SEX-BIASED DISPERSAL OBSCURES SPECIES BOUNDARIES IN INTEGRATIVE SPECIES DELIMITATION APPROACHES

Jonas Eberle[1], Erika Bazzato[1,2], Silvia Fabrizi[1], Michele Rossini[1,3], Mariastella Colomba[3], Davide Cillo[4], Marco Uliana[5], Ignazio Sparacio[6], Guido Sabatinelli[7], Rachel C M Warnock[8], Giuseppe Carpaneto[9], Dirk Ahrens[1]

[1] *Centre of Taxonomy and Evolutionary Research, Zoologisches Forschungsmuseum Alexander Koenig Bonn, Adenauerallee 160, 53113 Bonn, Germany;*

[2] *Via Madonna di Campiglio 22, 09045 Quartu Sant'Elena, Italy;*

[3] *Strada dei Guazzi, 1/3, 61122, Pesaro (PU), Italy;*

[4] *Via Zeffiro 8, 09130 Cagliari, Italy;*

[5] *Museo di Storia Naturale, Santa Croce 1730, 30135 Venezia, Italy;*

[6] *Via E. Notarbartolo 54 int. 13, 90145 Palermo, Italy;*

[7] *493, Route de la Fontaine, Le Grand Pré, Villa 12, 01280 Prévessin, France;*

[8] *ETH Zürich, D-BSSE, Mattenstrasse 26, 4058 Basel, Switzerland;*

[9] *Dipartimento di Scienze, Università Roma Tre, Viale Marconi 446, 00146 Roma, Italy.*

Corresponding author: Dirk Ahrens; Zoologisches Forschungsmuseum Alexander Koenig Bonn, Adenauerallee 160, 53113 Bonn, Germany; Tel.: +49 228 9122 286; Fax: +49 228 9122 212; Email: ahrens.dirk_col@gmx.de

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

Jonas Eberle[1], Erika Bazzato[1,2], Silvia Fabrizi[1], Michele Rossini[1,3], Mariastella Colomba[3], Davide Cillo[4], Marco Uliana[5], Ignazio Sparacio[6], Guido Sabatinelli[7], Rachel C M Warnock[8], Giuseppe Carpaneto[9], Dirk Ahrens[1]

[1] *Centre of Taxonomy and Evolutionary Research, Zoologisches Forschungsmuseum Alexander Koenig Bonn, Adenauerallee 160, 53113 Bonn, Germany;*

[2] *Via Madonna di Campiglio 22, 09045 Quartu Sant'Elena, Italy;*

[3] *Strada dei Guazzi, 1/3, 61122, Pesaro (PU), Italy;*

[4] *Via Zeffiro 8, 09130 Cagliari, Italy;*

[5] *Museo di Storia Naturale, Santa Croce 1730, 30135 Venezia, Italy;*

[6] *Via E. Notarbartolo 54 int. 13, 90145 Palermo, Italy;*

[7] *493, Route de la Fontaine, Le Grand Pré, Villa 12, 01280 Prévessin, France;*

[8] *ETH Zürich, D-BSSE, Mattenstrasse 26, 4058 Basel, Switzerland;*

[9] *Dipartimento di Scienze, Università Roma Tre, Viale Marconi 446, 00146 Roma, Italy.*

Corresponding author: Dirk Ahrens; Zoologisches Forschungsmuseum Alexander Koenig Bonn, Adenauerallee 160, 53113 Bonn, Germany; Email: ahrens.dirk_col@gmx.de

Abstract. — Accurate delimitation of species is crucial for a stable taxonomy, which provides the foundation for the study of evolutionary biology, ecology and essentially all biological disciplines. Several approaches towards impartial and repeatable taxonomic practices are available but all existing methods have potentially unacceptable shortcomings. In particular, problems can arise when the underlying model assumptions are violated, for instance in the presence of reduced gene flow. This is observed in the context of sex-biased dispersal, which is a common but underappreciated feature in many groups of organisms. Previously, simulations have indicated that sex-biased dispersal may lead to erroneous estimations of the true species numbers. However, this phenomenon has never been examined using empirical data. We evaluate the bias introduced by extreme female philopatry on a range of *de novo* (GMYC, PTP, ABGD, statistical parsimony, trinomial distribution of triplets model [tr2]) and validation (STACEY, iBPP) approaches to species delimitation in the scarab beetle genus *Pachypus*. Since female philopatry exhibited in this genus in particular can affect mitochondrial gene flow, we compared the results from analyses of single loci, mitochondrial loci, nuclear loci and combined data, as well as the performance of morphometric data as a secondary data source in a fully integrative Bayesian framework. Large overestimation of species numbers was observed across all analyses of combined and mitochondrial DNA datasets, suggesting specimens from nearly every sampling location as separate species. The use of nuclear data resulted in more reasonable estimations of species boundaries, which were largely supported by morphometrics of linear measurements, while geometric morphometrics of body outlines resulted in stronger splitting. Simulations of population divergence with migration, corresponding to the biology of *Pachypus*, showed that female philopatry strongly increases reciprocal monophyly of mitochondrial markers and may substantially contribute to over-splitting in species delimitation. Robust results

recovered using nuclear DNA and morphological data nevertheless enabled us to reach novel conclusions about species boundaries in *Pachypus*. Our findings suggest that mitochondrial DNA will be less suited to species delimitation in many cases, in particular in the presence of sex-biased dispersal.

[integrative species delimitation; sex-biased dispersal; deep mitochondrial coalescence; multispecies coalescent; mitochondrial DNA; nuclear DNA; morphometrics]

DNA-based approaches to species delimitation have undergone substantial development (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; Jones et al. 2014; Fujisawa et al. 2016) since a crisis in taxonomy was announced (Gewin 2002; Godfray 2002; Dayrat 2005). Many approaches have been widely applied to mitochondrial DNA, due to the success of the Barcoding initiative (Hebert et al. 2003b, 2003a; Ratnasingham and Hebert 2007), and proposals for rapid biodiversity assessment and turbo-taxonomy (Hebert and Gregory 2005; Butcher et al. 2012; Riedel et al. 2013). Unfortunately, many existing approaches can lead to major under- or overestimation of biodiversity (Meyer and Paulay 2005; Meier et al. 2006; Rubinoff et al. 2006; Balke et al. 2013; Katouzian et al. 2016). Since accurate delimitation of species is crucial for stable taxonomy as a foundation in all biological disciplines, there has been substantial criticism against the use of single mitochondrial markers for species delimitation (Seberg and Petersen 2009; Dupuis et al. 2012), and alternative integrative approaches, based on multiple sources of evidence, have been proposed for more robust species delimitation (Wiens and Penkrot 2002; Dayrat 2005; Will et al. 2005; Bond and Stockman 2008; Cardoso et al. 2009; Padial et al. 2009,

2010; Schlick-Steiner et al. 2010; Andújar et al. 2014). In particular, the inclusion of multi-locus species delimitation approaches based on multispecies coalescent theory (O'Meara 2010; Yang and Rannala 2010; Ence and Carstens 2011; Fujisawa et al. 2016; Jackson et al. 2017) and the simultaneous analysis of morphological and molecular data have provided promising alternatives to single marker approaches (Yeates et al. 2011; Solís-Lemus et al. 2015; Eberle et al. 2016; Huang and Knowles 2016).

However, methodological pitfalls are numerous. Species delimitation results may be confounded by shortcomings in upstream phylogenetic reconstruction and branch length estimation methods (Leache and Fujita 2010; Astrin et al. 2012; Tang et al. 2014) as well as variability in effective population size (N_e) among closely related species, which violates the assumptions of many speciation models (e.g., the mixed Yule–coalescence model) (Monaghan et al. 2009; Esselstyn et al. 2012; Fujisawa and Barraclough 2013; Ahrens et al. 2016). With use of the multispecies coalescent model, increasing attention has also been paid to the interplay of the population genetic parameters θ (population mutation rate) and τ (time between species divergences) in guaranteeing reliable species delimitation (Zhang et al. 2011; Esselstyn et al. 2012; Reid and Carstens 2012; Fujisawa and Barraclough 2013; Ahrens et al. 2016; Luo et al. 2018).

The birth-death model used to describe the speciation process in existing multispecies coalescent approaches also poses limitations (Sukumaran and Knowles 2017). By failing to model the natural protraction of the speciation process (Avise and Walker 1998; Avise et al. 1998; Nosil et al. 2009; Purvis et al. 2009; Rosindell et al. 2010; Etienne and Rosindell 2012; Etienne et al. 2014; Lambert et al. 2015), all lineage divergences along the history of a given

species may be evaluated as a speciation event rather than intraspecific population divergence. This may cause significant overestimation of species numbers.

In this context, mitochondrial data in particular may be prone to errors. Smaller N_e relative to autosomal DNA in diploids and higher mutation rates increase the genetic structure of the mitochondrial genome, accelerating the formation of reciprocal monophyly among subspecies (Palumbi et al. 2001). Reduced gene flow at the population level can cause further problems and may result in severe overestimation of true species diversity. In addition, the mode of mitochondrial inheritance can result in higher interspecific introgression, relative to autosomal markers, and confound species delimitation results (Bond and Stockman 2008; Hey 2009; Lohse 2009; Harrington and Near 2012; Andriollo et al. 2015; Sukumaran and Knowles 2017).

Of particular interest is the potential bias caused by sex-biased dispersal, where gene flow is reduced in the least dispersing sex. This common phenomenon can influence the potential for introgression of maternally versus paternally transmitted genes (Petit and Excoffier 2009) and can have a large impact on the genealogical patterns observed for the mitochondrial genome (Melnick and Hoelzer 1992; Palumbi and Baker 1994; Lyrholm et al. 1999; Castella et al. 2001; Kerth et al. 2002). The potential impact of sex-biased dispersal on species delimitation has already been explored using coalescent-based simulations (Dávalos and Russell 2014). Results showed that subtly restricted female-mediated gene flow between populations can lead to high false-positive error rates of distance- and tree-based species delimitation methods using mitochondrial-like markers. This phenomenon has never been assessed using empirical data.

Sex-biased dispersal is best known from vertebrates 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 present a case study based on the scarab beetle genus *Pachypus*,

which exhibits extreme male-mediated dispersal. Females of *Pachypus* lack elytra and hindwings and have an almost entirely hypogean (fossorial) life style (Crovetti 1969; Arnone and Sparacio 1990). *Pachypus* is the only genus in an ancient pleurostict lineage of chafers (Pachypodini) that originated around 80 Ma (Ahrens et al. 2014). The genus is exclusively distributed in the central Mediterranean (Crovetti 1969; Baraud 1985, 1992; Sparacio 2008), where it is known since at least the Oligocene (de Serres 1829). Taxonomists initially believed that *Pachypus* comprised only three extant species (Baraud 1985, 1992), but two more species were recently described from Sardinia (Sparacio 2008; Guerlach et al. 2013). Due to their hidden lifestyle, females are rarely found and almost all available data are based on males. Males actively disperse by flight in search of females that release pheromones from subterranean tunnels (Crovetti 1969; Arnone and Sparacio 1990). Since mitochondria are matrilineally inherited, we expect the dispersal of autosomal markers to be less affected (Prugnolle and de Meeus 2002), while mitochondrial lineages are expected to evolve quasi independently among spatially restricted populations.

Following recommendations to use external evidence such as morphology or ecology to verify the outcome of species delimitation hypotheses (Barley et al. 2017; Sukumaran and Knowles 2017), the present study is an empirical test case for such an integrative approach, in which lineage divergence is most likely confounded by the presence of extreme male-mediated sex-biased dispersal. We aim to verify the impact of methodological shortcomings and sex-biased dispersal on commonly used *de novo* species delimitation and species validation approaches (Dávalos and Russell 2014; Sukumaran and Knowles 2017), and use simulations of population divergence with migration corresponding to the biology of *Pachypus* to further examine the observed results. We investigate if the use of morphometric data can help to overcome these issues in a fully integrative Bayesian framework using the multispecies

coalescent. Specifically, we evaluate the species boundary related signal in linear measurements of body parts ("traditional morphometrics"), as well as that of outline shapes of body and genital parts ("geometric morphometrics"). Integrative species delimitation associated with different combinations of traits (morphological and/or molecular markers) should help to elucidate how the choice of data sampling (quality and intensity) may affect species delimitation and shed light on the obscure species boundaries in *Pachypus* beetles.

