INVITED REVIEWS AND META-ANALYSES
How to fail at species delimitation

BRYAN C. CARSTENS,* TARA A. PELLETIER,* NOAH M. REID† and JORDAN D. SATLER*

*Department of Evolution, Ecology and Organismal Biology, The Ohio State University, 318 W. 12th Avenue, Columbus, OH43210-1293, USA, †Department of Biological Sciences, Louisiana State University, Life Sciences Building, Baton Rouge, LA 70803, USA

Abstract

Species delimitation is the act of identifying species-level biological diversity. In recent years, the field has witnessed a dramatic increase in the number of methods available for delimiting species. However, most recent investigations only utilize a handful (i.e. 2–3) of the available methods, often for unstated reasons. Because the parameter space that is potentially relevant to species delimitation far exceeds the parameterization of any existing method, a given method necessarily makes a number of simplifying assumptions, any one of which could be violated in a particular system. We suggest that researchers should apply a wide range of species delimitation analyses to their data and place their trust in delimitations that are congruent across methods. Incongruence across the results from different methods is evidence of either a difference in the power to detect cryptic lineages across one or more of the approaches used to delimit species and could indicate that assumptions of one or more of the methods have been violated. In either case, the inferences drawn from species delimitation studies should be conservative, for in most contexts it is better to fail to delimit species than it is to falsely delimit entities that do not represent actual evolutionary lineages.

In the broadest sense, species delimitation is the act of identifying species-level biological diversity. As genetic data have become easier and less expensive to gather, the field of phylogeography has experienced an explosion in the number and variety of methodological approaches to species delimitation. In this study, we discuss the application of these methods to data collected from sexually reproducing organisms. Our aim is not to rank available methods or to evaluate their effectiveness under some narrow range of simulation conditions, as these exercises have been conducted previously (e.g. O'Meara 2010; Yang & Rannala 2010; Ence & Carstens 2011; Rittmeyer & Austin 2012). These studies generally find that methods are relatively accurate in simple cases (i.e. when attempting to delimit 2–3 lineages) using modest amounts of data (e.g. ~10–20 loci). Rather, we argue *any* of the existing methods are likely to be incapable of accurately delimiting evolutionary lineages under some plausible set of conditions and

that because of these limitations, researchers conducting species delimitation should analyse their data using a wide range of methods and place their trust in the observable congruence across the results. After presenting this argument, we broaden our discussion of species delimitation to place genetic approaches in a wider context that includes delimitation with nongenetic sources of data and various concepts of species.

The parameter space relevant to species delimitation

A useful thought experiment to consider before discussing the existing methods for species delimitation is to imagine the perfect methodological approach for delimiting independent evolutionary lineages using genetic data. Existing models range from nonparametric (e.g. Wiens & Penkrot 2002) to highly parameterized models (e.g. Yang & Rannala 2010). An idealized method would probably implement a parametric approach, because most methods for species delimitation operate by fitting some model of the historical diversification to

Correspondence: Bryan C. Carstens, Fax: +1 (614) 292 2030; E-mail: carstens.12@osu.edu

Box 1

Systems in which species delimitation methods are applied are emergent phylogenetic systems

Phylogenetic analysis is traditionally concerned with estimating relationships among species, while population genetics is concerned with adaptation and understanding the population-level forces that change allele frequencies. Since the modern synthesis, these fields largely developed in parallel, but without much interaction (Hull 1998; Felsenstein 2004). However, in recent decades, there has been increased activity at the interface of these disciplines. For example, divergence population genetics (Hey & Nielsen 2004) models divergence among populations, and a number of approaches concerned with estimating population divergence have been developed (Nielsen & Wakeley 2001; Hey & Nielsen 2004; Gutenkunst *et al.* 2009). Similarly, there are phylogenetic approaches to estimating population-level parameters (e.g. Heled & Drummond 2008). However, the most striking evidence of the merger between population genetics and phylogenetics has been the use of the multispecies coalescent model to estimate phylogeny (i.e. species trees, Edwards 2009). Seminal work by Maddison (1997) convinced a generation of researchers that they could not overlook the inherent stochasticity of genealogies when estimating phylogeny at the shallowest levels of diversification, and in time the basic coalescent model of Kingman (1982) was expanded to include multiple species (e.g. Rannala & Yang 2003). As multilocus data became common, and as curious statistical results such as anomalous zones (Degnan & Rosenberg 2006) were discovered, the multispecies coalescent (i.e. species trees) became the default option for phylogeny estimation within phylogeographical investigations. Concurrently, new algorithms for detecting population genetic structure (Pritchard *et al.* 2000; Huelsenbeck & Andolfatto 2007) were developed, and questions related to species delimitation returned to the forefront of systematics (Wiens 2007). Three particular developments have enabled new approaches to species delimitation. First, coalescent-based methods that model population-level processes such as genetic drift in combination with migration (Beerli & Felsenstein 2001), expansion (Kuhner *et al.* 1998), population divergence (Hey & Nielsen 2007) or combinations of these processes (Gutenkunst *et al.* 2009) are widely applied to phylogeographical data. Second, phylogenetic methods that estimate phylogeny while allowing for the action of some of these processes (Maddison & Knowles 2006; Liu *et al.* 2007; Kubatko *et al.* 2009; Heled & Drummond 2010) are increasingly applied to low-level phylogenetic systems. Third, the field is no longer data limited, as advances in DNA sequencing allow data from 1000s of loci to be collected from nearly any system (McCormack *et al.* 2013). In summary, the systems in which researchers conduct species delimitation exist at the interface of traditional population genetic and phylogenetic analyses. As such, they borrow methods from each but have a unique set of challenges.

the data collected from some natural system. However, the parameter space relevant to species delimitation is large, and existing tools limit their exploration to a subset of the potential parameter space by assuming a number of simplifications, with each method making a different set of simplifying assumptions. While this is a strategy inherent to model-based analysis, and not intended as a criticism, researchers should realize that all methods incorporate models that are imperfect imitations of the biological reality. At a minimum, this parameter space for species delimitation includes population genetic and phylogenetic parameters. Examples of population genetic parameters include gene flow and population size as well as other parameters that are perhaps not directly relevant to species delimitation, but which may influence patterns of genetic diversity (e.g. extrinsic rates of population size change, recombination). Phylogenetic parameters related to the pattern and timing of lineage diversification are also necessary because systems to which species delimitation methods are applied are inherently emergent phylogenetic

systems, even though the empirical data from such systems may approach the lower limits of appropriateness for phylogenetic analysis (see Box 1). Furthermore, an ideal method would be agnostic with regard to the assignment of samples to population and as such would not require that the researcher provide this information. Many existing methods implement a coalescent model (e.g. Kingman 1982) and thus also assume that the data meet the standard assumptions of this model (i.e. that the data are sampled from neutral loci that do not contain internal points of recombination, that populations are panmictic, etc.). The potential parameter space relevant to species delimitation is larger and far more complex than that considered by even the most heavily parameterized of existing methods. Given that there is a large gap between our *ideal method* and existing methods, how should researchers proceed when delimiting species?

A naive response to the above conundrum is to identify a single method that is demonstrably accurate in some simulation study and apply it alone to the data.

