



Bayesian species delimitation combining multiple genes and traits in a unified framework

Claudia Solís-Lemus,¹ L. Lacey Knowles,² and Cécile Ané^{1,3,4}

¹Department of Statistics, University of Wisconsin, 1300 University Avenue, Madison, Wisconsin 53706

²Department of Ecology and Evolutionary Biology, Museum of Zoology, University of Michigan, 1109 Geddes Ave, Ann Arbor, Michigan 48109

³Department of Botany, University of Wisconsin, 430 Lincoln Drive, Madison, Wisconsin 53706

⁴E-mail: ane@stat.wisc.edu

Received April 17, 2014

Accepted November 17, 2014

Delimitation of species based exclusively on genetic data has been advocated despite a critical knowledge gap: how might such approaches fail because they rely on genetic data alone, and would their accuracy be improved by using multiple data types. We provide here the requisite framework for addressing these key questions. Because both phenotypic and molecular data can be analyzed in a common Bayesian framework with our program iBPP, we can compare the accuracy of delimited taxa based on genetic data alone versus when integrated with phenotypic data. We can also evaluate how the integration of phenotypic data might improve species delimitation when divergence occurs with gene flow and/or is selectively driven. These two realities of the speciation process are ignored by currently available genetic approaches. Our model accommodates phenotypic characters that exhibit different degrees of divergence, allowing for both neutral traits and traits under selection. We found a greater accuracy of estimated species boundaries with the integration of phenotypic and genetic data, with a strong beneficial influence of phenotypic data from traits under selection when the speciation process involves gene flow. Our results highlight the benefits of multiple data types, but also draws into question the rationale of species delimitation based exclusively on genetic data.

KEY WORDS: Bayesian phylogenetic analysis, BPP, Brownian motion, gene flow, speciation, species delineation.

With the increased availability of multilocus datasets, genetic approaches are becoming a primary focus for delimiting species, as evidenced by the recent proliferation of methods (e.g., Knowles and Carstens 2007; Hausdorf and Hennig 2010; O'Meara 2010; Yang and Rannala 2010; Ence and Carstens 2011; Camargo et al. 2012). Emboldened by evidence from genetic simulations on the potential power of these model-based approaches, empirical applications that rely exclusively on genetic data to estimate species boundaries are becoming common (e.g., Pons et al. 2006; Leaché and Fujita 2010; Burbrink et al. 2011; Niemiller et al. 2012; Satler et al. 2013), in addition to DNA barcoding initiatives not considered here. Yet, despite studies that advocate for particular model-based approaches (e.g., Fujita et al. 2012; Carstens et al. 2013) or evaluate the performance of key steps in the

delimitation of taxa with genetic data (e.g., Olave et al. 2013; Rannala and Yang 2013), a critical issue in such systematic endeavors remains unknown. Specifically, would consideration of multiple data types in empirical studies provide more accurate estimates of species boundaries compared to relying exclusively on genetic data alone?

The theoretical ideals of some genetic model based approaches for species delimitation do not guarantee that these approaches will perform ideally in practice. With the limited expected divergence between newly derived species, species boundaries may go undetected because of insufficient genetic data (e.g., Camargo et al. 2012; Rittmeyer and Austin 2012; Olave et al. 2013). Likewise, the most relevant data associated with reproductive isolation will not be represented among neutral loci in

the case of selectively driven species divergence. The available approaches for delimiting species (see Carstens et al. 2013) will not use the very data that might prove to be the most informative. In these cases, failure to consider more than one data type may compromise the power to detect species boundaries. Failure to consider more than one data type may also exaggerate misleading inferences caused by model violations because gene-based approaches make strong assumptions. For example, neutral loci may be subject to gene flow across species boundaries with low levels of migration, in contrast to the traits that may distinguish taxa (e.g., particular morphological or ecological traits), causing genetic approaches to underestimate the number of taxa (Zhang et al. 2011). This means that genetic-based approaches do not effectively handle some of the biological realities of species divergence (e.g., divergence with gene flow, of selectively driven divergence). This includes all the fairly sophisticated methods recently developed (e.g., coalescent-based Bayesian delimitation, Yang and Rannala 2010).

Despite suggestions of seeking consensus across methods as a way to guard against errors in the delimitation of taxa with genetic data (Carstens et al. 2013), these genetic-based approaches share many assumptions, even though the details of particular methods differ (reviewed in Carstens et al. 2013). The current methods, for example, make very simplifying and implicit assumptions about the divergence process, like assuming that the unit delimited from neutral genetic variation corresponds to a species, or that divergence occurs without gene flow. As such, advocating for a consensus approach does not overcome the inherent limitations of relying on a single data type. Here, we propose a new method that explicitly addresses some of these concerns. Specifically, we present a Bayesian approach for the integration of multiple data types into a single model-based framework.

We acknowledge that incorporating information from diverse sources of data for inferring species boundaries has a long systematic tradition. However, the analysis of multiple data types has been constrained by the availability of methods that can actually integrate information from the different sources (Guillot et al. 2012; Edwards and Knowles 2014). Estimates of species boundaries from multiple data types typically rely upon a series of sequential analyses of each separate data type (e.g., Kergoat et al. 2011; Yeates et al. 2011). The goal of our work is to develop a species delimitation method to combine genetic and trait data into a common framework based on an explicit model of evolution. Specifically, we extend the Bayesian program BPP (Bayesian phylogenetics and phylogeography, Yang and Rannala 2010) to combine genetic and quantitative trait data in a single Bayesian framework, which we call iBPP (integrated BPP). Through simulations, as well as application to an empirical dataset of recently diverged Australian lizards (Edwards and Knowles 2014), we evaluate the general performance of the iBPP method, including

how the addition of trait data may improve the accuracy of estimated species boundaries when divergence occurs with gene flow.

NEW APPROACH

The iBPP method uses the same Bayesian framework as BPP for modeling molecular data across multiple genes. Among gene-based methods, BPP was shown to be superior to others in the simulation conditions in Camargo et al. (2012). As with BPP, iBPP assumes a known assignment of individuals to putative species and a guide tree of these putative species. Alternative hypotheses about delimited taxa are generated by collapsing one or more nodes in the guide tree, each of which is given an equal prior probability (but see Rannala and Yang 2013). The posterior probability (PP) of a particular number of taxa, as specified by a given species tree and its branch lengths, is calculated from independent gene trees according to the multispecies coalescent (Rannala and Yang 2003).

Our iBPP method, in addition to independent loci, can also take independent quantitative continuous traits, to evaluate the PP of a particular number of taxa. Each trait is assumed to have a normal distribution, with species means governed by a Brownian motion (BM) process along the species tree and individuals normally distributed around the species means. A parameter λ models the between-to-within species variance ratio, such that the expected variance of species means from the Brownian motion is $\lambda\sigma^2$ and the within-species variance is $(1 - \lambda)\sigma^2$ (Lynch 1991; Freckleton et al. 2002). Each trait is allowed to have its own λ parameter to model selection acting differently on different traits. Indeed, selection pressure is expected to reduce trait variation between individuals in the same species. A uniform prior is placed on λ , which is integrated using Markov chain Monte Carlo (MCMC). A conjugate prior distribution is placed on the trait mean at the root of the species tree and on the total variance σ^2 , which allows for analytical integration and avoidance of MCMC iterations for these parameters. Details are provided in the Materials and Methods section. The iBPP program and its companion simulation program are available open-source at <https://github.com/cecileane/iBPP/>.

Materials and Methods

METHODS FOR QUANTITATIVE TRAITS

Evolutionary model

Each quantitative trait is assumed to display between-species variation and within-species variation. We assume a Brownian motion model for the evolution of species means along the current species tree, with variance rate $\lambda\sigma^2$ on the tree normalized to have height 1. We then assume normally distributed within-species variation with variance $(1 - \lambda)\sigma^2$. With this model, λ measures the phylogenetic signal, as the ratio of the variation due to the phylogeny to the total variation expected from the root of the tree to individuals

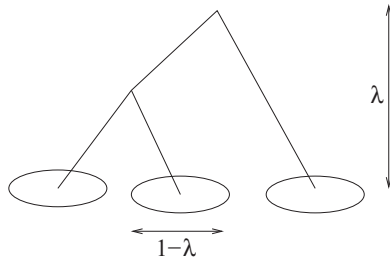


Figure 1. Model for trait variation, with Brownian motion along the species phylogeny for species means, and Gaussian within-species variation. The phylogenetic signal λ represents the expected proportion of between-species variation.

within species (see Fig. 1). In other words, the total expected variance is divided into a proportion λ between species, and $(1 - \lambda)$ within species (Lynch 1991; Pagel 1999).

More formally, this model can be written as $Y \sim \mathcal{N}(\mu, \sigma^2 \mathbf{V}_\lambda)$, where the covariance matrix \mathbf{V}_λ is determined by the species tree and the λ parameter of the particular trait. This matrix is obtained as the standard Brownian motion covariance matrix on a tree where each species is represented by a polytomy with k tips for k individuals. Each branch leading to an individual is assigned length $1 - \lambda$ to model within-species variation, whereas the initial species tree (with one tip per species) is scaled to a total height of λ to model between-species variation. Our model allows each trait to have its own parameters μ , σ^2 , and λ . Intuitively, a trait under selection is expected to show a high λ (large between-species variation compared to within-species variation), whereas a trait under neutral evolution is expected to show a low or intermediate λ .

Priors and marginal likelihood

In our Bayesian framework, we use a conjugate prior distribution for the ancestral species mean μ and the total variance σ^2 . This allows us to integrate these parameters out and to avoid the computational burden of tracking them in the MCMC algorithm. Namely, we use an inverse chi-squared distribution for σ^2 with prior variance (scale) σ_0^2 and ν_0 degrees of freedom (akin to a prior number of observations for the prior variance). For the ancestral species mean μ given σ^2 , we use a normal prior distribution with mean μ_0 and variance $(\mathbf{1}' \mathbf{V}_\lambda^{-1} \mathbf{1})^{-1} \sigma^2 n / \kappa_0$, where n is the total number of individuals for the trait of interest, prime denotes transposition, and $\mathbf{1}$ is a vector of ones. The hyperparameter κ_0 represents the weight for the prior mean μ_0 , or prior number of individual observations. This variance was chosen to be proportional to the sampling variance of the maximum likelihood estimator $\hat{\mu} = (\mathbf{1}' \mathbf{V}_\lambda^{-1} \mathbf{1})^{-1} \mathbf{1}' \mathbf{V}_\lambda^{-1} \mathbf{Y}$.

