methods for the analysis of comparative data in evolutionary biology. *Evolution* 45, 1065–1080.

MacLeod, N. and K. D. Rose (1993). Inferring locomotor behavior in Paleogene mammals via eigenshape analysis. *American Journal of Science* 293, 300–355. DOI: 10.2475/ajs.293.A.300.

MacLeod, N. (1999). Generalizing and extending the eigenshape method of shape space visualization and analysis. *Paleobiology* 25, 107–138.

Mallet, J. (1995). A species definition for the modern synthesis. *Trends in Ecology & Evolution* 10, 294–299.

Marvaldi, A. E., A. S. Sequeira, C. W. O'Brien, and B. D. Farrell (2002). Molecular and Morphological Phylogenetics of Weevils (Coleoptera , Curculionoidea): Do Niche Shifts Accompany Diversification? *Systematic biology* 51, 761–785. DOI: 10.1080/10635150290102465.

Mayr, E. (1995). "Species, classification, and evolution." In: *Biodiversity and Evolution*. Ed. by R. Arai, M. Kato, and Y. Doi. Tokyo: National Science Museum Foundation, Tokyo, pp. 3–12.

Mendelson, T. C. and K. L. Shaw (2005). Rapid speciation in an arthropod. *Nature* 433, 375–376.

Mitter, C., B. Farrell, and B. Wiegmann (1988). The phylogenetic study of adaptive zones: has phytophagy promoted insect diversification? *American Naturalist* 132, 107–128.

Newbold, T., L. N. Hudson, A. P. Arnell, S. Contu, A. De Palma, S. Ferrier, S. L. L. Hill, A. J. Hoskins, I. Lysenko, H. R. P. Phillips, V. J. Burton, C. W. T. Chng, S. Emerson, D. Gao, G. Pask-Hale, J. Hutton, M. Jung, K. Sanchez-Ortiz, B. I. Simmons, S. Whitmee, H. Zhang, J. P. W. Scharlemann, and A. Purvis (2016). Has land use pushed terrestrial biodiversity beyond the planetary boundary? A global assessment. *Science* 353, 288–291. DOI: 10.1126/science.aaf2201.

Oberprieler, R. G., A. E. Marvaldi, and R. S. Anderson (2007). Weevils, weevils, weevils everywhere. *Zootaxa* 1668, 491–520.

Olivier, G. A. (1789). *Entomologie ou Histoire naturelle des insectes, avec leurs caractères génériques et spécifiques, leur description, leur synonumie et leur figure enluminee ... Coléoptères Tomes I-V*. Vol. 1. Paris :Imp. Baudouin, p. 488.

Padial, J. M., A. Miralles, I. De la Riva, and M. Vences (2010). The integrative future of taxonomy. *Frontiers in zoology* 7, 16. DOI: 10.1186/1742-9994-7-16.

Panhuis, T. M., R. Butlin, M. Zuk, and T. Tregenza (2001). Sexual selection and speciation. *Trends in ecology & evolution* 16, 364–371. DOI: 10.1146/annurev.ecolsys.38.091206.095733.

Paterson, H. E. H. (1985). "The recognition concept of species." In: *Species and Speciation*. Ed. by E. S. Vrba. Pretoria: Transvaal Museum Monograph No. 4, Pretoria, pp. 21–29.

Pereira, H. M., P. W. Leadley, V. Proença, R. Alkemade, J. P. W. Scharlemann, J. F. Fernandez-Manjarrés, M. B. Araújo, P. Balvanera, R. Biggs, W. W. L. Cheung, L. Chini, H. D. Cooper, E. L. Gilman, S. Guénette, G. C. Hurtt, H. P. Huntington, G. M. Mace, T. Oberdorff, C. Revenga, P. Rodrigues, R. J. Scholes, U. R. Sumaila, and M. Walpole (2010). Scenarios for global biodiversity in the 21st century. *Science* 330, 1496–1501. DOI: 10.1126/science.1196624.

Polly, P. D. (2004). On the simulation of the evolution of morphological shape: multivariate shape under selection and drift. *Palaeontologia Electronica* 7, 1–28.

Pomiankowski, A. and Y. Iwasa (1998). Runaway ornament diversity caused by Fisherian sexual selection. *Proceedings of the National Academy of Sciences of the United States of America* 95, 5106–5111. DOI: 10.1073/pnas.95.9.5106.

Pope, R. (1960). A revision of the species of *Schizonycha* Dejean (Col.: Melolonthidae) from southern Africa. *Bulletin of the Natural History Museum* 9, 63–218.

Queiroz, K. de (2005). Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences* 102, 6600–6607. DOI: 10.1073/pnas.0502030102.

Rannala, B. (2015). The art and science of species delimitation. *Current Zoology* 61, 846–853. DOI: 10.1093/czoolo/61.5.846.

Reitter, E. (1898). Bestimmungs-Tabelle der Melolonthidae aus der europäischen Fauna und den angrenzenden Ländern. II Theil. Dynastini, Euchirini, Pachypodini, Cetonini, Valgini und Trichiini. *Verhandlungen des Naturforschenden Vereins, Brünn* 37, 21–111.

Ricklefs, R. E. (2006). Time, species, and the generation of trait variance in clades. *Systematic biology* 55, 151–159. DOI: 10.1080/10635150500431205.

Ritcher, P. O. (1958). Biology of Scarabaeidae. *Annual review of Entomology* 3, 311–334.

Ritchie, M. G. (2007). Sexual Selection and Speciation. *Annual Review of Ecology, Evolution, and Systematics* 38, 79–102. DOI: 10.1146/annurev.ecolsys.38.091206.095733.

Rohlf, F. J. and D. Slice (1990). Extensions of the Procrustes Method for the Optimal Superimposition of Landmarks. *Systematic Zoology* 39, 40–59. DOI: 10.2307/2992207.

Santos, C. M. D. and R. S. Capellari (2009). On reciprocal illumination and consilience in biogeography. *Evolutionary Biology* 36, 407–415. DOI: 10.1007/s11692-009-9070-y.

Schlick-Steiner, B. C., F. M. Steiner, B. Seifert, C. Stauffer, E. Christian, and R. H. Crozier (2010). Integrative Taxonomy: A Multisource Approach to Exploring Biodiversity. *Annual Review of Entomology* 55, 421–438. DOI: 10.1146/annurev-ento-112408-085432.

Scholtz, C. and S. Chown (1995). "The evolution of habitat use and diet in the Scarabaeoidea: a phylogenetic approach." In: *Biology, Phylogeny, and Classification of Coleoptera: Papers Celebrating the 80th Birthday of Roy A. Crowson.* Ed. by J. Pakaluk and S. A. Ślipiński. 1st ed. Warszawa: Muzeum i Instytut Zoologii PAN Warszawa, pp. 355–374.

Scholtz, C. H. and V. V. Grebennikov (2005). "Scarabaeoidea Latreille, 1802." In: *Coleoptera, beetles: Morphology and systematics (Archostemata, Adephaga, Myxophaga, Polyphaga partim), Band 1.* Ed. by R. Beutel and R. A. B. Leschen. Walter de Gruyter, p. 567.

Schoolmeesters, P. (2016). "Scarabs: World Scarabaeidae Database." In: *Species 2000 & ITIS Catalogue of Life.* Ed. by Y. Roskov, L. Abucay, l. T. Orrel, D. Nicolson, T. Kunze, C. Flann, N. Bailly, P. Kirk, T. Bourgoin, R. E. DeWalt, W. Decock, and A. De Wever. 28th July. Species 2000: Naturalis, Leiden, the Netherlands.

Simmons, L. W. (2014). Sexual selection and genital evolution. *Austral Entomology* 53, 1–17. DOI: 10.1111/aen.12053.

Simpson, G. G. (1962). *Principles of Animal Taxonomy*. New York: Columbia University Press, New York, p. 247.

Simpson, G. G. (1953). *The major features of evolution*. 1st ed. Columbia University Press, p. 434.

Solís-Lemus, C., L. L. Knowles, and C. Ané (2015). Bayesian species delimitation combining multiple genes and traits in a unified framework. *Evolution* 69, 492–507. DOI: 10.1111/evo.12582.14.

Sparacio, I. (2008). Una nuova specie di *Pachypus* Dejean di Sardegna. *Doriana* 8, 1–13.

Stork, N. E., J. McBroom, C. Gely, and A. J. Hamilton (2015). New approaches narrow global species estimates for beetles, insects, and terrestrial arthropods. *Proceedings of the National Academy of Sciences* 112, 7519–7523. DOI: 10.1073/pnas.1502408112.

Strong, D., J. Lawton, and R. Southwood (1984). *Insects on Plants: Community Patterns and Mechanisms*. Harvard University Press, p. 313.

Templeton, A. R. (1989). "The meaning of species and speciation: a genetic perspective." In: *Speciation and its consequences*. Ed. by D. Otte and J. A. Endler. Sunderland, MA: Sinauer Associates, Sunderland, MA, pp. 3–27.

Valen, L. van (1976). Ecological species, multispecies, and oaks. *Taxon* 25, 233–239. DOI: 10.2307/1219444.

West-Eberhard, M. J. (1983). Sexual Selection, Social Competition, and Speciation. *The Quarterly Review of Biology* 58, 155–183.

Wheeler, Q. (2005). Losing the plot: DNA "barcodes" and taxonomy. *Cladistics* 21, 405–407.

Whewell, W. (1847). *The philosophy of inductive sciences, founded upon their history*. 2nd ed. London: John W. Parker, London.

Will, K. W., B. D. Mishler, and Q. D. Wheeler (2005). The perils of DNA barcoding and the need for integrative taxonomy. *Systematic biology* 54, 844–51. DOI: 10.1080/10635150500354878.

Wilson, A. C., G. L. Bush, S. M. Case, and M.-C. King (1975). Social structuring of mammalian populations and rate of chromosomal evolution. *Genetics* 72, 5061–5065. DOI: 10.1073/pnas.72.12.5061.

Chapter II.

Sex-Biased Dispersal Obscures Species Boundaries in Integrative Species Delimitation Approaches

This chapter is intended for publication in Systematic Biology
Authors: Jonas Eberle, Erika Bazzato, Silvia Fabrizi, Michele Rossini, Mariastella Columba, Davide Cillo, Marco Uliana, Ignazio Sparacio, Guido Sabatinelli, Rachel C. M. Warnock, Giuseppe Carpaneto, and Dirk Ahrens

Authors' contributions to the original article:
Fieldwork collections: DA, SF, MR, MC, EB, DC, MU, GC, IS, GS; molecular lab work: JE, DA, SF; phylogenetic analyses, species delimitation analyses, figures: JE; manuscript design and writing: JE, RCMW, DA.

1. Introduction

Since the announced taxonomy crisis (Gewin 2002; Godfray 2002; Dayrat 2005) and proposed ways out (Sites and Marshall 2004) great progress has been made in the development of DNA-based approaches of species delimitation (e.g., Templeton et al. 1992; Meier et al. 2006; Pons et al. 2006; Puillandre et al. 2012b; Ratnasingham and Hebert 2013; Zhang et al. 2013; Jones 2014; Fujisawa et al. 2016). Many of these were widely applied using mitochondrial DNA, also due to the success of the Barcoding initiative (Hebert et al. 2003a; Hebert et al. 2003b; Ratnasingham and Hebert 2007), proposing solutions for rapid biodiversity assessment and turbo-taxonomy (Hebert and Gregory 2005; Butcher et al. 2012; Riedel et al. 2013a). Such studies often inferred major biodiversity underestimation (Pons et al. 2006; Monaghan et al. 2009; Balke et al. 2013; Riedel et al. 2013b; Katouzian et al. 2016; Thormann et al. 2016). Criticism against the use of single mitochondrial markers for species delimitation have been early acknowledged (Meyer and Paulay 2005; Meier et al. 2006; Rubinoff et al. 2006; Roe and Sperling 2007; Seberg and Petersen 2009; Dupuis et al. 2012) and found the answer in integrating various lines of evidence for taxonomy (Wiens and Penkrot 2002; Dayrat 2005; Will et al. 2005; Bond and Stockman 2008; Cardoso et al. 2009; Padial et al. 2009; Padial et al. 2010; Schlick-Steiner et al. 2010; Andújar et al. 2014). In particular the inclusion of multi-locus species delimitation approaches based on the multispecies coalescent theory (O'Meara 2010; Yang and Rannala 2010; Ence and Carstens 2011; Yang and Rannala 2014; Fujisawa et al. 2016) and simultaneous analysis of morphological and molecular data is promising (Yeates et al. 2011; Huang and Knowles 2015; Solís-Lemus et al. 2015; Eberle et al. 2016).

Accurate delimitation of species is crucial for a stable taxonomy which is the foundation for evolutionary, ecological, and basically all biological studies. Numerous cases prove to be challenging for species delimitation (Carstens and Dewey 2010; Hambäck et al. 2013). This is particularly true for species flocks with high morphological

but little genetic divergence (Moran and Kornfield 1993; Shaffer and McKnight 1996; Baldwin 1997; Freeland and Boag 1999; Petren et al. 2005; Wagner et al. 2013; Eberle et al. 2016) like often encountered in cases of recent divergences due to incomplete lineage sorting or hybridization (Witter and Carr 1988; Shaffer and Thomson 2007; Solís-Lemus et al. 2015; Eberle et al. 2016). But also little or no morphological but high genetic divergence may be problematic (Jockusch and Wake 2002; Holland and Hadfield 2004; Kozak et al. 2006; Wake 2006; Oliver et al. 2009; Riedel et al. 2010; Eberle et al. 2012; Niemiller et al. 2012; Arthofer et al. 2013; Barley et al. 2013; Cicconardi et al. 2013; Satler et al. 2013).

Incorrect species estimates can further be caused by upstream methodological shortcomings in phylogenetic reconstruction or branch smoothing methods (Leaché and Fujita 2010; Astrin et al. 2012; Tang et al. 2014) or the use of inappropriate molecular markers (e.g., Tang et al. 2012). Recently, increasing attention was paid to the ratio of the population genetic parameters θ (population mutation rate) and τ (time between species divergences, i.e., the speciation rate) (Zhang et al. 2011; Esselstyn et al. 2012; Reid and Carstens 2012), which has to be sufficiently high to allow accurate species delimitation. Particularly distinctly varying effective population sizes between nearly related species violate model assumptions and distort species delimitation results (Monaghan et al. 2009; Fujisawa and Barraclough 2013; Ahrens et al. 2016). Moreover, infections with *Wolbachia* can have tremendous effects on species delimitation (Whitworth et al. 2007).

While the effect of interspecific gene flow (introgression) on species' integrity and species delimitation was discussed in detail (Vogel and Johnson 2008; Petit and Excoffier 2009; Yang and Rannala 2010; Zhou et al. 2010; Zhang et al. 2011), the effect of reduced gene-flow among populations on species delimitation (Lohse 2009; Papadopoulou et al. 2009; Niemiller et al. 2012), received less attention so far. It may result in overestimations of actual species numbers by molecular approaches (Bond and Stockman 2008; Hey 2009; Lohse 2009; Harrington and Near 2012; Andriollo et al. 2015). One particular case is sex-biased dispersal, where gene flow is reduced in the least dispersing sex. This may influence the probability of introgression of maternally or paternally transmitted genes (Petit and Excoffier 2009) but also the genealogical patterns of the mitochondrial genome (Melnick and Hoelzer 1992; Palumbi and Baker 1994; Lyrholm et al. 1999; Castella et al. 2001; Kerth et al. 2002). The impact on

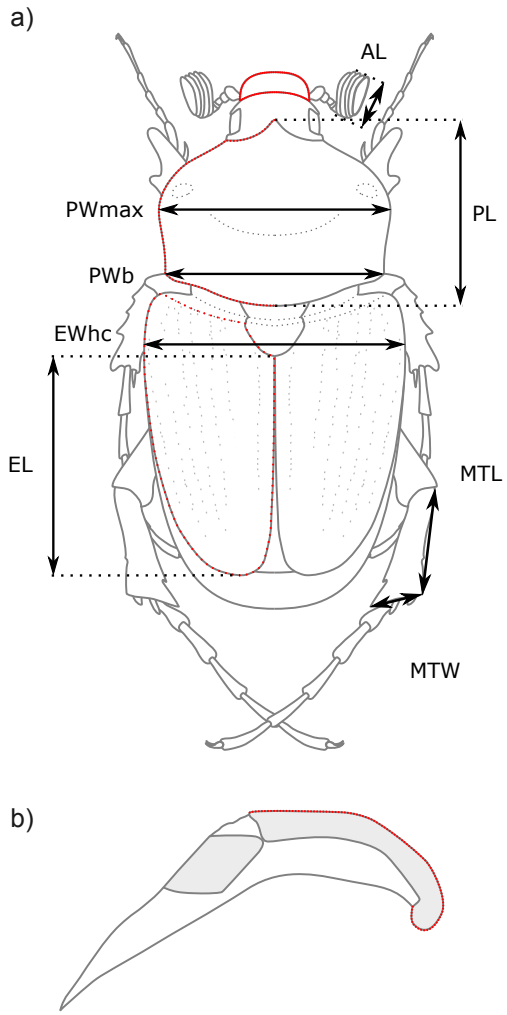

Figure II.1 Linear measurements and outlines used in morphometric analyses of (a) adult males (clypeus, pronotum (left half), elytron) and (b) their copulation organ (paramere, lateral view). MTL and MTW were measured in lateral view of the leg. Outlines were digitized as 100 equidistant semilandmarks (red dots). AL = length of antennal club, Ewhc = elytron width at humeral callus, EL = elytron length, Pwmax = maximum pronotum width, PWb = width of pronotum at base, PL = medial pronotum length, MTL = length of metatibia, MTW = maximum width of metatibia.

species delimitation is confirmed by high false-positive error rates of distance- and tree-based species delimitation methods, which were applied on data from coalescent simulations of mitochondrial data, including a subtle sex bias in dispersal (restricted female-mediated gene flow between populations) (Dávalos and Russell 2014).

Sex-biased dispersal is best known from vertebrates due to behavioral origin and is generally male-mediated in mammals and female-mediated in birds, while both modes are observed in insects (Petit and Excoffier 2009). Here, we investigate the case of the scarab beetle genus *Pachypus* which exhibits extreme (by orders of magnitude higher)

male-mediated dispersal. *Pachypus* is the only genus in an ancient pleurostict lineage of chafers (Pachypodini) that originated around 80 Ma (Ahrens et al. 2014). It is exclusively distributed in the western Mediterranean (Sparacio 2008) from where it is known since the Oligocene (Serres 1829). While taxonomists believed for decades that *Pachypus* is composed of three extant species (Baraud 1985; Baraud 1992), recently two more species were described from Sardinia (Sparacio 2008; Guerlach et al. 2013). Females of *Pachypus* lack elytra and hindwings and have a nearly entirely hypogean life style (Crovetti 1969; Arnone and Sparacio 1990).In contrast to their saprophagous larvae, adults have completely reduced mouth parts and do not feed at all. Due to their hidden lifestyle, females are rarely found and most available data is based on male specimens. Males actively disperse by flight in search of females that release pheromones from subterranean tunnels (Crovetti 1969; Arnone and Sparacio 1990). Therefore, along with unaffected dispersal of autosomal markers (Prugnolle and Meeus 2002), we expect mitochondrial maternal-lineages to evolve quasi independently within spatially strongly restricted areas.

The present study is an empirical test case for integrative species delimitation in the presence of extreme male-mediated sex-biased dispersal. Since female philopatry is assumed to affect at least gene flow of mitochondrial DNA, and possibly that of other sex-linked traits, we explore its impact on a variety of commonly used species delimitation methods, including *de novo* species delimitation and species validation approaches (Ence and Carstens 2011). We evaluate the introduced bias by comparison of results from exhaustive analyses of single loci, mitochondrial loci, nuclear loci, and combined data. Morphology proved to be highly informative for use in species delimitation (Huang and Knowles 2015; Solís-Lemus et al. 2015; Eberle et al. 2016). Therefore, we investigate the performance of morphometric data in the given scenario as a second data source in a fully integrative Bayesian framework. More precisely, we evaluate the species-boundary related signal in linear measurements of body parts ("traditional morphometrics"), as well as of outline shapes of body and genital parts ("geometric morphometrics").

2. Material and Methods

The study is based on a total sampling of 205 specimens of the genus *Pachypus* from 62 sampling localities (Supplementary Fig. A1; Electronic Supplement A1 and Supplement Table A1). Several outgroup taxa (all Scarabaeidae) were chosen from closely related lineages (Ahrens et al. 2014), such as *Oxyomus silvestris* (Aphodiinae), *Cetonia aurata* (Cetoniinae), *Buettikeria echinosa* (Pachydemini), *Chasmatopterus* sp. (Chasmatopterini), *Phyllopertha horticola* (Rutelinae). Maps were made in QGIS (`www.qgis.org`) using data from ETOPO1 (`http://www.ngdc.noaa.gov/mgg/global/relief/ETOPO1/data/`, accessed on 2015-10-21) and Natural Earth (`www.naturalearthdata.com`, accessed on 2013-09-24). 195 *Pachypus* specimens were suitable for DNA extraction and 164 specimens were used for morphometric studies; all others had to be excluded due to badly preserved DNA or missing or destroyed body parts.

2.1. Molecular Lab Procedures

Molecular data was prepared for two mitochondrial and two nuclear loci. Specimen collection, preservation and DNA extraction followed Ahrens and Vogler (2008). Qiagen®Multiplex PCR Kits were used for polymerase chain reaction. Mitochondrial loci were amplified using the primer pair stevPat and stevJerry (Timmermans et al. 2010) for the 3‘ end of cytochrome oxidase subunit 1 (*cox1*) and 16Sar and 16sB2 (Simon et al. 1994) for 16S ribosomal DNA (*rrnL*). Nuclear 28S ribosomal DNA (28S) was amplified using the primer pair FF and DD (Monaghan et al. 2007) and partial Arginine Kinase gene (ArgK) with AK183F and AK939R (Wild and Maddison 2008). If multiple products were amplified by the ArgK-primers, the band with the correct length was determined with *GeneRuler*^{TM} *100bp DNA Ladder Plus* and cut from the Agarose-gel. Forward and reverse strands were sequenced by Macrogen (Seoul, South

Korea) using the same primers. Sequences were manually edited in Geneious® 7.1.8. Specimen vouchers were deposited in the collections of the Zoological Research Museum A. Koenig, Bonn (ZFMK).

2.2. Sequence Alignment and Phylogenetic Inference

Sequences were aligned using the divide-and-conquer realignment techniques implemented in SATé-II (Liu et al. 2009; Liu et al. 2012). Sub-problems with a maximum size of 102 specimens were aligned with MAFFT (Katoh et al. 2002, 2005) and subsequently merged with MUSCLE (Edgar 2004). The tree estimator was set to RAxML (Stamatakis 2006). The resulting multiple sequence alignments were subsequently checked by eye. Gene trees for each locus were inferred with RAxML (Stamatakis 2014). Tree search parameters were adapted to the respective alignments by running multiple analyses with different settings on 5 initial randomized maximum parsimony trees: initial rearrangement settings and the optimal number of gamma categories for the GTRCAT model of nucleotide substitution were chosen based on likelihood scores under the gamma model of rate heterogeneity. Final tree searches under the GTRCAT model were conducted on 20 initial randomized maximum parsimony trees with the best inferred initial rearrangement setting and the optimal number of gamma categories, respectively. The best known tree was again chosen by the gamma-based likelihood score. Maximum likelihood estimation of base frequencies was applied. Node support was assessed with 1000 RELL bootstraps (Minh et al. 2013) and bipartitions were drawn on the best tree found before.

2.3. Morphometric analyses

For morphology-based species inference we used traditional morphometrics of eight linear distance measurements (Fig. II.1, Supplementary Table A2) and four shape traits (Fig. II.1). Distance measurements were directly taken from the specimen using an ocular grid on a Zeiss SM20 Stereomicroscope. Body parts were measured with the endpoints in focus so as to ensure horizontal orientation. Prior to further processing, the raw measurements were size corrected since most variation in biological datasets is

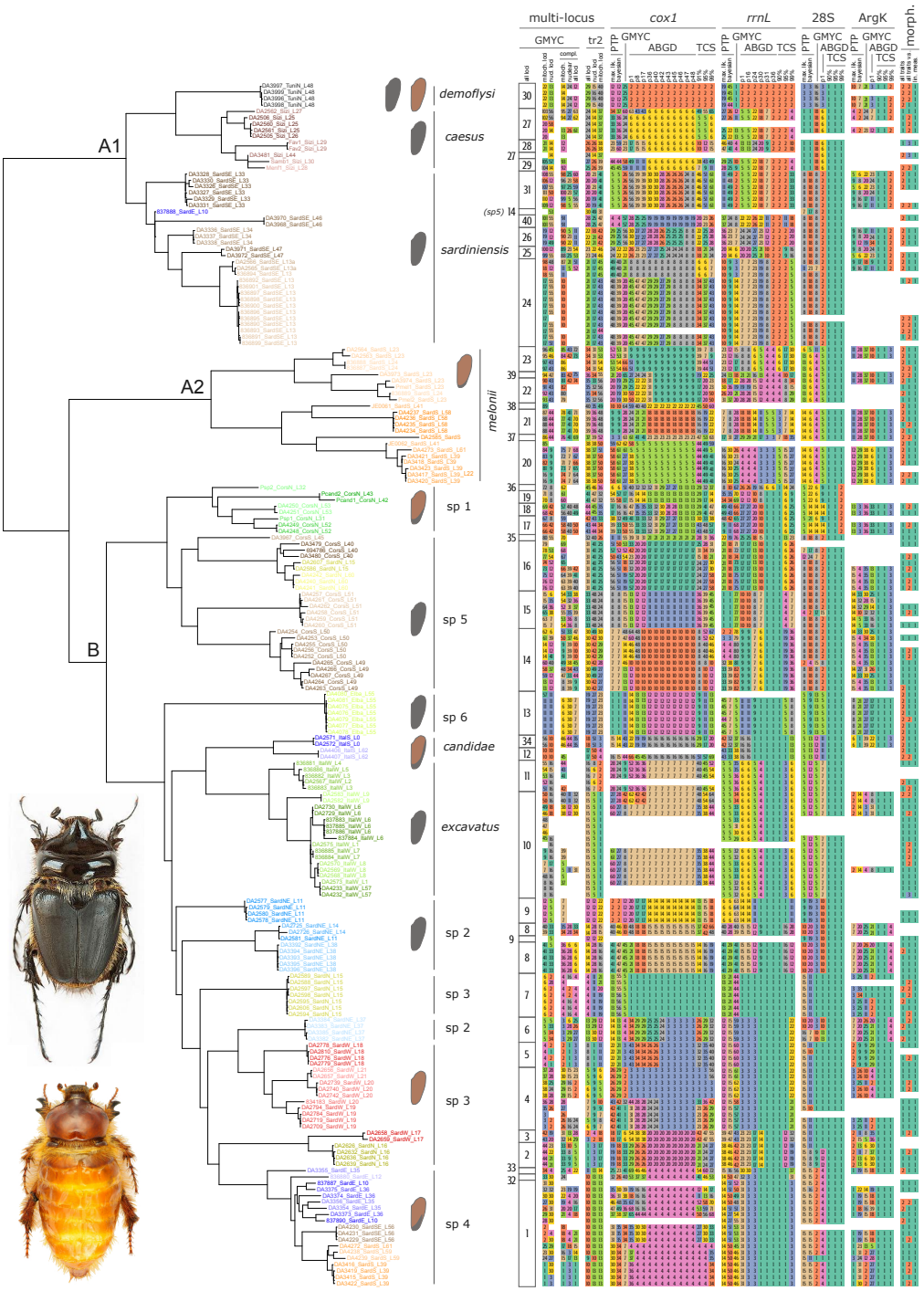

Figure II.2. ML tree of combined partitioned data and species delimitation results inferred by *'de novo'* approaches. Final delimitations from integrative analyses with iBPP are given next to the tree as are symbolic illustrations of the species right elytra coloration. (To be continued on next page.)

Figure II.2. (Continued.) The first column next to the tree depicts the results of the GMYC analysis of the combined nuclear and mitochondrial data that was used in species validation approaches (minimum clusters / MINCs). Numbers in the table indicate affiliation to the inferred putative species, table headings the respective data and species delimitation method. Specimens utilized in part 1 and 2 of subsequent Bayesian species delimitation with iBPP are indicated by A1 + A2 and B on the tree, respectively. Colors of tip labels refer to sampling localities in Figure II.6. Locality numbers in tip labels refer to Supplementary Figure A1 and Table A1. Colors of the delimitation table facilitate recognition of delimitation borders but may be inconclusive. Insets: male and unwinged female specimen of *Pachypus caesus* (photos by kind permission of Ballerio et al. 2014). Please find a scalable high resolution version of the figure in the pdf version on the CD.

usually introduced by differences in size (Jolicoeur 1963; Burnaby 1966; Ferrario et al. 1995). We applied Burnaby's Back Projection Method (Burnaby 1966) by projecting the log-transformed data on the isometric size vector and returning it to the original coordinate system (Adams and Rohlf 2000) using *R* code provided by Blankers et al. (2012). Shape traits were superimposed by Generalized Procrustes analyses on the outlines of the male copulation organ, the pronotum, the elytra, and the clypeus (Fig. II.1). For this purpose, images of each body part were captured with a Nikon digital camera DXM1200 (pronotum, elytra) or a Leica digital camera DFC 420, mounted on the Leica stereo-microscope SM-125 (aedeagus, clypeus). Objects were positioned during imaging with the outline in focus to ensure standardized orientation and 100 equidistant semilandmarks were digitized along the selected outline (Fig. II.1) in tpsDig (v. 2.17; Rohlf 2005). Generalized Procrustes analyses were performed using the *R* package *geomorph* (v. 3.0.0; Adams and Otarola-Castillo 2013). For closed curves (i.e., elytra and clypeus) all but the first landmark were defined as sliding landmarks (semilandmarks). For open curves (i.e., parameres and pronotum) all but the first and the last landmark were defined as sliding landmarks. Procrustes distance was used to optimize positions of the semilandmarks (Bookstein et al. 2002) and shapes were aligned by principal axes (Perez et al. 2006; Gunz and Mitteroecker 2013).

Dimensionality of measurements and outline data was reduced by principal component analyses (PCA) on the covariance matrix. The number of components to retain was identified by Horn's Parallel Analysis of principal components (Horn 1965) implemented in the *R* package *paran* (v. 1.5.1, Dinno 2009). This method adjusts for bias

in the retention of components that is introduced by finite sample sizes. Principal components with an adjusted Eigenvalue greater than 1 were used for downstream analyses. Morphological divergence among the minimum clusters (i.e., inferred species entities; MINCs) that were inferred by *de novo* delimitation approaches was visualized by mapping the entities on the first two principal components of three morphological datasets: linear measurements, all outline data, and all morphological trait data together. The variation within these datasets was summarized by principal component analyses on the correlation matrix. Size was not further regarded in any of the analyses.

2.4. *De Novo* Species Delimitation Approaches

On the basis of whether the samples are partitioned into species entities or not prior to the species delimitation analyses, delimitation methods were grouped into '*de novo*' delimitation and 'species validation' (Ence and Carstens 2011) approaches. The previously used term 'species discovery' (Ence and Carstens 2011) is rather misleading since species boundaries are discovered, and not species, and thus is confounded with the true discovery of an unknown new species. Therefore, we prefer to keep using the term '*de novo* species delimitation' (e.g., Zhang et al. 2013; Ahrens et al. 2016) for approaches which attempt to partition the samples into species without any a priori information regarding species membership. We applied four commonly used and one recently published *de novo* species delimitation method: (i) Automatic Barcode Gap Discovery (ABGD) (Puillandre et al. 2012a), (ii) Statistical Parsimony (Templeton 2001), (iii) Generalized Mixed Yule Coalescent (GMYC) modeling (Pons et al. 2006)), (iv) Poisson tree processes (PTP) modeling (Zhang et al. 2013), and (v) trinomial distribution of triplets (tr2) (Fujisawa et al. 2016). Mclust (Fraley and Raftery 2002; Fraley et al. 2012) was used to infer species boundaries based on morphological data. ABGD, GMYC, and PTP analyses were conducted without outgroup specimens and duplicate haplotypes to avoid false positives (Ahrens et al. 2016). Unique haplotype datasets were inferred with *haplotypes* in *R* (v. 1.0, Aktas 2015). Putative species assignments for excluded specimens were subsequently assigned to identical haplotypes.

ABGD relies on the assumption of a gap between smaller pairwise intraspecific and larger interspecific molecular distances (Puillandre et al. 2012a). We applied

ABGD (version of April 11th 2013) on distance matrices that were inferred with IQ-TREE (Nguyen et al. 2015) for each locus, correcting distances by the best fitting substitution models according the BIC (Supplementary Table A2). The ranges of prior intraspecific divergence per locus were adjusted by examining histograms of the pairwise distances: we evaluated 100 steps between 1% (Pmin) and 100% (Pmax) of the maximum observed distance since the approximate position of the barcode gap was unknown. This procedure was also applied for all loci including *cox1*, for which the barcode gap is regularly expected at about 1% to 3% pairwise sequence divergence (Puillandre et al. 2012a), since higher intraspecific divergence was expected due to the female philopatry of *Pachypus*. The relative gap width (X) regularly implements the assumption that the barcode gap is larger than any gap in the prior intraspecific divergence. Since no larger gap was found beyond any gap within prior intraspecific distance, X was set to 0.5. It thus allowed the interspecific gap to be half the size of gaps within intraspecific distances. This way, multiple species hypotheses were obtained which subsequently have to be evaluated based on further data or methods (Puillandre et al. 2012a). Additionally, *cox1* distances between specimen groups of the same locality were calculated using SpeciesIdentifier (Meier et al. 2006).

Statistical parsimony network analyses (Templeton 2001) were carried out in TCS (v1.21; Clement et al. 2000), with gaps treated as 5th character. This method infers potential species as statistically significant genetic differentiated entities based on the level of homoplasy in the data. It partitions the data into networks of closely related haplotypes connected by changes that are non-homoplastic with a given probability. It was shown that the 95% connection limit performs well on identifying insect species (Monaghan et al. 2006; Ahrens et al. 2007; Hart and Sunday 2007). Additionally, we applied a 90% and 99% connection limit to account for bias induced by philopatry. *Cox1* was analyzed with 91% instead of 90% because TCS hang up.

The GMYC combines stochastic lineage growth (a Yule process) and coalescent theory. It relies on the observation of a shift in branching rates at the transition point from speciation to coalescence, i.e., the putative species boundary which is detected as an increase in the accumulation of lineages through time (Pons et al. 2006). Maximum likelihood trees of each gene and combined partitioned genes, (i.e., mitochondrial genes, nuclear genes, and all genes) were inferred with RAxML using the same procedure as described above but using unique haplotypes or unique combina-

tions of haplotypes, respectively. For the combined data analyses we performed one run for each with all available data, and one run with complete data (i.e., only specimens without missing loci) only. Trees were made ultrametric in PathD8 (Britton et al. 2002; Britton et al. 2007) assigning the root an arbitrary age of one. PathD8 applies rapid rate smoothing and was shown to be suitable for use together with the GMYC method, producing stable species delimitation results (Monaghan et al. 2009; Astrin et al. 2012; Papadopoulou et al. 2013; Talavera et al. 2013; Tang et al. 2014). GMYC modeling was performed using *splits* (Ezard et al. 2014) in *R* applying the single threshold model which was shown to outperform the multiple threshold model, since the latter tends to overestimate the number of clusters (Fujisawa and Barraclough 2013; Dellicour and Flot 2015). The multi-model comparison approach (Burnham and Anderson 2002) was used to assess species supports.

In PTP, the Yule-coalescent transition points are modeled based on the change of substitution rates on the phylogenetic input tree (Zhang et al. 2013). The error prone step of tree ultrametrization, which is necessary for the GMYC, is omitted by directly analyzing mutational steps along the branches of the tree. PTP is thus intended for single-locus (or linked-loci) analyses (Zhang et al. 2013). The analyses using the likelihood-based (PTP) as well as the Bayesian approach (bPTP) were run using the web service (`http://species.h-its.org/ptp/`) for the same single loci RAxML trees that were used for the GMYC. MCMC chains for bPTP were run for 500,000 generations, sampling every 100 generations and discarding a burnin of 10%. Convergence and stationarity of bPTP runs was assessed by calculating the effective sampling size (ESS) with *coda* in *R* (Plummer et al. 2006), ensuring ESS values larger than 200.

The multispecies coalescent (Rannala and Yang 2003) provides a valuable framework for multi-locus species delimitation (O'Meara 2010; Yang and Rannala 2010; Ence and Carstens 2011; Yang and Rannala 2014; Fujisawa et al. 2016). The genome of sexual organisms is recombined in every generation and thus there is discordance between gene trees within panmictic populations. As a consequence of reproductive isolation between species, increased concordance is expected between gene trees beyond the species level (Degnan and Rosenberg 2009). tr2 infers the transition point of between species branching and within species branching by fitting the distribution of concordant triplet topologies (partial rooted trees consisting of three tips) among gene

trees to predefined trinomial distributions that are expected under the multispecies coalescent for the two branching regimes (Fujisawa et al. 2016). By decomposing the trees into rooted triplets, it is more efficient than earlier implementations of statistical delimitation based on topological congruence (O'Meara 2010) and thus suitable for large datasets. tr2 was applied on the gene-trees derived from a STACEY analysis (see below) without providing a guide tree or pre-defined species entities.

Putative species boundaries based on morphological data were inferred using model based finite mixture Gaussian clustering as implemented in the *R* package *mclust* 5.2 (Fraley and Raftery 2002; Fraley et al. 2012). The procedure evaluates the fit of the data to models with various combinations of geometric parameters (volume, shape, and orientation) for a predefined range of number of clusters. The best fitting number of clusters and model were chosen based on the BIC, which penalizes the log-likelihood for more complex models. While simpler models generally require more clusters for a good representation of the data, less clusters are needed with more complex models (Fraley and Raftery 2002). All available models on a predefined range of 1 to 100 clusters were investigated for the retained PCs of linear measurements and all four outline analyses in combination as well as for the retained PCs of linear measurements alone. A reduced optimal subset of variables was chosen with *clustvarsel* 2.2 in *R* (Scrucca and Raftery 2014) for an additional *mclust*-analysis on all combined traits. This was recently shown to improve clustering if only a subset of the available variables provide clustering information (Scrucca and Raftery 2014).

2.5. Species Delimitation by Validation

Algorithms modeling the multispecies coalescent (O'Meara 2010; Yang and Rannala 2010; Ence and Carstens 2011; Yang and Rannala 2014; Fujisawa et al. 2016) often supersede single-locus species delimitation approaches as they may accommodate incomplete lineage sorting (Rannala 2009; Camargo et al. 2012). Particularly the integration of multi-locus molecular data with morphological traits yields promising results (Solís-Lemus et al. 2015; Eberle et al. 2016). Due to their computational demand and the huge parameter space that has to be searched for simultaneous species tree estimation and species delimitation (Yang and Rannala 2014; Fujisawa et al. 2016), most methods require *a priori* designations of individuals to populations or

putative species. Here we employed STACEY (Boukaert et al. 2014; Jones 2014) and BPP (Yang and Rannala 2010, 2014) including its derivate iBPP (Solís-Lemus et al. 2015) to assess the support under the multispecies coalescent for initial species entity hypotheses obtained from the *de novo* species delimitation (i.e., GMYC clusters derived from the analysis of all loci including missing data). In order to decrease uncertainty in the results, specimens with large amounts of missing data were omitted from coalescent analyses.

For the STACEY analysis, data was partitioned for the 4 loci to allow independent site models, mutation rates, and gene trees for each. Site models were set according IQ-TREE (v1.3.13; Nguyen et al. 2015) inferences except for *cox1* which was set to the second best scoring model (Supplementary Table A2), since TIM2-based models are not available in BEAST. After confirming a rather clock-like distribution of rate heterogeneity (coefficient of rate variation of the lognormal clock ≈ 0.3) among branches in a preliminary analysis, we used a strict molecular clock. This simplification might be preferable and produces better rate estimates (Ho et al. 2005) and more accurate topologies (Drummond et al. 2006) in cases of low rate heterogeneity. Two variants of the analysis were conducted, describing the collapse weight prior (ω) by a beta-binomial distribution with (i) $\alpha = 1$ and $\beta = 1$, and (ii) $\alpha = 40$ and $\beta = 2$. While the former combination equals prior probability for all species numbers, the latter places the highest probability on five species which is way below the estimated numbers of potential species and puts a strain on data support for many species. Each run was repeated 4 times with 10^8 generations each, sampling every 10^4 generations. 10% of the samples were removed as burnin before assessing stability and convergence of the combined runs by examination of all parameters trace plots and ESS values (ESS > 200) with Tracer 1.6 (Rambaut et al. 2014). The parameter- and tree-log-files of all repeats were combined using *burntrees* (Nylander 2014) and summarized with the BEAST module *SpeciesDelimitationAnalyzer* (Jones et al. 2015) applying collapse weights of 10^{-4}, 10^{-3}, and 10^{-2}. Similarity matrices were visualized using the*R* code provided by Jones et al. (2015).

True progress towards integrative species delimitation was made by implementing the possibility of analyzing continuous trait data in the BPP (Yang and Rannala 2010, 2014) derivative iBPP (Solís-Lemus et al. 2015). iBPP requires as input a guide tree so that lumping of the *a priori* defined entities is restricted to predefined sister

clades (Leaché and Fujita 2010; Olave et al. 2014; Eberle et al. 2016). Where ever appropriate, alternative guide tree topologies should therefore be tested, which can for instance be informed by alternative molecular hypotheses, morphological data (Eberle et al. 2016), or geography. We used the species tree resulting from the STACEY analyses on MINCs as guide tree. Since the number of initial entities is limited in iBPP due to computational complexity, the guide tree was split in two parts that were analyzed separately (Supplementary Fig. A7). Since node support was generally low in the second part of the guide tree (node B; Fig. II.2, Supplementary Fig. A3) and the topology was inconsistent with the maximum likelihood tree inferred with RAxML on the combined partitioned data (Fig. II.2, Supplementary Fig. A3), an alternative guide tree topology based on relationships of nuclear 28S and geography was tested (Fig. A7c). We used six different combinations of molecular and morphological data to assess the influence of different data on species delimitation: (i) all loci and all traits (i.e., combined retained PCs of linear measurements and outline analyses of aedeagus, clypeus, elytron, and pronotum), (ii) all loci and linear measurements, (iii) nuclear loci and all traits, (iv) nuclear loci and linear measurements, (v) all traits without DNA, and (vi) linear measurements only. Additionally, we ran iBPP without data to evaluate the influence of the chosen priors. The priors for the population mutation rate (θ) and the root height (τ_0) can substantially alter the results and must be chosen carefully (Leaché and Fujita 2010; Eberle et al. 2016). We therefore applied nine pairwise combinations of three γ distributions G(α,β) for both θ and τ_0: (i) α=2 and β=20 with mean 0.1, (ii) α=2 and β=200 with mean 0.01, (iii) α=2 and β=20000 with mean 0.0001. The control files were created with BPPmulti (*perl* scripts available at `http://github.com/eberlejonas/BPPmulti`). We used the standard species delimitation algorithm which assigns equal probabilities to rooted species trees since the alternative algorithm which assigns equal probabilities to all labeled histories may tend to over-resolve large unbalanced guide trees (Yang and Rannala 2010). Results were visualized with BPPmultitool (*R* scripts available at `http://github.com/eberlejonas/BPPmultitool`).

3. Results

3.1. Phylogenetic inferences

Gene trees showed major topological discordance between all loci despite often high node supports (Supplementary Fig. A2). Likewise, topological differences between mitochondrial and nuclear trees were strong (Supplementary Figs. A2, A3), while the tree of all combined marker (Fig. II.2) appeared more similar to the tree of mitochondrial data. However, identical clades of few closely related specimens were inferred in all analyses (although often indistinguishable in 28S). These clades often reflected groups that were collected at the same or nearby localities. Only few localities (L15, L39, L59) were populated by more than one such stable clade. On the other hand, clades from localities at the north-eastern coast of Sardinia (L11, L14, L37, and L38) were scattered over the mitochondrial trees but closely related in nuclear trees where they are represented by only one haplotype (Supplementary Figs. A2, A3).

A close relationship that was recovered in all tree reconstructions (except with 28S) was inferred for Tunisian and Sicilian clades which are ascribed to *P. demoflysi* and *P. caesus*, respectively (Figs. II.2; Supplementary Figs. A2, A3). Together with a clade that is distributed at the south-eastern tip of Sardinia (*P. sardiniensis*), they constitute a basal lineage in the tree based on all available data. In the sister clade to these three previous forms, the clade of *P. melonii* from south-eastern Sardinia resulted sister to all remainder specimens, followed by a clade spanning the Strait of Bonifacio between Corsica and Sardinia. Further stable clades found in nearly all tree reconstructions were a very homogeneous clade from Elba and a lineage constituted of *P. excavatus* and *P. candidae* from western and southern Italian mainland, respectively. The relationships among the remaining clades, sampled from all over Sardinia, remain rather obscure and are highly inconsistent between gene trees which

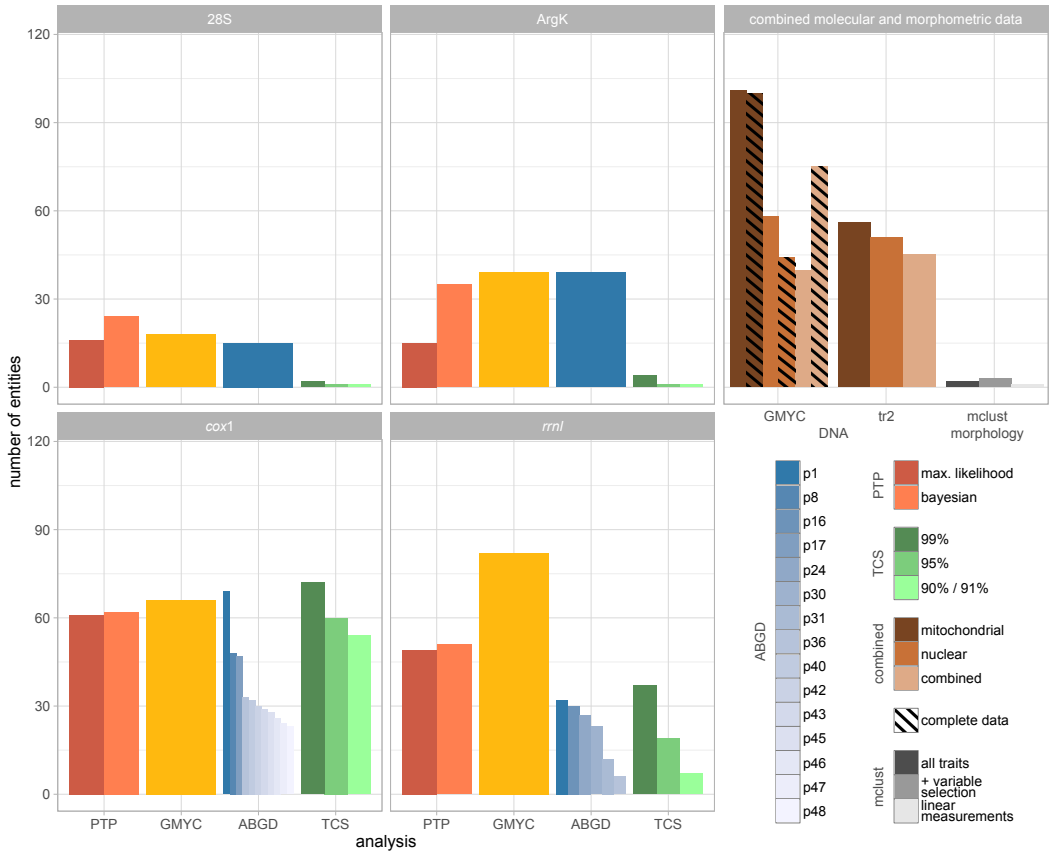

Figure II.3. Number of entities delimited by *'de novo'* approaches. Partitioning under differing prior interspecific distances is illustrated for ABGD if available.

is also reflected by low node support values (Supplementary Fig. A2) in this part of the tree based on all loci.

3.2. De Novo Species Delimitation Approaches

The range of putative species entities found by all *de novo* species delimitation approaches was huge, spanning from one entity, inferred for instance by statistical parsimony networks of nuclear markers, to 111 entities, inferred by the GMYC model on the tree of combined mitochondrial data (Figs. II.2, II.3). The number of inferred entities was highest if mitochondrial data was used. With these data GMYC assigned

even specimens from the same locality with only a few segregating sites in *cox1* to separate putative species (Fig. II.2). Most methods predominantly joined specimens from the same sampling locality together in one putative species entity. The raw genetic distances between specimen entities from the same locality was high, only 4 entities had less than 4% but about 60% of them had between 13 to 18% divergence for *cox1*. However, the distributions of pairwise distances that was used for ABGD revealed no obvious barcode gap (Supplementary Fig. A4) and thus resulted in unreliable results for even for the frequently used cox1. Depending on parameter choice, ABGD and TCS lumped increasing numbers of specimens without reaching stable estimates for a range of parameters (Fig. II.3). Similar problems arose with *rrnL*.

Similarly, irregular likelihood surfaces and low species support values of the GMYC (Supplementary Figs. A14–A23) indicated a bad fit of the model to all data sets. The GMYC relies on an increase of lineage accumulation at the species-population border; however, lineage through time plots showed no conspicuous and unambiguous increase at any point (Supplementary Figs.A14–A23). Compared to GMYC, bPTP and particularly the maximum likelihood version of PTP were less prone to over-splitting. However, posterior probabilities of bPTP rendered further over-splitting often more likely than lumping of the inferred entities; this was particularly true for *cox1*. Delimitation on combined molecular data (GMYC and tr2) had different effects particularly for the GMYC model: results between mitochondrial, nuclear, and combined analyses were very inconsistent among each other. With the combination of all markers, GMYC yielded 40 entities and more than halved the exaggerate number of entities obtained with the combined mitochondrial markers (n=111), Interestingly, also combining nuclear data consistently increased the number of inferred putative species (n=58) compared to the single nuclear gene analyses. tr2 yielded overall more consistent results between mitochondrial, nuclear, and combined analyses (Fig. II.3), with entity numbers (n=56, 51, 45; respectively) similar to those of the GMYC with combined all and nuclear data.

Mixed Gaussian Clustering with *mclust* did not recover the boundaries of clades that were found by any of the molecular methods (Fig. II.2). Only two diagonal clusters with equal volume and shape (Fraley and Raftery 2002) were inferred using all trait data (Fig. II.3). Using a subset of variables resulted in three ellipsoidal clusters with equal volume and equal shape while linear measurements alone showed no signal

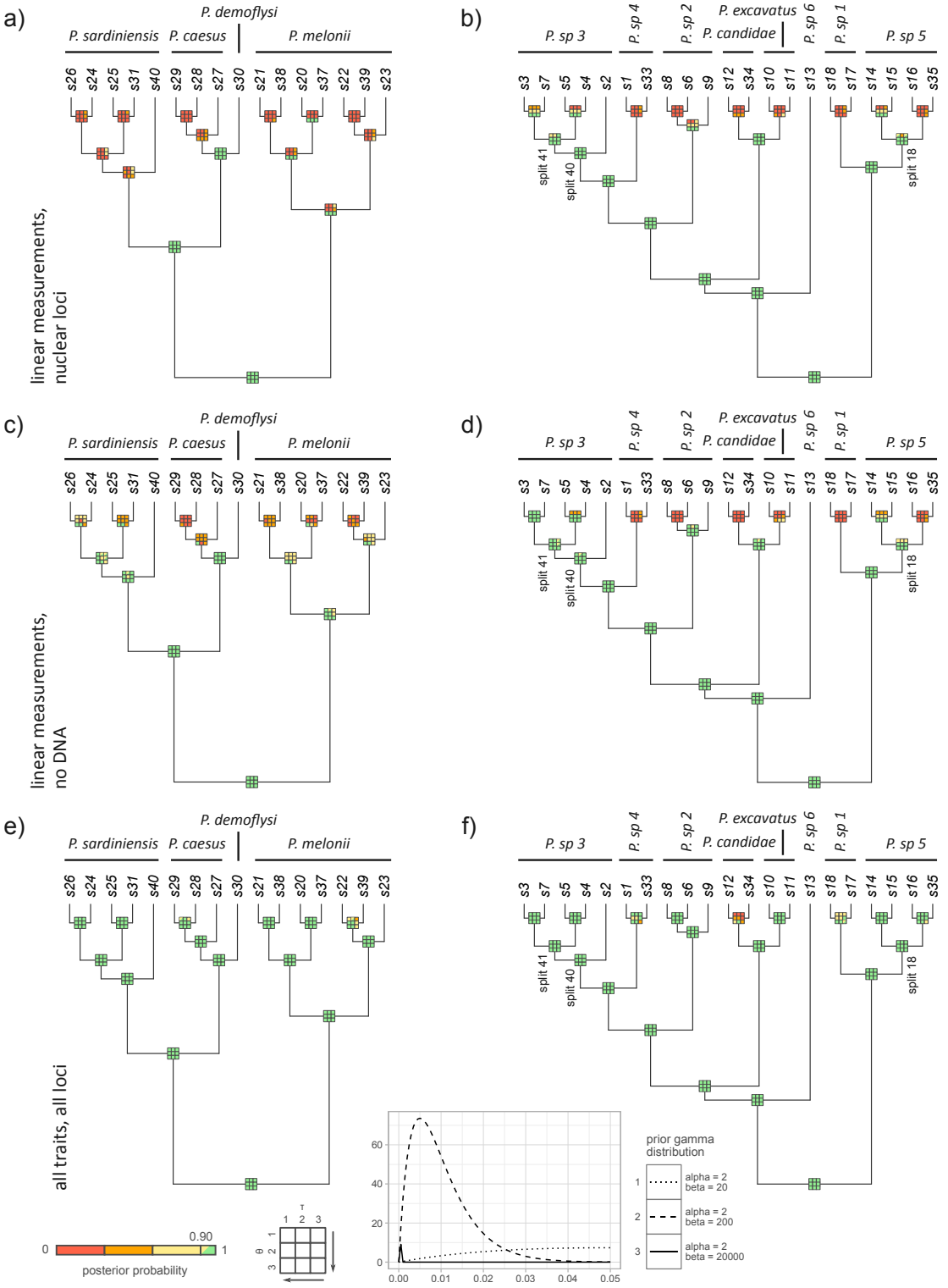

Figure II.4. Selected results of integrative Bayesian species delimitation with iBPP.
(To be continued on next page.)

Figure II.4. (Continued.) Part 1 (a, c, e) and modified geography-informed part 2 (b, d, f) of the guide tree are shown for analyses with linear measurements and nuclear loci (a, b), with linear measurements alone (c, d), and with all morphometric traits and all loci (e, f). Posterior probabilities of 3×3 combinations of τ_0 and θ prior distributions are illustrated at each node (referring to graphical legend; arrows point towards increased expected splitting). Minimum clusters from the GMYC analyses (Fig. II.2) are given at the tips of each tree along with the final species designations. The inset at the bottom shows the prior gamma distributions. Branch lengths are meaningless.

for any clusters (Fig. II.3). Cluster boundaries rarely matched those from molecular data. For instance, combined outline and linear measurement data separated the closely related *P. demoflysi* and *P. caesus* (Fig. II.2). However, in the probabilistic Bayesian framework of iBPP, resolution from morphometric data benefited from the a priori given hierarchical structure of the guide tree by restricting comparisons to molecular sister clades (see below).

3.3. Species Delimitation by Validation

Given the huge range of putative species numbers from *de novo* delimitation, we had to reduce subsequent validation approaches to one species model due to computational burden. We chose the putative species entities from the GMYC analysis of the combined data including all specimens (numbered s1–s40 and referred to as minimal clusters (MINCs) in the following; Supplementary Fig. A3) (see discussion for further explanation).

The STACEY analyses supported all MINCs as separate species (Supplementary Fig. A6). The two alternative applied prior expectations on the number of species (the first putting equal weight on all numbers of species and the second putting the highest weight on 5 species), yielded quasi equal results except for the highest tested collapse height ($\varepsilon = 0.01$). This setting, which still assumes two minimum clades to belong to one species if a certain degree of differentiation is already present, revealed highest similarity among MINCs from localities L33, L34, and L47 at the south-eastern tip of Sardinia (s25, s26, and s31; Supplementary Fig. A6). Only the more informative ω (collapse weight) prior, which puts a strain on less species, also found

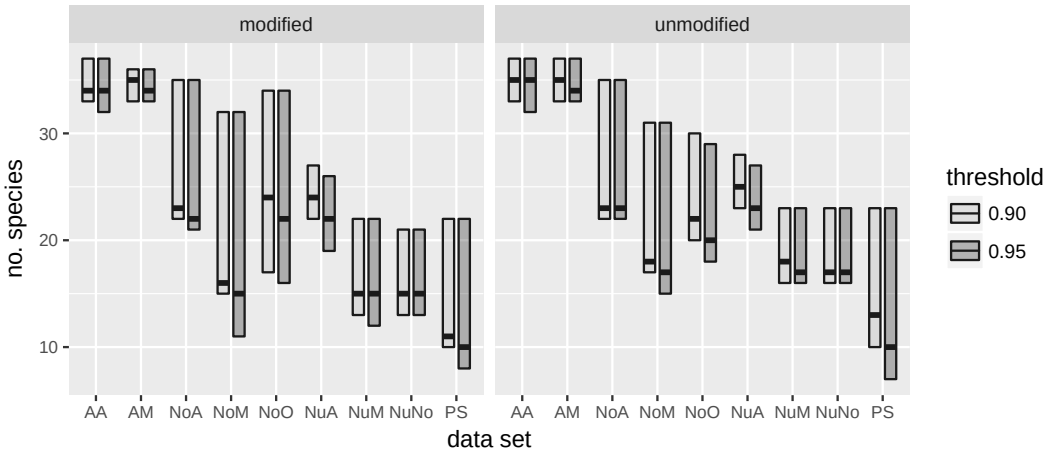

Figure II.5. Ranges of species numbers inferred by iBPP under different combinations of τ_0 and θ priors. Horizontal bars indicate median species numbers within each bar. Two thresholds of posterior probabilities for 'good species' are compared for the unmodified guide tree topology and the modified guide tree topology informed by geography and nuclear DNA for a range of data sets: AA = all loci, all traits, NoA = no DNA, all traits, NuA = nuclear DNA, all traits, AM = all loci, linear measurements, NoM = no DNA, linear measurements, NoO = no DNA, outline measurements, NuM = nuclear loci, linear measurements, NuNo = nuclear loci, no traits, PS = prior sampling.

increased probability to belong to the same species among MINCs from northern Corsica, north-western Sardinia, western Sardinia, south Sardinia, western Italian mainland, southern Apennine peninsula (Gulf of Taranto), and Sicilia (Supplementary Figs. A1, A6).

Mapping the MINCs on the three morphological datasets revealed large overlap among them (Supplementary Fig. A12). However, for outline data and all morphological data combined only 16.8% and 17.8% of the examined variation could be visualized in two dimensions (i.e., in principal component 1 and 2). The differentiation among these datasets was therefore much larger than apparent. Although large overlap among MINCs was visible in general, noticeable differentiation was found among certain nearly related molecular clades (e.g., *P. demoflysi* and *P. caesus*; i.e., s30 and s27 – s29; Supplementary Fig. A13).

iBPP was used with eight datasets of various combinations of molecular and morphometric data and prior sampling (Fig. A8), resulting in a total of 1215 analyses

(i.e., 9 prior combinations × 9 data sets × 5 repeats × 3 guide trees). Analyses including all molecular loci together with all morphological traits confirmed our previous findings. All MINCs were well separated except for s12 and s34 from southern Apennine peninsula (Fig. II.4e–f, Supplementary Figs. A9e–f, A10e–f, A11e–f). Support for many speciation splits decreased with the exclusion of mitochondrial loci (Figs. II.4a–d, II.5) or the exclusion of all molecular data (Fig. II.5, Supplementary Figs. A9a, b, h, A10a, b, h, A11a, b, h). Likewise, the inclusion of all morphological traits and of outline traits (Fig. II.5, Supplementary Figs. A9b, d, f, h, A10b, d, f, h, A11b, d, f, h) generally supported more splits than the inclusion of linear measurements only (Fig. II.5, Supplementary Figs. A9a, c, e, A10a, c, e, A11a, c, e). Strongest lumping of MINCs was found for nuclear loci, linear measurements, and their combination (Fig. II.5). Compared to other data combinations, nuclear loci (with or without linear measurements) yielded relatively high contrasts of node supports at putative species limits, i.e., well supported splits with posterior probabilities greater than 0.90 were immediately followed by highly unsupported splits with posterior probabilities smaller 0.3 (Supplementary Figs. A9c, A10c, A11c, green and red boxes). Consequently, the range of putative species numbers yielded from analyses with different prior combinations, was the smallest using the nuclear loci data set and the nuclear loci plus linear measurements data set (Fig. II.5), indicating a rather strong signal in the data for the respective species delimitation.

The choice of both θ and τ_0 priors influenced the posterior probabilities of splits. A general pattern was that prior distributions for θ with a higher mean tended to lumping while prior distributions for τ_0 with a higher mean tended to over-splitting (e.g., Fig. II.4b, d). However, this pattern was not consistent for all nodes in both analyses with and without data. Rarely was it even reversed (Fig. II.4a). While most of the posterior probabilities showed rather constant values with varying τ_0 priors, species support often varied decisively with changing θ priors. However, the majority of the speciation splits received very constant posterior probabilities independent of prior choice.

An alternative topology was tested for the second part of the guide tree (Fig. II.2, node B; Supplementary Fig. A7), where conflicting phylogenetic signal from mitochondrial and nuclear data between western and north-western Sardinian populations led to low node support values in combined data analyses (Supplementary Fig. A3).

After grouping MINCs from the Sardinian northern east-coast (MINCs: s6, s8, s9; localities: L11, L14, L37, and L38) and from the north-western and western localities (MINCs: s2–s5, s7; localities: L15-L21) in monophyletic clades according to their geographical vicinity, the posterior probabilities of nearly all splits strongly decreased (Fig. II.4b, d, f; Supplementary Figs. A11 vs. A10). Considerable – although decreased – molecular and morphological support was still found for splits in the latter clade (MINCs: s2–s5 + s7; Supplementary Fig. A11). Using nuclear loci and linear measurements with high θ- and high and medium τ_0-values, the posterior probabilities of the basal split of this clade (between s2 and s3–5,7) (Fig. II.4, split 40) decreased to 0.96 and 0.95, respectively (Supplementary Table A5). It lost support in several single repeat-analyses with high θ-values, also for the analyses of nuclear loci alone and of linear measurements alone.

4. Discussion

Due to generally uniform copulation organs and external morphology with strong color polymorphism and subtle differences even among spatially only little separated populations, the inference of species boundaries in *Pachypus* beetles represents a serious challenge. This finds also expression in the high number of synonymies (Olivier (1789), Fabricius (1792), Erichson (1840), Reitter (1898), Luigioni (1923), and Schoolmeesters (2016); source: Schoolmeesters (2016)) associated so far with the species *P. candidae* (Petagna 1787).

Our very comprehensive and integrative approach to species delimitation, using various molecular and morphological markers confirmed the difficulty to recognize robust species boundaries for all taxa without additional subjective assumptions (e.g., exclusion of data or priors). It suggests that these difficulties are linked with the low dispersal capacity of females since particularly mitochondrial data strongly contribute to the inconsistent patterns among the various delimitation approaches (i.e., in terms of data and different methods used).

locus	mean	range
cox1	0.83	0.34 – 1.00
rrnL	0.63	0.23 – 1.00
28S	0.56	0.19 – 1.00
ArgK	0.46	0.14 – 0.98

Table II.1 Summary of the putative species' posterior probabilities from Bayesian PTP

4.1. De Novo Species Delimitation Approaches

The vast range of putative species numbers inferred by the here employed *de novo* delimitation methods (Fig. II.3), together with low supports of inferred entities (GMYC and PTP; Supplementary Figs. A14–A23, Supplementary Table II.1), emphasized the difficulty of species delimitation in *Pachypus* using molecular data. Causes to be considered may be various: early concerns on the GMYC model were about susceptibility to strong geographic variation (Lohse 2009) like it is to expect for mitochondrial data in *Pachypus* due to extreme female philopatry. Due to the rapid coalescence within local demes compared to coalescence among different demes, the assumption of within-species panmixia, being made by coalescent based methods, is violated and might be wrongly detected at the transition to demes (Lohse 2009), or simply obscured. Nevertheless, these methods may be successfully applied in most cases (Papadopoulou et al. 2009). Furthermore, testing for panmixia in advance is impossible, since it requires knowledge of species boundaries which does not apply to species delimitation studies. The extreme case of *Pachypus* should sensitize researchers for the issues of these widely used methods associated to the common phenomenon of reduced gene-flow. This issue might not only affect *de novo* coalescent methods (PTP, GMYC, tr2) but also validation approaches (see below). A further critical factor is the variation of effective population size (N_e) and species divergence times which was shown to be problematic in DNA-based species delimitation (Esselstyn et al. 2012; Reid and Carstens 2012; Fujisawa and Barraclough 2013; Ahrens et al. 2016). Depending on topography and habitat availability, and due to limited dispersal capacity of females, it should be assumed that consensus and effective population size of *Pachypus* species vary considerably between different species, demes, and populations (Wright 1946; Barrowclough 1980) and might therefore explain shortcomings of the employed delimitation methods. In particular female contribution to effective population size in *Pachypus* species is expected to be very small due to philopatry and virtually local isolation. Since most species of *Pachypus* are quite rare they are expected to have a low N_e (Palstra and Fraser 2012; Ahrens et al. 2016). In consequence, populations might exhibit high relative differences in N_e. Since the population mutation rate (θ) depends on N_e, populations might diverge fast and at different pace. This might explain the tendency of mitochondrial but also of nuclear loci in the GMYC model to over-split species as shown for cases of low N_e (Esselstyn et al. 2012). In particular

the assumption of a uniform age of species' coalescence in the GMYC model makes it sensitive to varying N_e of the involved clades (Monaghan et al. 2009; Fujisawa and Barraclough 2013; Ahrens et al. 2016). On the other side it seems unlikely that shallow tree depth (Esselstyn et al. 2012; Reid and Carstens 2012; Fujisawa and Barraclough 2013) is contributing to these problems since *Pachypus* already exists for at least 24 Myr (Serres 1829; Krell 2000).

Increased introgression of loci that are associated with the least dispersing sex (Petit and Excoffier 2009) makes them less diagnostic for species delimitation. This is because of rapid introgression of alleles into colonizing individuals with insufficient gene flow, which renders them distant from their source population (Petit and Excoffier 2009). Loci experiencing strong genetic exchange are considered better suited for species delimitation since sufficient intraspecific gene flow likely prevents an increase in the frequency of allele introgression (Currat et al. 2008) and therefore "sharpens" species boundaries (Petit and Excoffier 2009). In case of extreme female philopatry like in *Pachypus*, introgression of mitochondrial DNA might therefore additionally obscure species specific mitochondrial patterns (Rubinoff et al. 2006).

Also other approaches, including those relying on genetic distances without taking the specimens relationships into account, might be prone to these issues (Ross et al. 2008; Zhang et al. 2013). ABGD was able to partition individuals at different levels of mitochondrial variation; yet an absolute indication of species boundaries remains uncertain due to the lack of a distinct gap between intra- and interspecific variation (Fig. II.3, Supplementary Fig. A4) which might again be obscured by varying θ among populations. Also tr2 which is based on the multi-species coalescence might have been affected. However, although more than 25 loci are supposed to be required by tr2 to delimit with 95% success rate (if species divergence time is half the effective population size) (Fujisawa et al. 2016), the method outperformed the GMYC by producing much more consistent results among data sets (Fig. II.3).

4.2. Species Delimitation by Validation Approaches

Due to their computational complexity, most current multi-species coalescent methods are time consuming and mostly only suited for small numbers of predefined minimum

clusters. A highly sensible task – particularly in cases of multiple competing hypotheses in result of presumed over-splitting – is to validate the results from *de novo* species delimitation methods by multispecies coalescent based methods (Zhang et al. 2013) that employ all available data in a simultaneous analysis. The use of small *a priori* entities is preferable, since validation approaches may only infer lumping of predefined minimal clusters but not further splitting. However, compromise is inevitable with large sampling size (i.e., high number of minimal clusters) due to computational intensity. Moreover, the influence of singletons, like they may often result from fine-grain partitioning of individuals, is often insufficiently studied (see also Lim et al. 2012; Ahrens et al. 2016) and should be avoided if possible. It was shown, that BPP needs exceedingly more data for correct delimitation if only one specimen is available for a species (Zhang et al. 2011). We chose the GMYC entities (n=40) resulting from the complete combined and partitioned molecular marker dataset as minimal clusters (MINCs; Fig. II.2, Supplementary Fig. A3). Alternative delimitations with even higher numbers of entities appeared highly unlikely and were dismissed since they often grouped even specimens from the same locality with only little genetic divergence into separate clusters. Entities from model-based clustering on morphological traits were not suited for use as MINCs since they grouped specimens nearly randomly without any link to their genealogy or provenience (Fig. II.2).

The conspicuous drop of species posterior probabilities in iBPP with nuclear data (Figs. II.4a, II.5, Supplementary Figs. A9c, d, A10c, d, and A11c, d) confirmed the expectation that nuclear loci are, compared to mitochondrial loci, not or at least much less prone to over-splitting caused by female philopatry. Indication for the improved support of speciation splits by nuclear data came from insensitivity of posterior probabilities to prior choice: low support of the data in favor or against a speciation split resulted in a large range of posterior probabilities per split across runs with different priors, while stable posterior probabilities were expected if the data was informative (Fig. II.5). In practice, nodes with 9 green boxes (i.e., high posterior probabilities; each box stands for a different θ and τ_0 prior combination; Fig. II.4) that are followed by nodes with 9 red boxes (i.e., very low posterior probabilities;) indicate a case of very good data support for the concerning putative species. Nodes composed of a mixture of green, yellow, and red boxes (i.e., high, medium, or low posterior probabilities) are an indicator for less decisive splits. Such an increase of 'sharpness' of

split support was increasingly observed in iBPP analyses of nuclear data alone and of nuclear data combined with linear measurements (Fig. II.5). However, to exclude the influence of other confounding factors like for instance sampling scheme, singletons, number of loci, or population structure, such assessments should be tested more thoroughly in simulation studies, which would be beyond the scope of this paper and are subject to future research.

BPP is to date supposed to be the most accurate species delimitation software under simulated conditions (Camargo et al. 2012). Especially its derivative iBPP seems to be highly suitable for purposes of integrative species delimitation (Huang and Knowles 2015; Solís-Lemus et al. 2015; Eberle et al. 2016; Pyron et al. 2016). Challenging remains the choice of τ_0 and θ prior combinations. Over-splitting behavior of iBPP and BPP was shown for low θ and for high τ_0 prior means (Leaché and Fujita 2010), although deviations from this expectancy have been reported (Leaché and Fujita 2010; Eberle et al. 2016). This behavior was largely confirmed in the present study but was not entirely consistent among nodes. In a few cases even contrary tendencies were observed with increased speciation probability for high θ and low τ_0 prior means (e.g., Fig. II.4a; MINCs s24–s26, s31, and s40). This highlights the importance of testing multiple prior combinations for each case study, besides the beneficial effect of assessing the signal's insensitivity to prior choice (see above). Likewise, the importance of correct assignment of samples to the guide tree in order to prevent over-splitting (Leaché and Fujita 2010; Zhang 2014; Eberle et al. 2016) is underlined by the present data: increased lumping strongly supported the geography-informed modified guide tree part 2 (Fig. II.4b, d, f, Supplementary Fig. A7c). Inferences based on the original guide tree led to a more extensive splitting. Alternative guide tree topologies based on molecular data might be inferred with DensiTree (Bouckaert 2010) or based on similarity matrices gained from STACEY analyses (Jones 2014; Jones et al. 2015)) with high ε values. For instance, STACEY matrices illustrated the similarity of MINCs s6 and s8 (Supplementary Fig. A6) that was also inferred from geographical and nuclear DNA and that was implemented in the alternative guide tree part 2 (Supplementary Fig. A7c). However, geographical origin and morphological similarity might provide less circular indications for testable sister entities.

4.3. Implications for Species Boundaries

Building on the above discussion, a conclusion on species boundaries seems not straightforward. One major issue is the configuration and the number of MINCs for the validation analysis in iBPP, another issue is the choice of the data and priors to be included in this analysis. It seems to be apparent, that major issues and difficulties in species delimitation rise not only from possibly insufficient informative data but also from the issue of the magnitude of Ne and its presumable variation among populations. While utilization of more genes promises more robust results in multi-species coalescent simulations (Zhang et al. 2011; Camargo et al. 2012; Fujisawa et al. 2016), N_e issues could still render challenging in extreme cases like that of *Pachypus*. Our study case showed that the inclusion of more data does not necessarily mean more robust or less ambiguous results (Fig. II.5), even with the inclusion of morphometric traits. The current hypotheses of the most comprehensive full evidence analysis would mean that nearly every current *Pachypus* population represents a separate species (Fig. II.4e, f)!

A more conservative approach to this discussion would consider *a priori* the exclusion of mtDNA from the iBPP analysis pursuing the combined analyses of nuclear DNA and morphological measurements which resulted in an estimated set of twelve *Pachypus* species (Figs. II.4, II.6). With these data, the analysis of the first part of the tree (Fig. II.4a, c, e) recovered all so far recognized nominal species of *Pachypus* as valid entities (except *P. candidae*; i.e., *P. sardiniensis*, *P. meloni*, *P. caesus*, *P. demoflys*i). The split between the strongly allopatric *P. sardiniensis*, *P. caesus* and *P. demoflysi* might be compatible with the end of the Messinian salinity crisis, and was further supported by phenological divergence that was observed for *P. caesus*: while all species of *Pachypus* have a summer emergence of adult specimens, all Sicilian populations (i.e., *P. caesus*) are autumnal. This would also be in line with the presence of color polymorphism in *P. demoflysi* (*i.e.*, specimens entirely black or black and brown; in this study not included in species delimitation analysis; Fig. II.2) and its absence in *P. sardiniensis* and *P. caesus* (specimens entirely black).

A more problematic situation was encountered in the clade comprising populations from Sardinia and Corsica which were so far merged under the species name *P. candidae*. There is no general consensus about the height of the posterior probability

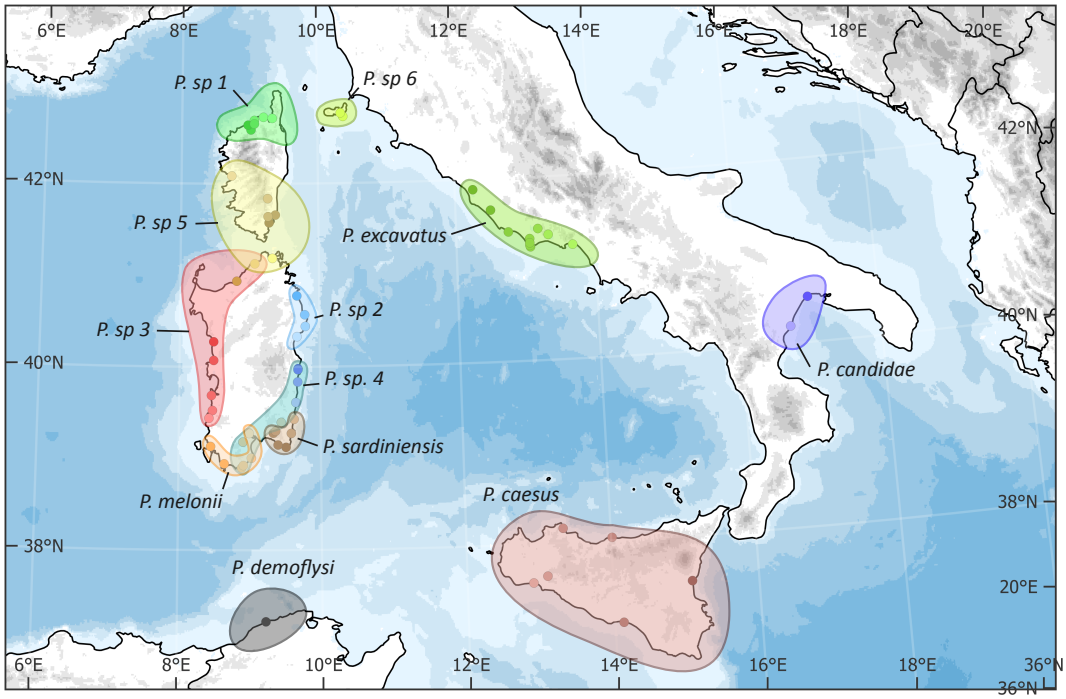

Figure II.6. Map of the central Mediterranean Sea including known distributions of the final inferred *Pachypus* species.

of a certain node to be a good split between valid species. Previous studies considered probabilities above 0.95 as support for a speciation event (Leaché and Fujita 2010; Eberle et al. 2016). However, the iBPP algorithm internally collapses nodes (in the result's tree output) with a posterior probability less than 0.90. Therefore, we generally adopted iBPPs internal behavior to consider a speciation split as supported with posterior probabilities > 0.90. While this decision had no impact in most cases (Fig. II.5), ancient splits of the presumptive species entities (*P.* sp3 and *P.* sp5; Fig. II.4b) were affected. Despite posterior probabilities > 0.90 in nearly all prior combinations (splits 18, 40, and 41, Fig. 4), we tend to consider the involved MINCs to belong to only two separate species. This decision was mainly based on the decrease of posterior speciation probabilities below 0.95 in the respective splits with certain prior combinations (Supplementary Table A5). This seems justifiable in the background of reduced dispersal due to philopatry in *Pachypus*, which makes artificial over-splitting due to strong differentiation of demes more likely.

It is furthermore to note, that the third gamma prior distribution ($\alpha = 2$, $\beta = 20{,}000$) is very extreme and might – in particular in combination with little informative data – lead to false positive and false negative splitting when applied for θ or τ_0, respectively. It was mainly chosen to test the robustness of results (as discussed above) and should not be thought as indicative for speciation probabilities.

The two entities *P. excavatus* and *P. candidae* were confirmed as valid, the taxonomic identity of these species was deducted from the geographic overlap of our samples with the respective type locality of the taxon. Other available species names so far grouped as synonyms of *P. candidae* were not considered here yet, as they require more detailed investigation of the type specimen material and lectotype designations due to often unprecise geographic collection label data.

A thorough taxonomic treatment of *Pachypus* species will be published separately in a forthcoming paper, as it requires the revision of type specimens, their embedding in the context or morphometric analyses, and detailed diagnosis of diagnostic shape traits including the formal description and illustration of the new species, which is beyond the scope of the present study.

4.4. Sex Biased Dispersal in Species Delimitation

Sex biased dispersal is a common feature (Petit and Excoffier 2009; Dávalos and Russell 2014) and prevalent in birds and mammals (Greenwood 1980). It might therefore affect species delimitation in many groups of organisms. Shortcomings of DNA barcoding (Hebert et al. 2003b) and the use of fragments of mitochondrial DNA for issues of species delimitation are basically well known (Meyer and Paulay 2005; Rubinoff et al. 2006; Roe and Sperling 2007; Seberg and Petersen 2009; Dupuis et al. 2012; Meik et al. 2015; Mallo and Posada 2016) and were reconfirmed here employing a very extreme case. Results largely confirmed the objections of Lohse (2009) in regard of sampling issues, which affect all methods of DNA-based species delimitation and identification. Besides vastly over-estimating the number of species, query sequences would likely not have been associated to cataloged ones if not sampled sufficiently dense or at exactly the same locality.

Our study revealed that also sophisticated validation approaches (with its various algorithms) are influenced by sex-biased dispersal and linked issues such as variation of N_e: all employed multi-species coalescent based methods supported exceedingly high species numbers (> 40) if all molecular data was used. The reduction of gene flow among populations of a species for whatever reason, violates the assumption of panmixia that is made by neutral coalescence models and consequently population boundaries might be mistaken for species boundaries (Hey 2009; O'Meara 2010; Niemiller et al. 2012; Hedin et al. 2015). Implicit modeling of reduced gene-flow and introgression might solve the problems in future models. An alternative might be seen in the collapse height (ε) in STACEY. It describes the expected number of mutations separating two species, i.e., MINCs were collapsed to one entity if separated by less than $1/2\varepsilon$ mutations (Jones et al. 2015). Thus, for $\varepsilon = 10^{-4}$, even species with only one mutation per 5,000 sites can be separated (Jones et al. 2015). If deep coalescences are expected, as for instance in cases of female philopatry or sessile species, an increase of ε may reflect prior knowledge of genetic structure and might thus be justifiable. However, very high values of ε drive the method ad absurdum since any lumping could be inferred arbitrarily. In *Pachypus* even $\varepsilon = 0.01$ resulted in little lumping (Supplementary Fig. A6), which illustrates the impact of the N_e issue but also that even sophisticated multi-species coalescent delimitation methods require "arbitrary" (prior) choices that have a huge impact of the so far widely believed "objective" species delimitation.

The great value of nuclear DNA for species delimitation in general is well recognized (Monaghan et al. 2005; Shaffer and Thomson 2007; Sonnenberg et al. 2007; Wiens 2007; Monaghan et al. 2009). In the case of sex-biased dispersal like in *Pachypus*, mixture of autosomal loci (i.e., most nuclear DNA) is expected to be less or not at all affected (Prugnolle and Meeus 2002) and renders them particularly useful. However, apart from the problems related to the size of N_e and its variation among demes and species (see above), species delimitation in *Pachypus* based on nuclear DNA might also be biased by other factors. The employed nuclear data might lack sufficient structure of variation for *de novo* approaches to work properly (particularly when applied on single markers). ABGD inferred either nearly every (ArgK) or every second (28S) haplotype as separate entity or collapsed all specimens into one putative species. Likewise, TCS inferred only one species with the commonly used 95% thresh-

old. In addition, sex-biased dispersal might indirectly also influence nuclear loci: in most cases of moderate sex-biased dispersal with a more or less similar male-female distribution, strong restrictions of female dispersal activity might have no impact on male dispersal. However, this might change in case of absence of females in large areas between populations, which could strongly reduce effective male dispersal (due to low or entirely missing reproduction success of males at higher distances from the population of origin). Even in case of potential very strict female philopatry and potential extremely strong male dispersal, a significant (although little) degree of nuclear genetic differentiation was observed in per se vagile bats (Kerth et al. 2002). Therefore, potentially unknown mechanisms likely appear to foster the differentiation also in nuclear loci – a pattern that might be exaggerated in *Pachypus* by the incapacity of females to disperse. Also morphological evidence supports this idea by not only showing distinct morphological divergence between putative species units but also between most MINCs and even at the level of local populations (Supplementary Fig. A12).

4.5. Conclusions

More than 150 years ago, Darwin (1859) mentioned the difficulty of a proper universal species definition. A multitude of definitions has been developed since (Simpson 1962; Valen 1976; Paterson 1985; Cracraft 1989; Templeton 1989; Mallet 1995; Mayr 1995), but controversy remained about the "right" concept which often hampered an universal and objective way of species delimitation. With the recent development of diverse methods for species delimitation using DNA and also other lines of evidence, this discussion keeps currency. Species concepts are often connected to particular mechanisms of achieving and maintaining genetic isolation between incipient species. As underlined by De Queiroz (2007), it is possible and also important to distinguish species delimitation from species definition or concept. To know the particular mechanism that led to isolation is not necessary to delimit a species but its delimitation should be based on detecting the ultimate outcome of speciation – genetic isolation on an evolutionary timescale (Rannala 2015). If species are viewed as independently evolving metapopulations (De Queiroz 2007; Yang and Rannala 2014) – a central property shared by all species concepts (De Queiroz 2007) – genetic data should be

expected to fit a species tree with the gene tree distributions described using the multispecies coalescent model (Rannala and Yang 2003; Yang and Rannala 2014). Modeling the immediate cease of gene flow after speciation, BPP reflects ideas of the biological species concept (Mayr 1995; Yang and Rannala 2010) although its probabilistic nature allows for a certain deviance from the underlying model.

Multi-locus species delimitation (Yang and Rannala 2010) in an integrative framework of population genetic and phylogenetic methods should be less sensitive to incomplete lineage sorting, trans-species polymorphism, hybridization, introgression (Leliaert et al. 2014), and varying N_e (Yang and Rannala 2016) than single-locus approaches and is increasingly used in species delimitation studies. Our results illustrate the advantage of an integrative approach to species delimitation with additional supporting data (morphology, phylogeography, ecology, development, or behavior) that are either used in simultaneous analyses or by reciprocal illumination (Henning 1966), to assess competing hypotheses for accurate species delimitation in many cases (Dayrat 2005; Ross et al. 2010; Sauer and Hausdorf 2012; Dávalos and Russell 2014; Solís-Lemus et al. 2015; Eberle et al. 2016) and variable effective population size (Ahrens et al. 2016) in result of various natural phenomena (e.g., sex-biased dispersal, geography) are inherent measures of living beings, the crucial questions is whether "subjective tuning" and a "qualitative judgement" by the authors can be avoided by the availability of more data (McFadden et al. 2011; Dowton et al. 2014; Coissac et al. 2016) as suggested by current multi-species coalescent simulations of "ideal" data (Zhang et al. 2011; Camargo et al. 2012).

The vastly varying results, depending on the choice of which method or data are used (Figs. II.3, II.5), illustrated also how complex it is to unify the different species concepts and delimitation algorithms within an integrative species delimitation approach towards a conclusive and robust species hypothesis. Deciding which method is the most accurate for a given problem is not trivial. "Statistical democracy" of all employed methods (Carstens et al. 2013b; Rannala 2015), i.e., trusting those delimitation that are found by the majority of methods, might suffer from a great loss of decisive power. It is thus preferable to choose well performing methods over poor performing ones (Rannala 2015). By its ability to simultaneously evaluate molecular and morphometric data and by making explicit and testable model assumptions, iBPP renders as the superior method in the present study.

Our results emphasize that reliance on single sources of data can be highly misleading in species delimitation analyses as well as using as much as possible data can be because of the introduction of more noise (Edwards and Knowles 2014). They once again underline the need of a critical discussion of species delimitation results (Carstens et al. 2013a; Rannala 2015), taking into account the biological background of the focal group since significant bias may exist, e.g., in cases like self-fertilizing (Prévot et al. 2013) or sex-biased dispersal (this study). In particular sex-biased dispersal is common in a variety of organisms and might lead to huge hidden errors as in our case or high uncertainty. Therefore, sensible choice of the data and priors, that reflect knowledge on the biology of the group of interest (Rannala 2015), was necessary in the present study, thus making qualitative and subjective judgments of the researcher yet unavoidable (Sites and Marshall 2004).

References

Adams, D. C. and E. Otarola-Castillo (2013). geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution* 4, 393–399.

Adams, D. C. and F. J. Rohlf (2000). Ecological character displacement in *Plethodon*: biomechanical differences found from a geometric morphometric study. *Proceedings of the National Academy of Sciences of the United States of America* 97, 4106–11.

Ahrens, D., T. Fujisawa, H.-J. Krammer, J. Eberle, S. Fabrizi, and A. P. Vogler (2016). Rarity and Incomplete Sampling in DNA-based Species Delimitation. *Systematic Biology* 65, 478–494.

Ahrens, D., M. T. Monaghan, and A. P. Vogler (2007). DNA-based taxonomy for associating adults and larvae in multi-species assemblages of chafers (Coleoptera: Scarabaeidae). *Molecular phylogenetics and evolution* 44, 436–49. DOI: 10.1016/j.ympev.2007.02.024.

Ahrens, D., J. Schwarzer, and A. P. Vogler (2014). The evolution of scarab beetles tracks the sequential rise of angiosperms and mammals. *Proceedings of the Royal Society B: Biological Sciences* 281, 20141470–20141470. DOI: 10.1098/rspb.2014.1470.

Ahrens, D. and A. P. Vogler (2008). Towards the phylogeny of chafers (Sericini): analysis of alignment-variable sequences and the evolution of segment numbers in the antennal club. *Molecular phylogenetics and evolution* 47, 783–98. DOI: 10.1016/j.ympev.2008.02.010.

Aktas, C. (2015). *haplotypes: Haplotype Inference and Statistical Analysis of Genetic Variation*, https://cran.r–project.org/package=haplotypes.

Andriollo, T., Y. Naciri, and M. Ruedi (2015). Two mitochondrial barcodes for one biological species: The case of European Kuhl's pipistrelles (chiroptera). *PLoS ONE* 10, 1–18. DOI: 10.1371/journal.pone.0134881.

Andújar, C., P. Arribas, C. Ruiz, J. Serrano, and J. Gómez-Zurita (2014). Integration of conflict into integrative taxonomy: fitting hybridization in species delimitation of *Mesocarabus* (Coleoptera: Carabidae). *Molecular Ecology* 23, 4344–4361. DOI: 10.1111/mec.12793.

Arnone, M. and I. Sparacio (1990). Il *Pachypus caesus* Erichson 1840: brevi note sulla biologia e la distribuzione in Sicilia (Coleoptera: Scarabaeoidea). *Naturalista Siciliano* 14, 63–71.

Arthofer, W., H. Rauch, B. Thaler-Knoflach, K. Moder, C. Muster, B. C. Schlick-Steiner, and F. M. Steiner (2013). How diverse is *Mitopus morio*? Integrative taxonomy detects cryptic species in a small-scale sample of a widespread harvestman. *Molecular ecology*. DOI: 10.1111/mec.12340.

Astrin, J. J., P. E. Stüben, B. Misof, J. W. Wägele, F. Gimnich, M. J. Raupach, and D. Ahrens (2012). Exploring diversity in cryptorhynchine weevils (Coleoptera) using distance-, character- and tree-based species delineation. *Molecular Phylogenetics and Evolution* 63, 1–14. DOI: 10.1016/j.ympev.2011.11.018.

Baldwin, B. G. (1997). "Adaptive radiation of the Hawaiian silversword alliance: congruence and conflict of phylogenetic evidence from molecular and non-molecular investigations." In: *Molecular evolution and adaptive radiation*. Ed. by T. J. Givnish and K. Systma. Cambridge, U.K.: Cambridge University Press, pp. 103–028.

Balke, M., L. Hendrich, E. F. A. Toussaint, X. Zhou, T. von Rintelen, and M. de Bruyn (2013). Suggestions for a molecular biodiversity assessment of South East Asian freshwater invertebrates. Lessons from the megadiverse beetles (Coleoptera). *Journal of Limnology* 72, 61–68. DOI: 10.4081/jlimnol.2013.s2.e4.

Ballerio, A., A. Rey, M. Uliana, M. Rastelli, S. Rastelli, M. Romano, and L. Colacurcio (2014). Coleotteri Scarabeoidei d'Italia. http://www.societaentomologicaitaliana.it/ColeotteriScarabeoidead'Italia2014/scarabeidi/home.htm. (Accessed on Mar. 22, 2017).

Baraud, J. (1985). *Tropinota (Epicometis) villiersi* nouvelle espèce du Moyen-orient (Coleoptera, Scarabaeoidea, Cetoniidae). *Revue française d'entomologie* 6, 61–63.

Baraud, J. (1992). *Coléoptères Scarabaeoidea d'Europe – Faune de France*. Paris ;Lyon: Fédération française des Sociétés de Sciences naturelles, Paris, Lyon, p. 856.

Barley, A. J., J. White, A. C. Diesmos, and R. M. Brown (2013). The challenge of species delimitation at the extremes: diversification without morphological change in philippine sun skinks. *Evolution* 67, 3556–3572. DOI: 10.1111/evo.12219.

Barrowclough, G. F. G. F. (1980). Gene flow, effective population sizes, and genetic variance components in birds. *Evolution* 34, 789–798.

Blankers, T., D. C. Adams, and J. J. Wiens (2012). Ecological radiation with limited morphological diversification in salamanders. *Journal of evolutionary biology* 25, 634–46. DOI: 10.1111/j.1420-9101.2012.02458.x.

Bond, J. E. and A. K. Stockman (2008). An integrative method for delimiting cohesion species: finding the population-species interface in a group of Californian trapdoor spiders with extreme genetic divergence and geographic structuring. *Systematic biology* 57, 628–646. DOI: 10.1080/10635150802302443.

Bookstein, F. L., A. P. Streissguth, P. D. Sampson, P. D. Connor, and H. M. Barr (2002). Corpus callosum shape and neuropsychological deficits in adult males with heavy fetal alcohol exposure. *Neuroimage* 15, 233–251. DOI: 10.1006/nimg.2001.0977[doi]\rS1053811901909772[pii].

Bouckaert, R. R. (2010). DensiTree: Making sense of sets of phylogenetic trees. *Bioinformatics* 26, 1372–1373. DOI: 10.1093/bioinformatics/btq110.

Boukaert, R., J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, A. Rambaut, and A. J. Drummond (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLOS computational biology* 10, 1–6. DOI: 10.1371/journal.pcbi.1003537.

Britton, T., C. L. Anderson, D. Jacquet, S. Lundqvist, and K. Bremer (2007). Estimating divergence times in large phylogenetic trees. *Systematic Biology* 56, 741–752. DOI: 10.1080/10635150701613783.

Britton, T., B. Oxelman, A. Vinnersten, and K. Bremer (2002). Phylogenetic dating with confidence intervals using mean path lengths. *Molecular phylogenetics and evolution* 24, 58–65.

Burnaby, T. P. (1966). Growth-invariant discriminant functions and generalized distances. *Biometrics* 22, 96–110.

Burnham, K. P. and D. R. Anderson (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Ed. by K. P. Burnham and D. R. Anderson. 2nd. New York, NY: Springer New York, p. 488. DOI: 10.1007/b97636.

Butcher, B. A., M. A. Smith, M. J. Sharkey, and D. L. J. Quicke (2012). A turbo-taxonomic study of Thai *Aleiodes (Aleiodes)* and *Aleiodes (Arcaleiodes)* (Hymenoptera: Braconidae: Rogadinae) based largely on COI barcoded specimens, with rapid descriptions of 179 new species. *Zootaxa* 232, 1–232.

Camargo, A., M. Morando, L. J. Avila, and J. W. Sites (2012). Species delimitation with ABC and other coalescent-based methods: a test of accuracy with simulations and an empirical example with lizards of the *Liolaemus darwinii* complex (Squamata: Liolaemidae). *Evolution* 66, 2834–2849. DOI: 10.5061/dryad.4409k652.

Cardoso, A., A. Serrano, and A. P. Vogler (2009). Morphological and molecular variation in tiger beetles of the *Cicindela hybrida* complex: Is an "integrative taxonomy' possible? *Molecular Ecology* 18, 648–664. DOI: 10.1111/j.1365-294X.2008.04048.x.

Carstens, B. C., R. S. Brennan, V. Chua, C. V. Duffie, M. G. Harvey, R. a. Koch, C. D. McMahan, B. J. Nelson, C. E. Newman, J. D. Satler, G. Seeholzer, K. Posbic, D. C. Tank, and J. Sullivan (2013a). Model selection as a tool for phylogeographic inference: An example from the willow *Salix melanopsis*. *Molecular Ecology* 22, 4014–4028. DOI: 10.1111/mec.12347.

Carstens, B. C. and T. A. Dewey (2010). Species Delimitation Using a Combined Coalescent and Information-Theoretic Approach: An Example from North American *Myotis* Bats. *Systematic Biology* 59, 400–414. DOI: 10.1093/sysbio/syq024.

Carstens, B. C., T. A. Pelletier, N. M. Reid, and J. D. Satler (2013b). How to fail at species delimitation. *Molecular ecology* 22, 4369–4383. DOI: 10.1111/mec.12413.

Castella, V., M. Ruedi, and L. Excoffier (2001). Contrasted patterns of mitochondrial and nuclear structure among nursery colonies of the bat *Myotis myotis*. *Journal of Evolutionary Biology* 14, 708–720. DOI: 10.1046/j.1420-9101.2001.00331.x.

Cicconardi, F., P. P. Fanciulli, and B. C. Emerson (2013). Collembola, the biological species concept and the underestimation of global species richness. *Molecular Ecology* 22, 5382–5396. DOI: 10.1111/mec.12472.

Clement, M., D. Posada, and K. A. Crandall (2000). TCS: a computer program to estimate gene genealogies. *Molecular Ecology* 9, 1657–1659. DOI: 10.1046/j.1365-294x.2000.01020.x.

Coissac, E., P. M. Hollingsworth, S. Lavergne, and P. Taberlet (2016). From barcodes to genomes: extending the concept of DNA barcoding. *Molecular Ecology* 25, 1423–1428. DOI: 10.1111/mec.13549.

Cracraft, J. (1989). "Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation." In: *Speciation and its consequences*. Ed. by D. Otte and J. A. Endler. Sunderland, MA: Sinauer Associates, Sunderland, MA, pp. 28–59.

Crovetti, A. (1969). Contributo alla conoscenza dei Coleotteri Scarabeidi I. Il genere *Pachypus* Serville (Coleoptera, Scarabaeidae, Pachypodinae). *Bollettino di Zoologia Agraria e di Bachicoltura* 9, 133–188.

Currat, M., M. Ruedi, R. J. Petit, and L. Excoffier (2008). The hidden side of invasions: Massive introgression by local genes. *Evolution* 62, 1908–1920. DOI: 10.1111/j.1558-5646.2008.00413.x.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection*. London: John Murray, London, p. 502. DOI: 10.1016/S0262-4079(09)60380-8.

Dávalos, L. M. and A. L. Russell (2014). Sex-biased dispersal produces high error rates in mitochondrial distance-based and tree-based species delimitation. *Journal of Mammalogy* 95, 781–791. DOI: 10.1644/14-MAMM-A-107.

Dayrat, B. (2005). Towards integrative taxonomy. *Biological Journal of the Linnean Society* 85, 407–415.

De Queiroz, K. (2007). Species Concepts and Species Delimitation. *Systematic Biology* 56, 879–886. DOI: 10.1080/10635150701701083.

Degnan, J. H. and N. A. Rosenberg (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* 24, 332–340. DOI: 10.1016/j.tree.2009.01.009.

Dellicour, S. and J.-F. Flot (2015). Delimiting Species-Poor Data Sets using Single Molecular Markers: A Study of Barcode Gaps, Haplowebs and GMYC. *Systematic Biology* 64, 900–908. DOI: 10.1093/sysbio/syu130.

Dinno, A. (2009). Implementing Horn's parallel analysis for principal component analysis and factor analysis. *The Stata Journal* 9, 291–298.

Dowton, M., K. Meiklejohn, S. L. Cameron, and J. Wallman (2014). A Preliminary Framework for DNA Barcoding, Incorporating the Multispecies Coalescent. *Systematic Biology* 63, 639–644. DOI: 10.1093/sysbio/syu028.

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut (2006). Relaxed Phylogenetics and Dating with Confidence. *PLoS Biology* 4. Ed. by D. Penny, e88. DOI: 10.1371/journal.pbio.0040088.

Dupuis, J. R., A. D. Roe, and F. A. H. Sperling (2012). Multi-locus species delimitation in closely related animals and fungi: One marker is not enough. *Molecular Ecology* 21, 4422–4436. DOI: 10.1111/j.1365-294X.2012.05642.x.

Eberle, J., R. Tänzler, and A. Riedel (2012). Revision and phylogenetic analysis of the Papuan weevil genus *Thyestetha* Pascoe (Coleoptera, Curculionidae, Cryptorhynchinae). *Zootaxa* 3355, 1–28.

Eberle, J., R. C. M. Warnock, and D. Ahrens (2016). Bayesian species delimitation in *Pleophylla* chafers (Coleoptera) – the importance of prior choice and morphology. *BMC Evolutionary Biology* 16, 94. DOI: 10.1186/s12862-016-0659-3.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32, 1792–1797. DOI: 10.1093/nar/gkh340.

Edwards, D. L. and L. L. Knowles (2014). Species detection and individual assignment in species delimitation: can integrative data increase efficacy? *Proceedings of the Royal Society B: Biological Sciences* 281, 20132765. DOI: 10.1098/rspb.2013.2765.

Ence, D. D. and B. C. Carstens (2011). SpedeSTEM: a rapid and accurate method for species delimitation. *Molecular ecology resources* 11, 473–80. DOI: 10.1111/j.1755-0998.2010.02947.x.

Erichson, W. F. (1840). Die Pachypoden, eine kleine Gruppe aus der Familie der Melolonthen. Entomographien, Untersuchungen in dem Gebiete der Entomologie mit besonderer Benutzung der Königl. Sammlung in Berlin. *Morin* 1, 29–43.

Esselstyn, J. A., B. J. Evans, J. L. Sedlock, F. a. Anwarali Khan, and L. R. Heaney (2012). Single-locus species delimitation: a test of the mixed Yule-coalescent model, with an

empirical application to Philippine round-leaf bats. *Proceedings of the Royal Society B: Biological Sciences* 279, 3678–3686. DOI: 10.1098/rspb.2012.0705.

Ezard, T., T. Fujisawa, and T. Barraclough (2014). *splits: SPecies' LImits by Threshold Statistics*.

Fabricius, J. C. (1792). *Entomologia systematica emendata et aucta. Secundum Classes, Ordines, Genera, Species adjectis Synonimis, Locis, Observationibus, Descriptionibus*. Hafniae, p. 538.

Ferrario, V. F., C. Sforza, J. H. Schmitz, A. J. Miani, and G. Taroni (1995). Fourier analysis of human soft tissue facial shape: sex differences in normal adults. *J. Anat.* 187, 593–602.

Fraley, C. and A. E. Raftery (2002). Model-Based Clustering, Discriminant Analysis, and Density Estimation. *Journal of the American Statistical Association* 97, 611–631. DOI: 10.1198/016214502760047131.

Fraley, C., A. E. Raftery, T. B. Murphy, L. Scrucca, M. Brendan, and L. Scrucca (2012). mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. *Technical Report 597, University of Washington*, 1–50.

Freeland, J. R. and P. T. Boag (1999). The mitochondrial and nuclear genetic homogeneity of the phenotypically diverse Darwin's ground finches. *Evolution* 53, 1553–1563.

Fujisawa, T., A. Aswad, and T. G. Barraclough (2016). A rapid and scalable method for multilocus species delimitation using Bayesian model comparison and rooted triplets. *Systematic Biology advance access*.

Fujisawa, T. and T. G. Barraclough (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, 707–724. DOI: 10.1093/sysbio/syt033.

Gewin, V. (2002). All living things, online. *Nature* 418, 362–363. DOI: 10.1038/418362a.

Godfray, H. C. J. (2002). Challenges for taxonomy. *Nature* 417, 17–19. DOI: 10.1038/417017a.

Greenwood, P. J. (1980). Mating systems, philopatry and dispersal in birds and mammals. *Animal Behaviour* 28, 1140–1162. DOI: 10.1016/S0003-3472(80)80103-5.

Guerlach, G., E. Bazzato, and D. Cillo (2013). Description d'une nouvelle espèce de *Pachypus* Dejean, 1821: *Pachypus sardiniensis* n.sp. *Lambillionea* 113, 73–76.

Gunz, P. and P. Mitteroecker (2013). Semilandmarks: A method for quantifying curves and surfaces. *Hystrix* 24, 103–109. DOI: 10.4404/hystrix-24.1-6292.

Hambäck, P. A., E. Weingartner, L. Ericson, L. Fors, A. Cassel-Lundhagen, J. A. Stenberg, and J. Bergsten (2013). Bayesian species delimitation reveals generalist and specialist parasitic wasps on *Galerucella* beetles (Chrysomelidae): sorting by herbivore or plant host. *BMC Evolutionary Biology* 13, 92. DOI: 10.1186/1471-2148-13-92.

Harrington, R. C. and T. J. Near (2012). Phylogenetic and coalescent strategies of species delimitation in snubnose darters (Percidae: *Etheostoma*). *Systematic Biology* 61, 63–79. DOI: 10.1093/sysbio/syr077.

Hart, M. W. and J. Sunday (2007). Things fall apart: biological species form unconnected parsimony networks. *Biology Letters* 3, 509–512. DOI: 10.1098/rsbl.2007.0307.

Hebert, P., S. Ratnasingham, and J. DeWaard (2003a). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London* 270, 96–99. DOI: 10.1098/rsbl.2003.0025.

Hebert, P. D. N., A. Cywinska, S. L. Ball, and J. R. DeWaard (2003b). Biological identifications through DNA barcodes. *Proceedings. Biological sciences / The Royal Society* 270, 313–21. DOI: 10.1098/rspb.2002.2218.

Hebert, P. D. N. and T. R. Gregory (2005). The promise of DNA barcoding for taxonomy. *Systematic Biology* 54, 852.

Hedin, M., D. Carlson, and F. Coyle (2015). Sky island diversification meets the multispecies coalescent - Divergence in the spruce-fir moss spider (*Microhexura montivaga*, Araneae, Mygalomorphae) on the highest peaks of southern Appalachia. *Molecular Ecology* 24, 3467–3484. DOI: 10.1111/mec.13248.

Henning, W. (1966). *Phylogenetic systematics*. Urbana: University of Illinois Press, pp. III+263.

Hey, J. (2009). "On the arbitrary identification of real species." In: *Speciation and Patterns of Diversity*. Ed. by R. K. Butlin, J. R. Bridle, and D. Schluter. Cambridge, UK: Cambridge University Press, Cambridge, UK., pp. 15–28.

Ho, S. Y. W., M. J. Phillips, A. J. Drummond, and A. Cooper (2005). Accuracy of Rate Estimation Using Relaxed-Clock Models with a Critical Focus on the Early Metazoan Radiation. *Molecular Biology and Evolution* 22, 1355–1363. DOI: 10.1093/molbev/msi125.

Holland, B. S. and M. G. Hadfield (2004). Origin and diversification of the endemic Hawaiian tree snails (Achatinellidae: Achatinellinae) based on molecular evidence. *Molecular Phylogenetics and Evolution* 32, 588–600. DOI: 10.1016/j.ympev.2004.01.003.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika* 30, 179–185. DOI: 10.1007/BF02289447.

Huang, J.-P. and L. L. Knowles (2015). The species versus subspecies conundrum: quantitative delimitation from integrating multiple data types within a single Bayesian approach in Hercules beetles. *Systematic Biology*, accepted.

Jockusch, E. L. and D. B. Wake (2002). Falling apart and merging: diversification of slender salamanders (Plethodontidae: *Batrachoseps*) in the American West. *Biological Journal of the Linnean Society* 76, 361–391. DOI: 10.1046/j.1095-8312.2002.00071.x.

Jolicoeur, P. (1963). 193. Note: The Multivariate Generalization of the Allometry Equation. *Biometrics*, 497–499.

Jones, G. (2014). *STACEY: species delimitation and phylogeny estimation under the multispecies coalescent*. Tech. rep. DOI: 10.1101/010199.

Jones, G., Z. Aydin, and B. Oxelman (2015). DISSECT: an assignment-free Bayesian discovery method for species delimitation under the multispecies coalescent. *Bioinformatics* 31, 991–998. DOI: 10.1093/bioinformatics/btu770.

Katoh, K., K.-i. Kuma, H. Toh, and T. Miyata (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic acids research* 33, 511–8. DOI: 10.1093/nar/gki198.

Katoh, K., K. Misawa, K.-i. Kuma, and T. Miyata (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research* 30, 3059–66.

Katouzian, A.-R., A. Sari, J. N. Macher, M. Weiss, A. Saboori, F. Leese, and A. M. Weigand (2016). Drastic underestimation of amphipod biodiversity in the endangered Irano-Anatolian and Caucasus biodiversity hotspots. *Scientific Reports* 6, 22507. DOI: 10.1038/srep22507.

Kerth, G., F. Mayer, and E. Petit (2002). Extreme sex-biased dispersal in the communally breeding, nonmigratory Bechstein's bat *(Myotis bechsteinii)*. *Molecular Ecology* 11, 1491–1498. DOI: 10.1046/j.1365-294X.2002.01528.x.

Kozak, K. H., D. W. Weisrock, and A. Larson (2006). Rapid lineage accumulation in a non-adaptive radiation: phylogenetic analysis of diversification rates in eastern North American woodland salamanders (Plethodontidae: *Plethodon*). *Proceedings. Biological sciences / The Royal Society* 273, 539–546. DOI: 10.1098/rspb.2005.3326.

Krell, F. T. (2000). The fossil record of Mesozoic and Tertiary Scarabaeoidea (Coleoptera: Polyphaga). *Invertebrate Taxonomy* 14, 871–905. DOI: 10.1071/IT00031.

Leaché, A. D. and M. K. Fujita (2010). Bayesian species delimitation in West African forest geckos *(Hemidactylus fasciatus)*. *Proceedings of the Royal Society B: Biological Sciences* 277, 3071–3077. DOI: 10.1098/rspb.2010.0662.

Lim, G. S., M. Balke, and R. Meier (2012). Determining species boundaries in a world full of rarity: singletons, species delimitation methods. *Systematic biology* 61, 165–169. DOI: 10.1093/sysbio/syr030.

Liu, K., S. Raghavan, S. Nelesen, C. R. Linder, and T. Warnow (2009). Rapid and Accurate Large-Scale Coestimation of Sequence Alignments and Phylogenetic Trees. *Science* 324, 1561–1564. DOI: 10.1126/science.1171243.

Liu, K., T. J. Warnow, M. T. Holder, S. M. Nelesen, J. Yu, A. P. Stamatakis, and C. R. Linder (2012). SATe-II: Very Fast and Accurate Simultaneous Estimation of Multiple Sequence Alignments and Phylogenetic Trees. *Systematic Biology* 61, 90–106. DOI: 10.1093/sysbio/syr095.

Lohse, K. (2009). Can mtDNA Barcodes Be Used to Delimit Species? A Response to Pons et al. (2006). *Systematic Biology* 58, 439–442. DOI: 10.1093/sysbio/syp039.

Luigioni, P. (1923). Le specie et le varieta del gen. *Pachypus* Serv. in Italia. *Memorie della Società Entomologica Italiana* 2, 50–64.

Lyrholm, T., O. Leimar, B. Johanneson, and U. Gyllensten (1999). Sex-biased dispersal in sperm whales: contrasting mitochondrial and nuclear genetic structure of global populations. *Proceedings of The Royal Society - Biological sciences* 266, 347–54. DOI: 10.1098/rspb.1999.0644.

Mallet, J. (1995). A species definition for the modern synthesis. *Trends in Ecology & Evolution* 10, 294–299.

Mallo, D. and D. Posada (2016). Multilocus inference of species trees and DNA barcoding. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 371. DOI: 10.1098/rstb.2015.0335.

Mayr, E. (1995). "Species, classification, and evolution." In: *Biodiversity and Evolution*. Ed. by R. Arai, M. Kato, and Y. Doi. Tokyo: National Science Museum Foundation, Tokyo, pp. 3–12.

McFadden, C. S., Y. Benayahu, E. Pante, N. Thoma, Jana, P. A. Nevarez, and S. C. France (2011). Limitations of mitochondrial gene barcoding in Octocorallia. *Molecular Ecology Resources* 11, 19–31. DOI: 10.1111/j.1755-0998.2010.02875.x.

Meier, R., K. Shiyang, G. Vaidya, and P. Ng (2006). DNA Barcoding and Taxonomy in Diptera: A Tale of High Intraspecific Variability and Low Identification Success. *Systematic Biology* 55, 715–728. DOI: 10.1080/10635150600969864.

Meik, J. M., J. W. Streicher, a. M. Lawing, and O. Flores-villela (2015). Limitations of Climatic Data for Inferring Species Boundaries: Insights from Speckled Rattlesnakes. *PLOS one* 10, e0131435. DOI: 10.5061/dryad.2rg12.Funding.

Melnick, D. and G. Hoelzer (1992). Differences in male and female macaque dispersal lead to contrasting distribution of nuclear and mitochondrial DNA variation. *International Journal of Primatology* 13, 379–393. DOI: 10.1007/BF02547824.

Meyer, C. P. and G. Paulay (2005). DNA barcoding: Error rates based on comprehensive sampling. *PLoS Biology* 3, 1–10. DOI: 10.1371/journal.pbio.0030422.

Minh, B. Q., M. A. T. Nguyen, and A. von Haeseler (2013). Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution* 30, 1188–1195. DOI: 10.1093/molbev/mst024.

Monaghan, M. T., M. Balke, J. Pons, and A. P. Vogler (2006). Beyond barcodes: complex DNA taxonomy of a South Pacific Island radiation. *Proceedings of the Royal Society B: Biological Sciences* 273, 887.

Monaghan, M. T., M. Balke, T. R. Gregory, and A. P. Vogler (2005). DNA-based species delineation in tropical beetles using mitochondrial and nuclear markers. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 360, 1925–1933. DOI: 10.1098/rstb.2005.1724.

Monaghan, M. T., D. J. G. Inward, T. Hunt, and A. P. Vogler (2007). A molecular phylogenetic analysis of the Scarabaeinae (dung beetles). *Molecular Phylogenetics and Evolution* 45, 674–692. DOI: 10.1016/j.ympev.2007.06.009.

Monaghan, M. T., R. Wild, M. Elliot, T. Fujisawa, M. Balke, D. J. G. Inward, D. C. Lees, R. Ranaivosolo, P. Eggleton, T. G. Barraclough, and A. P. Vogler (2009). Accelerated species inventory on Madagascar using coalescent-based models of species delineation. *Systematic biology* 58, 298–311. DOI: 10.1093/sysbio/syp027.

Moran, P. and I. Kornfield (1993). Retention of an ancestral polymorphism in the mbuna species flock (Teleostei: Cichlidae) of Lake Malawi. *Molecular Biology and Evolution* 10, 1015–1029.

Nguyen, L. T., H. A. Schmidt, A. Von Haeseler, and B. Q. Minh (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32, 268–274. DOI: 10.1093/molbev/msu300.

Niemiller, M. L., T. J. Near, and B. M. Fitzpatrick (2012). Delimiting Species Using Multilocus Data: Diagnosing Cryptic Diversity in the Southern Cavefish, *Typhlichthys subterraneus* (Teleostei: Amblyopsidae). *Evolution* 66, 846–866. DOI: 10.1111/j.1558-5646.2011.01480.x.

Nylander, J. A. A. (2014). *burntrees*. https://github.com/nylander/Burntrees/. (Accessed on July 10, 2014).

Olave, M., E. Solà, and L. L. Knowles (2014). Upstream Analyses Create Problems with DNA-Based Species Delimitation. *Systematic Biology* 63, 263–271. DOI: 10.1093/sysbio/syt106.

Oliver, P. M., M. Adams, M. S. Lee, M. N. Hutchinson, and P. Doughty (2009). Cryptic diversity in vertebrates: molecular data double estimates of species diversity in a radiation of Australian lizards (*Diplodactylus*, Gekkota). *Proceedings of the Royal Society B: Biological Sciences* 276, 2001–2007. DOI: 10.1098/rspb.2008.1881.

Olivier, G. A. (1789). *Entomologie ou Histoire naturelle des insectes, avec leurs caractères génériques et spécifiques, leur description, leur synonumie et leur figure enluminee ... Coléoptères Tomes I-V*. Vol. 1. Paris :Imp. Baudouin, p. 488.

O'Meara, B. C. (2010). New heuristic methods for joint species delimitation and species tree inference. *Systematic Biology* 59, 59–73. DOI: 10.1093/sysbio/syp077.

Padial, J. M., S. Castroviejo-Fisher, J. Köhler, C. Vilà, J. C. Chaparro, and I. De la Riva (2009). Deciphering the products of evolution at the species level: the need for an integrative taxonomy. *Zoologica Scripta* 38, 431–447. DOI: 10.1111/j.1463-6409.2008.00381.x.

Padial, J. M., A. Miralles, I. De la Riva, and M. Vences (2010). The integrative future of taxonomy. *Frontiers in zoology* 7, 16. DOI: 10.1186/1742-9994-7-16.

Palstra, F. P. and D. J. Fraser (2012). Effective/census population size ratio estimation: A compendium and appraisal. *Ecology and Evolution* 2, 2357–2365. DOI: 10.1002/ece3.329.

Palumbi, S. R. and C. S. Baker (1994). Contrasting Population Structure from Nuclear Intron Sequences and mtDNA of Humpback Whales. *Molecular Biology and Evolution* 11, 426–435.

Papadopoulou, A., A. Cardoso, and J. Gómez-Zurita (2013). Diversity and diversification of Eumolpinae (Coleoptera: Chrysomelidae) in New Caledonia. *Zoological Journal of the Linnean Society* 168, 473–495. DOI: 10.1111/zoj.12039.