Thus, one researcher might choose Structurama (Huelsenbeck *et al.* 2011) based on the results presented by Rittmeyer & Austin (2012), analyses their data using this method and delimits species based on the results. This course of action has two substantial shortcomings. First, to our knowledge, no simulation study has included every potentially useful method, so any evaluation of existing methods will necessarily be incomplete. Second, results from simulation studies are conditional on the specific attributes of the simulated data used in such studies (i.e. the assumed θ, number of variable sites, number of loci) and thus most relevant when tailored to specific systems. To illustrate why these details matter, consider the simulations presented by Camargo *et al.* (2012) in an investigation into *Liolaemus* lizards. They reported that BPP (Yang & Rannala 2010) and a custom approximate Bayesian computation method were more accurate in their simulations than the method spedeSTEM (Ence & Carstens 2011). Camargo *et al.* also found spedeSTEM to be less accurate in their simulations than in those reported by the program authors. This discrepancy resulted from differences in the sample size and numbers of variable sites in the simulated data, which influenced the quality of the gene tree estimates that are input into spedeSTEM. Because these studies were based on the levels of variation in data collected from different empirical systems, the simulation results differed slightly. For a researcher trying to delimit species in non-*Liolaemus* systems, the results from a simulation study such as Camargo *et al.* (2012) should be viewed as a rough guideline and not be considered a directly transferable prediction of accuracy.

The Camargo *et al.* (2012) study exemplifies a more sophisticated approach to delimiting species. By designing and conducting a simulation study that matches the characteristics of their *Liolaemus* lizard system, Camargo *et al.* (2012) establish an expectation of accuracy that provides them with some degree of confidence in their results. Because their simulations indicate that the methods used are generally able to delimit species using the amount of data that they have collected, they can confidently assert that their study has discovered independent evolutionary lineages within *Liolaemus*. Given that simulated data are far easier to generate than actual data, such an approach to species delimitation should be widely emulated by any empirical study that attempts to delimit species. Another aspect of their study is worth highlighting. Camargo *et al.* (2012) chose to focus on methods derived from the multispecies coalescent model (BPP and spedeSTEM), but because they were unwilling to assume that all shared polymorphism was due to incompletely sorted ancestral polymorphism (an assumption of the above methods), they also utilized approximate Bayesian computation (ABC; see

Csilléry *et al.* 2011) in an effort to model divergence with gene flow. While Camargo *et al.* (2012) delimit the southern and northern populations as separate lineages using all three approaches, their simulation study would have provided justification for choosing one set of results over the others in the event of incongruence. However, even though their simulations represent an additional level of rigour when compared to most empirical investigations, Camargo *et al.* (2012) limit their consideration to a restricted set of delimitation analyses and thus may not have identified the optimal method for their system. This is perhaps justifiable, because the focal system of Camargo *et al.* (2012) was clearly disjunct in two populations and sample assignment to putative lineages was unambiguous. In this way, their investigation represents the low-hanging fruit of species delimitation investigations because sample assignment was not in question and because the system represented either one or two lineages. However, many focal systems are more complex.

In systems where the populations are not clearly delineated, the assignment of samples to putative groups is essential. Sample assignment dramatically increases the difficulty in species delimitation (O'Meara 2010) because it adds dramatically to the complexity of the algorithms used in species delimitation. As a result, methods for species delimitation have approached sample assignment from a number of directions (Box 2). Genetic clustering approaches such as Structure (Pritchard *et al.* 2000) and Structurama (Huelsenbeck *et al.* 2011) operate by identifying the population assignments and level of clustering that minimizes Hardy–Weinberg disequilibrium. While clustering algorithms do not explicitly model population-level parameters (i.e. θ, migration rates or population divergence), they are flexible in terms of the data that are required and have been applied to a wide range of systems. Notably, the ability of clustering algorithms to cluster samples does not alone offer compelling proof that the delimited entities have a history of phylogenetic divergence because the population structure is inferred without consideration of historical patterns of diversification, and there is not always a clear correspondence between a given level of clustering and the branching pattern of a species tree (Jackson & Austin 2010; Kalinowski 2010). Furthermore, given enough data it may be possible to delimit extremely localized populations. For example, results from O'Dushlaine *et al.* (2010) indicate that population structure among individual villages in Europe can be identified using ~300 000 SNPs. While this is clearly more data than most existing species delimitation investigations utilize, it is likely that investigations into nonmodel systems will approach this number in the near future. Due to these potential issues, many authors use the results from Structure or Structu-

<div style="border:1px solid black; padding:1em;">

Box 2
Some useful methods for species delimitation
Species discovery approaches assign samples to groups without a priori information

Structurama (Huelsenbeck *et al.* 2011) implements the clustering algorithm first described by Pritchard *et al.* (2000; for their program Structure) that clusters samples into populations by minimizing Hardy–Weinberg disequilibrium for a given partitioning level. Structurama includes the addition of reversible jump MCMC to identify the optimal partitioning level. Nearly any type of genetic data can be input into Structurama, and the program can assign individuals to population with or without the admixture. One shortcoming of genetic clustering approaches is that they do not assess the evolutionary divergence of population clusters (http://cteg.berkeley.edu/~structurama/index. html).

Gaussian Clustering (Hausdorf & Hennig 2010) groups samples into populations using genotypic data by searching for clusters that can be attributed to being mixtures of normal allele frequency distributions. Like Structurama, the method is flexible in terms of the data that can be analysed. This approach is implemented in R using the prabclus (Hausdorf & Hennig 2010) and mclust (Fraley & Raftery 2006) packages. As in other clustering approaches, temporal divergence among putative groups is not explicitly estimated.

The **general mixed Yule coalescent model** (GMYC; Pons *et al.* 2006) takes an ultrametric genealogy estimated from a single genetic locus as input. The method attempts to model the transition point between cladogenesis and allele coalescence by utilizing the assumption that the former will occur at a rate far lower than the later. This results in a shift in the rate of branching of the genealogy that reflects the transition between species-level processes (such as speciation and extinction) and population-level processes (allele coalescence). Reid & Carstens (2012) proposed a version of the GMYC that accounts for phylogenetic uncertainty gene tree estimates using a Bayesian analysis. Both implementations of the GMYC are likely to delimit well-supported clades of haplotypes as independent lineages and as such may be prone to over delimitation (http://r-forge.r-project.org/projects/splits, https://sites.google. com/site/noahmreid/home/software).

Choi & Hey (2011) describe two new methods for jointly estimating population assignment along with the parameters of an isolation-with-migration model. **Joint demography and assignment** (JDA) is applicable to an island or two population models, while **joint demography and assignment of population tree** (JDAP) is applicable to more than two diverging populations. Each takes sequence data as input and is implemented within IMa2 (Hey & Nielsen 2007) (http://genfaculty.rutgers.edu/hey/software).

Unlike other methods described here, the unified model of Guillot *et al.* (2012) can analyse nongenetic data (phenotypical, geographical, behavioural) in addition to genetic data. Their approach implemented a Bayesian clustering algorithm that assumes that each cluster in a geographical domain can be approximated by polygons that are centred around points generated by a Poisson process. Guillot *et al.*'s model is flexible in terms of the genetic data that it can utilize and capable of accurately delimiting species. Their model is available as an extension of the R GENELAND package (Guillot *et al.* 2005) (http://www2.imm.dtu.dk/~gigu/Geneland/).

O'Meara's heuristic method (O'Meara 2010) of species delimitation takes gene trees from multiple loci as input and operates under a similar assumption to the GMYC (namely that allelic coalescence occurs more rapidly than speciation). Provided that this assumption is true, the longest branches of gene trees are likely to represent species-level differences, and thus, congruence across loci is indicative of both the species tree and the population assignments. O'Meara's method is implemented in the Brownie package (O'Meara 2008). Because this method takes gene trees as input, its accuracy will likely be correlated with the nodal support values in the gene trees (http://www. brianomeara.info/brownie).

Species delimitation analyses that use the multispecies coalescent model compare the probability of trees with differing numbers of OTUs to identify optimal partitions of the data (e.g. spedeSTEM, BPP). Salter *et al.* (2013) extend this strategy to its maximum extent by calculating the probability of the phylogeny that treats individual samples as putative lineages. The putative lineages are then sequentially collapsed on the basis of which samples are most closely related, the probability of the species tree is recalculated, and information theory (Burnham & Anderson 2002) is used to identify the optimal model of lineage composition. Thus, **spedeSTEM discovery** can be used to simultaneously delimit evolutionary lineages and assign samples to these lineages (http://carstenslab.org.ohio-state.edu/software.html).

