

# The influence of locus number and information content on species delimitation: an empirical test case in an endangered Mexican salamander

PAUL M. HIME,<sup>\*2</sup> SCOTT HOTALING,<sup>\*2</sup> RICHARD E. GREWELLE,<sup>\*</sup> ERIC M. O'NEILL,<sup>\*1</sup> S. RANDAL VOSS,<sup>\*</sup> H. BRADLEY SHAFFER<sup>†</sup> and DAVID W. WEISROCK<sup>\*</sup>

<sup>\*</sup>Department of Biology, University of Kentucky, Lexington, KY 40506, USA, <sup>†</sup>Department of Ecology and Evolutionary Biology and the La Kretz Center for California Conservation Science, University of California Los Angeles, Los Angeles, CA 90095, USA

## Abstract

Perhaps the most important recent advance in species delimitation has been the development of model-based approaches to objectively diagnose species diversity from genetic data. Additionally, the growing accessibility of next-generation sequence data sets provides powerful insights into genome-wide patterns of divergence during speciation. However, applying complex models to large data sets is time-consuming and computationally costly, requiring careful consideration of the influence of both individual and population sampling, as well as the number and informativeness of loci on species delimitation conclusions. Here, we investigated how locus number and information content affect species delimitation results for an endangered Mexican salamander species, *Ambystoma ordinarium*. We compared results for an eight-locus, 137-individual data set and an 89-locus, seven-individual data set. For both data sets, we used species discovery methods to define delimitation models and species validation methods to rigorously test these hypotheses. We also used integrated demographic model selection tools to choose among delimitation models, while accounting for gene flow. Our results indicate that while cryptic lineages may be delimited with relatively few loci, sampling larger numbers of loci may be required to ensure that enough informative loci are available to accurately identify and validate shallow-scale divergences. These analyses highlight the importance of striking a balance between dense sampling of loci and individuals, particularly in shallowly diverged lineages. They also suggest the presence of a currently unrecognized, endangered species in the western part of *A. ordinarium*'s range.

**Keywords:** *Ambystoma*, approximate likelihood, population structure, singular value decomposition, speciation, Trans-Mexican Volcanic Belt

Received 21 March 2016; revision received 19 September 2016; accepted 26 September 2016

## Introduction

An important recent advance in molecular systematics has been the development of refined evolutionary models and new analytical approaches for delimiting species

using multilocus DNA sequence data (Fujita *et al.* 2012). One of the most important decisions required by these new species delimitation methods involves the trade-off between the numbers of loci and individuals that should be sampled to accurately identify cryptic species lineages. Increased sampling of loci provides more precise and accurate inferences of population-level parameters (Harris *et al.* 2014), while increased sampling of individuals more completely captures variation within and between populations. Historically, phylogeographic investigations have relied on sampling genetic data

Correspondence: Paul M. Hime, Fax: 859-257-1717;

E-mail: paul.hime@uky.edu

<sup>1</sup>Present Address: Department of Biology, Geology, and Environmental Science, University of Tennessee at Chattanooga, Chattanooga, TN 37403, USA

<sup>2</sup>Authors contributed equally.

from relatively small numbers of loci, often from many individuals (e.g. hundreds). However, population genetic theory demonstrates that variation among gene histories can be enormous, often dwarfing variation among individuals, suggesting that increased gene sampling is key to accurate inferences (Irwin 2002). Now, advances in next-generation sequencing (NGS) are providing the opportunity to sample large numbers of loci (e.g. hundreds or thousands), although pragmatic (cost) considerations may limit the number of individuals that can be sampled. In practice, this results in a spectrum of potential data sets, ranging from small numbers of loci sampled from many individuals (e.g. Hotaling *et al.* 2016) to NGS-scale data sets sampled from fewer specimens (e.g. Smith *et al.* 2014; Rittmeyer & Austin 2015). To date, few studies have examined the influence of gene and individual sampling density on the detection of lineage divergence events. Similarly, an exploration of the information content of loci vs. the number of loci needed to detect lineage divergence events would be informative in guiding future delimitation studies that use NGS-scale data.

Equally important to the accurate delimitation of species from multilocus data sets is the use of analyses that account for the population-level forces structuring genetic variation within and between populations. Methods testing species divergence hypotheses using coalescent models have recently grown to include information-theoretic (Ence & Carstens 2011), Bayesian (Yang & Rannala 2010; Grummer *et al.* 2014) and approximate Bayesian (Camargo *et al.* 2012) frameworks. While these methods have the ability to explicitly incorporate population size and divergence time parameters into species delimitation tests, they do not allow for gene flow and its potential influence in structuring genetic variation. Alternatively, a number of methods explicitly estimate gene flow as a model parameter, permitting researchers to gauge the extent to which hypothesized lineages exchange genes. However, current methods either assess gene flow independently of divergence (e.g. MIGRATE-N; Beerli & Felsenstein 2001) or as part of a model that includes a divergence time parameter but does not assess alternative models lacking a history of divergence, which may offer closer fits to the data (e.g. IMA2; Hey 2010). Furthermore, extreme values of population genetic parameters (divergence, population size and migration rates) can all exert an influence of expected patterns of gene coalescence and affect species delimitation results (Rannala 2015). For example, small effective population sizes could make populations appear more diverged than they are. Consequently, species delimitation studies that aim to incorporate all of these parameters into decisions about lineage divergence must use multiple methods to

estimate different, and partially overlapping, sets of population genetic parameters relevant to the process of speciation (e.g. Jonsson *et al.* 2014; Reid *et al.* 2014).

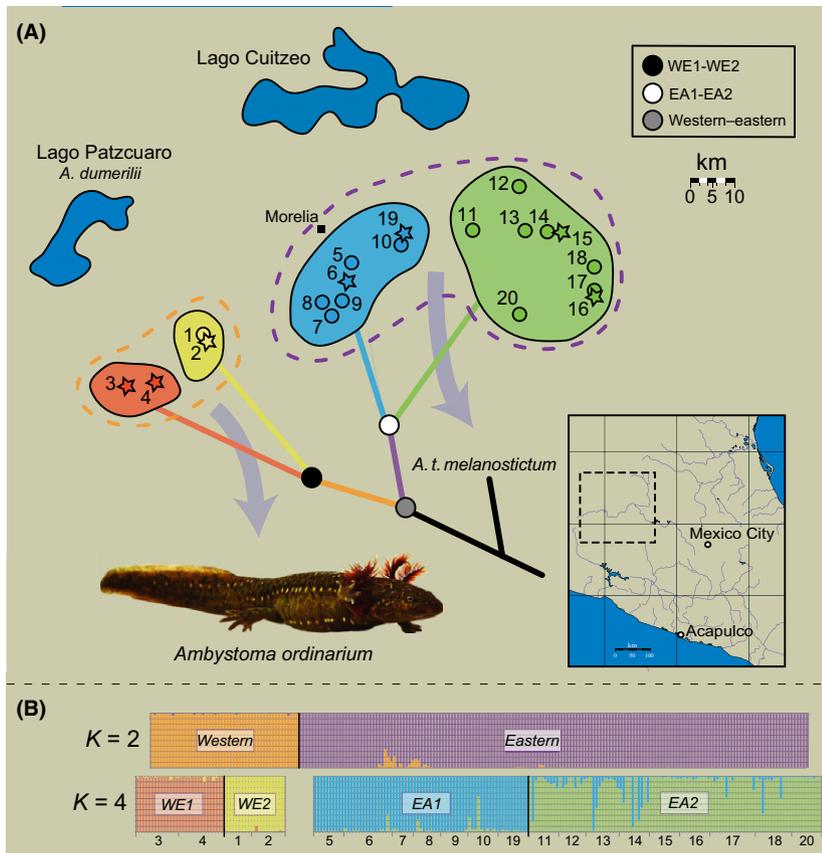
Here, we use an eight nuclear locus data set (referred to as the '8L' data set) sampled from a large number of individuals ( $n = 137$ ) and an 89 nuclear locus data set (referred to as the '89 L' data set) sampled from a small number of individuals ( $n = 7$ ) to empirically quantify the effects of gene and sampling density on species delimitation outcomes. Both data sets were generated from populations across the range of a narrowly distributed stream-dwelling salamander species, *Ambystoma ordinarium*. A member of the tiger salamander complex (Shaffer & McKnight 1996), *A. ordinarium*, has previously been diagnosed as a genealogically exclusive species based on patterns of monophyly in reconstructed gene trees from the 8L data set (Weisrock *et al.* 2006). The species has a small,  $\sim 120 \times 20$  km range across high elevations ( $>2200$  m) of the Mesa Central in the western Trans-Mexican Volcanic Belt (TMVB), a region characterized by rugged terrain created through a history of volcanism and tectonic uplift (Fig. 1). Furthermore, its range includes streams that flow into multiple independent drainage systems (Anderson & Worthington 1971), providing an opportunity for hydrologically mediated lineage divergence in allopatry and potential speciation, despite a small geographic scale.

