Phylogenetics and Molecular Evolution/Filogenética e Evolução Molecular

Octávio S. Paulo
Computational Biology and Population Genomics Group (CoBiG2)

Delimitação de espécies

Sumário:
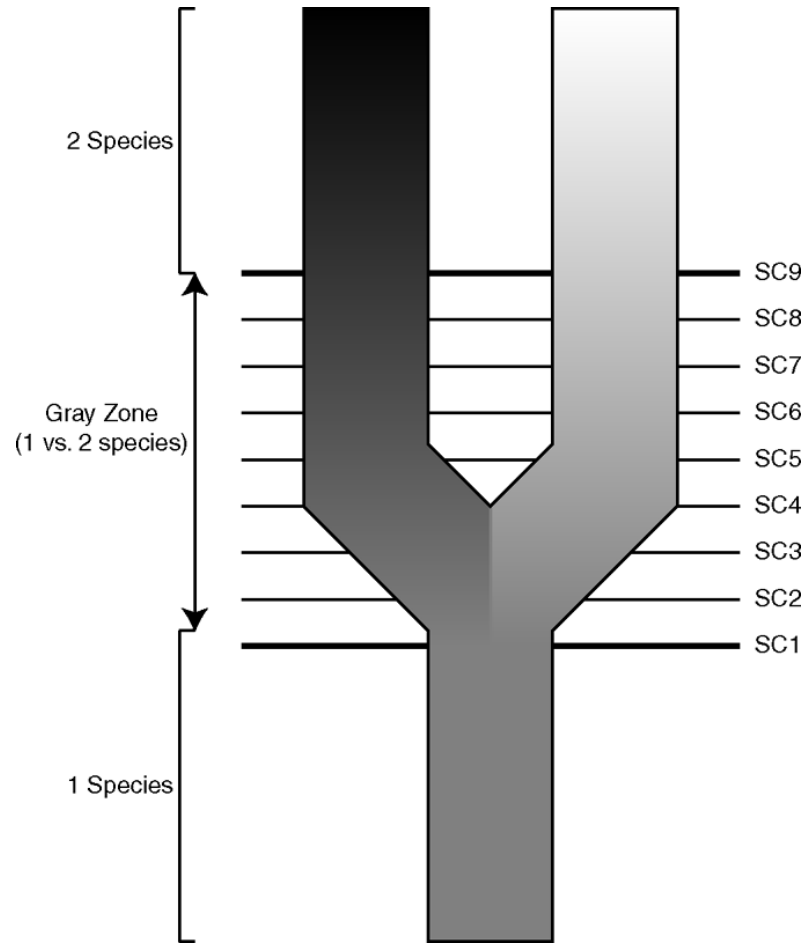Introdução à problemática da delimitação de espécies.

Figure 1 Lineage separation and divergence (speciation) and species concepts (after de Queiroz, 1998, 1999, 2005a). ...

de Queiroz 2007 SB

OXFORD
UNIVERSITY PRESS

Figure 1 Lineage separation and divergence (speciation) and species concepts (after de Queiroz, 1998, 1999, 2005a). ...

The general concept to which I refer equates species with separately evolving metapopulation lineages, or more specifically, with segments of such lineages. To clarify, here the term lineage refers to an ancestor-descendant series (Simpson, 1961; Hull, 1980) in this case of metapopulations or simply a metapopulation extended through time (cf. Simpson, 1951).

de Queiroz 2007 SB

OXFORD
UNIVERSITY PRESS
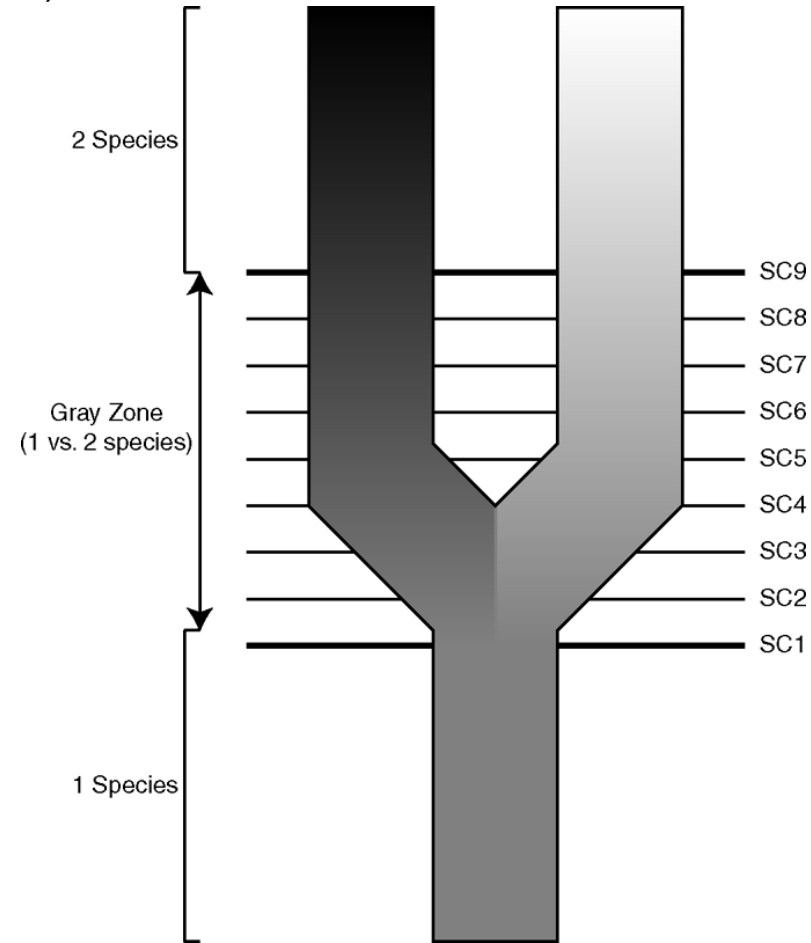
Species concepts

TABLE 1. Alternative contemporary species concepts (i.e., major classes of contemporary species definitions) and the properties upon which they are based (modified from de Queiroz, 2005). Properties (or the converses of properties) that represent thresholds crossed by diverging lineages and that are commonly viewed as necessary properties of species are marked with an asterisk (*). Note that under the proposal for unification described in this paper, the various ideas summarized in this table would no longer be considered distinct species concepts (see de Queiroz, 1998, for an alternative terminology). All of these ideas conform to a single general concept under which species are equated with separately evolving metapopulation lineages, and many of the properties (*) are more appropriately interpreted as operational criteria (lines of evidence) relevant to assessing lineage separation.

Species concept	Property(ies)	Advocates/references
Biological	Interbreeding (natural reproduction resulting in viable and fertile offspring)	Wright (1940); Mayr (1942); Dobzhansky (1950)
Isolation	*Intrinsic reproductive isolation (absence of interbreeding between heterospecific organisms based on intrinsic properties, as opposed to extrinsic [geographic] barriers)	Mayr (1942); Dobzhansky (1970)
Recognition	*Shared specific mate recognition or fertilization system (mechanisms by which conspecific organisms, or their gametes, recognize one another for mating and fertilization)	Paterson (1985); Masters et al. (1987); Lambert and Spencer (1995)
Ecological	*Same niche or adaptive zone (all components of the environment with which conspecific organisms interact)	Van Valen (1976); Andersson (1990)
Evolutionary	Unique evolutionary role, tendencies, and historical fate	Simpson (1951); Wiley (1978); Mayden (1997)
(some interpretations)	*Diagnosability (qualitative, fixed difference)	Grismer (1999, 2001)
Cohesion	Phenotypic cohesion (genetic or demographic exchangeability)	Templeton (1989, 1998a)
Phylogenetic	Heterogeneous (see next four entries)	(see next four entries)
Hennigian	Ancestor becomes extinct when lineage splits	Hennig (1966); Ridley (1989); Meier and Willmann (2000)
Monophyletic	*Monophyly (consisting of an ancestor and all of its descendants; commonly inferred from possession of shared derived character states)	Rosen (1979); Donoghue (1985); Mishler (1985)
Genealogical	*Exclusive coalescence of alleles (all alleles of a given gene are descended from a common ancestral allele not shared with those of other species)	Baum and Shaw (1995); see also Avise and Ball (1990)
Diagnosable	*Diagnosability (qualitative, fixed difference)	Nelson and Platnick (1981); Cracraft (1983); Nixon and Wheeler (1990)
Phenetic	*Form a phenetic cluster (quantitative difference)	Michener (1970); Sokal and Crovello (1970); Sneath and Sokal (1973)
Genotypic cluster (definition)	*Form a genotypic cluster (deficits of genetic intermediates; e.g., heterozygotes)	Mallet (1995)

Species concepts

Table 1 Summary of principal concepts of the species category, including the concepts of ESU and subspecies

Concept	Definition
Biological species concept (BSC)	"A biological species is an inclusive Mendelian population; it is integrated by the bonds of sexual reproduction and parentage." (Dobzhansky 1970, p. 354)
	"A species is a group of interbreeding natural groups that is reproductively isolated from other such groups." (Mayr and Ashlock 1991, p. 26)
Cohesion species concept (CSC)	"...the most inclusive population of individuals having the potential for phenotypic cohesion through intrinsic cohesion mechanisms." (Templeton 1989, p. 12)
	"…the most inclusive group of organisms having the potential for genetic and/or demographic exchangeability." (Templeton 1989, p. 25)
Differential species concept (DFC)	"Groups of individuals that are reciprocally characterized by features that would have negative fitness effects in other groups and that cannot be regularly exchanged between groups upon contact." (Hausdorf 2011, p. 927)
Evolutionary species concept (ESC)	"A species is a single lineage of ancestral-descendent populations which maintains its identity from other such lineages and which has its own evolutionary tendencies and historical fate." (Wiley 1978, p. 18)
	"...a lineage, comprised of organisms, whose history of individuation has manifested in its unique evolutionary trajectory through space and time." (Cotterill 2005, p. 115)
Genetic species concept (GSC)	"...population subdivisions concordantly identified by multiple independent genetic traits should constitute the population units worthy of recognition as phylogenetic taxa." (Avise and Ball 1990, p. 52)