</div>

Box 2 *Continued*
Species validation approaches require the user to assign samples to putative lineages

The popular validation method **BPP** (Yang & Rannala 2010) implements a reversible jump Markov chain Monte Carlo (rjMCMC) search of parameter space that includes θ, population divergence and estimated distributions of gene trees from multiple loci. The method takes sequence data as input and also requires the user to define the topology of the species tree. Given this information, the algorithm implemented in BPP then traverses the parameter space to compute the posterior probability of the proposed nodes of the species tree. While inaccurately specified guide trees can lead to false-positive delimitations, the accuracy of BPP does not generally appear dependent on its ability to estimate gene trees. As this manuscript was in review, an improvement to the rjMCMC was described (Rannala & Yang 2013) (http://abacus.gene.ucl.ac.uk/software.html).

The validation approach **spedeSTEM** was developed to test species boundaries in a system with existing subspecies taxonomy (Carstens & Dewey 2010). The approach computes the probability of the gene trees given the species tree for all hierarchical permutations of lineage grouping, and therefore, complex cases such as that described by Carstens and Dewey (four species with 1–4 described subspecies) can be evaluated. Because the $-\ln L$ (ST | GTs) is computed directly by STEM (Kubatko *et al.* 2009), rather than estimated, phylogenetic uncertainty in the species tree does not affect species delimitations. However, accuracy of spedeSTEM is dependent on the quality of the gene tree estimates (http://carstenslab.org.ohio-state.edu/software.html).

---

rama as a starting point for delimitation investigations (e.g. Leaché & Fujita 2010), and in particular, they complement methods that explicitly model population divergence.

Knowles & Carstens (2007) suggested that the multispecies coalescent model (Rannala & Yang 2003; Box 1) offers an important opportunity for species delimitation. Relative to the delimitation of evolutionary lineages, the important difference between species tree methods that implement this model and conventional methods of phylogenetic inference is the shift in what constitutes an operational taxonomic unit (OTU). Rather than using a single individual or several representative individuals as exemplars, the OTUs are explicitly evolutionary lineages with multiple samples contained within each lineage. Thus, species tree estimation methods model the membership of individuals to evolutionary lineages, in addition to the coalescent model of population processes used in the phylogeny inference. The species tree paradigm enables the relationships among lineages and the membership of individual samples in these lineages to be evaluated in a rigorous statistical framework (Edwards 2009). The species tree approach to species delimitation represents a fundamental shift in how genetic data can be used to delimit lineages (de Queiroz 2007). Rather than equating gene trees with a species tree or basing species status on some genetic threshold (e.g. Baker & Bradley 2006), the relationship between the gene trees and the lineage history is modelled probabilistically using coalescent theory (Hudson 1991; Rosenberg 2002). Adopting this explicit model-based approach also helps researchers to circumvent pernicious problems that result when

genetic thresholds are applied to genetic data—the detection biases that arise from the timing and method of speciation and failure of any threshold to take into account the stochastic variance associated with genetic processes.

By far the most popular of the species tree–based methods for species delimitation is the Bayesian method BPP (Yang & Rannala 2010). It uses a reversible jump Markov chain Monte Carlo approach to calculate the posterior probabilities of competing models that contain greater or fewer lineages. BPP takes sequence data from multiple loci as input and simulates posterior distributions of gene trees and species delimitations from these data. BPP does not consider species tree uncertainty but rather requires users to input a topology (i.e. species tree) to guide the Markov chain. Another method derived from the multispecies coalescent is spedeSTEM (Ence & Carstens 2011), a maximum-likelihood approach that uses STEM (Kubatko *et al.* 2009) to calculate the probability of models that contain differing numbers of evolutionary lineages and then uses information theory (see Burnham & Anderson 2002) to rank these models. SpedeSTEM takes gene trees as input (i.e. a single point estimate per locus) and thus does not consider uncertainty in the gene tree estimates, but does not require the user to input a guide topology. Both BPP and spedeSTEM assume that all of the shared polymorphism is the result of unsorted ancestral polymorphism. If gene flow has led to shared polymorphism across lineages, the underlying species tree is likely to be difficult to infer (Eckert & Carstens 2008), and thus, the accuracy of the species delimitation is likely to decrease (e.g. Camargo *et al.* 2012).

Species delimitation methods can be categorized on the basis of whether sample assignment is required. Cases where the data are input without a priori partitioning have been described as discovery methods and those that require the samples to be partitioned prior to analysis as validation approaches (Ence & Carstens 2011). In this dichotomy, discovery methods can be applied to any system, whereas validation approaches are limited to systems where either populations are clearly delineated (e.g. Camargo *et al.* 2012), where existing subspecific taxonomy can serve as the basis for lineage assignment (e.g. Carstens & Dewey 2010) or where other characters can be used to formulate a hypothesis for delimitation. While all existing validation approaches implement a multispecies coalescent model, some discovery approaches also estimate a species tree (e.g. O'Meara's heuristic method). Many authors have approached species delimitation in more complex systems by using combinations of these approaches to species delimitation, and we highlight several in the next section.

Strategies for complex empirical investigations

Many species delimitation investigations focus on understudied systems where the existing taxonomy is poor, and often these investigations analyse data collected from a single genetic locus as a first pass at estimating species diversity. For example, Esselstyn *et al.* (2012) describe one such system, the *Hipposideros* bats of the Philippines, where the most recent taxonomic work was conducted in 1963. Their approach was to collect data from a single mitochondrial locus and analyse them under the general mixed Yule coalescent (GMYC) model. This model, developed by Pons *et al.* (2006), is related to lineage through time plots, but the GMYC models the transition in a particular gene tree between allelic coalescence and cladogenesis. Once this transition point is identified, individuals can be assigned to species on the basis of whether they coalesce with a given individual before or after this transition point. Esselstyn *et al.* (2012) also conduct a simulation study prior to the analysis of their empirical data, in part to calibrate the prior distributions assumed in the analysis. Their analysis suggests that species diversity is dramatically under-described within Philippine *Hipposideros* bats, and they consider the lineages identified via their application of the GMYC as putative lineages and targets for future research.

In another system where sample assignment to putative lineage was not clear, Leaché & Fujita (2010) investigated forest geckos from Western Africa. They adopted a two-stage approach to species delimitation, using a species discovery approach (in this case,

Structurama) to assign individuals to putative groups and then subsequently validated these groups using BPP. Their approach alleviates substantial shortcomings of these methods when each is used independently, the inability of Structurama to model the pattern of population divergence and the requirement of BPP that samples are assigned to populations prior to the analysis. This pairing of a species discovery method with BPP has been followed by a number of authors (e.g. Barrett & Freudenstein 2011; Setiadi *et al.* 2011). A particularly useful approach may be the pairing of discovery methods that model the species tree to identify putative lineages. For example, Niemiller *et al.* (2012) use the heuristic method described by O'Meara (2010) to jointly estimate the number of species and the species tree topology in a system of cavefish from southeastern North America. O'Meara's method assumes that gene trees will tend to agree on interspecific history (as measured by gene tree parsimony score given a species tree) and disagree within species (as measured by lack of triplet overlap) and discovers putative lineages by minimizing the weighted sum of these across the possible species trees and assignments. Niemiller *et al.* (2012) were able to use this combined approach to delimit multiple lineages of cavefish.