MATERIAL & METHODS

The study is based on a total of 192 *Pachypus* specimens from 62 sampling localities (Supplementary Figure A1; Supplementary Tables A1, A2; doi: 10.5061/dryad.g97d432) of which 164 were suitable for morphometrics. Five outgroup taxa were chosen from closely related lineages (Ahrens et al. 2014; Supplementary Table A1). Maps were made in QGIS using data from ETOPO1 (http://www.ngdc.noaa.gov/mgg/global/relief/ETOPO1/data/) and Natural Earth (www.naturalearthdata.com).

Molecular Lab Procedures

We prepared molecular data 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 the ArgK-primers amplified multiple products, the band with the correct length was determined with *GeneRuler™ 100bp DNA Ladder Plus* and cut from the Agarose-gel. Macrogen (Seoul, South Korea) sequenced forward and reverse strands using the same primers. Sequences were manually edited in Geneious® 7.1.8. Vouchers were deposited in the collections of the Zoological Research Museum A. Koenig, Bonn (ZFMK).

Sequence Alignment and Phylogenetic Inference

Sequences were simultaneously aligned per marker using the divide-and-conquer realignment techniques implemented in SATé-II (Liu et al. 2009, 2012). Sub-problems with a maximum size of 102 specimens were aligned with MAFFT (Katoh et al. 2002, 2005, Katoh and Toh 2008, 2010) and subsequently merged with MUSCLE (Edgar 2004). The tree estimator was set to RAxML (Stamatakis 2006). Subsequently, we checked the resulting multiple sequence alignments 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 approximation of nucleotide substitution were chosen based on likelihood scores under the gamma model of rate heterogeneity (GTRGAMMA). Final tree searches 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 using 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.

Morphometric analyses

For morphology-based species inference we used “traditional morphometrics” of eight linear distance measurements (Fig. 1, Supplementary Table A4) and four shape traits (Fig. 1). Distance measurements were taken directly from the specimen using an ocular grid on a Zeiss SM20 Stereomicroscope. Body parts were measured with both endpoints in focus at the same time to ensure horizontal orientation. Raw measurements were size corrected, since most variation in biological data sets is usually attributable to 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 onto the isometric size vector and returning it to the original coordinate system (Adams and Rohlf 2000) using *R* code provided by Blankers et al. (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. 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. 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 (elytra, clypeus) all but the first landmark were defined as sliding landmarks (semilandmarks). For open curves (parameres, pronotum) all but the first and last landmarks were defined as sliding landmarks. Procrustes distance was used to optimize positions of the sliding landmarks (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* (Dinno 2009). This method adjusts for bias in the retention of components that is caused by finite sample sizes. Principal components with an adjusted Eigenvalue greater than 1 were used for downstream analyses. The variation within these data sets was summarized by PCA on the correlation matrix. Size was not further regarded in any of the analyses.

De Novo Species Delimitation Approaches

On the basis of whether the samples were partitioned into species entities prior to the species delimitation analyses or not, delimitation methods were grouped into "*de novo*" delimitation and "species validation" (Ence and Carstens 2011) approaches. The term "species discovery" (Ence and Carstens 2011) used previously in this context is misleading, since species boundaries are discovered, not species, and thus the term is confounded with the true discovery of an unknown new species. Therefore, we prefer to use the term "*de novo* species delimitation" (e.g., Zhang et al. 2013; Ahrens et al. 2016) for approaches that aim to partition samples into species without any *a priori* information regarding species membership. Molecular data were explored using four commonly employed and one recently published *de novo* species delimitation methods: (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 or duplicate haplotypes to avoid false

positives (Ahrens et al. 2016). Analyses were performed on unique haplotype data sets that were inferred with *haplotypes* in *R* (v. 1.0, Aktas 2015). Excluded specimens were subsequently assigned to putative species that shared identical haplotypes.

Molecular data.—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 dated April 11th 2013) on distance matrices that were inferred with IQ-TREE (Nguyen et al. 2015) for each locus, correcting distances using the best fit substitution models according to the BIC (Supplementary Table A1). ABGD applies a set of prior intraspecific divergences to detect the position of the barcode gap. Since the approximate position of the barcode gap was unknown, the prior intervals on intraspecific divergence for each locus were determined by examining histograms of the pairwise distances: we evaluated a set of 100 prior intraspecific divergences between 1 % (Pmin) and 100 % (Pmax) of the maximum observed distance. Since higher intraspecific divergence was expected due to the extreme philopatry of *Pachypus* females, this procedure was applied for all loci including *cox1,* for which the barcode gap is typically expected to be 1–3 % pairwise sequence divergence (Puillandre et al. 2012a). The default setting of 1.5 for the relative gap width (interspecific gap / intraspecific gaps; X) is based on the assumption that the barcode gap is larger than any gap observed at the intraspecific level. Since no barcode gap was observed with this setting, X was set to 0.5. This led to a partitioning of specimens at every observed gap in the distribution of pairwise distances that was at least half the extent of the preceding gaps at lower distance values. Despite the high genetic structure that is observed in the present case study that may assimilate intra- and interspecific gaps, the resulting partitions may still reflect species and were evaluated subsequently based on further data and comparison to alternative 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 1.21 (Clement et al. 2000), with gaps treated as a 5th character state. This method infers potential species as significantly differentiated genetic 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 assumed to be non-homoplastic based on a given probability. It has been shown that a 95 % probability performs well in insect species identification (Monaghan et al. 2006; Ahrens et al. 2007; Hart and Sunday 2007). Additionally, we applied a 90 % and 99 % connection limit to account for potential bias induced by philopatry. *Cox1* was analyzed using 91 % instead of 90 %, because the latter caused TCS to hang.

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 between speciation and coalescence, i.e. the putative species boundary is detected based on an increase in the accumulation of lineages through time (Pons et al. 2006). Maximum likelihood trees for each gene and combined partitions (i.e., mitochondrial genes, nuclear genes, and all genes) were inferred with RAxML using the same procedure as described above, using unique haplotypes or unique combinations of haplotypes. For the combined data sets we performed two analyses: one run including all available data and one run with complete alignments only (i.e., only specimens without missing loci). Trees were made ultrametric in PathD8 (Britton et al. 2002, 2007), assigning the root an arbitrary age of one. PathD8 applies rapid rate smoothing and has been shown to be suitable for use 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 has been 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 multimodel comparison approach (Burnham and Anderson 2002) was used to assess species support.

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 approach, 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). Both the likelihood (PTP) and Bayesian (bPTP) approaches were implemented, using the web service (http://species.h-its.org/ptp/) for the same single loci RAxML trees generated for the GMYC analyses. 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 were greater 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, 2014; Ence and Carstens 2011; Fujisawa et al. 2016). A greater degree of concordance between gene trees of recombining sexual organisms is expected between gene trees at and above the species level than the intraspecific level (Degnan and Rosenberg 2009). One approach based on this assumption is implemented in the program tr2, which can be used to infer species boundaries, i.e., the transition point between species branching and within species branching. It does so by fitting the distribution of concordant "triplet topologies" (rooted trees consisting of three tips) among gene

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

Morphological data.—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 clusters. The best fitting number of clusters and model were chosen based on the BIC, which penalizes the log-likelihood for more complex models. Simpler models generally require more clusters for a good representation of the data, while fewer clusters are needed with more complex models (Fraley and Raftery 2002). All available models for a predefined range of clusters (1 to 100) were investigated using the combined 25 informative principle components (PCs) of linear measurements (2 PCs; Supplementary Table A5) and all four outline analyses (aedeagus: 6, clypeus: 6, elytra: 6, and pronotum: 5 PCs), and in two alternative sets of analyses, the two most informative PCs of linear measurements. A reduced optimal subset of variables was chosen with *clustvarsel* 2.2 in *R* (Scrucca and Raftery 2014) for an additional *mclust*-analysis that combined all traits. This was shown to improve clustering if only a subset of the available variables are informative (Scrucca and Raftery 2014).

Species Delimitation by Validation

Multispecies coalescent models (O'Meara 2010; Yang and Rannala 2010, 2014; Ence and Carstens 2011; Fujisawa et al. 2016) can theoretically accommodate incomplete lineage sorting (Rannala 2009; Camargo et al. 2012) and variable N_e, which is a major advantage compared to single-locus delimitation approaches. In addition, the integration of multi-locus molecular data with morphological traits has yielded promising results (Solís-Lemus et al. 2015; Eberle et al. 2016). Due to their computational demand (Yang and Rannala 2014; Fujisawa et al. 2016), most methods require *a priori* designations of individuals to populations or putative species. Here we employed the programs STACEY (Boukaert et al. 2014; Jones 2015) and the BPP (Yang and Rannala 2010, 2014) derivate iBPP ("integrative BPP"; Solís-Lemus et al. 2015), to assess support for *de novo* species under the multispecies coalescent. We defined initial species entity hypotheses based on the results of the *de novo* species delimitation (i.e., GMYC clusters). In order to decrease uncertainty in the results, specimens with large amounts of missing data were omitted from these analyses.

For the STACEY analysis, the data was partitioned into 4 loci to allow independent site models, mutation rates and gene trees to be estimated. Mitochondrial genes (*cox1* and *rrnL*) shared one tree because they do not represent independent unlinked loci. Site models were set according to IQ-TREE (v1.3.13; Nguyen et al. 2015) inferences, with the exception of *cox1*, which was set to the second best scoring model (Supplementary Table A3). This is because TIM2-based models are not available in BEAST. After confirming a reasonably clock-like distribution of rate heterogeneity among branches of both nuclear partitions in a preliminary analysis (i.e., 95 % highest posterior density interval of the coefficient of rate variation of the lognormal clock included zero), we implemented strict molecular clock models. This simplification may be preferable and produces more accurate rate estimates (Ho et al. 2005) and

topologies (Drummond et al. 2006) in cases of low rate heterogeneity. A relaxed lognormal clock was set for mitochondrial partitions. Two variants of the analysis were conducted, using two alternative sets of priors to describe the collapse weight prior (ω): a beta-binomial distribution with (i) $\alpha = 1$ and $\beta = 1$, and (ii) $\alpha = 40$ and $\beta = 2$. The first combination places an equal prior probability on all species numbers, while the second places the highest probability on five species entities, which is much lower than the expected number of potential species. The scale parameters α and β were calculated using the R script from the STACEY documentation, which relies on the VGAM package (Yee and Wild 1996; Yee 2015). Each run was repeated 4 times with 10^8 generations each, sampling every 10^4 generations. During burnin 10 % of the samples were removed before assessing stability and convergence of the combined runs by examination of trace plots and ESS values (ESS > 200) for all parameters 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 using the BEAST module *SpeciesDelimitationAnalyzer* (Jones et al. 2015), applying collapse heights (ε) of 10^{-4}, 10^{-3}, and 10^{-2}. Similarity matrices were visualized using the R code provided by Jones et al. (2015).

Substantial progress towards integrative species delimitation was made by enabling the combined analysis of continuous trait data with molecular 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 the "lumping" of *a priori* defined entities is restricted to predefined sister clades (Leache and Fujita 2010; Olave et al. 2014; Eberle et al. 2016). However, alternative guide tree topologies should be tested, which can be informed using alternative sources of evidence, including molecular, morphological (Eberle et al. 2016), or geographic data. We used the species tree output from the STACEY analyses with the inferred species entities from the GMYC analysis

based on the combined dataset (in the following referred to as minimum clusters = 'MINCs') as a guide tree. Since the number of initial entities is limited in iBPP, the guide tree was split into two parts that were analyzed separately (Supplementary Fig. A7). Node support was generally low in the second part of the guide tree (node B; Fig. 2, Supplementary Fig. A3) and the topology was inconsistent with the maximum likelihood tree inferred with RAxML using the combined partitioned data (Fig. 2, Supplementary Fig. A3). Therefore, an alternative guide tree topology based on the relationships established using nuclear 28S and geography was tested (Supplementary Fig. A7c). We used eight different combinations of molecular and morphological data sets to assess the influence of different data on species delimitations: (i) all loci and all traits (i.e., combined retained PCs of linear measurements and outline analyses of the 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 excluding DNA, (vi) linear measurements only, (vii) outline traits only, and (viii) nuclear DNA only. Additionally, we ran iBPP without data to evaluate the priors. The priors for the population mutation rate (θ) and the root height (τ_0) can have a large impact on the results and must be chosen carefully (Leache and Fujita 2010; Eberle et al. 2016). Because there is a great deal of uncertainty associated with these parameters we applied nine pairwise combinations of three gamma distributions $G(\alpha,\beta)$ 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 third gamma prior distribution (α = 2, β = 20000) is extreme and may lead to false positives or negatives (i.e., splitting a true species or lumping two or more true species) when applied to θ and τ_0, respectively, especially if the data is uninformative. It was chosen to test the robustness of results (see discussion) and should not be considered biologically realistic.

