Syst. Biol. 66(5):799–812, 2017

DOI:10.1093/sysbio/syw117
Advance Access publication March 11, 2017

Species Delimitation with Gene Flow

NATHAN D. JACKSON[1,*], BRYAN C. CARSTENS[2], ARIADNA E. MORALES[2], AND BRIAN C. O'MEARA[1]

[1]*Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, 442 Hesler Biology Building, Knoxville, TN 37996, USA and*
[2]*Department of Evolution, Ecology and Organismal Biology, Ohio State University, 318 W. 12th Avenue, Columbus, OH 43210, USA*
[*]*Correspondence to be sent to: Center for Genes, Environment, and Health, National Jewish Health, Denver, CO 80206, USA; E-mail: jacksonN@njhealth.org.*

Received 7 December 2015; reviews returned 9 December 2016; accepted 13 December 2009
Associate Editor: John McCormack

Abstract.—Species are commonly thought to be evolutionarily independent in a way that populations within a species are not. In recent years, studies that seek to identify evolutionarily independent lineages (i.e., to delimit species) using genetic data have typically adopted multispecies coalescent approaches that assume that evolutionary independence is formed by the differential sorting of ancestral alleles due to genetic drift. However, gene flow appears to be common among populations and nascent species, and while this process may inhibit lineage divergence (and thus independence), it is usually not explicitly considered when delimiting species. In this article, we apply Phylogeographic Inference using Approximate Likelihoods (PHRAPL), a recently described method for phylogeographic model selection, to species delimitation. We describe an approach to delimiting species using PHRAPL that attempts to account for both genetic drift and gene flow, and we compare the method's performance to that of a popular delimitation approach (BPP) using both simulated and empirical datasets. PHRAPL generally infers the correct demographic-delimitation model when the generating model includes gene flow between taxa, given a sufficient amount of data. When the generating model includes only isolation in the recent past, PHRAPL will in some cases fail to differentiate between gene flow and divergence, leading to model misspecification. Nevertheless, the explicit consideration of gene flow by PHRAPL is an important complement to existing delimitation approaches, particularly in systems where gene flow is likely important. [approximate likelihoods; coalescent simulations; genealogical divergence index; *Homo sapiens*; isolation-with-migration; multispecies coalescent; *Sarracenia; Scincella*.]

Species are a foundational unit of analysis in biology. Whether the goal is to retrace the evolution of traits on a phylogenetic tree or to prioritize taxonomic groups that are most in need of special management status, inferences from the fields of evolutionary biology, ecology, and conservation depend on how species units are defined. Despite this practical importance, assigning groups of organisms to species—species delimitation—is one of the most challenging objectives of phylogenetic and population genetic research (Coyne and Orr 2004). This is in large part because most biologists consider species to be separate evolutionary lineages but often disagree regarding the best criteria for recognizing them (de Queiroz 2007). In addition, some of these criteria, such as reproductive isolation can be difficult to establish in many taxa.

Multilocus genetic data provide a powerful line of evidence for delimiting species by allowing us to identify distinct evolutionary lineages within a sample (Fujita et al. 2012). In recent years, several analytical approaches have been developed that use coalescent models (Kingman 1982; Rannala and Yang 2003) to infer the most probable model of species limits given a multilocus dataset (e.g., Knowles and Carstens 2007a; O'Meara 2010; Yang and Rannala 2010; Ence and Carstens 2011; Grummer et al. 2014; Jones et al. 2015). While the specifics of these methods vary, all identify independent lineages by modeling the differential sorting of ancestral alleles in isolated populations due to genetic drift. When genetic drift is the primary evolutionary process that influences allele frequencies, simulation results suggest that these methods can accurately delimit species in many cases (e.g., Camargo et al. 2012; Rittmeyer and Austin 2012). However, the evolutionary process of gene flow should also be explicitly considered when inferring species limits (e.g., Ence and Carstens 2011; Camargo et al. 2012), as this process (like incomplete lineage sorting) can result in shared polymorphism across lineages. Gene flow can halt or reduce genetic divergence that accumulates due to population isolation (Wright 1931), and theoretical work shows that when this process is present, but excluded from coalescent species tree inference, phylogenetic accuracy can be reduced (Eckert and Carstens 2008; Leaché et al. 2014). That divergence with gene flow has been observed in multiple groups (Pinho and Hey 2010) suggests that the accuracy of species delimitation methods may also decline when ongoing gene flow is ignored (e.g., Ence and Carstens 2011; Camargo et al. 2012).

In large part due to the computational complexity introduced by adding migration parameters to the already complex models used in species delimitation, it has thus far only been possible to consider gene flow using customized modeling in an ABC approach (e.g., Camargo et al. 2012). Thus, the current strategy taken by most researchers is either to ignore gene flow when delimiting species or to infer species boundaries and rates of gene flow (and often, the species tree) in separate steps. While these approaches may work well in some systems (e.g., Zhang et al. 2011; Camargo et al. 2012; Heled et al. 2013; Burbrink and Guiher 2015), given covariance among the species phylogeny (including divergence times), species boundaries, and migration rates, these parameters would ideally be estimated in a single analysis.

Jackson et al. (in press) have proposed an analytical framework (Phylogeographic Inference using Approximate Likelihoods; PHRAPL) for exploring the relative fit of phylogeographic datasets to a broad and flexible model space. By analyzing gene tree topologies rather than full sequence datasets, PHRAPL is able to consider a large number of complex models that incorporate gene flow in addition to other parameters involved in the speciation process, potentially making the method well suited to inferring species limits. PHRAPL estimates the likelihood of an observed dataset under a particular demographic model by first simulating a distribution of coalescent gene trees under that model using the program ms (Hudson 2002). The inputted dataset consists of a set of gene tree topologies along with *a priori* assignments of individuals to populations or species. The demographic model contains parameters that describe the timing of coalescent events among populations (i.e., the species tree), migration rates (M), population growth rates (g), and/or differences in effective population size ($\theta = 4N_e$). All parameters can be specified among tips and/or ancestral populations. The proportion of simulated gene tree topologies that match the observed gene tree is then used to approximate the log-likelihood of the model given the observed data. Because approximate likelihoods can be calculated quickly by PHRAPL, a large number of demographic models can be compared and ranked using Akaike information criterion (AIC), facilitating model selection. PHRAPL also optimizes parameters for each model by estimating log-likelihoods across a user-specified grid of values (called a "grid search"), which contains all possible combinations of values specified for each imposed parameter. Thus, model selection (i.e., inferring the set of parameters) and parameter optimization (i.e., inferring parameter values) are conducted jointly. Simulations have demonstrated that PHRAPL is generally effective at identifying the optimal model given moderate amounts of data (e.g., on the order of 10–100 loci), although parameter estimation is not likely to be as accurate as with methods that apply a Markov chain Monte Carlo (MCMC) approach (Jackson et al. in press).

In this article, we extend PHRAPL to address the challenge of delimiting species while accounting for the possibility of gene flow. We first describe how delimitation models can be constructed and compared using PHRAPL, and then test the performance of this approach using datasets simulated under a variety of sizes, coalescent times, and migration rates. To illustrate the application of PHRAPL to empirical data, we apply the method to datasets from lizards and pitcher plants. We also apply the method to humans, which we know to comprise a single species, as a way to test for propensity to Type 1 error (e.g., oversplitting). Finally, we compare the performance of PHRAPL to that of a popular method for species delimitation (BPP; Yang and Rannala 2010, 2015), which uses a full Bayesian framework for inferring species, but which does not account for gene flow among taxa.

Methods

Specifying Delimitation Models in PHRAPL

If two sampled populations represent distinct evolutionary lineages, then gene trees from these populations are expected to better fit a model that includes divergence among populations than a model that excludes divergence (i.e., when $t = 0$, where t is the divergence time). Thus, in PHRAPL, which compares the fit of gene trees to a set of demographic models, nested species delimitation models can be constructed from any given model of species history (i.e., species tree) by systematically collapsing the nodes of the tree. Similarly, non-nested species delimitation models can be compared by first constructing nested sets of models for different species histories (e.g., different species trees) and then combining these models into a single analysis. In this way, all possible species relationships and species delimitations—both with and without accompanying migration—can be compared using AIC weights calculated for each model (e.g., Carstens and Dewey 2010), although the practical number of models considered may be limited by computational resources and/or the patience of the researchers.

Simulation Testing

Performance of PHRAPL for species delimitation.—We used simulated data to test the ability of PHRAPL to infer the correct delimitation-migration model. We simulated two rooted, three-species histories where (1) species A and B coalesce at time t_{AB} in the past, followed by coalescence of ancestral species AB with species C at time t_{ABC} (Fig. 1a) and where (2) the same branching history above occurs, but with constant symmetrical migration between species A and B until t_{AB} (Fig. 1c). We simulated 20 individuals for each of the three populations. We also simulated a two-species history where A and B were merged, becoming a single evolutionary lineage in the present (Fig. 1b). These simulated datasets allowed us to assess rates of accuracy for PHRAPL. Specifically, we inferred rates of false negatives (i.e., failing to detect two species where two are present) and false positives (i.e., falsely inferring two species when the data have evolved within a single lineage).