Using these two different data sets, we explore patterns of lineage divergence within *A. ordinarium*, and in the process, address the impact of locus sampling and individual sampling on the delimitation of cryptically diverging lineages. We specifically investigate the impact of both the number and information content of loci on our results using random subsamples of our data and using increasing numbers of loci ordered by their phylogenetic information content. As part of this process, we also implement a novel approach towards species delimitation using a recently developed species tree reconstruction method, SVDQUARTETS (Chifman & Kubatko 2014). This method is restricted to a coalescent model that does not currently parameterize gene flow, but which scales well to multilocus sequence data. Drawing on this species delimitation work, we then extend our species delimitation tests through model selection across models that simultaneously consider divergence and gene flow.

## Materials and methods

### 8L sequence data

The 8L data used here are those published in Weisrock *et al.* (2006), with the exception that the mtDNA data



**Fig. 1** Geographic sampling of *Ambystoma ordinarius* populations used in this study and patterns of population genetic structure. (A) Localities 1–20 represent sampling for the 8L data set. Stars denote sampling localities used in the 89L data set. Locality numbers match those of Weisrock *et al.* (2006). Dashed outlines enclose population genetic clusters identified from STRUCTURE analyses at  $K = 2$  (western and eastern), while solid coloured areas represent hierarchical population genetic structure identified within the western and eastern clusters. Nodes on the hypothesized species tree topology represent the putative divergence events tested in this study. Arrows denote distinct southward-flowing drainages of streams in the western and eastern portions of the range. (B)  $K = 2$  is best supported in analyses of the full 8L data set, but when western and eastern populations were analysed separately, each had a  $\Delta K$  favouring a  $K = 2$ . Photograph of *A. ordinarius* courtesy of Sebastian Voitel. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)].

were excluded from all analyses given the clear signatures of mitochondrial DNA (mtDNA) introgression involving *Ambystoma ordinarius* and other *Ambystoma tigrinum* complex species (Weisrock *et al.* 2006). Analyses were focused on phased Sanger DNA sequence data from eight nuclear loci (*col1a1*, *dlx3*, *ctg1506*, *ctg1908*, *g1d6*, *g1f1*, *g1c12* and *g3d7*) generated from 217 paedomorphic or young larval specimens of *A. ordinarius* sampled from 20 geographic localities distributed across the known species range (Fig. 1; Table S1, Supporting information). We used this 217-individual data set for all population structure analyses. For all subsequent species delimitation tests, we reduced the total 8L sequence data set down to 137 individuals that contained no missing data for all eight loci. Sequence data from a single *Ambystoma tigrinum melanostictum* individual was used as an out-group.

Measures of genetic variation within *A. ordinarius*, including numbers of haplotypes, substitution and indel variation, and estimates of heterozygosity were previously reported in Weisrock *et al.* (2006), and are included in Table S2 (Supporting information). Nuclear DNA sequence data for all 217 *A. ordinarius* individuals are available in GenBank (*col1a1*: DQ252580–252937; *ctg1506*, DQ252938–253365; *ctg1908*, DQ254388–254797;

*dlx3*, DQ248436–248859; *g1c12*, DQ254798–255197; *g1d6*; DQ255356–255783; *g1f1*, DQ253450–253837; *g3d7*, DQ253924–254301).

#### 89L data

The *A. ordinarius* NGS-scale data used here are a subset of the loci published in O'Neill *et al.* (2013) as part of a larger study of the *A. tigrinum* complex. These data were generated via parallel tagged amplicon sequencing (PTAS) on a Roche 454 sequencing platform. Full details of marker selection, field sampling and sequence generation can be found in O'Neill *et al.* (2013). For this study, phased sequence data from seven *A. ordinarius* individuals and one *A. tigrinum melanostictum* individual were extracted from this larger data set. The *A. ordinarius* individuals represent samples from seven different localities broadly covering the geographic range. Of the 95 total loci in the original study, 81 contained one or more variable sites and had data present for at least five *A. ordinarius* individuals and the *A. t. melanostictum* out-group. We combined these 81 loci with the 8L loci (using the same set of individuals present in the PTAS data) to produce the 89L data set used here (details of the 81 additional loci are provided

in Table S3, Supporting information). Both the 8L and 89L data sets analysed are included with this study's Dryad accession (doi:10.5061/dryad.td997).

#### *Species hypothesis generation – population structure*

Genetically based species delimitation is typically a multipart process (Carstens *et al.* 2013), beginning with a discovery phase aimed at identifying hypothetical lineages and assigning sampled individuals to these lineages. Inaccuracy in this 'discovery phase' can lead to biased downstream results (Edwards & Knowles 2014; Olave *et al.* 2014; but see Zhang *et al.* 2014). Therefore, we utilized several discovery approaches to identify consensus patterns of population structure. For the full 217-individual 8L data set, we used the program STRUCTURE v2.3.1 (Pritchard *et al.* 2000) to estimate the relative assignments of individuals into  $K$  populations (allowing for admixture) for all  $K$  from 1 to 20. We used plots of the log probability of the data [ $\ln \Pr(X|K)$ ] and  $\Delta K$  (Evanno *et al.* 2005) to determine the optimal number of population clusters and to detect hierarchical structure, respectively. We also used the program STRUCTURAMA2 (Huelsenbeck *et al.* 2011), treating  $K$  as a random variable, although results were largely concordant with those from STRUCTURE. Finally, we also used the program SPLITSTREE v4.13.1 (Huson & Bryant 2006) to assess the degree to which haplotypes were shared among vs. within putative clusters. Additional details of STRUCTURE and full details for the STRUCTURAMA and SPLITSTREE analyses are provided in Supporting Information (Figs S2 and S4, Supporting information, respectively).

#### *Delimitation hypothesis testing – SPEDESTEM*

Using lineage hypotheses for *A. ordinarium* from the discovery phase, we tested a two-lineage hypothesis, which splits *A. ordinarium* into western and eastern lineages, and a four-lineage hypothesis, which further subdivides the western clade into two lineages (WE1 and WE2) and the eastern clade into two lineages (EA1 and EA2). The two-lineage model roughly corresponds to separate stream drainages on either side of a mountain ridge, and although the four-lineage model does not correspond to any known geographic barriers, its utility is supported by results of the discovery phase. We tested these hypotheses using an information-theoretic model selection approach implemented in SPEDESTEM v0.9.5 (Ence & Carstens 2011). SPEDESTEM uses point estimates of gene trees, but considers all or many of the possible lumpings and splittings of hypothesized lineages.

Attempts to perform SPEDESTEM analyses on the 89L data presented computational challenges; therefore, we

focused our analyses on the 8L data. For both lineage hypotheses, we performed tip subsampling using two, three, five and 10 alleles from each lineage in reconstructed gene trees, with 1000 replicates each. For the two-lineage hypothesis, it was also computationally feasible to sample 25 tips from each lineage for 100 replicates. Likelihood scores were averaged across replicates, and Akaike information criterion (AIC) scores and relative model probabilities were calculated for each species tree model.

#### *Delimitation hypothesis testing – BPP*

We also used the program BPP3 (Yang & Rannala 2010) to test our two- and four-lineage species delimitation hypotheses. Although this popular method no longer requires an a priori guide tree, we chose to utilize the fixed tree topology (Fig. 1) which was strongly supported by our discovery approaches. BPP uses reversible-jump Markov chain Monte Carlo sampling to compare nested species models by collapsing or failing to collapse nodes in the user-specified guide tree. In doing so, BPP calculates posterior probabilities (PPs) of those nodes in the guide tree and the relative model probability of competing delimitation models.

We used BPP v3.0 (Yang & Rannala 2014) and BPP v3.1, respectively, to analyse the 8L and 89L data sets using nine combinations of priors for  $\theta$  and  $\tau$ , corresponding to small (Gamma distribution set with  $\alpha = 2$ ,  $\beta = 1000$ ), medium (2, 100) or large (2, 10) population sizes, and shallow (2, 1000), intermediate (2, 100) or deep (2, 10) divergence times. In addition, we performed a set of analyses using priors that reflected empirical estimations of  $\theta$  (3, 1250) and  $\tau$  (25, 1149), which were most similar to the small population size and shallow divergence time prior (see Supporting Information for additional details of BPP analyses).

#### *Delimitation hypothesis testing – SVDQUARTETS*

We applied a recently developed species tree reconstruction method using the program SVDQUARTETS (Chifman & Kubatko 2014) implemented in PAUP v4.0a146 (Swofford 2015). For these analyses, we analysed our 89L data as a 43-locus subset (referred to as the '43L' data set), which contained complete data sampling across all seven *A. ordinarium* individuals and the *A. t. melanostictum* out-group. We analysed our 43L data two ways, each taking a different approach for assessing support for our species delimitation hypotheses. First, we estimated a 'lineage tree' following Chifman & Kubatko (2014), where tips in the tree represent the random pairing of gene copies across loci for a diploid individual. Here, SVDQUARTETS can provide

support for species divergence without a priori identification of species hypotheses; branches separating populations that are part of the same species are not expected to be reconstructed with high branch support in the lineage tree. This approach may be viewed as a species discovery approach. Second, we performed analyses using a five-tip species tree model that corresponded to putative species (WE1, WE2, EA1 and EA2) within *A. ordinarium*, and an *A. t. melanostictum* out-group. We treated these analyses as a species validation test, where placement of tips into their expected clades with high bootstrap support is interpreted as evidence for species-level entities. For our 8L data, we performed SVDQUARTETS analyses solely within a species tree framework, as analysis of a lineage tree using 137 individuals proved computationally intractable. For all SVDQUARTETS analyses, we performed exhaustive sampling of all possible quartets (every combination of four tips was examined). Branch support for the inferred trees was estimated using 100 nonparametric bootstrap replicates.