The control files were created using BPPmulti (*perl* scripts available at http://github.com/eberlejonas/BPPmulti). We used the standard species delimitation algorithm that assigns equal probabilities to rooted species trees, since the alternative algorithm assigns equal probabilities to all labeled histories and can over-resolve large unbalanced guide trees (Yang and Rannala 2010). Generally, there is no consensus about the value of the posterior probability for a given node to be considered a "good" split between two valid species. Previous studies considered probabilities above 0.95 as support for a speciation event (Leache 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 have adopted iBPPs internal behavior and considered support for a speciation event to be good for posterior probabilities > 0.90. Results were visualized with BPPmultitool (*R* scripts available at http://github.com/eberlejonas/BPPmultitool). MINCs that were inferred by *de novo* delimitation approaches were mapped onto the first two principal components of three morphological data sets (linear measurements, all outline data, and all morphological trait data together) to visualize morphological divergence among the clusters.

Population Demographic Simulation with Sex-Biased Dispersal

We simulated autosomal and mitochondrial DNA in a two population model to verify that the high genetic structure in mitochondrial loci observed in *Pachypus* could be explained by sex-biased dispersal rather than matrilineal inheritance or higher mutation rates in mitochondrial relative to autosomal loci. We replicated the simulations of Dávalos and Russell (2014) and extended them to incorporate parameter values that reflect the biology of *Pachypus*. Specifically, we included simulations for one- and two-year life cycles (since no explicit information is available for *Pachypus*) and further reduced female migration rates ($m=1.25\times10^{-5}$, 1.25×10^{-6}, and

1.25×10^{-7}). Ancestral population size was varied from 10 to 10^5 and population divergence time was varied from 5×10^3 to 5×10^6 years (Supplementary Table A10), which covers the time since the end of the Messinian salinity crisis to the present. Trees and sequences were simulated with ms (version from March 4, 2014; Hudson 2002) and seq-gen (v1.3.2; Rambaut and Grass 1997) as implemented in phyclust (v0.1-22; Chen 2011). Median genetic distance between populations and percentage reciprocal population monophyly were calculated as summary statistics for 100 replicates per parameter set using *ape* (Paradis et al. 2004) and illustrated with *lattice* (Sarkar and Andrews 2016) using perceptual uniform *viridis* colors (Garnier 2018).

RESULTS

Phylogenetic inferences

Multiple sequence alignment of 28S, ArgK, *cox1,* and *rrnL* resulted in matrices of 651, 724, 808, and 521 bp, respectively. Gene trees showed major topological discordance between all loci, despite overall high node support, particularly between the mitochondrial and the nuclear gene trees (Supplementary Figs. A2, A3). The tree obtained using all markers combined (Fig. 2) reflected the topology of the tree obtained using the mitochondrial data. There were several clades of closely related specimens that were consistently recovered in all analyses (although these specimens were often indistinguishable on the basis of 28S). These clades often represented groups that were collected at the same or nearby localities. Only three localities in northern (L15) and southern Sardinia (L39, L59) were populated by more than one of these stable clades. In contrast, individuals from the northeastern coast of Sardinia (L11, L14, L37, and

L38) were distributed throughout the mitochondrial trees but were represented by only a single haplotype from each locality (Supplementary Figs. A2, A3).

A close relationship between a Tunisian, a Sicilian, and a south-east Sardinian clade that are ascribed to *P. demoflysi, P. caesus*, and *P. sardiniensis,* respectively, was recovered in all tree reconstructions (Fig. 2; Supplementary Figs. A2, A3). They constitute one of the earliest branches in the tree based on the combined data (Fig. 2). Additional stable clades, found in all tree reconstructions, included one from south-western Sardinia (*P. melonii*), one from Elba, one from the eastern coast of the Tyrrhenian sea (*P. excavatus*), one from the Ionian coastal plain (*P. candidae*), and one sampled from across Sardinia and Corsica. However, the internal relationships within clades remained obscure, as these nodes were highly inconsistent across gene trees and had low node support values (Supplementary Figs. A2, A3).

De Novo Species Delimitation Approaches

The range of putative species entities found by different *de novo* species delimitation approaches was large, ranging from one entity (statistical parsimony networks with nuclear markers) to 111 entities (GMYC results based on the tree of the combined mitochondrial data) (Figs. 2, 3). The largest number of entities was inferred with mitochondrial DNA. Analysis of this data assigned specimens even from the same locality to independent putative species (GMYC; Fig. 2). Most methods predominantly joined specimens from the same sampling locality together in a single putative species entity. The uncorrected genetic distances between specimens was high, only 4 pairwise comparisons among MINCs were estimated to be less than 4 % divergent for *cox1*, while ~60 % were estimated to be 13-18 % divergent, which is approximately 10-fold the expected distance between beetle species. However, the distribution of pairwise distances revealed no clear barcoding gap (Supplementary Fig. A4) and thus resulted in

unreliable results in the ABGD analysis (Supplementary Table A6), even for *cox1*. Depending on parameter choice, ABGD and TCS lumped together large numbers of specimens (Fig. 3). No consistent estimate for the number of species was recovered across a range of initial parameter values that may indicate a species-boundary related signal was present in the data.

Irregular likelihood surfaces and low species support values for the GMYC entities (Supplementary Figs. A16-A25) indicated a bad fit of this model to all data sets. The GMYC uses an abrupt increase in lineage accumulation at the species-population boundary to infer species entities; however, lineage through time plots did not exhibit this signal at any point for any marker (Supplementary Figs. A16-A25). Compared to the GMYC approach, bPTP and in particular, the maximum likelihood version of PTP, resulted in less over-splitting. However, the posterior probabilities (PPs) for the bPTP entities (and the equivalent support values calculated for the maximum likelihood PTP analyses) suggested low support for majority of the inferred putative species (Fig. A26). Delimitation using the combined molecular data (GMYC and tr2) produced very different results in comparison to the analysis using single markers, particularly for the GMYC model. In addition, results between mitochondrial, nuclear, and combined analyses were very inconsistent (Fig. 3). The GMYC approach yielded 40 entities based on the analysis of all markers, which was less than half the number obtained using the combined mitochondrial markers (n=111). Interestingly, the combined nuclear data set consistently inferred a larger number of putative species (n=58), compared to the single nuclear gene analyses (n_{28S}=18, n_{ArgK}=39). tr2 yielded overall more consistent results between nuclear, and combined analyses (Fig. 3), with entity numbers (n=51 and 43, respectively) being similar to those obtained with the GMYC analysis of the combined data sets.

Mixed Gaussian clustering with *mclust* (Fraley and Raftery 2002) rarely recovered the boundaries between clades that were recovered by any of the DNA-based approaches (Fig. 2). A large overlap in the morphological variation obscured most putative species boundaries and resulted in low numbers of clusters. Only two clusters were inferred using all trait data (Fig. 3). Using the subset of variables inferred by *clustvarsel* resulted in three clusters, while analysis of linear measurements alone showed no signal for more than a single cluster (Fig. 3).

Species Delimitation by Validation

Given the huge range of putative species numbers from the *de novo* delimitation and computational constraints, we reduced subsequent validation approaches to a single hypothesis to test. We chose the putative species entities (s1-s40) from the GMYC analysis of the combined data that included all specimens (referred to as minimal clusters [MINCs] in the following text; Supplementary Fig. A3).

STACEY analyses supported all 40 MINCs as separate species (Supplementary Fig. A6). Alternative prior expectations on the number of species (i.e., the uninformative collapse weight prior ω, with equal weight on all possible species numbers, versus the informative ω prior, specifying 5 species to be the most likely) yielded nearly identical results (Supplementary Fig. A6). The estimated species number only reduced with the highest applied collapse height (ε=0.01) to 38 with the uninformative ω prior (a beta-binomial distribution with $\alpha = 1$ and $\beta = 1$) and to 35 with the informative ω prior ($\alpha = 40$, $\beta = 2$).

iBPP was used with eight combinations of molecular and morphometric data sets as well as nine different prior parameter choices and was repeated for a partly geography-informed species tree (Figs. 4, 5; Supplementary Fig. A9-A11). The majority of splits (bifurcations in the guide tree) were robust to prior parameter choice (i.e., all PPs from analyses with different prior

combinations were well above or below 0.95/0.90), meaning it was possible to draw unequivocal conclusions for these nodes. However, about 10 % of the splits were sensitive to the choice of both θ and τ_0 (population mutation rate and root height) priors (excluding the most extreme prior combinations, which affected all splits). Prior distributions that placed a higher probability on large θ values tended to lump a larger number of MINCs together, while prior distributions that placed a higher probability on large τ_0 values tended to result in more splits (e.g., Fig. 5b, d). However, this pattern was not consistent throughout the guide trees (Fig. 5a). The results were less sensitive to the τ_0 prior than the θ prior, as demonstrated by the variable consistency of PPs obtained using different prior combinations.

All analyses including all molecular loci, together with all morphological traits confirmed our previous findings with regards to over-splitting. All 37 initial MINCs (3 MINCs excluded due to incomplete morphological data) were supported as separated species, with the exception of s12 and s34. However, support for many speciation events decreased when the mitochondrial loci or all molecular data were excluded. The inclusion of all morphological traits tended to support more speciation events than the analysis of linear measurements only. The smallest number of species was found using nuclear loci, linear measurements, or the combination of these data. Compared to other data combinations, nuclear loci (with or without linear measurements) showed the sharpest species boundaries and were the least dependent on the choice of priors. They resulted in the highest number of well-supported speciation events, which are indicated by having strong support (e.g. > 0.90 PP) for a given speciation event across all 9 (3×3) prior combinations, and low support (e.g. < 0.30 PP) for the descendent (nested) nodes (green vs. red boxes).

Testing an alternative species tree topology corresponding to the geographical vicinity of MINCs (Fig. 5; subtree at node B of Fig. 2, Supplementary Fig. A7c) considerably decreased PPs at "intraspecific" nodes of the affected species (Fig. 5b, d, f, *sp3* and *sp5*; Supplementary Figs. A11 vs. A10). However, strong support (>0.95) for separate species based on molecular and morphological data remained for the deepest splits in the respective clades (Fig. 5, splits 18, 40, and 41).

Mapping the MINCs as well as iBBP species hypothesis (from linear measurements and nuclear loci) onto the morphological data sets showed a large degree of overlap between most clusters (Supplementary Figs. A12, A13). Noticeable differentiation was found among several closely related molecular clades (e.g., *P. demoflysi, P. candidae,* and *P. caesus*; Supplementary Fig. A13), which were also distinctly separated in the canonical variate analyses of all morphological data (Supplementary Fig. A14). Linear measurements showed less distinction among MINCs and putative species (Supplementary Fig. A15).

Population Demographic Simulation with Sex-Biased Dispersal

Simulations of population divergence incorporating variable migration rates supported the hypothesis that sex-biased dispersal drives the observed extreme mitochondrial divergence of populations (Fig. 6, Supplementary Figs. A27, A28; Supplementary Table A10). In simulations incorporating a two-year life cycle, populations had a nearly 90 % probability to be reciprocally monophyletic after only 100,000 years, assuming a standard mitochondrial mutation rate (10^{-8} mutations/year) and 0.00025 % female migration rate per generation (1 female per 4000 males), which is reasonable in the case of *Pachypus*. Median genetic distances between populations were 1 % after 5 Ma and thus concordant with values observed from real data for the most closely related populations, if we assume a separation of populations with the end of the Messinian

salinity crisis. Further lowering female migration rates strongly increased distances between populations. Divergence was lower with a generation time of one year, only reaching 5 % divergence after 10^7 years given very small ancestral population sizes and lowest female migration. Simulated autosomal loci showed neither reciprocal monophyly nor population divergence, which is expected considering higher N_e, male migration, and their mode of inheritance.