We varied model parameters to represent a wide range of empirical systems for which the question of species boundaries is relevant. We varied t_{AB} across six values: 0.05, 0.125, 0.25, 1, 2, and 4, where values are in units of $4N$ (diploid effective population size), and t_{ABC} was set to 2.5, 2.5, 2.5, 2.5, 5, and 10, respectively (Fig. 1). For the two-species history t_{ABC} was set to 2.5. We also varied migration, M (0.5, 2, 5, and 10), where M is in units of $4Nm$ (the number of migrants per generation), and varied the number of loci simulated per dataset (1, 4, 10, and 50). Each treatment combination (96 in total) was simulated 50 times using different starting seeds. To approximate empirical sequence data, we simulated genealogies using the program ms and

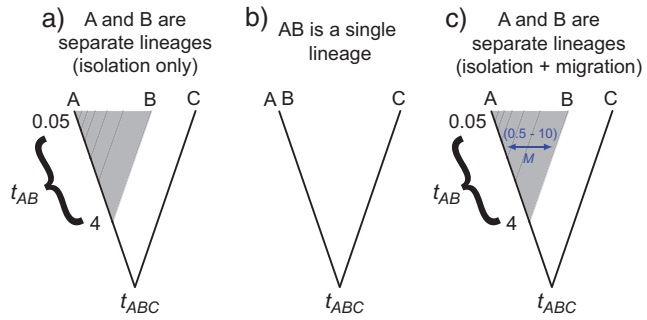

FIGURE 1. Three histories underlying simulated datasets that were analyzed using PHRAPL: a) taxa A and B diverged at time t_{AB} in the past; b) taxa A and B are a single panmictic lineage; and c) taxa A and B diverged at time t_{AB} in the past, but continued to share migrants at rate M. In histories (a) and (c), species coalescent times t_{AB} were varied between 0.05 and 4 (shown in the shaded region), where times t are in units of $4N$. In history (c), M was varied between 0.5 and 10, where $M = 4Nm$. Note that the branch lengths preceding t_{ABC} in these trees were adjusted according to the t_{AB}simulated (see text); however, this scaling is not shown here.

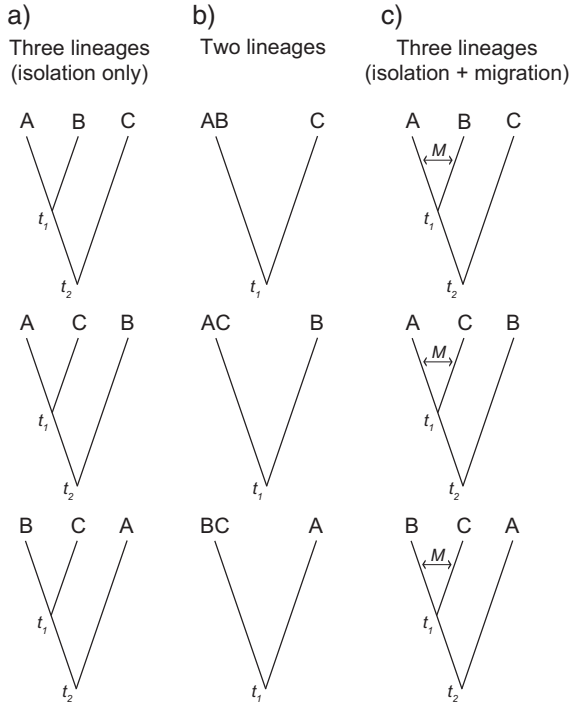

FIGURE 2. The set of models fitted to simulated datasets using PHRAPL. $M = 4Nm$ and t_1 and t_2 are population coalescent times in units of $4N$.

subsequently evolved DNA sequences along branches using the program Seq-Gen (Rambaut and Grassly 1997). Matching settings used in earlier simulation testing (Knowles and Carstens 2007b; Ence and Carstens 2011), sequences were evolved using the HKY model, 500 bp per locus, base pair frequencies = 0.3, 0.2, 0.2, and 0.3 (for A, C, G, and T), and transition/transversion ratio = 3. Two different levels of genetic diversity were simulated: $\theta = 0.005$ and 0.025. We then inferred gene trees from sequence datasets using RAxML 7.2.6 with five replicate searches, rapid hill-climbing, and the GTRGAMMA model (Stamatakis 2006).

We constructed two sets of models to explore the accuracy of PHRAPL in delimiting species using these simulated datasets. The first model set (6 models) contained the true underlying topological history, as well as all other possible histories in which species A and B were and were not collapsed into one (Fig. 2a, b). As discussed above, two-species models were implemented here as special cases of the three-species models, but where t_{AB} was set to zero. These models assume that divergence occurs only due to processes taking place within taxa (i.e., excluding migration) and are equivalent to the models considered by BPP (Yang and Rannala 2015). The second model set (9 models) included all the models in the first model set plus the three possible three-species histories to which symmetrical migration was added between sister populations A and B (Fig. 2a–c). Prior to model comparison using PHRAPL, we randomly subsampled all datasets 10 times where each subsample replicate included four tips per species (12 tips per tree). Parameter optimization was performed using a grid of values. We used eight values for t (0.10, 0.22, 0.46, 1.00, 2.15, 4.64, 10.00, and 15.00) and seven values for M(0.10, 0.22, 0.46, 1.00, 2.15, 4.64, 10.00). We set nTrees (the number of trees simulated per parameter combination) to 10,000.

Comparing the performance of PHRAPL with that of BPP.— We compared the performance of PHRAPL to that of BPP, one of the leading software packages for species delimitation under the multispecies coalescent model (Yang and Rannala 2015). BPP uses a full Bayesian approach with reversible-jump Markov chain Monte Carlo to compare the fit of models positing different ways of collapsing putative species to sequence data. A recent extension of the method also allows for testing delimitation hypotheses across different tree topologies, such that species limits and relationships can be inferred simultaneously (Yang and Rannala 2015). However, unlike PHRAPL, BPP does not explicitly model gene flow, although arguments have been put forward that BPP nonetheless behaves well in the face of migration (Zhang et al. 2011, 2014).

We analyzed all simulated datasets (treatments shown in Fig. 1a–c) using BPP. For all datasets we placed a gamma prior on $\theta \sim G(2, 400)$, with a mean equal to our simulated value ($\theta = 0.005$; the higher diversity sequence dataset was not analyzed). Our prior on root age was based on the simulated value for t_{ABC}, with $\tau \sim G(2, 160)$, $\tau \sim G(2, 80)$, and $\tau \sim G(2, 40)$ for $t_{ABC} = 2.5$, 5, and 10, respectively. We allowed fine-tune parameters to be automatically adjusted, enabled the nearest neighbor interchange algorithm to allow inference of the phylogeny, set the species model prior to zero, used delimitation algorithm 0 (with default e value), and sampled every two of 100,000 iterations after a burn-in of 10,000 iterations. We analyzed each dataset 10 times to assess consistency of results.

Empirical Examples

Scincella lateralis.—We analyzed data from North American ground skinks (*Scincella lateralis*) using PHRAPL. Although *S. lateralis* is currently classified as a single widespread species, recent phylogeographic studies (Jackson and Austin 2010, 2012) suggest that the species is composed of at least three cryptic parapatric lineages ("Eastern," "Central," and "Western", all largely distributed in the southeastern United States). For this dataset, as well as for the other empirical datasets analyzed in this article, we followed the original papers when assigning individuals to populations. The data consists of eight nuclear loci (ranging from 442 to 837 bp) collected from between 68 and 80 individuals. Maximum likelihood trees were estimated using 10 replicate searches of rapid hill-climbing using the GTRGAMMA model in RAxML.

We assessed the fit of the skink data to 55 demographic models (see these models in Supplementary Fig. S1, available on Dryad at http://dx.doi.org/10.5061/dryad.t8016) for three lineages. This model set included all possible relationships and delimitation scenarios without migration (7 models) and with symmetrical migration (24 models). For models with migration, we not only considered full migration models (migration among all populations), but also all possible partial migration scenarios (e.g., including a model with migration only between Eastern and Western or a model with migration between all populations except Central and Eastern). As above, we also implemented all isolation-with-migration (IM) models as secondary contact models, where migration ceases after $t = 0.1$ (24 models). We subsampled four tips per population, set nTrees to 100,000, and performed 10 replicate analyses. Grid values used were $t = 0.30$, 0.58, 1.11, 2.12, 4.07 and $M = 0.10$, 0.22, 0.46, 1.00, 2.15, 4.64.

For comparison, we also analyzed the ground skink dataset using BPP. We set the prior on root τ to G(2, 1000) and the prior on θ was set to G(2, 400), with a mean (0.005) equal to the estimated nucleotide diversity for this group (Jackson and Austin 2010). All other aspects of the analysis were kept the same as used for the simulated datasets above except that we sampled 500,000 iterations and increased burn-in to 50,000.

Sarracenia alata.—We also applied PHRAPL to sequence data collected from pitcher plants distributed in the southeastern United States (*Sarracenia alata*). *Sarracenia alata* has traditionally been classified as a single species. However, Carstens and Satler (2013) presented evidence for the existence of two cryptic species corresponding to two disjunct populations on opposite sides of the Mississippi River near the Gulf Coast. We further evaluate species limits here using PHRAPL. In addition to including the two disjunct populations ("Eastern" and "Western"), we also considered a third population, previously inferred to be distinct using a STRUCTURE analysis (Zellmer et al. 2012), as a putative lineage. This population is isolated from other Western samples by the Red River, which could act as a strong dispersal barrier. Thus we analyzed this three-population dataset (20 loci, ranging from 164 to 403 bp, from 80 samples, averaging 47 samples per locus) with PHRAPL using the same model set, grid, and analysis specifications used for ground skinks.

