

Historical Species Distribution Models Predict Species Limits in Western *Plethodon* Salamanders

TARA A. PELLETIER¹, CHARLIE CRISAFULLI², STEVE WAGNER³, AMANDA J. ZELLMER⁴, AND BRYAN C. CARSTENS^{1,*}

¹Department of Evolution, Ecology and Organismal Biology, Ohio State University, Columbus, OH 43201; ²U.S. Forest Service, Pacific Northwest Research Station, Olympia, WA 98512; ³Department of Biological Sciences, Central Washington University, Ellensburg, WA 98926; and ⁴Department of Biology, Occidental College, Los Angeles, CA 90041, USA

*Correspondence to be sent to: Department of Evolution, Ecology and Organismal Biology, Ohio State University, Columbus, OH 43201, USA; E-mail: carstens.12@osu.edu.

Received 13 August 2014; reviews returned 30 October 2014; accepted 10 November 2014
Associate Editor: Jean-Francois Flot

Abstract.—Allopatry is commonly used to predict boundaries in species delimitation investigations under the assumption that currently allopatric distributions are indicative of reproductive isolation; however, species ranges are known to change over time. Incorporating a temporal perspective of geographic distributions should improve species delimitation; to explore this, we investigate three species of western *Plethodon* salamanders that have shifted their ranges since the end of the Pleistocene. We generate species distribution models (SDM) of the current range, hindcast these models onto a climatic model 21 Ka, and use three molecular approaches to delimit species in an integrated fashion. In contrast to expectations based on the current distribution, we detect no independent lineages in species with allopatric and patchy distributions (*Plethodon vandykei* and *Plethodon larselli*). The SDMs indicate that probable habitat is more expansive than their current range, especially during the last glacial maximum (LGM) (21 Ka). However, with a contiguous distribution, two independent lineages were detected in *Plethodon idahoensis*, possibly due to isolation in multiple glacial refugia. Results indicate that historical SDMs are a better predictor of species boundaries than current distributions, and strongly imply that researchers should incorporate SDM and hindcasting into their investigations and the development of species hypotheses. [Coalescent; niche model; Pacific Northwest; range expansion; species delimitation.]

The geographic range occupied by a species provides clues into its evolutionary history, including phenomena such as ring species (e.g., Wake 1997), hybridization (e.g., Riley et al. 2003), and local adaptation (e.g., Stephens et al. 1999). Data from field collections document the geographic range of a species, and are thus vital to systematic biology (e.g., Mayr 1942), and provide key insights into speciation (e.g., Coyne and Orr 2004). As such, distributions of species, particularly allopatric distributions within nominal species, often motivate species delimitation studies (e.g., Rosell et al. 2010; Weisrock et al. 2010; Leaché and Fujita 2010; Camargo et al. 2012). For example, under the biological species concept (Mayr 1942), allopatry can be interpreted as evidence in favor of reproductive isolation (and thus species status). Data that establish allopatric distributions are especially useful to compliment genetic data when little or no morphological or ecological variation is present (Fujita et al. 2012; Carstens et al. 2013). Molecular investigations into species limits will necessarily utilize geographic data, and can incorporate climatic data and species distribution modeling (SDM) in order to quantify geographic distributions.

SDM (Austin 2002; Guisan and Thuiller 2005), also known as ecological niche modeling (ENM; Soberon 2005; Lozier et al. 2009), is an approach that takes as input a set of sampling localities and a set of climatic data to identify the environmental conditions that best predict the current distribution of the focal taxon. SDMs of current conditions have been utilized by several species delimitation investigations. For example, Rissler and Apodaca (2007) used this approach to quantify differences in the environmental niche of

two putative lineages to demonstrate the correlation between genetic and environmental divergence in the salamander *Aneides flavipunctatus*. Other investigations have followed this approach (e.g., Burbrink et al. 2011; Florio et al. 2012; Zhou et al. 2012), strengthening the overall case for cryptic species. Modeling of these allopatric lineages does not provide causal evidence that ecological differentiation is promoting speciation, but instead indicates that the environmental conditions in the disjunct regions occupied by putative lineages are different, thus implying that dispersal between regions is difficult.

Estimates of the current distribution of a focal taxon do not necessarily reflect the conditions under which species arose, especially in modern times where species' habitats are being lost from the globe at an unprecedented rate (McKee et al. 2004; Butchart et al. 2010). It is therefore imperative to consider the dynamic nature of species ranges as distributions may change dramatically in size and location over time (Kirkpatrick and Barton 1997; Tingley and Beissinger 2009). Although identifying temporally dynamic species distributions may be difficult, an explicit consideration of changes in the distribution of taxa over time may inform species delimitation investigations. It is vital in temperate species that inhabit regions impacted by glaciation, because such species have experienced dramatic shifts in their range within a relatively recent geologic timeframe. Changes in the distributions of species can be estimated using paleodistribution modeling methods. SDMs make use of GIS layers of climate data that are collected from global weather stations. These data also serve as the basis for GIS layers that describe the historical climate

at certain times in the past, notably the last glacial maximum (LGM) (Otto-Bliesner et al. 2006). SDMs of the current distribution can thus be projected onto the historical layers to estimate the historical range of the focal taxon. Paleodistribution models have provided critical information to comparative phylogeographic studies (e.g., Hugall et al. 2002) and served as the basis for phylogeographic hypotheses (Carstens and Richards 2007; Collevatti et al. 2012), but are generally not directly applied to investigations that explicitly seek to discover or validate species boundaries. This represents a missed opportunity because short of paleopollen or fossils, which are available for few taxa, SDMs projected into the past represent the only method for estimating the historical range of many species. We demonstrate here that intriguing inferences are possible when SDMs, paleodistribution models, and molecular approaches to species delimitation are combined, by using current SDMs to validate molecular results, and also using past SDMs as a tool for developing hypotheses about species limits that can be validated with molecular data.