The prior distribution on (μ, σ^2) is then conjugate, in that the posterior distribution of (μ, σ^2) given the trait data has the same form. Namely, the posterior distribution of μ given σ^2 is

Gaussian with mean $\mu_n = (\kappa_0 \mu_0 + n \hat{\mu}) / (\kappa_0 + n)$ and variance $(\mathbf{1}' \mathbf{V}_\lambda^{-1} \mathbf{1})^{-1} \sigma^2 n / \kappa_n$, where $\kappa_n = \kappa_0 + n$. Also, the posterior distribution of σ^2 is inverse chi-squared with $\nu_n = \nu_0 + n$ degrees of freedom and scale $\sigma_n^2 = \text{postSS} / \nu_n$. The posterior sum of squares is defined by

$$\text{postSS} = \nu_0 \sigma_0^2 + (\mathbf{Y} - \hat{\mu} \mathbf{1})' \mathbf{V}_\lambda^{-1} (\mathbf{Y} - \hat{\mu} \mathbf{1}) + (\mu_0 - \hat{\mu})^2 \frac{\kappa_0}{\kappa_n} \mathbf{1}' \mathbf{V}_\lambda^{-1} \mathbf{1}.$$

For each trait, μ and σ^2 can then be integrated out, leading to the following marginal likelihood $p(Y | \mathbf{V}_\lambda)$ for each trait Y :

$$-2 \log p(Y | \mathbf{V}_\lambda) = n \log \pi + \log \frac{\kappa_n}{\kappa_0} - 2 \log \frac{\Gamma(\nu_n/2)}{\Gamma(\nu_0/2)} \tag{1}$$

$$- \nu_0 \log(\nu_0 \sigma_0^2) + \nu_n \log(\text{postSS}) + \log |\mathbf{V}_\lambda|.$$

A fast algorithm was used to calculate the terms involving the large matrix \mathbf{V}_λ (Ho and Ané 2014). After standardization of traits, reasonable choices are $\sigma_0^2 = 1$ and $\mu_0 = 0$. We also use $\nu_0 = 0$ and $\kappa_0 = 0$ for noninformative priors. These are improper prior distributions, but the marginal likelihood is well defined up to a constant independent of the trait data and of the current species delimitation model.

For simplicity, multiple traits are considered independent in what follows and in our implementation (but see Appendix), with independent phylogenetic signals λ a priori. A noninformative uniform prior between 0 and 1 was placed on λ for each trait. Specific proposal moves were implemented in iBPP to estimate the posterior distribution of λ with MCMC. When the current model contains a single species, between-species variation is undefined and λ is dropped from the model to avoid overparametrization. To deal with this change of dimension, a reversible jump MCMC move was implemented in iBPP when proposing to move between a one-species model and a two-species model.

Joint Bayesian inference for genes and traits

For the simultaneous analysis of genes and traits, both data types are considered independent conditional on the species tree. In other words, the sampled traits are assumed to evolve independently of the sampled genes, beyond the correlation that comes from sharing the same species tree. The joint likelihood is then the product of the genetic data likelihood and trait data likelihood. This framework thus provides an equal footing to weigh the evidence from both data sources. Namely, with k molecular alignments A_1, \dots, A_k and p traits Y_1, \dots, Y_p , the likelihood of the data and of gene trees G_1, \dots, G_k given a species tree S and $\lambda_1, \dots, \lambda_p$ is the product $\mathbb{P}(Y_1, \dots, Y_p | \lambda_1, \dots, \lambda_p, S) * \mathbb{P}\{A_1, \dots, A_k | G_1, \dots, G_k\} * \mathbb{P}\{G_1, \dots, G_k | S, \theta, \tau\}$. The first term for traits is calculated as $\prod_i p(Y_i | \mathbf{V}_{\lambda_i})$ from (1) above. The second term is the sequence likelihood based on a substitution model along gene trees. It can be decomposed as $\prod_i \mathbb{P}\{A_i | G_i\}$

with substitution and rate parameters omitted to simplify notations. The last term uses the coalescent model for gene trees with branch lengths (Rannala and Yang 2003). The gene trees and phylogenetic signals are then integrated out using MCMC. Our implementation is an extension of the BPP package version 2.0 (Yang and Rannala 2010) and allows the user to easily choose to analyze the genetic data only, the trait data only, or both data types combined.

SIMULATIONS TO DELINEATE ONE VERSUS TWO SPECIES WITH TRAITS

The MCcoal program in the BPP package (Rannala and Yang 2003; Yang and Rannala 2010) was extended to simulate trait data in addition to DNA sequence data. The Perl script used for all of our simulations is available at <http://dx.doi.org/10.5061/dryad.4gf03>. A first set of simulations was conducted with a two-species guide tree, to quantify the power from trait data in a simple case. Ten individuals from each of the two putative species were simulated. In the first scenario, all 20 individuals were simulated from the same single species. In the second scenario, the two putative species were truly distinct and each trait was simulated with an expected 70% between-species variation ($\lambda = 0.7$). Either one or three traits and 500 replicates were simulated in each case. The PP for the two-species model was obtained from iBPP.

SIMULATIONS WITH MIGRATION

The sequence data were simulated either with or without migration between sister species, even though the analysis with iBPP requires the assumption of no migration. This study aimed to simulate neutral loci susceptible to gene flow between species, and traits under selection within species. The sequence simulation model with migration was described by Zhang et al. (2011). The simulation model for traits matched the model assumed for analysis, with a Brownian motion process determining each true species trait mean, and a normally distributed within-species variation among individuals. Each trait had its own λ parameter, which represents the ratio of among-species variation (as expected from the root of the species tree to the species averages) to the total variation (including within-species variation).

We simulated data on a three-species tree (species AB, C, and DE, Fig. 2), and then analyzed these data using a guide tree with five putative species A, B, C, D, and E, so that the number of species could be underestimated or overestimated by 2. The baseline simulation consisted of five individuals per putative species, three traits, $\lambda = 0.7$, four loci with 600 base pairs each, $u = 1$ coalescent units from root to tip, and rescaled population size $\theta = 0.001$. For analysis with iBPP, we used a Gamma(1,100) prior for θ , a noninformative prior ($\kappa_0 = \nu_0 = 0$) for the traits, and algorithm 0 described in Yang and Rannala (2010). Several

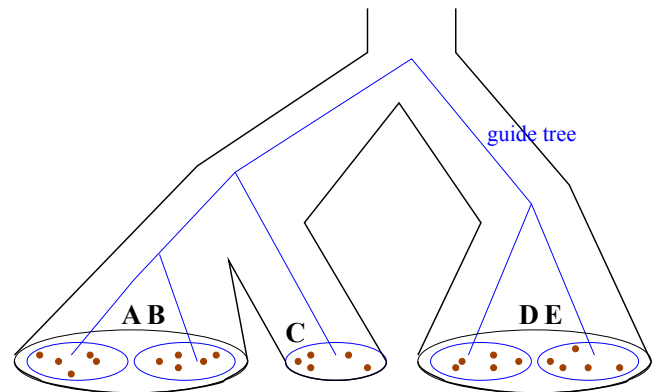


Figure 2. The true species tree used for simulations had three species: AB, C, and DE, with AB and C sister to each other (black). The guide tree (blue) used for analysis had five putative species: A, B, C, D, and E.

migration rates were tested. When present, only two types of migration were allowed: between the common ancestor of ABC and the common ancestor of DE, and migration between AB and C. Both correspond to migration between sister species. All migration rates were set to be equal and four cases were simulated: $M = 0.1, 1, 5,$ and 10 expected migrant individuals from population i to population j per generation (as in Zhang et al. 2011). Note that we kept λ constant at 0.7 as migration M was increased, that is, we assumed that the sampled traits were under selection and unaffected by the flow of neutral genes. Several combinations of parameters were tried by varying the number of traits, number of loci, expected between-species trait variance (λ), and migration rates. For the case of no migration, the total tree height (in coalescent units) and the number of sites were also varied with $u = 0.5, 1, 2$ and loci of 600–1000 base pairs. With each combination of parameters, 100 replicates were simulated and we calculated the PP of the true tree, the number of false positives (species falsely split into two), and the number of false negatives (species pairs falsely lumped) in the maximum a posteriori tree. These numbers were then averaged over all 100 simulated datasets. We also explored the computational requirements with an increased number of genes. For this, we used the same baseline settings as above but with six traits, $M = 5,$ and 10 to 100 genes. For 10 and 20 genes, we simulated 40 datasets, and for 50 and 100 genes, we simulated 10 datasets only.

SIMULATIONS WITH PHENOTYPIC MODEL VIOLATIONS

To explore the effects of violating the assumption of the phenotypic model, we extended the simulation software with two options: trait plasticity and non-BM evolution. We measured the level of plasticity as the proportion of traits for which two populations of a given species have different means. This was to

model a plastic response to an environmental variable that differs between the two populations. In our simulations with plasticity (unaccounted for during inference), we used the same proportion of plastic traits between populations A and B and between populations D and E (see Fig. 2). This plasticity level was varied between 0 and 0.10, with other parameters set to six traits at $\lambda = 0.7$, four genes with or without migration ($M = 0$ or 5), and 100 datasets in each case.

To simulate a violation of the mode of evolution, we simulated an Ornstein–Uhlenbeck (OU) process for the species mean evolution, with λ still controlling the between-to-within species variance. Under the OU process, the species mean is affected by random drift (like the BM), but is also attracted to an optimum μ . After time T , the variance of the process is of the form $\gamma(1 - e^{-2\alpha T})$, which is taken as the between-species variance when T is the height of the species tree. Our simulation tool can simulate the OU process with a constant optimum μ across the species tree, or with μ varying from branch to branch to simulate a phenotypic landscape with varying adaptive peaks (see, e.g., Butler and King 2004). We simulated traits under BM and under this OU model with $\alpha = 277$ such that the time taken by the process to move half-way toward its optimum is 50% of the species tree height ($\tau = 0.005$). We simulated values μ_i on branch i according to a centered normal distribution with a low SD of 0.1, relative to the trait's scale. For this set of simulations, we analyzed the trait data only and varied the following parameters: from 3 to 100 traits and λ from 0.3 to 0.7. Again, 100 replicates were simulated under each condition.