We carried out a BPP analysis for *S. alata* using the approach used for ground skinks except that the prior on θ was set to G(2, 500), with a mean equal to the mean estimate of nucleotide diversity ($\pi = 0.004$) calculated using PopGenome (Pfeifer et al. 2014).

Homo sapiens.—There exists no agreed-upon threshold of population divergence or migration beyond which we define a group as one species or two (de Queiroz 2007). Consequently, it can be challenging to assess the propensity of a species delimitation method to oversplitting (e.g., to falsely delimiting structured populations within a species). Whether a genetic pattern signals biological species is in fact the question to which we turn to delimitation methods for an answer, and thus there tends to be some circularity to the testing of delimitation methods using datasets for which we know that divergence has occurred, but where we must rely on the method to ascertain whether this amounts to "species-level divergence." One solution is to analyze data from a system that contains population genetic structure, but where this structure is well understood to be intraspecific. Since there is no species that has been the focus of more genetic investigation than our own, we analyzed data from *Homo sapiens* as a sort of "sanity check": if PHRAPL delimits multiple species of humans, this would be strong evidence that PHRAPL is overly prone to Type 1 statistical error, where populations are being wrongly split into species.

We thus applied PHRAPL to a human dataset comprising DNA sequences from 50 loci (ranging from 415 to 960 bp) available for four widely sampled, geographically defined populations, from which participants were identified as having heritage: 10 samples from Africa, 10 samples from Europe, 10 samples from Asia, and 12 samples from South and Central America (Yu et al. 2002; Fagundes et al. 2007). Here we simply use the full datasets and groupings of the original studies. We chose this dataset as it was the best available in terms of the sample sizes and number of sequenced loci. It is important to point out that these four groups are ethnically diverse and are not "populations" in any biological sense. In fact, the major finding of the paper from which most of these data originated was that more genetic diversity exits within the African sample than between samples (Yu et al. 2002) and it is well understood that most non-African genetic diversity is a subset of African diversity (Tishkoff and Kidd 2004; Jakobsson 2008; Li et al. 2008). Thus, there is no question that all the samples in these "groups" originate from a single species, which is why this is a good dataset for testing for Type 1 error in delimitation methods.

We aligned sequences using MUSCLE (Edgar 2004) and inferred haplotype phase using PHASE 2.1.1 (Stephens et al. 2001). We inferred gene trees using RAxML with 10 replicate searches, rapid hill-climbing, and the GTRGAMMA model. Trees were rooted using midpoint rooting. We analyzed 100 replicate subsamples per locus, each subsample comprising three tips per population. Our model set contained all possible topologies and delimitation scenarios involving the four populations, with and without full symmetrical migration among the tips (87 models; see these models in Supplementary Fig. S2). We based the range of grid values used in these analyses on parameter estimates from Fagundes et al. (2007), which were derived from these datasets. The minimum coalescent time was set to be the lower bound of the highest posterior density (HPD) interval for the estimated time at which the Americas were colonized (7647 years) divided by the upper bound of the HPD interval for θ within the Native American population (13,740), the least diverse of the populations (thus, $t_{\min} = T/4N_e = 7647/[4*13,740] = 0.14$). Similarly, the maximum coalescent time was set to be the upper bound of the HPD interval for the estimated time at which humans dispersed from Africa (70,937 years) divided by the lower bound of the HPD interval for θ within the ancestral population (6604), prior to diversification ($t_{\max} = 70,937/[4*6604] = 2.69$). We used four grid values within this range ($t = 0.14$, 0.38, 1.00, 2.69). We also selected four grid values for migration ($M = 0.10$, 0.27, 0.74, 2.00), which incorporate all median Nm estimates from Fagundes et al. (2007). The analysis was repeated 10 times, with nTrees set to 100,000 in all runs.

We also analyzed the human dataset using BPP. We used a θ gamma prior of G(2, 2975) whose mean equals the mean estimate of nucleotide diversity ($\pi = 0.00067$) calculated across the 50 loci using PopGenome. Our prior on root τ was G(1, 535), with a mean of 0.00187, which is based on a 170,000 year coalescence time for humans (Ingman et al. 2000) and an average mutation rate of 1.1×10^{-8} mutations per site per generation (Roach et al. 2010). We also tried a root τ prior of G(1, 3500), which was used by Yang and Rannala (2010). All other aspects of the analysis followed those used with the empirical datasets above.

Inferring Species Delimitation

The operational criterion that PHRAPL uses for identifying candidate species (i.e., evolutionary lineages) with model selection is similar to that used by other approaches that apply a multispecies coalescent framework for species delimitation (O'Meara 2010; Yang and Rannala 2010; Ence and Carstens 2011). Typically, the process of speciation proceeds when genetic isolation among populations is sufficiently strong and long lived such that the rate at which mutations differentially accrue among populations exceeds the rate at which gene flow disperses them. In PHRAPL, we thus identify groups as candidate species when the genetic divergence resulting from this process becomes statistically demonstrable in a model comparison framework: if a two-species model garners substantially more support than a single-species model, two species are inferred.

However, complications to delimitation decisions arise when a statistically favored two-species model also contains a migration parameter (an IM model). First, the detection of gene flow among groups suggests that reproductive isolation may not be complete. Under a strict biological species concept, this would cause one to conclude that a single species is present, despite there being two evolutionary lineages in the inferred model. However, many if not most biologists would argue that speciation may still proceed in the face of some gene flow under certain conditions (Coyne and Orr 2004). Moreover, although gene flow was important over the course of divergence, reproductive isolation may have recently developed. So if some gene flow may be permissible among diverging species, inferences about species boundaries will often depend on the *amount* of gene flow in the supported model. For example, lineages that are inferred to share migrants at a very high effective rate (e.g., $Nm \gg 10$) should likely be considered a single species (Wright 1931). Further difficulties arise because both the gradual process of allele sorting due to drift and allele sharing due to migration can result in similar patterns of gene tree nonmonophyly (Funk and Omland 2003), potentially confounding these processes. This may bias estimates of migration and coalescence time under some conditions, resulting in poor model choice (discussed later on in this article).

For these reasons, discrete model selection alone may not always provide the best estimate of lineage independence when both gene flow and genetic drift are modeled. When inferring species limits using PHRAPL, it is thus important to also consider parameter estimates derived from a grid search, as these values are informative of the species boundaries we aim to infer. To facilitate this, we developed the genealogical divergence index (*gdi*), which can be calculated from estimates of migration rate and coalescence time obtained from a PHRAPL analysis. This index provides an estimate of the overall degree of genetic divergence between two taxa due to the combined effects of genetic isolation and gene flow and is useful in the interpretation of the results from model selection and parameter estimation. If one samples two gene copies from species 1 and one gene copy from species 2, then let G_1 be the resulting genealogy in which the two gene copies from species 1 are sister to each other. We define the unscaled GDI_u to be

$$GDI_u = \mathbb{P}(G_1 | M_1, M_2, t)$$

where M_1 and M_2 are bi-directional migration rates because the divergence of the species at time t. Rather than analytically calculating GDI_u, we approximate it using ms (Hudson 2002) such that

$$gdi_u = \text{observed}(GDI_u)$$

For a given M_1,M_2, and t, we iteratively simulate coalescent trees with three gene copies under the two-taxon species tree model described above, and then calculate the proportion of simulated trees in which the two gene copies originating from species 1 are sisters. The index is then scaled to be between 0 (panmixia) and 1 (strong divergence) using

$$gdi = [\text{observed}(gdi_u) - \min(gdi_u)]/[\max(gdi_u) - \min(gdi_u)]$$

where $\min(gdi_u) \approx 1/3$ (with three tips, under panmixia, species 1 monophyly is expected ~1/3 of the time) and $\max(gdi_u) = 1$ (with extreme isolation, species 1 will always be monophyletic). The *gdi*, along with confidence intervals, can be calculated using the CalculateGdi function within PHRAPL.

The *gdi* is similar to the genealogical sorting index (*gsi*; Cummings et al. 2008) in that it calculates the degree of nonmonophyly in a set of gene trees, and in fact these two indexes are highly correlated ($R^2 = 0.9$) and perform similarly if used to delimit species based on a range of theoretical cutoffs (see Supplementary Figs S3a, b). However, the two indexes differ in two important ways. First, PHRAPL aims to delimit species while simultaneously understanding those aspects of demographic history that have given rise to these groups, and the *gdi* explicitly incorporates these inferred processes (in the form of estimated parameter values). The *gsi*, in contrast to this, measures divergence directly from genetic data, and thus does not presuppose any information about the underlying causes of divergence. Secondly, the *gdi* measures divergence between two focal sister populations or groups, which is the level at which delimitation questions arise. However, the *gsi* measures the exclusivity of a focal taxon relative to the entire tree, and thus the degree of divergence inferred for that taxon can depend on patterns of genetic structure within other parts of the tree (Winter et al. 2016).