Potential shortcomings of validation approaches

Validation approaches, such as BPP and spedeSTEM, are often given more weight by empirical investigations because they explicitly model the process of lineage diversification. However, there are shortcomings inherent to these methods, so results should be interpreted with caution. For example, spedeSTEM does not estimate θ when it computes the ML species tree (θ is supplied by the user) and assumes that this parameter is unchanging across the phylogeny. In practice, θ influences the absolute probability of the ST rather than the relative probability, and thus, AIC model probabilities are not influenced by imprecision in this parameter. The accuracy of spedeSTEM is dependent on the quality of the gene tree estimates, and while it is straightforward to assess the quality of such estimates (e.g. using bootstrap values or Bayesian posterior probabilities), identification of a single threshold of nodal support below which results from spedeSTEM are suspect requires a simulation study tailored to the specifics of a given system. Conversely, BPP simulates a posterior distribution of gene trees, estimates θ on each branch of the species tree and calculates the posterior probabilities of the species delimitations. One shortcoming of BPP, first illustrated by Leaché & Fujita (2010), is its dependency on the accuracy of the guide tree. BPP requires that the user specifies the guide tree topology (i.e. the

species tree), which is used by the program to guide the reversible jumps in the Markov chain. Leaché & Fujita (2010) conducted simulations to explore the effect of inaccuracies in the guide tree on the species delimitations and demonstrate that if this guide topology is specified incorrectly, BPP is likely to delimit each of the putative lineages (i.e. each OTU of the species tree). This is due to the artificial increase in genetic divergence between sister lineages. This shortcoming can be partially mitigated by directly estimating the topology of the guide tree by grouping samples assigned to populations into OTUs for a species tree analysis. For example, Leaché and Fujita use *Beast (Heled & Drummond 2010) to directly estimate the species tree, which is subsequently used as a guide topology. In their case, they were able to produce a well-supported estimate of the species tree topology when assuming the population assignments produced by Structurama. However, species trees with low levels of divergence are difficult to estimate, particularly when the membership of samples to lineages is unknown (O'Meara 2010), and estimates of the species tree topology are not always clear (e.g. Carstens & Satler 2013). In addition, Bayesian approaches to species tree estimation generate a posterior distribution of species trees, and a consensus tree from this posterior distribution is often used as a guide topology for the subsequent BPP analysis (e.g. Leaché & Fujita 2010). It is important to recognize that there is no guarantee that the consensus tree will share the topology of the maximum-likelihood estimate (MLE) of the species tree (Felsenstein 2004) or that the posterior distribution of a Bayesian analysis will contain the MLE. Leaché & Fujita (2010) caution readers to carefully consider the guide tree in a BPP analysis; if this topology does not reflect the true history of diversification, BPP analyses are prone to falsely delimiting species.

Incongruence across results

Many species delimitation investigations report discordance across results from various methods (Table 2). One explanation might be that the assumptions of one or more methods are violated, leading to an incorrect result. However, the nature of the speciation process leads us to expect incongruence if two conditions are met: first, most researchers have followed Darwin (1859) in considering speciation as a gradual process, with exceptions for some particular modes such as allopolyploid speciation (e.g. Kim *et al.* 2008). If speciation in most cases is gradual, then incongruent results would be expected if the methods applied had differing degrees of statistical power to detect independent lineages. The results of simulation studies such as that by Camargo *et al.* (2012) suggest that this is the case, but until such

studies are conducted for a broad range of empirical systems we cannot determine whether certain methods have better or worse statistical power in all cases or only under specific conditions. Clearly, researchers should proceed by conducting simulation testing and choosing methods that have complementary shortcomings.

For example, spedeSTEM and BPP are complementary validation approaches because they use different strategies for simplifying the parameter space of species delimitation. BPP takes sequences as input and uses genealogical sampling and reversible jump MCMC to evaluate species delimitations given a guide tree. Conversely, spedeSTEM takes previously estimated gene trees as input, calculates the maximum-likelihood species tree for species delimitation and identifies the best of many possible delimitations using information theory. Because spedeSTEM computes the likelihood of species trees (i.e. it does not estimate this parameter) for all possible permutations of the putative lineages, it is robust to phylogenetic error in the species tree. However, it is clear from simulation studies that inaccuracy in spedeSTEM is largely conservative; it fails to delimit what are in reality separate lineages (Ence & Carstens 2011), rather than falsely delimiting the putative lineages as independent, as may occur in BPP if the guide tree is mis-specified. In this way, spedeSTEM and BPP are complementary approaches to species validation because each appears prone to failure in an opposite manner. When both programs are used to validate population assignments and the results are congruent, researchers can reasonably infer that BPP has not been mislead due to an inaccurately designated guide tree and that spedeSTEM has not failed to detect independent lineages due to poorly estimated gene trees. Conversely, incongruence in the results probably indicates that one of these problems is likely evident.

Another potential shortcoming of species delimitation investigation is the limited sample sizes common to many multilocus investigations (see Fujita *et al.* 2012). Before multilocus data became common, phylogeographical investigations often collected data from hundreds of samples (e.g. Avise 2000 and references therein). However, while the need for multilocus data for both phylogeography (e.g. Brumfield *et al.* 2003) and species delimitation (e.g. Dupuis *et al.* 2012) is by now firmly established, the addition of multiple loci to phylogeographical investigations has too often come at the expense of sampling. For the papers reviewed here, there is a negative trend between number of samples and the number of loci, but the correlation is not significant ($r = -0.26$; $P = 0.182$). We attribute this phenomenon to the expense of collecting data using Sanger sequencing and anticipate that recently developed strategies for collecting phylogeographical data using high-

throughput sequencing (e.g. McCormack *et al.* 2013) will enable researchers to collect hundreds of loci from hundreds of individual samples. At an absolute minimum, researchers should collect at least 10 samples from all putative lineages because this number leads to a reasonably high probability (>90%) of sampling the deepest coalescent events in each population (Saunders *et al.* 1994), thus ensuring that most meaningful genetic variation (at least to the question of species boundaries) is sampled. However, optimal sampling levels are highly dependent on both the study system and the analyses used to delimit species—another justification for simulation testing (e.g. Camargo *et al.* 2012; Esselstyn *et al.* 2012).

How to approach species delimitation

A shortcoming of empirical investigations to date is that researchers limit the number and scope of their species delimitation analyses. In 2006, a symposium organized by John Wiens (sponsored by the Society of Systematic Biologists) at the Evolution Annual meeting highlighted the need for new approaches to species delimitation (see Wiens 2007). Since then, dozens of novel methods have been developed (see Box 2). Despite this recent abundance of methods, researchers commonly present results from only a handful (mean 2.25 ± 1.2; Table 1). When viewed in the context of the difference in parameterization between an *ideal method* for species delimitation and existing methods, this is problematic because any existing method is forced to make a series of simplifying assumptions, any one of which could possibly be violated in a particular empirical system. Viewed in this light, incongruence in the results across methods is evidence of the individual shortcomings of one or more of approaches used to delimit species, given the data. By limiting the number and scope of the analyses applied to the data, researchers do not allow for the possibility that there are inherent differences in statistical power across analyses and thus limit their ability to identify incongruence across methods. Conversely, researchers who apply a wide range of species delimitation analyses to their data can have one of two outcomes. First, they may find that all of their results are generally in agreement, which strengthens their confidence in these results and the resulting species delimitations. Second, in cases where there is incongruence across results, researchers may be prompted to explore the possible causes of this incongruence and are likely to be conservative in the inferences drawn from the analysis. Either of these outcomes is preferable to conclusions about species limits that are based on a single or a limited number of analyses and thus more likely to be inaccurate. In our reading of the literature, most workers either state explicitly or assume that the goal of their work is to identify evolutionary lineages, essentially evolutionarily significant units *sensu* Moritz (1994). Species delimitation thus has important implications to the conservation of biodiversity.