Species concepts

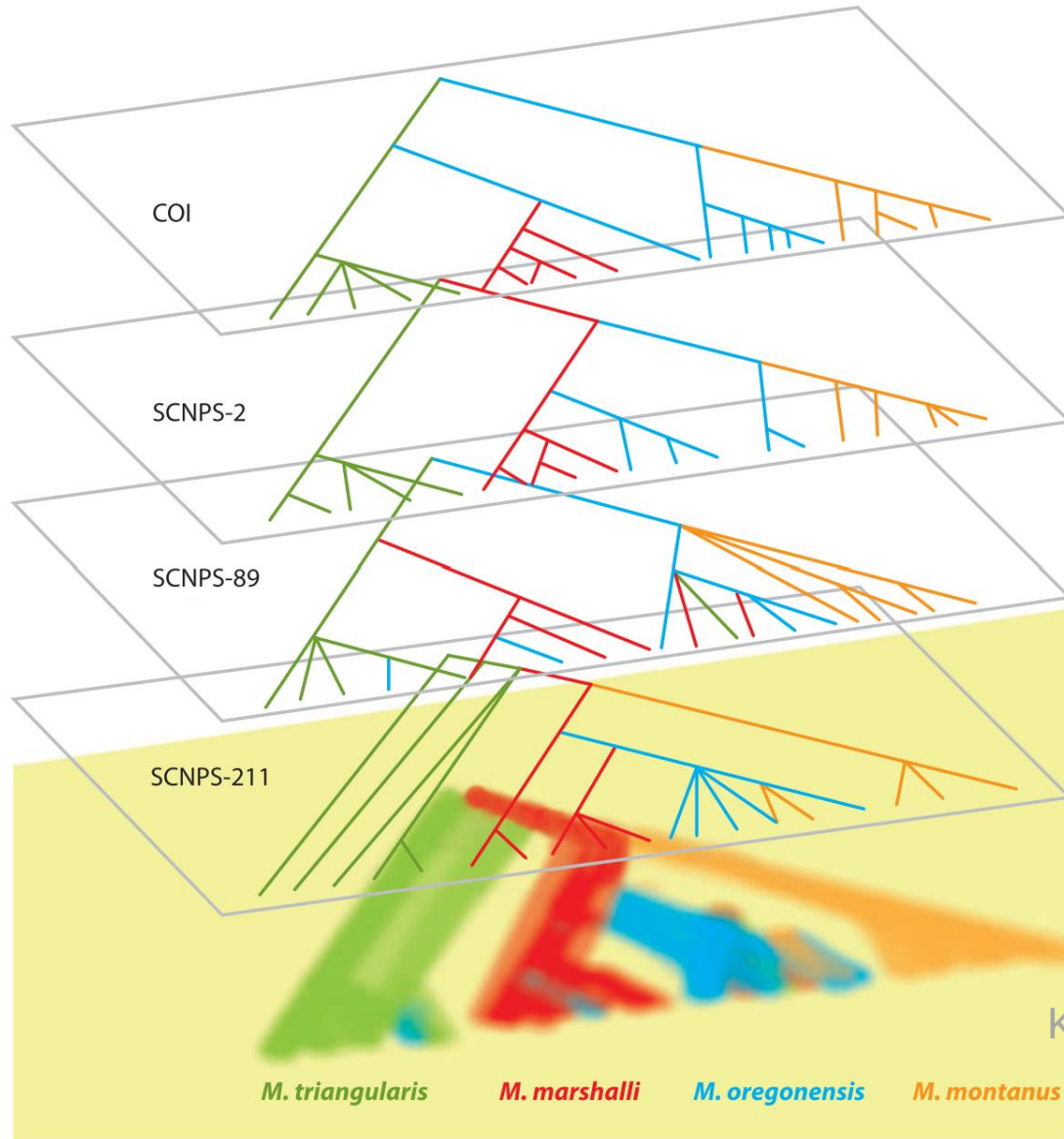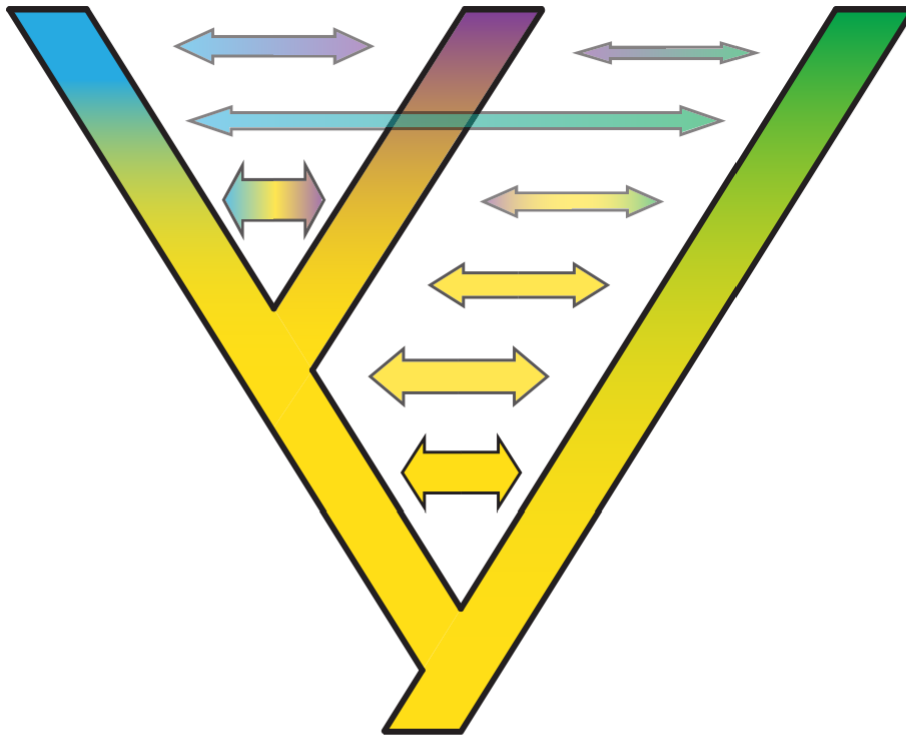|	
Phylogenetic species concept diagnosable version (PSC1)	"The smallest diagnosable cluster of individual organisms within which there is a parental pattern of ancestry and descent." (Cracraft 1983, p. 170)
Phylogenetic species concept monophyly version (PSC2)	"...a geographically constrained group of individuals with some unique apomorphous character, is the unit of evolutionary significance" (Rosen 1978, p. 176). Equates with ISC (Internodal Species Concept) of Kornet (1993), which was termed the CSC by Brooks and McLennan (1999)
Phylogenetic species concept diagnosable/monophyly version (PSC3)	"...the smallest diagnosable cluster of individual organisms forming a monophyletic group within which there is a parental pattern of ancestry and descent" (McKitrick and Zink 1988, defined by Mayden 1997, p. 407). This version of the PSC was termed the PSC2 by Brooks and McLennan (1999)
Recognition species concept (RSC)	"...the most inclusive population of individual biparental organisms that share a common fertilization system." (Paterson 1985, p. 25)

subSpecies concepts

Table 1 (continued)

Concept	Definition
Superspecies	"A superspecies is a monophyletic group of closely related and largely or entirely allopatric species that are too distinct to be included in a single species or that demonstrate their reproductive isolation in a zone of contact." (Mayr and Ashlock 1991, p. 430)
Subspecies	"An aggregate of phenotypically similar populations of a species, inhabiting a geographic subdivision of the range of a species, and differing taxonomically from other populations of the species." (Mayr 1969, p. 41)
Evolutionary significant unit (ESU)	"...a population (or group of populations) that (1) is substantially reproductively isolated from other conspecific population units, and (2) represents an important component in the evolutionary legacy of species." (Waples 1991) "ESUs should be reciprocally monophyletic for mtDNA alleles and show significant divergence of allele frequencies at nuclear loci." (Moritz 1994, p. 373) "A lineage demonstrating highly restricted gene flow from other such lineages within the higher organizational level (lineage) of a species." (Fraser and Bernatchez 2001, p. 2742)

Species delimitation

COI

SCNPS-2

SCNPS-89

SCNPS-211

M. triangularis   *M. marshalli*   *M. oregonensis*   *M. montanus*

Knowles 2009 ARES

Bravo et. al 2019 PeerJ

Figure 6 **Gradual speciation, or isolation-with-migration.** After starting to split, gene flow between species decreases gradually. Such a gradual decrease in the extent of gene flow between species might present an especially useful extension of the standard multispecies coalescent model. Colors depict different gene pools and their gradual change along branches describes how species gradually differentiate despite the existence of migration over time. Thickness and color intensity of arrows show that gene flow becomes weaker as species gradually isolate. Full-size 🖼 DOI: 10.7717/peerj.6399/fig-6

The single-gene species delimitation

Poisson Tree Processes model (Zhang et al., 2013),

      Poisson tree processes (PTP), the original likelihood implementation, the Bayesian Poisson tree processes (bPTP) and the multirate Poisson tree processes (mPTP) (Kapli et al., 2017).
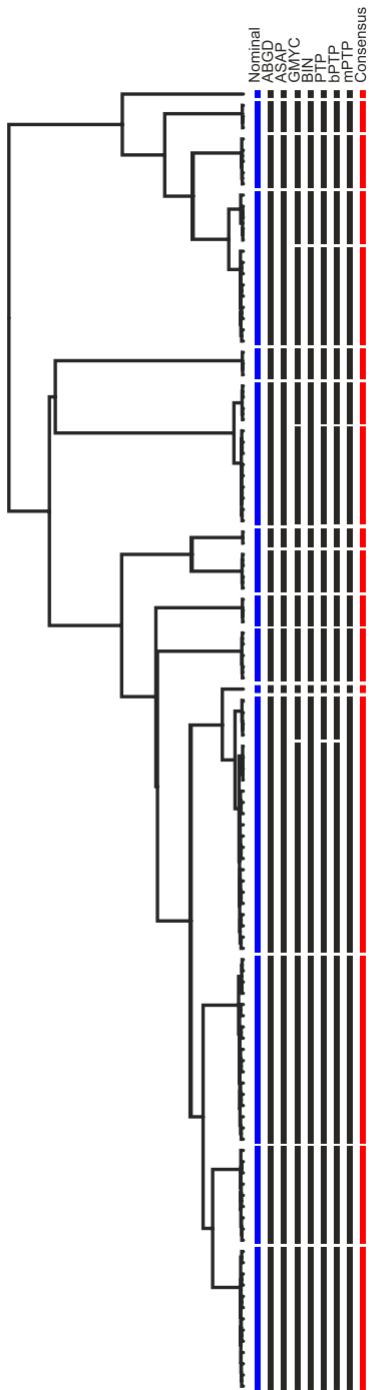
Generalized Mixed Yule Coalescent (GMYC) (Pons et al., 2006),

Automatic Barcode Gap Discovery (ABGD) (Puillandre et al., 2012)

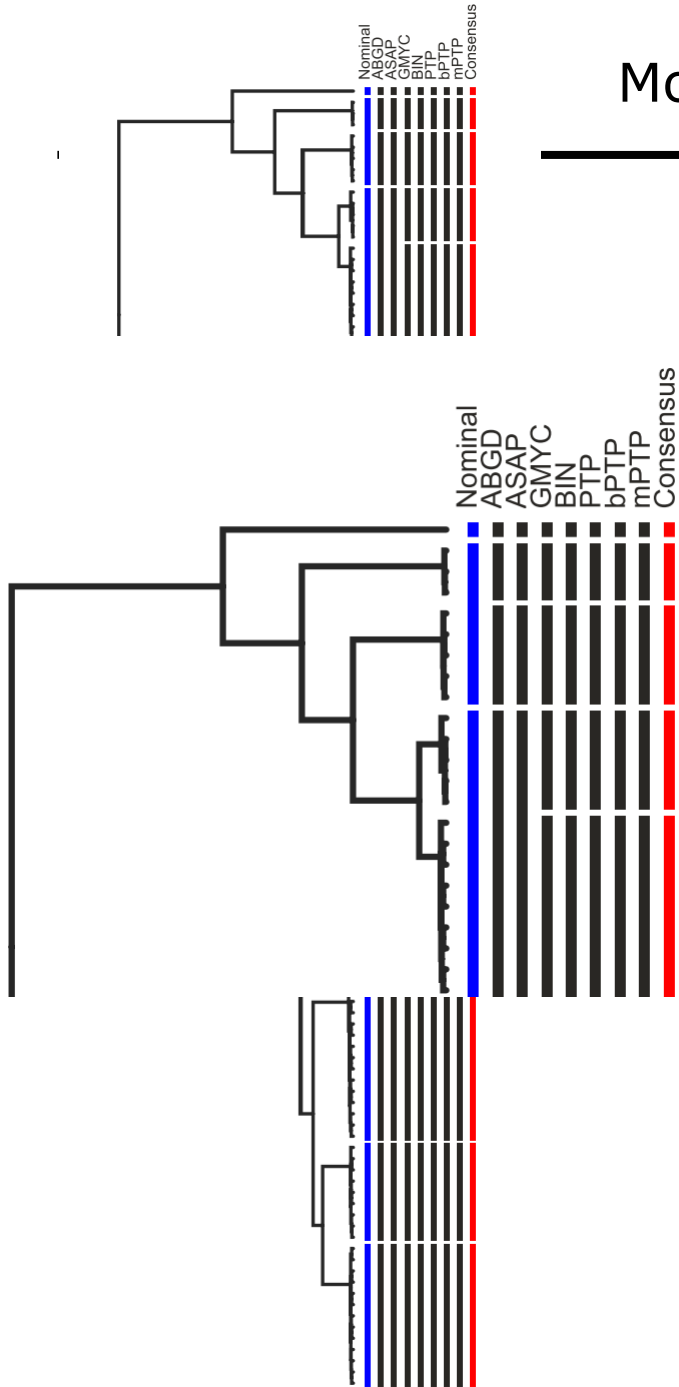Assemble Species by Automatic Partitioning (ASAP) (Puillandre et al., 2021).