The *gdi* is continuous (as is the speciation process itself), and thus, while informative of where a taxon lies on the path to speciation, it is not an ideal metric by which to delimit species. It is, however, a useful way to explore how accurately PHRAPL can delimit species when this delimitation is solely based on parameter estimates rather than on model selection. We thus calculated the *gdi* for each simulated treatment using model-averaged estimates of coalescence time and migration rate. We then compared species delimitation based on these values, assuming a range of *gdi* "delimitation cutoffs," with *gdi*-based delimitation derived from the generating models' parameter values.

Availability

PHRAPL is written in R and Perl and can be downloaded from https://github.com/omeara/phrapl. A tutorial is included in Supplementary Materials.

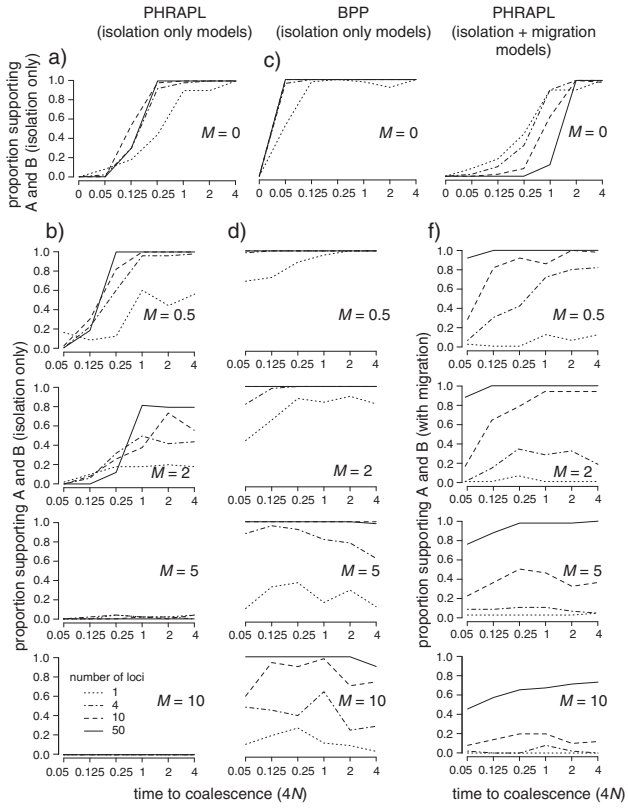

FIGURE 3. Model selection results for simulated datasets using PHRAPL (a, b, e, and f) and BPP (c, d) where $\theta = 0.005$. Panels (a, b) show the proportion of replicate PHRAPL analyses supporting divergence between species A and B when datasets were analyzed against isolation-only models (models in Fig. 2a, b). Proportions are shown across datasets produced under increasing coalescence time t (x-axis), increasing migration rate M (top to bottom), and increasing numbers of loci (line solidity). Panels (c, d) show these same proportions when the datasets are analyzed using BPP. Panels (e, f) show results when datasets were analyzed under an expanded model set that includes migration (models in Fig. 2a–c). In contrast to the other plots, plots in panel (f) give the proportion of replicate analysis in which the true isolation plus migration model was supported.

RESULTS

Simulation Testing

Performance of PHRAPL for species delimitation.—(a) Analyzing data simulated with and without migration using isolation-only models.

In the case where true coalescent time (t) is zero, PHRAPL supported the single-species model in 96% of cases with one locus and in 100% of cases with more than one locus (where the inferred model is the one receiving the highest AIC weight). Thus, when only isolation is considered, PHRAPL does not tend to infer two species when a single panmictic lineage is present (Fig. 3a). A single species was also typically favored when t was small ($t \leqslant 0.125$ when $\theta = 0.005$; Fig. 3a; $t \leqslant 0.05$ when $\theta = 0.025$; Supplementary Fig. S4a) or regardless of t when migration rate (M) was high ($M \geqslant 5$; Fig. 3b and Supplementary Fig. S4b). The number of loci analyzed did not strongly affect performance, aside from

slightly poorer sensitivity when data from a single locus were analyzed.

(b) Analyzing data simulated with and without migration using both isolation-only and IM models.

When data generated under IM models were analyzed using both isolation-only and IM models, the true IM model was supported in ~100% of replicates, if 50 loci were sampled and migration was not very high ($M < 10$) (Fig. 3f and Supplementary Fig. S4d). When fewer loci are analyzed, the true model was identified in cases where migration was moderate to low ($M \leqslant 2$) and divergence was moderate to high ($t \geqslant 0.125$). However, results were poor in all cases when data from a single locus were analyzed (Fig. 3f and Supplementary Fig. S4d).

When data were generated under isolation-only ($M = 0$), but analyzed using both isolation-only and IM models, the amount of population divergence required by PHRAPL to delimit lineages increased from ~0.125–0.25 $4N$ generations (obtained with an isolation-only model set; Fig. 3a) to ~0.25–2 $4N$ generations, depending on the number of loci (Fig. 3e) or value of θ (Supplementary Fig. S4c). At lower values of divergence, IM models that pair high rates of migration and overestimated coalescence times were usually favored over models that posit either isolation-only or a single AB lineage. Even in the case where the generating model includes $t = 0$, an IM model—rather than the single-lineage model—was supported in between 4% (1 locus) and 44% (50 loci) of replicates. This support for spurious IM models likely results in part from similarity in gene tree patterns produced by genetic drift and gene flow. For example, a model positing high migration and inflated divergence may yield a degree of tree nonmonophyly similar to that produced by the true generating model of recent divergence in isolation.

Because parameter estimates can be biased for IM models under some conditions, we generally recommend that the *gdi* (or some other metric that provides a good measure of the overall level of genetic divergence across taxa) be used to help interpret PHRAPL results when an IM model is inferred. When species delimitation was based on the *gdi* calculated from simulated and estimated parameter values, they were generally in agreement across a range of cutoffs (Fig. 4 and Supplementary S5), suggesting that *gdi* values based on PHRAPL parameter estimates are good approximations of *gdi* values derived from the generating model. When considering cutoffs, we selected an upper value (above which two lineages are inferred) and a lower value (below which a single lineage is inferred) to allow for an ambiguous inference when intermediate values are observed. This accommodates the gray zone that is inherent to the continuous speciation process (de Queiroz 2007). However, note that the selected cutoffs used are completely arbitrary and meant only to explore sensitivity of performance to the chosen set of values. When delimitation inference from simulated and estimated indexes disagreed, this was usually in the form of a tendency toward either excessive or insufficient confidence in the number of species. Categorical inference of the wrong number of species did not occur (Fig. 4 and Supplementary Fig. S5).

Comparing the performance of PHRAPL with that of BPP.—With 50 loci, BPP delimited species A and B in a high proportion of replicates (i.e., at or near 100%) regardless of the amount of migration or the depth of coalescence time, as long as $t > 0$ (Fig. 3C, D and Supplementary S4C, D). With fewer loci, the frequency of support for delimitation of A and B dropped with increasing migration. When $t = 0$ in the generating model, BPP always supported a single lineage. Note that results were similar regardless whether the proportion of analyses supporting the true model or the average posterior probability of the true model was used (data not shown).

When comparing BPP delimitation inference under the best-case scenario (i.e., the case of 50 loci) with the true underlying delimitation for each dataset (as defined by the range of *gdi* cutoffs), BPP increasingly overestimated the number of species with diminishing time to coalescence and with growing rates of migration (Fig. 4). Once migration is high ($M \geqslant 5$), nearly all BPP analyses infer two species (for A and B) whereas nearly all *gdi* values from the generating models indicate a single species. Thus, BPP is effective at identifying population isolation, even given levels of gene flow that are expected to be homogenizing (Wright 1931).

Empirical Examples

Scincella lateralis.—Two similar models contained most of the AIC weight in the PHRAPL analysis: both models include three species with a sister relationship between the adjacent Western and Central populations, including either constant migration (average wAIC = 0.28) or migration under secondary contact (average wAIC = 0.25) between nonsister Central and Eastern populations (Fig. 5; Supplementary Table S6). These inferred relationships and migration models are similar to those inferred in a previous study using species tree (*BEAST) and migration-divergence (IMa2) analyses (Jackson and Austin 2012). In that study, migration was also found to only occur among geographically adjacent populations—appearing to be particularly strong between Central and Eastern groups—and to be especially important in the recent past.

In contrast, all BPP analyses supported three species with a sister relationship between the Central and Eastern populations. Average posterior probability was 1.0 for both the delimitation and the topology. This alignment between Central and Eastern populations could in part be due to high gene flow between them, which is registered as recent coalescence by BPP.