To illustrate the importance of basing species delimitations on a conservative consensus across a wide range of methods, consider a recent investigation by Salter *et al.* (2013). They investigated species boundaries in a clade of trapdoor spiders from the west coast of North America using six different approaches to species delimitation (Fig. 1). Results from these approaches delimited between 3 and 18 lineages, and Satler *et al.* interpreted this incongruence in a conservative manner by recognizing three of these lineages as species. It is possible that additional lineages identified by some of the analyses represent species-level diversity, in which case Satler *et al.* failed at species delimitation by being overly conservative in their interpretation of the results. However, each of the methods used by Satler *et al.* contain evolutionary models with inherent simplifications of parameter space, and because these simplifications are nonoverlapping in terms of their assumptions and the parameters estimated from the data, the use of multiple approaches forced Satler *et al.* into reconciling their results into unified conceptual (although not quantitative) evolutionary model that is the basis for the requisite interpretations pertaining to species assignment. This cautious approach to delimitation assumes that failing to delimit species is preferable to falsely delimiting entities that do not represent actual evolutionarily lineages, particularly when the goal of the analysis is species description (Box 3). This approach is consistent with the spirit of integrative taxonomy, which argues that taxonomic inference should be based on congruence across analyses that utilize multiple sources of data (e.g. Padial *et al.* 2010; Schlick-Steiner *et al.* 2010). Species delimitation is perhaps the product of phylogeographical research that has the most tangible implications to the world beyond academia and as a field we should exercise care and caution when drawing taxonomic conclusions from our investigations. At the same time, it is important to have confidence in the data and the methods used to analyse these data and to make taxonomic recommendations in systems where they are warranted.

Integrating genetic and nongenetic sources of data

Inferences regarding species boundaries based on genetic data alone are likely inadequate, and species delimitation should be conducted with consideration of the life history, geographical distribution, morphology

Table 1 Properties of recent species delimitation studies. Shown for each study are the number of assumed species prior to analysis, the number of inferred lineages after the analysis, the number of samples and number of loci used, as well as a list of the discovery and validation approaches used

Reference	Existing species	Species after delimitation	No. of samples	No. of loci	Discovery approaches	Validation approaches
Avila *et al.* (2006)	6	>12	293	1	Statistical parsimony (NCA)	None
Barrett & Freudenstein (2011)	3	3	162	5	Morphological cluster analysis, PCA	BPP
Burbrink *et al.* (2011)	1	1	45	3	Structurama	BPP
Camargo *et al.* (2012)	3	1	505	4	None	spedeSTEM, BPP, ABC
Carstens & Dewey (2010)	3	7	42	6	None	spedeSTEM, Bayes Factors
Carstens & Satler (2013)	1	2	82	21	Structurama, Gaussian Clustering	spedeSTEM, BPP
Duminil *et al.* (2012)	Unknown	Unstated	103	7	Morphometric clustering; structure	None
Esselstyn *et al.* (2012)	13	18–19	413	1	GMYC	None
Florio *et al.* (2012);	1	2	111	1	Canonical variates analysis	None
Flot *et al.* (2010)	1		74	3	Haplowebs	None
Hamilton *et al.* (2011)	4	3	147	1	Combo WP and barcoding gap, monophyly, GMYC	
Kelly *et al.* (2008)	39	1	114	1	WP	None
Leaché & Fujita (2010)	1	3	51	6	Structure	BPP
Leavitt *et al.* (2012)	19	2	414	6	Structure	BPP, mean genetic distance
Leliart *et al.* (2009)	19	13	175	1	GMYC, statistical parsimony (NCA)—clades that exceed 95% cut off	None
Niemiller *et al.* (2012)	1	19	135	9	O'Meara clustering	BPP
Pons *et al.* (2006)	24	54	468	1	Parsimony network, PAA, CHA, WP, GMYC	None
Puillandre *et al.* (2009)	1	4	44	2	Elliptic Fourier analysis on shape to the mollusc shell; qualitative phylogenetic evidence	None
Puillandre *et al.* (2012)	43	27	1000	2	GMYC, ABGD	None
Rielly *et al.* (2012)	1	3	50	14	Structurama	Monophyly
Rosell *et al.* (2010)	23	Unstated	83	4	Discriminant analysis on morphological data; qualitative phylogenetic analysis	None
Salter *et al.* (2013)	1	2	142	6	Structure, Brownie heuristic, bGMYC, step up STEM	spedeSTEM, BPP
Setiadi *et al.* (2011)	4	11	683	2	Genealogical-concordance criterion	BPP
Stech *et al.* (2013)	9	9	70	2	None	Comparison with morphology
Stielow *et al.* (2011)	Unknown	24	142	1	Qualitative phylogenetic analysis	None
Weisrock *et al.* (2010)	Unknown	16	216	5	Structure; gsi; phylogeny estimation	None
Welton *et al.* (2013)	6	8	53	5	Structure	BPP
Zhou *et al.* (2012)	3	1	394	6	None	BPP

and behaviour (where applicable) of the focal system (Knowles & Carstens 2007; Schlick-Steiner *et al.* 2010). There are several approaches for this integration. Some researchers interpret phenotypical variation in the context of delimitations from genetic data (e.g. Setiadi *et al.* 2011; Esselstyn *et al.* 2012; Stech *et al.* 2013) as a way to buttress the results of delimitation analysis. In some systems, morphological differences serve as the basis for taxonomic hypotheses that are validated using genetic data (e.g. Carstens & Dewey 2010; Welton *et al.* 2013). When results from morphological and genetic data are incongruent, it is reasonable to exercise cau-
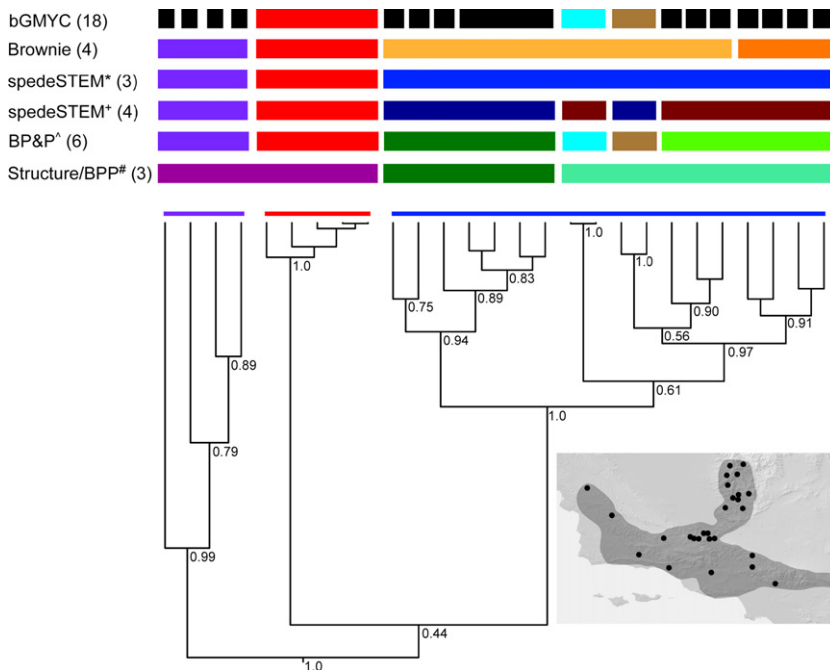
Fig. 1 Incongruent results from an empirical system. Six loci were used to delimit lineages in *Aliatypus* species complex, a group of trapdoor spiders from southern California (localities shown by dots in the inset map). Results from six methods (named at left) are depicted, with bars coloured to highlight congruence across methods. Numbers in parentheses represent the number of lineages delimited using each method. Also shown is an estimate of the species tree made using *Beast, with posterior probabilities shown under each node. Immediately above the species tree, thin coloured lines are used to identify the three species described by Satler *et al.* This figure is redrawn from Salter *et al.* (2013).