Plethodon salamanders are an ideal system to explore the relative influence of current and historical ranges on species limits. Often, there is little morphological or ecological variation among closely related taxa (e.g., Mueller et al. 2004; Kozak et al. 2006; Wiens et al. 2006; Wake 2009). Dermal respiration limits plethodontids to cool and moist habitat; thus ecologically constrained, they tend to exhibit niche conservatism (Kozak et al. 2006; Wake 2009). Although vicariance is often invoked as the cause of species formation in Plethodontidae, morphological homoplasy and/or conservatism makes incipient species difficult to detect (Mueller et al. 2004; Wake 2009). Genetic studies have substantially aided cryptic species identification in some taxa. For example, genetic data revealed that at least two-dozen speciation events in eastern *Plethodon* occurred in the Appalachian Mountains during the warm, dry climates of the late Miocene and Pliocene (7.2–2.5 Ma; Highton and Larson 1979; Highton 1995; Shepard and Burbrink 2009; Wake 2009; Highton et al. 2012). The inability of a species to traverse unsuitable habitat may constrain migration among populations leading to speciation (Kozak and Wiens 2006); however, this widely accepted concept is not well studied.

We focus here on plethodontids from the Pacific Northwest of North America (PNW), because this region has experienced dramatic climate driven shifts in habitat since the end of the Pleistocene (Brunsfeld et al. 2001). The PNW contains multiple contact zones and phylogeographic breaks (Swenson and Howard 2005), likely induced by some combination of climatic variation and shifting species ranges in response to glaciation. Because glaciations during the Pleistocene were cyclical, climatic conditions at the LGM likely represent the greatest contrast to present-day climatic conditions that we can accurately evaluate using SDMs, and thus those from the present and the LGM represent the extremes of range shifts in species from this region.

The current ranges of three western *Plethodon* species present an opportunity to apply recently developed species delimitation methods for taxa with contrasting distribution patterns (Fig. 1). *Plethodon vandykei* occurs in three discrete, isolated regions in Washington State (WA): the crest to foothills of the southern Cascade Range (hereafter Cascades), the Olympic Peninsula, and the Willapa Hills, though allozymes suggest that the Olympic and Willapa regions are nearly identical (Howard et al. 1993). *Plethodon larselli* is extremely fragmented, with a patchy distribution from the southern margin of the Columbia River in Oregon (OR) to as far north as Wenatchee National Forest (Aubry et al. 1987). If these current distributions are predictive of lineage boundaries, we expect to detect two or three independent lineages in both *P. vandykei* and *P. larselli* (see species tree hypotheses in Fig. 2). Evidence for cryptic diversity is found in both mitochondrial DNA (mtDNA) sequence data and RAPDs, which indicate divergence among groups on the north and south banks of the Columbia River and western Washington (Wagner et al. 2005). In contrast, *P. idahoensis* has an expansive and continuous range throughout the inland temperate rainforest of Idaho, although most of its range was covered by glacial ice as recently as approximately 21 Ka. According to this distribution, we expect to detect only one lineage in this species. Signal in the mtDNA is consistent with recent population expansion and the presence of two haplotype clades suggests population structure due to multiple glacial refugia (Carstens et al. 2004; Carstens and Richards 2007).

Our investigation proceeds along two fronts. First, we collect multilocus sequence data from across the range of each species and investigate species limits using several molecular species delimitation methods. Species limits are defined as lineages with their own unique evolutionary history as in de Queiroz (2007). We approach species delimitation from discovery and validation standpoints (Ence and Carstens 2011), while considering geographic breaks within the current range of the nominal species to develop our hypotheses. We also explore the robustness of some of these methods as each has their own set of assumptions and limitations (Carstens et al. 2013). Second, SDM is used to estimate the current and historical ranges based on existing locality information. Taken together, these complementary approaches allow us to explore the extent to which the possible range of each species may have changed during the late Pleistocene/Holocene, and how such changes might inform investigations of species boundaries.

METHODS

Genetic Data Collection

Sequence data were collected from several loci from all three species covering their current geographic distributions (average 7.5 individuals per potential lineage; Table 1). Primers were identified either from