Sarracenia alata.—In PHRAPL, most of the AIC weight (average wAIC = 0.89) supported one of three two-species models, where the two Western populations were collapsed into a single lineage: one model included constant migration between the two lineages (average

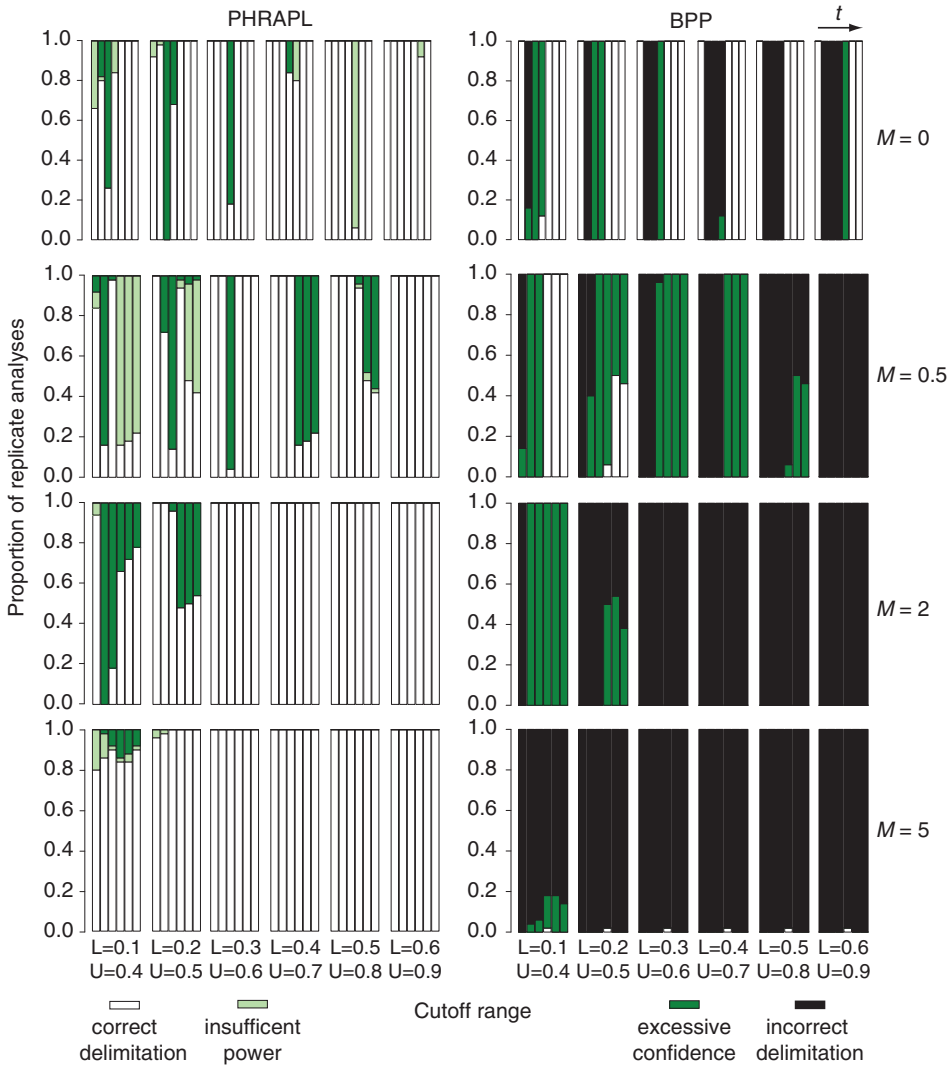

FIGURE 4. Accuracy of species delimitation when using the *gdi* in conjunction with PHRAPL and BPP. Each bar shows the frequency with which species delimitation based on the *gdi* from the true parameter values agrees with species delimitation based on, (left), the *gdi* from the estimated parameter values from PHRAPL or, (right), results from BPP. Only performance using 50 loci is shown. Along the *x*-axis, each group of bars gives results for different *gdi* cutoffs. These cutoffs consist of a lower bound (L), below which populations A and B are defined as a single species, and an upper bound (U), above which populations A and B are defined as separate species. Index values between bounds signal when the number of species is ambiguous. Bars within bar groups represent treatments of increasing coalescence time and each row represents a different migration rate. The frequency with which each treatment resulted in the correct, or partially correct, delimitation is shown using different bar shading: white = the inferred delimitation outcome matched the true outcome; light green/gray = ambiguity was inferred when the true delimitation is known (insufficient power); dark green/gray = delimitation was inferred (whether one or two species) when the truth was ambiguous (excessive confidence); and black = one species was inferred when there were two, or vice versa.

wAIC = 0.25), another model included migration upon secondary contact (average wAIC = 0.25), and the last model excluded migration (average wAIC = 0.40; Fig. 5). BPP supported a three-species model in all replicate analyses, with the two Western populations placed sister to one another (Fig. 5; Supplementary Table S8). Average posterior probability for the tree and number of species was 1.0 across replicates.

Homo sapiens.—The single species model garnered the most support in PHRAPL (Fig. 5; Supplementary Table S8). The ratio of the average wAIC of the single species model (0.14) to that of the next best model was 2.64, indicating that this model had over twice the support of any other. In contrast, BPP supported a four-species model in all 10 replicate runs, regardless of which of the two root τ priors we used, with an average posterior probability (pp) of 0.99 or 1.0 for the small or large root τ prior, respectively. All but one replicate run supported (((Asia, America), Europe), Africa) with an average pp of 0.92. The remaining replicate supported (((Europe, America), Asia), Africa) with a pp of 0.93.

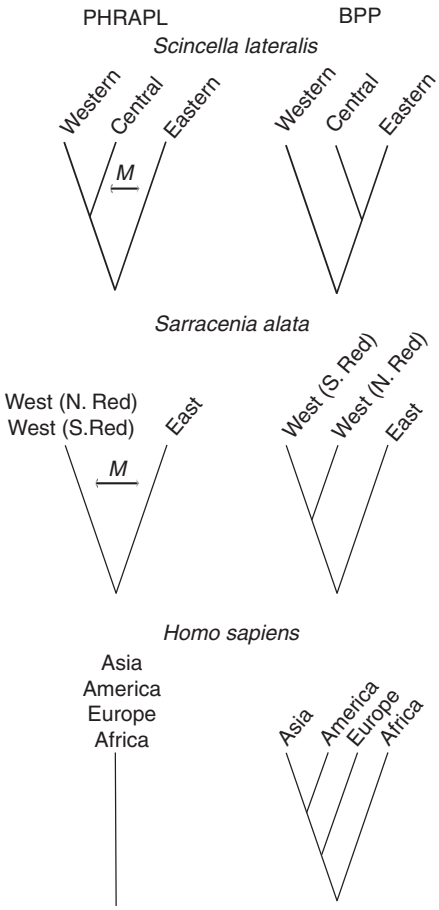

FIGURE 5. Inferred best models for three empirical datasets based on analysis using PHRAPL (left) and BPP (right). For *S. lateralis* and *S. alata*, the PHRAPL models shown are composites constructed from the top models inferred for each.

DISCUSSION

Populations that are considered to be candidate species will typically exhibit some evidence of genetic isolation, either in the form of phenotypic/behavioral/genetic differences or geographic separation. However, in recent years, it has become clear that species divergence can occur in the presence of gene flow (Hey 2006; Nosil 2008) and that taxa which have diverged in geographic isolation may come into secondary contact and resume genetic exchange (e.g., Zamudio and Savage 2003; Noonan and Gaucher 2006). Such opportunities for gene flow among groups not only can obscure the true evolutionary trajectory signaled by an observed level of divergence, but also can challenge our notions of what constitutes a species.

Despite the prevalence of gene flow in nature and its relevance to speciation, most methods aimed at inferring species boundaries do not explicitly account for it. Consequently, if two distinct species exchange alleles, a model that ignores migration may attribute the resulting increase in genetic similarity to either more recent divergence or higher effective population size in comparison to the actual values (e.g., Leaché et al. 2014). This may cause an analytical method to falsely lump two species into one. Alternatively, if the genetic signature of substantial gene flow between two once-divergent taxa is ignored or misattributed in a model, a single lineage composed of two highly connected populations may be falsely split into two species. Thus, given the central role of gene flow in the conceptual diagnosis of species, failing to account for it when delimiting species in practice may either obscure the mechanisms that underlie an observed level of divergence among taxa, or even result in high confidence for the wrong number of species.

Using PHRAPL for species delimitation is therefore appealing because this method can consider non-nested delimitation models that include both isolation and migration parameters. When isolation-only and IM models were considered, we found that PHRAPL was generally successful at inferring the history of divergence with gene flow, given low to moderate migration ($M \leqslant 2$) and adequate (>10 loci) data. Correct detection when migration rates are high requires more (>50 loci) data. Model selection was less accurate when the generating model excluded migration, unless coalescence time was deep ($t \geqslant 2$). In these instances, PHRAPL tended to attribute nonmonophyly that was actually caused by incomplete lineage sorting to nonmonophyly produced in part by gene flow, and thus was biased in favor of IM models with high migration over true models of recent isolation. It is not clear why the more complex spurious model is favored over the true simpler model in this portion of the parameter space, but if one blindly uses model selection to delimit species in these cases, one would wrongly infer distinct species. This bias can be somewhat reduced by capping migration at a moderate or low value. For example, capping M at 2 in the parameter grid results in 100% support of a single-species model when true $t = 0$. However, at moderate values of t, a bias toward selecting IM models remains (Supplementary Fig. S9). The practical nonidentifiability of incomplete lineage sorting and gene flow, while a difficult challenge whenever genetic divergence is modeled (e.g., Heled et al. 2013), is particularly expected when branch lengths are excluded from the data (as in PHRAPL), given that a lot of information concerning the degree of incomplete lineage sorting is contained in the branch lengths (Pamilo and Nei 1988; Maddison 1997). Thus, when jointly inferring species delimitation, migration, and divergence, an undertaking that is currently not available in a full likelihood or Bayesian framework, some regions of practical nonidentifiability will likely exist.