#### *Investigating the role of data scale and content on species delimitation*

Using our 89L data, we explored the degree to which the amount of sequence data influenced the results of coalescent-based species delimitation. We focused these analyses on BPP and also explored the effects of the number of sampled loci and their phylogenetic information content on delimitation inferences. To address the influence of the number of loci, we generated nine data sets varying in the number of sampled loci by increments of 10 (10, 20, 30, etc.) up to 89 loci. For each subsampling increment, we generated 10 replicate jackknifed data sets, each using random locus sampling (89L data set analysed only once). To examine the influence of information content, we ranked the 89 loci in order of their parsimony-informative sites (based on *A. ordinarium* and the *A. t. melanostictum* out-group). We again generated a series of nine data set sizes (10 through 89) that increased (starting with the most informative loci) in the number of sampled loci by increments of 10. Due to the ordered nature of these data sets, only a single round of analysis was performed for each data set.

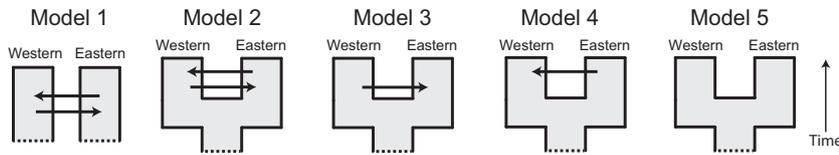
We applied these two data assembly strategies to three different sets of species hypotheses within *A. ordinarium*: (1) testing the split between western and eastern lineages, (2) testing the split between two western lineages (WE1 and WE2) and (3) testing the split between two eastern lineages (EA1 and EA2). All analyses were performed using empirically estimated priors and the same run conditions mentioned above (fully described in Supporting Information).

#### *Phylogeographic model selection and parameter estimation – PHRAPL*

We extended our species delimitation tests to include model selection-based phylogeographic inference implemented in PHRAPL v0.53 (O'Meara *et al.* 2015). PHRAPL employs a heuristic exploration of model space to define a set of the most plausible models given a set of gene trees estimated from multilocus sequence data. Using empirically estimated gene trees and a maximum number of free model parameters, PHRAPL uses approximate likelihoods to infer the model (or models) best supported by the data. This method uses simulated gene trees [generated in the program ms (Hudson 2002)] and compares the simulated gene tree topologies to those from empirical data. Over many replicates, the number of exact matches can be used to calculate likelihoods across a wide range of predefined models. AIC is then used to rank models and to identify the best-fitting model. This method has a natural fit to the process of species delimitation as it allows for the assessment of divergence models that also include migration parameters, providing a more complete assessment of the divergence history of a set of hypothesized species.

For both the 8L and 43L data sets, we analysed five possible models for a two-lineage scenario with one or two free parameters (Fig. 2). Model parameters included divergence time ( $\tau$ ), and rates and direction of migration ( $m$ ). Bidirectional  $m$  were constrained to be symmetric. For each of the 8L loci, we partitioned individuals into western and eastern lineages and randomly sampled six gene copies per lineage across 50 replicates, increasing the probability of sampling each gene copy at least once. We defined all five models corresponding to a two-lineage, two-parameter scenario with one free parameter for  $m$  and  $\tau$ , respectively. Following O'Meara *et al.* (2015), we conducted grid searches across  $\tau$  and/or  $m$  reflecting arbitrary (but realistic) values ( $\tau = 0.3, 0.58, 1.11, 2.12, 4.07, 7.81, 15.0$ ;  $m = 0.10, 0.22, 0.46, 1.00, 2.15, 4.64, 10.0$ ).

For each combination of parameters for each model, we simulated 100 000 balanced 12-tip gene trees in ms and compared the topologies of the observed (subsampling) empirical gene trees to the simulated gene trees. We sought exact topological matches with the caveat that the labelling of individuals drawn from the same population was arbitrary. We defined the approximate likelihood of a given model with a given set of parameter values to be equal to the number of matches between the empirical and simulated trees divided by the number of replicates. Log likelihoods of models were summed across loci, and an AIC score was defined as  $-2 \times \ln(L(\text{model}_i | \text{Data})) + 2K$ , where  $K$  is the number of free parameters in a given model. We computed model likelihoods for each model, and final



**Fig. 2** Demographic and phylogeographic models for the western and eastern *A. ordinarius* lineages tested in PHRAPL analyses. Horizontal arrows indicate gene flow between lineages.

model selection was performed by ranking models by increasing AIC and observing the plot of  $\Delta$ AIC across models. PHRAPL analysis of the 43L data followed the same approach described above, except that five gene copies from both the western and eastern lineage were sampled, and fewer replicates (five) were needed to ensure that all gene copies were sampled from all gene trees at least once.

For all PHRAPL analyses, we also explored models including five or fewer free parameters which allowed for changes in effective population size and asymmetric migration. However, PHRAPL was unable to discriminate among this larger set of models as evidenced by low  $\Delta$ AIC ( $<0.25$ ) values, likely due to an insufficient number of loci and/or variable sites per locus, and these more complex models were not considered further. To generate model averaged estimates of parameters for each data set, we calculated the likelihood-weighted arithmetic mean of each parameter across all models using the CalculateModelAverages function in PHRAPL. Additional details of PHRAPL analyses are provided in Supporting Information.

#### Demographic model selection and parameter estimation – MIGRATE-N

We used MIGRATE-N v3.6 (Beerli 2006) to estimate gene flow under a coalescent framework for a range of two-population models using the 8L data as limited individual sampling in the 89L data set precluded its use. We tested: (i) a ‘panmixia’ model treating eastern and western lineages as a single population, (ii) a two-population model with bidirectional gene flow, (iii) a two-population model with unidirectional gene flow from the western lineage into the eastern lineage and (iv) a two-population model with no migration (Fig. S1, Supporting information). Initial parameter values were calculated using  $F_{ST}$ , and we employed model averaging to estimate migration rate ( $m$ ) and  $\theta$ . Full details of MIGRATE-N analyses are provided in Supporting Information.

## Results

### Data summary

The 8L data set contained a total of 4176 nucleotide sites. Including the *Ambystoma tigrinum melanostictum* out-group, the number of parsimony-informative sites

(PIS) across loci ranged from 4 to 25, with a mean of 14.9. Within *Ambystoma ordinarius* the number of PIS across loci ranged from 3 to 25, with a mean of 13.9. The 81 PTAS loci from O’Neill *et al.* (2013) contained a total of 20 006 nucleotide sites. Including the *A. t. melanostictum* out-group, PIS across loci ranged from 0 to 12, with a mean of 4.06. Within *A. ordinarius*, PIS across loci ranged from 0 to 7 with a mean of 1.07 (Tables S2 and S3, Supporting information). Maximum-likelihood gene trees for the 8L and 89L loci are included in Fig. S2 (Supporting information) and in this study’s Dryad accession, respectively.

### Population structure and hypothesis generation

Analysis of the 8L data using STRUCTURE resulted in a  $\Delta K$  that supported a  $K = 2$  model separating western and eastern populations of *A. ordinarius* with low levels of admixture (Fig. 1). Separate analyses on each of these groups identified a  $K = 2$  level of population structure within each (hereafter referred to as WE1 and WE2 across western populations and EA1 and EA2 across eastern populations). In some cases, further STRUCTURE analysis within these groups suggested additional population structure, but with high degrees of admixture, and these clusters were not explored further. Population structure results were not method-dependent, and additional population clustering results for STRUCTURE, STRUCTURAMA (Fig. S3, Supporting information) and SPLITSTREE (Fig. S4, Supporting information) are provided in Supporting Information.

### Delimitation hypothesis testing – SPEDESTEM

For the two-lineage model, SPEDESTEM analysis of the 8L data only supported divergence between western and eastern lineages when 25 alleles were sampled. Under a four-lineage model, significant divergence was detected when sampling five or ten gene copies (Table 1); however, divergence was restricted to the splitting of WE2 from all other hypothesized lineages (WE1, EA1 and EA2). In addition, we tested models that fixed divergence between western and eastern lineages, and then tested for divergence within either the western (WE1 and WE2) or eastern lineage (EA1 and EA2). In both cases, SPEDESTEM supported models that lacked divergence within western and eastern lineages (Table S4, Supporting information).