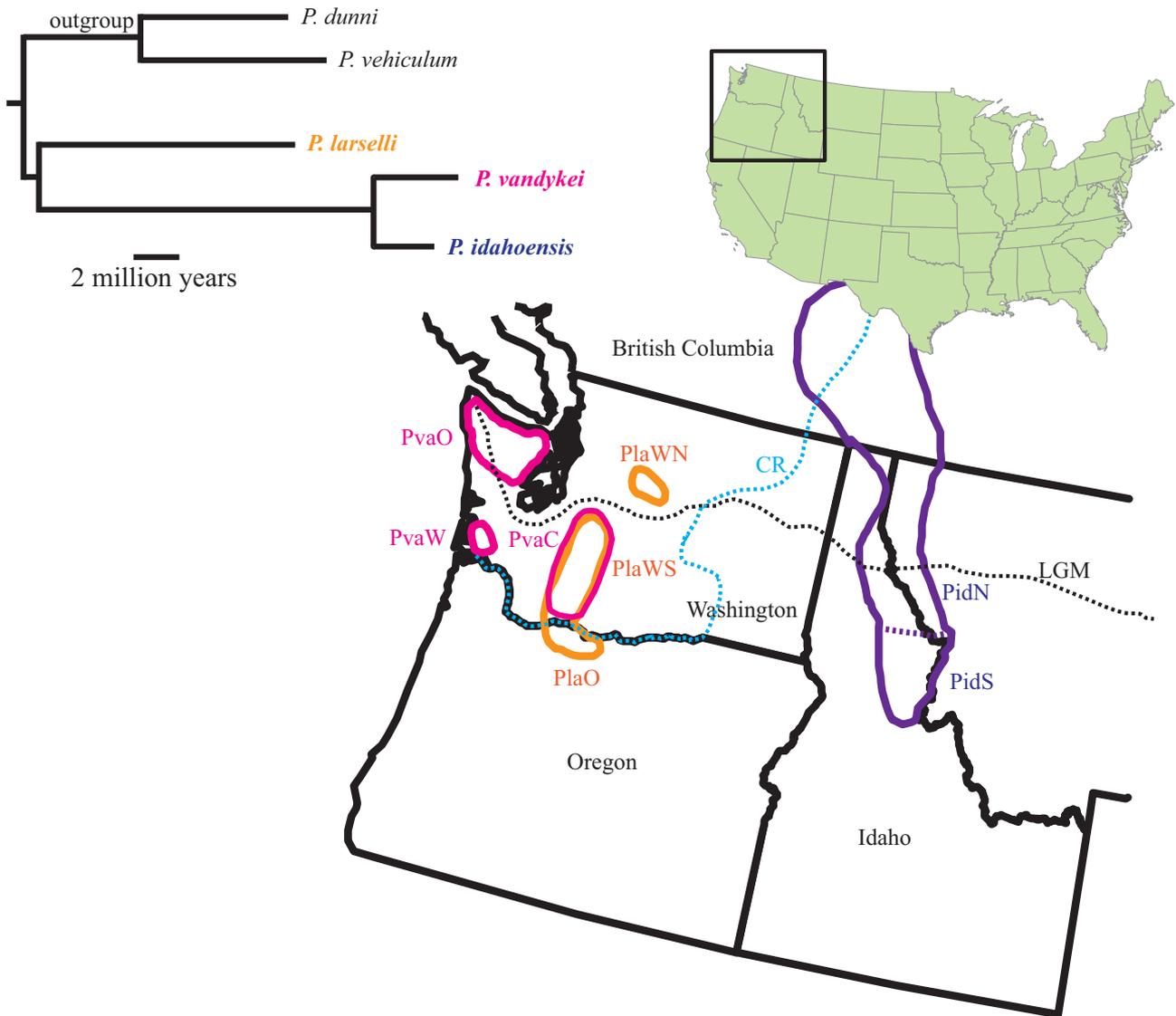


FIGURE 1. Distribution map and species tree estimate for the western *Plethodon* (*P. vandykei*, *P. idahoensis*, and *P. larselli*) in the Pacific Northwest of North America. *Plethodon vandykei*, is shown in pink, *P. idahoensis* in purple (dotted line indicates split between northern and southern river drainages), and *P. larselli* in orange. The black dotted line represents the extent of the ice sheet at the LGM (~21kya), and the blue dotted line is the Columbia River (CR). Labeling corresponds to that in text and Table 1, and represents potential independent lineages. *Plethodon vandykei*: PvaC=Cascade Mountains, PvaO = Olympic Peninsula, PvaW = Willapa Hills; *P. larselli*: PlaWN=Washington north, PlaWS=Washington south, PlaOR=Oregon; and *P. idahoensis*: PidN=northern river drainages, PidS=southern river drainages.

the literature or developed using a genomic library (see Table 2 for details). Genomic DNA libraries were constructed from *Plethodon vehiculum/dunni* and *Ensatina sp.*, using restriction digest and subsequent subcloning using the Qiagen PCR cloning kit. In total, 70 primer pairs were tested across *P. idahoensis*, *P. vandykei*, and *P. larselli* individuals. Many primers did not amplify across all three species and of those that did, many contained multiple bands or peaks in the electropherograms. Given the large genomes of these species (*P. vandykei*, C-value=69.3 pg [Mizuno and Macgregor 1974]; *P. larselli*=49.5 pg [Sessions and Larson 1987]), it seems likely that many primers amplified paralogous genetic regions, and thus were discarded. Additional information regarding the tested primers is available

as Supplementary Material (S1, available on Dryad at <http://dx.doi.org/10.5061/dryad.55137>).

Ultimately, we confidently sequenced up to nine loci in each species, with five homologous loci across species (Table 2). These include the mtDNA cytochrome b gene (Cyt b) and the nDNA recombination activating gene 1 (RAG1), glyceraldehyde-3-phosphate dehydrogenase gene (GAPD), the internal transcribed spacer subunit 1 (ITS1), ribosomal protein L12 (RPL12), heat shock 70 protein 8 (HSPA8), and three anonymous nuclear loci. Conditions for PCR programs are available as Supplementary Material (S1 on Dryad at <http://dx.doi.org/10.5061/dryad.55137>). Sanger sequencing was carried out with BigDye Terminator v3.1 on an ABI 3130XL Genetic Analyzer

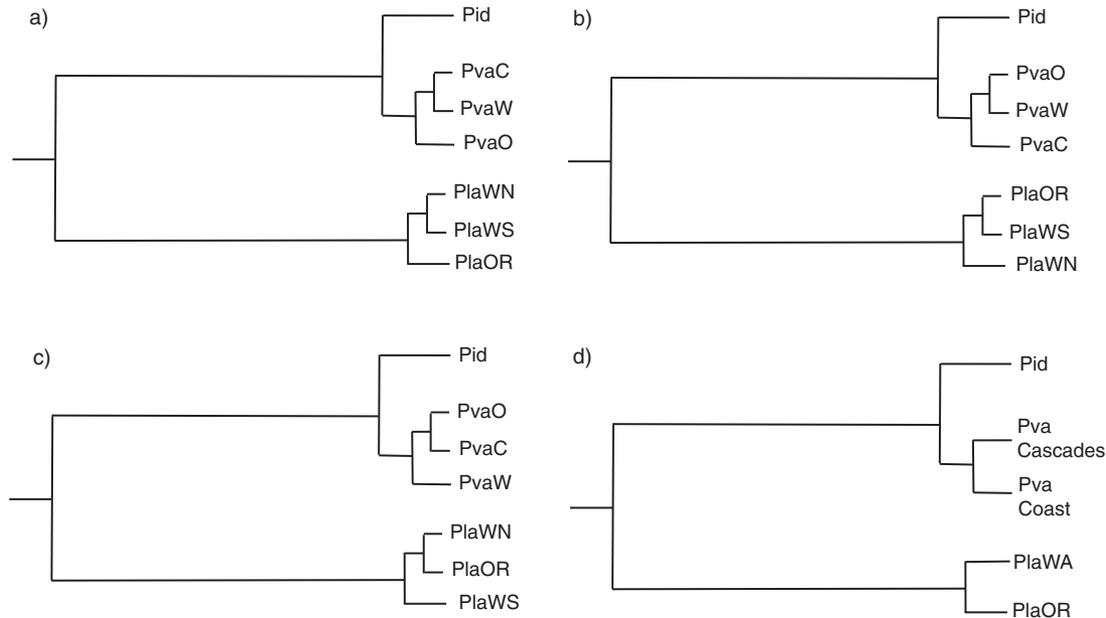


FIGURE 2. Species trees. (a–c) All possible species trees used as guide trees in the BPP analyses from the seven putative lineages hypothesized. (a) Topology recovered from *BEAST analysis. (d) Alternative hypothesis for independent lineages in *P. vandykei* and *P. larselli*, where there are two independent lineages within each species instead of three. Population labels correspond to those in Table 1 and Figure 1.