Box 3

Species concepts and species delimitation

Many recent investigations into species limits either explicitly state or imply that a primary goal of the research is to identify independent evolutionary lineages, which are equivalent to metapopulation lineages through time (Simpson 1951; Wiley 1978; de Queiroz 2005) and commonly interpreted as evolutionarily significant units (Moritz 1994). Thus, the goal of most investigations that collect genetic data and utilize recently developed methods is to discover or validate units of biodiversity with their own unique evolutionary history. Much of this work is enabled by Kevin de Queiroz's reformulation of species concepts (de 1998; de Queiroz 2005, 2007). Prior to this, species delimitation was directly tied to species concepts, to the extent that one researcher was required to assume a species concept and collect data that could directly speak to the criterion used to delimit species given this concept. Thus, morphological differentiation was required to delimit species under the morphological species concept, allelic coalescence was a requirement under the genealogical species concept (Baum & Shaw 1995), and reproductive isolation was a requirement under the biological species concept (Mayr 1942). While this made species delimitation less ambiguous, it also produced a long-standing argument that lead many researchers to ignore species delimitation as a potential research goal because such an exercise required the adoption of one of a number of competing species concepts and thus forced the researcher to what had become an interminable debate.

de Queiroz offered a practical solution to this dilemma by redefining the criteria inherent to species concepts (de 1998; de Queiroz 2005, 2007). Rather than treating criteria such as morphological differences, monophyly or reproductive isolation as *the* single indicator of species-level differentiation, de Queiroz argued that *each* of these is a property of the evolutionary divergence of lineages. Because all species concepts (at least in de Queiroz's view) assume that species represent independent metapopulation lineages through time, the solution to the species problem is to treat the traditional criteria used to demarcate species as attributes that accumulate during the process of lineage diversification (de Queiroz 2005). This generalized lineage concept (GLC) has been broadly adopted by recent investigations into species limits (Table 2) and indirectly promoted the development of recent approaches to species delimitation (e.g. Knowles & Carstens 2007). However, it is a mistake to characterize the entire field as adopting the GLC or to assume that the GLC is a prerequisite for species delimitation (e.g. Rosell *et al.* 2010; Barrett & Freudenstein 2011; Duminil *et al.* 2012). However, it is good practice to define some species concept when reporting investigation that includes a substantial species delimitation component, if only because this articulation enforces the need for a clear argument regarding the criteria used to recognize species.

Table 2 Strategies adopted by recent species delimitation studies. For the studies cited in Table 1, we show the species concept and geographical context of the investigation and report if/how nongenetic data were used and how the authors handled incongruence across results, and the taxonomic implications of their study

Reference	Species concept	Geographical context	Nongenetic data used?	Interpretation of incongruence	Taxonomic treatment
Avila et al. (2006)	Inertial	Sympatry/allopatry	No	n/a	None
Barrett & Freudenstein (2011)	PSC	Sympatry/allopatry	Morphological	Conservative consensus	None
Burbrink et al. (2011)	GLC	Allopatry	ENM	Consensus	Elevate subspecies
Camargo et al. (2012)	None	Allopatry	No	Results congruent	None
Carstens & Dewey (2010)	None	Sympatry/allopatry	Morphology as basis for validation	Favour one (spedeSTEM)	None
Carstens & Satler (2013)	GLC	Allopatry	No	Conservative - delimitations shared across results	Recommend new species
Duminil et al. (2012)	Morphological implied	Partly	Morphological	None offered	None
Esselstyn et al. (2012)	None	Sympatry/allopatry	Behaviour, echolocation	Attribute lack of congruence to small sample sizes.	None
Florio et al. (2012)	GLC	Allopatry (w/ contact zone)	Morphology and ENM	Results congruent	New species described
Flot et al. (2010)	GLC	Unclear	No	n/a	Recommend new species
Hamilton et al. (2011)	None	Sympatry/allopatry	Geographical	Consensus - conservative	None
Kelly et al. (2008)	None	Sympatry/allopatry?	No	Results congruent	None
Leaché & Fujita (2010)	GLC	Allopatry	No	Structure as basis for BPP	New species described
Leavitt et al. (2012)	None	Not clear	No	Existing taxonomy guide genetic analyses	None
Leliart et al. (2009)	Phylogenetic	Mostly allopatric	No	Consensus - conservative	None
Niemiller et al. (2012)	GLC	Allopatry	No	Conservative	None
Pons et al. (2006)	No	Sympatry/allopatry	Morphology, geography	Integrate PAA and GMYC	None
Puillandre et al. (2009)	Morphological	Allopatry	Morphology	n/a	None
Puillandre et al. (2012)	None	Sympatry/allopatry	Bathymetric data, morphology	Klee diagrams, comparison with morphology	None
Rielly et al. (2012)	GLC implied	Parapatry	No	General congruence across results	None
Rosell et al. (2010)	Phylogenetic	Allopatry	Yes	Suggested convergence in morphology	No
Salter et al. (2013)	GLC	Allopatry	No	Conservative consensus	New species described
Setiadi et al. (2011)	GLC	Sympatry/allopatry	Morphology	Favour morphology - do not use BPP results	None
Stech et al. (2013)	None	Sympatry/allopatry	Morphology	Reassign samples to species based on genetic data	None
Stielow et al. (2011)	None	Unclear	Morphology	Authors favour results from morphology	New species described
Weisrock et al. (2010)	GLC	Allopatry	No	Invoke incompletely lineage sorting	None
Welton et al. (2013)	None	Mostly allopatric	Morphology and geography as basis for validation	Express caution in areas of incongruence	Recommend new species
Zhou et al. (2012)	None	Allopatry	ENM	None	None

PSC: Phylogenetic species concept; GLC: General lineage concept; ENM: Ecological niche models.

tion in species delimitation (e.g. Leliart et al. 2009; Barrett & Freudenstein 2011) while allowing for the possibility of morphologically cryptic species (e.g. Salter et al. 2013). While ecological niche modelling (i.e. Peterson 2001; Hugall et al. 2002) is generally underutilized in species delimitation, several researchers have com-

bined these approaches. This is particularly appealing because the environmental modelling enables an assessment of the environmental differentiation between putative lineages (e.g. Florio *et al.* 2012; Zhou *et al.* 2012). For example, Zhou *et al.* (2012) demonstrated that the major lineages of *Rana* frogs in central China were environmentally differentiated, supporting their inference of a cryptic species in this clade. Species distribution models can also provide valuable information regarding historical changes in species range, and preliminary results from our laboratory indicate that hindcasted species distribution models are a better predictor of species boundaries in *Plethodon* salamanders than their current ranges demonstrate (TA Pelletier, C Crisafulli, S Wagner, AJ Zellmer, BC Carstens in preparation). Inferences about species limits are best made using an approach that integrates across many data types and analyses.

Do researchers trust in the results of their species delimitation analyses?

Fewer than 30% of the studies reviewed here made taxonomic recommendations and only 25% describe new species. This could indicate a lack of confidence in the results of the species delimitation analyses (perhaps caused by a lack of adequate training in taxonomy), an inability to reconcile incongruence across methods, or it could be an implicit admission of the inadequacy of the data. These reluctances are exacerbated by selection pressures imposed by the academic system; species descriptions are not always viewed as equivalent to research papers, and it is optimal for researchers at the early stages of their career to publish in general interest journals such as Molecular Ecology rather than taxon-focused journals. Nevertheless, formal species descriptions have a lasting impact that will ultimately exceed that of all but the most highly cited papers in general interest journals, and we encourage researchers to generate these descriptions.

In our review of the literature, it also because clear that many researchers conduct species delimitation investigation without a clear statement of their goals and without defining the relevance of the identification of independent evolutionary lineages to their system. Some studies are apparently designed to document current species diversity, while others seek to identify independent lineages in order to better understand evolutionary processes. It is imperative on researchers to clearly state the goals of their investigations and incumbent on reviewers to demand that such statements are present. Such statements can also establish the appropriate level of conservativeness in delimitating species. For example, studies that seek to conduct taxonomic

revisions and describe new species are less justified in favouring the results of a method such as the GMYC that tends to delimit a larger number of lineages than studies that seek to use the delimited lineages in downstream ecological analyses (e.g. Stevens *et al.* 2012).