When delimiting species using PHRAPL, particularly if an IM model is inferred, it is important to not only consider the best supported model, but also parameter estimates, as these help one to understand the degree of divergence within a favored model and the processes that underlie this divergence. For example, if a supported IM model includes a high

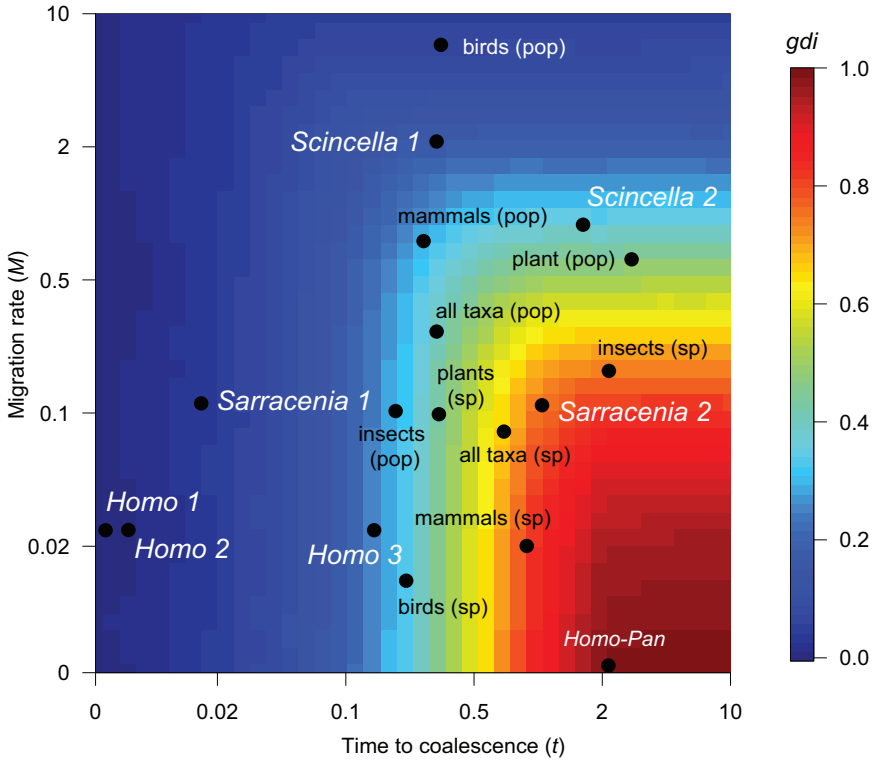

FIGURE 6. Contour plot showing the relationship between parameter values (coalescence time and migration rate) and the *gdi*. For the three empirical datasets analyzed in this study, model averaged parameter estimates from PHRAPL based on the first (1), second (2), and third (3) splits in the best supported tree are shown in enlarged font. Because ancestral migration was not included in the analyzed models, when calculating the *gdi* for the root node, ancestral migration was assumed to equal contemporary migration. For comparison, estimated values for several other taxa were also plotted. First, a point for the split between Homo and Pan is shown; coalescence time (t) for this split is based on 8.5 million years of separation (Benton and Donoghue 2007), an N_e of 47,500 (Schrago 2014), and a 20-year generation time ($(8.5\times10^6)/(4\times47{,}500)$) and migration rate is assumed to be zero. Four taxonomic groups were also plotted using data culled in a meta-analysis (Pinho and Hey 2010). Data were divided based on species rank: "sp," where populations represent recognized species, and "pop," where populations are not recognized as species. Font color differences are only meant to enhance readability of the text. Each point shows the median *gdi* calculated for datasets in each group based on estimated coalescence time (converted to units of $4N$) and migration rate (converted to $4Nm$). Sample sizes for bird, mammal, insect, and plant populations were 14, 18, 8, and 12, respectively; samples sizes for species of these same groups were 8, 9, 13, and 8, respectively.

rate of migration and a recent divergence time, most researchers will consider the taxa to be a single lineage. Conversely, if that IM model is accompanied by a low migration rate and a more ancient divergence time, most would instead infer separate lineages. This follows the argument presented by Hey and Pinho (2012), that divergence time and migration rate should jointly be considered when delimiting species. Furthermore, it is important to estimate parameters in a manner that accounts for the possibility that gene flow and genetic drift will be confounded. One approach toward this is to co-estimate divergence and gene flow parameters using a full likelihood or Bayesian method such as IMa2 (Hey 2010). Another approach, which can easily be carried out as part of a PHRAPL analysis, is to calculate the *gdi* for pairs of taxa, which ignores the relative contributions of genetic drift and gene flow to topological patterns. By combining estimates of coalescence time and migration rate, this index yields a metric of genetic divergence that can help to inform delimitation inference.

As a way to observe the behavior of the *gdi* across parameter space, we plotted *gdi* values for the three empirical datasets analyzed on a contour plot, where *gdi* values were calculated for both nodes in each best supported tree (i.e., t_1 and t_2) (Fig. 6). For comparison, we have also plotted estimated parameter values from the split between *Pan* and *Homo*, as well as median values of the *gdi* based on migration rates and coalescence times for several taxonomic groups culled in a meta-analysis of 178 datasets (Pinho and Hey 2010). Values from recognized species were plotted separately from values estimated from "populations" not currently recognized as species. Although there exists no definitive boundary between *gdi* values for "populations" and "species" (Fig. 6 and Supplementary Table S10), which reflects the continuous nature of the speciation process, some guidance in regards to using the *gdi* to help infer species boundaries can be deduced. First, datasets from all animal populations and species have median *gdi* values of 0.30 and 0.68, respectively. This suggests a median *gdi* threshold $\sim 0.3-0.7$ dividing populations

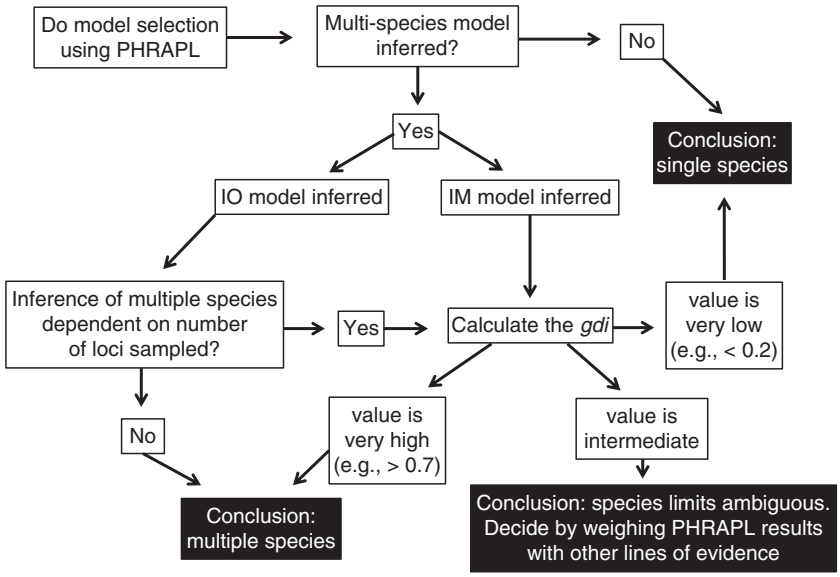

FIGURE 7. An example decision tree for using PHRAPL to delimit species with genetic data.

and species in animals. For plants, the median *gdi* for populations (0.46) is actually a little higher than the median value for species (0.41), offering little evidence for a definitive threshold. When analyzing as many as 50 simulated loci, the signature of a nonzero divergence time disappears when the actual divergence time $t <$ 0.25 (upper left plot in Fig. 3), which corresponds with a *gdi* around 0.4 (Fig. 6). When gene flow is added to the generating model (bottom three plots on the left in Fig. 3), this divergence time threshold at which nonzero coalescence time is detected increases, such that two species are never inferred with high AIC weight when the *gdi* is below ~0.3. Finally, although there is broad *gdi* overlap between populations and species when looking at individual datasets (Supplementary Table S10), the upper quartile range of *gdi* values observed for groups identified as "populations" in the 178 empirical datasets never rises above 0.66, suggesting that a *gdi* value above ~0.7 signals that speciation has likely occurred. Thus, as a rule of thumb, *gdi* values less than 0.2 suggest that a single species exists; *gdi* values above 0.7 suggest there are two species. Values in between indicate ambiguous delimitation (but of course, with values near 0.2 and 0.7 providing stronger or weaker evidence for a single species, respectively), which reflects the reality that there exists a speciation gray zone, where a definitive answer cannot easily be found.