DISCUSSION

Sex-biased dispersal and the pitfalls of species delimitation

Sex-biased dispersal is widespread in nature, although it is substantially less well characterized for invertebrates than for vertebrates. It can have profound impacts on the distribution of mitochondrial and nuclear genotypes and is thought to impact species delimitation (Petit and Excoffier 2009; Dávalos and Russell 2014). Here, using a case study of *Pachypus* beetles that exhibit extreme female philopatry, we demonstrate that it can also affect species delimitation, especially for any analysis that includes mitochondrial DNA. In particular, we find that *de novo* species delimitation approaches (i.e. methods that make no prior assumptions about true species identities) recover a large, biologically unreasonable number of species (Fig. 3). These results are in part driven by extreme divergence of mitochondrial DNA, observed even for conspecific specimens collected at different localities. Furthermore, we show using simulations based on the biology of *Pachypus* that sex-biased dispersal can explain this high degree of divergence (Fig. 6). A presumptive null-hypothesis that the exaggerated genetic structure of mitochondrial DNA was simply observed due to its faster mutation rate, smaller N_e, and

matrilineal inheritance (Ballard and Whitlock 2004), could be rejected. Mean distances and reciprocal monophyly under the assumption of unbiased dispersal of females and males (μ:10^{-8}, migration rate: 1 %; Supplementary Figs A27, A28) were distinctly lower than empirically observed.

The results of our analyses highlight the potential issues associated with relying on mtDNA for species delimitation. Alternatively, nuclear DNA is considered valuable and may identify "sharper" more robust boundaries between species because it undergoes recombination and is expected to have a higher effective population size (N_e) (Prugnolle and de Meeus 2002; Currat et al. 2008; Petit and Excoffier 2009). Slower mutation rates of nDNA compared to mtDNA may be disadvantageous but with the increasing availability of genome scale data this issue may become negligible and is outweighed by the strong advantage of biparental inheritance. However, autosomal DNA will also be impacted by sex-biased dispersal, since the females of the species are not involved in dispersing autosomal genotypes, leading to increased geographical structuring and smaller N_e (Wright 1946), relative to species that do not exhibit female philopatry. This means autosomal markers may also be prone to over-splitting in species delimitation analyses. Indeed, we show that while nuclear DNA produces more reasonable results, the number of putative species predicted by *de novo* approaches was still large (Fig. 3). Interestingly, although morphology shows considerable overlap between species and produced inconsistent species delimitations (Fig. 2, Supplementary Fig. A5), morphological data show evidence of geographical structuring below the species level (Supplementary Fig. A12, A13) which likely reflects the structure of autosomal DNA.

One of the most important factors impacting *de novo* delimitation approaches is effective population size (N_e, and thus the population mutation rate, $\theta = 4N_e\mu$), which must be sufficiently

large, relative to the age of species, and uniform across species to result in identifiable species boundaries (Esselstyn et al. 2012; Reid and Carstens 2012; Fujisawa and Barraclough 2013; Ahrens et al. 2016). This may be especially relevant in the present case study, where multiple factors are expected both to reduce and drive variation in N_e. (i) Populations with limited dispersal capacity tend to have smaller N_e, which in the case of *Pachypus*, will be driven by limited dispersal among females and will affect mtDNA in particular. (ii) The divergent lifestyle between males and females will also drive increased selection among nuclear loci, which will result in decreased N_e, an effect that is further amplified by linkage disequilibrium (Liu and Mittler 2008; Slatkin 2008). (iii) Rugged topography and variable habitat availability across the range of *Pachypus* may cause N_e to vary considerably between different species, populations, and demes (Barrowclough 1980). (iv) Several species are expected to have low N_e simply because they are rare (Palstra and Fraser 2012; Ahrens et al. 2016). All of these features may explain the tendency of the *de novo* delimitation methods employed here to overestimate the number of species among *Pachypus*, using mitochondrial and – to a lesser extent – nuclear loci.

Incomplete sampling presents a challenge to any approach to species delimitation (Lohse 2009), but may be more problematic in combination with geographic structuring driven by sex-biased or other types of limited dispersal. Incomplete sampling means that some localities of a given species' distribution will not be represented in available molecular datasets. If a species exhibits geographic structuring, failure to sample some localities will lead to underestimation of the true molecular diversity of the species and consequently to the erroneous proposal of more than one species. Since limited dispersal drives geographic structuring this may be especially important for groups of species such as *Pachypus*. The effects of incomplete sampling have been studied in detail for the GMYC model (Reid and Carstens 2012; Fujisawa and Barraclough 2013;

Ahrens et al. 2016) and will also apply to other *de novo* delimitation methods (Ross et al. 2008; Zhang et al. 2013; Yang and Rannala 2016).

There remains a great deal of uncertainty in the *Pachypus* species recovered using different *de novo* methods, preventing a "statistical democracy" (i.e. identifying species that are recovered in the majority of analysis). An alternative to *de novo* approaches is to use parametric species validation approaches based on the multispecies coalescent (STACEY, iBPP) that can incorporate variation in N_e. We find some combinations of data (nuclear DNA, morphometrics of linear measurements, or their combination) result in stable estimates of species, evident from the analysis in iBPP using different prior combinations (Fig. 5). However, we also find that sex-biased dispersal in *Pachypus* is likely to cause problems for these methods as well. Overall, the results of the analysis using iBPP showed that this method has the potential to over-split. For example, the results based on total evidence analysis suggest that nearly every sampling locality of *Pachypus* is associated with a separate species (Fig. 5e, f). One relevant criticism of the birth-death model implemented in iBPP is that is assumes gene flow ceases immediately following speciation, reflecting the assumptions of the biological species concept (Mayr 1995; Yang and Rannala 2010). However, temporarily isolated populations may appear divergent despite their ability to merge back into the parent species' genepool if isolation barriers were to disappear prior to speciation (i.e. protracted speciation). Previously, Sukumaran and Knowles (2017) demonstrated that can lead to over-splitting using simulations and the results obtained here show that this is likely to be a problem for empirical datasets.

A critical step in these analyses involves selecting the initial set of minimal clusters (MINCs). In the absence of an alternative objective approach, validating the output of *de novo* species delimitation analysis is considered a valuable strategy (Zhang et al. 2013). Since

validation approaches are only capable of evaluating splits between MINCs, but not further subdividing these clusters, the use of the smallest possible MINCs may be desirable. This may be impractical for several reasons. First, due to their complexity, Bayesian multispecies coalescent methods are computationally intensive and not easily scalable (Fujisawa et al. 2016) and currently BPP is limited to 20 MINCs. Second, a large number of singletons may be problematic. BPP requires substantially more data for accurate delimitation if only a single specimen is available for a given species (Zhang et al. 2011) and their influence on results is not well understood (Ahrens et al. 2016). However, the natural rareness of many species means that some species will always only be represented by a small number or even a single individual(s) (singletons) (Lim et al. 2012). Third, in the case of *Pachypus*, *de novo* delimitations differed widely in their results and recovered a large number of total species (Fig. 3). The GMYC entities (MINCs, n=40) resulting from the complete combined partitioned molecular data set (Fig. 2, Supplementary Fig. A3) were considered an appropriate starting point for validation, separated across two guide trees for analysis using iBPP. Alternative delimitations with higher numbers of entities were considered biologically unlikely, and would have introduced a large number of singletons. Morphology did not produce results that had a clear link to genealogy or provenience (Fig. 2) and were thus also not considered as a reasonable starting point. However, as the results of our integrative analysis show, morphological data still has a valuable role to play in species validation analysis.

Guide tree choice is another critical aspect of setting up analysis using iBPP (Leache and Fujita 2010; Zhang et al. 2014; Eberle et al. 2016). In particular, we confirmed that specifying accurate phylogenetic relationships among MINCs is important in minimizing over-splitting. In addition, we show that exploring the use of alternative guide trees that are derived from different

methods or on the basis of different data can identify support for species. For example, the conspecificity of MINCs s6, s8, and s9 (*P. sp2*) was supported on the basis of a geographical-informed modified guide tree (Fig. 5b, d, f vs. Supplementary Figs. 7, A10, A11). Similarly, the conspecificity of MINCs s6 and s8 was supported on the basis of the similarity matrices obtained from the STACEY analyses. Nevertheless, we emphasize the need to interpret these results critically and highlight the challenge in identifying non-circular guide tree topologies based on independent evidence (e.g., morphology, phenology, or geographical origin).

Finally, our results confirmed the importance of examining the use of different prior parameter choices on the robustness of species delimitations using BPP and iBPP (Leache and Fujita 2010; Eberle et al. 2016). In general, nodes with high posterior probabilities (PPs) (i.e. nodes with 9 green boxes in Fig. 5, where each box represents a different θ and τ_0 prior combination) that have descendent nodes with low PPs (i.e., nodes with 9 red boxes) indicate a strong signal in the data for a given putative species. Nodes that receive mixed support under different prior settings (i.e., green, yellow, and red boxes) would indicate that a split is not strongly supported by the data. This consideration is independent of the number of sampled loci, although more loci are likely to produce more robust results. A strong contrast between support for putative speciation events and descendent nodes representing populations (i.e. green versus red boxes) was observed for the iBPP analyses based on the combined and individual analyses of nuclear and linear measurements data (e.g., Fig. 5a, b). These data can thus be considered to reflect putative species boundaries better than the mitochondrial DNA. This supports the initial hypothesis that nuclear data should be less affected by sex-biased dispersal and gains additional support from congruent results based on morphology.

Implications for Species Boundaries in Pachypus

Given the above considerations, deriving conclusions about species boundaries in *Pachypus* is not straightforward. Addressing the major issues in the iBPP analyses (configuration of initial entities [MINCs], their placement in the guide tree, robustness of results to different priors, and choice of data), the combined analyses based on nuclear DNA and morphological measurements resulted in a conservative estimate of twelve *Pachypus* species, recovering all six nominal species of *Pachypus* recognized to date as valid entities (Figs. 5, 7).

Results are translated for taxonomic conclusions as follows: despite deep population differentiation, *P. melonii* was confirmed as single entity. Strong phenological evidence also supports the integrity of *P. caesus* and its distinctiveness from *P. demoflysi*: the Sicilian *P. caesus* is autumnal (Crovetti 1969), while all other *Pachypus* species have summer adults. The taxonomic identity of validated *P. excavatus* and *P. candidae* from the Italian mainland, (s12/s34 and s10/s11, respectively) is also supported by the geographic overlap or close vicinity of our samples with the respective type localities. Support of the newly hypothesized species (*P. sp1-6*), so far classified as *P. candidae* (Sparacio 2008; Guerlach et al. 2013), was non-ambiguous for three of these entities (*P. sp1*, *4*, *6*) (Fig. 5), while the situation is less clear for the remaining three (*P. sp2, 3*, *5*). The absence of monophyly of *P. sp2* and *P. sp3* was addressed using a nDNA and geography-informed guide tree, which enforced monophyly (Fig. 5) and by excluding mtDNA and outline data. In this scenario the entities *P. sp2*. *P. sp3* and *P. sp5* were supported as distinct from other species based on the iBPP results. Despite support from iBPP to further subdivide *P. sp3* and *P. sp5*, we chose not to separate these into multiple species to avoid inflated species numbers and potentially invalid taxonomic names. This conservative approach is justified given the aforementioned tendency of all data sources to have exaggerated structure and thus to over-split species. Further support for their species status comes from the CVA analyses

based on all morphological evidence, which clearly separates them from their sister species but does not provide evidence that they should be further subdivided (Supplement Fig. A14). Here, future availability of genome scale data might help to clarify the species status. For the unnamed entities of *Pachypus* (*P. sp1-6*), older available species names currently ranked as synonyms with *P. candidae* (Bezdek 2016) have not yet been considered. This requires a more detailed investigation of the type specimen material, including lectotype or possibly neotype designations, which is challenging due to multiple type localities, as well as yet non-located or lost type material and imprecise geographic collection label data. A thorough taxonomic treatment of *Pachypus* species will therefore be published in a forthcoming paper.

Concluding Remarks

Recent methodological developments in species delimitation methods have resulted in a revival on the subject of how a species should be defined and our general understanding of the speciation process (e.g., Freudenstein et al. 2017). Species concepts are often linked to the underlying mechanisms of speciation but these processes are not straightforward to identify. Instead, we aim to recognize the ultimate outcome of the speciation process: genetic isolation of independently evolving metapopulations (Rannala 2015). Our results show that this is extremely challenging in the case of *Pachypus* using *de novo* approaches alone. “Statistical democracy” (Carstens et al. 2013; Rannala 2015), which places confidence in delimitations recovered by the majority of employed methods, was problematic since no consensus could be found. The datasets and species delimitation approaches employed in this study resulted in substantial over-splitting. Furthermore, population divergence simulations using parameters consistent with the biology of *Pachypus* supported extreme philopatry as the principle cause, at least in the case of mtDNA

data. However, careful parameter choice and the integration of different datasets were shown to be valuable in this difficult case study.