Conclusions

Species delimitation is a vital enterprise within evolutionary biology; it bridges the historically independent disciplines of phylogenetics and population genetics and identifies the point when population-level processes begin to produce phylogenetic patterns. Researchers have never had a better collection of methodological tools for delimiting species, and high-throughput sequencing enables extremely large amounts of data to be collected from nearly any empirical system. The appropriate way to conduct a species delimitation investigation is to analyse these data with a wide variety of methods and to delimit lineages that are consistent across the results. If the discipline of phylogeography is to do our part to document and protect biodiversity, it is incumbent on researchers to design and conduct effective investigations into species boundaries (Fujita *et al.* 2012), to trust in their data and results and to follow through on these results by proposing the requisite taxonomic revisions.

References

Avila LJ, Morando M, Sites JW Jr (2006) Congeneric phylogeography: hypothesizing species limits and evolutionary processes in Patagonian lizards of the *Liolaemus boulengeri* group (Squamata: Liolaemini). *Biological Journal of the Linnean Society*, **89**, 241–275.

Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, Massachusetts.

Baker RJ, Bradley RD (2006) Speciation in mammals and the genetic species concept. *Journal of Mammalogy*, **87**, 643–662.

Barrett CF, Freudenstein JV (2011) An integrative approach to delimiting species in a rare but widespread mycoheterotrophic orchid. *Molecular Ecology*, **20**, 2771–2786.

Baum DA, Shaw KL (1995) Genealogical perspectives on the species problem. In: *Experimental and Molecular Approaches to Plant Biosystematics* (eds Hoch PC, Stephenson AG), pp. 289–303. Monographs in Systematic Botany from the Missouri Botanical Garden, St Louis, Missouri.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective populations sizes in n subpopulations by using a coalescent approach. *Proceedings National Academy Science USA*, **98**, 4563–4568.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology and Evolution*, **18**, 249–256.

Burbrink FT, Yao H, Ingrasci M, Bryson RW Jr, Guiher TJ, Ruane S (2011) Speciation at the Mogollon Rim in the Arizona Mountain Kingsnake (*Lampropeltis pyromelana*). *Molecular Phylogenetics and Evolution*, **60**, 445–454.

Burnham KP, Anderson DA (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd edn, 488 pp. Springer-Verlag, New York.

Camargo A, Morando M, Avila LJ, Sites JW (2012) Species delimitation with ABC and other coalescent-based methods: a test of accuracy with simulations and an empirical example with lizards of the *Liolaemus darwinii* complex (Squamata: Liolaemidae). *Evolution*, **66**, 2834–2849.

Carstens BC, Dewey TA (2010) Species delimitation using a combined coalescent and information-theoretic approach: an example from North American *Myotis* bats. *Systematic Biology*, **59**, 400–414.

Carstens BC, Satler JD (2013) The carnivorous plant described as *Sarracenia alata* contains two cryptic species. *Biological Journal of the Linnean Society*, doi:10.1111/bij.12093. in press.

Choi SC, Hey J (2011) Joint inference of population assignment and Demographic history. *Genetics*, **189**, 561–577.

Csilléry K, Blum MGB, Gaggiotti OE, Francois O (2011) Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology and Evolution*, **25**, 410–418.

Darwin C (1859) *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray Press, London, UK.

de Queiroz K (1998) The general lineage concept of species, species criteria, and the process of speciation. In: *Endless Forms: Species and Speciation* (eds. Howard SJ, Berlocher SH), pp. 57–78. Oxford University Press, New York.

Degnan JH, Rosenberg NA (2006) Discordance of species trees with their most likely gene trees. *PLoS Genetics* **2 e68**, 762–768.

Duminil J, Kenfack D, Viscosi V, Grumiau L, Hardy OJ (2012) Testing species delimitation in sympatric species complexes: the case of an African tropical tree, *Carapa* spp. (Meliaceae). *Molecular Phylogenetics and Evolution*, **62**, 275–285.

Dupuis JR, Roe AD, Sperling FA (2012) Multi-locus species delimitation in closely related animals and fungi: one marker is not enough. *Molecular Ecology* **21**, 4422–4436.

Eckert AJ, Carstens BC (2008) Does gene flow destroy phylogenetic signal? The performance of three methods for estimating species phylogenies in the presence of gene flow. *Molecular Phylogenetics & Evolution*, **49**, 832–842.

Edwards SV (2009) Is a new and general theory of molecular systematics emerging? *Evolution*, **63**, 1–19.

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

Esselstyn JA, Evans BJ, Sedlock JL, Anwarali Khan FA, Heaney LR (2012) Single-locus species delimitation: a test of the mixed Yule-coalescent model, with an empirical application to Philippine round-leaf bats. *Proceedings of the Royal Society of London B*, **279**, 3678–3686.

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

Florio AM, Ingram CM, Rakotondravony HA, Louis EE Jr, Raxworthy CJ (2012) Detecting cryptic speciation in the widespread and morphologically conservative carpet chameleon (*Furcifer lateralis*) of Madagascar. *Journal of Evolutionary Biology*, **25**, 1399–1414.

Flot J-F, Couloux A, Tillier S (2010) Haplowebs as a graphical tool for delimiting species: a revival of Doyle's "field for recombination" approach and its application to the coral genus Pocillopora in Clipperton. *BMC Evolutionary Biology*, **10**, 372.

Fraley C, Raftery AE.(2006). MCLUST Version 3 for R: Normal mixture modeling and model-based clustering. Technical Report no. 504, Department of Statistics, University of Washington.

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

Guillot G, Mortier F, Estoup A (2005) Geneland: a program for landscape genetics. *Molecular Ecology Notes*, **5**, 712–715.

Guillot G, Renaud S, Ledevin R, Michaux J, Claude J (2012) A unifying model for the analysis of phenotypic, genetic, and geographic data. *Systematic Biology*, **61**, 897–911.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, **5**, 1–11.

Hamilton CA, Formanowicz DR, Bond JE (2011) Species delimitation and phylogeography of *Aphonopelma hentzi* (Araneae, Mygalomorphae, Theraphosidae): cryptic diversity in North American Tarantulas. *PLoS ONE*, **6**, e26207.

Hausdorf B, Hennig C (2010) Species delimitation using dominant and codominant multilocus markers. *Systematic Biology*, **59**, 491–503.

Heled J, Drummond AJ (2008) Bayesian inference of population size history from a multiple loci. *BMC Evolutionary Biology*, **8**, 289.

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

Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, **167**, 747–760.

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences USA*, **104**, 2785–2790.

Hudson RR (1991) Gene genealogies and the coalescent process. In: *Oxford Surveys in Evolutionary Biology* (eds Futuyma D & Antonovics J), pp. 1–44. Oxford University Press, New York.

Huelsenbeck JP, Andolfatto P (2007) Inference of population structure under a Dirichlet process model. *Genetics*, **175**, 1787–1802.

Huelsenbeck JP, Andolfatto P, Huelsenbeck ET (2011) Structurama: Bayesian inference of population structure. *Evolutionary Bioinformatics*, **7**, 55–59.

Hugall A, Moritz C, Moussalii A, Stanisic J (2002) Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail Arosophia bellendenkerensis (Brazier 1875). *Proceedings of the National Academy of Sciences USA*, **99**, 6112–6117.

Hull DL (1998) *Science as a Process*. University of Chicago Press, Chicago, Illinois.

Jackson ND, Austin CC (2010) The combined effects of rivers and refugia generate extreme cryptic fragmentation within the common ground skink (*Scincella lateralis*). *Evolution*, **64**, 409–428.

Kalinowski ST (2010) The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population genetic structure. *Heredity*, **106**, 625–632.