ANALYSIS OF LIZARD DATA

We reanalyzed a dataset from 81 males across 11 Australian amphibolurine lizard species within the *Ctenophorus maculatus* species complex (Squamata: Agamidae) (Edwards and Knowles 2014). We followed these authors for the assignment to and detection of putative species, which is required for downstream species delimitation with iBPP (like BPP). Specifically, individuals were assigned to putative taxa using a combination of multivariate and clustering techniques for analysis of genetic, morphological, and ecological data (see supplement for summary of assignment approach used by Edwards and Knowles 2014). Only males were analyzed because of sexual dimorphism, and because males typically show more distinctiveness between species in lizards. The level of divergence in the different data types differs (see Edwards and Knowles 2014, for details), with some groups of species showing shallow genetic divergence with marked morphological differences, whereas other groups of species show no morphological differentiation but deep genetic divergences. Biological processes responsible for this pattern might include hybridization and gene flow. The five phased loci, 12 morphometric characters, and five meristic characters

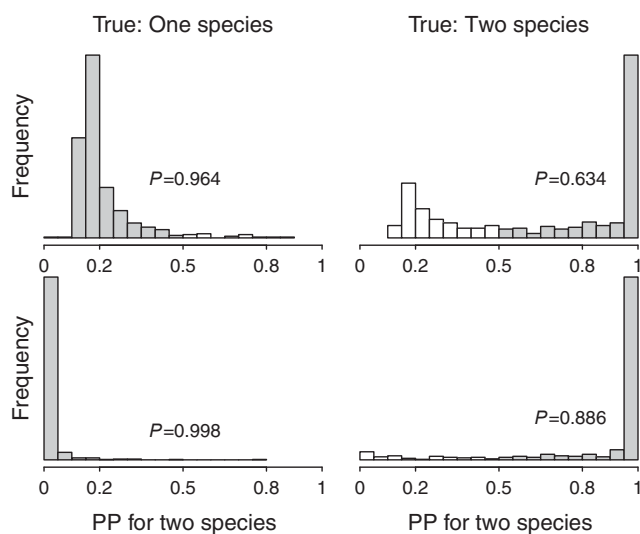


Figure 3. Posterior probability (PP) for the two-species model given one trait (top) or three traits (bottom), when the true model has one species (left) or two species (right). p is the proportion of replicates for which the PP of the true model was 0.50 or higher.

measured in adult specimens with no missing values are detailed in Edwards and Knowles (2014) and the data are available on Dryad (dx.doi.org/10.5061/dryad.mm11q). To meet the independence assumption in our model, morphometric data were corrected for allometric growth as in Lleonart et al. (2000). Snout-vent length (SVL) was used as a proxy for body size as confirmed by a principal components analysis done with the *prcomp* function in R: SVL had the largest correlation (0.96) with principal component 1. The morphometric trait transformation consisted of regressing the log-transformed values of each character against the log of SVL and using the residuals instead of the original variables. The R^2 values between pairs of transformed morphometric characters were greatly reduced after transformation. The majority fell below 0.025, and only 25% of them were above 0.15. The largest (0.575) was between interorbital width and arm length and the second largest (0.567) was between head width and arm length. For this reason, the arm length character was eliminated from the analysis, reducing the trait data to 16 characters. For the Bayesian species delimitation, three separate iBPP analyses were performed: using multilocus genetic data only, using morphological data only, and finally using both sources of data. Each analysis was run for 2.5 million rjMCMC generations after a burn-in period of 100,000 generations, sampling every 250 generations. This resulted in a final sample size of 10,000. The prior distributions for θ (ancestral population size) and τ (root age) were both set to the gamma distribution $G(2, 400)$. A noninformative conjugate prior distribution was used for the trait variances and ancestral means with parameters $\nu_0 = 0$ and $\kappa_0 = 0$. The guide tree topology (Fig. 10) was obtained from

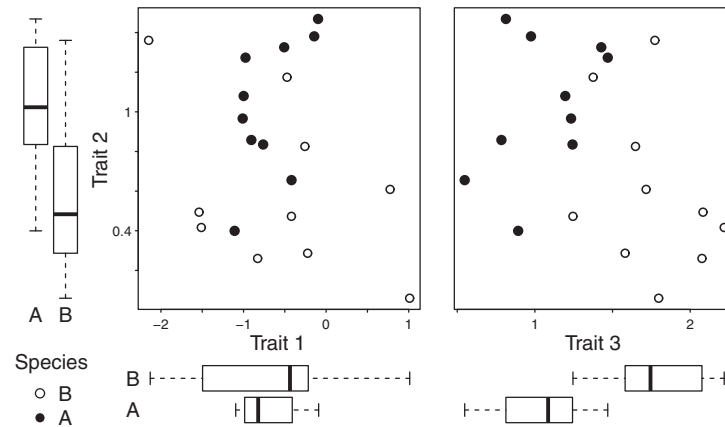


Figure 4. Degree of overlap in trait values between species for three traits simulated with $\lambda = 0.7$. Boxplots along the axes margins summarize the distribution of each trait within each species separately (median, first, and third quartile and range of trait values). Analysis with iBPP gives PP 0.96 for two species.

Edwards and Knowles (2014), who used *BEAST. Analyses were performed using algorithm 0 with $\epsilon = 15$ described in Yang and Rannala (2010). Convergence was assessed by comparing results from two separate independent runs. The separate runs were then combined. A small number of runs started with all nodes collapsed in the guide tree (one-species model) and were stuck in that state, lacking convergence. This issue was documented by Yang and Rannala (2010) and Rannala and Yang (2013). These runs were thus discarded and new runs were conducted.

The total tree height in coalescent units was evaluated as $2\tau/\theta$ from the genetic data. For τ , we used the age of the root as estimated by iBPP (in substitutions per site). The population size θ was taken either as the mean or as the median of the θ values estimated along non-root edges. Both resulting tree height estimates (0.851 and 1.02 coalescent units) were close to what was used in the simulations with migration ($u = 1$).

Analyses of the same phenotypic and genetic data were conducted with the R package Geneland (Guillot et al. 2012). For comparison with iBPP analyses, a nonspatial model was fit with uncorrelated allele frequencies. All analyses used a maximum possible number of clusters $K_{\max} = 20$. Each analysis with Geneland was run for 140,000 iterations, sampling every 100 generations and with a burn-in period of 400 after thinning, which resulted in a final sample size of 1000. Convergence assessment was done by inspecting log-likelihood profiles and comparing the results from two separate independent runs with each set.

Results and Discussion

MODEL-BASED DELIMITATION WITH PHENOTYPIC DATA

In simulations with one versus two species, the power to detect the correct species number was already very high with a single

trait: 0.964 when all individuals belonged to a single species, and 0.634 under two species when the expected between-species trait variation was 70% (Fig. 3, top). In this situation, the PP for the two-species tree correlates with the variance in the trait values explained by the grouping in the two putative species, as obtained by an analysis of variance for instance. When the two species grouping explains 40% of the total trait variance, the two-species tree receives about 80% PP, and 60% trait variance explained corresponds to a nearly full support for the two-species tree. With three traits, the power to detect the correct species model was 0.998 under a single-species model and 0.886 under a two-species model and $\lambda = 0.7$ (Fig. 3, bottom).

To illustrate the applicability of these simulations to empirical data, we show in Figure 4 an example simulated from $\lambda = 0.7$ from three traits and two true species. The species groups explained 1%, 26%, and 59% of the variance in traits 1, 2, and 3, respectively. Overlap between species is present in single traits, but the PP for two species was 0.96 with the three traits combined.

Even with limited data considered here (i.e., one to three traits), there is sufficient power to delimit taxa with a model-based analysis of phenotypic data (Fig. 3). For example, by comparing the inferred number of taxa to the true number of taxa (i.e., the number of taxa used in the simulations), with three traits we were able to detect the splitting of lineages and determine when there had been no divergence of species lineages in 88% and 98% of the replicates, respectively (Fig. 3). In other words, a model of phenotypic evolution provides the necessary context for interpreting patterns of morphological variance within and between putative species for determining the probability that trait data correspond to hypothesized species boundaries. This finding parallels the conclusions drawn about the utility of model-based frameworks more generally, such as with genetic approaches that account for variation among loci arising from coalescent variance

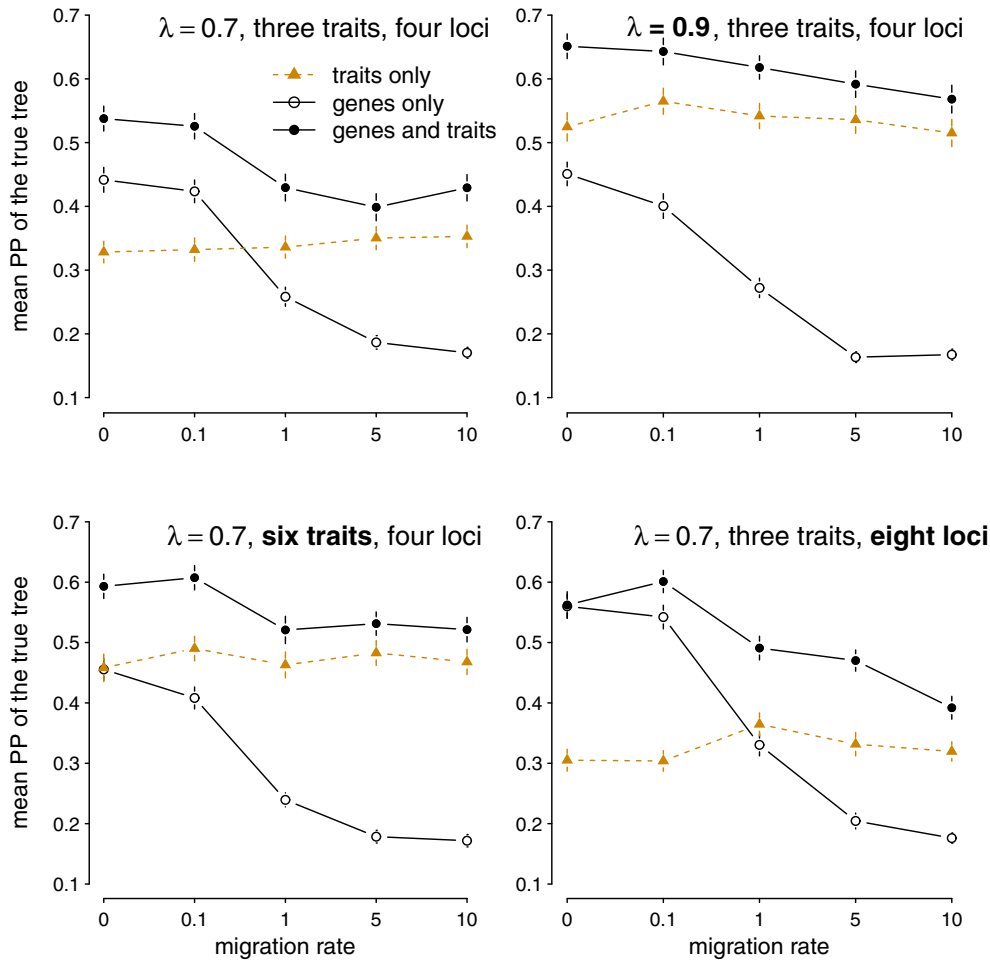


Figure 5. Mean posterior probability (PP) of the tree with the three true species with increasing migration rates. Each point is the mean over 100 replicates for each scenario. Lines show SEs. Simulations included five individuals per putative species, a total tree height of 1 coalescent units, $\theta = 0.001$, and 600 bp loci.