Despite issues associated with the definition of species and the application of the multispecies coalescent (Sukumaran and Knowles 2017), the positive outcome of our analysis in iBPP demonstrates the benefits of this integrative modelling framework. The ability to simultaneously analyze molecular and morphometric data and allowing the user to test explicit species hypotheses, has the potential to overcome problems created by small or variable N_e, sex-biased dispersal, geographical isolation, or rarity.

The final outcome regarding species boundaries in *Pachypus* was influenced by the choice of data. In particular, the exaggerated intraspecific divergence observed for mtDNA violates the assumptions of applied delimitation methods and consequently, the more robust signal recovered using morphometric and nuclear data is likely to better reflect the true species boundaries. A conservative approach to species delimitation may consider the *a priori* exclusion of mtDNA or any loci associated with the least dispersing sex from species delimitation analysis. The availability of large-scale genomic data, failure of nuclear DNA to recover species boundaries on the basis of low mutation rate or too few sampled loci may become negligible, but other problems associated with the nature of species will remain for DNA-based species delimitation. Further development of species delimitation methods should allow for the integration of multiple independent sources of data and for more complex models of species' evolution that better reflect the true nature of species. Among important phenomena to consider will be protracted speciation, rareness, and – as shown in the present study – philopatry, to ensure meaningful utility of these methods in future species delimitation studies.

FUNDING

The study was in part supported by the German Science Association (DFG; AH/175/3 and AH/175/6).

ACKNOWLEDGEMENTS

We are grateful to Liliana M. Dávalos for help with the population demographic simulations, to Alberto Ballerio and Marcello Romano for kind permission to use their specimen photographs. John McCormack and two anonymous referees helped to improve the manuscript with their valuable comments. Alexander Donath (ZFMK) kindly supported us in the use of the ZFMK computing cluster.

REFERENCES

Adams D.C., Otarola-Castillo E. 2013. geomorph: an R package for the collection and analysis of geometric morphometric shape data. Methods Ecol. Evol. 4:393–399.

Adams D.C., Rohlf F.J. 2000. Ecological character displacement in Plethodon: biomechanical differences found from a geometric morphometric study. Proc. Natl. Acad. Sci. U. S. A. 97:4106–4111.

Ahrens D., Fujisawa T., Krammer H.-J., Eberle J., Fabrizi S., Vogler A.P. 2016. Rarity and Incomplete Sampling in DNA-Based Species Delimitation. Syst. Biol. 65:478–494.

Ahrens D., Monaghan M.T., Vogler A.P. 2007. DNA-based taxonomy for associating adults and larvae in multi-species assemblages of chafers (Coleoptera: Scarabaeidae). Mol. Phylogenet. Evol. 44:436–449.

Ahrens D., Schwarzer J., Vogler A.P. 2014. The evolution of scarab beetles tracks the sequential rise of angiosperms and mammals. Proc. R. Soc. B Biol. Sci. 281:20141470.

Ahrens D., Vogler A.P. 2008. Towards the phylogeny of chafers (Sericini): analysis of alignment-variable sequences and the evolution of segment numbers in the antennal club. Mol. Phylogenet. Evol. 47:783–798.

Aktas C. 2015. haplotypes: Haplotype Inference and Statistical Analysis of Genetic Variation. Available from https://cran.r-project.org/package=haplotypes.

Andriollo T., Naciri Y., Ruedi M. 2015. Two mitochondrial barcodes for one biological species: The case of European Kuhl's pipistrelles (chiroptera). PLoS One. 10:1–18.

Andújar C., Arribas P., Ruiz C., Serrano J., Gómez-Zurita J. 2014. Integration of conflict into integrative taxonomy: fitting hybridization in species delimitation of Mesocarabus (Coleoptera: Carabidae). Mol. Ecol. 23:4344–4361.

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

Astrin J.J., Stüben P.E., Misof B., Wägele J.W., Gimnich F., Raupach M.J., Ahrens D. 2012. Exploring diversity in cryptorhynchine weevils (Coleoptera) using distance-, character- and tree-based species delineation. Mol. Phylogenet. Evol. 63:1–14.

Avise J.C., Walker D. 1998. Pleistocene phylogeographic effects on avian populations and the speciation process. Proc. R. Soc. B Biol. Sci. 265:457–463.

Avise J.C., Walker D., Johns G.C. 1998. Speciation durations and Pleistocene effects on vertebrate phylogeography. Proc. R. Soc. B Biol. Sci. 265:1707–1712.

Balke M., Hendrich L., Toussaint E.F.A., Zhou X., von Rintelen T., de Bruyn M. 2013. Suggestions for a molecular biodiversity assessment of South East Asian freshwater invertebrates. Lessons from the megadiverse beetles (Coleoptera). J. Limnol. 72:61–68.

Ballard J.W.O., Whitlock M.C. 2004. The incomplete natural history of mitochondria. Mol. Ecol. 13:729–744.

Ballerio A., Rey A., Uliana M., Rastelli M., Rastelli S., Romano M., Colacurcio L. 2014. Coleotteri Scarabeoidei d'Italia. Available from http://www.societaentomologicaitaliana.it/Coleotteri Scarabeoidea d'Italia 2014/index.htm (accessed Sept. 15, 2017).

Baraud J. 1985. *Tropinota (Epicometis) villiersi* nouvelle espèce du Moyen-orient (Coleoptera, Scarabaeoidea, Cetoniidae). Rev. 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.

Barley A.J., Brown J.M., Thomson R.C. 2017. Impact of Model Violations on the Inference of Species Boundaries Under the Multispecies Coalescent. Syst. Biol.:syx073.

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

Bezdek A. 2016. Pachypodini: 249. In: Löbl I., Smetana A., editors. Catalogue of Palaearctic Coleoptera. Volume 3. Scarabaeoidea – Scirtoidea – Dascilloidea – Buprestoidea – Byrrhoidea. Revised and Updated Edition (2nd Edition). Leiden: Brill.

Blankers T., Adams D.C., Wiens J.J. 2012. Ecological radiation with limited morphological diversification in salamanders. J. Evol. Biol. 25:634–646.

Bond J.E., Stockman A.K. 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. Syst. Biol. 57:628–646.

Bookstein F.L., Streissguth A.P., Sampson P.D., Connor P.D., Barr H.M. 2002. Corpus callosum shape and neuropsychological deficits in adult males with heavy fetal alcohol exposure. Neuroimage. 15:233–251.

Boukaert R., Heled J., Kühnert D., Vaughan T., Wu C.-H., Xie D., Suchard M.A., Rambaut A., Drummond A.J. 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLOS Comput. Biol. 10:1–6.

Britton T., Anderson C.L.L., Jacquet D., Lundqvist S., Bremer K. 2007. Estimating divergence times in large phylogenetic trees. Syst. Biol. 56:741–752.

Britton T., Oxelman B., Vinnersten A., Bremer K. 2002. Phylogenetic dating with confidence intervals using mean path lengths. Mol. Phylogenet. Evol. 24:58–65.

Burnaby T.P. 1966. Growth-invariant discriminant functions and generalized distances. Biometrics. 22:96–110.

Burnham K.P., Anderson D.R. 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. New York, NY: Springer.

Butcher B.A., Smith M.A., Sharkey M.J., Quicke D.L.J. 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., Morando M., Avila L.J., Sites J.W. 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 (N. Y). 66:2834–2849.

Cardoso A., Serrano A., Vogler A.P. 2009. Morphological and molecular variation in tiger beetles of the Cicindela hybrida complex: Is an "integrative taxonomy" possible? Mol. Ecol. 18:648–664.

Carstens B.C., Pelletier T.A., Reid N.M., Satler J.D. 2013. How to fail at species delimitation. Mol. Ecol. 22:4369–4383.

Castella V., Ruedi M., Excoffier L. 2001. Contrasted patterns of mitochondrial and nuclear structure among nursery colonies of the bat Myotis myotis. J. Evol. Biol. 14:708–720.

Chen W.-C. 2011. Overlapping Codon Model, Phylogenetic Clustering, and Alternative Partial Expectation Conditional Maximization Algorithm. Ph.D. Diss., Iowa Stat Univ.

Clement M., Posada D., Crandall K.A. 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9:1657–1659.

Crovetti A. 1969. Contributo alla conoscenza dei Coleotteri Scarabeidi I. Il genere *Pachypus* Serville (Coleoptera, Scarabaeidae, Pachypodinae). Boll. di Zool. Agrar. e di Bachic. 9:133–188.

Currat M., Ruedi M., Petit R.J., Excoffier L. 2008. The hidden side of invasions: Massive introgression by local genes. Evolution (N. Y). 62:1908–1920.

Dávalos L.M., Russell A.L. 2014. Sex-biased dispersal produces high error rates in mitochondrial distance-based and tree-based species delimitation. J. Mammal. 95:781–791.

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

Degnan J.H., Rosenberg N.A. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24:332–340.

Dellicour S., Flot J.-F. 2015. Delimiting Species-Poor Data Sets using Single Molecular Markers: A Study of Barcode Gaps, Haplowebs and GMYC. Syst. Biol. 64:900–908.

Dinno A. 2009. Implementing Horn's parallel analysis for principal component analysis and factor analysis. Stata J. 9:291–298.

Drummond A.J., Ho S.Y.W., Phillips M.J., Rambaut A. 2006. Relaxed Phylogenetics and Dating with Confidence. PLoS Biol. 4:e88.

Dupuis J.R., Roe A.D., Sperling F.A.H. 2012. Multi-locus species delimitation in closely related animals and fungi: One marker is not enough. Mol. Ecol. 21:4422–4436.

Eberle J., Warnock R.C.M.R.C.M., Ahrens D. 2016. Bayesian species delimitation in *Pleophylla* chafers (Coleoptera) – the importance of prior choice and morphology. BMC Evol. Biol. 16:94.

Edgar R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Ence D.D., Carstens B.C. 2011. SpedeSTEM: a rapid and accurate method for species delimitation. Mol. Ecol. Resour. 11:473–480.

Esselstyn J.A., Evans B.J., Sedlock J.L., Anwarali Khan F. a., Heaney L.R. 2012. Single-locus species delimitation: a test of the mixed Yule-coalescent model, with an empirical application to Philippine round-leaf bats. Proc. R. Soc. B Biol. Sci. 279:3678–3686.

Etienne R.S., Morlon H., Lambert A. 2014. Estimating the duration of speciation from phylogenies. Evolution (N. Y). 68:2430–2440.

Etienne R.S., Rosindell J. 2012. Prolonging the past counteracts the pull of the present: Protracted speciation can explain observed slowdowns in diversification. Syst. Biol. 61:204–213.

Ezard T., Fujisawa T., Barraclough T. 2014. splits: SPecies' LImits by Threshold Statistics. Available from https://r-forge.r-project.org/projects/splits/. R package version 1.0-19/r51.

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

Fraley C., Raftery A.E. 2002. Model-Based Clustering, Discriminant Analysis, and Density Estimation. J. Am. Stat. Assoc. 97:611–631.

Fraley C., Raftery A.E., Murphy T.B., Scrucca L., Brendan M., Scrucca L. 2012. mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. Tech. Rep. 597, Univ. Washingt.:1–50.

Freudenstein J. V., Broe M.B., Folk R.A., Sinn B.T. 2017. Biodiversity and the Species Concept—Lineages are not Enough. Syst. Biol. 66:644–656.

Fujisawa T., Aswad A., Barraclough T.G. 2016. A Rapid and Scalable Method for Multilocus Species Delimitation Using Bayesian Model Comparison and Rooted Triplets. Syst. Biol. 65:759–771.

Fujisawa T., Barraclough T.G. 2013. Delimiting Species Using Single-Locus Data and the Generalized Mixed Yule Coalescent Approach: A Revised Method and Evaluation on Simulated Data Sets. Syst. Biol. 62:707–724.

Garnier S. 2018. viridisLite: Default Color Maps from “matplotlib” (Lite Version). .

Gewin V. 2002. All living things, online. Nature. 418:362–363.

Godfray H.C.J. 2002. Challenges for taxonomy. Nature. 417:17–19.

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

Gunz P., Mitteroecker P. 2013. Semilandmarks: A method for quantifying curves and surfaces. Hystrix. 24:103–109.

Harrington R.C., Near T.J. 2012. Phylogenetic and coalescent strategies of species delimitation in snubnose darters (Percidae: Etheostoma). Syst. Biol. 61:63–79.