Didelot et. al 2021 MBE

Molecular operational taxonomic units (MOTU)
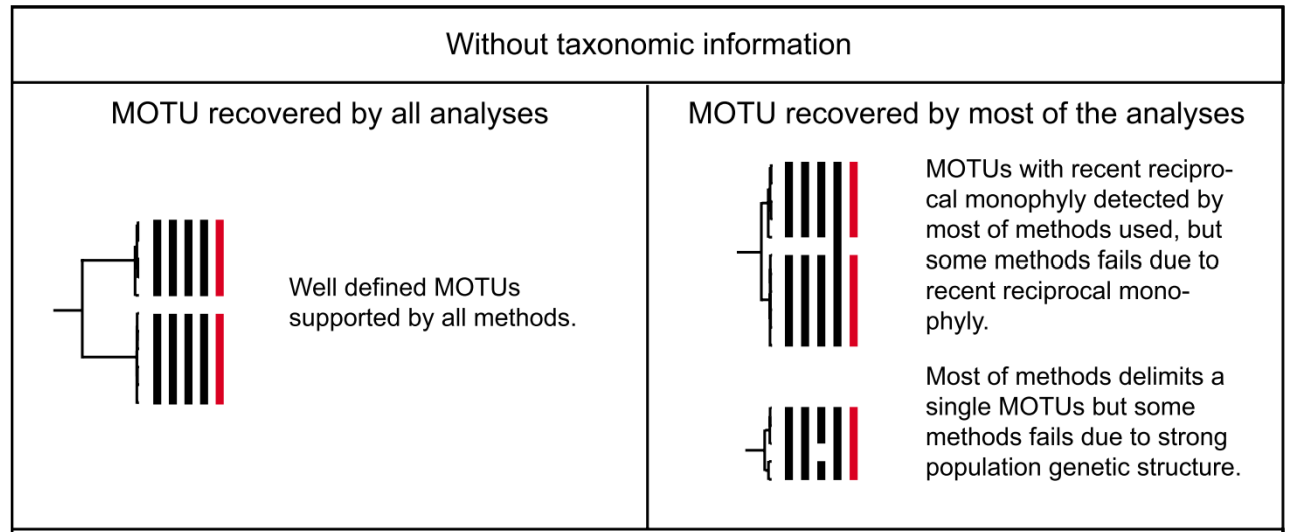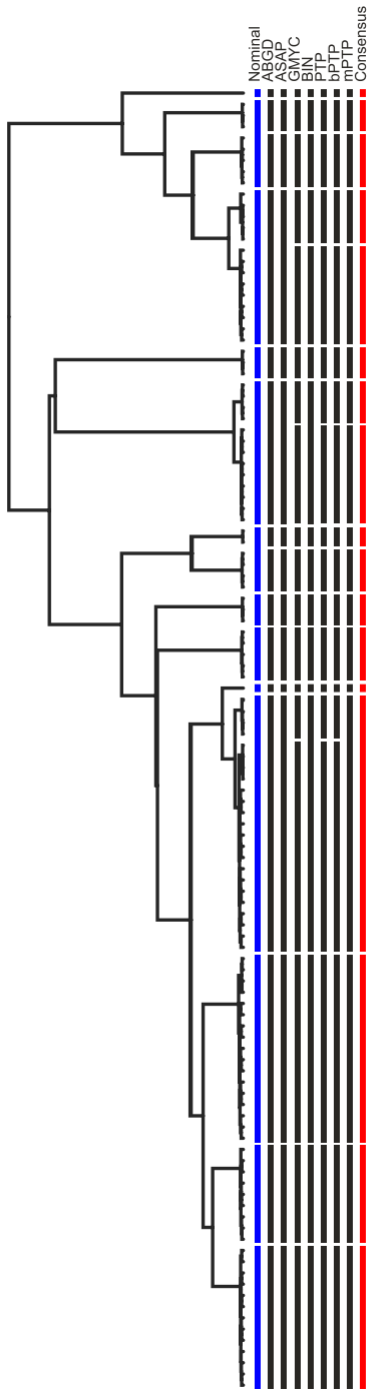
Ramirez et. al 2023 MER

Molecular operational taxonomic units (MOTU)

	Mean intra-group distance	Maximum intra-group distance	Nearest neighbour (NN)	Distance to NN
Megaleporinus brinco	0	0	M. obtusidens	6.78
M. conirostris	2.13	3.99	M. obtusidens	5.6
M. muyscorum	7.66	15.31	M. trifasciatus	4.51
M. elongatus	0.0371	0.16695	M. obtusidens	2.73593
M. garmani	0	0	M. obtusidens	7.68126
M. macrocephalus	0.903538	1.85854	M. trifasciatus	4.51779
M. obtusidens	1.937511	6.71724	M. elongatus	2.73593
M. piavussu	0.266286	1.00758	M. obtusidens	2.90372
M. reinhardti	0.316828	0.70177	M. conirostris	6.14484
M. trifasciatus	3.618149	6.33176	M. macrocephalus	4.51779

Ramirez et. al 2023 MER

Molecular operational taxonomic units (MOTU)

| Without taxonomic information	
MOTU recovered by all analyses	**MOTU recovered by most of the analyses**
Well defined MOTUs supported by all methods.	MOTUs with recent reciprocal monophyly detected by most of methods used, but some methods fails due to recent reciprocal monophyly. Most of methods delimits a single MOTUs but some methods fails due to strong population genetic structure.

Ramirez et. al 2023 MER
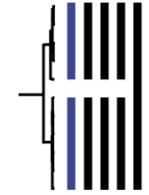
Species delimitation
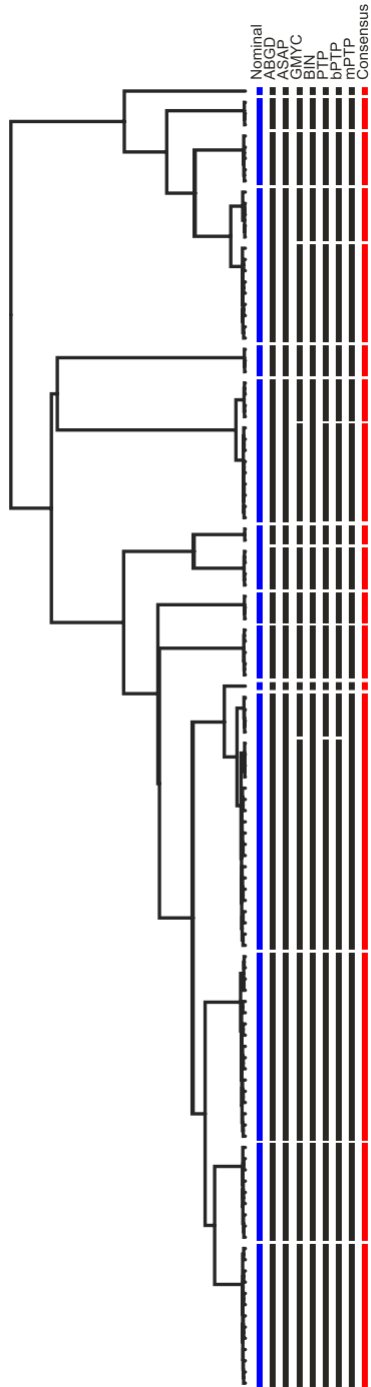
With taxonomic reference

MOTU totally matching the taxonomy
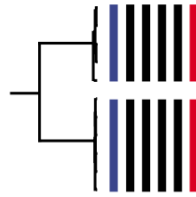
All analyses support taxonomic previous delimitation.
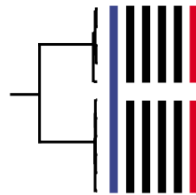
MOTU mostly matching the taxonomy

Most of methods delimits MOTUs supporting nominal species but some methods fails due to recent speciation.

Most of methods delimits a single MOTUs supporting nominal species but some methods fails due to strong population genetic structure.

MOTU totally mismatching the taxonomy

Potential criptic species, MOTUs well supported within a nominal species.

Synonymous species, hybridization or very recent speciation (with few time to reach monophyly).

MOTUs mostly mismatching the taxonomy
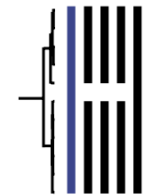
Potential criptic species with recent reciprocal monophyly detected by most of methods used or a single species with strong population genetic structure.
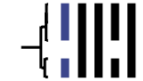
Very recent speciation not detected by most of methods used.

Multilocus species delimitation

Bayesian Phylogenetic and Phylogeography (BPP) (Yang and Rannala 2010 PNAS; Yang and Rannala 2014 MBE; Yang 2015 CZ; Flouri et al 2018 MBE)

STACEY (Jones 2017 JMB)

SNAPP (Bryant et al 2012 MBE)

Species Tree Inference with BPP Using Genomic Sequences and the *Multispecies Coalescent*

Tomáš Flouri,[1] Xiyun Jiao,[1] Bruce Rannala,*[2] and Ziheng Yang*[1]

[1]Department of Genetics, Evolution and Environment, University College London, London, United Kingdom
[2]Department of Ecology and Evolution, University of California, Davis, CA

*Corresponding authors: E-mails: brannala@ucdavis.edu; z.yang@ucl.ac.uk.
Associate editor: Anne D. Yoder

Abstract

The multispecies coalescent provides a natural framework for accommodating ancestral genetic polymorphism and coalescent processes that can cause different genomic regions to have different genealogical histories. The Bayesian program BPP includes a full-likelihood implementation of the multispecies coalescent, using transmodel Markov chain Monte Carlo to calculate the posterior probabilities of different species trees. BPP is suitable for analyzing multilocus sequence data sets and it accommodates the heterogeneity of gene trees (both the topology and branch lengths) among loci and gene tree uncertainties due to limited phylogenetic information at each locus. Here, we provide a practical guide to the use of BPP in species tree estimation. BPP is a command-line program that runs on LINUX, MACOSX, and WINDOWS. This protocol shows how to use both BPP 3.4 (http://abacus.gene.ucl.ac.uk/software/) and BPP 4.0 (https://github.com/bpp/).

Key words: BPP, MCMC, multispecies coalescent, species tree inference.

Fig. 1. A species tree for three species (A, B, and C) to illustrate the parameters of the MSC model, with a gene tree for five sequences (a_1 and a_2 sampled from species A, b_1 and b_2 from species B, and c_1 from species C) running inside the species tree. Within each species/population, sequences coalesce at random at the rate determined by the population size (or θ parameter), generating a gene tree with branch lengths (coalescent times), conditioned on the species tree. Note that θ_C is not estimable if there is at most one sequence from species C at each locus.