(e.g., Knowles and Carstens 2007; O'Meara 2010), including the Bayesian program BPP in particular (Yang and Rannala 2010)—the basis of the framework used in our iBPP approach for species delimitation across multiple data types.

Despite differences in the number of characters between the phenotypic and molecular data, the results show consistent improvements in the accuracy of delimited taxa when both data types are used (Fig. 5). In the absence of migration, the PP for the true species tree increased when combining trait and genetic data. In some situations, traits alone could provide more information than genes. Trait data were particularly beneficial when traits were under selection (high λ), or when the gene trees were highly discordant or less resolved (from a lower tree height, or shorter sequences), even under no migration.

Our results based on phenotypic data alone, with or without gene flow should not be interpreted as an advocacy on our part for species delimitation based exclusively on phenotypic traits. Instead, we include these results to illustrate the effects of data

quantity and quality on the accuracy of delimited species, which provide general guidelines for empiricists. For example, it is clear that both the ratio of variance between species to within species as reflected in the λ parameter and the number of traits impact the probability of recovering the true history of species divergence (Fig. 5).

DIVERGENCE WITH GENE FLOW

In the presence of neutral loci affected by gene flow, the mean PP of the true tree deteriorated substantially with increasing migration rates when using genetic data only. The addition of data from traits under selection (unaffected by gene flow) always increased the PP for the true tree, and rescued a high PP of recovering the correct species even in the presence of high migration rates (Fig. 5).

In the absence of migration, both data sources showed a slight trend to oversplit, leading to slightly more false positives than false negatives (Figs. 6, 7). Note that the prior probabilities

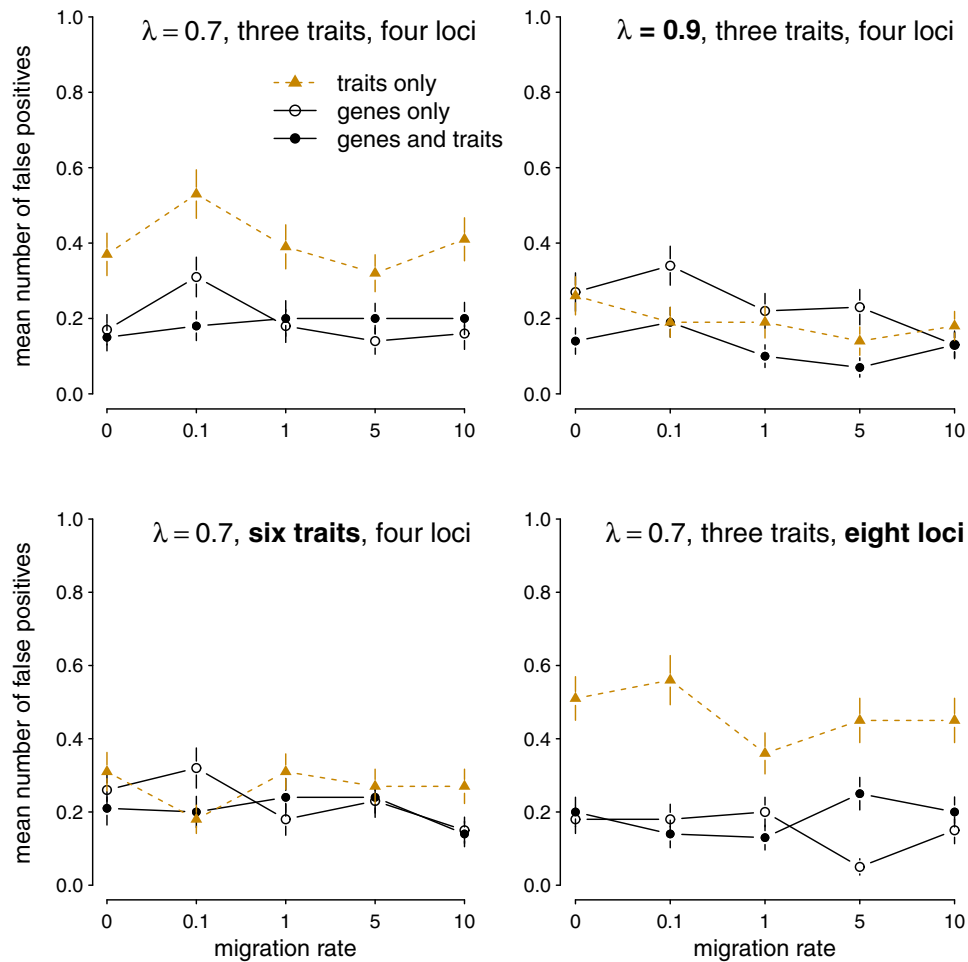


Figure 6. Mean number of false positives (oversplitting) in the most probable tree a posteriori. Points and bars as in Figure 5.

given to undersplitting and oversplitting were not equal, because the uniform prior on all trees derived from the guide tree results in a prior average of 5/7 or 0.71 false positives, and 4/7 or 0.57 false negatives. This slight imbalance might explain the slight tendency toward oversplitting in the absence of migration. In the presence of gene flow between sister species, the genetic data tended to lump separate species together instead, as expected from migration. This trend to undersplit was largely overcome with the addition of traits under selection, even with the highest level of migration considered here (Figs. 6, 7).

What is notable about results from analyses in which phenotypic data alone are used for delimiting taxa is that they outperform the analyses based on genetic data alone when divergence occurs with gene flow for any appreciable level of migration (i.e., values above 0.1; Fig. 5). It was already clear that coalescent-based approaches were sensitive to migration rates above 0.1, resulting in underestimates of species numbers (as documented in Zhang et al. 2011). In fact, the parameter values used here correspond to the levels of migration and species divergence times used in past simulation studies. The failure to detect species lineages under

gene flow with analyses based on genetic data alone compared to phenotypic data alone (Fig. 7) is what drives the higher accuracy of delimitation with phenotypic compared to genetic data (Fig. 5). Indeed, our results show that taxa delimited only with phenotypic data will be more subject to oversplitting of species lineages, especially with limited numbers of traits and high within-to-between lineage variance in those traits, that is, when λ is small (Fig. 6).

ASSUMPTIONS OF THE EVOLUTIONARY MODEL IN iBPP

Violation of the underlying assumptions regarding the molecular data (i.e., distribution of gene trees and of substitutions along gene trees) may compromise the reliability of inferences with iBPP, just like it does with BPP (e.g., see Zhang et al. 2011). However, by considering multiple data types, we show how the accuracy of species delimitation can be improved when incorporating phenotypic data because it can capture the signal of some of the processes involved in the speciation process—namely, selectively driven divergence and divergence with gene flow—that actually compromise the results based on molecular data.

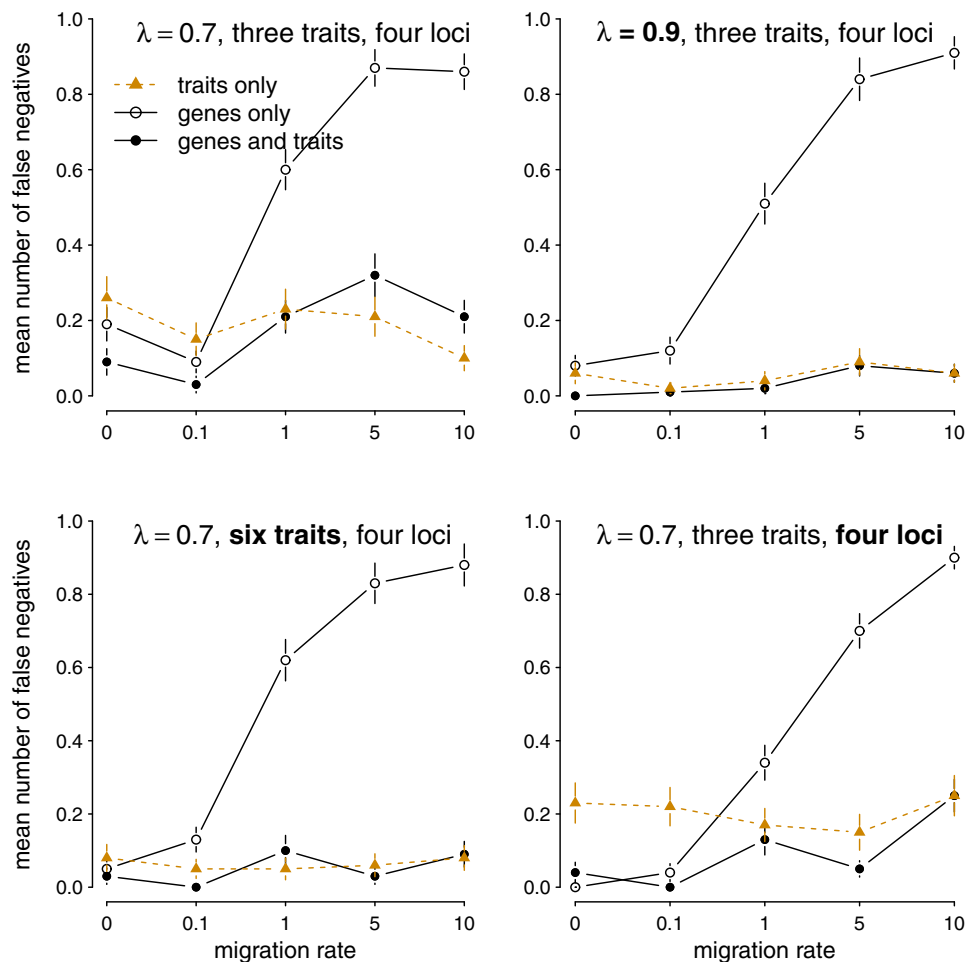


Figure 7. Mean number of false negative (undersplitting) in the most probable tree a posteriori. Points and bars as in Figure 5.