Papadopoulou, A., M. T. Monaghan, T. G. Barraclough, and A. P. Vogler (2009). Sampling Error Does Not Invalidate the Yule-Coalescent Model for Species Delimitation. A Response to Lohse (2009). *Systematic Biology* 58, 442–444. DOI: 10.1093/sysbio/syp038.

Paterson, H. E. H. (1985). "The recognition concept of species." In: *Species and Speciation*. Ed. by E. S. Vrba. Pretoria: Transvaal Museum Monograph No. 4, Pretoria, pp. 21–29.

Perez, S. I., V. Bernal, and P. N. Gonzalez (2006). Differences between sliding semi-landmark methods in geometric morphometrics, with an application to human craniofacial and dental variation. *Journal of Anatomy* 208, 769–784.

Petagna, V. (1787). *Specimen insectorum ulterioris Calabriae*. Francofurti et Moguntiae, Apud Varrentrapp et Wenner, p. 62.

Petit, R. J. and L. Excoffier (2009). Gene flow and species delimitation. *Trends in Ecology and Evolution* 24, 386–393. DOI: 10.1016/j.tree.2009.02.011.

Petren, K., P. R. Grant, B. R. Grant, and L. F. Keller (2005). Comparative landscape genetics and the adaptive radiation of Darwin's finches: The role of peripheral isolation. *Molecular Ecology* 14, 2943–2957. DOI: 10.1111/j.1365-294X.2005.02632.x.

Plummer, M., N. Best, K. Cowles, and K. Vines (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News* 6, 7–11. DOI: 10.1159/000323281.

Pons, J., T. G. Barraclough, J. Gomez-Zurita, A. Cardoso, D. P. Duran, S. Hazell, S. Kamoun, W. D. Sumlin, and A. P. Vogler (2006). Sequence-Based Species Delimitation for the DNA Taxonomy of Undescribed Insects. *Systematic Biology* 55, 595–609. DOI: 10.1080/10635150600852011.

Prévot, V., K. Jordaens, G. Sonet, and T. Backeljau (2013). Exploring Species Level Taxonomy and Species Delimitation Methods in the Facultatively Self-Fertilizing Land Snail Genus *Rumina* (Gastropoda: Pulmonata). *PLoS ONE* 8, e60736. DOI: 10.1371/journal.pone.0060736.

Prugnolle, F. and T. de Meeus (2002). Inferring sex-biased dispersal from population genetic tools: a review. *Heredity* 88, 161–165. DOI: 10.1038/sj/hdy/6800060.

Puillandre, N., A. Lambert, S. Brouillet, and G. Achaz (2012a). ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology* 21, 1864–1877. DOI: 10.1111/j.1365-294X.2011.05239.x.

Puillandre, N., M. V. Modica, Y. Zhang, L. Sirovich, M. C. Boisselier, C. Cruaud, M. Holford, and S. Samadi (2012b). Large-scale species delimitation method for hyperdiverse groups. *Molecular Ecology* 21, 2671–2691. DOI: 10.1111/j.1365-294X.2012.05559.x.

Pyron, R. A., F. W. Hsieh, A. R. Lemmon, E. M. Lemmon, and C. R. Hendry (2016). Integrating phylogenomic and morphological data to assess candidate species-delimitation models in brown and red-bellied snakes *(Storeria)*. *Zoological Journal of the Linnean Society* 177, 937–949. DOI: 10.1111/zoj.12392.

Rambaut, A., M. Suchard, D. Xie, and A. Drummond (2014). Tracer v1.6.

Rannala, B. (2009). The art and science of species delimitation. 61, 846–853. DOI: 10.1093/czoolo/61.5.846.

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

Rannala, B. (2015). The art and science of species delimitation. *Current Zoology* 61, 846–853. DOI: 10.1093/czoolo/61.5.846.

Ratnasingham, S. and P. Hebert (2007). BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes* 7, 355–364. DOI: 10.1111/j.1471-8286.2006.01678.x.

Ratnasingham, S. and P. D. N. Hebert (2013). A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System. *PLoS ONE* 8. DOI: 10.1371/journal.pone.0066213.

Reid, N. M. and B. C. Carstens (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, 196. DOI: 10.1186/1471-2148-12-196.

Reitter, E. (1898). Bestimmungs-Tabelle der Melolonthidae aus der europäischen Fauna und den angrenzenden Ländern. II Theil. Dynastini, Euchirini, Pachypodini, Cetonini, Valgini und Trichiini. *Verhandlungen des Naturforschenden Vereins, Brünn* 37, 21–111.

Riedel, A., D. Daawia, and M. Balke (2010). Deep *cox1* divergence and hyperdiversity of *Trigonopterus* weevils in a New Guinea mountain range (Coleoptera, Curculionidae). *Zoologica Scripta* 39, 63–74. DOI: 10.1111/j.1463-6409.2009.00404.x.

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

Riedel, A., K. Sagata, S. Surbakti, R. Tänzler, and M. Balke (2013b). One hundred and one new species of *Trigonopterus* weevils from New Guinea. *ZooKeys* 280, 1–150. DOI: 10.3897/zookeys.280.3906.

Roe, A. D. and F. A. Sperling (2007). Patterns of evolution of mitochondrial cytochrome *c* oxidase I and II DNA and implications for DNA barcoding. *Molecular Phylogenetics and Evolution* 44, 325–345. DOI: 10.1016/j.ympev.2006.12.005.

Rohlf, F. (2005). *tpsDig, digitize landmarks and outlines*. Department of Ecology and Evolution, State University of New York at Stony Brook.

Ross, H., S. Murugan, and W. L. S. Li (2008). Testing the Reliability of Genetic Methods of Species Identification via Simulation. *Systematic Biology* 57, 216–230. DOI: 10.1080/10635150802032990.

Ross, K. G., D. Gotzek, M. S. Ascunce, and D. D. Shoemaker (2010). Species Delimitation: A Case Study in a Problematic Ant Taxon. *Systematic Biology* 59, 162–184. DOI: 10.1093/sysbio/syp089.

Rubinoff, D., S. Cameron, and K. Will (2006). A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. *The Journal of heredity* 97, 581–94. DOI: 10.1093/jhered/esl036.

Satler, J. D., B. C. Carstens, and M. Hedin (2013). Multilocus species delimitation in a complex of morphologically conserved trapdoor spiders (Mygalomorphae, Antrodiaetidae, *Aliatypus*). *Systematic Biology* 62, 805–823. DOI: 10.1093/sysbio/syt041.

Sauer, J. and B. Hausdorf (2012). A comparison of DNA-based methods for delimiting species in a Cretan land snail radiation reveals shortcomings of exclusively molecular taxonomy. *Cladistics* 28, 300–316. DOI: 10.1111/j.1096-0031.2011.00382.x.

Schlick-Steiner, B. C., F. M. Steiner, B. Seifert, C. Stauffer, E. Christian, and R. H. Crozier (2010). Integrative Taxonomy: A Multisource Approach to Exploring Biodiversity. *Annual Review of Entomology* 55, 421–438. DOI: 10.1146/annurev-ento-112408-085432.

Schoolmeesters, P. (2016). "Scarabs: World Scarabaeidae Database." In: *Species 2000 & ITIS Catalogue of Life*. Ed. by Y. Roskov, L. Abucay, l. T. Orrel, D. Nicolson, T. Kunze, C. Flann, N. Bailly, P. Kirk, T. Bourgoin, R. E. DeWalt, W. Decock, and A. De Wever. 28th July. Species 2000: Naturalis, Leiden, the Netherlands.

Scrucca, L. and A. E. Raftery (2014). clustvarsel: A Package Implementing Variable Selection for Model-based Clustering in R. *Journal of Statistical Software*, pre–print. arXiv: 1411.0606.

Seberg, O. and G. Petersen (2009). How many loci does it take to DNA barcode a crocus? *PloS one* 4, e4598. DOI: 10.1371/journal.pone.0004598.

Serres, M. de (1829). *Géognosie des terrains tertiaires: ou, Tableau des principaux animaux invertébrés des terrains marins tertiaires, du midi de la France*. Montpellier, Paris: Pomathio-Durville, Montpellier, Paris, pp. xcii, 1–277.

Shaffer, H. B. and M. L. McKnight (1996). The Polytypic Species Revisited: Genetic Differentiation and Molecular Phylogenetics of the Tiger Salamander *Ambystoma tigrinum* (Amphibia: Caudata) Complex. *Evolution* 50, 417–433. DOI: 10.2307/2410811.

Shaffer, H. B. and R. C. Thomson (2007). Delimiting Species in Recent Radiations. *Systematic Biology* 56, 896–906. DOI: 10.1080/10635150701772563.

Simon, C., F. Frati, A. 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. *Annals of the Entomological Society of America* 87, 651–701.

Simpson, G. G. (1962). *Principles of Animal Taxonomy*. New York: Columbia University Press, New York, p. 247.

Sites, J. W. and J. C. Marshall (2004). Operational criteria for delimiting species. *Annual review of Ecology, Evolution and Systematics* 35, 199–227. DOI: 10.2307/annurev.ecolsys.35.112202.30000009.

Solís-Lemus, C., L. L. Knowles, and C. Ané (2015). Bayesian species delimitation combining multiple genes and traits in a unified framework. *Evolution* 69, 492–507. DOI: 10.1111/evo.12582.14.

Sonnenberg, R., A. W. Nolte, and D. Tautz (2007). An evaluation of LSU rDNA D1-D2 sequences for their use in species identification. *Frontiers in Zoology* 4, 6. DOI: 10.1186/1742-9994-4-6.

Sparacio, I. (2008). Una nuova specie di *Pachypus* Dejean di Sardegna. *Doriana* 8, 1–13.

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.

— (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxford, England)* 30, 1312–1333. DOI: 10.1093/bioinformatics/btu033.

Talavera, G., V. Dincă, and R. Vila (2013). Factors affecting species delimitations with the GMYC model: insights from a butterfly survey. *Methods in Ecology and Evolution* 4. Ed. by E. Paradis, 1101–1110. DOI: 10.1111/2041-210X.12107.

Tang, C. Q., A. M. Humphreys, D. Fontaneto, and T. G. Barraclough (2014). Effects of phylogenetic reconstruction method on the robustness of species delimitation using single-locus data. *Methods in Ecology and Evolution* 5. Ed. by E. Paradis, 1086–1094. DOI: 10.1111/2041-210X.12246.

Tang, C. Q., F. Leasi, U. Obertegger, A. Kieneke, T. G. Barraclough, and D. Fontaneto (2012). The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. *Proceedings of the National Academy of Sciences* 109, 16208–16212. DOI: 10.1073/pnas.1209160109.

Templeton, A. R., K. A. Crandall, and C. F. Sing (1992). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 132, 619–633. DOI: c:\Docs\Bayes.

Templeton, A. R. (1989). "The meaning of species and speciation: a genetic perspective." In: *Speciation and its consequences*. Ed. by D. Otte and J. A. Endler. Sunderland, MA: Sinauer Associates, Sunderland, MA, pp. 3–27.

— (2001). Using phylogeographic analyses of gene trees to test species status and processes. *Molecular ecology* 10, 779–791.

Thormann, B., D. Ahrens, D. Marín Armijos, M. K. Peters, T. Wagner, and J. W. Wägele (2016). Exploring the Leaf Beetle Fauna (Coleoptera: Chrysomelidae) of an Ecuadorian Mountain Forest Using DNA Barcoding. *Plos One* 11, e0148268. DOI: 10.1371/journal.pone.0148268.

Timmermans, M. J. T. N., S. Dodsworth, C. L. Culverwell, L. Bocak, D. Ahrens, D. T. J. Littlewood, J. Pons, and a. P. Vogler (2010). Why barcode? High-throughput multiplex sequencing of mitochondrial genomes for molecular systematics. *Nucleic acids research* 38, e197. DOI: 10.1093/nar/gkq807.

Valen, L. van (1976). Ecological species, multispecies, and oaks. *Taxon* 25, 233–239. DOI: 10.2307/1219444.

Vogel, L. S. and S. G. Johnson (2008). Estimation of Hybridization and Introgression Frequency in Toads (Genus: *Bufo*) Using DNA Sequence Variation at Mitochondrial and Nuclear Loci. *Journal of Herpetology* 42, 61–75.

Wagner, C. E., I. Keller, S. Wittwer, O. M. Selz, S. Mwaiko, L. Greuter, A. Sivasundar, and O. Seehausen (2013). Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology* 22, 787–798. DOI: 10.1111/mec.12023. arXiv: 1302.3274.

Wake, D. B. (2006). Problems with species: patterns and processes of species formation in salamanders. *Annals of the Missouri Botanical Garden* 93, 8–23. DOI: 10.3417/0026-6493(2006)93[8:PWSPAP]2.0.CO;2.

Whitworth, T. L., R. D. Dawson, H. Magalon, and E. Baudry (2007). DNA barcoding cannot reliably identify species of the blowfly genus *Protocalliphora* (Diptera: Calliphoridae). *Proceedings of the Royal Society B-Biological Sciences* 274, 1731–1739. DOI: 10.1098/rspb.2007.0062.

Wiens, J. J. (2007). Species delimitation: new approaches for discovering diversity. *Systematic biology* 56, 875–878. DOI: 10.1080/10635150701748506.

Wiens, J. J. and T. A. Penkrot (2002). Delimiting Species Using DNA and Morphological Variation and Discordant Species Limits in Spiny Lizards *(Sceloporus)*. *Systematic Biology* 51, 69–91. DOI: 10.1080/106351502753475880.

Wild, A. L. and D. R. Maddison (2008). Evaluating nuclear protein-coding genes for phylogenetic utility in beetles. *Molecular Phylogenetics and Evolution* 48, 877–891. DOI: 10.1016/j.ympev.2008.05.023.

Will, K. W., B. D. Mishler, and Q. D. Wheeler (2005). The perils of DNA barcoding and the need for integrative taxonomy. *Systematic biology* 54, 844–51. DOI: 10.1080/10635150500354878.

Witter, M. S. and G. D. Carr (1988). Adaptive radiation and genetic differentiation in the Hawaiian Silversword Alliance (Compositae: Madiinae). *Evolution* 42, 1278–1287.

Wright, S. (1946). Isolation by distance under diverse systems of mating. *Genetics* 31, 39–59.

Yang, Z. and B. Rannala (2010). Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America* 107, 9264–9269. DOI: 10.1073/pnas.0913022107.

— (2014). Unguided Species Delimitation Using DNA Sequence Data from Multiple Loci. *Molecular Biology and Evolution* 31, 3125–3135. DOI: 10.1093/molbev/msu279.

Yeates, D. K., A. Seago, L. Nelson, S. L. Cameron, L. Joseph, and J. W. H. Trueman (2011). Integrative taxonomy, or iterative taxonomy? *Systematic Entomology* 36, 209–217. DOI: 10.1111/j.1365-3113.2010.00558.x.

Zhang, C., D.-X. Zhang, T. Zhu, and Z. Yang (2011). Evaluation of a Bayesian Coalescent Method of Species Delimitation. *Systematic Biology* 60, 747–761. DOI: 10.1093/sysbio/syr071.

Zhang, J., P. Kapli, P. Pavlidis, and A. Stamatakis (2013). A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* 29, 2869–2876. DOI: 10.1093/bioinformatics/btt499.

Zhang, J. (2014). "Models and Algorithms for Phylogenetic Marker Analysis." PhD thesis. University of Lübeck and Max Planck Institute for Evolutionary Biology.

Zhou, Y. F., R. J. Abbott, Z. Y. Jiang, F. K. Du, R. I. Milne, and J. Q. Liu (2010). Gene flow and species delimitation: a case study of two pine species with overlapping distributions in southeast China. *Evolution* 64, 2342–2352. DOI: 10.1111/j.1558-5646.2010.00988.x.

Chapter III.

Bayesian species delimitation in *Pleophylla* chafers (Coleoptera) – the importance of prior choice and morphology

This chapter is published in:
Eberle J., R. C. M. Warnock, D. Ahrens (2016). Bayesian species delimitation in *Pleophylla* chafers (Coleoptera) – the importance of prior choice and morphology. *BMC Evolutionary Biology* 16, 94. DOI: 10.1186/s12862-016-0659-3

Note that some text passages of the Material and Methods and Results sections have been submitted as supplementary material to fit the journals requirements.

Authors' contributions to the original article:
Clustering analyses, DNA-based species delimitation, integrative species delimitation: JE; molecular lab work, sequence assembly and alignments, phylogenetic inference, DNA-based species delimitation, JML-analyses: RCMW; fieldwork collections: DA; morphometric analyses, manuscript design and writing: JE, RCMW, DA.

1. Introduction

The identification and delimitation of species is one of the most crucial exercises in the assessment of biodiversity and in understanding the Tree of Life, because species occupy a central role in nearly all disciplines of biology. Species delimitation therefore has broad implications, from biological and ecological conservation, to comparative evolutionary analyses (Daugherty et al. 1990; Agapow et al. 2004; Isaac 2004; Padial and De la Riva 2006). Despite the challenge and importance of defining species units, methods for delimiting species using independent sources of data (e.g., DNA and phenetic data) have only recently been proposed (e.g., Puorto et al. 2001; Wiens and Penkrot 2002; Pons et al. 2006; Knowles and Carstens 2007; Leache et al. 2009; Carstens and Dewey 2010; Ezard et al. 2010; Guillot et al. 2012; Carstens et al. 2013; Edwards and Knowles 2014; Solís-Lemus et al. 2015). Nevertheless, at least since Sneath and Sokal (1962), there has been an extensive use of quantitative methods to infer similarity based on morphological traits. Broadly defined as "numerical taxonomy", or phenetics, these methods have traditionally been used (and criticized) for inferring phylogenetic relationships (e.g., Blackwelder 1967; Sterner 2014). However, integrative approaches to taxonomy shed new light on the utility of these methods, which have the potential to offer an independent, more reproducible way of inferring species limits (Yeates et al. 2011).

In addition to controversy over the application of different species concepts and their impact for delimiting species (Yang and Rannala 2014), delimitation is expected to be especially challenging during the earliest stages of divergence, or speciation, when both molecular and morphological characters exhibit low levels of differentiation (De Queiroz 2007). At this stage it can be extremely difficult to detect genetic isolation (i.e., the ultimate outcome of speciation) due to gene flow among populations and incomplete lineage sorting between species (Hudson and Coyne 2002; Degnan and Rosenberg 2009). Although molecular data can be useful for the rapid identification

and delimitation of species, these processes can compromise the interpretation of the results. Incomplete lineage sorting – shared ancestral polymorphisms between species – can lead to perceived genetic similarity among phenotypically divergent species. Consequently, gene flow and incomplete lineage sorting can result in similar patterns among inferred gene trees (Maddison 1997; Slowinski et al. 1997; Shaffer and Thomson 2007). To further complicate matters, introgressive hybridization – secondary gene flow between species – can also produce similar patterns among inferred gene trees (e.g., Wu and Campbell 2005; Bossu and Near 2009; Keck and Near 2010).

A suite of new methods have been proposed that can incorporate incomplete lineage sorting in a multilocus framework for the estimation of species trees (Edwards et al. 2007; Kubatko et al. 2009; Heled and Drummond 2010; O'Meara 2010) and/or species delimitation (O'Meara 2010; Yang and Rannala 2010, 2014). Although these methods rely on the a priori assignment of individuals to pre-defined units (species or populations; Yang and Rannala 2014), they can be used to test explicit hypotheses of species delimitations. However, studies of recent radiations, or speciation in a young species, will be characterized by uncertain species designations, and are likely to remain challenging.

In contrast to DNA-based taxonomy, common practice for the traditional taxonomic treatment of taxa is an assessment of the organism's entire morphology. In most groups of insects this includes detailed examination of the copulation organs, which often undergo rapid morphological divergence, driven by sexual selection (Simmons 2014). However, quantitative data on insect genitalia are rarely obtained for the purposes of integrative taxonomy, and so methods for combining this type of morphological information with molecular data are still underdeveloped (Yeates et al. 2011). Previously, the only available methods for delimiting species on the basis of morphology were clustering approaches (Fraley and Raftery 2002; Ezard et al. 2010; Fraley et al. 2012; Guillot et al. 2012). Unfortunately, these methods quickly loose power when too many species are included, or when dealing with specimens whose closest phylogenetic relatives are unknown (Edwards and Knowles 2014; Solís-Lemus et al. 2015). Here we use morphometric and molecular data in an integrative framework, to delimit species in the scarab beetle genus *Pleophylla* Erichson, 1847. Following the recommendation of Carstens et al. (2013), we implemented a suite of methods, including a recently developed approach that incorporates continuous morphological

trait data with the multispecies coalescent (Yang and Rannala 2010; Solís-Lemus et al. 2015).

Pleophylla is a highly conspicuous genus, found only in isolated parts of the South African escarpment and the East African highlands. The genus belongs to the tribe Sericini (Coleoptera: Scarabaeidae), a highly diverse clade of herbivorous beetles with nearly 4,000 described species. The adults feed polyphagously on a variety of angiosperms, while the larva feed on humus and plant roots in the upper soil layers. Morphological and molecular evidence has shown that the genus belongs to one of the most ancestral-branching lineages of the Sericini, together with its presumptive sister group, *Omaloplia*, in the eastern Mediterranean (Ahrens 2006; Eberle et al. 2016b). Members of the genus exhibit extreme homogeneity in external morphology, and identification of species usually relies on examination of the male genitalia – a trait used to commonly distinguish between homogenous species of insects (Eberhard 1985), including most members of the tribe Sericini (Ahrens and Lago 2008). Current taxonomic classification recognizes only three valid species (Dalla Torre 1912; `globalspecies.org/ntaxa/2359831`; accessed Dec 13, 2015), however, an extensive survey and taxonomic revision of museum collections has identified 24 distinct morphospecies (Beckett 2012; Eberle et al. 2016a). The aim of our study was to provide a primer for the clarification of the taxonomy of this group, and to explore power and limitations of morphological, molecular and combined approaches to species delimitation in an integrative framework for an apparent "complex" case study.

2. Material and Methods

2.1. Taxon sampling and molecular data collection

A total of 110 individuals of eight putative morphospecies of the genus *Pleophylla* were collected from eight localities in South Africa (Appendix Table B1–B2, Fig. III.1). So far, all known species are endemic to South Africa and represent a limited selection of the morphological diversity of *Pleophylla* (Eberle et al. 2016a). Four of these species have not been described yet, therefore we refer to all putative morphospecies using the same numerical format throughout the text for consistency. *Omaloplia nigromarginata* and *O. ruricola* from the putative sister lineage of *Pleophylla* (Ahrens 2006) were included as outgroup taxa. We assessed support for the monophyly of putative morphospecies using standard molecular markers – the nuclear ribosomal rRNA 28S gene, the nuclear internal transcribed spacer 1 (ITS1), and the mitochondrial cytochrome oxidase subunit 1 *(cox1)* and 16S rRNA *(rrnL)* genes.

DNA was extracted non-destructively from leg or thorax tissue using the Promega Wizard SV96 Plate extraction kit, as per the manufacturer's protocol. An 826 bp fragment of the 3' end of the mitochondrial gene cytochrome oxidase subunit 1 *(cox1)* was amplified using primers C1-J-2183 (Jerry) and TL2-N-3014 (Pat) (Simon et al. 1994). A 469–471 bp fragment of the mitochondrial 16S rRNA gene *(rrnL)* was amplified using the primer LR-N-13398 (16Sar) paired with either N1-J-12585 (ND1A) or LR-J-12961 (16Sb2) (Simon et al. 1994). A 645 bp fragment of the nuclear ribosomal rRNA 28S gene was amplified using the primers 28SFF and 28SDD (Monaghan et al. 2007). The nuclear internal transcribed spacer 1 (ITS1) (636–723 bp) was amplified using the primers 5'GTAGGTGAACCTGCAGAAGG and 5'GCGTTCGAARTGCGATGATCAA (Vogler and Desalle 1994). These primers are sometimes referred to as ITS1R and ITS1F, respectively (e.g., Kerdelhué et al. 2002; Santos et al. 2007),

but note that ITS1R binds to the 5' 18S region, while ITS1F binds to the 3' 5.8S region of this product.

Amplified products were sequenced in both directions using ABI BigDye technology and an AB1 PRISM 3730 DNA Analyzer (Applied Biosystems) at the Natural History Museum (London). Contiguous sequences were assembled from both strands and edited using Sequencher 4 (Gene Codes Corporation, Ann Harbor, MI, USA). All sequences were deposited in GenBank. Accession numbers, identifications and specimen vouchers are provided in Appendix Table B1.

2.2. Morphometric analysis

The partial outline of the male's left paramere (part of the intromittent genital organs, in dorsal view) (Fig. B1) was digitized from images captured on a microscope. The partial outline was extracted from 68 male specimens where the paramere was well preserved. The outlines were re-sampled as a set of 150 semi-landmarks using tpsDig (v. 2.1, `http://www.life.bio.sunysb.edu/morph/`). Standard Eigenshape analysis (MacLeod and Rose 1993; MacLeod 1999) was performed in Eigenshape 2.6, as implemented in *morpho-tools* (v. 4.0, `http://www.morpho-tools.net`), using the covariance for calculation of the similarity matrix and φ: sets of outline coordinate points were converted from the Cartesian (x,y) form to the φ-form of Zahn and Roskies' (Zahn and Roskies 1972) shape function, thereby removing size information. The resultant shape functions were expressed as the raw net angular deviation between outline coordinates. The shapes were mean centered and no standardization was applied. Of the 67 eigenaxes produced, further analysis was performed on the first four eigenaxes that together explained 75% of the variation in the samples. Based on these informative eigenaxes we performed a canonical variate analysis (CVA), grouping the samples according to the morphospecies assignments.

Model-based hierarchical clustering (Fraley and Raftery 2002; Fraley and Raftery 2007) was applied to identify groups of individuals that resemble each other, independent of other evidence or *a priori* assignments, using the R package *mclust* (v. 4.4, Fraley and Raftery 2002; Fraley and Raftery 2007). The function *mclust* was used to evaluate the fit of all available clustering models to the morphometric

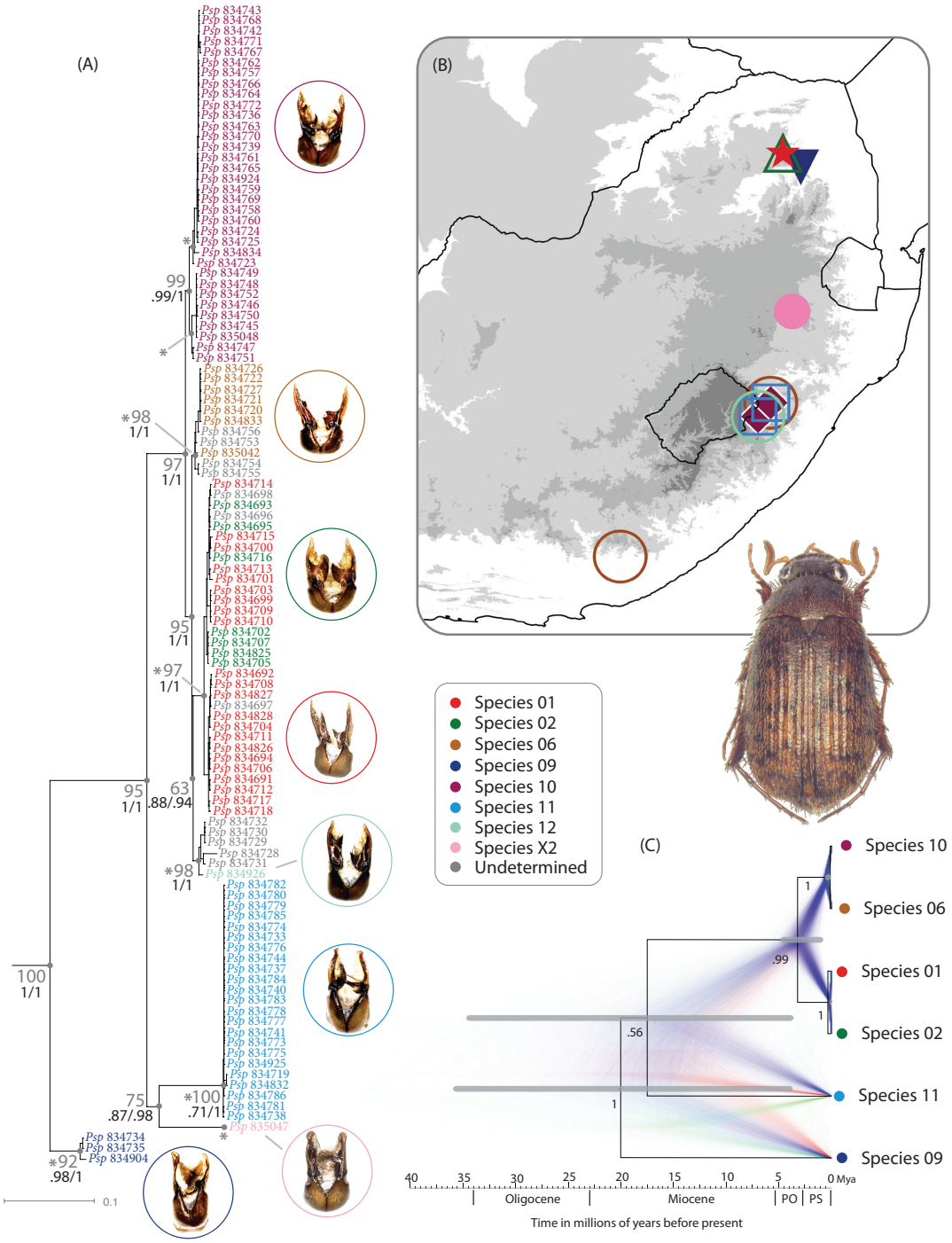

Figure III.1. (A) Maximum likelihood (RAxML) tree of *Pleophylla* for the combined molecular dataset. Specimens are colored according to morphospecies (B1). Branch length corresponds substitutions per site. (To be continued on next page.)

Figure III.1. (Continued.) Support values for ML and Bayesian posterior probabilities are shown next to branches in grey (RAxML) or indicated below (PhyML/MrBayes). ITS1 GMYC clusters are indicated by an asterisk (*). (B) Map of South African sampling localities (B2). (C) Bayesian species tree obtained using *BEAST. Clade posterior probabilities are indicated next to branches. Confidence intervals (grey bars) show the upper limits of the 95% HPDs obtained using a divergence rate for *cox1* of 2% My^{-1}, and the lower limits obtained using a rate of 4%. Mean node ages arbitrarily correspond to the mean estimates obtained using a rate of 2% My^{-1} (Table B6).

data that explained 75% (Eigenshape axes 1–4) and 95% (Eigenshape axes 1–14) of total paramere shape variance. This method uses expectation maximization (EM) to estimate the maximum likelihood of alternative multivariate mixture models that describe shape variation in the morphometric data (McLachlan and Basford 1988; Celeux and Govaert 1995), and estimates the optimal number of clusters based on the Bayesian Information Criterion (BIC) (Schwarz 1978). All models were evaluated for a predefined number of 1 to 20 clusters and the best-fit result was used for further analyses.

To assess the fit of the *a priori* morphospecies assignments and the hierarchical clusters found using *mclust* to the data, we performed a linear discriminant analysis based on the respective specimen groupings and calculated the probability of group membership for each individual. This was done using the R package *MASS* (v. 7.3.35 Venables and Ripley 2002). The prior probability that a specimen belonged to a given group was set to be equal for all individuals and groups.

Finally, to investigate the impact of phylogeny on the inferred morphospace, the RAxML tree topology (based on the partitioned combined molecular dataset) was projected onto the paramere morphospace (eigenaxes 1 and 2) using the function phylomorphospace in the R package *phytools* (Revell 2012). This function estimates the positions of the ancestral nodes using a maximum likelihood approach (Revell 2012). In addition, a three-dimensional version of this plot was produced based on eigenaxes 1, 2 and 3 using the function *phylomorphospace3d*. The code was modified to make coloration for species group affiliation possible.

2.3. Phylogenetic analysis

Cox1, rrnL and 28S sequences were aligned unambiguously using ClustalW 2 (Larkin et al. 2007), with default gap opening and extension penalties (15 and 6.66 respectively). ITS1 was aligned using ClustalW, with further refinement using the refine option in MUSCLE (Edgar 2004). The ITS1 alignment was further edited, and regions that could not be unambiguously aligned were removed. For all remaining sequences, the optimal partition schemes and substitution models were simultaneously identified using the Bayesian Information Criterion implemented in PartitionFinder (Lanfear et al. 2012), with branch lengths linked across partitions. It has been demonstrated that it is difficult to estimate the gamma distribution and invariant sites parameters simultaneously (Yang 1993; Sullivan et al. 1999; Mayrose et al. 2005). To avoid over-parameterization, these parameters were never combined in the same model.

Phylogenetic analyses of individual and combined markers were performed using likelihood and Bayesian methods. Each analysis was run with the substitution model and partitions selected using PartitionFinder (Lanfear et al. 2012) (Table B3).

Unpartitioned maximum likelihood analysis was performed using the subtree pruning and regrafting (SPR) algorithm implemented in PhyML 3.0 (Guindon et al. 2010). All model parameters were estimated, and where applicable, we implemented 5 gamma rate categories. The Bayesian-like transformation of the approximate likelihood ratio test (aBayes) was used to assess branch support (Anisimova et al. 2011).

Partitioned maximum likelihood analysis was performed using RAxML 7.3 (Stamatakis 2006; Stamatakis et al. 2008), implementing the rapid bootstrapping algorithm and a subsequent search for the best scoring ML tree. The number of bootstrap replicates for each dataset was determined using the bootstopping criteria (Pattengale et al. 2009) with a maximum of 1,000 replicates.

Bayesian phylogenetic analysis was performed using the MPI version of MrBayes (v3.1.2 Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003; Altekar et al. 2004). Branch lengths were linked across partitions but partitions were allowed to evolve under different rates. All other substitution model parameters were unlinked across partitions. It has been demonstrated that the default prior on branch lengths implemented in MrBayes can lead to spuriously large estimates of internal branch

lengths (Marshall 2010), and consequently the 95% posterior credibility intervals can exclude the maximum likelihood estimate for tree length (Rannala et al. 2012). The default branch length prior in MrBayes is an exponential distribution with rate parameter $\lambda = 10$, where the mean = 0.1 substitutions/site. Analysis of our dataset using the default branch length prior in MrBayes produced estimates of total tree length (sum of branch lengths) that were an order of magnitude greater than the maximum likelihood estimate obtained in RAxML and PhyML. Because the GMYC approach to species delineation is sensitive to estimates of branch lengths, we ran four sets of analyses using an exponential prior on the branch lengths with mean = 0.1 (default), 0.05, 0.01 or 0.005. Each analysis included two parallel runs of 10 million generations, using one cold and three incrementally heated Markov chains (λ = 0.1), sampling every 1,000 steps and discarding the first 3,000 trees as burn-in. Convergence and mixing were assessed using standard diagnostics (standard deviation of split frequencies, effective sample size, and visual inspection of trace plots). All MCMC output was examined visually using Tracer 1.5 (Drummond et al. 2007).

2.4. Bayesian species tree estimation

The species tree and individual gene trees (*cox1* and ITS1) were co-estimated using the multispecies coalescent model implemented in *BEAST 1.75 (Drummond et al. 2006; Heled and Drummond 2010; Drummond et al. 2012). This analysis requires species (or operational taxonomic units) to be defined *a priori*. We used the putative morphospecies to define taxonomic units and omitted all 14 individuals for which there was ambiguity (Appendix Table B1). Outgroups were not included in this analysis. We implemented the partition scheme and partition-specific models selected for both markers using PartitionFinder for analysis in MrBayes (*cox1* [P1 vs. P2 vs. P3] and ITS1). Substitution model parameters were estimated separately for each partition. Base frequencies were calculated from the data. The transition-transversion ratio parameter of the HKY substitution model was specified using a lognormal prior with mean and standard deviation equal to 1 and 1.25, respectively, LN(1, 1.25). The exchange-ability rate parameters of the GTR model were specified using a gamma prior, with shape and scale parameters equal to 1, G(1,1). The invariant sites parameter was specified using a uniform prior, U(0, 1). Rate heterogeneity across sites

was modeled using the discrete gamma model, with four independent rate categories and the shape parameter was specified using an exponential prior, with mean equal E(0.5).

The uncorrelated lognormal relaxed clock model was implemented to estimate divergence times (Drummond et al. 2006). The mean substitution rate of *cox1* was fixed, and a diffuse exponential prior (E(1/3)) was used to specify the parameter that describes variation in the substitution rate (ucld.stdev). Clock model parameters were unlinked across genes, and the rate of ITS1 was estimated relative to that of *cox1*. We applied a range of mean branch rates, in five independent sets of analyses (2, 2.5, 3, 3.5 or 4% My^{-1}). The parameter used to describe rate variation across branches in the relaxed clock model – the standard deviation of the lognormal distribution of rates – was specified using an exponential prior, E(0.3). The mean of *cox1* was fixed as described above, and the mean rate of ITS1 was specified using an exponential prior, E(1).

We assumed a Yule speciation process, with a constant speciation rate and population size through time. The population size and birth rate parameters were specified using the non-informative improper prior, $\frac{1}{x}$. Two independent MCMC runs were performed for each analysis, each consisting of 500 million iterations, sampling every $5{,}000^{\text{th}}$ generation and discarding the first 50 million steps, resulting in 90,000 samples post burn-in. As above, convergence was assessed using Tracer 1.5. Mean node ages and credibility intervals were calculated using TreeAnnotator 1.7.5.

2.5. Distinguishing incomplete lineage sorting from hybridization

To assess whether low genetic variation observed among morphospecies could be attributed to incomplete lineage sorting, we used the posterior predictive checking approach developed by Joly et al. (2009) and implemented in the software JML (Joly 2012). This approach uses simulated datasets of gene trees and sequence alignments generated under a coalescent model that assumes no migration (or hybridisation) for a given species tree. The proportion of simulated datasets for which the minimum pairwise distance is lower than the observed, can be interpreted as the posterior prob-

ability *(P)* that the model is correct. A small P value therefore suggests that a model that assumes no hybridization does not fit the data well (e.g., the observed minimum genetic distances are lower than expected). To account for uncertainty, simulations were performed for individual partitions using 10,000 trees from the posterior distribution of species tree output by *BEAST, which include estimates of population size and branch lengths. Analysis in JML was performed individually for each partition specified in the *BEAST analysis (*cox1* [P1 vs. P2 vs. P3] and ITS1). The substitution model and rate parameters were specified using the parameters estimated by *BEAST. The relative rates for each *cox1* partition were calculated by multiplying the relative rate of each codon by the mean branching rate of *cox1*. Appropriate heredity scalars were selected for *cox1* (= 0.5) and ITS1 (= 2.0).

2.6. DNA-based species delimitation

For single marker species delimitation (*cox1* and ITS1) we used four widely implemented approaches: statistical parsimony analysis (Templeton et al. 1992), automated barcode gap detection (ABGD) (Puillandre et al. 2012a), the generalized mixed Yule-coalescent (GMYC) model (Pons et al. 2006; Fontaneto et al. 2007; Ezard et al. n.d.), and the Poisson tree processes (PTP) model (Zhang et al. 2013). Outgroup species *(Omaloplia)* and specimens with duplicate haplotypes were pruned from the dataset (or tree) prior to analysis, otherwise some methods have been shown to produce false positives (Ahrens et al. 2016).

Haplotype networks for each individual marker were generated using statistical parsimony analysis (Templeton et al. 1992) implemented in TCS 1.2 (Clement et al. 2000). Statistical parsimony analysis partitions the data into networks of closely related haplotypes connected by changes that are non-homoplastic with a 95% probability; if applied to mtDNA, the inferred networks have been found to be largely congruent with Linnaean species (Hart and Sunday 2007). The GMYC model (Pons et al. 2006; Fontaneto et al. 2007) was used to estimate species boundaries with the trees obtained from MrBayes and RAxML using in the R package splits (Ezard et al. n.d.), with single and multiple threshold options. This method is based on the phylogenetic species concept and identifies species clusters by recognising the apparent increase in the branching rate from interspecific diversification to population-level

coalescence, and defining the threshold based on an ultrametric tree. Trees were converted to ultrametric using PATHd8 (Britton et al. 2007) and the penalized likelihood method implemented in r8s 1.7 (Sanderson 2003), with the optimal smoothing parameter selected using the cross-validation procedure. The age of the ingroup was assigned an arbitrary age of 1, and the resultant trees were fully resolved using TreeEdit 1.0 using an arbitrary branch length of 4×10^{-6}. Finally, we estimated uncertainty in the number of GMYC species clusters based on the Akaike Information Criterion (AIC), using the method outlined in (Powell 2012). This approach uses a modified AIC score, corrected for sample size (AICc), to assess the relative support for alternative (single and multiple threshold) models, versus the maximum likelihood model, and the null model (no change in the branching rate). Akaike weights (the relative support for each model) are assigned to each model based on the AICc scores. Model-averaged estimates of the number of GMYC species are obtained from the models within $\delta AICc = 2$. The phylogenetic species concept also underlies the Poisson tree processes (PTP) model for species delimitation (Zhang et al. 2013). However, in contrast to the GMYC approach, the PTP infers speciation events based on a shift in the number of substitutions at internal nodes. We employed the maximum likelihood variant of PTP using the RAxML trees. For the ABGD approach we used the online version (last modified on Oct. 29, 2015 and accessed on Jan. 23, 2016, `http://wwwabi.snv.jussieu.fr/public/abgd/abgdweb.html`, Puillandre et al. 2012b). This method is based on the assumption that divergence among organisms belonging to the same species will be less than the divergence observed among organisms of different species. The first significant gap in the distribution of sequence distances beyond intraspecific sequence divergence can thus be used to infer operational taxonomic units (OTU) that may be related to species (e.g., Vogler and Monaghan 2007). ABGD analyses were performed on matrices of pairwise sequence divergence, calculated for each marker using MEGA (v. 6.06, Tamura et al. 2013). Distances were corrected using the best fitting substitution models. Prior maximum divergence of intraspecific diversity was set to 0.01, which has previously been demonstrated to recover species accurately (Puillandre et al. 2012a).

Finally, the results of competing approaches to species delimitation were compared using the "entities counts" (i.e. inferred species counts) and the match ratio $= 2 * Nmatch/(Ni+Nmorph)$, where Nmatch is the number of species with exact matches

(i.e., all specimens of a given morphospecies – and only those – belong to a single GMYC entity) and Ni and Nmorph are the number of inferred molecular operational taxonomic units (MOTUs) and morphospecies, respectively (Ahrens et al. 2016). If the there is complete congruence between the GMYC entities and the morphospecies the match ratio = 1, otherwise the ratio will be <1.

2.7. Total-evidence species delimitation

We assessed support for the *a priori* morphospecies assignments using a total-evidence-based Bayesian approach, implemented in the programs iBPP 2.1.2 (Solís-Lemus et al. 2015) and BPP 3.0, (Yang and Rannala 2010, 2014). Briefly, this method uses a multispecies coalescent model to assess competing hypotheses of species delimitations, allowing for conflict between gene and species trees. The results are conditioned on a user specified guide tree and depend on estimates of the species divergence times (τ) and population sizes (θ). Individuals are assigned to independent populations and alternative delimitation hypotheses are proposed by collapsing one or more internals nodes in the guide tree. In the original implementation, the likelihood calculation is based on molecular data (Yang and Rannala 2010), while iBPP includes an extension of the model that allows continuous trait data to be included in the likelihood calculation (Solís-Lemus et al. 2015). This latter approach therefore enables both molecular and morphological data to be combined in the assessment of *a priori* species assignments.

It has been demonstrated that the results of this method can be sensitive to both prior parameter and guide tree choice (Leaché and Fujita 2010). For example, for high values of θ the model tends to (over-) split species, and for low values of θ the model tends to lump species together. To assess the robustness of our results, we compared the results obtained under variable combinations of the specified priors on the root age (τ_0) and the population mutation rate (θ) (Table III.1). To assess the influence of the guide tree, we compared the results obtained using three alternative input trees: (a) the topology estimated using *BEAST, (b) the topology estimated from the concatenated DNA matrix using RAxML/MrBayes, (c) a modified version of the *BEAST topology based on morphological similarity among species (Fig. B2). All combinations of prior parameter (Table III.1) and guide tree choices were per-

Table III.1. α and β parameters, describing the prior distributions of the population mutation rate (θ) and the root age (τ_0) that were combined in the iBPP analyses.

prior	α	β	mean	variation
θ_1	1	10	0.1	0.01
θ_2	1	20	0.05	$2.5e^{-3}$
θ_3	2	2000	0.001	$5e^{-7}$
τ_{0-1}	1	10	0.1	0.01
τ_{0-2}	1	20	0.05	$2.5e^{-3}$
τ_{0-3}	2	2000	0.001	$5e^{-7}$

formed in iBPP (a) without data, to evaluate the impact of the priors, and using the following three datasets: (b) molecular data only, (c) morphometric data only, and (d) molecular and morphometric data. The analysis sometimes got stuck in a single species model, resulting in poor overall convergence, and so all analyses were repeated 10 times with different random seeds to ensure stability of the results.

In an additional set of analyses, we implemented unguided species delimitation using the program BPP (Yang and Rannala 2014). This method accounts for uncertainty in the guide tree, by proposing changes to the species tree topology using nearest-neighbor interchange (NNI), as well as proposing changes to species assignments. Morphometric data cannot be analyzed in BPP, so this analysis was performed for the molecular dataset only. The analyses were performed using the above combinations of priors and initial guide tree choices.

To explore the impact of distinct single-marker genotypes within the same morphospecies, in combination with the morphological trait data, we also analyzed an additional guide tree with guided and unguided BPP, in which *sp10* was specified as two species entities (This split received strong support in several single marker delimitations, see section Results.

3. Results

3.1. Sequence data, alignment and model selection

We obtained 438 new sequences for 110 individuals (Supplementary Table B1). There is remarkably low molecular variation among the members of the genus. Excluding outgroups, the final alignments for each individual marker contained 173, 18, 3 and 194 parsimony informative characters for *cox1* (826 bp), *rrnL* (469–471 bp), 28S (645 bp) and ITS1 (726–812 bp), respectively. Previously published *cox1, rrnL* and 28S sequences for the outgroup species (*Omalopia nigromarginata* and *O. ruricola*) were included in our analysis (Ahrens and Vogler 2008). We also obtained ITS1 sequences for *O. nigromarginata* and *O. ruricola*, however it was not possible to align unambiguously these sequences with the ingroup taxa, so these sequences were excluded from further analysis. The final concatenated alignment contained 2,795 characters, including 3.57% missing data and gaps. The partition schemes and optimal substitution models were selected using PartitionFinder for each dataset and for use in different programs are presented in Supplementary Table B3.

3.2. Phylogenetic analysis and the monophyly of morphospecies

Phylogenetic analysis of independent and combined datasets using different approaches and parameter choices (PhyML, RAxML, and MrBayes) produced overall similar topologies (Fig. III.1, Figs. B3 – B5). Changing the branch length prior implemented in MrBayes had no impact on the inferred topology but had a large impact on tree length (the sum of branch lengths) (Table B4). Analysis of different datasets (mitochondrial, nuclear or combined) mainly differed in their degree of tree resolution,

and the level of support for the monophyly of individual morphospecies and/or interspecific relationships. There is remarkably low interspecific molecular variation observed across the entire genus. The trees produced using the ribosomal markers (*rrnL* and 28S) were poorly resolved. The *cox1* data provided better resolution and supported the monophyly of two out of eight putative morphospecies. ITS1 provided the best resolution and supported the monophyly of all but two morphospecies (*sp01* and *sp02*) (Figs. B3 – B5).

The topology obtained using the combined dataset that included all four markers was identical to the ITS1 gene tree (Fig. III.1), but support values for most nodes were greater than those obtained using individual genes. In the combined analyses of all four markers, the monophyly of all putative morphospecies was strongly supported with the aforementioned exception. Morphospecies *sp01* and *sp02* were never recovered as monophyletic, although these groups occupied distinct areas of the morphospace in the morphometric analysis of the genitalia (Figs. III.2, B1).

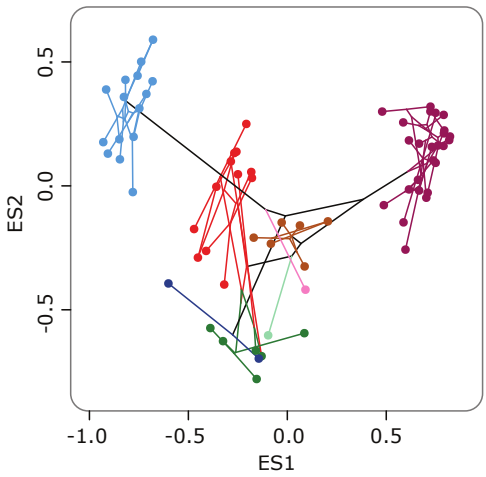

Figure III.2. Plot of the 2D phylomorphospace using the RAxML tree topology (based on the partitioned combined molecular dataset) projected onto the paramere morphospace explained by eigenaxes 1 and 2.

Changing the branch length prior implemented in MrBayes had no impact on the inferred topology but had a large impact on tree length (the sum of branch lengths) (Supplementary Table B4). Note that in some cases the Bayesian 95% posterior intervals do not contain the maximum likelihood estimate. A large difference was observed between the estimates of tree length obtained using the Bayesian analysis, under the default branch length prior in MrBayes, and the maximum likelihood estimates obtained using RAxML and PhyML (Supplementary Table B4). Changing the branch length prior to favour shorter branch lengths reduced this discrepancy, and for analyses of *cox1* actually produced shorter lengths than the maximum likelihood analyses. Partitioned maximum likelihood analysis produced longer tree lengths. In par-

ticular, when *cox1* is included in the analysis, partitioning by codon had a large impact on the branch lengths.

The Bayesian species tree estimated using *BEAST for the combined cox1 and ITS1 dataset resulted in strong support for the interspecific relationships estimated using the *cox1* data, rather than the ITS1 data. Although the species tree topology differed to that obtained using alternative phylogenetic methods (PhyML, RAxML and MrBayes), the individual gene trees (for *cox1* and ITS1) obtained using *BEAST were not different. The age of the most recent common ancestor of the sampled members of the genus, was estimated to be 2.64 – 35.97 and 3.69 – 17.88 Mya, based on the 95% highest posterior density intervals for the slowest and fastest *cox1* substitution rates (2 and 4% My^{-1}), respectively (Table B5). The use of a higher *cox1* substitution rate produced younger and, unexpectedly, more precise posterior age estimates. The ages for the two youngest divergence events (*sp01* + *sp02* and *sp06* + *sp10*) were estimated to be no older than 0.17 My and 0.65 My, respectively (Table B6).

Evidence of hybridization was assessed using the posterior predictive checking approach as implemented in the software JML (Joly 2012), based on the minimum pairwise sequence distances among morphospecies for each marker partition (*cox1* [P1 vs. P2 vs. P3], and ITS1), and the resulting posterior probability (P) of observing these distances under the multispecies coalescent model assuming no hybridization (Table B7). In all cases the observed pairwise distances between individuals of all morphospecies were not lower than expected at the 5% level ($P > 0.05$), given the null model (the coalescent with no migration or hybridization) across all partitions (*cox1* P1 and P2, $P > 0.1$; P3, $P > 0.05$; ITS1, $P > 0.2$). The distances observed between individuals of the two species pairs that could not be resolved using *cox1* (*sp06* + *sp10*) or both *cox1* and ITS1 (*sp01* + *sp02*) were not lower than expected for either marker (i.e., *sp06* + *sp10*, $P > 0.2$; *sp01* + *sp02*, $P > 0.6$). The tests performed thus suggest that incomplete lineage sorting is sufficient to explain the observed genetic variation.

However, the third *(cox1)* codon partition (P3) produced anomalous results for *sp09* and *sp11*. The probabilities of obtaining the observed distances under the null model was lower than expected at the 10% level ($P < 0.1$; $P > 0.1$ for all other pairwise comparisons). Note that the third codon position evolves at one and two orders of magnitude faster than the first and second positions respectively (Supplementary

Table B5). For the ITS1 data, the probabilities of obtaining the observed sequence distances under the null model were not lower than expected ($P > 0.2$ for all pairwise comparisons).

3.3. Molecular tree- and character-based species delimitation

We investigated DNA-based species delimitation and associated uncertainty using (i) statistical parsimony, (ii) the GMYC model, (iii) the PTP model, and (iv) ABGD approach. The analyses using the *rrnL* and 28S data did not provide support for any of the putative morphospecies (results not shown). Of the 13 resulting *cox1* networks, three matched exclusively a single putative morphospecies. ITS1 networks provided a closer correspondence to the morphospecies. Of the 9 ITS1 networks, four matched exclusively a single putative morphospecies: *sp09, sp11, sp12* and *spX2*. Individuals of morphospecies *sp06* shared two networks, and individuals of morphospecies *sp10* shared two networks. Individuals of morphospecies *sp01* and *sp02* shared a single network. Together these results suggest that there is a higher degree of incomplete lineage sorting among *cox1* than ITS1, and that species *sp01* and *sp02* cannot be distinguished on the basis of the molecular markers used here.

The GMYC results obtained using *cox1* were very sensitive to the input tree, but there were no obvious differences in the GMYC output that could be attributable to the trees generated using MrBayes versus RAxML, or PATHd8 versus r8s (Table B8). Bayesian trees with longer branch lengths tended to result in more GMYC entities (species clusters + singletons), but not ubiquitously. Consequently, the *cox1* trees produced very variable results. In most cases several (up to 8) models contributed to a majority of the Akaike weight (> 0.5), suggesting that no single model best represented the data. Accounting for uncertainty in model selection resulted in the number of entities ranging between 3.00 ($\sigma^2 = 0$) and 16.54 ($\sigma^2 = 0.89$), depending on the input tree; these GMYC units were widely incongruent with the a priori morphospecies assignments (further details therefore not shown here). There was less variation in the GMYC results obtained using the ITS1 trees – the single threshold models were always preferred to the multiple threshold models. In the majority of cases only one single threshold model was found within $\delta AICc = 2$, suggesting that the preferred model provided an appreciably better fit to the data than the alterna-

Table III.2. DNA based species delimitation results. The number of delimited entities and the match ratio ($2 * N_{match}/(N_{GMYC} + N_{morph})$) (Ahrens et al. 2016) after removing undetermined specimens is given.

	cox1				ITS1			
	PTP	GMYC	TCS	ABGD	PTP	GMYC	TCS	ABGD
Entities	7	13	13[1]	11	8	8	9	8
Match ratio	0.27	0.29	0.30	0.42	0.63	0.63	0.47	0.63

[1]Contained one MOTU composed of only female specimens, this unit was not considered for match ratio estimation.

tives. The ITS1 data resulted in a minimum of 8 ($\sigma^2 = 0$) and a maximum of 10.99 ($\sigma^2 = 4.05$) entities, depending on the input tree. In 8 out of 10 cases, the preferred model resulted in eight entities, corresponding to morphospecies *sp01 + sp02, sp06, sp09, sp11, sp12, spX2,* and two clusters of morphospecies *sp10*.

In general, congruence between the inferred MOTUs and the morphospecies was more dependent on marker choice than species delimitation method (Table III.2, B2). For *cox1* the number of MOTUs ranged from 7 (PTP) to 13 (GMYC), while the analyses based on ITS1 resulted in 8 (GMYC, PTP, ABGD), 9 (TCS) and 10 (GMYC) entities. The PTP and ABGD analyses largely confirmed the results of the GMYC model for the ITS1 data; five of the eight MOTUs were fully congruent with the morphospecies *(sp11, spx2, sp9, sp6, sp12)*. Finally, the match ratios obtained for *cox1* were consistently lower (0.27–0.42) than those obtained using ITS1 (0.47–0.63) (Table III.2).

3.4. Morphometric evidence for species delimitation

We first assessed quantitative support for the eight putative morphospecies assignments among *Pleophylla* based on an open shape outline of the left paramere of the male genitalia, using (i) standard eigenshape analysis, (ii) canonical variate analysis (CVA), (iii) hierarchical clustering, and (iv) linear discriminant analysis. The first four eigenshape axes represented 75% of the cumulative variation of the outline shape (Table B9, Fig. B6). Eigenaxis 1, 2, 3 and 4 represented 51.5%, 15.6%, 6.8% and 6.0% of the variation, respectively. The first 14 eigenaxes account for 95% of

the cumulative variation. The plots of the 2D and 3D phylomorphospace (Fig. III.2, electronic supplementary Fig. B1) showed clear separation between all but one of the morphospecies, with no intermediate states between the morphospecies. The only exception was *sp12*, which overlapped in morphospace with *sp02*. CVA on eigenaxes 1–4 (Fig. B1) revealed a clear distinction between five of the eight morphospecies *(sp01, sp02, sp06, sp10* and *sp11)*, with the exception of those for which only one or two specimens were available for analysis *(sp09, sp12* and *spX2)*. This was in contrast to the DNA-based tree topology and species delimitation, where specimens of two species pairs *(sp01 + sp02*, and *sp06 + sp10)* could not be distinguished based on the analysis of *cox1* and/or ITS1.

Hierarchical model-based cluster analysis (Fraley and Raftery 2002) can identify unique morphological clusters of individuals without requiring *a priori* species assignments (e.g., Ezard et al. 2010). The results of this analysis were extremely sensitive to the model choice (Fig. III.3). Different mixture models favoured strikingly different numbers of clusters (e.g., 9, 7, 5, and 3 clusters were found for eigenaxes 1–4 under different models) (BIC, Fig. III.3A). The best model obtained for eigenaxes 1–4 (the ellipsoidal, equal shape model; VEV) resulted in 3 clusters, but only morphospecies *sp11* and *sp10* (with the exception of one individual) were recovered as independent unique clusters. The best-fit model obtained for eigenaxes 1–14 (the diagonal, varying volume, equal shape model; VEI) resulted in 12 clusters (Fig. III.3B), with all morphospecies recovered in more than one group, with the exception of the singletons and *sp6*; the latter was recovered together with individuals of *sp9* and *spX2*.

Linear discriminant analysis (LDA) with respect to the *a priori* defined morphospecies recovered one of the eight species (*sp11*, 100% of individuals) based on eigenaxes 1–4 (Fig. III.3, Table B10). Two of the eight morphospecies were recovered with the LDA based on eigenaxes 1–14 (*sp10, sp11,* 100% of individuals; the remaining morphospecies were recovered for 50–92% of individuals). LDA with respect to groups identified by the model-based cluster analysis recovered all three clusters correctly based on eigenaxes 1–4 (Fig. III.3, Table B11). Finally, LDA on clusters from the second analysis based on eigenaxes 1–14 recovered all but two of the groups for 100% of individuals.

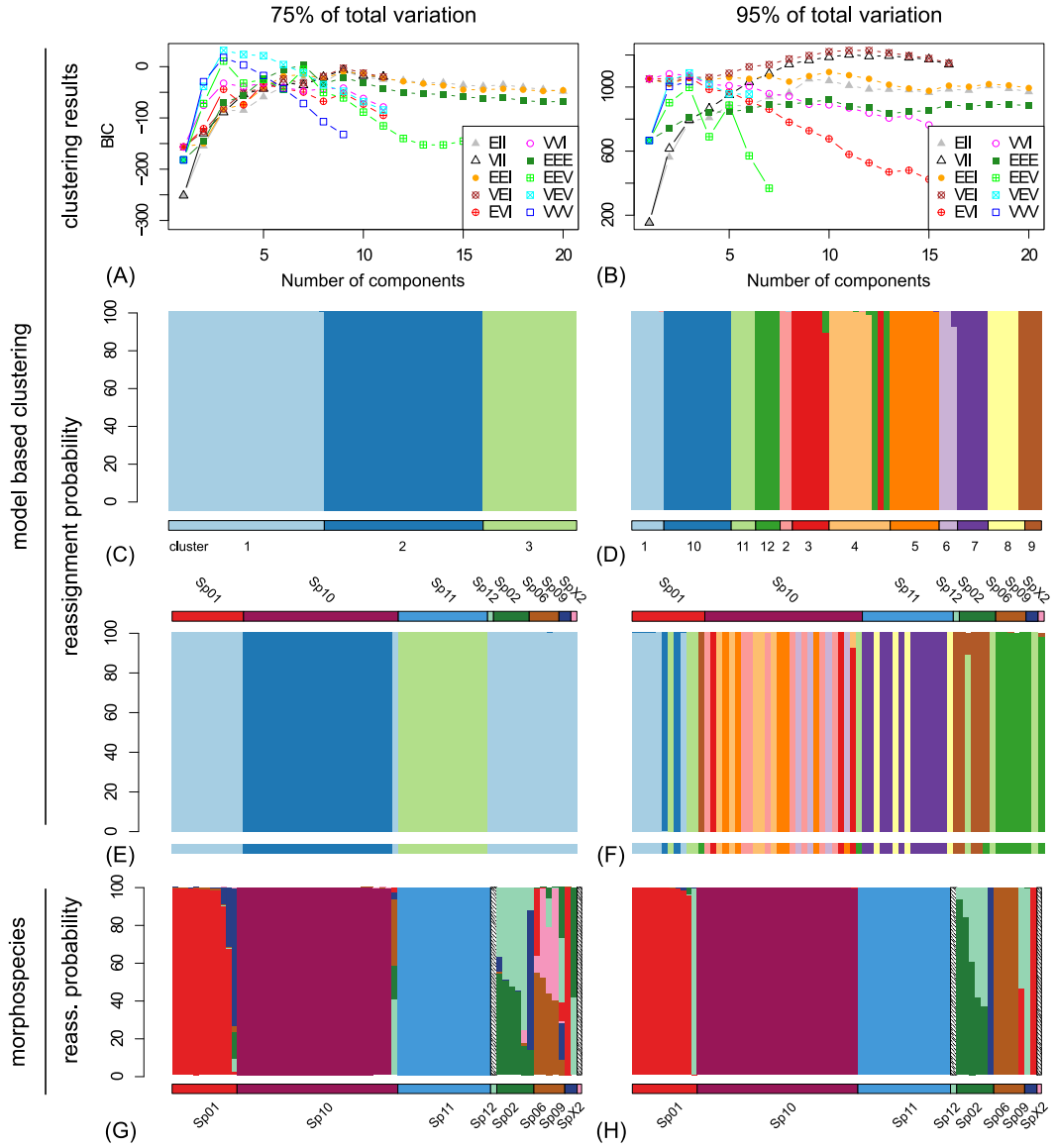

Figure III.3. Species estimates of hierarchical clustering and confidence evaluation of *a priori* defined morphospecies and morphoclusters by LDA. Columns show results for 75% and 95% of total variation in the morphometric data. (A, B): Choice of the best fitting cluster model by BIC. Reassignment probabilities to the clusters from hierarchical clustering with individuals ordered by (C, D) clusters and (E, F) by *a priori* defined morphospecies, and (G, H) reassignment probabilities to *a priori* defined morphospecies. Bars below plots C-H indicate prior group assignment for LDA, bars above plots E and F indicate affiliation to a priori defined morphospecies. Individuals in plots E-H are ordered identically.

3.5. Bayesian species delimitation

The total-evidence approach to Bayesian species delimitation (Yang and Rannala 2010; Solís-Lemus et al. 2015) provided strong support for the *a priori* defined morphospecies, however, for independent data types (molecular versus morphometric), the results were sensitive to the priors on the root age (τ_0) and population size (θ) parameters (Fig. , Table S13). Broadly, posterior probabilities (i.e., support for species delimitations) increased in the integrative analyses that combined molecular and morphological trait data (Fig. III.4). While results were sensitive to both the choice of τ_0 and θ, the choice of θ seemed to be more influential. The most consistent pattern that emerged was that low values of θ sometimes lead to low support for species delimitations. Species remained relatively well supported with high prior values of τ_0. When the model was run under the prior (e.g., without data), with exception of the deepest divergences *(sp09)*, the model did not result in any support ($P > 0.95$) for the *a priori* species assignments. This indicates that although the results were sensitive to the priors, the data contained informative signal.

Based on morphometric data alone, the divergence between *sp01* + *sp02* and *sp12* in tree A was strongly supported ($P > 0.95$), however, the combination with low θ values reduced support at these nodes (Fig. III.4). The analysis based on molecular data alone provided overall support for the *a priori* morphospecies assignments. Exceptions occur for all nodes given low values of θ with all data sets. For example, *sp01* and *sp02* were strongly supported in analyses with higher values of θ ($P > 0.95$), while there was low support for this divergence in analyses with the lowest value of θ ($P < 0.32$). The delimitation between *sp02* and *sp12* (tree C) was the only split that consistently received low support under all θ prior values and with all data sets.

As expected, the results were also sensitive to the guide tree choice. For example, when *sp02* and *sp12* were specified as belonging to separate groups of species, they were always strongly supported with high posterior probabilities (tree A, B). However, when the guide topology was modified to accommodate the observed high morphological similarity between *sp02* and *sp12* (guide tree C), they were almost never recovered as independent species (Fig. III.4). Interestingly, none of the *a priori* defined species gained high support for all prior combinations across all guide trees, even using the integrative total evidence approach (Fig. III.4D).

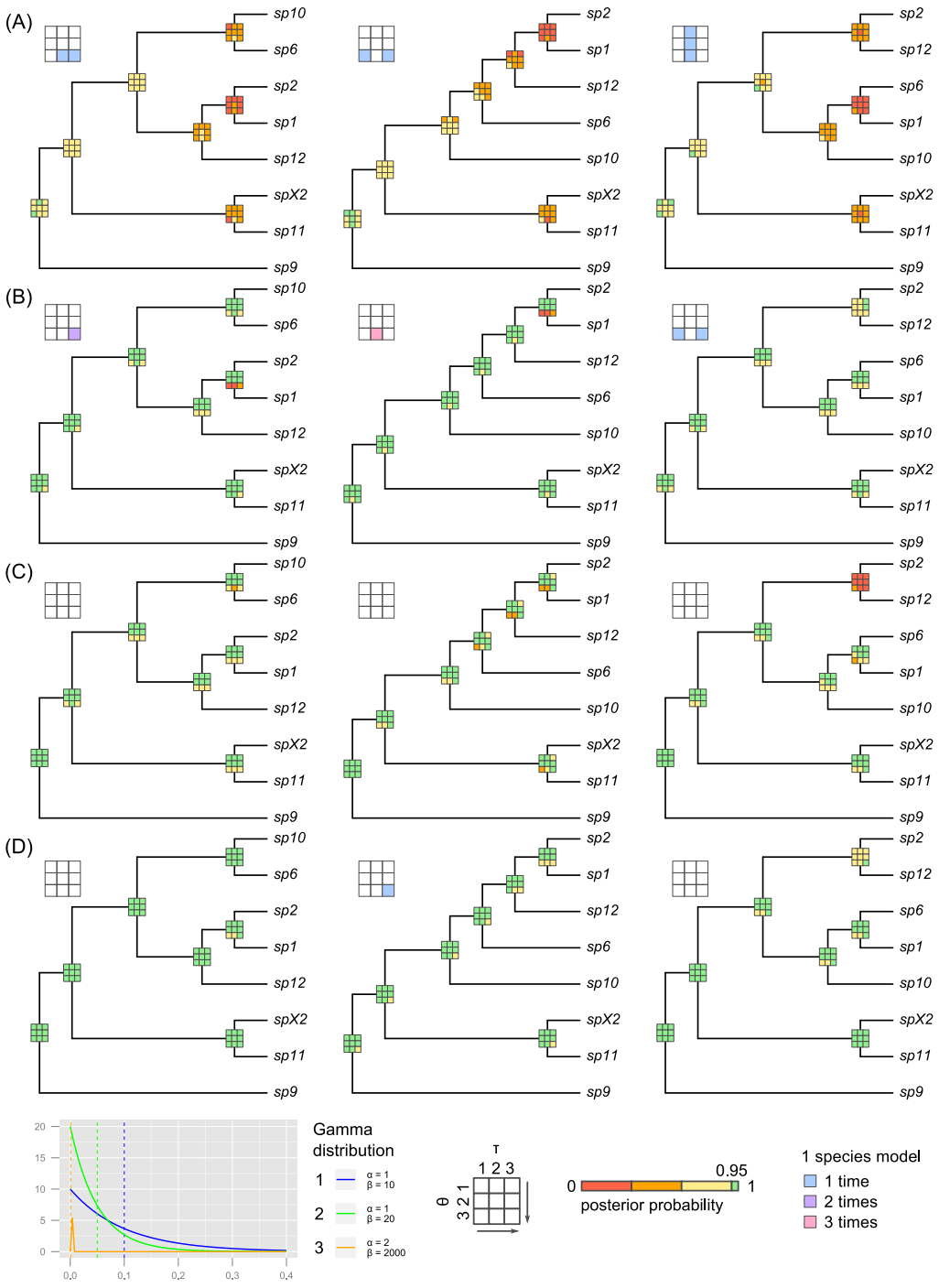

Figure III.4. Mean posterior probabilities of Bayesian species delimitations from 10 repeated runs with commonly used priors. Means inferred under 9 different τ and θ prior combinations are color-coded in 3×3 boxes on each putative speciation split of the guide trees. The arrows in the legend point to the direction of more conservative prior choices. (Continued on next page.)

Figure III.4. (Continued.) Columns from left to right: 3 alternative guide tree topologies from *BEAST, ML and MrBayes analyses, and a modified *BEAST topology based on morphological similarity of the species; rows: analyses using (A) no data (prior only), (B) molecular data, (C) morphometric trait data, and (D) both data sources. The colors of the large 3×3 inset boxes indicate the number of repeat-analyses that were stuck in the one species model. Gamma distribution densities of τ and θ priors 1-3 are depicted in the bottom left corner. Dashed lines indicate the respective distribution means.

The unguided analyses (molecular data only) that applied nearest neighbor interchanges (NNI) to the initial guide tree topologies largely confirmed the results of the guided (iBPP) analyses. While the initial guide tree and the choice of the τ_0 prior did not alter the results, the choice of the θ prior had strong influence on the posterior probabilities of the speciation splits. All *a priori* defined morphospecies were well supported under θ_1 and θ_2 (Table III.1) however, under the narrow and small θ_3 prior, in particular *sp1* and *sp2*, but also *sp1, sp10, sp12, sp2,* and *sp6*, were lumped into one species (Electronic Supplement Tables B4–B5).

In the final set of analyses, in which *sp10* was specified as two separate entities, corresponding to two distinct genotypes (Fig. B7), this split was not supported based on the analysis of the morphometric data alone, as expected. However, this split received strong support based on the analysis of both the molecular only and combined datasets (Fig. B7).

4. Discussion

4.1. Congruence between single DNA markers and morphometric evidence

Using a wide range of morphometric and phylogenetic tools, we tested for congruence between morphological, molecular, and integrative approaches (i.e., iterative *sensu* Yeates et al. 2011) to species delimitation in the chafer beetle genus *Pleophylla*. Morphometric analysis (eigenshape analysis) of the left paramere of the male genitalia, as well as subsequent CVA and LDA provided quantitative support for the majority species assignments based on morphology. In contrast, model based hierarchical clustering showed much less congruence with the morphospecies (Fig. III.3E, F), indicating that this approach may not be suitable for delimitation at the level of species.

Molecular-based species delimitation resulted in a wide range of support for morphospecies, based on the analysis of standard markers used among beetles (e.g., Ahrens et al. 2007; Hunt et al. 2007; Ahrens and Vogler 2008; Bocak et al. 2014), from zero (28S and *rrnL*) to moderate or high (*cox1* and ITS1). The ribosomal markers were insufficiently informative to support any of the putative morphospecies (Fig. B3–B5; Table B8), due to the remarkably low interspecific molecular variation observed across the entire genus. This is less surprising for the slowly evolving 28S rRNA marker, but *rrnL* has previously provided reasonable resolution at the species level among scarabs (e.g., Ahrens et al. 2007). The mitochondrial gene *cox1* and the nuclear region ITS1 were more informative, while the latter provided the best resolution. There was overall congruence between the morphospecies and the ITS1 MOTUs (GMYC, ABGD, PTP), despite the fact that ITS1 had fewer haplotypes than *cox1* (23 versus 53) and a lower relative substitution rate (Table B5). A wide range of tree building methods, parameters and tree linearization approaches did not

improve the results of the GMYC model using cox1. In particular, there were three putative morphospecies that were difficult to distinguish on the basis of molecular data alone (*sp01* vs *sp02; sp6* vs *sp10; sp02* vs *sp12*). At one extreme, individuals belonging to a single morphospecies *(sp10)* were assigned to two MOTUs on the basis of two distinct ITS1 genotypes. The genotypes had a total of 31 segregating sites, including one 2-base-deletion, two 4-base-deletions, and one 2-base-insertion, indicating that a single mutation is unlikely to the cause of the molecular variation, although this pattern was not recovered by any other marker. At the other extreme, individuals belonging to two distinct morphospecies were assigned to a single MOTU and shared identical cox1 and ITS1 haplotypes *(sp01, sp02)*, which may be attributed to introgressive hybridization or incomplete lineage sorting.

Distinguishing between secondary gene flow and incomplete lineage sorting is difficult because both processes produce similar phylogenetic patterns (Joly et al. 2009). JML analyses (Joly 2012) indicated that incomplete lineage sorting may be sufficient to explain the observed level genetic variation across species with independent data partitions and species – with the exception of the fast evolving *cox1* third codon (Table B5; *sp09* and *sp11*), the monophyly of these species was otherwise well supported. The basic substitution model implemented in JML may not be sufficient to account for hidden substitutions at this site and may underestimate the genetic distance for this partition (Tables B5, B7). Overall, the JML results provide support for an incomplete lineage sorting scenario, however, this test cannot be treated as definitive against secondary gene flow. The method implemented in JML can only be used to detect hybridization events for sequences that have a coalescence time younger than the speciation event (Joly et al. 2009), and this approach can result in false negatives (Heled et al. 2013).

4.2. Bayesian species delimitation using an integrative taxonomy framework

In concordance with our results from the Bayesian species delimitation, Solís-Lemus et al. (2015) have shown that the integration of morphological evidence together with molecular data may greatly enhance the discriminative power of species delimitation

models. However, it has also been shown that errors and uncertainties in upstream analyses (e.g., guide tree inference, individual-species assignment) and prior parameter choice may impact the accuracy of results (Leaché and Fujita 2010; Olave et al. 2014; Zhang et al. 2014). Here, we assessed the impact of a wide range of parameter combinations, including prior parameter and guide tree choice.

Leaché and Fujita (2010) previously demonstrated the significant impact of using randomly generated guide tree topologies. Rannala (2015) questioned the practicality of exhaustive guide tree manipulation, with respect to the increased computation time associated with popular phylogenetic inference methods. In addition, a random set of guide trees will include some unreasonable or unlikely topologies, which can lead to inaccurate delimitations (e.g., over-splitting; Zhang et al. 2014). Here, we limited our guide tree choice to three options, justified on the basis of evidence of independent molecular and morphometric evidence, in order to further evaluate incongruences between both data sources (Fig. B2). The use of alternative guide trees had a large impact on the results. For example, the use of guide tree C (based on morphological similarity) allowed us to identify support for a putative species pair (*sp02/sp12*), which was otherwise not identified using alternative molecular based approaches, including the unguided (NNI) approach in BPP (Electronic Supplement Tables B4–B5). In an additional set of experiments, we used a fourth guide tree topology based on the support for a putative case of cryptic diversity obtained using alternative single-marker delimitation approaches (*sp10*, Table B2; Fig. III.5). This experiment, however, cannot provide definitive support for these species entities, because the units were inferred on the basis of non-independent evidence. Manual inspection of the alignments for *sp10* revealed 2 ITS1 genotypes with 43 segregating sites represented by *sp10a* and *sp10b*. This is a very strong signal compared to a total of 44 segregating sites in both mitochondrial markers, which did not exhibit any diverging signal between *sp10a* and *sp10b*. Only a single site was polymorphic for *sp10b* in 2 of the 4 *sp10b* specimens. However, these analyses serve to demonstrate that the results obtained using this model can be extremely sensitive to the signal present in single molecular markers, even in presence of data that provide strong evidence for morphological similarity (Fig. B7).

The use of alternative prior combinations for the population size (θ) and root age (τ_0) priors each had a large impact on the results. These analyses indicate that these

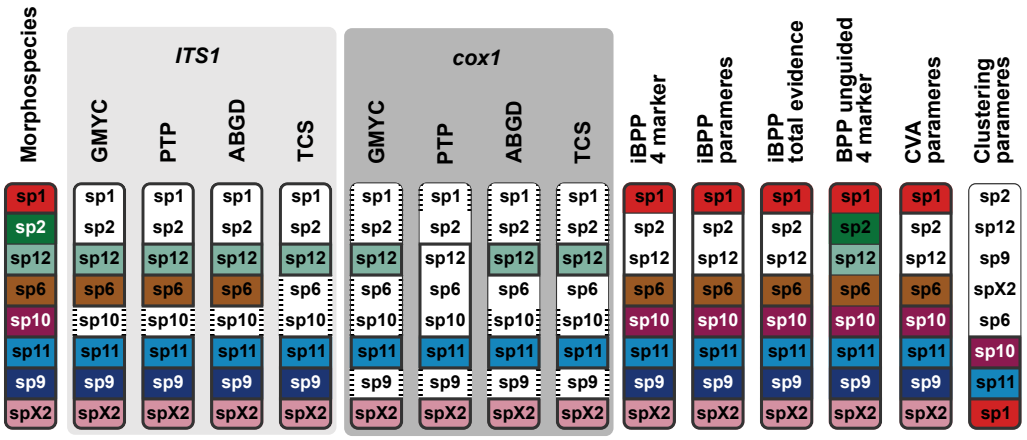

Figure III.5. Overview of the results from the different species delimitation methods and data. Inferred entities that were fully congruent with the *a priori* morphospecies assignments are indicated by the bold circumscribed coloured squares, incongruent units remain white; sub-splitting within morphospecies is indicated by horizontal dashes. Additional sub-splitting within morphospecies that share overlapping MOTUs are circumscribed by a narrow line. Uncertain delimitations are indicated by thin lines between *a priori* morphospecies.

parameters must either be chosen using extreme caution (using independent empirical evidence) or multiple analyses should be performed to assess the robustness of species delimitations to these parameters, such as the analyses performed here. We found that phylogenetically younger species *(sp01, sp02, sp06, sp10, sp12)* and analyses that employed less data (e.g. single versus combined traits) were typically more sensitive to the results. It has also been demonstrated that strong variation in mutation rate and population size among populations or species can also decrease the accuracy of alternative coalescent-based delimitation models (Fujisawa and Barraclough 2013).

The inclusion of more individuals (and/or data) can lead to more accurate and precise parameter estimates (Yang 2015), but increased taxon sampling is sometimes not possible due to the natural rarity of some species (Ahrens et al. 2016). The development of better approaches to account for this uncertainty may be important, because in reality many biodiversity studies will be subject to limited taxon sampling. Further research using empirical and simulated data are required to fully assess the impact of guide tree, prior parameter choice, model violation and taxon sampling.

Here, we demonstrate that the inclusion of morphological data can lead to more robust estimates of species delimitations. The results obtained using the combined dataset are less sensitive to prior parameter choice, than the analysis based on molecular or morphological dataset alone (Fig. III.4; Electronic Supplement Tables B3–B5). Overall, nearly all morphospecies received strong support based on the analysis of the combined dataset (Fig. III.4D). All sequence-based inference methods, including tree inference using concatenated data or coalescent-based approaches such as *BEAST and BPP, may be impacted by putative incomplete lineage sorting or introgression. An integrative approach to taxonomy enables all available evidence to be utilized and may be particularly useful for delimiting very young species, which will always be difficult to distinguish on the basis of molecular or morphological data alone.

4.3. Conclusions, species concepts, and implications for integrative taxonomy

The earliest stages of speciation will be the point at which it will be the hardest to establish a boundary between population and species level divergence. However, such cases (and their solution) are the "holy grail" of taxonomy and provide an exemplar for investigating the intermediate stages of the "Darwinian continuum" from varieties to species (Mallet 2008) and inevitably create problems for the definition of species. Integrative or multiple strategies may be necessary in such cases where conflicts are most likely to exist (Yeates et al. 2011; Carstens et al. 2013). Together with previous studies (Edwards and Knowles 2014; Solís-Lemus et al. 2015) we have confirmed that morphology can be a highly informative trait within an integrative approach, such as iBPP, to species delimitation.

Complex cases of species delimitation, such as those among *Pleophylla* species, demonstrate the sensitivity of delimitation approaches to prior parameter choices and are thus useful for investigating the performance of new methodologies. We have highlighted the importance of examining the effect of prior choice on species delimitation results in BPP and iBPP, especially if highly informative prior distributions ($\alpha > 1$) are used. Previously, specifying a high θ and a low τ_0 value was intended to constitute a conservative prior combination that should not lead to over-splitting

(Leaché and Fujita 2010). However, we found that this combination actually led to higher support for more splits, which was attributable to the strong influence of the θ parameter. For a conservative estimate of species delimitations, we recommend using a low value of θ to avoid species over-splitting.

The incongruence between trait- or gene-based species delimitations (Fig. III.5) may have multiple independent causes. First, sampling issues and the ability to capture statistically significant entities may be problematic, particularly for trait-based inference (Edwards and Knowles 2014) (see also above). For example, trait-based clustering algorithms quickly loose power when including too many or too poorly sampled species, or when variation is distributed over too many dimensions, resulting in more noise (Hausdorf and Hennig 2010; Solís-Lemus et al. 2015). These problems may also pose a challenge for combined approaches to species delimitation, however their impacts have not been fully explored. Second, the incongruence among independent methods, employed for the analysis of different data types (molecules versus discrete or continuous morphological traits), may be attributed to the use of competing species concepts (Fujita et al. 2012; Carstens and Satler 2013). Model based clustering applied on morphological traits is simply based on the morphological species concept; tree-based species inference methods (e.g., GMYC, PTP) are based on the phylogenetic species concept in (Sites and Marshall 2004; Zhang et al. 2014), which rely on the assumption of reciprocal monophyly across gene trees. The assumption of monophyly among independent markers may problematic because this assumption is known to be violated for closely related species. De Queiroz redefined the criteria inherent to most species concepts (De Queiroz 1998; Queiroz 2005; De Queiroz 2007) that species represent independent metapopulation lineages through time. Instead, in the generalized lineage concept (GLC) the criteria used to demarcate species (e.g., morphological differences, monophyly or reproductive isolation) are treated as attributes that accumulate during the process of lineage diversification (Queiroz 2005). This concept has been broadly adopted by coalescent-based approaches to species delimitation (O'Meara et al. 2006; Knowles and Carstens 2007; Yang and Rannala 2010; Ence and Carstens 2011; Camargo et al. 2012; Carstens et al. 2013; Edwards and Knowles 2014; Jones 2014; Yang and Rannala 2014; Jones et al. 2015), which model the lineage diversification process using multiple markers to delimit species (e.g., Edwards 2009). Several studies have delimited species successfully using these

approaches (Carstens and Dewey 2010; Leaché and Fujita 2010; Kubatko et al. 2011; Fujita et al. 2012; Niemiller et al. 2012; Carstens and Satler 2013).

BPP (and iBPP) treat species as hypotheses in a probabilistic framework, using objective tests to delineate independent evolutionary lineages (i.e., species), therefore satisfying numerous species concepts (Fujita et al. 2012). Caution should always be taken when interpreting the results of a single dataset (Carstens et al. 2013; Edwards and Knowles 2014); however, an integrative model-based approach to detecting species is likely to have more utility and could result in more robust species delimitations, especially when divergence varies across different phenotypic, genetic or ecological parameters (Edwards and Knowles 2014).

Finally, based on the outcome of the integrative iBPP analysis (Figs. III.4, III.5), which was broadly congruent with the single trait evidence, we conclude that in our *Pleophylla* data set *sp1, sp6, sp9, sp10, sp11,* and *spX2* are valid species, while *sp2* and *sp12* very likely belong to the same taxon. The results of alternative molecular delimitation methods provided support for potential cryptic species *(sp10).* However, this signal comes from one of the four markers only (which we demonstrated can overwhelm the signal of other data in the BPP/iBPP analyses, Fig. B7) and this result is not corroborated by morphological or geographical evidence (the two MOTUs occur in the same location). Therefore, at this stage we do not consider these as two separate species. (These conclusions will be further developed by formal taxonomic treatment, type material and taxonomic revision that will be presented in a separate upcoming study; Eberle et al. 2016a). Additional information about the structure of a population or species complex, based on much broader individual, geographical and DNA sequence sampling would very likely have improved our case study. However, natural rarity (linked with the time constraints of most biodiversity studies) will always have an impact on the number of available samples and may strongly bias the results (Lim et al. 2012).

Simulations have suggested that the number of loci required for robust Bayesian species delimitation may be large (Knowles and Carstens 2007). Here, we demonstrate that the signal from a single marker can influence the outcome of a fully integrative analysis, even given the inclusion of morphology. These results further underlay the necessity for upgrading the globally successful Barcoding initiatives to include a broader range of universal markers (Collins and Cruickshank 2014). De-

spite numerous disadvantages (Collins and Cruickshank 2014; Dowton et al. 2014), this approach would help to overcome some of the major challenges to accurate species delimitation (Dupuis et al. 2012). Future directions in integrative taxonomy will need to further address these issues, including integrative study design and the interpretation of frequently incongruent results. In addition the development of new tools for integrating disparate types of specimen-based data in taxonomic studies offer an exciting opportunity to free taxonomy from subjectivity.

4.4. Availability of supporting data

Voucher specimens have been deposited in the Zoological Research Museum A. Koenig (Bonn). All molecular sequences generated for this study were deposited in GenBank (Table B1). Sequence alignments, program input files and phylogenetic trees were deposited on Zenodo (DOI: `10.1186/s12862-016-0659-3`). The perl script used for running (i)BPP with multiple prior combinations, along with all input files, is available at `https://github.com/eberlejonas/BPPmulti.git`.

4.5. Acknowledgments

We would like to thank Simon Joly for advice using the JML software, Norman MacLeod for advice with the Eigenshape analysis, the referees for their helpful comments on the early draft of the manuscript, Silvia Fabrizi for assisting with the lab work, and Pia Addison and Cate Bazelet for help with the collection permit for the Cape Province. This project was supported by a studentship from the Natural Environment Research Council to R.C.M.W (NE/E522891/1), and grants from the German Science Association to D.A. (DFG/AH175/1 and AH175/3). For providing D.A. with research and collection permits, we thank the various governmental institutions and departments in Eastern Cape (Permit No.: WRO 122/07WR and WRO123/07WR), Gauteng (Permit No.: CPF6 1281), Limpopo (Permit No.: CPM-006-00001), Mpumalangma (Permit No.: MPN-2009-11-20-1232), Cape Province (Permit No: AAA0007-00097-0056), and KwaZulu-Natal (Permit Nos.: OP3752/2009, 1272/2007, 3620/2006). This work was partially supported by the computational fa-

cilities of the Advanced Computing Research Centre, University of Bristol and of the Zoological Research Museum A. Koenig, Bonn.

References

Agapow, P. M., O. R. P. Binida-Emonds, K. A. Crandall, J. C. Gittleman, G. M. Mace, J. C. Marshall, and A. Purvis (2004). The Impact of Species Concept on Biodiversity Studies. *The Quarterly Review of Biology* 79, 161–179.

Ahrens, D. and P. K. Lago (2008). Directional asymmetry reversal of male copulatory organs in chafer beetles (Coleoptera: Scarabaeidae): implications on left–right polarity determination in insect terminalia. *Journal of Zoological Systematics and Evolutionary Research* 46, 110–117. DOI: 10.1111/j.1439-0469.2007.00449.x.

Ahrens, D. (2006). The phylogeny of Sericini and their position within the Scarabaeidae based on morphological characters (Coleoptera: Scarabaeidae). *Systematic Entomology* 31, 113–144. DOI: 10.1111/j.1365-3113.2005.00307.x.

Ahrens, D., T. Fujisawa, H.-J. Krammer, J. Eberle, S. Fabrizi, and A. P. Vogler (2016). Rarity and Incomplete Sampling in DNA-based Species Delimitation. *Systematic Biology* 65, 478–494.

Ahrens, D., M. T. Monaghan, and A. P. Vogler (2007). DNA-based taxonomy for associating adults and larvae in multi-species assemblages of chafers (Coleoptera: Scarabaeidae). *Molecular phylogenetics and evolution* 44, 436–49. DOI: 10.1016/j.ympev.2007.02.024.

Ahrens, D. and A. P. Vogler (2008). Towards the phylogeny of chafers (Sericini): analysis of alignment-variable sequences and the evolution of segment numbers in the antennal club. *Molecular phylogenetics and evolution* 47, 783–98. DOI: 10.1016/j.ympev.2008.02.010.

Altekar, G., S. Dwarkadas, J. P. Huelsenbeck, and F. Ronquist (2004). Parallel metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* 20, 407–415.

Anisimova, M., M. Gil, J. Dufayard, C. Dessimoz, and O. Gascuel (2011). Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Systematic biology* 60, 685–699.

Beckett, M. (2012). "The distribution patterns in *Pleophylla* species (Coleptera: Scarabaeidae) – indicators of ancient forest distributions." PhD thesis. Bonn: Rheinische Friedrich-Wilhelms-Universität Bonn.

Blackwelder, R. E. (1967). A Critique of Numerical Taxonomy. *Systematic Zoology* 16, 64. DOI: 10.2307/2411518.

Bocak, L., C. Barton, A. Crampton-Platt, D. Chesters, D. Ahrens, and A. P. Vogler (2014). Building the Coleoptera tree-of-life for >8000 species: composition of public DNA data and fit with Linnaean classification. *Systematic Entomology* 39, 97–110. DOI: 10.1111/syen.12037.

Bossu, C. M. and T. J. Near (2009). Gene Trees Reveal Repeated Instances of Mitochondrial DNA Introgression in Orangethroat Darters (Percidae: Etheostoma). *Systematic Biology* 58. DOI: 10.1093/sysbio/syp014.

Britton, T., C. L. Anderson, D. Jacquet, S. Lundqvist, and K. Bremer (2007). Estimating divergence times in large phylogenetic trees. *Systematic Biology* 56, 741–752. DOI: 10.1080/10635150701613783.

Camargo, A., M. Morando, L. J. Avila, and J. W. Sites (2012). Species delimitation with ABC and other coalescent-based methods: a test of accuracy with simulations and an empirical example with lizards of the *Liolaemus darwinii* complex (Squamata: Liolaemidae). *Evolution* 66, 2834–2849. DOI: 10.5061/dryad.4409k652.

Carstens, B. C. and T. A. Dewey (2010). Species Delimitation Using a Combined Coalescent and Information-Theoretic Approach: An Example from North American *Myotis* Bats. *Systematic Biology* 59, 400–414. DOI: 10.1093/sysbio/syq024.

Carstens, B. C., T. A. Pelletier, N. M. Reid, and J. D. Satler (2013). How to fail at species delimitation. *Molecular ecology* 22, 4369–4383. DOI: 10.1111/mec.12413.

Carstens, B. C. and J. D. Satler (2013). The carnivorous plant described as *Sarracenia alata* contains two cryptic species. *Biological Journal of the Linnean Society* 109, 737–746. DOI: 10.1111/bij.12093.

Celeux, G. and G. Govaert (1995). Gaussian parsimonious clustering models. *Pattern Recognition* 28, 781–793. DOI: 10.1016/0031-3203(94)00125-6.

Clement, M., D. Posada, and K. A. Crandall (2000). TCS: a computer program to estimate gene genealogies. *Molecular Ecology* 9, 1657–1659. DOI: 10.1046/j.1365-294x.2000.01020.x.

Collins, R. A. and R. H. Cruickshank (2014). Known Knowns, Known Unknowns, Unknown Unknowns and Unknown Knowns in DNA Barcoding: A Comment on Dowton et al. *Systematic Biology* 63, 1005–1009. DOI: 10.1093/sysbio/syu060.

Dalla Torre, K. W. (1912). "Scarabaeidae: Melolonthinae I." In: *Coleopterorum Catalogus 45*.

Daugherty, C. H., A. Cree, J. M. Hay, and M. B. Thompson (1990). Neglected taxonomy and continuing extinctions of tuatara *(Sphenodon). Nature* 347, 177–179. DOI: 10.1038/347177a0.

De Queiroz, K. (1998). "The General Lineage Concept of Species, Species Criteria, and the Process of Speciation." In: *Endless Forms: Species and Speciation*. Ed. by D. J. Howard and S. H. Berlocher. New York: Oxford University Press; New York. Chap. 5, pp. 57–75. DOI: 10.1080/10635150701701083.

— (2007). Species Concepts and Species Delimitation. *Systematic Biology* 56, 879–886. DOI: 10.1080/10635150701701083.

Degnan, J. H. and N. A. Rosenberg (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* 24, 332–340. DOI: 10.1016/j.tree.2009.01.009.

Dowton, M., K. Meiklejohn, S. L. Cameron, and J. Wallman (2014). A Preliminary Framework for DNA Barcoding, Incorporating the Multispecies Coalescent. *Systematic Biology* 63, 639–644. DOI: 10.1093/sysbio/syu028.

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut (2006). Relaxed Phylogenetics and Dating with Confidence. *PLoS Biology* 4. Ed. by D. Penny, e88. DOI: 10.1371/journal.pbio.0040088.

Drummond, A. J., S. Y. W. Ho, N. Rawlence, and A. Rambaut (2007). *A Rough Guide to BEAST 1.4*.

Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular biology and evolution* 29, 1969–1973. DOI: 10.1093/molbev/mss075.

Dupuis, J. R., A. D. Roe, and F. A. H. Sperling (2012). Multi-locus species delimitation in closely related animals and fungi: One marker is not enough. *Molecular Ecology* 21, 4422–4436. DOI: 10.1111/j.1365-294X.2012.05642.x.

Eberhard, W. G. (1985). *Sexual selection and animal genitalia*. Cambridge: Harvard University Press.

Eberle, J., M. Beckett, A. Özguel-Siemund, J. Frings, S. Fabrizi, and D. Ahrens (2016a). Afromontane forests hide nineteen new species of an ancient chafer lineage (Coleoptera: Scarabaeidae): *Pleophylla* – phylogeny and taxonomic revision. *Zoological Journal of the Linnean Society – in press*.

Eberle, J., S. Fabrizi, P. Lago, and D. Ahrens (2016b). A historical biogeography of megadiverse Sericini – another story "out of Africa"? *Cladistics*, article first published online. DOI: 10.1111/cla.12162.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32, 1792–1797. DOI: 10.1093/nar/gkh340.

Edwards, D. L. and L. L. Knowles (2014). Species detection and individual assignment in species delimitation: can integrative data increase efficacy? *Proceedings of the Royal Society B: Biological Sciences* 281, 20132765. DOI: 10.1098/rspb.2013.2765.

Edwards, S. V., L. Liu, and D. K. Pearl (2007). High-resolution species trees without concatenation. *Proceedings of the National Academy of Sciences* 104, 5936–5941. DOI: 10.1073/pnas.0607004104.

Edwards, S. V. (2009). Is a New and General Theory of Molecular Systematics Emerging? *Evolution* 63, 1–19. DOI: 10.1111/j.1558-5646.2008.00549.x.

Ence, D. D. and B. C. Carstens (2011). SpedeSTEM: a rapid and accurate method for species delimitation. *Molecular ecology resources* 11, 473–80. DOI: 10.1111/j.1755-0998.2010.02947.x.

Ezard, T. H. G., T. Fujisawa, and T. G. Barraclough. *SPLITS: Species' Limits by Threshold Statistics R package. 2009. http://barralab.bio.ic.ac.uk. Accessed May 2012.*

Ezard, T. H. G., P. N. Pearson, and A. Purvis (2010). Algorithmic approaches to aid species' delimitation in multidimensional morphospace. *BMC evolutionary biology* 10, 175. DOI: 10.1186/1471-2148-10-175.

Fontaneto, D., E. a. Herniou, C. Boschetti, M. Caprioli, G. Melone, C. Ricci, and T. G. Barraclough (2007). Independently evolving species in asexual bdelloid rotifers. *PLoS biology* 5. Ed. by M. A. F. Noor, e87. DOI: 10.1371/journal.pbio.0050087.

Fraley, C. and A. E. Raftery (2002). Model-Based Clustering, Discriminant Analysis, and Density Estimation. *Journal of the American Statistical Association* 97, 611–631. DOI: 10.1198/016214502760047131.

Fraley, C. and A. E. Raftery (2007). Bayesian Regularization for Normal Mixture Estimation and Model-Based Clustering. *Journal of Classification* 24, 155–181. DOI: 10.1007/s00357-007-0004-5.

Fraley, C., A. E. Raftery, T. B. Murphy, L. Scrucca, M. Brendan, and L. Scrucca (2012). mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. *Technical Report 597, University of Washington*, 1–50.

Fujisawa, T. and T. G. Barraclough (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, 707–724. DOI: 10.1093/sysbio/syt033.

Fujita, M. K. M. M. K., A. D. Leaché, F. F. T. Burbrink, J. J. a. McGuire, and C. Moritz (2012). Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology and Evolution* 27, 480–488. DOI: 10.1016/j.tree.2012.04.012.

Guillot, G., S. Renaud, R. Ledevin, J. Michaux, and J. Claude (2012). A Unifying Model for the Analysis of Phenotypic, Genetic, and Geographic Data. *Systematic biology* 61, 897–911. DOI: 10.1093/sysbio/sys038.

Guindon, S., J.-F. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology* 59, 307–321. DOI: 10.1093/sysbio/syq010.

Hart, M. W. and J. Sunday (2007). Things fall apart: biological species form unconnected parsimony networks. *Biology Letters* 3, 509–512. DOI: 10.1098/rsbl.2007.0307.

Hausdorf, B. and C. Hennig (2010). Species Delimitation Using Dominant and Codominant Multilocus Markers. *Systematic Biology* 59, 491–503. DOI: 10.1093/sysbio/syq039.

Heled, J., D. Bryant, and A. J. Drummond (2013). Simulating gene trees under the multispecies coalescent and time-dependent migration. *BMC Evolutionary Biology* 13, 44. DOI: 10.1186/1471-2148-13-44.

Heled, J. and A. J. Drummond (2010). Bayesian inference of species trees from multilocus data. *Molecular biology and evolution* 27, 570–580. DOI: 10.1093/molbev/msp274.

Hudson, R. R. and J. A. Coyne (2002). Mathematical Consequences of the Genealogical Species Concept. *Evolution* 56, 1557–1565. DOI: 10.1111/j.0014-3820.2002.tb01467.x.

Huelsenbeck, J. P. and F. Ronquist (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics (Oxford, England)* 17, 754–5.

Hunt, T., J. Bergsten, Z. Levkanicova, A. Papadopoulou, O. S. John, R. Wild, P. M. Hammond, D. Ahrens, M. Balke, M. S. Caterino, J. Gomez-Zurita, I. Ribera, T. G. Barraclough, M. Bocakova, L. Bocak, A. P. Vogler, J. Gómez-Zurita, I. Ribera, T. G. Barraclough, M. Bocakova, L. Bocak, and A. P. Vogler (2007). A Comprehensive Phylogeny of Beetles Reveals the Evolutionary Origins of a Superradiation. *Science* 318, 1913–1916. DOI: 10.1126/science.1146954.

Isaac, N. (2004). Taxonomic inflation: its influence on macroecology and conservation. *Trends in Ecology & Evolution* 19, 464–469. DOI: 10.1016/j.tree.2004.06.004.

Joly, S. (2012). JML: testing hybridization from species trees. *Molecular Ecology Resources* 12, 179–184. DOI: 10.1111/j.1755-0998.2011.03065.x.

Joly, S., P. A. McLenachan, and P. J. Lockhart (2009). A Statistical Approach for Distinguishing Hybridization and Incomplete Lineage Sorting. *The American Naturalist* 174, E54–E70. DOI: 10.1086/600082.

Jones, G. (2014). *STACEY: species delimitation and phylogeny estimation under the multispecies coalescent*. Tech. rep. DOI: 10.1101/010199.

Jones, G., Z. Aydin, and B. Oxelman (2015). DISSECT: an assignment-free Bayesian discovery method for species delimitation under the multispecies coalescent. *Bioinformatics* 31, 991–998. DOI: 10.1093/bioinformatics/btu770.

Keck, B. P. and T. J. Near (2010). A young clade repeating an old pattern: diversity in Nothonotus darters (Teleostei: Percidae) endemic to the Cumberland River. *Molecular Ecology* 19, 5030–5042. DOI: 10.1111/j.1365-294X.2010.04866.x.

Kerdelhué, C., G. Roux-Morabito, J. Forichon, J.-M. Chambon, A. Robert, and F. Lieutier (2002). Population genetic structure of *Tomicus piniperda* L. (Curculionidae: Scolytinae) on different pine species and validation of *T. destruens* (Woll.) *Molecular Ecology* 11, 483–494.

Knowles, L. L. and B. C. Carstens (2007). Delimiting species without monophyletic gene trees. *Systematic biology* 56, 887–895. DOI: 10.1080/10635150701701091.

Kubatko, L. S., B. C. Carstens, and L. L. Knowles (2009). STEM: species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics* 25, 971–973. DOI: 10.1093/bioinformatics/btp079.

Kubatko, L. S., H. L. Gibbs, and E. W. Bloomquist (2011). Inferring Species-Level Phylogenies and Taxonomic Distinctiveness Using Multilocus Data in Sistrurus Rattlesnakes. *Systematic Biology* 60, 393–409. DOI: 10.1093/sysbio/syr011.

Lanfear, R., B. Calcott, S. Y. W. Ho, and S. Guindon (2012). PartitionFinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. *Molecular biology and evolution* 29, 1695–1701. DOI: 10.1093/molbev/mss020.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. DOI: 10.1093/bioinformatics/btm404.

Leaché, A. D. and M. K. Fujita (2010). Bayesian species delimitation in West African forest geckos *(Hemidactylus fasciatus)*. *Proceedings of the Royal Society B: Biological Sciences* 277, 3071–3077. DOI: 10.1098/rspb.2010.0662.

Leache, A. D., M. S. Koo, C. L. Spencer, T. J. Papenfuss, R. N. Fisher, and J. A. McGuire (2009). Quantifying ecological, morphological, and genetic variation to delimit species in the coast horned lizard species complex (Phrynosoma). *Proceedings of the National Academy of Sciences* 106, 12418–12423. DOI: 10.1073/pnas.0906380106.

Lim, G. S., M. Balke, and R. Meier (2012). Determining species boundaries in a world full of rarity: singletons, species delimitation methods. *Systematic biology* 61, 165–169. DOI: 10.1093/sysbio/syr030.

MacLeod, N. and K. D. Rose (1993). Inferring locomotor behavior in Paleogene mammals via eigenshape analysis. *American Journal of Science* 293, 300–355. DOI: 10.2475/ajs.293.A.300.

MacLeod, N. (1999). Generalizing and extending the eigenshape method of shape space visualization and analysis. *Paleobiology* 25, 107–138.

Maddison, W. P. (1997). Gene Trees in Species Trees. *Systematic Biology* 46, 523–536. DOI: 10.1093/sysbio/46.3.523.

Mallet, J. (2008). Mayr's view of Darwin: was Darwin wrong about speciation? *Biological Journal of the Linnean Society* 95, 3–16. DOI: 10.1111/j.1095-8312.2008.01089.x.

Marshall, D. C. (2010). Cryptic Failure of Partitioned Bayesian Phylogenetic Analyses: Lost in the Land of Long Trees. *Systematic Biology* 59, 108–117. DOI: 10.1093/sysbio/syp080.

Mayrose, I., N. Friedman, and T. Pupko (2005). A Gamma mixture model better accounts for among site rate heterogeneity. *Bioinformatics* 21, 151–158.

McLachlan, G. J. and K. E. Basford (1988). *Mixture Models: Inference and Applications to Clustering*. New York: Marcel Dekker.

Monaghan, M. T., D. J. G. Inward, T. Hunt, and A. P. Vogler (2007). A molecular phylogenetic analysis of the Scarabaeinae (dung beetles). *Molecular Phylogenetics and Evolution* 45, 674–692. DOI: 10.1016/j.ympev.2007.06.009.

Niemiller, M. L., T. J. Near, and B. M. Fitzpatrick (2012). Delimiting Species Using Multilocus Data: Diagnosing Cryptic Diversity in the Southern Cavefish, *Typhlichthys subterraneus* (Teleostei: Amblyopsidae). *Evolution* 66, 846–866. DOI: 10.1111/j.1558-5646.2011.01480.x.

Olave, M., E. Solà, and L. L. Knowles (2014). Upstream Analyses Create Problems with DNA-Based Species Delimitation. *Systematic Biology* 63, 263–271. DOI: 10.1093/sysbio/syt106.

O'Meara, B. C. (2010). New heuristic methods for joint species delimitation and species tree inference. *Systematic Biology* 59, 59–73. DOI: 10.1093/sysbio/syp077.

O'Meara, B. C., M. J. Sanderson, and P. C. Wainwright (2006). Testing for different rates of continuous trait evolution using likelihood. *Bio One* 60, 922–933.

Padial, J. and I. De la Riva (2006). Taxonomic Inflation and the Stability of Species Lists: The Perils of Ostrich's Behavior. *Systematic Biology* 55, 859–867. DOI: 10.1080/10635150600081588.

Pattengale, N. D., M. Alipour, O. R. Bininda-Emonds, B. M. Moret, and A. Stamatakis (2009). "How many bootstrap replicates are necessary?" In: *Annual International Conference on Research in Computational Molecular Biology*. Ed. by S. Batzoglou. Springer, pp. 184–200.

Pons, J., T. G. Barraclough, J. Gomez-Zurita, A. Cardoso, D. P. Duran, S. Hazell, S. Kamoun, W. D. Sumlin, and A. P. Vogler (2006). Sequence-Based Species Delimitation for the DNA Taxonomy of Undescribed Insects. *Systematic Biology* 55, 595–609. DOI: 10.1080/10635150600852011.

Powell, J. R. (2012). Accounting for uncertainty in species delineation during the analysis of environmental DNA sequence data. *Methods in Ecology and Evolution* 3, 1–11. DOI: 10.1111/j.2041-210X.2011.00122.x.

Puillandre, N., A. Lambert, S. Brouillet, and G. Achaz (2012a). ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology* 21, 1864–1877. DOI: 10.1111/j.1365-294X.2011.05239.x.

Puillandre, N., M. V. Modica, Y. Zhang, L. Sirovich, M. C. Boisselier, C. Cruaud, M. Holford, and S. Samadi (2012b). Large-scale species delimitation method for hyperdiverse groups. *Molecular Ecology* 21, 2671–2691. DOI: 10.1111/j.1365-294X.2012.05559.x.

Puorto, G., M. Da Graça Salomão, R. D. G. Theakston, R. S. Thorpe, D. A. Warrell, and W. Wüster (2001). Combining mitochondrial DNA sequences and morphological data to infer species boundaries: phylogeography of lanceheaded pitvipers in the Brazilian Atlantic forest, and the status of Bothrops pradoi (Squamata: Serpentes: Viperidae). *Journal of Evolutionary Biology* 14, 527–538. DOI: 10.1046/j.1420-9101.2001.00313.x.

Queiroz, K. de (2005). Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences* 102, 6600–6607. DOI: 10.1073/pnas.0502030102.

Rambaut, A. and M. Charleston. *TreeEdit 1.0. http://tree.bio.ed.ac.uk/software/treeedit/. Accessed May 2012.*

Rannala, B. (2015). *Are molecular taxonomists lost upstream?* http://phylogenyetc.tumblr.com/post/78791524128/are-molecular-taxonomists-lost-upstream. (Accessed on Dec. 1, 2015).

Rannala, B., T. Zhu, and Z. Yang (2012). Tail Paradox, Partial Identifiability, and Influential Priors in Bayesian Branch Length Inference. *Molecular Biology and Evolution* 29, 325–335. DOI: 10.1093/molbev/msr210.

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3, 217–223. DOI: 10.1111/j.2041-210X.2011.00169.x.

Ronquist, F. and J. P. Huelsenbeck (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574. DOI: 10.1093/bioinformatics/btg180.

Sanderson, M. J. (2003). r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19, 301–302. DOI: 10.1093/bioinformatics/19.2.301.

Santos, H., J. Rousselet, E. Magnoux, M.-R. Paiva, M. Branco, and C. Kerdelhue (2007). Genetic isolation through time: allochronic differentiation of a phenologically atypical population of the pine processionary moth. *Proceedings of the Royal Society B: Biological Sciences* 274, 935–941. DOI: 10.1098/rspb.2006.3767.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics* 6, 461–464. DOI: 10.1214/aos/1176344136.

Shaffer, H. B. and R. C. Thomson (2007). Delimiting Species in Recent Radiations. *Systematic Biology* 56, 896–906. DOI: 10.1080/10635150701772563.

Simmons, L. W. (2014). Sexual selection and genital evolution. *Austral Entomology* 53, 1–17. DOI: 10.1111/aen.12053.

Simon, C., F. Frati, A. 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. *Annals of the Entomological Society of America* 87, 651–701.

Sites, J. W. and J. C. Marshall (2004). Operational criteria for delimiting species. *Annual review of Ecology, Evolution and Systematics* 35, 199–227. DOI: 10.2307/annurev.ecolsys.35.112202.30000009.

Slowinski, J. B., A. Knight, and A. P. Rooney (1997). Inferring Species Trees from Gene Trees: A Phylogenetic Analysis of the Elapidae (Serpentes) Based on the Amino Acid Sequences of Venom Proteins. *Molecular Phylogenetics and Evolution* 8, 349–362. DOI: 10.1006/mpev.1997.0434.

Sneath, P. H. A. and R. R. Sokal (1962). Numerical Taxonomy. *Nature* 193, 855–860. DOI: 10.1038/193855a0.

Solís-Lemus, C., L. L. Knowles, and C. Ané (2015). Bayesian species delimitation combining multiple genes and traits in a unified framework. *Evolution* 69, 492–507. DOI: 10.1111/evo.12582.14.

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.

Stamatakis, A., P. Hoover, and J. Rougemont (2008). A Rapid Bootstrap Algorithm for the RAxML Web Servers. *Systematic Biology* 57, 758–771. DOI: 10.1080/10635150802429642.

Sterner, B. (2014). "Well-Structured Biology – Numerical Taxonomy's Epistemic Vision for Systematics." In: *Patterns of Nature*. Ed. by A. Hamilton. California: University of California Press, pp. 213–244.

Sullivan, J., D. L. Swofford, and G. J. Naylor (1999). The effect of taxon sampling on estimating rate heterogeneity parameters of maximum-likelihood models. *Molecular Biology and Evolution* 16, 1347–1356.

Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* 30, 2725–2729. DOI: 10.1093/molbev/mst197.

Templeton, A. R., K. A. Crandall, and C. F. Sing (1992). A Cladistic-Analysis of Phenotypic Associations with Haplotypes Inferred from Restriction Endonuclease Mapping and DNA-Sequence Data. 3. Cladogram Estimation. *Genetics* 132, 619–633.

Venables, W. N. and B. D. Ripley (2002). *Modern Applied Statistics with S*. 4th ed. New York: Springer.

Vogler, A. P. and R. Desalle (1994). Evolution and Phylogenetic Information Content of the ITS-1 Region in the Tiger Beetle Cicindela dorsalis. *Molecular biology and evolution* 11, 393–405.

Vogler, A. P. and M. T. Monaghan (2007). Recent advances in DNA taxonomy. *Journal of Zoological Systematics and Evolutionary Research* 45, 1–10. DOI: 10.1111/j.1439-0469.2006.00384.x.

Wiens, J. J. and T. A. Penkrot (2002). Delimiting Species Using DNA and Morphological Variation and Discordant Species Limits in Spiny Lizards *(Sceloporus)*. *Systematic Biology* 51, 69–91. DOI: 10.1080/106351502753475880.

Wu, C. A. and D. R. Campbell (2005). Cytoplasmic and nuclear markers reveal contrasting patterns of spatial genetic structure in a natural Ipomopsis hybrid zone. *Molecular Ecology* 14, 781–792. DOI: 10.1111/j.1365-294X.2005.02441.x.

Yang, Z. (1993). Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Molecular biology and evolution* 10, 1396–1401.

— (2015). The BPP program for species tree estimation and species delimitation. *Current Zoology* 61, 854–865. DOI: 10.1093/czoolo/61.5.854.

Yang, Z. and B. Rannala (2010). Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America* 107, 9264–9269. DOI: 10.1073/pnas.0913022107.

— (2014). Unguided Species Delimitation Using DNA Sequence Data from Multiple Loci. *Molecular Biology and Evolution* 31, 3125–3135. DOI: 10.1093/molbev/msu279.

Yeates, D. K., A. Seago, L. Nelson, S. L. Cameron, L. Joseph, and J. W. H. Trueman (2011). Integrative taxonomy, or iterative taxonomy? *Systematic Entomology* 36, 209–217. DOI: 10.1111/j.1365-3113.2010.00558.x.

Zahn, C. T. and R. Z. Roskies (1972). Fourier Descriptors for Plane Closed Curves. *IEEE Transactions on Computers* C-21, 269–281.

Zhang, C., B. Rannala, and Z. Yang (2014). Bayesian Species Delimitation Can Be Robust to Guide-Tree Inference Errors. *Systematic biology* 63, 993–1004. DOI: 10.1093/sysbio/syu052.

Zhang, J., P. Kapli, P. Pavlidis, and A. Stamatakis (2013). A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* 29, 2869–2876. DOI: 10.1093/bioinformatics/btt499.

Chapter IV.

Afromontane forests hide nineteen new species of an ancient chafer lineage (Coleoptera: Scarabaeidae: Melolonthinae): *Pleophylla* Erichson, 1847 – phylogeny and taxonomic revision

This chapter is accepted for publication in:
Eberle, J., M. Beckett, A. Özgül-Siemund, J. Frings, S. Fabrizi, and D. Ahrens (in press). Afromontane forests hide 19 new species of an chafer lineage (Coleoptera: Scarabaeidae: Melolonthinae): *Pleophylla* Erichson, 1847 – phylogeny and taxonomic revision. *Zoological Journal of the Linnean Society.* DOI:10.1111/zoj.12489

Authors' contributions to the original article:
Molecular lab work: JE, SF, DA; sequence assembly and alignments, molecular phylogeny: JE; morphological phylogeny: DA; fieldwork collections: DA, SF, JE; specimen preparation and label data collection: JE, MB, AÖS, JF; distribution maps: JE; SEM and stack photography: MB, AÖS, JF, SF; taxonomic descriptions, identification key: DA; manuscript design and writing: JE, DA.

1. Introduction

Among phytophagous scarab beetles, *Pleophylla* is a highly conspicuous genus, found only in isolated parts of the South African escarpment and the East African highlands where it seems to be closely associated with indigenous Afromontane forests. Little is known about the biology, the taxonomy, and distribution of the genus. *Pleophylla* belongs to the tribe Sericini (Coleoptera: Scarabaeidae), a highly diverse clade of herbivorous beetles with nearly 4,000 described species. Sericini adults are polyphagous and feed on a variety of angiosperms, while their larvae feed on humus and plant roots in the upper soil layers. Morphological and molecular evidence has shown that *Pleophylla* belongs to one of the more basal-branching lineages of Sericini, together with its presumptive sister group, *Omaloplia* (Ahrens 2006; Eberle et al. 2016a). Their split dates back to ca 79 Mya, consequently *Pleophylla* represents one of the most ancient Sericini lineages (Eberle et al. 2016a).

Pleophylla was established by Erichson (1847) without any nominal species included. The first subsequently included species was *Pleophylla fasciatipennis* Blanchard 1850. Nearly contemporarily, Burmeister (1855) and Boheman (1857) added further taxa, *P. navicularis* and *P. pilosa* / *P. maculipennis*, respectively. In the later revision of Schaufuss (1871), another two nominal species were described, *P. flavicornis* and *P. opalina*. The last taxonomic treatment of the genus dates back to Péringuey (1904) who established one new species (*P. tongaatsana*) but synonymised all species, except *P. navicularis*, with *P. fasciatipennis*. This status was adopted by all subsequent authors (Dalla Torre 1912; Krajčík 2012; The Global Biodiversity Information Facility 2014).