Flouri et. al 2018 MBE

Octávio S. Paulo – FEM

17

Mathematical Biology

CrossMark

Algorithmic improvements to species delimitation and phylogeny estimation under the multispecies coalescent

Graham Jones[1] iD

Abstract The focus of this article is a Bayesian method for inferring both species delimitations and species trees under the multispecies coalescent model using molecular sequences from multiple loci. The species delimitation requires no a priori assignment of individuals to species, and no guide tree. The method is implemented in a package called STACEY for BEAST2, and is a extension of the author's DISSECT package. Here we demonstrate considerable efficiency improvements by using three new operators for sampling from the posterior using the Markov chain Monte Carlo algorithm, and by using a model for the population size parameters along the branches of the species tree which allows these parameters to be integrated out. The correctness of the moves is demonstrated by tests of the implementation. The practice of using a pipeline approach to species delimitation under the multispecies coalescent, has been shown to have major problems on simulated data (Olave et al. in Syst Biol 63:263–271. doi:10.1093/sysbio/syt106, 2014). The same simulated data set is used to demonstrate the accuracy and improved convergence of the present method. We also compare performance with *BEAST for a fixed delimitation analysis on a large data set, and again show improved convergence.

In the context of species delimitation using STACEY, the species tree has tips which represent **minimal clusters** of individuals (Jones et al. 2014). These minimal clusters may be merged but not split to form potential species. At its most flexible, there is just one individual in each minimal cluster, so the possible number of species ranges from one to the number of individuals. Thus 'species tree' is not a good name for this tree, and instead we will refer to it a the **SMC-tree**, as a shorthand for 'species or minimal clusters tree'.
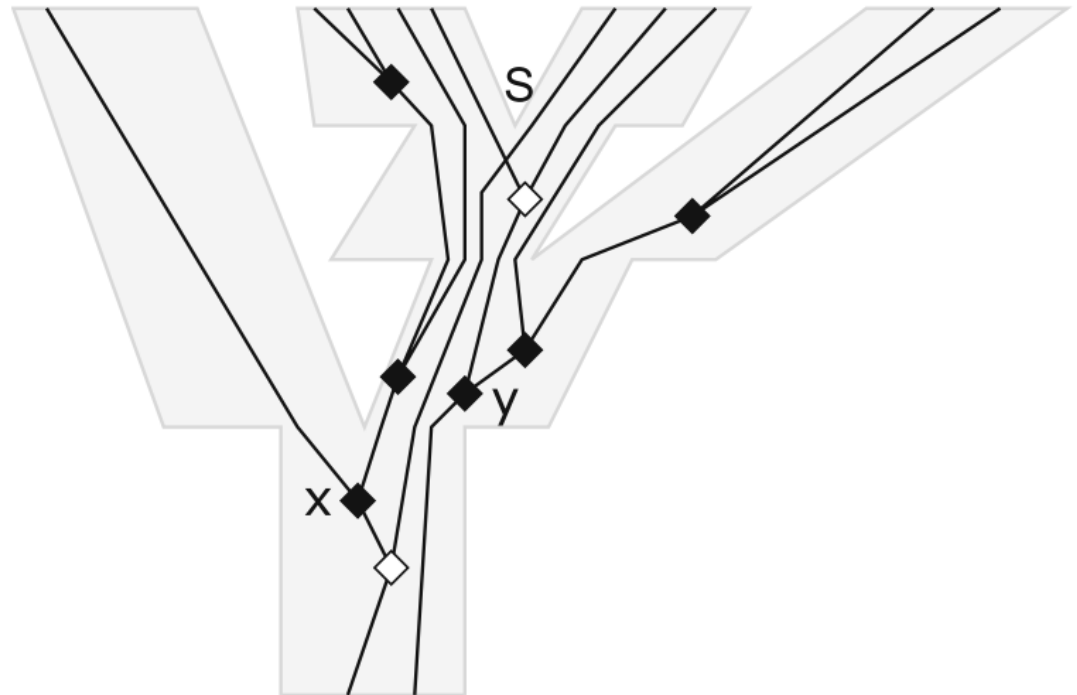
Fig. 1 Example of hitched nodes. The SMC-tree is *pale gray*. A gene tree is shown inside it. Gene tree nodes which are hitched to the SMC-tree node *S* are shown as *white diamonds*, and other nodes as *black diamonds*
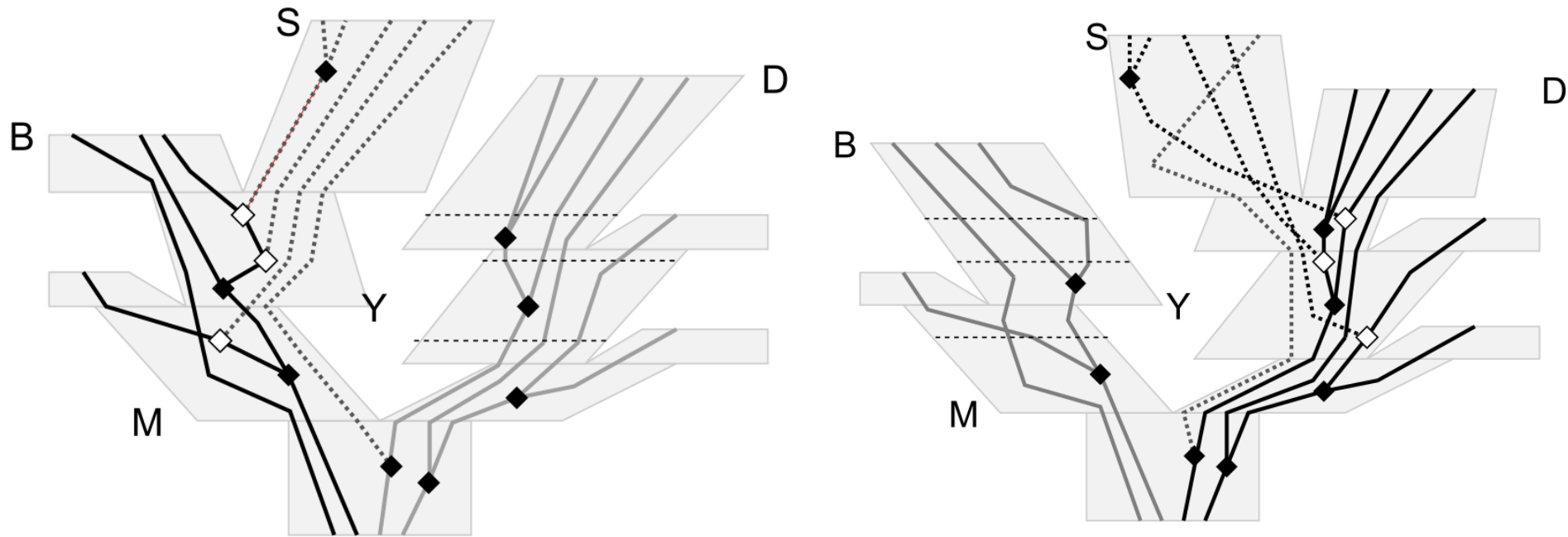
Fig. 4 Example of the CoordinatedPruneRegraft move. The state before the move is at the *top*, and after the move *below*. The SMC-tree is *pale gray*. Gene tree branches whose sequences all belong to $I(S)$ are *dotted*. Before the move, gene tree branches whose sequences are descendants of the same (*left*) child of M as S but do not belong entirely to $I(S)$ are *black*. The *white nodes* are the origins (anc(s) in the text) of subtrees that need to be pruned and regrafted, and the *thin horizontal dotted lines* cut across the available destination branches. After the move, the *colors* and styles are reversed to illustrate the reverse move

Inferring Species Trees Directly from Biallelic Genetic Markers: Bypassing Gene Trees in a Full Coalescent Analysis

David Bryant,*,[1] Remco Bouckaert,[2] Joseph Felsenstein,[3] Noah A. Rosenberg,[4] and Arindam RoyChoudhury[5]

Abstract

The multispecies coalescent provides an elegant theoretical framework for estimating species trees and species demographics from genetic markers. However, practical applications of the multispecies coalescent model are limited by the need to integrate or sample over all gene trees possible for each genetic marker. Here we describe a polynomial-time algorithm that computes the likelihood of a species tree directly from the markers under a finite-sites model of mutation effectively integrating over all possible gene trees. The method applies to independent (unlinked) biallelic markers such as well-spaced single nucleotide polymorphisms, and we have implemented it in SNAPP, a Markov chain Monte Carlo sampler for inferring species trees, divergence dates, and population sizes. We report results from simulation experiments and from an analysis of 1997 amplified fragment length polymorphism loci in 69 individuals sampled from six species of *Ourisia* (New Zealand native foxglove).

Didelot et. al 2021 MBE

Octávio S. Paulo – FEM

21

Inferring Species Trees Directly from Biallelic Genetic Markers: Bypassing Gene Trees in a Full Coalescent Analysis

David Bryant,*,[1] Remco Bouckaert,[2] Joseph Felsenstein,[3] Noah A. Rosenberg,[4] and Arindam RoyChoudhury[5]

Our approach makes the following assumptions of the data:

(A1) Each marker is a single biallelic character (e.g., a biallelic SNP or AFLP banding pattern);

(A2) The genealogies for separate markers are conditionally independent given the species tree. In practice, this assumption applies to unlinked markers or linked markers that have so little linkage that they do not possess a discernible excess of linkage disequilibrium.

Second, we assume that gene dynamics within populations are well described by the (neutral) Wright–Fisher model, approximated by a coalescent process. In some cases,

Third, we assume that there is no gene flow between populations. Incorporating gene flow will be difficult mainly because the dynamic programming algorithm used within SNAPP relies on a lack of gene flow between descendent populations. Here, approaches based on diffusion processes

We have presented a method that takes biallelic markers sampled from multiple individuals from multiple species and computes the likelihood of a species tree topology together with population genetic parameters. Our approach implements a full multispecies coalescent model without having to explicitly integrate or sample the gene trees at each locus. With our MCMC sampler, SNAPP, we can concentrate on the parameters of interest: the species tree, population sizes, and divergence times rather than on the problem of traversing through the space of potential gene trees. The likelihood values we compute are exact up to numerical error and do not require a simplification or approximation of the full coalescent model.