Hart M.W., Sunday J. 2007. Things fall apart: biological species form unconnected parsimony networks. Biol. Lett. 3:509–512.

Hebert P., Ratnasingham S., DeWaard J.R. 2003a. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc. R. Soc. London. 270:96–99.

Hebert P.D.N., Cywinska A., Ball S.L., DeWaard J.R. 2003b. Biological identifications through DNA barcodes. Proc. Biol. Sci. 270:313–321.

Hebert P.D.N., Gregory T.R. 2005. The promise of DNA barcoding for taxonomy. Syst. Biol. 54:852.

Hey J. 2009. On the arbitrary identification of real species. In: Butlin R.K., Bridle J.R., Schluter D., editors. Speciation and Patterns of Diversity. Cambridge, UK: Cambridge University Press, Cambridge, UK. p. 15–28.

Ho S.Y.W., Phillips M.J., Drummond A.J., Cooper A. 2005. Accuracy of Rate Estimation Using Relaxed-Clock Models with a Critical Focus on the Early Metazoan Radiation. Mol. Biol. Evol. 22:1355–1363.

Horn J.L. 1965. A rationale and test for the number of factors in factor analysis. Psychometrika. 30:179–185.

Huang J.-P., Knowles L.L. 2016. The Species versus Subspecies Conundrum: Quantitative Delimitation from Integrating Multiple Data Types within a Single Bayesian Approach in Hercules Beetles. Syst. Biol. 65:685–699.

Hudson R.R. 2002. Generating samples under a Wright–Fisher neutral model of genetic variation. Bioinforma. Appl. Note. 18:337–338.

Jackson N.D., Carstens B.C., Morales A.E., O'Meara B.C. 2017. Species Delimitation with Gene Flow. Syst. Biol. 66:799–812.

Jolicoeur P. 1963. The multivariate generalization of the allometry equation. Biometrics. 19:497–499.

Jones G., Aydin Z., Oxelman B. 2015. DISSECT: an assignment-free Bayesian discovery method for species delimitation under the multispecies coalescent. Bioinformatics. 31:991–998.

Jones G.R. 2015. STACEY: species delimitation and phylogeny estimation under the multispecies coalescent. bioRxiv.

Katoh K., Kuma K., Toh H., Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33:511–518.

Katoh K., Misawa K., Kuma K., Miyata T. 2002. MAFFT: a novel method for rapid multiple

sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066.

Katoh K., Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief. Bioinform. 9:286–298.

Katoh K., Toh H. 2010. Parallelization of the MAFFT multiple sequence alignment program. Bioinformatics. 26:1899–1900.

Katouzian A.-R., Sari A., Macher J.N., Weiss M., Saboori A., Leese F., Weigand A.M. 2016. Drastic underestimation of amphipod biodiversity in the endangered Irano-Anatolian and Caucasus biodiversity hotspots. Sci. Rep. 6:22507.

Kerth G., Mayer F., Petit E. 2002. Extreme sex-biased dispersal in the communally breeding, nonmigratory Bechstein's bat (Myotis bechsteinii). Mol. Ecol. 11:1491–1498.

Lambert A., Morlon H., Etienne R.S. 2015. The reconstructed tree in the lineage-based model of protracted speciation. J. Math. Biol. 70:367–397.

Leache A.D., Fujita M.K. 2010. Bayesian species delimitation in West African forest geckos (Hemidactylus fasciatus). Proc. R. Soc. B Biol. Sci. 277:3071–3077.

Lim G.S., Balke M., Meier R. 2012. Determining species boundaries in a world full of rarity: singletons, species delimitation methods. Syst. Biol. 61:165–169.

Liu K., Raghavan S., Nelesen S., Linder C.R., Warnow T. 2009. Rapid and Accurate Large-Scale Coestimation of Sequence Alignments and Phylogenetic Trees. Science (80-.). 324:1561–1564.

Liu K., Warnow T.J., Holder M.T., Nelesen S.M., Yu J., Stamatakis A.P., Linder C.R. 2012. SATe-II: Very Fast and Accurate Simultaneous Estimation of Multiple Sequence

Alignments and Phylogenetic Trees. Syst. Biol. 61:90–106.

Liu Y., Mittler J.E. 2008. Selection dramatically reduces effective population size in HIV-1 infection. BMC Evol. Biol. 8:133.

Lohse K. 2009. Can mtDNA Barcodes Be Used to Delimit Species? A Response to Pons et al. (2006). Syst. Biol. 58:439–442.

Luo A., Ling C., Ho S.Y.W., Zhu C. 2018. Comparison of Methods for Molecular Species Delimitation Across a Range of Speciation Scenarios. Syst. Biol. 0:1–13.

Lyrholm T., Leimar O., Johanneson B., Gyllensten U. 1999. Sex-biased dispersal in sperm whales: contrasting mitochondrial and nuclear genetic structure of global populations. Proc. R. Soc. - Biol. Sci. 266:347–354.

Mayr E. 1995. Species, classification, and evolution. In: Arai R., Kato M., Doi Y., editors. Biodiversity and Evolution. Tokyo: National Science Museum Foundation, Tokyo. p. 3–12.

Meier R., Shiyang K., Vaidya G., Ng P. 2006. DNA Barcoding and Taxonomy in Diptera: A Tale of High Intraspecific Variability and Low Identification Success. Syst. Biol. 55:715–728.

Melnick D., Hoelzer G. 1992. Differences in male and female macaque dispersal lead to contrasting distribution of nuclear and mitochondrial DNA variation. Int. J. Primatol. 13:379–393.

Meyer C.P., Paulay G. 2005. DNA barcoding: Error rates based on comprehensive sampling. PLoS Biol. 3:1–10.

Minh B.Q., Nguyen M.A.T., von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. 30:1188–1195.

Monaghan M.T., Balke M., Pons J., Vogler A.P. 2006. Beyond barcodes: complex DNA taxonomy of a South Pacific Island radiation. Proc. R. Soc. B Biol. Sci. 273:887.

Monaghan M.T., Inward D.J.G., Hunt T., Vogler A.P. 2007. A molecular phylogenetic analysis of the Scarabaeinae (dung beetles). Mol. Phylogenet. Evol. 45:674–692.

Monaghan M.T., Wild R., Elliot M., Fujisawa T., Balke M., Inward D.J.G., Lees D.C., Ranaivosolo R., Eggleton P., Barraclough T.G., Vogler A.P. 2009. Accelerated species inventory on Madagascar using coalescent-based models of species delineation. Syst. Biol. 58:298–311.

Nguyen L.T., Schmidt H.A., Von Haeseler A., Minh B.Q. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32:268–274.

Nosil P., Harmon L.J., Seehausen O. 2009. Ecological explanations for (incomplete) speciation. Trends Ecol. Evol. 24:145–156.

Nylander J.A.A. 2014. burntrees v. 0.2.2. Available from https://github.com/nylander/Burntrees/.

O'Meara B.C. 2010. New heuristic methods for joint species delimitation and species tree inference. Syst. Biol. 59:59–73.

Olave M., Solà E., Knowles L.L. 2014. Upstream Analyses Create Problems with DNA-Based Species Delimitation. Syst. Biol. 63:263–271.

Padial J.M., Castroviejo-Fisher S., Köhler J., Vilà C., Chaparro J.C., De la Riva I. 2009. Deciphering the products of evolution at the species level: the need for an integrative taxonomy. Zool. Scr. 38:431–447.

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

Palstra F.P., Fraser D.J. 2012. Effective/census population size ratio estimation: A compendium and appraisal. Ecol. Evol. 2:2357–2365.

Palumbi S.R., Baker C.S. 1994. Contrasting Population Structure from Nuclear Intron Sequences and mtDNA of Humpback Whales. Mol. Biol. Evol. 11:426–435.

Palumbi S.R., Cipriano F., Hare M.P. 2001. Predicting Nuclear Gene Coalescence From Mitochondrial Data: the Three-Times Rule. Evolution (N. Y). 55:859–868.

Papadopoulou A., Cardoso A., Gómez-Zurita J. 2013. Diversity and diversification of Eumolpinae (Coleoptera: Chrysomelidae) in New Caledonia. Zool. J. Linn. Soc. 168:473–495.

Paradis E., Claude J., Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 20:289–290.

Perez S.I., Bernal V., Gonzalez P.N. 2006. Differences between sliding semi-landmark methods in geometric morphometrics, with an application to human craniofacial and dental variation. J. Anat. 208:769–784.

Petit R.J., Excoffier L. 2009. Gene flow and species delimitation. Trends Ecol. Evol. 24:386–393.

Plummer M., Best N., Cowles K., Vines K. 2006. CODA: Convergence Diagnosis and Output Analysis for MCMC. R News. 6:7–11.

Pons J., Barraclough T.G., Gomez-Zurita J., Cardoso A., Duran D.P., Hazell S., Kamoun S., Sumlin W.D., Vogler A.P. 2006. Sequence-Based Species Delimitation for the DNA

Taxonomy of Undescribed Insects. Syst. Biol. 55:595–609.

Prugnolle F., de Meeus T. 2002. Inferring sex-biased dispersal from population genetic tools: a review. Heredity (Edinb). 88:161–165.

Puillandre N., Lambert A., Brouillet S., Achaz G. 2012a. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. Mol. Ecol. 21:1864–1877.

Puillandre N., Modica M. V., Zhang Y., Sirovich L., Boisselier M.C., Cruaud C., Holford M., Samadi S. 2012b. Large-scale species delimitation method for hyperdiverse groups. Mol. Ecol. 21:2671–2691.

Purvis A., Orme C.D.L., Toomey N.H., Pearson P.N. 2009. Temporal patterns in diversification rates. In: Butlin R., Schluter D., Bridle J., editors. Speciation and patterns of diversity. Cambridge, U.K: Cambridge Univ. Press. p. 278–300.

Rambaut A., Grass N.C. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Bioinformatics. 13:235–238.

Rambaut A., Suchard M.A., Xie D., Drummond A.J. 2014. Tracer v.1.6. Available from http://beast.bio.ed.ac.uk/Tracer.

Rannala B. 2009. The art and science of species delimitation. 61:846–853.

Rannala B. 2015. The art and science of species delimitation. Curr. Zool. 61:846–853.

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

Ratnasingham S., Hebert P.D.N. 2007. BOLD: The Barcode of Life Data System (http://www. barcodinglife. org). Mol. Ecol. Notes. 7:355–364.

Ratnasingham S., Hebert P.D.N. 2013. A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System. PLoS One. 8.

Reid N.M., Carstens B.C. 2012. Phylogenetic estimation error can decrease the accuracy of species delimitation: a Bayesian implementation of the general mixed Yule-coalescent model. BMC Evol. Biol. 12:196.

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

Rohlf F. 2005. tpsDig, digitize landmarks and outlines. .

Rosindell J., Cornell S.J., Hubbell S.P., Etienne R.S. 2010. Protracted speciation revitalizes the neutral theory of biodiversity. Ecol. Lett. 13:716–727.

Ross H.A., Murugan S., Li W.L.S. 2008. Testing the reliability of genetic methods of species identification via simulation. Syst. Biol. 57:216–230.

Rubinoff D., Cameron S., Will K. 2006. A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. J. Hered. 97:581–594.

Sarkar D., Andrews F. 2016. latticeExtra: Extra Graphical Utilities Based on Lattice. .

Schlick-Steiner B.C., Steiner F.M., Seifert B., Stauffer C., Christian E., Crozier R.H. 2010. Integrative Taxonomy: A Multisource Approach to Exploring Biodiversity. Annu. Rev. Entomol. 55:421–438.

Scrucca L., Raftery A.E. 2014. clustvarsel: A Package Implementing Variable Selection for Model-based Clustering in R. J. Stat. Softw.:pre-print.

Seberg O., Petersen G. 2009. How many loci does it take to DNA barcode a crocus? PLoS One.

4:e4598.

de Serres M. 1829. Géognosie des terrains tertiaires: ou, Tableau des principaux animaux invertébrés des terrains marins tertiaires, du midi de la France. Montpellier, Paris: Pomathio-Durville.

Simon C., Frati F., Beckenbach A., Crespi B., Liu H., Flook P. 1994. Evolution, Weighting, and Phylogenetic Utility of Mitochondrial Gene Sequences and a Compilation of Conserved Polymerase Chain Reaction Primers. Ann. Entomol. Soc. Am. 87:651–701.

Slatkin M. 2008. Linkage disequilibrium – understanding the evolutionary past and mapping the medical future. Nat. Rev. Genet. 9:477–485.