*Delimitation hypothesis testing – BPP*

BPP analysis of the 8L data produced strong support (PPs = 1.0) for divergence between western and eastern lineages across all combinations of priors for  $\Theta$  and  $\tau$  (Fig. 3A). There was no difference in delimitation results between algorithms 0 and 1. Analyses using randomized tip labelling produced low posterior support for divergence between the western and eastern lineage, indicating that results were not biased by our choice of priors (Fig. 3B). Support for divergence between WE1 and WE2 varied across prior combinations. Small and intermediate population size priors produced strong support for divergence (PPs = 1.0), regardless of divergence time prior, while larger population size priors weakly supported divergence between WE1 and WE2. Divergence between EA1 and EA2 received weak support across all prior combinations (Fig. 3A).

BPP analysis of the 89L data set produced PPs = 1.0 for the split between the western and eastern lineages under all prior combinations (Fig. 3C). Randomized tip labelling generally yielded low PP support for the western–eastern split (Fig. 3D). For intermediate and large population size priors, posterior support varied for the EA1-EA2 and WE1-WE2 splits (Fig. 3C). Across replicates, support ranged from as little as PP = 0 to PP = 1 for these divergence events. Prior combinations featuring small population size (including our empirical-based priors) produced PPs close to one for both the EA1-EA2 and WE1-WE2 splits. Under all prior

combinations, randomized tip labelling produced PPs close to 0 for these splits.

*Delimitation hypothesis testing – SVDQUARTETS*

SVDQUARTETS analysis of the 43L data resulted in a lineage tree with a strongly supported split between the western and eastern lineages (Fig. 4A), with each forming a separate clade of haplotypes with high bootstrap support (>99%). The WE1 and WE2 splits were similarly well supported in the lineage tree, with bootstrap values of 92.6% and 99.4%, respectively. In contrast, the EA1 and EA2 groups were not resolved as reciprocally monophyletic (Fig. 4A). Branches within the eastern lineage generally received low bootstrap support. SVDQUARTETS analyses of the 8L and 43L data using a species tree framework generated a tree supporting the split between the western and eastern lineages with high levels of bootstrap support (Fig 4B; 8L: 96% and 99%, respectively; 43L: 100% for both lineages).

*Influence of data scale and content*

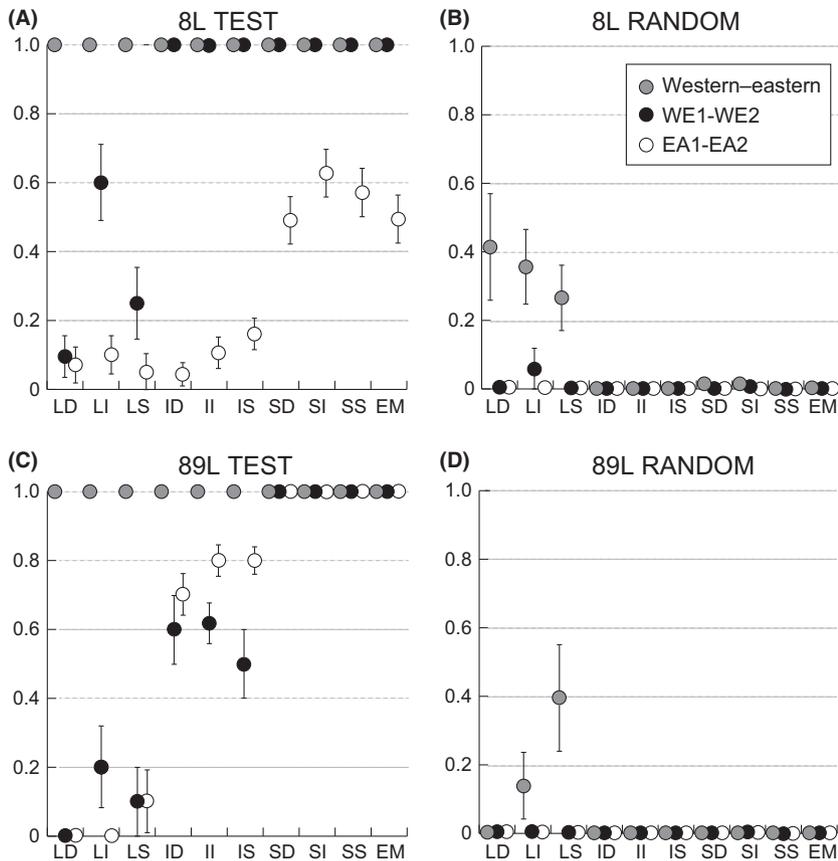
Using BPP, we achieved strong PP support for divergence between the western and eastern lineage with as few as 10 of our 89L loci (Fig. 5A) with minimal variation across replicates. When sampling 20 or more loci, support for this divergence received PPs = 1.0 across all replicates. For more shallow divergences, a greater effect of locus sampling was detected. Strong support for divergence between WE1 and WE2 was detected with as few as 30 loci (Fig. 5B); however, not all replicates provided strong support for this split, with at least one 30-locus replicate yielding PP = 0.05. This large difference in maximum and minimum posterior support persisted with increasing numbers of sampled loci, with a mean PP  $\geq$  0.95 achieved with 80 loci (minimum PP = 0.83). Similar results were obtained for analyses of the EA1-EA2 divergence (Fig. 5C). High levels of posterior support for divergence were produced with as few as 10 sampled loci (maximum PP = 0.94); however, large differences in PPs were detected across replicates, a pattern observed for most levels of locus subsampling. A total of 70 loci were required to produce a mean PP  $\geq$  0.95 (minimum PP = 0.83). Overall, for these shallower divergences, variance in support across replicates decreased with greater locus sampling.

Analysis of 89L loci with the highest number of PIS had a strong effect on support for the more shallow divergence events. Whereas at least 80 randomly sampled 89L loci were needed to produce mean PP  $\geq$  0.95 for the WE1-WE2 split, only 40 of the most informative 89L loci were needed to produce a similar level of support (Fig. 5D). As few as 30 of the most informative 89L

**Table 1** Results of SPEDESTEM analyses under two- and four-lineage scenarios. Underscores in the delimitation result indicate grouping; commas indicate divergence.  $\omega_i$  indicates the relative model probability

Gene copies sampled	Replicates	$\omega_i$	Delimitation result
<b>Two-lineage</b>			
2	1000	0.72	One species (WE_EA)
3	1000	0.72	One species (WE_EA)
5	1000	0.69	One species (WE_EA)
10	1000	0.66	One species (WE_EA)
25	100*	0.62	Two species (WE, EA)
<b>Four-lineage</b>			
2	1000	0.22	One species (WE1_WE2_EA1_EA2)
3	1000	0.20	One species (WE1_WE2_EA1_EA2)
5	1000	1.00	Two species (WE1_EA1_EA2, WE2)
10	1000	1.00	Two species (WE1_EA1_EA2, WE2)

\*Fewer replicates were performed due to computational limits.



**Fig. 3** Results from BPP analyses of the 8L and 89L data sets. Circle coloration corresponds with hypothesized divergences in Fig. 1. The x-axis is labelled with two-letter designations for prior combinations of  $\Theta$  and  $\tau$ , with  $\Theta$  designated as large (L), intermediate (I), or small (S) and  $\tau$  designated as deep (D), intermediate (I) or shallow (S). Results are also presented for empirically (EM) derived priors. In total, ten different combinations of priors for were tested and mean posterior probability and standard error is reported for 10 replicates per prior for the 89L data and 20 replicates per prior for the 8L data. Points are shaded according to nodes in the hypothesized species topology shown in Fig. 1. Figure panels correspond to (A) 8L data, tips assigned to hypothesized species, (B) 8L data, tips randomly assigned, (C) 89L data, tips assigned to hypothesized species, (D) 89L data, tips randomly assigned.

loci produced a PP > 0.95 for the WE1-WE2 split. A similar pattern was observed for the EA1-EA2 split, which required as many as 70 randomly sampled 89L loci to produce a mean PP  $\geq$  0.95, but which required only 50 of the most informative 89L loci to produce the same level of support (Fig. 5D). For the western-eastern split, as few as 10 of the most informative 89L loci produced posterior support of 1.0.