TABLE 1. Sampling distribution from putative lineages

Population ID	Species	Geographic region	<i>n</i>
PvaC	<i>P. vandykei</i>	Lower Cascades Washington	11
PvaO	<i>P. vandykei</i>	Olympic Peninsula Washington	6
PvaW	<i>P. vandykei</i>	Willapa Hills Washington	3
PlaWN	<i>P. larselli</i>	Kittitas Co Washington	7
PlaWS	<i>P. larselli</i>	Skamania Co Washington	6
PlaOR	<i>P. larselli</i>	Multnomah Co Oregon	6
PidN	<i>P. idahoensis</i>	Northern River drainages of Idaho	11
PidS	<i>P. idahoensis</i>	Southern River drainages of Idaho	10
Total			60

Notes: Population labels correspond to those in Figures 1 and 2. *Plethodon vandykei*: PvaO = Olympic Peninsula, PvaW = Willapa Hills, PvaC = Cascade Mountains; *P. larselli*: PlaOR = Oregon, PlaWS = Washington south, PlaWN = Washington north; and *P. idahoensis*: PidN = northern river drainages, PidS = southern river drainages.

(Applied Biosystems). Sequence editing and alignment were conducted using Geneious v5.4 (Drummond et al. 2011); alignments were generally unambiguous due to a paucity of indels. Sequence data were phased to alleles using PHASEV2.1 (Stephens et al. 2001) and SeqPHASE (Flot 2010) to generate files; most alleles were phased at high probability (> 0.9), otherwise individuals were sub-cloned using the Qiagen PCR cloning kit to determine phase. The GAPD locus included heterozygous indels so CHAMPURUV1.0 (Flot 2007) was used to determine phase for some individuals. Sequences are deposited in GenBank and alignments from this study in Dryad <http://dx.doi.org/10.5061/dryad.55137>.

Because many methods described below are derived from coalescent theory and thus assume that the loci used are not under selection and there is no recombination, we conducted several tests prior to analysis. The SBP and GARD methods implemented in

Hy-Phy (Pond and Frost 2005; Pond et al. 2006) were used to test for recombination. Tajima's *D* (Tajima 1989) and Fu and Li's *F* statistic (Fu and Li 1993) were calculated in DnaSP (Rozas et al. 2003) for each locus to assess neutrality. DnaSP was also used to calculate the number of segregating sites and nucleotide diversity for each locus.

Species Discovery

Bayesian clustering, as implemented in Structurama (Huelsenbeck et al. 2011) aims to discover cryptic diversity without *a priori* assignment of samples to putative groups. We chose this approach because a recent simulation study suggests that it is the most effective of several multilocus discovery approaches aimed at detecting cryptic diversity (Rittmeyer and Austin 2012). Structurama analyses were conducted

TABLE 2. Loci used for analyses

Locus	Source	Forward primer	Reverse primer	Bp	DT-ModSel	r ^a	Tajima's D	Fu and Li's F
RPL12	T. Devitt	5' ATTCCACTGGGCTATTGAT-	-CCCAAGTTTGACCCTACAGAGAT 3'	453	K80	92	1.349 ns	1.588 ns
HSPA8	T. Devitt	5' ATTCAGGATACCGTTAGCATCAATGT-	-TGCCAAAGCTAGATAAAATTCAGATCC 3'	492	K80+I	66	-0.239 ns	1.109 ns
CYT B	Carstens et al. (2004)	5' GAACATAATGGCCACACWWTACGNA-	-AGGAGTGAAGTAGAGTAAAGTA 3'	662	HKY+I+G	57	0.649 ns	-0.042 ns
GAPD	Dolman and Phillips (2004)	5' ACCTTTATGGGGTGTGGCATTGC-	-CATCAAGTCCACAACACGGTTGCTGTA 3'	653	K80+I	94	0.327 ns	0.710 ns
RAG1	Wiens et al. (2006)	5' AGYCARTAYCAAYARAATGTA-	-GTGGTGCTTCAGAACATCCCTC 3'	1210	K80+I	108	0.834 ns	1.584 ns
		5' AGAACCTGGAGCGCTATGAGATGTGGCG-	-TTCTTCCTCAAGTCTTGTCG 3'	3470	total			
PL1	This study	5' TACCACAAGGGGAGGACTTC-	-CCCCAGATCTTTTCCCATT 3'	215	JC	28 ^b	1.032 ns	0.828 ns
PL14	This study	5' GAATAGCGCCAATCCTGGTA-	-CCCCCTGTAGAAATCCCATT 3'	468	F81	16 ^b	-0.033 ns	0.374 ns
				464		34 ^c	0.773 ns	0.731 ns
PL96	This study	5' AGTGGTGGTTCGCTTTCAC-	-CCTCGTTCAGCCCAATCAICT 3'	271	F81	22 ^b	1.471 ns	0.984 ns
ITS	Hillis and Dixon (1991)	5' GAGGGTCGCTTGAACATCAAT-	-TGATCTGAGGTGCTAGTCGAGA 3'	254	JC+I	34 ^c	1.246 ns	1.066 ns
		5' GAGTGCAGCACCTCAAGGAC-	-GTCGTAAACAAGGTTCCCGTAGG 3'	884	F81+I	26 ^b	0.443 ns	1.031 ns

Note: Loci in bold are aligned across all three species; ns, not significant.

^aNumber of alleles.

^b*Plethodon idahoensis*.

^c*Plethodon larselli*.

by treating the level of clustering (K) as a random variable using the Dirichlet process gamma priors (0.1, 1) and (1, 10) to explore various clustering permutations; 10,000,000 generations sampling every 1000 steps with a burn-in of 10% were run on all available loci for each species.