Solís-Lemus C., Knowles L.L., Ané C. 2015. Bayesian species delimitation combining multiple genes and traits in a unified framework. Evolution (N. Y). 69:492–507.

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.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30:1312–1333.

Sukumaran J., Knowles L.L. 2017. Multispecies coalescent delimits structure, not species. Proc. Natl. Acad. Sci. 114:1607–1612.

Talavera G., Dincă V., Vila R. 2013. Factors affecting species delimitations with the GMYC model: insights from a butterfly survey. Methods Ecol. Evol. 4:1101–1110.

Tang C.Q., Humphreys A.M., Fontaneto D., Barraclough T.G. 2014. Effects of phylogenetic

reconstruction method on the robustness of species delimitation using single-locus data. Methods Ecol. Evol. 5:1086–1094.

Templeton A.R. 2001. Using phylogeographic analyses of gene trees to test species status and processes. Mol. Ecol. 10:779–791.

Templeton A.R., Crandall K.A., Sing C.F. 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.

Timmermans M.J.T.N., Dodsworth S., Culverwell C.L., Bocak L., Ahrens D., Littlewood D.T.J., Pons J., Vogler A.P. 2010. Why barcode? High-throughput multiplex sequencing of mitochondrial genomes for molecular systematics. Nucleic Acids Res. 38:e197.

Wiens J.J., Penkrot T.A. 2002. Delimiting Species Using DNA and Morphological Variation and Discordant Species Limits in Spiny Lizards (Sceloporus). Syst. Biol. 51:69–91.

Wild A.L., Maddison D.R. 2008. Evaluating nuclear protein-coding genes for phylogenetic utility in beetles. Mol. Phylogenet. Evol. 48:877–891.

Will K.W., Mishler B.D., Wheeler Q.D. 2005. The perils of DNA barcoding and the need for integrative taxonomy. Syst. Biol. 54:844–51.

Wright S. 1946. Isolation by distance under diverse systems of mating. Genetics. 31:39–59.

Yang Z., Rannala B. 2010. Bayesian species delimitation using multilocus sequence data. Proc. Natl. Acad. Sci. U. S. A. 107:9264–9269.

Yang Z., Rannala B. 2014. Unguided Species Delimitation Using DNA Sequence Data from Multiple Loci. Mol. Biol. Evol. 31:3125–3135.

Yang Z., Rannala B. 2016. Species Identification by Bayesian Fingerprinting: A Powerful Alternative to DNA Barcoding. bioRxiv.:041608.

Yeates D.K., Seago A., Nelson L., Cameron S.L., Joseph L., Trueman J.W.H. 2011. Integrative taxonomy, or iterative taxonomy? Syst. Entomol. 36:209–217.

Yee T.W. 2015. Vector Generalized Linear and Additive Models: With an Implementation in R. New York, USA: Springer.

Yee T.W., Wild C.J. 1996. Vector Generalized Additive Models. J. R. Stat. Soc. Ser. B. 58:481–493.

Zhang C., Rannala B., Yang Z. 2014. Bayesian Species Delimitation Can Be Robust to Guide-Tree Inference Errors. Syst. Biol. 63:993–1004.

Zhang C., Zhang D.-X., Zhu T., Yang Z. 2011. Evaluation of a Bayesian Coalescent Method of Species Delimitation. Syst. Biol. 60:747–761.

Zhang J., Kapli P., Pavlidis P., Stamatakis A. 2013. A general species delimitation method with applications to phylogenetic placements. Bioinformatics. 29:2869–2876.

SUPPLEMENTARY MATERIAL

Supplementary material available from Dryad data repository at http://datadryad.org, http://dx.doi.org/10.5061/dryad.g97d432.

Trees and DNA data available from TreeBASE: http://dx.doi.org/[NNNN] and GenBank (accession numbers KY238355 – KY239022; see Supplementary Table A1).

FIGURE CAPTIONS

Figure 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 (bold black 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.

Figure 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. 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 (a zoomable version of the figure is available in the electronic pdf). 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. Locality numbers in tip labels refer to Supplementary Fig. A1 and Tab. A2. Colors/shades of the delimitation table facilitate recognition of delimitation boundaries but may be inconclusive. Insets: male and non-winged female specimen of *Pachypus caesus* (from Ballerio et al. 2014).

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

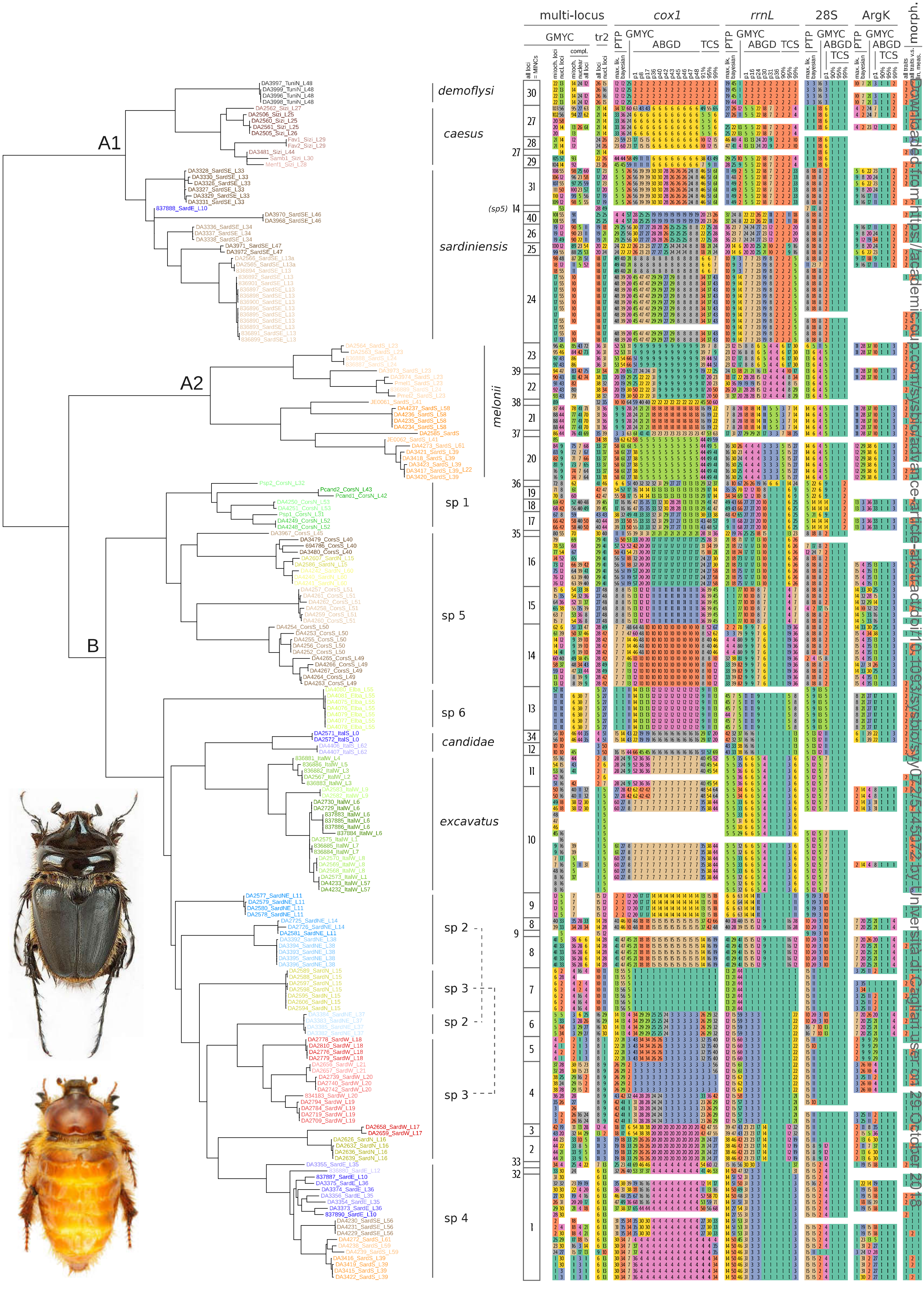
Figure 4. Ranges of species numbers inferred by iBPP under different combinations of τ_0 and θ priors. Horizontal bars indicate median species numbers within each bar. Two thresholds

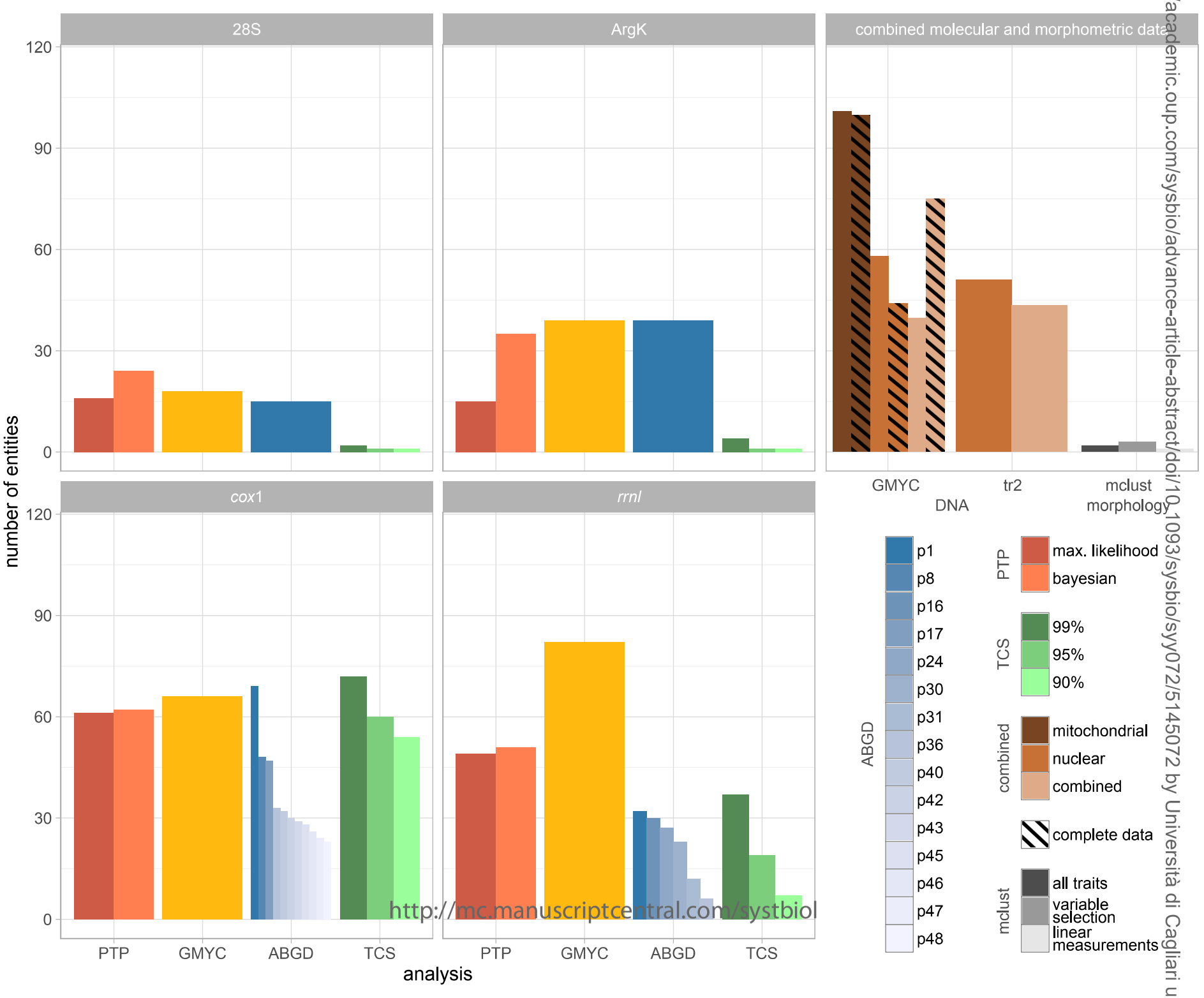
of posterior probabilities for 'good species' are compared for the unmodified guide tree topology and the modified guide tree topology informed by geography and nuclear DNA (Supplementary Fig. A7) for a range of data sets: AA = all loci, all traits; NoA = no DNA, all traits; NuA = nuclear DNA, all traits; AM = all loci, linear measurements; NoM = no DNA, linear measurements; NuM = nuclear loci, linear measurements; NuNo = nuclear loci, no traits; PS = prior sampling.