Kelly CMR, Barker NP, Villet MH, Broadley DG, Branch WR (2008) The snake family Psammophiidae (Reptilia: Serpentes): phylogenetics and species delimitation in the African sand snakes (*Psammophis Boie*, 1925) and allied genera. *Molecular Phylogenetics and Evolution*, **47**, 1045–1060.

Kim S-T, Sultan SE, Donoghue MJ (2008) Allopolyploid speciation in Persicaria (Polygonaceae): insights from a low-copy nuclear region. *Proceedings of the National Academy of Sciences USA*, **105**, 12370–12375.

Kingman JFC (1982) One the genealogy of large populations. *Journal of Applied Probability*, **19**, 27–43.

Knowles LL, Carstens BC (2007) Delimiting species without monophyletic gene trees. *Systematic Biology*, **56**, 887–895.

Kubatko LS, Carstens BC, Knowles LL (2009) STEM: species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics*, **25**, 971–973.

Kuhner MK, Yamato J, Felsenstein J (1998) Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics*, **149**, 429–434.

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

Leavitt SD, Johnson LA, Goward T, Clair LL (2012) Species delimitation in taxonomically difficult lichen-forming fungi: an example from morphologically and chemically diverse Xanthoparmelia (Parmeliaceae) in North America. *Molecular Phylogenetics and Evolution*, **60**, 317–332.

Leliart F, Verbruggen H, Wysor B, De Clerk O (2009) DNA taxonomy in morphologically plastic taxa: Algorithmic species delimitation in the *Boodlea* complex (Chlorophyta: Cladophorales). *Molecular Phylogentics and Evolution*, **53**, 122–133.

Liu L, Pearl DK, Brumfield RT, Edwards SV (2007) Estimating species trees using multiple-allele DNA sequence data. *Evolution*, **62**, 2080–2091.

Maddison WP (1997) Gene trees in species trees. *Systematic Biology*, **46**, 523–536.

Mayr E (1942) *Systematics and the Origin of Species, from the Viewpoint of a Zoologist (No. 13)*. Harvard University Press, Cambridge, Massachusetts.

Maddison WP, Knowles LL (2006) Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology*, **55**, 21–30.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, **66**, 526–538.

Moritz C (1994) Defining "Evolutionarily Significant Units" for conservation. *Trends in Ecology and Evolution*, **9**, 373–375.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov Chain Monte Carlo approach. *Genetics*, **158**, 885–896.

Niemiller ML, Near TJ, Fitzpatrick BM (2012) Delimiting species using multilocus data: diagnosing cryptic diversity in the southern cavefish *Typhlichthys subterraneus* (Teleostei: Amblyopsidae). *Evolution*, **66**, 846–866.

O'Dushlaine C, McQuillan R, Weale M *et al.* (2010) Genes predict village of origin in rural Europe. *European. Journal of Human Genetics* **18**, 1269–1270.

O'Meara BC (2008) Using Trees: Myrmecocystus phylogeny and character evolution and new methods for investigating trait evolution and species delimitation. PhD Dissertation, Nature Proceedings http://dx.doi.org/10.1038/npre.2008.2261.1

O'Meara BC (2010) New heuristic methods for joint species delimitation and species tree inference. *Systematic Biology*, **59**, 59–73.

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

Peterson AT (2001) Predicting species geographic distributions based on ecological niche modeling. *Condor*, **103**, 599–605.

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

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Puillandre N, Baylac M, Boisselier M-C, Cruaud C, Samadi S (2009) An integrative approach to species delimitation in Benthomangelia (Mollusca: Conoidea). *Biological Journal Linnean Society*, **96**, 696–708.

Puillandre N, Modica MV, Zhang Y *et al.* (2012) Large-scale species delimitation method for hyperdiverse groups. *Molecular Ecology*, **21**, 2671–2691.

de Queiroz K (2005) Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences USA*, **102**, 6600–6607.

de Queiroz K (2007) Species concepts and species delimitation. *Systematic Biology*, **56**, 879–886.

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.

Rannala B, Yang Z (2013) Improved reversible jump algorithms for Bayesian Species delimitation. *Genetics*, **194**, 245–253.

Reid NM, Carstens BC (2012) Phylogenetic estimation error can decrease the accuracy species delimitation: a Bayesian implementation of the General Mixed Yule Coalescent model. *BMC Evolutionary Biology*, **12**, 196.

Rielly SB, Marks SB, Jennings WB (2012) Defining evolutionary boundaries across parapatric ecomorphs of Black salamanders (*Aneides flavipunctatus*) with conservation implications. *Molecular Ecology*, **21**, 5745–5761.

Rittmeyer EN, Austin CC (2012) The effects of sampling on delimiting species from multi- locus sequence data. *Molecular Phylogenetics and Evolution*, **65**, 451–463.

Rosell JA, Olson ME, Weeks A *et al.* (2010) Diversification in species complexes: tests of species origin and delimitation in the *Bursera simaruba* clade of tropical trees (Burseraceae). *Molecular Phylogenetics and Evolution*, **57**, 798–811.

Rosenberg NA (2002) The probability of topological concordance of gene trees and species trees. *Theoretical Population Biology*, **61**, 225–247.

Salter JD, Carstens BC, Hedin M (2013) Multilocus species delimitation in a complex of morphologically conserved trapdoor spiders (Mygalomorphae, Antrodiaetidae, *Aliatypus*). *Systematic Biology*, in press.

Saunders IW, Tavare S, Watterson GA (1994) On the genealogy of nested subsamples from a haploid population. *Advances in Applied Probability*, **16**, 471–491.

Schlick-Steiner BC, Steiner FM, Seifert B, Stauffer C, Christian E, Crozier RH (2010) Integrative Taxonomy: a multisource approach to exploring biodiversity. *Annual Reviews of Entomology*, **55**, 421–438.

Setiadi MI, McGuire JA, Brown RM *et al.* (2011) Adaptive radiation and ecological opportunity in Sulawesi and Philippine Fanged Frog (*Limnonectes*) Communities. *American Naturalist*, **178**, 221–240.

Simpson GG (1951) The species concept. *Evolution*, **5**, 285–298.

Stech M, Veldman S, Larrain J *et al.* (2013) Molecular species delimitation in the *Racomitrium canescens* complex (Grimmiaceae) and implications for DNA barcoding of species complexes in Mosses. *PLoS ONE*, **8**, e15614.

Stevens RD, Gavilanez MM, Tello JS, Ray DA (2012) Phylogenetic structure illuminates the mechanistic role of environmental heterogeneity in community organization. *Journal of Animal Ecology*, **81**, 455–462.

Stielow B, Bratek Z, Orczan AKI *et al.* (2011) Species delimitation in taxonomically difficult Fungi: the case of Hymenogaster. *PLoS ONE*, **6**, e15614.

Weisrock DW, Rasoloarison RM, Fiorentino I *et al.* (2010) Delimiting species without nuclear monophyly in Madagascar's Mouse lemurs. *PLoS ONE*, **5**, e9883.

Welton L, Siler C, Oaks J, Diesmos A, Brown R (2013) Multilocus phylogeny and Bayes estimates of species boundaries reveal hidden evolutionary relationships and cryptic diversity in Southeast Asian monitor lizards. *Molecular Ecology*, in press.

Wiens JJ (2007) Species delimitation: new approaches for discovering diversity. *Systematic Biology*, **56**, 875–878.

Wiens JJ, Penkrot TA (2002) Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (Sceloporus). *Systematic Biology*, **51**, 69–91.

Wiley EO (1978) The evolutionary species concept reconsidered. *Systematic Zoology*, **27**, 17–26.

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

Zhou W-W, Wen Y, Fu J *et al.* (2012) Speciation in the *Rana chensinensis* species complex and its relationship to the uplift of the Qinghai–Tibetan Plateau. *Molecular Ecology*, **21**, 960–973.