Species Validation

In contrast to Structurama, both *spedeSTEM* (Ence and Carstens 2011) and BPP (Yang and Rannala 2010) model the divergence of lineages using the multispecies coalescent and require the assignment of samples to putative lineages. These methods allow for gene-tree heterogeneity across loci because the putative lineages are treated as OTUs (rather than exemplar samples as in traditional phylogenetics [Liu et al. 2009]), and therefore allow cryptic lineages to be detected early in the diversification process (Carstens and Dewey 2010). However, due to the complexity of parameter space, each method uses a different simplification to account for uncertainty in phylogenetic parameter space. *SpedeSTEM* takes gene trees as input, assumes that they are estimated without error, and calculates the species phylogeny while computing the probability of a particular model of lineage composition given the data. Alternatively, BPP uses a Bayesian approach to integrate over the uncertainty in gene tree space, but does not attempt to estimate the species tree. Rather, BPP assesses species limits by collapsing nodes on a guide tree, which is assumed by the analysis to be the true species tree. Although the phylogenetic uncertainty in the species tree is not accounted for by BPP, some users (e.g., Leaché and Fujita 2010) have circumvented this issue by using multiple guide trees. We consider these approaches as complementary because each considers uncertainty in one of the two relevant regions of parameter space, and so we apply both to our data.

SpedeSTEM, a species delimitation extension of *STEMv2.0* (Kubatko et al. 2009) was used to analyze data from *P. vandykei*, *P. larselli*, and *P. idahoensis*. For each analysis, we removed redundant alleles within populations in order to standardize the number of alleles across loci at four per putative lineage. A maximum-likelihood search in PAUP* (Swofford 2002) was used to estimate the ML gene tree for all loci (see Table 2 for models of sequence evolution estimated using DT-ModSel [Minin et al. 2003] and Supplementary Material, S3 available on Dryad at <http://dx.doi.org/10.5061/dryad.55137>), which were then saved with midpoint rooting and branch lengths optimized under the molecular clock for input into *STEM*. Theta ($\theta=4N_e\mu$) was estimated using *Migrate-n* v3.3.2 (Beerli 2009) for each locus treating each species as a single population. *SpedeSTEM* analyses were conducted using version 2.0 written in Python and available at: <http://carstenslab.org.ohio-state.edu/software.html>. *STEM* produces an analytical calculation of the probability of a particular model of lineage composition given the data, so the

resulting species tree is that which maximizes the likelihood given the gene trees. Once these probabilities are calculated for all models of lineage composition, information theory (Anderson 2008) is used to calculate the model probabilities of each species tree.

BPP was used also to analyze each species. Ambiguous sites were removed and $\theta=4N_e\mu$ was allowed to vary among loci. Six sets of gamma priors were run independently ranging from small to large effective population size, and shallow to deep divergence: (2, 1000), (2, 1000); (2, 100), (2, 1000); (2, 100), (2, 100); (2, 100), (2, 10); (2, 10), (2, 100); and (2, 10), (2, 10) were used for the population size parameters (θ) and divergence time (τ_0), respectively; the Dirichlet process is used to assign the non-ancestral divergence time priors (Yang and Rannala 2010). All possible guide trees within each species (Fig. 2a–c) were tested and each analysis was run with two different starting trees and both algorithms implemented in BPP to confirm convergence, totaling 168 independent runs. Runs were conducted with all available loci for each species. For all runs, the first 10% of samples were discarded as burn-in and then 100,000 samples retained in the posterior with a sampling frequency of 5.

Species Tree Estimation

Bayesian species trees were estimated using *BEASTv1.7.1 (Drummond and Rambaut 2007) for the three species using *P. vehiculum*, *P. dumni*, and *P. elongatus* as outgroups (three loci), and then all eight putative lineages based on geography (Table 1) within the three species (five loci). After trial runs to assess prior settings, a strict clock was used for the anonymous loci C31C32, C109C110, and GAPD, and a relaxed clock was used for *Cytb* and *RAG1*; two independent runs of 2×10^8 generations were conducted with a burn-in of 10%. Besides models of sequence evolution and using a piecewise constant model, default settings from BEAUTI v1.7.1 were used. Convergence and ESS values were confirmed using Tracer v1.5 (Rambaut and Drummond 2007) and runs were combined (LogCombiner v1.7.1) to estimate the maximum clade credibility tree (TreeAnnotator v1.7.1).

Species Distribution Modeling

To evaluate the contemporary niches and historic distributions, we created a series of niche models based on georeferenced sampling localities for each species in conjunction with global environmental data. GPS coordinates were obtained from the sampling localities of this study, the HerpNet database (<http://herpnet.org>), and Garvey-Darda P.A. (SW personal communication), totaling 142 *P. vandykei*, 52 *P. larselli*, and 329 *P. idahoensis* GPS points. After removal of localities within the same 1 km grid cell to prevent pseudo-replication, a total of 76 *P. vandykei*, 42 *P. larselli*, and 73 *P. idahoensis* GPS points were used for

niche modeling (Supplementary Material, S2 available on Dryad at <http://dx.doi.org/10.5061/dryad.55137>). Good predictive power is generally possible with ≥ 15 points (Papes and Gaubert 2007).

Global environmental data used for constructing niche models for current distributions included both climatic and vegetation data. For contemporary SDMs, a total of 19 bioclimatic variables are available at 1 km resolution (Bio1-19) in addition to four vegetation variables (NDVI, NDVISTD, QSCAT, and TREE) and elevation (SRTM) (see Supplementary Material, S2 available on Dryad at <http://dx.doi.org/10.5061/dryad.55137>, for descriptions and citations for each variable). For historical distribution predictions, a total of 14 bioclimatic variables are available at 1 km resolution (Supplementary Material, S2 available on Dryad at <http://dx.doi.org/10.5061/dryad.55137>). Since only a subset of the variables is available for the historic predictions, we generated two sets of contemporary SDMs: one with the full set of data and a second using only the subset of data available for historic predictions for direct comparison. The historical model represents hypothesized species distributions at approximately 21 kya, the LGM. ArcMap v10.1 (ESRI 2011) was used for data layer manipulation. Niche models were created using MAXENTv3.3.3 (Phillips et al. 2006). A total of 80% of the localities for each species were used to predict the model and 20% of the localities were used in a jackknife analysis to test the models. The models were evaluated using area under the curve (AUC) scores. Each niche model was estimated with both the full set of variables available and with pruned data sets that excluded highly correlated variables ($r \leq -0.9$ and ≥ 0.9). All SDMs were predicted in the gray area shown on the map (i.e., the maps were clipped and are demonstrated by the area shown on the maps in gray). In *P. idahoensis*, the map was clipped further to prevent over-prediction (see Supplementary Material, S5 available on Dryad at <http://dx.doi.org/10.5061/dryad.55137>).