#### Phylogeographic model selection and parameter estimation – PHRAPL

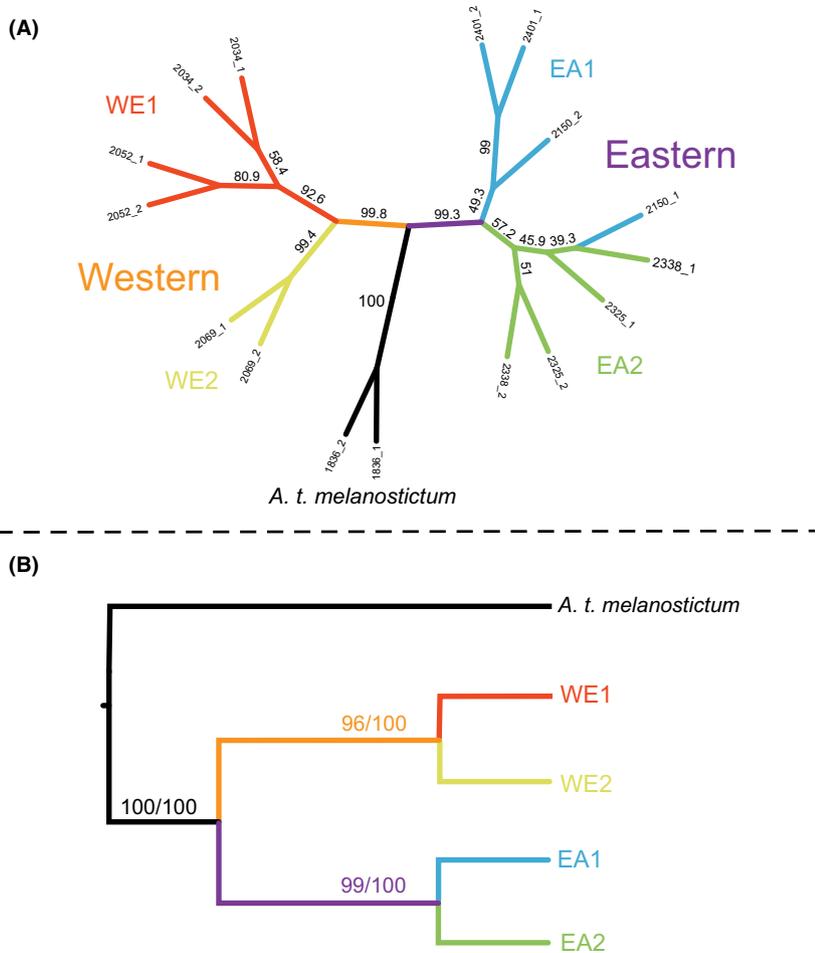
PHRAPL analysis of the 8L data resulted in the greatest support for model 4 (Fig. 2), which specified divergence between the western and eastern lineages, along with gene flow from the eastern lineage into the western lineage (Table 2; model probability = 0.67). A model treating *A. ordinarium* as a single lineage received the next highest support ( $\Delta$ AIC = 1.77, model probability = 0.28), while support for distinct western and eastern lineages with no gene flow was lowest ( $\Delta$ AIC = 35.21, model probability =  $1.51 \times 10^{-8}$ ). Model averaged parameter estimates suggest a relatively deep divergence with low-level gene flow from western populations into eastern populations and relatively high

post-divergence gene flow from eastern lineages into western lineages (Table 2).

PHRAPL analysis of the 43L data indicated the greatest support for a model of divergence between the western and eastern lineage with no gene flow (Table 2; model probability = 0.63). The next best-supported model was one of no divergence between eastern and western lineages ( $\Delta$ AIC = 3.24, model probability = 0.12). Model averaged parameter estimates suggest a relatively deep divergence with near zero post-divergence gene flow in either direction between western and eastern populations (Table 2).

#### Demographic model selection and parameter estimation – MIGRATE-N

MIGRATE-N analysis of the 8L data best supported a bidirectional migration model between the western and eastern lineages (Table 3; model 2, model probability > 0.99). The next best model included unidirectional gene flow from the eastern lineage into the western lineage; however, this model received a very low probability (log Bayes factor  $\geq$  81.6, model probability =  $1.9 \times 10^{-18}$ ). A model combining the western and eastern populations into a single



**Fig. 4** Relationships among *A. ordinarium* inferred with SVDQuartets. a) ‘Lineage tree’ for 43L data set using exhaustive sampling of quartets over 1000 bootstrap replicates. b) Species tree inferred under a four-lineage constraint for the 8L and 43L data sets using exhaustive quartet sampling and 100 bootstrap replicates. Branch support values to the left of the backslash are for the 8L data set and those to the right are for the 43L data set. Branch colours correspond to Fig. 1. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)].

population (model 1) received the lowest model probability ( $3.5 \times 10^{-105}$ ).

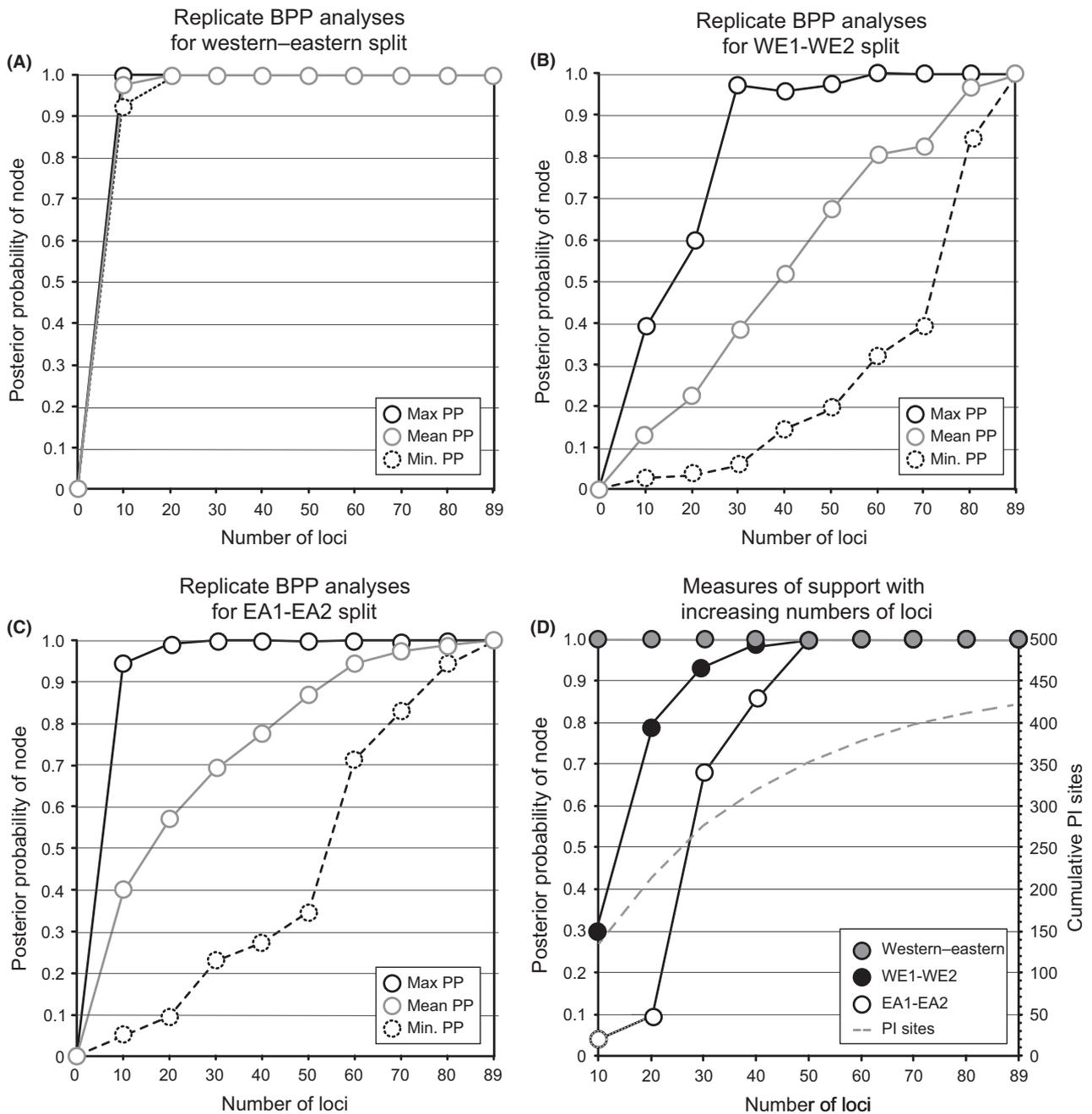
Estimates of the number of migrants per generation were significantly skewed towards migration from the western into the eastern lineage (Table S5, Supporting information), with a mean  $N_m$  of 0.44 (95% confidence interval = 0–1.29) in this direction vs. a mean  $N_m$  of 0.16 (95% confidence interval = 0–0.77) for the opposite direction.

## Discussion

### *Data sampling in species delimitation*

Species delimitation is in a state of transition in terms of the data used for analysis, with systematists facing important choices regarding the numbers of loci and individuals to sample. Recent studies have investigated the role of locus number and gene copy sampling in the performance of genetically based species delimitation methods (e.g. Hird *et al.* 2010; Camargo *et al.* 2012). However, no empirical study has compared the

influence of locus number and information content on species delimitation results. One general conclusion from this study is that both small and large data sets have the potential to resolve cryptic species boundaries between recently diverged species. Across data sets, species discovery methods (e.g. STRUCTURE) highlight the same candidate species, and species validation tests (e.g. BPP, PHRAPL) provide similarly strong support for the western–eastern divergence event within *Ambystoma ordinarium*. These results are, in part, encouraging for the broader molecular systematic community, suggesting that large-scale data sets, for example those generated with NGS methods, may not always be necessary for the delimitation of morphologically cryptic species. This may be particularly true for older and well-differentiated species, where species delimitation is expected to be straightforward (Shaffer & Thomson 2007). Within *A. ordinarium*, the split between the western and eastern lineage was recovered with strong posterior support in all BPP analyses of the 8L data and in every subsampled 10-locus 89L data set (Fig. 3A). Furthermore, given that all subsampled 10-locus 89L data sets resulted in strong



**Fig. 5** Effects of locus subsampling on BPP node support for 89L *A. ordinarium* data sets. Randomly selected loci were sampled without replacement in increments of 10 (i.e. 10, 20, 30... 89) across 10 independent replicates. The maximum, mean and minimum posterior probability (PP) for nodes in the guide tree are shown for (A) western-eastern, (B) WE1-WE2 and (C) EA1-EA2. All runs used the empirically estimated priors described in the text. (D) When loci are rank ordered by number of parsimony-informative (PI) sites and analysed in multiples of 10 loci, support for eastern-western split is unanimous and support for EA1-EA2 and WE1-WE2 both reach PP = 1.0 by 50 loci. Points in (D) are shaded according to nodes in Fig. 1.

support for the western-eastern split and that these loci were drawn from a pool of loci with a wide range of variability, for smaller data sets it may not be as important to make highly informed choices about which loci to use in the delimitation of deeper divergence events.