Under what conditions should the *gdi* be used when delimiting species with PHRAPL? (Fig. 7). The initial step should always be to perform a grid search to infer the best demographic/delimitation model(s) as well as parameter values. If an isolation-only or single species model is inferred, delimitation conclusions can typically follow directly from model selection. However, if inference of a multispecies model depends on using a very large number of loci, one can calculate the *gdi* to ensure that its value is consistent with a multispecies model being not just statistically significant, but biologically significant as well. If an IM model is inferred, delimitation conclusions are less straightforward. In these cases, one should calculate the *gdi* among sister groups of interest (which requires both the species tree and parameter estimates from the grid search) to inspect the overall level of divergence produced by isolation and gene flow. If that value is very small or large (e.g., <0.2 or >0.7), this could be interpreted as genetic evidence for no speciation or speciation, respectively. However, if the value is intermediate, the conclusion will necessarily be ambiguous and researchers should consider other sources of data (e.g., ecological, morphological, etc.). For example, in the case of *S. lateralis*, for which an IM model was supported, *gdi* values were relatively low (0.17 for t_1). Thus, although genetic isolation (with migration) was inferred for these lineages, in part due to relatively high rates of migration, the overall level of divergence estimated is more similar to levels observed within populations of a single species than to levels observed among different species (Fig. 6). This, in combination with the fact that no morphological differences have been observed across the range of *S. lateralis* (e.g., Lewis 1951; Johnson 1953; Brooks 1967), suggests that these distinct populations may best be considered a single lineage. In contrast, in the carnivorous plant *S. alata*, where an IM model was also inferred, *gdi* values for $t_2$are somewhat high (0.75), and above values observed in most plant species (Fig. 6), providing genetic evidence for the delimitation of these two groups.

Comparisons with BPP

As observed here and elsewhere (Zhang et al. 2011; Camargo et al. 2012; Yang and Rannala 2015), BPP is a very powerful method for detecting genetic isolation, even when that isolation has occurred recently and in the face of high gene flow. This power likely comes from the full likelihood-based multispecies coalescent model that BPP implements, which allows the program to harness information regarding both species relationships and coalescence times while accounting for uncertainty in estimated gene trees. But detecting genetic isolation is not the same thing as delimiting species. With enough loci, it is possible that even a short history of genetic isolation (e.g., $t = 0.05$) combined with a high rate of genetic exchange (e.g., $M = 10$) will yield an inference of speciation when using BPP (Fig. 3d), yet when migration rates are so high, in what sense has genetic isolation taken place? Once M is greater than ~2, phylogenetic divergence does not appear to form (Fig. 6), nor is it expected to (Wright 1931). Thus, that BPP delimits taxa that share genes in excess of this rate suggests that the method is prone to oversplitting in the face of gene flow, a tendency illustrated by BPP inference from a human dataset. Gene flow estimates among human populations vary, but can be high among non-African populations (e.g., Garrigan et al. 2007; Li et al. 2008), contributing to high amounts of shared variation among human populations (Yu et al. 2002; Wall et al. 2008). Moreover, all the available genetic evidence shows that modern humans are members of a single, relatively young species, for which there has not been nearly enough temporal or geographical separation for species-level differences to form. Nevertheless, BPP wrongly delimits four distinct human lineages, demonstrating that the program is detecting population-level divergence rather than true lineage independence. Given these results from human data, we should treat with caution BPP's delimitation of the two Western pitcher plant populations, as the estimated *gdi* between these populations was extremely low ($gdi = 0.0014$). To reduce the risk of oversplitting using BPP, one might consult posterior distributions of parameters (e.g., θ and τ) or use the "τ threshold" approach (Yang and Rannala 2010), whereby species are only delimited if there is a high posterior probability that a divergence time, τ, is above an assumed threshold value. When Yang and Rannala (2010) used the τ threshold approach to delimit human populations (which, to our knowledge is not the typical BPP algorithm used), they inferred only a single species. When we analyzed our human dataset using the τ threshold method, we inferred a single species as well (see Supplementary Methods S11). While a τ threshold can help to circumvent BPP's tendency to oversplit, assuming a satisfactory threshold can be devised, it is not likely to perform well when gene flow is present in the dataset, as this will likely drive estimates of τ downward. The *gdi* has an advantage over τ in these cases as it incorporates both isolation and gene flow.

In a previous study of the effects of migration on delimitation inference using BPP, Zhang et al. (2011) reported that the method tended to lump species when migration rates were high, in contrast to our study, which reports the opposite. The discrepancy likely results from the fact that the amount of migration simulated by Zhang et al. that resulted in consistent lumping of species by BPP was four times higher than the highest amount of migration we simulated here ($4Nm = 10$ in our article rather than $4Nm = 40$). We did not simulate higher values because once $4Nm > 2$, we found that divergence of populations (as measured from topologies) effectively never formed, regardless of how long ago divergence commenced (Fig. 6). Another discrepancy is that Zhang et al. only simulated up to 10 loci, in contrast to our 50. Given that the likelihood that BPP will delimit species increases with the number of loci (Fig. 3), had Zhang et al. simulated more loci, they may have inferred distinct species even in the case where migration was set to $4Nm = 40$. Given that our results are generally consistent with those of Zhang et al. (2011) in the areas over which our parameter spaces overlap, we assume that we also would have observed lumping of species by BPP with 10 loci and $4Nm = 40$. The important point is that a delimitation method should be lumping species not only at $4Nm = 40$, but at lower migration rates as well (e.g., at $4Nm = 10$, a value which was not simulated by Zhang et al. 2011), and regardless of the number of loci used.

PHRAPL is clearly not as powerful as BPP in detecting genetic isolation. For example, when PHRAPL applied the same models considered by BPP to simulated data, the method only supported lineages A and B as two species once underlying divergence reached a certain depth ($\sim t \geqslant 0.25$) or once migration rate fell below a certain value ($M = 0.5$) (Fig. 3a). Although the specific thresholds will change to some extent with the parameter grid used to analyze a dataset, this inability to infer distinct species given extremely recent isolation or high migration is a desirable property for any method that delimits species (Carstens et al. 2013), and it suggests that PHRAPL errs on the side of failing to delimit actual species rather than falsely splitting a single species into multiple lineages. This conservative inference may be appropriate given the legal ramifications of the species category (e.g., Fujita et al. 2012) and its use in a wide range of biological disciplines. However, it also means that truly separate species will sometimes be missed, which can also have negative conservation implications. As with all species delimitation methods, we recommend using PHRAPL in concert with natural history information and other available approaches when doing alpha taxonomy (e.g., Bacon et al. 2012; Hendrixson et al. 2015).

Conclusion

When the goal is to delimit species, one usually seeks a simple binary answer to the question, "one species or

two?" However, this result may not always be achievable. First, this is counter to the spirit of multimodal inference (e.g., Burnham and Anderson 2002), which underlies the model selection framework in PHRAPL (Jackson et al. in press). Multimodel inference seeks to quantify the support for the models given the data. In some cases, the optimal model will receive the vast majority of the support, and in these cases the inferences regarding species limits may be straightforward. However, in cases where support across models is equivocal, researchers should clearly assess their sampling (of individuals and/or loci) and ask whether it is adequate for the question at hand. Second, an understanding of the evolutionary processes giving rise to divergence is critical when delimiting species (Hey and Pinho 2012), and thus decisions about species boundaries resulting from model selection should not be made without also considering the relative and absolute importance of isolation and migration within an inferred model.

One limitation of the method is that PHRAPL assumes that both the gene trees and population assignments are specified without error (i.e., it is a validation approach, *sensu* Ence and Carstens 2011). The degree to which PHRAPL delimitation and model selection are robust to data error and ambiguity is currently unknown and represents an important area of further research (although note that the analyses here used gene trees inferred from simulated sequence data, not true gene trees). One should thus ensure that accurate gene trees and population assignments are used.

Although in this study we have focused on populations shaped by isolation and symmetrical migration, additional parameters will likely be necessary to adequately model the history of many species. PHRAPL models can be constructed that include additional complexities such as different population sizes, population growth or contraction, and asymmetrical migration. The performance of these models will be a worthwhile subject of future research.

PHRAPL is a valuable new approach for investigating species boundaries in that it allows one to compare a large number of demographic models while jointly considering both gene flow and population divergence. This broadens the relevant biological complexity that can be considered, and will particularly be a useful delimitation method for biological systems in which divergence with gene flow has likely been important.

Supplementary Materials

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.t8016.

Funding

This work was supported by the National Science Foundation (grant numbers DEB 1257669, DEB 1257784).

Acknowledgements

We thank members of the Carstens and O'Meara labs for discussions related to PHRAPL and species delimitation. We thank participants at the PHRAPL workshops held in May of 2014 at OSU and June of 2015 at the Evolution Annual Meeting for their feedback. We also thank the Society of Systematic Biologists and the National Science Foundation (grant number DEB 1500774) for funding the 2015 workshop.

References

Bacon C.D., McKenna M.J., Simmons M.P., Wagner W.L. 2012. Evaluating multiple criteria for species delimitation: an empirical example using Hawaiian palms (Arecaceae: *Pritchardia*). BMC Evol. Biol. 12–23.

Brooks G.R. 1967. Population ecology of the ground skink, *Lygosoma laterale* (Say). Ecol. Monogr. 37:71–87.

Burbrink F.T., Guiher T.J. 2015. Considering gene flow when using coalescent methods to delimit lineages of North American pitvipers of the genus *Agkistrodon*. Zool. J. Linn. Soc. 173:505–526.

Burnham K.P., Anderson D.R. 2002. Model selection and multimodel inference: a practical infromation-theoric approach. New York: Springer-Verlag.