Recent morphometric and molecular evidence from an integrative analysis (Eberle et al. 2016b) revealed, that the current taxonomic classification, which recognised only three valid species (Dalla Torre 1912; `globalspecies.org/ntaxa/2359831`; accessed March 5, 2014), is largely outdated. Based on a morphometric shape analysis of the

parameres, Eberle et al. (2016b) showed that shape differences of male copulation organs are highly significant amongst the different morphospecies, being also congruent with the DNA taxonomy of a fast evolving nuclear molecular marker (ITS1) and with results from Bayesian species delimitation using four genes and morphometric genital traits (parameres). For young speciation events (between 0.17 and 0.65 Mya), shape divergence of male genitalia was found to be even faster than genetic lineage sorting of some fast-evolving mitochondrial DNA markers such as *cox1*, that were often not able to distinguish between highly different morphotypes due to incomplete lineage sorting (Eberle et al. 2016b). These findings confirmed male genital morphological characteristics (Ahrens and Lago 2008) as reliable markers for species taxonomy in *Pleophylla*.

Based on the results of Eberle et al. (2016b), a vast diversity of morphologically highly distinct species was found among the material that was available for the present study. We investigate the taxonomy of *Pleophylla* species by thorough revision of type specimens, and reconstruct a phylogenetic tree of the species using their adult morphological characters. Based on the specimens examined from a selection of natural history collections, we explore distribution patterns and co-distributions of the species to enhance taxonomic hypotheses. We present an updated and extended DNA taxonomy of *Pleophylla*. Additionally, to accomplish an integrative taxonomic approach, female type specimens were matched with specimens that were identified by DNA using morphological evidence from morphometric analysis of female copulation organs (Özgül-Siemund and Ahrens 2015).

2. Material and Methods

2.1. DNA sequencing

A total of 320 individuals of 10 putative morphospecies of the genus *Pleophylla* were collected from 26 localities in South Africa (Supplementary Table IV.1-IV.2). *Omaloplia nigromarginata* (Herbst, 1785) and *O. ruricola* (Fabricius, 1775) from the putative sister group *Omaloplia* (Ahrens 2006) were included as outgroup taxa. Samples were preserved in absolute ethanol and, following DNA extraction, individuals were dry mounted and preserved for morphological study. Voucher specimens have been deposited in the Zoologisches Forschungsmuseum Alexander Koenig (ZFMK, Bonn).

DNA was extracted non-destructively from leg or thorax tissue using the Promega Wizard SV96 Plate extraction kit according to the manufacturer's protocol. An 826 bp fragment of the 3' end of the mitochondrial gene cytochrome oxidase subunit 1 (*cox1*) was amplified using primers C1-J-2183 (Jerry) and TL2-N-3014 (Pat) (Simon et al. 1994). The nuclear internal transcribed spacer 1 (ITS1) (636-723 bp) was amplified using the primers 5'-GTAGGTGAACCTGCAGAAGG-3' and 5'-GCGTTCGAARTGCGATGATCAA-3' (Vogler and Desalle 1994). These primers are sometimes referred to as ITS1R and ITS1F, respectively (e.g., Kerdelhué et al. 2002; Santos et al. 2007), but note that ITS1R binds to the 5' 18S region, while ITS1F binds to the 3' 5.8S region of this product. Amplified products were sequenced in both directions using ABI BigDye technology and an AB1 PRISM 3730 DNA Analyzer (Applied Biosystems) at the Natural History Museum (London) (Eberle et al. 2016b); later sequences were produced at the ZFMK with Macrogen (Seoul, South Korea) using the same primers. Contiguous sequences were assembled from both strands and edited using Sequencher 4 (Gene Codes Corporation, Ann Harbor, MI, USA) and Geneious ® 7.1.8 (Kearse et al. 2012). All sequences were deposited in GenBank. Accession numbers, identifications and specimen vouchers are provided

in Supplementary Table IV.1. All vouchers of DNA specimens are deposited in the collection of the Zoological Research Museum A. Koenig, Bonn (ZFMK).

2.2. Sequence alignment and DNA-based phylogenetic analysis

Sequences were aligned for each marker separately using MAFFT (v7.017; Katoh et al. 2002) and subsequently checked by eye in Geneious® 7.1.8. Inferences of gene trees were conducted with Maximum Likelihood (ML) using RAxML (Stamatakis 2014) and with Bayesian Methods using MrBayes (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003). Nucleotide substitution models and optimal partition schemes for *cox1*-codon-positions were set according to PartitionFinder (Lanfear et al. 2012) analyses, which were conducted separately for each setting, choosing from the set of substitution models available for the respective program (see below). The RAxML tree search parameters were adapted to the respective alignments by running multiple analyses with different settings on 5 initial randomized maximum parsimony trees: initial rearrangement settings and the optimal number of gamma categories were chosen based on likelihood scores under the gamma model of rate heterogeneity. Final tree searches under the GTRCAT model of nucleotide substitution were conducted on 20 initial randomized maximum parsimony trees with an initial rearrangement setting of 10. For *cox1* and ITS1, 55 and 25 rate categories were applied, respectively. The *cox1* alignment was split in two partitions (codon positions (1+2),3). The best known tree was again chosen by the gamma-based likelihood score. For Bayesian phylogenetic inference with MrBayes, the *cox1* alignment was divided into 3 codon position partitions (cp1–3) and analysed under the substitution models that were inferred with PartitionFinder (cp1: GTR+G, cp2: GTR+I, cp3: F81+I). ITS1 was analysed under the K80+G model. Each analysis was run with one heated (temperature = 0.1) and three cold chains with a temperature of 0.1. The analysis of *cox1* was run for 50 million generations, sampling every 5,000 generations, and the analysis of ITS1 was run for 10 million generations, sampling every 1,000 generations. Convergence and stationarity of runs was assessed with Tracer 1.6 (Drummond et al. 2012, available from `http://tree.bio.ed.ac.uk/software/tracer/`). A burnin of 10% was removed before summarizing the trees.

2.3. Morphology based cladistic analysis

Terms and methods used for measurements, specimen dissection and preparation of genitalia are as used by Ahrens (2004). The data matrix was assembled in NDE (Page 2001). The matrix for the cladistic analysis contains 40 morphological characters of external and genital morphology. Female genital characters were excluded since they were not available for all species of *Pleophylla*. Inapplicable characters were coded as '–', whereas unknown character states were coded as '?' (Strong and Lipscomb 1999). The parsimony analysis was performed in NONA 2.0 (Goloboff 1999) using the parsimony ratchet (Nixon 1999). WINCLADA 1.00.08 (Nixon 2002) served as a frontend to NONA. In a first run, all characters were treated equally weighted and non-additive. Additionally, implied weighting was used in a second run. The number of characters to be sampled for re-weighting during the parsimony ratchet was determined to be ten. Two hundred iterations were performed (one tree hold per iteration) per analysis. All searches were run under the collapsing option 'ambiguous', which collapses every node whose minimum length is 0. State transformations were considered to be apomorphies of a given node only if they were unambiguous (i.e. without arbitrary selection of accelerated or delayed optimization) and if they were shared by all dichotomized most-parsimonious trees. Bootstrap values were calculated using NONA to evaluate the trees. Character changes were mapped on the consensus tree using WINCLADA. Additionally, Bayesian inference (BI) was performed using MrBayes 3.2 (Huelsenbeck and Ronquist 2001) with 1.2×10^6 generations using the Mk model (Markov k model; Lewis 2001) (further settings: preset applyto=(all) ratepr=variable; lset rates= gamma). We sampled every 1000 generations and discarded the first 300 samples (25%) of the trees to calculate the consensus trees (halfcompat and allcompat option). The morphological data matrix is presented in Table IV.3.

2.4. Taxonomic revision

Data of specimens examined are cited in the text with original label contents given in quotation marks, multiple labels are separated by a "/". Due to the extensive material examined here (ca 1000 specimens), we present in the main paper only locality data

of examined primary type material (and short paratype citations), while all other locality data, including the non-type material, are provided as electronic supplementary material C1. Measurements refer to the maximum extension of the specimen or the named structure. Diagnostic descriptions were generated automatically from the NDE character matrix omitting synapomorphies of *Pleophylla* and some subsequent manual editing. Male genitalia were examined from all male specimens for identification and glued to a small pointed card which was attached to the specimen's pin. Genitalia (in both lateral and dorsal views) and habitus of type specimens were photographed using a stereomicroscope Leica M125 with a Leica DC420C digital camera. In the image stacking software Leica Application Suite (V3.3.0), a number of single focused images were combined in order to obtain an entirely focused image. The resulting images were subsequently digitally edited using Artweaver® 0.5 to remove the background. The distribution maps were generated using the QGIS 2.0.1 software (`www.qgis.org`).

Abbreviations used in the text for collection depositories are as follows:

CPPB	Collection Petr Pacholátko, Brno, Czech Republic;
MLUH	Martin-Luther-Universität, Wissenschaftsbereich Zoologie, Halle (Saale), Germany;
NHRS	Naturhistoriska Riksmuseet Stockholm, Sweden;
NME	Naturkundemuseum Erfurt, Germany;
RMCA	Musee Royal de l'Afrique Centrale, Tervuren, Belgium;
SAMC	South African Museum, Cape Town, South Africa;
SANC	South African National Insect collection, Pretoria, South Africa;
SMNS	Stuttgarter Museum für Naturkunde, Stuttgart, Germany;
TMSA	Transvaal Museum (now Ditsong Museum), Pretoria, South Africa;
USNM	Smithsonian National Museum of Natural History, Washington D.C., USA;
ZFMK	Zoolgisches Forschungsmuseum A. Koenig, Bonn, Germany;
ZMAN	Universiteit van Amsterdam, Zoologisch Museum Amsterdam (now: Naturalis Leiden), Netherlands;
ZMHB	Museum für Naturkunde, Berlin, Germany.

3. Results

3.1. Taxonomy

3.1.1. *Pleophylla* Erichson, 1847

Pleophylla Erichson 1847: 695 (type species *Omaloplia fasciatipennis* Blanchard 1850; by monotypy); Blanchard 1850: 83; Burmeister 1855: 180; Schaufuss 1871 231; Brenske 1899: 83; Péringuey 1904: 6; Dalla Torre 1912: 44; Machatschke 1959: 744; Krajčík 2012: 216.

Diagnosis. Body of medium size (7.5-13.5 mm), elongate, dark brown, sometimes with greenish shine, moderately shiny; dorsal surface with double pilosity, dense white adpressed setae mixed with long and robust erect setae. With a few exceptions labroclypeus wide and trapezoidal, with fine dense punctures densely interspersed with coarser ones; anterior margin of labroclypeus straight, or rarely weakly to distinctly sinuate. Mentum flat, both ligular lobes fused medially (Fig. IV.1B). Galea (maxilla) with six teeth (Fig. IV.1C, D). Lacinia reduced in size, nearly round (as in *Omaloplia*) (Fig. IV.1D,E). Mandibles with a large molar lobe which bears numerous rows of tubercles (Fig. IV.1F-H). Antenna composed of 10 antennomeres, club in male with 6 antennomeres (e.g., Fig. IV.2D, H), slightly to strongly reflexed and nearly twice as long as the remaining antennomeres combined, in female club with 4 antennomeres, short, nearly as long as the remaining antennomeres combined. Pronotum with the basal marginal line nearly complete. Elytra with complete epipleural margin, external apical angle convex, apical margin chitinous, without membranous rim of microtrichomes. Anal vein of hind wings abruptly bent (Fig. IV.1M). Ventral surface including metacoxa densely covered with adpressed and moderately long setae. Legs moderately long, shiny, with dense pilosity. Metatibia dorsally with two groups of spines and longitudinally convex; at apex bluntly truncate and concavely sinuate

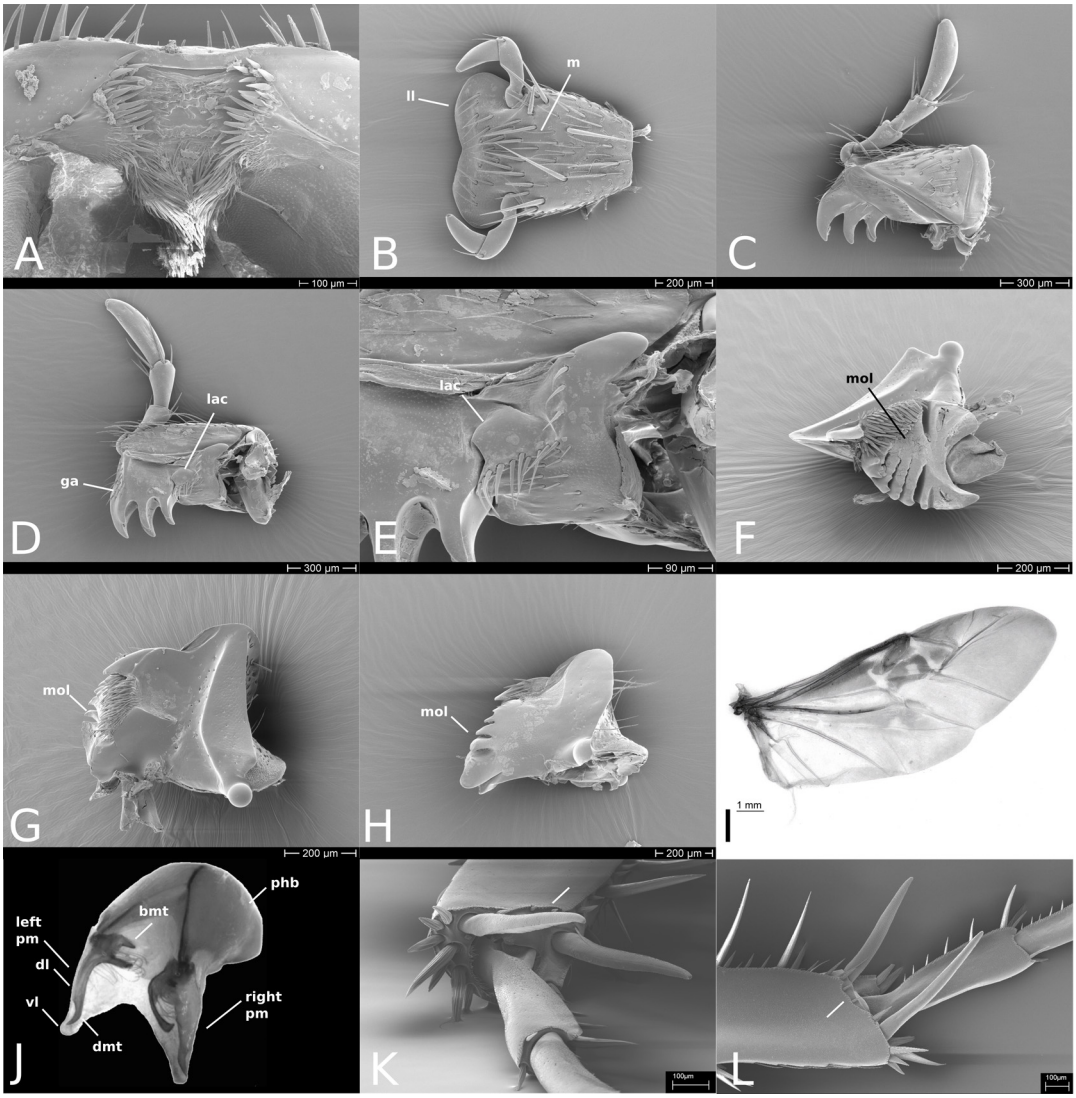

Figure IV.1. Scanning electron microscope images of general morphology of *Pleophylla* (all *P. fasciatipennis* Bl. except J, *P. navicularis* Burm.). A: Labrum; B: Labium; C: Maxilla; D: Maxilla; E: Maxilla (detail of lacinia); F: right mandible; G: Mandible; H: Mandible; I: Hind wing; J: Aedeagus; K: Apex of metatibia; L: Apex of metatibia. A, B, C, G: ventral view; D, E, H, I: dorsal view; F, L: mesal view; J: dorsolateral view; K: distal view. Abbreviations: bmt – basal median tooth, dl – dorsal lobe, dmt – distal median tooth, ga – galea, lac – lacinia, ll – ligular lobes, m – mentum, mol – molar lobe, pm – parameres, phb – phallobase, vl – ventral lobe.

near tarsal articulation (Fig. IV.1L). Protibia bidentate. Pygidium nearly twice as wide as long, flat, with double pilosity at dorsal surface. Parameres composed of a ventral and a dorsal lobe (Fig. IV.1J), the first simple, the latter is composed by two to three distinct branches (sometimes tooth-like) and bears long setae at the medial face. The female copulation organ comprises several sclerotised structures: the basal piece of the ductus bursae, with a number of small lateral protuberances, and the more distal triangular sclerite (Fig. IV.5; Özgül-Siemund and Ahrens 2015). The common duct of the accessory glands is short, and the single glands are small and compact in shape.

Remarks. *Pleophylla* was established by Erichson (1847) without any nominal species included. The first subsequently included species was *Pleophylla fasciatipennis* Blanchard 1850, which is therefore type species of the genus.

3.1.2. Key to *Pleophylla* Erichson, 1847 species (males)

1 Mesosternal process large. Dorsal and ventral lobe of parameres fused (Fig. IV.4Z, Ad). Lateral margin of labroclypeus and ocular canthus produce a distinct angle. Anterior margin of labroclypeus distinctly sinuate medially. **2**

- Mesosternal process small. Dorsal and ventral lobe of parameres separate (Fig. IV.2B). Lateral margin of labroclypeus and ocular canthus only produce an indistinct angle. Anterior margin of labroclypeus straight or shallowly sinuate. **3**

2 (1) Metatarsi dorsally punctate and with fine setae. Parameres almost straight and flat
(Fig. IV.4Z). Hypomeron carinate. . . . . ***P. tongaatsana***

- Metatarsi dorsally smooth, impunctate. Parameres almost curved and not flat
(Fig. IV.4Ad). Hypomeron not carinate. . . . . ***P. kruegeri*** sp. n.

3 (1) Species from South Africa. **4**

- Species from Eastern Africa (Kenya, Burundi, Tanzania, Rwanda). **19**

4 (3) Erect dorsal pilosity sparse (less than 10 setae per interval). **5**

- Erect dorsal pilosity dense (more than 10 setae per interval). **6**

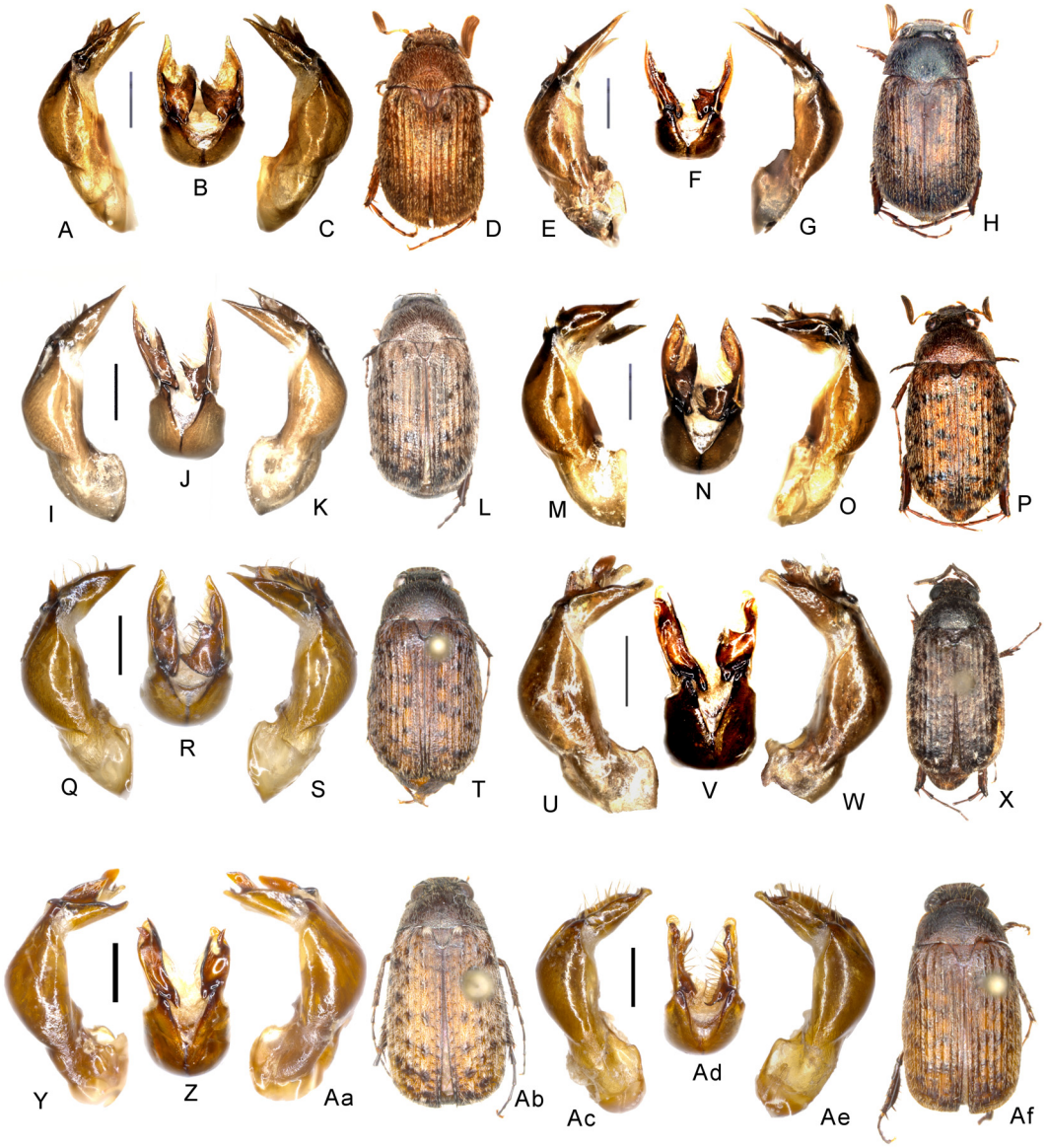

Figure IV.2. A-D: *Pleophylla fasciatipennis* Blanchard (RSA: George); E-H: *P. pilosa* Boheman (RSA: Karkloof for.); I-L: *P. pseudopilosa* sp. n. (holotype); M-P: *P. nelshoogteensis* sp. n. (holotype); Q-T: *P. mlilwaneensis* sp. n. (holotype); U-X: *P. burundiensis* sp. n. (holotype); Y-Ab: *P. congoensis* sp. n. (holotype); Ac-Af: *P. settentrionalis* sp. n. (holotype). A, E, I, M, Q, U, Y, Ac: Aedeagus left side lateral view; C, G, K, O, S, W, Aa, Ae: Aedeagus right side lateral view; B, F, J, N, R, V, Z, Ad: Parameres, dorsal view; D, H, L, P, T, X, Ab, Af: Habitus, dorsal view. Scale: 0.5 mm. Habitus not to scale.

Figure IV.3. A-D: *Pleophylla ferruginea* Burmeister (RSA: Tygerkloof); E-H: *P. murzini* sp. n. (holotype); I-L: *P. navicularis* Burmeister (RSA: Weza); M-P: *P. silvatica* sp. n. (holotype); Q-T: *P. transkeiensis* sp. n. (holotype); U-X: *P. ruthae* sp. n. (holotype); Y-Ab: *P. warnockae* sp. n. (holotype); Ac-Af: *P. harrisoni* sp. n. (holotype). A, E, I, M, Q, U, Y, Ac: Aedeagus left side lateral view; C, G, K, O, S, W, Aa, Ae: Aedeagus right side lateral view; B, F, J, N, R, V, Z, Ad: Parameres, dorsal view; D, H, L, P, T, X, Ab, Af: Habitus, dorsal view. Scale: 0.5 mm. Habitus not to scale.

Figure IV.4. A-D: *Pleophylla mpumalanga* sp. n. (holotype); E-H: *P. charlyi* sp. n. (holotype); I-L: *P. linzleri* sp. n. (holotype); M-P: *P. taitaensis* sp. n. (holotype); Q-T: *P. stalsi* sp. n. (holotype); U-X: *P. wakkerstroomensis* sp. n. (holotype); Y-Ab: *P. tongaatsana* Péringuey (RSA: Sjonajona); Ac-Af: *P. kruegeri* sp. n. (holotype). A, E, I, M, Q, U, Y, Ac: Aedeagus left side lateral view; C, G, K, O, S, W, Aa, Ae: Aedeagus right side lateral view; B, F, J, N, R, V, Z, Ad: Parameres, dorsal view; D, H, L, P, T, X, Ab, Af: Habitus, dorsal view. Scale: 0.5 mm. Habitus not to scale.

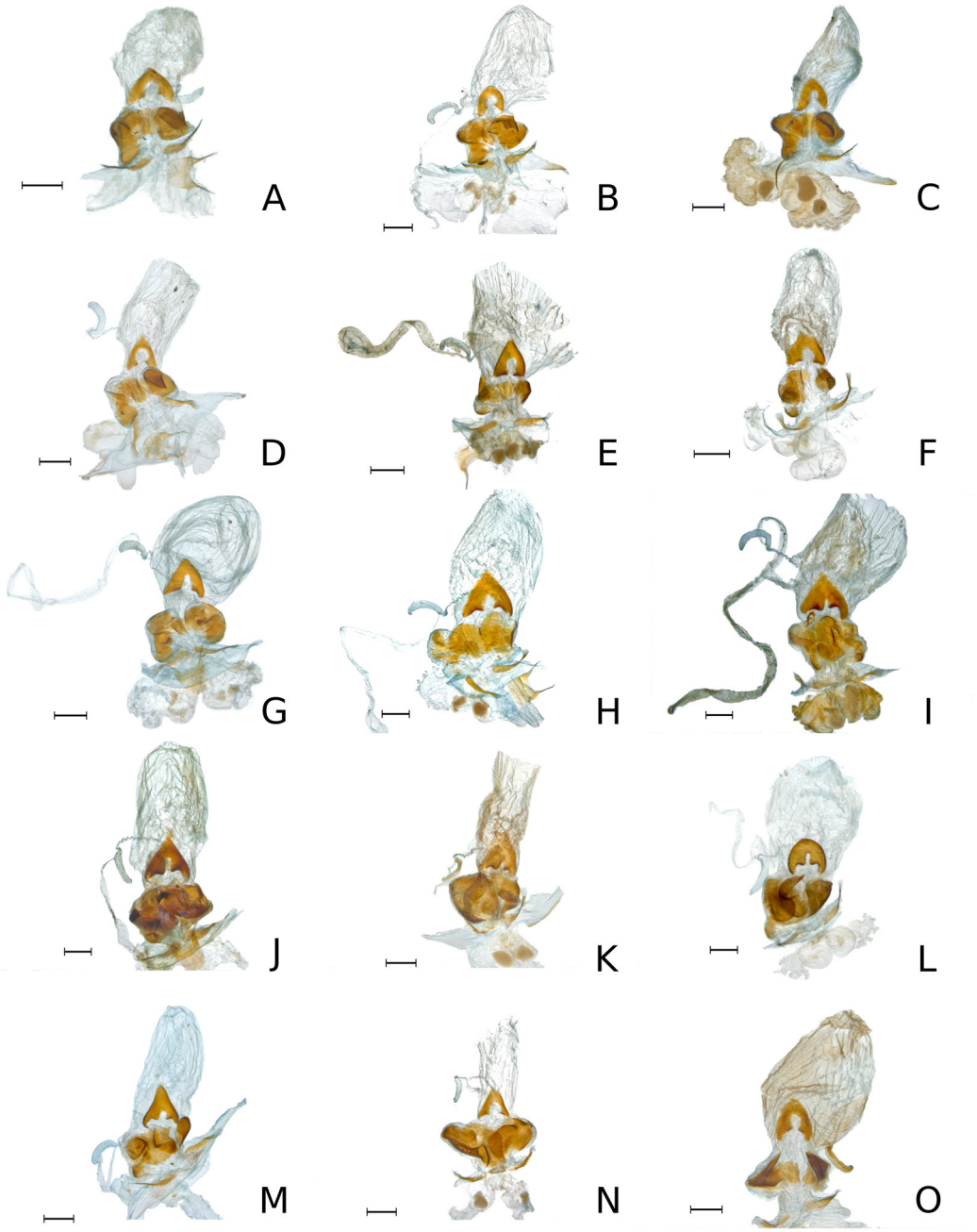

Figure IV.5. Female genitalia of *Pleophylla* species. A: *Pleophylla fasciatipennis*; B: *P. pilosa*; C: *P. pseudopilosa* sp. n.; D: *P. nelshoogteensis* sp. n.; E: *P. burundiensis* sp. n.; F: *P. settentrionalis* sp. n.; G: *P. ferruginea*; H: *P. murzini* sp. n.; I: *P. navicularis*; J: *P. silvatica*; K: *P. transkeiensis* sp. n.; L: *P. ruthae* sp. n.; M: *P. warnockae* sp. n.; N: *P. harrisoni* sp. n.; O: *P. stalsi* sp. n. Scale bars = 0.5 mm

5 (4) Pronotum apparently glabrous, only a few long setae anteriorly, otherwise with moderately dense minute adpressed setae. Elytra without dark spots.
P. wakkerstroomensis sp. n.

- Pronotum with dense and thick adpressed setae. Elytra with more or less distinct dark spots. . . . . ***P. stalsi*** sp. n.

6 (4) Body size >11.0 mm. **7**

- Body size <11.0 mm. **10**

7 (6) Basal margin of pronotum complete. . . . . ***P. murzini*** sp. n.

- Basal margin of pronotum medially narrowly interrupted. **8**

8 (7) Thick, white pilosity of elytra condensed to longitudinal patches with very dense setae. . . . . ***P. mpumalanga*** sp. n.

- Thick, white pilosity of elytra not condensed to patches, evenly dense. **9**

9 (8) The two narrow basal lobes of both parameres subequal in length (Fig. IV.3N).
P. silvatica sp. n.

- The two narrow basal lobes of both parameres very different in length (Fig. IV.3J).
P. navicularis

10 (6) Sutural interval delimitated in apical declivity of elytra by a sharp carina. Larger punctures on pronotum very coarse. **12**

- Sutural interval convex in apical declivity of elytra, not delimitated by a sharp carina. Larger punctures on pronotum moderately coarse. **11**

11 (10) Metatibia slender, ratio length / maximum width: 3.8. Margins of labroclypeus moderately reflexed. . . . . ***P. ruthae*** sp. n.

- Metatibia short, ratio length / maximum width: 3.6. Margins of labroclypeus strongly reflexed. . . . . ***P. warnockae*** sp. n.

12 (10) Smooth area in front of the eye 1.5 times as wide as long. . ***P. fasciatipennis***

- Smooth area in front of the eye at least twice as wide as long. **13**

13 (12) Smooth area in front of eye three times as wide as long. Eyes smaller, ratio diameter/interocular width: 0.7. Species smaller than 10 mm.
P. harrisoni sp. n.

- Smooth area in front of eye nearly twice as wide as long. **14**

14 (13) Body size <10 mm. Metatibia ratio length / maximum width = 4. . . . . **15**

- Body size >10 mm. Metatibia ratio length / maximum width >4.3. . . . . **18**

15 (14) Parameres (ventral lobe) in lateral view wider and shorter. **16**

- Parameres (ventral lobe) in lateral view narrower. Basis of dorsal lobe of left paramere with two medial teeth. . . . . ***P. pilosa***

16 (15) Base of left paramere narrow. Basis of dorsal lobe of left paramere with a single narrow medial tooth. . . . . ***P. pseudopilosa*** sp. n.

- Base of left paramere widened. Basis of dorsal lobe of left paramere with a double medial tooth. **17**

17 (16) Both basomedial teeth in basal third of left paramere (Fig. IV.2N).
P. nelshoogteensis sp. n.

- Distal basomedial tooth in apical third of left paramere (Fig. IV.2R).
P. mlilwaneensis sp. n.

18 (14) Labroclypeus narrow. Parameres almost symmetrical (Fig. IV.3B).
P. ferruginea

- Labroclypeus wide. Parameres distinctly asymmetrical (Fig. IV.3R).
P. transkeiensis sp. n.

19 (3) Anterior margin of labroclypeus straight medially. Body size >11 mm. Centre of intervals impunctate. **22**

- Anterior margin of labroclypeus weakly sinuate medially. Body size <10.5 mm).
20

20 (19) Centre of intervals in major extent coarsely punctate. **21**

- Centre of intervals impunctate. . . . . ***P. settentrionalis*** sp. n.

21 (20) Basal lobe of left paramere ends before basal half of paramere (Fig. IV.2V).
P. burundiensis sp. n.

- Basal lobe of left paramere produced beyond distal half of paramere (Fig. IV.2Z).
P. congoensis sp. n.

22 (19) Ventral lobes of parameres not extended dorsally, symmetrical (Fig. IV.4N). ***P. taitaensis*** sp. n.

- Ventral lobes of parameres extended dorsally as asymmetrical lobes (Fig. IV.4F, J). **23**

23 (22) Metatibia ratio length / maximum width: ≈ 3.8. Eyes smaller, ratio diameter/-interocular width: 0.7. . . . . ***P. lizleri*** sp. n.

- Metatibia ratio length / maximum width: ≈ 4.3. Eyes larger, ratio diameter/interocular width: 0.78. . . . . ***P. charlyi*** sp. n.

3.1.3. *Pleophylla fasciatipennis* Blanchard, 1850

(Figs IV.1A-L, IV.2A-D, IV.5A, IV.6)

Omaloplia fasciatipennis Blanchard 1850: 83 [type locality: Cape of Good Hope]. *Pleophylla fasciatipennis* Schaufuss 1871: 231; Brenske 1899: 84; Péringuey 1904: 8. *Pleophylla maculipennis* Boheman 1857: 124 [type locality: Gariep]; syn. by Péringuey 1904: 8; Schaufuss 1871: 232; Brenske 1899: 83.

Type material examined: Lectotype (here designated): ♂ "*P. fasciatipennis* Cat. Mus. Cap de B. Esp./ Museum Paris Afrique Hottentot Delalande/ *Pleophylla fasciatipennis* Type Blanch. [unknown hand writing]/ Afriq. Hottentot Delalande/ IMG0032" (MNHN). Paralectotypes: 1 ♀ "Museum Paris Cap de Bonne Esperance/ *Pleophylla fasciatipennis* Type Blanch." (MNHN), 1 ♀ "Museum Paris Afrique Hottentot Delalande/ *Pleophylla fasciatipennis* Type Blanch." (MNHN). Lectotype (*maculipennis*, here designated): ♂ "Caffraria./ J. Wahlb./ Type./ Typus [red printed label]/3674E91/ Naturhistoriska Riksmuseet Stockholm Loan no. 1743/07" (NHRS). Paralectotypes (*pilosa*): 1 ♂ "Cap. B. Spei./ Victorin./ 3676 E91+/ Naturhistoriska Riksmuseet Stockholm Loan no. 1745/07 "(NHRS), 1 ♂ "Cap. B. Spei./ Victorin./ 3675 E91+/ Naturhistoriska Riksmuseet Stockholm Loan no. 1744/07 "(NHRS).
Additional material examined: see electronic supplementary material C1.

Redescription. Body length: 7.9 mm, elytral length: 5.9 mm, maximum width: 4.2 mm.
Angle between lateral margins of labroclypeus and ocular canthus indistinct; anterior margin of labroclypeus medially straight; margins of labroclypeus moderately

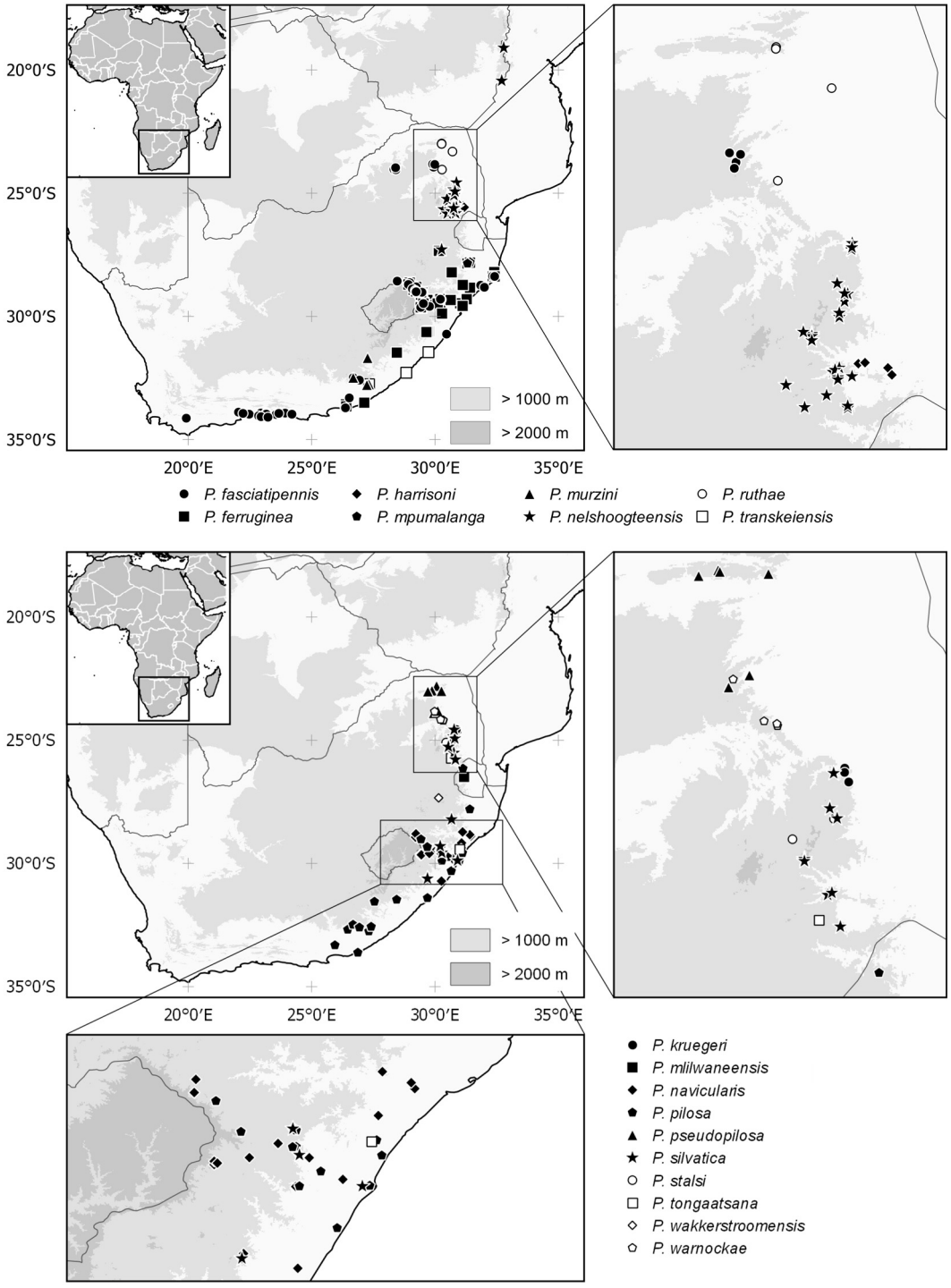

Figure IV.6. Distribution of *Pleophylla* species in Africa.

reflexed; smooth area in front of eyes 1.5 times as wide as long; ratio of eye diameter / interocular width: 0.5. Pronotum with dense and thick erect setae; unicoloured; basal margin of pronotum complete; larger punctures on pronotum very coarse. Hypomeron not carinate. Elytra with dark spots; erect dorsal pilosity on elytra dense (more than 10 setae per interval); sutural interval in apical declivity of elytra delimited by a sharp carina; white, fine pilosity not condensed to longitudinal patches; intervals punctate at least laterally; adpressed white pilosity long, 1/3 to 1/2 of width of elytral intervals; external apical elytral angle evenly convex. Mesosternal process short. Ratio of length of metepisternum/metacoxa: 1/1.21. Metatibia, ratio of width/length: 1/3.59. Metatarsi dorsally smooth; first metatarsomere as long as subsequent one. Aedeagus: Fig. IV.2A–C. Habitus: Fig. IV.5D. Female genitalia: Fig. IV.5A.

Remarks. In previous publications we referred to this species as "sp2" (for *P. pilosa*) and "sp12" (for *P. maculipennis*) (Özgül-Siemund and Ahrens 2015; Eberle et al. 2016b).

3.1.4. *Pleophylla pilosa* Boheman, 1857

(Figs IV.2E-H, IV.5B, IV.6)

Pleophylla pilosa Boheman 1857: 125 [Limpopo river]; Schaufuss 1871: 232, Brenske 1899: 84; Péringuey 1904: 8.
Pleophylla opalina Schaufuss 1871: 232 [Caffraria interia]; Brenske 1899: 84; Péringuey 1904: 8; **syn. nov.**

Type material examined: Lectotype (here designated): ♂ "Caffraria./ J. Wahlb./ Type./ *Pleophylla pilosa*/ *pilosa* Bhn. Ent. Kaffr./ 3680E91+/ Naturhistoriska Riksmuseet Stockholm Loan no. 1746/07" (NHRS). Lectotype (opalina, here designated): ♂ "Coll. L.W. Schaufuss/ Nunqn at pag. 232/ opalina Schauf. Caffr. inter./ Type./ *Pleophylla opalina* m. ~~navicularis Burm~~. [sic! canceled] Natal" (ZMHB).

Additional material examined: see electronic supplementary material C1.

Redescription. Body length: 9.0 mm, elytral length: 6.1 mm, maximum width: 4.7 mm.
Angle between lateral margins of labroclypeus and ocular canthus indistinct; anterior margin of labroclypeus medially straight; margins of labroclypeus moderately

reflexed; smooth area in front of eyes twice as wide as long; ratio of eye diameter/ interocular width: 0.68. Pronotum with dense and thick erect setae; unicoloured; basal margin of pronotum complete; larger punctures on pronotum moderately large. Hypomeron not carinate. Elytra with dark spots; erect dorsal pilosity on elytra dense (more than 10 setae per interval); sutural interval in apical declivity of elytra delimited by a sharp carina; white, fine pilosity not condensed to longitudinal patches; intervals punctate at least laterally; adpressed white pilosity long, 1/3 to 1/2 of width of elytral intervals; external apical elytral angle evenly convex. Mesosternal process short. Ratio of length of metepisternum/metacoxa: 1/1.33. Metatibia, ratio of width/length: 1/3.87. Metatarsi dorsally smooth; first metatarsomere as long as subsequent one. Aedeagus: Fig. IV.2E–G. Habitus: Fig. IV.2H.

Remarks. The shape of the parameres of *P. opalina* are virtually identical in shape with those of the lectotype of *P. pilosa*. In previous publications we referred to this species as "sp6" (Özgül-Siemund and Ahrens 2015; Eberle et al. 2016b).

3.1.5. *Pleophylla pseudopilosa* Ahrens, Beckett, Eberle & Fabrizi sp. nov.

(Figs IV.2I-L, IV.5C, IV.6)

Type material examined: Holotype: ♂ "834699 X-DA0394 South Africa Cheerio Farm, Haenertsburg, ca 20km W of Tzaneen, 1492 m 023°53'42,7"S 029°57'09,7"E 20–22.XI.2006 leg. D. Ahrens & S. Fabrizi, *Pleophylla* sp1" (TMSA). Paratypes: see electronic supplementary material C1.

Description. Body length: 8.3 mm, elytral length: 6.0 mm, maximum width: 4.6 mm.
Angle between lateral margins of labroclypeus and ocular canthus indistinct; anterior margin of labroclypeus medially straight; margins of labroclypeus moderately reflexed; smooth area in front of eyes twice as wide as long; ratio of eye diameter / interocular width: 0.65. Pronotum with dense and thick erect setae; unicoloured; basal margin of pronotum complete; larger punctures on pronotum very coarse. Hypomeron not carinate. Elytra with dark spots; erect dorsal pilosity on elytra dense (more than 10 setae per interval); sutural interval in apical declivity of elytra delimited by a sharp carina; white, fine pilosity not condensed to longitudinal patches; intervals punctate at least laterally; adpressed white pilosity long, 1/3 to 1/2 of width of elytral inter-

vals; external apical elytral angle evenly convex. Mesosternal process short. Ratio of length of metepisternum/metacoxa: 1/1.34. Metatibia, ratio of width/length: 1/4. Metatarsi dorsally smooth; first metatarsomere as long as subsequent one.
Aedeagus: Fig. IV.2I-K. Habitus: Fig. IV.2L. Female genitalia: Fig. IV.5C.

Diagnosis. *Pleophylla pseudopilosa* sp. nov. differs from *P. pilosa* in the parameres by having the ventral lobe in lateral view wider and shorter; the base of the dorsal lobe of left paramere has a single narrow medial tooth (not two as in *P. pilosa*).

Variation. Body length: 6.8–8.5 mm, elytral length: 4.2–6.1 mm, maximum width: 3.8-4.8 mm. Colour varies from reddish brown with dark spot to entirely yellowish brown without spots. Female: Antennal club of female straight, composed of 5 antennomeres, slightly longer than the remaining antennomeres combined; 6^{th} antennomere subequal a quarter of club length. Eyes smaller than in male: Ratio of eye diameter/interocular width: 0.5.

Etymology. The name of the new species is composed of the Latinised Greek prefix pseudo- (false) and pilosa (species name of I, with reference to its general similarity with this species) (noun in apposition).

Remarks. In previous publications we referred to this species as "sp1" (Özgül-Siemund and Ahrens 2015; Eberle et al. 2016b).

3.1.6. *Pleophylla nelshoogteensis* Ahrens, Beckett, Eberle & Fabrizi sp. nov.

(Figs IV.2M–P, IV.5D, IV.6)

Type material examined: Holotype: ♂ "S. Afr; Tv. Nelshoogte Forest Station 25.50 S – 30.50 E/ 2.12.1986 E-Y: 2346 UV-light collection leg. Endrödy-Younga" (TMSA). Paratypes: see electronic supplementary material C1.

Description. Body length: 9.6 mm, elytral length: 6.3 mm, maximum width: 4.6 mm.
Angle between lateral margins of labroclypeus and ocular canthus indistinct; anterior margin of labroclypeus medially straight; margins of labroclypeus moderately reflexed; smooth area in front of eyes twice as wide as long; ratio of eye diameter / interocular width: 0.77. Pronotum with dense and thick erect setae; unicoloured; basal

margin of pronotum narrowly interrupted medially (distinctly less than scutellum width); larger punctures on pronotum very coarse. Hypomeron not carinate. Elytra with dark spots; erect dorsal pilosity on elytra dense (more than 10 setae per interval); sutural interval in apical declivity of elytra delimited by a sharp carina; white, fine pilosity not condensed to longitudinal patches; intervals punctate at least laterally; adpressed white pilosity long, 1/3 to 1/2 of width of elytral intervals; external apical elytral angle evenly convex. Mesosternal process short. Ratio of length of metepisternum/metacoxa: 1/1.5. Metatibia, ratio of width/length: 1/4.14. Metatarsi dorsally smooth; first metatarsomere as long as subsequent one.
Aedeagus: Fig. IV.2M–O. Habitus: Fig. IV.2P. Female genitalia: Fig. IV.5D.

Diagnosis. *Pleophylla nelshoogteensis* sp. nov. differs from *P. pilosa* in the parameres by having the ventral lobe in lateral view wider and shorter, and from P. pseudopilosa by the base of left paramere being widened, the base of dorsal lobe of left paramere with a double medial tooth.

Variation. Body length: 7.1–9.6 mm, elytral length: 5.6–6.5 mm, maximum width: 3.9–4.8 mm. Density of dark spots may vary slightly, but never completely without spots. Female: Antennal club of female straight, composed of 5 antennomeres, slightly longer than the remaining antennomeres combined; 6^{th} antennomere subequal one to three quarters of club length. Eyes smaller than in male: Ratio of eye diameter/interocular width: 0.5.

Etymology. The new species is named after its type locality, Nelshoogte Forest Station.

Remarks. In a previous publication we referred to this species as "sp1A" (Özgül-Siemund and Ahrens 2015).

3.1.7. *Pleophylla mlilwaneensis* Ahrens, Beckett, Eberle & Fabrizi sp. nov.

(Figs IV.2Q–T, IV.6)

Type material examined. Holotype ♂ "Swaziland, Mlilwane Wildlife Sanctuary 26°29,22'S, 31°11'E 800mNN, 17.–18.XI.1996, leg. M. Hartmann" (NME).

Description. Body length: 7.7 mm, elytral length: 6.8 mm, maximum width: 4.0 mm.
Angle between lateral margins of labroclypeus and ocular canthus indistinct; anterior margin of labroclypeus medially straight; margins of labroclypeus strongly reflexed; smooth area in front of eyes twice as wide as long; ratio of eye diameter / interocular width: 0.74. Pronotum with dense and thick erect setae; unicoloured; basal margin of pronotum complete; larger punctures on pronotum moderately large. Hypomeron not carinate. Elytra with dark spots; erect dorsal pilosity on elytra dense (more than 10 setae per interval); sutural interval in apical declivity of elytra flat and not delimited by a sharp carina; white, fine pilosity not condensed to longitudinal patches; intervals punctate at least laterally; adpressed white pilosity long, 1/3 to 1/2 of width of elytral intervals; external apical elytral angle evenly convex. Mesosternal process short. Ratio of length of metepisternum/metacoxa: 1/1.44. Metatibia, ratio of width/length: 1/3.71. Metatarsi dorsally smooth; first metatarsomere slightly shorter than subsequent one.
Aedeagus: Fig. IV.2Q–S. Habitus: Fig. IV.2T.

Diagnosis. *Pleophylla mlilwaneensis* sp. nov. is in the shape of parameres rather similar to *P. nelshoogteensis* sp. nov. It differs from the latter in having the distal basomedial tooth in the apical third of the left paramere (rather than having both in basal half).

Etymology. The new species is named after its type locality, the Mlilwane Wildlife Sanctuary in Swaziland.

3.1.8. *Pleophylla burundiensis* Ahrens, Beckett, Eberle & Fabrizi sp. nov.

(Figs IV.2U–X, IV.5E, IV.7)

Type material examined. Holotype: ♂ "Burundi Ngozi" (ZFMK). Paratypes: see electronic supplementary material C1.

Description. Body length: 10.3 mm, elytral length: 7.1 mm, maximum width: 4.6 mm.
Angle between lateral margins of labroclypeus and ocular canthus indistinct; anterior margin of labroclypeus medially weakly sinuate; margins of labroclypeus strongly

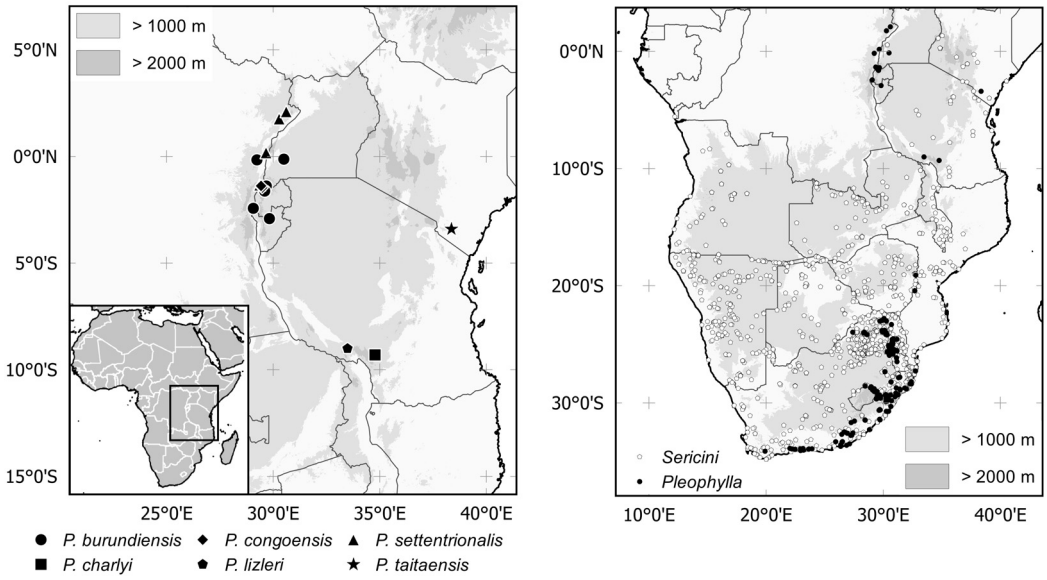

Figure IV.7. Distribution of *Pleophylla* species (left side) and absence records (right side), represented by sampling localities of all other sampled Sericini species without records of *Pleophylla* (white dots).

reflexed; smooth area in front of eyes twice as wide as long; ratio of eye diameter / interocular width: 0.7. Pronotum with dense and thick erect setae; unicoloured; basal margin of pronotum complete; larger punctures on pronotum very coarse. Hypomeron not carinate. Elytra with dark spots; erect dorsal pilosity on elytra dense (more than 10 setae per interval); sutural interval in apical declivity of elytra flat and not delimited by a sharp carina; white, fine pilosity condensed to transverse patches; intervals punctate at least laterally; adpressed white pilosity long, 1/3 to 1/2 of width of elytral intervals; external apical elytral angle evenly convex. Mesosternal process short. Ratio of length of metepisternum/metacoxa: 1/1.33. Metatibia, ratio of width/length: 1/4. Metatarsi dorsally smooth; first metatarsomere as long as subsequent one.
Aedeagus: Fig. IV.2U–W. Habitus: Fig. IV.2X. Female genitalia: Fig. IV.5E.

Diagnosis. *Pleophylla burundiensis* sp. nov. resembles in external shape *P. pilosa*, however, the punctures on dorsal surface are denser and coarser, and the sutural interval is flat and not limited by a sharp carina in apical declivity of elytra.

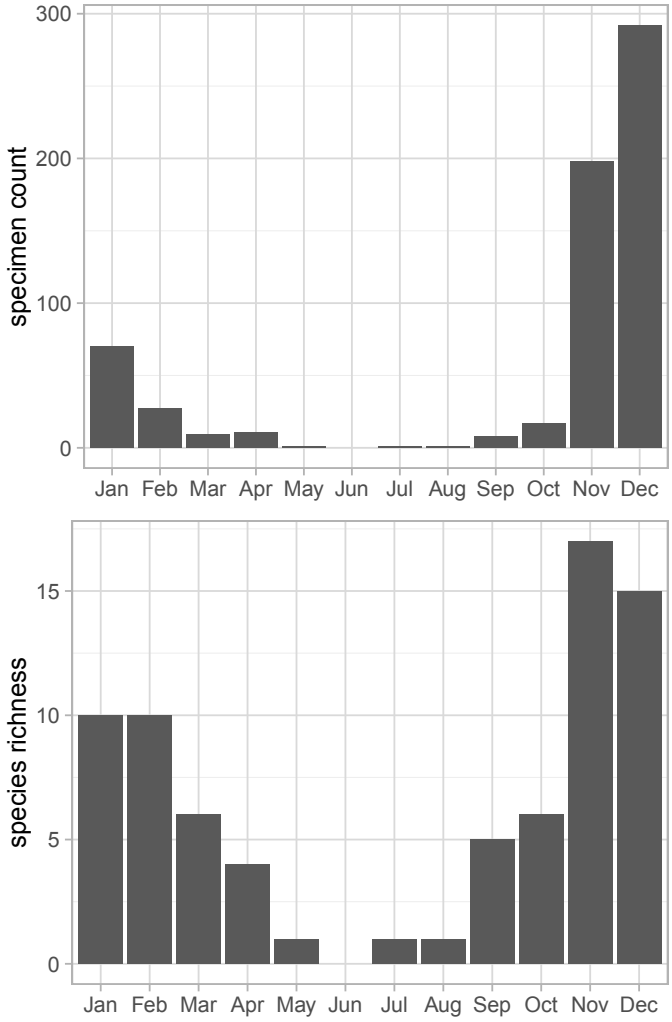

Figure IV.8. Phenology of *Pleophylla* in terms of species richness (below) and numbers of recorded adult specimens (above).

This new species is in genital shape similar to *P. maculipennis*, however, the right paramere is basally tooth-like and extended, and the basal dorsal portion of the left paramere bears two small teeth.

Variation. Body length: 8.0-10.3 mm, elytral length: 6.5-7.1 mm, maximum width: 4.2-4.8 mm. Density of dark spots may vary slightly, but never completely without spots. Female: Antennal club of female straight, composed of 5 antennomeres, as long as the remaining antennomeres combined; 6th antennomere very short, less than 1/5 of club length. Eyes smaller than in male: Ratio of eye diameter/interocular width: 0.52.

Etymology. The new species is named after is occurrence in Burundi.

Remarks. In a previous publication we referred to this species as "SpM-21" (Özgül-Siemund and Ahrens 2015).

3.1.9. *Pleophylla congoensis* Ahrens, Beckett, Eberle & Fabrizi sp. nov.

(Figs IV.2Y-Ab, IV.7)

Type material examined. Holotype ♂ "Coll. Mus. Congo Bukima IV-1948 J.V. Leroy" (RMCA). Paratypes: 1 ♂ "1083, S. Bishoke (2400), (Kibga), 8/19-II-1935, G. F. de Witte, PARC NAT. ALBERT" (RMCA), 1 ♀ "Rwanda, Bisoke, 1,46S/ 29,48E, 2600-3300m 19.IX.02, Th. Wagner leg." (ZFMK), 1 ♀ "Coll. I.R.Sc.N.B. Rwanda P.N. Nyungwe IX-2007 Leg. E. Vingerhoedt" (ISNB).

Description. Body length: 9.1 mm, elytral length: 7.0 mm, maximum width: 4.7 mm.
Angle between lateral margins of labroclypeus and ocular canthus indistinct; anterior margin of labroclypeus medially weakly sinuate; margins of labroclypeus moderately reflexed; smooth area in front of eyes twice as wide as long; ratio of eye diameter / interocular width: 0.73. Pronotum with dense and thick erect setae; unicoloured; basal margin of pronotum complete; larger punctures on pronotum very coarse. Hypomeron not carinate. Elytra with dark spots; erect dorsal pilosity on elytra dense (more than 10 setae per interval); sutural interval in apical declivity of elytra distinctly elevated but convex; white, fine pilosity not condensed to longitudinal patches; intervals punctate at least laterally; adpressed white pilosity long, 1/3 to 1/2 of width of elytral

intervals; external apical elytral angle evenly convex. Mesosternal process short. Ratio of length of metepisternum/metacoxa: 1/1.27. Metatibia, ratio of width/length: 1/4.31. Metatarsi dorsally smooth; first metatarsomere as long as subsequent one. Aedeagus: Fig. IV.2Y-Aa. Habitus: Fig. IV.2Ab.

Diagnosis. *Pleophylla congoensis* sp. nov. is similar to *P. burundiensis* sp. nov., but it can be distinguished by the basal lobe of the left paramere which is produced beyond the distal half of the paramere.

Variation. Body length: 8.8-10.6 mm, elytral length: 6.9-8.1 mm, maximum width: 4.7-5.6 mm. Density of dark spots may vary slightly, but never completely without spots. Female: Antennal club of female straight, composed of 5 antennomeres, slightly longer than the remaining antennomeres combined; 6^{th} antennomere short, slightly shorter than half of club length. Eyes smaller than in male: Ratio of eye diameter/interocular width: 0.52.

Etymology. The new species is named "*congoensis*" following its occurrence in the eastern Democratic Republic of Congo (Virunga Mountains).

3.1.10. *Pleophylla settentrionalis* Ahrens, Beckett, Eberle & Fabrizi sp. nov.

(Figs IV.2Ac-Af, IV.5F, IV.7)

Type material examined. Holotype ♂ "Coll. Mus. Congo N. Lac Kivu Rwankwi XI-1951 J.V. Leroy" (RMCA). Paratypes: see electronic supplementary material C1.

Description. Body length: 7.7 mm, elytral length: 5.6 mm, maximum width: 4.0 mm.
Angle between lateral margins of labroclypeus and ocular canthus indistinct; anterior margin of labroclypeus medially weakly sinuate; margins of labroclypeus strongly reflexed; smooth area in front of eyes twice as wide as long; ratio of eye diameter / interocular width: 0.77. Pronotum with dense and thick erect setae; unicoloured; basal margin of pronotum complete; larger punctures on pronotum very coarse. Hypomeron not carinate. Elytra with dark spots; erect dorsal pilosity on elytra dense (more than 10 setae per interval); sutural interval in apical declivity of elytra flat and not delimited by a sharp carina; white, fine pilosity not condensed to longitudi-

nal patches; intervals punctate at least laterally; adpressed white pilosity long, 1/3 to 1/2 of width of elytral intervals; external apical elytral angle evenly convex. Mesosternal process short. Ratio of length of metepisternum/metacoxa: 1/1.35. Metatibia, ratio of width/length: 1/4. Metatarsi dorsally smooth; first metatarsomere as long as subsequent one.
Aedeagus: Fig. IV.2Ac-Ae. Habitus: Fig. IV.2Af. Female genitalia: Fig. IV.5F.

Diagnosis. *Pleophylla settentrionalis* sp. nov. differs from the previous two species, *P. congoensis* sp. nov. and *P. burundiensis* sp. nov. by having the centre of the elytral intervals impunctate rather than densely punctate.

Variation. Body length: 7.6-9.0 mm, elytral length: 5.6-7.0 mm, maximum width: 4.0-4.8 mm. Density of dark spots may vary, sometimes completely without spots. Female: Antennal club of female straight, composed of 5 antennomeres, slightly longer than the remaining antennomeres combined; 6^{th} antennomere short, slightly shorter than a quarter of club length. Eyes smaller than in male: Ratio of eye diameter/interocular width: 0.53.

Etymology. This new species is named "*settentrionalis*" with reference to its occurrence in the northernmost part of the known range of *Pleophylla* species.

Remarks. In a previous publication we referred to this species as "sp27" (Özgül-Siemund and Ahrens 2015).

3.1.11. *Pleophylla ferruginea* Burmeister, 1855

(Figs IV.3A–D, IV.5G, IV.6)

Pleophylla ferruginea Burmeister 1855: 181 [type locality: Weihnachtsbai [i.e. Durban](Ecklon)]; Schaufuss 1871: 231; Brenske 1899:84; Péringuey 1904: 8.

Type material examined. Lectotype (here designated): ♂ "ferruginea Eckl. Pt. nat. Eck./ Martin-Luther Universitat Halle-Wittenberg Zentralmagazin Naturw. Sammlungen 08/03/12 03.Jul. 2014" (MLUH).

Additional material examined. see electronic supplementary material C1.

Description. Body length: 10.3 mm, elytral length: 7.0 mm, maximum width: 5.0 mm.

Angle between lateral margins of labroclypeus and ocular canthus indistinct; anterior margin of labroclypeus medially straight; margins of labroclypeus moderately reflexed; smooth area in front of eyes twice as wide as long; ratio of eye diameter / interocular width: 0.76. Pronotum with dense and thick erect setae; unicoloured. Hypomeron not carinate. Basal margin of pronotum narrowly interrupted medially (distinctly less than scutellum width); larger punctures on pronotum very coarse. Elytra with dark spots; erect dorsal pilosity on elytra dense (more than 10 setae per interval); sutural interval in apical declivity of elytra delimited by a sharp carina; white, fine pilosity not condensed to longitudinal patches; intervals punctate at least laterally; adpressed white pilosity long, 1/3 to 1/2 of width of elytral intervals; external apical elytral angle evenly convex. Mesosternal process short. Ratio of length of metepisternum/metacoxa: 1/1.43. Metatibia, ratio of width/length: 1/4.4. Metatarsi dorsally smooth; first metatarsomere as long as subsequent one.
Aedeagus: Fig. IV.3A–C. Habitus: Fig. IV.3D. Female genitalia: Fig. IV.5G.

Remarks. *Pleophylla transvaalica* is an invalid 'in litteris' name of a specimen without a locality label that is housed in ZMHB (ex coll. Brenske) but originally came from Péringuey (SAMC). In previous publications we referred to this species as "sp10" (Özgül-Siemund and Ahrens 2015; Eberle et al. 2016b).

3.1.12. *Pleophylla murzini* Ahrens, Beckett, Eberle & Fabrizi sp. nov.

(Figs IV.3E–H, IV.5H, IV.6)

Type material examined. Holotype: ♂ "S. Africa Eastern Cape pr Katberg pass, h=1250m; 09.1.2002 S. Murzin leg" (ZFMK). Paratypes: see electronic supplementary material C1.

Description. Body length: 11 mm, elytral length: 8.14 mm, maximum width: 6.14 mm.
Angle between lateral margins of labroclypeus and ocular canthus indistinct; anterior margin of labroclypeus medially straight; margins of labroclypeus moderately reflexed; smooth area in front of eyes twice as wide as long; ratio of eye diameter / interocular width: 0.63. Pronotum with dense and thick erect setae; unicoloured; basal margin of pronotum complete; larger punctures on pronotum very coarse. Hypomeron not carinate. Elytra with dark spots; erect dorsal pilosity on elytra dense (more than

10 setae per interval); sutural interval in apical declivity of elytra distinctly elevated but convex; white, fine pilosity not condensed to longitudinal patches; intervals punctate at least laterally; adpressed white pilosity long, 1/3 to 1/2 of width of elytral intervals; external apical elytral angle evenly convex. Mesosternal process short. Ratio of length of metepisternum/metacoxa: 1/1.44. Metatibia, ratio of width/length: 1/4.11. Metatarsi dorsally smooth; first metatarsomere as long as subsequent one. Aedeagus: Fig. IV.3E–F. Habitus: Fig. IV.3H. Female genitalia: Fig. IV.5H.

Diagnosis. The new species is in shape of male genitalia rather similar to *P. ferruginea*. *Pleophylla murzini* differs from the latter by the slightly larger body size and the shape of parameres: the basal basomedian branches (tooth) of the left and right paramere are much larger than the distal basomedian branches (while in *P. ferruginea* they are subequal in size). From the other large species (following below) it differs by the complete basal marginal line on the pronotum.

Variation. Body length: 8.9-11 mm, elytral length: 6.8-8.5 mm, maximum width: 4.8-6.5 mm. Density of dark spots may vary, sometimes completely without spots. Female: Antennal club of female straight, composed of 5 antennomeres, slightly longer than the remaining antennomeres combined; 6^{th} antennomere short, half as long as club length. Eyes smaller than in male: Ratio of eye diameter/interocular width: 0.53.

Etymology. The new species is named after its collector, Sergej Murzin (Moscow).

Remarks. In a previous publication we referred to this species as "SpM-16" (Özgül-Siemund and Ahrens 2015).

3.1.13. *Pleophylla navicularis* Burmeister, 1855

(Figs IV.3I–L, IV.5I, IV.6)

Pleophylla navicularis Burmeister 1855: 181 [type locality: Weihnachtsbai (Höppig)]; Schaufuss 1871: 231; Brenske 1899: 83; Péringuey 1904: 7.
Pleophylla flavicornis Schaufuss 1871: 232 [type locality: Caffr., New Germany]; Brenske 1899: 83, syn. nov.

Type material examined. Lectotype (*navicularis*, here designated): ♂ "navicularis Nab. subcylindr. Pt. Nat. Gm." (MLUH). Paralectoptype: 1 ♀ [without labels,

provided with a copy of the original collection label] (MLUH). Lectotype (*flavicornis*, here designated): ♀ "Type./ flavicornis m. N. Germ./ Nunqu at. pag. 232/ Coll. L.W. Schaufuss/ flavicornis Schauf. Caffraria/ Pleophylla" (ZMHB).

Additional material examined: see electronic supplementary material C1.

Redescription. Body length: 13.0 mm, elytral length: 9.4 mm, maximum width: 6.6 mm.
Angle between lateral margins of labroclypeus and ocular canthus indistinct; anterior margin of labroclypeus medially straight; margins of labroclypeus moderately reflexed; smooth area in front of eyes three times as wide as long; ratio of eye diameter / interocular width: 0.68. Pronotum with dense and thick erect setae; unicoloured; basal margin of pronotum narrowly interrupted medially (distinctly less than scutellum width); larger punctures on pronotum very coarse. Hypomeron not carinate. Elytra with dark spots; erect dorsal pilosity on elytra dense (more than 10 setae per interval); sutural interval in apical declivity of elytra distinctly elevated but convex; white, fine pilosity not condensed to longitudinal patches; intervals punctate at least laterally; adpressed white pilosity long, 1/3 to 1/2 of width of elytral intervals; external apical elytral angle evenly convex. Mesosternal process short. Ratio of length of metepisternum/metacoxa: 1/1.47. Metatibia, ratio of width/length: 1/4.3. Metatarsi dorsally smooth; first metatarsomere as long as subsequent one.
Aedeagus: Fig. IV.3I–K. Habitus: Fig. IV.3L. Female genitalia: Fig. IV.5I.

Remarks. In previous publications we referred to this species as "sp11" (Özgül-Siemund and Ahrens 2015; Eberle et al. 2016b).

3.1.14. *Pleophylla silvatica* Ahrens, Beckett, Eberle & Fabrizi sp. nov.

(Figs IV.3M–P, IV.5J, IV.6)

Type material examined. Holotype: ♂ "S. Afr., E. Transvaal Berlin Karst plat. 25.31 S - 30.46 E/ 08.12.1986 E-Y: 2363 fungous Pinus logs leg. Endrödy-Younga" (TMSA). Paratypes: see electronic supplementary material C1.

Description. Body length: 13.4 mm, elytral length: 9.3 mm, maximum width: 7.1 mm.
Angle between lateral margins of labroclypeus and ocular canthus indistinct; anterior

margin of labroclypeus medially straight; margins of labroclypeus moderately reflexed; smooth area in front of eyes three times as wide as long; ratio of eye diameter / interocular width: 0.67. Pronotum with dense and thick erect setae; unicoloured; basal margin of pronotum narrowly interrupted medially (distinctly less than scutellum width); larger punctures on pronotum very coarse. Hypomeron not carinate. Elytra with dark spots; erect dorsal pilosity on elytra dense (more than 10 setae per interval); sutural interval in apical declivity of elytra distinctly elevated but convex; white, fine pilosity not condensed to longitudinal patches; intervals punctate at least laterally; adpressed white pilosity long, 1/3 to 1/2 of width of elytral intervals; external apical elytral angle evenly convex. Mesosternal process short. Ratio of length of metepisternum/metacoxa: 1/1.54. Metatibia, ratio of width/length: 1/4.14. Metatarsi dorsally smooth; first metatarsomere as long as subsequent one.
Aedeagus: Fig. IV.3M–O. Habitus: Fig. IV.3P. Female genitalia: Fig. IV.5J.

Diagnosis. *Pleophylla silvatica* sp. nov. is in external and genital shape similar to *P. navicularis*. It differs from the latter in having the two narrow basal branches of both parameres (dorsal lobe) subequal in length, while in *P. navicularis* the basal branches of both parameres (dorsal lobe) are different in length.

Variation. Body length: 8.6-13.4 mm, elytral length: 6.4-9.4 mm, maximum width: 4.6-7.1 mm. Density of dark spots may vary, but never completely without spots. Female: Antennal club of female straight, composed of 5 antennomeres, slightly longer than the remaining antennomeres combined; 6^{th} antennomere short, half as long as club length. Eyes smaller than in male: Ratio of eye diameter/interocular width: 0.5.

Etymology. The new species is named "*silvatica*" following its strict occurrence in native forests.

Remarks. The northern and southern populations are slightly differentiated in shape of parameres. In a previous publication we referred to this species as "SpeciesM-1" (Özgül-Siemund and Ahrens 2015).

3.1.15. *Pleophylla transkeiensis* Ahrens, Beckett, Eberle & Fabrizi sp. nov.

(Figs IV.3Q–T, IV.5K, IV.6)

Type material examined. Holotype: ♂ "S. Afr., Transkei Ntsubane forest 31.27 S - 29.44 E/ 25.11.1987 E-Y: 2537 Fungi & for. litter leg Endrödy-Younga" (TMSA). Paratypes: see electronic supplementary material C1.

Description. Body length: 10.6 mm, elytral length: 7.4 mm, maximum width: 5.7 mm.
Angle between lateral margins of labroclypeus and ocular canthus indistinct; anterior margin of labroclypeus medially straight; margins of labroclypeus moderately reflexed; smooth area in front of eyes twice as wide as long; ratio of eye diameter / interocular width: 0.6. Pronotum with dense and thick erect setae; unicoloured; basal margin of pronotum narrowly interrupted medially (distinctly less than scutellum width); larger punctures on pronotum very coarse. Hypomeron not carinate. Elytra with dark spots; erect dorsal pilosity on elytra dense (more than 10 setae per interval); sutural interval in apical declivity of elytra delimited by a sharp carina; white, fine pilosity not condensed to longitudinal patches; intervals punctate at least laterally; adpressed white pilosity long, 1/3 to 1/2 of width of elytral intervals; external apical elytral angle evenly convex. Mesosternal process short. Ratio of length of metepisternum/metacoxa: 1/1.53. Metatibia, ratio of width/length: 1/4.38. Metatarsi dorsally smooth; first metatarsomere as long as subsequent one.
Aedeagus: Fig. IV.3Q–S. Habitus: Fig. IV.3T. Female genitalia: Fig. IV.5K.

Diagnosis. *Pleophylla transkeiensis* sp. nov. is in external morphology similar to *P. ferruginea*. *Pleophylla transkeiensis* sp. nov. differs from the latter by the wider labroclypeus and the distinctly asymmetric parameres.

Variation. Body length: 9.7-10.6 mm, elytral length: 7.2-7.4 mm, maximum width: 5.1-5.7 mm. Density of dark spots may vary, sometimes completely without spots. Female: Antennal club of female straight, composed of 5 antennomeres, as long as the remaining antennomeres combined; 6^{th} antennomere short, a quarter as long as club length. Eyes smaller than in male: Ratio of eye diameter/interocular width: 0.51.

Etymology. The new species is named "*transkeiensis*", after its occurrence in the Transkei Region of South Africa.

Remarks. In a previous publication referred to this species as "SpM-16" (Özgül-Siemund and Ahrens 2015).

3.1.16. *Pleophylla ruthae* Ahrens, Beckett, Eberle & Fabrizi sp. nov.

(Figs IV.3U–X, IV.5L, IV.6)

Type material examined. Holotype: ♂ "South Africa, TVL Entabeni Forest Res. 23.00 S 30.16 E. Nov. 1978. G.L. Prinsloo & L. van Luik" (SANC). Paratypes: see electronic supplementary material C1.

Description. Body length: 10.7 mm, elytral length: 7.7 mm, maximum width: 5.7 mm.
Angle between lateral margins of labroclypeus and ocular canthus indistinct; anterior margin of labroclypeus medially weakly sinuate; margins of labroclypeus strongly reflexed; smooth area in front of eyes three times as wide as long; ratio of eye diameter / interocular width: 0.7. Pronotum with dense and thick erect setae; unicoloured; basal margin of pronotum narrowly interrupted medially (distinctly less than scutellum width); larger punctures on pronotum very coarse. Hypomeron not carinate. Elytra with dark spots; erect dorsal pilosity on elytra dense (more than 10 setae per interval); sutural interval in apical declivity of elytra distinctly elevated but convex; white, fine pilosity condensed to longitudinal patches; intervals punctate at least laterally; adpressed white pilosity long, 1/3 to 1/2 of width of elytral intervals; external apical elytral angle evenly convex. Mesosternal process short. Ratio of length of metepisternum/metacoxa: 1/1.44. Metatibia, ratio of width/length: 1/3.84. Metatarsi dorsally smooth; first metatarsomere as long as subsequent one.
Aedeagus: Fig. IV.3U–W. Habitus: Fig. IV.3X. Female genitalia: Fig. IV.5L.

Diagnosis. *Pleophylla ruthae* sp. nov. is in external morphology quite similar to *P. fasciatipennis*. *Pleophylla ruthae* sp. nov. differs in the sutural interval being convex in apical declivity of elytra and not delimitated by a sharp carina. In shape of parameres, *P. ruthae* sp. nov. is similar to *P. transkeiensis*, but the left paramere is basally less produced medially and the right paramere lacks a blunt median tooth.

Variation. Body length: 9.0-10.7 mm, elytral length: 6.6-8.0 mm, maximum width: 4.4-5.7 mm. Density of dark spots may vary, never completely without spots. Female: Antennal club of female straight, composed of 5 antennomeres, as long as the remaining antennomeres combined; 6^{th} antennomere shorter than a quarter of club length. Eyes smaller than in male: Ratio of eye diameter/interocular width: 0.5.

Etymology. This new species is dedicated to Ruth Müller (TMSA), collector of many *Pleophylla* species, in gratitude for her hospitality and long lasting support of our research in South Africa.

Remarks. In a previous publication we referred to this species as "SpM-23" (Özgül-Siemund and Ahrens 2015).

3.1.17. *Pleophylla warnockae* Ahrens, Beckett, Eberle & Fabrizi sp. nov.

(Figs IV.3Y–Ab, IV.5M, IV.6)

Type material examined. Holotype: ♂ "Balloon Forest Tzaneen 24°11'S 30°20'E 6-9 XI 1980 D.H. Jacobs" (TMSA). Paratypes: see electronic supplementary material C1.

Description. Body length: 10.0 mm, elytral length: 6.9 mm, maximum width: 4.9 mm.
Angle between lateral margins of labroclypeus and ocular canthus indistinct; anterior margin of labroclypeus medially straight; margins of labroclypeus strongly reflexed; smooth area in front of eyes three times as wide as long; ratio of eye diameter / interocular width: 0.73. Pronotum with dense and thick erect setae; unicoloured; basal margin of pronotum narrowly interrupted medially (distinctly less than scutellum width); larger punctures on pronotum very coarse. Hypomeron not carinate. Elytra with dark spots; erect dorsal pilosity on elytra dense (more than 10 setae per interval); sutural interval in apical declivity of elytra distinctly elevated but convex; white, fine pilosity condensed to longitudinal patches; intervals punctate at least laterally; adpressed white pilosity long, 1/3 to 1/2 of width of elytral intervals; external apical elytral angle evenly convex. Mesosternal process short. Ratio of length of metepisternum/metacoxa: 1/1.6. Metatibia, ratio of width/length: 1/3.62. Metatarsi dorsally

smooth; first metatarsomere as long as subsequent one.
Aedeagus: Fig. IV.3Y–Aa. Habitus: Fig. IV.3Ab. Female genitalia: Fig. IV.5M.

Diagnosis. *Pleophylla warnockae* sp. nov. is in external morphology quite similar to *P. ruthae* sp. nov. but it differs in the slightly shorter metatibia (ratio length / maximum width: 3.6) and the strongly reflexed margins of the labroclypeus; the basomedial teeth are widely separated from each other, while in *P. ruthae* sp. nov. they are adjacent.

Variation. Body length: 8.1-10.0 mm, elytral length: 6.3-7.1 mm, maximum width: 4.4–5.8 mm. Density of dark spots may vary, never completely without spots. Female: Antennal club of female straight, composed of 5 antennomeres, as long as the remaining antennomeres combined; 6^{th} antennomere shorter than a quarter of club length. Eyes smaller than in male: Ratio of eye diameter/interocular width: 0.5.

Etymology. The new species is dedicated to Rachel Warnock; her Master's thesis contributed to a better understanding of the integrative taxonomy of *Pleophylla*.

Remarks. In previous publications we referred to this species as "sp9" (Özgül-Siemund and Ahrens 2015; Eberle et al. 2016b).

3.1.18. *Pleophylla harrisoni* Ahrens, Beckett, Eberle & Fabrizi sp. nov.

(Figs IV.3Ac–Af, IV.5N, IV.6)

Type material examined. Holotype ♂ "S. Afr, Tvl. Nelspruit Nat. Res, rivulet val. 25.29 S - 30.55 E/ 9.2.1987 E-Y: 2433 beating leg Endrödy-Younga" (TMSA). Paratypes: see electronic supplementary material C1.

Description. Body length: 9.0 mm, elytral length: 5.7 mm, maximum width: 4.6 mm.
Angle between lateral margins of labroclypeus and ocular canthus indistinct; anterior margin of labroclypeus medially straight; margins of labroclypeus strongly reflexed; smooth area in front of eyes three times as wide as long; ratio of eye diameter / interocular width: 0.74. Pronotum with dense and thick erect setae; unicoloured. Hypomeron not carinate. Basal margin of pronotum narrowly interrupted medially (distinctly less than scutellum width); larger punctures on pronotum very coarse; Elytra with dark spots; erect dorsal pilosity on elytra dense (more than 10 setae

per interval); sutural interval in apical declivity of elytra delimited by a sharp carina; white, fine pilosity not condensed to longitudinal patches; intervals punctate at least laterally; adpressed white pilosity long, 1/3 to 1/2 of width of elytral intervals; external apical elytral angle evenly convex. Mesosternal process short. Ratio of length of metepisternum/metacoxa: 1/1.41. Metatibia, ratio of width/length: 1/3.87. Metatarsi dorsally smooth; first metatarsomere as long as subsequent one.
Aedeagus: Fig. IV.3Ac–Ae. Habitus: Fig. IV.3Af. Female genitalia: Fig. IV.5N.

Diagnosis. *Pleophylla harrisoni* sp. nov. is in external morphology rather similar to *P. fasciatipennis*, but it differs from the latter by the smooth area in front of the eyes being three times as wide as long. Parameres are quite similar to *P. warnockae* sp. nov. and *P. ruthae* sp. nov., but the dorsal lobe of the left paramere has in *P. harrisoni* sp. nov. only a single basal tooth.

Variation. Body length: 6.9-9.0 mm, elytral length: 5.2-5.7 mm, maximum width: 3.6-4.6 mm. Density of dark spots may vary, never completely without spots. Female: Antennal club of female straight, composed of 5 antennomeres, as long as the remaining antennomeres combined; 6^{th} antennomere shorter than a quarter of club length. Eyes smaller than in male: Ratio of eye diameter/interocular width: 0.5.

Etymology. This new species is dedicated to James du Guesclin Harrison (Pretoria). James supports our research in South Africa and recollected this species for our molecular analyses.

Remarks. In a previous publication we referred to this species as "SpM-17" (Özgül-Siemund and Ahrens 2015).

3.1.19. *Pleophylla mpumalanga* Ahrens, Beckett, Eberle & Fabrizi sp. nov.

(Figs IV.4A–D, IV.6)

Type material examined. Holotype: ♂ "S. Afr.: KwaZuluNatal Tygerkloof 27.51 S 31.19 E/ 28.10.2002: E-Y: 3563 grassland & forest leg. TMSA staff" (TMSA). Paratype: 1 ♂ "S. Afr.: KwaZulu Natal Ngome State forest/ 27.48 S - 31.25 E; 18-22.1.1993 Krüger & Dombrowsky" (ZFMK).

Description. Body length: 12.1 mm, elytral length: 8.9 mm, maximum width: 6.1 mm.
Angle between lateral margins of labroclypeus and ocular canthus indistinct; anterior margin of labroclypeus medially straight; margins of labroclypeus moderately reflexed; smooth area in front of eyes twice as wide as long; ratio of eye diameter / interocular width: 0.61. Pronotum with dense and thick erect setae; unicoloured; basal margin of pronotum narrowly interrupted medially (distinctly less than scutellum width); larger punctures on pronotum very coarse. Hypomeron carinate. Elytra with dark spots; erect dorsal pilosity on elytra dense (more than 10 setae per interval); sutural interval in apical declivity of elytra distinctly elevated but convex; white, fine pilosity condensed to longitudinal patches; intervals punctate at least laterally; adpressed white pilosity long, 1/3 to 1/2 of width of elytral intervals; external apical elytral angle evenly convex. Mesosternal process short. Ratio of length of metepisternum/metacoxa: 1/1.28. Metatibia, ratio of width/length: 1/4.2. Metatarsi dorsally smooth; first metatarsomere as long as subsequent one.
Aedeagus: Fig. IV.4A–C. Habitus: Fig. IV.4D.

Diagnosis. *Pleophylla mpumalanga* sp. nov. differs from *P. navicularis* and *P. silvatica* sp. nov. in the thick, white pilosity of elytra beeing condensed to longitudinal patches rather than being nearly evenly distributed.

Variation. Body length: 11.0-12.1 mm, elytral length: 8.7-8.9 mm, maximum width: 6.0-6.1 mm. Female: unknown.

Etymology. The new species is named "*mpumalanga*", after its occurrence in the Mpumalanga Province of South Africa.

3.1.20. *Pleophylla charlyi* Ahrens, Beckett, Eberle & Fabrizi sp. nov.

(Figs IV.4E–H, IV.7)

Type material examined. Holotype: ♂ "Tanzania 14.XII.1998 Njombe Southern Highland Werner & Lizler leg." (ZFMK).

Description. Body length: 12.1 mm, elytral length: 8.4 mm, maximum width: 6.0 mm.
Angle between lateral margins of labroclypeus and ocular canthus indistinct; anterior

margin of labroclypeus medially straight; margins of labroclypeus strongly reflexed; smooth area in front of eyes twice as wide as long; ratio of eye diameter / interocular width: 0.7. Pronotum with dense and thick erect setae; unicoloured; basal margin of pronotum narrowly interrupted medially (distinctly less than scutellum width); larger punctures on pronotum very coarse. Hypomeron not carinate. Elytra with dark spots; erect dorsal pilosity on elytra dense (more than 10 setae per interval); sutural interval in apical declivity of elytra flat and not delimited by a sharp carina; white, fine pilosity not condensed to longitudinal patches; intervals punctate at least laterally; adpressed white pilosity long, 1/3 to 1/2 of width of elytral intervals; external apical elytral angle evenly convex. Mesosternal process short. Ratio of length of metepisternum/metacoxa: 1/1.39. Metatibia, ratio of width/length: 1/4.28. Metatarsi dorsally smooth; first metatarsomere slightly longer than subsequent one. Female unknown. Aedeagus: Fig. IV.4E–F. Habitus: Fig. IV.4H.

Diagnosis. *Pleophylla charlyi* sp. nov. is in external morphology quite similar to *P. navicularis* and *P. silvatica* sp. nov. *Pleophylla charlyi* sp. nov. differs from both in the strongly asymmetrical ventral lobes of the parameres being distinctly produced dorsally. From *P. mpumalanga* sp. nov., which is similar in shape of parameres, *P. charlyi* sp. nov. differs in the evenly distributed white pilosity on the elytra and the large lobes produced dorsally by the extensions of the ventral lobes.

Etymology. The new species is named after one of its collectors, Karl Werner (his nickname was "Charly"). He passed away too early.

3.1.21. *Pleophylla lizleri* Ahrens, Beckett, Eberle & Fabrizi sp. nov.

(Figs IV.4I-L, IV.7)

Type material examined. Holotype: ♂ "Tanzania 7.XII.1999 Porote Mts. Mbeya prov. Werner & Lizler leg." (ZFMK). Paratypes: 2 ♂ "Tanzania 7.XII.1999 Porote Mts. Mbeya prov. Werner & Lizler leg." (ZFMK).

Description. Body length: 11.4 mm, elytral length: 7.6 mm, maximum width: 5.6 mm.
Angle between lateral margins of labroclypeus and ocular canthus indistinct; anterior margin of labroclypeus medially straight; margins of labroclypeus strongly reflexed;

smooth area in front of eyes twice as wide as long; ratio of eye diameter / interocular width: 0.78. Pronotum with dense and thick erect setae; unicoloured; basal margin of pronotum narrowly interrupted medially (distinctly less than scutellum width); larger punctures on pronotum very coarse. Hypomeron not carinate. Elytra with dark spots; erect dorsal pilosity on elytra dense (more than 10 setae per interval); sutural interval in apical declivity of elytra flat and not delimited by a sharp carina; white, fine pilosity not condensed to longitudinal patches; intervals punctate at least laterally; adpressed white pilosity long, 1/3 to 1/2 of width of elytral intervals; external apical elytral angle evenly convex. Mesosternal process short. Ratio of length of metepisternum/metacoxa: 1/1.18. Metatibia, ratio of width/length: 1/3.84. Metatarsi dorsally smooth; first metatarsomere as long as subsequent one.
Aedeagus: Fig. IV.4I-K. Habitus: Fig. IV.4L.

Diagnosis. *Pleophylla lizleri* sp. nov. differs from the quite similar *P. charlyi* sp. nov. and *P. mpumalanga* sp. nov. in the less extended lobe produced by the dorsolateral extension of the ventral lobe of the right paramere. Furthermore, *P. lizleri* sp. nov. differs from *P. charlyi* sp. nov. in the robust basomedial teeth of the left paramere which is basally more widened, and from *P. mpumalanga* sp. nov. by the wider parameres (lateral view).

Variation. Body length: 9.9-11.4 mm, elytral length: 7.1-7.6 mm, maximum width: 4.8-5.6 mm. Female unknown.

Etymology. The new species is named after one of its collectors, Robert Lízler (Hradec Králové).

3.1.22. *Pleophylla taitaensis* Ahrens, Beckett, Eberle & Fabrizi sp. nov.

(Figs IV.4M–P, IV.7)

Type material examined. Holotype ♂ "Kenya Taita Hills Wundanyi 18.3.-22.3.1997 Lgt. M. Snizek [sic]" (ZFMK). Paratypes: 2 ♂♂ "Coll. I.R.Sc.N.B. Kenya, Taita Hills Mbololo, III.2001 I.G. 31.839 Leg. S. Sabari" (ISNB).

Description. Body length: 11.5 mm, elytral length: 7.4 mm, maximum width: 5.4 mm. Angle between lateral margins of labroclypeus and ocular canthus indistinct; anterior margin of labroclypeus medially straight; margins of labroclypeus strongly

reflexed; smooth area in front of eyes 1.5 times as wide as long; ratio of eye diameter / interocular width: 0.83. Pronotum with dense and thick erect setae; unicoloured; basal margin of pronotum complete; larger punctures on pronotum very coarse. Hypomeron not carinate. Elytra with dark spots; erect dorsal pilosity on elytra dense (more than 10 setae per interval); sutural interval in apical declivity of elytra flat and not delimited by a sharp carina; white, fine pilosity not condensed to longitudinal patches; intervals punctate at least laterally; adpressed white pilosity long, 1/3 to 1/2 of width of elytral intervals; external apical elytral angle evenly convex. Mesosternal process short. Ratio of length of metepisternum/metacoxa: 1/1.4. Metatibia, ratio of width/length: 1/4.13. Metatarsi dorsally smooth; first metatarsomere slightly longer than subsequent one.
Aedeagus: Fig. IV.4M–O. Habitus: Fig. IV.4P.

Diagnosis. *Pleophylla taitaensis* sp. nov. differs from the very similar *P. lizleri* sp. nov. and *P. charlyi* sp. nov. by the ventral lobes of parameres being symmetrical and not extended dorsally.

Variation. Body length: 10.4-11.5 mm, elytral length: 7.4-7.8 mm, maximum width: 5.4-5.6 mm. Female unknown.

Etymology. The new species is named after its occurrence in the Taita Hills of Kenya.

3.1.23. *Pleophylla stalsi* Ahrens, Beckett, Eberle & Fabrizi sp. nov.

(Figs IV.4Q–T, IV.5, IV.6)

Type material examined. Holotype: ♂ "South Africa: MPU Graskop 5km on Pilgrimsrest rd 24°56'S 30.47E 12.XI.2005 1639m W. Breytenbach/ Collected from *Protea caffra* (Proteaceae) many leaves were beaten" (SANC). Paratypes: see electronic supplementary material C1.

Description. Body length: 9.9 mm, elytral length: 6.7 mm, maximum width: 5.1 mm.
Angle between lateral margins of labroclypeus and ocular canthus indistinct; anterior margin of labroclypeus medially straight; margins of labroclypeus moderately reflexed; smooth area in front of eyes twice as wide as long; ratio of eye diameter

/ interocular width: 0.64. Pronotum apparently glabrous (with only a few erect setae anteriorly, otherwise with only minute setae); unicoloured; basal margin of pronotum complete; larger punctures absent on pronotum. Hypomeron not carinate. Elytra with dark spots; erect dorsal pilosity on elytra sparse (less than 10 setae per interval); sutural interval in apical declivity of elytra distinctly elevated but convex; white, fine pilosity not condensed to longitudinal patches; intervals punctate at least laterally; adpressed white pilosity short, less than 1/4 of width of elytral intervals; external apical elytral angle evenly convex. Mesosternal process short. Ratio of length of metepisternum/metacoxa: 1/1.38. Metatibia, ratio of width/length: 1/3.94. Metatarsi dorsally smooth; first metatarsomere as long as subsequent one.
Aedeagus: Fig. IV.4Q–S. Habitus: Fig. IV.4T. Female genitalia: Fig. IV.5.

Diagnosis. *Pleophylla stalsi* sp. nov. differs from all other *Pleophylla* species by the short white pilosity on the dorsal surface, and the almost absent erect pilosity on the pronotum.

Variation. Body length: 9.9-10.4 mm, elytral length: 6.6-6.7 mm, maximum width: 4.6-5.1 mm. Density of dark spots may vary, often body reddish brown and completely without spots. Female: Antennal club of female straight, composed of 5 antennomeres, as long as the remaining antennomeres combined; 6^{th} antennomere shorter than a quarter of club length. Eyes smaller than in male: Ratio of eye diameter/interocular width: 0.5.

Etymology. This new species is dedicated to Riaan Stals (SANC) in gratitude for his support of our research in South Africa.

3.1.24. *Pleophylla wakkerstroomensis* Ahrens, Beckett, Eberle & Fabrizi sp. nov.

(Figs IV.4U–X, IV.6)

Type material examined. Holotype: ♂ "835047 X-DA1676 South Africa Free State: Wakkerstro[o]m, (Wetland Lodge env.); 1802m 27°20'19,6"S 30°09'11,1"E 14.-15.XII.2007 leg. D. Ahrens & S. Fabrizi, Pleophylla spSAX2" (ZFMK).

Description. Body length: 10.4 mm, elytral length: 7.6 mm, maximum width: 5.4 mm.

Angle between lateral margins of labroclypeus and ocular canthus indistinct; anterior margin of labroclypeus medially straight; margins of labroclypeus strongly reflexed; smooth area in front of eyes 1.5 times as wide as long; ratio of eye diameter / interocular width: 0.55. Pronotum apparently glabrous (with only a few erect setae anteriorly, otherwise with only minute setae); unicoloured. Hypomeron not carinate. Basal margin of pronotum complete; larger punctures absent on pronotum. Elytra without dark spots; erect dorsal pilosity on elytra sparse (less than 10 setae per interval); sutural interval in apical declivity of elytra delimited by a sharp carina; white, fine pilosity not condensed to longitudinal patches; intervals punctate at least laterally; adpressed white pilosity short, less than 1/4 of width of elytral intervals; external apical elytral angle evenly convex. Mesosternal process short. Ratio of length of metepisternum/metacoxa: 1/1.39. Metatibia, ratio of width/length: 1/4.06. Metatarsi dorsally smooth; first metatarsomere slightly longer than subsequent one. Aedeagus: Fig. IV.4U–W. Habitus: Fig. IV.4X.

Diagnosis. *Pleophylla wakkerstroomensis* sp. nov. is in external appearance rather similar to *P. stalsi* sp. nov., however, the dorsal pilosity is nearly completely absent in the former and parameres differ significantly in shape: ventral lobes in lateral view widened and convexly rounded at apex (while being sharply pointed in *P. stalsi* sp. nov.).

Etymology. The new species is named after its type locality, Wakkerstroom, in South Africa.

Remarks. In a previous publication we referred to this species as "spX2" (Eberle et al. 2016b).

3.1.25. *Pleophylla tongaatsana* Péringuey, 1904

(Figs IV.4Y–Ab, IV.6)

Pleophylla tongaatsana Péringuey 1904: 9.

Type material examined. Lectotype (here designated): ♂ "Upper Tongaat N. [sic] 11/01/ *Pleophylla tongaatsana* Type Py/ Type SAM/ Ent 31/4" (SAMC). Additional material examined: 1 ♂ "S. Afr.: Mpumalanga, Badplaas 25.44 S- 30.40 E/ 11.11.2002 E-Y: 3566 general collect 1410m leg. TMSA staff" (TMSA).

Description. Body length: 9 mm, elytral length: 6.3 mm, maximum width: 4.8 mm. Angle between lateral margins of labroclypeus and ocular canthus distinct; anterior margin of labroclypeus medially distinctly sinuate; margins of labroclypeus moderately reflexed; smooth area in front of eyes 1.5 times as wide as long; ratio of eye diameter / interocular width: 0.69. Pronotum with dense and thick erect setae; bicoloured; basal margin widely interrupted medially (equal to or more than scutellum width); larger punctures on pronotum moderately large. Hypomeron carinate. Elytra with dark spots; erect dorsal pilosity on elytra dense (more than 10 setae per interval); sutural interval in apical declivity of elytra flat and not delimited by a sharp carina; white, fine pilosity not condensed to longitudinal patches; intervals punctate at least laterally; adpressed white pilosity long, 1/3 to 1/2 of width of elytral intervals; external apical elytral angle evenly convex. Mesosternal process long. Ratio of length of metepisternum/metacoxa: 1/1.45. Metatibia, ratio of width/length: 1/4.07. Metatarsi dorsally punctate; first metatarsomere as long as subsequent one.
Aedeagus: Fig. IV.4Y–Aa. Habitus: Fig. IV.4Ab.

3.1.26. *Pleophylla kruegeri* Ahrens, Beckett, Eberle & Fabrizi sp. nov.

(Figs IV.4Ac-Af, IV.6)

Type material examined. Holotype: ♂ "S. Afr.: Mpumalanga Marie[p]skop, 1700m 12.12.1998; Kruger/ DNA voucher BMNH 837896" (TMSA). Paratypes: 1 ♂, 1 ♀ "RSA: Mpumalanga Prov., Mariepskop Forest Reserve, 1300-1600 m, 21.-26.XI.2008, leg. W. Schawaller" (SMNS), 2 ♀ "S. Afr.; Mpumalanga Mariepskop forest, 24.34 S – 30.52 E/ 23.11.2008, E-Y:3802 night with torch, 1613m leg. Ruth Müller" (TMSA).

Description. Body length: 10.6 mm, elytral length: 7.1 mm, maximum width: 5.1 mm.
Angle between lateral margins of labroclypeus and ocular canthus distinct; anterior margin of labroclypeus medially distinctly sinuate; margins of labroclypeus strongly reflexed; smooth area in front of eyes 1.5 times as wide as long; ratio of eye diameter / interocular width: 0.63. Pronotum apparently glabrous (with only a few erect setae anteriorly, otherwise with only minute setae); unicoloured; basal margin widely interrupted medially (equal to or more than scutellum width); larger punctures on

pronotum very coarse. Hypomeron carinate. Elytra with dark spots; erect dorsal pilosity on elytra dense (more than 10 setae per interval); sutural interval in apical declivity of elytra flat and not delimited by a sharp carina; white, fine pilosity condensed to longitudinal patches; impunctate (only striae punctate); adpressed white pilosity long, 1/3 to 1/2 of width of elytral intervals; external apical elytral angle blunt; Mesosternal process long; Ratio of length of metepisternum/metacoxa: 1/1.33. Metatibia, ratio of width/length: 1/4.33. Metatarsi dorsally smooth; first metatarsomere slightly longer than subsequent one.
Aedeagus: Fig. IV.4Ac-Ae. Habitus: Fig. IV.4Af.

Diagnosis. This species differs from all other known *Pleophylla* species by having the ventral and dorsal lobes of the parameres completely fused with each other, and without setae on the parameres.

Variation. Body length: 10.6-11.5 mm, elytral length: 7.1-8.1 mm, maximum width: 5.1-5.6 mm. Density of dark spots may vary, but never completely without spots. Female: Antennal club of female straight, composed of 5 antennomeres, as long as the remaining antennomeres combined; 6th antennomere shorter than a quarter of club length. Eyes smaller than in male: Ratio of eye diameter/interocular width: 0.5.

Etymology. The new species is named after one of its collectors, Martin Krüger (TMSA).

Remarks. The systematic placement of this species is still uncertain. The species differs clearly from all other *Pleophylla* species which might justify a separate genus. Many symplesiomorphies of *Pleophylla* and *Omaloplia* are absent from this species, such as the separate dorsal and ventral lobes on the parameres or the pilosity of parameres. However, *P. kruegeri* shares with *Omaloplia* only two plesiomorphies and therefore it does not occupy a basal position among the *Pleophylla* species in the cladistic analysis (Fig. IV.9), a prerequisite for the placement in a separate genus.
In a previous publication we referred to this species as "SpM-24" (Özgül-Siemund and Ahrens 2015).

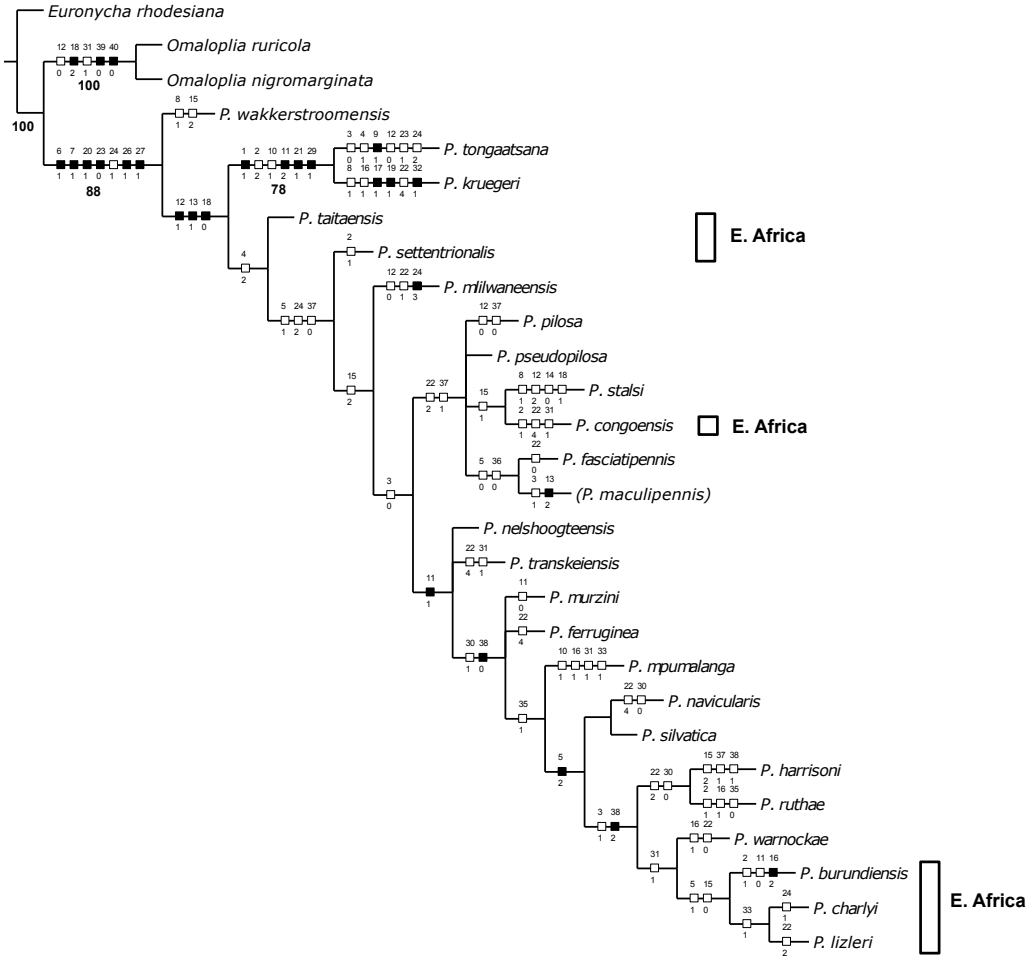

Figure IV.9. Strict consensus tree of the Winclada/NONA tree search (parsimony ratchet). Black squares – unambiguous apomorphies, white squares – unambiguous but to some degree homoplastic apomorphies. Small numbers above squares are characters, small numbers below squares are character states. Bootstrap values > 50 are shown in large and bold font under branches.

3.2. Patterns of distribution and phenology

Although the distribution of *Pleophylla* in South Africa is focused to the eastern and southern parts of the country, its northern occurrences extend patchily along the African Eastern Arc Mountains to the Democratic Republic of Congo and Uganda as well as Kenya. Between these areas records are lacking over long distances. Current distribution data suggest that the species of *Pleophylla* are closely associated with the Afrotropical forest, but as far as known, they lack in the tropical lowland forests of western Central Africa and western Africa. The ranges of many species overlap, at least partly, and numerous taxa co-occur syntopically (Table IV.1); so far we found up to five species in the same locality (in South Africa, Natal, Karkloof forest). The species with the widest range is *P. fasciatipennis* which is found from northeastern South Africa south to the Cape.

The main temporal occurrence in terms of numbers of species and individuals is during November and December, after February species richness and abundance drops considerably (Fig. IV.8). During May to August records are very rare, except in June where records are entirely lacking.

3.3. Phylogeny of the genus *Pleophylla*

Table IV.2. Morphological characters used for the cladistic analysis of *Pleophylla*.

1 *Clypeus, angle between lateral margins of labroclypeus and ocular canthus*: (0) indistinct; (1) distinct.

2 *Anterior margin of labroclypeus medially*: (0) straight; (1) weakly sinuate; (2) distinctly sinuate.

3 *Margins of labroclypeus*: (0) moderately reflexed; (1) strongly reflexed.

4 *Eyes, ratio diameter/interocular width*: (0) 0.5< x <0.65; (1) 0.65< x <0.72; (2) >0.72.

5 *Smooth area in front of eyes*: (0) 1.5 times as wide as long; (1) twice as wide as long; (2) 3 times as wide as long.

6 *Antennal club (males)*: (0) composed of three antennomeres; (1) composed of six antennomeres.

7 *Antennal club (females)*: (0) composed of three antennomeres; (1) composed of 4 antennomeres.

8 *Pronotum*: (0) with dense and thick erect setae; (1) apparently glabrous (with only a few erect setae anteriorly, otherwise with only minute setae); (2) with fine adpressed setae only.

9 *Pronotum*: (0) unicoloured; (1) bicoloured (sides yellow, disc darker).

10 *Hypomeron*: (0) non carinate; (1) carinate.

11 *Basal margin of pronotum*: (0) complete; (1) narrowly interrupted medially (distinctly less than scutellum width); (2) widely interrupted medially (equal to or more than scutellum width); (3) completely absent.

12 *Larger punctures on pronotum*: (0) moderately large; (1) very coarse; (2) absent.

13 *Elytra*: (0) without dark spots; (1) with dark spots; (2) with and without (when without dark spots elytra always yellowish).

14 *Erect dorsal pilosity on elytra*: (0) sparse (less than 10 setae per interval); (1) dense (more than 10 setae per interval); (2) absent.

15 *Sutural interval in apical declivity of elytra*: (0) flat and not delimited by a sharp carina; (1) distinctly elevated but convex; (2) delimited by a sharp carina.

16 *White, fine pilosity on elytra*: (0) not condensed to longitudinal patches; (1) condensed to longitudinal patches; (2) condensed to transverse patches.

17 *Elytral intervals*: (0) punctate at least laterally; (1) impunctate (only striae punctate).

18 *Elytra, adpressed white pilosity*: (0) long, 1/3 to 1/2 of width of elytral intervals; (1) short, less than 1/4 of width of elytral intervals; (2) absent.

19 *Elytra, external apical angle*: (0) evenly convex; (1) blunt.

20 *Mesosternum*: (0) without process; (1) with process.

21 *Mesosternal process*: (0) short; (1) long.

22 *Metatibia, ratio length / maximum width*: (0) $< 1/3.658$; (1) $1/3.65 < x < 1/3.8$; (2) $1/3.8 < x < 1/3.99$; (3) $1/3.99 < x < 1/4.25$; (4) $>1/4.25$.

23 *Metatarsi dorsally*: (0) smooth; (1) punctate.

24 *First metatarsomere*: (0) distinctly longer than subsequent one; (1) slightly longer than subsequent one; (2) as long as subsequent one; (3) slightly shorter than subsequent one.

25 *Pygidium*: (0) nearly as wide as long; (1) nearly 1.5 times as wide as long; (2) twice as wide as long.

26 *Ductus bursae*: (0) completely membranous; (1) irregularly sclerotised basally and with a triangular sclerite.

27 *Vaginal palps*: (0) large, half as long as bursa copulatrix wide; (1) small, one third as long as bursa copulatrix wide.

28 *Parameres*: (0) symmetrical; (1) nearly symmetrical; (2) distinctly asymmetrical.

29 *Parameres, dorsal and ventral lobe*: (0) separate; (1) fused.

30 *Parameres, ventral lobe in lateral view*: (0) narrower; (1) wide and short.

31 *Apex of phallobase*: (0) subsymmetrical (insertion of parameres at almost same level, or less than half of width of paramere displaced); (1) distinctly asymmetrical (insertion of parameres at one side distinctly displaced distally by more than half of paramere width).

32 *Parameres*: (0) with long dense setae; (1) glabrous.

33 *Ventral lobes of parameres*: (0) symmetrical, without lobes on dorsal face; (1) strongly asymmetrical, with large lobes extending to dorsal face.

34 *Ventral lobes of parameres*: (0) adjacent to phallobase as plate-like structures; (1) on apex of phallobase as motile, paramere-like appendix.

35 *Ventral lobes of parameres at apex (lateral view)*: (0) pointed; (1) convexly enlarged and rounded.

36 *Dorsal lobe of parameres*: (0) reduced in length (at maximum half as long as ventral lobe); (1) subequal (at least 2/3 as long) in length to ventral lobe.

37 *Dorsal lobe of left paramere*: (0) with one distal and two basal branches/ teeth; (1) with one distal and one basal branch/ tooth.

38 *Dorsal lobe of right paramere*: (0) with one distal and two basomedial branch/ teeth; (1) with one distal and one basomedial branch/teeth; (2) with two distal and one basal branch/ teeth; (3) with two distal and no basal branch/teeth.

39 *Ligular lobes*: (0) separate; (1) medially fused.

40 *Prementum before apex*: (0) elevated; (1) flat.

Table IV.1. Number of pairwise co-occurrences of the *Pleophylla* species compared to the number of sites with only one recorded species (i.e., spatially unique samples).

	P. transkeiensis	*P. murzini*	*P. pseudopilosa*	*P. ruthae*	*P. warnockae*	*P. fasciatipennis*	*P. ferruginea*	*P. navicularis*	*P. pilosa*	*P. silvatica*	*P. mpumalanga*	*P. wakkerstroomensis*	*P. nelshoogteensis*	*P. kruegeri*	*P. stalsi*	*P. tongaatsana*	*P. burundiensis*	*P. charlyi*	*P. congoensis*	*P. harrisoni*	*P. lizleri*	*P. mlilwaneensis*	*P. settentrionalis*	*P. taitaensis*	spatially unique samples
P. transkeiensis		1							1																3
P. murzini	1					1		1	2																4
P. pseudopilosa				1		1																			8
P. ruthae			1																						4
P. warnockae						1																			5
P. fasciatipennis		1	1		1		16	7	5	2															59
P. ferruginea						16		13	11	4	2	1													48
P. navicularis		1				7	13		6	5															24
P. pilosa	1	2				5	11	6		4	1														29
P. silvatica						2	4	5	4				6												14
P. mpumalanga							2		1																2
P. wakkerstroomensis							1																		1
P. nelshoogteensis										6				2	1	1									29
P. kruegeri													2												3
P. stalsi													1			1									3
P. tongaatsana													1		1										2
P. burundiensis																									7
P. charlyi																									1
P. congoensis																									4
P. harrisoni																									7
P. lizleri																									1
P. mlilwaneensis																									1
P. settentrionalis																									4
P. taitaensis																									2

For the morphology-based phylogenetic analysis, we were able to score a total of 40 characters for all species (Tables IV.2, IV.3). Parsimony ratchet tree searches (Winclada, NONA) with and without implied weighting produced much better resolved trees compared to BI (Appendix Fig. C1). Analysis of the known *Pleophylla* species resulted in nearly complete polytomies from BI with the exception of *P. wakkerstroomensis* + *P. stalsi* at the ancestral node. The unweighted parsimony ratchet resulted in four equally long trees (tree length 133 steps; consistency index: 0.45, retention index: 0.6). In the strict consensus tree (Fig. IV.9), the monophyly of *Pleophylla* is supported by six unambiguous apomorphies. Unfortunately, bootstrap support from most nodes was rather low for the unweighted and for the weighted parsimony ratchet. Surprisingly from the viewpoint of genital morphology, the most dissimilar *Pleophylla* species (*P. kruegeri*) was not the most ancestral lineage; instead this position was occupied by *P. wakkerstroomensis*. *Pleophylla kruegeri* and *P. tongaatsana* were sister species and sister to all other *Pleophylla* species except *P. wakkerstroomensis*. The eastern African and southern African species did not form separate clusters, but the first were nested as three separate clades within the southern African species (Fig. IV.9). Compared to the unweighted parsimony tree (Fig. IV.9), the consensus tree of the parsimony ratchet analysis with implied weighting differed only slightly in topology (Appendix Fig. C1). Besides the more ancestral position of *P. fasciatipennis*, the association of closely related species was similar, including the placement of *P. wakkerstroomensis* as the most ancestral species.

Single gene trees (ITS1, *cox1*) differed greatly in their general topology (Figs IV.10, IV.11). In contrast to the morphology based trees, DNA based trees (coalescence based species tree (Eberle et al. 2016b); species tree from four concatenated genetic markers (Eberle et al. 2016b); single gene trees) always had *P. warnockae* as the most ancestral taxon, with *P. wakkerstroomensis* nested among the more derived taxa. The only species that was not clearly distinguished by ITS1 was *P. pseudopilosa* that incompletely sorted with some specimens of *P. fasciatipennis*. Furthermore, we found a deep and well supported split within *P. ferruginea* caused by two ITS1 genotypes; however, this split was neither consistent with the *cox1* divergences nor with any morphological/ morphometric differentiation.

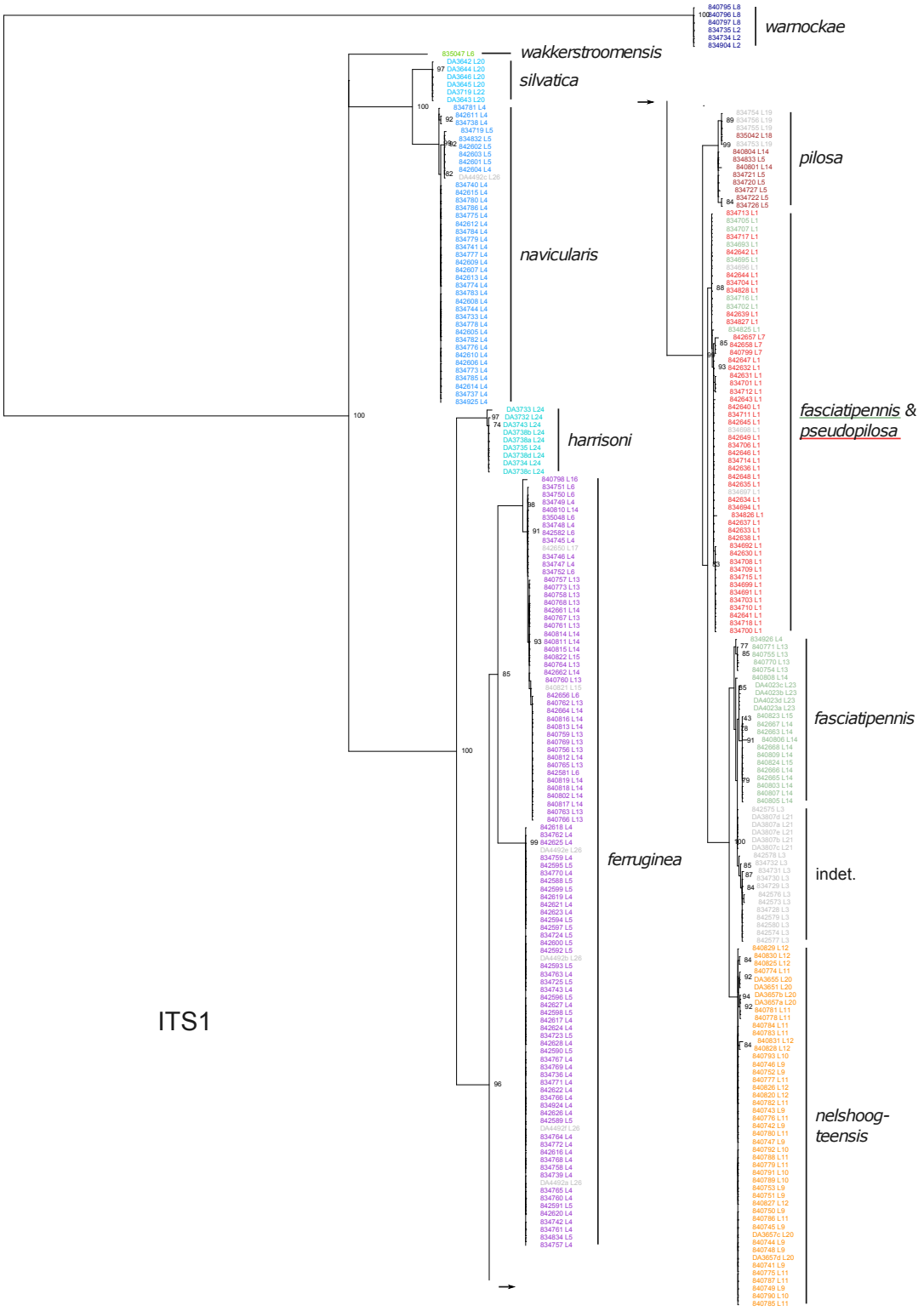

Figure IV.10. RAxML tree for 10 of the 24 *Pleophylla* species (each species with a different colour) based on ITS1 sequences. Branch support values > 50 are shown (tree divided in upper and lower parts).