In contrast, our investigations of information content (i.e. PIS sites) show that all loci are not equal in their ability to recover signatures of shallower divergence events. In the case of the WE1-WE2 and EA1-EA2 splits, while at least one 30-locus data set provided strong

**Table 2** Results of PHRAPL analyses of the 8L and 43L data for five two-lineage, two-parameter phylogeographic models

Model	AIC	Log likelihood	ΔAIC	ωAIC	Rank	$\tau_{WE-EA}$	$m_{WE \rightarrow EA}$	$m_{EA \rightarrow WE}$
<i>8L</i>						4.10	0.72	3.81
1	354.22	-176.11	1.77	0.28	2			
2	357.49	-176.74	5.04	0.05	3			
3	363.86	-179.93	11.42	$2.22 \times 10^{-3}$	4			
4	352.45	-174.22	-	0.67	1			
5	387.66	-192.83	35.21	$1.51 \times 10^{-8}$	5			
<i>43L</i>						5.33	0.03	0.03
1	2069.24	-1033.62	3.24	0.12	2			
2	2071.22	-1033.61	5.23	0.05	5			
3	2069.87	-1032.94	3.88	0.09	4			
4	2069.52	-1032.76	3.52	0.11	3			
5	2065.99	-1032.00	-	0.63	1			

**Table 3** Model descriptions and selection results for a range of two-species migration models tested in MIGRATE-N. BAS: Bezier approximation score (log marginal likelihood) for all loci. LBF: log Bayes factor. LBFs and model probabilities calculated following Beerli and Palczewski (2010) and Kass and Rafferty (1995)

Model	Description	BAS	LBF	Probability	Choice
1	Panmixia	-7184.5	481.0	$3.5 \times 10^{-105}$	4
2	Full migration	-6944.0	-	1	1
3	Unidirectional: western into eastern	-7007.4	126.7	$3.1 \times 10^{-28}$	3
4	Unidirectional: eastern into western	-6984.8	81.6	$1.9 \times 10^{-18}$	2

support for these divergences, other 30-locus data sets did not (Fig. 5B,C). While this discrepancy was also observed in increasingly larger data sets, support variance declined as locus number increased. There is a wide range of phylogenetic information across our 89L loci, with the 20 most informative loci having approximately the same total number of PISs ( $n = 214$ ) as the next 50 ( $n = 184$ ; Fig. 5D), and this range in information content is likely driving the large swing in support for these more shallow divergences when randomly selecting loci. For example, the mean posterior support for the WE1-WE2 split from a randomly sampled 30-locus data set was ~0.4 (Fig. 5B), while the 30 most informative loci produced posterior support of 0.94 (Fig. 5D). Similar to conclusions derived from studies focused on factors influencing species tree reconstruction at shallow tree depths (Harris *et al.* 2014; Huang *et al.* 2010; Knowles *et al.* 2012; Lanier *et al.* 2014), our results indicate that the phylogenetic information content of loci is a primary factor in the delimitation of species separated by shallow depths of divergence.

Overall, our results provide a mixed message to the systematist considering how to generate data for a species delimitation study. A small number of loci may be sufficient to both discover and validate many cryptic species, allowing researchers with the ability to generate relatively small data sets to continue the identification of

new lineages and taxa. However, this work shows that many shallowly diverged species may go undiscovered, or not pass statistical validation via coalescent tests, when using these smaller data sets. While some cryptic species may only require a small number of loci to be detected, this is impossible to know a priori, and as a result, systematists are most likely to be assured of clarifying the boundary between cryptic species and structured populations when analysing large multilocus data sets. With the growing accessibility of genome-scale data sets for phylogeography and species delimitation (e.g. Lemmon & Lemmon 2012; Smith *et al.* 2014), an increasing number of studies are likely to include sufficient numbers of loci for drawing boundaries between intraspecific and interspecific variation. However, in the case of salamanders, large genome size (often >30 Gb) may preclude genome-scale data generation for species delimitation using standard NGS approaches, although recent advances (e.g. McCartney-Melstad *et al.* 2016) may reduce this bottleneck in the future.

#### *Phylogeographic model selection and species delimitation*

To date, most coalescent-based species delimitation studies have been restricted to the parameterization of  $N_e$  and divergence time in models meant to capture the

history of populations. Yet, data that include signatures of gene flow are likely to have impacts on likelihood calculations, with potentially important ramifications for accurate species delimitation. A few studies (Leache 2009; Jackson & Austin 2012; Ruane *et al.* 2014) have examined the impacts of gene flow on species tree reconstruction at the phylogeographic level, demonstrating the importance of considering the effects of introgression and low-level gene flow. Here, we extended these efforts using a model selection approach in PHRAPL to consider gene flow as a parameter in our species validation tests. In doing so, we rejected the hypothesis that *A. ordinarium* represents a single lineage. Our 8L and 43L data sets contrasted, however, in support for a history of gene flow between lineages, with the 8L data favouring a model that included unidirectional gene flow and the 43L data favouring a model of no gene flow. This difference may be related to the much greater individual sampling of the 8L data, which may have included individuals bearing the signatures of gene flow. Indeed, the 8L data, but not the 43L data, contained sampling from localities 7 and 8 of the eastern lineage, both of which contain individuals with signatures of admixture within the western lineage (Fig. 1B). In addition, the 8L loci are longer and more variable, on average, than the 81 PTAS loci, which may make them more informative in detecting historical, low-level gene flow. Finally, it is worth noting that estimates of gene flow between these two lineages are low, with MIGRATE-N estimates of  $N_m$  much less than one (Table S5, Supporting information). In any case, the consideration of both gene flow and divergence using a model selection approach are concordant with results from other analyses in supporting divergence between the western and eastern lineages.

We note, however, that our model selection approach faced substantial limitations in the complexity of models and the numbers of free parameters that could be considered. For example, we were unable to confidently perform analyses with additional parameters that allowed for changing  $N_e$  through time or asymmetric migration rates. Similarly, expanding PHRAPL analyses to account for models with three or four lineages, which also required an expanded set of free parameters, proved challenging. While our data appear to be informative in recovering the deeper western–eastern divergence event while accounting for gene flow, testing more complex models that take these additional parameters into account would likely require increased sampling of both individuals and loci to produce credible estimates of parameters such as migration rates and effective population sizes, and accordingly, to distinguish more complex models of gene flow and divergence from each other.

### *Species boundaries*

Collectively, our genetic results strongly support eastern and western populations of *A. ordinarium* as independently evolving population-level lineages which likely represent distinct species. Population structure results strongly delineate these as separate clusters with limited evidence for admixture, and the 8L SVDQUARTETS lineage tree reconstructs the western and eastern populations as two strongly supported clades. Coalescent-based tests using BPP validate this hypothesis; divergence between western and eastern populations was strongly supported with both our 8L and 89L data across all explored prior combinations and across different subsets of loci. Beyond genetic evidence, these two species also occur in separate headwater systems of southward-flowing streams in the TMVB (Fig. 1).

Within both putative species, support for additional levels of lineage divergence varied markedly across analyses, and our analyses suggest that these more shallow-scale divergence events do not represent species-level divergence. This conclusion is principally derived from the inconsistent patterns of support in BPP validation analyses for both data sets and varied prior combinations (Fig. 3). This was particularly true for divergence within the eastern lineage, which was poorly supported under all priors in analyses of the 8L data, and only received high posterior support in analyses of the 89L data that featured small population size priors. Similar patterns of inconsistent support in BPP validation tests were also seen within the western lineage. In addition, a subset of SPEDESTEM results split WE2 while lumping all other hypothesized lineages, in contrast to all species discovery results, which found a clear division between eastern and western populations. This latter result could indicate an inapplicability of SPEDESTEM to this particular study (Camargo *et al.* 2012; Carstens *et al.* 2013). However, while population structure is clearly evident within and between the eastern and western lineages, given the lack of consistent support for the delimitation of additional lineages, and with an aim to not promote taxonomic instability (e.g. Turtle Taxonomy Working Group 2007), we refrain from describing additional species-level taxa within this group at present.