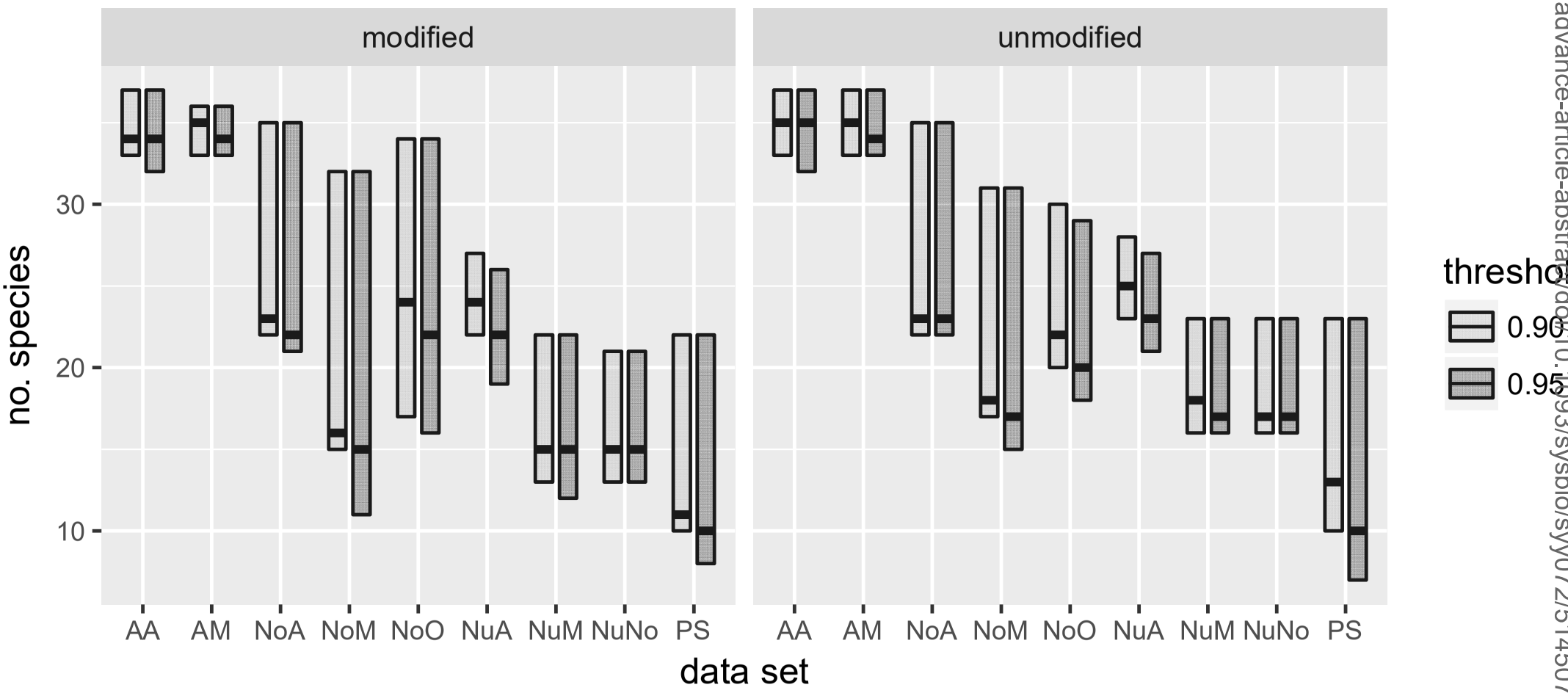
Figure 5. Selected results of the integrative Bayesian species delimitation with iBPP. Part 1 (a, c, e) and modified geography-informed part 2 (b, d, f) of the guide tree are shown for analyses with linear measurements and nuclear loci (a, b), with linear measurements alone (c, d), and with all morphometric traits and all loci (e, f). Posterior probabilities of 3×3 combinations of τ_0 and θ prior distributions are illustrated at each node (arrows point towards increased *a priori* splitting). Minimum clusters from the GMYC analyses (Fig. 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.

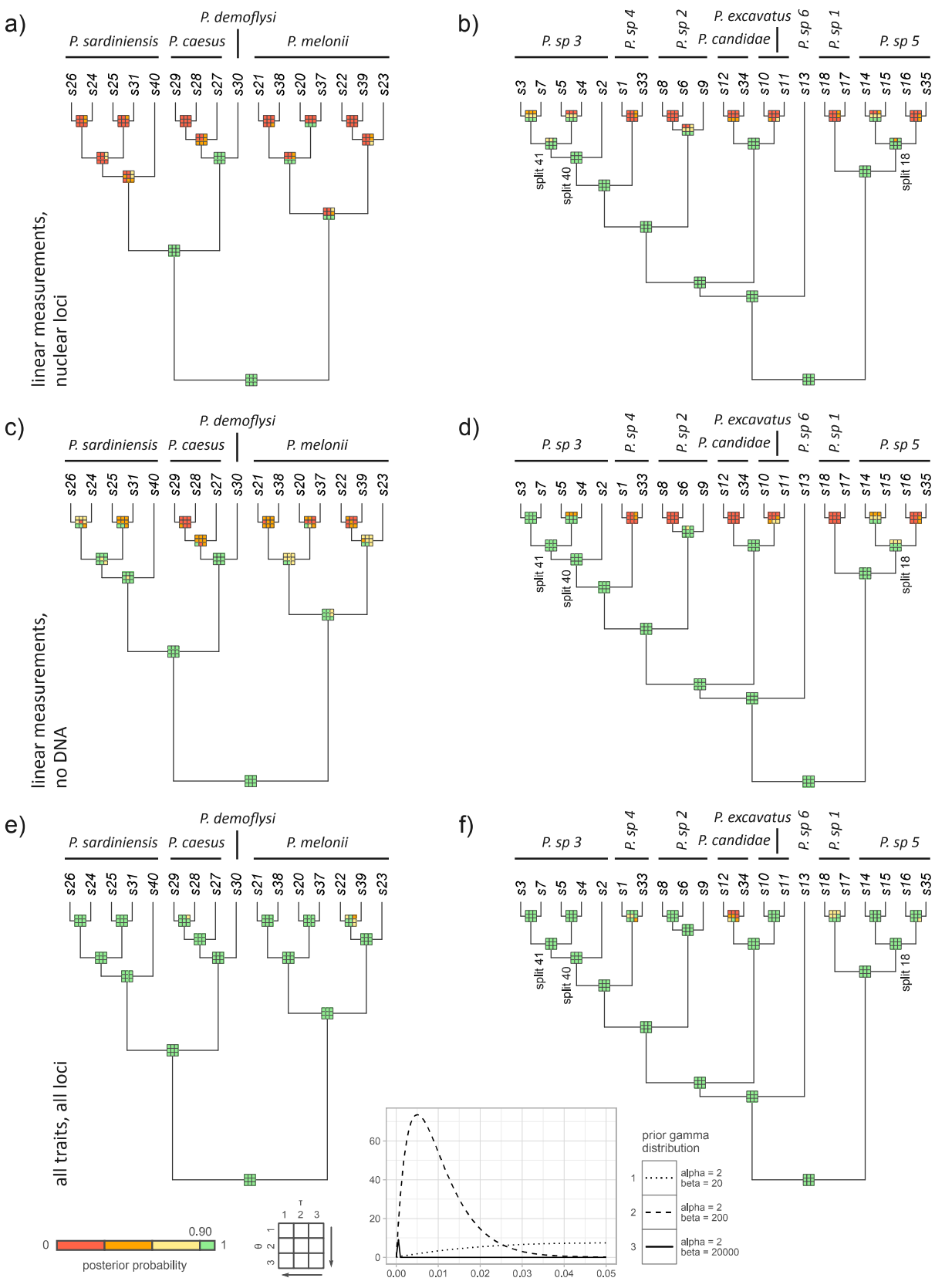
Figure 6. Simulated impact of sex-biased dispersal on nuclear autosomal and mitchondrial loci of two populations in the context of species delimitation. Panel headers indicate origin of locus, mutation rate, and migration rate between populations per generation. Panels a) – c) illustrate median raw genetic distance; panels d) – f) the percentage of reciprocal monophyly of populations in 100 simulation replicates. Generation time was set to two years (see Supplement Figures A27 – A28 for one year generation time).

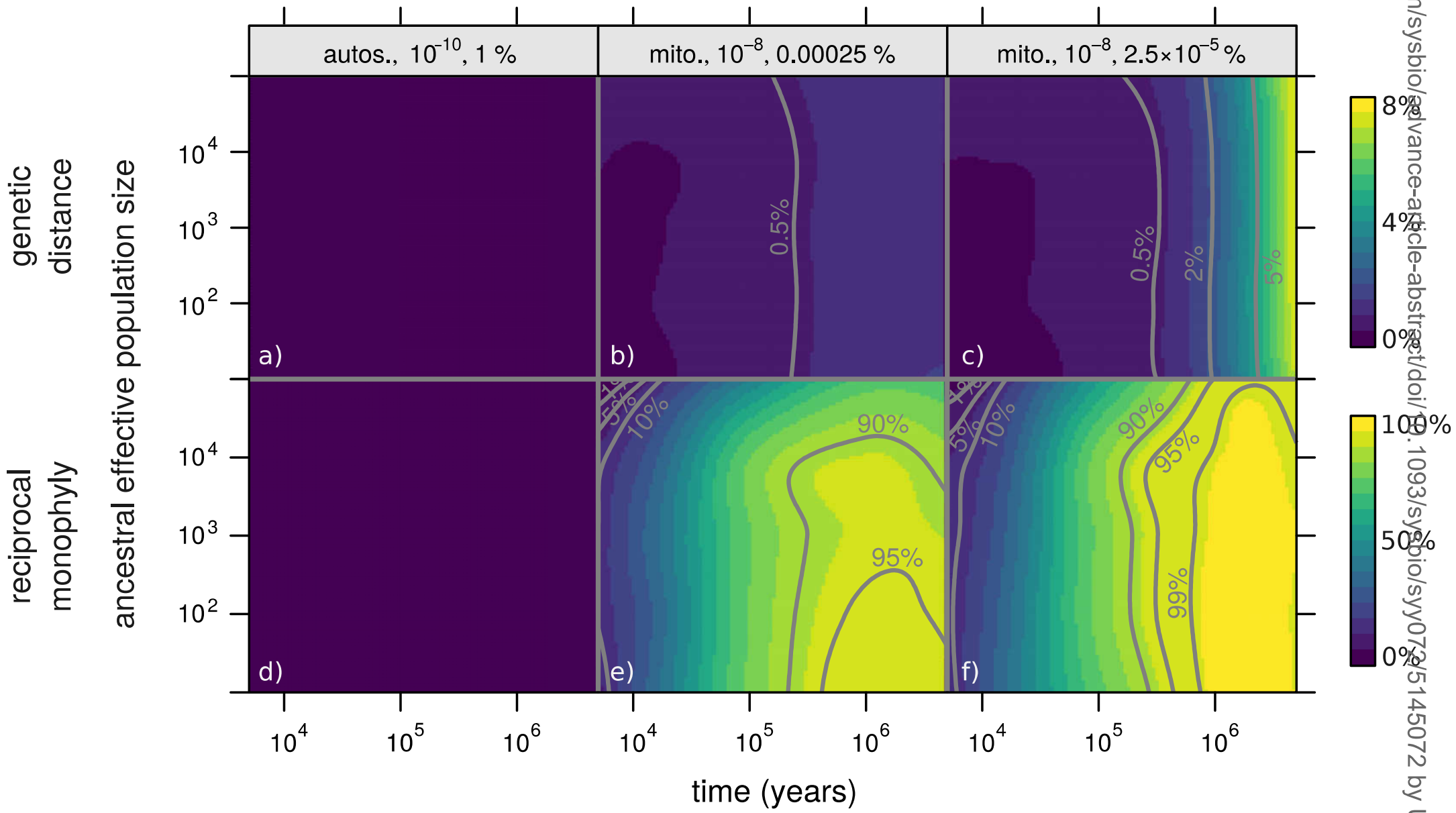
Figure 7. Map of known distributions of the inferred *Pachypus* species.










autos., 10^{-10}, 1 %
mito., 10^{-8}, 0.00025 %
mito., 10^{-8}, 2.5×10^{-5} %
genetic distance
reciprocal monophyly
ancestral effective population size
time (years)
a)
b)
c)
d)
e)
f)
0.5%
2%
5%
10%
90%
95%
99%
8%
4%
0%
100%
50%
0%