Additionally, current and historical niche models for the putative lineages based on geography in this study were compared within each nominal species. We did this to look for evidence of ecological differentiation among any of the potential lineages. This includes three niches in *P. vandykei*, three niches in *P. larselli*, and two niches (see below) in *P. idahoensis*. These models are hereafter referred to as regional distribution models (RDMs).

Based on results from the genetic data, we further tested for niche divergence between the two niches of *P. idahoensis* using a multivariate niche analysis (McCormack et al. 2010). This method first directly compares differences in environmental variation between lineages and then compares the extent of these differences to a null distribution of differences in the background environment available to each lineage. The method proceeds by describing environmental variation using a principal components analysis to generate multiple niche axes for each of the georeferenced *P. idahoensis* localities and separately for randomly chosen background locations within the distribution of

each lineage. For each retained principal components axis, the average difference is calculated between the northern and southern lineages, a *t*-test is conducted to test for significance, and finally the differences are evaluated against jackknife comparisons from the background data. Niche differences greater than the distribution of background differences suggest significant divergence, whereas niche differences less than the distribution of background differences suggest significant conservatism. Niche differences that fall within the range of generated background differences fail to reject the null hypotheses of divergence or conservatism.

Statistical analyses were conducted in Stata v11 (StataCorp 2009). All PC axes with eigenvalues greater than 1 were retained. Significance in niche differences were calculated using two tailed *t*-tests and the Bonferroni correction was applied to control for multiple comparisons. Null background distributions were created using a random sample of 25% of the background points with 1000 jackknife repetitions. Significance in niche differences relative to the null background distribution was evaluated by determining whether the observed differences fell outside the 98% confidence intervals of the null distribution.

RESULTS

Collection of Genetic Data

We obtained sequence data from five loci in 20 *P. vandykei* individuals, seven loci in 19 *P. larselli* individuals, and nine loci in 21 *P. idahoensis* individuals (Table 2). Recombination was not detected in any locus using both the SBP and GARD methods. Tajima's *D* and Fu and Li's *F* values were not significant for all loci. BLAST found no significant similarity for the three anonymous loci. Summary statistics are presented in Table 3.

Species Discovery

The results from Structurama were similar using both sets of priors (Supplementary Material, S4 available on Dryad at <http://dx.doi.org/10.5061/dryad.55137>) so we report results using just one set of priors (0.1, 1) for simplicity (Table 4). In each of the three species, the posterior probability is highest for two populations, with moderate support for three, and little to no support for a single population or more than three. However, within *P. vandykei* and *P. larselli*, these populations do not correspond to geographic regions (see Supplementary Material, S4 available on Dryad at <http://dx.doi.org/10.5061/dryad.55137> for population assignment plots). In *P. vandykei*, the probability of all individuals from the Cascades belonging to the same population is 0.87, but lower for the Willapa Hills ($P=0.75$) and Olympics ($P=0.49$) regions. If the two coastal regions are pooled (the Olympics and Willapa Hills), the probability is 0.46. In *P. larselli*, the

TABLE 3. Summary statistics for each locus

	ss	ss/1000 bp	π
<i>P. vandykei</i> loci			
Cytb	12	18.1	0.0072
RPL12	6	13.4	0.0017
HSPA8	2	4.1	0.0020
GAPD	6	10.2	0.0035
RAG1	5	4.0	0.0009
<i>P. larselli</i> loci			
Cytb	26	39.3	0.0155
RPL12	7	15.7	0.0063
HSPA8	8	16.3	0.0048
GAPD	17	28.2	0.0119
RAG1	4	3.2	0.0006
PL14	1	2.2	0.0008
PL96	2	7.9	0.0031
<i>P. idahoensis</i> loci			
Cytb	43	65.0	0.0106
RPL12	4	8.8	0.0019
HSPA8	8	16.3	0.0052
GAPD	9	15.4	0.0030
RAG1	13	10.4	0.0019
PL14	5	10.7	0.0032
PL96	1	3.7	0.0019
ITS	13	14.7	0.0046
PL1	1	4.7	0.0012

Note: ss, segregating sites, bp, base pairs, and π , nucleotide diversity.

probability that northern Washington samples belong to the same population is high ($P=0.99$), similarly high for the Oregon samples ($P=0.93$), but low for the southern Washington samples ($P=0.47$). When the two WA regions or the southern Washington and Oregon regions are pooled together the probability that individuals are from the same population is only 0.5 or 0.46, respectively. Taken together, the assumption of a strict correlation between the current boundaries of geographic distributions and genetic populations is not strongly supported. Notably however, in *P. idahoensis* the optimal clustering level is 2, and this largely corresponds to a division of samples into a northern ($P=0.93$) and southern ($P=0.91$) set of river drainages in the northern Rocky Mountains. Only in *P. idahoensis* do the clusters assigned by Structurama correspond with high support to discrete geographic regions (see Supplementary Material, S4 available on Dryad at <http://dx.doi.org/10.5061/dryad.55137>, for population assignment plot) so we use this partition for species validation (below), following Leaché and Fujita (2010).

Species Validation

SpedeSTEM analyses do not support cryptic lineages in any species (Table 5). In each species, the model that is consistent with the current taxonomy (i.e., that all populations are constituents of the same evolutionary lineage) has the highest support. All runs from the BPP analyses had ESS values well above 200 suggesting the Markov chain is sampling from a stationary region of the posterior distribution, but results were not always consistent across runs. The mean probability for

TABLE 4. Species delimitation results: Structurama and BPP

Structurama	k	P
<i>P. vandykei</i>	1	0.03
	2	0.73
	3	0.20
<i>P. larselli</i>	1	0
	2	0.68
	3	0.24
<i>P. idahoensis</i>	1	0
	2	0.67
	3	0.25
BPP	Number of lineages	Mean P
<i>P. vandykei</i>	1	0.36
	2	0.08
	3	0.57
<i>P. larselli</i>	1	0.40
	2	0.01
	3	0.59
<i>P. idahoensis</i>	1	0.04
	2	0.96

Notes: Structurama results show the probability of $k=1$ to $k=3$ lineages for each species; P =posterior probability. BPP results are the mean probability of 1–3 lineages in each species; mean P is the mean probability of the species tree with a certain number of lineages using several sets of priors, algorithms, starting trees, and guide trees. See Supplementary Material, S4 available on Dryad at <http://dx.doi.org/10.5061/dryad.55137> for additional details.