The model is based on the coalescent process and yet no explicit sampling of gene trees takes place.
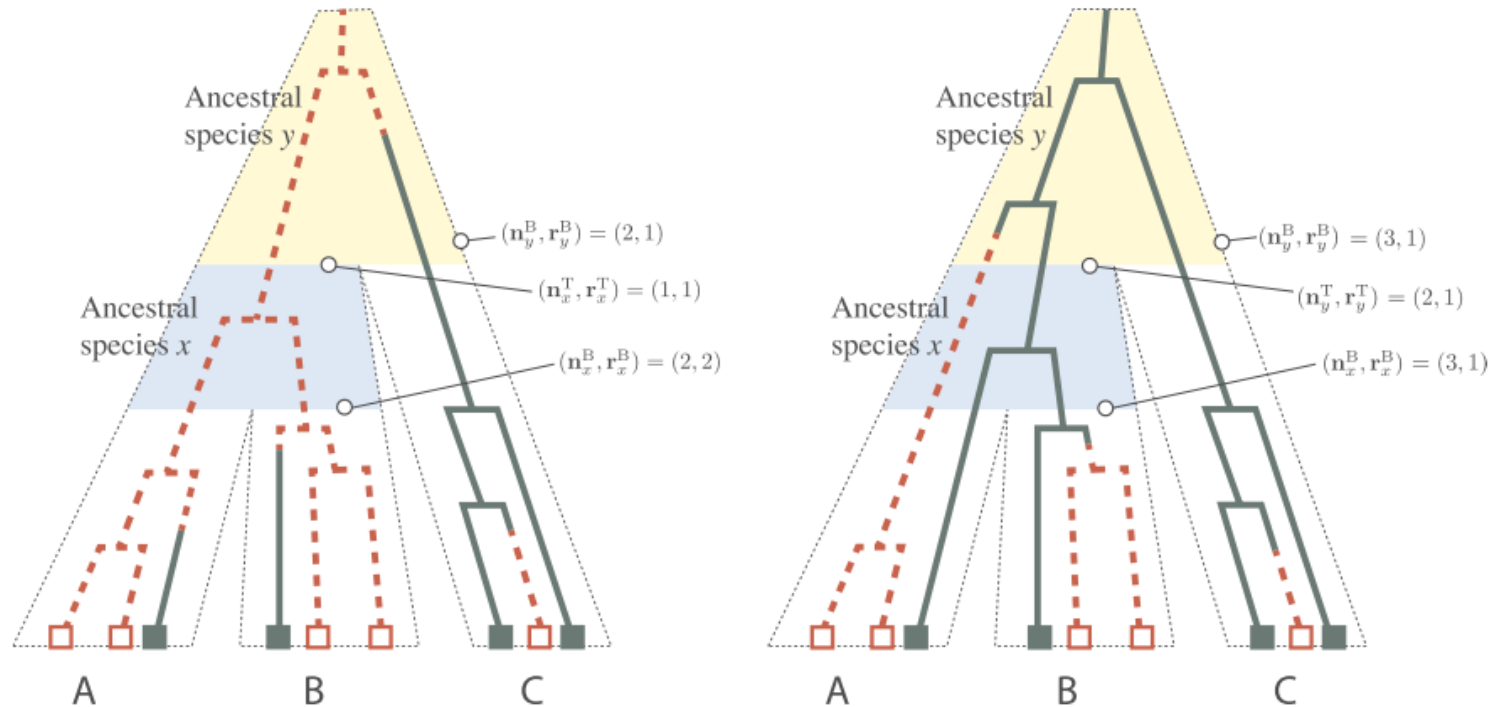
FIG. 1. Gene trees in species trees. Each branch in the species trees corresponds to a species that is either contemporary (A,B,C) or ancestral (x, y). The present-day samples are represented by green (solid) and red (hollow) squares along the lower edge of the tree. The red (dashed) and green (solid) lines trace out two possible gene trees for these individuals, the red–green coloring indicating which allele is carried by a lineage at any particular time. The random variables $\mathbf{n}_x^B$ and $\mathbf{r}_x^B$ equal the number of lineages and the number of red lineages, respectively, at the bottom of the branch for ancestral species x. The corresponding values at the top of this branch are denoted $\mathbf{n}_x^T$ and $\mathbf{r}_x^T$, respectively.

SNAPP
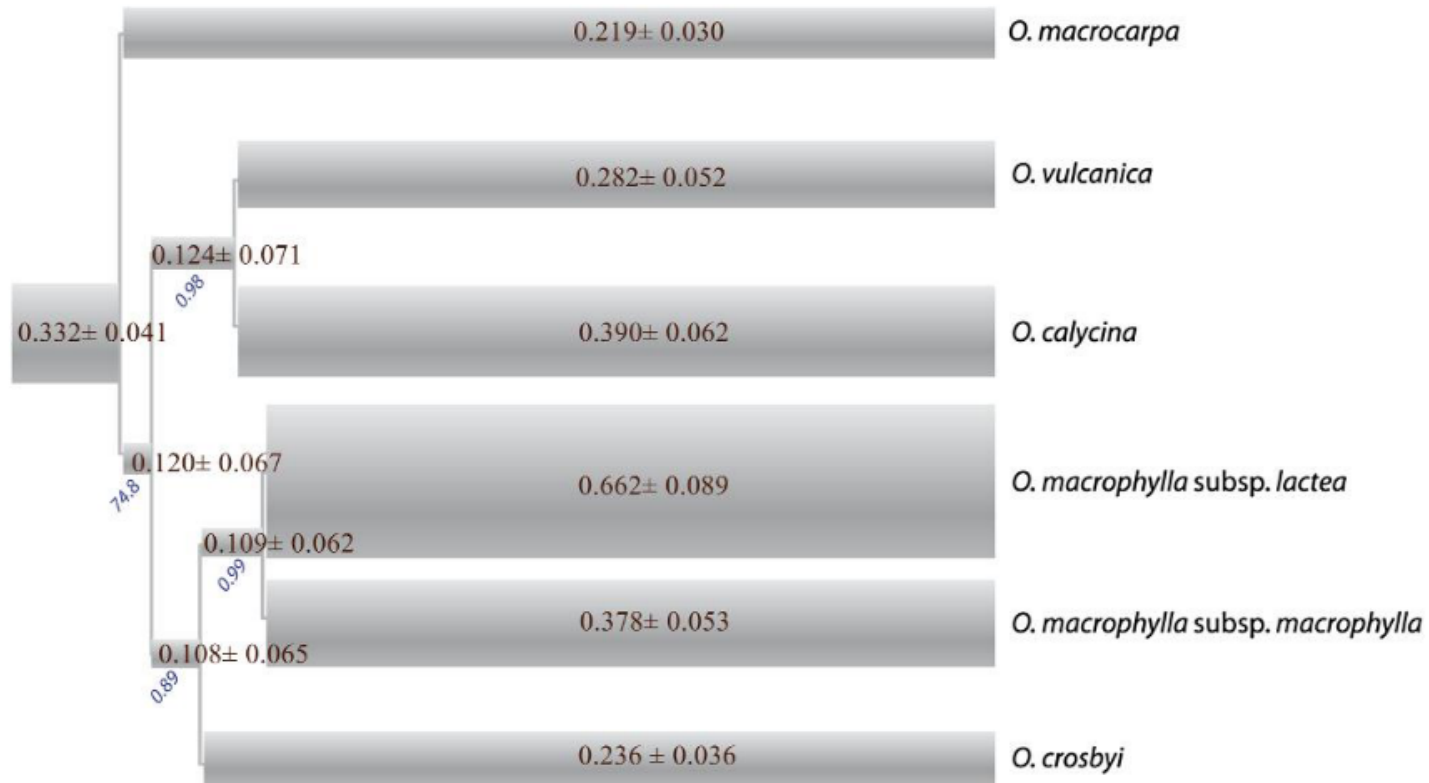
FIG. 5. Species tree with the highest posterior probability (73%) for six "large-leaf" *Ourisia* species. The thicknesses of bars are proportional to θ values for the respective populations. θ values for each population are printed on the pipes. The posterior probabilities for internal nodes are printed on an angle.

Species delimitation STACEY & BPP

Species limits in butterflies (Lepidoptera: Nymphalidae): reconciling classical taxonomy with the multispecies coalescent

PÁVEL MATOS-MARAVÍ[1,2,3], NIKLAS WAHLBERG[4],
ALEXANDRE ANTONELLI[1,2,5] and CARLA M. PENZ[6]

[1]Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, Sweden, [2]Gothenburg Global Biodiversity Centre, Gothenburg, Sweden, [3]Institute of Entomology, Biology Centre CAS, České Budějovice, Czech Republic, [4]Department of Biology, Lund University, Lund, Sweden, [5]Royal Botanical Gardens Kew, Richmond, U.K. and [6]Department of Biological Sciences, University of New Orleans, New Orleans, LA, U.S.A.
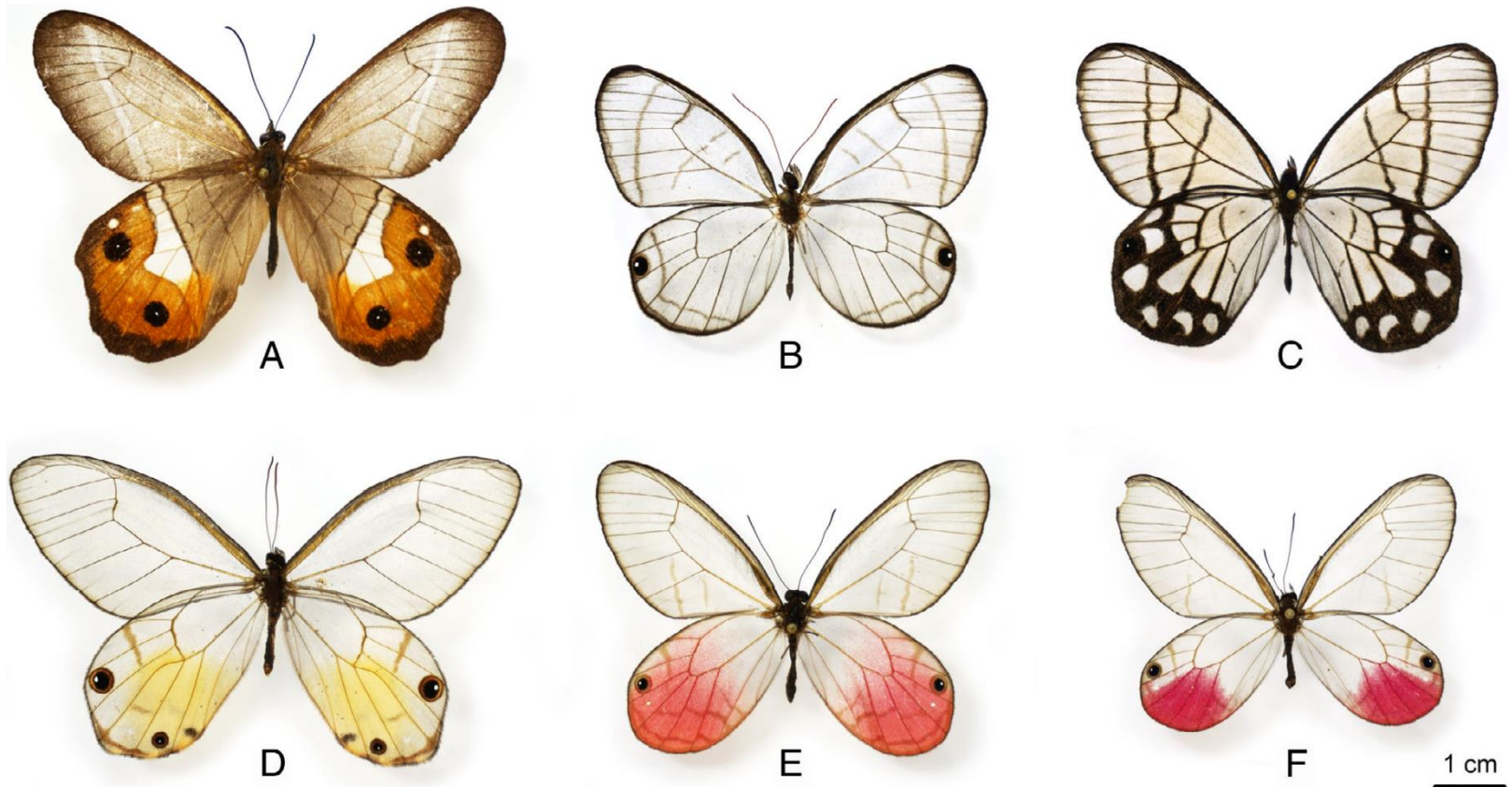
Fig. 1. Representatives of the tribe Haeterini. (A) *Pierella nereis* (Brazil, Minas Gerais, Santa Barbara; Milwaukee Public Museum); (B) *Dulcedo polita* (Costa Rica, Sarapiqui, Tirimbina Biological Station; Phil DeVries Collection); (C) *Pseudohaetera mimica* (Peru, Junin, Satipo; Natural History Museum of Los Angeles County); (D) *Haetera piera* (Ecuador, Napo, Garza Cocha; Phil DeVries Collection); (E) *Cithaerias cliftoni* (Ecuador, Oriente; Natural History Museum of Los Angeles County); (F) *Cithaerias aurora tambopata* (Peru, Madre de Dios, Pakitza, Manu National Reserve; Smithsonian Institution). Photographs: Carla M. Penz. [Colour figure can be viewed at wileyonlinelibrary.com].