Our model uses a uniform prior distribution for λ , the relative between-species variance of a trait. Although we could have used an informed prior derived from population genetic theory (e.g., $\lambda/(1 - \lambda)$ approximately centered around the tree height $2\tau/\theta$ in coalescent units), this expectation only applies to neutrally evolving traits and ignores environmental variance (Lynch and Hill 1986). Therefore, we opted for a noninformative prior for λ . Our approach also assumes that traits evolved independently, to facilitate the use of a trait-specific λ . To accommodate this assumption in practice, we, for example, transformed the traits to remove correlations due to allometry in the lizard dataset. Alternatively, a more complex model with correlated traits might be implemented and is described (see Appendix). However, its full implementation would incur a higher computational cost and is left to future work.

Although very promising, the accuracy of iBPP using phenotypic data should be explored when the actual model of phenotypic evolution is violated. One such violation is phenotypic plasticity. For example, different populations of the same species might differ phenotypically because of a plastic response to environmental

differences between their respective geographic areas. We found that a modest level of plasticity (up to 10% of traits) only slightly affected the inference accuracy (Fig. 8). The effects of gene flow between species and phenotypic plasticity within species were expected to pull the estimated boundaries in opposite directions, with gene flow causing too few recognized species and phenotypic plasticity causing too many. Indeed, inference based on both genes and traits was always more accurate than when using either data type alone (Fig. 8).

We also explored violations of the mode of trait evolution. Figure 9 shows that our method is very robust when species means evolved according to an OU process, instead of BM. In fact, the major factor controlling accuracy, when using trait data only, is the between-species variation as reflected by λ . Our simulations involved slight variations between optima along separate species lineages. Large variations between lineage-specific optima are expected to cause more variation between species, and we expect our model to account for this by a high estimated λ and a total variance σ^2 mostly representative of the variation in selective optima between species. More generally, any model of trait

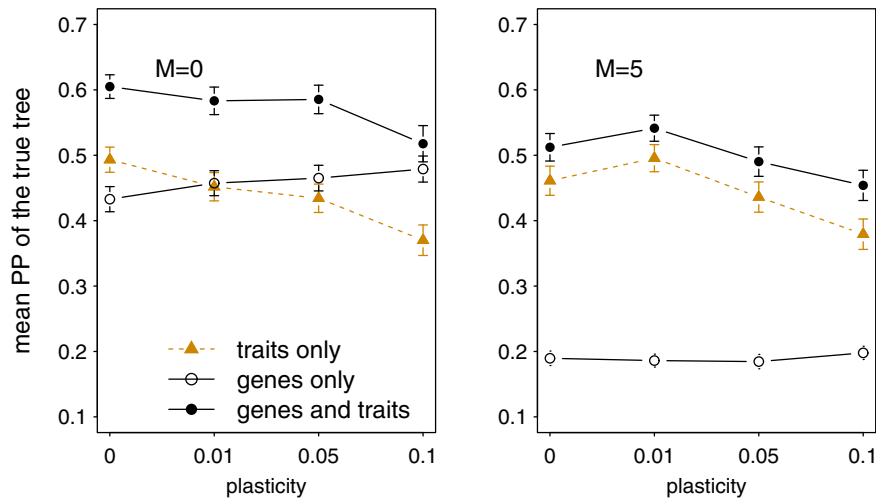


Figure 8. Mean posterior probability (PP) for the true three species as in Figure 5 with phenotypic plasticity, without (left) or with (right) gene flow. The plasticity level is the proportion of traits with different means for two populations in the same species.

evolution that would increase the trait variance between species relative to the within-lineage variance would be expected to increase the accuracy of delimited taxa based on phenotypic data.

ANALYSIS OF AUSTRALIAN LIZARD DATA

All 11 putative species (Fig. 10) were supported by iBBP analyses of each of the three datasets (genes only, traits only, and genes and traits). All nodes from the guide tree were present in the maximum PP tree and were supported as divergent nodes with PP of 1.0 in all analyses, except for three nodes when only the traits were analyzed (Fig. 10). All the iBBP analyses showed convergence with similar results from separate runs; the mean absolute difference in node PP between runs was 0, 0.06, and 0 using genes, traits, and both genes and traits, respectively.

The general consistency of the delimited taxa across iBBP analyses of morphological and genetic datasets, as well as the correspondence with species boundaries from traditional, qualitative taxonomic criteria used to describe species (Edwards and Knowles 2014), suggests that the results from our quantitative tests are robust to undersplitting of taxa associated with limited information or divergence processes that compromise inferences based on single data types (see McKay et al. 2013). It is worth noting that the putative lizard taxa exhibit differences in the level of genetic and morphological differentiation among taxa, conditions that can make species delimitation difficult. Because the maximum number of taxa was constrained by the initial 11 putative species, it is possible that additional taxa have gone undiscovered. This issue relates to the upstream assignment of individuals, which is a difficult task (see below). To the extent that the delimited taxa are indeed accurate, it would appear the evolutionary model used in iBBP that accommodates different rates of

evolution across dataset types is effective (but see discussion above about assumptions of the model in iBBP).

In contrast, the number of putative species appears to be significantly underestimated, with only a few clusters recognized, with the alternative method Geneland (Guillot et al. 2012) for combining genetic and phenotypic data (Fig. 11). The estimated number of clusters \hat{K} was 2 when using genetic haplotype data only (with PP 1) or both haplotype and phenotypic data (with PP 0.95), and $\hat{K} = 3$ with PP 0.5 when using traits only (95% credibility interval of [3–5]). Proper mixing was reached in these three analyses, based on log-likelihood plots and consistent results across two separate runs. Using traits only, no individual was confidently assigned to any cluster and the three clusters did not seem to correspond to any particular species or groups of species (Fig. 11, top). Using haplotypes only, the two clusters did not correspond to any particular species or groups of species either (Fig. 11, center). Interestingly, the combined analysis of haplotype and phenotypic data inferred clusters that matched a clade in the guide tree exactly (Fig. 11, bottom). One cluster was formed by all individuals in the *C. femoralis* + *C. fordi* clade and the other cluster contained the remaining individuals. This result corroborates our claim that combining different sources of data can improve species delimitation, although the number of species was still underestimated with Geneland.

This empirical application to Australian lizards contrasts the strengths and weaknesses of two methods—iBBP and Geneland—for integrating multiple data types. The former has the advantage of using an explicit evolutionary model with a branching structure for population relationships, a substitution model for genetic variation, and a model of phenotypic divergence. The latter has the advantage of not relying on a preassignment of individuals to putative species, or on a guide tree for these species. Geneland is

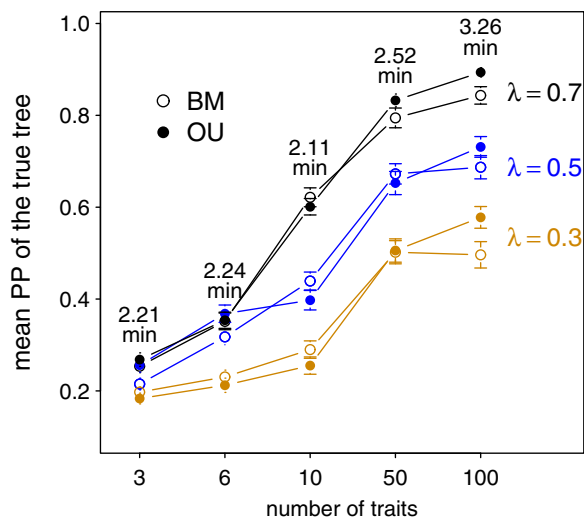


Figure 9. Mean posterior probability (PP) for the true three species as in Figure 5 from trait data only, with a BM (empty circles) or OU (filled circles) mode of evolution. Computing times are indicated, per replicate dataset.

also able to integrate geographic data (but see below). However, this method is not based on any explicit evolutionary model. It relies instead upon statistical clustering for detecting species clusters (i.e., a tessellation model; Møller and Stoyan 2014), which is not evolutionary and does not recognize the hierarchy of species grouping as expected from a species phylogeny.

A number of factors may have contributed to the limited utility of Geneland to delimit the *Ctenophorus* taxa. Most likely the variable, but linked sites contained within the sequenced loci, provides limited information for inferring individual ancestries from allele frequencies (Guillot et al. 2012), given that genealogical information contained by the haplotypes is not used (as with other approaches based on allele frequency, like the program STRUCTURE; Pritchard et al. 2000). Moreover, given that the relevant information for such inferences depends upon allele frequencies, as the diversity of haplotypes increases, the performance of such approaches will be compromised (see Olave et al. 2013). This probably explains why a mere two species of *Ctenophorus* were recognized in the Geneland analyses.

CONTINUED CHALLENGE OF ASSIGNING INDIVIDUALS TO PUTATIVE SPECIES

Assigning individuals to putative populations is a preprocessing step for downstream analysis with iBPP. This step is potentially error-prone and currently, there is no standard method for individual assignment. An obviously more satisfying approach would be to consider individual assignment and species tree estimation simultaneously. However, doing so is computationally very challenging (but see O'Meara 2010; Jones and Oxelman 2014) and outside the scope of this article.

Individual assignment (i.e., individual-putative species associations) can be estimated with programs such as STRUCTURE (Pritchard et al. 2000), STRUCTURAMA (Huelsenbeck and Andolfatto 2007), or Geneland (Guillot et al. 2005, 2012). However, these methods ignore the phylogenetic relationships between species, as mentioned earlier. With the potential for errors at the initial steps in the delimitation process (i.e., assigning individuals to putative species) to be magnified in the downstream process of delimiting species with iBPP (or BPP; see Olave et al. 2013), approaches that integrate across multiple data types may be preferable at the initial steps in the process as well (see Edwards and Knowles 2014).

CHALLENGES WITH LARGE DATA SETS

Although powerful conclusions can be reached with a few loci in some systems, current sequencing trends enable large genomic datasets. The computational burden of iBPP increases with the number of loci (Fig. 12), as it does with BPP, but stays reasonable in the situation considered here with a five-species guide tree. Very large SNP data are also increasingly available, but require specific methods to analyze under the coalescent. They have been used for species delimitation with SNAPP (Bryant et al. 2012) and Bayes factor delimitation (Leaché et al. 2014). The theory developed here could readily be used to integrate trait data with SNAPP (or other programs such as DISSECT; Jones and Oxelman 2014), and therefore integrate trait data with large-scale SNP data along a species tree.

The number of traits that might be used in species delimitation can also become very large (e.g., gene expression data). Our simulations show that iBPP scales very well to many traits (Fig. 9), but that power is limited when between-species variation is low (i.e., low λ values). This does not bode well for gene expression data, which are expected to have a large within-individual and within-species variance. Phylogenetic signal and large between-species variation has been observed in some gene expression studies, however (e.g., across mammals; Brawand et al. 2011).