### *Evidence for lineage divergence within A. ordinarium*

Although we do not have a firm divergence time estimate for the split between the western and eastern lineages, we expect that it does not coincide with the geological evolution of the TMVB. The majority of tectonic and volcanic activity producing the TMVB occurred in the mid-Miocene, approximately 7–

11 million years ago (Ferrari *et al.* 1999, 2000), likely predating the common ancestor of the entire *A. tigrinum* species complex (Shaffer & McKnight 1996). Phylogeographic studies of birds (McCormack *et al.* 2008), lizards (Zarza *et al.* 2008) and toads (Mulcahy *et al.* 2006) support the role of the TMVB in Pliocene-Pleistocene species divergence, a time of active, but less extreme uplift (Ferrusquía-Villafranca & González-Guzmán 2005). Tectonic and volcanic activity in the TMVB also substantially changed its hydrology over time (Israde-Alcantara & Garduno-Monroy 1999), leading to the divergence of fish species (Mateos *et al.* 2002; Schönhuth & Doadrio 2003; Doadrio & Dominguez 2004; Hulseley *et al.* 2004; Dominguez-Dominguez *et al.* 2008). Many of these species divergences have been dated to the Pliocene with the most recent described divergence among them occurring 0.6–0.8 Ma between two allopatric species of the genus *Allotoca* found in Lakes Patzcuaro and Zirahuén (Dominguez-Dominguez *et al.* 2006).

The Late Pleistocene in central Mexico was characterized by cooler and drier conditions (Metcalfe *et al.* 2000), and palynological data indicate an absence of pine forest across upper elevations (>2500 m) of much of central Mexico at this time (Lozano-García & Vazquez-Selem 2005). In contrast, pollen studies of Lago Patzcuaro, which is at a lower elevation than contemporary *A. ordinarium* populations (~2000 m), reveal stable pine forest over the last 48 000 years, indicating that lower elevations of central Mexico were not as strongly impacted by drier conditions (Bradbury 2000). Given this environmental history, it is possible that *A. ordinarium* populations, which are facultative in their ability to metamorphose, tracked the movement of available pine forest into lower elevations during the late Pleistocene and into the Holocene, and that these distributional shifts, perhaps into refugia representative of the current drainage basins occupied by *A. ordinarium*, initiated lineage divergence. The strong signature of recent mtDNA introgression between the western lineage of *A. ordinarium* and the Lago Patzcuaro endemic paedomorph *A. dumerilii* (Weisrock *et al.* 2006; Fig. 1), combined with their current allopatric distribution, further supports this lower elevation refugium hypothesis.

### Conservation implications

We anticipate that evidence for an additional, cryptic species within the endangered and range-restricted *A. ordinarium* may have immediate conservation implications for this group of ambystomatid salamanders. *Ambystoma ordinarium* already has an IUCN 'Endangered' listing due to its limited distribution and disappearing forest habitat (Shaffer *et al.* 2004). If *A. ordinarium* is actually composed of two distinct species, each by necessity has an even

more restricted range and must be considered even more fragile and threatened. As an example in ambystomatid salamander conservation, the discovery and recognition of two cryptic species (*A. cingulatum* and *A. bishopi*) within the former *A. cingulatum* species (Pauly *et al.* 2007) was quickly adopted by the U.S. Fish and Wildlife Service, with both species upgraded to endangered status. As such, the recognition of cryptic and recently diverged species using the methods outlined here may be especially important beyond salamanders in the conservation of biodiversity in recently derived, endangered taxa.

### Acknowledgements

We thank Brian and Shonna Storz for their efforts in the field to sample *A. ordinarium* across its range. This study was greatly improved through discussions with participants of a Species Delimitation Workshop at the National Institute of Mathematical and Biological Synthesis, and through additional discussions with Laura Kubatko and Tara Pelletier. The design and implementation of PHRAPL analyses were greatly aided through discussions with PHRAPL workshop participants at the Ohio State University. We thank the University of Kentucky Center for Computational Sciences and Lipscomb High Performance Computing Cluster, as well as Jeremiah Smith, for access to supercomputing resources. We thank Sebastian Voitel for providing a photograph of *A. ordinarium*. We also thank associate editor Bryan Carstens and two anonymous reviewers for valuable comments that improved the final manuscript. This research was supported by the National Science Foundation through awards DEB-1257648 and DEB-1457832 (to HBS), DEB-0949532 (to DWW and EMO), and DEB-1355000 (to DWW). This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 3048109801 to PMH. SH was supported by a University of Kentucky Department of Biology Gertrude Flora Ribble Fellowship.

### References

- Anderson JD, Worthington RD (1971) The life history of the Mexican salamander *Ambystoma ordinarium*. *Herpetologica*, **27**, 165–176.
- Beerli P (2006) Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics*, **22**, 341–345.
- Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 4563–4568.
- Beerli P, Palczewski M (2010) Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics*, **185**, 313–326.
- Bradbury J (2000) Limnologic history of Lago de Patzcuaro, Michoacan, Mexico for the past 48,000 years: impacts of climate and man. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **163**, 69–95.

- Camargo A, Morando M, Avila LJ, Sites JW 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 darwini* complex (Squamata: Liolaemidae). *Evolution*, **66**, 2834–2849.
- Carstens BC, Pelletier TA, Reid NM, Satler JD (2013) How to fail at species delimitation. *Molecular Ecology*, **22**, 4369–4383.
- Chifman J, Kubatko L (2014) Quartet inference from SNP data under the coalescent model. *Bioinformatics*, **30**, 3317–3324.
- Doadrio I, Dominguez O (2004) Phylogenetic relationships within the fish family Goodeidae based on cytochrome *b* sequence data. *Molecular Phylogenetics and Evolution*, **31**, 416–430.
- Dominguez-Dominguez O, Doadrio I, de Leon GPP (2006) Historical biogeography of some river basins in central Mexico evidenced by their goodeine freshwater fishes: a preliminary hypothesis using secondary Brooks parsimony analysis. *Journal of Biogeography*, **33**, 1437–1447.
- Dominguez-Dominguez O, Alda F, de Leon GPP, Garcia-Garitaotia JL, Doadrio I (2008) Evolutionary history of the endangered fish *Zoogoneticus quitzeoensis* (Bean, 1898) (Cyprinodontiformes: Goodeidae) using a sequential approach to phylogeography based on mitochondrial and nuclear DNA data. *BMC Evolutionary Biology*, **8**, 161.
- Edwards DL, Knowles LL (2014) Species detection and individual assignment in species delimitation: can integrative data increase efficacy? *Proceedings of the Royal Society B*, **281**, 20132765.
- Ence DD, Carstens BC (2011) SpedeSTEM: a rapid and accurate method for species delimitation. *Molecular Ecology Resources*, **11**, 473–480.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.
- Ferrari L, Lopez-Martinez M, Aguirre-Diaz G, Carrasco-Nunez G (1999) Space-time patterns of Cenozoic arc volcanism in central Mexico: from the Sierra Madre Occidental to the Mexican Volcanic Belt. *Geology*, **27**, 303–306.
- Ferrari L, Conticelli S, Vaggelli G, Petrone CM, Manetti P (2000) Late Miocene volcanism and intra-arc tectonics during the early development of the Trans-Mexican Volcanic Belt. *Tectonophysics*, **318**, 161–185.
- Ferrusquía-Villafranca I, González-Guzmán LI (2005) Northern Mexico's landscape, part II: the biotic setting across time. In: *Biodiversity, Ecosystems, and Conservation in Northern Mexico* (eds Cartron J-LE, Ceballos G, Flegler RS), pp. 39–51. Oxford University Press, Oxford, UK.
- Fujita MK, Leache AD, Burbrink FT, McGuire JA, Moritz C (2012) Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution*, **27**, 480–488.
- Grummer JA, Bryson RW Jr, Reeder TW (2014) Species delimitation using Bayes factors: simulations and application to the *Sceloporus scalaris* species group (Squamata: Phrynosomatidae). *Systematic Biology*, **63**, 119–133.
- Harris RB, Carling MD, Lovette IJ (2014) The influence of sampling design on species tree inference: a new relationship for the New World chickadees (Aves: *Poecile*). *Evolution*, **68**, 501–513.
- Hey J (2010) Isolation with migration models for more than two populations. *Molecular Biology and Evolution*, **27**, 905–920.
- Hird S, Kubatko L, Carstens B (2010) Rapid and accurate species tree estimation for phylogeographic investigations using replicated subsampling. *Molecular Phylogenetics and Evolution*, **57**, 888–898.
- Hotaling S, Foley ME, Lawrence NM *et al.* (2016) Species discovery and validation in a cryptic radiation of endangered primates: coalescent-based species delimitation in Madagascar's mouse lemurs. *Molecular Ecology*, **25**, 2029–2045.
- Huang H, He Q, Kubatko LS, Knowles LL (2010) Sources of error inherent in species-tree estimation: impact of mutational and coalescent effects on accuracy and implications for choosing among different methods. *Systematic Biology*, **59**, 573–583.
- Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, **18**, 337–338.
- Huelsenbeck JP, Andolfatto P, Huelsenbeck ET (2011) Structurama: Bayesian inference of population structure. *Evolutionary Bioinformatics*, **7**, 55–59.
- Hulsey CD, de Leon FJG, Johnson YS, Hendrickson DA, Near TJ (2004) Temporal diversification of Mesoamerican cichlid fishes across a major biogeographic boundary. *Molecular Phylogenetics and Evolution*, **31**, 754–764.
- Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.
- Irwin DE (2002) Phylogeographic breaks without geographic barriers to gene flow. *Evolution*, **12**, 2383–2394.
- Israde-Alcantara I, Garduno-Monroy VH (1999) Lacustrine record in a volcanic intra-arc setting: the evolution of the Late Neogene Cuitzeo basin system (central-western Mexico, Michoacan). *Palaeogeography Palaeoclimatology Palaeoecology*, **151**, 209–227.
- Jackson ND, Austin CC (2012) Inferring the evolutionary history of divergence despite gene flow in a lizard species, *Scincella lateralis* (Scincidae), composed of cryptic lineages. *Biological Journal of the Linnean Society*, **107**, 192–209.
- Jonsson H, Schubert M, Seguin-Orlando A *et al.* (2014) Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 18655–18660.
- Kass RE, Raftery RE (1995) Bayes factors. *Journal of the American Statistical Association*, **90**, 773–795.
- Knowles LL, Lanier HC, Klimov PB, He Q (2012) Full modeling versus summarizing gene-tree uncertainty: method choice and species-tree accuracy. *Molecular Phylogenetics and Evolution*, **65**, 501–509.
- Lanier HC, Huang H, Knowles LL (2014) How low can you go? The effects of mutation rate on the accuracy of species-tree estimation. *Molecular Phylogenetics and Evolution*, **70**, 112–119.
- Leache AD (2009) Species tree discordance traces to phylogeographic clade boundaries in North American fence lizards (*Sceloporus*). *Systematic Biology*, **58**, 547–559.
- Lemmon AR, Lemmon EM (2012) High-throughput identification of informative nuclear loci for shallow-scale phylogenetics and phylogeography. *Systematic Biology*, **61**, 745–761.
- Lozano-Garcia S, Vazquez-Selem L (2005) A high-elevation Holocene pollen record from Iztaccihuatl volcano, central Mexico. *The Holocene*, **15**, 329–338.