Table IV.3. Morphological character matrix used for the cladistic analysis of *Pleophylla*.

	1 1111111112 2222222223 3333333334
	1234567890 1234567890 1234567890 1234567890
Euronycha rhodesiana	0202100200 0202000100 -01000000- 0000----11
Omaloplia ruricola	0010000000 0002000200 -11010020- 1011-0--00
O. nigromarginata	0010000000 0002000200 -11010020- 1011-0--00
P. fasciatipennis	0000011000 0111200001 0002211200 0001001311
P. maculipennis	0011011000 0121200001 0202211200 0001001311
P. navicularis	0001211000 1111100001 0402211200 0001110011
P. silvatica	0001211000 1111100001 0302211201 0001110011
P. pilosa	0001111000 0011200001 0202211200 0001010111
P. pseudopilosa	0001111000 0111200001 0202211200 0001011111
P. nelshoogteensis	0002111000 1111200001 0302211200 0001010111
P. mlilwaneensis	0012111000 0011200001 0103211200 0001010311
P. tongaatsana	1201011011 2011000001 1312211010 000101--11
P. warnockae	0012211000 1111110001 0002211201 1001110211
P. murzini	0000111000 0111100001 0302211101 0001010011
P. ferruginea	0002111000 1111200001 0402211101 0001010011
P. wakkerstroomensis	0010011100 0200200101 0301211201 0011111111
P. mpumalanga	0000111001 1111110001 0302211201 1011110011
P. transkeiensis	0000111000 1111200001 0402211200 1001010111
P. harrisoni	0012211000 1111200001 0202211200 0001111111
P. charlyi	0011111000 1111000001 0301211201 1011110211
P. lizleri	0012111000 1111000001 0202211201 1011110211
P. burundiensis	0111111000 0111020001 0302211201 1001110211
P. stalsi	0000111100 0210100101 0202211200 0001011311
P. ruthae	0111211000 1111110001 0202211200 0001010211
P. kruegeri	1210011101 2111011011 140121111- 01-1----11
P. congoensis	0102111000 0111100001 0402211200 1001011311
P. settentrionalis	0112111000 0111000001 0302211200 0001010311
P. taitaensis	0012011000 0111000001 0301211200 0001011311

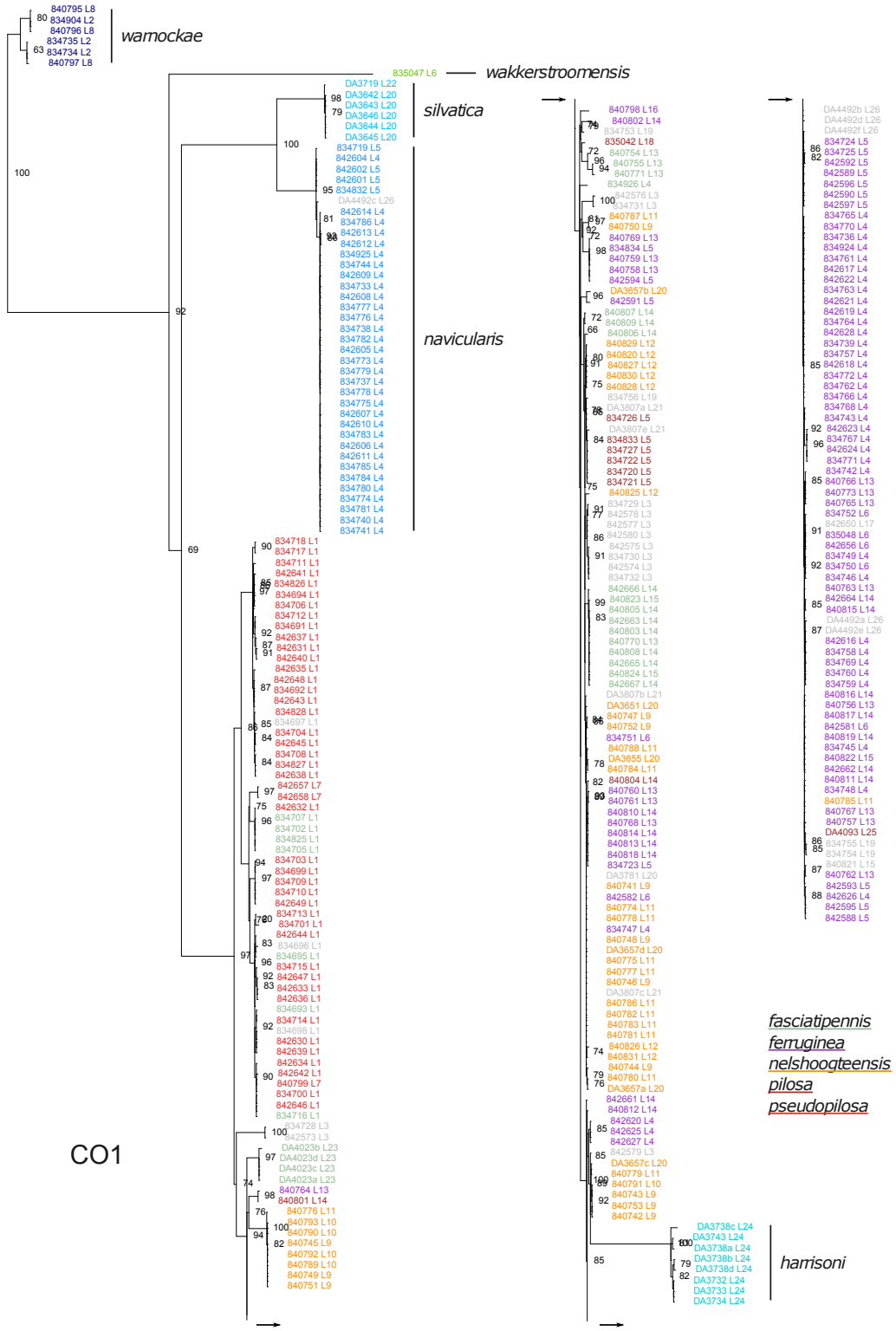

Figure IV.11. RAxML tree for 10 of the 24 *Pleophylla* species (each species with a different colour) based on *cox1* sequences. Branch support values > 50 are shown (tree divided in upper and lower parts).

4. Discussion

The chafer genus *Pleophylla* is an ancient lineage of the highly diverse tribe Sericini. The Sericini most likely originated in Africa (Eberle et al. 2016a), where its sister group Ablaberini exclusively occurs (Ahrens 2006). The eastern European and Mediterranean genus *Omaloplia* Schönherr, 1817 was identified earlier as the sister group of *Pleophylla* and their divergence was estimated to date back ca 80 My (Eberle et al. 2016a). Although sampling for the current DNA-based phylogeny did not cover all Sericini lineages, morphological evidence supports the monophyly of *Omaloplia* + *Pleophylla* (Ahrens 2006) and excludes the hypotheses that other genera might belong to this clade. This relationship thus represents an interesting case of southern African- Mediterranean disjunction (Bologna et al. 2008; Carpaneto 2008); however, most cases of southern African–Mediterranean disjunctions are younger and date back only up to the Paleogene and the Neogene (e.g., Audisio et al. 2008; Bologna et al. 2008). The onset of the crown group diversification of *Pleophylla* around 22 Mya (Eberle et al. 2016b) was possibly as a result of the aridification at the Eocene – Oligocene boundary (Feakins and Demenocal 2010).

Given the lack of molecular data for most of the species, the molecular and morphological phylogenies are only hardly comparable with each other. Discussions on the phylogeny of the entire group have to rely so far on the complete morphology-based tree (Fig. IV.9). This tree topology (with three east African clades nested within South African species) made three independent dispersals from southern Africa to the African Eastern Arc Mountains more likely than one single northward dispersal. An eastern African origin of the genus with a subsequent southward expansion was rendered even less likely by this tree. However, this hypothesis has not yet been assessed with more sophisticated methodologies (see Eberle et al. 2016a) and conclusions have to be done with some caution, also because sampling in eastern Africa is still very poor and the Sericini in this area are still extremely poorly known.

Despite *Pleophylla*'s age, some speciations are considered young (Eberle et al. 2016b). This is reflected in the lack of genetic resolution that was observed in *P. pseudopilosa* and *P. fasciatipennis* even in fast evolving genetic markers. Apart from the possibility of genetic introgression by hybridisation, incomplete lineage sorting is therefore a sufficient explanation for the observed patterns (Eberle et al. 2016b). Morphological differentiation of male and female genitalia, on the other hand, was shown to reliably distinguish even between recently diverged species (Özgül-Siemund and Ahrens 2015; Eberle et al. 2016b). It was therefore also prioritised in the decision to regard the genetic divergence of ITS1 in *P. ferruginea* as intraspecific variation. However, both molecular and morphological phylogenetic approaches are currently not completely satisfying: on the one hand sampling for the DNA-based tree search is still highly incomplete and fast fast-evolving genetic markers (ITS1 and *cox1*) show incongruent results; on the other hand, character incongruence in morphological cladistic analyses is high. As already shown earlier (Eberle et al. 2016b), ITS1 recovered morphospecies well, whereas many of the phylogenetically younger species could not be resolved with *cox1*. Due to other standard markers (*rrnL*, 28S) being uninformative in previous studies (Eberle et al. 2016b), we refrained from doing a tree search with combined genetic markers for the scope of this paper.

This paper can bee seen only as a primer for further research on this interesting group, many areas of the eastern African arc with its afrotropical forests are widely unexplored, in terms of the knowledge of the Sericini. Therefore, we expect many more new taxa and, in consequence, a better understanding of the evolution of *Pleophylla*, in particular in regard to their historical biogeography and of the faunal interchange between the patches of the afrotropical forests from Eastern Africa to South Africa – questions that can not be answered definitively at the moment.

5. Acknowledgements

We are grateful to P. Pacholátko (Brno), K. Schneider (MLUH), M. Uhlig and J. Frisch (both ZMHB), A. Mayekiso and S. van Noort (SAMC), M. Hartmann (NME), R. Stals (SANC), R. Müller (TMSA), W. Schawaller (SMNS), K.A. Johanson (NHRS), M. De Meyer (RMCA), B. Ratcliffe and M. Paulsen (Nebraska/ USNM) for the loan of the *Pleophylla* material in their collections. We are furthermore grateful to J. du G. Harrison and R. Müller for collecting additional specimens for DNA analysis and to D. Rohwedder for technical assistance for some images. J. du G. Harrison improved the English of the final version of this manuscript. This project was supported by grants from the German Science Association to D.A. (DFG/AH175/1 and AH175/3) and by SYNTHEYS (SE-TAF-3424). For providing D.A. with research and collection permits, we thank the various South African governmental institutions and departments in Eastern Cape (Permit No.: WRO 122/07WR and WRO123/07WR), Gauteng (Permit No.: CPF6 1281), Limpopo (Permit No.: CPM-006-00001), Mpumalangma (Permit No.: MPN-2009-11-20-1232), and Kwazulu-Natal (Permit Nos.: OP3752/2009, 1272/2007, 3620/2006).

References

Ahrens, D. and P. K. Lago (2008). Directional asymmetry reversal of male copulatory organs in chafer beetles (Coleoptera: Scarabaeidae): implications on left–right polarity determination in insect terminalia. *Journal of Zoological Systematics and Evolutionary Research* 46, 110–117. DOI: 10.1111/j.1439-0469.2007.00449.x.

Ahrens, D. (2004). *Monographie der Sericini des Himalaya (Coleoptera, Scarabaeidae)*. Dissertation.de - Verlag im Internet GmbH, Berlin, p. 534.

— (2006). The phylogeny of Sericini and their position within the Scarabaeidae based on morphological characters (Coleoptera: Scarabaeidae). *Systematic Entomology* 31, 113–144. DOI: 10.1111/j.1365-3113.2005.00307.x.

Audisio, P., A. D. E. Biase, A. H. Kirk-Spriggs, A. R. Cline, M. Trizzino, G. Antonini, and E. Mancini (2008). Molecular biogeography of Mediterranean and southern African disjunctions as exemplified by pollen beetles of the *Meligethes planiusculus* species-group and related taxa (Coleoptera: Nitidulidae; Meligethinae). *Biogeographica* 29, 45–65.

Blanchard, M. É. (1850). *Catalogue de la collection Entomologique. Classes des Insectes. Ordre des Coléoptères. part.: Melolonthidae*. Tome I. Muséum d'Histoire Naturelle de Paris, pp. 1–128.

Boheman, C. H. (1857). *Insecta Caffrariae. Annis 1838-1845. Part II. Holmiae*, pp. 1–395.

Bologna, M., P. Audisio, M. Biondi, and A. Casale (2008). The biogeographical patterns of disjunct distribution with special emphasis on the Mediterranean and southern African model. *Biogeographica* 29, 7–17.

Brenske, E. (1899). Die Serica-Arten der Erde. III. *Berliner Entomologische Zeitschrift* 44, 161–272.

Burmeister, H. (1855). *Handbuch der Entomologie. 4. Band. Besondere Entomologie, Fortsetzung. 2. Abteilung, Coleoptera Lamellicornia Phyllophaga chaenochela*. Berlin: Theodor Christian Friedrich Enslin, Berlin, pp. 1–467.

Carpaneto, G. M. (2008). The Mediterranean-southern African disjunct distribution pattern in the scarab beetles: a review (Coleoptera Scarabaeoidea). *Biogeographia* 29, 68–79.

Dalla Torre, K. W. (1912). "Scarabaeidae: Melolonthinae I." In: *Coleopterorum Catalogus 45*.

Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular biology and evolution* 29, 1969–1973. DOI: 10.1093/molbev/mss075.

Eberle, J., S. Fabrizi, P. Lago, and D. Ahrens (2016a). A historical biogeography of megadiverse Sericini – another story "out of Africa"? *Cladistics*, article first published online. DOI: 10.1111/cla.12162.

Eberle, J., R. C. M. Warnock, and D. Ahrens (2016b). Bayesian species delimitation in *Pleophylla* chafers (Coleoptera) – the importance of prior choice and morphology. *BMC Evolutionary Biology* 16, 94. DOI: 10.1186/s12862-016-0659-3.

Erichson, W. (1847). *Naturgeschichte der Insecten Deutschlands. Erste Abtheilung. Coleoptera.* 3, Lfrg. 4. Berlin: Nicolaische Buchhandlung, Berlin, pp. 552–986.

Feakins, S. J. and P. B. Demenocal (2010). "Global and african regional climate during the Cenozoic." In: *Cenozoic Mammals of Africa.* Ed. by L. Werdelin. Berkeley: University of California Press, Berkeley, pp. 45–56.

Goloboff, P. A. (1999). *NONA, version 2.0.*

Huelsenbeck, J. P. and F. Ronquist (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics (Oxford, England)* 17, 754–5.

Katoh, K., K. Misawa, K.-i. Kuma, and T. Miyata (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research* 30, 3059–66.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Mentjies, and A. Drummond (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 1647–1649.

Kerdelhué, C., G. Roux-Morabito, J. Forichon, J.-M. Chambon, A. Robert, and F. Lieutier (2002). Population genetic structure of *Tomicus piniperda* L. (Curculionidae: Scolytinae) on different pine species and validation of *T. destruens* (Woll.) *Molecular Ecology* 11, 483–494.

Krajčík, M. (2012). *Checklist of the World Scarabaeoidea.* Animma.x - Supplement 5, pp. 1–278.

Lanfear, R., B. Calcott, S. Y. W. Ho, and S. Guindon (2012). PartitionFinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. *Molecular biology and evolution* 29, 1695–1701. DOI: 10.1093/molbev/mss020.

Lewis, P. O. (2001). A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic biology* 50, 913–925. DOI: 10.1080/106351501753462876.

Machatschke, J. W. (1959). Phylogenetische Untersuchungen über die Sericini (sensu Dalla Torre 1912) (Coleoptera: Lamellicornia: Melolonthidae). *Beiträge zur Entomologie1* 9, 730–746.

Nixon, K. C. (2002). *Winclada (BETA), version 1.00.08.*

Nixon, K. C. (1999). The Parsimony Ratchet, a New Method for Rapid Parsimony Analysis. *Cladistics* 15, 407–414. DOI: 10.1111/j.1096-0031.1999.tb00277.x.

Özgül-Siemund, A. and D. Ahrens (2015). Taxonomic utility of female copulation organs in Sericini chafers (Coleoptera, Scarabaeidae), with special reference to asymmetry. *Contributions to Zoology* 84, 167–178.

Page, R. D. M. (2001). *NDE [Nexus Data Editor].* http://taxonomy.zoology.gla.ac.uk/rod/NDE/.

Péringuey, L. (1904). Descriptive catalogue of the Coleoptera of South Africa (Lucanidae and Scarabaeidae). *Transactions of the South African Philosophical Society* 13, 1–293.

Ronquist, F. and J. P. Huelsenbeck (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574. DOI: 10.1093/bioinformatics/btg180.

Santos, H., J. Rousselet, E. Magnoux, M.-R. Paiva, M. Branco, and C. Kerdelhue (2007). Genetic isolation through time: allochronic differentiation of a phenologically atypical population of the pine processionary moth. *Proceedings of the Royal Society B: Biological Sciences* 274, 935–941. DOI: 10.1098/rspb.2006.3767.

Schaufuss, L. W. (1871). Die Arten der Gattung *Pleophylla* Er. *Nunquam otios I* 6, 231–233.

Simon, C., F. Frati, A. 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. *Annals of the Entomological Society of America* 87, 651–701.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxford, England)* 30, 1312–1333. DOI: 10.1093/bioinformatics/btu033.

Strong, E. and D. Lipscomb (1999). Character coding and inaplicable data. *Cladistics* 15, 363–371.

The Global Biodiversity Information Facility (2014). *GBIF*. http://www.gbif.org/. (Accessed on Aug. 30, 2014).

Vogler, A. P. and R. Desalle (1994). Evolution and Phylogenetic Information Content of the ITS-1 Region in the Tiger Beetle Cicindela dorsalis. *Molecular biology and evolution* 11, 393–405.

Chapter V.

The evolution of morphospace in phytophagous scarab chafers: no competition – no divergence?

This chapter is published in:
Eberle, J., R. Myburgh, D. Ahrens (2014). The Evolution of Morphospace in Phytophagous Scarab Chafers: No Competition – No Divergence? *PLoS One* 9. DOI: 10.1371/journal.pone.0098536

Authors' contributions to the original article:
Morphometric analyes, comparative analyses: JE; morphometric measurements: RM; manuscript design and writing: JE, DA.

1. Introduction

Phytophagous scarabs (Coleoptera: Scarabaeidae) are a very diverse group of some 25,000 described species of beetles (Scholtz and Grebennikov 2005) which includes more than two thirds of all species in the superfamily Scarabaeoidea. Their monophyly is supported by a number of distinct morphological synapomorphies (Ritcher 1958; Balthasar 1963; Browne and Scholtz 1998). Early in taxonomic history, they were recognized as a group called Pleurosticti (Erichson 1847). Pleurosticts are usually subdivided into four major subfamilies including: Dynastinae, Rutelinae, Melolonthinae, and Cetoniinae, plus several other small groups (Smith 2006). Most species are highly polyphagous, with the adults generally feeding on leaves, flowers or pollen of a wide range of plant taxa, and the larvae primarily feeding on soil humus, living roots, or decaying wood. Because of this polyphagy, their tremendous diversity cannot be explained by insect-host plant co-diversification, a widely accepted hypothesis for the great species diversity in phytophagous insects (Ehrlich and Raven 1964; Mitter et al. 1991; Farrell 1998), hence alternative hypotheses are needed that may explain their successful diversification.

For many groups of organisms it was argued that niche partitioning, as a result of competition, leads to a positive relationship between species richness and the ecomorphological diversity of animal assemblages (Ricklefs and Miles 1994). It is well known and widely accepted that resource partitioning is one of the most important factors in scarab biology leading to profound structural changes and adaptations for particular feeding functions or foraging behavior (Ritcher 1958).

Scholtz and Chown (1995) proposed a substantial shift in Scarabaeoidea biology with the use of living plant material as a food resource instead of dead or decayed organic matter. They assumed that the massive radiation of the main pleurostict lineages (Melolonthinae, Adoretini, Anomalini, Dynastinae, and Cetoniinae) followed the rapid diversification of the angiosperms during the Late Cretaceous – Early

Palaeogene. Unlike in dung beetles, considered a model group for comparative studies of niche partitioning and functional structure (Peck and Forsyth 1982; Hanski and Cambefort 1991; Giller and Doube 1994; Finn and Gittings 2003; Horgan and Fuentes 2005; Inward et al. 2011), food resources of the phytophagous pleurosticts are less patchy in distribution and less ephemeral. While the food resources of the saprophagous ancestors of the pleurostict scarabs (Scholtz and Chown 1995) were restricted to a relatively limited two-dimensional stratum (upper soil layers), with the rise of angiosperms a vast food space became available (Jermy 1985). It is assumed that this new third dimension of food availability generally reduced competition among herbivores (Jermy 1985; Kaplan and Denno 2007) and the exploitation of different parts of the plants, like roots, stems, leaves, and florescences (Ritcher 1958; Scholtz and Chown 1995) provided further possibilities for avoiding competition. The presence of various aggregation mechanisms (volatiles, pheromones) for host location and/or mate finding (Ritcher 1958; Loughrin et al. 1995; Potter et al. 1996) and a highly complex chemical ecology (Meinecke 1975; Leal 1998) seem to support this hypothesis.

If competition for food resources triggered morphospace diversification and assemblage structure in pleurosticts, we would, as in the closely related dung beetles, observe an increased divergence of morphospace among similar feeding types. Alternatively, if we would observe less or no divergence in morphospace between similarly feeding lineages, we would expect little or no directed selection on morphological traits. However, further environmental pressures may also cause divergence.

While actual inter-specific competition is difficult to measure and needs to be explored at the assemblage level (Inward et al. 2011), competition in the past that no longer exists due to partitioning of the species niche or extinction of less competitive species ('ghost of competition past', (Connell 1980)) may be reflected in the morphospace. However, phylogenetic lineages will differ in morphospace if their common ancestors did, because members of a lineage share a greater similarity in their morphology as a result of the lingering legacy of a common ancestor (Felsenstein 1985; Harvey and Pagel 1991; Richman and Price 1992; Polly 2001). Competition that led to niche partitioning in the ancestors of extant lineages is therefore also visible at a phylogenetic level. However, under the hypothesis of reduced competition we might also encounter divergence between different feeding types (herbivorous, floricolous,

etc.), despite their spatial avoidance of competition, due to subsequent adaptation to the life style in relation to the use of a new food resource.

Here we used a multivariate analysis of body length measurements of external body morphology that was linked to a phylogenetic hypothesis of the group (Ahrens and Vogler 2008) to investigate the evolution of scarab morphospace. Our ecomorphological approach followed Wainwright and Reilly (1994) assuming that body size and allometric shape variation reflect differences in the species ecology and behavior (Ricklefs and Travis 1980; Travis and Ricklefs 1983; Douglas and Matthews 1992; Ribera et al. 1999; Melville et al. 2006). Additionally, we explored the presence of evolutionary key innovations that were possibly linked with quantitative traits of body shape and that might have promoted the diversification of certain lineages (Heard and Hauser 1995). We explored the morphospace divergence in a twofold approach: 1) Searching for simple phenetic divergence at a nested level and detecting which traits contribute most to the observed divergences. I.e. searching simply for differences between the different feeding types, major lineages, and sister clades. Our null assumption was "no divergence – no competition" such that among species of the same feeding type that do exhibit no or very little divergence in morphospace no competition occurs. 2) Exploring morphospace divergence in relation to molecular rates of evolution: Through the link with the molecular branch lengths we were able to infer directed selection that is linked with significant divergence of body morphospace at any phylogenetic level. For traits under neutral evolution and therefore stochastic drift, rates of morphological change are correlated with those of molecular evolution. Observed divergence and uncorrelated rates among similar feeding types would provide insight, whether (and what kind of) directed selection (likely as result of competition avoidance) had an impact on pleurostict morphospace divergence, which would allow to identify key factors of the successful scarab diversification.

2. Material and Methods

2.1. Ethics statement

We obtained permission from the Zoological Research Museum A. Koenig, Bonn (ZFMK) to access, loan and dissect the material in the collections.

2.2. Taxon sampling and morphometric measurements

Based on the phylogenetic analysis of Ahrens and Vogler (2008), we sampled a single specimen of 182 species of all principal lineages of phytophagous Scarabaeidae from Neotropical, Palearctic, Afrotropical, Oriental, and Australasian regions (see Table D1). Vouchers are deposited in the collection of the Zoological Research Museum A. Koenig, Bonn (ZFMK).

Twenty linear distance measurements were performed on adult beetles to capture the complexity of body shape (Fig. V.1A–C). Characters subjected to a strong sexual dimorphism were not used because female and male specimens were included in the molecular phylogenetic analysis (Ahrens and Vogler 2008). The measurements were taken directly (where possible) from the sequenced voucher specimens of Ahrens and Vogler (2008) with the help of an ocular grid on a Zeiss SM20 Stereomicroscope, and values were converted to millimeters for the different magnifications. Measurements were taken in such a way that the endpoints were in focus. In order to reduce the variance introduced by several sources of subjective measurement errors (Claude 2008), the measurements of all specimens were repeated 5 times and subsequent analyses were conducted with the means of the measured values.

Table V.1. F-values from non-parametric MANOVA of the complete sampling (excluding singletons) regarding 95% of the total variation. The values for the size-corrected data set are shown in the upper triangle, those for the uncorrected in the lower one. Significant differences ($p < 0.05$) are highlighted in bold. Higher F-values for the same significant pairings are shaded in gray in the respective triangle.

	Adoretini	Anomalini	Aphodiinae	Cetoniini	Clade B	Dynastinae	Glaphyridae	Hopliinae	Hybosoridae	Scarabaeinae	Sericini A	Sericini B	Sericini C	SWM[1]	Valgini
Adoretini		4.96*	9.45*	12.34*	**6.80**	12.64*	3.71*	3.80*	3.49*	6.34*	**16.90**	**24.50**	**21.45**	**5.88**	6.42*
Anomalini	5.22*		12.78*	6.07*	**6.04**	**5.47**	3.61*	4.55*	2.64*	6.96*	**10.22**	**13.50**	**10.46**	1.96	4.13*
Aphodiinae	39.30*	47.75*		11.95*	**8.43**	11.87*	6.02	12.79*	3.10	4.96	16.32*	**31.88**	**25.77**	6.86*	7.28
Cetoniini	7.73*	1.29	33.83*		**13.97**	9.67*	4.07*	8.71*	5.72*	6.16*	**12.61**	**20.71**	**14.15**	**4.17**	3.55*
Clade B	3.96	0.88	**21.07**	0.55		**9.91**	3.75*	**5.97**	1.95	8.66*	**24.02**	**57.48**	**36.40**	**5.15**	8.58*
Dynastinae	13.83*	3.12	49.12*	0.45	0.98		9.10*	13.32*	1.91	7.90*	**12.24**	**23.52**	**18.27**	**5.24**	5.12*
Glaphyridae	0.39	1.41	24.58	2.10	1.01	4.37		1.95	6.58	3.71	10.24*	**12.05**	10.05*	2.47	6.17
Hopliinae	1.30	8.40*	17.05*	9.73*	6.21*	15.44*	1.18		5.63*	6.42*	**19.17**	**24.37**	**20.56**	5.06*	7.57*
Hybosoridae	2.87	6.29*	13.29	5.92	3.02	9.76*	5.89	0.56		2.99	7.88*	**11.11**	9.72*	2.10	4.57
Scarabaeinae	5.13	0.81	30.33	0.47	0.38	1.13	1.93	5.16	8.10		14.28*	**19.66**	15.85*	4.87*	2.44
Sericini A	8.01*	**20.27**	17.92*	**22.65**	**19.12**	**30.91**	4.81*	3.77*	1.63	11.26*		**5.32**	4.51*	**6.47**	6.59*
Sericini B	2.58	9.99*	**24.01**	**16.18**	**32.11**	**21.02**	1.56	2.64	1.51	5.19*	3.90*		1.76	**11.46**	9.54*
Sericini C	5.51*	**18.89**	**17.44**	**23.88**	**30.56**	**30.92**	3.17	2.51	1.24	8.68*	0.47	4.64*		**7.24**	7.56*
SWM1	0.43	3.28	5.76*	5.05*	7.61*	6.50*	0.35	0.24	0.14	1.47	1.10	1.19	1.22		3.37*
Valgini	4.83	9.63*	0.82	8.25	8.61*	10.72*	1.59	2.09	0.69	2.69	2.82	7.14*	3.56	1.57	

*Significant without sequential Bonferroni correction
[1]Southern World Melolonthinae

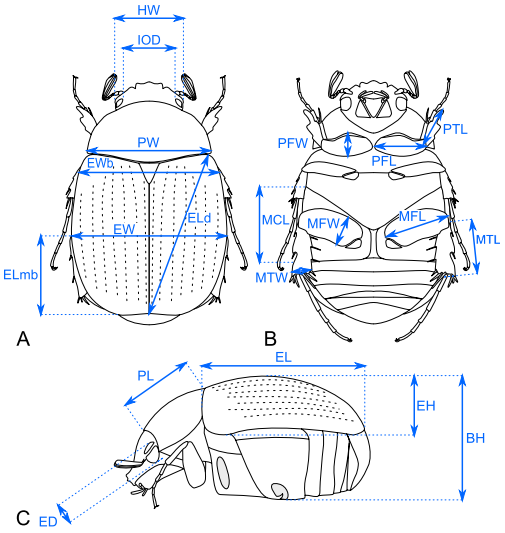

Figure V.1 Illustration of the measured traits. Schematic drawings of a Sericini beetle in (A) dorsal, (B) ventral, and (C) lateral aspect. Body: BH - maximal body height, EH - maximal elytra height, EL - maximal elytra length, Eld - maximal diagonal elytra length, Elmb - length from maximal body width to elytral apex, EW - maximal elytra width, Ewb - elytral width at middle of scutellum, PL - maximal pronotum length, PW - maximal pronotum width; Head: ED - maximal eye diameter, HW - maximal head with including eyes, IOD - minimal interocular distance (dorsal view); Legs: MCL - maximal length of metacoxa, MFL - maximal length of metafemur, MFW - maximal width of metafemur, MTL - maximal length of metatibia, MTW - maximal width of metatibia, PFL - maximal length of profemur, PFW - maximal width of profemur, PTL - maximal length of protibia.

2.3. Analysis of morphospace

Analyses of morphospace were implemented based on the Bayesian phylogenetic tree (Ahrens and Vogler 2008) on the preferred alignment as a backbone. Morphospace was explored for the complete data set of all specimens and for five subsets that compare major sister clades. Comparisons between sister clades with low support values are omitted except those with relationships that are also well established in traditional morphology-based systematics. All calculations for the analysis of morphospace were made within the R statistics environment version 2.15 [40] (R Development Core Team 2012) unless otherwise stated. Obtained linear measurements were log10-transformed to render more linear relations among variables and to obtain a similar dimension of variance (Ricklefs et al. 1981; Klingenberg 1996). Generally, the major component of variance in morphometric data sets of biological specimens is explained through size (Jolicoeur 1963; Burnaby 1966; Ferrario et al. 1995). To avoid a strong bias of size over variation of shape, we employed approaches that separate size from shape information. In landmark-based geometric morphometrics, this is achieved using “two point registration” methods (Zelditch et al. 2004; Berner 2011),

but for linear measurements there is still some debate regarding how to perform this separation (Adams and Rohlf 2000; Berner 2011). Here, we employed the Burnaby Back Projection Method (BBPM) (Burnaby 1966) by projecting the log-transformed data on the isometric size vector and returning it to the original coordinate system (Adams and Rohlf 2000; Blankers et al. 2012) as implemented in an R-code provided by Blankers et al. (2012). This method has the advantage of deriving a composite measure of size from all traits and considering shape as the projection onto the orthogonal space of this isometric vector. Data treated in this manner are subsequently referred to as size-corrected data (set). Correcting the data for size can strongly affect the results depending on the method used and must be considered well. Therefore, we compared the results from the BBPM with shape data derived from a linear regression (residuals) against overall body length (Jolicoeur et al. 1984; Reist 1985) which was chosen to be representative of the beetles' body size. Because a high error is introduced to the total body length measure through the motility of the prothorax against the pterothorax, a proxy was used by calculating the logarithm of the sum of pronotal and elytral length (log(PL+EL)). The impact of size (percentage of variation that is explained by size alone) was assumed to be represented by the percentage of variation explained by the first principal component of the uncorrected data set.

Patterns of morphometric covariation were analyzed with standard principal component analysis (PCAs; Jolicoeur and Mosimann 1960; Teissier 1960) on uncorrected and size-corrected data. Results were visualized with the help of the ade4 package (Dray and Dufour 2007). Additionally, the molecular phylogeny was projected onto the morphospace explained by PCs 1 and 2 using the function *phylomorphospace* in the R package phytools (Revell 2012). The program therefore estimates the positions of the ancestral nodes using a maximum likelihood approach.

Statistical evaluation of group differentiation in morphospace was done by MANOVA and linear discriminant analysis (LDA). To avoid confusion through noise introduced from measurement errors or minor unspecific variation (Gauch Jr 1982; Peres-Neto et al. 2005; Ezard et al. 2010), we only used the principal components that explained 95% of total variation. Non-parametric MANOVA (Anderson 2001) was performed for the complete data set and each sister clade subset in PAST 2.17 (Hammer et al. 2001) to test for significant differentiation between lineages. Sequential Bonferroni (Holm 1979) correction was applied. LDA was conducted on the

same groupings to evaluate group discrimination by the reassignment probabilities (McLachlan 2004) which were evaluated by leave-one-out cross-validation using the MASS-package (Venables and Ripley 2002) in R. Lineages represented only by a single species, i.e. Ablaberini, were included in the PCA but had to be excluded from LDA and MANOVA.

2.4. Feeding habits and morphospace

Inference of the potential influence of the food resource on morphospace variation was done by mapping feeding habits of each species onto morphospace. Details on feeding behavior were taken from the literature and were complemented with personal observations (Table V.1). Coprophagous (COP) and saprophagous (SAP) species were represented by Aphodiinae/ Scarabaeinae and Hybosoridae, respectively. Anthophilous (ANT) species exclusively forage on flowers, feeding on pollen and nectar, whereas herbivorous species (HERB) devour various plant materials, including foliage, twigs, and petals. Dynastinae species examined here are sap / fluid utilizers (SFU) feeding under ground on stems or roots in order to gain access to fluids from the wounds (Ritcher 1958). Adults of *Pachypus* do not feed (NF).

A correlation analysis between morphospace and feeding types was performed employing phylogenetic generalized least squares regression in the package caper using the *pgls* function (Orme et al. 2012). The assigned feeding types were used as independent variables and (standard) principal components explaining 95% of cumulative variation as dependent variables representing the morphospace. To improve the fit of the data to the tree, Pagel's branch length transformation variable λ (internal branch lengths are multiplied with λ; Pagel (1999)) was set to be estimated by maximum likelihood. κ (each branch length is raised to the power κ, (Pagel 1999)) and δ (the node heights are raised to the power δ, (Pagel 1999)) were set at 1.

A possible correlation between molecular and morphological distances between the specimens was estimated by Mantel-tests, performing 10,000 permutations of Pearson correlations with the R package vegan (Oksanen et al. 2013). The analysis was made for size-corrected data for all members of each feeding type separately.

2.5. Detecting driven selection and key innovations

Reduced correlation between molecular and multivariate morphometric distances is likely to indicate decoupling of molecular and morphological rates of evolution, with accelerated or decelerated rates of evolution in either of the traits, i.e. directed selection on morphospace evolution. Therefore, Mantel-tests with Pearson correlation were performed on distance matrices of patristic distances (calculated with the *cophenetic*-function in the R-package ape (Paradis et al. 2004) from the molecular tree (Ahrens and Vogler 2008)) and Euclidean distances of the respective morphological data sets. To infer individual traits that underlay directed selection, i.e. that deviate from Brownian Motion, the descriptive K statistic of Blomberg et al. (2003) was calculated for every trait over the complete size-corrected data set using the R package phytools (Revell 2010, 2012). A K value greater than one implies that close relatives are more similar than expected under Brownian motion evolution (Blomberg et al. 2003).

Branches in the phylogeny, where the molecular and the morphological distances between nodes deviate from each other, were detected by projecting both the uncorrected and the size-corrected data set on the constrained topology of the phylogenetic tree (Ahrens and Ribera 2009; Cooper and Purvis 2009). The branch lengths were inferred with the *optim.phylo.ls*-function from the phytools package (Revell 2012) using Euclidean distance matrices of the respective data sets. Negative branches were set to zero. For both the size-corrected and the uncorrected data set, ratios of morphological and molecular branch lengths were calculated for each branch. Values above and below the 95% confidence interval of the ratios were considered as significantly different in their branch lengths, i.e. indicating an extraordinary decoupling of molecular and morphological rates and consequently directed selection at the respective 'outlier' branch. Because the lengths of internal branches and tips often largely differed, they were evaluated separately. (cf. Fig. V.2)

Additionally, we calculated standardized phylogenetic independent contrasts (Garland Jr. 1992; McPeek 1995a,b) in order to compare evolutionary rates of morphospace divergence between clades. For this objective, we used the multivariate approach introduced by McPeek et al. (2008) and applied it to both data sets. The method of McPeek et al. (2008) was implemented in R and the script is provided in supplemental file D2. The ultrametric tree necessary for this approach was calculated based on the

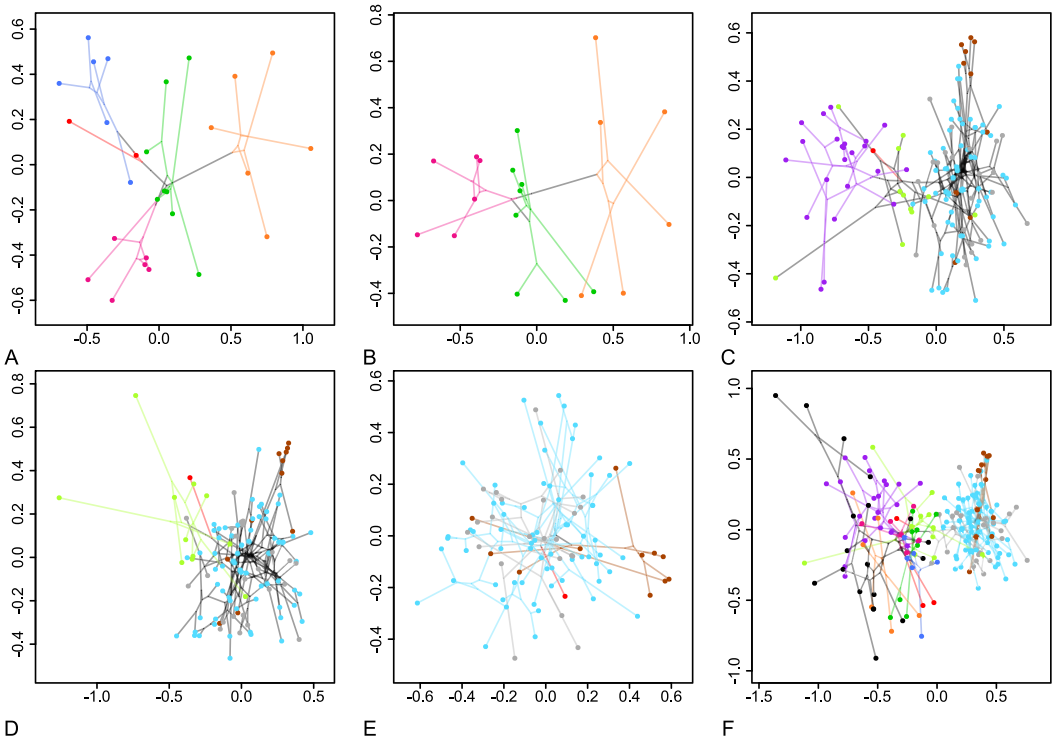

Figure V.2. Lineage diversifications in morphospace. Phylomorphospace projections of the molecular phylogenetic tree [31] for the sister clade subsets 1-5 (A-E) and the complete data set (F) showing the first two PC axes of the size-corrected data set.

preferred alignment (2513 bp) of Ahrens and Vogler (2008) using PathD8 (Britton et al. 2007), with the root of an arbitrary age of one. Ancestral linear size measurements of traits possibly linked with presumptive key innovations were reconstructed with the function *fastAnc* in phytools (Revell 2012) using a Maximum Likelihood approach.

3. Results

Results for the two methods for removing isometric size from the data, the Burnaby Back Projection Method (BBPM) and linear regression against a size metric, were quite similar (Figs. V.3,V.4). Therefore most results for the latter method are presented in the supplement information only (Tables D9,D10,D11, Fig. D5) and were compared concisely to those of BBPM in the discussion.

3.1. Feeding habits and morphospace

The phylogenetic generalized least squares analysis did not recover any significant correlation between the feeding types and the morphospace. The r^2-values of the regression were low for size-corrected and uncorrected data (adjusted $r^2 = 0.031$ and 0.025, respectively). Plots of the PCA-scores of PCs 1 and 2 of the complete data set analysis were very similar for the size-corrected and the uncorrected data set (Fig. V.3D–F), showing a large overlap of all phytophagous groups (anthophilous, herbivorous, and sap / fluid feeders). Non-feeders showed no separation from these groups. Coprophagous and saprophagous feeders appeared somewhat divergent from the phytophagous groups, although an overlap in particular with the herbivores, which occupied a very vast morphospace, was also present. Separate clusters became evident for all feeding types except juicy feeders when projecting feeding types on the morphospace of major lineages (Fig. V.3G–I).

Significant correlations between molecular and morphological distances could not be found within any of the feeding types.

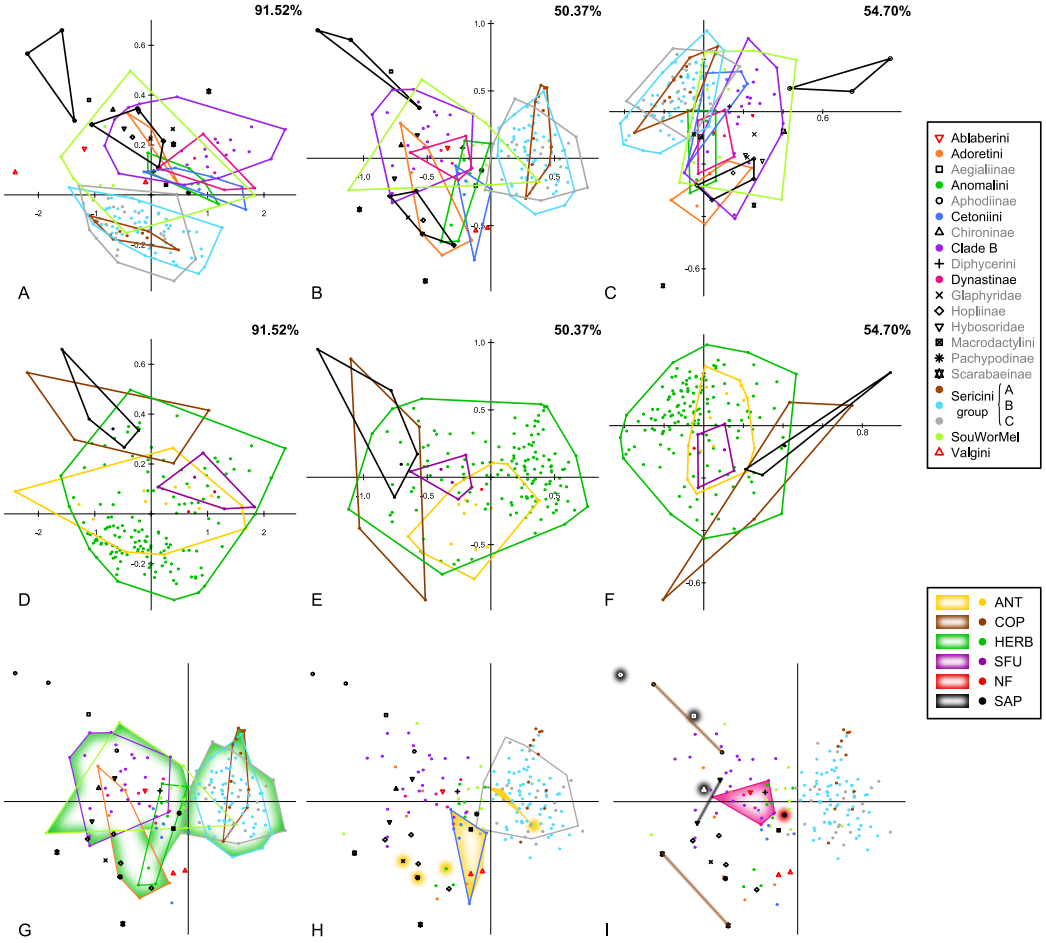

Figure V.3. Patterns of morphospace covariation between major phylogenetic lineages and feeding types. Scatterplots of the principal component scores from the analysis of the complete sampling of (A, D) the uncorrected and the size-corrected data sets from (B, E) the Burnaby Back Projection Method (BBPM) and (C, F) the linear regression method with (A-C) major phylogenetic lineages and (D-F) feeding types projected on it (ANT = anthophilous, COP = coprophagous, HERB = herbivorous, SFU = sap / fluid utilizers, NF = not feeding, SAP = saprophagous). The percentage of variance explained by principal component 1 and 2 is given in each upper right corner. Taxa with more than 2 members are surrounded by a similarly colored hull. (G-I) Morphospace divergence within the feeding types projected on scatterplots of the principal component scores from size corrected data (BBPM): (G) Herbivores, (H) anthophilous, and (I) the remaining feeding types. Dots are color-coded in the molecular phylogeny (Fig. VI.3A) for phylogenetic lineages. x-axis: PC1, y-axis: PC2.

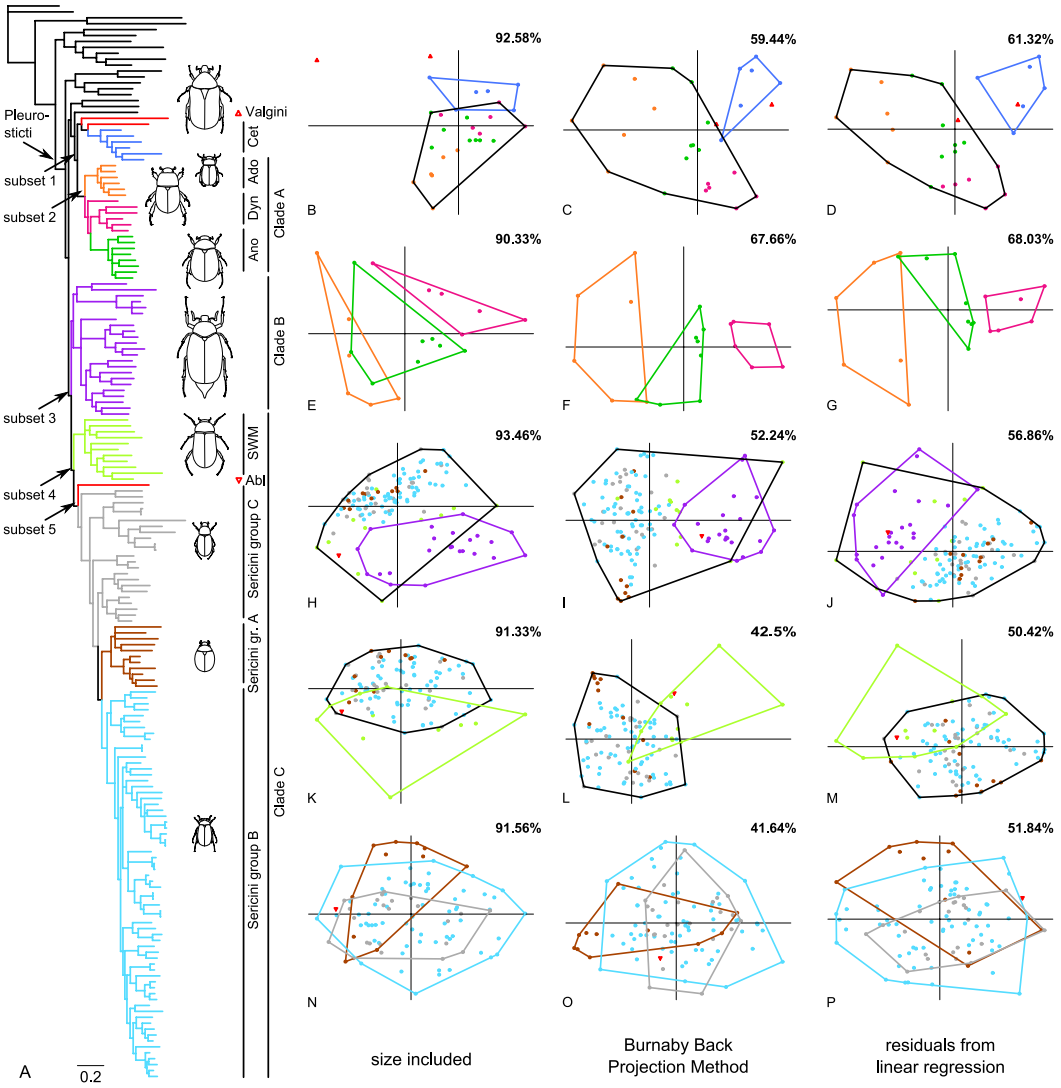

Figure V.4. Backbone phylogeny and morphospace covariation between sister clade subsets of the complete sampling. (A) Phylogenetic tree of major scarab lineages from the Bayesian analysis of Ahrens and Vogler (2008). Scatterplots of the principal component scores for the uncorrected and the size-corrected data sets: (B-D) Cetoniini (Cet) + Valgini and Adoretini (Ado) + Anomalini (Ano) + Dynastinae (Dyn), (E-G) Adoretini, Anomalini, and Dynastinae, (H-J) Clade B and Southern World Melolonthinae (SWM) + Ablaberini (Abl) + Sericini, (K-M) SWM and Ablaberini + Sericini, and (N-P) Ablaberini and Sericini subgroups. The groups are color-coded in the phylogeny and the scatterplots. The percentage of variance explained by principal component 1 and 2 is given in the top right corner. Groups with more than 2 members are surrounded by a similarly colored hull. Black hulls border sister lineages for illustration of divergence. x-axis: PC1, y-axis: PC2.

3.2. Morphospace divergence of phylogenetic lineages

The impact of size (the variation of uncorrected data explained by PC1 alone), was high among the complete sampling and all subsets (82.4% to 89.5%, Table D8). A large overlap of most lineages was visible from the scatterplots of the scores of principal components 1 and 2 for both the size-corrected and the uncorrected data set (Fig. V.3 A–B). However, in morphospace Aphodiinae and Sericini were quite separate from the other lineages in both data sets. Despite the large overlap of the lineages in the PCA scatterplots, the MANOVAs showed generally significant differentiations between the major lineages within the size-corrected as well as the uncorrected data set (size-corrected data set: $F = 13.42$, $p<0.0001$; uncorrected data set: $F = 8.49$, $p<0.0001$). However, only 38 of the 105 pair-wise comparisons yielded significant results for the size-corrected data and only 14 were significant for the uncorrected data (Table V.1). Most of the significant results involved one of the three Sericini subgroups. The lineages of Adoretini, Glaphyridae, Hopliinae, Hybosoridae, Scarabaeinae, and Southern World Melolonthinae were distinguished only by the size-corrected data set from at least one other clade. Valgini did not yield significant results. MANOVA of the uncorrected data gained higher F-values in comparison to those from the analysis of the size-corrected data set (6 out of 14 with higher F-values) for lineages that were represented by many large species (EL + PL>15 mm), such as Cetoniini, Clade B, and Dynastinae.

The comparison of sister lineages (Fig. V.4B–P) was not influenced by the interference of variation from other lineages. While subsets 1 and 2 showed an improved differentiation compared to the analysis of the complete sampling, the patterns of morphospace-distribution changed only marginally for subsets 3, 4 and 5. However, the comparisons of sister lineages generally showed more specific information about which part of body shape (PC vectors of the subset) represent the morphological divergence of sister clades. Size correction improved the outcome only slightly in these comparisons (Fig. V.4).

The two major lineages of subset 1 (Cetoniinae vs. Clade A; Figure 3B-C) are well differentiated for the uncorrected and the size-corrected data set although a slight overlap was present. These results were congruent with the results of the MANOVAs of subset 1 where the analysis of the size-corrected data set resulted in a nearly 8-fold

Table V.2. F-values from non-parametric MANOVA (Anderson 2001) of each subset (ss1-ss5, excluding singletons) regarding 95% of the total variation. The values for the size-corrected data set are shown in the upper triangle, those for the uncorrected in the lower one. Significant differences ($p < 0.05$) are highlighted in bold.

Subset 1	Cetoniinae	Clade A	
Cetoniinae		**6.79**	
Clade A	0.87		
Subset 2	Adoretini	Anomalini	Dynastinae
Adoretini		**5.13**	**12.96**
Anomalini	**0.02**		**5.61**
Dynastinae	**0.00**	0.10	
Subset 3	Clade B	Clade C	
Clade B		**47.90**	
Clade C	**41.85**		
Subset 4	Sericini	SWM1	
Sericini		**11.02**	
SWM[1]	1.08		
Subset 5	Sericini A	Sericini B	Sericini C
Sericini A		**5.29**	**4.83**
Sericini B	3.82		**2.11**
Sericini C	0.56	4.58	

[1]Southern World Melolonthinae

higher F-value (Table V.2), while F-values of uncorrected data were not significant. LDA incorrectly reassigned five specimens with the uncorrected data set (85.71% correctly reassigned, Fig. D1A), but only three specimens with the size-corrected data set (89.29% correctly reassigned, Fig. D1B). *Microvalgus* was only correctly assigned to Valgini with the size-corrected data set.

For subset 2, both data sets reveal a differentiation between groups, although a slight overlap was present in the uncorrected data set (Fig. V.4 E–G). Anomalini take an intermediate position in morphospace between Adoretini and Dynastinae. MANOVA showed only slight separation for the uncorrected data compared to the size-corrected data (Table V.2). LDA correctly reassigned 90% of the specimens to the respective groups for the size-corrected data set, but still 80% for the uncorrected data set (Fig. D3C–D). Scatterplots of the first two PC axes of subset 3 (comprising all 'Melolonthinae') show a large overlap of Clade B and Clade C in the size-corrected and the uncorrected data set. This is mainly caused by the Southern World Melolonthinae that widely "invade" the morphospace of Clade B (Fig. V.4H–J; green dots). MANOVA, which considers multiple PCA dimensions of the morphospace, suggests a distinct separation of the two sister lineages. The F-value was slightly higher for the size-corrected data set (Table V.2) what was consistent with the reassignment probabilities from LDA (correctly reassigned for uncorrected data: 95.52%, for size-corrected data: 96.27%). Although the number of specimens correctly reassigned by the discriminant function was nearly equal in subset 3, the number of specimens with reassignment probabilities over 95% decreased from 96% to 94% (Fig. D1E–F). In contrast to all others, for this data subset the membership to the predefined groups could be recovered more unambiguously with the uncorrected data.

For data subset 4, the correction for size resulted in no marked differences in patterns of the specimen-distribution in morphospace, and Ablaberini + Sericini and Southern World Melolonthinae showed a considerable overlap in morphospace (Fig. V.4K–M). MANOVA's F-values resulting from the size-corrected data set are about 10-fold higher, and those of uncorrected data were not significant (Table V.2). LDA on the size-corrected data set correctly reassigned 92.79% of the specimens (63.64% of Southern World Melolonthinae) whereas 90.99% (only 18.18% of SWM) were correctly reassigned for the uncorrected data set.

Subset 5 (Sericini) was sampled more in detail (Ahrens and Vogler 2008) and was further subdivided into three groups (Fig. V.4N–P): Group A (subtribe Trochalina) is monophyletic and partly characterized by a more or less spherical body shape (genus *Trochalus*), group B (subtribe Sericina) is the monophyletic sister clade to the Trochalina and is characterized by a more oval body shape, while group C represents the paraphyletic remainder of Sericini basal to group A + B. Based on our measurements, Sericini subgroups differed only slightly in morphospace (Fig. V.4N–P). MANOVA on the size-corrected data set supported these results suggesting a differentiation of group A (Trochalina) from both other subgroups of Sericini (Table V.2). MANOVA on uncorrected data revealed no significant differentiation. The reassignment of specimens to the predefined groups by the discriminant function was improved through the size-correction of the data set (correctly reassigned for uncorrected data set: 69.0%, for size-corrected data set: 75.0%; Fig. D1I–J). The projection of the phylogenetic tree onto PCs 1 and 2 revealed the divergence within the subsets and clades (Fig. V.2). Analysis of subsets 1 and 2 revealed a shift in morphospace of the ancestors of Anomalini, Adoretini, Cetoniinae, and Dynastinae, with subsequent diversification within the respective morphospace units. A similar result was observed for the genus *Trochalus* (Fig. V.2E, lower half, brown dots).

3.3. Which traits shape the morphospace divergence?

Because the directions of morphospace divergences between the lineages were much more evident from the size-corrected data, we used these to investigate trait behavior in the context of the measured divergence. PCA vector loadings allowed us to draw conclusions about the contribution of traits to the morphological divergence of the lineages (Tables D1–D6, Fig. D2). Scarab morphospace (PC axes 1 and 2) inferred from the complete data set (Fig. V.3B, Table D1) was principally influenced by traits of limb length (PTL, MTL, PFL, MFL) and elytral height (EH). The strongest influence in total was metacoxal length (MCL).

Observed principal components (PCs) of variation from sister clade comparisons (subsets 1–5) were not influenced by the interference of variation with other lineages, and thus we were able to detect morphological divergence linked with the divergence of sister lineages. In comparison to its sister clade A (including Adoretini, Anomalini,

and Dynastinae), Cetoniinae were mainly characterized by traits that are equivalent for a relatively smaller head (HW, IOD), smaller eyes (ED), a longer pronotum (PL), and a dorso-ventrally flattened body (EH; Table D2, Fig. D2B, C). Adoretini were found to have a wider head and larger eyes (HW, IOD, ED), a shorter pronotum (PL) compared to Dynastinae (Table D3, Fig. D2E, F). Dynastinae had shorter and stouter extremities. Anomalini had an intermediate position in morphospace between Adoretini and Dynastinae. Within Melolonthinae (subset 3) there was an overwhelming influence of metacoxal length (MCL, Table D4). Also the width of hind limbs contributed to the differentiation of the sister clades. The influence of MCL was distinctly reduced when Clade B was excluded (subset 4, Table D5, Fig. D2K, L). However, the major part of Sericini was still found to be divergent in longer metacoxa and broader hind limbs in general. Within Sericini (subset 5), specimens with a more spherical appearance (higher values for EW, EWb, PW, EH, and BH but also PL; Table D6) were located in the left side of the plot (Figures V.4O, D2N–O), whereas more elongate specimens were located in the right side of the plot. Therefore, a significant shape divergence must have occurred within Sericini group A with *Trochalus* appearing on the extreme left side of total variation along the x-axis while *Allokotarsa*, *Idaeserica*, and *Ablaberoides* are more centered in the plot.

The inference of the influence of the measured traits on scarab morphospace (Table D1) is complemented by the estimated phylogenetic signal (descriptive K-statistics, Table D1, (Blomberg et al. 2003)) of every trait from the size-corrected data set. All traits except MCL have K-values below 1, indicating that they tend to exhibit a weaker signal than expected under Brownian motion model (Blomberg et al. 2003). The K-value found for the metacoxal length (MCL; K = 3.21) was with distance the highest value, indicating a higher conservatism for this trait, with close relatives being more similar than expected under Brownian motion evolution and though possibly indicating directed selection. However, the absolute amount of K was highly influenced by the biased sampling towards Sericini in our study. In fact, if we simulated a stepwise decreased amount of Sericini by pruning species of this lineage from the tree and the morphometric data set, the K-value went below 1 (with 3 sampled species of Sericini, Fig. D3C). The K-value for MCL was, however, always the highest or second highest value in total. A subsequent maximum likelihood reconstruction of the

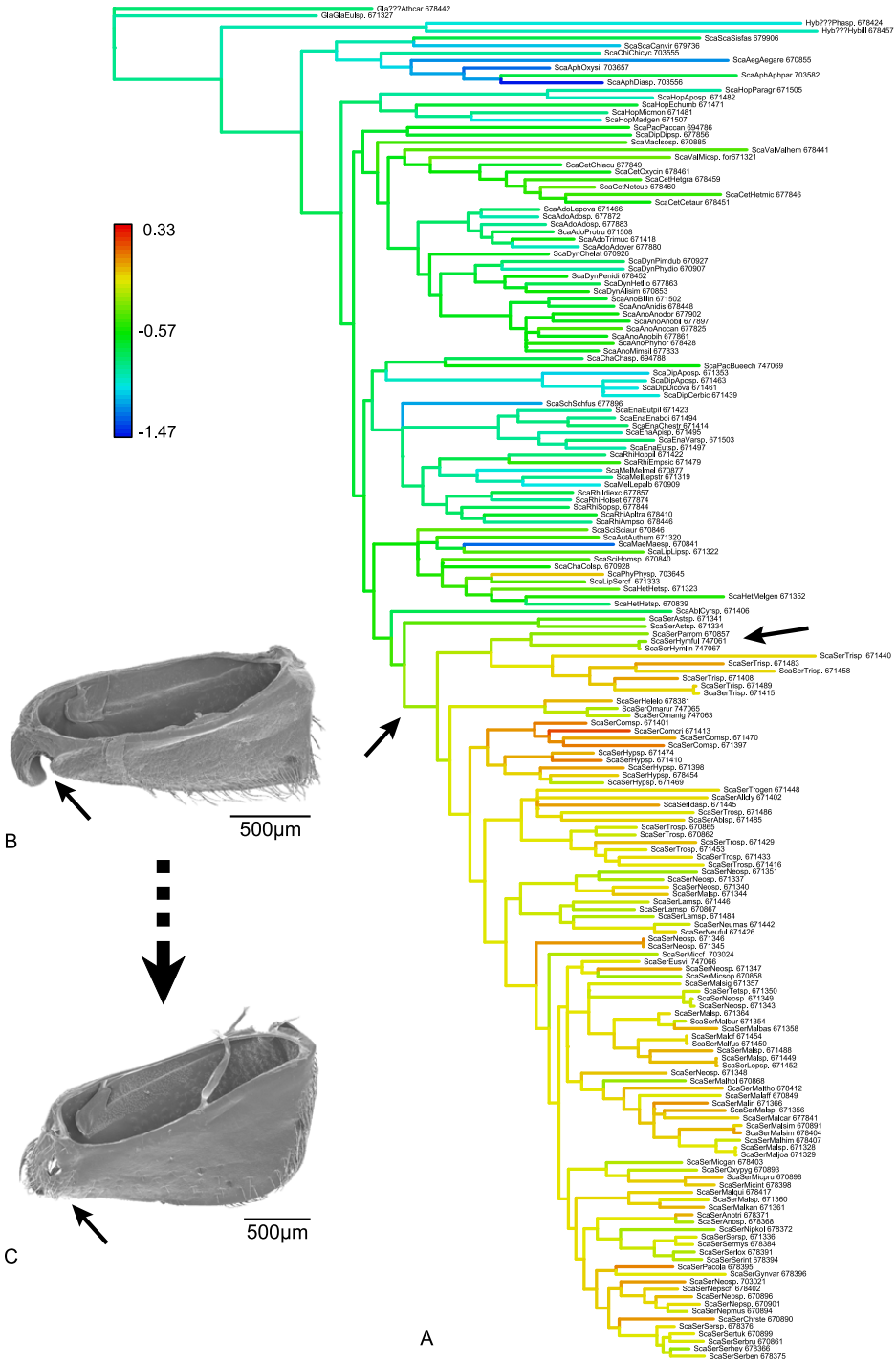

Figure V.5. Correlated evolution of metacoxal length and the secondary metacoxal ostium. (To be continued on next page.)

Figure V.5. (Continued.) (A) Reconstruction of relative metacoxal length in ancestral nodes of the molecular phylogeny (Ahrens and Vogler 2008). The left hand arrow shows the internal branch where ancestral relative metacoxal length strongly increases and where the secondary ostium of metacoxa is closed by the medial apophysis (Garland et al. 1992). The right hand arrow points to the clade of *Hymenoplia* and *Paratriodonta* (see text for explanation). (B) *Chasmatopterus* spec., metacoxa, dorsal view: secondary ostium open (arrow). (C) *Hymenoplia castilliana*, metacoxa, dorsal view: secondary ostium closed (arrow). The numbers in the legend correspond to the size-corrected values of metacoxal length.

size-corrected MCL (Fig. V.5A) on the tree revealed a strong shift of MCL's relative length at branches of ancestral Sericini (Fig. V.5A left hand arrow).

3.4. Morphological vs. molecular rates of evolution

The Mantel tests between molecular and morphological distance matrices were significant only for the size-corrected data of the complete sampling. Correlation was low ($r = 0.11$, $p = 0.01$).

Optimization of morphospace variation onto the phylogenetic tree provided a better measure of the relative morphological divergence of phylogenetic lineages (Fig. V.6B, C), allowing a more general assessment of morphological change at diverse phylogenetic levels, especially when extraordinary rate decoupling was identified through rate ratio outliers (Figs. V.5B, C, D4). Given that variation in the uncorrected data set was mainly induced by size-differences of the species, long branches in the respective tree should be mainly attributed to change in size of the hypothetical ancestor of the group. Long branches that result from the size-corrected data set, however, indicated a higher rate of change in shape. Internal branches were of highest interest for the inference of morphological lineage divergence because they represented change in morphospace common to a whole clade. Terminal optimized branches, however, were generally longer than internal ones, suggesting that only a very few morphospace shifts exceeded interspecific variation of extant taxa within selected lineages.

Only a few specimens within Sericini group C and none of the internal branches coincidently showed significant branch length differences, thus supporting increased morphological change and presumed directed selection (Fig. V.6B, C; in both, the

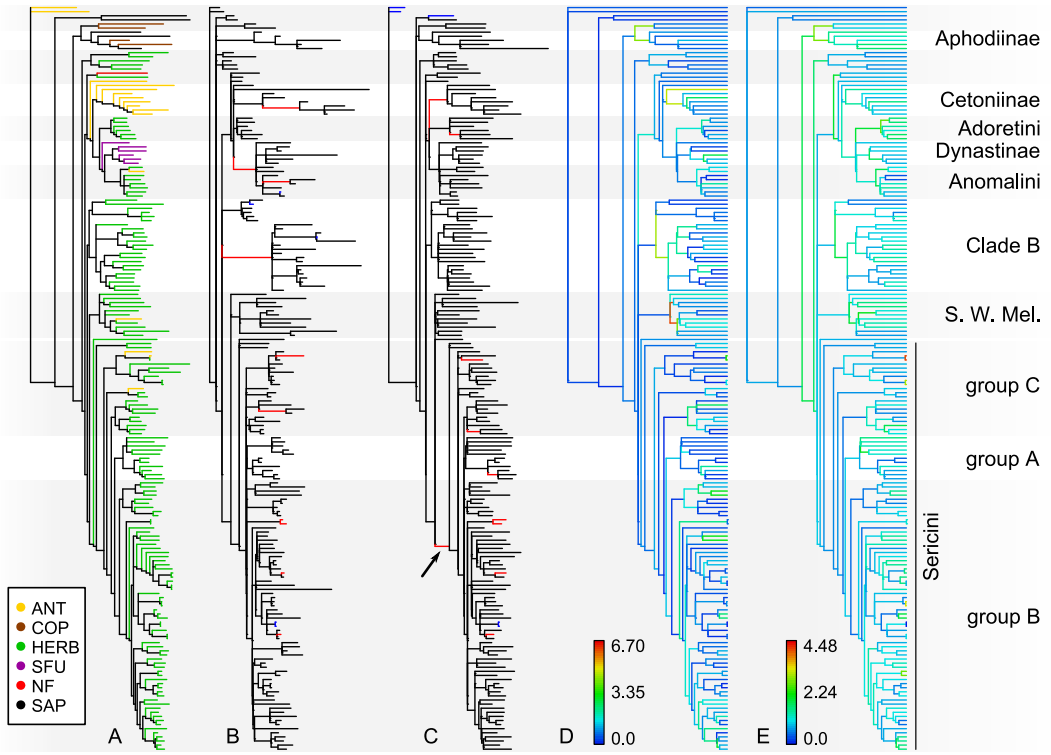

Figure V.6. Morphological divergence in multivariate space and rates of morphological divergence. (A) Molecular phylogenetic tree [31], trees with optimized branch lengths by (B) the uncorrected and (C) the size-corrected data set, and rates of morphological divergence (multivariate standardized phylogenetic independent contrasts) for (D) the uncorrected and (E) the size-corrected data set mapped on the ultrametric phylogenetic tree showing relative divergence times. The tips of the molecular tree (A) are color-coded for feeding habits (ANT = anthophilous, COP = coprophagous, HERB = herbivorous, SFU = sap / fluid utilizers, NF = not feeding, SAP = saprophagous). Branches in (B) and (C) with significantly lower (blue) and higher (red) morphological rates of evolution are colored respectively. Background shading indicates clade affiliation.

size-corrected and the uncorrected data set). Most lineages either diverged primarily in size (i.e. the uncorrected data set) or shape (i.e. in the size-corrected data). Cetoniinae were divergent from Adoretini + Anomalini + Dynastinae in shape (Fig. V.6C), whereas the clade of Anomalini + Dynastinae clearly showed a common divergence in size from Adoretini (Fig. V.6B). Clade B was found to have common (although little) divergence in shape, whereas two subordinate lineages strongly differed in size (Fig. V.6B, C). Rates of evolution were significantly different for shape and the molecular markers in the branch leading to the most recent common ancestor of Sericini. The genus *Trochalus* showed conspicuous divergence in shape within the Sericini group A (Fig. V.6C).

Phylogenetic independent contrasts revealed, in part, strong changes of rate of morphological evolution for both the uncorrected and the size-corrected data (i.e. shape, Fig. V.6D, E). Among larger clades, we found strong rate shifts among both data sets between Aphodiinae and Scarabaeinae, within the Southern World Melolonthinae, but also within Clade B (slightly more distinct pattern in uncorrected data, Fig. V.6D). For several clades with low contrasts for uncorrected data we found elevated contrasts for the size-corrected data. For the size-corrected data, deeper branches (e.g. between saprophagous and coprophagous Scarabaeidae and Pleurosticti) were found to exhibit major morphological change whereas the uncorrected data set revealed stronger change and frequently accelerated rates of morphological divergence on more recent time scales (Fig. V.6 B–E). Several closely related taxa (i.e. terminal species pairings), especially within the densely sampled Sericini, exhibit higher rates of divergent evolution.

4. Discussion

4.1. Morphospace divergence in the light of feeding habits

Because morphological traits are an important expression of the species niche, a partitioning of the niche as a consequence of interspecific competition between coexisting species may lead to divergence in ecomorphospace (Richman and Price 1992; Ricklefs and Miles 1994). A directed selective force on external morphology is likely to alter the rate of morphological evolution. Rates are then unlinked from the Brownian Motion model (Felsenstein 1985; Garland et al. 1992) which is assumed by our approach in the molecular phylogenetic framework based on 16S, 28S and *Cox1* (Fontaneto et al. 2007; Ahrens and Vogler 2008). While actual inter-specific competition is difficult to measure and needs to be explored at the assemblage level, competition in the past that led to a partitioning of the species niche or extinction of less competitive species ('ghost of competition past', (Connell 1980)) may be reflected in morphospace; phylogenetic lineages will differ in morphospace based on historical constraints of their common ancestors and because members of a lineage share a greater similarity in their morphology as a result of the lingering legacy of a common ancestor (Felsenstein 1985; Harvey and Pagel 1991; Richman and Price 1992; Polly 2001). Competition that led to niche partitioning in the ancestors of extant lineages is, to some degree, also visible at the phylogenetic level. Evolution of dung beetles, the nearest relatives of Pleurosticti, was markedly influenced by strong competition for their food resource: the resulting resource partitioning led to divergence in morphospace which presumably triggered the diversification of the dung beetles (Ricklefs and Miles 1994). To assume an analogous situation for pleurosticts would be hard to prove since competition between adults has not been shown yet in literature. We therefore assumed as null hypothesis the reverse: if no divergence in morphospace is found, it should be concluded that there is no resource competition among species of the same feeding

Table V.3. Correlation from Mantel-test between molecular and morphometric distance-matrices for specimens within one feeding type and the complete sampling. Sample size, coefficient of determination, and p-values are given.

	sample size	r	p
anthophilous	16	−0.01	0.51
coprophagous	4	−0.01	0.51
herbivorous	150	0.04	0.15
sap/fluid utilizers	6	−0.06	0.52
saprophagous	5	−0.61	0.97
complete sampling	182	0.11	0.01

type and that there is any other selective pressure. However, the opposite does not necessarily mean that competition is the cause of divergence.

For example, different feeding type assemblages of adult pleurosticts are generally spatially separated (e.g. flowers, leaves, wood), but different locomotion behavior may be required besides other adaptations. Body traits such as legs and parts of the flight apparatus are therefore also likely to cause divergence in morphospace, as found for Dynastinae and rose chafers.

Results indicate directed selection in pleurostict chafers that explains morphological expansion of feeding types. A vast portion of the wide overlap between the different groups of feeding behavior can be explained by convergence of feeding behavior for most of the feeding types (Fig. V.3G–I). In particular, repeated shifts from herbivory to anthophagy (e.g. in Sericini, Hopliini, Southern World Melolonthines; Fig. V.3), linked with the rise of the angiosperms, offered large amounts of new nutritious resources.

Also within one feeding type, rates of morphological evolution departed from Brownian motion. Besides multiple rate shifts (Fig. V.3D, E), we observed an indication of decoupling of morphological rates of evolution from Brownian motion in herbivores (Table V.3) that favored a hypothesis of directed selection on morphospace. The general difficulty to detect missing correlation of morphological and molecular rates could not be overcome even by better sampling, as results of well-sampled herbivore chafers show. Multiple shifts of rates of morphological evolution suggested that directed selection on morphospace took place within herbivores. Especially one trait, metacoxal length, caused distinct divergence of the herbivores morphospace (Tab. D1).

This would reject at least in part the idea that the vast Angiosperm feeding resource available to phytophagous scarabs would result in lacking divergence (i.e. stochastic morphological drift in morphospace dimensions).

4.2. Morphospace divergence of phylogenetic lineages

When discussing lineage divergence and all other topology related issues, uncertainty of the phylogenetic hypothesis is a major bias. Therefore, the majority of our analyses focused on well-supported clades, which were retrieved also in other studies (Hunt et al. 2007; Ahrens et al. 2011; Bocak et al. 2014). While major phylogenetic lineages of Scarabaeidae generally showed large overlap for principal components 1 and 2 in both data sets (Fig. V.3A, B), significant differentiation was found by MANOVA between a number of lineages (Table V.1). The high number of non-significant pairwise comparisons in the MANOVA is likely to be caused by the limited sampling of certain clades in the phylogenetic tree (Ahrens and Vogler 2008).

Intermediate positions in morphospace found for certain lineages were highly indicative of the role of shape in evolution. Representatives of both Ablaberini + Sericini and species of Clade B are lacking in southern world continents (particularly in Australia), and the latter have invaded the Australian region (likely late during Tertiary) being present there only with a few species. Obviously, morphospace of Southern World Melolonthinae (i.e. Australian, as in Ahrens and Vogler (2008) which included mainly Australian representatives) expanded due to the lack of these competitors. Wide overlap in morphospace with Clade B and Ablaberini + Sericini (Figs. V.3 A, B, V.3F, G) was the result. Accelerated rates of morphological divergence were observed with both data sets within this lineage. Early, fast divergence in body size of two major lineages preceded lower rates (Fig. V.6E), accompanied by medium to high rates of divergence in shape (Figure 6D). The increased rates of morphological divergence in Southern World Melolonthinae (Fig. V.6D, E) fit the scenario of rapid convergence in a framework of an 'adaptive' radiation in the Southern World, where occupation of ecological licenses may have been similar to Australian Marsupialia (Springer et al. 1997).

Rates of morphological divergence that were inferred from the densely sampled clade of Sericini might easily be influenced by a node-density effect (Hugall and Lee 2007), where lineages in less densely sampled clades appear to have lower rates of molecular evolution (Lanfear et al. 2010). As phylogenetic independent contrasts are standardized over branch lengths, the rates of morphological divergence (Fig. V.6D, E) are also affected and should be considered with care.

4.3. What shapes morphospace evolution?

Knowledge about the drivers of scarab shape divergence will greatly enhance our understanding of the evolutionary biology of this group of beetles. Whereas the analysis of the complete sampling allowed conclusions about general trends within Scarabaeidae, the investigation of subsets of the complete data revealed information about diverging traits between sister lineages. Although some measurements were likely to be correlated, as a whole they allowed differentiation between major lineages. Generally, conclusions that are drawn from the size-corrected data set are congruent with those from the uncorrected one, where size is contained. However, patterns of directed selection and rates of morphological divergence of uncorrected data showed a quite different and plausible signal from that of shape (Fig. V.6) indicating that body size itself has an important role in morphospace evolution.

Scarab shape morphospace (size-corrected data) was highly influenced by measurements of extremities and features linked with flight apparatus (EH; Fig. V.6). These traits are highly adaptive for burrowing and locomotion behavior and have undergone major changes in the evolution of pleurostict scarabs (Fig. D2). Shorter and stouter forelegs in Dynastinae are suitable for burrowing in soil and organic matter; dorsoventral flattening of the body in Cetoniinae could be connected with the particular hovering flight behavior of the group (in particular among Cetoniini). Cetoniini beetles are able to target flowers in flight and land on them with high precision, an essential adaptation to anthophily. This ability is linked with key innovations of the elytral articulation and a lateral concave sinuation of the elytra for flight (Ahrens 2006).

It is widely agreed that key innovations in phenotypic characters show evolutionary importance (Hunter 1998) and intensify diversification of a lineage (Levinton 1988; Heard and Hauser 1995). Metacoxal length (MCL) is conspicuously increased in Sericini, separating the mainly herbivorous lineage of Sericini from other herbivore scarabs (Fig. V.5). This lineage is significantly more speciose (ca 4000 species) compared to its presumed sister lineage, the Ablaberini (ca 200 species). Additionally, the Old World Sericini clade (ca 3800 species) is more speciose compared to the Neotropical clade (ca 200 species; here represented by *Astaena*). The influence of MCL on morphospace was conspicuous (Figs. V.4G, D2H, I, Tables D1 and D4), and the strong phylogenetic signal it exhibited, together with the decoupling of rates of morphological and molecular evolution, might possibly indicate an evolutionary shift and accompanying impact on morphospace evolution. Our results also showed that the K-value depends not only on the tree size (Blomberg et al. 2003), but also on sampling within the tree (Fig. D3). It might, therefore, be questionable how useful it is to investigate the phylogenetic signal in order to infer directed selection on a certain trait. Nevertheless, even with only three sampled Sericini species, the K-value of MCL was the second highest value and in conjunction with the reconstruction of ancestral trait measures (Fig. V.5 5), it suggests strongly driven selection in relation to other traits towards a stabilization of an increased MCL within the lineage (Blomberg et al. 2003).

The link between high phylogenetic signal, and directed selection, possibly in combination with morphological key innovations, is not always evident. However, MCL was the only trait for which we found, based on evidence from a previous study (Ahrens 2006), a trace of a physiologically linked counterpart. A subsequent maximum likelihood reconstruction of the size-corrected MCL (Fig. V.5A) on the tree revealed a strong shift of MCL's relative length (Fig. V.5A, left hand arrow) being linked with the secondary closure of the posterior opening of the metacoxal operculum, produced by the mesal metacoxal process and the posterior margin of the metacoxal plate ((Ahrens 2006), Fig. V.5B, C, arrows). The presence of a mesal metacoxal process that originated among pleurostict chafers (Ahrens 2006) allows a broader rotation of the hind limbs and a progressive enlargement of the MCL among the pleurosticts which could be explained with improved statics of the exoskeleton, in particular in context of the burrowing behavior. Evolutionary key innovations may

have strong diverging influence on lineages in multivariate morphospace because they can promote evolutionary change in other traits (Hunter 1998). The secondary closure of the posterior metacoxal opening produced by the extended mesal metacoxal process and the posterior margin of the metacoxal plate (Fig. V.5B, C, left side, (Ahrens 2006)) is very likely an evolutionary key innovation (Hunter 1998). This hypothesis is strongly supported by the subsequent increased rate of morphospace evolution (Fig. V.6C, arrow) and diversification. Its linkage with a substantial functional advantage enables Sericini to occupy new ecological space. Sericini species can burrow rapidly into sandy ground in case of danger by flapping their hind legs about 180°forwards and backwards (personal observation). The complete closure of the metacoxal ostium (Fig. V.5B,C; left hand arrow) enables the beetles to rotate the hind limb more anteriorly and, in combination with the increased metacoxal length, the locomotion statics of the body are improved for this burrowing behavior. A functional dependence of the metacoxal ostium and MCL is further supported as a reversal towards a slightly open metacoxal ostium, which occurred in the lineage of Hymenoplia + Paratriodonta ((Ahrens 2006); Fig. V.5A, right hand arrow). This is linked with a recurring slight reduction of MCL. Further functional consequences that increased locomotion statics might also be the observed reduction of sclerotization of the exoskeleton that presumably reduces body weight and possibly also physiological efficiency. Other morphological characters seem to support a hypothesis of this trend, such as the reduction of the elytral shelf in Sericini (Ahrens 2006). These hypotheses need further investigation and might be the subject of future research.

4.4. The influence of size correction

Two different methods for removing isometric size from the data, the Burnaby Back Projection Method (BBPM) and linear regression against a size metric, only minor differed in morphospace patterns (Figs. V.3,V.4) and comparisons of relative shape divergence (Figs. V.5,D5). The portion of total variation explained by the first two principal components was always slightly lower in the BBPM-data than in the data derived from linear regression. MANOVA on the BBPM-data recovered more significant pairwise lineage comparisons and mostly higher F-values (Tables V.1,V.2,D9).

However, significant results for the phylogenetic regression of feeding types in morphospace only resulted from the linear regression data (Table D12).

In the uncorrected data, divergent patterns of shape were less evident due to strong convergence of size and uneven distribution of variation (Table D7). In a few cases, size data improved the differentiation between groups or sister lineages (Tables V.2,V.3). Size correction appears to be valuable for inference of patterns of shape variation (Burnaby 1966). As our study case has shown, it is informative to include size data, particularly when inferring rates of shape evolution, because patterns may be revealed that are in the same way relevant to niche formation and that may explain morphospace evolution from another perspective.

5. Conclusion

Vast resources associated with angiosperm biomass seem to favor a hypothesis of reduced competition between adults. This is supported by the highly developed chemical communication of pleurosticts (Meinecke 1975; Leal 1998; Leal et al. 1998; Leal 1999) that is used for aggregation, host location, and/or mate finding. But directed selection within the feeding types and strong rate shifts for some lineages indicated the opposite, at least for parts of the pleurostict tree. Significant shape divergence found between major lineages, combined with a lack of strong differentiation among younger and more closely related lineages such as the Sericini subgroups, indicated that the interpretation of results for pleurostict morphospace evolution, triggered by driven selection and competition, needs to be addressed at different time scales. The trend of convergence of feeding habits in multiple lineages indicate an evolutionary tendency that might be interpreted as resource partitioning which not in all cases is necessarily linked with morphospace divergence (e.g. Sericini). Striking morphospace divergence between some sister lineages with divergent feeding habits reveals that at least in the past (at the origin of these lineages) strong directed selection on morphospace was also likely to be linked with resource partitioning although being catalyzed by other factors such as feeding related locomotion behavior. But the same is true for scarabaeine dung beetles (Inward et al. 2011). However, poor autecological knowledge of most pleurostict species and lacking community studies on assemblage level (competition acts only on individuals of all developmental stages in local assemblages) make it hard to investigate the linkage between divergence and competition in more detail. Therefore, further studies are needed to examine morphological divergence of pleurosticts and community composition at local scales to more rigorously investigate the question of competition. Conceivable hypothetical scenarios of competition might also include the issue of larval foraging and their competition for food and

space, such as for dung beetles (Finn and Gittings 2003), because the larvae occupy an environment (soil) that is much more reduced in its dimensionality.

6. Acknowledgements

We thank the two anonymous referees who helped to improve the final version of the manuscript, and Mary-Liz Jameson for helpful comments and improving the English of the manuscript.

References

Adams, D. C. and F. J. Rohlf (2000). Ecological character displacement in *Plethodon*: biomechanical differences found from a geometric morphometric study. *Proceedings of the National Academy of Sciences of the United States of America* 97, 4106–11.

Ahrens, D. (2006). The phylogeny of Sericini and their position within the Scarabaeidae based on morphological characters (Coleoptera: Scarabaeidae). *Systematic Entomology* 31, 113–144. DOI: 10.1111/j.1365-3113.2005.00307.x.

Ahrens, D. and I. Ribera (2009). Inferring speciation modes in a clade of Iberian chafers from rates of morphological evolution in different character systems. *BMC evolutionary biology* 9, 234. DOI: 10.1186/1471-2148-9-234.

Ahrens, D., M. Scott, and A. P. Vogler (2011). The phylogeny of monkey beetles based on mitochondrial and ribosomal RNA genes (Coleoptera: Scarabaeidae: Hopliini). *Molecular phylogenetics and evolution* 60, 408–415. DOI: 10.1016/j.ympev.2011.04.011.

Ahrens, D. and A. P. Vogler (2008). Towards the phylogeny of chafers (Sericini): analysis of alignment-variable sequences and the evolution of segment numbers in the antennal club. *Molecular phylogenetics and evolution* 47, 783–98. DOI: 10.1016/j.ympev.2008.02.010.

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology* 26, 32–46.

Balthasar, V. (1963). *Monographie der Scarabaeidae und Aphodiidae der palaearktischen und orientalischen Region. Coleoptera: Lamellicornia.* Band 1. Prag: Verlag der Tschechoslowakischen Akademie der Wissenschaften, p. 391.

Berner, D. (2011). Size correction in biology: how reliable are approaches based on (common) principal component analysis? *Oecologia* 166, 961–71. DOI: 10.1007/s00442-011-1934-z.

Blankers, T., D. C. Adams, and J. J. Wiens (2012). Ecological radiation with limited morphological diversification in salamanders. *Journal of evolutionary biology* 25, 634–46. DOI: 10.1111/j.1420-9101.2012.02458.x.

Blomberg, S. P., T. Garland, and A. R. Ives (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution; international journal of organic evolution* 57, 717–45.

Bocak, L., C. Barton, A. Crampton-Platt, D. Chesters, D. Ahrens, and A. P. Vogler (2014). Building the Coleoptera tree-of-life for >8000 species: composition of public DNA data and fit with Linnaean classification. *Systematic Entomology* 39, 97–110. DOI: 10.1111/syen.12037.

Britton, T., C. L. Anderson, D. Jacquet, S. Lundqvist, and K. Bremer (2007). Estimating divergence times in large phylogenetic trees. *Systematic Biology* 56, 741–752. DOI: 10.1080/10635150701613783.

Browne, J. and C. Scholtz (1998). Evolution of the scarab hind wing articulation and wing base: a contribution toward the phylogeny of the Scarabaeidae (Scarabaeoidea: Coleoptera). *Systematic Entomology* 23, 307–326.

Burnaby, T. P. (1966). Growth-invariant discriminant functions and generalized distances. *Biometrics* 22, 96–110.

Claude, J. (2008). *Morphometrics with R*. Ed. by R. Gentleman, K. Hornik, and G. Parmigiani. Springer Science+Business Media, LLC, p. 316.

Connell, J. H. (1980). Diversity and the Coevolution of Competitors, or the Ghost of Competition Past. *Oikos* 35, 131–138.

Cooper, N. and a. Purvis (2009). What factors shape rates of phenotypic evolution? A comparative study of cranial morphology of four mammalian clades. *Journal of Evolutionary Biology* 22, 1024–1035. DOI: 10.1111/j.1420-9101.2009.01714.x.

Douglas, M. E. and W. J. Matthews (1992). Does morphology predict ecology? Hypothesis testing within a freshwater stream fish assemblage. *Oikos*, 213–224.

Dray, S. and A. Dufour (2007). The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software* 22, 1–20.

Ehrlich, P. and P. Raven (1964). Butterflies and plants: A study in coevolution. *Evolution* 18, 586–608.

Erichson, W. (1847). *Naturgeschichte der Insecten Deutschlands. Erste Abtheilung. Coleoptera.* 3, Lfrg. 4. Berlin: Nicolaische Buchhandlung, Berlin, pp. 552–986.

Ezard, T. H. G., P. N. Pearson, and A. Purvis (2010). Algorithmic approaches to aid species' delimitation in multidimensional morphospace. *BMC evolutionary biology* 10, 175. DOI: 10.1186/1471-2148-10-175.

Farrell, B. (1998). "Inordinate Fondness" explained: Why are there so many beetles? *Science* 281, 555–559.

Felsenstein, J. (1985). Phylogenies and the comparative method. *The American Naturalist* 125, 1–15.

Ferrario, V. F., C. Sforza, J. H. Schmitz, A. J. Miani, and G. Taroni (1995). Fourier analysis of human soft tissue facial shape: sex differences in normal adults. *J. Anat.* 187, 593–602.

Finn, J. A. and T. Gittings (2003). A review of competition in north temperate dung beetle communities. *Ecological Entomology* 28, 1–13.

Fontaneto, D., E. a. Herniou, C. Boschetti, M. Caprioli, G. Melone, C. Ricci, and T. G. Barraclough (2007). Independently evolving species in asexual bdelloid rotifers. *PLoS biology* 5. Ed. by M. A. F. Noor, e87. DOI: 10.1371/journal.pbio.0050087.

Garland Jr., T. (1992). Rate tests for phenotypic evolution using phylogenetically independent contrasts. *American Naturalist* 140, 509–519.

Garland, T., P. Harvey, and A. Ives (1992). Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Systematic Biology* 41, 18–32.

Gauch Jr, H. G. (1982). Noise reduction by eigenvector ordinations. *Ecology* 63, 1643–1649.

Giller, P. S. and B. M. Doube (1994). Spatial and Temporal Co-Occurrence of Competitors in Southern African Dung Beetle Communities. *Journal of Animal Ecology* 63, 629–643.

Hammer, O., D. A. T. Harper, and P. D. Ryan (2001). Paleontological statistics software package for education and data analysis. *Paleontologica Electronica* 4, 1–9.

Hanski, I. and Y. Cambefort (1991). *Dung beetle ecology*. Princeton University Press.

Harvey, P. and M. Pagel (1991). *The comparative method in evolutionary biology*. Vol. 239. Oxford university press Oxford.

Heard, S. B. and D. L. Hauser (1995). Key innovations and their ecological mechanisms. *Historical Biology* 10, 151–173.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics* 6, 65–70.

Horgan, F. G. and R. C. Fuentes (2005). Asymmetrical competition between Neotropical dung beetles and its consequences for assemblage structure. *Ecological Entomology* 30, 182–193.

Hugall, A. and M. Lee (2007). The likelihood node density effect and consequences for evolutionary studies of molecular rates. *Evolution* 61, 2293–2307.

Hunt, T., J. Bergsten, Z. Levkanicova, A. Papadopoulou, O. S. John, R. Wild, P. M. Hammond, D. Ahrens, M. Balke, M. S. Caterino, J. Gomez-Zurita, I. Ribera, T. G. Barraclough, M. Bocakova, L. Bocak, A. P. Vogler, J. Gómez-Zurita, I. Ribera, T. G. Barraclough, M. Bocakova, L. Bocak, and A. P. Vogler (2007). A Comprehensive Phylogeny of Beetles Reveals the Evolutionary Origins of a Superradiation. *Science* 318, 1913–1916. DOI: 10.1126/science.1146954.

Hunter, J. P. (1998). Key innovations and the ecology of macroevolution. *Trends in ecology & evolution* 13, 31–6.

Inward, D. J. G., R. G. Davies, C. Pergande, A. J. Denham, and A. P. Vogler (2011). Local and regional ecological morphology of dung beetle assemblages across four biogeographic regions. *Journal of Biogeography* 38, 1668–1682. DOI: 10.1111/j.1365-2699.2011.02509.x.

Jermy, B. T. (1985). Is there competition between phytophagous insects? *Journal of Zoological Systematics and Evolutionary Research* 23, 275–285.

Jolicoeur, P. (1963). The multivariate generalization of the allometry equation. *Biometrics* 19, 497–499.

Jolicoeur, P. and J. E. Mosimann (1960). Size and shape variation in the painted turtle. A principal component analysis. *Growth* 24, 339–354.

Jolicoeur, P., P. Pirlot, G. Baron, and H. Stephan (1984). Brain structure and correlation patterns in insectivora, chiroptera, and primates. *Systematic Zoology* 33, 14–29.

Kaplan, I. and R. F. Denno (2007). Interspecific interactions in phytophagous insects revisited: a quantitative assessment of competition theory. *Ecology Letters* 10, 977–994.

Klingenberg, C. P. (1996). "Multivariate allometry." In: *Advances in Morphometrics*. Ed. by M. et al. New York: Plenum Press, pp. 23–49.

Lanfear, R., J. J. Welch, and L. Bromham (2010). Watching the clock: studying variation in rates of molecular evolution between species. *Trends in ecology & evolution* 25, 495–503.

Leal, W. S. (1998). Chemical ecology of phytophagous scarab beetles. *Annual review of entomology* 43, 39–61. DOI: 10.1146/annurev.ento.43.1.39.

Leal, W. (1999). "Mechanisms of chemical communication in scarab beetles." In: *Environmental entomlology: behaviour, physiology, and chemical ecology*. Ed. by T. Hidaka, Y. Matsumoto, K. Honda, H. Honda, and K. Tatsuki. Tokyo: University of Tokyo Press, pp. 464–478.

Leal, W., H. Wojtasek, and M. Miyazawa (1998). Pheromone-binding proteins of scarab beetles. *Annals of the New York Academy of Sciences* 855, 301–305.

Levinton, J. S. (1988). *Genetics, Paleontology, and Macroevolution*. 2nd ed. Cambridge University Press, Cambridge, UK.

Loughrin, J. H., D. A. Potter, and T. R. Hamilton-Kemp (1995). Volatile compounds induced by herbivory act as aggregation kairomones for the Japanese beetle (Popillia japonica Newman). *Journal of Chemical Ecology* 21, 1457–1467.

McLachlan, G. (2004). *Discriminant analysis and statistical pattern recognition*. John Wiley & Sons, p. 526.

McPeek, M. A. (1995a). Morphological evolution mediated by behavior in the damselflies of two communities. *Evolution* 49, 749–769.

— (1995b). Testing hypotheses about evolutionary change on single branches of a phylogeny using evolutionary contrasts. *American Naturalist* 145, 686–703.

McPeek, M. A., L. Shen, J. Z. Torrey, and H. Farid (2008). The Tempo and Mode of Three-Dimensional Morphological Evolution in Male Reproductive Structures. *The American Naturalist* 171, E158–E178. DOI: 10.1086/587076.

Meinecke, C. (1975). Riechsensillen und Systematik der Lamellicornia. *Zoomorphologie* 82, 1–42.

Melville, J., L. J. Harmon, and J. B. Losos (2006). Intercontinental community convergence of ecology and morphology in desert lizards. *Proceedings. Biological sciences / The Royal Society* 273, 557–563. DOI: 10.1098/rspb.2005.3328.

Mitter, C., B. Farrell, and D. J. Futuyma (1991). Phylogenetic Studies of Insect-Plant Interactions: Insights into the Genesis of Diverstiy. *Trends in ecology & evolution* 6, 290–293.

Oksanen, J., F. Blanchet, R. Kindt, P. Legendre, P. Minchin, R. O'Hara, G. Simpson, P. Solymos, H. Henry, M. Stevens, and H. Wagner (2013). *vegan: Community Ecology Package*.

Orme, D., R. Freckleton, G. Thomas, T. Petzoldt, S. Fritz, N. Isaac, and W. Pearse (2012). *caper: Comparative Analyses of Phylogenetics and Evolution in R*.

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature* 401, 877–84. DOI: 10.1038/44766.

Paradis, E., J. Claude, and K. Strimmer (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290.

Peck, S. B. and A. Forsyth (1982). Composition, structure, and competitive behaviour in a guild of Ecuadorian rain forest dung beetles (Coleoptera; Scarabaeidae). *Canadian Journal of Zoology* 60, 1624–1634.

Peres-Neto, P. R., D. a. Jackson, and K. M. Somers (2005). How many principal components? stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics & Data Analysis* 49, 974–997. DOI: 10.1016/j.csda.2004.06.015.

Polly, P. D. (2001). On morphological clocks and paleophylogeography: towards a timescale for *Sorex* hybrid zones. *Genetica* 112–113, 339–357.

Potter, D. A., J. H. Loughrin, W. J. Rowe, and T. R. Hamilton-Kemp (1996). Why do Japanese beetles defoliate trees from the top down? *Entomologica Experimentalis er Applicata* 80, 209–212.

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. http://www.r-project.org/.

Reist, J. D. (1985). An empirical evaluation of several univariate methods that adjust for size variation in morphometric data. *Canadian Journal of Zoology* 63, 1429–1439.

Revell, L. J. (2010). Phylogenetic signal and linear regression on species data. *Methods in Ecology and Evolution* 1, 319–329. DOI: 10.1111/j.2041-210X.2010.00044.x.

— (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3, 217–223. DOI: 10.1111/j.2041-210X.2011.00169.x.

Ribera, I., G. N. Foster, I. S. Downie, D. I. Mccracken, and V. J. Abernethy (1999). A comparative study of the morphology and life traits of Scottish ground beetles (Coleoptera, Carabidae). *Annales Zoologici Fennici* 36, 21–37.

Richman, A. D. and T. Price (1992). Evolution of ecological differences in the Old World leaf warblers. *Nature* 355, 817–821.

Ricklefs, R. E. and J. Travis (1980). A morphological approach to the study of avian community organization. *The Auk*, 321–338.

Ricklefs, R. E., D. Cochran, and E. R. Pianka (1981). A morphological analysis of the structure of communities of lizards in desert habitats. *Ecology* 62, 1474–1483.

Ricklefs, R. E. and D. B. Miles (1994). "Ecological and Evolutionary Interences from Morphology: An Ecological Perspective." In: *Ecological Morphology: Integrative Organismal Biology*. Ed. by P. C. Wainwright and S. M. Reilly. Chicago: University Of Chicago Press, pp. 13–41.

Ritcher, P. O. (1958). Biology of Scarabaeidae. *Annual review of Entomology* 3, 311–334.

Scholtz, C. and S. Chown (1995). "The evolution of habitat use and diet in the Scarabaeoidea: a phylogenetic approach." In: *Biology, Phylogeny, and Classification of Coleoptera: Papers Celebrating the 80th Birthday of Roy A. Crowson*. Ed. by J. Pakaluk and S. A. Ślipiński. 1st ed. Warszawa: Muzeum i Instytut Zoologii PAN Warszawa, pp. 355–374.

Scholtz, C. H. and V. V. Grebennikov (2005). "Scarabaeoidea Latreille, 1802." In: *Coleoptera, beetles: Morphology and systematics (Archostemata, Adephaga, Myxophaga, Polyphaga partim), Band 1*. Ed. by R. Beutel and R. A. B. Leschen. Walter de Gruyter, p. 567.

Smith, A. B. T. (2006). A Review of the Family-group Names for the Superfamily Scarabaeoidea (Coleoptera) with Corrections to Nomenclature and a Current Classification. *Coleopterists Society Monograph* 5, 144–204.

Springer, M., J. Kirsch, and J. Case (1997). "The chronicle of marsupial evolution." In: *Molecular Evolution and Adaptive Radiation*. Ed. by T. Givnish and K. Sytsma. Vol. 35. Cambridge University Press, pp. 129–162.

Teissier, G. (1960). "Relatice growth." In: *The physiology of crustacea I. Metabolism and growth*. Ed. by T. H. Waterman. New York: Academic Press, pp. 537–560.

Travis, J. and R. E. Ricklefs (1983). A morphological comparison of island and mainland assemblages of Neotropical birds. *Oikos*, 434–441.

Venables, W. N. and B. D. Ripley (2002). *Modern Applied Statistics with S*. 4th ed. New York: Springer.

Wainwright, P. C. and S. M. Reilly (1994). Chicago: University of Chicago Press, p. 367.

Zelditch, M., D. Swiderski, H. Sheets, and W. Fink (2004). *Geometric Morphometrics for Biologists. A Primer*. San Diego, CA: Elsevir Academic Press, p. 416.

Chapter VI.

Asymmetry in genitalia does not increase the rate of their evolution

This chapter is published in:
Eberle, J., W. Walbaum, R. C. M. Warnock, S. Fabrizi, D. Ahrens (2015). Asymmetry in genitalia does not increase the rate of their evolution. *Molecular Phylogenetics and Evolution* 93, 180–187. DOI: 10.1016/j.ympev.2015.08.005

Authors' contributions to the original article:
molecular lab work: JE, SF, DA; sequence assembly and alignments, phylogenetic inferences: JE; morphometric measurements: JE, WW; morphometric analyses: JE; manuscript design and writing: JE, DA.

1. Introduction

The diversity of shape in the copulation organs of insects is one of the most fascinating subjects in biology. Although great advances in elucidating the driving processes for this variety have been made, many mysteries remain. Copulation organs are a central focus for many evolutionary biologists and systematists (Arnqvist 1998), and the richness and complexity of their characters offer great utility in systematic studies ((Sharp and Muir 1912; Jeannel 1955; D'Hotman and Scholtz 1990; Scholtz 1990). Evolutionary biologists have argued that sexual selection, species isolation, and rapid evolution of the genitalia might be closely linked (Eberhard 1985; Arnqvist et al. 2000; Gage et al. 2002; Hosken and Stockley 2004; Simmons 2014). Asymmetry (i.e. directional asymmetry; (Palmer 1996)) is part of the complexity of copulation organs in otherwise nearly completely bilateral organisms, and it is known that asymmetric genitalia, in insects, have multiple origins (Huber et al. 2007).

Within the chafer beetle genus *Schizonycha* (Coleoptera: Scarabaeidae: Schizonychini) we find species with symmetric and asymmetric male genitalia, which makes them a good model for testing hypotheses regarding asymmetric/symmetric morphologies. *Schizonycha* comprises 370 species with most of them (349) occurring in the Afro-tropical region (Lacroix 2010). It thus includes 87% of all known species of the tribe Schizonychini. Adults of these small to medium sized chafers (7.5–23 mm) feed polyphagously on plants, while their larvae are soil dwellers and feed on humus or plant roots. While *Schizonycha* have mostly gained attention due to their appearance as crop pests (e.g. Pollard 1956; Kulkarni et al. 2007; Harrison and Wingfield 2016)), the taxonomy of the genus is still poorly understood (Pope 1960). The phylogenetic position of *Schizonycha* within Scarabaeidae is not yet fully resolved (Ahrens and Lago 2008) – its phylogenetic position varies among various tree reconstruction approaches, although it is recovered consistently within a monophyletic chafer clade

comprising Melolonthini, Enariini, and Rhizotrogini. Within these tribes species with asymmetric genitalia also occur, however, examples are rare.

Several hypotheses for the causes of asymmetric genitalia have been proposed (Huber et al. 2007; Schilthuizen 2007; Huber 2010; Schilthuizen 2013). However, sexual selection seems to be the most plausible explanation in the case of *Schizonycha*: morphological compensation for changes in mating position must be rejected, because all species mate in the same position (male above female). Spatial and ecological constraints are also unlikely, because only subtle morphological modifications occur between species, and the left and right sides retain similar functions (both parameres form a functional unit as clasping organ and do not perform different tasks) (Sharp and Muir 1912). However, the lacking knowledge of female genital structures in *Schizonycha* makes the distinction between sexual selection via cryptic female choice or via antagonistic coevolution difficult, if not impossible (Huber 2010; Schilthuizen 2013).

The role of asymmetry in the diversification of genital shape has so far not been addressed, and the investigation of this aspect might provide new insights into the mechanisms that determine genital shape (Palmer 1996; Ahrens and Lago 2008). In *Schizonycha*, as in all other scarab beetles, parameres function as clasping structures, which are inserted into the female genital tract and aid in mate recognition and copulation. The presence of complex microsensilla in female insect genitalia and high sensitivity to modifications of these (e.g. Acebes et al. 2003) raise the question of whether asymmetry, i.e. the dissimilar development of paired and previously symmetric components in the copulation organ, might influence mate recognition and the courtship process. Morphological divergence of genitalia between species is a prerequisite for the female's recognition of mating partners and therefore, also for sexual selection. In the case of asymmetric parameres, two structures coincidentally evolve dissimilar morphological variation in terms of size and shape. They might therefore in sum accumulate more morphological variation and hence diverge further and faster in morphospace compared to symmetric parameres, which evolve equally. Here, we examined the link between the occurrence and degree of asymmetry, and the evolutionary rate of divergence of paramere shape in *Schizonycha* in a molecular phylogenetic framework.

In particular, we investigated the evolutionary plasticity of genital asymmetry. We tested the hypothesis of multiple independent origins of asymmetric genitalia during the evolution of *Schizonycha* by comparing it with scenarios of a single origin of asymmetry or symmetry. To assess whether genital asymmetry increases the rate of morphological divergence, we investigated the morphospace occupancy of phylogenetic lineages with asymmetric and symmetric genitalia, and tested for correlation between the degree of asymmetry in ancestral nodes and the rate of divergent morphological evolution in the respective clade. In the case of asymmetric parameres as a driver of morphological divergence of overall paramere shape, we expect lineages with many asymmetric species to occupy a larger portion of the overall paramere morphospace than predominantly symmetric lineages.

2. Material and Methods

2.1. Taxon sampling, DNA extraction and DNA sequencing

A sample of 99 *Schizonycha* specimens comprising 34 morphospecies was collected from 30 sites in eastern South Africa (Table E1).

These taxa represent about a third of the South African fauna and 7.5% of the World fauna (Lacroix 2010). Five closely related outgroup taxa were included in the analyses, representing species of the South African genera *Asactopholis* Brenske, *Achloa* Erichson, *Hypopholis* Erichson and *Rhabdopholis* Burmeister (Ahrens et al., unpublished data). Female specimens and 10 males with badly preserved genitalia were excluded from the analysis, resulting in a total of 76 *Schizonycha* specimens. DNA was extracted with the Promega Wizard® SV96 Plate extraction kit using muscle tissue of one thoracic leg. After DNA extraction, genitalia of male specimens were extracted and dry mounted on the same pin. Diagnostic characters used to distinguish adult morphospecies were those used in traditional taxonomic studies of the group, including body size and shape, coloration, surface sculpture and pilosity, as well as male genital morphology (Pope 1960). Voucher specimens are deposited in the collections of the Zoological Research Museum A. Koenig Bonn (ZFMK).

Two mitochondrial and one nuclear gene region were amplified and sequenced for the analyses. Mitochondrial gene regions included cytochrome oxidase subunit 1 (*cox1*) and 16S ribosomal RNA (*rrnL*). PCR and sequencing was performed using primers Pat and Jerry (*cox1*), and 16Sar and 16SB2 (16S) (Simon et al. 1994). Nuclear 28S rRNA, containing the variable domains D3–D6, was amplified using primers FF and DD (Monaghan et al. 2007). Sequencing was performed on both strands using BigDye v. 2.1 and an ABI3730 automated sequencer. Sequences were edited

manually using BioEdit v7.0.9 (Hall 1999) and Geneious 5.3.4 (Drummond et al. 2010). GenBank accession numbers are provided in Electronic Supplement Table E1.

2.2. Alignments and phylogenetic analysis

DNA sequences of each marker were aligned independently with MAFFT (Katoh et al. 2002, 2005) and checked by eye. Maximum likelihood (ML) tree searches on the concatenated matrix were performed in PhyML v3.0 (Guindon and Gascuel 2003). Modeltest (Posada and Crandall 1998) was used to select the GTR+I+G model based on the AIC score (Akaike 1974). All parameters (base frequencies, proportion of variable sites and gamma distribution parameter) were estimated from the data. Bayesian analysis was conducted using MrBayes 3.2 (Huelsenbeck and Ronquist 2001; Ronquist et al. 2012), partitioning the data for *rrnL*, 28S, and the three codon positions of *cox1* (Nylander et al. 2004; Brandley et al. 2005). Substitution models were selected using PartitionFinder (v.1.1.1; (Lanfear et al. 2012); *cox1*, codon position (cp) 1: SYM+I+G; *cox1*, cp 2: HKY+I+g; *cox1*, cp 3: HKY+G; *rrnL*: GTR+I+G; 28S: K80+I). Standard deviation of split frequencies and visualization of the output in Tracer 1.6 (Rambaut and Drummond 2003) were used to assess stationarity and convergence of runs. Tree searches were conducted for 5×10^7 generations, using a random starting tree and two runs of three heated and one cold Markov chains (heating parameter $\lambda = 0.1$). Chains were sampled every 5,000 generations and 5×10^6 generations were discarded as burn-in, based on the average standard deviation of split frequencies, as well as the plots of $-\ln L$ against generation time. All trees were rooted with *Asactopholis* sp.

Two alternative tree hypotheses, with regard to the number of origins of symmetric versus asymmetric parameres, were tested by site bootstrapping as implemented in CONSEL (Shimodaira and Hasegawa 2001): a) all putative asymmetric species were constrained to be monophyletic, and b) all putative symmetric species were constrained to be monophyletic. Site bootstrapping identifies the top ranking topology for alternative tree hypotheses under the likelihood criterion and assesses support for each topology based on p-values calculated for the approximately unbiased test (AU), bootstrap probability tests (NP, BP and PP), the Shimodaira–Hasegawa test (SH), and the weighted Shimodaira–Hasegawa test (WSH). We used the default scaling fac-

tors of 0.5–1.4, with 10,000 pseudoreplicates. Individual site likelihoods used for the CONSEL analysis were calculated for alternative topological hypotheses (constrained and unconstrained) in PhyML.

2.3. Species tree and relative divergence times

A species tree was estimated with *BEAST 2.1.3 (Bouckaert and Heled 2014) using the same models and partitions implemented in MrBayes. Trees were unlinked across nuclear and mitochondrial markers, and independent lognormal relaxed clock models (Drummond et al. 2006) were applied to the 3 markers. Due to a complete lack of fossil calibration points, relative divergence times were estimated by fixing the mean clock rate of *rrnL* to 1, with the rates of *cox1* and 28S estimated relative to *rrnL*. We ran 2 independent analyses with 8×10^8 generations; parameters were sampled every 10,000 and trees every 20,000 generations. Convergence of parameters was assessed for each run in Tracer 1.6. Log files and species trees from independent runs were combined using LogCombiner 2.1.3, after removing a burnin of 10%. Trees were summarized using TreeAnnotator 2.1.2. The `.xml`-file for a single MCMC-run is provided as Electronic Supplement File E3.

2.4. Shape analysis

We captured the genital shape of the left parameres in dorsal view as a partial outline from images taken separately 5 times for each specimen using a Zeiss Discovery V20 stereomicroscope with 20× magnification. Outlines were digitized with tpsDIG v2.10 (Rohlf 2006) as an open curve of 100 semilandmarks. The shape of the right parameres was acquired in the same way from mirrored images in order to allow for a simultaneous analysis and comparison with the left parameres. Raw data is provided in Electronic Supplement file E4. Outlines were aligned with Generalized Procrustes analysis (GPA: Gower 1975; Rohlf and Slice 1990) as implemented in the R package *geomorph* (v2.1.2, Adams and Otarola-Castillo 2013). Semilandmarks 2–99 were allowed to slide along the outline curve under the Procrustes distance criterion. For subsequent analyses of asymmetry (see below), GPA was done twice, once with and

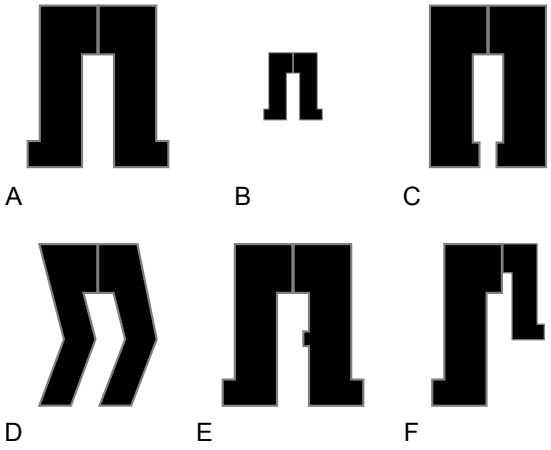

Figure VI.1 Simulated shapes used to infer the impact of scaling during GPA on asymmetry estimation: (A)–(C) perfectly symmetric shapes of different size, (D) strongly and (E) slightly asymmetric shapes, and (F) a shape which is only asymmetric due to size differences of the left and the right side.

once without scaling the paramere outlines during superimposition. The latter was achieved by multiplying each shape coordinate by its centroid size and dividing it by the mean shape's centroid size of all included samples. A principal components analysis of shape variation was performed and informative components summing up to 95% of total variation were retained.

2.5. Quantifying asymmetry

During Procrustes superimposition, shapes are centered, rotated, and scaled (Zelditch et al. 2004). To assess the influence of scaling on the quantification of asymmetry, we performed (i) full (with scaling) and (ii) partial (without scaling, see paragraph above) Procrustes superimposition on artificial paramere shapes (notation follows Zelditch et al. 2004). Artificial parameres included perfectly symmetric shapes of different size, strongly and slightly asymmetric shapes (Fig. VI.1A–E), and a case in which parameres were asymmetric only due to differences in size (Fig. VI.1F). The degree of asymmetry of each artificial shape from both full and partial Procrustes superimposition (i.e. the divergence between the superimposed left and right parameres) was determined using Euclidean distances of the informative principal components between the left and right parameres of the same shape. The same procedure was applied to the biological samples, using partial Procrustes superimposition. Subjective

errors from imaging of the biological samples were minimized by selecting the repeat for each specimen, which resulted in the smallest asymmetry values per specimen, for further analyses.

2.6. Integrative analysis of shape evolution

In order to infer the evolutionary history of asymmetric parameres in *Schizonycha*, we reconstructed the absolute degree of asymmetry between the parameres in the ancestral nodes of the MrBayes tree using maximum likelihood reconstruction in *ace* (R package ape 3.2, Paradis et al. 2004; Paradis 2012). In addition, the number of reversals between asymmetry and symmetry of the parameres was independently inferred by reconstructing asymmetry as a discrete binary character at the ancestral nodes of the species tree, also using maximum likelihood reconstruction in *ace*. For this purpose, we coded asymmetric parameres as those recognized as such by eye.

We subsequently assessed the influence of the degree of asymmetry on the overall rate of shape divergence of the parameres. Overall phenotypic variance of the parameres was summarized by a principal component analysis on the covariance matrix of the combined informative principal components of the left and right parameres from the full Procrustes superimposition. We applied a two-fold approach for the integrative analysis of asymmetry and morphological divergence: (i) The Bayesian consensus tree was projected onto the resulting first two principal component scores plot of overall variation, using the *phylomorphospace*-function in the R package phytools (Sidlauskas 2008; Revell 2012). The shape divergence of asymmetric specimens relative to symmetric individuals was inspected visually in the *phylomorphospace*. (ii) To evaluate the dependence of evolutionary rates on paramere asymmetry, we calculated the multivariate standardized phylogenetic independent contrasts (MSPIC) for overall phenotypic variance of the parameres, following the method of McPeek et al. (2008). See Electronic Supplement E5 or Appendix D2 for the R code). The MSPIC can be interpreted as the overall evolutionary rates of divergent evolution for the respective clade (Garland Jr. 1992; McPeek 1995a,b). Finally, we performed a linear regression on the reconstructed degree of ancestral asymmetry for each node in the species tree (independent variable) and the resulting MSPIC (dependent variable).