With large datasets (i.e., hundreds to thousands of loci), it is possible that the gene(s) underlying a particular trait be used in the analysis, in which case our conditional independence assumption between genes and traits is dubious. The expression of a single gene, for instance, might be more accurately modeled along the tree for that gene, rather than along the species tree. Nevertheless, our independence assumption might be adequate if the gene tree does not conflict strongly with the species tree.

Theoretically, analysis of very large datasets (hundreds of genes or traits) would require the development of new prior distributions for gene-specific parameters (such as substitution rate) or trait-specific parameters (such as λ). Currently, these parameters are given independent priors. With many genes or many traits, the

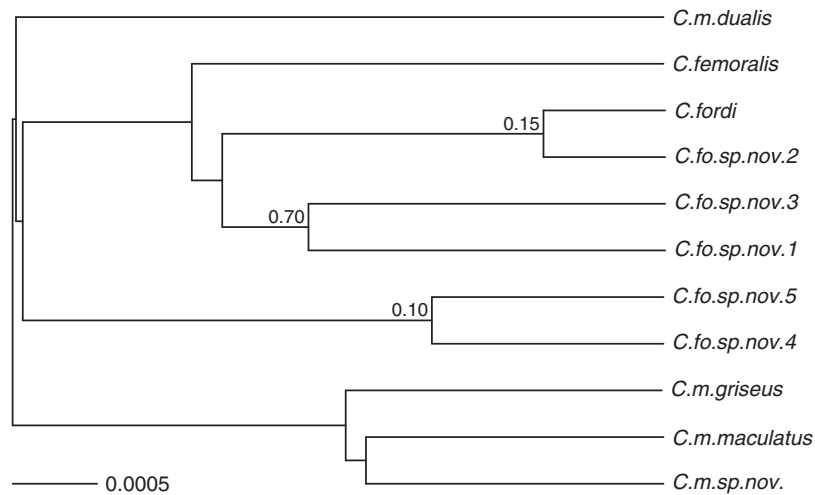


Figure 10. All 11 putative species in the *Ctenophorus maculatus* species complex were delimited with full support (i.e., 1.0 PP) in each of the iBBP analyses of the three datasets (traits only, genes only, and genes and traits), except when only the traits were analyzed (nodes PP as marked). The guide tree used in iBBP analyses was from Edwards and Knowles (2014), which is from a species tree estimated from multilocus data with *BEAST. Branches lengths (τ) are shown as estimated from genes and traits. Estimated rescaled population sizes (θ) ranged from 0.0014 to 0.019, excluding the value at the root (0.130).

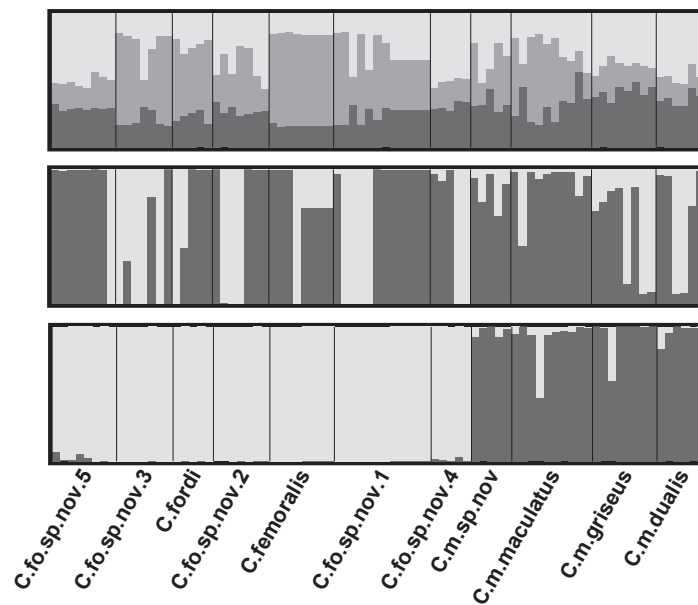


Figure 11. Posterior probabilities of individuals assignment to the estimated \hat{K} clusters from Geneland analyses using traits only (top, $\hat{K} = 3$), haplotypes only (center, $\hat{K} = 2$), and both traits and haplotypes (bottom, $\hat{K} = 2$).

mean of these parameters across all genes or all traits becomes narrow-peaked and informative. Similar phenomena have been discovered recently for other models, with an undue influence of prior choices on very large trees (Rannala et al. 2012; Dos Reis et al. 2014).

FUTURE DEVELOPMENTS FOR iBBP (OR OTHER MODEL-BASED INTEGRATIVE APPROACHES)

Given the promising results from the analyses of phenotypic traits with continuous variation, extending the framework to accommodate additional phenotypic trait data would increase the general

applicability of the approach, as well as provide access to potentially informative traits that are not yet accommodated. This includes the development of models for discrete traits, as well as models of evolution that account for dimorphism, for example, by augmenting within-species variation with known covariates such as sex, age, or environmental variables known to cause plasticity. As noted in the analysis of the Australian lizard datasets, only males were analyzed to avoid variance attributable to sexual dimorphism, which decreased the number of individuals available for estimating species boundaries.

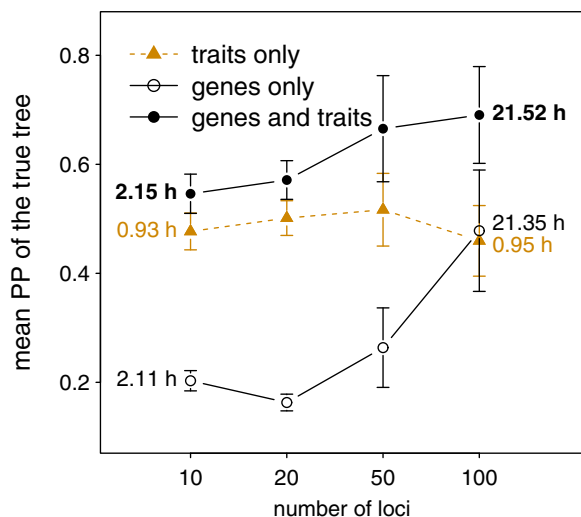


Figure 12. Mean posterior probability (PP) for the true three species as in Figure 5. Simulated data included six traits, between 10 and 100 genes with migration rate $M = 5$. For 10 and 20 genes, we simulated 40 datasets, and for 50 and 100 genes, we simulated only 10 datasets. Computing times are indicated, per replicate dataset.

A challenging but nonetheless interesting prospect would be to extend the model-based framework beyond genetic and phenotypic data to geographic data. Geographic data are already used to inform assignment of individuals to putative taxa. For instance, the programs STRUCTURE and Geneland (Hubisz et al. 2009; Guillot et al. 2012) use geographic proximity in their prior distribution as evidence for clustering closely located individuals in the same population. Indeed, geographic proximity is expected to correspond to coancestry: geographically proximate individuals are expected to be potentially interbreeding. Deviations from this expectation might also provide valuable information for inferring species boundaries. In particular, geographic proximity of lineages showing genetic distinctiveness might be interpreted as additional evidence for these lineages to be separate species. In analyses of genetic data alone, it is arguably unclear what has been delimited: distinct species versus populations of the same species showing genetic differentiation due to geographic isolation (see Hey and Pinho 2012). If geographic data can be leveraged, demonstration of genetic distinctiveness despite the potential for gene flow (i.e., geographic connectedness in past or present; He et al. 2013) would provide key information for interpreting the processes underlying observed genetic differences, such as a barrier to reproduction that is not simply a function of geographic isolation. In addition to geography, ecological divergence might also be integrated into the model-based framework given the relevant information it might contain about species boundaries (e.g., Wiens and Penkrot 2002; Rissler and Apodaca 2007). Recent developments for the characterization of divergent ecological niches

for allopatric populations (see Warren et al. 2008; McCormack et al. 2010; Warren et al. 2010) would provide a way to avoid the confounding effects of nonoverlapping ranges in the characterization of ecological niches, especially for those based on ecological niche modeling (see Alvarado-Serrano and Knowles 2013).

Lastly, as when using BPP, expanding the iBPP approach to search over multiple guide trees would guard against a key factor that contributes to failed delimitation: errors in the guide tree. Ideally, the analysis would not rely on a single guide tree, but would also include alternative species relationships and maximally split putative species, because the number of estimated species is currently limited by the number of putative species included in the guide tree. With respect to this latter point, the trade-off in assigning individuals to subgroups within species lineages that are not in conflict with actual patterns of ancestry may be difficult to overcome (Olave et al. 2013).

BENEFITS OF AN INTEGRATED EVOLUTIONARY MODEL

In this work, we provide a framework to address a critical knowledge gap about the limitation of species delimitation methods based on genetic data alone. Specifically, because both phenotypic and molecular data are analyzed in a common Bayesian framework with our program iBPP, not only can we compare the accuracy of delimited taxa based on genetic data alone versus when integrated with phenotypic data, but we can also evaluate how the integration of phenotypic data might improve the accuracy of estimates when divergence occurs with gene flow and/or is selectively driven. These two realities of the speciation process can compromise the accuracy of inferred species boundaries based on genetic data alone because these methods assume selective neutrality and no gene flow. Based on the improvements associated with the integration of phenotypic and genetic data, not to mention the limitations we demonstrate of using genetic data alone (see also Bauer et al. 2010; Harrington and Near 2012; McKay et al. 2013; Olave et al. 2013; Edwards and Knowles 2014), it is difficult to justify the practice of focusing exclusively on genetic data for species delimitation (see also Guillot et al. 2012).

Our study explicitly shows how the benefits of an integrated evolutionary model for species delimitation can be realized by (1) narrowing the parameter space where species are undetected or oversplit (e.g., Bauer et al. 2010; Harrington and Near 2012), even when species divergence corresponds to the specified model, as well as (2) avoiding misleading conclusions when genetic divergence violates the assumptions of the coalescent model used to statistically delimit taxa (Hey and Pinho 2012). Because it is model-based, iBPP provides a context for interpreting patterns of differentiation, even when the level of divergence among putative taxa differs across loci and/or morphological characters. For

example, limited divergence in phenotypic traits may not be in conflict with detected genetic differentiation, but instead consistent with expectations based on recent divergence times or low rates of evolution in the phenotypic characters. Note that the model used in iBPP accommodates trait-specific parameter values (as captured by the λ parameter)—that is, it does not assume equal within-to-between species variance/covariance across phenotypic characters. With the estimation of λ , iBPP can accommodate phenotypic traits that evolve at different rates, including those subject to selection, and uncertainty in the actual rate of evolution and between-species contribution are taken into account by the specified prior. This means that a diversity of traits can be studied even when the investigator does not have background information on whether the trait is likely or not to be diverging under selection.