- Mateos M, Sanjur OI, Vrijenhoek RC (2002) Historical biogeography of the livebearing fish genus *Poeciliopsis* (Poeciliidae: Cyprinodontiformes). *Evolution*, **56**, 972–984.
- McCartney-Melstad E, Mount GG, Shaffer HB (2016) Exon capture optimization in large-genome amphibians. *Molecular Ecology Resources*, **16**, 1084–1094.
- McCormack JE, Peterson AT, Bonaccorso E, Smith TB (2008) Speciation in the highlands of Mexico: genetic and phenotypic divergence in the Mexican jay (*Aphelocoma ultramarina*). *Molecular Ecology*, **17**, 2505–2521.
- Metcalfe SE, O'Hara SL, Caballero M, Davies SJ (2000) Records of Late Pleistocene-Holocene climatic change in Mexico - a review. *Quaternary Science Reviews*, **19**, 699–721.
- Mulcahy DG, Morrill BH, Mendelson JR (2006) Historical biogeography of lowland species of toads (*Bufo*) across the Trans-Mexican Neovolcanic Belt and the Isthmus of Tehuantepec. *Journal of Biogeography*, **33**, 1889–1904.
- Olave M, Sola E, Knowles LL (2014) Upstream analyses create problems with DNA-based species delimitation. *Systematic Biology*, **63**, 263–271.
- O'Meara BC, Jackson ND, Morales-Garcia AE, Carstens BC (2015) Phylogeographic inference using approximate likelihoods. *BioRxiv*, doi:10.1101/025353
- O'Neill EM, Schwartz R, Bullock CT *et al.* (2013) Parallel tagged amplicon sequencing reveals major lineages and phylogenetic structure in the North American tiger salamander (*Ambystoma tigrinum*) species complex. *Molecular Ecology*, **22**, 111–129.
- Pauly GB, Piskurek O, Shaffer HB (2007) Phylogeographic concordance in the southeastern United States: the flatwoods salamander, *Ambystoma cingulatum*, as a test case. *Molecular Ecology*, **16**, 415–429.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- Rannala B (2015) The art and science of species delimitation. *Current Zoology*, **61**, 846–853.
- Reid NM, Hird SM, Brown JM *et al.* (2014) Poor fit to the multispecies coalescent is widely detectable in empirical data. *Systematic Biology*, **63**, 322–333.
- Rittmeyer EN, Austin CC (2015) Combined next-generation sequencing and morphology reveal fine-scale speciation in Crocodile Skinks (Squamata: Scincidae: *Tribolonotus*). *Molecular Ecology*, **24**, 466–483.
- Ruane S, Bryson RW Jr, Pyron RA, Burbrink FT (2014) Coalescent species delimitation in milksnakes (genus *Lampropeltis*) and impacts on phylogenetic comparative analyses. *Systematic Biology*, **63**, 231–250.
- Schönhuth S, Doadrio I (2003) Phylogenetic relationships of Mexican minnows of the genus *Notropis* (Actinopterygii, Cyprinidae). *Biological Journal of the Linnean Society*, **80**, 323–337.
- Shaffer HB, McKnight ML (1996) The polytypic species revisited: genetic differentiation and molecular phylogenetics of the tiger salamander *Ambystoma tigrinum* (Amphibia: Caudata) complex. *Evolution*, **50**, 417–433.
- Shaffer HB, Thomson RC (2007) Delimiting species in recent radiations. *Systematic Biology*, **56**, 896–906.
- Shaffer HB, Flores-Villela O, Parra-Olea G, Wake DB (2004) *Ambystoma ordinarius*. IUCN 2009. IUCN Red List of Threatened Species. Version 2009.1
- Smith BT, Harvey MG, Faircloth BC, Glenn TC, Brumfield RT (2014) Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Systematic Biology*, **63**, 83–95.
- Swofford DL (2015) *PAUP\* Version 4.0a146. Phylogenetic Analysis Using Parsimony (\*and Other Methods)*. Sinauer Associates, Sunderland, Massachusetts.
- Turtle Taxonomy Working Group (2007) Turtle taxonomy: methodology, recommendations, and guidelines. *Chelonian Research Monographs*, **4**, 73–84.
- Weisrock DW, Shaffer HB, Storz BL, Storz SR, Voss SR (2006) Multiple nuclear gene sequences identify phylogenetic species boundaries in the rapidly radiating clade of Mexican ambystomatid salamanders. *Molecular Ecology*, **15**, 2489–2503.
- Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 9264–9269.
- Yang Z, Rannala B (2014) Unguided species delimitation using DNA sequence data from multiple Loci. *Molecular Biology and Evolution*, **31**, 3125–3135.
- Zarza E, Reynoso VH, Emerson BC (2008) Diversification in the northern neotropics: mitochondrial and nuclear DNA phylogeography of the iguana *Ctenosaura pectinata* and related species. *Molecular Ecology*, **17**, 3259–3275.
- Zhang C, Rannala B, Yang Z (2014) Bayesian species delimitation can be robust to guide-tree inference errors. *Systematic Biology*, **63**, 993–1004.

## Data accessibility

All DNA sequence alignments are available in this study's Dryad accession (doi:10.5061/dryad.td997).

P.M.H., S.H., H.B.S. and D.W.W. designed the study. H.B.S. and S.R.V. provided samples. E.M.O. performed the laboratory work for the 89L data set. P.M.H., S.H., R.E.G. and D.W.W. analysed the data. P.M.H., S.H. and D.W.W. wrote the manuscript. All authors read and approved the final version.

## Supporting information

Additional supporting information may be found in the online version of this article.

**Appendix S1** Additional details of Methods.

**Figure S1** Demographic and phylogeographic models for the western and eastern *A. ordinarius* lineages tested in MIGRATE-N analyses.

**Figure S2** Additional STRUCTURE results for non-hierarchical  $K = 3$  and  $K = 4$ , and for optimal  $K$  (3) from STRUCTURAMA treating  $K$  as a random variable.

**Figure S3** Gene trees estimated for 8L loci.

**Figure S4** Generalized multilocus haplotype networks for *A. ordinarium* inferred in SPLITSTREE for the (A) 8L and (B) 89L data sets. Uncorrected P-distances were plotted using convex hull representation.

**Table S1** Sampling information for 8L and 43L/89L data sets.

**Table S2** Summary statistics for all loci included in the 8L data set.

**Table S3** Summary statistics of the 89 loci included in the 43L and 89L data sets.

**Table S4** Results of SPEDESTEM analyses under two-three-lineage scenarios: (A) the eastern populations are fixed as a lineage and the split between WE1 and WE2 is tested, and (B) the western populations are fixed as a single lineage and the split between EA1 and EA2 is tested.

**Table S5** Rate of migration ( $M$ ), direction,  $\theta$  (mutation-scaled effective population size), and  $m$  (number of immigrants per generation) for the best-fit model (Model 2) between eastern and western *A. ordinarium* lineages estimated using MIGRATE-N.