The model was validated by testing the residuals for being normally distributed using the Kolmogorov-Smirnov test (R package *nortest* v1.0-3).

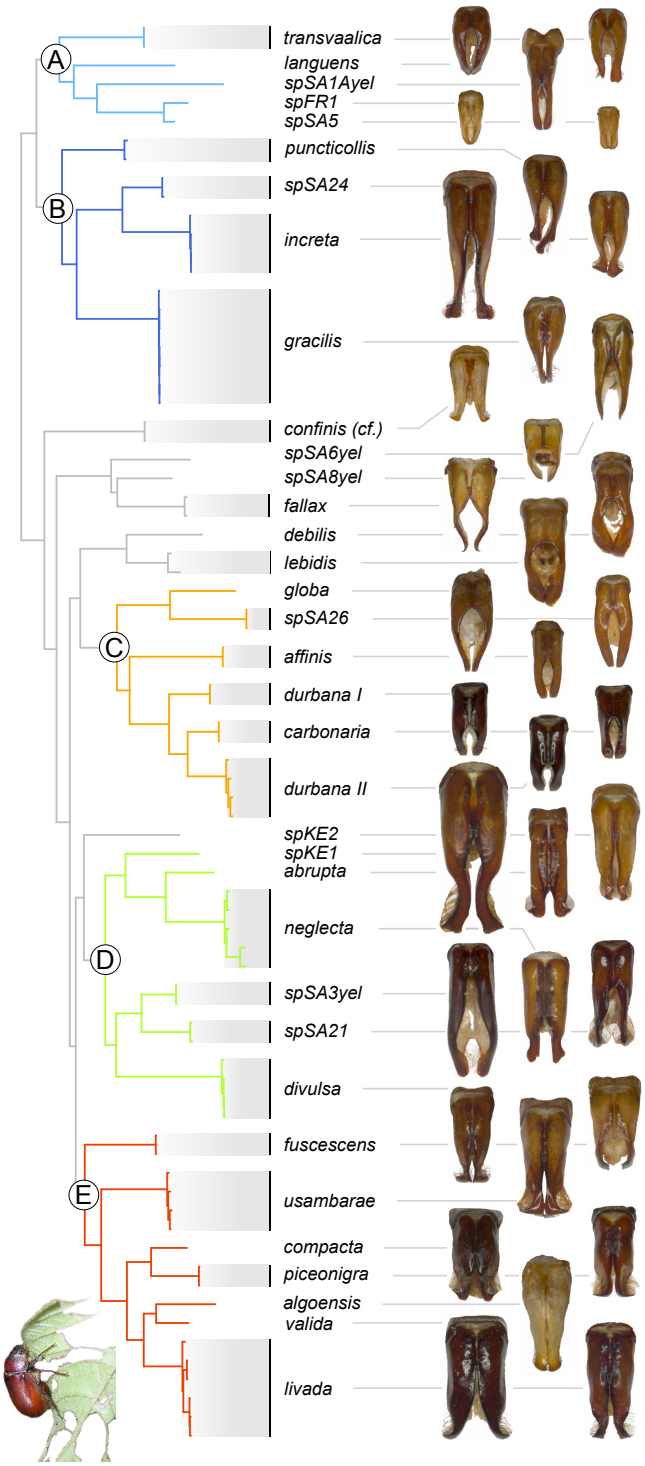

Figure VI.2 Consensus tree from Bayesian tree inference. The main clades that were found throughout all analyses are colored respectively. Exemplary images of the parameres are provided for each species.

3. Results

3.1. Schizonycha phylogeny

The sequence alignment resulted in an 826 bp fragment of *cox1*, while the *rrnL* and 28S fragments had a length of 563 bp and 658 bp, respectively. The combined data matrix included 2047 bp, of which 522 characters were parsimony informative. *Schizonycha* was monophyletic under both tree searchers with high support (Maximum likelihood, aRLT = 97; Bayesian posterior probability: pp = 1.0). In both trees we recovered five major clades A, B, C, D, and E, with high branch support in both analyses (Fig. VI.2; E1). Within these clades the species topologies were identical between the two trees. However, inter-clade relationships and the positions varied in species that were not in the main clades.

3.2. Quantification of asymmetry

Analysis of simulated shapes showed that scaling during Procrustes superimposition distorted the calculation of continuous asymmetry values, if asymmetry was partly caused by size differences between the parameres (Fig. VI.3). Therefore, partial Procrustes superimposition (without scaling) was used to quantify asymmetry. Four principal components were retained (Table VI.1) for the estimation of asymmetry values. Euclidean distances as a measure of asymmetry were consistent with asymmetry identified by eye. The highest values for between paramere distance were assigned to 8 species with putatively asymmetric parameres: *S. transvaalica*, *S. puncticollis*, *S. increta*, *S. gracilis*, *S. debilis*, *S. lebidis*, *spSA21*, and *spSA8yel* (Fig. VI.4C). However, we found a continuous transition in the degree of asymmetry among species.

Table VI.1. Variance explained by principal components from analysis on superimposed paramere outlines from Generalized Procrustes Analysis with and without scaling (Var. = Total Variance in percent; Cum. = Total cumulative variance in percent).

	Original size		Scaled	
	Var.	Cum. var.	Var.	Cum. var.
PC1	87.88	87.88	31.70	31.70
PC2	4.76	92.63	19.74	51.43
PC3	1.63	94.26	13.88	65.31
PC4	1.43	95.69	7.82	73.13
PC5	1.08	96.76	4.87	78.00
PC6	0.55	97.32	3.53	81.52
PC7	0.44	97.75	2.42	83.94
PC8	0.29	98.04	1.95	85.89
PC9	0.23	98.27	1.44	87.33
PC10	0.17	98.44	1.31	88.64
PC11	0.16	98.59	1.01	89.66
PC12	0.13	98.72	0.87	90.52
PC13	0.11	98.83	0.74	91.26
PC14	0.09	98.93	0.65	91.91
PC15	0.09	99.01	0.59	92.49
PC16	0.07	99.09	0.55	93.04
PC17	0.07	99.16	0.50	93.54
PC18	0.06	99.22	0.42	93.96
PC19	0.05	99.27	0.38	94.35
PC20	0.05	99.32	0.37	94.72

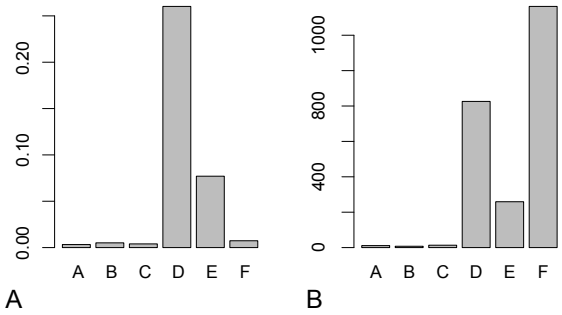

Figure VI.3 Barplots illustrating the estimated proportional degree of asymmetry of simulated shapes (corresponding Fig. VI.1) using (A) full (with scaling) and (B) partial (without) Procrustes superimposition. The strong influence of scaling in evident from shape F.

Based on ancestral state reconstruction, when asymmetry was coded as a discrete character state, we found five separate origins of paramere asymmetry in clades A, B, D, and in sister lineages of clades C and E. Reversals of asymmetry were found in clade B (S. *spSA24*) (Fig. E2). A trend or directionality of the evolution of genital asymmetry was not evident in *Schizonycha*. The site bootstrapping analysis performed in CONSEL supported these findings by significantly favoring the unconstrained analysis with multiple origins of asymmetry, against a single origin of asymmetry or symmetry (Table VI.2).

Full Procrustes superimposition of paramere outlines resulted in 20 informative principal components (Table VI.1). The projection of the Bayesian tree onto the morphospace of the combined left and right paramere variation (Fig. VI.5) suggested no obvious link between the absolute degree of asymmetry of the parameres and clade divergence. For example, strongly asymmetric species were observed close to symmetric species within the same clade. The highest degree of asymmetry was found in Clade B, which showed no exceptional divergence within the morphospace.

Similar evidence was obtained from the multivariate standardized phylogenetic independent contrasts, which can be interpreted as the overall divergence rates of shape for the respective clades (Garland Jr. 1992; McPeek 1995a,b) (Fig. VI.4B). Rates of overall divergence for all clades ranged from 3.95 to 16.77. The highest rate of divergence (14.79) among the main clades was found for Clade E, and can be attributed to *S. fuscescens*, the only Asian species in the study. Clade A that comprised the largest number of asymmetric species showed intermediate rates across the clade (max. 12.01), whereas all other clades exhibited highly increased rates of divergence for at least one lineage. A correlation between the degree of asymmetry and the

Table VI.2. Site bootstrapping test performed in CONSEL, identifying the multiple reversals between asymmetric and symmetric genitalia as most likely (scenarios: unconstrained PhyML analysis; AM: apparently asymmetric genitalia constrained to be monophyletic; SM: apparently symmetric genitalia constrained to be monophyletic). Columns show the observed log-likelihood difference of trees (obs), the results from approximately unbiased tests (au); bootstrap probability of item/hypothesis (np); non-scaled bootstrap probability (bp) and the Bayesian posterior probability (pp) calculated by the Bayesian Information Criterion approximation, the (weighted) Kishino-Hasegawa test (kh and wkh) and the (weighted) Shimodaira–Hasegawa test (sh and wsh).

rank	scenario	obs	au	np	bp	pp	kh	sh	wkh	wsh
1	unconstrained	-58.8	0.999	0.999	1.000	1.000	1.000	1.000	1.000	1.000
2	AM	216.6	2.00E-13	2.00E-08	0	8.00E-95	0	0	0	0
3	SM	233.2	6.00E-62	8.00E-18	0	5.00E-102	0	0	0	0

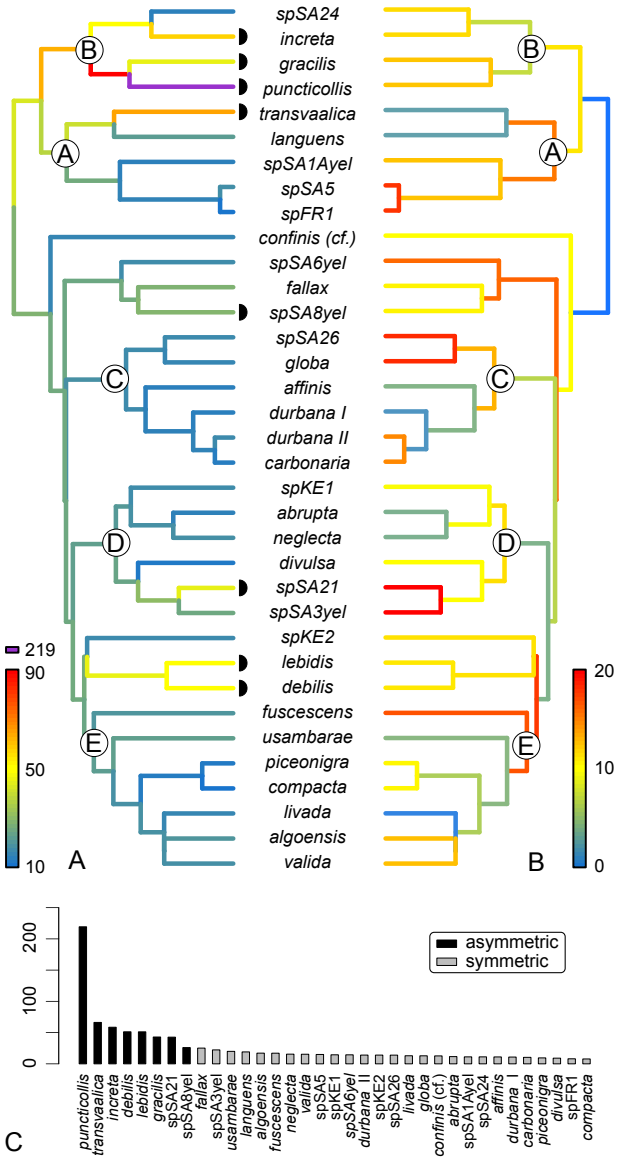

Figure VI.4 Juxtaposition of the species tree from the *BEAST analyses with (A) the quantitative degree of paramere asymmetry (red=high; blue=low) at tips and ancestral nodes and (B) the overall rate (red=high; blue=low) of morphological divergence of parameres of each clade projected onto it. Apparent asymmetric species (bye-eye inspection) are marked with a black semicircle. (C) Species mean asymmetry values sorted decreasingly. Apparent asymmetric species are highlighted black.

divergence rates of the respective clades was not evident (r^2: -0.03, p-value: 0.97, Fig. VI.6). Residuals were normally distributed according the Kolmogorov-Smirnov test (p-value = 0.39).

4. Discussion

4.1. Origins of asymmetry

The five independent origins of putative asymmetric parameres found within *Schizonycha* demonstrate a high degree of evolutionary plasticity of asymmetry, and this is further demonstrated by the instance of inferred reversal from asymmetry to symmetry (*spSA24*, Fig. VI.4A, Fig. E2). The wide distribution of *Schizonycha*, the large number of species, and the large number of undescribed taxa make it unlikely that complete sampling can be achieved in the near future, and thus we have to consider these results preliminary. However, the hypothesis of multiple independent origins of asymmetry seems favorable despite limited sampling of *Schizonycha* chafers, representing only 7.5% of the known species, since a single origin of asymmetry or symmetry was inferred to be significantly less likely (based on the CONSEL analyses). Directional asymmetry of genital structures has evolved independently many times within Insects and Coleoptera (Huber et al. 2007; Ahrens and Lago 2008). Phylogenetic patterns of the occurrence of asymmetry differ even among closely related taxa – for example, whereas the genitalia of *Serica* or *Peltonotus* (Scarabaeidae; Ahrens 2005; Huber et al. 2007; Breeschoten et al. 2013) are asymmetric throughout the genus, asymmetry has developed repeatedly in others (e.g., *Cyclocephala*; Breeschoten et al. 2013), as was demonstrated here for *Schizonycha*.

4.2. Symmetry–asymmetry transition and measurement errors

While subjective inspection by-eye may imply an obvious distinction between species with asymmetric versus symmetric parameres (Fig. E2), our quantitative morpho-

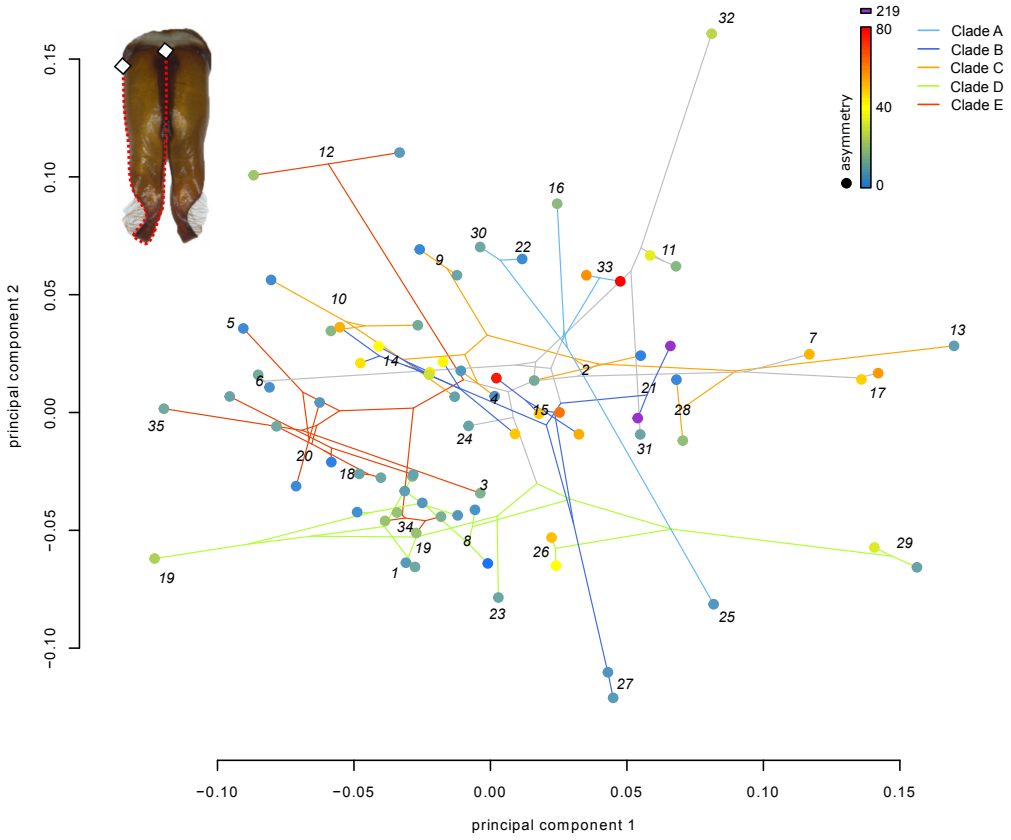

Figure VI.5. Phylomorphospace projection of paramere shape for all specimens. The dots are colored according to the degree of asymmetry; the branch colors indicate clade affiliation. The numbers encode the following *Schizonycha*species: 1 *abrupta*, 2 *affinis*, 3 *algoensis*, 4 *carbonaria*, 5 *compacta*, 6 *confinis* (cf.), 7 *debilis*, 8 *divulsa*, 9 *durbana I*, 10 *durbana II*, 11 *fallax*, 12 *fuscescens*, 13 *globa*, 14 *gracilis*, 15 *increta*, 16 *languens*, 17 *lebidis*, 18 *livada*, 19 *neglecta*, 20 *piceonigra*, 21 *puncticollis*, 22 *spFR1*, 23 *spKE1*, 24 *spKE2*, 25 *spSA1Ayel*, 26 *spSA21*, 27 *spSA24*, 28 *spSA26*, 29 *spSA3yel*, 30 *spSA5*, 31 *spSA6yel*, 32 *spSA8yel*, 33 *transvaalica*, 34 *usambarae*, 35 *valida*. The inset on the left shows an exemplar outline of a paramere, digitized in tpsDig.

metric analysis revealed a continuous transition in the degree of asymmetry between symmetric and asymmetric species (Fig. VI.4C). Measurements of species with putatively symmetric parameres also deviated from perfect symmetry (i.e. zero distance between the left and right parameres), which is not unexpected. The phenomenon of fluctuating asymmetry (Van Valen 1962), representing the inability of an individual to develop perfectly symmetric organs due to impairing factors, may obscure a distinct gap between symmetric and asymmetric species. However, fluctuating asymmetry is non- directional and its influence on the value of species' asymmetry, averaged over several specimens, should be minimized – at least for well-sampled species. Since sexual selection seems to discriminate against imperfectly symmetric males in vertebrates and invertebrates (Møller and Pomiankowski 1993; Watson and Thornhill 1994; Møller and Swaddle 1997; Huber et al. 2007), this raises the question of how the adaptive valley between perfectly symmetric and strongly asymmetric species is bypassed (Otronen 1998; Huber et al. 2007; Schilthuizen 2013). Cases of slight paramere asymmetry found in *Schizonycha* might represent an intermediate stage between symmetry and the establishment of strong directed asymmetry, which may have arisen despite sexual discrimination against asymmetry (Huber et al. 2007). However, other factors that may obscure the inference of symmetry-asymmetry transitions include digitalization errors, e.g. cases where parameres are not fully in plain when photographed. Erroneous rotations of the genitalia along the median axis can cause symmetric objects to appear asymmetric. We accounted for this phenomenon by capturing each specimen 5 times, and selecting the replicate that resulted in the lowest asymmetry value for the specimen for further processing. Accidental rotation can make objects appear more asymmetric than they really are, but never more symmetric. Therefore, the lowest asymmetry values are most likely to approximate the true value.

4.3. Asymmetry and Rates of Divergence

Asymmetry of the parameres implies non-identical trait evolution of the right and left side. Under these circumstances, morphological divergence of the left and right parameres may accumulate in different amounts, directions (vectors), and dimensions (number of axes) of the described morphospace. This has the potential to increase the amount of interspecific divergence of asymmetric species, relative to symmetric

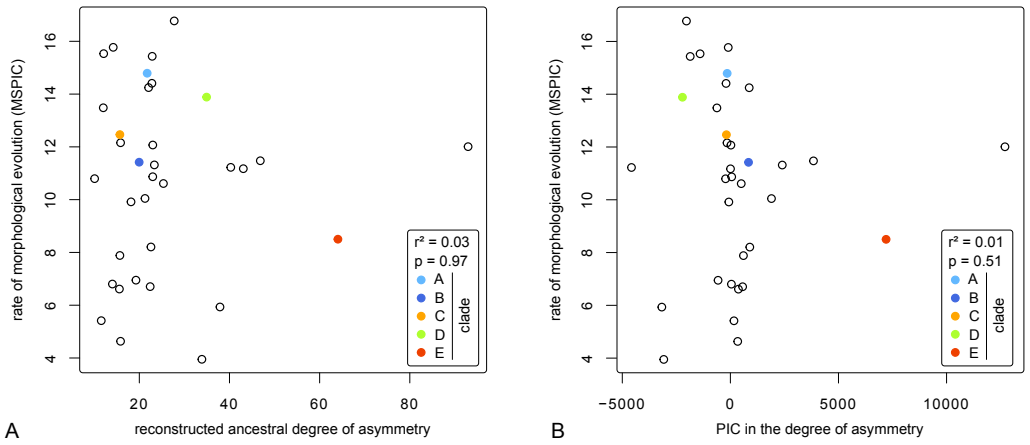

Figure VI.6. Regression analysis plot of the reconstructed degree of paramere asymmetry at the ancestral nodes and the overall rate of divergent morphological evolution in the respective clade. The values of the main clades' ancestral nodes (A-E) are color-coded. The coefficient of determination (r^2) and the significance-value (p) of the regression analysis are given in the bottom-right corner.

species. While closely related species are expected to appear close to each other in the morphospace plots, due to common ancestry (Richman and Price 1992), deviations may occur under the influence of strongly divergent selection. If asymmetry is responsible for an increase in the divergence among species, a higher distance is expected between asymmetric versus symmetric species. However, in the case of *Schizonycha*, increased paramere asymmetry did not result in a higher divergence between species' paramere shape (Fig. VI.5). The morphological variation in overall paramere shape was higher (or equal) than the gain in variation that was induced by asymmetry. Furthermore, divergent evolution of the left and right side did not lead to an increase in the rate of divergence among species (Fig. VI.6). Random divergent evolution has been proposed as explanation for the high variation in genitals on the ground that they are internal structures and not opposed to natural selection (Mayr 1963). However, this is questionable since the intromittent external secondary copulatory organs of odonates underlie equally high variation (Lloyd 1979; Waage 1979; Simmons 2014). Rates of morphological divergence depend on the density of sampling of phylogenetic lineages, due to the node density effect (Hugall and Lee 2007) . Rates of molecular evolution will appear higher in more densely sampled lineages (Lanfear et al.

2010). Phylogenetic independent contrasts, being standardized over branch lengths and rates of morphological divergence (Fig. VI.4B) can also be affected. Therefore, it will be useful to extend the analyses presented here to a more complete sample of *Schizonycha* (if ever achievable), which in the case of our study was not possible due to reasonable time and funding constraints.

5. Conclusions

The frequent occurrence of asymmetry in the parameres and its parallel evolution suggests selection in favor of a transition from symmetry to asymmetry (Stern 2013) and, given the reversal in *spSA24*, potentially vice versa. As previously mentioned, a hypothetical functional advantage of paramere asymmetry is not linked with the copulation position. The lack of correlation between the degree of asymmetry of the parameres and rates of morphological divergence also suggests that asymmetry is not causally linked to intensified genital divergence in *Schizonycha* (Fig. VI.6). Therefore it remains questionable whether paramere asymmetry is linked to any functional advantage. The high frequency of changes in symmetry in *Schizonycha* suggests that symmetry can alter rapidly. Whether subtle changes in genital symmetry are more easily perceived by females of these beetles than symmetric changes, and consequently have a higher impact under sexual selection by cryptic female choice (Arnqvist 1998; Hosken and Stockley 2004), is hard to say. The situation is even more complex however, as illustrated by the many cases of apparent symmetric genitalia in combination with asymmetric internal structures (e.g. folded endophallus with asymmetric spines; Breeschoten et al. 2013). This underlines the highly complex evolution of copulation organs (Eberhard 1985; Simmons 2001; Hosken and Stockley 2004; Simmons et al. 2009; Kamimura and Iwase 2010), even more so, when asymmetry is involved. To complete the interpretation of the observed evolutionary patterns in morphology detailed knowledge on copulation, courtship, speciation mechanisms and infraspecific variation of the internal copulation organs is required, but this data is yet lacking for the group studied here. Nevertheless, the frequent evolution of genital asymmetry brought out a variety of distinct shapes in this exceptionally species-rich genus, which makes *Schizonycha* a well-suited model system for investigating the evolutionary mechanisms that lead to genital asymmetry, and the impact of genital asymmetry on species evolution.

6. Acknowledgements

We are grateful to J. Harrison, R. Müller, and M. Snížek, for their help with collecting specimens, H. J. Krammer for his assistance in the course of this project, and M. A. McPeek for kindly providing his java program for the calculation of multivariate phylogenetic independent contrasts. For providing us with research and collection permits, we thank the various South African governmental institutions and departments in Eastern Cape (Permit Nos: WRO 122/07WR and WRO123/07WR), Gauteng (Permit No.: CPF6 1281), Limpopo (Permit No.: CPM-006-00001), Mpumalangma (Permit No.: MPN-2009-11-20-1232), and Kwazulu-Natal (Permit Nos: OP3752/2009, 1272/2007, 3620/2006). We thank M. Schilthuizen and the anonymous referees for their helpful comments on the earlier versions of the manuscript. The study was supported by the German Science Foundation (DFG; AH175/1 and AH175/3).

References

Acebes, A., M. Cobb, and J.-F. Ferveur (2003). Species-specific effects of single sensillum ablation on mating position in Drosophila. *Journal of Experimental Biology* 206, 3095–3100.

Adams, D. C. and E. Otarola-Castillo (2013). geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution* 4, 393–399.

Ahrens, D. and P. K. Lago (2008). Directional asymmetry reversal of male copulatory organs in chafer beetles (Coleoptera: Scarabaeidae): implications on left–right polarity determination in insect terminalia. *Journal of Zoological Systematics and Evolutionary Research* 46, 110–117. DOI: 10.1111/j.1439-0469.2007.00449.x.

Ahrens, D. (2005). *A taxonomic review on the Serica (s. str.) MacLeay, 1819 species of Asian mainland (Coleoptera, Scarabaeidae, Sericini)*. 1st ed. Keltern: Goecke & Evers, p. 163.

Akaike, H. (1974). 'A new look at the statistical model identification'. *IEEE Transactionson Automatic Control* 19, 716–723.

Arnqvist, G., M. Edvardsson, U. Friberg, and T. Nilsson (2000). Sexual conflict promotes speciation in insects. *Proceedings of the National Academy of Sciences of the United States of America* 97, 10460–4.

Arnqvist, G. (1998). Comparative evidence for the evolution of genitalia by sexual selection. *Nature* 393, 784–786.

Bouckaert, R. and J. Heled (2014). *DensiTree 2: Seeing Trees Through the Forest*. Tech. rep., http://biorxiv.org/lookup/doi/10.1101/012401. DOI: 10.1101/012401.

Brandley, M. C., A. Schmitz, and T. W. Reeder (2005). Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. *Systematic biology* 54, 373–90. DOI: 10.1080/10635150590946808.

Breeschoten, T., D. R. Clark, and M. Schilthuizen (2013). Evolutionary patterns of asymmetric genitalia in the beetle tribe Cyclocephalini (Coleoptera: Scarabaeidae: Dynastinae). *Contributions to Zoology* 82, 95–106.

D'Hotman, D. and C. Scholtz (1990). Phylogenetic significance of the structure of the external male genitalia in the Scarabaeoidea (Coleoptera). *Entomology Memoir Department of Agricultural Development* 77, 1–51.

Drummond, A., B. Ashton, M. Cheung, J. Heled, M. Kearse, R. Moir, S. Stones-Havas, T. Thierer, and A. Wilson (2010). *Geneious 5.3*.

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut (2006). Relaxed Phylogenetics and Dating with Confidence. *PLoS Biology* 4. Ed. by D. Penny, e88. DOI: 10.1371/journal.pbio.0040088.

Eberhard, W. G. (1985). *Sexual selection and animal genitalia*. Cambridge: Harvard University Press.

Gage, M. J. G., G. a. Parker, S. Nylin, and C. Wiklund (2002). Sexual selection and speciation in mammals, butterflies and spiders. *Proceedings. Biological sciences / The Royal Society* 269, 2309–16. DOI: 10.1098/rspb.2002.2154.

Garland Jr., T. (1992). Rate tests for phenotypic evolution using phylogenetically independent contrasts. *American Naturalist* 140, 509–519.

Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika* 40, 33–51. DOI: 10.1007/BF02291478.

Guindon, S. and O. Gascuel (2003). A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Systematic Biology* 52, pp. 696–704.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41, 95–98.

Harrison, J. G. and M. Wingfield (2016). A taxonomic review of white grubs and leaf chafers (Coleoptera: Scarabaeidae: Melolonthinae) recorded from forestry and agricultural crops in Sub-Saharan Africa. *Bulletin of Entomological Research* 106, 141–153. DOI: 10.1017/S0007485315000565.

Hosken, D. J. and P. Stockley (2004). Sexual selection and genital evolution. *Trends in ecology & evolution* 19, 87–93. DOI: 10.1016/j.tree.2003.11.012.

Huber, B. a. (2010). Mating positions and the evolution of asymmetric insect genitalia. *Genetica* 138, 19–25. DOI: 10.1007/s10709-008-9339-6.

Huber, B. A., B. J. Sinclair, and M. Schmitt (2007). The evolution of asymmetric genitalia in spiders and insects. *Biological reviews of the Cambridge Philosophical Society* 82, 647–98. DOI: 10.1111/j.1469-185X.2007.00029.x.

Huelsenbeck, J. P. and F. Ronquist (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics (Oxford, England)* 17, 754–5.

Hugall, A. and M. Lee (2007). The likelihood node density effect and consequences for evolutionary studies of molecular rates. *Evolution* 61, 2293–2307.

Jeannel, R. G. (1955). L'édéage: initiation aux recherches sur la systématique des coléoptères. *Publications du Muséum d'Histoire Naturelle Paris* 16, 1–155.

Kamimura, Y. and R. Iwase (2010). Evolutionary genetics of genital size and lateral asymmetry in the earwig *Euborellia plebeja* (Dermaptera: Anisolabididae). *Biological Journal of the Linnean Society* 101, 103–112.

Katoh, K., K.-i. Kuma, H. Toh, and T. Miyata (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic acids research* 33, 511–8. DOI: 10.1093/nar/gki198.

Katoh, K., K. Misawa, K.-i. Kuma, and T. Miyata (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research* 30, 3059–66.

Kulkarni, N., K. Chandra, P. N. Wagh, K. C. Joshi, and R. B. Singh (2007). Incidence and management of white grub, Schizonycha ruficollis on seedlings of teak (*Tectona grandis* Linn. f.) *Insect Science* 14, 411–418. DOI: 10.1111/j.1744-7917.2007.00168.x.

Lacroix, M. (2010). *Melolonthinae afrotropicaux (Scarabaeoidea, Melolonthidae): genera et catalogue commenté*. Paris: Lacroix, p. 277.

Lanfear, R., J. J. Welch, and L. Bromham (2010). Watching the clock: studying variation in rates of molecular evolution between species. *Trends in ecology & evolution* 25, 495–503.

Lanfear, R., B. Calcott, S. Y. W. Ho, and S. Guindon (2012). PartitionFinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. *Molecular biology and evolution* 29, 1695–1701. DOI: 10.1093/molbev/mss020.

Lloyd, J. E. (1979). Mating Behavior and Natural Selection. *The Florida Entomologist* 62, 17. DOI: 10.2307/3494039.

Mayr, E. (1963). *Animal species and evolution*. 1st ed. The Belknap Press of Harvard University Press, Cambridge, Massachusetts, p. 811.

McPeek, M. A. (1995a). Morphological evolution mediated by behavior in the damselflies of two communities. *Evolution* 49, 749–769.

— (1995b). Testing hypotheses about evolutionary change on single branches of a phylogeny using evolutionary contrasts. *American Naturalist* 145, 686–703.

McPeek, M. A., L. Shen, J. Z. Torrey, and H. Farid (2008). The Tempo and Mode of Three-Dimensional Morphological Evolution in Male Reproductive Structures. *The American Naturalist* 171, E158–E178. DOI: 10.1086/587076.

Møller, A. E. and A. Pomiankowski (1993). Fluctuating asymmetry and sexual selection. *Genetica* 89, 267–279.

Møller and Swaddle (1997). "1.9 Why is adaptive external asymmetry not more common?" In: *Asymmetry, Developmental Stability and Evolution*. Oxford, New York, Tokyo: Oxford University Press, Oxford, pp. 21–23.

Monaghan, M. T., D. J. G. Inward, T. Hunt, and A. P. Vogler (2007). A molecular phylogenetic analysis of the Scarabaeinae (dung beetles). *Molecular Phylogenetics and Evolution* 45, 674–692. DOI: 10.1016/j.ympev.2007.06.009.

Nylander, J., F. Ronquist, J. Huelsenbeck, and J. Nieves-Aldrey (2004). Bayesian Phylogenetic Analysis of Combined Data. *Systematic Biology* 53, 47–67. DOI: 10.1080/10635150490264699.

Otronen, M. (1998). Male asymmetry and postcopulatory sexual selection in the fly *Dryomyza anilis*. *Behavioral Ecology and Sociobiology* 42, 185–191. DOI: 10.1007/s002650050430.

Palmer, A. R. (1996). From symmetry to asymmetry: phylogenetic patterns of asymmetry variation in animals and their evolutionary significance. *Proceedings of the National Academy of Sciences of the United States of America* 93, 14279–86.

Paradis, E., J. Claude, and K. Strimmer (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290.

Paradis, E. (2012). *Definition of Formats for Coding Phylogenetic Trees in R*.

Pollard, D. G. (1956). The Control of Chafer Grubs (Schizonycha Sp., Coleoptera, Melolonthinae) in the Sudan. *Bulletin of Entomological Research* 47, 347–360.

Pope, R. (1960). A revision of the species of *Schizonycha* Dejean (Col.: Melolonthidae) from southern Africa. *Bulletin of the Natural History Museum* 9, 63–218.

Posada, D. and K. A. Crandall (1998). MODELTEST : testing the model of DNA substitution. *Bioinformatics applications note* 14, 817–818.

Rambaut, A. and A. J. Drummond (2003). *Tracer: MCMC trace analysis tool*.

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3, 217–223. DOI: 10.1111/j.2041-210X.2011.00169.x.

Richman, A. D. and T. Price (1992). Evolution of ecological differences in the Old World leaf warblers. *Nature* 355, 817–821.

Rohlf, F. J. and D. Slice (1990). Extensions of the Procrustes Method for the Optimal Superimposition of Landmarks. *Systematic Zoology* 39, 40–59. DOI: 10.2307/2992207.

Rohlf, J. F. (2006). A comment on phylogenetic correction. *Evolution* 60, 1509–1515. DOI: 10.1002/(SICI)1097-0177(199909)216:1<1::AID-DVDY1>3.0.CO;2-T.

Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. a. Suchard, and J. P. Huelsenbeck (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology* 61, 539–42. DOI: 10.1093/sysbio/sys029.

Schilthuizen, M. (2007). The evolution of chirally dimorphic insect genitalia. *Tijdschrift voor Entomologie* 150, 347–354.

— (2013). Something gone awry: unsolved mysteries in the evolution of asymmetric animal genitalia. *Animal Biology* 63, 1–20. DOI: 10.1163/15707563-00002398.

Scholtz, C. (1990). Phylogenetic trends in Scarabaeoidea (Coleoptera). *Journal of Natural History* 24, 1027–1066.

Sharp, D. and F. Muir (1912). The comparative anatomy of the male genital tube in Coleoptera. *Transactions of the Entomological Society of London* 3, 477–641.

Shimodaira, H. and M. Hasegawa (2001). CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics applications note* 17, 1246–1247.

Sidlauskas, B. (2008). Continuous and arrested morphological diversification in sister clades of characiform fishes: a phylomorphospace approach. *Evolution; international journal of organic evolution* 62, 3135–56. DOI: 10.1111/j.1558-5646.2008.00519.x.

Simmons, L. W. (2001). *Sperm competition and its evolutionary consequences in the insects*. 1st ed. Princeton and Oxford: Princeton University Press, p. 435.

— (2014). Sexual selection and genital evolution. *Austral Entomology* 53, 1–17. DOI: 10.1111/aen.12053.

Simmons, L. W., C. M. House, J. Hunt, and F. García-González (2009). Evolutionary response to sexual selection in male genital morphology. *Current biology : CB* 19, 1442–6. DOI: 10.1016/j.cub.2009.06.056.

Simon, C., F. Frati, A. 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. *Annals of the Entomological Society of America* 87, 651–701.

Stern, D. L. (2013). The genetic causes of convergent evolution. *Nature reviews. Genetics* 14, 751–64. DOI: 10.1038/nrg3483.

Van Valen, L. (1962). A Study of Fluctuating Asymmetry. *Evolution* 16, 125–142.

Waage, J. K. (1979). Dual Function of the Damselfly Penis: Sperm Removal and Transfer. *Science* 203, 916–918.

Watson, P. J. and R. Thornhill (1994). Fluctuating asymmetry and sexual selection. *Trends in ecology & evolution* 9, 21–25.

Zelditch, M., D. Swiderski, H. Sheets, and W. Fink (2004). *Geometric Morphometrics for Biologists. A Primer*. San Diego, CA: Elsevir Academic Press, p. 416.

Chapter VII.

A historical biogeography of megadiverse Sericini – another story "Out of Africa"?

This chapter is published in:
Eberle J, S Fabrizi, P Lago, D Ahrens (2016) A historical biogeography of megadiverse Sericini – another story "out of Africa"? *Cladistics*, article first published online. DOI: 10.1111/cla.12162

Authors' contributions to the original article:
Manuscript conception and design: JE, DA; data acquisition/drafting, revising, and approving the article: JE, DA, SF, PL; analysis and interpretation of data: JE, DA.

1. Introduction

With nearly 4,000 described species, Sericini chafers (Coleoptera: Scarabaeidae) represent a megadiverse tribe of beetles with a nearly worldwide distribution which is absent only in Australia and circumpolar regions (Ahrens 2006c). Most extant species are found in a monophyletic lineage ("modern Sericini"; Ahrens 2006c; Ahrens and Vogler 2008). It comprises two large palaeotropic subtribes, Sericina and Trochalina, which number about 3,000 species and 370 species, respectively. The phytophagous Sericini belong to the lineage of Pleurostict chafers (Scarabaeidae) that are thought to have greatly diversified with the rise of angiosperms around 108 million years ago (Mya) (Farrell 1998; Ahrens et al. 2014b; Eberle et al. 2014). Compared to the soil dwelling larval stage, the emergence of adults is short. The beetles are generally fully winged but commonly exhibit restricted distribution patterns or even high endemism (e.g., Ahrens 2004; Liu et al. 2015). Their poor dispersal capacity is underlined by their absence on most oceanic islands and archipelagos (e.g. Lesser Antilles, Papua New Guinea, Canary Islands) and by high regional endemism (Ahrens 2004), and makes them highly interesting for the study of biogeographic patterns.

The origin of Sericini was proposed to be in Africa because of the exclusively African distribution of Ablaberini, the sister group of Sericini (Ahrens 2006b). The divergence of Sericini from Ablaberini dates back to ca 100 Mya (Ahrens et al. 2014b), which makes the group suitable for exploring biogeographical processes since the breakup of Gondwana, which deeply shaped distribution patterns of organisms in the southern hemisphere (Bossuyt et al. 2006; Waters and Craw 2006). Knowledge of the timing and dispersal modes and routes from Africa to other regions is crucial for a deeper understanding of the evolution of this group. In particular, the colonization of the Asian subcontinent, where Sericini are exceptionally diverse, is of major interest to provide a primer for investigating the causes of their exceptional species richness. The signal of past biogeographical processes that is observed in distribution patterns of extant

species may be overlain by subsequent dispersal events and their inference requires robust phylogenetic hypotheses (Givnish and Renner 2004; McGlone 2005). Previous considerations of Sericini biogeography were based on morphology-based phylogenetic hypotheses (Ahrens 2006a,b,c,d, 2007a,c) that were expected to be strongly influenced by homoplasy (Ahrens and Vogler 2008).

Most current knowledge of historical biogeography comes from vertebrates and plants with well-known distributions (Holt et al. 2013) and considerable fossil records. Evidence based on invertebrate data is rather rare despite their enormous species richness and ecological diversity (Sanmartín and Ronquist 2004; Monaghan et al. 2007; Schaefer and Renner 2008; Kodandaramaiah and Wahlberg 2009; Kodandaramaiah et al. 2010; Bukontaite et al. 2014; Struempher et al. 2014). Among the reasons for this may be poor knowledge of species phylogenetic relationships and taxonomy. Fragmented knowledge of species and their distributions may hamper capturing their diversity in space and time, and this is particularly true in megadiverse groups with significant numbers of undescribed species. Externally homogeneous morphology and strong homoplasy in the few existing diagnostic features (e.g. genitalia) aggravates the issue. In our study group, the scarab tribe Sericini, about 60% of the genera were erected as monotypic (Ahrens 2007b,c, ; unpublished data), particularly in Island faunas. In contrast, large and long-recognized genera of Sericini have been found to be para- or polyphyletic, with many smaller genera nested within these larger collective groups (Ahrens and Vogler 2008; Liu et al. 2015). Therefore, sampling based on such an error-prone taxonomy could easily lead to biased *a priori* assumptions. This is predictably true when only few species of wide-spread genera are included in an analysis or if certain clades, particularly those containing nested genera, are inadvertently over-sampled.

On the other hand, megadiverse groups are a valuable resource of information for biogeographic studies in that they provide an enormous diversity of species with comparable vagility and dispersal capability (e.g., Bossuyt et al. 2006; Linder 2008; Albert et al. 2011). Therefore, these groups are well suited not only for the detection of historically common but, in particular, of less frequented biogeographic dispersal routes. Sampling as many species as possible within a group without discarding congeneric species a priori, not only overcomes the above mentioned taxonomic sampling issues but also naturally increases the probability of biogeographic events being reflected in

its phylogenetic relationships. Observed biogeographic patterns should be spatially more clearly resolved than they would in a species-poor clade, in particular if high endemism prevails within the group. Multiple occurrences of a specific event will provide stronger evidence and the lack of a dispersal event, on the other hand, will be a stronger indication for a less likely route than might be seen in a species poor clade.

In the present study, we examine historical biogeographic processes that led to the current distribution of Sericini lineages in the context of a newly developed time-calibrated molecular phylogenetic hypothesis. We attempt to reconstruct major global scale migrations of the main lineages of Sericini, with the objective of capturing parallels or dissimilarities compared to current knowledge which is mainly derived from better-known vertebrates. We further investigate the dynamics of the extensive diversification of Sericini in the context of their historical biogeography. Efficient algorithms and increasing computational power enables us to overcome biases of phylogeny dependent analyses from a partly artificial generic classification and from large numbers of undescribed species by extensive sampling of all lineages without discarding congeneric species *a priori*.

2. Material and Methods

2.1. Sampling and molecular lab procedures

The present study greatly extends the sampling of previous molecular phylogenies of Sericini (Ahrens and Vogler 2008; Liu et al. 2015) in terms of taxa and geography (Electronic Supplement Table F4; Fig. F1). It includes 872 specimens from 46 countries representing 665 morphospecies of all major lineages of Sericini (Electronic Supplement Table F4, Table F2). Specimen collection, preservation and DNA extraction followed Ahrens and Vogler (2008). Vouchers are deposited in the collections of the Zoological Research Museum A. Koenig, Bonn (ZFMK). Two mitochondrial markers, the 3‘ end of cytochrome oxidase subunit 1 (*cox1*) and 16S ribosomal DNA (*rrnL*), and a fragment of nuclear 28S rDNA, containing the variable domains D3–D6 were used in our analysis. Fieldwork in South Africa was enabled by the following collection permits: Eastern Cape (Permit No.: WRO 122/07WR and WRO123/07WR), Gauteng (Permit No.: CPF6 1281), Limpopo (Permit No.: CPM-006-00001), Mpumalangma (Permit No.: MPN-2009-11-20-1232), and Kwazulu-Natal (Permit Nos OP3752/2009, 1272/2007, 3620/2006).

Specimens were preserved in 96% ethanol and identified by examining male genitalia. Species were sorted to morphospecies if identification to a described species was impossible. DNA was extracted from the left mid-leg and from thoracic flight muscles of ethanol-preserved specimens with *Qiagen®DNeasy Blood & Tissue* Kits using standard protocols. Subsequently, the genitalia were glued on a card and dry mounted on the same pin as the specimen. The mitochondrial markers and nuclear DNA fragments, as described above, were amplified with polymerase chain reaction (PCR, see Table F3 for PCR protocols). *Qiagen®Multiplex PCR Kits* were used with primers stevPat and stevJerry for *cox1* (Timmermans et al. 2010), 16Sar and 16sB2 for *rrnL* (Simon et al. 1994), and FF and DD (Monaghan et al. 2007) for 28S. For-

ward and reverse strands were sequenced by Macrogen (Seoul, South Korea) using the same primers. Sequences were manually edited in Geneious 7.1.8.

2.2. Multiple sequence alignment and phylogenetic inference

Since multiple sequence alignment can be problematic for large datasets, especially when markers with highly variable regions like *rrnL* are included, we employed the divide-and-conquer realignment technique implemented in SATé-II (version 2.2.7, Liu et al. 2012). This method simultaneously estimates a phylogenetic tree and the alignment in multiple iterations and can lead to great improvements in hard-to-align data sets by deconstructing the alignment to smaller, closely related subsets of sequences (subproblems), which are separately aligned and subsequently merged. We ran 10 iterations on the multilocus data set, aligning subproblems with a maximum size of 200 individuals with MAFFT (version 6.717, Katoh and Toh 2008, 2010). Subproblems were generated by the centroid strategy and remerged with Muscle (version 3.7, Edgar 2004a,b). The simultaneous tree estimation was done with FastTree (version 2.1.4, Price et al. 2010).

Phylogenetic relationships were inferred using maximum likelihood in RAxML (version 8.0.20, Stamatakis 2014). The combined matrix was partitioned for the three markers and the tree was estimated under the GTR+CAT model (Stamatakis 2006) with final optimization under the GTR+Γ model. Base frequencies were estimated for each partition. Branch support was assessed by the nonparametric Shimodaira–Hasegawa–like implementation (SHL, Guindon et al. 2010) of the approximate likelihood-ratio test (aLRT, Anisimova and Gascuel 2006), which is much faster than traditional bootstrapping (Anisimova and Gascuel 2006; Guindon et al. 2010; Anisimova et al. 2011) and more robust against model violations (Anisimova et al. 2011). We adopt a conservative approach by considering branches with SHL-values > 85 as strongly supported (Guindon et al. 2010; Anisimova et al. 2011; Pyron and Wiens 2011; Pyron 2014).

Well supported phylogenetic information is mandatory for correct reconstruction of historical biogeography (Santos and Amorim 2007). Only small deviations in the topology of the Sericini phylogeny would imply different biogeographical conclusions

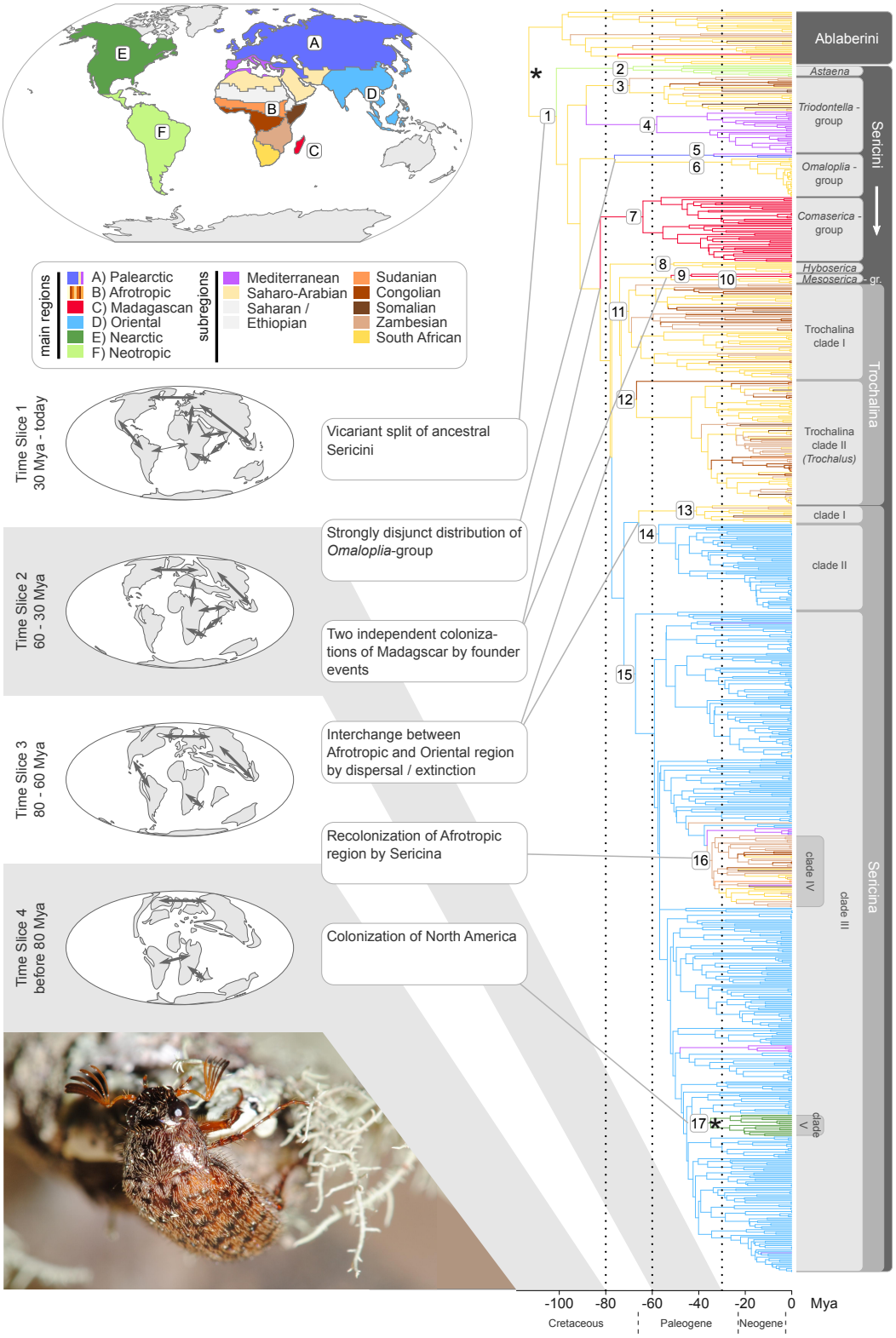

Figure VII.1. Caption on next page.

Figure VII.1. Time calibrated phylogeny of Sericini chafers and biogeographical inferences. Color codes for ancestral area reconstruction are explained upper left. Branches of the tree are colored according to the maximum likelihood ancestral range estimations of the subregions. Time slices with the respective paleogeography (based on R. Blakey's maps, `www.jan.ucc.nau.edu`) including the stratification settings used in the BioGeoBEARS analysis are shown on the left: High dispersal probabilities (1.0) are indicated by thick and lower probabilities (0.5) by narrow arrows. Numbers on the tree mark nodes discussed more in detail in the text. Calibration points are indicated by an asterisk.

for the colonization of the Oriental region. Therefore, site bootstrapping was performed in CONSEL (Shimodaira and Hasegawa 2001) on two alternative tree hypotheses that were inferred by constrained RAxML analyses of the same dataset. Scenarios of monophyly (i) of the Oriental Sericina clades II and III and (ii) of African "Modern Sericini" (i.e., Trochalina + Sericina clade I; Fig. VII.1) were compared to the unconstrained scenario.

2.3. Divergence time estimation

Divergence times were estimated on the fixed topology of the RAxML tree with BEAST (version 1.7.5, Drummond et al. 2012). We used two calibration points: one fossil of *Serica antediluviana* Wickham 1912 (Wickham 1912; Krell 2000) from Florissant, USA (37.2–33.9 Mya, http://fossilworks.org, accessed May 8th, 2015), was used to calibrate the only clade of Nearctic Sericini (Fig. VII.1, node 19) with an exponential prior, since exponential distributions require only one parameter and no further information about the mode of the prior distribution was available (Ho and Phillips 2009). The geographical occurrence of *S. antediluviana* makes it the only unambiguously assignable fossil of the group, since it is impossible to reliably determine clade affiliations of fossilized Sericini based on their homogeneous external morphology. A second calibration point was applied to the most recent common ancestor of Ablaberini and Sericini from a previous study (Ahrens et al. 2014b) with a lognormal distribution (mean=90.28, stdev=0.0635, offset=8.48) since divergence times estimated from molecular data typically exhibit lognormal distributions (Morrison 2008). PartitionFinder (Lanfear et al. 2012, 2014) was used to infer optimal partition schemes and substitution models. *Cox1* was divided into its codon positions because

they are known to differ substantially in their substitution rates (Ho and Lanfear 2010). Due to convergence issues with more complex substitution models (GTR+I+Γ and SYM+I+Γ) that were inferred with PartitionFinder, the simpler model HKY+Γ was set for all partitions. The data were subdivided into 5 partitions (*rrnL*, 28S, 3 codon positions of *cox1*) to agree with the PartitionFinder results as closely as possible. The uncorrelated lognormal relaxed clock model (Drummond et al. 2006) was used to estimate branch rates and the Yule process was set as tree prior. Two independent runs with 108 generations each, sampling every 5,000 generations, were performed on the high performance computing cluster at the ZFMK. Convergence and stationarity of all model parameters were assessed with Tracer (v1.6; Rambaut et al. 2014) by checking ESS values and visually inspecting the log-likelihood vs. generation plots. Based on the latter, a burnin of 10^6 generations was discarded. The sampled trees of the individual runs were combined using LogCombiner and the maximum clade credibility tree with mean node heights was calculated using TreeAnnotator (Drummond et al. 2012).

2.4. Biogeographic Analyses

Biogeographical regions have mostly been defined by high levels of endemism and distributional dissimilarity of organisms for quite some time (Wallace 1876). Recent studies based on cluster analyses and dissimilarity of subregions (Linder et al. 2012; Procheş and Ramdhani 2012), which also include phylogenetic turnover of assemblages of species (Holt et al. 2013), have shown that zoogeographical patterns are specific for individual groups of organisms, regardless of scale, whether it be global or continental. However, the principal borders of zoogeographical regions are consistent between groups (Linder et al. 2012; Procheş and Ramdhani 2012; Holt et al. 2013). Ancestral ranges have mostly been inferred using ancestral state reconstruction under Maximum Parsimony or Maximum Likelihood criteria without explicit models for range evolution. Meanwhile, several models describing biogeographical events (Ronquist 1997; Ree and Smith 2008; Landis et al. 2013; Matzke 2013a) have been developed and these provide valuable extensions of ancestral range inference methods. The recently developed BioGeoBEARS (Matzke 2013a) provides extensive models that incorporate all previously mentioned models extending them for founder-

event speciation (long distance or trans-oceanic dispersal), which previously has not been explicitly considered (Matzke 2012, 2013b). BioGeoBEARS therefore provides a well-suited framework for testing the fit of different models to the data.

Since all methods used to reconstruct the historical biogeography of the group have certain deficiencies, especially regarding model shortcomings and limitations for the number of areas under consideration, we employed multiple approaches to infer ancestral ranges. In contrast to reconstructions with maximum likelihood (ML) and parsimony (MP), model based ancestral area estimation with BioGeoBEARS allows the inference of processes underlying a biogeographic event. It also allows stratification of the analysis over geological time, which may improve the inference by taking into account paleogeography and continental drift (Buerki et al. 2011). Analyses were conducted in the R statistics environment using (i) Maximum Parsimony under the MPR (Hanazawa et al. 1995) and the ACCTRAN (Farris 1970; Swofford and Maddison 1987) criterion in *phangorn* (Schliep 2011), (ii) Maximum Likelihood ancestral states estimation in APE (Paradis et al. 2004), and (iii) ancestral area estimation in BioGeoBEARS (Matzke 2013a,b).

In total, six main regions and nine sub-regions were defined (Fig. VII.1) which were largely inherited from statistically defined biogeographical regions of previous studies (Linder et al. 2012; Holt et al. 2013). Species were considered to occur only in one region at once. Major zoogeographical regions were derived from realms postulated by Holt et al. (2013); however, modifications were made to fit specific biogeographical patterns of Sericini. The Oriental region and the Sino-Japanese subregion, which belong to the Palaearctic, were never isolated and had similar climates throughout their histories (Cox and Moore 2010). As a result, they are very similar in species assemblages and are not more differentiated than the subregions of the Oriental realm (e.g. Indian vs. Indochinese subregion) such that a clear separation was not possible. The Sino-Japanese and the Oriental region were therefore treated as one unit to satisfy computational limitations. Likewise, the Saharo-Arabian realm (Holt et al. 2013), which is represented by only 3 specimens in our sampling, was combined with the Palaearctic region. For the ancestral area reconstruction methods, which are not susceptible to the number of areas that are considered in the analysis (see below), the Mediterranean Basin and the Saharo-Arabian subregion were coded separately from the Palaearctic, and the Afrotropical region was further subdivided following Linder

et al. (2012). *Maladera affinis* from Réunion Island was coded as an Oriental species because it was most likely introduced with sugar cane or moved by Indian immigrants (Ahrens 2003).

The BioGeoBEARS analysis was stratified by setting different dispersal probabilities between regions for four geological time slices according to continental drift and ocean currents following Buerki et al. (2011) (Fig. VII.1, Electronic Supplement Table F5), with the best model chosen by its AIC score (Table F1). Seven models were compared and the best one was chosen by its AIC score: DEC (Ree and Smith 2008), DEC+J, DIVA-like (Ronquist 1997), DIVA-like+J, BayArea-like (Landis et al. 2013), BayArea-like+J, and a modified DEC+J model that only allowed simultaneous occupation of continuous landmasses (0–30 Mya: AD and EF, 30–60 Mya AD and AE, 60–80 Mya: AD, and 80–120 Mya: AD and BF; Fig. VII.1). This way, migrations between continents that were not adjacent in the respective time slices were possible only by long distance over sea dispersal (founder events). Due to computational limitations, single species were maximally allowed to occupy two regions at one time.

2.5. Analysis of Diversification

Sampled lineages of Sericini through time (LTT) of the BEAST tree were plotted with APE (v3.3, Paradis et al. 2004). LTT plots of 100 evenly spaced subsamples of the BEAST post-burnin tree samples indicated the range of node divergence time. The same procedure was applied to the subclades of Sericina (node 14, 15, and 16), African Sericina clade IV (node 17), and Trochalina (node 11 and 12).

The dynamics of species diversification through time were inferred with BAMM (v2.5.0; Rabosky et al. 2013; Rabosky 2014; Shi and Rabosky 2015). The fraction of the species sampled from the major sericine clades was roughly estimated based on a world species database of Sericini. The backbone sampling fraction was set as the ratio of species sampled to the total number of known Sericini species. These settings, and subsequent analysis of the BAMM results, have to be made with care since the systematic assignment of species to the various defined lineages is problematic in some cases (Liu et al. 2015). Furthermore, previous taxonomic revisions revealed large proportions of undescribed species (e.g., Ahrens 2004, 2005; Ahrens et al. 2014a,b;

Fabrizi and Ahrens 2014; Liu et al. 2014). Many could not be considered for yet unrevised groups. Ablaberini, as well as clearly duplicate species were pruned from the tree prior to the analysis to avoid an overestimation of diversification rates. According to the number of species that were included in the analysis, the number of expected shifts was set to 5. Priors were optimized for the data with BAMMtools (v2.1.0; Rabosky et al. 2014b). Four MCMC chains were run for 30 million generations each, at a temperature increment parameter of 0.01, sampling every 15,000 generations. Speciation rate shifts on tips were not considered (mincladesize=2). Convergence and stationarity of the run was assessed with CODA (v0.18-1; Plummer et al. 2006) in R by visually inspecting the traces of the log-likelihood and the number of shifts and by calculating the post-burnin effective sample size. The output was summarized and visualized with BAMMtools, including speciation rate through time plots for the above mentioned clades (LTT). Differences in the speciation rate dynamics of Sericini lineages were illustrated using a cohort analysis (Rabosky et al. 2014a).

3. Results

3.1. Phylogenetic inference

The phylogenetic relationships of main Sericini lineages (Fig. VII.1) based on the alignment of 2063 base pairs largely coincided with previous hypotheses (Ahrens and Vogler 2008; Liu et al. 2015): Ablaberini was recovered as a sister group to Sericini. South American Sericini, represented by *Astaena* (i.e., 7 of about 170 species and 1 out of 4 genera; node 2) was consistently recovered as sister to all other Sericini. The remaining species grouped in the following clades (Fig. VII.1, Table VII.1): the *Triodontella*-group (nodes 3 + 4), the *Omaloplia*-group (nodes 5 + 6), the Madagascan *Comaserica* group comprising also *Hyposerica* (node 7), *Hyboserica* (node 8), Trochalina as sister to the *Mesoserica*-group (nodes 9 + 10) , and Sericina. Trochalina and Sericina are the only two subtribes so far defined for Sericini (Machatschke 1959). Tests of alternative constrained topologies of the oldest Sericina clades with CONSEL highly supported the unconstrained analysis (Table VII.2). Concordant with previous studies (Ahrens 2004; Liu et al. 2015), the traditional generic classification of many taxa was found to be inconsistent with the tree. Besides many undescribed species, many genera were highly para- or polyphyletic and scattered over the tree.

3.2. Biogeographic inferences and divergence times

The widely and intensely sampled and well-resolved phylogenetic tree allowed exploration of the historical timescale of biogeographic patterns in Sericini (Electronic Supplement Fig. F1). The different methods of ancestral area inference (Fig. VII.1, Electronic Supplement Figs. F2–F3) yielded similar results, except for the exact origin (biogeographic subregion) of the subtribe Trochalina and the colonization mode

Table VII.1. Timing of stem group origins of principal Sericini lineages with the respective biogeographic event inferred by BioGeoBEARS and current distribution.

Node	Clade	Distribution	Biogeographic event	Origin of stem group [Mya]	95% HPD [Mya]
1	Sericini	worldwide except Australia	-	111.60	101.19–121.13
2	Neotropical Sericini	Neotropic	vicariance	102.85	92.53–112.45
3	*Triodontella* group part.	Afrotropicc	dispersal-extinction	89.51	79.37–100.63
4	*Triodontella* group part.	Mediterranean	dispersal-extinction	"	"
5	*Omaloplia* group part.	Palaearctic	dispersal-extinction	78.79	63.61–90.98
6	*Omaloplia* group part.	Southern Africa	dispersal-extinction	"	"
7	*Comaserica* group	Madagascar	founder event	82.28	73.49–90.61
8	*Hyboserica*	South Africa	-	79.19	71.43–88.00
9	*Mesoserica* group part.	Madagascar	founder event	53.6	42.64–64.69
10	*Mesoserica* group part.	South Africa	-	"	"
11	Trochalina clade I	Afrotropic	-	70.61	63.13–78.78
12	Trochalina clade II	Afrotropic	-	"	"
13	modern Sericini	P, O, A, N	-	73.45	65.52–81.73
14	Sericina clade I	Afrotropic	-	67.64	59.72–76.76
15	Sericina clade II	SE Asia	dispersal-extinction	"	"
16	Sericina clade III	P, O, A, N	dispersal-extinction	73.50	65.11–81.18
17	Sericina clade IV	Afrotropic	dispersal-extinction	39.91	35.01–45.03
18	*Euserica*	Mediterranean	founder event	46.84	40.68–83.70
19	Sericina clade V	Nearctic	founder event	43.16	39.13–47.69

A = Afrotropic, N = Nearctic, O = Orient, P = Palaearctic

Table VII.2. Parametric bootstrapping support of the CONSEL analysis for topologies from constrained analyses that imply alternative biogeographic conclusions for the colonization of Asia. The *p*-values of the approximately unbiased test (au) and the bootstrap probability (np) are shown.

Topology constraints	au	np
unconstrained	1.000	1.000
Sericina clade II and III monophyletic	<0.001	<0.001
Trochalina and African Sericina clade I monophyletic	<0.001	<0.001

of the Oriental region (see below). Inferences considering selected biogeographical subregions (Africa and the Mediterranean) were based on ancestral state inference (ML and MP) methods only (Fig. VII.1, Electronic Supplement Fig. F3). ML and MP analyses consistently resulted in the southern African subregion as the area of origin of Sericini and all of its African lineages, except for Sericina clade IV (node 16; origin in the Zambesian subregion), as well as the ancestor of Trochalina. For the latter, however, MP reconstruction with MPR was ambiguous while ACCTRAN optimization predicted the origin in the Zambesian subregion (Electronic Supplement Fig. F3).

BioGeoBEARS analyses on the six main regions (Fig. VII.1, Electronic Supplement Fig. F2) fitted the data best under the DEC+J model (Table F1). Models including cladogenetic dispersal by founder events (+J, Matzke 2012) always fitted the data better than the respective alternatives without founder events (Table F1). The common ancestor of all Sericini occurred on the West Gondwanan landmass (Africa and South America; VII.1). The divergence between the Afrotropical Ablaberini and Sericini is dated back to 111.6 Mya (Table VII.1). The most basal split of Sericini (node 1, 93–112 Mya 95% HPD), separating the Neotropical genus *Astaena* from the remaining Sericini, was inferred as a vicariance process and coincides with the breakup of West Gondwana at about 105 Mya (Sanmartín and Ronquist 2004; Cox and Moore 2010).

The subsequent biogeographic history of the remaining Sericini was characterized by repeated migrations out of Africa and back again. The two oldest lineages of Old World Sericini (i.e. the *Triodontella*-group (nodes 3 + 4) and the *Omaloplia*-group (nodes 5 + 6)) showed similar distribution patterns of their two respectively vicariant clades. Ancestors of both lineages were inferred to have occurred in both the

Palaearctic and Africa (64–101 Mya). In contrast to the African lineage of the *Triodontella*-group (node 3), which is today distributed throughout Sub-Saharan Africa, *Pleophylla* (node 6) occurs only in forest remnants of southern Africa northward to the Albertine Rift Mountains in Rwanda, so that the extant distribution of the *Omaloplia*-group has to be considered strongly disjunctive. Madagascar appeared to be invaded twice by a founder event (i.e., long distance trans-oceanic dispersal) by two independent lineages: the *Comaserica*-group (node 7) and the Madagascan clade of the *Mesoserica*-group (node 9). Invasion by the latter species-poor clade took place about 30 My after the first invasion (Table VII.1).

The mode of colonization of the Oriental region by Sericina could not be unambiguously resolved. Since relationships between basal Sericina were found to be stable according to SH-like aLRT branch support and the outcome of the CONSEL analysis, two scenarios of the colonization are plausible: (1) an ancestor of all Sericina populated Africa, and the Orient was invaded twice, or (2) the ancestor of Sericina dispersed to the Orient only once and Sericina clade I reinvaded Africa (result of the ML reconstruction, however, with low certainty, and MP reconstruction under ACCTRAN, Electronic Supplement Fig. F3). BioGeoBEARS inferred (with high uncertainty) scenario (1) with a vicariant split of the basal Sericina lineages (nodes 14 + 15 and 16) and subsequent jump dispersal (Matzke 2014) to the Orient (node 15) (Electronic Supplement Fig. F2).

Greater certainty is seen in the recolonization of Africa by Sericina clade IV (node 17). However, SH-like aLRT supports are low at node 17 and its ancestral nodes (Electronic Supplement Fig. F3) and one species of *Maladera* (DA3821) from Africa seemed to distort the BioGeoBEARS analyses seriously. It is nested in the neighborhood of the previously mentioned clades containing Asian species and led BioGeoBEARS to spurious results (i.e., much earlier dispersal to Africa implying 5 re-dispersals to Asia and the Eastern Mediterranean). The sister group of Sericina clade IV is *Maladera* (subgenus *Macroserica*), which has its extant distribution in the Eastern Mediterranean subregion.

The colonization of North America was inferred to occur about 43 Mya (HPD: ca. 39–48 Mya), at a time when the Bering Strait was mostly traversable, while the still present Turgai Sea and the expanding Norwegian Sea inhibited interchange between Asia, Europe, and Eastern North America (Cox and Moore 2010). An eastward

migration along the Bering Strait of today's Nearctic Sericina (node 19) is likely. Members of the genus *Euserica* (node 18), all inhabiting the Mediterranean subregion, diverged from its Asian sister group around 47 Mya, when the Turgai Sea still existed.

3.3. Analysis of Diversification

Four rate shifts were found in the highest probability rate shift configuration (posterior probability = 0.13) with BAMM (Fig. VII.2). Medium high speciation rates were inferred for ancient Sericini lineages including *Astaena*, the *Triodontella*-group and the *Omaloplia*-group, followed by a slowdown of speciation rate. Distinct shifts to higher speciation rates were found for the genus *Trochalus* (Trochalina clade II, node 12, without basal lineage), the derived genus *Tetraserica* which is nested within Sericina clade II (node 15), and the Asian Sericina clade III (node 16) except its basal lineage. Mean speciation rates increased conspicuously from ca 62 to 50 Mya, in particular in Sericina clade III (node 16, Fig. VII.3) which is concordant with a remarkable slowdown in the accumulation of lineages through time from 65 to 60 Mya (Fig. VII.3) after the Cretaceous-Paleogene mass extinction (Schulte et al. 2010). While the speciation rate decreased in Sericina clades III and IV since 50 Mya, they slightly increased in Trochalina clades I and II (Fig. VII.3). At 30 Mya a subtle slowdown of lineage aggregation is visible for all clades examined. A corresponding signal in the diversification rates (Fig. VII.3) is very weak but still present.

The cohort analysis clearly identified differing macroevolutionary dynamics for lineages with an inferred rate shift (Fig. F2), but also identified lineages with subtler deviations in rate dynamics, for instance, Sericina clade IV, which reinvaded Africa in the Middle Eocene. Examining this clade alone revealed exceptionally high but steadily decreasing rates of speciation (Fig. VII.3).

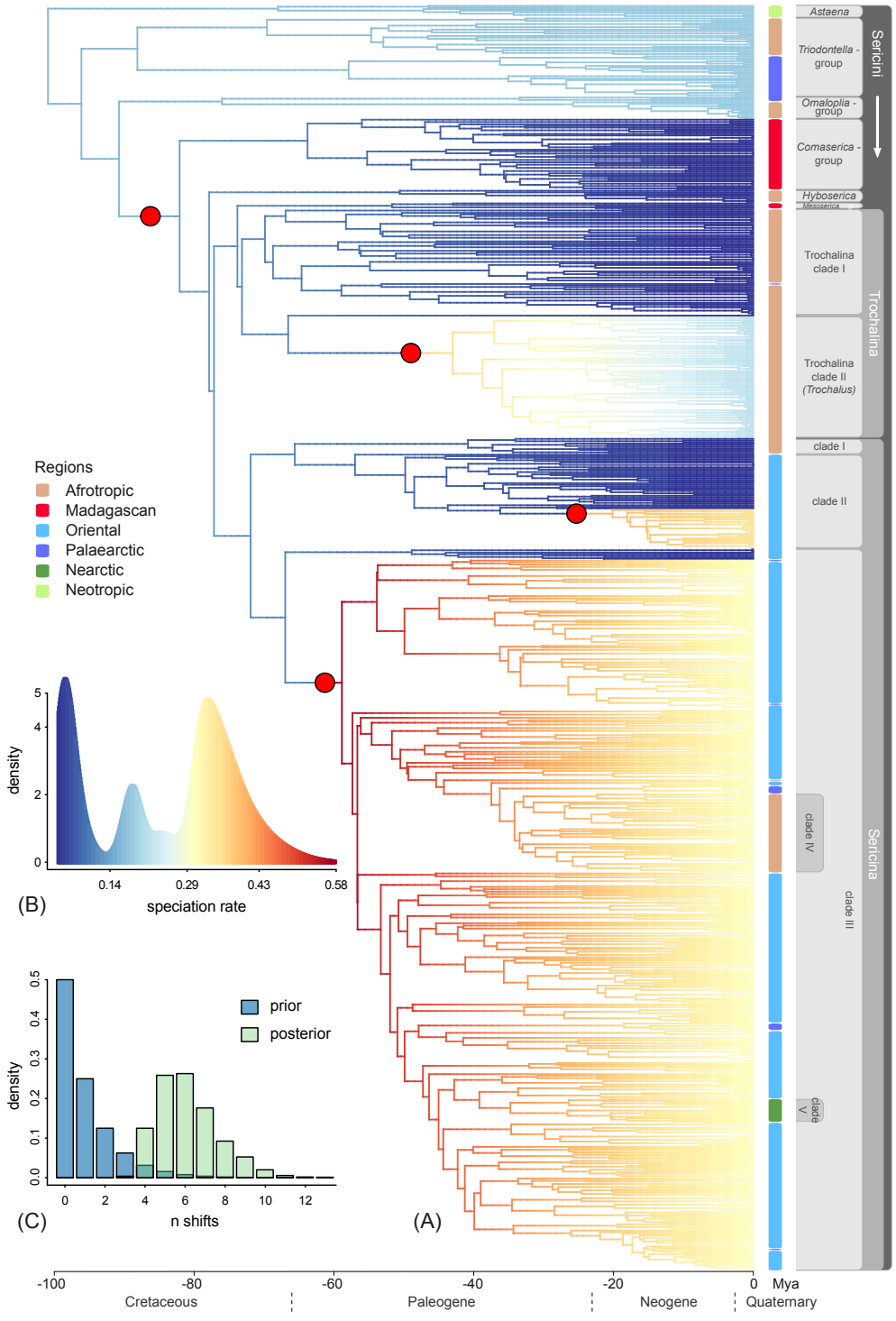

Figure VII.2. Caption on next page.

Figure VII.2. Speciation rate dynamics of Sericini. (A) Phylorate plot of the rate shift configuration with the maximum *a posteriori* probability that was inferred with BAMM. Major rate shifts are marked by red dots. Warmer colors indicate higher rates of speciation and correspond to (B) the histogram of speciation rates. (C) Prior and posterior distribution of the number of shifts. Geographic origin and taxonomy of the specimens under study are indicated by the columns next to the tree.

4. Discussion

The present study is the first comprehensive biogeographic study of Sericini chafers, a megadiverse group of beetles with an as yet poorly resolved taxonomy. In addition to incomplete sampling and data support, and therefore possibly uncertain phylogenetic information, choice of biogeographic reconstruction methods and models, and unknown extinction events may obscure crucial biogeographic events (Heath et al. 2008; Crisp et al. 2011; Matzke 2013b; Matzke 2014). Our results demonstrate how extensive sampling may help to overcome sampling induced biases that originate from poor classification and taxonomic knowledge in phylogeny-dependent biogeographic analyses. This 'as-complete-as-reasonably-possible sampling' approach also helped to minimize long-branch attraction effects (Bergsten 2005) and to overcome problems of sequence alignment (Philippe et al. 2011). The divide-and-conquer realignment technique with simultaneous tree estimation (SATé-II, Liu et al. 2012) that we employed here may lead to great improvements, especially in hard-to-align data sets. Branch support was reasonable for the major clades that were used for biogeographic inferences. The basal position of the Neotropical genus *Astaena* had low aLRT support but is backed up by previous molecular and morphological studies (Ahrens 2006c; Ahrens and Vogler 2008; Liu et al. 2015). The same applies to the sister group relationships of lineages within the *Triodontella*- and the *Omaloplia*-group (nodes 3, 4, 5, and 6); however, the position of *Hyboserica* (node 8) has to be treated with care. All of this helped to facilitate the inference of the complex phylogenetic history of the large radiation of Sericini beetles, for which many of the commonly used algorithms would reach computational limits (Varón et al. 2010; Ronquist et al. 2012).

The historical biogeography of Sericini was discovered to be characterized by repeated migrations out of Africa with a huge radiation in Southeast Asia, from where they (re-)colonized distant regions, including Africa. The West-Gondwanan (i.e., African) origin of Sericini revealed here by several methods is backed by the exclu-

sively Afrotropical distribution of Ablaberini, the sister tribe of Sericini. The age of Sericini (ca 112 Mya) was inferred to be about 10 My older than previously estimated (Ahrens et al. 2014b), scarcely pre-dating the separation of South America from Africa (ca 105 Mya, Sanmartín and Ronquist 2004; Cox and Moore 2010). The divergence of the South American Sericini (ca 103 Mya) appeared to be more likely the result of vicariance (Electronic Supplement Fig. F2) rather than that of long distance over-sea dispersal (Rage 1988; Mourer-Chauviré 1999) which was hypothesized previously (Ahrens 2006c). Since morphological evidence strongly supports the monophyly of the South American Sericini (Ahrens 2006c), this hypothesis is unlikely to be rejected based on a more representative sampling of the Neotropical fauna. While movements of vertebrates between post-Gondwanan fragments were rare and mainly dispersals "out-of-Africa", interchanges between Africa and Laurasia were numerous and bidirectional during the Cretaceous and the Paleogene (Gheerbrant and Rage 2006). Dispersal over diverse routes (e.g., trans-Tethyan dispersals via the Mediterranean Tethyan Sill) by amphibians, dinosaurs, and mammals were described from the earliest Cretaceous (Gheerbrant and Rage 2006), even for those considered to be poor dispersers. Likewise, all of the six global scale migrations of Sericini during the Cretaceous and Early Paleogene were dispersals "out of Africa". More recent trans-Tethyan migrations of vertebrates in the Eocene were bidirectional (Gheerbrant and Rage 2006) or happened in the Northern Hemisphere (e.g. nodes 18 and 19). This is supported by the re-colonization of Africa by Sericini (node 17) and also by a dispersal of nymphalid butterflies from Africa to Asia around the same time (Kodandaramaiah and Wahlberg 2007; Aduse-Poku et al. 2009).

While vicariance sometimes displaced trans-oceanic dispersal as an explanation for observed disjunct distributions, the latter has been resurrected by various authors in an attempt to explain the colonization of distant regions (Givnish and Renner 2004; De Queiroz 2005; Matzke 2012; Pyron 2014). Similarly, in the present study, biogeographic models including founder event speciation always fit the data best (Table F1). In particular, the colonization of Madagascar was recently explained by trans-oceanic dispersal from Africa (Ali and Huber 2010; Tolley et al. 2013), as was also found in the present study for both Madagascan lineages. This was consistent with other studies showing that founder event speciation is an important biogeographic process (i.e., colonization) for oceanic islands, but is also apparent in terrestrial or global systems

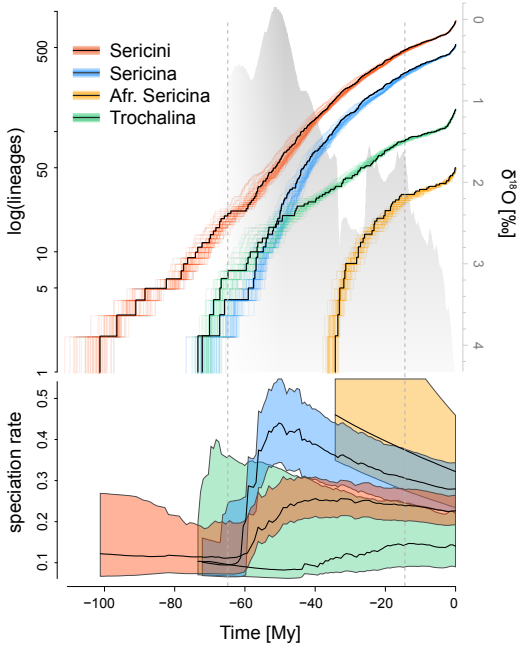

Figure VII.3 Diversification of Sericini through time. Lineage and rate through time plots of the time calibrated phylogeny from Bayesian inference of Sericini and subclades. Colored lines in the upper half indicate the divergence time range of the Bayesian post-burnin tree sample, black lines the mean node heights. Deep sea oxygen isotope records (δ^{18}O) as a proxy for climate change (Zachos et al. 2001) are shaded in grey. The lower half illustrates speciation rates for the same clades that were inferred using BAMM. Vertical lines at 65 Mya and 15 Mya depict the beginning Cretaceous-Paleogene mass extinction and the Mid-Miocene climate cooling, respectively.

(Matzke 2013a,b; Matzke 2014; Pyron 2014). Although land bridges to Madagascar, such as the Davie Ridge (Taquet 1982) have been suggested, the divergence dates of these dispersal events more likely overlap with the timing of the prevailing west to east (Africa to Madagascar) palaeo-currents (Ali and Huber 2010). Similar evidence also exists for vertebrates and other invertebrates (Vences et al. 2003; Esser and Cumberlidge 2011; Samonds et al. 2012; Tolley et al. 2013). Since the Madagascan lineages considered in the present study only cover a small part of the genera of the highly endemic fauna, some uncertainty about the exact number of such independent colonizations remains. Based on preliminary morphological evidence we may definitively exclude any of the Madagascan species as belonging to either Sericina or Trochalina, which has a significant impact on the reconstruction of the colonization of Asia.

Several routes have been proposed for the colonization of Southeast Asia from Africa, including trans-Tethyan dispersal, long distance oceanic dispersal and "out of India" rafting (Karanth 2006). Dispersal over the Tethyan Sill would be the only scenario fitting the BioGeoBEARS inference for early Sericini (Fig. VII.1, Electronic Supplement Fig. F2); this route was still available for Sericini in the Late Cretaceous. Generally, different patterns of distributions of the two main Asian clades with preva-

lence for the Indomalay archipelago and Indochina (node 15) and the Asian mainland (node 16), respectively, make a second subsequent colonization of Southeast Asia of Sericina clade II (node 15) more likely than a re-colonization of Africa by clade I composed of mostly southern African species. Competitive exclusion of earlier and later colonizers might be a plausible cause shaping such patterns (Hardin 1960; Waters 2011). With the Somalian route, a second variant has been proposed for an "out of India"-scenario (Chatterjee and Scotese 1999; Mehrotra 2003). Since this fits the timing of the Asian invasion as well, the "out-of-India" hypotheses cannot be convincingly rejected without more comprehensive sampling of the southern Indian Sericini fauna, represented in our study by only a few *Maladera* species (Fig. F1, Electronic Supplement Table F4). In the same way, this might help to overcome the ambiguity for the precise reconstruction of the colonization of Asia (see above).

By the Middle Eocene, the reinvasion of Africa by Sericina happened very likely via the Iranian route that connected Africa with southeastern Europe and southwestern Asia (Gheerbrant and Rage 2006) since the sister of the African clade IV (node 17) is distributed in the eastern Mediterranean. Discounting some uncertainty due to the phylogenetic misplacement of one Angolan *Maladera* species, the success of this re-colonization is impressive compared to other groups (Monaghan et al. 2007): clade IV is present in Africa with ca 300 species, nearly the same amount as that of autochthonous Trochalina.

During the period of North American colonization, the Bering Strait was mostly traversable, while the Turgai Sea was still present and the expanding Norwegian Sea inhibited an interchange between Asia, Europe, and Eastern North America (Cox and Moore 2010). Given that *Euserica* (node 18) is not directly related to the North America species (clade V), an eastward-directed migration from Asia along the Bering Strait of today's Nearctic Sericina (node 19) is most likely. The few Asian species in the European fauna have very wide ranges and very likely invaded Europe after the Pleistocene from eastern Siberia. In contrast, exchange between North and South America never occurred, most likely due to the young age of the land connection and the Central American deserts (Howden 1966).

Analyses considering the various subregions placed the origin of many older lineages (nodes 3, 6, 8, 10, 11, and 12) and that of Sericini itself in the southern African subregion. Strong exchange apparent between Afrotropical subregions (Fig. VII.1)

makes it likely that the ancestors of these lineages were widely distributed and became extinct in other regions. This hypothesis is supported by the disjunct distribution of old lineages (e.g., *Triodontella* group, *Omaloplia* group), a pattern more commonly observed in ancient lineages that faced ecological pressures over long time periods, followed by extinction in intervening areas (Pyron 2014). For example, the African lineage of the *Omaloplia* group *Pleophylla* (node 6), is presently only distributed in forest remnants in South Africa and the Albertine Rift Valley mountains in East Africa (unpublished data). Climatic refugia in southern Africa during the Last Glacial Maximum are evident for mammals, reptiles, and insects (Lawes et al. 2007; Lorenzen et al. 2010; Lorenzen et al. 2012; Barlow et al. 2013; Huntley et al. 2014; Switala et al. 2014) and low rates of extinction through climatic long-term stability in the Cape region (Schnitzler et al. 2011) might have promoted persistence of old lineages. Ancient relics (e.g., *Hyboserica*, *Pleophylla*) highlight the evolution of the African Sericini fauna as a series of large taxon pulses (Erwin 1981; Erwin 1985) whose general patterns might be highly relevant for identification of high priority conservation zones with reference to phylogenetic diversity (Sechrest et al. 2002).

The two clades of "modern Sericini", Trochalina and, in particular, Sericina include exceptionally species-rich radiations, which arose contemporaneously about 70 Mya. A rapid diversification was not evident in the early history of these clades, but younger lineages within them (*Trochalus* and Sericina clade III without basal lineage) did exhibit a strong diversity burst. Possible causes of extensive diversifications of Asian Sericina are radiations into new areas with strong tectonic dynamics (Besse and Courtillot 1988; Ramstein et al. 1997; Tapponnier et al. 2001; Hall 2002) or the development of evolutionary key innovations (Hunter 1998) such as the mechanical defense mechanisms used against ants by members of the genus *Trochalus*. Testing these hypotheses is, however, a subject for future research. Remarkably, the increase of the number of African and Asian Sericini lineages through time stagnated for about 5 My after the K-T boundary (Fig. VII.3). This period is known for the Cretaceous-Paleogene mass extinction whose causes (e.g., a single or multiple asteroid impacts, Deccan volcanism) that globally affected conditions for life and climate are still debated (Schulte et al. 2010). Studies on other phytophagous insect groups showed similar patterns which were attributed to a distinct reduction of actual lineages followed by subsequent diversifications (Wahlberg et al. 2009). The increase

of diversification rate in Sericini a few million years after the K-T-event supports this hypothesis. The slowdown of diversification within African Sericini lineages during the late Miocene-Pliocene coincides with hypotheses of a globally cooling climate (Zachos et al. 2001) and with a changing environment in Africa. The cold upwelling Benguela current along the Namibian coast (Siesser 1980) additionally cooled the southern African continental climate resulting in aridification and marked seasonality (Goldblatt and Manning 2000; Stuut et al. 2004). Open habitats, vegetated with more arid-adapted C4 grasses, and savannas expanded (Cerling et al. 1997; Verboom et al. 2014) and were colonized by Sericini, among which better dispersers and ant-adapted lineages were obviously favored. Consequences for the Asian lineages might have been less strong. Possibly the impact of climate oscillations was buffered or compensated by the presence of large mountain ranges throughout Asia. Future studies with more detailed geographical sampling should shed more light on this aspect of Sericine evolution. Refinement of paleogeographic scenarios by more intense geographical sampling, together with the interpretation of morphological evolution through more detailed morphological studies (e.g., Eberle et al. 2014), will likely help to further clarify the biogeographic history and the evolution of Sericini.

5. Acknowledgements

For providing us with research and collection permits in South Africa, we thank the various governmental institutions (electronic supplement material). We are grateful to Ruth Müller and James Harrison (Ditsong Museum Pretoria) for their logistic help and the numerous collectors for providing additional specimens. Reese Worthington provided valuable comments on an earlier version of this manuscript.

The study was supported by the German Science Association (DFG; AH175/1-2 and AH/175/3) and with a travel grant to JE by the A. Koenig Gesellschaft, Bonn.

References

Aduse-Poku, K., E. Vingerhoedt, and N. Wahlberg (2009). Out-of-Africa again: A phylogenetic hypothesis of the genus *Charaxes* (Lepidoptera: Nymphalidae) based on five gene regions. *Molecular Phylogenetics and Evolution* 53, 463–478. DOI: 10.1016/j.ympev.2009.06.021.

Ahrens, D. (2003). *Maladera affinis* (Blanchard, 1850) comb. n. (Coleoptera, Scarabaeoidea, Sericini), an oriental faunal element in the Malagasy region. *Deutsche Entomologische Zeitschrift* 50, 133–142. DOI: 10.1002/mmnd.20030500113.

— (2004). *Monographie der Sericini des Himalaya (Coleoptera, Scarabaeidae)*. Dissertation.de - Verlag im Internet GmbH, Berlin, p. 534.

— (2005). *A taxonomic review on the Serica (s. str.) MacLeay, 1819 species of Asian mainland (Coleoptera, Scarabaeidae, Sericini)*. 1st ed. Keltern: Goecke & Evers, p. 163.

— (2006a). Evolution of Asian 'lowland' taxa in relation to the Alpine-Himalayan Tertiary orogenic belt - Insight from a cladistic analysis of *Maladera (Cycloserica)* (Coleoptera: Scarabaeidae: Sericini). *Zoologischer Anzeiger* 244, 193–203. DOI: 10.1016/j.jcz.2005.10.002.

— (2006b). Phylogenetic analysis of *Anomalophylla* Reitter, 1887 (Coleoptera, Scarabaeidae: Sericini). *Insect Systematics & Evolution* 37, 1–16. DOI: 10.1163/187631206788831533.

— (2006c). The phylogeny of Sericini and their position within the Scarabaeidae based on morphological characters (Coleoptera: Scarabaeidae). *Systematic Entomology* 31, 113–144. DOI: 10.1111/j.1365-3113.2005.00307.x.

— (2006d). The phylogeny of the genus *Lasioserica* inferred from adult morphology - implications on the evolution of montane fauna of the south Asian orogenic belt (Coleoptera: Scarabaeidae: Sericini). *Journal of Zoological Systematics and Evolutionary Research* 44, 34–53. DOI: 10.1111/j.1439-0469.2005.00340.x.

— (2007a). Beetle evolution in the Asian highlands: Insight from a phylogeny of the scarabaeid subgenus *Serica* (Coleoptera, Scarabaeidae). *Systematic Entomology* 32, 450–476. DOI: 10.1111/j.1365-3113.2006.00373.x.

— (2007b). Taxonomic changes and an updated catalogue for the Palaearctic Sericini (Coleoptera: Scarabaeidae: Melolonthinae). *Zootaxa* 1504, 1–51.

— (2007c). Type species designations of Afrotropical Ablaberini and Sericini genera (Coleoptera: Scarabaeidae: Melolonthinae). *Zootaxa* 1496, 53–62.

Ahrens, D., W.-g. Liu, S. Fabrizi, M. Bai, and X.-k. Yang (2014a). A taxonomic review of the *Neoserica* (sensu lato) *abnormis* group (Coleoptera, Scarabaeidae, Sericini). *ZooKeys* 439, 27–82. DOI: 10.3897/zookeys.439.8055.

Ahrens, D., J. Schwarzer, and A. P. Vogler (2014b). The evolution of scarab beetles tracks the sequential rise of angiosperms and mammals. *Proceedings of the Royal Society B: Biological Sciences* 281, 20141470–20141470. DOI: 10.1098/rspb.2014.1470.

Ahrens, D. and A. P. Vogler (2008). Towards the phylogeny of chafers (Sericini): analysis of alignment-variable sequences and the evolution of segment numbers in the antennal club. *Molecular phylogenetics and evolution* 47, 783–98. DOI: 10.1016/j.ympev.2008.02.010.

Albert, J., P. Petry, and R. Reis (2011). "Major Biogeographic and Phylogenetic Patterns." In: *Historical Biogeography of Neotropical Freshwater Fishes*. Ed. by J. Albert and R. Reis. University of California Press Berkeley, Los Angeles, London. Chap. 2, pp. 21–58.

Ali, J. R. and M. Huber (2010). Mammalian biodiversity on Madagascar controlled by ocean currents. *Nature LETTERS* 463, 653–656. DOI: 10.1038/nature08706.

Anisimova, M., M. Gil, J. Dufayard, C. Dessimoz, and O. Gascuel (2011). Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Systematic biology* 60, 685–699.

Anisimova, M. and O. Gascuel (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic biology* 55, 539–552. DOI: 10.1080/10635150600755453.

Barlow, A., K. Baker, C. R. Hendry, L. Peppin, T. Phelps, K. a. Tolley, C. E. Wüster, and W. Wüster (2013). Phylogeography of the widespread African puff adder *(Bitis arietans)* reveals multiple Pleistocene refugia in southern Africa. *Molecular Ecology* 22, 1134–1157. DOI: 10.1111/mec.12157.

Bergsten, J. (2005). A review of long-branch attraction. *Cladistics* 21, 163–193. DOI: 10.1111/j.1096-0031.2005.00059.x.

Besse, J. and V. Courtillot (1988). Paleogeographic maps of the continents bordering the Indian ocean since the early Jurassic. *Journal of Geophysical Research* 93, 791–808.

Bossuyt, F., R. M. Brown, D. M. Hillis, D. C. Cannatella, and M. C. Milinkovitch (2006). Phylogeny and Biogeography of a Cosmopolitan Frog Radiation: Late cretaceous Diversification Resulted in Continent-Scale Endemism in the Family Ranidae. *Systematic biology* 55, 579–594. DOI: 10.1080/10635150600812551.

Buerki, S., F. Forest, N. Alvarez, J. a. Nylander, N. Arrigo, and I. Sanmartín (2011). An evaluation of new parsimony-based versus parametric inference methods in biogeography: A case study using the globally distributed plant family Sapindaceae. *Journal of Biogeography* 38, 531–550. DOI: 10.1111/j.1365-2699.2010.02432.x.

Bukontaite, R., K. B. Miller, and J. Bergsten (2014). The utility of CAD in recovering Gondwanan vicariance events and the evolutionary history of Aciliini (Coleoptera: Dytiscidae). *BMC evolutionary biology* 14, 5.

Cerling, T. E., J. M. Harris, B. J. MacFadden, M. G. Leakey, J. Quade, V. Eisenmann, and J. R. Ehleringer (1997). Global vegetation change through the Miocene / Pliocene boundary. *Nature* 389, 153–158. DOI: 10.1038/38229.

Chatterjee, S. and C. R. Scotese (1999). The breakup of Gondwana and the evolution and biogeography of the Indian plate. *Proceedings of the Indian National Science Academy Part A* 65, 397–426.

Cox, C. B. and P. D. Moore (2010). *Biogeography – An Ecological and Evolutionary Approach*. 8th ed. John Wiley & Sons, Inc., p. 498.

Crisp, M. D., S. a. Trewick, and L. G. Cook (2011). Hypothesis testing in biogeography. *Trends in ecology & evolution* 26, 66–72. DOI: 10.1016/j.tree.2010.11.005.

De Queiroz, A. (2005). The resurrection of oceanic dispersal in historical biogeography. *Trends in Ecology and Evolution* 20, 68–73. DOI: 10.1016/j.tree.2004.11.006.

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut (2006). Relaxed Phylogenetics and Dating with Confidence. *PLoS Biology* 4. Ed. by D. Penny, e88. DOI: 10.1371/journal.pbio.0040088.

Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular biology and evolution* 29, 1969–1973. DOI: 10.1093/molbev/mss075.

Eberle, J., R. Myburgh, and D. Ahrens (2014). The Evolution of Morphospace in Phytophagous Scarab Chafers: No Competition – No Divergence? *PLOS one* 9. DOI: 10.1371/journal.pone.0098536.

Edgar, R. C. (2004a). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32, 1792–1797. DOI: 10.1093/nar/gkh340.

— (2004b). No Title. *BMC Bioinformatics* 5, 113. DOI: 10.1186/1471-2105-5-113.

Erwin, T. L. (1981). Taxon pulses, vicariance, and dispersal: an evolutionary synthesis illustrated by carabid beetles. *Vicariance Biogeography: A Critique*. Ed. by G. Nelson and D. Rosen, 159–196.

Erwin, T. L. (1985). *The taxon pulse: a general pattern of lineage radiation and extinction among carabid beetles*. Ed. by G. E. Ball. Dordrecht: Dr W. Junk Publishers, pp. 437–472.

Esser, L. J. and N. Cumberlidge (2011). Evidence that salt water may not be a barrier to the dispersal of Asian freshwater crabs (Decapoda: Brachyura: Gecarcinucidae and Potamidae). *Raffles Bulletin of Zoology* 59, 259–268.

Fabrizi, S. and D. Ahrens (2014). A Monograph of the Sericini of Sri Lanka (Coleoptera: Scarabaeidae). *Bonn Zoological Bulletin, Supplements* 61, 1–124.

Farrell, B. (1998). "Inordinate Fondness" explained: Why are there so many beetles? *Science* 281, 555–559.

Farris, J. S. (1970). Methods for Computing Wagner Trees. *Systematic Zoology* 19, 83–92.

Gheerbrant, E. and J.-C. Rage (2006). Paleobiogeography of Africa: How distinct from Gondwana and Laurasia? *Palaeogeography, Palaeoclimatology, Palaeoecology* 241, 224–246. DOI: 10.1016/j.palaeo.2006.03.016.

Givnish, T. J. and S. S. Renner (2004). Tropical Intercontinental Disjunctions: Gondwana Breakup, Immigration from the Boreotropics, and Transoceanic Dispersal. *International Journal of Plant Sciences* 165, S1–S6. DOI: 10.1086/424022.

Goldblatt, P. and J. Manning (2000). *Cape plants – a conspectus of the Cape flora of South Africa*. Ed. by P. Goldblatt and J. C. Manning. Pretoria: National Botanical Institute of South Africa, p. 743.

Guindon, S., J.-F. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology* 59, 307–321. DOI: 10.1093/sysbio/syq010.

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, 353–431. DOI: 10.1016/S1367-9120(01)00069-4.

Hanazawa, M., H. Narushima, and N. Minaka (1995). Generating most parsimonious reconstructions on a tree: A generalization of the Farris-Swofford-Maddison method. *Discrete Applied Mathematics* 56, 245–265. DOI: 10.1016/0166-218X(94)00089-V.

Hardin, G. (1960). The competitive exclusion principle. *Science (New York, N.Y.)* 131, 1292–1297. DOI: 10.1126/science.131.3409.1292.

Heath, T. A., S. M. Hedtke, and D. M. Hillis (2008). Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution* 46, 239–257. DOI: 10.3724/SP.J.1002.2008.08016.

Ho, S. Y. W. and R. Lanfear (2010). Improved characterisation of among-lineage rate variation in cetacean mitogenomes using codon-partitioned relaxed clocks. *Mitochondrial DNA* 21, 138–146. DOI: 10.3109/19401736.2010.494727.

Ho, S. Y. W. and M. J. Phillips (2009). Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic biology* 58, 367–380. DOI: 10.1093/sysbio/syp035.

Holt, B. G., J.-P. Lessard, M. K. Borregaard, S. a. Fritz, M. B. Araújo, D. Dimitrov, P.-H. Fabre, C. H. Graham, G. R. Graves, K. a. Jønsson, D. Nogués-Bravo, Z. Wang, R. J. Whittaker, J. Fjeldså, and C. Rahbek (2013). An update of Wallace's zoogeographic regions of the world. *Science (New York, N.Y.)* 339, 74–78. DOI: 10.1126/science.1228282.

Howden, H. F. (1966). Some possible effects of the Pleistocene on the distributions of North American Scarabaeidae (Coleoptera). *The Canadian Entomologist* 98, 1177–1190. DOI: 10.4039/Ent981177-11.

Hunter, J. P. (1998). Key innovations and the ecology of macroevolution. *Trends in ecology & evolution* 13, 31–6.

Huntley, B., G. F. Midgley, P. Barnard, and P. J. Valdes (2014). Suborbital climatic variability and centres of biological diversity in the Cape region of southern Africa. *Journal of Biogeography* 41. Ed. by J. Stewart, 1338–1351. DOI: 10.1111/jbi.12288.

Karanth, P. (2006). Out-of-India Gondwanan origin of some tropical Asian biota. *Current Science* 90, 789–792.

Katoh, K. and H. Toh (2008). Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics* 9, 286–298. DOI: 10.1093/bib/bbn013.

— (2010). Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics* 26, 1899–1900. DOI: 10.1093/bioinformatics/btq224.

Kodandaramaiah, U. and N. Wahlberg (2007). Out-of-Africa origin and dispersal-mediated diversification of the butterfly genus *Junonia* (Nymphalidae: Nymphalinae). *Journal of Evolutionary Biology* 20, 2181–2191. DOI: 10.1111/j.1420-9101.2007.01425.x.

Kodandaramaiah, U., D. C. Lees, C. J. Müller, E. Torres, K. P. Karanth, and N. Wahlberg (2010). Phylogenetics and biogeography of a spectacular Old World radiation of butterflies: the subtribe Mycalesina (Lepidoptera: Nymphalidae: Satyrini). *BMC evolutionary biology* 10, 172. DOI: 10.1186/1471-2148-10-172.

Kodandaramaiah, U. and N. Wahlberg (2009). Phylogeny and biogeography of *Coenonympha* butterflies (Nymphalidae: Satyrinae) - patterns of colonization in the Holarctic. *Systematic Entomology* 34, 315–323.

Krell, F. T. (2000). The fossil record of Mesozoic and Tertiary Scarabaeoidea (Coleoptera: Polyphaga). *Invertebrate Taxonomy* 14, 871–905. DOI: 10.1071/IT00031.

Landis, M. J., N. J. Matzke, B. R. Moore, and J. P. Huelsenbeck (2013). Bayesian analysis of biogeography when the number of areas is large. *Systematic Biology* 62, 789–804. DOI: 10.1093/sysbio/syt040.

Lanfear, R., B. Calcott, S. Y. W. Ho, and S. Guindon (2012). PartitionFinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. *Molecular biology and evolution* 29, 1695–1701. DOI: 10.1093/molbev/mss020.

Lanfear, R., B. Calcott, D. Kainer, C. Mayer, and A. Stamatakis (2014). Selecting optimal partitioning schemes for phylogenomic datasets. *Evolutionary Biology* 14, 1–14. DOI: 10.1186/1471-2148-14-82.

Lawes, M. J., H. A. C. Eeley, N. J. Findlay, and D. Forbes (2007). Resilient forest faunal communities in South Africa: A legacy of palaeoclimatic change and extinction filtering? *Journal of Biogeography* 34, 1246–1264. DOI: 10.1111/j.1365-2699.2007.01696.x.

Linder, H. P., H. M. de Klerk, J. Born, N. D. Burgess, J. Fjeldså, and C. Rahbek (2012). The partitioning of Africa: Statistically defined biogeographical regions in sub-Saharan Africa. *Journal of Biogeography* 39, 1189–1205. DOI: 10.1111/j.1365-2699.2012.02728.x.

Linder, H. (2008). Plant species radiations: where, when, why? *Philosophical Transactions of the Royal Society B: Biological Sciences* 363, 3097–3105. DOI: 10.1098/rstb.2008.0075.

Liu, K., T. J. Warnow, M. T. Holder, S. M. Nelesen, J. Yu, A. P. Stamatakis, and C. R. Linder (2012). SATe-II: Very Fast and Accurate Simultaneous Estimation of Multiple Sequence Alignments and Phylogenetic Trees. *Systematic Biology* 61, 90–106. DOI: 10.1093/sysbio/syr095.

Liu, W.-G., J. Eberle, M. Bai, X.-K. Yang, and D. Ahrens (2015). A phylogeny of Sericini with particular reference to Chinese species using mitochondrial and ribosomal DNA (Coleoptera: Scarabaeidae). *Organisms Diversity & Evolution* 15, 343–350. DOI: 10.1007/s13127-015-0204-z.

Liu, W.-g., S. Fabrizi, M. Bai, X.-k. Yang, and D. Ahrens (2014). A taxonomic revision of the *Neoserica* (sensu lato) *pilosula* group (Coleoptera, Scarabaeidae, Sericini). *ZooKeys* 440, 89–113. DOI: 10.3897/zookeys.440.8126.

Lorenzen, E. D., R. Heller, and H. R. Siegismund (2012). Comparative phylogeography of African savannah ungulates. *Molecular Ecology* 21, 3656–3670. DOI: 10.1111/j.1365-294X.2012.05650.x.

Lorenzen, E. D., C. Masembe, P. Arctander, and H. R. Siegismund (2010). A long-standing Pleistocene refugium in southern Africa and a mosaic of refugia in East Africa: Insights from mtDNA and the common eland antelope. *Journal of Biogeography* 37, 571–581. DOI: 10.1111/j.1365-2699.2009.02207.x.

Machatschke, J. W. (1959). Phylogenetische Untersuchungen über die Sericini (sensu Dalla Torre 1912) (Coleoptera: Lamellicornia: Melolonthidae). *Beiträge zur Entomologie1* 9, 730–746.

Matzke, N. J. (2014). Model Selection in Historical Biogeography Reveals that Founder-event Speciation is a Crucial Process in Island Clades. *Systematic Biology* 63, 951–970. DOI: 10.1093/sysbio/syu056.

Matzke, N. J. (2012). Founder-event speciation in BioGeoBEARS package dramatically improves likelihoods and alters parameter inference in Dispersal-Extinction-Cladogenesis (DEC) analyses. *Frontiers of Biogeography* 4, 210.

— (2013a). "Probabilistic historical biogeography: new model for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing." PhD thesis. University of California, Berkeley (CA).

Matzke, N. J. (2013b). Probabilistic historical biogeography: new model for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers of Biogeography* 5, 251.

McGlone, M. S. (2005). Goodbye Gondwana. *Journal of Biogeography* 32, 739–740. DOI: 10.1111/j.1365-2699.2005.01278.x.

Mehrotra, R. C. (2003). Status of plant megafossils during the early Paleogene in India. *Special papers - Geological society of America* 369, 413–423. DOI: 10.1130/0-8137-2369-8.413.

Monaghan, M. T., D. J. G. Inward, T. Hunt, and A. P. Vogler (2007). A molecular phylogenetic analysis of the Scarabaeinae (dung beetles). *Molecular Phylogenetics and Evolution* 45, 674–692. DOI: 10.1016/j.ympev.2007.06.009.

Morrison, D. A. (2008). How to summarize estimates of ancestral divergence times. *Evolutionary bioinformatics online* 4, 75–95.

Mourer-Chauviré, C. (1999). Les relations entre les avifaunes du Tertiaire inférieur d'Europe et d'Amérique du Sud. *Bull. Soc. Géol. Fr.* 170, 85–90.

Paradis, E., J. Claude, and K. Strimmer (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290.

Philippe, H., H. Brinkmann, D. V. Lavrov, D. T. J. Littlewood, M. Manuel, G. Wörheide, and D. Baurain (2011). Resolving difficult phylogenetic questions: Why more sequences are not enough. *PLoS Biology* 9, e1000602. DOI: 10.1371/journal.pbio.1000602.

Plummer, M., N. Best, K. Cowles, and K. Vines (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News* 6, 7–11. DOI: 10.1159/000323281.

Price, M. N., P. S. Dehal, and A. P. Arkin (2010). FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5, e9490. DOI: 10.1371/journal.pone.0009490.

Proches, Ş. and S. Ramdhani (2012). The World's Zoogeographical Regions Confirmed by Cross-Taxon Analyses. *BioScience* 62, 260–270. DOI: 10.1525/bio.2012.62.3.7.

Pyron, R. A. (2014). Biogeographic Analysis Reveals Ancient Continental Vicariance and Recent Oceanic Dispersal in Amphibians. *Systematic biology* 63, 779–797. DOI: 10.1093/sysbio/syu042.

Pyron, R. A. and J. J. Wiens (2011). A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. *Molecular phylogenetics and evolution* 61, 543–583. DOI: 10.1016/j.ympev.2011.06.012.

Rabosky, D. L. (2014). Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS ONE* 9, e89543. DOI: 10.1371/journal.pone.0089543. arXiv: 1401.6602.

Rabosky, D. L., S. C. Donnellan, M. Grundler, and I. J. Lovette (2014a). Analysis and visualization of complex Macroevolutionary dynamics: An example from Australian Scincid lizards. *Systematic Biology* 63, 610–627. DOI: 10.1093/sysbio/syu025.

Rabosky, D. L., M. Grundler, C. Anderson, P. Title, J. J. Shi, J. W. Brown, H. Huang, and J. G. Larson (2014b). BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution* 5, 701–707. DOI: 10.1111/2041-210X.12199.

Rabosky, D. L., F. Santini, J. Eastman, S. A. Smith, B. Sidlauskas, J. Chang, and M. E. Alfaro (2013). Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nature Communications* 4, 1958. DOI: 10.1038/ncomms2958.

Rage, J. (1988). "Gondwana, Tethys, and terrestrial vertebrates during the Mesozoic and Cainozoic." In: *Gondwana and Tethys. Geol. Soc. Spec. Paper, vol. 37*. Ed. by M. Audley-Charles and A. Hallam. Oxford Univ. Press, pp. 235–273.

Rambaut, A., M. Suchard, D. Xie, and A. Drummond (2014). Tracer v1.6.

Ramstein, G., F. Fluteau, J. Besse, and S. Joussaume (1997). Effect of orogeny, plate motion and land–sea distribution on Eurasian climate change over the past 30 million years. *Nature* 386, 788–795. DOI: 10.1038/386788a0.

Ree, R. H. and S. A. Smith (2008). Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic biology* 57, 4–14. DOI: 10.1080/10635150701883881.

Ronquist, F. (1997). Dispersal-Vicariance Analysis: A New Approach to the Quantification of Historical Biogeography. *Systematic Biology* 46, 195–203. DOI: 10.2307/2413643.

Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. a. Suchard, and J. P. Huelsenbeck (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology* 61, 539–42. DOI: 10.1093/sysbio/sys029.

Samonds, K. E., L. R. Godfrey, J. R. Ali, S. M. Goodman, M. Vences, M. R. Sutherland, M. T. Irwin, and D. W. Krause (2012). Spatial and temporal arrival patterns of Madagascar's vertebrate fauna explained by distance, ocean currents, and ancestor type. *Proceedings of the National Academy of Sciences* 109, 5352–5357. DOI: 10.1073/pnas.1113993109.

Sanmartín, I. and F. Ronquist (2004). Southern hemisphere biogeography inferred by event-based models: plant versus animal patterns. *Systematic biology* 53, 216–243. DOI: 10.1080/10635150490423430.

Santos, C. M. D. and D. S. Amorim (2007). Why biogeographical hypotheses need a well supported phylogenetic framework: a conceptual evaluation. *Papéis Avulsos de Zoologia* 47, 63–73.

Schaefer, H. and S. S. Renner (2008). A phylogeny of the oil bee tribe Ctenoplectrini (Hymenoptera: Anthophila) based on mitochondrial and nuclear data: evidence for early Eocene divergence and repeated out-of-Africa dispersal. *Molecular phylogenetics and evolution* 47, 799–811. DOI: 10.1016/j.ympev.2008.01.030.

Schliep, K. P. (2011). phangorn: Phylogenetic analysis in R. *Bioinformatics* 27, 592–593. DOI: 10.1093/bioinformatics/btq706.

Schnitzler, J., T. G. Barraclough, J. S. Boatwright, P. Goldblatt, J. C. Manning, M. P. Powell, T. Rebelo, and V. Savolainen (2011). Causes of plant diversification in the cape biodiversity hotspot of South Africa. *Systematic Biology* 60, 343–357. DOI: 10.1093/sysbio/syr006.

Schulte, P., L. Alegret, I. Arenillas, J. a. Arz, P. J. Barton, P. R. Bown, T. J. Bralower, G. L. Christeson, P. Claeys, C. S. Cockell, G. S. Collins, A. Deutsch, T. J. Goldin, K. Goto, J. M. Grajales-Nishimura, R. a. F. Grieve, S. P. S. Gulick, K. R. Johnson, W. Kiessling, C. Koeberl, D. a. Kring, K. G. MacLeod, T. Matsui, J. Melosh, A. Montanari, J. V. Morgan, C. R. Neal, D. J. Nichols, R. D. Norris, E. Pierazzo, G. Ravizza, M. Rebolledo-Vieyra, W. U. Reimold, E. Robin, T. Salge, R. P. Speijer, A. R. Sweet, J. Urrutia-Fucugauchi, V. Vajda, M. T. Whalen, and P. S. Willumsen (2010). The Chicxulub asteroid impact and mass extinction at the Cretaceous-Paleogene boundary. *Science (New York, N.Y.)* 327, 1214–1218. DOI: 10.1126/science.1177265.

Sechrest, W., T. M. Brooks, G. A. B. da Fonseca, W. R. Konstant, R. A. Mittermeier, A. Purvis, A. B. Rylands, and J. L. Gittleman (2002). Hotspots and the conservation of evolutionary history. *Proceedings of the National Academy of Sciences of the United States of America* 99, 2067–2071. DOI: 10.1073/pnas.251680798.

Shi, J. J. and D. L. Rabosky (2015). Speciation dynamics during the global radiation of extant bats. *Evolution* 69, 1528–1545. DOI: 10.1111/evo.12681.

Shimodaira, H. and M. Hasegawa (2001). CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics applications note* 17, 1246–1247.

Siesser, W. G. (1980). Late Miocene Origin of the Benguela Upswelling System off Northern Namibia. *Science (New York, N.Y.)* 208, 283–285. DOI: 10.1126/science.208.4441.283.

Simon, C., F. Frati, A. 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. *Annals of the Entomological Society of America* 87, 651–701.

Stamatakis, A. (2006). Phylogenetic models of rate heterogeneity: A high performance computing perspective. *20th International Parallel and Distributed Processing Symposium, IPDPS 2006*, 8 pp. DOI: 10.1109/IPDPS.2006.1639535.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxford, England)* 30, 1312–1333. DOI: 10.1093/bioinformatics/btu033.

Struempher, W. P., C. L. Sole, M. H. Villet, and C. H. Scholtz (2014). Phylogeny of the family Trogidae (Coleoptera: Scarabaeoidea) inferred from mitochondrial and nuclear ribosomal DNA sequence data. *Systematic Entomology* 39, 548–562.

Stuut, J. B. W., X. Crosta, K. van der Borg, and R. Schneider (2004). Relationship between Antarctic sea ice and southwest African climate during the late Quaternary. *Geology* 32, 909–912. DOI: 10.1130/G20709.1.

Switala, A. K., C. L. Sole, and C. H. Scholtz (2014). Phylogeny, historical biogeography and divergence time estimates of the genus *Colophon* Gray (Coleoptera: Lucanidae). *Invertebrate Systematics* 28, 326–336.

Swofford, D. L. and W. P. Maddison (1987). Reconstructing ancestral character states under Wagner parsimony. *Mathematical Biosciences* 87, 199–229. DOI: 10.1016/0025-5564(87)90074-5.

Tapponnier, P., X. Zhiqin, F. Roger, B. Meyer, N. Arnaud, G. Wittlinger, and Y. Jingsui (2001). Oblique Stepwise Rise and Growth of the Tibet Plateau. *Science* 294, 1671–1677. DOI: 10.1126/science.105978.

Taquet, P. (1982). Une connexion continentale entre afrique et madagascar au crétacé supérieur: données géologiques et paléontologiques. *Geobios* 15, 385–391. DOI: 10.1016/S0016-6995(82)80127-7.

Timmermans, M. J. T. N., S. Dodsworth, C. L. Culverwell, L. Bocak, D. Ahrens, D. T. J. Littlewood, J. Pons, and a. P. Vogler (2010). Why barcode? High-throughput multiplex sequencing of mitochondrial genomes for molecular systematics. *Nucleic acids research* 38, e197. DOI: 10.1093/nar/gkq807.

Tolley, K. A., T. M. Townsend, and M. Vences (2013). Large-scale phylogeny of chameleons suggests African origins and Eocene diversification. *Proceedings of the Royal Society B: Biological Sciences* 280, 20130184. DOI: 10.5061/dryad.11350.

Varón, A., L. S. Vinh, and W. C. Wheeler (2010). POY version 4: Phylogenetic analysis using dynamic homologies. *Cladistics* 26, 72–85. DOI: 10.1111/j.1096-0031.2009.00282.x.