TABLE 5. Species delimitation results: spedeSTEM

Species tree	Lineages	AIC	w_i
(PlaWN+PlaWS+PlaOR, (PidN+PidS, PvaC+PvaO+PvaW))	3	1444.20	1
(PlaWN+PlaWS+PlaOR, (PvaC+PvaO+PvaW, (PidN, PidS)))	4	4909.38	0
((PlaOR, PlaWN+PlaWS), (PidN+PidS, PvaC+PvaO+PvaW))	4	4920.97	0
(PlaWN+PlaWS+PlaOR, (PidN+PidS, (PvaC, PvaO+PvaW)))	4	5326.52	0
((PlaWN+PlaOR, PlaWS), (PidN+PidS, PvaC+PvaO+PvaW))	4	5582.67	0

Notes: List of the top possible species trees for *P. vandykei*, *P. larselli*, and *P. idahoensis*; w_i is the model probability calculated from the likelihood score of each species tree. See Supplementary Material, S4 available on Dryad at <http://dx.doi.org/10.5061/dryad.55137> for all possible species trees AIC values.

splitting *P. vandykei* and *P. larselli* into three lineages was similar (0.57 and 0.59, respectively; Table 4). The mean probability supporting two lineages in *P. idahoensis* is 0.96 (Table 4). Although these analyses do not seem to be heavily influenced by the guide tree (Supplementary Material, S4 available on Dryad at <http://dx.doi.org/10.5061/dryad.55137>), the prior distribution appears to have some influence. Results most often indicate a single lineage when the priors for θ are large. Conversely, lineages are most often split when the priors for θ are small. For example, in *P. larselli*, the mean probability of splitting all lineages under priors (2, 10) (2, 10) is 0.26 (median=0) whereas under priors

(2, 100) (2, 100) it is 0.78 (median = 0.995) (Mann-Whitney U $P = 0.00046$).

Species Tree Estimation

The *BEAST analysis using *P. vehiculum*, *P. dunni*, and *P. elongatus* as outgroups (Supplementary Material, S3 available on Dryad at <http://dx.doi.org/10.5061/dryad.55137>) is consistent with previous *Plethodon* phylogenetic investigations (Wiens et al. 2006; Kozak et al. 2009). Divergence is deep between most groups relative to the potential lineages within *P. vandykei*, *P. larselli*, and *P. idahoensis*. *Plethodon vandykei* and *P. idahoensis* are sister taxa with *P. larselli* sister to this pair. The topology is similar when nominal species are divided into subsets on the basis of geographic location (topology in Fig. 2a): within *P. vandykei*, the Olympic population and Cascade population are more closely related to each other than the Willapa Hills population, and within *P. larselli*, the two WA populations are more closely related to one another than to the OR population.

Species Distribution Models

In order to directly compare the historical and current niche models, the SDMs discussed below are based on the reduced set of uncorrelated variables only available at 21 Ka; however, the SDMs built using the full sets of environmental variables are similar to those discussed below (Supplementary Material, S2 and S5 available on Dryad at <http://dx.doi.org/10.5061/dryad.55137>). Environmental variable contributions for all models can be found in Supplementary Material, S5 available on Dryad at <http://dx.doi.org/10.5061/dryad.55137>.

The historical SDM (21 Ka) for *P. vandykei* predicts a single area of high suitability in the valley between the Coastal range and Cascade Mountains and somewhat suitable habitat extended as far south as northern California. In contrast, the current SDM (AUC = 0.987) predicts its distribution throughout the Coastal range and foothills of the Cascades; the current three disjunct groups are nested with the current SDM and do not cross the Columbia River as do the predicted SDMs (Fig. 3a). A similar pattern is found in *P. larselli* where the current SDM (AUC = 0.990; Fig. 3b) predicts a slightly larger niche than is their known distribution. The current known distribution and SDM of this species is greatly reduced compared with the historical model (Fig. 3b). The historical model of *P. idahoensis* (Fig. 3c) is shifted to the southwest and there is indication of only slightly suitable habitat in its current predicted niche (AUC = 0.975; Fig. 3c), which matches its known distribution, except for patchiness following current streambeds.

There is extensive overlap among the RDMs of both *P. vandykei* and *P. larselli* (Fig. 4). This is evident for the models from the Cascades (AUC = 0.993) and Olympics (AUC = 0.979) in *P. vandykei*, whereas the Willapa Hills (AUC = 0.997) region is the most restrictive.

The Washington south (AUC = 0.992) population of *P. larselli* predicts the widest area with relatively even probability, whereas the Washington north (AUC = 0.993) and Oregon (AUC = 0.980) predictions are smaller. The historical RDMs for all the potential lineages in both species show slight shifts in distribution but again a lot of overlap among all potential lineages (Supplementary Material, S5 available on Dryad at <http://dx.doi.org/10.5061/dryad.55137>).

In *P. idahoensis*, there is little overlap between the two current northern (AUC = 0.972) and southern (AUC = 0.996) RDMs (Fig. 5). The northern RDM slightly projects into the southern range but the southern RDM does not project into the north. Intriguingly, the historical RDMs are very different from the current distribution of *P. idahoensis* (Fig. 5). Both the northern and southern historical RDMs do not show any areas with high probability. Within the current range there are some very faint areas of suitability and two more large solid areas to the southwest. The RDMs using the full set of variables are also similar (Supplementary Material, S5 available on Dryad at <http://dx.doi.org/10.5061/dryad.55137>).