The improved accuracy afforded by the use of multiple data types has obvious implications for species delimitation in practice. By not relying on a singular data type, there is a lower likelihood that diversity is misrepresented (McKay et al. 2013; Edwards and Knowles 2014). Again, by considering both phenotypic and genetic data, the characters related to the speciation process (e.g., selected phenotypes affecting reproductive isolation that evolve faster than neutral loci) may be considered in analyses, and statistical approaches for delimitation are less likely to be misled when the assumptions of the genetic model or phenotypic model are violated. The consideration of multiple data types using a framework such as iBPP also provides connections between species detection and description in an integrative taxonomy (Yeates et al. 2011). This contrasts with the criticisms leveled at investigations where the delimitation of taxa relies solely on genetic data, leaving unanswered the question of what might distinguish taxa phenotypically and therefore impeding the study of such taxa if the only means used to identify a putative taxon is by characterizing its genetic distinctiveness relative to other sampled taxa (see Bauer et al. 2010). It is also worth noting that applications of integrative modeling of multiple data types can make connections with taxonomic traditions of identifying distinguishing phenotypes while maintaining the appeal of statistical rigor and objectivity—two aspects that have contributed to the increased popularity of genetic-based delimitation, while diminishing the support of classic taxonomic treatments (Hey and Pinho 2012).

ACKNOWLEDGMENTS

We thank D. Weisrock and the NIMBioS working group on species delimitation, which fostered initial discussions on this work. We also thank anonymous reviewers for their insightful comments. This work was supported by the National Science Foundation (DEB-0949121, DMS-1106483, and DEB 1354793 to CA, and DEB-0715487 to LLK).

DATA ARCHIVING

The doi for our scripts is [dx.doi.org/10.5061/dryad.4gf03](https://doi.org/10.5061/dryad.4gf03), and for our lizard data is [dx.doi.org/10.5061/dryad.mm11q](https://doi.org/10.5061/dryad.mm11q).

LITERATURE CITED

- Alvarado-Serrano, D. F., and L. L. Knowles. 2013. Ecological niche models in phylogeographic studies: applications, advances and precautions. *Mol. Ecol. R* 14:233–248.
- Bauer, A. M., J. F. Parham, R. M. Brown, B. L. Stuart, L. Grismer, T. J. Papenfuss, W. Böhme, J. M. Savage, S. Carranza, J. L. Grismer, et al. 2010. Availability of new Bayesian-delimited gecko names and the importance of character-based species descriptions. *Proc. R. Soc. B Biol. Sci.* 278:490–492.
- Brawand, D., M. Soumillon, A. Necșulea, P. Julien, G. Csardi, P. Harrigan, M. Weier, A. Liechti, A. Aximu-Petri, M. Kircher, et al. 2011. The evolution of gene expression levels in mammalian organs. *Nature* 478:343–348.
- Bryant, D., R. Bouckaert, J. Felsenstein, N. A. Rosenberg, and A. RoyChoudhury. 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Mol. Biol. Evol.* 29:1917–1932.
- Burbrink, F., H. Yao, M. Ingrasci, R. J. Bryson, T. Guiher, and S. Ruane. 2011. Speciation at the mogollon rim in the arizona mountain kingsnake (*Lampropeltis pyromelana*). *Mol. Phylogenet. Evol.* 60:445–454.
- Butler, M. A., and A. A. King. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *Am. Nat.* 164:683–695.
- Camargo, A., M. Morando, L. J. Avila, and J. W. Sites. 2012. Species delimitation with abc and other coalescent-based methods: a test of accuracy with simulations and an empirical example with lizards of the *Liolaemus darwini* complex (Squamata: Liolaemidae). *Evolution* 66:2834–2849.
- Carstens, B. C., T. A. Pelletier, N. M. Reid, and J. D. Satler. 2013. How to fail at species delimitation. *Mol. Ecol.* 22:4369–4383.
- Dos Reis, M., T. Zhu, and Z. Yang. 2014. The impact of the rate prior on Bayesian estimation of divergence times with multiple loci. *Syst. Biol.* 63:555–565.
- Edwards, D. L., and L. L. Knowles. 2014. Species detection and individual assignment in species delimitation: can integrative data increase efficacy? *Proc. R. Soc. B Biol. Sci.* 281:20132765.
- Ence, D., and B. Carstens. 2011. SpedeSTEM: a rapid and accurate method for species delimitation. *Mol. Ecol. Res.* 11:473–480.
- Felsenstein, J. 2008. Comparative methods with sampling error and within-species variation: contrasts revisited and revised. *Am. Nat.* 171:713–725.
- Freckleton, R. 2012. Fast likelihood calculations for comparative analyses. *Methods Ecol. Evol.* 3:940–947.
- Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. *Am. Nat.* 160:712–726.
- Fujita, M. K., A. D. Leaché, F. T. Burbrink, J. A. McGuire, and C. Mortiz. 2012. Coalescent-based species delimitation in an integrative taxonomy. *Trends Ecol. Evol.* 27:480–488.
- Guillot, G., F. Mortier, and A. Estoup. 2005. Geneland: a program for landscape genetics. *Mol. Ecol. Notes* 5:712–715.
- Guillot, G., S. Renaud, R. Ledevin, J. Michaux, and J. Claude. 2012. A unifying model for the analysis of phenotypic, genetic and geographic data. *Syst. Biol.* 61:897–911.
- Harrington, R. C., and T. J. Near. 2012. Phylogenetic and coalescent strategies of species delimitation in snubnose darters (Percidae: *Etheostoma*). *Syst. Biol.* 61:63–79.
- Hausdorf, B., and C. Hennig. 2010. Species delimitation using dominant and codominant multilocus markers. *Syst. Biol.* 59:491–503.
- He, Q., D. L. Edwards, and L. L. Knowles. 2013. Integrative testing of how environments from the past to the present shape genetic structure across landscapes. *Evolution* 67:3386–3402.
- Hey, J., and C. Pinho. 2012. Population genetics and objectivity in species diagnosis. *Evolution* 66:1413–1429.

- Ho, L. S. T., and C. Ané. 2014. A linear-time algorithm for gaussian and non-gaussian trait evolution models. *Syst. Biol.* 63:397–408. R package: phylolm.
- Hubisz, M. J., D. Falush, M. Stephens, and J. K. Pritchard. 2009. Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Res.* 9:1322–1332.
- Huelsenbeck, J. P., and P. Andolfatto. 2007. Inference of population structure under a Dirichlet process model. *Genetics* 175:1787–1802.
- Jones, G., Aydin, Z., and B. Oxelman. 2014. DISSECT: an assignment-free Bayesian discovery method for species delimitation under the multi-species coalescent. *Bioinformatics*. doi: 10.1093/bioinformatics/btu770. Available at <http://bioinformatics.oxfordjournals.org/content/early/2014/12/10/bioinformatics.btu770>.
- Kergoat, G., B. P. L. Ru, G. Genson, C. Cruaud, A. Couloux, and A. Delobel. 2011. Phylogenetics, species boundaries and timing of resource tracking in a highly specialized group of seed beetles (Coleoptera: Chrysomelidae: Bruchinae). *Mol. Phylogenet. Evol.* 59:746–760.
- Knowles, L. L., and B. Carstens. 2007. Delimiting species without monophyletic gene trees. *Syst. Biol.* 56:887–895.
- Leaché, A., and M. Fujita. 2010. Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proc. R. Soc. B Biol. Sci.* 277:3071–3077.
- Leaché, A. D., M. K. Fujita, V. N. Minin, and R. R. Bouckaert. 2014. Species delimitation using genome-wide SNP data. *Syst. Biol.* 63:534–542.
- Lleonart, J., J. Salat, and G. Torres. 2000. Removing allometric effects of body size in morphological analysis. *J. Theor. Biol.* 205:85–93.
- Lynch, M. 1991. Methods for the analysis of comparative data in evolutionary biology. *Evolution* 45:1065–1080.
- Lynch, M., and W. G. Hill. 1986. Phenotypic evolution by neutral mutation. *Evolution* 40:915–935.
- McCormack, J. E., A. J. Zellmer, and L. L. Knowles. 2010. Does niche divergence accompany allopatric divergence in Aphelocoma jays as predicted under ecological speciation?: insights from tests with niche models. *Evolution* 64:1231–1244.
- McKay, B. D., H. L. Mays, Y. Wu, H. Li, C.-t. Yao, I. Nishiumi, and F. Zou. 2013. An empirical comparison of character-based and coalescent-based approaches to species delimitation in a young avian complex. *Mol. Ecol.* 22:4943–4957.
- Møller, J., and D. Stoyan. 2014. Stochastic geometry and random tessellations. In R. van de Weijngaert, G. Vegter, V. Icke, and J. Ritzerveld, eds. *Tessellations in the sciences: virtues, techniques and applications of geometric tilings*. Springer-Verlag. To appear.
- Niemiller, M. L., T. J. Near, and B. M. Fitzpatrick. 2012. Delimiting species using multilocus data: diagnosing cryptic diversity in the southern cavefish, *Typhlichthys subterraneus* (Teleostei: Amblyopsidae). *Evolution* 66:846–866.
- Olave, M., E. Solà, and L. L. Knowles. 2013. Upstream analyses create problems with dna-based species delimitation. *Syst. Biol.* 63:263–271.
- O’Meara, B. 2010. New heuristic methods for joint species delimitation and species tree inference. *Syst. Biol.* 59:59–73.
- Pagel, M. 1999. Inferring the historical patterns of biological evolution. *Nature* 401:877–884.
- Pons, J., T. G. Barraclough, J. Gomez-Zurita, A. Cardoso, D. P. Duran, S. Hazell, S. Kamoun, W. D. Sumlin, and A. P. Vogler. 2006. Sequence based species delimitation for the DNA taxonomy of undescribed insects. *Syst. Biol.* 55:595–609.
- Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
- Rannala, B., and Z. Yang. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* 164:1645–1656.
- . 2013. Improved reversible jump algorithms for Bayesian species delimitation. *Genetics* 194:245–253.
- Rannala, B., T. Zhu, and Z. Yang. 2012. Tail paradox, partial identifiability, and influential priors in Bayesian branch length inference. *Mol. Biol. Evol.* 29:325–335.
- Rissler, L. J., and J. J. Apodaca. 2007. Adding more ecology into species delimitation: ecological niche models and phylogeography help define cryptic species in the black salamander (*Aneides flavipunctatus*). *Syst. Biol.* 56:924–942.
- Rittmeyer, E., and C. Austin. 2012. The effects of sampling on delimiting species from multi-locus sequence data. *Mol. Phylogenet. Evol.* 65:451–463.
- Satler, J. D., B. C. Carstens, and M. Hedin. 2013. Multilocus species delimitation in a complex of morphologically conserved trapdoor spiders (Mygalomorphae, Antrodiaetidae, Aliatypus). *Syst. Biol.* 62:805–823.
- Warren, D. L., R. E. Glor, and M. Turelli. 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* 62:2868–2883.
- . 2010. ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography* 33:607–611.
- Wiens, J. J., and T. A. Penkrot. 2002. Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (*sceloporus*). *Syst. Biol.* 51:69–91.
- Yang, Z., and B. Rannala. 2010. Bayesian species delimitation using multilocus sequence data. *Proc. Natl. Acad. Sci. USA* 107:9264–9269.
- Yeates, D., A. Seago, L. Nelson, S. Cameron, L. Joseph, and J. Trueman. 2011. Integrative taxonomy, or iterative taxonomy? *Syst. Ent.* 36:209–217.
- Zhang, C., D.-X. Zhang, T. Zhu, and Z. Yang. 2011. Evaluation of a Bayesian coalescent method of species delimitation. *Syst. Biol.* 60:747–761.