Vences, M., D. R. Vieites, F. Glaw, H. Brinkmann, J. Kosuch, M. Veith, and A. Meyer (2003). Multiple overseas dispersal in amphibians. *Proceedings of the Royal Society of London B: Biological Sciences* 270, 2435–2442.

Verboom, G. A., H. P. Linder, F. Forest, V. Hoffmann, N. G. Bergh, and R. M. Cowling (2014). "Cenozoic assembly of the greater Cape flora." In: *Fynbos: Ecology, Evolution, and Conservation of a Megadiverse Region*. Ed. by N. Allsopp, J. F. Colville, and G. A. Verboom. 1st ed. New York: Oxford University Press. Chap. 5, pp. 93–118.

Wahlberg, N., J. Leneveu, U. Kodandaramaiah, C. Peña, S. Nylin, A. V. L. Freitas, and A. V. Z. Brower (2009). Nymphalid butterflies diversify following near demise at the Cretaceous/Tertiary boundary. *Proceedings of the Royal Society B* 276, 4295–4302. DOI: 10.1098/rspb.2009.1303.

Wallace, A. R. (1876). *The Geographical Distribution of Animals*. Cambridge: Cambridge Univ. Press.

Waters, J. M. (2011). Competitive exclusion: Phylogeography's 'elephant in the room'? *Molecular Ecology* 20, 4388–4394. DOI: 10.1111/j.1365-294X.2011.05286.x.

Waters, J. M. and D. Craw (2006). Goodbye Gondwana? New Zealand biogeography, geology, and the problem of circularity. *Systematic biology* 55, 351–356. DOI: 10.1080/10635150600681659.

Wickham, H. F. (1912). A report on some recent collections of fossil Coleoptera from the Miocene shales of Florissant. *Bulletin of the State University of Iowa, Bulletin from the Laboratories of Natural History* 6, 3–38.

Zachos, J., M. Pagani, L. Sloan, E. Thomas, and K. Billups (2001). Trends, Rhythms, and Aberrations in Global Climate 65 Ma to Present. *Science* 292, 686–693. DOI: 10.1126/science.1059412.

Chapter VIII.

Landscape genetics indicate recently increased habitat fragmentation in African forest-associated chafers

This chapter is published in:
Authors: Eberle J., D. Rödder, M. Beckett, D. Ahrens (2017). Landscape genetics indicate recently increased habitat fragmentation in African forest-associated chafers. *Global Change Biology* 23, 1988–2004. DOI: 10.1111/gcb.13616

Authors' contributions to the original article:
Molecular lab work, sequence assembly and alignment, population genetic analyses, habitat connectivity analyses (spatial and climatic model based): JE; distribution data: JE, MB; climatic niche modeling: DR; manuscript concept and design: JE, DA; drafting, revising, and approving the manuscript: JE, DR, DA.

1. Introduction

Forests cover about one quarter of earth's surface (Bartholomé and Belward 2005) but harbor more than half of terrestrial biodiversity (Millennium Ecosystem Assessment 2005a). They provide important ecosystem services on a global and a regional scale, including climate regulation, carbon storage, and erosion control (Foley et al. 2005; Millennium Ecosystem Assessment 2005b; Newbold et al. 2015). Although there is evidence that South Africa's land surface was covered by extended forests in the past (Deacon et al. 1983), it is today predominantly covered by open habitats like grassland, Fynbos, Karoo, and savannah biomes (Mucina and Rutherford 2006; Huntley et al. 2016). Less than 0.6% of the area of South Africa is covered by indigenous forests (Low and Rebelo 1996; Mucina and Rutherford 2006) which are predominantly found along the Great Escarpment as well as the South and South-Eastern coasts. The vast majority of the forests is highly fragmented with 78.5% of the recorded forest patches being smaller than 1 km^2 (estimate based on data of Mucina and Rutherford 2006). Diversity and patchiness of the forests imply their relictual character (Deacon et al. 1983; Mucina and Rutherford 2006). Although past climatic fluctuations in southern Africa repeatedly caused forests to expand and retreat again (Geldenhuys 1997; Dupont et al. 2011; Huntley et al. 2016), anthropogenic influence is major cause for an extreme forest retreat within the last centuries (Fourcade 1889; Bews 1913, 1920; King 1941; Acocks 1953; Bond et al. 2003). Records of first explorers (e.g., Vasco da Gama 15^{th} century; Ravenstein 1898) might be interpreted in a way that coastal indigenous forests were widely expanded. Also field surveys, palaeo-environmental evidence, and climatic niche modeling proposed that areas in south-eastern South Africa had and have the potential to be widely covered by forest (Acocks 1953; Eeley et al. 1999; Bond et al. 2003; Chase and Meadows 2007; Quick et al. 2011).

Grasslands are thought to have expanded at the expense of forests in many parts of the world since they withstand conditions that limit the establishment or survival of

woody species (Bond 2008; Edwards et al. 2010). Besides grazing of large herbivores and xeric climatic conditions, fires are one of the most important natural drivers of vegetation structure in southern Africa (Phillips 1930; Little et al. 2013), but man-made fire regimes with more frequent burning of smaller areas have replaced natural ones (Deacon 1983; Archibald et al. 2013) and sometimes even threat fire-adapted ecosystems (Reside et al. 2012). There is evidence that man used fire for vegetation management for at least 100 ky (Deacon 1983; Deacon and Deacon 1999) and cleared large portions of forest particularly at the east coast of southern Africa (Fourcade 1889; Bews 1913, 1920; King 1941; Acocks 1953; Castley and Kerley 1996). When Vasco da Gama sailed along the southern African coast in 1498, burning fires along the coast probably led him to name places "Ponta das Queimadas" (Gulf of forest fires, St. Francis Bay) or "Terra dos Fumos" (Land of smoke, around todays Maputo) (Ravenstein 1898; Bews 1913). Today, native forest-ecosystems are invaded by fire-adapted alien plants (Brooks et al. 2004) and agricultural areas are retained by fires. Currently, South African conservation management controls alien invaders' expansion by burning of uprising woodland to preserve fire adapted Fynbos and grassland biomes (Wilgen 2009; Wilgen et al. 2012), which also inhibits restoration of indigenous forests (Luger and Moll 1993). The crucial question in this context is, whether or not and to which extent indigenous forest is the potential natural vegetation (vs. grassland). Besides that, indigenous forests are also threatened by their exploitation to satisfy needs for building material, fuel wood, food, and medicine (Mucina and Rutherford 2006).

Most forest plants and associated insects disappear quite abruptly with forest clearance. However, depending on land use intensity and abiotic factors like shading, forest soils retain their original properties for several years (Balesdent et al. 1988; Lemenih et al. 2005) and thus stay suitable for most of its soil fauna. One such element of soil fauna is *Pleophylla*, a genus of soil dwelling scarab chafers (Coleoptera: Scarabaeidae) which occurs predominantly in the isolated forest patches throughout South Africa. It expands with a few species in the Afromontane forests along the Eastern Arc up north to Uganda and D.R. Congo (Eberle et al. 2016a), a pattern that is also observed in other forest associated species (e.g., Huber 2003). It is one of the oldest lineages of the highly diverse tribe of Sericini (Ahrens 2006; Eberle et al. 2016b) that originated ca 79 Mya and showed a burst of speciation since the Miocene (Eberle et al.

2016c). Most available records of *Pleophylla* are located in or in close vicinity to forest remnants; therefore the genus is suspected to be forest associated. However, their polyphagous feeding habits makes them quite independent from specific forest plant species: the fully winged adults of *Pleophylla* feed, as most Sericini, polyphagously on leaves of a variety of angiosperms including many allochthonous ones, while their larvae develop in the upper soil strata feeding on organic matter and plant roots. Therefore, these beetles can still exist when primary indigenous forest plants have gone and thus may serve as a proxy for past or potential forest distribution in South Africa.

In the present study, we address the question on the natural extent of South African forests by combining population genetics, spatial distribution, and climatic niche modeling (Fig. VIII.1a) of populations of nine species of *Pleophylla*, to assess present and past connectivity of forest habitats. We investigated the degree of genetic isolation of populations since the last glacial period, using fast evolving mitochondrial and nuclear loci. Good genetic admixture is expected if forests are currently or were recently more extended or at least well connected through migration corridors. The opposite (i.e., poor admixture and extreme endemism) is expected for poorly connected, long term isolated and little extended forests. The demographic histories of the species were examined in order to detect population size alterations that might be related to habitat expansion or fragmentation. Current and past distribution models of the species were used to infer potential fluctuations of distribution ranges which might be linked to forest distribution and to further explore the hypothesis of historically more extended forest. Genetic population structure inference was backed by landscape connectivity analyses based on species distribution models (SDM) inferred from climatic data. Three different scenarios were evaluated by using the unrestricted SDM (F0), by additional consideration of forest patches (forest-accounting SDM; F1), and by excluding potentially suitable soils for *Pleophylla* larvae outside forest occurrences (forest-restricted SDM; F2) (Fig. VIII.1). Knowledge about the connectivity among *Pleophylla* and other forest dwelling species' populations (Fig. VIII.1a, F1) appears to be crucial to identify areas for high-priority conservation of forest faunas. Based on the combined evidence from *Pleophylla*, we discuss the potential distribution of forests and the anthropogenic influence on forest habitats in South Africa.

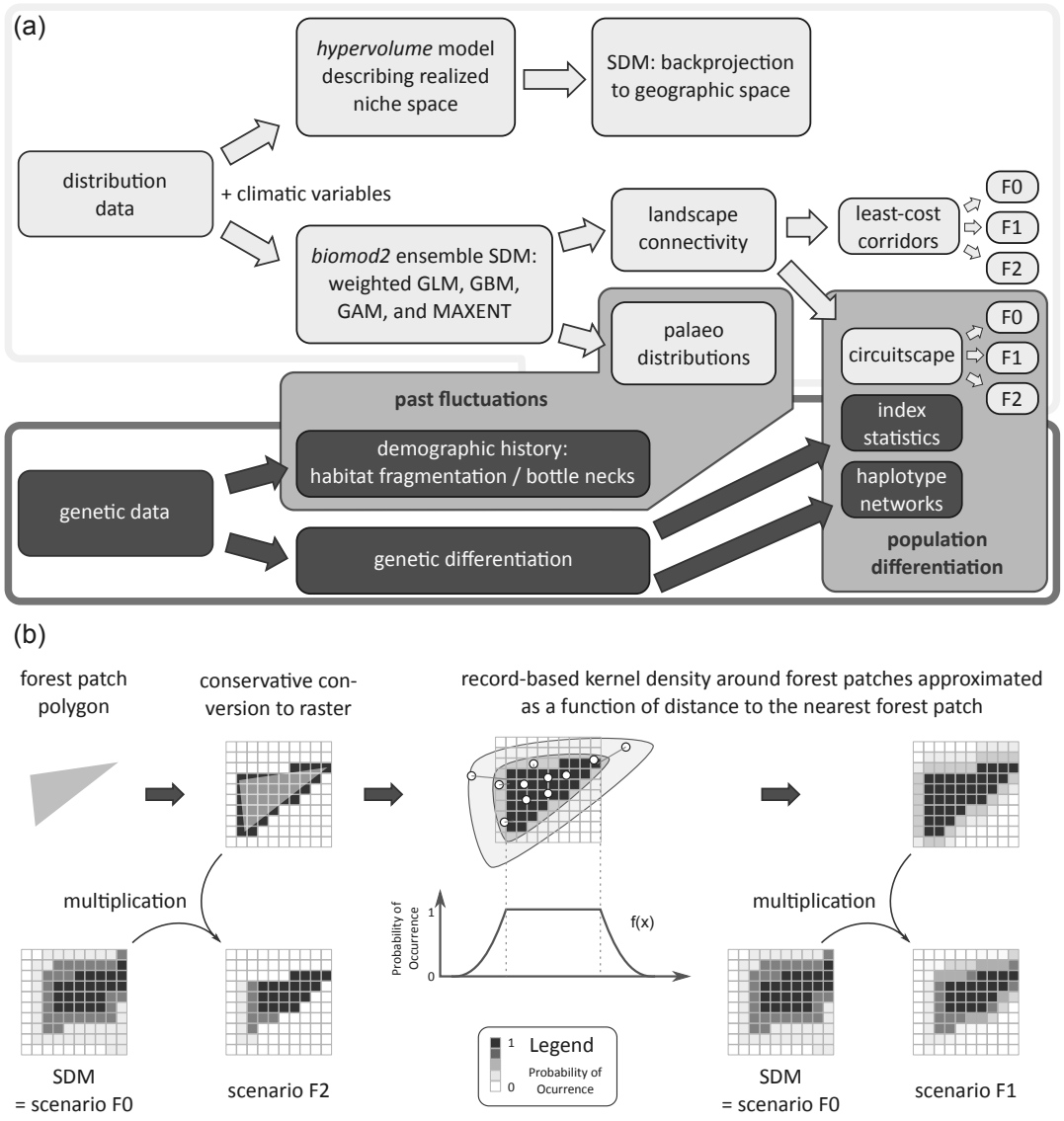

Figure VIII.1. Integration of climatic niche modeling and molecular analyses. (a) Flow chart illustrating batched analysis steps in two lines of evidence (dark grey: genetic and light grey: niche modeling). Medium grey background boxes highlight complementary results. (b) Diagrammed steps to SDMs based on actual forest distribution (scenarios F1 and F2). Please refer to the main text for detailed explanation.

2. Material and Methods

2.1. Sampling and assessment of forest association

The distribution data of *Pleophylla* included 319 specimens of which DNA-data was available (Tables G24, G1) and 828 dry specimens from 12 different museum collections (Eberle et al. 2016a). Most specimens were collected with light traps at 172 unique localities. They were identified by examining the dissected male genitalia, females partly through match of species-specific DNA markers (Eberle et al. 2016a,c). Collection locations of museum specimens without GPS data were localized using the GeoNames geographical database (`www.geonames.org`) and Google Maps (`www.google.com/maps`). The forest-association hypothesis of *Pleophylla* was assessed using the distribution data of all southern African Sericini species (20,000 specimens comprising ca 400 species; Eberle et al., unpublished data). Collection sites, where no species of *Pleophylla* were recorded, are highly informative in this context, since all Sericini are collected with the same method during the same season (predominantly with light traps). Therefore, these localities can be likely considered absence data on a rough scale. The forest association of *Pleophylla* is a fundamental assumption of this study. Its reliability was evaluated by measuring the distance of the sampling points from forest patch polygons of Mucina and Rutherford (2006) in QGIS (v2.10, `www.qgis.org`, NNjoin plugin). Eight records of *Pleophylla* specimens with spuriously large distances to forest (>20 km) were checked on satellite images (CNES/Astrium, Feb. and Apr. 2016; DigitalGlobe®, Sep. 2014; accessed via Google Earth) and not considered in this analysis, since very small forest patches were found nearby that were not digitized by Mucina and Rutherford (2006). An analysis of variance (ANOVA) of all specimens of *Pleophylla*, all other southern African Sericini, and 1,000 randomly distributed points in South Africa was performed in R (R Devel-

opment Core Team 2015), followed by pairwise t-tests; p-values were adjusted by the Holm method (Holm 1979).

2.2. Analyses of genetic variation

Specimen collection, preservation, and DNA extraction followed Ahrens and Vogler (2008). A total of 347 specimens were used for molecular analyses. Qiagen®Multiplex PCR Kits were used to amplify the 3‘ end of cytochrome oxidase subunit 1 (*cox1*) with the primer pairs stevPat and stevJerry (Timmermans et al. 2010). Primer pairs ITS1F and ITS1R (Hillis and Dixon 1991; Vogler and Desalle 1994) were used for the amplification of the internal transcribed spacer ITS-1. Sequencing was done by Macrogen (Seoul, South Korea). Sequences were aligned per marker using MAFFT (v7.017; Katoh et al. 2002) and subsequently checked by eye in Geneious® 7.1.8. All DNA voucher specimens are deposited in the collections of the Zoological Research Museum A. Koenig, Bonn (ZFMK). GenBank Accessions are listed in Electronic Supplement Table G24. The genetic variation of South African *Pleophylla* species was investigated on the basis of haplotype networks in *pegas* (v0.8-2; Paradis 2010; R Development Core Team 2015. Distances between haplotypes were calculated under pairwise deletion of missing data. The networks were colored according to the sampling localities in order to visualize genetic differentiation in geographical context. Spatial isolation might indicate limited gene flow since the last glacial due to range fragmentation (Templeton et al. 2001). Therefore we calculated the standardized measure of genetic differentiation G'_{ST} (Hedrick 2005) between the sampling localities of each species separately for *cox1* (826 bp length) and ITS1 (853 bp length) using *diveRsity* (v1.9.73; Keenan et al. 2013).

Genetic bottlenecks may occur by strong reduction or fragmentation of a species range and leave traces in the DNA of populations. Indications for such reductions in *Pleophylla*, most likely linked to the loss of habitat (i.e., forests), were inferred by the reconstruction of the species' demographic histories using Extended Bayesian Skyline Plots (EBSP; Heled and Drummond 2008) of sufficiently sampled species (i.e., with a sample size > 10 specimens). The analyses were conducted with BEAST (v1.8.1; Drummond et al. 2012) on both markers. The substitution models, clocks, and trees were unlinked for *cox1* and ITS1 partitions. Optimal substitution models

were inferred with PartitionFinder (Lanfear et al. 2012, 2014) (Table G2). Since low levels of rate variation are expected in intraspecific data sets, a strict molecular clock was used for each partition (Brown and Yang 2011). The analyses were time calibrated by setting the pairwise divergence rate of *cox1* to 3.54% My^{-1} (Papadopoulou et al. 2010). The rate of ITS1 was estimated relative to that rate under a uniform prior. The analyses were all run twice for 60 million generations and subsequently combined with *burntrees* (v0.2.2; Nylander 2014) after removing a burnin of 10%. Stationarity of repeated runs at similar values and convergence was assessed with Tracer (v1.6.0; Rambaut et al. 2014) before conducting the demographic analyses on the combined output.

2.3. Species distribution modeling

Environmental predictors for species distribution models (SDMs) were compiled from a set of 19 bioclimatic variables (`http://www.worldclim.org/bioclim`) at a spatial resolution of 30 arc seconds available from WorldClim (Hijmans et al. 2005). In order to restrict the overall environmental background, a bioclim model (Busby 1991) based on all sampling points of *Pleophylla* was calculated and original variables were clipped to it (Table G4). This initial step was necessary to reduce computation time without omitting potentially suitable areas, as hypervolume models (see below) are per definition nested within an overall bioclim model. Following Blonder et al. (2014), a spatial principle component analysis was performed based on the clipped background in order to create an orthogonal niche space, only retaining components with Eigenvalues > 1. This step is crucial as the hypervolume analyses require an orthogonal parameter space in order to avoid pseudo-replication.

Niches were quantified following Hutchinson's original niche concept of n-dimensional hypervolumes (Hutchinson 1957) enclosing all environmental conditions which allow infinite existence of populations. Recently Blonder et al. (2014; Blonder 2015) provided the R package *hypervolume* (Blonder 2015; R Development Core Team 2015) allowing for the first time to compute even high dimensional hypervolumes that are based on multidimensional kernel density estimators to derive a density distribution of species records in PCA space. This density distribution is used to compute the total volume of the species' realized niche space and allows geometric operations of

multiple hypervolumes including intersection, unique proportions etc. (Blonder et al. 2014). These hypervolumes were projected back in geographic space indicating those geographic areas that provide suitable conditions for populations of the species. However, with increasing dimensionality the required minimum number of species records increases with this method exponentially. Therefore, it was possible to include in this analysis only nine of the 13 species with a minimum of 5 unique sampling locations. For comparisons between hypervolumes the Sørensen Index was calculated as a measure of niche overlap (Sørensen 1948) based on shared and unique proportions of two hypervolumes following Blonder et al. (2014).

Evidence for past distribution and landscape connectivity of *Pleophylla* species came from *biomod2* ensemble SDMs (Thuiller 2003; Thuiller et al. 2013) which were modeled using a subset of the 19 previously mentioned bioclimatic variables: in order to remove possible negative effects of spatial autocorrelation, inter-correlation structure among the variables throughout the study area was assessed by computing pairwise squared Spearman's rank correlation coefficients. In cases, where r^2 exceeded 0.75, only the putatively biologically most important variables were chosen. Using this strategy the following variables were retained: mean diurnal range (BIO2), temperature annual range (BIO7), mean temperature of warmest quarter (BIO10), mean temperature of coldest quarter (BIO11), annual precipitation (BIO12), precipitation of wettest quarter (BIO16), precipitation of driest quarter (BIO17), precipitation of warmest quarter (BIO18).

Modeling techniques employed in *biomod2* ensembles were the Generalized Linear Model (GLM), the Generalized Boosting Model (GBM), the Generalized Additive Model (GAM), and Maximum Entropy (MAXENT) (Thuiller 2003; Thuiller et al. 2013). As environmental background two different sets of each 10,000 pseudo-absence records were created within a circular buffer of 200 km enclosing the respective species records, but not closer than 100 km. We preferred to use pseudo-absences here as it is generally difficult to proof the absence of a *Pleophylla* species at a given site, especially given the varying degrees of sampling effort and focus taxa in the data set of Sericini species. All models were repeated 5 times for each set of pseudo-absences randomly splitting the species records in 80% used for model training and 20% used for model evaluation resulting in 40 single SDMs per species (2 × pseudo-absences × 4 algorithms × 5 repetitions). As evaluation measures we computed the area under

the receiver operating characteristic curve (ROC; Swets 1988), Cohen's Kappa and the True Skills Statistic (TSS) (Allouche et al. 2006). For the calculation of the final ensemble model the best fitting models (with ROC>0.7) were proportionally weighted according to their fit, as recommended in the *biomod2* manual (Thuiller et al. 2016). When projecting, areas requiring extrapolation beyond the environmental training range of the SDMs were discarded. Species with less than 24 spatially unique records were excluded from the *biomod2* approach (retaining *P. fasciatipennis*, *P. ferruginea*, *P. navicularis*, *P. nelshoogteensis*, and *P. pilosa*).

In order to evaluate past habitat expansion of *Pleophylla*, potential distributions of *Pleophylla* species in the Last Glacial Maximum (LGM, 21 kya) and the Holocene Altithermal (HA, 6 kya) were inferred. The species' weighted ensemble models were used to assess past potential distributions based on 11 different global circulation models of the Paleoclimate Modelling Intercomparison Project (PMIP) 3 (Braconnot et al., 2011, 2012): bcc-csm1-1, CCSM4, CNRM-CM5, COSMOS-ASO, CSIRO-Mk3-6-0, FGOALS-g2, GISS-E2-R, IPSL-CM5A-LR, MIROC-ESM, MPI-ESM-P, and MRI-CGCM3 (Electronic Supplement Table G25). Original monthly outputs of the global circulation models run with r1i1p1 initial conditions were downscaled to a resolution of 2.5 arc min (approximately 4 km in the study area) using the delta method proposed by Peterson and Nyári (2008). Subsequently, the respective bioclim variables were computed using the relevant functions of the *dismo* package for cran R (Hijmans et al. 2015).

2.4. Landscape connectivity analyses

Two methods for modeling the potential connectivity of the sampled *Pleophylla* populations and for identifying important dispersal corridors and pinch points were applied: *Circuitscape* (Shah and McRae 2008) and *Least Cost Corridors* (LCCs, Adriaensen et al. 2003; Verbeylen et al. 2003). *Circuitscape* adapts concepts from electric circuit theory since many parallels exist between organism movement and electric current flow (McRae 2006; Shah and McRae 2008). It is able to assess the amount of gene flow in complex landscapes and seems particularly suited to evaluate the isolation of populations since species movement (i.e., current) over long distances and high resistances is allowed to end by death of the moving individual (analog to ground-

ings). In contrast to *LCCs* models it can incorporate the effects of wider habitat swaths and of independent, parallel pathways connecting samples (McRae 2006). We used as current sources in the resistance landscape all sampling localities of a species which was modeled with *biomod2* (landscape resistance model F0). Additionally, two derivatives of F0 were used which were informed by the actual distribution of indigenous forests (Fig. VIII.1b). These derivatives were employed to model the distribution of *Pleophylla*, not only considering climatic factors but also the actual forest occurrences. The first forest-accounting derivate (F1) models the current potential distribution while the second (F2) describes a scenario without potentially suitable soils for *Pleophylla* larvae outside forest occurrences (forest-restricted). Therefore, the vector format forest patch polygons of Mucina and Rutherford (2006) were transformed into a binary raster layer of forest patches in QGIS (v2.10), by assigning all pixels the value 1 (i.e., forest present) which fully or partially overlapped the polygons (Fig. VIII.1b). This approach artificially enlarges the forest patches (i.e., the potential habitat of *Pleophylla*) slightly, leading to a more conservative approach of habitat fragmentation inference than the alternative approach of selecting only fully overlapped pixels would. To consider also the occurrence of *Pleophylla* individuals in yet humic soils outside forest patches (forest-accounting scenario F1), its probability of occurrence outside forests was approximated as a function of distance to the nearest forest patch based on a kernel density of all available sampling points (Fig. VIII.1b). For scenario F1, gradients of decreasing occurrence probability around forest patches (with values ranging from 1 to 0) were added to the above mentioned raster layer prior to multiplying with the *biomod2* ensemble SDMs. The forest-restricted scenario F2 was produced by multiplying the raster-layer without gradients of decreasing occurrence probability with the *biomod2* ensemble SDMs.

Least Cost Corridors (LCCs) and *Paths* (LCPs) were inferred with *SDMtoolbox* (v1.1c; Brown 2014) in ArcGIS® 10.2.2. Due to computational limitations, occurrence data of each species was spatially rarified with *SDMtoolbox*, i.e. records with high spatial autocorrelation were removed (Brown 2014). *Biomod2* SDMs and its derivatives were resampled to 50% of the original resolution. LCCs were calculated in a pairwise manner between sampling sites. The results were visualized with the *raster*-package in R (v2.4-15; Hijmans 2015; R Development Core Team 2015) and with QGIS.

An overview of the entire pipeline of species distribution modeling and landscape connectivity analyses is shown in Fig. VIII.1.

Additionally, landscape connectivity metrics based on forest patch characteristics were calculated with FRAGSTATS (McGarigal et al. 2012) employing the binary raster layer of indigenous forest patches (see above). Besides the number of forest patches and forest patch density, the edge to edge Euclidean distances between all nearest neighboring patches (McGarigal and Marks 1995), the connectance index (CI, McGarigal et al. 2012), and the degree of landscape division (LDI, Jaeger 2000) were calculated. The CI gives the percentage of pairwise patch-comparisons that are expected to be connected under a given threshold. The threshold was set to the maximum distance of all *Pleophylla* records to the nearest neighboring forest patches which also includes occurrences in forest plantations. Additionally, the CI was calculated for 1 km and 5 km thresholds. The LDI is interpreted as the probability that two randomly chosen pixels in the landscape are not situated in the same patch. The eight cell neighborhood rule was applied for all calculations.

Species richness was estimated using the TomBio-plugin (v2.5.0; `http://www.tombio.uk/qgisplugin`) for QGIS, by counting the number of species per 100 km grid cell in all available records of *Pleophylla*.

Fire frequencies in South Africa were inferred using the MODIS Burned Area Product (Collection 5.1, MCD45; Roy et al. 2002; Roy et al. 2005, 2008) which covers 13 years from 2001 to 2013. It counts no more than one burning event for a given pixel per month if a fire was detected. We summarized the data in a raster layer giving the number of burning events in 13 years using the *raster*-package in R (Hijmans et al. 2015). Since fires are largely of anthropogenic origin, this data was not used for modeling purposes.

3. Results

3.1. Assessment of forest association

Pleophylla species were shown to be strictly forest-associated, although differences were found in the occurrence of *Pleophylla* species in the different forest subtypes (Fig. G23). The mean distance for randomly chosen points, sample sites of other Sericini, and *Pleophylla* species to the nearest forest patch were 161.5 km, 58.5 km, and 3.5 km, respectively (Fig. VIII.2). The ANOVA of distances to nearby forest patches found highly significant differences ($p < 0.001$) among the examined groups and in pairwise t-tests ($p<0.001$ in all pairwise comparisons).

3.2. Present and past distribution models

Dimension reduction of the *bioclim* model retained three principal components for the n-dimensional hypervolume approach so that distributions of species with more than five spatially independent records could be considered (Table G3). The climatic elements most driving divergence in *Pleophylla* bio-climatic records were found along principal components (PCs) one and two: annual precipitation and precipitation in warmest and wettest (BIO 12, 13, 16, and 18) versus precipitation in coldest and driest periods (BIO 14, 17, and 19) as well as mean and extreme temperatures (BIO 1, 5, 6, 8, 9, 10, and 11) versus annual and diurnal temperature ranges (BIO 2 and 7) (Electronic Supplement Fig. G1, Table G4).

Test statistics in terms of ROC, TSS and Cohen's Kappa indicate an overall good discrimination ability of both hypervolume and biomod2 ensembles (Tables G3 and G6).

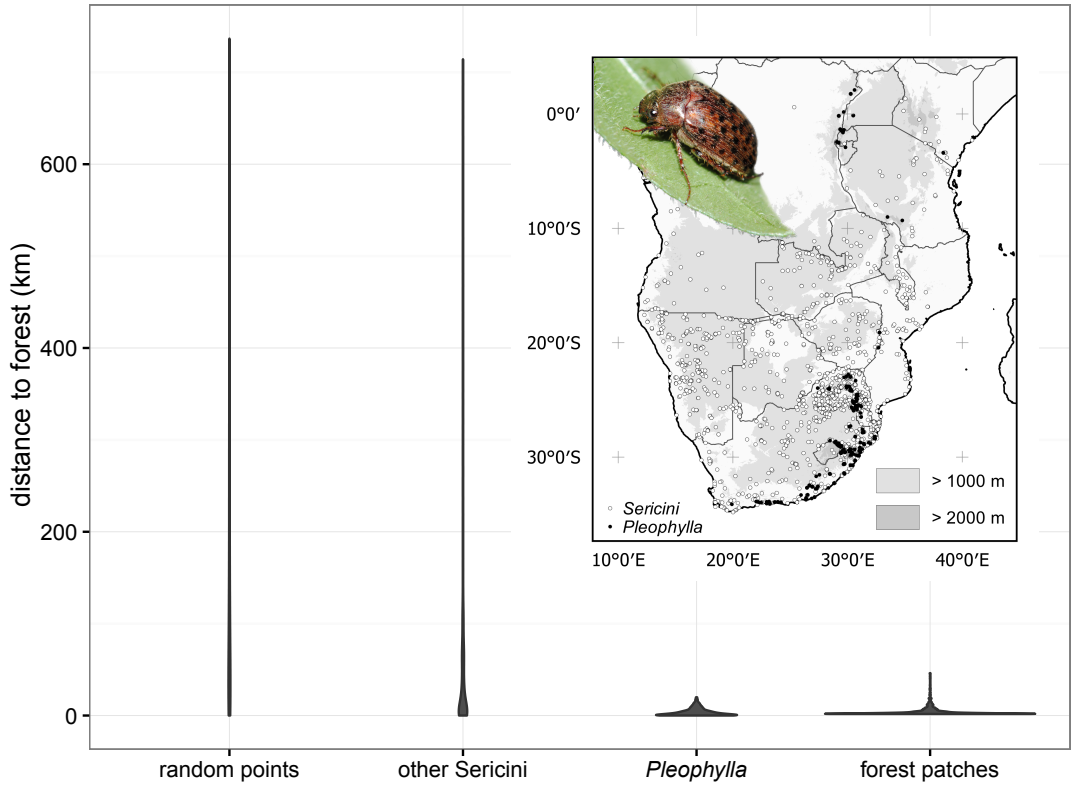

Figure VIII.2. Dependence of *Pleophylla* species on forest habitat. Equal area violin plots illustrating the distance of all available *Pleophylla* locality data to the closest forest edge compared to other Sericini occurrences, randomly distributed points in South Africa, and nearest neighbor distances of all indigenous forest patches. The widths of the violins depict the probability of occurrence density at a given distance. The inset shows the geographical distribution of all specimens under study including absence records (Eberle et al. 2016a) and *P. fasciatipennis* on a leaf.

Highest niche similarity was found among species that predominantly occurred in KwaZulu-Natal (*P. fasciatipennis*, *P. ferruginea*, *P. navicularis*, and *P. pilosa*) and among northern South African species (*P.harrisoni*, *P. pseudopilosa*, and *P. warnockae*) (Fig. G2, Table G5). Climatic niches of *P. nelshoogteensis* and *P. silvatica* slightly overlapped with the latter but were more similar to the southern species. The divergence of hypervolumes of northern and southern South African *Pleophylla* species was mainly driven by PC2 which was dominated by variables of precipitation. Although the estimated species distribution models distinctly differed in their extent, they were all restricted to the southern, south-eastern, and eastern parts of South Africa, enclosed by the Great Escarpment and the coastline (Fig. VIII.3). The models predicted by far larger areas to be climatically suitable for *Pleophylla* than is currently covered by forests.

Compared to the hypervolume models, potential distributions under current climate conditions as suggested by the *biomod2* ensembles were slightly larger for *P. ferruginea*, *P. nelshoogteensis*, and *P. pilosa*, in particular in the Eastern and Western Cape provinces (Electronic Supplement Figs. G3, G7, G11, G19). The predicted potential distribution (Electronic Supplement Fig. G3) was smaller in the *biomod2* ensemble for *P. fasciatipennis*, showing a more fragmented pattern than the hypervolume SDMs (Fig. VIII.3a). According to the *biomod2* ensembles, climatic niches of all species were mainly defined by precipitation variables (Table G6), followed by annual and mean diurnal temperature ranges, being congruent to the results from the hypervolume approach.

Palaeo-distribution models vastly differed in all species for the different PMIP3 global circulation models, mostly caused by differing reconstructions of precipitation among the alternative climate models (PMIP3 synthesis maps, pmip3.lsce.ipsl.fr; accessed March 20, 2016) (Varela et al. 2015). For instance, the distribution of *P. fasciatipennis* in the LGM (Electronic Supplement Fig. G5) in southern Africa ranged from few small and less suited patches along the eastern coast and in the Soutpansberg area of northern South Africa (model IPSL-CM5A-LR) to an extensive and well suited area similar todays range (model MIROC-ESM).

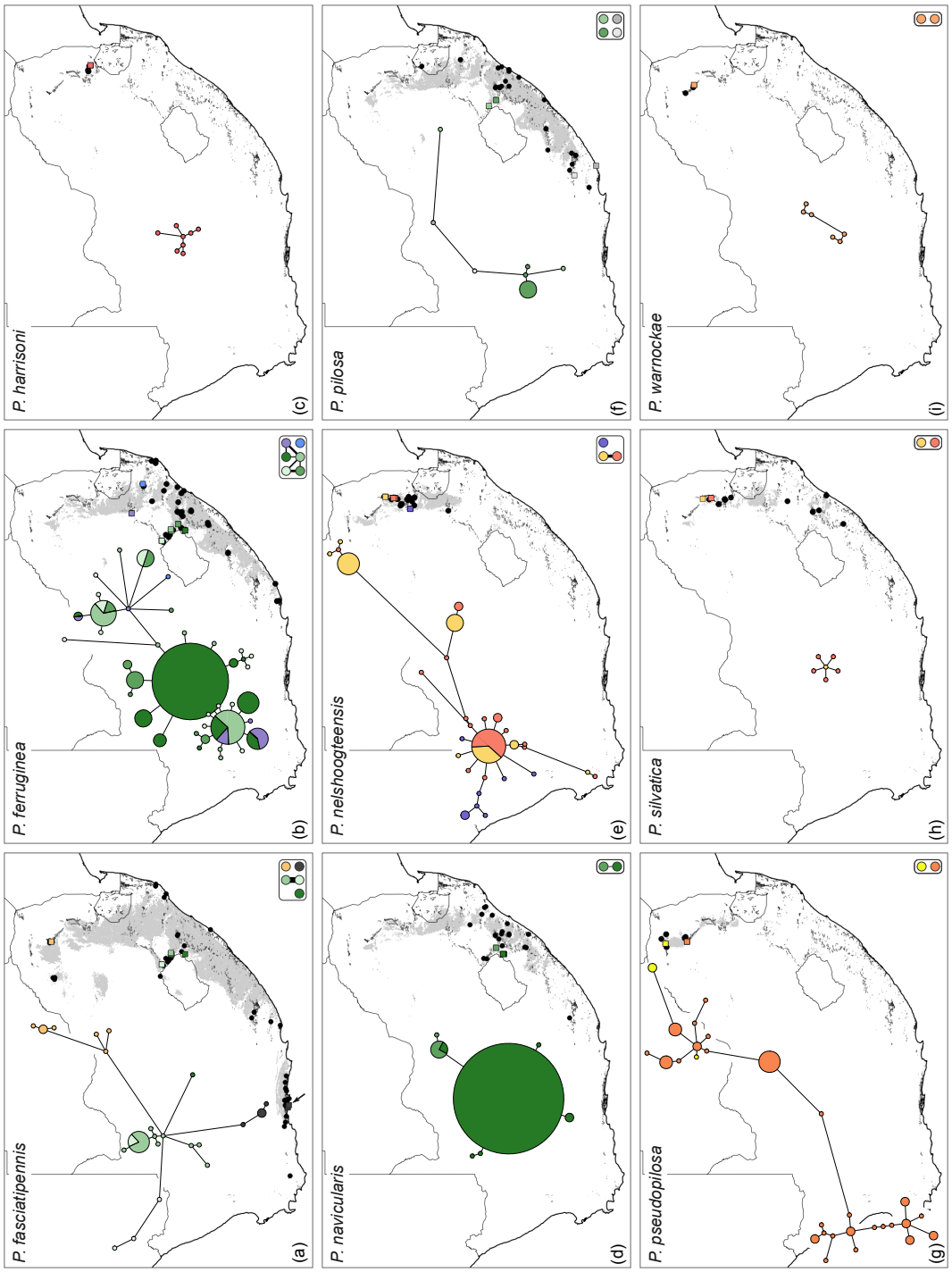

Figure VIII.3. Mitochondrial genetic structure *(cox1)* and population differentiation of nine *Pleophylla* species and their modeled potential distributions (SDM). (To be continued on next page.)

Figure VIII.3. (Continued.) Pie chart sizes in haplotype networks correspond to the number of haplotypes, colors indicate geographic origin of haplotypes like coded on the respective maps, and branch lengths indicate the amount of mutational change. Non-DNA sampling sites that were used for SDM are shown as black dots. Back-projected SDM from n-dimensional hypervolumes are shaded in gray. Where applicable, boxes at the bottom-right corner show genetic admixture among sampling sites measured by G'_{ST} values with thick, thin, or missing lines between color-coded localities respectively depicting high or low genetic exchange or isolation. (a) *P. fasciatipennis*, (b) *P. ferruginea*, (c) *P. harrisoni*, (d) *P. navicularis*, (e) *P. nelshoogteensis*, (f) *P. pilosa*, (g) *P. pseudopilosa*, (h) *P. silvatica*, (i) *P. warnockae*.

3.3. Genetic differentiation and demographic history

Generally, good admixture of haplotypes was found for molecular data with some exceptions as outlined in detail below (Figs. VIII.3, VIII.4). An exceptional pattern was observed in *P. fasciatipennis* which showed remarkable concordance in geographical and mitochondrial genetic differentiation between populations in Limpopo, KwaZulu-Natal, and Western Cape (Fig. VIII.3a). It is the only species of the genus that is distributed from the Cape to the north of the country at Soutpansberg (Fig. VIII.3).

Despite apparent admixture, which was evident from haplotype networks, population differentiation statistics found high genetic differentiation among many sampling sites (Fig. VIII.3, VIII.4, Table G26). Particularly southern populations were strongly isolated (Fig. VIII.3a, f, VIII.4f). Further isolated populations were inferred at northern KwaZulu-Natal (Ngome forest; Fig. VIII.3b, VIII.4f; blue symbols), and in northern South Africa (Soutpansberg; Fig. VIII.3g, VIII.4g; yellow symbols). Good admixture was found among sampling sites at the Drakensberge (Fig. VIII.3a,b,d, VIII.4b,d; green squares) and among sites in northern Mpumalanga (Fig. VIII.3e, VIII.4e, reddish, yellow, and purple squares). Shared haplotypes over long distances between the Drakensberge- and southern Mpumalanga were found for *P. ferruginea* (Fig. VIII.3b, VIII.4b; green and purple squares; *cox1*: $G'_{ST} = 0.78$ and 0.87, ITS1: $G'_{ST} = 0.62$–0.89; Table G26).

Haplotype network analyses revealed repeated reciprocal exchange of related haplotypes between two or more sampling sites (Fig. VIII.3, VIII.4), which was not detected by differentiation statistics but likewise indicated gene flow, although potentially more ancient than evident from shared haplotypes. A local differentiation of

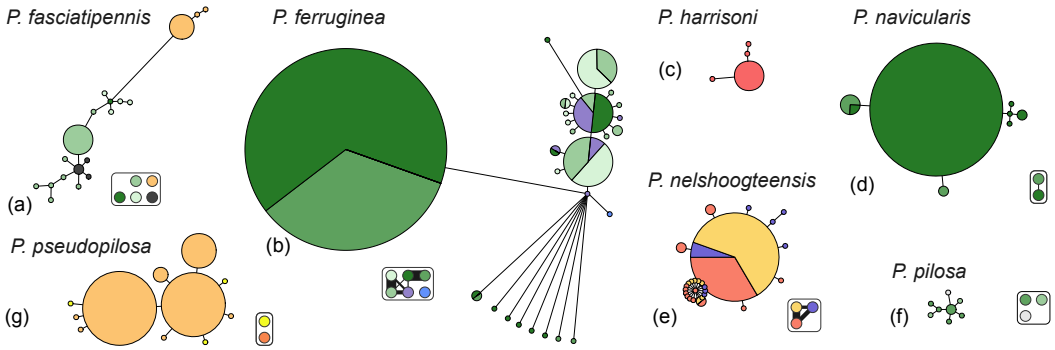

Figure VIII.4. Nuclear genetic structure (ITS1) and population differentiation of seven *Pleophylla* species. Pie chart sizes in haplotype networks correspond to the number of haplotypes and is the same as in Fig. VIII.3. Colors indicate geographic origin of haplotypes like coded on the respective maps in Fig. VIII.3 and branch lengths indicate the amount of mutational change. Where applicable, boxes at the bottom-right of the networks show genetic admixture among sampling sites measured by G'_{ST} values with thick, thin, or missing lines between color-coded localities respectively depicting high or low genetic exchange or isolation. (a) *P. fasciatipennis*, (b) *P. ferruginea*, (c) *P. harrisoni*, (d) *P. navicularis*, (e) *P. nelshoogteensis*, (f) *P. pilosa*, (g) *P. pseudopilosa*.

haplotypes, indicating limited gene flow, was found in *P. ferruginea*, *P. nelshoogteensis*, and *P. pseudopilosa* (Fig. VIII.3). These populations appeared to be completely differentiated according to G'_{ST} due to the lack of shared haplotypes. ITS1 sequences of *P. silvatica* and *P. warnockae* all belong to the same haplotype.

We had sufficient data to infer the demographic history (EBSP) of six species (*P. fasciatipennis*, *P. ferruginea*, *P. navicularis*, *P. nelshoogteensis*, *P. pilosa*, and *P. pseudopilosa*). A recent increase of mean population mutation rate over the last 5–10 ky was a basic pattern observed in all species (Fig. VIII.5). This implied an increase of effective population size (N_e) since the mutation rate was constant over time. There were no fluctuations prior to the LGM (21 kya), indicating a loss of demographic signal in the utilized markers.

3.4. Landscape connectivity

Patch-based landscape measures indicated a highly fragmented distribution of indigenous forests in South Africa. A forest patch density of 0.054 patches per 100 ha

was calculated with a mean nearest neighbor distance between patches of 4.17 km (median: 2.42 km). For comparison, the mean distance of *Pleophylla* records to the nearest forest patch was 3.47 km (median 1.94 km) (Fig. VIII.2). The connectance index for the 1 km, 5 km, and 20 km migration capability thresholds were 0.00%, 0.21%, and 0.97%, respectively. That is, assuming that specimens of *Pleophylla* may disperse up to 20 km outside indigenous forest patches (based on present records; Eberle et al. 2016a), only less than 1% of the pairwise forest patch evaluations resulted connected for *Pleophylla*. Likewise, the Landscape Division Index indicated strong fragmentation of indigenous forest patches. The probability that two randomly chosen points were not in the same patch was 98.2%.

Circuitscape models allow migration to end between sampling points due to high resistances or too long distances and are therefore supposed to reflect genetic admixture. In case of *Pleophylla*, they were strongly affected by the reduction of potential distributions to actual forest patches (Electronic Supplement Figs. G6, G10, G14, G18, G22). In all species, potential migration routes through wider habitat swaths were narrowed to corridors in the resistance landscape informed by forest-accounting scenario (F1). The migration intensity was distinctly increased in these corridors. Migration was restricted to within forest patches in the forest-restricted scenario (F2), completely isolating nearly all sampling localities from each other. Areas with the highest potential loss of connectivity (from F0/F1 to F2) were found on the slopes of the Drakensberg east of the border to Lesotho and in Central KwaZulu-Natal (*P. fasciatipennis*, *P. ferruginea*, and *P. navicularis*), along the cost between Port Elisabeth and Durban (*P. pilosa*) and north of Swaziland between Mbombela to the Motlatse River Canyon (*P. nelshoogteensis*) (Electronic Supplement Figs. G6, G10, G14, G18, G22). In concordance with the genetic differentiation, populations of *P. fasciatipennis* in Limpopo, KwaZulu-Natal and Eastern Cape were not connected in any model (Fig. VIII.3a, Electronic Supplement Fig. G6). The connection between the Drakensberg mountains and southern Mpumalanga (purple square), which was inferred by the haplotype networks and the genetic differentiation index (G'_{ST}) for *P. ferruginea* (Fig. VIII.3b, VIII.4b), was observed along the Great Escarpment (Electronic Supplement Fig. G10, F1). Localities of intermediate *cox1*-haplotypes of *P. nelshoogteensis* and *P. pilosa* (Fig. VIII.3e, f) were also connected in the circuitscape models (scenario F1; Electronic Supplement Figs. G18, G22, arrows). The populations of *P.*

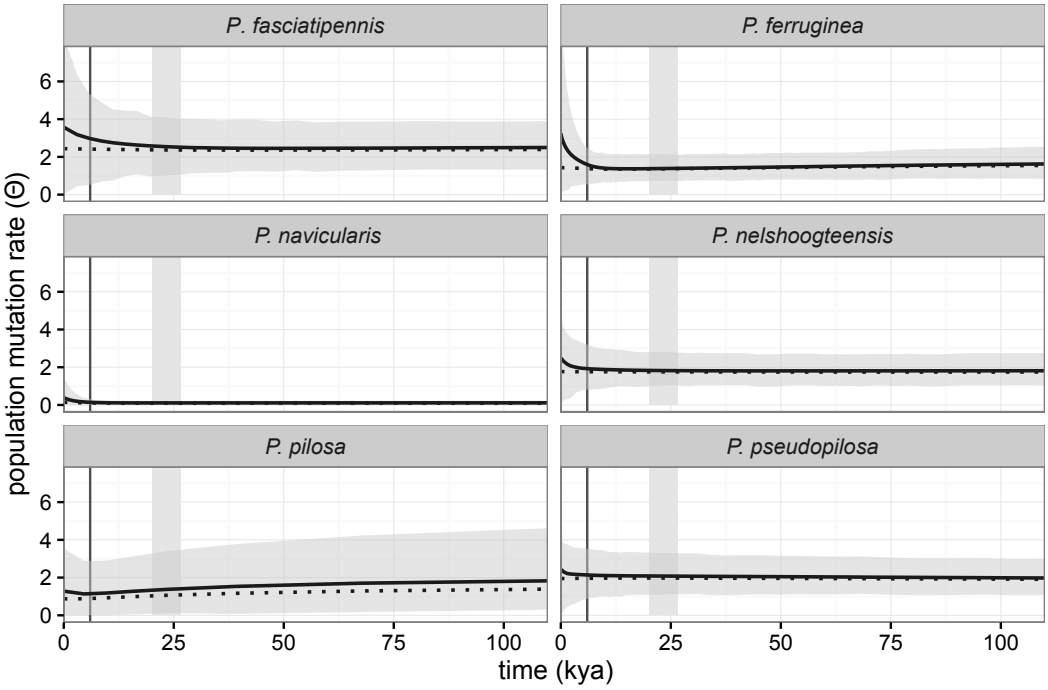

Figure VIII.5. Extended Bayesian Skyline plots of demographic histories. Solid and dotted lines are mean and median population mutation rates (θ) through time, respectively. Gray shading indicates 95% confidence intervals. The vertical black line and the gray bar mark the HA (6 kya) and the LGM (26.5 kya–20 kya), respectively.

fasciatipennis that appeared isolated from DNA data were also disconnected in the circuitscape analyses (Fig. VIII.3a, Electronic Supplement Fig. G6).

LCC/LCP models always spanned even long distances between sampling points (Electronic Supplement Figs. G6, G10, G14, G18, G22). Restrictions of potential migration routes by forest-informed *biomod2* derivatives F1 and F2 altered the results only marginally. Since LCCs/LCPs also connected localities that were inferred to be isolated by circuitscape and molecular methods, they marked connecting areas between those populations that had the highest density of indigenous forest patches and good climatic suitability (Fig. VIII.6).

4. Discussion

4.1. Forest association of *Pleophylla* species

Our distance-to-forest analysis of sampling plots confirmed *Pleophylla* as strictly forest associated (Fig. VIII.2). Due to its polyphagous feeding style of larvae and adults, *Pleophylla* is not restricted to specific plant species or forest subtypes (Fig. G23) and may therefore serve as a proxy for potential forest distribution in South Africa. The proven link between forests and *Pleophylla* established the basis for our use of *Pleophylla* species as a proxy for indigenous forests' distribution in a landscape genetics context, an approach that is for the first time applied to South African forest remains. Records of specimens outside forest patches are reconcilable by the ability of light traps to attract insects over certain distances. However, although short-distance dispersal out of forests at least in some species may rarely occur, the mean and maximum collection distance from forests of 3.5 and 20 km, is better explained by suitable replacement habitats that may often exist in sufficient number in the vicinity of current forest patches (Fig. VIII.2). Some records used for our analyses date back more than 20–30 years, in which the extension and quality of forest habitats might have undergone significant changes due to human land management (e.g., controlled burning, expansion of industrial forestry, etc.) and thus discrepancies between specimen records and recently digitized forest patches might have become even stronger.

Further support for *Pleophylla* being a forest related species is found in the high influence of precipitation variables in all species' models (Electronic Supplement Figs. G1, G2), since precipitation is a crucial factor for forest development (Sankaran et al. 2005). The exclusive occurrence in forests might be attributable to the beetles' dependence on humic forest soils for larval development. Depending on the climatic conditions, these soils remain suitable for considerable time after deforestation since its degradation takes several years (Lemenih et al. 2005).

The distribution models likely improved by past and recent records outside current forest patches, which reveal suitable areas that would have been disregarded by other approaches. The beetles' ability to fly prevents extreme genetic structuring between only little separated forest patches which might lead to overly strong conclusions of a general connectivity breakdown.

4.2. Past development of *Pleophylla* species ranges

Miocene — The Miocene is known for a general cooling and aridification (Zachos et al. 2001) and for the onset of intensified diversification in *Pleophylla* (Eberle et al. 2016c) and other forest-associated organisms (Measey and Tolley 2011; Mlambo et al. 2011; Menegon et al. 2014; Eberle et al. 2016b). This was argued to be attributable to fragmentation and isolation of previously widespread species in forest remains (Maley 1996). Furthermore, the interaction of general long-term climatic stability since the Miocene and complex microclimates were suggested to be the driver of the exceptional plant diversity in the Cape region of southern Africa (Schnitzler et al. 2011) which might also apply for faunal elements. Highly regional forest endemics, like they are for instance found in many flightless and little vagile dung beetles (Davis et al. 2001; Medina and Scholtz 2005; Deschodt and Scholtz 2008; Mlambo et al. 2011), rose chafers (Šípek and Malec 2016), or long-horned grasshoppers (Naskrecki et al. 2008; Samways et al. 2012), might derive from such past subdivisions. However, in most cases these are very rare species which are reported from only a tiny fraction of existing forest patches. Being much more narrowly adapted to forest habitats (shadow, host plants, mammal dung), it seems unlikely that they could persist in only a few tiny and interrupted areas of a few hectares since the Miocene. Also in *Pleophylla* we encounter several highly endemic species (Eberle et al. 2016a) which are scattered over numerous isolated forest patches whose distances exceed by far the dispersal capacity of the species, indicating that forests were until the recent past more extended than currently observed.

Pleistocene — *Pleophylla*'s population demography showed, at least for the species for which we had sufficient data, no fluctuations prior to the LGM, indicating a loss of demographic signal in the utilized molecular markers (Fig. VIII.5). However, late Pleistocene regional paleo-environmental data for eastern South Africa revealed

complex climatic mosaics and frequent changes in vegetation cover with periodically more extended forest patches all over South Africa (Scott 1999; Parkington et al. 2000; Finch and Hill 2008; Chevalier and Chase 2015; Quick et al. 2016), which is also confirmed by climatic models (Huntley et al. 2016). The lack of demographic signal in the EBSP (Fig. VIII.5) that was observed for all investigated *Pleophylla* species, may indicate a strong bottleneck during the LGM in result of habitat reduction (e.g., Heled and Drummond 2008) that erased earlier signals in the utilized markers (Ho and Shapiro 2011). The EBSP therefore supported a strong decline of populations due to the diminution of forests during dry periods around the LGM and only slow recovery of populations. This is largely concordant with paleo-environmental evidence which supposes a general cooling and drying for South Africa (Scott 1989; Partridge et al. 1999; Finch and Hill 2008) except for the western part (current winter rainfall zone) that appears to have been moister than today (Chase and Meadows 2007). Contemporary evidence for moister conditions inland (Free State province and Drakensberg escarpment) (Scott 1989; Norström et al. 2014) strengthen the impression of a highly complex mosaic of climatic conditions.

Concordant with the paleontological record, the SDMs based on the climatic models IPSL-CM5A-LR, CCSM4, and CNRM-CM5 (Table G25) inferred highly reduced distributions and drier conditions in south-eastern South Africa for all investigated *Pleophylla* species during the LGM (Figs. G5, G9, G13, G17, G21). The highly contrasting and contradicting inferences among PMIP3 models (Braconnot et al. 2011) for South Africa (e.g., Fig. G5) may reflect instable climatic conditions and frequent fluctuations that are also evident from paleo-environmental records throughout the country. Recently, long-term precipitation trends in northern South Africa were explained by sea-surface and continental temperature trends while a main influence of the Southern Hemisphere westerlies was deduced for central South African precipitation cycles (Chevalier and Chase 2015), placing South Africa in a transition zone of multiple climatic influences. A dynamic mosaic of microclimates, providing numerous refugial areas over long time time-spans and sustaining ancient phylogenetic diversity like in the case of *Pleophylla*, seems thus likely in the light of fossil records and modeling approaches.

Our data renders potential glacial refugia for *Pleophylla* in coastal areas of today's KwaZulu-Natal and northern South Africa most likely. The best fitting LGM-models

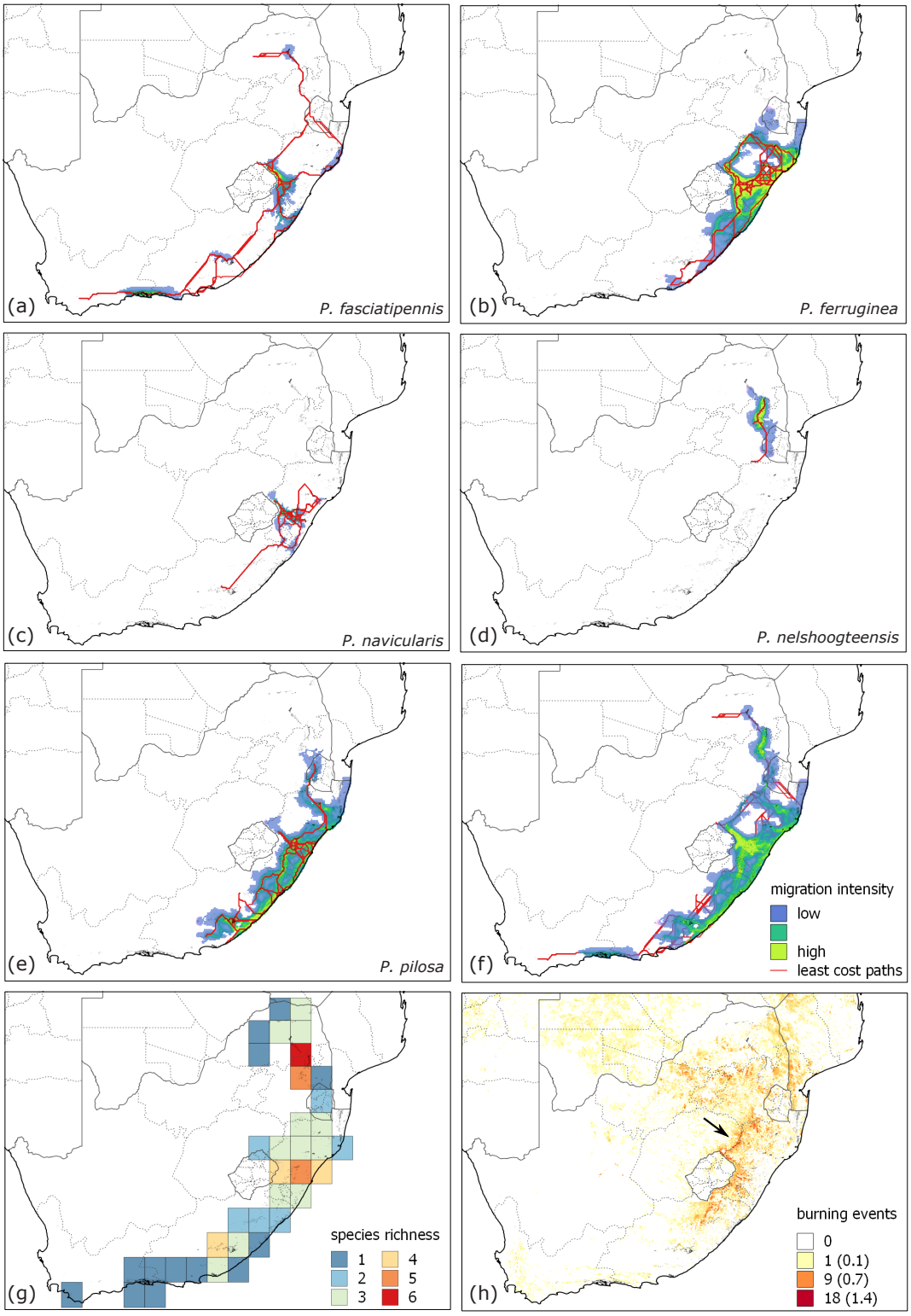

Figure VIII.6. Caption on next page.

Figure VIII.6. Connectivity among sampling sites for five species of *Pleophylla*, fire frequency, and species richness in the investigated area. (a–f) Migration intensity that was inferred with circuitscape based on the forest-accounting species distribution models (F1) (light green: high migration density, blue: low migration density) and the least cost paths (LCPs, red lines) among available sampling sites are overlaid. LCPs roughly illustrate areas that connect current occurrences of *Pleophylla* most parsimoniously. (f) Cumulative migration intensity of the above five species and all LCPs. These areas are climatically best suited for afforestation and intensified conservation of existing forest patches. (g) Regional species richness is depicted as number of species occurring per 100 km grid cell based on all available *Pleophylla* records. (h) Distribution of fire frequency (legend: summarizes number of months with burning events over 13 years; numbers in brackets are averaged burning events per year). The arrow points to areas with high burning frequency along the Great Escarpment which coincide with LCPs and high migration intensity.

(IPSL-CM5A-LR, the CCSM4, and the CNRM-CM5) all showed a shift of species' distributions toward the south-eastern coast (Electronic Supplement Figs. G5, G9, G13, G17, G21), which was also proposed for eastern forests (Scott 1989; Finch and Hill 2008). More northern occurring species, such as *P. nelshoogteensis* and *P. fasciatipennis*, appeared to have persisted in northern refugia (Electronic Supplement Figs. G5, G17). Such potential long term refugia (Ibrahim et al. 1996; Nistelberger et al. 2014) of *Pleophylla* in the northern parts of the Great Escarpment were also supported by the high haplotype diversity that was recovered for the northern species in this study (Figs. VIII.3e,g, VIII.4e,f) as well as the occurrence of locally endemic species (e.g., *P. warnockae*, *P. pseudopilosa*, *P. ruthae*; Eberle et al. 2016a).

Holocene — The late increase of some species' population size after the HA (< 6 kya; Fig. VIII.5) might indicate forest expansions from refuges at the eastern coastal areas and southern KwaZulu-Natal, where a more humid climate and increased forest cover were documented (Eeley et al. 1999; Neumann et al. 2010, 2014; Chevalier and Chase 2015). Otherwise drier conditions in the HA (Jolly et al. 1998) likely prevented earlier expansions. It is also to consider that time-calibration of population-level analyses based on mutation rates inferred from interspecific analyses – although being common practice – can be problematic (Ho et al. 2005; Ho and Larson 2006; Grant 2015). Mutation rates can be an order of magnitude higher than substitution rates that are observed among species (Hoareau 2016). The time estimates from the EBSP might therefore be biased towards older ages, i.e. the observed increase of population size

in all species (Fig. VIII.5) might have occurred even more recently. This fits further palynological evidence suggest rather recently (> 3 kya) more widespread forests and a steady decrease from this time on (Finch and Hill 2008; Neumann et al. 2008; Finch et al. 2009; Neumann et al. 2010). Increasing pollen of neophytes (e.g., *Zea mays* and *Pinus*) and Poaceae in concert with a drastic decline of *Podocarpus* and other trees ca 700 years ago marks the appearance of the first iron age settlers at the coast of KwaZulu-Natal and in todays northern Limpopo province (Scott 1987; Neumann et al. 2010).

4.3. Current population connectivity

For landscape connectivity modeling of current *Pleophylla* populations, we used three resistance layers circumscribing various probabilities for specimen dispersal and occurrence (F0, F1, and F2; Fig. VIII.1b). The models were completed by population genetic inferences which showed good concordance with circuitscape landscape connectivity models based on forest-accounting scenario F1, i.e. circuitscape inferred no or low dispersal between genetically isolated populations (e.g., Figs. VIII.3a, VIII.4a, Electronic Supplement Fig. G6). Complete isolation of sampling sites that was often found by G_{ST}-, G'_{ST}-, and D_{Jost}-statistics (Table G26), which all depend on shared haplotypes, is likely to be caused by limited sampling in some cases. However, a re-evaluation using increased specimen sampling and thorough inference of panmictic populations, which was impossible in the framework of the present study, may reveal more genetic mixture than is currently evident. Despite this potential underestimation, considerable genetic mixture was evident. Therefore, the hypothesis that South African forests have been highly fragmented to isolated patches over long time spans is not supported by our data. This argumentation holds despite a possible alternative explanation for the observed haplotype network pattern for *P. fasciatipennis*: a long distance dispersal from Kwazulu Natal (green) to Limpopo (orange) might have occurred in the past (the yellow haplotypes already diverged and diversified; Figs. VIII.3a, VIII.4a). However, the climatic niche models which show a broad suited connection (Figs. VIII.3a, G3) as well as the hypothesis of a northern glacial refuge in Limpopo render a relictual population in the past more likely. For *P. ferruginea, P. pilosa*, and *P. pseudopilosa*, two or more of the naturally rare long distance

dispersal events would have been necessary to explain the observed patterns, rendering this scenario unlikely as well. Given the high mutation rates (as discussed above) in the employed markers, considerable genetic exchange should thus be assumed at least over the last 5 ky. However, in the present study we found high geographical distances between forest patches that were already slightly larger than the maximum distance that *Pleophylla* species were supposed to migrate between forest patches. Strong fragmentation and low connectance of forest patches (FRAGSTATS analyses, Fig. VIII.2) supported this reasoning. The current data cannot exclude that gene flow among many populations already ceased during the last centuries by anthropogenic influence. The observed genetic mixture of the markers used in this study bear the signature of the recent past (< 5 ky) when possibly most populations were still better connected. Comparing the mean record distances of *Pleophylla* specimens from forests with distances among forest patches (3.5 vs. 4.4 km; see also Electronic Supplement Fig. G2) revealed that the maximum tolerated migration distance between forest patches is reached in many cases.

For all species that were suitable for landscape connectivity modeling, sampling localities were found that were completely isolated from neighboring populations under current conditions (forest-accounting SDMs; F1; Electronic Supplement Figs. G6, G10, G14, G18, G22). The models inferred potential connections for some of them under optimal conditions (i.e., unrestricted SDMs; F0) so that they might be reconnected to larger populations (Fig. VIII.6, Electronic Supplement Fig. G10). Those regions are very likely to be disconnected from other populations by anthropogenic influence. For the reconnection of such isolated populations, burning and intensive forestry in those regions should be reconsidered and re-establishment of indigenous forests in suited areas should be promoted. Well suited areas for afforestation and high priority conservation of existing forest patches might be found along the LCPs that were inferred in this study based on the forest-accounting scenario F1 (Fig. VIII.6). Since they trace the most parsimonious path between larger populations that has the highest forest patch density and the highest climatic suitability, best chances for the reconnection of isolated populations are given there, although being partly under strong impact of man-made fire management (Fig. VIII.6h). Restricting specimen movements to forest patches (forest-restricted SDMs; F2) resulted in a drastic deterioration of predictions of population connectivity (Figs. G6, G10, G14, G18, G22).

4.4. Implications for conservation management and future research

Our results improve the understanding of the forests' natural extension, contributing to potential solutions to the long lasting conservational dilemma whether fire-adapted grasslands and fynbos or forest should be fostered in specific areas. Reliance on two lines of evidence from independent sources of data strengthened the results, which were further backed up by reflections on past development of forest cover in South Africa. Conclusions are closely linked with the conservational importance of South Africa as a cradle of evolution (Pickford 2004; Eberle et al. 2016b), promoting the persistence of a rich and valuable (phylogenetic) diversity (Sechrest et al. 2002; Schnitzler et al. 2011; Huntley et al. 2016). Frequent natural climate fluctuations that led to repeated forest retreats and expansions in southern Africa (Deacon 1983; Eeley et al. 1999) might have acted as speciation pump (Terborgh 1992; Voelker et al. 2010; McDonald and Daniels 2012) producing a unique and divers flora and fauna. Besides *Pleophylla*, other highly diverse insect groups like canthonine dung beetles (Canthonini), which bear many flightless taxa, exclusively occur in South African forest remains (Davis et al. 2001; Medina and Scholtz 2005; Deschodt and Scholtz 2008; Mlambo et al. 2011). Likewise, considerable diversity of flightless species is found in grassland biomes (e.g., Pope 1960; Naskrecki et al. 2008), indicating a certain stability of both biomes in the region. This supports the fossil- and modeling-based idea of a long term mosaic of sufficiently connected grassland and forests that was able to persist in refugia provided by a variety of geological features and different climatic influences during glacial periods. However, with intensified human land use, urban development, and fire management (Fig. VIII.6h), the ability of forests to track environmental change is seriously limited (Eeley et al. 1999) which might result in precarious habitat loss, particularly in times of global climate change. Although more detailed field observations of the specimens' migration between forest patches are necessary for a more detailed insight of today's patterns, a breakdown of many current gene-flow corridors is likely (forest-restricted SDMs, F2), since intensive land use and fires degrade soil organic matter within few decades (Mills and Fey 2003), with serious consequences for forest fauna relics. With further degradation of soils, stepping stone populations that currently connect populations between indigenous forest remains

will disappear and further cease gene flow. Such potential stepping stone populations are frequently found in vicinity of forest plantations, along certain river valleys, or suburban sites where soils stay suitable for many years. However, most plantations are usually burned after logging to remove decaying wood which inevitably impoverishes the soil fauna. It was also shown for other species like the red colobus monkey (*Procolobus gordonorum*) that burning and urban development strongly diminishes gene-flow among populations in forest patches (Ruiz-Lopez et al. 2016). It is therefore important to protect forest remains and to ensure connectivity among them by reforestation of suited connecting areas. Our results support previous findings that current climatic conditions support a much wider forest extent in South Africa (Eeley et al. 1999). The inference of extremely recent drops of population sizes, like they may have occurred by anthropogenic deforestation over the last few hundred years, has to be one major aspect of future studies. Molecular studies employing extremely fast evolving genetic loci like microsatellites and a careful calibration of the molecular clock (Ho and Larson 2006) could provide improved evidence. Shotgun sequencing or restriction site associated DNA analysis (Davey et al. 2011; Hohenlohe et al. 2011) may deliver large datasets for more analyses at finer time scales. In this context it will be fruitful to take into consideration other factors like grazing of large herbivores and the putative influence of wild-fires, also in the light of the risk of invasive plants that endanger ancient grasslands (Bond 2016). This future research should also include other alternative taxa, in particular less mobile model groups like wingless insects that do not have passive dispersal.

Conservation area connectivity is one major issue in the light of global climate change, however, its understanding requires the recognition of current and past patterns. Our results can be used as a step towards the identification of concrete areas where re-establishment and protection of existing indigenous forest could be more effective for connecting forest species populations. It can be seen as a primer to identify areas or problematic regions where additional research at the local scale needs to be conducted (Fig. VIII.6) in order to apply adequate conservation management (e.g., reforestation vs. burning) and thus to ensure the protection of ancient and evolutionary distinct species.

5. Acknowledgements

We are grateful to P. Pacholátko (Brno), C. Schneider (MLUH), M. Uhlig and J. Frisch (both ZMHB), A. Mayekiso and S. van Noort (SAMC), M. Hartmann (NME), R. Stals (SANC), R. Müller (TMSA), W. Schawaller (SMNS), K.A. Johanson (NHRS), M. De Mayer (RMCA), B. Ratcliffe and M. Paulsen (Nebraska/ USNM), M. Barclay (NHM), J. Hájek (NMPC) for the loan of the *Pleophylla* material in their collections. We are furthermore grateful to J. du G. Harrison and R. Müller for collecting additional specimens for DNA analysis. This project was supported by grants from the German Science Association to D.A. (DFG/AH175/1 and AH175/3) and by SYNTHEYS (SE-TAF-3424). For providing D.A. with research and collection permits, we thank the various South African governmental institutions and departments in Eastern Cape (Permit No.: WRO 122/07WR and WRO123/07WR), Gauteng (Permit No.: CPF6 1281), Limpopo (Permit No.: CPM-006-00001), Mpumalangma (Permit No.: MPN-2009-11-20-1232), and Kwazulu-Natal (Permit Nos.: OP3752/2009, 1272/2007, 3620/2006).

6. Data Accessibility

DNA sequences: A detailed listing of Genbank accessions is available in Electronic Supplementary Table G24 — Sampling locations are listed in electronic supplementary table G1 — Species distribution models and results from landscape connectivity analyses for use in Geographic Information Systems are available as electronic supplement file G27 or online at Zenodo (DOI: 10.5281/zenodo.58181).

References

Acocks, J. (1953). Veld Types of South Africa. *Memoir of the Botanical Survey of South Africa* 28, 1–192.

Adriaensen, F., J. Chardon, G. De Blust, E. Swinnen, S. Villalba, H. Gulinck, and E. Matthysen (2003). The application of 'least-cost' modelling as a functional landscape model. *Landscape and Urban Planning* 64, 233–247. DOI: 10.1016/S0169-2046(02)00242-6.

Ahrens, D. (2006). The phylogeny of Sericini and their position within the Scarabaeidae based on morphological characters (Coleoptera: Scarabaeidae). *Systematic Entomology* 31, 113–144. DOI: 10.1111/j.1365-3113.2005.00307.x.

Ahrens, D. and A. P. Vogler (2008). Towards the phylogeny of chafers (Sericini): analysis of alignment-variable sequences and the evolution of segment numbers in the antennal club. *Molecular phylogenetics and evolution* 47, 783–98. DOI: 10.1016/j.ympev.2008.02.010.

Allouche, O., A. Tsoar, and R. Kadmon (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* 43, 1223–1232. DOI: 10.1111/j.1365-2664.2006.01214.x.

Archibald, S., C. E. R. Lehmann, J. L. Gómez-dans, and R. A. Bradstock (2013). Defining pyromes and global syndromes of fire regimes. *Proceedings of the National Academy of Sciences of the United States of America* 110, 6445–6447. DOI: 10.1073/pnas.1211466110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1211466110.

Balesdent, J., G. Wagner, and A. Mariotti (1988). Soil Organic Matter Turnover in Long-term Field Experiments as Revealed by Carbon-13 Natural Abundance. *Soil Science Society of America Journal* 52, 118–124.

Bartholomé, E. and A. Belward (2005). GLC2000: A new approach to global land cover mapping from Earth observation data. *International Journal of Remote Sensing* 26, 1959–1977. DOI: 10.1080/01431160412331291297.

Bews, J. W. (1913). An Œcological Survey of the Midlands of Natal, with Special Reference to the Pietermaritzburg District. *Annals of the Natal Museum* 2, 485–543.

— (1920). The plant ecology of the Coast Belt of Natal. *Annals of the Natal Museum* 4, 367–469.

Blonder, B. (2015). *hypervolume: High-Dimensional Kernel Density Estimation and Geometry Operations*.

Blonder, B., C. Lamanna, C. Violle, and B. J. Enquist (2014). The n-dimensional hypervolume. *Global Ecology and Biogeography* 2011, 1–15. DOI: 10.1111/geb.12146.

Bond, W. J., G. F. Midgley, and F. I. Woodward (2003). What controls South African vegetation – climate or fire? *South African Journal of Botany* 69, 79–91.

Bond, W. J. (2008). What Limits Trees in C_4 Grasslands and Savannas? *Annual Review of Ecology, Evolution, and Systematics* 39, 641–659. DOI: 10.1146/annurev.ecolsys.39.110707.173411.

Bond, W. J. (2016). Ancient grasslands at risk: Highly biodiverse tropical grasslands are at risk from forest-planting efforts. *Science* 351, 120–122.

Braconnot, P., S. P. Harrison, B. Otto-Bliesner, A. Abe-Ouchi, J. Jungclaus, and J.-Y. Peterschmitt (2011). The paleoclimate modeling intercomparison project contribution to CMIP5. *CliVAR Exchanges* 56, 2.

Brooks, M. L., C. M. D'Antonio, D. M. Richardson, J. B. Grace, J. E. Keeley, J. M. DiTomaso, R. J. Hobbs, M. Pellant, and D. Pyke (2004). Effects of Invasive Alien Plants on Fire Regimes. *BioScience* 54, 677–688.

Brown, J. L. (2014). SDMtoolbox: a python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *Methods in Ecology and Evolution* 5, 694–700. DOI: 10.1111/2041-210X.12200.

Brown, R. P. and Z. Yang (2011). Rate variation and estimation of divergence times using strict and relaxed clocks. *BMC Evolutionary Biology* 11, 271. DOI: 10.1186/1471-2148-11-271.

Busby, J. R. (1991). "BIOCLIM – a bioclimatic analysis and prediction system." In: *Nature conservation: cost effective biological surveys and data analysis*. Ed. by C. R. Margules and M. P. Austin. CSIRO, Melbourne, pp. 64–68.

Castley, J. G. and G. I. H. Kerley (1996). The paradox of forest conservation in South Africa. *Forest Ecology and Management* 85, 35–46. DOI: 10.1016/S0378-1127(96)03748-6.

Chase, B. M. and M. E. Meadows (2007). Late Quaternary dynamics of southern Africa's winter rainfall zone. *Earth-Science Reviews* 84, 103–138. DOI: 10.1016/j.earscirev.2007.06.002.

Chevalier, M. and B. M. Chase (2015). Southeast African records reveal a coherent shift from high- to low-latitude forcing mechanisms along the east African margin across last glacial-interglacial transition. *Quaternary Science Reviews* 125, 117–130. DOI: 10.1016/j.quascirev.2015.07.009.

Davey, J. W., P. a. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen, and M. L. Blaxter (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature reviews. Genetics* 12, 499–510. DOI: 10.1038/nrg3012.

Davis, A. L. V., C. H. Scholtz, and J. D. G. Harrison (2001). Cladistic, phenetic and biogeographical analysis of the flightless dung beetle genus, *Gyronotus* van Lansberge (Scarabaeidae: Scarabaeinae), in threatened eastern Afrotropical forests. *Journal of Natural History* 35, 1607–1625. DOI: 10.1080/00222930131709235l.

Deacon, H. J. (1983). Another look at the Pleistocene climates of South Africa. *South African Journal of Science* 79, 325–328.

Deacon, H. J., Q. B. Hendey, and J. J. N. Lambrechts, eds. (1983). *Fynbos palaeoecology: a preliminary synthesis. South African National Scientific Programmes Report No 75.* Cape Town: South African National Scientific Programmes, p. 216.

Deacon, H. and J. Deacon (1999). *Human Beginnings in South Africa: Uncovering the Secrets of the Stone Age*. Cape Town: David Philip Publishers, p. 215.

Deschodt, C. M. and C. H. Scholtz (2008). Systematics of South African forest-endemic dung beetles: new genera and species of small Canthonini (Scarabaeidae: Scarabaeinae). *African Entomology* 16, 91–106. DOI: 10.4001/1021-3589-16.1.91.

Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular biology and evolution* 29, 1969–1973. DOI: 10.1093/molbev/mss075.

Dupont, L. M., T. Caley, J.-H. Kim, I. Castañeda, B. Malaizé, and J. Giraudeau (2011). Glacial-interglacial vegetation dynamics in South Eastern Africa coupled to sea surface temperature variations in the Western Indian Ocean. *Climate of the Past* 7, 1209–1224. DOI: 10.5194/cp-7-1209-2011.

Eberle, J., M. Beckett, A. Özguel-Siemund, J. Frings, S. Fabrizi, and D. Ahrens (2016a). Afromontane forests hide nineteen new species of an ancient chafer lineage (Coleoptera: Scarabaeidae): *Pleophylla* – phylogeny and taxonomic revision. *Zoological Journal of the Linnean Society – in press*.

Eberle, J., S. Fabrizi, P. Lago, and D. Ahrens (2016b). A historical biogeography of megadiverse Sericini – another story "out of Africa"? *Cladistics*, article first published online. DOI: 10.1111/cla.12162.

Eberle, J., R. C. M. Warnock, and D. Ahrens (2016c). Bayesian species delimitation in *Pleophylla* chafers (Coleoptera) – the importance of prior choice and morphology. *BMC Evolutionary Biology* 16, 94. DOI: 10.1186/s12862-016-0659-3.

Edwards, E. J., C. P. Osborne, C. a. E. Strömberg, S. a. Smith, W. J. Bond, P.-A. Christin, A. B. Cousins, M. R. Duvall, D. L. Fox, R. P. Freckleton, O. Ghannoum, J. Hartwell, Y. Huang, C. M. Janis, J. E. Keeley, E. a. Kellogg, A. K. Knapp, A. D. B. Leakey, D. M. Nelson, J. M. Saarela, R. F. Sage, O. E. Sala, N. Salamin, C. J. Still, and B. Tipple (2010). The origins of C_4 grasslands: integrating evolutionary and ecosystem science. *Science (New York, N.Y.)* 328, 587–591. DOI: 10.1126/science.1177216.

Eeley, H. A. C., M. J. Lawes, and S. E. Piper (1999). The influence of climate change on the distribution of indigenous forest in KwaZulu-Natal, South Africa. *Journal of Biogeography* 26, 595–617.

Finch, J. M. and T. R. Hill (2008). A late Quaternary pollen sequence from Mfabeni Peatland, South Africa: Reconstructing forest history in Maputaland. *Quaternary Research* 70, 442–450. DOI: 10.1016/j.yqres.2008.07.003.

Finch, J., M. J. Leng, and R. Marchant (2009). Late Quaternary vegetation dynamics in a biodiversity hotspot, the Uluguru Mountains of Tanzania. *Quaternary Research* 72, 111–122. DOI: 10.1016/j.yqres.2009.02.005.

Foley, J. A., R. DeFries, G. P. Asner, C. Barford, G. Bonan, S. R. Carpenter, F. S. Chapin, M. T. Coe, G. C. Daily, H. K. Gibbs, J. H. Helkowski, T. Holloway, E. a. Howard, C. J. Kucharik, C. Monfreda, J. a. Patz, I. C. Prentice, N. Ramankutty, and P. K. Snyder (2005). Global consequences of land use. *Science* 309, 570–574. DOI: 10.1126/science.1111772. arXiv: arXiv:1011.1669v3.

Fourcade, H. (1889). *Report on the Natal Forests*. Ed. by W. Watson. Pietermaritzburg: Printer to the Natal Government, p. 197.

Geldenhuys, C. J. (1997). Composition and biogeography of forest patches on the inland moutains of the southern Cape. *Bothalia* 27, 57–74.

Grant, W. S. (2015). Problems and Cautions With Sequence Mismatch Analysis and Bayesian Skyline Plots to Infer Historical Demography. *Journal of Heredity* 106, 333–346. DOI: 10.1093/jhered/esv020.

Hedrick, P. W. (2005). A standardized genetic differentiation measure. *Evolution* 59, 1633–1638. DOI: DOI10.1111/j.0014-3820.2005.tb01814.x.

Heled, J. and A. J. Drummond (2008). Bayesian inference of population size history from multiple loci. *BMC evolutionary biology* 8, 289. DOI: 10.1186/1471-2148-8-289.

Hijmans, R., S. Phillips, J. Leathwick, and J. Elith (2015). *dismo: Species Distribution Modeling*.

Hijmans, R. J. (2015). *raster: Geographic Data Analysis and Modeling*. R package version 2.4-15.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25, 1965–1978. DOI: 10.1002/joc.1276.

Hillis, D. M. and M. T. Dixon (1991). Ribosomal DNA: Molecular Evolution and Phylogenetic Inference. *The Quarterly Review of Biology* 66, 411–446.

Ho, S. and G. Larson (2006). Molecular clocks: when times are a-changin'. *Trends in Genetics* 22, 79–83. DOI: 10.1016/j.tig.2005.11.006.

Ho, S. Y. W., M. J. Phillips, A. Cooper, and A. J. Drummond (2005). Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution* 22, 1561–1568. DOI: 10.1093/molbev/msi145.

Ho, S. Y. W. and B. Shapiro (2011). Skyline-plot methods for estimating demographic history from nucleotide sequences. *Molecular Ecology Resources* 11, 423–434. DOI: 10.1111/j.1755-0998.2011.02988.x.

Hoareau, T. B. (2016). Late Glacial Demographic Expansion Motivates a Clock Overhaul for Population Genetics. *Systematic Biology* 65, 449–464. DOI: 10.1093/sysbio/syv120.

Hohenlohe, P. A., S. J. Amish, J. M. Catchen, F. W. Allendorf, and G. Luikart (2011). Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular ecology resources* 11 Suppl 1, 117–22. DOI: 10.1111/j.1755-0998.2010.02967.x.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics* 6, 65–70.

Huber, B. A. (2003). Southern African pholcid spiders: revision and cladistic analysis of *Quamtana* gen. nov. and *Spermophora* Hentz (Araneae: Pholcidae), with notes on male-female covariation. *Zoological Journal of the Linnean Society* 139, 477–527. DOI: 10.1046/j.0024-4082.2003.00082.x.

Huntley, B., Y. C. Collingham, J. S. Singarayer, P. J. Valdes, P. Barnard, G. F. Midgley, R. Altwegg, and R. Ohlemüller (2016). Explaining patterns of avian diversity and endemicity: climate and biomes of southern Africa over the last 140,000 years. *Journal of Biogeography* 43, 874–886. DOI: 10.1111/jbi.12714.

Hutchinson, G. E. (1957). Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology* 22, 415–427. DOI: 10.1101/SQB.1957.022.01.039.

Ibrahim, K. M., R. a. Nichols, and G. M. Hewitt (1996). Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity* 77, 282–291. DOI: 10.1038/hdy.1996.142.

Jaeger, J. A. G. (2000). Landscape division, splitting index, and effective mesh size: New measures of landscape fragmentation. *Landscape Ecology* 15, 115–130. DOI: 10.1023/A:1008129329289. arXiv: 0005074v1 [arXiv:astro-ph].

Jolly, D., I. C. Prentice, R. Bonnefille, A. Ballouche, M. Bengo, P. Brenac, G. Buchet, D. Burney, J. P. Cazet, R. Cheddadi, T. Edorh, H. Elenga, S. Elmoutaki, J. Guiot, F. Laarif, H. Lamb, A. M. Lezine, J. Maley, M. Mbenza, O. Peyron, M. Reille, I. Reynaud-Farrera, G. Riollet, J. C. Ritchie, E. Roche, L. Scott, I. Ssemmanda, H. Straka, M. Umer, E. Van Campo, S. Vilimumbalo, A. Vincens, and M. Waller (1998). Biome reconstruction

from pollen and plant macrofossil data for Africa and the Arabian peninsula at 0 and 6000 years. *Journal of Biogeography* 25, 1007–1027. DOI: 10.1046/j.1365-2699.1998.00238.x.

Katoh, K., K. Misawa, K.-i. Kuma, and T. Miyata (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research* 30, 3059–66.

Keenan, K., P. McGinnity, T. F. Cross, W. W. Crozier, and P. A. Prodöhl (2013). diveRsity: An R package for the estimation of population genetics parameters and their associated errors. *Methods in Ecology and Evolution* 4, 782–788. DOI: 10.1111/2041-210X.12067.

King, N. (1941). The exploitation of the indigenous forests of South Africa. *Journal of the South African Forestry Association* 6, 26–48. DOI: 10.1080/03759873.1941.9631098.

Lanfear, R., B. Calcott, S. Y. W. Ho, and S. Guindon (2012). PartitionFinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. *Molecular biology and evolution* 29, 1695–1701. DOI: 10.1093/molbev/mss020.

Lanfear, R., B. Calcott, D. Kainer, C. Mayer, and A. Stamatakis (2014). Selecting optimal partitioning schemes for phylogenomic datasets. *Evolutionary Biology* 14, 1–14. DOI: 10.1186/1471-2148-14-82.

Lemenih, M., E. Karltun, and M. Olsson (2005). Soil organic matter dynamics after deforestation along a farm field chronosequence in southern highlands of Ethiopia. *Agriculture, Ecosystems and Environment* 109, 9–19. DOI: 10.1016/j.agee.2005.02.015.

Little, I. T., P. a. R. Hockey, and R. Jansen (2013). A burning issue: Fire overrides grazing as a disturbance driver for South African grassland bird and arthropod assemblage structure and diversity. *Biological Conservation* 158, 258–270. DOI: 10.1016/j.biocon.2012.09.017.

Low, A. and A. Rebelo, eds. (1996). *Vegetation of South Africa, Lesotho and Swaziland*. Pretoria: Vegetation of South Africa, Lesotho and Swaziland, p. 84.

Luger, A. D. and E. J. Moll (1993). Fire Protection and Afromontane Forest Expansion in Cape Fynbos. *Biological Conservation* 64, 51–56.

Maley, J. (1996). "The African rain forest – main characteristics of changes in vegetation and climate from the Upper Cretaceous to the Quaternary." In: *Essays on the Ecology of the Guinea-Congo rain forest*. Ed. by I. Alexander, M. Swaine, and R. Watling. 104B. Proceedings of the Royal Society of Edinburgh, pp. 31–73.

McDonald, D. E. and S. R. Daniels (2012). Phylogeography of the Cape velvet worm (Onychophora: *Peripatopsis capensis*) reveals the impact of Pliocene/Pleistocene climatic oscillations on Afromontane forest in the Western Cape, South Africa. *Journal of Evolutionary Biology* 25, 824–835. DOI: 10.1111/j.1420-9101.2012.02482.x.

McGarigal, K., S. A. Cushman, and E. Ene (2012). *FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps*. http://www.umass.edu/landeco/research/fragstats/fragstats.html. Computer software program produced by the authors at the University of Massachusetts, Amherst.

McGarigal, K. and B. J. Marks (1995). *FRAGSTATS: spatial pattern analysis program for quantifying landscape structure. Gen. Tech. Report PNW-GTR-351, USDA Forest Service, Pacific Northwest Research Station, Portland, OR.* Tech. rep.

McRae, B. (2006). Isolation by resistance. *Evolution* 60, 1551–1561.

Measey, G. J. and K. A. Tolley (2011). Sequential fragmentation of Pleistocene forests in an East Africa biodiversity hotspot: Chameleons as a model to track forest history. *PLoS ONE* 6, e26606. DOI: 10.1371/journal.pone.0026606.

Medina, C. and C. Scholtz (2005). Systematics of the southern African genus *Epirinus* Reiche (Coleoptera: Scarabaeinae: Canthonini): descriptions of new species and phylogeny. *Insect Systematics & Evolution* 36, 1–16. DOI: 10.1163/187631205788838500.

Menegon, M., S. Loader, S. Marsden, W. Branch, T. Davenport, and S. Ursenbacher (2014). The genus *Atheris* (Serpentes: Viperidae) in East Africa: Phylogeny and the role of rifting and climate in shaping the current pattern of species diversity. *Molecular Phylogenetics and Evolution* 79, 12–22. DOI: 10.1016/j.ympev.2014.06.007.

Millennium Ecosystem Assessment (2005a). *Ecosystems and Human Well-being: Biodiversity Synthesis*. Washington, DC.: World Resources Institute.

— (2005b). *Ecosystems and Human Well-being: Synthesis*. Washington, DC.: Island Press, p. 155.

Mills, A. J. and M. V. Fey (2003). Declining soil quality in South Africa : effects of land use on soil organic matter and surface crusting. *South African Journal of Science* 99, 429–436.

Mlambo, S., C. L. Sole, and C. H. Scholtz (2011). Phylogeny of the African ball-rolling dung beetle genus *Epirinus* Reiche (Coleoptera:Scarabaeidae:Scarabaeinae). *Invertebrate Systematics* 25, 197–207. DOI: 10.1071/IS10032.

Mucina, L. and M. C. Rutherford, eds. (2006). *The vegetation of South Africa, Lesotho and Swaziland*. Pretoria: Strelitzia 19, South African National Biodiversity Institute.

Naskrecki, P., C. S. Bazelet, and L. A. Spearman (2008). New species of flightless katydids from South Africa (Orthoptera: Tettigoniidae: Meconematinae). *Zootaxa* 32, 19–32.

Neumann, F. H., G. A. Botha, and L. Scott (2014). 18,000 years of grassland evolution in the summer rainfall region of South Africa: evidence from Mahwaqa Mountain, KwaZulu-Natal. *Vegetation History and Archaeobotany* 23, 665–681. DOI: 10.1007/s00334-014-0445-3.

Neumann, F. H., L. Scott, C. B. Bousman, and L. van As (2010). A Holocene sequence of vegetation change at Lake Eteza, coastal KwaZulu-Natal, South Africa. *Review of Palaeobotany and Palynology* 162, 39–53. DOI: 10.1016/j.revpalbo.2010.05.001.

Neumann, F. H., J. C. Stager, L. Scott, H. J. T. Venter, and C. Weyhenmeyer (2008). Holocene vegetation and climate records from Lake Sibaya, KwaZulu-Natal (South Africa). *Review of Palaeobotany and Palynology* 152, 113–128. DOI: 10.1016/j.revpalbo.2008.04.006.

Newbold, T., L. N. Hudson, S. L. L. Hill, S. Contu, I. Lysenko, R. A. Senior, L. Borger, D. J. Bennett, A. Choimes, B. Collen, J. Day, A. De Palma, S. Diaz, S. Echeverria-Londono, M. J. Edgar, A. Feldman, M. Garon, M. L. K. Harrison, T. Alhusseini, D. J. Ingram, Y. Itescu, J. Kattge, V. Kemp, L. Kirkpatrick, M. Kleyer, D. L. P. Correia, C. D. Martin, S. Meiri, M. Novosolov, Y. Pan, H. R. P. Phillips, D. W. Purves, A. Robinson, J. Simpson, S. L. Tuck, E. Weiher, H. J. White, R. M. Ewers, G. M. Mace, J. P. W. Scharlemann, and A. Purvis (2015). Global effects of land use on local terrestrial biodiversity. *Nature* 520, 45–50. DOI: 10.1038/nature14324.

Nistelberger, H., N. Gibson, B. Macdonald, S.-L. Tapper, and M. Byrne (2014). Phylogeographic evidence for two mesic refugia in a biodiversity hotspot. *Heredity* 113, 454–463. DOI: 10.1038/hdy.2014.46.

Norström, E., F. H. Neumann, L. Scott, R. H. Smittenberg, H. Holmstrand, S. Lundqvist, I. Snowball, H. S. Sundqvist, J. Risberg, and M. Bamford (2014). Late Quaternary

vegetation dynamics and hydro-climate in the Drakensberg, South Africa. *Quaternary Science Reviews* 105, 48–65. DOI: 10.1016/j.quascirev.2014.09.016.

Nylander, J. A. A. (2014). *burntrees*. https://github.com/nylander/Burntrees/. (Accessed on July 10, 2014).

Papadopoulou, A., I. Anastasiou, and A. P. Vogler (2010). Revisiting the insect mitochondrial molecular clock: The mid-aegean trench calibration. *Molecular Biology and Evolution* 27, 1659–1672. DOI: 10.1093/molbev/msq051.

Paradis, E. (2010). Pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics* 26, 419–420. DOI: 10.1093/bioinformatics/btp696.

Parkington, J., C. Cartwright, R. Cowling, A. Baxter, and M. Meadows (2000). Palaeovegetation at the last glacial maximum in the western Cape, South Africa: wood charcoal and pollen evidence from Elands Bay Cave. *Aouth African Journal of Science* 96, 543–546.

Partridge, T. C., L. Scott, and J. E. Hamilton (1999). Synthetic reconstructions of southern African environments during the Last Glacial Maximum (21-18 kyr) and the Holocene Altithermal (8-6 kyr). *Quaternary International* 57/58, 207–214. DOI: 10.1016/S1040-6182(98)00061-5.

Peterson, A. T. and Á. S. Nyári (2008). Ecological niche conservatism and pleistocene refugia in the thrush-like Mourner, *Schiffornis* sp., in the neotropics. *Evolution* 62, 173–183. DOI: 10.1111/j.1558-5646.2007.00258.x.

Phillips, J. F. (1930). Fire: its influence on biotic communities and physical factors in South and East Africa. *South African Journal of Science* 28, 352–367.

Pickford, M. (2004). Southern Africa: A cradle of evolution. *South African Journal of Science* 100, 205–214.

Pope, R. (1960). A revision of the species of *Schizonycha* Dejean (Col.: Melolonthidae) from southern Africa. *Bulletin of the Natural History Museum* 9, 63–218.

Quick, L. J., B. M. Chase, M. E. Meadows, L. Scott, and P. J. Reimer (2011). A 19.5 kyr vegetation history from the central Cederberg Mountains, South Africa: Palynological evidence from rock hyrax middens. *Palaeogeography, Palaeoclimatology, Palaeoecology* 309, 253–270. DOI: 10.1016/j.palaeo.2011.06.008.

Quick, L. J., M. E. Meadows, M. D. Bateman, K. L. Kirsten, R. Mäusbacher, T. Haberzettl, and B. M. Chase (2016). Vegetation and climate dynamics during the last glacial period in the fynbos-afrotemperate forest ecotone, southern Cape, South Africa. *Quaternary International* 404, 136–149. DOI: 10.1016/j.quaint.2015.08.027.

R Development Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. http://www.r-project.org/.

Rambaut, A., M. Suchard, D. Xie, and A. Drummond (2014). Tracer v1.6.

Ravenstein, E. (1898). *A journal of the first vayage of Vasco da Gama 1497–1499*. Ed. by E. Ravenstein. London: Printed for the Hakluyt Society at Bedford Press, Bedfordbury, W.C., p. 324.

Reside, A. E., J. Vanderwal, A. Kutt, I. Watson, and S. Williams (2012). Fire regime shifts affect bird species distributions. *Diversity and Distributions* 18, 213–225. DOI: 10.1111/j.1472-4642.2011.00818.x.

Roy, D. P., L. Boschetti, C. O. Justice, and J. Ju (2008). The collection 5 MODIS burned area product – Global evaluation by comparison with the MODIS active fire product. *Remote Sensing of Environment* 112, 3690–3707. DOI: 10.1016/j.rse.2008.05.013.

Roy, D. P., Y. Jin, P. E. Lewis, and C. O. Justice (2005). Prototyping a global algorithm for systematic fire-affected area mapping using MODIS time series data. *Remote Sensing of Environment* 97, 137–162. DOI: 10.1016/j.rse.2005.04.007.

Roy, D. P., P. E. Lewis, and C. O. Justice (2002). Burned area mapping using multi-temporal moderate resolution spatial resolution data - a bi-directional reflectance model-based expectation approach. *Remote Sensing of Environment* 83, 263–286.

Ruiz-Lopez, M. J., C. Barelli, F. Rovero, K. Hodges, C. Roos, W. E. Peterman, and N. Ting (2016). A novel landscape genetic approach demonstrates the effects of human disturbance on the Udzungwa red colobus monkey *(Procolobus gordonorum). Heredity* 116, 167–176. DOI: 10.1038/hdy.2015.82.

Samways, M. J., M. Hamer, and R. Veldtman (2012). "Insect Conservation: Past, Present and Prospects." In: *Insect Conservation: Past, Present and Prospects*. Ed. by T. R. New. Dordrecht: Springer Netherlands, pp. 245–278. DOI: 10.1007/978-94-007-2963-6.

Sankaran, M., N. P. Hanan, R. J. Scholes, J. Ratnam, D. J. Augustine, B. S. Cade, J. Gignoux, S. I. Higgins, X. Le Roux, F. Ludwig, J. Ardo, F. Banyikwa, A. Bronn, G. Bucini, K. K. Caylor, M. B. Coughenour, A. Diouf, W. Ekaya, C. J. Feral, E. C. February, P. G. H. Frost, P. Hiernaux, H. Hrabar, K. L. Metzger, H. H. T. Prins, S. Ringrose, W. Sea, J. Tews, J. Worden, and N. Zambatis (2005). Determinants of woody cover in African savannas. *Nature* 438, 846–849. DOI: 10.1038/nature04070.

Schnitzler, J., T. G. Barraclough, J. S. Boatwright, P. Goldblatt, J. C. Manning, M. P. Powell, T. Rebelo, and V. Savolainen (2011). Causes of plant diversification in the cape biodiversity hotspot of South Africa. *Systematic Biology* 60, 343–357. DOI: 10.1093/sysbio/syr006.

Scott, L. (1987). Late quaternary forest history in Venda, Southern Africa. *Review of Palaeobotany and Palynology* 53, 1–10. DOI: 10.1016/0034-6667(87)90008-X.

Scott, L. (1989). Climatic conditions in Southern Africa since the last glacial maximum, inferred from pollen analysis. *Palaeogeography, Palaeoclimatology, Palaeoecology* 70, 345–353. DOI: Doi:10.1016/0031-0182(89)90112-0.

— (1999). Vegetation history and climate in the Savanna biome South Africa since 190,000 ka: A comparison of pollen data from the Tswaing Crater (the Pretoria Saltpan) and Wonderkrater. *Quaternary International* 57-58, 215–223. DOI: 10.1016/S1040-6182(98)00062-7.

Sechrest, W., T. M. Brooks, G. A. B. da Fonseca, W. R. Konstant, R. A. Mittermeier, A. Purvis, A. B. Rylands, and J. L. Gittleman (2002). Hotspots and the conservation of evolutionary history. *Proceedings of the National Academy of Sciences of the United States of America* 99, 2067–2071. DOI: 10.1073/pnas.251680798.

Shah, V. B. and B. H. McRae (2008). Circuitscape : A Tool for Landscape Ecology. *Proceedings of the 7th Python in Science Conference*, 62–65.

Šípek, P. and P. Malec (2016). On the cetoniine fauna of Eastern Cape (EC) and KwaZulu-Natal (KZN) and the basic guidelines to captive breeding of these beetles (Coleoptera, Scarabaeidae, Cetoniinae). *Cetoniimania* 9, 54–80.

Sørensen, T. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Kongelige Danske Videnskabernes Selskab* 5, 1–34.

Swets, J. A. (1988). Measuring the Accuracy of Diagnostic Systems. *Science* 240, 1285–1293. DOI: 10.1126/science.3287615.

Templeton, A., R. Robertson, J. Brisson, and J. Strasburg (2001). Disrupting evolutionary processes: the effect of habitat fragmentation on collared lizards in the Missouri Ozarks. *Proceedings of the National Academy of Sciences of the United States of America* 98, 5426–5432. DOI: 10.1073/pnas.091093098.

Terborgh, J. (1992). *Diversity and the Tropical Rain Forest*. New York: Freeman, New York.

Thuiller, W. (2003). BIOMOD: Optimising predictions of species distributions and projecting potential future shift under global change. *Global Change Biology* 9, 1353–1362.

Thuiller, W., D. Georges, and R. Engler (2013). *biomod2: Ensemble platform for species distribution modeling*.

Thuiller, W., D. Georges, R. Engler, and F. Breiner (2016). *biomod2: Ensemble Platform for Species Distribution Modeling. R package version 3.3-7*.

Timmermans, M. J. T. N., S. Dodsworth, C. L. Culverwell, L. Bocak, D. Ahrens, D. T. J. Littlewood, J. Pons, and a. P. Vogler (2010). Why barcode? High-throughput multiplex sequencing of mitochondrial genomes for molecular systematics. *Nucleic acids research* 38, e197. DOI: 10.1093/nar/gkq807.

Varela, S., M. S. Lima-Ribeiro, and L. C. Terribile (2015). A short guide to the climatic variables of the last glacial maximum for biogeographers. *PLoS ONE* 10. DOI: 10.1371/journal.pone.0129037.

Verbeylen, G., L. De Bruyn, F. Adriaensen, and E. Matthysen (2003). Does matrix resistance influence Red squirrel (*Sciurus vulgaris* L. 1758) distribution in an urban landscape? *Landscape Ecology* 18, 791–805.

Voelker, G., R. K. Outlaw, and R. C. K. Bowie (2010). Pliocene forest dynamics as a primary driver of African bird speciation. *Global Ecology and Biogeography* 19, 111–121. DOI: 10.1111/j.1466-8238.2009.00500.x.

Vogler, A. P. and R. Desalle (1994). Evolution and Phylogenetic Information Content of the ITS-1 Region in the Tiger Beetle Cicindela dorsalis. *Molecular biology and evolution* 11, 393–405.

Wilgen, B. W. van (2009). The evolution of fire management practices in savanna protected areas in South Africa. *South African Journal of Science* 105, 343–349.

Wilgen, B. W. van, G. G. Forsyth, and P. Prins (2012). The management of fire-adapted ecosystems in an urban setting: The case of table mountain National Park, South Africa. *Ecology and Society* 17, 8. DOI: 10.5751/ES-04526-170108.

Zachos, J., M. Pagani, L. Sloan, E. Thomas, and K. Billups (2001). Trends, Rhythms, and Aberrations in Global Climate 65 Ma to Present. *Science* 292, 686–693. DOI: 10.1126/science.1059412.

Chapter IX.

General Discussion

The research comprising the present thesis acted at the interface of multiple disciplines including species delimitation, taxonomy, phylogenetics, biogeography, ecological niche modeling, evolutionary biology, and population genetics. By comparing the latest methods to well established ones, thoroughly testing them, and evaluating their benefits and pitfalls, it advanced commonly applied procedures and contributed considerably to the knowledge of scarab chafer biology and the mechanisms driving their exceptional diversity.

Sericini chafers were in the focus of the majority of the studies. Working with such a mega-diverse group was highly beneficial in terms of the large and dense information content provided. On the other hand, the still widely unknown taxonomy of Sericini posed challenges for investigations comprising the whole tribe. The problems that arose from mega-diversity of the focal group were incomplete sampling in general, lacking knowledge of higher systematics which is necessary for sensible reduction of the sampling, and computational constraints. Proper species delimitation (Chapter III) and the taxonomic revision of the genus *Pleophylla* (Chapter IV) facilitated detailed investigation of habitat connectivity at the regional level for a range of ecologically well comparable species (Chapter VIII). However, this procedure is time consuming and not yet feasible for the whole tribe. Extensive sampling helped to overcome the above-mentioned problems for the historical biogeography of Sericini at least partly (Chapter VII). The study shed light on the tribes' distribution patterns and major factors driving their diversification by combining phylogenetics, fossil data, and geography. In particular, the invasion of the Asian continent by the subtribe Sericina seems to have boosted species formation. This may be interpreted as entry into a new adaptive zone (ecological speciation) (Simpson 1953; Mitter et al. 1988). Although other insect species might have occupied the same or very similar ecological niches like those of the first Sericini species that invaded the Asian continent, competitive superiority of Sericini and vast availability of food resources (Jermy 1985; Kaplan and Denno 2007) might have promoted their evolutionary success. Such competitive superiority might for instance be explained by evolutionary key innovations (Hunter 1998) or the related concept of correlated progression (Thomson 1992). Developments of the Sericini locomotory system, i.e., secondary closure and extension of the mesal process in concert with the enlargement of the metacoxa in ancient lineages of Sericini (Chapter V) are likely responsible for the ability of Sericini to quickly dig themselves

into soil to escape predators (Chapter V, VII). Also, the adaptation to predatory ants in the African genus *Trochalus* are candidates for trait systems that promoted intensified diversification. However, proper evidence is hard to obtain, in particular if a trait is only known from one phylogenetic lineage (Hunter 1998).

Further causes of diversification were investigated for the related herbivore scarab genus *Schizonycha* (Chapter VI). Comprising 370 species, it is one of the largest known genera of animals. The molecular phylogenetic hypothesis of the genus revealed multiple shifts to asymmetry of male genitalia, making *Schizonycha* an interesting model case for sexual selection as diversifying factor. Although there was no measurable evidence for increased morphological divergence by asymmetry, its frequent occurrence likely indicates a significant function. Females may differentiate among symmetric and asymmetric genitalia by different mechanisms that are not easily measurable by morphological divergence. The arrangement of mechanical receptors in the female genital tract might be one explanation. Asymmetry therefore likely provides a working point for sexual selection which might foster diversification of the group (Hosken and Stockley 2004). However, further research also regarding the female genital tract morphology is necessary to draw firm conclusions.

Summarizing, within herbivore scarabs, three of the major known mechanisms for diversification of life were identified to have had potential impact: evolutionary key innovations, entry into new adaptive zones, and sexual selection. The former two of them were found in a single tribe of herbivore scarabs, the Sericini. The vast variability in male genitalia, including asymmetry, renders a diversifying impact of sexual selection likely as well. The hyper-diversity of herbivore scarabs may thus be explained by a series of major diversifying events and mechanisms, enabled and sustained by the foundation of a huge availability of food resources.

The integration of multiple data sources played a central role in the studies comprising this thesis and proved highly valuable to address diverse evolutionary and taxonomic questions. In particular, to species delimitation, the integration of molecular and morphometric data was highly beneficial. Both molecular and morphometric data profited from each other. While morphometric data helped to overcome certain problems that occurred when delimiting species using molecular data, the well-defined hierarchical framework given by molecular data (guide trees in iBPP; Solís-Lemus et al. 2015) strongly improved the utility of morphometric data in species delimitation

(Chapter II and III). Our studies foster the development of the still underdeveloped (Yeates et al. 2011) methods for the integration of molecular and continuous morphological data, by thoroughly testing recently developed methods (Solís-Lemus et al. 2015) and highlighting its utility. We showed that morphological divergence predates divergence of molecular loci that are commonly used for species delimitation *(cox1)* and may resemble species boundaries even more consistently than extremely fast and erratically evolving molecular markers (ITS1; Chapter III). Usage of the latter can easily lead to false positive conclusions (i.e., over-splitting; Chapter III), in particular when only few molecular loci are available. In the case of the Mediterranean genus *Pachypus*, inferences based on mitochondrial DNA data were highly distorted by sex-biased and thus limited dispersal (Chapter II). Here, morphometrics based on linear measurements were only little influenced and resembled species boundaries well. Morphometric data is thus not only beneficial in cases of young radiations but also in cases of erroneous over-splitting (Chapter II) which might not only occur by constraint gene flow but also by under-sampling (Lohse 2009). However, like for molecular data, the choice of well suited "partitions", i.e., molecular loci or body parts, was also necessary for morphological data (morphometrics of linear body length measurements vs. geometric morphometrics of body outlines; Chapter II). Therefore, the simultaneous analysis of multiple morphological partitions and their sensible integration in a multi-data approach might be subject to future research.

Apart from choosing suited partitions other considerations might be relevant for future improvements of morphological species recognition. Delimitation of species by morphology does not necessarily rely on the sheer amount of mathematically measurable variation. Species may be identified by subtle, but constant and thus potentially well suited characteristics (e.g., Figure VI.1 on page 232). However, geometric morphometrics measure the amount of variation in shape. Subtle characteristics like the dentation of an insects leg may easily be obscured by other, less decisive variation. The high-dimensional mathematical description of morphological variation that is gained from geometric morphometrics – particularly from measurements of outlines – is commonly summarized with principal component analyses, which collects the highest amount of variation in the first components. Subsequent components are discarded to avoid problems of extremely high dimensionality and to reduce uninformative noise (Gauch Jr 1982; Zelditch et al. 2004). Along with these components,

relevant information for species boundaries might be lost. Researchers may focus their measurements to the region of interest to overcome the issue which is, however, not always possible. In this case the development of more specialized and sensitive methods are needed that are able to identify species related signal in dimensions describing minor amounts of variation and can distinguish it from random noise variation.

It is widely acknowledged that genital morphology strikingly differs among many insect species and that it is particularly useful for systematics and taxonomy (Eyer 1924; Dirsh 1956; Tuxen 1970; Simmons 2014), although almost exclusively applied to males (Ah-King et al. 2014). Potential causes are discussed for more than a century and range from "key and lock" hypotheses (Dufour 1948), random divergence due to decoupling from natural selection (pleiotropy, Mayr 1963) to sexual selection (Lloyd 1979). Most authors now agree on the latter with Fisherian female choice often raised as the driving mechanism (Eberhard 1985, 1996; Hosken and Stockley 2004; Eberhard 2010; Simmons 2014). In Chapter III, we showed that genital morphology is well suited for use with modern species delimitation methods even in very recent radiations. Offering a vast variety of shapes makes them interesting for many applications; however, their complexity may also pose challenges to morphometric methods. Particularly variation in the third dimension is hard to capture with the most commonly used method, i.e., two dimensional photography. Rotations in space might distort the actual shape and introduce artificial variation among samples (Chapter VI). Increased availability of 3D-imaging technologies (e.g., micro-CT, synchrotron) and computational power facilitates the application of morphometrics in three dimensions for a broad community of researchers and offers great perspectives for geometric morphometrics of complex genital characters. Hitherto subjective evaluation of morphological traits of genitalia may thus be increasingly superseded by impartial statistics on continuous trait data.

High dimensional continuous trait data results not only from morphometric analyses but also for instance from ecological niche modeling. The application of these kind of data and its challenges is thus not restricted to morphometric analyses but is used in many disciplines which will benefit from methodological advances. In case of ecological niche modeling, evolutionary biologists might for instance investigate the degree of niche conservatism in a group of species. Niche shifts might be correlated with diversification (e.g., Marvaldi et al. 2002; Schnitzler et al. 2012) and

differing niches of sister species can help to identify species boundaries (e.g., Bond and Stockman 2008). Using morphometric data, the responses of characters to environmental change maybe studied. Since multiple species are not statistically independent from each other, assumptions of standard statistical methods will most likely be violated (Revell 2010). Comparative evolutionary analyses take into account the non-independence of data from biological species (Felsenstein 1985; Harvey and Pagel 1991) by integrating the species' phylogenetic covariation into the analyses. However, when it comes to highly multivariate data, in particular in the response variable, the suite of available methods diminishes considerably and sufficient sample sizes are hard to obtain since all specimens need to be incorporated in the phylogenetic framework. The implementation of multivariate standardized phylogenetic independent contrasts (McPeek et al. 2008) into the open source statistics environment R (Chapter VI, V; D2; R Development Core Team 2015) is a first step towards greater availability of such methods to a broad community. However, it is only capable to analyze traits evolving along Brownian Motion, i.e., in a steady manner without accelerations and decelerations like it is expected under directed or stabilizing selection, respectively. More complex evolutionary models may be implemented for instance with generalized estimating equations (Paradis and Claude 2002). This field of research has great potential for the investigation of evolutionary mechanisms leading to an intensified diversification of biological species.

The basis for many data used in this thesis is based on natural history collections which are thus highlighted as highly valuable documents of natural history. The present research brings together the vast information inherent in the collections with molecular data from freshly sampled specimens. Modern techniques facilitated the accession of morphological information with morphometrics and discrete morphological characters were used for phylogenetic inference (Chapter III and IV). Distribution data from the specimens' labels enabled inferences of phylogeography, of species' current and paleo-distributions based on climatic niche modeling, and ultimately of population level landscape connectivity analyses, directly supplying conservation biologists with valuable information (Chapter VIII). The description of nineteen new species, synonymization of two species, and the removal of one species from synonymy (Chapter IV) heavily depended on material from natural history collections. It is a

first step towards the comprehensive cataloging of the southern African Sercini fauna and provided a proper basis for further research on the species level.

References

Bond, J. E. and A. K. Stockman (2008). An integrative method for delimiting cohesion species: finding the population-species interface in a group of Californian trapdoor spiders with extreme genetic divergence and geographic structuring. *Systematic biology* 57, 628–646. DOI: 10.1080/10635150802302443.

Dirsh, V. M. (1956). The phallic complex in Acridoidea (Orthoptera) in relation to taxonomy. *Transactions of the Entomological Society of London* 108, 223–270.

Dufour, L. (1948). Anatomie générale des Dipteres. *Annuare de Science Naturelle* 1, 244–264.

Eberhard, W. G. (1985). *Sexual selection and animal genitalia*. Cambridge: Harvard University Press.

— (1996). *Female Control: Sexual Selection by Cryptic Female Choice*. Princeton, New Jersey, USA: Princeton University Press, Princeton, New Jersey, USA, p. 501.

Eberhard, W. G. (2010). Evolution of genitalia: Theories, evidence, and new directions. *Genetica* 138, 5–18. DOI: 10.1007/s10709-009-9358-y.

Eyer, J. R. (1924). The comparative morphology of the male genitalia of the primitive Lepidoptera. *Annals of the Entomological Society of America* 17, 275–328.

Felsenstein, J. (1985). Phylogenies and the comparative method. *The American Naturalist* 125, 1–15.

Gauch Jr, H. G. (1982). Noise reduction by eigenvector ordinations. *Ecology* 63, 1643–1649.

Harvey, P. and M. Pagel (1991). *The comparative method in evolutionary biology*. Vol. 239. Oxford university press Oxford.

Hosken, D. J. and P. Stockley (2004). Sexual selection and genital evolution. *Trends in ecology & evolution* 19, 87–93. DOI: 10.1016/j.tree.2003.11.012.

Hunter, J. P. (1998). Key innovations and the ecology of macroevolution. *Trends in ecology & evolution* 13, 31–6.

Jermy, B. T. (1985). Is there competition between phytophagous insects? *Journal of Zoological Systematics and Evolutionary Research* 23, 275–285.

Kaplan, I. and R. F. Denno (2007). Interspecific interactions in phytophagous insects revisited: a quantitative assessment of competition theory. *Ecology Letters* 10, 977–994.

Ah-King, M., A. B. Barron, and M. E. Herberstein (2014). Genital Evolution: Why Are Females Still Understudied? *PLoS Biology* 12, e1001851. DOI: 10.1371/journal.pbio.1001851.

Lloyd, J. E. (1979). Mating Behavior and Natural Selection. *The Florida Entomologist* 62, 17. DOI: 10.2307/3494039.

Lohse, K. (2009). Can mtDNA Barcodes Be Used to Delimit Species? A Response to Pons et al. (2006). *Systematic Biology* 58, 439–442. DOI: 10.1093/sysbio/syp039.

Marvaldi, A. E., A. S. Sequeira, C. W. O'Brien, and B. D. Farrell (2002). Molecular and Morphological Phylogenetics of Weevils (Coleoptera , Curculionoidea): Do Niche Shifts

Accompany Diversification? *Systematic biology* 51, 761–785. DOI: 10.1080/10635150290102465.

Mayr, E. (1963). *Animal species and evolution*. 1st ed. The Belknap Press of Harvard University Press, Cambridge, Massachusetts, p. 811.