Matos-Moravi *et al.* 2019 SysEntol

Matos-Moravi *et al.* 2019 SysEntol
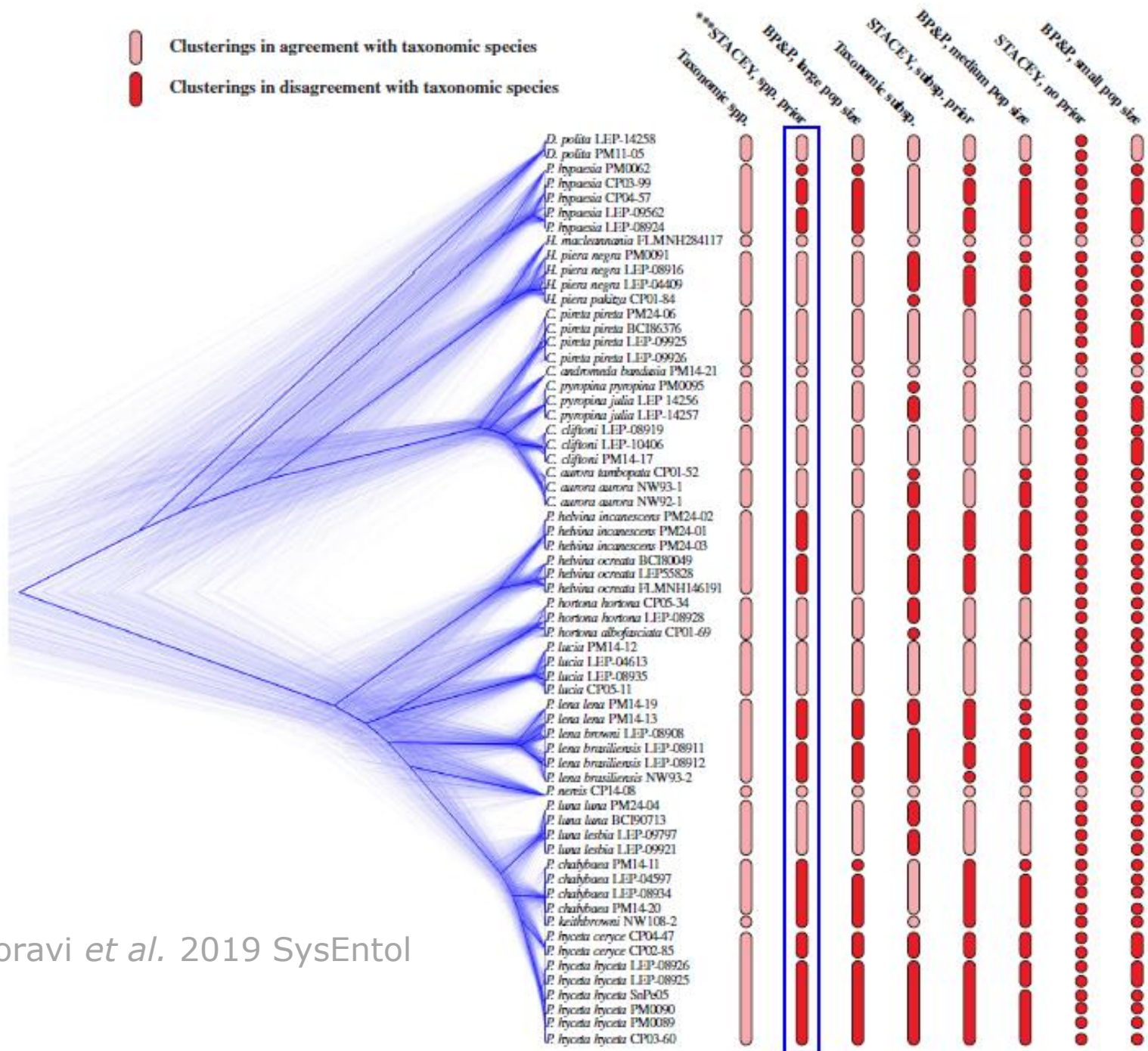
Species delimitation STACEY & BPP

Fig. 2. Evaluated species delimitation hypotheses using Bayes factors. The eight scenarios are: Taxonomic species (spp., 18 lineages) or subspecies raised to species (subsp., 26 lineages), STACEY's delimited species under prior accounting for number of taxonomic species (spp. prior, 22 lineages) or number of subspecies raised to species (subsp. prior, 24 lineages), as well as with prior not informed by taxonomy (no prior, 63 lineages), and BP&P's delimited species under prior for ancestral population size as large (21 lineages), medium (28 lineages) or small (53 lineages). The 'cloudogram', which is a diagram representing phylogenetic uncertainty of the 63 Haeterini specimens, is based on 500 posterior trees from STACEY analysis (thicker blue line represents the consensus phylogeny). The delimitation model STACEY under prior accounting for taxonomic species (outlined by a surrounding box) receives significant support based on Bayes factors over all other models, and is also the one that fits well with morphological and geographic differences; it is thus the classification that we propose here. [Colour figure can be viewed at wileyonlinelibrary.com].

Table 1. Characteristics of the molecular dataset used in this study.

Genes	Specimens	Length (bp)	Variable sites	Missing data (%)	GC content (%)
CAD	29 (46%)	850	157	25.8	33.5
COI	60 (95%)	1475	474	26.5	29.5
EF1α	49 (78%)	1240	230	17.5	48.6
GAPDH	21 (33%)	691	129	2.6	45.4
RpS5	55 (87%)	617	124	1.9	45.2
wingless	58 (92%)	412	96	9.3	58.6
TOTAL	63	5285	1210	39.2	41.4

Matos-Moravi *et al.* 2019 SysEntol

Species delimitation STACEY & BPP

A   **Haeterini species tree,STACEY, spp prior**

Dulcedo polita
Pseudohaetera [sp.n. #1]
Pseudohaetera [sp.n. #2]
Pseudohaetera hypaesia
Haetera macleannania
Haetera piera
Cithaerias pireta
Cithaerias andromeda
Cithaerias pyropina
Cithaerias cliftoni
Cithaerias aurora
Pierella brasiliensis **stat. rev.**
Pierella lena
Pierella ocreata **stat. rev.***
Pierella incanescens **stat. rev.***
Pierella lucia
Pierella hortona
Pierella nereis
Pierella luna
Pierella chalybaea
Pierella ceryce **stat. rev.**
Pierella hyceta

30   25   20   15   10   5   2   0

B   **Haeterini (sub)species tree,BP&P, medium pop size**

Dulcedo polita
Pseudohaetera [sp.n. #1]
Pseudohaetera hypaesia
Haetera macleannania
Haetera piera [subsp.n.]
Haetera piera negra
Haetera piera pakitza
Cithaerias pireta pireta
Cithaerias andromeda bandusia
Cithaerias pyropina julia
Cithaerias cliftoni
Cithaerias aurora tambopata
Cithaerias aurora aurora
Pierella brasiliensis **stat. rev.**
Pierella lena lena
Pierella lena [subsp.n.]
Pierella lena browni **stat. rev.**
Pierella ocreata **stat. rev.***
Pierella incanescens **stat. rev.***
Pierella lucia
Pierella hortona hortona
Pierella nereis
Pierella luna
Pierella chalybaea [subsp.n.]
Pierella chalybaea
Pierella ceryce **stat. rev.**
Pierella hyceta
Pierella hyceta [subsp.n.]
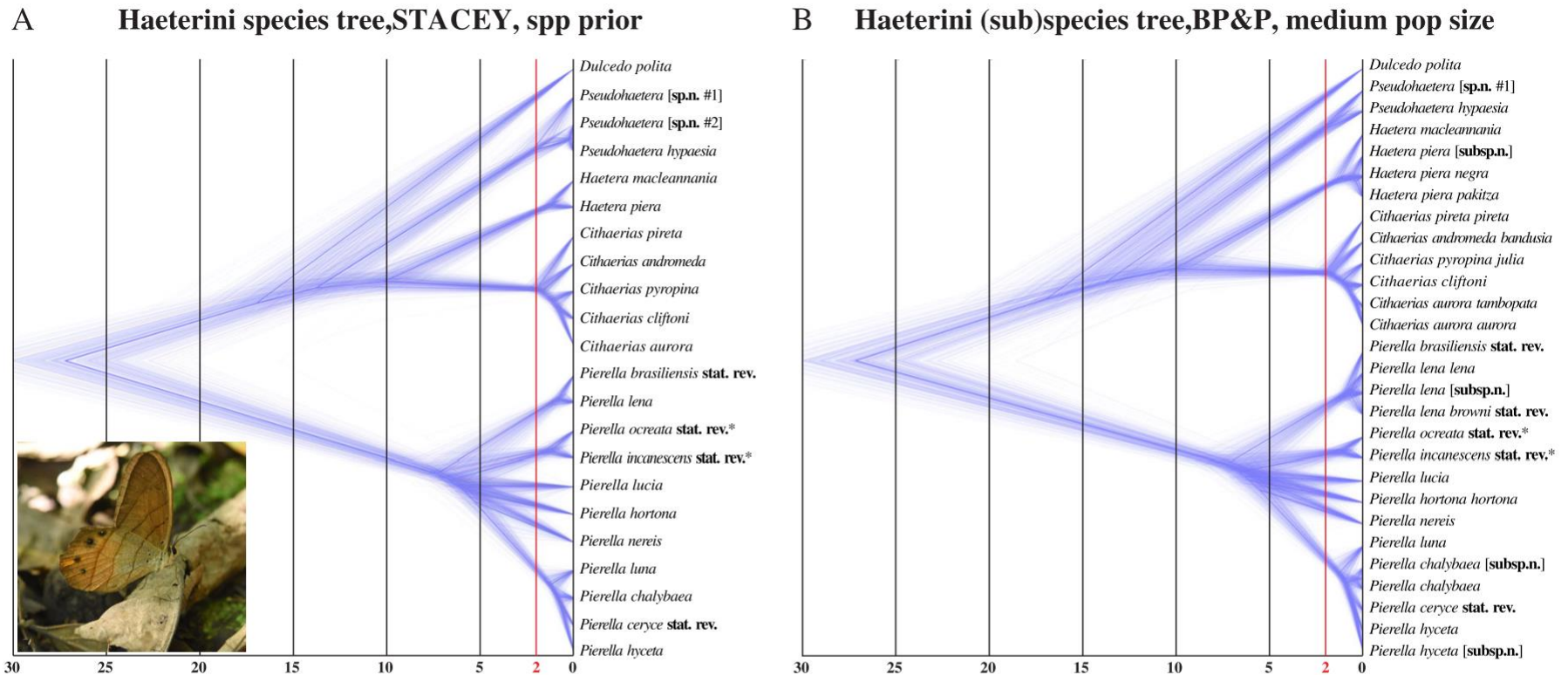
30   25   20   15   10   5   2   0

Fig. 3. Time-calibrated phylogenetic hypotheses of models that best approximate species and subspecies in Haeterini. (A) 'Cloudogram' of the best-fit species delimitation model based on Bayes factors, STACEY under prior accounting for taxonomic species. (B) 'Cloudogram' of the delimitation model that best approximate described subspecies, BP&P under prior for medium ancestral population size. Time axes in both panels are scaled to Ma. *The species status of *Pierella helvina ocreata* and *Pierella helvina incanescens* may change with the inclusion of *Pierella helvina helvina*, but it is likely that *P. helvina ocreata* and *P. helvina incanescens* are separate species. Inset: butterfly *Pierella hyceta hyceta* (Peru, Pasco, Cañón de Huancabamba, 1200 masl, 29.vii.2017). Photograph: Markéta Aubrechtová. [Colour figure can be viewed at wileyonlinelibrary.com].