Associate Editor: R. Brumfield

Handling Editor: R. Shaw

Appendix: Multiple Traits

Our model assumes independent traits, but can be generalized to correlated traits. A complex model uses covariance matrix $\mathbf{C}_b \otimes \mathbf{V}_1 + \mathbf{C}_w \otimes \mathbf{V}_0$ between all traits in all individuals, with evolutionary covariance \mathbf{C}_b between traits along the phylogeny, and phenotypic covariance \mathbf{C}_w within species (Felsenstein 2008). Here, \otimes represents the Kronecker product.

This model can be simplified to $\mathbf{C} \otimes \mathbf{V}_\lambda$ if the phylogenetic and phenotypic covariances are assumed to be proportional with the same phylogenetic signal λ for all traits. More formally, this model can be written as $\text{vec}(\mathbf{Y}) \sim \mathcal{N}(\boldsymbol{\mu}' \otimes \mathbf{1}, \mathbf{C} \otimes \mathbf{V}_\lambda)$, where $\text{vec}(\mathbf{Y})$ is a column vector of size nd (n individuals, d traits) made by stacking up all the individual trait data vectors, $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_d)$ contains the ancestral species mean μ_j for trait j , $\mathbf{C} = \mathbf{C}_b = \mathbf{C}_w$ is the covariance between traits, and \mathbf{V}_λ is the phylogenetic covariance between individuals. The likelihood p is then given by

$$-2 \log p(\mathbf{Y} | \boldsymbol{\mu}, \mathbf{C}, \mathbf{V}_\lambda) = nd \log(2\pi) + n \log |\mathbf{C}| + d \log |\mathbf{V}_\lambda| \\ + \text{tr}(\mathbf{C}^{-1}(\mathbf{Y} - \mathbf{1}\boldsymbol{\mu})' \mathbf{V}_\lambda^{-1}(\mathbf{Y} - \mathbf{1}\boldsymbol{\mu})),$$

where $\text{tr}(\mathbf{M})$ is the trace of matrix \mathbf{M} and $\mathbf{1}$ is a vector of ones of size n . For our Bayesian framework, we use a conjugate prior distribution for the ancestral means $\boldsymbol{\mu}$ given \mathbf{C} and \mathbf{V}_λ : $\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\mu}_0, \frac{n}{\kappa_0}(\mathbf{1}'\mathbf{V}_\lambda^{-1}\mathbf{1})^{-1}\mathbf{C})$, which allows us to integrate $\boldsymbol{\mu}$ out analytically. Given the data, \mathbf{C} and \mathbf{V}_λ , $\boldsymbol{\mu}$ is Gaussian with mean $\boldsymbol{\mu}_n = (\kappa_0\boldsymbol{\mu}_0 + n\hat{\boldsymbol{\mu}})/(\kappa_0 + n)$ and variance $n/\kappa_n(\mathbf{1}'\mathbf{V}_\lambda^{-1}\mathbf{1})^{-1}\mathbf{C}$, where $\kappa_n = \kappa_0 + n$. Given \mathbf{C} and \mathbf{V}_λ , the density of the data then becomes

$$-2 \log p(\mathbf{Y}|\mathbf{C}, \mathbf{V}_\lambda) = nd \log(2\pi) + n \log |\mathbf{C}| + d \log |\mathbf{V}_\lambda| + d \log(\kappa_n/\kappa_0) + \text{tr}(\mathbf{C}^{-1}S(\hat{\boldsymbol{\mu}})), \quad (\text{A1})$$

where

$\hat{\boldsymbol{\mu}} = \{(\mathbf{I} \otimes \mathbf{1})'(\mathbf{C} \otimes \mathbf{V}_\lambda)^{-1}(\mathbf{I} \otimes \mathbf{1})\}^{-1}(\mathbf{I} \otimes \mathbf{1})'(\mathbf{C} \otimes \mathbf{V}_\lambda)^{-1}\text{vec}(\mathbf{Y})$ is the MLE of $\boldsymbol{\mu}$ and

$$S(\hat{\boldsymbol{\mu}}) = (\mathbf{Y} - \mathbf{1}\hat{\boldsymbol{\mu}})' \mathbf{V}_\lambda^{-1} (\mathbf{Y} - \mathbf{1}\hat{\boldsymbol{\mu}}) + \frac{\kappa_0}{\kappa_n} (\mathbf{1}' \mathbf{V}_\lambda^{-1} \mathbf{1}) (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)' (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)$$

is a multivariate sum of squares. Because $\hat{\boldsymbol{\mu}}$ depends on \mathbf{C} , there does not seem to be any tractable conjugate prior for \mathbf{C} , unfortunately. However, we can approximate $\hat{\boldsymbol{\mu}}$ by $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \dots, \tilde{\mu}_d)$, where $\tilde{\mu}_j$ is the estimated ancestral mean for trait j when ignoring the other traits, that is, when assuming that traits are independent. Estimates $\hat{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\mu}}$ are expected to be close unless correlations in \mathbf{C} are very large. Because $\tilde{\boldsymbol{\mu}}$ does not depend on \mathbf{C} , plugging in $S(\tilde{\boldsymbol{\mu}})$ for $S(\hat{\boldsymbol{\mu}})$ in (A1) allows us to recognize a conjugate distribution. Namely, we use an inverse Wishart distribution for \mathbf{C} with prior covariance matrix $\boldsymbol{\Lambda}_0$ and ν_0 degrees of freedom. Given the data and \mathbf{V}_λ , \mathbf{C} is then approximately inverse Wishart with $\nu_n = \nu_0 + n$ degrees of freedom and covariance $\boldsymbol{\Lambda}_n = \boldsymbol{\Lambda}_0 + S(\tilde{\boldsymbol{\mu}})$. Thus, we obtain the following approximate marginal likelihood $\tilde{p}(\mathbf{Y}|\mathbf{V}_\lambda)$, where $\boldsymbol{\mu}$ and \mathbf{C} have been integrated out:

$$-2 \log \tilde{p}(\mathbf{Y}|\mathbf{V}_\lambda) = nd \log \pi + d \log \frac{\kappa_n}{\kappa_0} - 2 \log \frac{\Gamma_d(\nu_n/2)}{\Gamma_d(\nu_0/2)} - \nu_0 \log |\boldsymbol{\Lambda}_0| + \nu_n \log |\boldsymbol{\Lambda}_n| + d \log |\mathbf{V}_\lambda|.$$

The assumption of a shared λ across all traits is unreasonable, however. Alternatively, trait k can have phylogenetic signal λ_k , with covariance $C_{kl}\mathbf{V}_{\sqrt{\lambda_k\lambda_l}}$ between trait k and trait l (Freckleton 2012; Ho and Ané 2014). In this model, we can rewrite $\text{cov}(Y_{ik}, Y_{jl}) = C_{kl}(\sqrt{\lambda_k\lambda_l}V_{1,ij} + (1 - \sqrt{\lambda_k\lambda_l})V_{0,ij})$. Here, \mathbf{V}_1 is calculated as when $\lambda = 1$ (between-species variance only) and $\mathbf{V}_0 = \mathbf{I}$ as when $\lambda = 0$ (within-species variance only). Under this model, $\text{vec}(\mathbf{Y})$ has covariance $\mathbf{W} = \mathbf{C}_1 \otimes \mathbf{V}_1 + \mathbf{C}_2 \otimes \mathbf{I}$, where $\mathbf{C}_1 = \mathbf{D}\mathbf{C}\mathbf{D}$, $\mathbf{C}_2 = \mathbf{C} - \mathbf{C}_1$, and \mathbf{D} is a diagonal matrix with terms $\sqrt{\lambda_k}$ on the diagonal. We can again set a conjugate prior for $\boldsymbol{\mu}$ given \mathbf{W} : $\boldsymbol{\mu} \sim N(\boldsymbol{\mu}_0, \frac{n}{\kappa_0}\tilde{\mathbf{W}})$. Here, $\tilde{\mathbf{W}} = \sum_{i,j=1}^n \tilde{\mathbf{W}}_{ij}$ is constructed by splitting the matrix \mathbf{W}^{-1} of size $nd \times nd$ into $d \times d$ submatrices $\tilde{\mathbf{W}}_{ij}$, one for each pair of individuals $i, j = 1, \dots, n$. Given \mathbf{W} and the data, $\boldsymbol{\mu}$ has a Gaussian posterior distribution with mean $\boldsymbol{\mu}_n = (\kappa_0\boldsymbol{\mu}_0 + n\hat{\boldsymbol{\mu}})/(\kappa_0 + n)$ and variance $n/\kappa_n\tilde{\mathbf{W}}$. Given \mathbf{C} , the tree, and λ_k values, the likelihood becomes

$$-2 \log p(\mathbf{Y}|\mathbf{C}, \mathbf{V}_1, \lambda_1, \dots, \lambda_d) = nd \log(2\pi) + nd \log |\mathbf{W}| + d \log \frac{\kappa_n}{\kappa_0} + (\text{vec}(\mathbf{Y}) - \hat{\boldsymbol{\mu}}' \otimes \mathbf{1})' \mathbf{W}^{-1} (\text{vec}(\mathbf{Y}) - \hat{\boldsymbol{\mu}}' \otimes \mathbf{1}) + \frac{\kappa_0}{\kappa_n} (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)' \tilde{\mathbf{W}} (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)'$$

Given the complexity of \mathbf{W} , we could not find a conjugate prior that would allow us to integrate out \mathbf{C} analytically. This causes some extra computational burden for the MCMC to sample \mathbf{C} and this model was not implemented.