McPeek, M. A., L. Shen, J. Z. Torrey, and H. Farid (2008). The Tempo and Mode of Three-Dimensional Morphological Evolution in Male Reproductive Structures. *The American Naturalist* 171, E158–E178. DOI: 10.1086/587076.

Mitter, C., B. Farrell, and B. Wiegmann (1988). The phylogenetic study of adaptive zones: has phytophagy promoted insect diversification? *American Naturalist* 132, 107–128.

Paradis, E. and J. Claude (2002). Analysis of comparative data using generalized estimating equations. *Journal of Theoretical Biology* 218, 175–85. DOI: 10.1006/yjtbi.3066.

R Development Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. http://www.r-project.org/.

Revell, L. J. (2010). Phylogenetic signal and linear regression on species data. *Methods in Ecology and Evolution* 1, 319–329. DOI: 10.1111/j.2041-210X.2010.00044.x.

Schnitzler, J., C. H. Graham, C. F. Dormann, K. Schiffers, and H. Peter Linder (2012). Climatic niche evolution and species diversification in the Cape flora, South Africa. *Journal of Biogeography*. Ed. by S. Higgins, n/a–n/a. DOI: 10.1111/jbi.12028.

Simmons, L. W. (2014). Sexual selection and genital evolution. *Austral Entomology* 53, 1–17. DOI: 10.1111/aen.12053.

Simpson, G. G. (1953). *The major features of evolution*. 1st ed. Columbia University Press, p. 434.

Solís-Lemus, C., L. L. Knowles, and C. Ané (2015). Bayesian species delimitation combining multiple genes and traits in a unified framework. *Evolution* 69, 492–507. DOI: 10.1111/evo.12582.14.

Thomson, K. S. (1992). Macroevolution: The Morphological Problem. *Integrative and Comparative Biology* 32, 106–112. DOI: 10.1093/icb/32.1.106.

Tuxen, S. L., ed. (1970). *Taxonomist's Glossary of Genitalia in Insects*. 2nd ed. Copenhagen: Munksgaard, Copenhagen, p. 359.

Yeates, D. K., A. Seago, L. Nelson, S. L. Cameron, L. Joseph, and J. W. H. Trueman (2011). Integrative taxonomy, or iterative taxonomy? *Systematic Entomology* 36, 209–217. DOI: 10.1111/j.1365-3113.2010.00558.x.

Zelditch, M., D. Swiderski, H. Sheets, and W. Fink (2004). *Geometric Morphometrics for Biologists. A Primer*. San Diego, CA: Elsevir Academic Press, p. 416.

Danksagung

Während der Anfertigung meiner Dissertation begleiteten mich sowohl beruflich als auch privat viele Personen, die an dieser Stelle erwähnt sein sollen. Allen voran möchte ich Dirk Ahrens danken, der mich bei meiner Arbeit betreut hat und das Projekt überhaupt erst ermöglicht hat. Er war immer für Diskussionen und Problemlösungen zur Stelle und seine jahrelange Vorarbeit vor allem an der Taxonomie der Sericini schaffte eine essentielle Grundlage für die vorliegende Arbeit.

Bernhard Misof hat sich schon sehr früh bereit erklärt, das Erstgutachten für meine Arbeit zu schreiben, brachte neue Ideen in das Projekt und war immer bereit zu helfen. Vielen Dank dafür! Ebenso möchte ich den weiteren Kommissionsmitgliedern, Susanne Dobler, Maximilian Weigend und Torsten Wappler dafür danken, dass sie sich bereit erklärt haben, diese vielschichtige und interdisziplinäre Arbeit zu begutachten.

Ich danke ebenso den Koautoren der Artikel für ihre wertvollen Beiträge zu den begleitenden Publikationen und für die Bereitstellung der von ihnen gesammelten Tiere. Auch den Kuratoren der von mir besuchten naturhistorischen Sammlungen möchte ich für ihr Vertrauen mir gegenüber danken, mit dem sie mir Tiere zur Bestimmung und Präparation geliehen haben. Unter diesen ist ganz besonders auch Ruth Müller zu erwähnen, die uns bei unserer Exkursion nach Südafrika herzlich bei sich aufnahm.

Meine Arbeit im Molekularlabor wäre ohne Claudia Etzbauer um ein Vielfaches schwieriger gewesen. Vielen Dank für die immer sehr nette Einweisung und Hilfe bei Problemen! Auch dem Rest vom Laborteam und der Arbeitsgruppe Coleoptera gebührt Dank für ihre ständige Bereitschaft, bei Problemen zu helfen und für ein sehr gutes Arbeitsklima!

Finanzielle Unterstützung erhielt ich von der Deutschen Forschungsgemeinschaft (Grant AH175/3), dem SYNTHESYS Projekt und der Alexander Koenig Gesellschaft. Dem Zoologischen Forschungsmuseum A. Koenig danke ich für die Bereitstellung der Forschungsinfrastruktur und für meinen guten Arbeitsplatz.

Zu guter Letzt schließe ich mit Dank an meine Familie, die immer bedingungslos zu mir gehalten und mir den Rücken gestärkt hat. Ganz besonders meine Frau Steffi hat während dieser Zerreißprobe für das soziale Leben trotz zu vieler entfallener Wochenenden und Urlauben immer zu mir gehalten.

Chapter X.

Appendix

A. Appendix to chapter II

A.1. Supplementary Figures

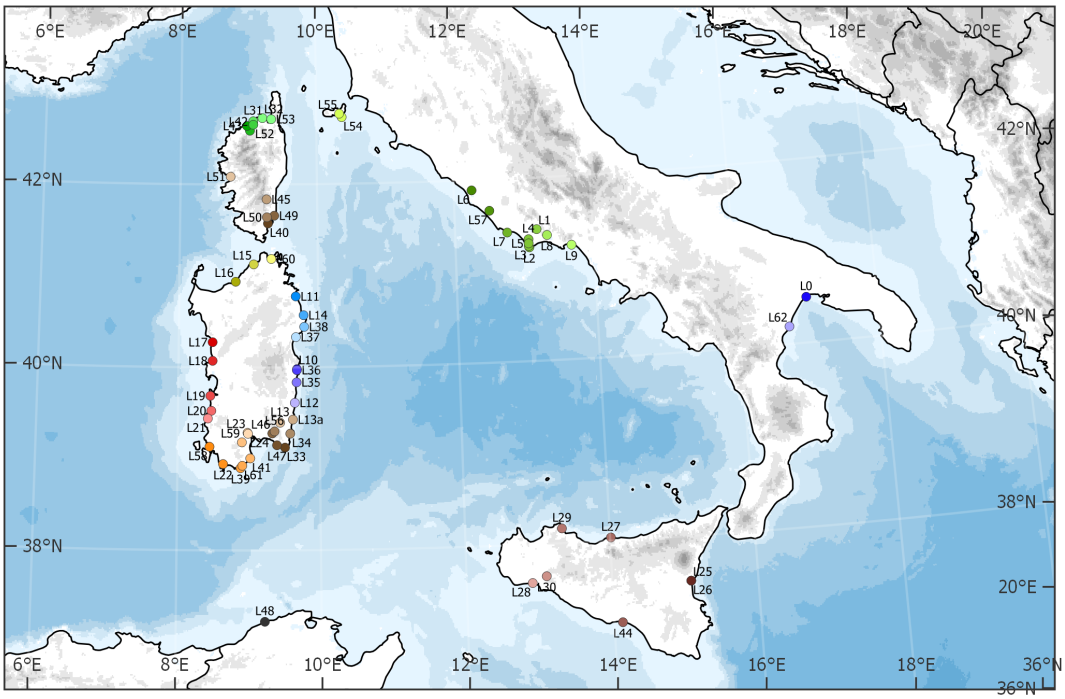

Figure A1. Map of the central Mediterranean Sea depicting sampling localities of all studied *Pachypus* specimen. Abbreviations refer to Table A1.

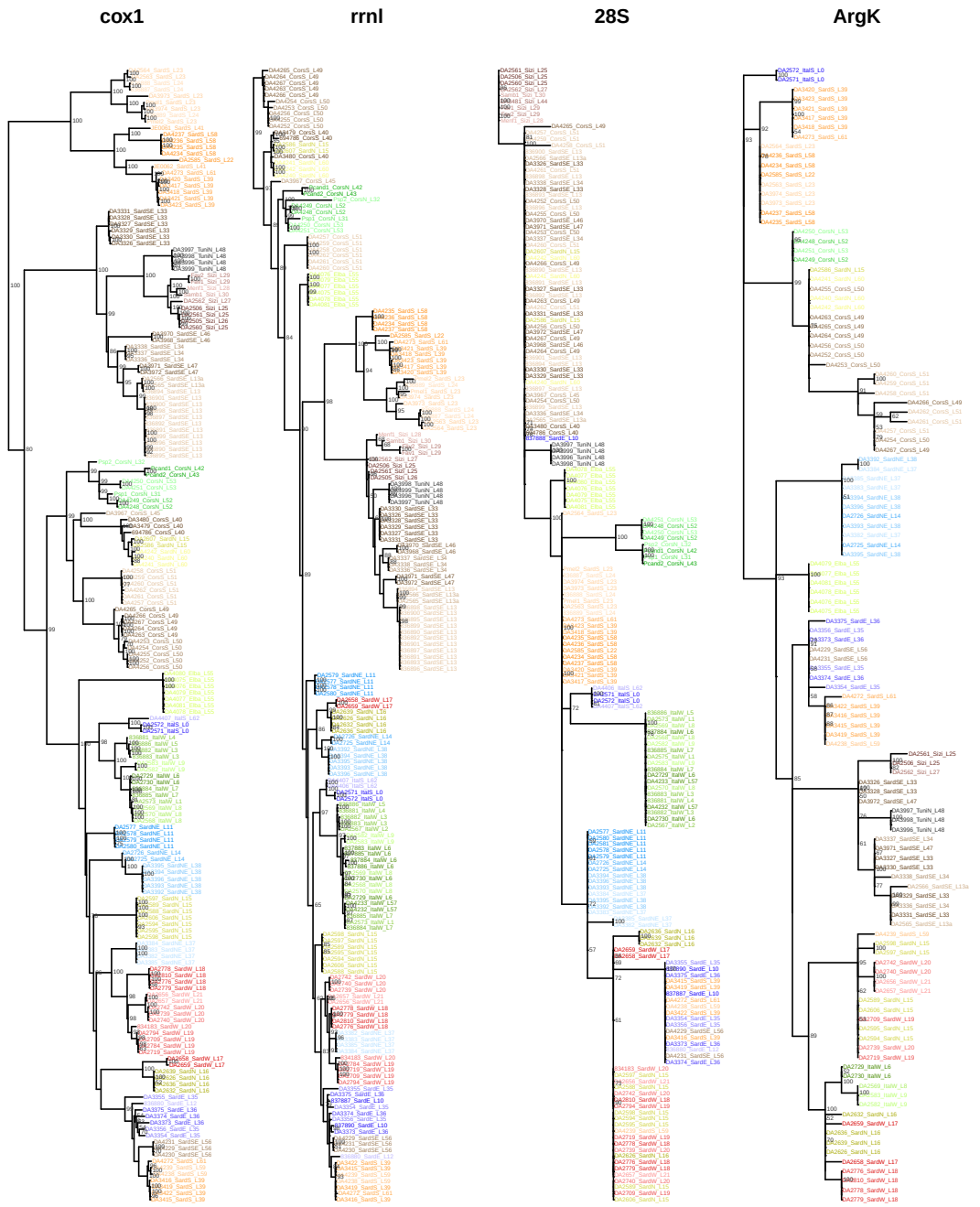

Figure A2. Maximum likelihood trees from RAxML analyses on single markers. The trees were rooted with outgroup specimens which are not shown. RELL-bootstrap support values >50 are shown at the nodes. Sampling localities are color-coded on the maps (Fig. II.5, A1) and given at the tip-labels.

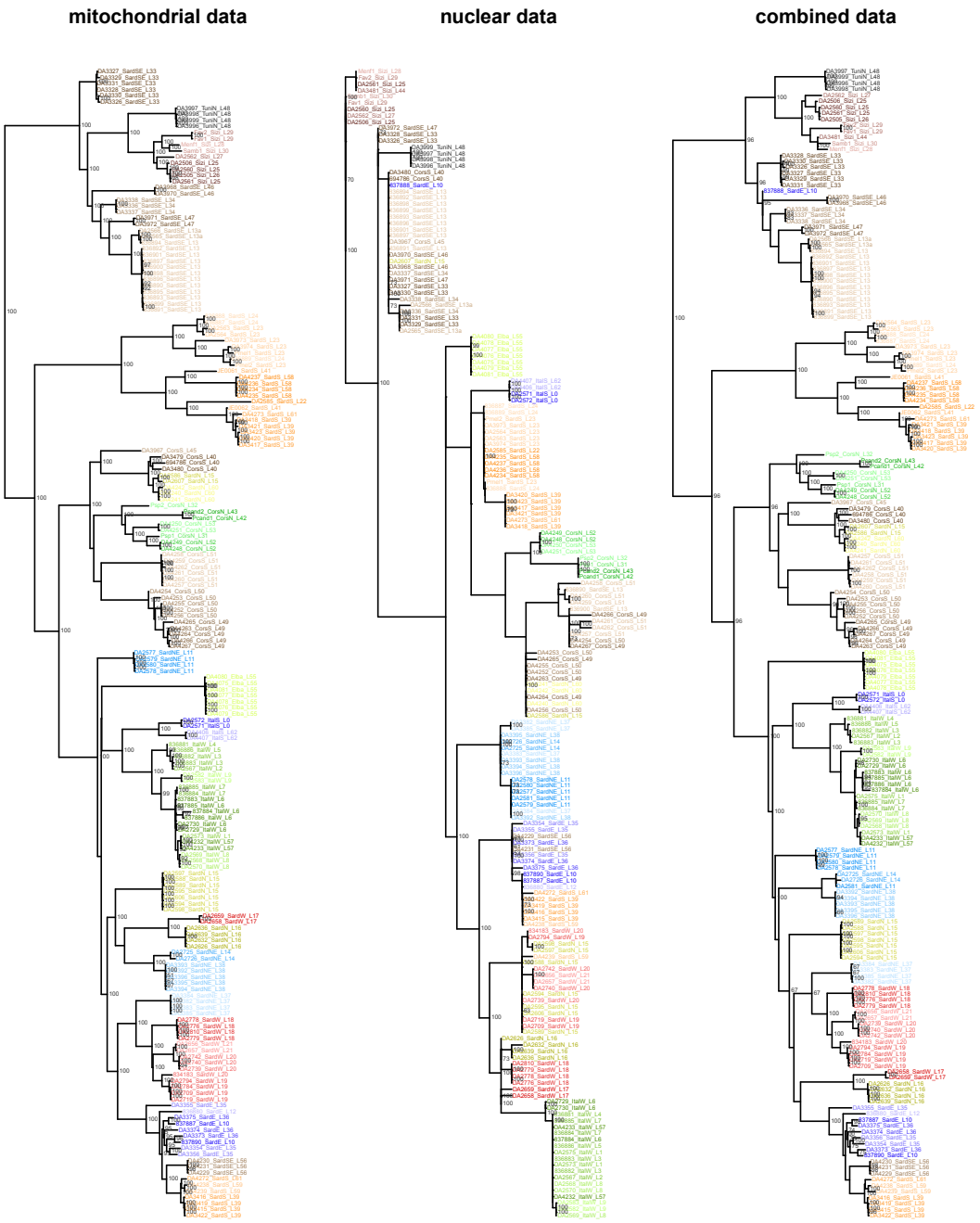

Figure A3. Maximum likelihood trees from RAxML analyses on combined partitioned mitochondrial *(cox1, rrnL)*, combined partitioned nuclear (28S, ArgK), and all combined partitioned loci. The trees were rooted with outgroup specimens which are not shown. RELL-bootstrap support values >50 are shown at the nodes. Sampling localities are color-coded on the maps (Fig. II.5, A1) and given at the tip-labels.

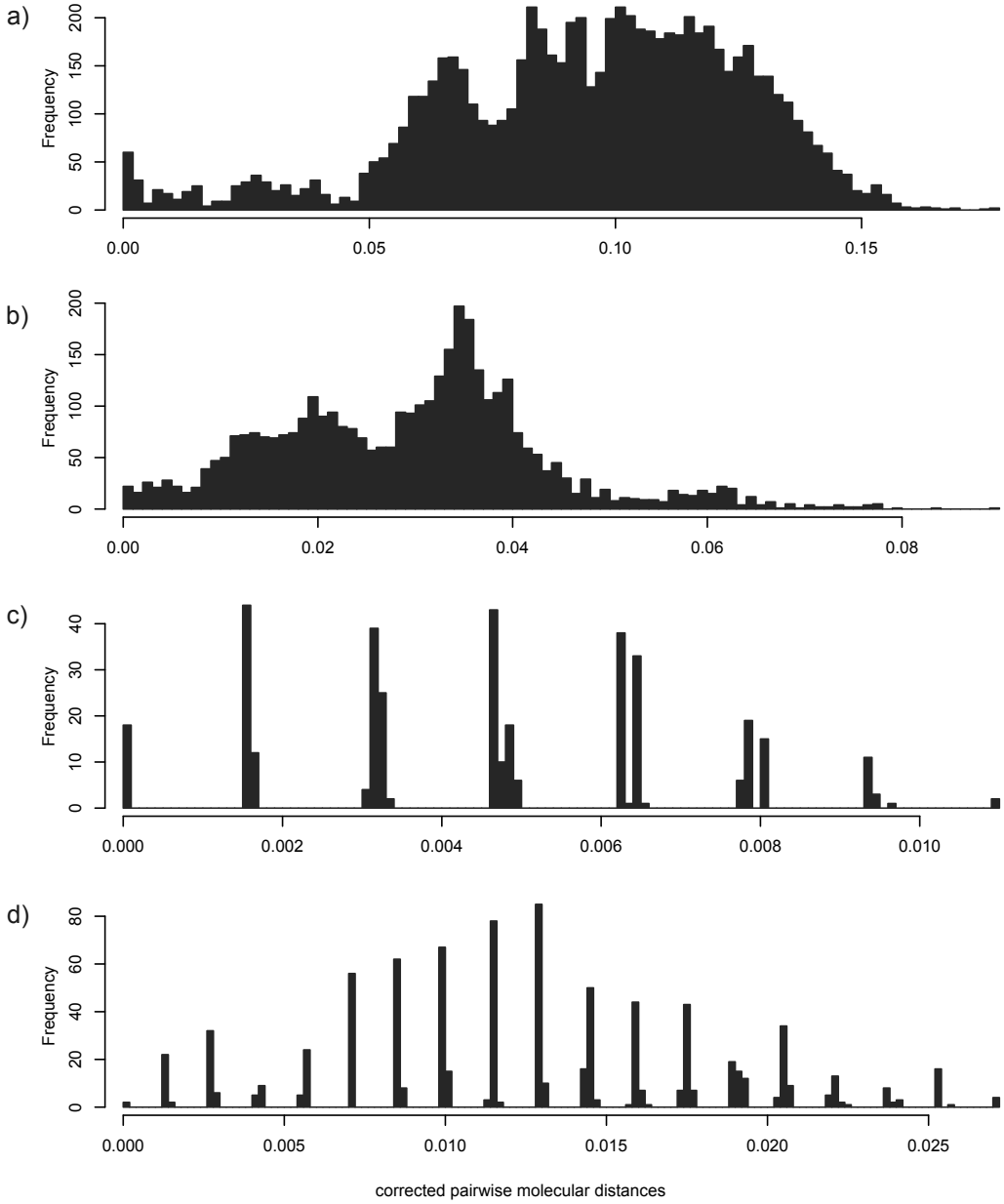

Figure A4. Histograms of corrected pairwise distances that were used as input for ABGD for markers (a) *cox1*, (b) *rrnL*, (c) 28S, and (d) ArgK. The distances were corrected by the best fitting substitution model that was inferred with IQ-TREE (Table A2).

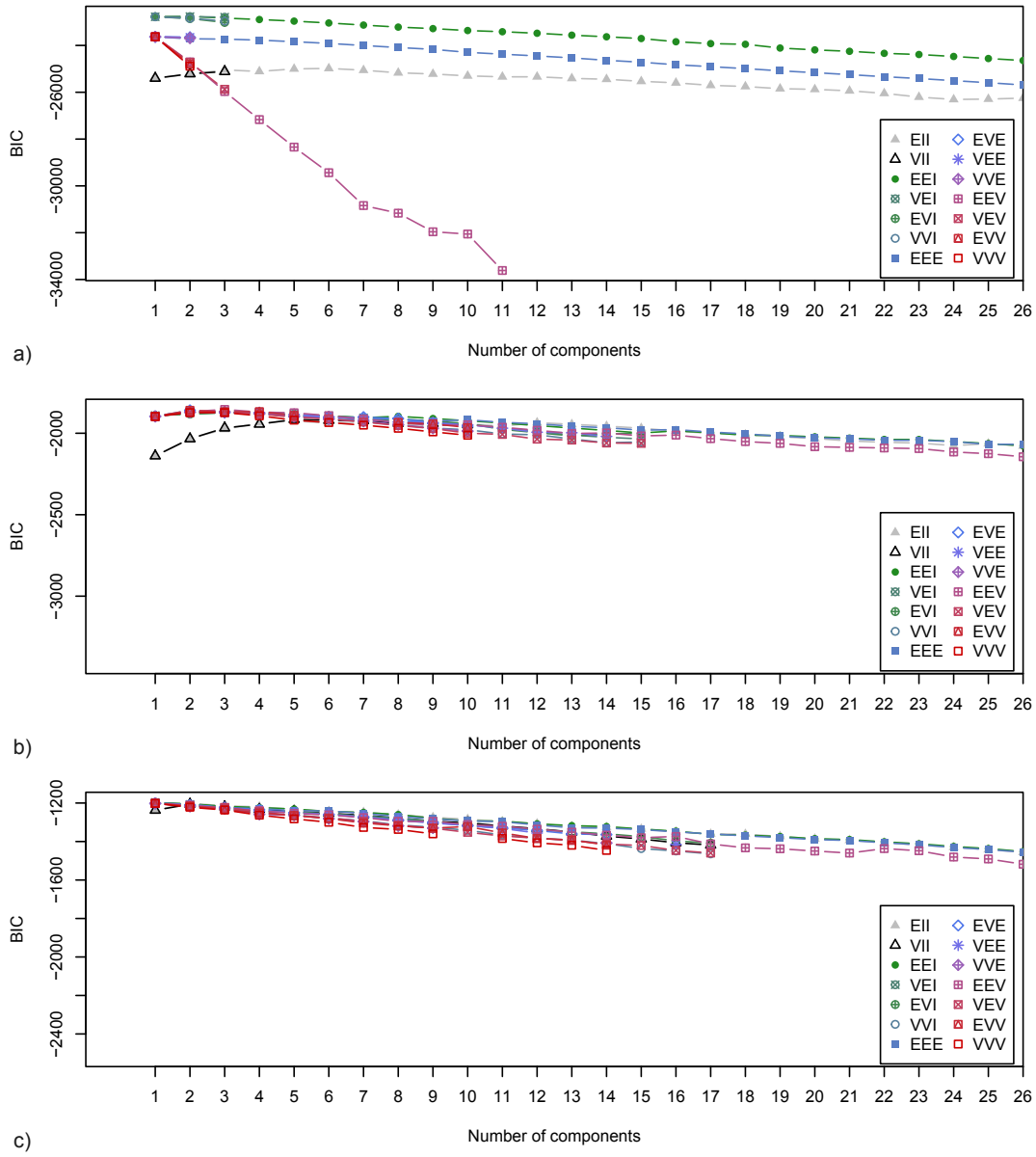

Figure A5. BIC scores of multivariate mixture models evaluated by Gaussian clustering with Mclust for (a) all trait data, (b) all data with prior variable selection, and (c) linear measurements. EII = spherical, equal volume; VII = spherical, unequal volume; EEI = diagonal, equal volume and shape; VEI = diagonal, varying volume, equal shape; EVI = diagonal, equal volume, varying shape; VVI = diagonal, varying volume and shape; EEE = ellipsoidal, equal volume, shape, and orientation; EVE = ellipsoidal, equal volume and orientation; VEE = ellipsoidal, equal shape and orientation; VVE = ellipsoidal, equal orientation; EEV = ellipsoidal, equal volume and equal shape; VEV = ellipsoidal, equal shape; EVV = ellipsoidal, equal volume; VVV = ellipsoidal, varying volume, shape, and orientation.

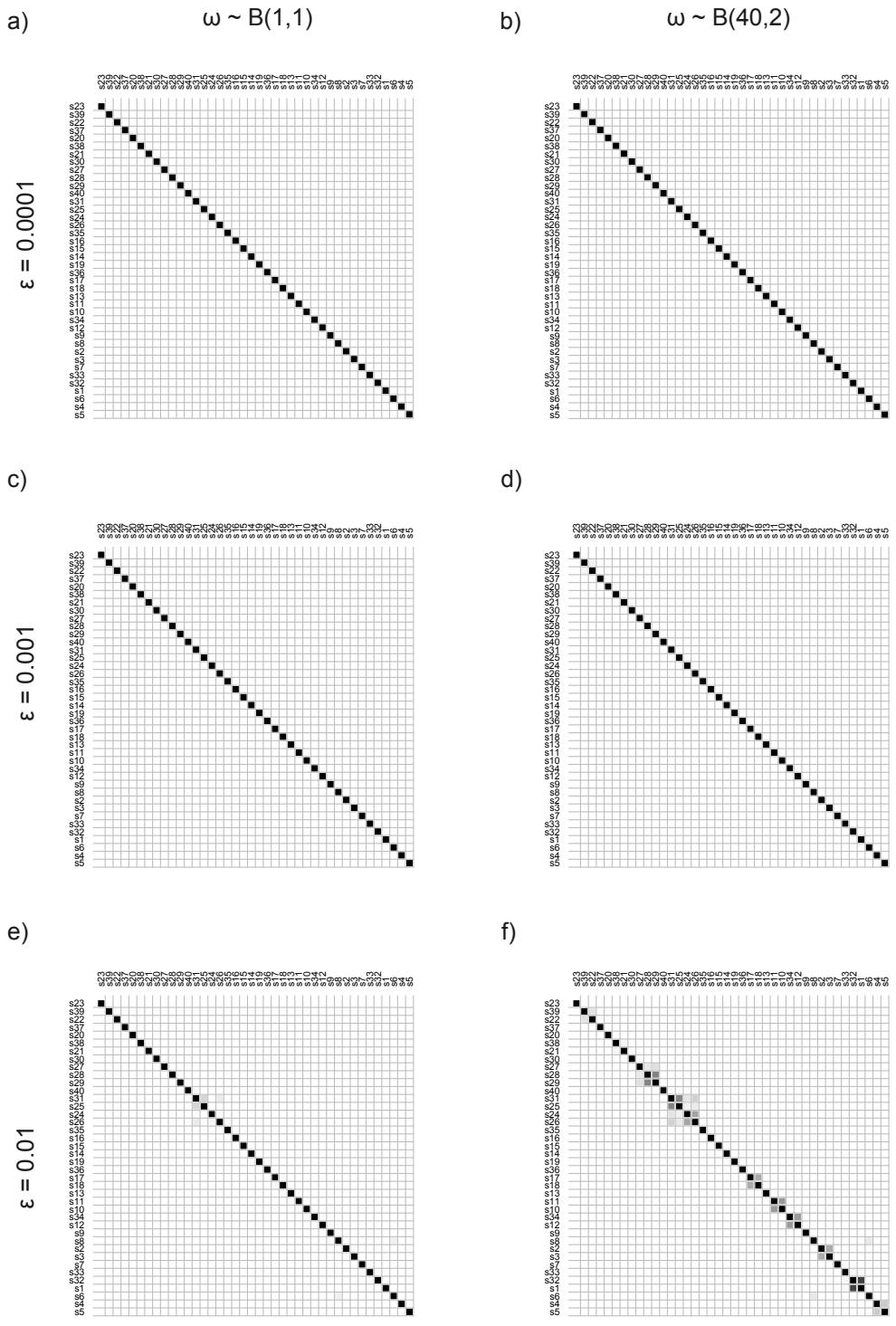

Figure A6. Similarity matrices illustrating the results from STACEY analyses of MINCs (GMYC clusters) under different collapse weights (a, c, e vs. b, d, f) and increasing collapse heights (a, b vs. c, d vs. e, f). Darker shaded pairs are more similar to each other.

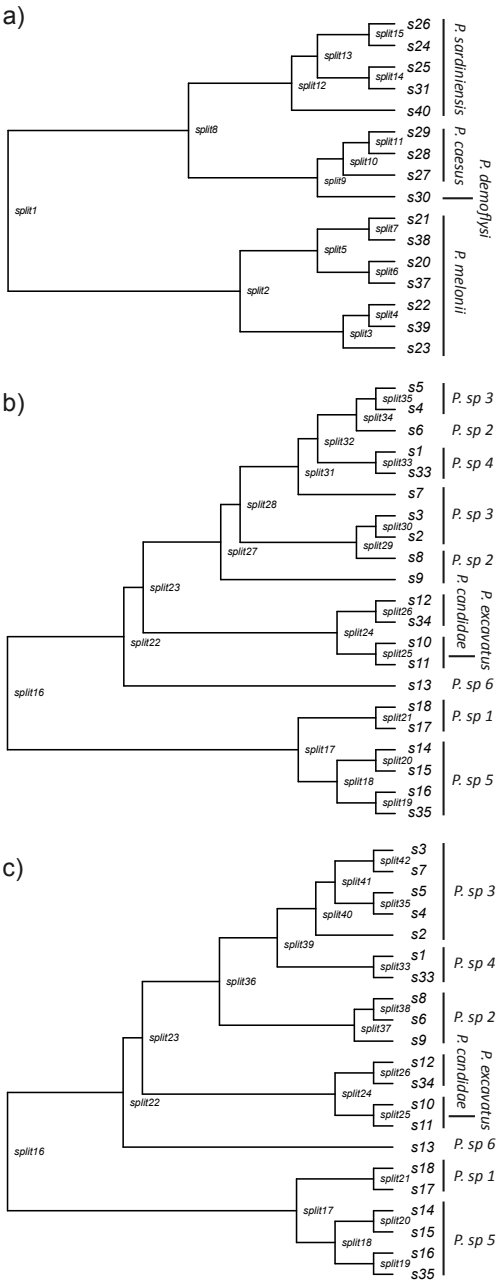

Figure A7. Guide trees used for iBPP analyses. (a) Guide tree part 1 (Fig. II.2, nodes A1+A2). (b) Unmodified guide tree part 2 (Fig. II.2, node B). (c) Geography-informed guide tree part 2.

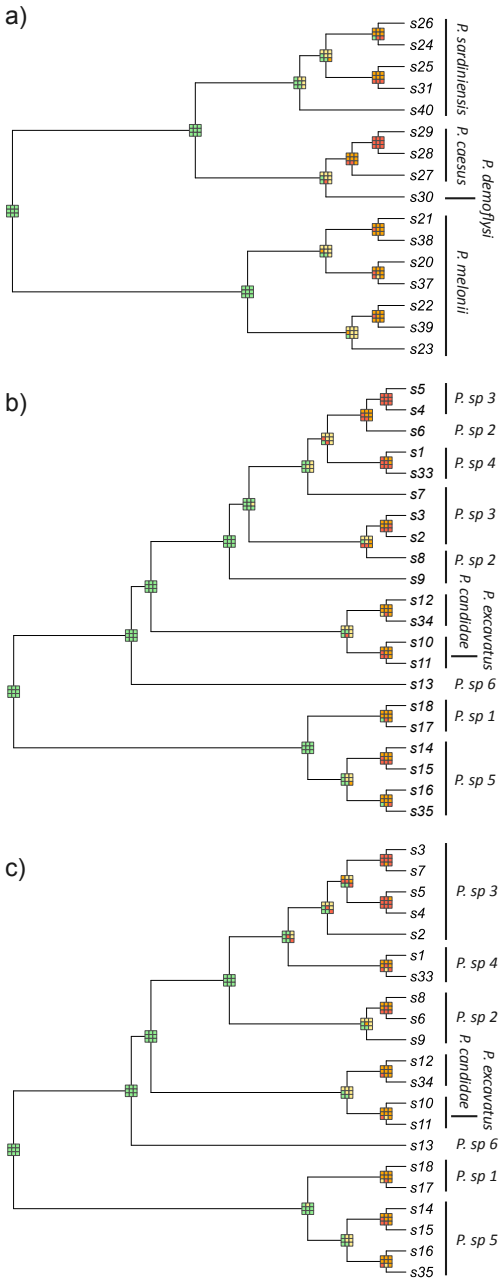

Figure A8. Results of iBPP analyses without data, i.e., prior sampling for all guide trees. (a) Guide tree part 1, (b) unmodified guide tree part 2, and (c) geography-informed guide tree part 2.

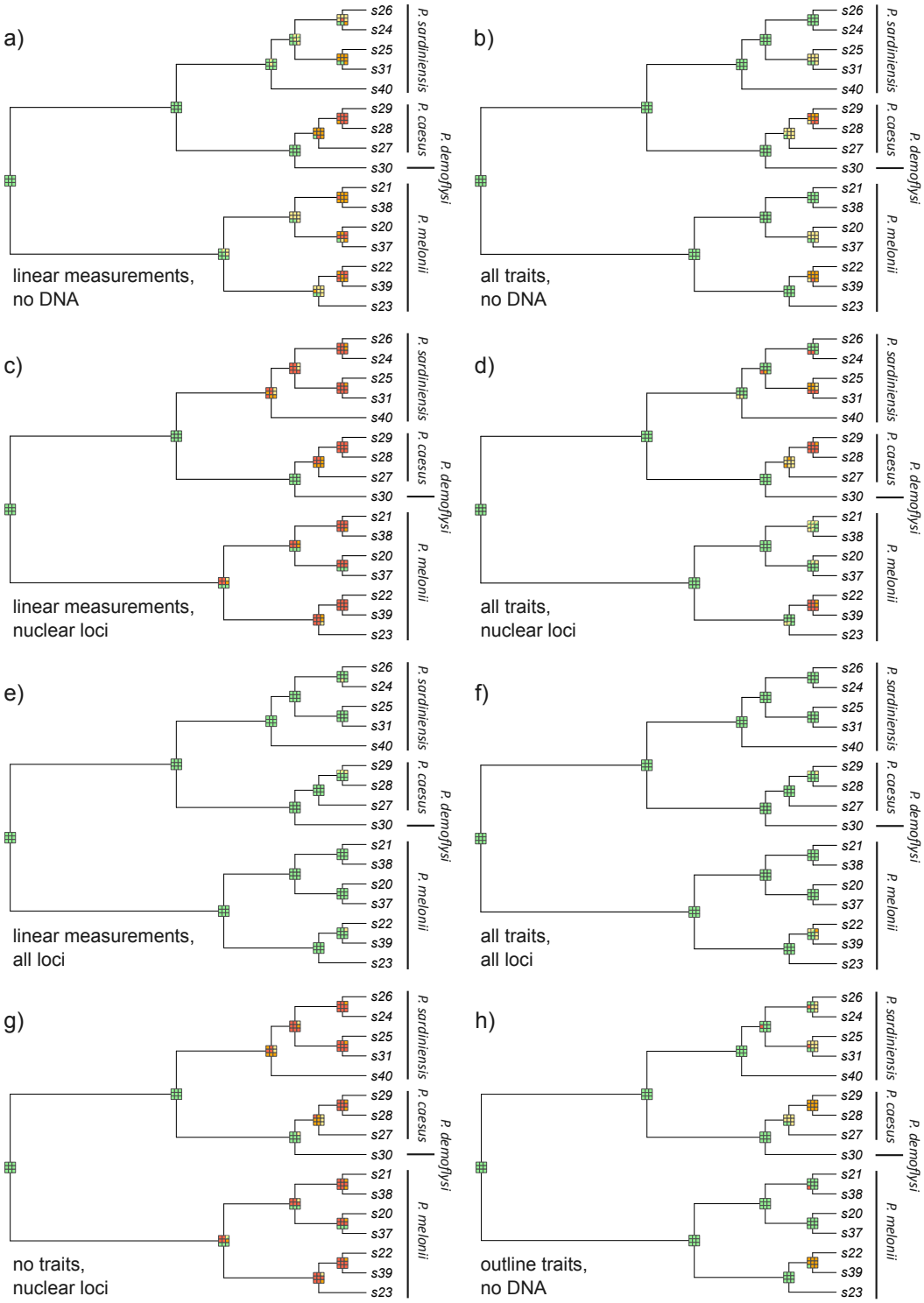

Figure A9. Comprehensive results of integrative Bayesian species delimitation with iBPP using eight datasets for guide tree part 1. Posterior probabilities of 3×3 combinations of τ_0 and θ prior distributions are illustrated at each node (see graphical legend in Fig. II.4). Minimum clusters from the GMYC analyses (Fig. II.2) are given at the tips of each tree along with the final species designations.

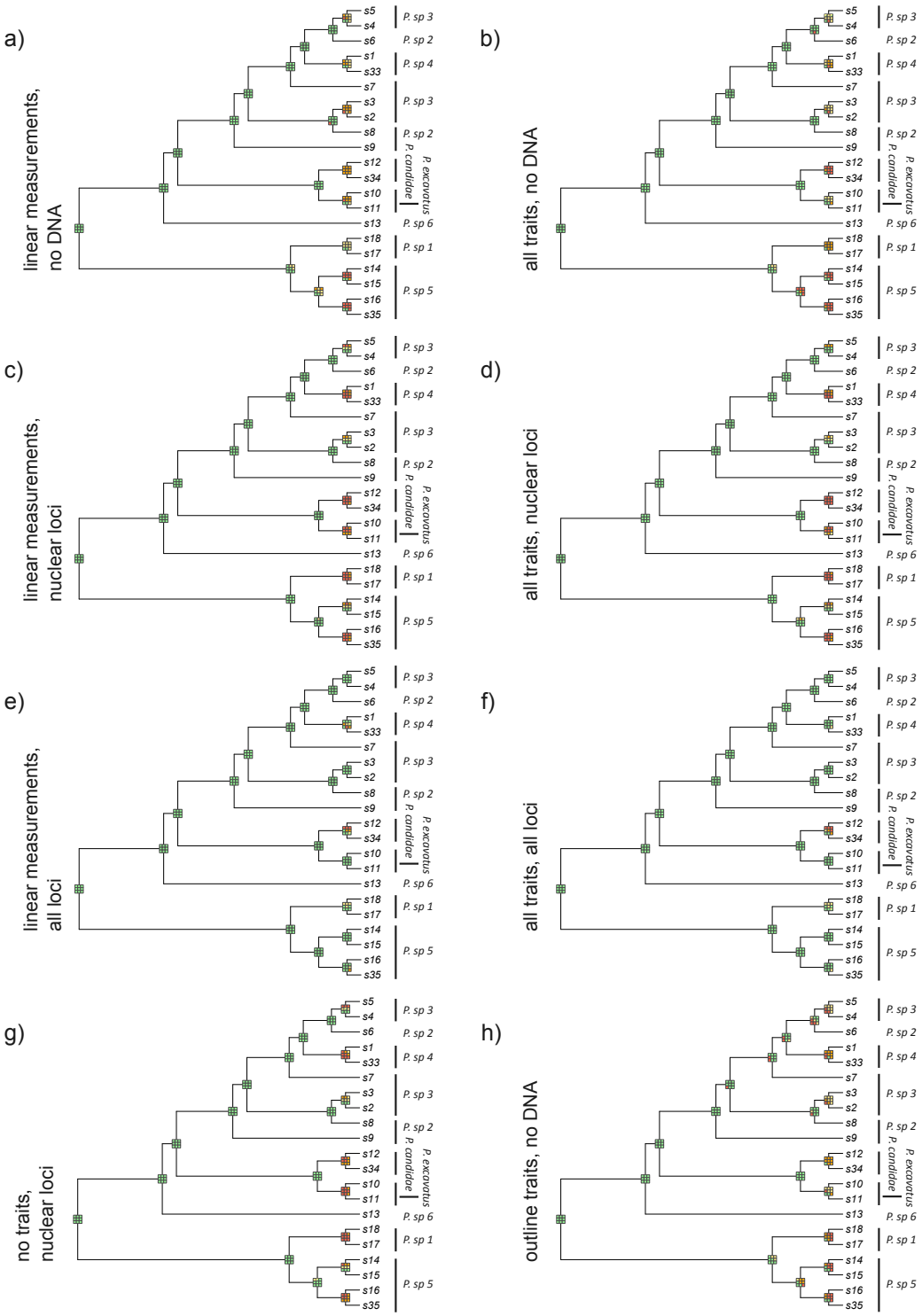

Figure A10. Comprehensive results of integrative Bayesian species delimitation with iBPP using eight datasets sampling for the unmodified guide tree part 2. Posterior probabilities of 3×3 combinations of τ_0 and θ prior distributions are illustrated at each node (see graphical legend in Fig. II.4). Minimum clusters from the GMYC analyses (Fig. II.2) are given at the tips of each tree along with the final species designations.

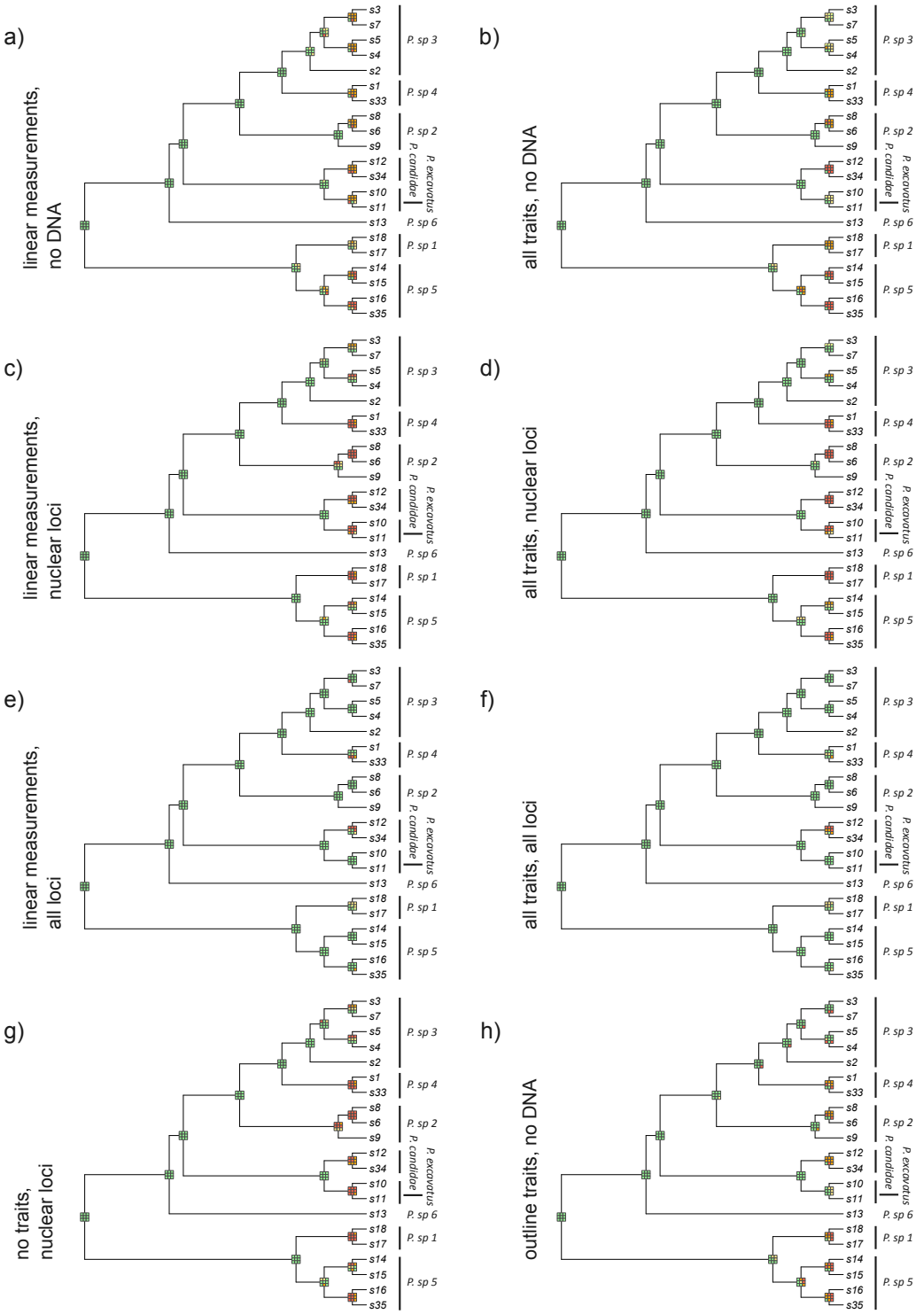

Figure A11. Comprehensive results of integrative Bayesian species delimitation with iBPP using eight datasets for the geography-informed guide tree part 2. Posterior probabilities of 3×3 combinations of τ_0 and θ prior distributions are illustrated at each node (see graphical legend in Fig. II.4). Minimum clusters from the GMYC analyses (Fig. II.2) are given at the tips of each tree along with the final species designations.

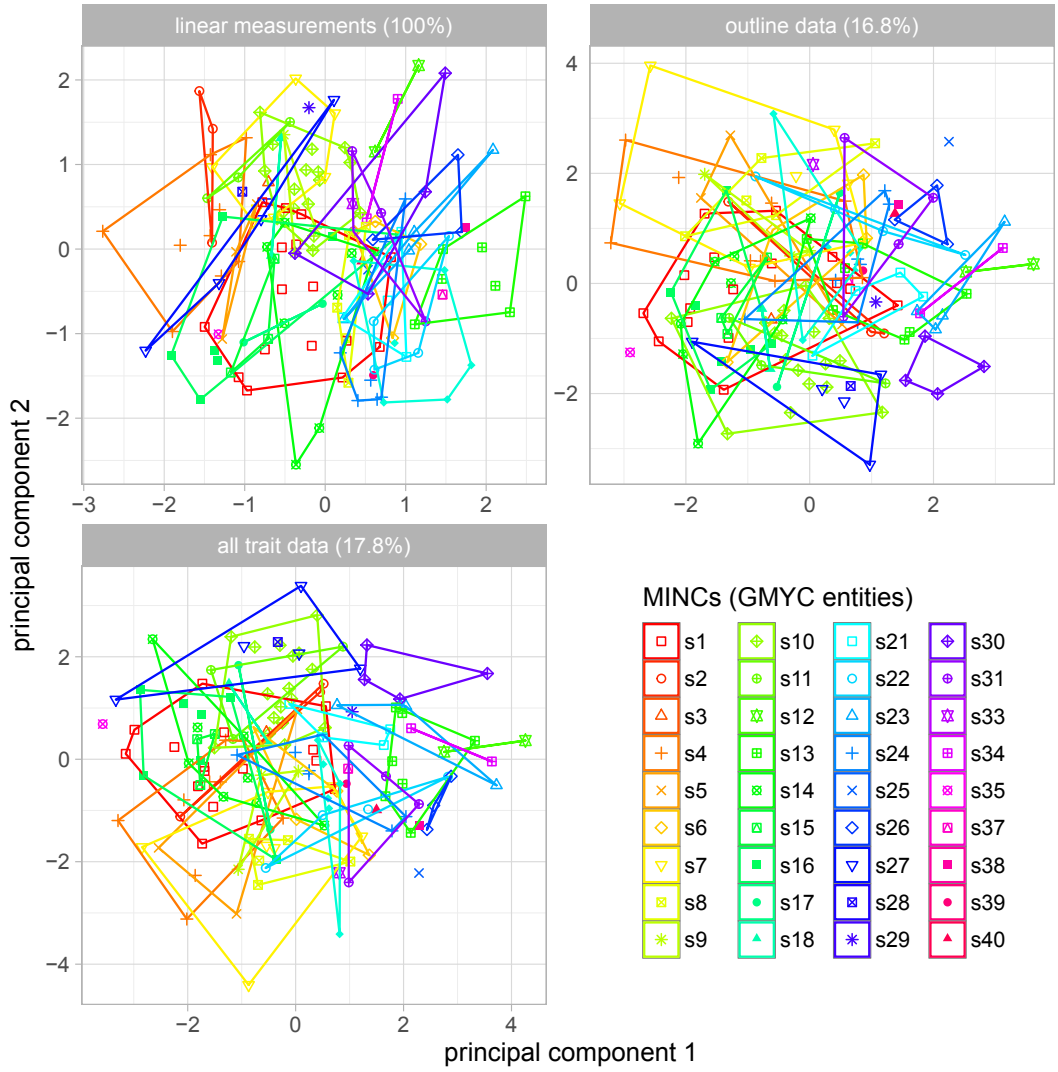

Figure A12. Minimum clusters (MINCs) from the GMYC analysis of the combined molecular data (Fig. II.2) that were used for species delimitation by validation approaches mapped on the three morphological datasets employed. Percentages of variation depicted by principal components 1 and 2 are given in subfigure headings.

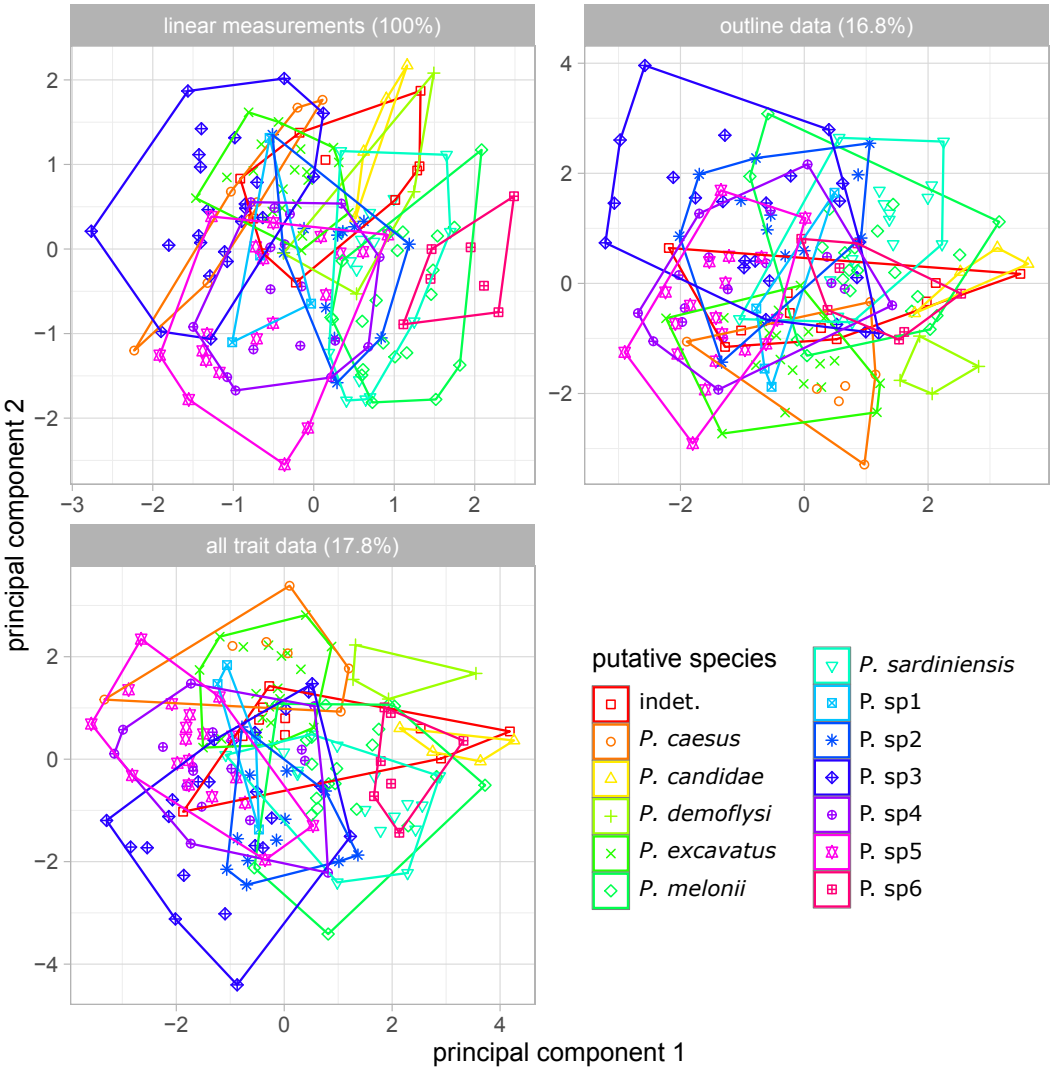

Figure A13. Putative species that were inferred with iBPP mapped on the three morphological datasets employed. Percentages of variation depicted by principal components 1 and 2 are given in subfigure headings.

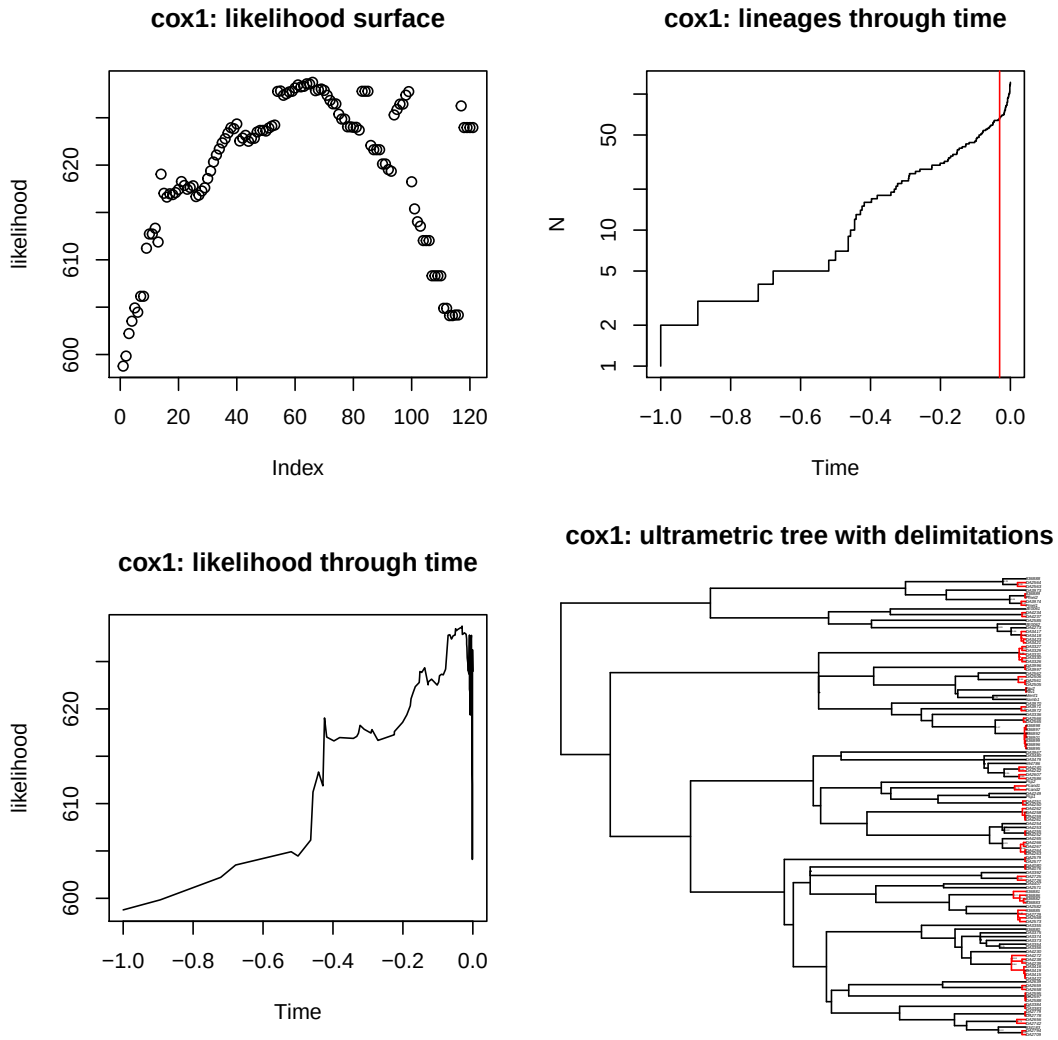

Figure A14. GMYC results for *cox1*.

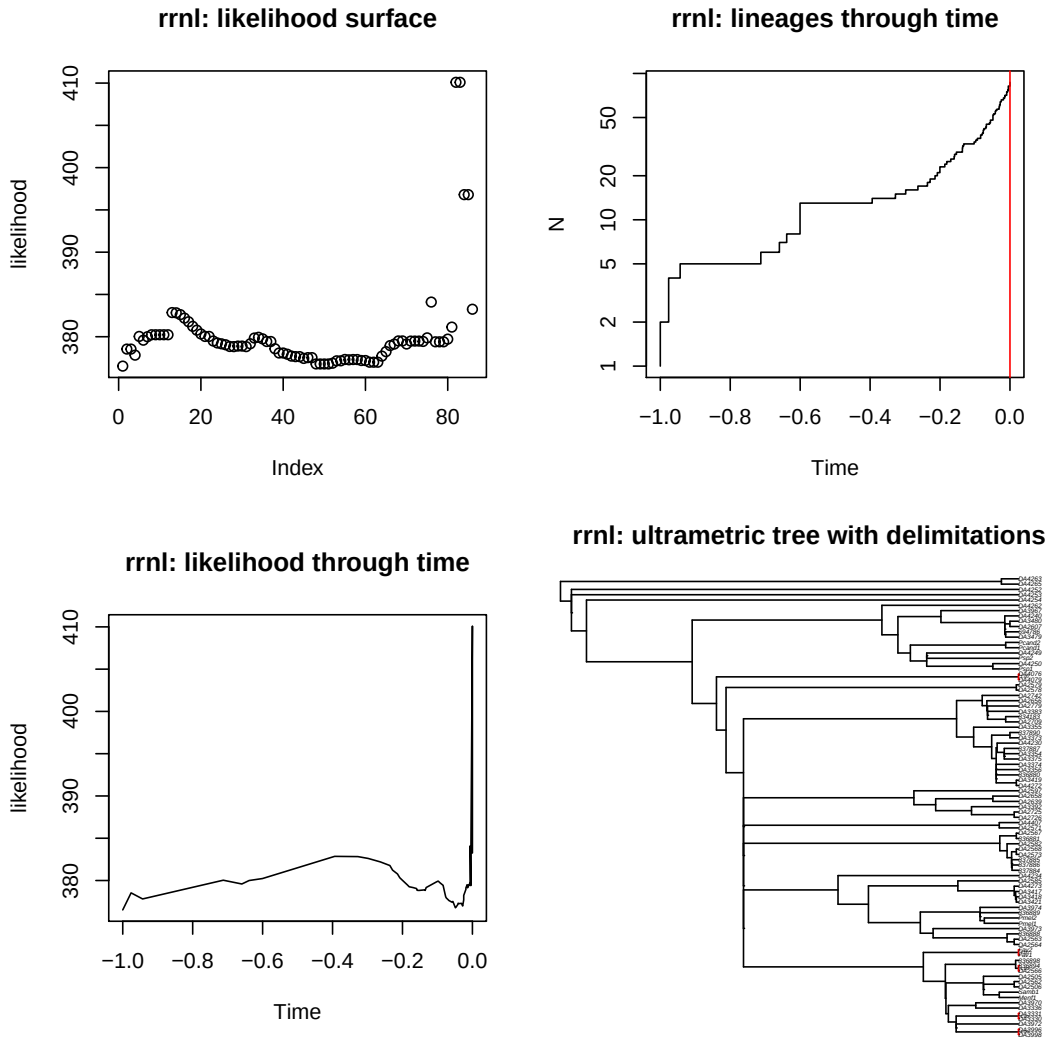

Figure A15. GMYC results for *rrnL*.

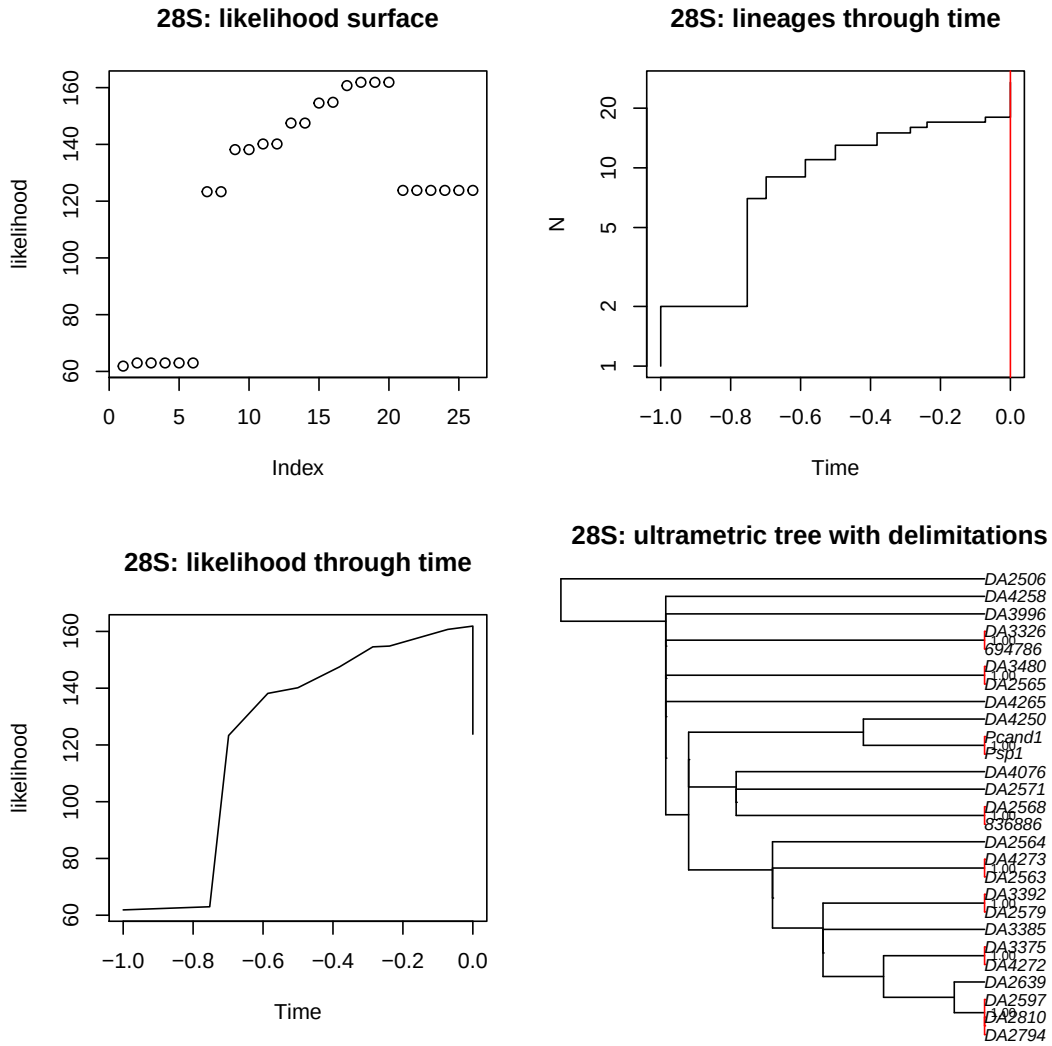

Figure A16. GMYC results for 28S.

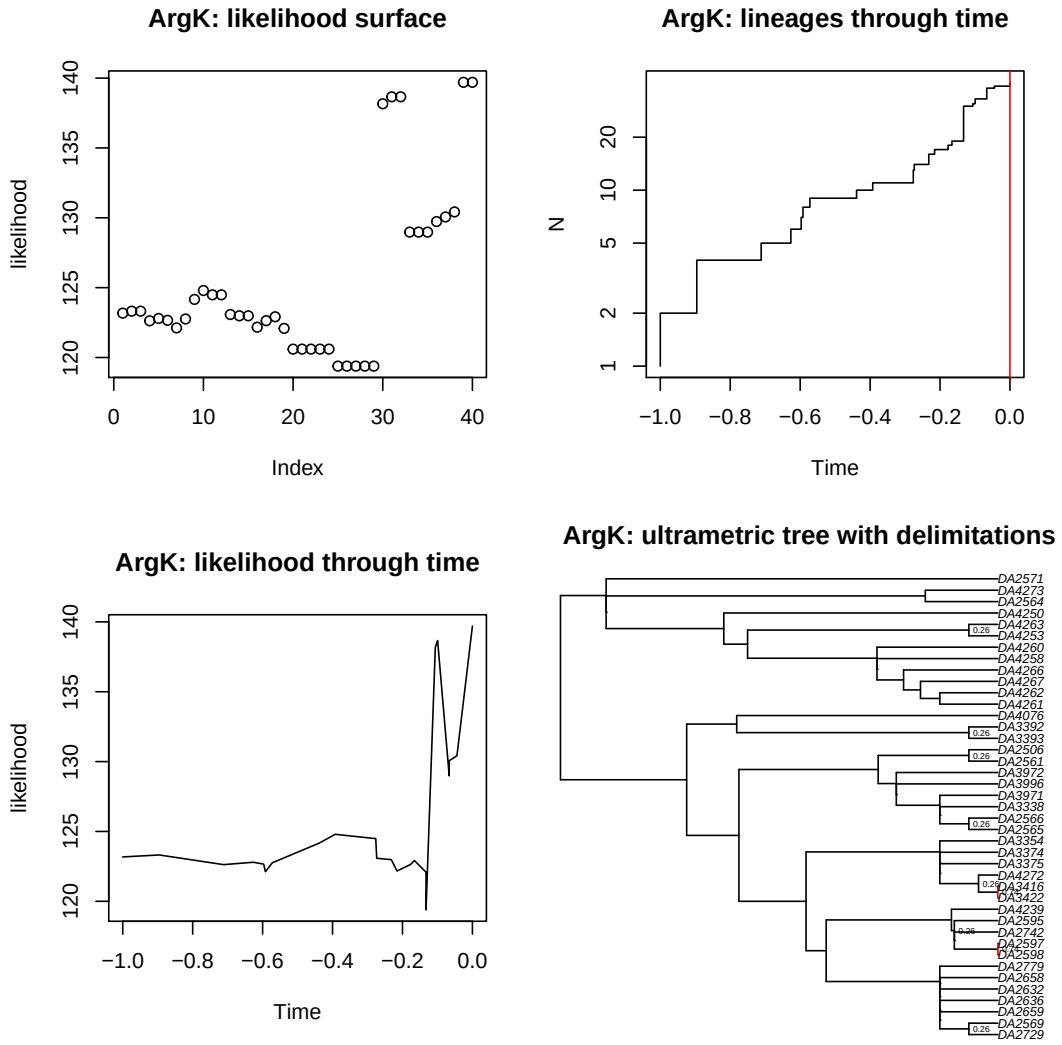

Figure A17. GMYC results for ArgK.

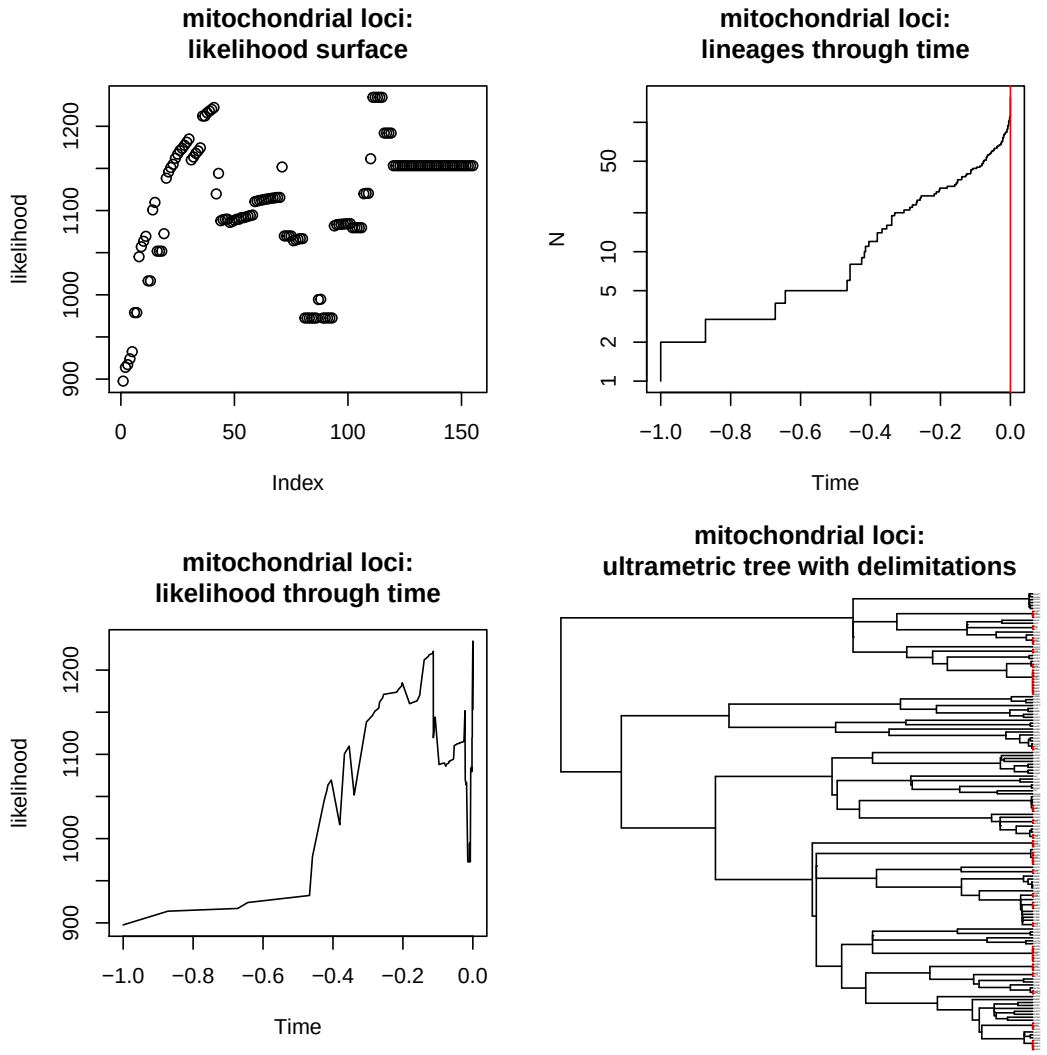

Figure A18. GMYC results for combined mitochondrial loci *(cox1 + rrnL)*.

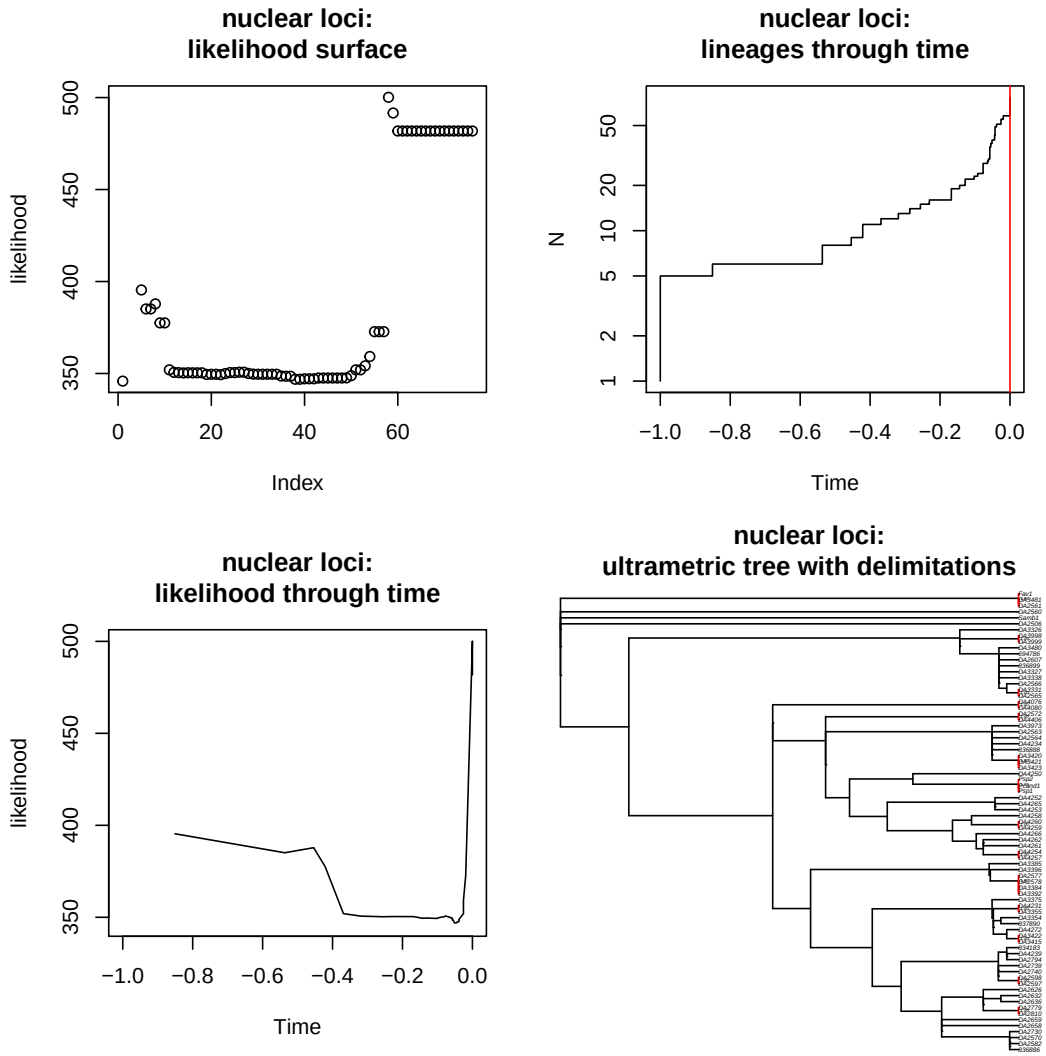

Figure A19. GMYC results for combined nuclear loci (28S + ArgK).

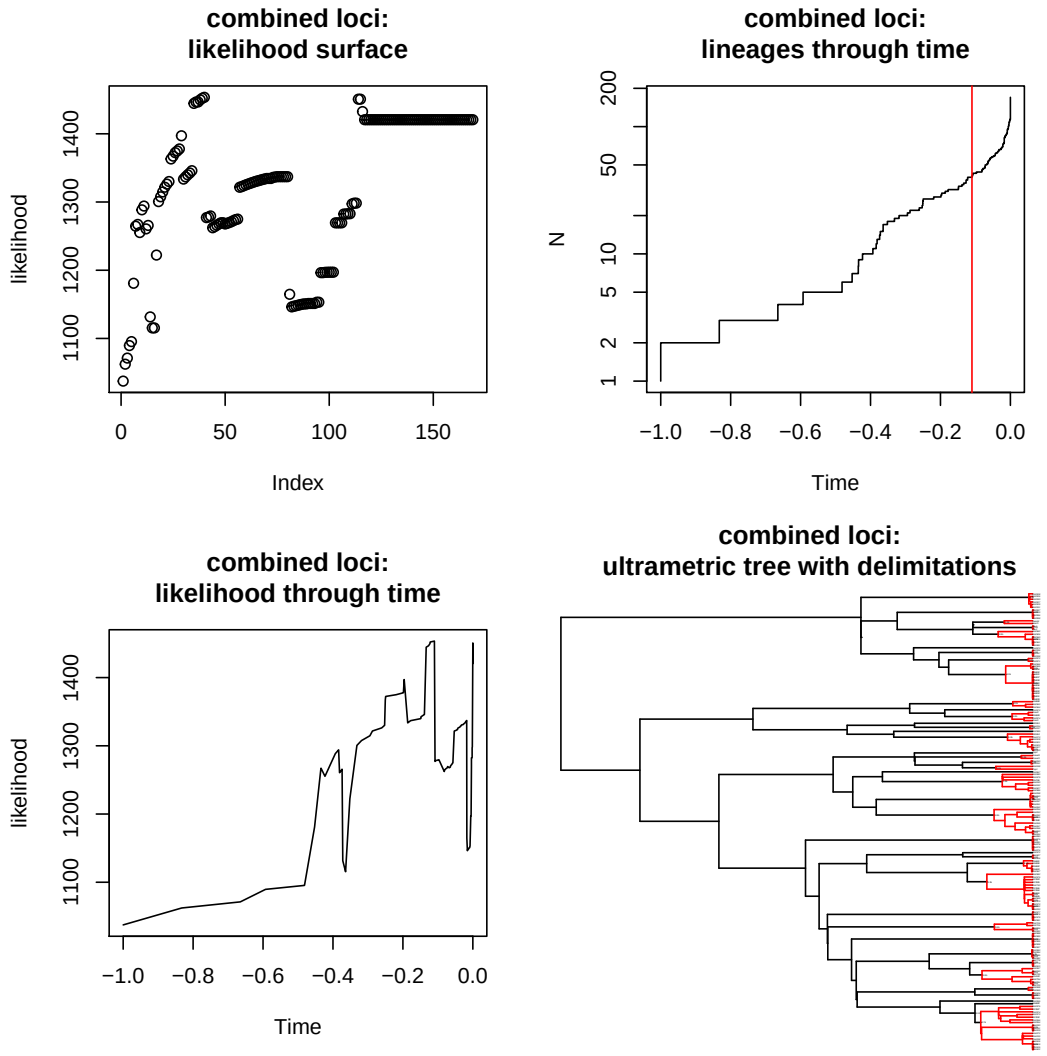

Figure A20. GMYC results for all loci combined.

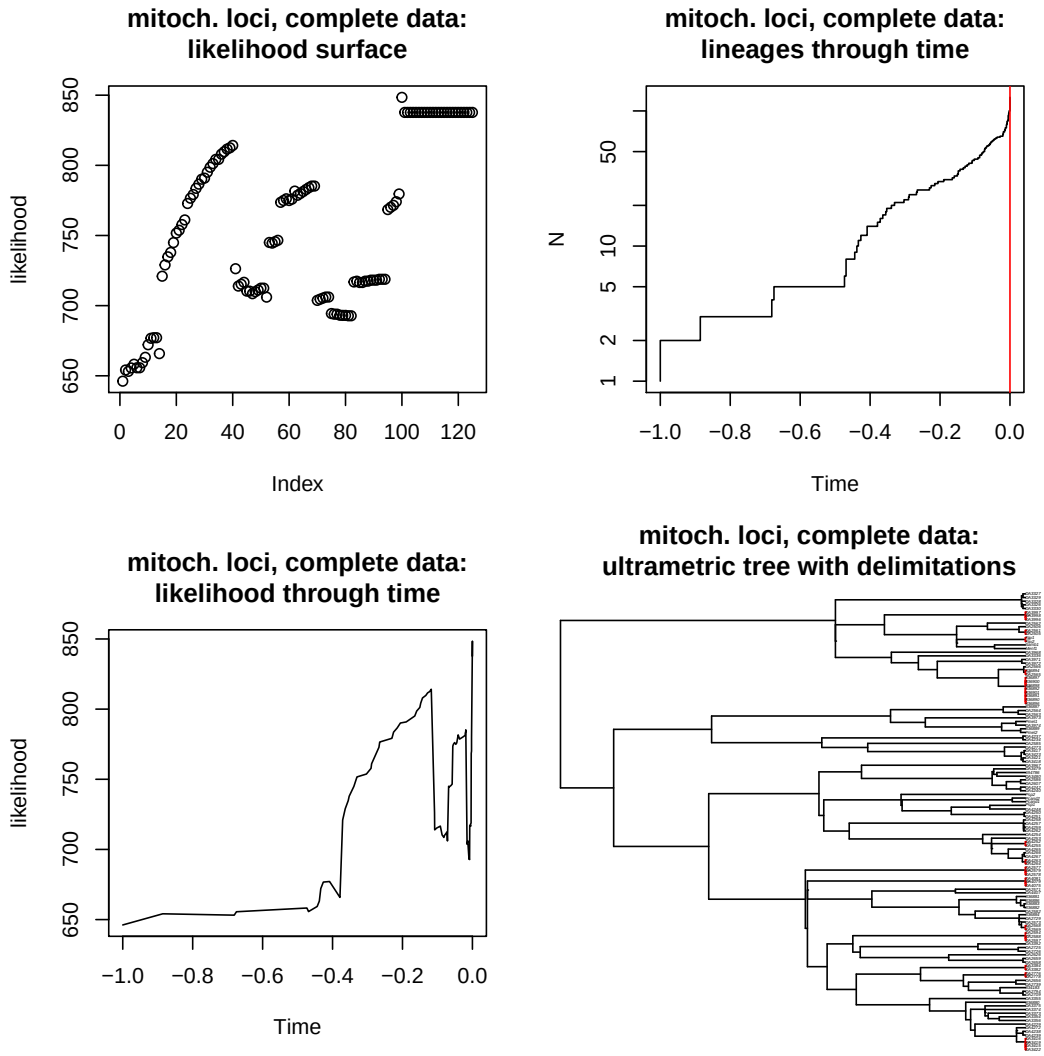

Figure A21. GMYC results for combined mitochondrial loci *(cox1 + rrnL)*, only using specimens for which both loci were available.

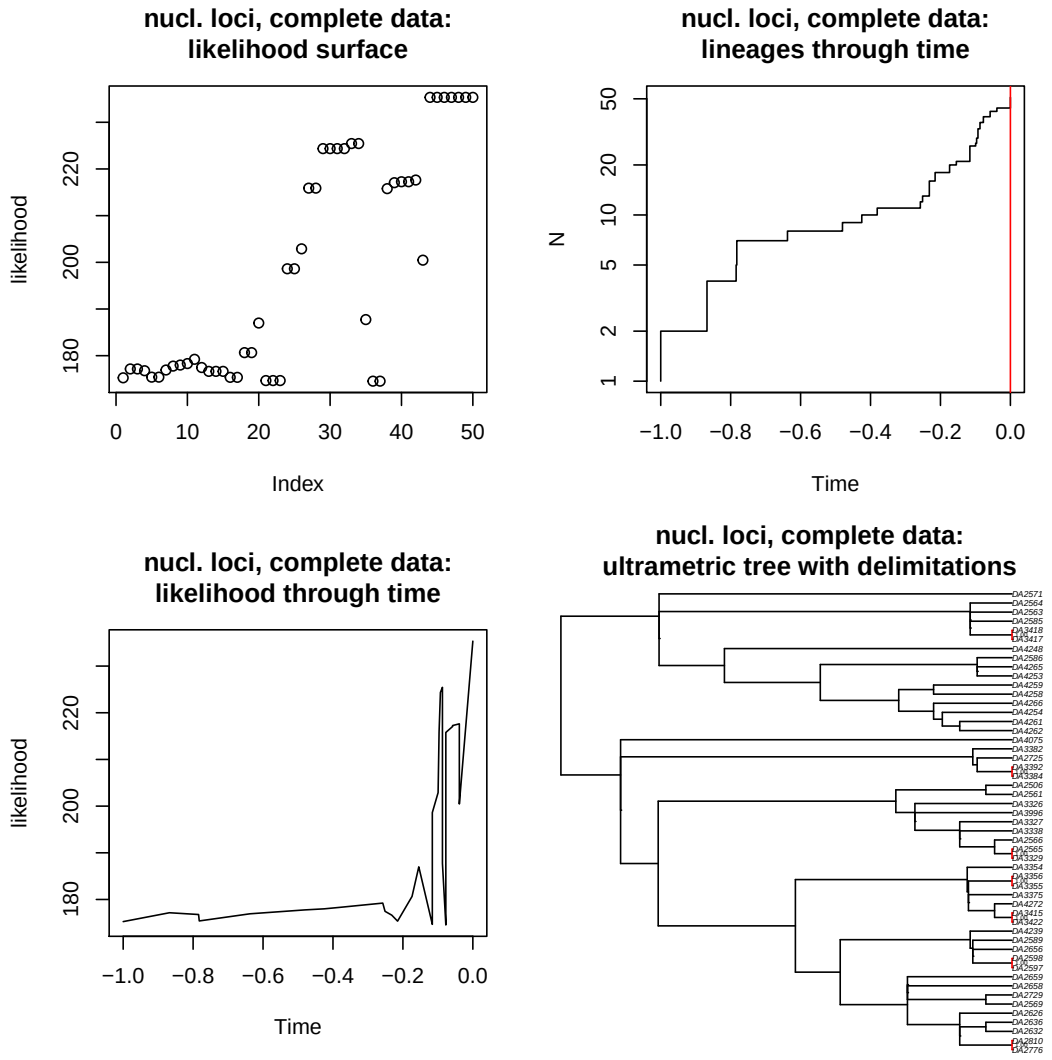

Figure A22. GMYC results for combined nuclear loci (28S + ArgK), only using specimens for which both loci were available.

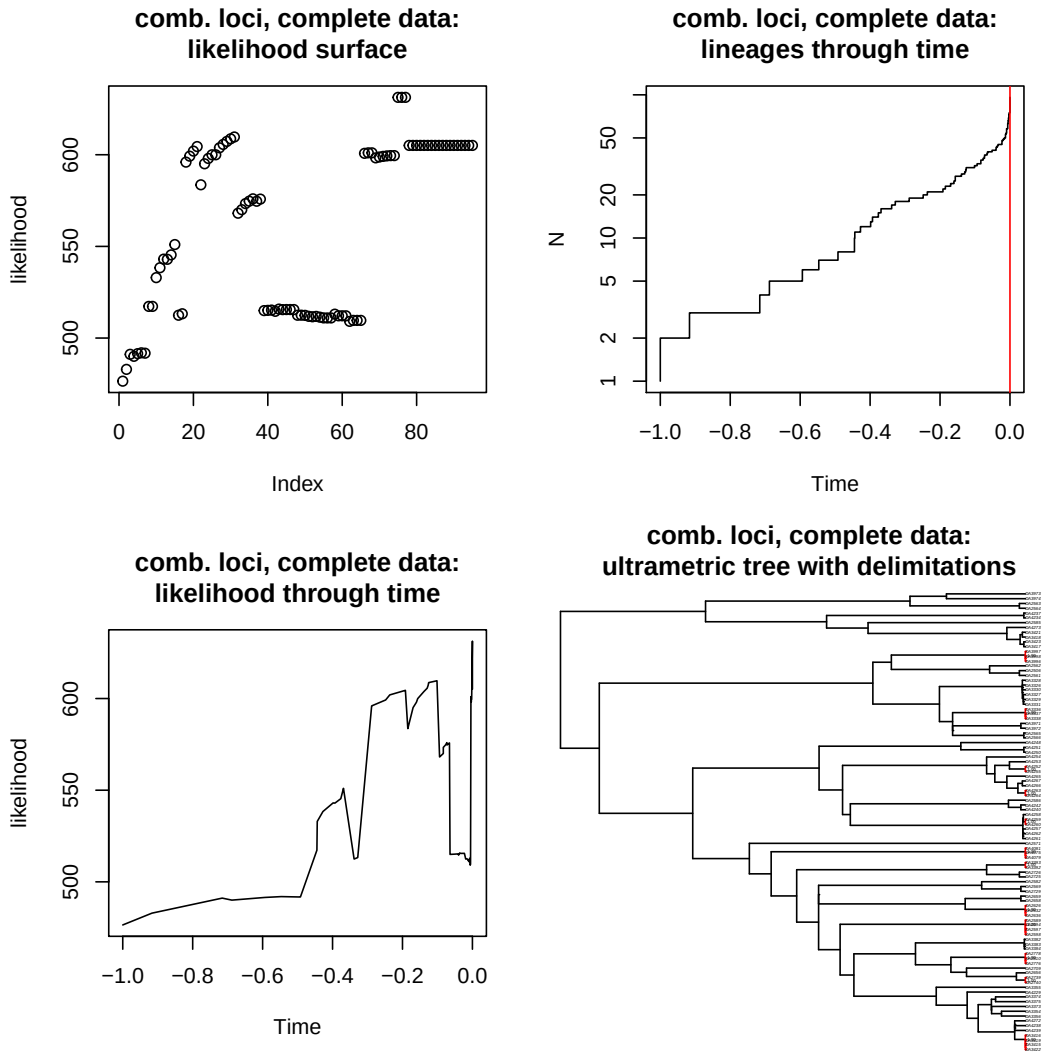

Figure A23. GMYC results for all loci combined, only using specimens for which both loci were available.

A.2. Supplementary Tables

Table A1. Numbers of sampling localities that were used in trees (Fig. II.2, A2, A3) and the locality map (Fig. A1). Coordinates are given in decimal format.

no	sampling locality	latitude	longitude
L0	Italy: Puglia: Fiume Lato	40.495833	16.990833
L1	Italy: (Frosinone Prov.) Bosco Polverino, 40 m	41.436111	13.187222
L2	Italy: (Latina Prov.) Circeo National Park, Mt. Circeo, Quarto Freddo, Peretto (Conecofor site), 120 m	41.240932	13.069646
L3	Italy: Fonte di Lucullo (W Sabaudia)	41.257222	13.061389
L4	Italy: Selva del Circeo, 5 km N of Sabaudia	41.330556	13.062778
L5	Italy: W of Lago di Sabaudia	41.284722	13.061667
L6	Italy: Castel di Guido	41.889167	12.263889
L7	Italy: Pineta di Torre Astura (S Nettuno)	41.413611	12.758333
L8	Italy: (Latina Prov.) Monti Ausoni, Sugherete di San Vito e Valle Marina, San Vito 70 m,	41.366667	13.333333
L9	Italy: Torre Gianola Scauri, 6km E Formia	41.246837	13.680725
L10	Sardinia: 2 km S Santa Maria Navarese	39.980000	9.686667
L11	Sardinia: (OT) San Teodoro, Cala d'Ambra, 2 m,	40.775000	9.677222
L12	Sardinia: S Barusia (5 km S of Marina di Tertenia)	39.611389	9.653056
L13	Sardinia: Foce del Fiumendosa (Muravera)	39.395000	9.443889
L13a	Sardinia: (CA) Muravera	39.428333	9.625000
L14	Sardinia: Cala Pineta, near St. Lucia (Nuoro), 12 m	40.568611	9.788056
L15	Sardinia: Camping Vignola Mare	41.126944	9.073056
L16	Sardinia: Valledoria	40.937000	8.812998
L17	Sardinia: Bosa Marina, 16 m	40.275278	8.484722
L18	Sardinia: Is Arenas, 15 m	40.071667	8.486111
L19	Sardinia: dunes between Torre dei Corsari and Pistis, 40 m	39.688889	8.456944
L20	Sardinia: (W coast): Camping (Scioppadrosciu) 2 km W of Irgutosu, 62 m	39.525278	8.472222
L21	Sardinia: Portixeddu, 0-42 m	39.440556	8.423333
L22	Sardinia: Porto Pino env. (3 km S)	38.943333	8.640556
L23	Sardinia: (CA) Assemini	39.285556	8.983333
L24	Sardinia: betw. Uta and Assemini (fiume Mannu)	39.275000	8.990556
L25	Sicily: Catania (CT): Giardino Bellini	37.512222	15.083056
L26	Sicily: Catania, Piazza S. Maria di Gesu	37.512222	15.083056
L27	Sicily: Cefalù	38.034444	14.014167
L28	Sicily: Menfi (AG) Porto Palo	37.578333	12.908611
L29	Sicily: Palermo, Parco della Favorita	38.160000	13.343889
L30	Sicily: Sambuca di Sicilia (AG) Lago Arancio	37.646111	13.104722
L31	Corsica: Desert des Agriates baie de l'Aacciolu	42.691568	9.068386
L32	Corsica: Desert des Agriates loc Magtazzini Rues	42.725420	9.203274
L33	Sardinia: Villasimius	39.123056	9.511111
L34	Sardinia: Piscina Rei env.	39.274722	9.585556

L35	Sardinia: Torre Bari env.	39.835556	9.680278
L36	Sardinia: Lido delle Rose	39.969167	9.685556
L37	Sardinia: Caletta di Osalla	40.330000	9.675278
L38	Sardinia: Cala Ginepro (Cala Liberotto)	40.441389	9.795000
L39	Sardinia: Torre Chia	38.899167	8.886111
L40	Corsica: 2 km S Porto-Vecchio	41.573719	9.281359
L41	Sardinia: Pula	39.005779	9.021993
L42	Corsica: Belgodere	42.587782	9.018395
L43	Corsica: Tour Saleccia	42.639768	8.975615
L44	Sicily: Manfria	37.103863	14.116609
L45	Corsica: Arggiavara	41.837339	9.262047
L46	Sardinia: Marcalagonis (CA) Vill. Dei Giggli	39.278211	9.332063
L47	Sardinia: Torre delle Stelle	39.148966	9.397931
L48	Tunisia: Gov. Beja Cap Serrat, 50 km E. Tabarka	37.218847	9.221964
L49	Corsica: Camping Villata (10 km N Porto Vecchio)	41.658333	9.372833
L50	Corsica: Casteddu d'Araghju (vistor parking place of the castle); 5 km NW Porto Vecchio	41.641667	9.266167
L51	Corsica: Estuary env. of river Liamone, 4 km S Sagone	42.085000	8.735962
L52	Corsica: L'Ostriconi	42.653333	9.068667
L53	Corsica: Camping "U Sole Marinu", Farinole env., river Albine	42.713333	9.333500
L54	Elba: Pareti	42.728226	10.379419
L55	Elba: Portoferraio, loc. Norsi	42.766328	10.342791
L56	Sardinia: Sinnai (CA); S. Gregorio	39.300266	9.364815
L57	Italy: Lazio, (RM), Sughereta di Pomezia	41.659255	12.511969
L58	Sardinia: Matzaccara	39.132291	8.449044
L59	Sardinia: Gutturu Mannu (Assemini)	39.181125	8.904419
L60	Sardinia: Palau, Porto Pollo	41.185695	9.327950
L61	Sardinia: S. Margherita di Pula	38.925129	8.910341
L62	Italy: Basilicata: Policoro	40.190807	16.715527

Table A2. Best fitting substitution models that were inferred with IQ-TREE (used with ABGD) and simplified models for use with STACEY (see main text). Furthermore, the minimum and maximum intraspecific divergence used in ABGD (P_{min}, P_{max}) is given per locus.

locus	for ABGD	P_{min}	P_{max}	for STACEY
cox1	TIM+I+G4	0.0018	0.1777	GTR+I+G4
rrnL	HKY+I+G4	0.0009	0.0895	HKY+I+G4
28S	JC+I	0.0001	0.0110	JC+I
ArgK	K2P+G4	0.0003	0.0271	K2P+G4

Table A3. Summary of principal component analysis on the covariance matrix of linear measurements.

component	variance	cum. variance
PC1	46.94%	46.94%
PC2	15.65%	62.59%
PC3	10.32%	72.90%
PC4	9.07%	81.98%
PC5	7.43%	89.41%
PC6	6.30%	95.71%
PC7	4.29%	100.00%
PC8	0.00%	100.00%

Table A4. Details of recursive partitions inferred with ABGD. Partition ranges, the respective resulting number of entities, and the respective prior intraspecific divergence are given.

locus	partitions	no. entities	P-range
cox1	p1– p7	69	0.0018 – 0.0023
	p8 – p16	48	0.0025 – 0.0036
	p17 – p35	47	0.0037 – 0.0086
	p36 – p39	33	0.0091 – 0.0104
	p40 – p41	32	0.0109 – 0.0114
	p42	30	0.0120
	p43 – p44	29	0.0125 – 0.0131
	p45	28	0.0138
	p46	26	0.0144
	p47	24	0.0151
	p48 – p55	23	0.0158 – 0.0219
	p56	1	0.0230
rrnL	p1– p15	32	0.0009 – 0.0017
	p16 – p23	30	0.0018 – 0.0025
	p24 – p29	27	0.0026 – 0.0033
	p30	23	0.0035
	p31 – p35	12	0.0036 – 0.0044
	p36 – p42	6	0.0046 – 0.0060
	p43	1	0.0063
28S	p1 – p100	15	<0.0000
ArgK	p1 – p100	39	<0.0000

Table A5. Effective sampling sizes of bPTP log-likelihoods that were inferred with the R package coda.

	cox1	*rrnL*	28S	ArgK
ESS	2063.732	1456.121	4501.00	3801.69

A.3. Electronic Supplements

Electronic Supplement A1 Tabular listing of specimen assignment to the putative species that were delimited in this study, sampling locality (refers to table A2), and GenBank accession numbers.

Please find this table on the attached CD at
`./electronic_supplement/Chapter_II/Electronic_Supplement_A1.pdf`

Electronic Supplement A2 Tabular listing of linear measurements of body parts in millimeter. AL = length of antennal club, Ewhc = elytron width at humeral callus, EL = elytron length, Pwmax = maximum pronotum width, PWb = width of pronotum at base, PL = medial pronotum length, MTL = length of metatibia, MTW = maximum width of metatibia, 'na' = not available.

Please find this table on the attached CD at
`./electronic_supplement/Chapter_II/Electronic_Supplement_A2.pdf`

Electronic Supplement A3 Tabular listing of single and median posterior probabilities of all iBPP analyses on guide tree part 1. For better readability, the tables are arranged according to the 3×3 prior combinations in Figs. II.4 and A9 – A11 (see graphical legend of Fig. II.4).

Please find this table on the attached CD at
`./electronic_supplement/Chapter_II/Electronic_Supplement_A3.txt`

Electronic Supplement A4 Tabular listing of single and median posterior probabilities of all iBPP analyses on the unmodified guide tree part 2. For better readability, the tables are arranged according to the 3×3 prior combinations in Figures II.4 and A9 – A11 (see graphical legend of Fig. II.4).

Please find this table on the attached CD at
`./electronic_supplement/Chapter_II/Electronic_Supplement_A4.txt`

Electronic Supplement A5 Tabular listing of single and median posterior probabilities of all iBPP analyses on the geography-informed guide tree part 2. For better readability, the tables are arranged according to the 3×3 prior combinations in Figures II.4 and A9 – A11 (see graphical legend of Fig. II.4).

Please find this table on the attached CD at
`./electronic_supplement/Chapter_II/Electronic_Supplement_A5.txt`

B. Appendix to chapter III

B.1. Supplementary Figures

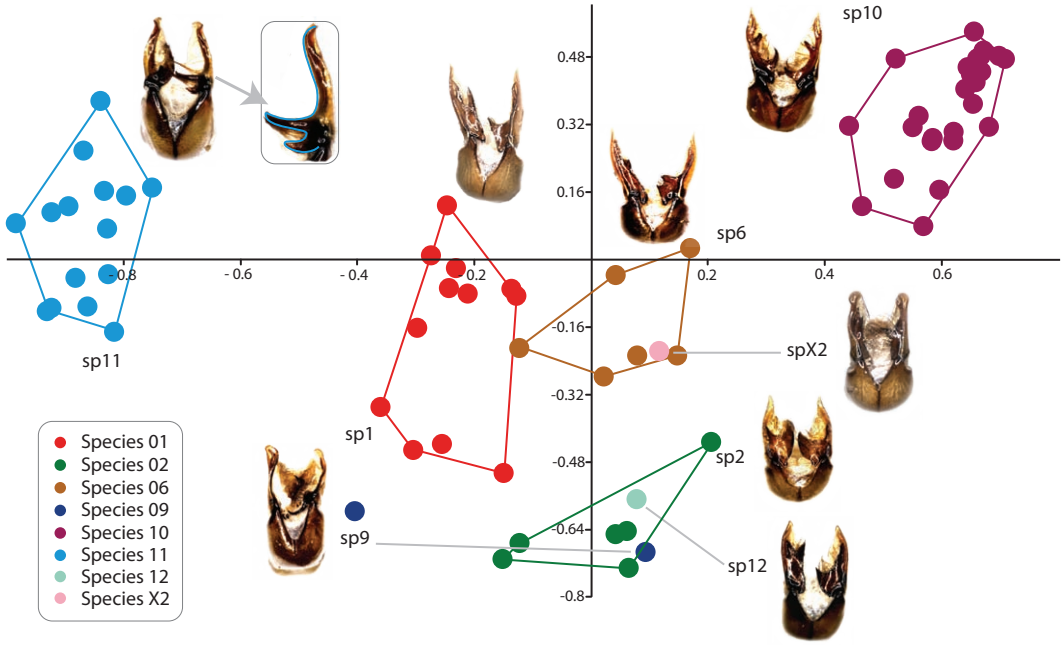

Figure B1. Plots of axes 1 and 2 from canonical variate analysis of the left paramere of male *Pleophylla* specimens. Specimens are colored according to morphospecies assignments (Table B1). The digitized outline of the left paramere is indicated in blue line for *sp11* (see arrow).

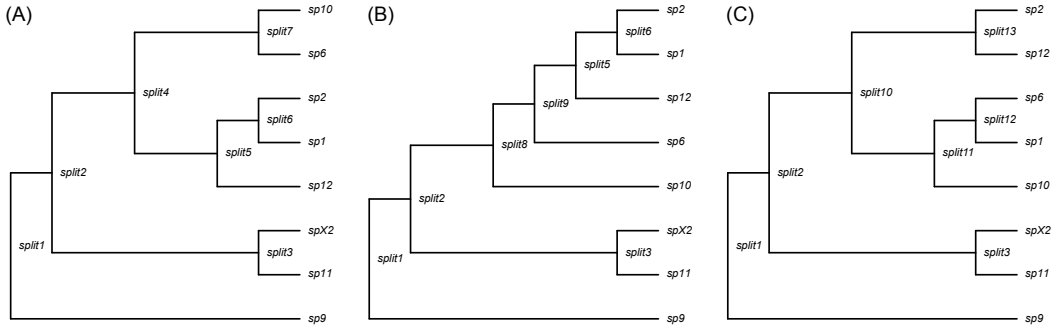

Figure B2. Guide tree topologies for BPP analyses and speciation split labeling. (A) Tree A: *BEAST topology, *sp12* and *spX2* were inserted according to the RAxML topology, (B) RAxML and MrBayes topology, and (C) topology derived from morphological similarity. Identical splits are labeled with the same number.

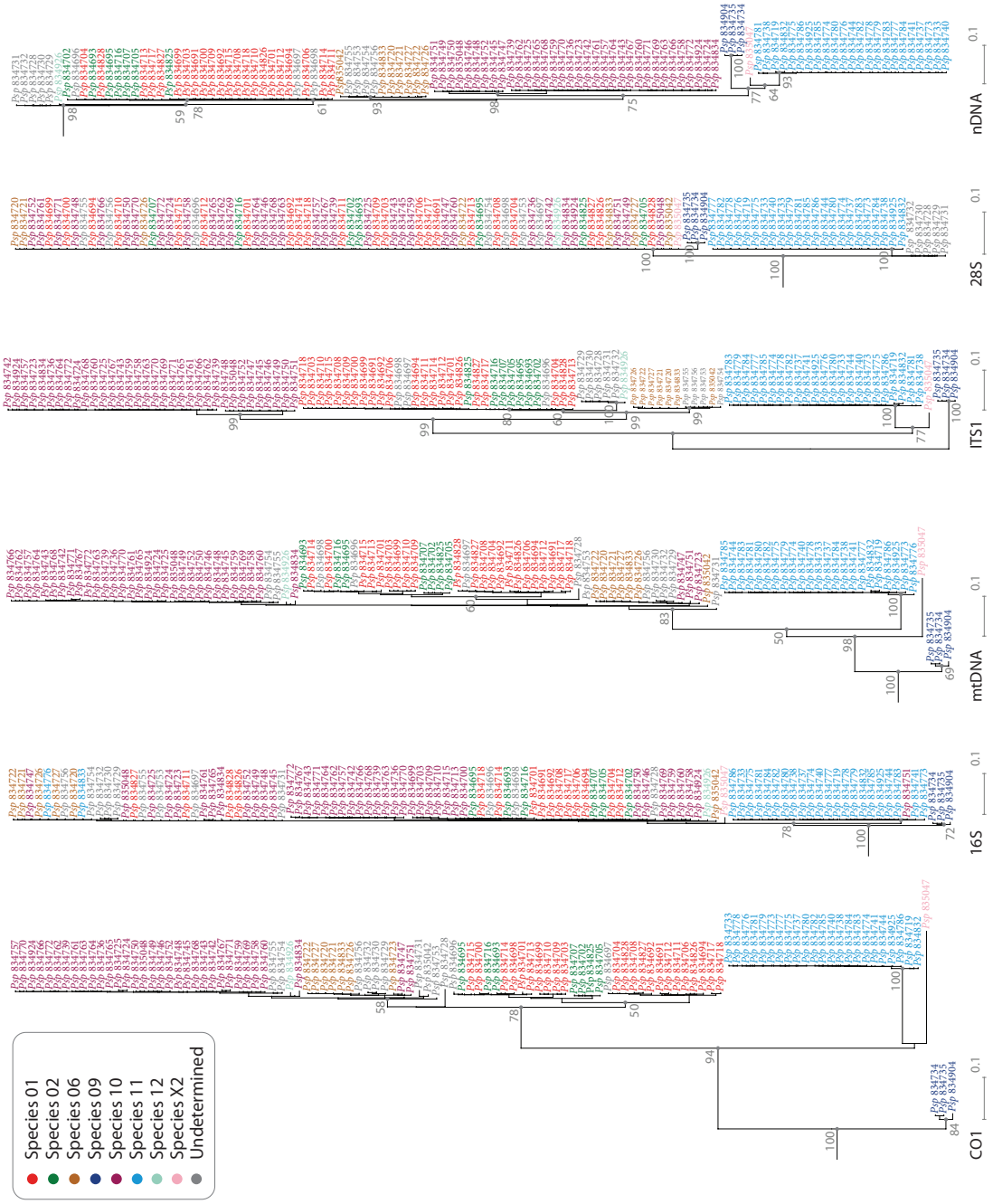

Figure B3. Maximum likelihood (RAxML) trees of *Pleophylla* for independent and combined molecular datasets (mt: mitochondrial, n: nuclear). Specimens are colored according to morphospecies assignments (Table B1). Branch length corresponds to inferred numbers of substitutions per site. Bootstrap support values are indicated for interspecific divergences. Values less than 50% are not shown.

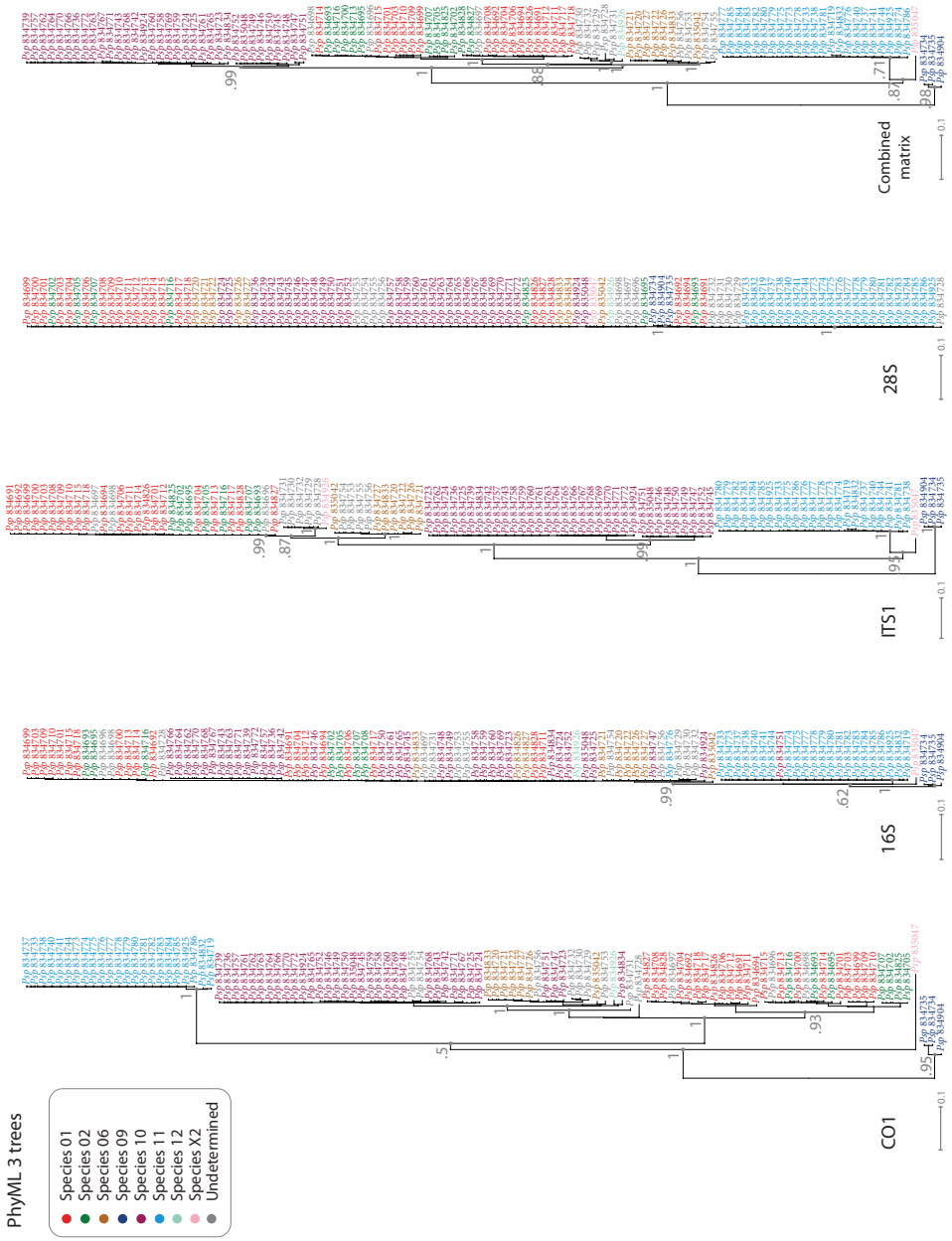

Figure B4. Maximum likelihood (PhyML) trees of *Pleophylla* (Psp) for independent and combined molecular datasets. Specimens are colored according to morphospecies assignments (Table B1). Branch length corresponds to inferred numbers of substitutions per site. Bayes support values are indicated for interspecific divergences. Values less than 0.5 are not shown.

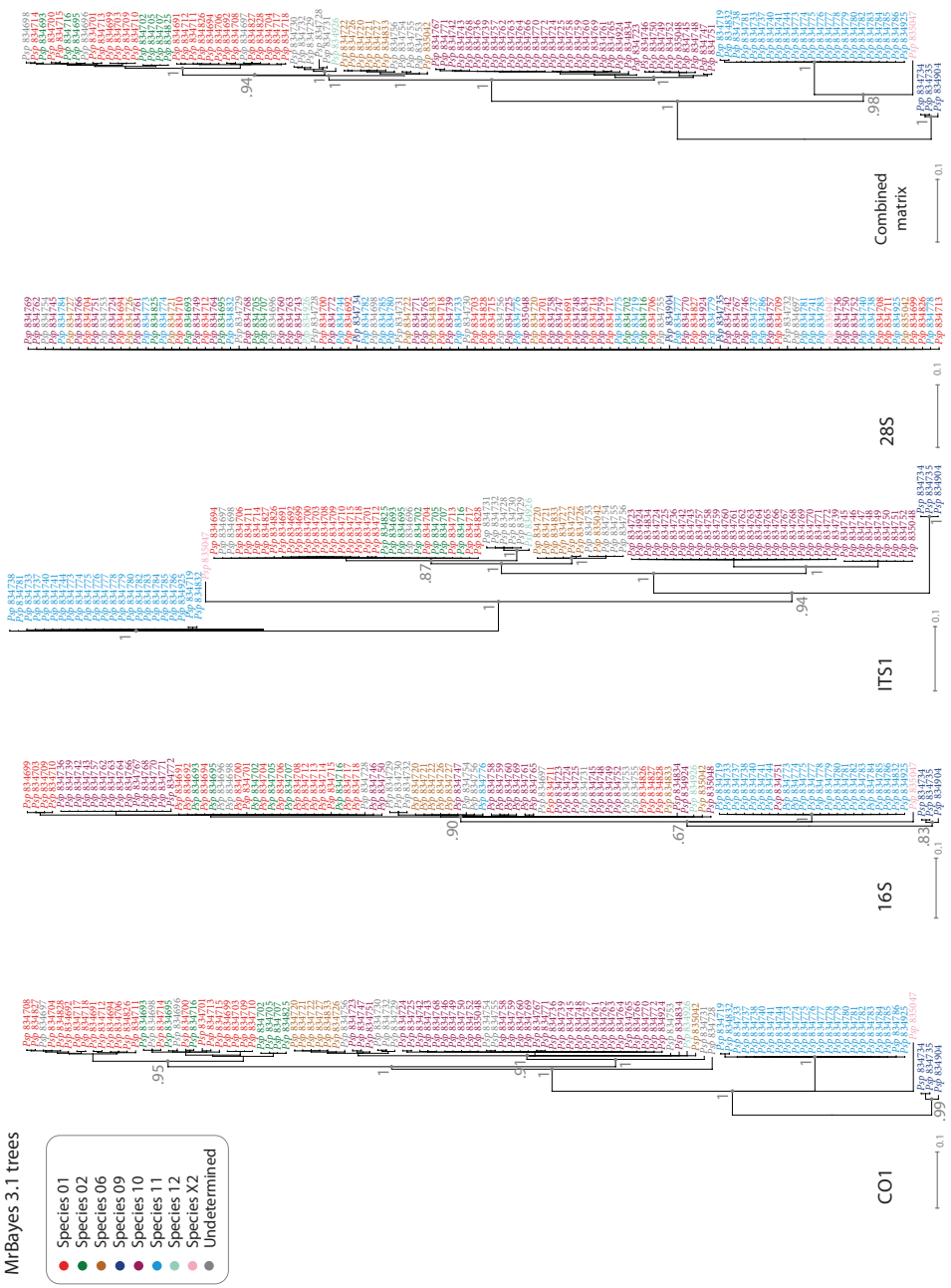

Figure B5. Bayesian (MrBayes) trees of *Pleophylla* (Psp) for independent and combined molecular datasets. Specimens are colored according to morphospecies assignments (Table B1). Branch length corresponds to inferred numbers of substitutions per site. Bayesian posterior probabilities are indicated for interspecific divergences. Values less than 0.5 are not shown.

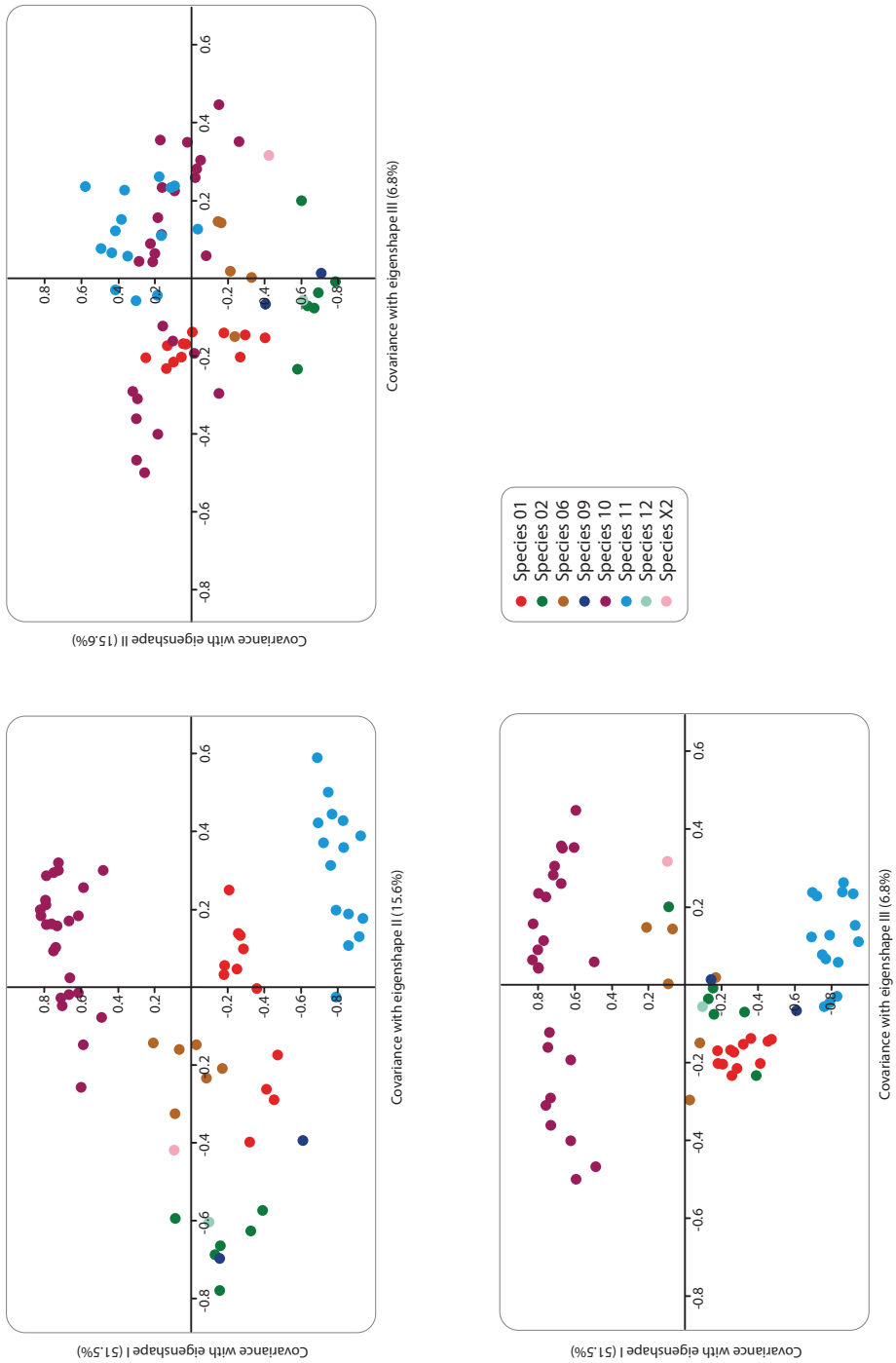

Figure B6. Pairwise plots of Eigenshape axes 1–3 from the Eigenshape analysis of partial paramere outlines.