Octávio S. Paulo – FEM

Matos-Moraví *et al.* 2019 SysEntol

29

A comparative approach for species delimitation based on multiple methods of multi-locus DNA sequence analysis: A case study of the genus *Giraffa* (Mammalia, Cetartiodactyla)

Alice Petzold[1,2]**, Alexandre Hassanin**[1,2]*

1 Institut de Systématique, Évolution, Biodiversité (ISYEB), Sorbonne Université, MNHN, CNRS, EPHE, Paris, France, **2** Muséum national d'Histoire naturelle, CP51, Paris, France
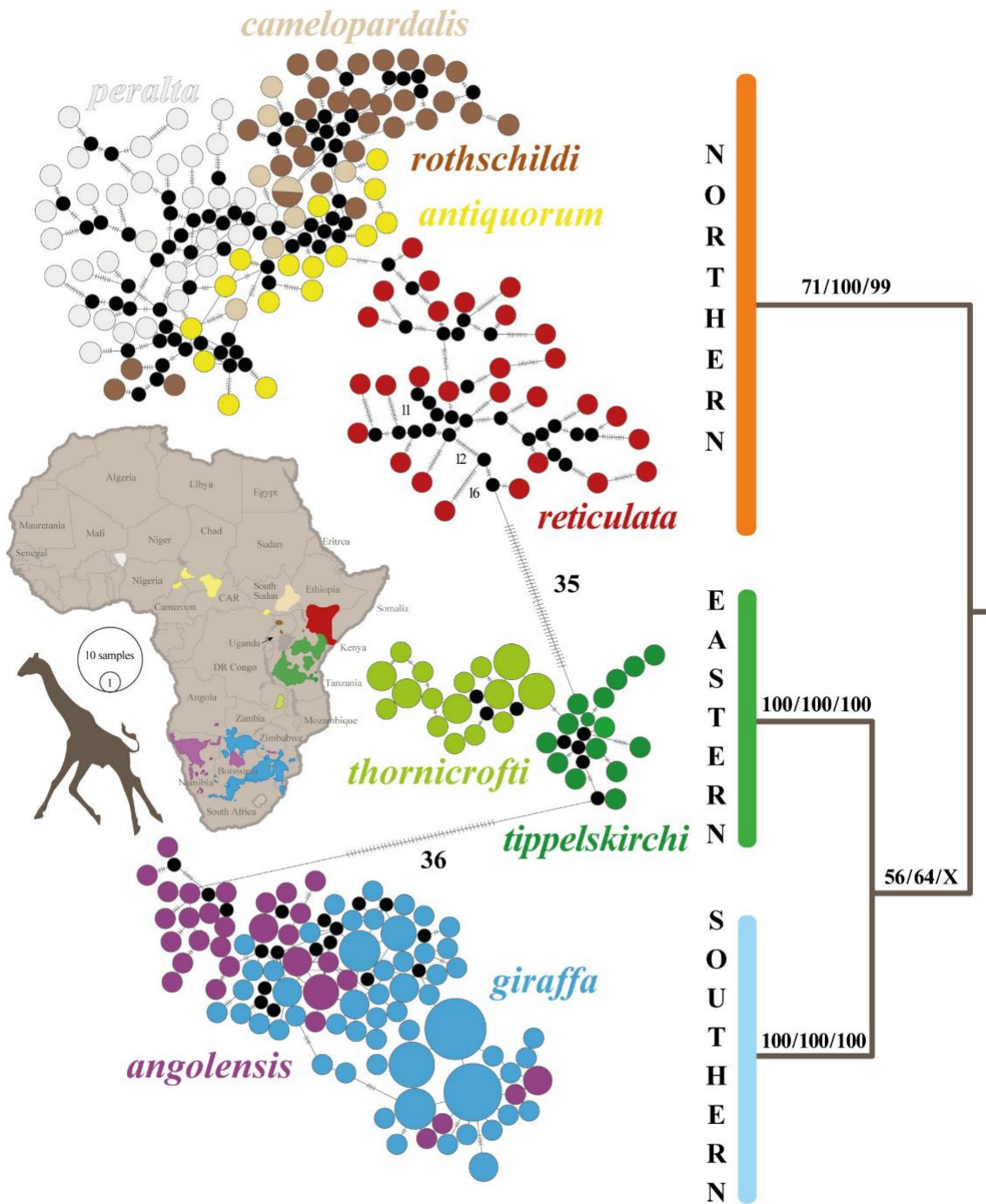
camelopardalis

peralta

rothschildi

antiquorum

reticulata

thornicrofti

tippelskirchi

giraffa

angolensis

10 samples
1

35

36

NORTHERN

71/100/99

EASTERN

100/100/100

56/64/X

SOUTHERN

100/100/100

31

Table 1. Species delimitation methods based on multi-locus nuDNA sequences used in this study.

Method	Input	Category	Reference	Description	SD Criteria
STRUCTURE	alignment of phased alleles	**PG**	Pritchard et al. [26] Falush et al. [53]	Bayesian clustering method based on the estimation of allele frequencies	ΔK and plateau methods
Haplowebs	alignment of phased alleles	**PG**	Flot et al. [30]	Allele-sharing method, in which groups of individuals sharing a unique pool of alleles are connected by curves.	Group of individuals constituting a field for recombination [FFR]
Conspecificity matrix (CM)	Conspecificity scores	**PG**	Debortoli et al. [31]	The conspecificity matrix was generated by calculating for each pair of individuals a conspecificity score, i.e., the number of markers supporting conspecificity in haploweb analyses.	Number of markers supporting the hypothesis of conspecificity
BPP (Bayesian Phylogenetics and Phylogeography)	alignment of consensus sequences	**MSC**	Yang and Rannala [64]	Bayesian method based on the MSC model, in which a reversible-jump Markov chain Monte Carlo algorithm is used to calculate the posterior probabilities of species delimitations.	Probability ≥ 0.95
STACEY (Species Tree and Classification Estimation, Yarely)	alignment of phased alleles	**MSC**	Jones [24] ----- Individual assignment of alleles: present study	Bayesian method implemented in BEAST 2 [65] for the inference of a "species or minimal clusters tree" (SMC) under the birth-death-collapse tree prior and without the requirement of a guide tree.	Probability ≥ 0.95
*****BEAST** (Species Tree Ancestral Reconstruction in BEAST)	alignment of phased alleles	**MSC + P**	Program: Heled and Drummond [19]; Individual assignment of alleles: present study	Bayesian method implemented in BEAST 2 [65] based on the MSC model	Probability ≥ 0.95
Bootstrap Analysis of Haplotypes	alignment of phased haplotypes	**PG+ + P**	Present study	Bootstrap consensus tree reconstructed with ML, MP or NJ methods	Bootstrap ≥ 90
Supermatrix MrBayes	alignment of consensus sequences	**P**	Ronquist et al. [46]	Bayesian inference of phylogeny	Probability ≥ 0.95
PhyML			Guindon et al. [47]	ML method for tree construction	Bootstrap ≥ 90
SuperTRI (SuperTree with Reliability Indices)	Weighted binary matrix of node support for each locus	**P**	Ropiquet et al. [20]	Three measures are calculated to estimate the reliability of the nodes (SBP, MPP and NRep) using the branch support values (PP) of all phylogenetic hypotheses produced during the separate Bayesian analyses of the 21 introns	(1) SBP ≥ 90 (2) MPP ≥ 0.1 (3) NRep ≥ 2

SD: Species delimitation; PG: Population Genetics; MSC: Multispecies Coalescent; P: Phylogenetic methods; SBP: Supertree Boostrap Percentage; MPP: Mean Posterior Probability