Camargo A., Morando M., Avila L.J., Sites J.W., Jr. 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 B.C., Dewey T.A. 2010. Species delimitation using a combined coalescent and information-theoretic approach: an example from North American *Myotis* Bats. Syst. Biol. 59:400–414.

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

Carstens B.C., Satler J.D. 2013. The carnivorous plant described as *Sarracenia alata* contains two cryptic species. Biol. J. Linn. Soc. 109:737–746.

Coyne J.A., Orr H.A. 2004. Speciation. Sunderland (MA): Sinauer Associates.

Cummings M.P., Neel M.C., Shaw K.L. 2008. A genealogical approach to quantifying lineage divergence. Evolution 62:2411–2422.

de Queiroz K. 2007. Species concepts and species delimitation. Syst. Biol. 56:879–886.

Eckert A.J., Carstens B.C. 2008. Does gene flow destroy phylogenetic signal? The performance of three methods for estimating species phylogenies in the presence of gene flow. Mol. Phylogenet. Evol. 49:832–842.

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

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

Fagundes N.J.R., Ray N., Beaumont M., Neuenschwander S., Salzano F.M., Bonatto S.L., Excoffier L. 2007. Statistical evaluation of alternative models of human evolution. Proc. Natl Acad. Sci. USA 104:17614–17619.

Fujita M.K., Leache A.D., Burbrink F.T., McGuire J.A., Moritz C. 2012. Coalescent-based species delimitation in an integrative taxonomy. Trends Ecol. Evol. 27:480–488.

Funk D.J., Omland K.E. 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Ann. Rev. Ecol. Evol. Syst. 34:397–423.

Garrigan D., Kingan S.B., Pilkington M.M., Wilder J.A., Cox M.P., Soodyall H., Strassmann B., Destro-Bisol G., de Knijff P., Novelletto A., Friedlaender J., Hammer M.F. 2007. Inferring human population sizes, divergence times and rates of gene flow from mitochondrial, X and Y chromosome resequencing data. Genetics 177:2195–2207.

Grummer J.A., Bryson R.W., Jr., Reeder T.W. 2014. Species delimitation using Bayes factors: simulations and application to the *Sceloporus*

scalaris species group (Squamata: Phrynosomatidae). Syst. Biol. 63:119–133.
Heled J., Bryant D., Drummond A.J. 2013. Simulating gene trees under the multispecies coalescent and time-dependent migration. BMC Evol. Biol. 13:44.
Hendrixson B.E., Guice A.V., Bond J.E. 2015. Integrative species delimitation and conservation of tarantulas (Araneae, Mygalomorphae, Theraphosidae) from a North American biodiversity hotspot. Insect Conserv. Divers. 8:120–131.
Hey J. 2006. Recent advances in assessing gene flow between diverging populations and species. Curr. Opin. Genet. Dev. 16:592–596.
Hey J. 2010. Isolation with migration models for more than two populations. Mol. Biol. Evol. 27:905–920.
Hey J., Pinho C. 2012. Population genetics and objectivity in species diagnosis. Evolution 66:1413–1429.
Hudson R.R. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics 18:337–338.
Ingman M., Kaessmann H., Paabo S., Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. Nature 408:708–713.
Jackson N.D., Austin C.C. 2010. The combined effects of rivers and refugia generate extreme cryptic fragmentation within the common ground skink (*Scincella lateralis*). Evolution 64:409–428.
Jackson N.D., O'Meara B.C., Morales A.E., Carstens B.C. In press. PHRAPL: Phylogeographic Inference using approximated likelihoods. Syst. Biol.
Jackson N.D., Austin C.C. 2012. Inferring the evolutionary history of divergence despite gene flow in a lizard species, *Scincella lateralis* (Scincidae), composed of cryptic lineages. Biol. J. Linn. Soc. 107: 192–209.
Johnson R.M. 1953. A contribution on the life history of the lizard *Scincella laterale*(Say). Tulane Stud. Zool. 1:10–27.
Jakobsson M., Scholz S.W., Scheet P., Gibbs J.R., VanLiere J.M., Fung H.C., Szpiech Z.A., Degnan J.H., Wang K., Guerreiro R., Bras J.M., Schymick J.C., Hernandez D.G., Traynor B.J., Simon-Sanchez J., Matarin M., Britton A., van de Leemput J., Rafferty I., Bucan M., Cann H.M., Hardy J.A., Rosenberg N.A., Singleton A.B. 2008. Genotype, haplotype and copy-number variation in worldwide human populations. Nature. 451:998–1003.
Jones G., Aydin Z., Oxelman B. 2015. DISSECT: an assignment-free Bayesian discovery method for species delimitation under the multispecies coalescent. Bioinformatics 31:991–998.
Kingman J.F.C. 1982. On the genealogy of large populations. J. Appl. Probab. 19:27–43.
Knowles L.L., Carstens B.C. 2007a. Delimiting species without monophyletic gene trees. Syst. Biol. 56:887–895.
Knowles L.L., Carstens B.C. 2007b. Estimating a geographically explicit model of population divergence. Evolution 61:477–493.
Leaché A.D., Harris R.B., Rannala B., Yang Z. 2014. The influence of gene flow on species tree estimation: a simulation study. Syst. Biol. 63:17–30.
Lewis T.H. 1951. The biology of *Leiolopisma laterale* (Say). Am. Midl. Nat. 45:232–240.
Li J.Z., Absher D.M., Tang H., Southwick A.M., Casto A.M., Ramachandran S., Cann H.M., Barsh G.S., Feldman M., Cavalli-Sforza L.L., Myers R.M. 2008. Worldwide human relationships inferred from genome-wide patterns of variation. Science 319: 1100–1104.
Maddison W.P. 1997. Gene trees in species trees. Syst. Biol. 46:523–536.
Noonan B.P., Gaucher P. 2006. Refugial isolation and secondary contact in the dyeing poison frog, *Dendrobates tinctorius*. Mol. Ecol. 15: 4425–4435.
Nosil P. 2008. Speciation with gene flow could be common. Mol. Ecol. 17:2103–2106.
O'Meara B.C. 2010. New heuristic methods for joint species delimitation and species tree inference. Syst. Biol. 59:59–73.
Pamilo P., Nei M. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol., 5:568–583.
Pfeifer B., Wittelsbuerger U., Ramos-Onsins S.E., Lercher M.J. 2014. PopGenome: an efficient Swiss Army Knife for population genomic analyses. R. Mol. Biol. Evol. 31:1929–1936.
Pinho C., Hey J. 2010. Divergence with gene flow: models and data. Ann. Rev. Ecol. Evol. Syst. 41:215–230.
Rambaut A., Grassly N.C. 1997. An application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees, version 1.2.5. Comput. Appl. Biosci. 13:235–238.
Rannala B., Yang Z.H. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics 164:1645–1656.
Rittmeyer E.N., Austin C.C. 2012. The effects of sampling on delimiting species from multi-locus sequence data. Mol. Phylogenet. Evol. 65:451–463.
Roach J.C., Glusman G., Smit A.F.A., Huff C.D., Hubley R., Shannon P.T., Rowen L., Pant K.P., Goodman N., Bamshad M., Shendure J., Drmanac R., Jorde L.B., Hood L., Galas D.J. 2010. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science 328:636–639.
Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.
Stephens M., Smith N.J., Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68:978–989.
Tishkoff S.A., Kidd K.K. 2004. Implications of biogeography of human populations for 'race' and medicine. Nature Genetics. 36: S21–S27.
Wall J.D., Cox M.P., Mendez F.L., Woerner A., Severson T., Hammer M.F. 2008. A novel DNA sequence database for analyzing human demographic history. Genome Res. 18:1354–1361.
Winter D.J., Trewick S.A., Waters J.M., Spencer H.G. 2016. The genealogical sorting index and species delimitation. bioRxiv. doi: http://dx.doi.org/10.1101/036525.
Wright S. 1931. Evolution in Mendelian populations. Genetics 16: 97–159.
Yang Z., Rannala B. 2015. Unguided species delimitation using DNA sequence data from multiple loci. Mol. Biol. Evol. 31:3125–3135.
Yang Z.H., Rannala B. 2010. Bayesian species delimitation using multilocus sequence data. Proc. Natl. Acad. Sci. USA 107:9264–9269.
Yu N., Chen F.C., Ota S., Jorde L.B., Pamilo P., Patthy L., Ramsay M., Jenkins T., Shyue S.K., Li W.H. 2002. Larger genetic differences within Africans than between Africans and Eurasians. Genetics 161:269–274.
Zamudio K.R., Savage W.K. 2003. Historical isolation, range expansion, and secondary contact of two highly divergent mitochondrial lineages in spotted salamanders (*Ambystoma maculatum*). Evolution 57:1631–1652.
Zellmer A.J., Hanes M.M., Hird S.M., Carstens B.C. 2012. Deep phylogeographic structure and environmental differentiation in the carnivorous plant *Sarracenia alata*. Syst. Biol. 61:763–777.
Zhang C., Rannala B., Yang Z. 2014. Bayesian species delimitation can be robust to guide-tree inference errors. Syst. Biol. 63: 993–1004.
Zhang C., Zhang D.-X., Zhu T., Yang Z. 2011. Evaluation of a Bayesian coalescent method of species delimitation. Syst. Biol. 60:747–761.