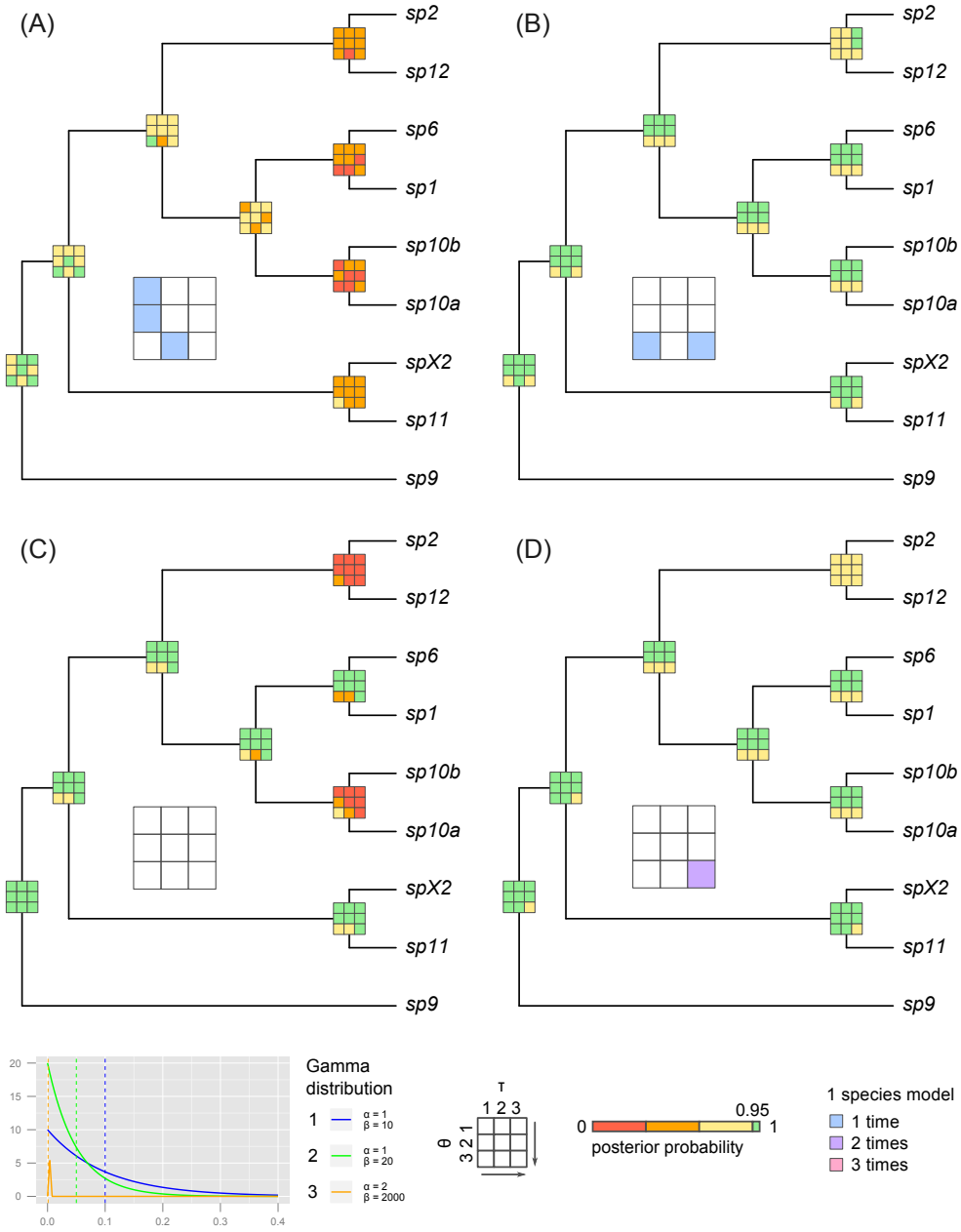

Figure B7. Mean posterior probabilities of Bayesian species delimitations from 10 repeated runs with commonly used priors using the additional guide tree. Means inferred under 9 different θ and τ_0 prior combinations are color-coded in 3×3 boxes on each putative speciation split of the guide trees. The arrows in the legend point to the direction of more conservative prior choices. Results are shown for analyses using (A) no data (prior only), (B) molecular data, (C) morphometric trait data, and (D) both data sources. The colors of the large 3×3 inset boxes indicate the number of repeat-analyses that were stuck in the one species model. Gamma distribution densities of θ and τ_0 priors 1–3 are depicted in the bottom left corner. Dashed lines indicate the respective distribution means.

B.2. Supplementary Tables

Table B1. Accession numbers for specimens included in the morphometric and phylogenetic analyses, along with voucher numbers and geographical origin. All specimens were included in the phylogenetic analyses. Specimens that were included in the morphometric analysis are indicated by an asterisk (*).

Morphospecies	Voucher	Loc.	*cox1*	*rrnl*	ITS1	28S
Omaloplia nigromarginata	747065	-	EF487770	EF487791	EU084255	NA
Omaloplia ruricola	747063	-	EF487771	EF487790	EU084256	NA
Pleophylla sp1	834691	L1	KC904098	KC964429	KC964210	KC964320
Pleophylla sp1	834692*	L1	KC904099	KC964430	KC964211	KC964321
Pleophylla sp1	834694*	L1	KC904101	KC964432	KC964213	KC964323
Pleophylla sp1	834699*	L1	KC904106	KC964437	KC964218	KC964328
Pleophylla sp1	834700	L1	KC904107	KC964438	KC964219	KC964329
Pleophylla sp1	834701*	L1	KC904108	KC964439	KC964220	KC964330
Pleophylla sp1	834703	L1	KC904110	KC964441	KC964222	KC964332
Pleophylla sp1	834704	L1	KC904111	KC964442	KC964223	KC964333
Pleophylla sp1	834706	L1	KC904113	KC964444	KC964225	KC964335
Pleophylla sp1	834708*	L1	KC904115	KC964446	KC964227	KC964337
Pleophylla sp1	834709*	L1	KC904116	KC964447	KC964228	KC964338
Pleophylla sp1	834710	L1	KC904117	KC964448	KC964229	KC964339
Pleophylla sp1	834711	L1	KC904118	KC964449	KC964230	KC964340
Pleophylla sp1	834712*	L1	KC904119	KC964450	KC964231	KC964341
Pleophylla sp1	834713	L1	KC904120	KC964451	KC964232	KC964342
Pleophylla sp1	834714	L1	KC904121	KC964452	KC964233	KC964343
Pleophylla sp1	834715*	L1	KC904122	KC964453	KC964234	KC964344
Pleophylla sp1	834717*	L1	KC904124	KC964455	KC964236	KC964346
Pleophylla sp1	834718	L1	KC904125	KC964456	KC964237	KC964347
Pleophylla sp1	834826*	L1	KC904195	KC964523	KC964307	KC964416
Pleophylla sp1	834827	L1	KC904196	KC964524	KC964308	KC964417
Pleophylla sp1	834828*	L1	KC904197	KC964525	KC964309	KC964418
Pleophylla sp10	834723*	L2	KC904130	KC964461	KC964242	NA
Pleophylla sp10	834724*	L2	KC904131	KC964462	KC964243	KC964352
Pleophylla sp10	834725*	L2	KC904132	KC964463	KC964244	KC964353
Pleophylla sp10	834736*	L3	KC904143	KC964474	KC964255	KC964364
Pleophylla sp10	834739*	L3	KC904146	KC964477	KC964258	KC964367
Pleophylla sp10	834742	L3	KC904149	KC964480	KC964261	KC964370
Pleophylla sp10	834743*	L3	KC904150	KC964481	KC964262	KC964371
Pleophylla sp10	834745*	L3	KC904152	KC964483	KC964264	KC964373
Pleophylla sp10	834746	L3	KC904153	KC964484	KC964265	KC964374
Pleophylla sp10	834747	L3	KC904154	KC964485	KC964266	KC964375
Pleophylla sp10	834748*	L3	KC904155	KC964486	KC964267	KC964376

Pleophylla sp10	834749	L3	KC904156	KC964487	KC964268	KC964377
Pleophylla sp10	834750*	L6	KC904157	KC964488	KC964269	KC964378
Pleophylla sp10	834751	L6	KC904158	NA	KC964270	KC964379
Pleophylla sp10	834752	L6	KC904159	KC964489	KC964271	KC964380
Pleophylla sp10	834757*	L3	KC904164	KC964494	KC964276	KC964385
Pleophylla sp10	834758*	L3	KC904165	KC964495	KC964277	KC964386
Pleophylla sp10	834759*	L3	KC904166	KC964496	KC964278	KC964387
Pleophylla sp10	834760*	L3	KC904167	KC964497	KC964279	KC964388
Pleophylla sp10	834761	L3	KC904168	KC964498	KC964280	KC964389
Pleophylla sp10	834762	L3	KC904169	KC964499	KC964281	KC964390
Pleophylla sp10	834763*	L3	KC904170	KC964500	KC964282	KC964391
Pleophylla sp10	834764*	L3	KC904171	KC964501	KC964283	KC964392
Pleophylla sp10	834765*	L3	KC904172	KC964502	KC964284	KC964393
Pleophylla sp10	834766*	L3	KC904173	KC964503	KC964285	KC964394
Pleophylla sp10	834767*	L3	KC904174	KC964504	KC964286	KC964395
Pleophylla sp10	834768*	L3	KC904175	KC964505	KC964287	KC964396
Pleophylla sp10	834769	L3	KC904176	KC964506	KC964288	KC964397
Pleophylla sp10	834770*	L3	KC904177	KC964507	KC964289	KC964398
Pleophylla sp10	834771*	L3	KC904178	KC964508	KC964290	KC964399
Pleophylla sp10	834772*	L3	KC904179	KC964509	KC964291	KC964400
Pleophylla sp10	834834*	L2	KC904200	KC964528	KC964312	KC964421
Pleophylla sp10	834924*	L3	KC904202	KC964530	KC964314	KC964423
Pleophylla sp10	835048*	L6	KC904207	KC964535	KC964319	KC964428
Pleophylla sp11	834719*	L2	KC904126	KC964457	KC964238	KC964348
Pleophylla sp11	834733*	L4	KC904140	KC964471	KC964252	KC964361
Pleophylla sp11	834737*	L3	KC904144	KC964475	KC964256	KC964365
Pleophylla sp11	834738	L3	KC904145	KC964476	KC964257	KC964366
Pleophylla sp11	834740*	L3	KC904147	KC964478	KC964259	KC964368
Pleophylla sp11	834741*	L3	KC904148	KC964479	KC964260	KC964369
Pleophylla sp11	834744*	L3	KC904151	KC964482	KC964263	KC964372
Pleophylla sp11	834773*	L3	KC904180	KC964510	KC964292	KC964401
Pleophylla sp11	834774*	L3	KC904181	KC964511	KC964293	KC964402
Pleophylla sp11	834775*	L3	KC904182	KC964512	KC964294	KC964403
Pleophylla sp11	834776*	L3	KC904183	NA	KC964295	KC964404
Pleophylla sp11	834777*	L3	KC904184	KC964513	KC964296	KC964405
Pleophylla sp11	834778*	L3	KC904185	KC964514	KC964297	KC964406
Pleophylla sp11	834779	L3	KC904186	KC964515	KC964298	KC964407
Pleophylla sp11	834780	L3	KC904187	KC964516	KC964299	KC964408
Pleophylla sp11	834781	L3	KC904188	KC964517	KC964300	KC964409
Pleophylla sp11	834782	L3	KC904189	KC964518	KC964301	KC964410
Pleophylla sp11	834783	L3	KC904190	KC964519	KC964302	KC964411
Pleophylla sp11	834784	L3	KC904191	KC964520	KC964303	KC964412
Pleophylla sp11	834785*	L3	KC904192	KC964521	KC964304	KC964413
Pleophylla sp11	834786	L3	KC904193	KC964522	KC964305	KC964414
Pleophylla sp11	834832*	L2	KC904198	KC964526	KC964310	KC964419
Pleophylla sp11	834925*	L3	KC904203	KC964531	KC964315	KC964424

Pleophylla sp12	834926*	L3	KC904204	KC964532	KC964316	KC964425
Pleophylla sp2	834693*	L1	KC904100	KC964431	KC964212	KC964322
Pleophylla sp2	834695	L1	KC904102	KC964433	KC964214	KC964324
Pleophylla sp2	834702*	L1	KC904109	KC964440	KC964221	KC964331
Pleophylla sp2	834705*	L1	KC904112	KC964443	KC964224	KC964334
Pleophylla sp2	834707*	L1	KC904114	KC964445	KC964226	KC964336
Pleophylla sp2	834716*	L1	KC904123	KC964454	KC964235	KC964345
Pleophylla sp2	834825*	L1	KC904194	NA	KC964306	KC964415
Pleophylla sp6	834720*	L2	KC904127	KC964458	KC964239	KC964349
Pleophylla sp6	834721*	L2	KC904128	KC964459	KC964240	KC964350
Pleophylla sp6	834722*	L2	KC904129	KC964460	KC964241	KC964351
Pleophylla sp6	834726*	L2	KC904133	KC964464	KC964245	KC964354
Pleophylla sp6	834727	L2	KC904134	KC964465	KC964246	KC964355
Pleophylla sp6	834833*	L2	KC904199	KC964527	KC964311	KC964420
Pleophylla sp6	835042*	L8	KC904205	KC964533	KC964317	KC964426
Pleophylla sp9	834734	L4	KC904141	KC964472	KC964253	KC964362
Pleophylla sp9	834735*	L4	KC904142	KC964473	KC964254	KC964363
Pleophylla sp9	834904*	L4	KC904201	KC964529	KC964313	KC964422
Pleophylla spX2	835047*	L6	KC904206	KC964534	KC964318	KC964427
Pleophylla sp(indet.)	834696	L1	KC904103	KC964434	KC964215	KC964325
Pleophylla sp(indet.)	834697	L1	KC904104	KC964435	KC964216	KC964326
Pleophylla sp(indet.)	834698	L1	KC904105	KC964436	KC964217	KC964327
Pleophylla sp(indet.)	834728	L5	KC904135	KC964466	KC964247	KC964356
Pleophylla sp(indet.)	834729	L5	KC904136	KC964467	KC964248	KC964357
Pleophylla sp(indet.)	834730	L5	KC904137	KC964468	KC964249	KC964358
Pleophylla sp(indet.)	834731	L5	KC904138	KC964469	KC964250	KC964359
Pleophylla sp(indet.)	834732	L5	KC904139	KC964470	KC964251	KC964360
Pleophylla sp(indet.)	834753	L7	KC904160	KC964490	KC964272	KC964381
Pleophylla sp(indet.)	834754	L7	KC904161	KC964491	KC964273	KC964382
Pleophylla sp(indet.)	834755	L7	KC904162	KC964492	KC964274	KC964383
Pleophylla sp(indet.)	834756	L7	KC904163	KC964493	KC964275	KC964384

Table B2. Collection localities with their geographical coordinates.

ID	Name	Lat	Long
L1	South Africa: Cheerio Farm, Haenertsburg, ca. 20 km W of Tzaneen, 1492 m	-23.895	29.953
L2	South Africa: Drakensberge, Highmoor, 1582 m, at light	-29.331	29.691
L3	South Africa: Kwazulu-Natal: Drakensberge Lodge 15 km NW of Himeville, 1622 m, at light	-29.631	29.419
L4	South Africa: Limpopo: Legalameetse Nature Reserve, guest house camp (Murchson range), 814 m, at light	-24.200	30.337
L5	South Africa: Kwazulu-Natal: Nature Farm (Louwsburg), 1100 m, at light	-27.567	31.300
L6	South Africa: Free State: Wakkerstrom (Wetland Lodge env.), 1802 m	-27.339	30.153
L7	South Africa: Eastern Cape: Morgans Bay env. (Yellowwood Tree Park), 4 m	-32.696	28.334
L8	South Africa: Fort Fordyce, 1000 m	-32.410	26.280

Table B3. Optimal partition schemes and substitution models selected for each phylogenetic program using PartitionFinder (Lanfear et al., 2012) under the Bayesian Information Criterion (BIC). Note that when invariant (I) and gamma distributed (Γ) sites both featured in the optimal model, we implemented only the Γ parameter (see materials and method for details).

Dataset	Software	Partition scheme (model)
cox1	PhyML	Unpartitioned (HKY+I+Γ)
	RAxML	Codon 1+2 (GTR+Γ), Codon 3 (GTR+Γ)
	MrBayes	Codon 1 (GTR+I), Codon 2 (F80+I), Codon 3 (GTR+Γ)
MtDNA	PhyML	Unpartitioned (HKY+Γ)
	RAxML	Codon 1+2 (GTR+Γ), Codon 3 (GTR+Γ), 16S (GTR+Γ)
	MrBayes	Codon 1 (GTR+I), Codon 2 (F80+I), Codon 3 (GTR+Γ), 16S (HKY+Γ)
NucDNA	PhyML	Unpartitioned (K80+Γ)
	RAxML	ITS1 (GTR+I+Γ), 28S (GTR+Γ)
	MrBayes	ITS1 (K80+Γ), 28S (K80+I)
Mt+Nuc	PhyML	Unpartitioned (HKY+I+Γ)
	RAxML	Codon 1 (GTR+Γ), Codon 2 (GTR+Γ), Codon 3 (GTR+Γ), 16S (GTR+Γ), ITS1 (GTR+I+Γ), 28S (GTR+Γ)
	MrBayes	Codon 1 (GTR+Γ), Codon 2 (F81), Codon 3 (GTR+Γ), 16S (HKY+Γ), ITS1 (K80+Γ), 28S (K80+I), 28S (K80+I)

Table B4. Maximum likelihood estimates of tree length obtained for each marker using maximum likelihood phylogenetic analysis (PhyML and RAxML) and Bayesian phylogenetic analysis under different branch length priors (MrBayes). Posterior means and intervals are presented for the Bayesian estimates.

Analysis	MLE[a]	Mean (95% HPDs)[b]	MLE	Mean (95% HPDs)	MLE	Mean (95% HPDs)	MLE	Mean (95% HPDs)
	cox1		rrnL[u]		mtDNA			
MrBayes Exp(10)[c]	14.98	17.12 (14.25, 19.99)	23.2	21.85 (19.13, 24.84)	23.41	21.77 (19.12, 24.67)		
MrBayes Exp(20)	6.47	7.34 (6.18, 8.58)	10	10.9 (9.39, 12.3)	9.91	10.93 (9.59, 12.41)		
MrBayes Exp(100)	1.54	1.43 (1.25, 1.61)	1.45	1.82 (1.49, 2.15)	1.17	1.25 (1.09, 1.41)		
MrBayes Exp(200)	1.051	1 (0.9, 1.1)	0.65	0.77 (0.65, 0.89)	0.9	0.88 (0.79, 0.96)		
PhyML	1.43[u]	-	0.28	-	1.14[u]	-		
RAxML	2.12	-	0.27	-	1.38	-		
	ITS1[u]		28S[u]		nucDNA		Combined matrix	
MrBayes Exp(10)	19.31	21.33 (18.48, 24.43)	6.7	10.22 (8.2, 12.36)	19.9	21.8 (18.88, 24.84)	NA[d]	NA
MrBayes Exp(20)	8.99	10.47 (9.13, 11.95)	3.95	5.48 (4.53, 6.4)	9.61	10.76 (9.35, 10.24)	10.7	10.7 (9.29, 12.1)
MrBayes Exp(100)	0.7	0.84 (0.69, 0.98)	1.19	1.56 (1.34, 1.78)	0.93	1.59 (1.13, 1.13)	1.14	1.14 (0.96, 1.33)
MrBayes Exp(200)	0.55	0.59 (0.52, 0.66)	0.71	0.86 (0.74, 0.98)	0.44	0.75 (0.63, 0.79)	0.78	0.78 (0.7, 0.86)
PhyML	0.4	-	0.01	-	0.23[u]	-	0.69[u]	-
RAxML	0.4	-	0.01	-	0.27	-	1.05	-

[a] Maximum likelihood estimate of tree length (sum of branch lengths)
[b] Posterior means and 95% highest posterior density intervals for tree length
[c] Exponential prior distribution Exp(λ), where λ is the rate parameter
[d] NA = analysis failed to converge
[u] Unpartitioned analysis

Table B5. The estimated mean rates (± SD) of *cox1* partitions and ITS1 estimated using *BEAST, relative to the fixed rate of *cox1*.

cox1 ucld.mean[a]	*cox1* mean rate[b]	*cox1*: 1st codon[c]	*cox1*: 2nd codon	*cox1*: 3rd codon	ITS1 ucld.mean	ITS1 mean rate
0.01 (2%)[d]	0.0094 ± 0.001	0.0011 ± 0.032	0.0004 ± 0.019	0.0267 ± 0.039	0.0017 ± 0.001	0.0024 ± 0.001
0.0125 (2.5%)	0.0118 ± 0.002	0.0014 ± 0.032	0.0005 ± 0.019	0.0336 ± 0.039	0.002 ± 0.001	0.0029 ± 0.001
0.015 (3%)	0.014 ± 0.002	0.0016 ± 0.033	0.0006 ± 0.019	0.04 ± 0.04	0.0029 ± 0.011	0.0035 ± 0.004
0.0175 (3.5%)	0.0165 ± 0.002	0.0019 ± 0.032	0.0007 ± 0.019	0.0477 ± 0.039	0.0029 ± 0.001	0.0042 ± 0.002
0.02 (4%)	0.019 ± 0.002	0.0022 ± 0.033	0.0008 ± 0.019	0.0541 ± 0.039	0.0033 ± 0.001	0.0047 ± 0.002

[a]Mean of the uncorrelated log-normal relaxed clock, equivalent to the mean branch rate [b]Mean rate of evolution across the whole tree [c]The rate of evolution for individual codons is obtained by multiplying the mean rate of cox1 by the relative codon rate [d]Percentage divergence My^{-1}

Table B6. Interspecific divergence times estimated using *BEAST under variable cox1 substitution rates.

Split	2% My^{-1} Mean (95% HPDs)	2.5% My^{-1} Mean (95% HPDs)	3% My^{-1} Mean (95% HPDs)	3.5% My^{-1} Mean (95% HPDs)	4% My^{-1} Mean (95% HPDs)
01 + 02	0.39 (0.01, 0.17)	0.07 (0.01, 0.14)	0.06 (0.01, 0.12)	0.05 (0, 0.1)	0.04 (0, 0.09)
06 + 10	0.09 (0.17, 0.65)	0.31 (0.14, 0.51)	0.26 (0.11, 0.43)	0.23 (0.1, 0.37)	0.19 (0.09, 0.32)
01 + 02, 06 + 10	3.17 (1.51, 4.82)	2.52 (1.22, 3.82)	2.11 (1.02, 3.24)	1.81 (0.85, 2.73)	1.56 (0.77, 2.34)
01 + 02 + 06 + 10, 11	17.53 (7.17, 34.73)	13.48 (5.33, 26.12)	11.73 (1.75, 21.04)	9.77 (4.05, 18.87)	8.32 (3.64, 15.09)
01 + 02 + 06 + 10 + 11, 9 (root)	20.23 (2.64, 35.97)	15.56 (1.96, 26.77)	13.42 (1.58, 24.43)	11.26 (1.56, 19.84)	9.55 (3.69, 17.88)

Table B7. Results of the JML analysis based on *cox1* codon partitions [P1/P2/P3] and ITS1: minimum distances (below the diagonal) and *P*-values (above the diagonal, shaded in grey). Values are presented for trees obtained using the *cox1* rate 2% My^{-1}, but note the results are equivalent for trees generated under different substitution rates.

		Species 01	*Species 02*	*Species 06*	*Species 09*	*Species 10*	*Species 11*
cox1							
	Species 01	-	1 / 1 / 0.66	0.43 / 0.49 / 0.16	0.1 / 0.26/ 0.03	0.17 / 0.49 / 0.21	0.3 / 0.14 / 0.03
	Species 02	0 / 0 / 0	-	0.42 / 0.48 / 0.14	0.17 / 0.43 / 0.03	0.17 / 0.48/ 0.19	0.29 / 0.3 / 0.04
	Species 06	0 / 0 / 0.11	0 / 0 / 0.11	-	0.24 / 0.43 / 0.03	0.75 / 0.89 / 0.26	0.2 / 0.3 / 0.04
	Species 09	0.02 / 0.01 / 0.29	0.02 / 0.01 / 0.29	0.03 / 0.01 / 0.29	-	0.17 / 0.46 / 0.03	0.42 / 0.42 / 0.03
	Species 10	0 / 0 / 0.11	0 / 0 / 0.11	0 / 0 / 0.01	0.02 / 0.01 / 0.29	-	0.2 / 0.29 / 0.04
	Species 11	0.03 / 0 / 0.3	0.03 / 0.01 / 0.3	0.02 / 0.01 / 0.29	0.03 / 0.01 / 0.28	0.02 / 0.01 / 0.29	-
ITS1							
	Species 01	-	0.97	0.29	0.89	0.92	0.76
	Species 02	0	-	0.26	0.89	0.92	0.75
	Species 06	0.01	0.01	-	0.88	0.99	0.48
	Species 09	0.12	0.12	0.12	-	0.9	0.99
	Species 10	0.04	0.04	0.02	0.12	-	0.78
	Species 11	0.09	0.09	0.08	0.12	0.09	-

Table B8. My caption

	Tree[a]	logLNull[b]	$\log L^{c}_{GMYC-Single}$	$\log L^{d}_{GMYC-Multi}$	AIC^{e}_{c}	Akaike weight[f]	Entities[g]	Variance (σ^2)[h]
cox1								
	MrBayes Exp(10)i + PATHd8	285.49	293.61***[j]	293.75	-575.93	0.22	15.12 (3)[k]	0.92
	MrBayes Exp(20) + PATHd8	283.39	288.93*	289.77	-566.59	0.17	16.43 (4)	0.89
	MrBayes Exp(100) + PATHd8	264.44	267.84	526.59***	-1039.35	1	3.00 (1)	0.00
	MrBayes Exp(200) + PATHd8	258.45	265.51	268.17***	-522.52	0.54	15.00 (1)	0.00
	RAxML + PATHd8	321.56	386.43	840.18***	-1663.87	0.5	4.00 (2)	0.00
	MrBayes Exp(10) + R8S	268.04	278.24	280.61***	-547.48	0.22	16.33 (2)	0.44
	MrBayes Exp(20) + R8S	260.21	269.56***	270.29	-527.84	0.15	11.92 (7)	1.54
	MrBayes Exp(100) + R8S	245.94	253.77	255.10**	-496.38	0.13	12.73 (8)	2.36
	MrBayes Exp(200) + R8S	239.96	245.02	247.89**	-481.95	0.37	12.00 (1)	0.00
	RAxML + R8S	283.23	290.25**	290.71	-569.22	0.23	16.54 (3)	0.89
	Statistical Parsimony	-	-	-	-		13.00	-
ITS1								
	MrBayes Exp(10) + PATHd8	70.02	75.70**	75.76	-137.41	0.34	10.99 (4)	4.05
	MrBayes Exp(20) + PATHd8	69.85	77.29**	77.29	-140.57	0.7	8.00 (1)	0.00
	MrBayes Exp(100) + PATHd8	67.84	74.45**	74.45	-134.91	0.59	8.00 (1)	0.00
	MrBayes Exp(200) + PATHd8	65.89	72.08**	72.08	-130.16	0.43	8.00 (1)	0.00
	RAxML + PATHd8	70.50	77.33**	77.49	-140.66	0.6	8.00 (1)	0.00
	MrBayes Exp(10) + R8S	65.95	72.02**	72.02	-130.05	0.49	8.00(1)	0.00
	MrBayes Exp(20) + R8S	65.60	71.60**	71.60	-129.21	0.48	8.00 (1)	0.00
	MrBayes Exp(100) + R8S	63.45	69.02*	69.02	-124.04	0.43	9.34 (3)	5.22
	MrBayes Exp(200) + R8S	64.70	71.16**	71.35	-128.32	0.58	8.00 (1)	0.00
	RAxML + R8S	71.36	79.76***	79.76	-145.52	0.75	8.00 (1)	0.00
	Statistical Parsimony	-	-	-	-		9	-
rrnL								
	Statistical Parsimony	-	-	-	-		1	-
28S								
	Statistical Parsimony	-	-	-	-		1	-

[a] Tree building and (+) tree linearization method, with the exception of statistical parsimony
[b] The likelihood of the null model
[c] The likelihood of the GMYC single threshold model
[d] The likelihood of the GMYC multiple threshold model
[e] The AIC_c score of the preferred model
[f] Akaike weight of the preferred model
[g] Model averaged entities within $\delta AIC_c = 2$
[h] Variance in the total number of entities
[i] Exponential prior distribution $Exp(\lambda)$, where λ is the rate parameter
[j] Preferred model, likelihood ratio test $^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05$
[k] Number of models within $\delta AIC_c = 2$

Table B9. Results of the Eigenshape analysis: eigenvalues, total variance and total cumulative variation expressed by the first fourteen Eigenshapes.

Eigenshape	Eigenvalue	Total variance (%)	Cumulative variance (%)
ES1	23.93	51.48	51.48
ES2	7.25	15.6	67.08
ES3	3.15	6.78	73.86
ES4	2.79	6.01	79.87
ES5	1.79	3.86	83.73
ES6	1.28	2.75	86.48
ES7	1.07	2.31	88.79
ES8	0.78	1.69	90.48
ES9	0.49	1.06	91.54
ES10	0.46	0.98	92.52
ES11	0.39	0.84	93.36
ES12	0.38	0.81	94.17
ES13	0.35	0.75	94.92
ES14	0.25	0.54	95.46

Table B10. Results of LDA reflecting the fit of morphometric data to *a priori* defined morphospecies showing for each species sample size (n) and percentage of correctly reassigned individuals.

		% corr.	
Species	n	ES 1-4	ES 1-14
total	68	84.85	89.39
sp1	12	91.67	91.67
sp2	6	33.33	50
sp6	5	60	80
sp9	2	0	0
sp10	26	96.15	100
sp11	15	100	100
sp12	1	0	0
spX2	1	0	0

Table B11. Results of LDA reflecting the fit of morphometric data to the groups recognized with cluster analysis showing for each species sample size (n) and percentage of correctly reassigned individuals.

	ES 1-4		ES 1-14	
Cluster	n	% corr.	n	% corr.
total	68	100	68	94.12
Cluster 1	28	100	6	100
Cluster 2	25	100	2	100
Cluster 3	15	100	6	100
Cluster 4	-	-	10	70
Cluster 5	-	-	8	100
Cluster 6	-	-	3	100
Cluster 7	-	-	5	100
Cluster 8	-	-	6	83.33
Cluster 9	-	-	3	100
Cluster 10	-	-	11	100
Cluster 11	-	-	4	100
Cluster 12	-	-	4	100

B.3. Electronic Supplement

Electronic Supplement B1 Plot of the 3D phylomorphospace using the RAxML tree topology (based on the partitioned combined molecular dataset) projected onto the paramere morphospace explained by eigenaxes 1–3.

Please find this table on the attached CD at
`./electronic_supplement/Chapter_III/Electronic_Supplement_B1`

Electronic Supplement B2 Single gene species delimitation analyses output.

Please find this table on the attached CD at
`./electronic_supplement/Chapter_III/Table_S9.xlsx`

Electronic Supplement B3 Posterior probabilities of speciation splits of all guide trees for prior only analyses and 3 data sets, 9 τ_0 and θ prior combinations, and 10 repeats of each analysis.

Please find this table on the attached CD at
`./electronic_supplement/Chapter_III/Electronic_Supplement_B3.xls`

Electronic Supplement B4 Lumping behaviour of BPP analyses with simultaneous species tree estimation. Posterior probabilities for species delimitation scenarios are shown for all initial guide trees and all τ_0 and θ prior combinations. Lumped species are indicated by parentheses. Scenarios with pp < 0.05 in all analyses are not shown.

Please find this table on the attached CD at
`./electronic_supplement/Chapter_III/Electronic_Supplement_B4.xls`

Electronic Supplement B5 Posterior probabilities for *a priori* defined morphospecies from BPP analyses with simultaneous species tree estimation for all initial guide trees. Species that were never sampled are denoted with *'na'*.

Please find this table on the attached CD at
`./electronic_supplement/Chapter_III/Electronic_Supplement_B5.xls`

C. Appendix to chapter IV

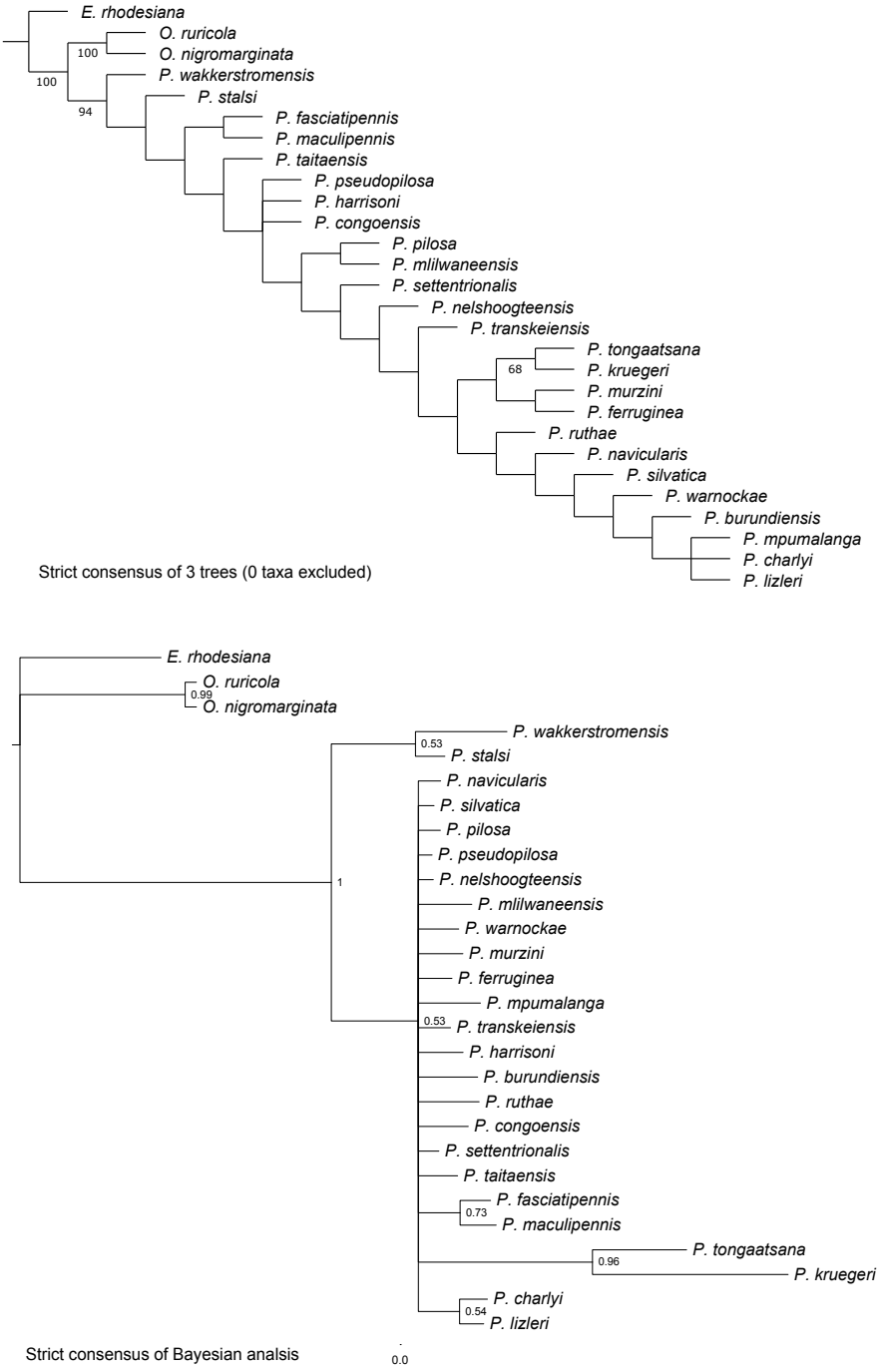

Figure C1. Strict consensus of implied weighting (k=3) resulting from 3 equally parsimonious trees (above) and strict consensus tree of the Bayesian tree inference (below).

Electronic Supplement C1 Comprehensive label data of all specimens included in the study.

Please find the document on the attached CD at
`./electronic_supplement/Chapter_IV/Electronic_Supplement_C1.pdf`

D. Appendix to chapter V

D.1. Supplementary Figures

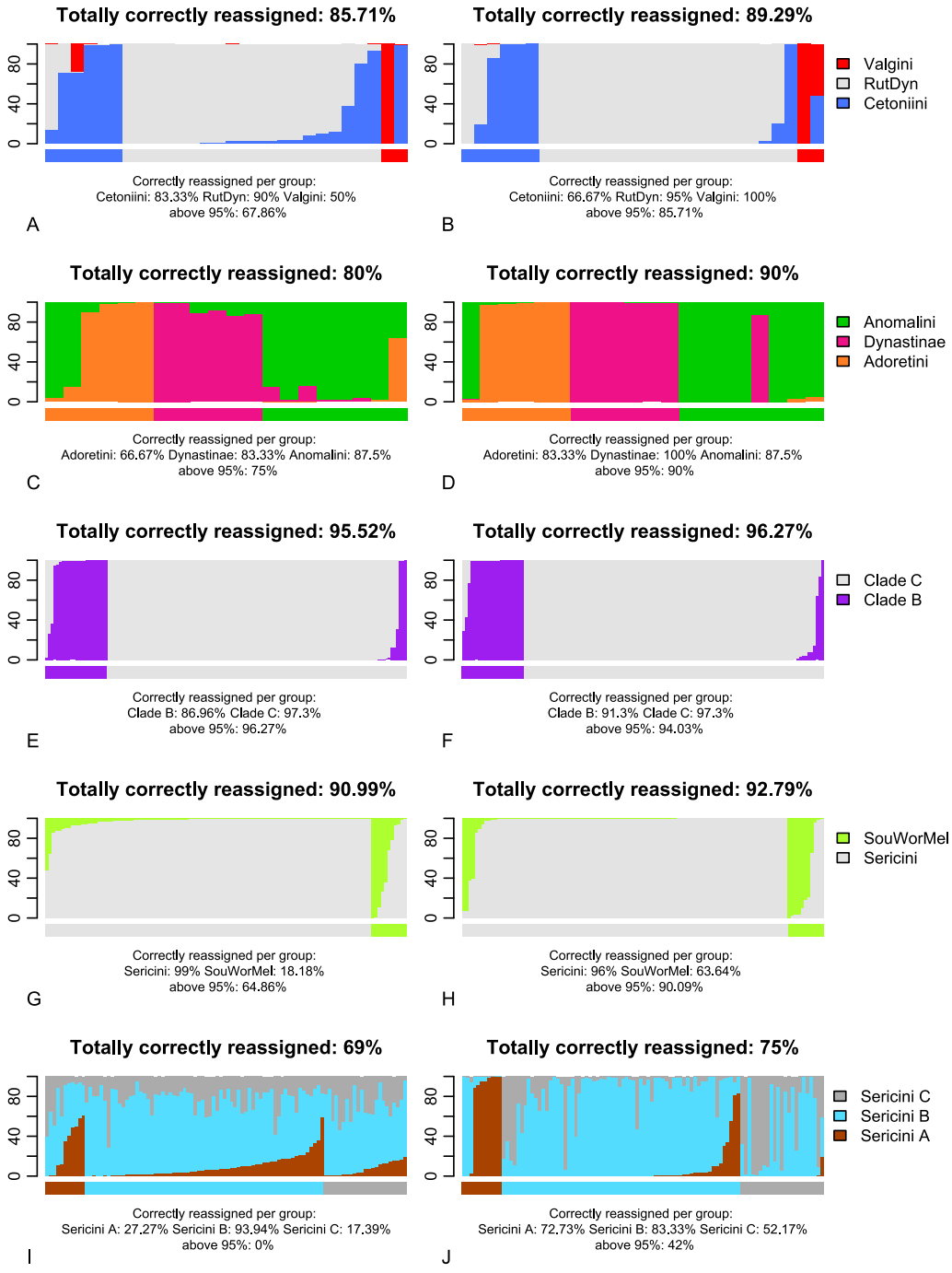

Figure D1. Discrimination of phylogenetic sister clade lineages. Barplots of the individual reassignment probabilities [%] from the discriminant analyses. Group membership priors are given under the plot by horizontal color bars. Rows refer to sister lineage subsets 1–5, columns show values for the uncorrected (left) and the size-corrected (BBPM; right) data sets.

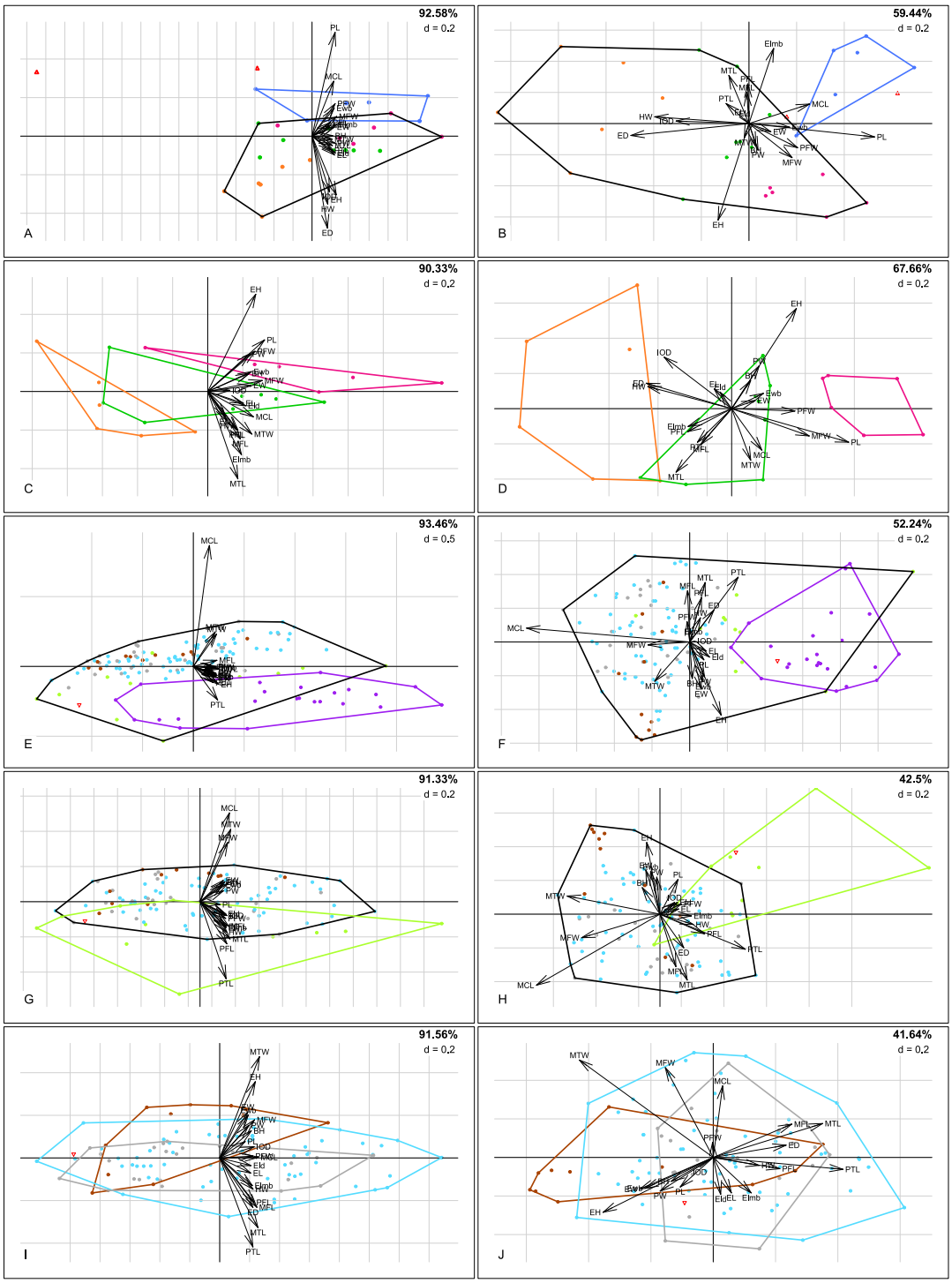

Figure D2. The drivers of morphospace divergence. Biplots of PCA scores and loadings for the uncorrected and the size-corrected data sets: (A–C) Cetoniini + Valgini and Adoretini + Anomalini + Dynastinae, (D–F) Adoretini, Anomalini, and Dynastinae, (G–I) Clade B and Clade C, (J–L) Southern World Melolonthinae and Ablaberini + Sericini, and (M–O) Sericini subgroups. The groups are color-coded in the molecular phylogeny (Fig. V.4A). The percentage of variance explained by principal component 1 and 2 is given in the top right corner. Groups with more than 2 members are surrounded by a similarly colored hull. x-axis: PC1, y-axis: PC2. d = mesh of the grid.

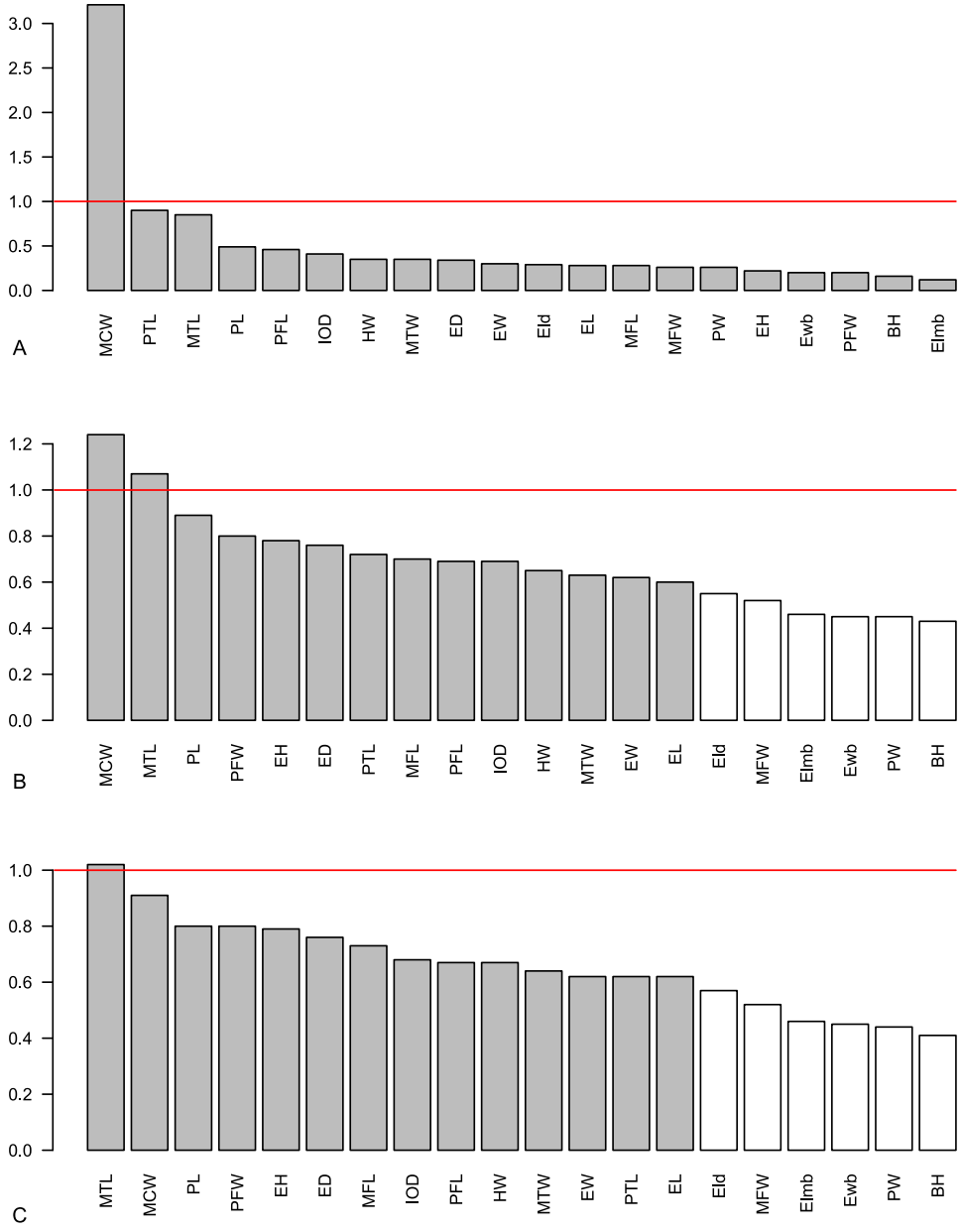

Figure D3. Dependence of Blomberg et al. (2003) descriptive K-statistic from the sampling. Barplots of the K-values for all traits were calculated from the size-corrected data set for (A) the complete sampling (100 Sericini specimens) and reduced Sericini samplings with (B) 10 Sericini specimens and (C) 3 Sericini specimens. White bars indicate non-significance.

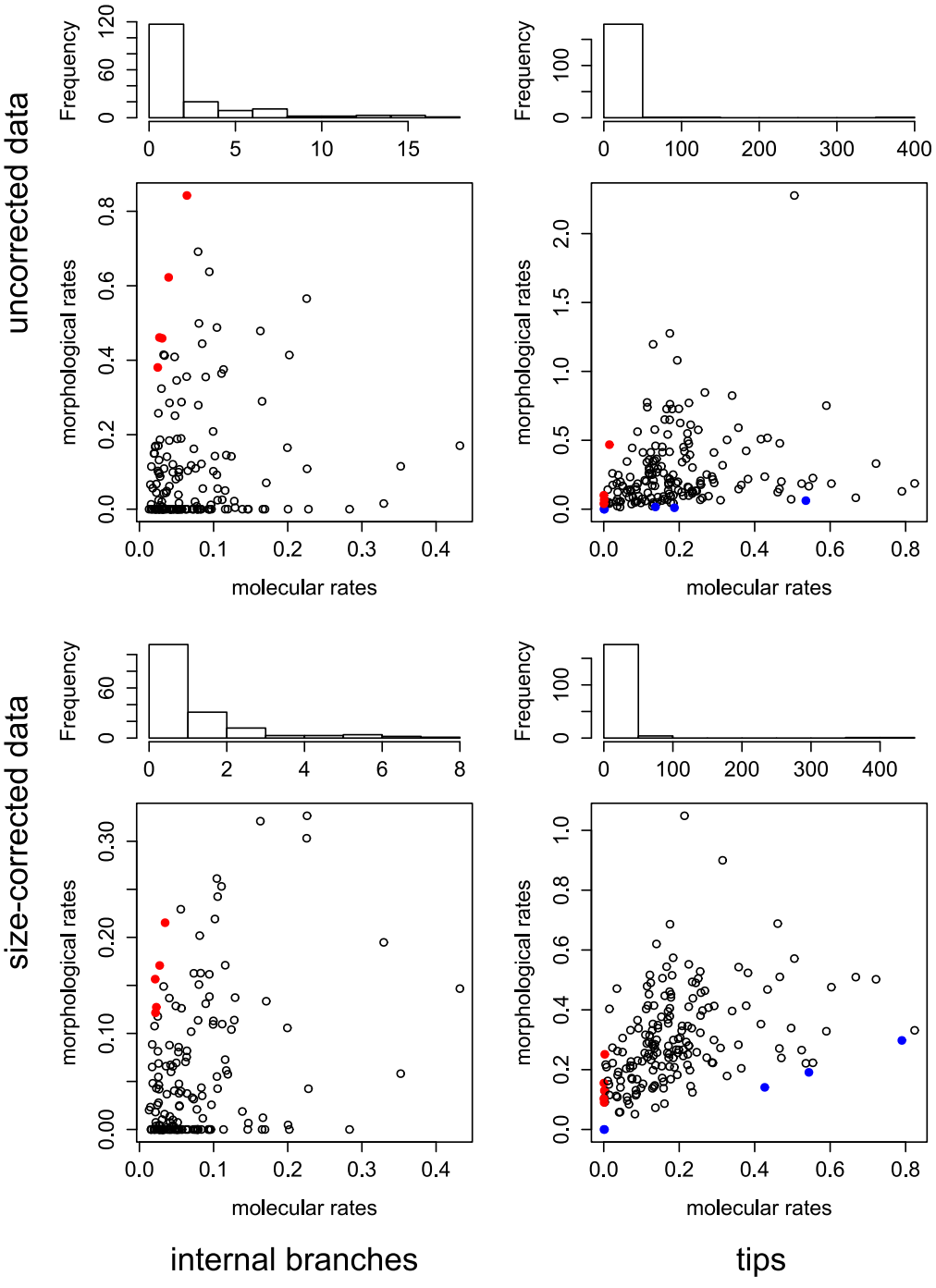

Figure D4. Inference of branches in the trees where directed selection on the morphospace occurred. The scatterplots illustrate morphological and molecular branch lengths for each branch in the trees. The ratios calculated from morphological to molecular branch lengths are quantified in the histograms above. Columns show values for internal branches and tips, rows show uncorrected and size-corrected data (BBPM). Dots of branches with significantly higher and lower morphological rates are indicated in red and blue, respectively.

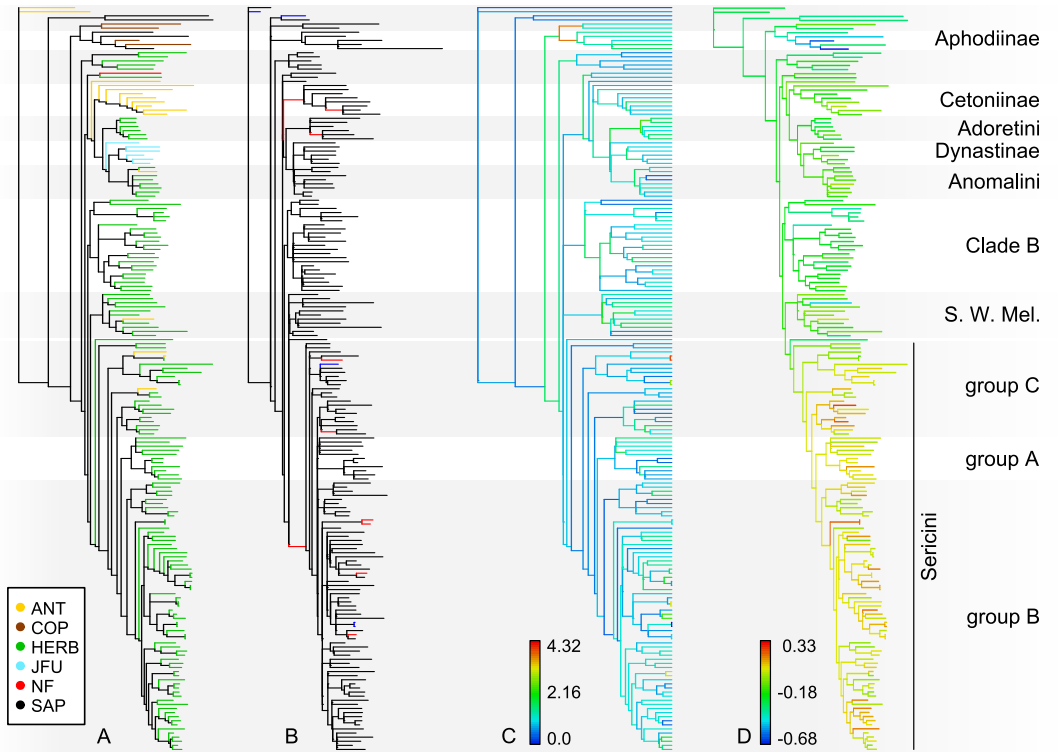

Figure D5. Main results with size corrected data from the linear regression method. (A) Molecular phylogenetic tree (Ahrens and Vogler 2008), (B) tree with optimized branch lengths by the size-corrected data set from the linear regression method, (C) rates of morphological divergence (multivariate standardized phylogenetic independent contrasts) for the size-corrected data set mapped on the ultrametric phylogenetic tree showing relative divergence times, and (D) reconstruction of relative metacoxal length in ancestral nodes of the molecular phylogeny. The tips of the molecular tree (A) are color-coded for feeding habits (ANT = anthophilous, COP = coprophagous, HERB = herbivorous, SFU = sap / fluid utilizers, NF = not feeding, SAP = saprophagous). Branches in (B) with significantly lower (blue) and higher (red) morphological rates of evolution are colored respectively. Background shading indicates clade affiliation.

D.2. Supplementary Tables

Table D1. PCA-loadings for PCs 1–3 of the analysis of the complete sampling. BBPM-size-corrected (corr.) and uncorrected dataset (uncorr.). The last column shows K-values (phylogenetic signal) for every trait (size corrected data) for the complete sampling.

uncorr.	PC1	PC2	PC3	corr.	PC1	PC2	PC3	K-statistic
EL	0.23	0.01	-0.19	EL	-0.02	0.17	0.23	0.28
PL	0.21	0.15	0.01	PL	-0.14	0.02	-0.36	0.49
Eld	0.23	0.03	-0.21	Eld	-0.04	0.19	0.19	0.29
Elmb	0.23	-0.02	-0.02	Elmb	0.01	0.01	0.26	0.12
EW	0.22	0	-0.2	EW	0	0.19	0	0.3
Ewb	0.23	0.05	-0.11	Ewb	-0.06	0.1	-0.09	0.2
PW	0.22	0.07	-0.18	PW	-0.07	0.18	-0.07	0.26
BH	0.22	-0.03	-0.21	BH	0.02	0.2	0.01	0.16
EH	0.23	0.08	-0.57	EH	-0.08	0.55	-0.02	0.22
HW	0.21	0.06	0.01	HW	-0.06	0.01	0.18	0.35
IOD	0.21	0.11	-0.13	IOD	-0.1	0.16	-0.06	0.41
ED	0.22	0.05	0.06	ED	-0.06	-0.06	0.34	0.34
PTL	0.24	0.27	0.33	PTL	-0.29	-0.33	0.03	0.9
PFL	0.23	0.09	0.27	PFL	-0.1	-0.26	0.09	0.46
PFW	0.22	0.07	0.11	PFW	-0.07	-0.09	-0.34	0.2
MCL	0.19	-0.89	0.04	MCL	0.89	-0.05	0.12	3.21
MTL	0.26	0.08	0.37	MTL	-0.11	-0.4	0.21	0.85
MTW	0.24	-0.14	0.13	MTW	0.12	-0.16	-0.48	0.35
MFL	0.23	0	0.3	MFL	-0.01	-0.31	0.12	0.28
MFW	0.23	-0.19	0.1	MFW	0.17	-0.12	-0.35	0.26

Table D2. PCA-loadings for PCs 1–3 of the analysis of subset 1. BBPM-size-corrected (corr.) and uncorrected dataset (uncorr.).

uncorr.	PC1	PC2	PC3	corr.	PC1	PC2	PC3
EL	0.24	0.1	-0.1	EL	0.04	0.04	0.28
PL	0.24	-0.54	0.2	PL	-0.53	-0.09	-0.15
Eld	0.25	0.09	-0.09	Eld	0.02	0.03	0.28
Elmb	0.26	-0.06	-0.56	Elmb	-0.1	0.48	0.53
EW	0.24	-0.05	0.04	EW	-0.09	-0.05	0.13
Ewb	0.23	-0.15	0.03	Ewb	-0.17	-0.02	0.16
PW	0.24	0.04	0.14	PW	-0.04	-0.17	0.22
BH	0.22	0	0.14	BH	-0.03	-0.14	0.08
EH	0.25	0.3	0.53	EH	0.13	-0.62	0.21
HW	0.17	0.35	-0.08	HW	0.4	0.04	-0.09
IOD	0.18	0.29	-0.05	IOD	0.31	0.02	0.03
ED	0.16	0.48	0.02	ED	0.5	-0.08	-0.07
PTL	0.21	0.07	-0.13	PTL	0.1	0.13	-0.21
PFL	0.19	-0.07	-0.22	PFL	0.01	0.25	-0.16
PFW	0.24	-0.17	0.19	PFW	-0.2	-0.16	-0.15
MCL	0.23	-0.29	-0.06	MCL	-0.25	0.13	-0.16
MTL	0.21	0.05	-0.32	MTL	0.08	0.31	-0.2
MTW	0.21	0.02	0.12	MTW	0.02	-0.09	-0.4
MFL	0.19	-0.06	-0.16	MFL	0.01	0.2	-0.2
MFW	0.26	-0.1	0.22	MFW	-0.18	-0.22	-0.12

Table D3. PCA-loadings for PCs 1–3 of the analysis of subset 2. BBPM-size-corrected (corr.) and uncorrected dataset (uncorr.).

uncorr.	PC1	PC2	PC3	corr.	PC1	PC2	PC3
EL	0.2	0.06	-0.23	EL	0.07	0.11	0.17
PL	0.32	-0.27	0.5	PL	-0.48	-0.19	-0.23
Eld	0.22	0.07	-0.23	Eld	0.05	0.09	0.12
Elmb	0.19	0.33	-0.21	Elmb	0.18	-0.1	0.29
EW	0.25	-0.03	-0.06	EW	-0.1	0.05	0.04
Ewb	0.25	-0.1	-0.04	Ewb	-0.13	0.09	0.02
PW	0.24	-0.19	-0.17	PW	-0.11	0.24	0.03
BH	0.24	-0.09	-0.16	BH	-0.08	0.16	0.14
EH	0.27	-0.5	-0.34	EH	-0.26	0.57	0.07
HW	0.11	0.15	-0.26	HW	0.34	0.13	-0.07
IOD	0.13	-0.01	-0.32	IOD	0.27	0.29	-0.41
ED	0.1	0.13	-0.27	ED	0.35	0.14	0.28
PTL	0.17	0.2	0.13	PTL	0.14	-0.19	-0.29
PFL	0.16	0.2	0.05	PFL	0.18	-0.13	-0.33
PFW	0.27	-0.21	0.23	PFW	-0.26	-0.01	-0.33
MCL	0.26	0.13	0.17	MCL	-0.12	-0.24	0.05
MTL	0.17	0.45	0.08	MTL	0.23	-0.36	-0.02
MTW	0.25	0.22	0.14	MTW	-0.08	-0.29	0.47
MFL	0.19	0.25	0.07	MFL	0.12	-0.2	-0.1
MFW	0.31	-0.06	0.21	MFW	-0.32	-0.16	0.09

Table D4. PCA-loadings for PCs 1–3 of the analysis of subset 3. BBPM-size-corrected (corr.) and uncorrected dataset (uncorr.).

uncorr.	PC1	PC2	PC3	corr.	PC1	PC2	PC3
EL	0.23	0.06	-0.05	EL	-0.09	0.05	-0.28
PL	0.2	0.12	-0.09	PL	-0.08	0.11	-0.02
Eld	0.24	0.08	-0.09	Eld	-0.11	0.09	-0.25
Elmb	0.22	0.01	0.04	Elmb	-0.02	-0.03	-0.36
EW	0.22	0.06	-0.28	EW	-0.06	0.28	-0.09
Ewb	0.22	0.07	-0.23	Ewb	-0.07	0.24	-0.06
PW	0.22	0.08	-0.2	PW	-0.08	0.21	-0.08
BH	0.22	0	-0.22	BH	-0.01	0.22	-0.13
EH	0.25	0.13	-0.44	EH	-0.17	0.43	0.02
HW	0.22	0.05	0.15	HW	-0.06	-0.14	0.04
IOD	0.22	0.02	0	IOD	-0.03	-0.01	0.28
ED	0.23	0.11	0.19	ED	-0.13	-0.19	-0.06
PTL	0.24	0.24	0.39	PTL	-0.26	-0.38	0.06
PFL	0.22	0.05	0.26	PFL	-0.06	-0.26	0.04
PFW	0.2	0	0.13	PFW	0.01	-0.12	0.31
MCL	0.16	-0.87	0.08	MCL	0.87	-0.08	-0.27
MTL	0.26	0.02	0.34	MTL	-0.08	-0.35	-0.04
MTW	0.23	-0.23	-0.25	MTW	0.19	0.23	0.6
MFL	0.23	-0.04	0.3	MFL	0.01	-0.3	0.02
MFW	0.21	-0.25	-0.03	MFW	0.22	0.02	0.27

Table D5. PCA-loadings for PCs 1–3 of the analysis of subset 4. BBPM-size-corrected (corr.) and uncorrected dataset (uncorr.).

uncorr.	PC1	PC2	PC3	corr.	PC1	PC2	PC3
EL	0.22	-0.08	0.02	EL	-0.09	0.03	0.28
PL	0.17	-0.02	0.1	PL	-0.08	0.21	0.13
Eld	0.22	-0.07	0.06	Eld	-0.08	0.07	0.24
Elmb	0.22	-0.15	-0.01	Elmb	-0.15	-0.01	0.25
EW	0.2	0.12	0.19	EW	0.07	0.26	0.14
Ewb	0.2	0.1	0.2	Ewb	0.04	0.25	0.08
PW	0.2	0.06	0.17	PW	0.01	0.22	0.13
BH	0.21	0.11	0.11	BH	0.08	0.16	0.19
EH	0.23	0.1	0.51	EH	0.06	0.43	-0.08
HW	0.22	-0.17	-0.01	HW	-0.16	-0.06	-0.06
IOD	0.23	-0.08	0.18	IOD	-0.07	0.07	-0.46
ED	0.22	-0.14	-0.2	ED	-0.11	-0.2	0.23
PTL	0.22	-0.44	-0.11	PTL	-0.4	-0.21	-0.13
PFL	0.23	-0.24	-0.03	PFL	-0.21	-0.12	-0.16
PFW	0.22	-0.1	0.11	PFW	-0.1	0.06	-0.28
MCL	0.25	0.5	-0.55	MCL	0.58	-0.42	0.27
MTL	0.25	-0.21	-0.3	MTL	-0.13	-0.39	-0.09
MTW	0.26	0.41	0.17	MTW	0.44	0.11	-0.46
MFL	0.24	-0.14	-0.25	MFL	-0.07	-0.31	-0.1
MFW	0.24	0.34	-0.19	MFW	0.37	-0.14	-0.1

Table D6. PCA-loadings for PCs 1–3 of the analysis of subset 5. BBPM-size-corrected (corr.) and uncorrected dataset (uncorr.).

uncorr.	PC1	PC2	PC3	corr.	PC1	PC2	PC3
EL	-0.21	-0.07	0.11	EL	-0.06	-0.18	-0.17
PL	-0.17	0.08	-0.06	PL	0.11	-0.15	-0.18
Eld	-0.21	-0.04	0.13	Eld	-0.02	-0.19	-0.13
Elmb	-0.22	-0.13	0.16	Elmb	-0.12	-0.18	-0.03
EW	-0.19	0.22	0.04	EW	0.24	-0.16	-0.17
Ewb	-0.19	0.2	0.07	Ewb	0.22	-0.16	-0.09
PW	-0.19	0.15	0.09	PW	0.17	-0.17	-0.09
BH	-0.21	0.13	0.08	BH	0.13	-0.12	-0.15
EH	-0.24	0.36	0.53	EH	0.35	-0.28	0.27
HW	-0.22	-0.15	0.05	HW	-0.15	-0.04	0.04
IOD	-0.23	0.05	0.28	IOD	0.04	-0.05	0.57
ED	-0.22	-0.23	-0.22	ED	-0.23	0.06	-0.51
PTL	-0.22	-0.42	0.06	PTL	-0.42	-0.06	0.07
PFL	-0.23	-0.21	0.16	PFL	-0.21	-0.06	0.25
PFW	-0.22	0.01	-0.02	PFW	0	0.08	0.3
MCL	-0.26	0	-0.3	MCL	-0.03	0.37	-0.02
MTL	-0.26	-0.33	-0.08	MTL	-0.35	0.18	0.09
MTW	-0.27	0.48	-0.34	MTW	0.43	0.51	0.04
MFL	-0.25	-0.23	-0.1	MFL	-0.25	0.17	0.08
MFW	-0.23	0.18	-0.51	MFW	0.15	0.47	-0.16

Table D7. Percentage of total variation explained by principal components summing up to $\geq 95\%$. BBPM-size-corrected and uncorrected dataset.

	uncorrected					size-corrected										
PC axis	1	2	3	4	5	1	2	3	4	5	6	7	8	9	10	11
Complete	86.9	4.6	2.0	1.6		34.9	15.5	11.9	8.9	7.6	4.2	4.0	3.0	2.1	1.9	1.4
Subset 1	88.5	4.1	3.0			36.0	23.5	14.4	8.6	4.4	3.5	2.7	1.9	1.4		
Subset 2	85.5	4.9	3.7	2.0		47.1	20.5	10.8	6.5	4.5	3.6	2.0				
Subset 3	89.5	3.9	1.5			38.5	13.7	11.3	7.6	6.2	4.4	4.1	3.8	2.0	1.8	1.7
Subset 4	88.5	2.8	2.2	1.4	1.1	23.4	19.1	12.4	9.3	8.5	6.1	5.3	3.9	2.6	2.2	1.9
Subset 5	88.5	3.1	1.9	1.5	1.1	24.9	16.8	12.7	9.9	8.2	7.3	5.6	3.5	2.5	2.2	1.7

Table D8. The impact of size. Percentage of variation explained by size alone (PVESA) within the subsets.

Taxa subset	PVESA (%)
1	88.4
1*	82.4
2	85.2
3	89.5
4	88.5
5	88.4

*without *Microvalgus*

Table D9. Alternative size correction with linear regression: F-values from non-parametric MANOVA of the complete sampling (excluding singletons) regarding 95% of total variation. Values for the size-corrected dataset are shown in the upper triangle, those for the uncorrected in the lower one. Significant differences ($p < 0.05$) are highlighted in bold. Higher F-values for the same significant pairings are underlined in the respective triangle.

	Adoretini	Anomalini	Aphodiinae	Cetoniini	Clade B	Dynastinae	Glaphyridae	Hopliinae	Hybosoridae	Scarabaeinae	Sericini A	Sericini B	Sericini C	SWM	Valgini
Adoretini		5.34*	15.35*	15.18*	**8.93**	12.65*	5.05*	3.95*	4.06*	5.42*	**13.49**	**19.60**	**19.12**	6.34*	6.37*
Anomalini	5.22*		19.02*	**7.86**	**6.17**	**4.60**	4.53*	5.50*	3.00*	6.97*	**6.28**	**8.22**	**6.94**	2.30	3.86*
Aphodiinae	39.30*	47.75*		11.78*	**10.50**	15.75*	5.72	15.36*	3.94	7.72	19.39*	**35.60**	**30.30**	7.89*	8.08
Cetoniini	7.73*	1.29	33.83*		**11.15**	9.18*	3.49*	10.67*	5.33*	8.25*	9.44*	**15.75**	**10.90**	3.16*	3.06
Clade B	3.96	0.88	**21.07**	0.55		**8.35**	3.00*	**6.63**	1.55	10.18*	**14.36**	**35.66**	**21.80**	2.42*	5.76*
Dynastinae	13.83*	3.12	49.12*	0.45	0.98		8.47*	12.72*	2.57	7.47*	**6.92**	**14.28**	**11.91**	3.84*	3.98*
Glaphyridae	0.39	1.41	24.58	2.10	1.01	4.37		2.83*	5.43	3.84	7.39*	**8.84**	7.58*	1.60	5.75
Hopliinae	1.30	8.40*	17.05*	9.73*	6.21*	15.44*	1.18		4.43*	5.78*	**13.97**	**18.53**	**16.35**	4.31*	5.81*
Hybosoridae	2.87	6.29*	13.29	5.92	3.02	9.76*	5.89	0.56		2.82	5.88*	**8.34**	7.70*	1.45	4.31
Scarabaeinae	5.13	0.81	30.33	0.47	0.38	1.13	1.93	5.16	8.10		10.67*	**16.29**	14.56*	5.32*	2.42
Sericini A	8.01*	20.27*	17.92*	**22.65**	**19.12**	**30.91**	4.81*	3.77*	1.63	11.26*		3.81*	3.93*	4.98*	5.71*
Sericini B	2.58	9.99*	**24.01**	**16.18**	**32.11**	**21.02**	1.56	2.64	1.51	5.19*	3.90*		1.76	**9.74**	7.79*
Sericini C	5.51*	**18.89**	17.44	**23.88**	**30.56**	**30.92**	3.17	2.51	1.24	8.68*	0.47	4.64*		**5.48**	6.78*
SWM	0.43	3.28	5.76*	5.05*	7.61*	6.50*	0.35	0.24	0.14	1.47	1.10	1.19	1.22		2.45*
Valgini	4.83	9.63*	0.82	8.25	8.61*	10.72*	1.59	2.09	0.69	2.69	2.82	7.14*	3.56	1.57	

*Significant without sequential Bonferroni correction

Table D10. Alternative size correction with linear regression: F-values from non-parametric MANOVA (Anderson 2001) of each subset (ss1–ss5, excluding singletons) regarding 95% of total variation. Values for the size-corrected dataset are shown in the upper triangle, those for the uncorrected in the lower one. Significant differences ($p < 0.05$) are highlighted in bold

Subset 1	Cetoniinae	Clade A	
Cetoniinae		**8.66**	
Clade A	0.87		
Subset 2	Adoretini	Anomalini	Dynastinae
Adoretini		**5.60**	**12.97**
Anomalini	**0.02**		**4.96**
Dynastinae	**0.00**	0.10	
Subset 3	Clade B	Clade C	
Clade B		**29.17**	
Clade C	**41.85**		
Subset 4	Sericini	SWM	
Sericini		**9.27**	
SWM	1.08		
Subset 5	Sericini A	Sericini B	Sericini C
Sericini A		**3.73**	**3.97**
Sericini B	3.82*		2.03
Sericini C	0.56	4.58*	

*Significant without sequential Bonferroni correction

Table D11. Alternative size correction with linear regression: Correlation between molecular and morphometric distance-matrices for specimens within one feeding type and the complete sampling. Coefficients of determination and p-values from Mantel-tests.

	r	p
complete	0.06	< 0.01
anthophilous	0.25	< 0.01
coprophagous	0.48	0.21
herbivorous	0.43	< 0.01
sap/fluid utilizers	0.29	0.21
saprophagous	−0.30	0.82

Table D12. Results of the phylogenetic least squares analyses. Coefficients of determination and p-values are given for the uncorrected and both size-corrected data sets.

data set	r^2	p
uncorrected	0.03	0.65
size-corrected		
BBPM	0.03	0.06
linear regression residuals	0.06	< 0.01

D.3. Electronic Supplement files

Electronic Supplement D1 Full list of specimens in the study. BMNH-shortcut, group affiliation, and assigned feeding habit are given.

Please find this table on the attached CD at
`./electronic_supplement/Chapter_V/Electronic_Supplement_D1.pdf`

Electronic Supplement D2 R function used to calculate the multivariate standardized phylogenetic independent contrasts (McPeek et al. 2008). The package ape (Paradis et al. 2004) is required. Rownames of the data (x, matrix or data.frame of continuous variables) must perfectly fit the tip labels of the tree (phy, ape format). Please find the R-script on the CD at `./electronic_supplement/Chapter_V/Electronic_Supplement_D2.R`

```
mpic <- function (x, phy)
{
contr <- c()
nb.tax <- length(phy$tip.label)
phy <- makeNodeLabel(phy, prefix="")
x <- as.matrix(x)
while(phy$Nnode >= 1){
nb.tip <- length(phy$tip.label)
nb.nodes <- phy$Nnode
k <- c()
# identify terminal adjacent tips and save their mrca (k)
for(i in phy$node.label){
if(length(extract.clade(phy, i)$tip.label) == 2)
k <- c(k, i)
}
# calculate contrast for all tips descending from nodes in k
# and drop tips
for(j in 1:length(k)){
tips <- extract.clade(phy, k[j])$tip.label
# get branch lengths of branches...
vi <- phy$edge.length[which.edge(phy, tips[1])]
# ...leading to tips
vj <- phy$edge.length[which.edge(phy, tips[2])]
# formula (6) from McPeek et al. (standardized multivariate contrast)
c <- sqrt(sum(diff(x[tips,])^2)) / sqrt(vi+vj)
names(c) <- k[j]
contr <- c(contr, c)
# calculate values for internal node that is now a tip...
Xk <- (vi*x[tips[2],] + vj*x[tips[1],]) / (vi + vj)
x <- rbind(x, Xk)
row.names(x) <- c(rownames(x)[1:nrow(x)-1], k[j])
if(phy$Nnode > 1) phy <- ape::drop.tip(phy, tips, trim.internal=F)
else phy$Nnode <- 0
# ... and scale branch
phy$edge.length[which.edge(phy, as.numeric(k[j]))] <-
phy$edge.length[which.edge(phy, as.numeric(k[j]))] + vi*vj/(vi+vj)
}
}
names(contr) <- as.numeric(names(contr)) + nb.tax
contr[order(as.numeric(names(contr)))]
}
```

E. Appendix to chapter VI

E.1. Supplementary Figures

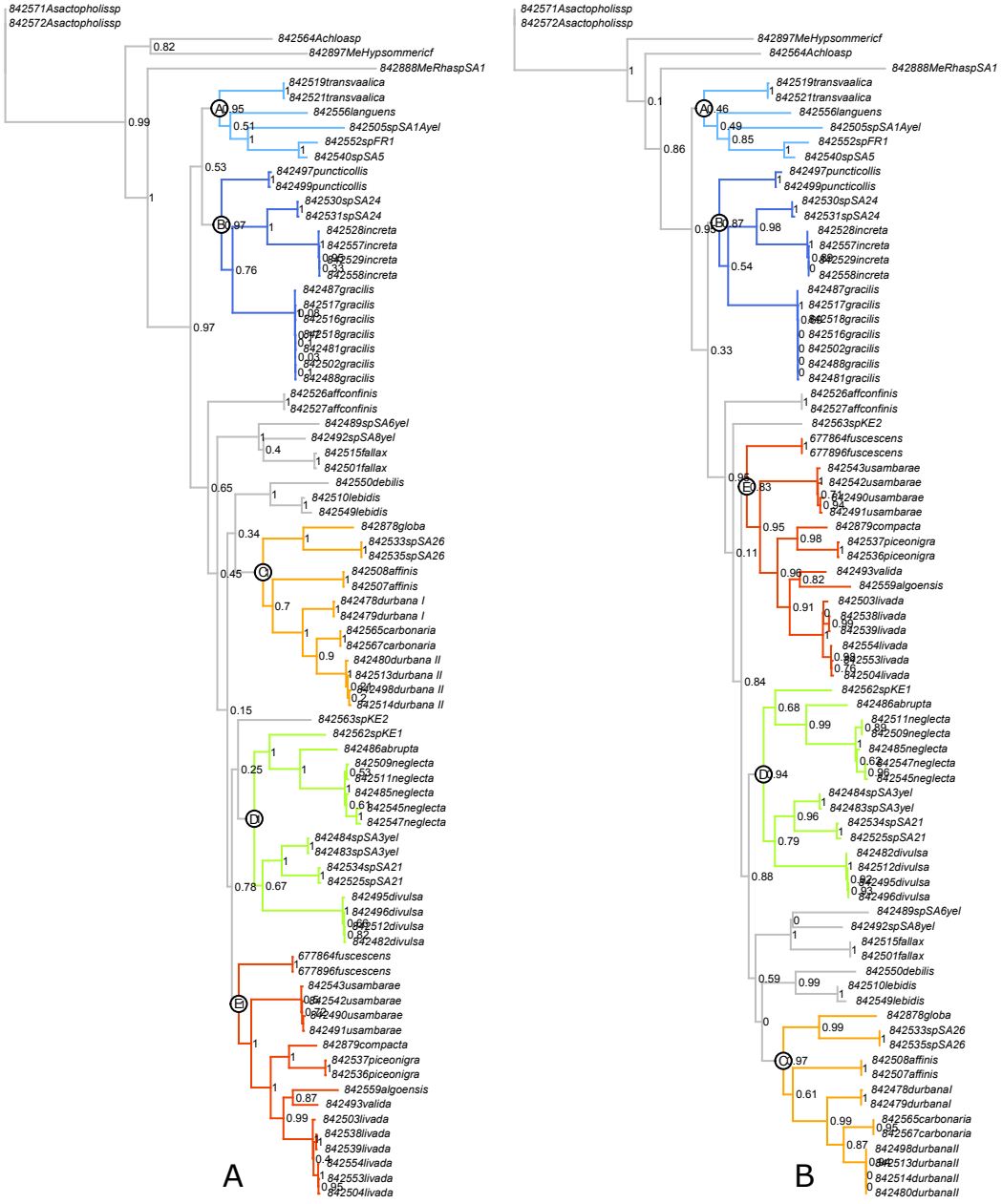

Figure E1. Phylogenetic trees from (A) Bayesian and (B) maximum likelihood inference. Node values represent posterior probabilities from Bayesian inference and SH-like supports, respectively. Clade colors and numberings correspond to the main text.

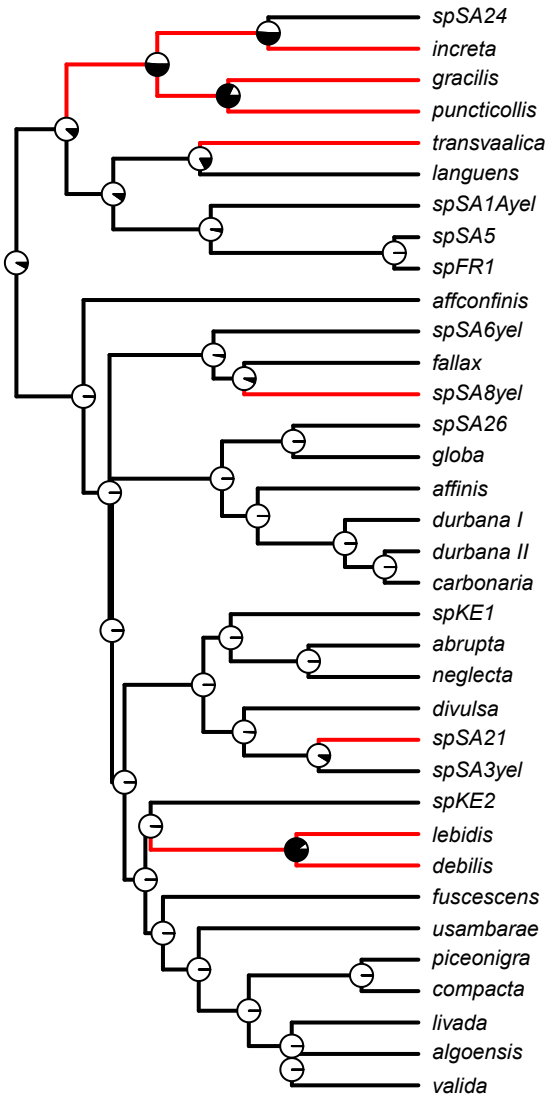

Figure E2. Maximum likelihood reconstruction of asymmetry as a discrete trait in ancestral nodes of the species tree inferred with *BEAST. Pie charts in the respective nodes indicate the probability of an asymmetric (black) or symmetric (white) state. Branches leading to nodes with higher probability for an asymmetric state are colored in red.

E.2. Supplementary Tables

Table E1. Sampling localities of *Schizonycha* specimens under study. Geographical coordinates are given in decimal degrees.

No	Locality	Lat	Long	Acronym
1	KENYA, Eastern E of THIKA 28.12. SW KANGONDE Lgt. Snížek 2007	-1.1833	38.0167	XL8
2	Cheerio Farm, Haenertsburg, ca 20km W of Tzaneen	-23.8952	29.9527	SAF7
3	Legalameetse Nature Reserve, guest house camp (Murchson range)	-24.1999	30.3374	SAF11
4	Klasserie Dam Caravan Park	-24.5299	31.0588	SAF17
5	Rust de Winter Nature Reserve	-25.2231	28.4913	SAF5
6	Groenkloof Nature Reserve, Pretoria	-25.7871	28.2030	SAF3
7	Irene, S Pretoria	-25.8750	28.2167	SAF2
8	Rietvlei Nature Reserve, Pretoria env.	-25.8994	28.2854	SAF1
9	Rietvlei Nature Reserve, Pretoria env.	-25.8994	28.2854	SAF36
10	Ithala Game Reserve, Doornkraal camp (Louwsburg env.)	-27.5125	31.2038	SAF19
11	Ithala Game Reserve, Ntshodwe camp (Louwsburg env.)	-27.5433	31.2834	SAF18
12	Sodwana Bay camping site (St. Lucia Wetland Park)	-27.5514	32.6717	SAF25
13	Sodwana Bay camping site (St. Lucia Wetland Park)	-27.5514	32.6717	SAF27
14	Sodwana Bay camping site (St. Lucia Wetland Park)	-27.5514	32.6717	SAF28
15	Nature Farm (Louwsburg)	-27.5669	31.3001	SAF24
16	Mkuze Game Reserve	-27.5958	32.2194	SAF29
17	Drakensberge, Lodge 15 km NW Himeville	-29.6309	29.4191	SAF32
18	Soutpansberg, 10km NW Makhado	-22.9709	29.8697	SA2009-3
19	NE Nylstroom N of Johannesburg, 14.1.2008	-24.5367	28.8000	XL6
20	Kwazulu-Natal: NE Ndumo, W border Tembe Elephant Park, 80m, 29.12.2007-9.1.2008 (Maputoland)	-26.9297	32.4706	XL7
21	Kwazulu-Natal: Vryheid Mt. Hill Nat. Res., JH254, 27.750621S, 30.800424E	-27.7506	30.8004	XL2
22	Monk's Cowl Forest Reserve (Drakensberge)	-29.0509	29.4028	SA2009-17
23	O.F.S., Orange riv. S of Philippolis, 26.12.2007 lgt. M. Snižek	-30.2797	25.2839	XL4
24	Dombietersfontein	-31.3736	23.1156	XL1
25	Silaka Nature Reserve	-31.6510	29.5086	SA07-8
26	Toorberg E, 32.10S, 24.02E	-32.1667	24.0333	XL3
27	Cwebe Nature Reserve (The Haven)	-32.2405	28.9120	SA07-7
28	Morgans Bay env., (Yellowwood Tree Park)	-32.6965	28.3343	SA07-6
29	North West, Vaal riv., 1250m, 22.12.2007 lgt. M. Snizek	not localized		XL5

E.3. Electronic Supplements

Electronic Supplement E1 Full list of specimen available for this study. BMNH identifier, GenBank submission numbers for each sequenced marker, sampling site acronym, sex, and specimen shortcut are provided if available.

Please find this table on the attached CD at
`./electronic_supplement/Chapter_VI/Electronic_Supplement_E1.pdf`

Electronic Supplement E2 Euclidean distance values between informative principal components of left and right paramere, representing the degree of paramere asymmetry of each specimen and species means.

Please find this table on the attached CD at
`./electronic_supplement/Chapter_VI/Electronic_Supplement_E2.pdf`

Electronic Supplement E3 *BEAST setup `.xml`-file that was used for a single run.

Please find a `.xml`-file on the attached CD at:
`./electronic_supplement/Chapter_VI/Electronic_Supplement_E3.xml`

Electronic Supplement E4 Raw coordinate data of paramere outlines for geometric morphometrics.

Please find a `.zip`-file on the attached CD at:
`./electronic_supplement/Chapter_VI/Electronic_Supplement_E4.zip`

Electronic Supplement E5 R-function that was used for the calculation of multivariate phylogenetic independent contrasts. (See also Appendix D.)

Please find a `.R`-scriptfile on the attached CD at:
`./electronic_supplement/Chapter_VI/Electronic_Supplement_E5.R`

F. Appendix to chapter VII

F.1. Supplementary Figures

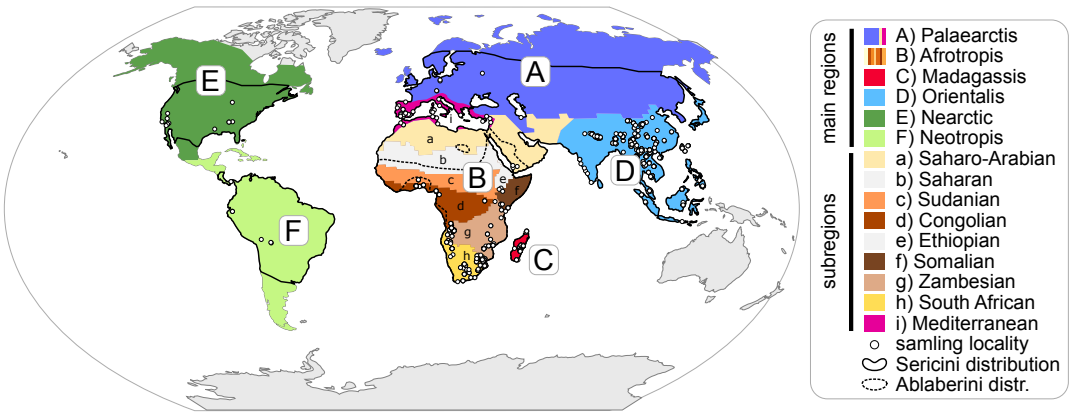

Figure F1. Sampling localities, Ablaberini and Sericini distributions, and areas used for ancestral area reconstructions and estimation.

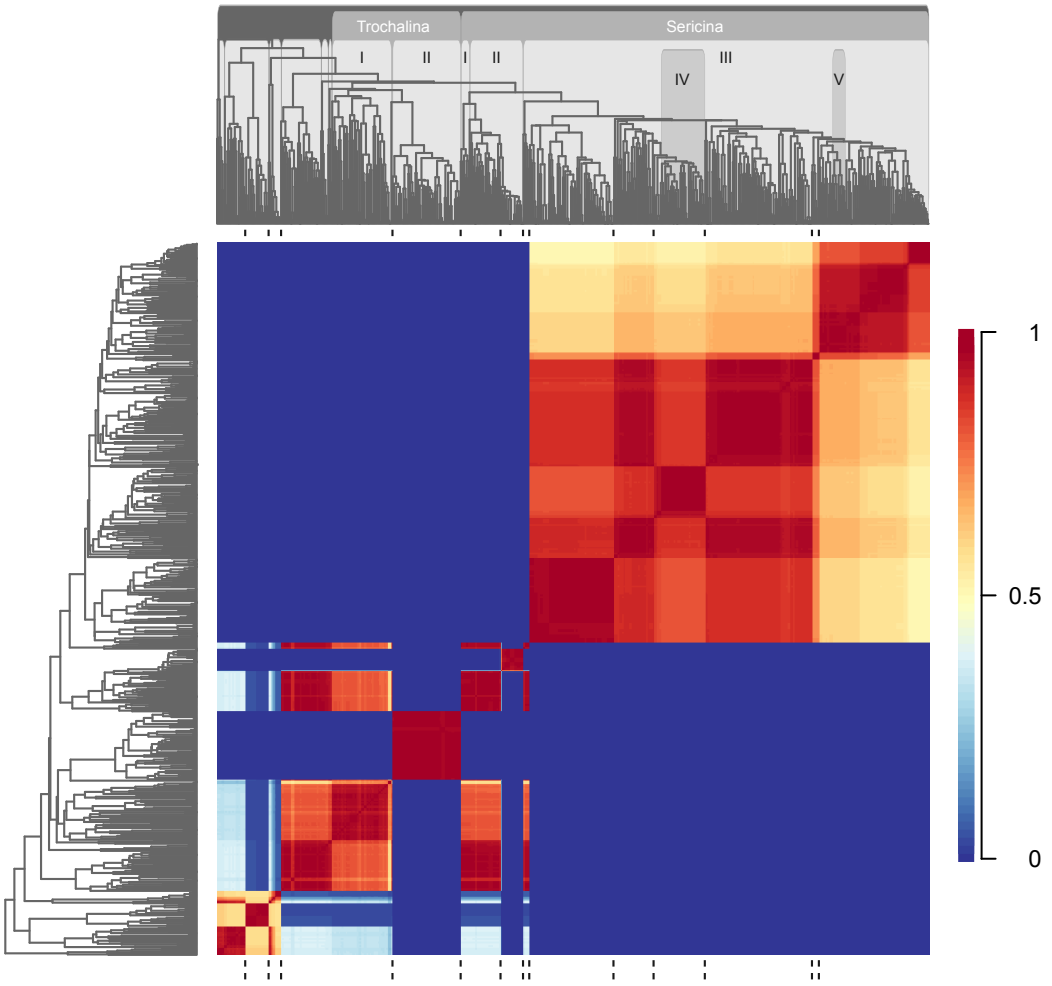

Figure F2. Macroevolutionary cohort matrix for Sericini based on the BAMM analysis. Values between 0 and 1 represent the pairwise probability of species for sharing macroevolutionary dynamics.

F.2. Supplementary Tables

Table F1. Likelihoods and AIC scores of 7 models tested in BioGeoBEARS.

	LnL	no. parameters	AIC
DEC+J	-118.24	3	242.5
DIVALIKE+J	-134.99	3	276.0
BAYAREALIKE+J	-135.40	3	276.8
modified DEC+J	-138.82	3	283.64
DEC	-156.89	2	317.8
DIVALIKE	-160.19	2	324.4
BAYAREALIKE	-220.00	2	444.0

Table F2. Lineages discussed in the text and their generic composition. The node numbers correspond to the main text (Fig. VII.1, Table VII.1).

Node	Clade	Genera
3	*Triodontella* group part.	*Euronycha, Triodontella*
4	*Triodontella* group part.	*Apotriodonta, Hymenoplia, Paratriodonta, Triodontella*
5	*Omaloplia* group part.	*Hellaserica, Omaloplia*
6	*Omaloplia* group part.	*Pleophylla*
7	*Comaserica* group	*Comaserica, Glaphyserica, Hyposerica*
9	*Mesoserica* group part.	*Eriphoserica, Heteroserica*
10	*Mesoserica* group part.	*Mesoserica*
11	*Trochalina* clade I	*Ablaberoides, Allokotarsa, Arraphytarsa, Aulacoserica, Doleroserica, Dolerotarsa, Etiserica, Idaeserica, Microtrochalus, Pseudotrochalus, Trochaloserica, Trochalus*
12	*Trochalina* clade II	*Bilga, Pseudotrochalus, Trochalus*
14	*Sericina* clade I	*Gryphonycha, Lamproserica, Nedymoserica, Neuroserica, Philoserica*
15	*Sericina* clade II	*Archohomaloplia, Microserica, Neoserica, Tetraserica, Trioserica*
16	*Sericina* clade III	*Amiserica, Anomalophylla, Aulacoserica, Calloserica, Chrysoserica, Eumaladera, Euphoresia, Euserica, Gastroserica, Gynaecoserica, Lasioserica, Lepidoserica, Lepiserica, Leuroserica, Maladera, Microserica, Neomaladera, Neoserica, Nepaloserica, Nipponoserica, Oxyserica, Pachyserica, Paramaladera, Serica, Sericania, Taiwanoserica, Trichomaladera, Xenoserica*
17	*Sericina* clade IV	*Aulacoserica, Euphoresia, Lepiserica, Maladera, Neomaladera, Neoserica*
19	*Sericina* clade V	*Serica*

Table F3. PCR Protocols.

		cox1		*rrnl* & 28S	
		Temp. [℃]	Time [min]	Temp. [℃]	Time [min]
1	Initial denaturation	95	15:00	95	15:00
2	Denaturation	94	0:35	95	0:35
3	Annealing	55 / 50	1:30	49	1:00
4	Elongation	72	1:30	72	1:00
5	Final Elongation	72	10:00	72	10:00
	Repeats step 2-4	15x (-1 per cycle) / 25x		35x	

F.3. Electronic Supplement

Electronic Supplement F1 Figure of the BEAST tree with mean node ages and 95% highest posterior density intervals depicted as bar at each node.

Please find this figure on the attached CD at
`./electronic_supplement/Chapter_VII/Electronic_Supplement_F1.pdf`

Electronic Supplement F2 (Left) Ancestral area estimation results from BioGeoBEARS under the best fitting model (DEC+j) and (right) the respective percentage ancestral state likelihoods depicted as pie charts at each node and branch.

Please find this figure on the attached CD at
`./electronic_supplement/Chapter_VII/Electronic_Supplement_F2.pdf`

Electronic Supplement F3 Alternative ancestral range reconstructions on the BEAST tree: (left) maximum likelihood (SHL supports >50 are given at nodes) and maximum parsimony under (middle) the MPR optimality criterion and (right) the ACCTRAN optimality criterion. Colors correspond to Fig. F1 and to the main text.

Please find this figure on the attached CD at
`./electronic_supplement/Chapter_VII/Electronic_Supplement_F3.pdf`

Electronic Supplement F4 Species identification, voucher number, GenBank accession numbers, and sampling locations.

Please find this table on the attached CD at
`./electronic_supplement/Chapter_VII/Electronic_Supplement_F4.pdf`

Electronic Supplement F5 BioGeoBEARS dispersal matrices for 4 time slices.

Please find this table on the attached CD at
`./electronic_supplement/Chapter_VII/Electronic_Supplement_F5.pdf`

G. Appendix to chapter VIII

G.1. Supplementary Tables

Table G1. Location identity with geographical coordinates (referring to Supplementary Table G24).

ID	Name	Lat	Long
L1	South Africa: Cheerio Farm, Haenertsburg, ca. 20 km W of Tzaneen, 1492 m	−25.8994	28.2854
L2	South Africa: Drakensberge, Highmoor, 1582 m, at light	−25.8750	28.2167
L3	South Africa: Kwazulu-Natal: Drakensberge Lodge 15 km NW of Himeville, 1622 m, at light	−25.7871	28.2030
L4	South Africa: Limpopo: Legalameetse Nature Reserve, guest house camp (Murchson range), 814 m, at light	−25.8761	28.2975
L5	South Africa: Kwazulu-Natal: Nature Farm (Louwsburg), 1100 m, at light	−25.2231	28.4913
L6	South Africa: Free State: Wakkerstrom (Wetland Lodge env.), 1802 m	−23.8952	29.9527
L7	South Africa: Eastern Cape: Morgans Bay env. (Yellowwood Tree Park), 4 m	−24.1999	30.3374
L8	South Africa: Fort Fordyce, 1000 m	−24.5299	31.0588
L9	South Africa Mpumalanga Marieskop Forest Reserve (Head quarter station), 1310 m	−24.5855	30.8635
L10	South Africa Mpumalanga Marieskop Forest Reserve (Reserve entrance gate close to indigenous forest, ca 1 km NW head quarter), 1350 m	−24.5829	30.8623
L11	South Africa Mpumalanga Mac Mac forest retreat, ca. 8 km N of Sabie,1278 m	−24.9931	30.8083
L12	South Africa Mpumalanga Waterfal-Boven env., 1489 m	−25.6521	30.3469
L13	South Africa Kwazulu Natal The Ledges, NE of Royal Natal National Park (Drakensberge), 1401 m	−28.6307	28.9761
L14	South Africa Kwazulu Natal Dragon Peaks Park (Drakensberge), 1233 m	−29.0185	29.4367
L15	South Africa Kwazulu Natal Monk's Cowl Forest Reserve (Drakensberge), 1519 m	−29.0509	29.4028
L16	South Africa Kwazulu Natal Ngome forest (Natal), 1123 m	−27.8167	31.4167
L17	Sout Africa Free State Wakkerstrom (Wetland Lodge env.), 1802 m	−27.3388	30.1531
L18	South Africa Fort Fordyce, 1000 m	−32.6833	26.4667
L19	South Africa Eastern Cape Morgans Bay env., (Yellowwood Tree Park), 4 m	−32.6965	28.3343
L20	South Africa Mpumalanga: Graskop env., 10 km before Pilgrim's Rest; "Zur Alten Mine" Lodge, 1525 m, UV-light trap, inside dense forest remains	−24.9296	30.8083
L21	South Africa Mpumalanga: Hlumu Mts. Near Badplaas (5 km W); Hlumu lodge, 1239 m	−25.9459	30.5296
L22	South Africa Mpumalanga: Blyderiverspoort, Forever Resort, 1194 m	−24.5731	30.7804
L23	South Africa Western Cape: Goukamma Nature Reserve; Dunes near Rondavell Chalet, 10 m	−34.0675	22.9474
L24	South Africa Mpumalanga: WITS' Pullen Farm, light trap, 934 m	−25.5719	31.1814
L25	South Africa Port Alfred, Bretton, UV light trap, Coastal dune-scrub forest, 18 m	−33.6177	26.8761
L26	South Africa KwaZulu Natal prov. uMgungundlovu district; Wartburg Local Municipality, Karkloof forest, 1149 m	−29.2971	30.3008

Table G2. Substitution models of nucleotide evolution that were used for Extended Bayesian Skyline inference. All models were inferred by PartitionFinder along with a two-partition-scheme.

species	*cox1*	ITS1
P. fasciatipennis	TrN+I	K80+I
P. ferruginea	TrN+I+G	K80+I
P. navicularis	HKY	K80
P. nelshoogteensis	TrN+I	K80+I+G
P. pilosa	HKY+I	K80
P. pseudopilosa	TrN+I	SYM+I+G

Table G3. Model fit and characterization of hypervolume SDMs. Columns PC1 – PC3 give the contributions of principal components to the hypervolumes.

Species	ROC	kappa	TSS	Volume	PC1	PC2	PC3
P. fasciatipennis	0.86	0.29	0.72	290.04	9.98	10.16	10.77
P. ferruginea	0.84	0.29	0.69	99.71	5.13	3.85	4.14
P. nelshoogteensis	0.90	0.44	0.81	21.36	1.43	1.53	2.22
P. pilosa	0.86	0.18	0.72	102.16	5.58	5.38	5.11
P. navicularis	0.78	0.20	0.57	46.60	3.49	2.88	2.73
P. silvatica	0.83	0.27	0.65	13.92	1.49	1.53	1.80
P. harrisoni	0.57	0.22	0.14	0.05	0.01	0.01	0.02
P. pseudopilosa	0.63	0.12	0.26	15.76	2.52	2.50	2.58
P. warnockae	0.50	0.00	0.00	0.00	0.00	0.00	0.00

Table G4. Summary of the spatial principle component analysis that was performed based on the clipped environmental background of 19 bioclimatic variables.

	Variable	PC1	PC2	PC3
BIO1	Annual Mean Temperature	−0.79	−0.60	0.08
BIO2	Mean Diurnal Range	0.75	−0.11	0.56
BIO5	Max Temp. of Warmest Month	−0.60	−0.55	0.47
BIO6	Min Temp. of Coldest Month	−0.92	−0.23	−0.22
BIO7	Temperature Annual Range	0.71	−0.08	0.56
BIO8	Mean Temp. of Wettest Quarter	−0.35	−0.85	0.29
BIO9	Mean Temp. of Driest Quarter	−0.91	−0.27	−0.08
BIO10	Mean Temp. of Warmest Quarter	−0.74	−0.60	0.23
BIO11	Mean Temp. of Coldest Quarter	−0.88	−0.44	−0.06
BIO12	Annual Precipitation	0.29	−0.36	−0.84
BIO13	Precipitation of Wettest Month	0.56	−0.56	−0.58
BIO14	Precipitation of Driest Month	−0.72	0.52	−0.34
BIO16	Precipitation of Wettest Quarter	0.59	−0.56	−0.57
BIO17	Precipitation of Driest Quarter	−0.68	0.56	−0.34
BIO18	Precipitation of Warmest Quarter	0.60	−0.58	−0.53
BIO19	Precipitation of Coldest Quarter	−0.67	0.58	−0.34
Eigenvalues		7.71	4.09	3.05
Explained Variance [%]		48.17	25.59	19.07

Table G5. Hypervolume overlap statistics. The Sørensen Index is shown in the upper triangle of the pairwise matrix while the lower triangle gives the geometric intersection of the hypervolumes.

	P. fasciatipennis	*P. ferruginea*	*P. nelshoogteensis*	*P. pilosa*	*P. navicularis*	*P. silvatica*	*P. harrisoni*	*P. pseudopilosa*	*P. warnockae*
P. fasciatipennis		0.50	0.09	0.50	0.27	0.09	NA	0.02	NA
P. ferruginea	97.39		0.15	0.68	0.59	0.23	0.00	0.02	NA
P. nelshoogteensis	13.24	8.95		0.12	0.24	0.26	0.00	0.24	NA
P. pilosa	98.18	69.75	8.03		0.52	0.21	0.00	0.00	NA
P. navicularis	45.90	43.16	8.58	38.19		0.43	0.00	0.03	NA
P. silvatica	13.60	13.26	4.62	12.33	13.28		0.00	0.03	0.00
P. harrisoni	NA	0.00	0.00	0.00	0.00	0.00		0.00	0.00
P. pseudopilosa	2.58	1.35	4.46	0.21	0.71	0.47	0.00		0.00
P. warnockae	NA	NA	NA	NA	NA	0.00	0.00	0.00	

Table G6. Model fit and the percentage of bioclimatic variables contribution of *biomod2* ensemble SDMs.

species	kappa	TSS	ROC	BIO2	BIO7	BIO10	BIO11	BIO12	BIO16	BIO17	BIO18
P. faciatipennis	0.44	0.81	0.95	6.04	8.82	8.87	15.38	17.88	5.35	21.68	15.98
P. ferruginea	0.35	0.70	0.89	8.55	7.36	4.09	4.60	19.26	7.43	16.62	32.10
P. navicularis	0.26	0.79	0.91	7.34	2.33	5.70	2.74	22.28	33.71	14.51	11.39
P. nelshoogteensis	0.43	0.93	0.97	16.44	17.48	12.96	10.22	20.47	6.48	5.24	10.73
P. pilosa	0.09	0.64	0.82	4.45	12.43	4.17	7.71	22.44	3.20	42.35	3.26
mean	-	-	-	8.56	9.68	7.16	8.13	20.47	11.23	20.08	14.69

G.2. Electronic Supplement

Electronic Supplement G1 Results from spatial principle component analysis for hypervolume models performed on the clipped climatic background. Principal component scores and loadings are shown in the upper and lower triangle of the plot, respectively. Percentage of total variance explained by principal components is given.

Please find this figure on the CD at:
`./electronic_supplement/PartVII/Electronic_Supplement_G1--G23.pdf`

Electronic Supplement G2 N-dimensional hypervolumes of *Pleophylla* species. Convex expectations of hypervolumes and variable (i.e., principal component) contributions to hypervolumes are shown in the upper and lower triangle of the plot, respectively.

Please find this figure on the CD at:
`./electronic_supplement/PartVII/Electronic_Supplement_G1--G23.pdf`

Electronic Supplement G3 *Biomod2* ensemble distribution models of *P. fasciatipennis* for current, Holocene Altithermal (HA, 6 kya), and Last Glacial Maximum (LGM, 21 kya) conditions.

Please find this figure on the CD at:
`./electronic_supplement/PartVII/Electronic_Supplement_G1--G23.pdf`

Electronic Supplement G4 Distribution models of *P. fasciatipennis* for Holocene Altithermal (HA, 6 kya) conditions inferred by single PMIP3 experiments.

Please find this figure on the CD at:
`./electronic_supplement/PartVII/Electronic_Supplement_G1--G23.pdf`

Electronic Supplement G5 Distribution models of *P. fasciatipennis* for Last Glacial Maximum (LGM, 21 kya) conditions inferred by single PMIP3 experiments.

Please find this figure on the CD at:
`./electronic_supplement/PartVII/Electronic_Supplement_G1--G23.pdf`

Electronic Supplement G6 Circuitscape and Least Cost Corridor landscape connectivity models of *P. fasciatipennis* based on 3 *biomod2* distribution models: (F0) the unrestricted ensemble, (F1) the ensemble restricted by actual forest patches including a probability of occurrence gradient zone, and (F2) the ensemble strictly restricted to forest patches. The legend explains the degree of migration (current flow). for F1 and F2 were considered as models of current and future connectivity. White dots mark sampling localities used for modeling.

Please find this figure on the CD at:
`./electronic_supplement/PartVII/Electronic_Supplement_G1--G23.pdf`

Electronic Supplement G7 *Biomod2* ensemble distribution models of *P. ferruginea* for current, Holocene Altithermal (HA, 6 kya), and Last Glacial Maximum (LGM, 21 kya) conditions.

Please find this figure on the CD at:
`./electronic_supplement/PartVII/Electronic_Supplement_G1--G23.pdf`

Electronic Supplement G8 Distribution models of *P. ferruginea* for Holocene Altithermal (HA, 6 kya) conditions inferred by single PMIP3 experiments.

Please find this figure on the CD at:
`./electronic_supplement/PartVII/Electronic_Supplement_G1--G23.pdf`

Electronic Supplement G9 Distribution models of *P. ferruginea* for Last Glacial Maximum (LGM, 21 kya) conditions inferred by single PMIP3 experiments.

Please find this figure on the CD at:
`./electronic_supplement/PartVII/Electronic_Supplement_G1--G23.pdf`

Electronic Supplement G10 Circuitscape and Least Cost Corridor landscape connectivity models of *P. ferruginea* based on 3 *biomod2* distribution models: (F0) the unrestricted ensemble, (F1) the ensemble restricted by actual forest patches including a probability of occurrence gradient zone, and (F2) the ensemble strictly restricted to forest patches. The legend explains the degree of migration (current flow). F1 and F2 were considered as models of current and future connectivity. White dots mark sampling localities used for modeling.

Please find this figure on the CD at:
`./electronic_supplement/PartVII/Electronic_Supplement_G1--G23.pdf`

Electronic Supplement G11 *Biomod2* ensemble distribution models of *P. navicularis* for current, Holocene Altithermal (HA, 6 kya), and Last Glacial Maximum (LGM, 21 kya) conditions.

Please find this figure on the CD at:
`./electronic_supplement/PartVII/Electronic_Supplement_G1--G23.pdf`

Electronic Supplement G12 Distribution models of *P. navicularis* for Holocene Altithermal (HA, 6 kya) conditions inferred by single PMIP3 experiments.

Please find this figure on the CD at:
`./electronic_supplement/PartVII/Electronic_Supplement_G1--G23.pdf`

Electronic Supplement G13 Distribution models of *P. navicularis* for Last Glacial Maximum (LGM, 21 kya) conditions inferred by single PMIP3 experiments.

Please find this figure on the CD at:
`./electronic_supplement/PartVII/Electronic_Supplement_G1--G23.pdf`

Electronic Supplement G14 Circuitscape and Least Cost Corridor landscape connectivity models of *P. navicularis* based on 3 *biomod2* distribution models: (F0) the unrestricted ensemble, (F1) the ensemble restricted by actual forest patches including a probability of occurrence gradient zone, and (F2) the ensemble strictly restricted to forest patches. The legend explains the degree of migration (current flow). F1 and F2 were considered as models of current and future connectivity. White dots mark sampling localities used for modeling.

Please find this figure on the CD at:
`./electronic_supplement/PartVII/Electronic_Supplement_G1--G23.pdf`

Electronic Supplement G15 *Biomod2* ensemble distribution models of *P. nelshoogteensis* for current, Holocene Altithermal (HA, 6 kya), and Last Glacial Maximum (LGM, 21 kya) conditions.

Please find this figure on the CD at:
`./electronic_supplement/PartVII/Electronic_Supplement_G1--G23.pdf`

Electronic Supplement G16 Distribution models of *P. nelshoogteensis* for Holocene Altithermal (HA, 6 kya) conditions inferred by single PMIP3 experiments.

Please find this figure on the CD at:
`./electronic_supplement/PartVII/Electronic_Supplement_G1--G23.pdf`

Electronic Supplement G17 Distribution models of *P. nelshoogteensis* for Last Glacial Maximum (LGM, 21 kya) conditions inferred by single PMIP3 experiments.

Please find this figure on the CD at:
`./electronic_supplement/PartVII/Electronic_Supplement_G1--G23.pdf`

Electronic Supplement G18 Circuitscape and Least Cost Corridor landscape connectivity models of *P. nelshoogteensis* based on 3 *biomod2* distribution models: (F0) the unrestricted ensemble, (F1) the ensemble restricted by actual forest patches including a probability of occurrence gradient zone, and (F2) the ensemble strictly restricted to forest patches. The legend explains the degree of migration (current flow). F1 and F2 were considered as models of current and future connectivity. White dots mark sampling localities used for modeling.

Please find this figure on the CD at:
`./electronic_supplement/PartVII/Electronic_Supplement_G1--G23.pdf`

Electronic Supplement G19 *Biomod2* ensemble distribution models of *P. pilosa* for current, Holocene Altithermal (HA, 6 kya), and Last Glacial Maximum (LGM, 21 kya) conditions.

Please find this figure on the CD at:
`./electronic_supplement/PartVII/Electronic_Supplement_G1--G23.pdf`

Electronic Supplement G20 Distribution models of *P. pilosa* for Holocene Altithermal (HA, 6 kya) conditions inferred by single PMIP3 experiments.

Please find this figure on the CD at:
`./electronic_supplement/PartVII/Electronic_Supplement_G1--G23.pdf`

Electronic Supplement G21 Distribution models of *P. pilosa* for Last Glacial Maximum (LGM, 21 kya) conditions inferred by single PMIP3 experiments.

Please find this figure on the CD at:
`./electronic_supplement/PartVII/Electronic_Supplement_G1--G23.pdf`

Electronic Supplement G22 Circuitscape and Least Cost Corridor landscape connectivity models of *P. pilosa* based on 3 *biomod2* distribution models: (F0) the unrestricted ensemble, (F1) the ensemble restricted by actual forest patches including a probability of occurrence gradient zone, and (F2) the ensemble strictly restricted to forest patches. The legend explains the degree of migration (current flow). F1 and F2 were considered as models of current and future connectivity. White dots mark sampling localities used for modeling.

Please find this figure on the CD at:
`./electronic_supplement/PartVII/Electronic_Supplement_G1--G23.pdf`

Electronic Supplement G23 Percentual occurrence of *Pleophylla* species in forest subtypes defined by Mucina and Rutherford (2006).

Please find this figure on the CD at:
`./electronic_supplement/PartVII/Electronic_Supplement_G1--G23.pdf`

Electronic Supplement G24 GenBank accession numbers of *Pleophylla* specimens included in DNA based analyses, along with voucher numbers and geographical origin.

Please find this table on the CD at:
`./electronic_supplement/PartVII/Electronic_Supplement_G24.pdf`

Electronic Supplement G25 Characterization of individual PMIP3 paleo-climate models that were used in the present study.

Please find this table on the CD at:
`./electronic_supplement/PartVII/Electronic_Supplement_G25.pdf`

Electronic Supplement G26 Characterization of individual PMIP3 paleo-climate models that were used in the present study.

Please find this table on the CD at:
`./electronic_supplement/PartVII/Electronic_Supplement_G26.xls`

Electronic Supplement G27 All species distribution- and landscape connectivity models that were generated in the present study.

Please find a `.zip`-file on the CD at:
`./electronic_supplement/PartVII/Electronic_Supplement_G27.zip`

H. Erklärung

Ich versichere, dass ich diese Arbeit selbständig verfasst, keine anderen Quellen und Hilfsmittel als die angegebenen benutzt und die Stellen der Arbeit, die anderen Werken dem Wortlaut oder Sinn nach entnommen sind, kenntlich gemacht habe. Beiträge von Koautoren zur Originalpublikation sind am jeweiligen Kapitelanfang gelistet.

Diese Arbeit hat in dieser oder ähnlichen Form keiner anderen Prüfungsbehörde vorgelegen.

18. April 2017