Rep: Reproducibility Index; SPR: Subtree Pruning and Re-grafting
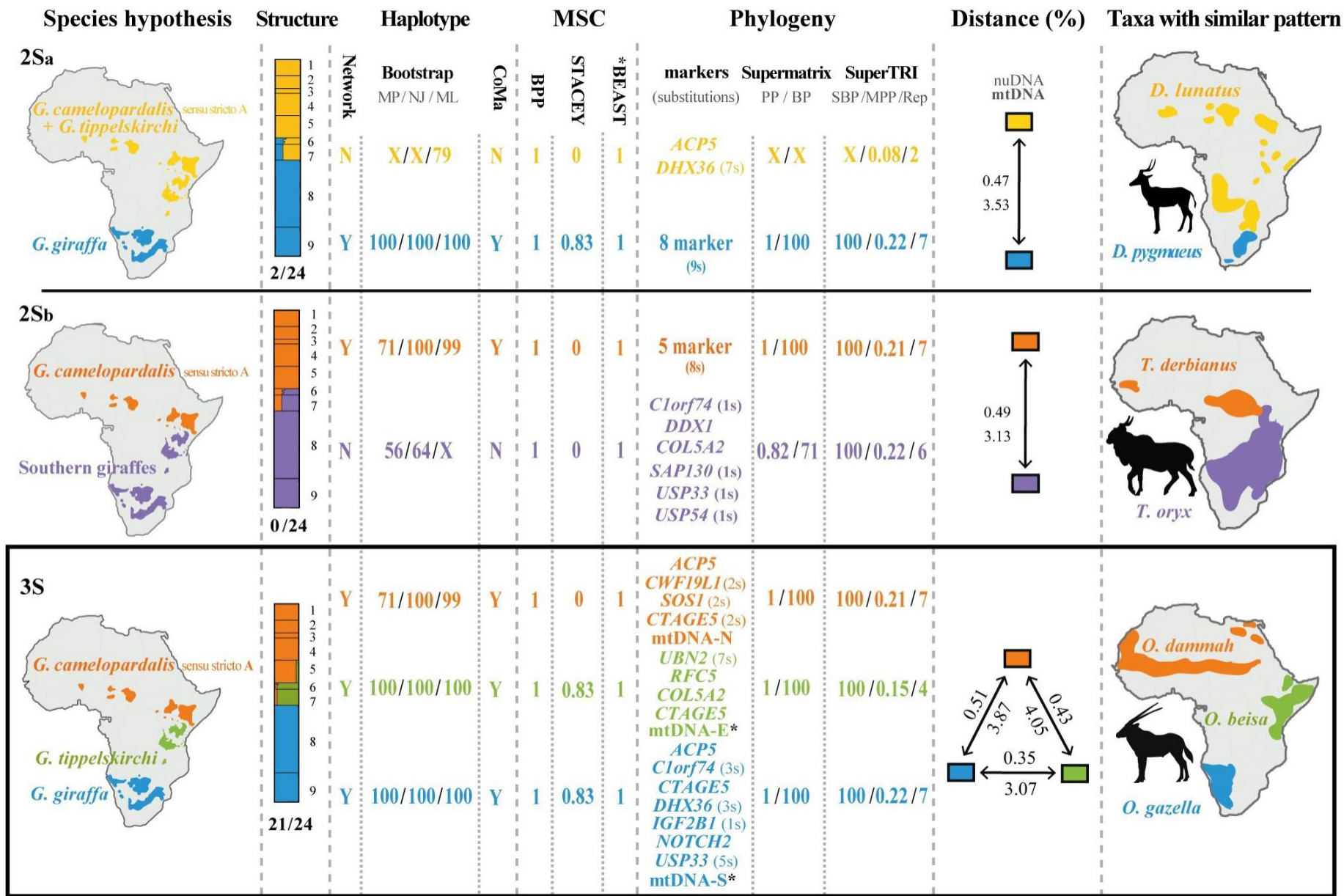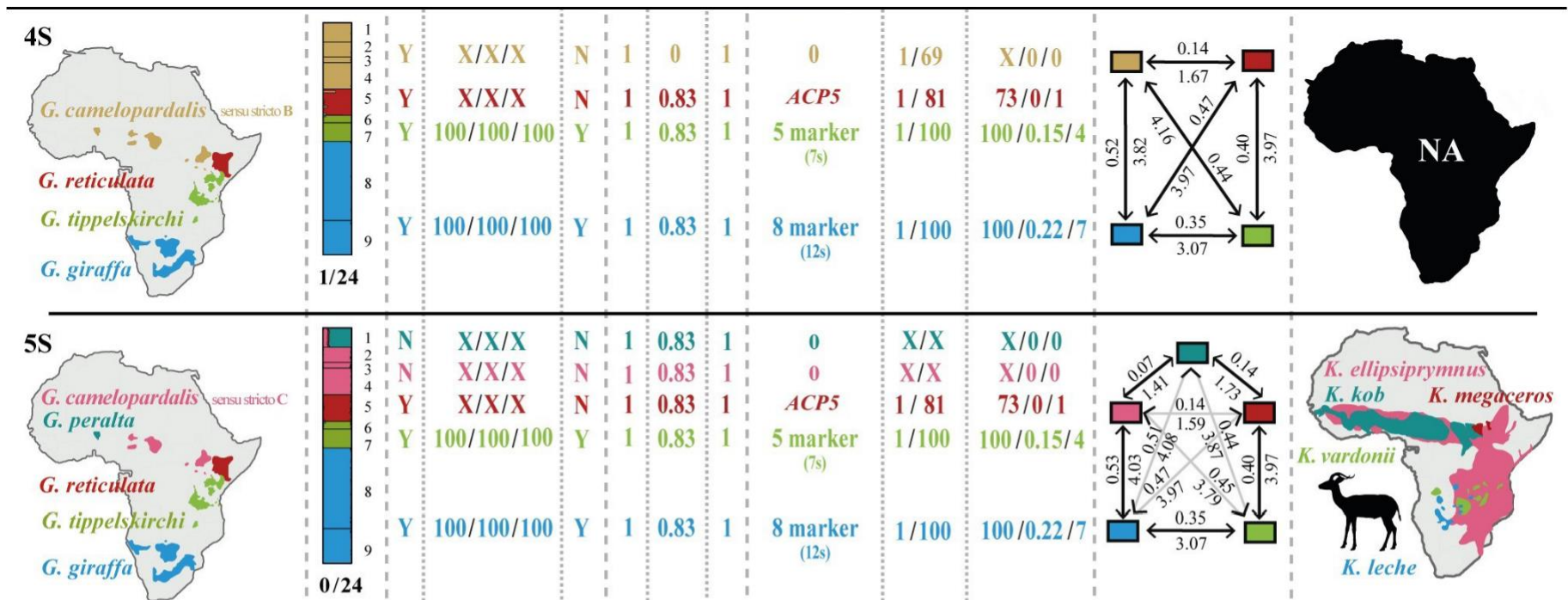
Species delimitation STACEY & BPP

Fig 6. The five molecular hypotheses for giraffe taxonomy. The five taxonomic hypotheses that received some support from our analyses on giraffes show the existence of two species, with two possible geographic patterns (2Sa and 2Sb hypotheses), three species (3S hypothesis), i.e. *G. camelopardalis* sensu stricto A, *G. giraffa* and *G. tippelskirchi*, four species (4S hypothesis), i.e. *G. camelopardalis* sensu stricto B, *G. giraffa*, *G. reticulata*, and *G. tippelskirchi*, or five species (5S hypothesis), i.e. *G. camelopardalis* sensu stricto C, *G. giraffa*, *G. peralta*, *G. reticulata*, and *G. tippelskirchi*. In the first column are drawn the geographic distributions of giraffe species for each of the five taxonomic hypotheses. In the second column are summarized the results obtained from STRUCTURE analyses. Barplots were illustrated with DISTRUCT (1 = peralta, 2 = antiquorum, 3 = camelopardalis, 4 = rothschildi, 5 = reticulata, 6 = tippelskirchi, 7 = thornicrofti, 8 = giraffa, 9 = angolensis) and number of analyses supporting each taxonomic hypothesis (in total 24, see Table 3) is indicated beneath barplots. In the third column are illustrated the results obtained in the different haplotype analyses, including the network analysis (Y = yes, the species represents a cluster; N = no, the species is not found as a cluster), the bootstrap values obtained with the phylogenetic analyses based on the Maximum Parsimony (MP), Distance (NJ) and Maximum Likelihood (ML) criterion ("X": support < 50) and the conspecificity matrix (CoMa) (Y = yes, the species is supported by the analysis; N = no, the species is not supported by the analysis). In the fourth column are shown the support values provided by the three Multispecies coalescent (MSC) methods, i.e. BPP, STACEY and *BEAST. In the fifth column are listed the results

Multispecies coalescent delimits structure, not species

Jeet Sukumaran[a,1,2] and L. Lacey Knowles[a,1]

[a]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor MI 48109-1079

The multispecies coalescent model underlies many approaches used for species delimitation. In previous work assessing the performance of species delimitation under this model, speciation was treated as an instantaneous event rather than as an extended process involving distinct phases of speciation initiation (structuring) and completion. Here, we use data under simulations that explicitly model speciation as an extended process rather than an instantaneous event and carry out species delimitation inference on these data under the multispecies coalescent. We
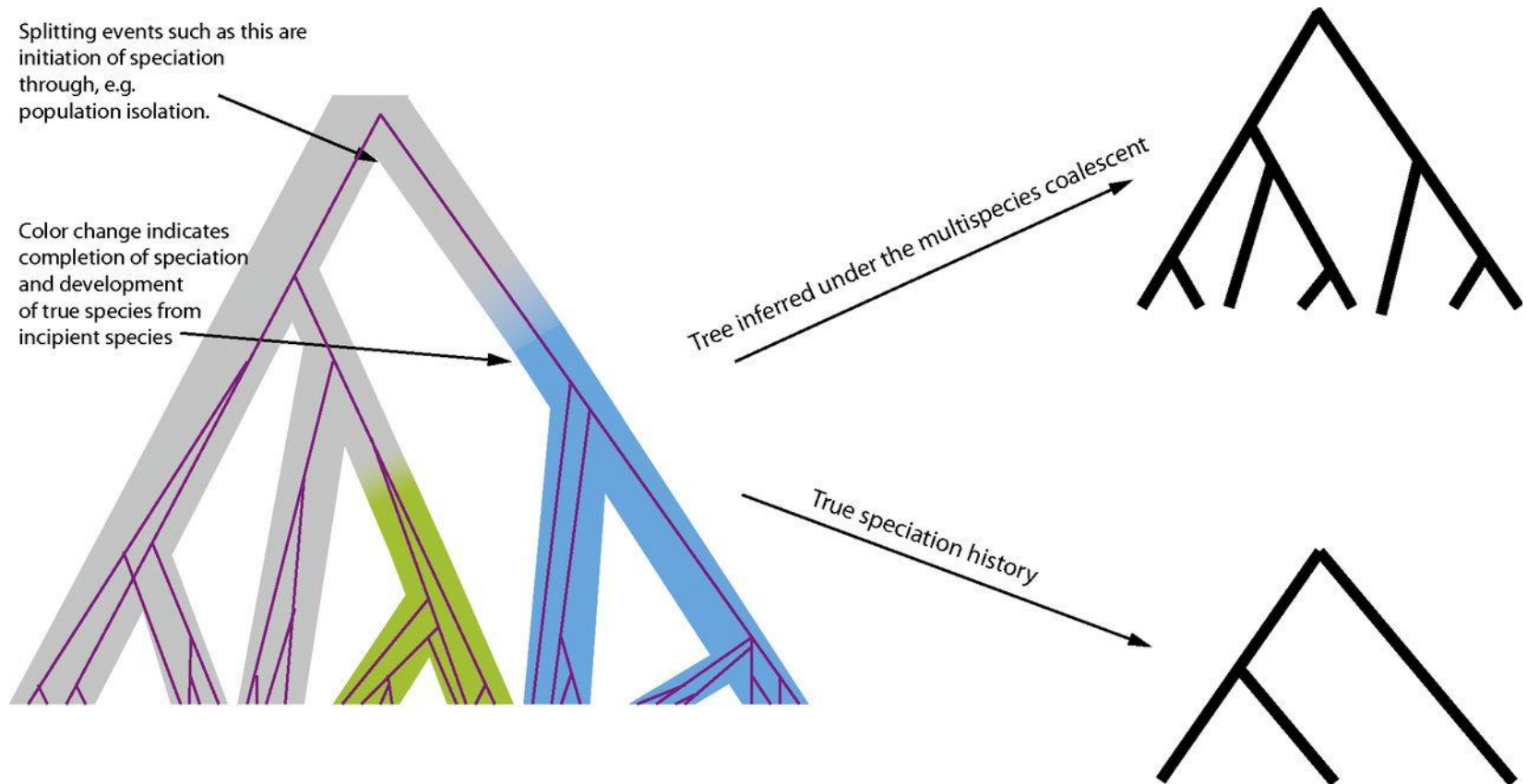
consequence, the increased resolution of genomic data makes it possible to not only detect divergent species lineages, but also local population structure within them—that is, a fractal hierarchy of divergences.

Misidentification of population structure as putative species is therefore emerging as a key issue (8) that has received insufficient attention, especially with respect to methodologies for delimiting taxa based on genetic data alone. Because species delimitation is inextricably linked to patterns of species diversity.

Significance

Despite its widespread application to the species delimitation problem, our study demonstrates that what the multispecies coalescent actually delimits is structure. The current implementations of species delimitation under the multispecies coalescent do not provide any way for distinguishing between structure due to population-level processes and that due to species boundaries. The overinflation of species due to the misidentification of general genetic structure for species boundaries has profound implications for our understanding of the generation and dynamics of biodiversity, because any ecological or evolutionary studies that rely on species as their fundamental units will be impacted, as well as the very existence of this biodiversity, because conservation planning is undermined due to isolated populations incorrectly being treated as distinct species.

(Left) The tree represents the true history upon which the gene genealogies (shown by thin purple lines) are conditioned, with the colors representing species.

Splitting events such as this are initiation of speciation through, e.g. population isolation.

Color change indicates completion of speciation and development of true species from incipient species

Tree inferred under the multispecies coalescent

True speciation history

Jeet Sukumaran, and L. Lacey Knowles PNAS
2017;114:7:1607-1612

PNAS