The multivariate niche analyses resulted in four PC axes with eigenvalues greater than 1, accounting for a total of 80% of environmental variation (Table 6). The first three axes were significantly different between lineages (PC1: $u = 2.97$, $P < 0.0001$; PC2: $u = 2.31$, $P < 0.0001$; PC3: $u = 1.77$, $P = 0.0001$), whereas the fourth axis showed no significant differences between lineages (PC4: $u = 0.15$, $P = 0.72$). Moreover, PC axes 1 and 3 were significantly more divergent compared with the null distribution generated by random background points ($P < 0.01$), indicating significant evidence for niche divergence of the northern and southern *P. idahoensis* lineages across these two axes. The highest loadings for these axes are maximum temperature of the warmest month, precipitation of the driest quarter, temperature seasonality, and greenness seasonality. PC axis 2 was significantly less divergent than the null distribution generated by random background points ($P < 0.01$), indicating significant evidence for niche conservatism across this axis. The highest loading for this axis was mean diurnal temperature range.

DISCUSSION

Species delimitation investigations are dependent on an understanding of the geographic range of the focal taxon. In particular, disjunct distributions within a species motivate many molecular investigations because this allopatry provides a way to partition samples to test species limits that are external to the genetic data (e.g., Leaché and Fujita 2010; Burbrink et al. 2011; Camargo et al. 2012). However, the distributions of species are not stable through time and change in part because they are dependent on climatic conditions that are themselves dynamic (Brown et al. 1996). In temperate regions heavily impacted by glaciation, such as the PNW, this dynamism in the geographic distribution is pronounced. Consequently, species delimitation investigations that

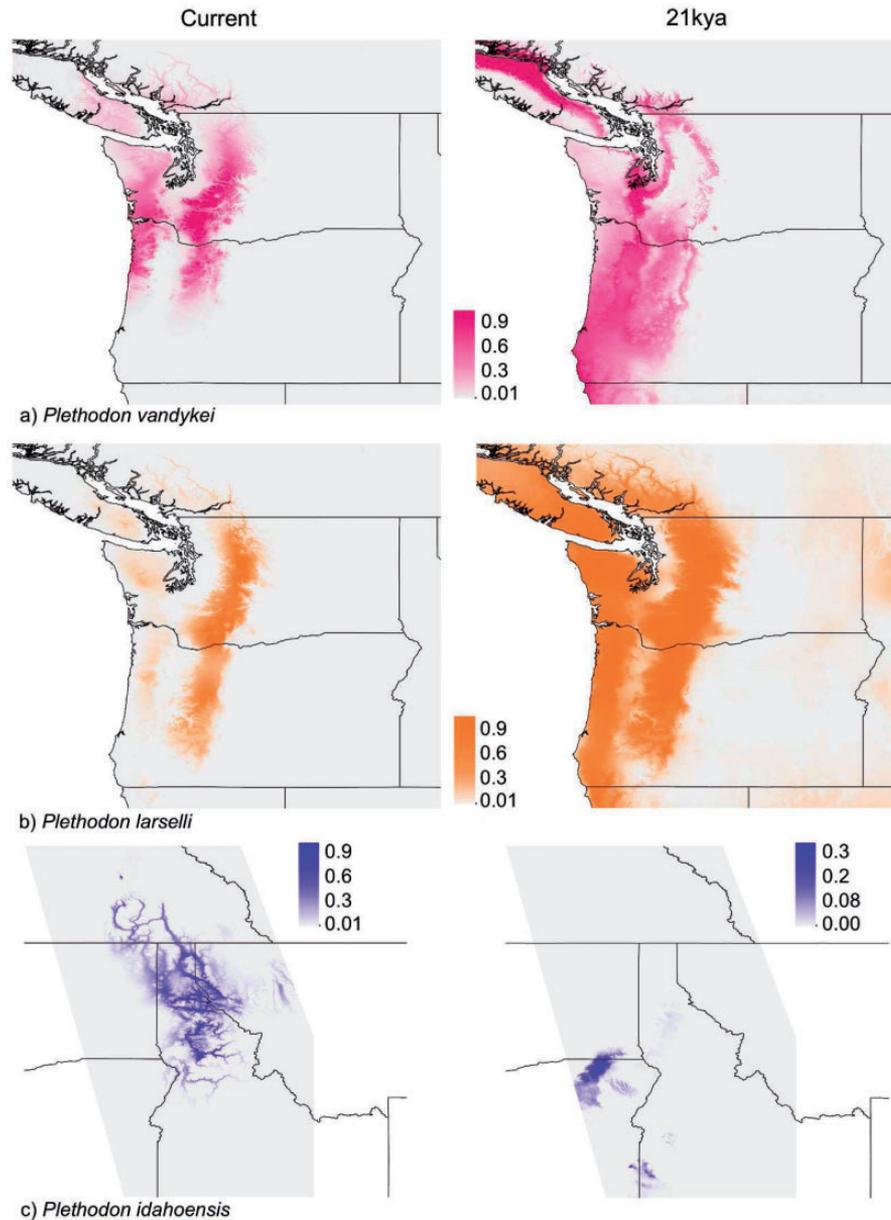


FIGURE 3. Current and historical (21 kya) SDMs. These models are based on the reduced data set with correlated environmental variables removed (Supplementary Material, S2 available on Dryad at <http://dx.doi.org/10.5061/dryad.55137>). (a) *Plethodon vandykei*. (b) *Plethodon larselli*. (c) *Plethodon idahoensis*—this model was also constructed on a reduced geographic area to prevent overprediction into the far southwest (Supplementary Material, S5 available on Dryad at <http://dx.doi.org/10.5061/dryad.55137>). Note the difference in probability score between the current and historical models for *P. idahoensis*.

incorporate an estimate of the temporal range shifts have great potential. Our comparative exploration into species boundaries includes examples of three different types of ranges (disjunct, patchy, and continuous). We discuss each of these in turn before integrating the main findings into a broader context.

Plethodon vandykei

The distribution of *P. vandykei* exemplifies the types of geographic distributions that often motivate molecular

species delimitation investigations. Populations of this species are currently isolated into three allopatric regions that are separated by approximately 170 km. Due to the apparent lack of dispersal ability in plethodontid salamanders, it was easy to speculate prior to this work that *P. vandykei* might contain cryptic evolutionary lineages, however, results were opposite our expectations. The Olympic Peninsula, Willapa Hills, and Cascades do not contain cryptic lineages in this species, according to the results of Structurama, BPP, and SpedeSTEM (Fig. 6). Given the current range,

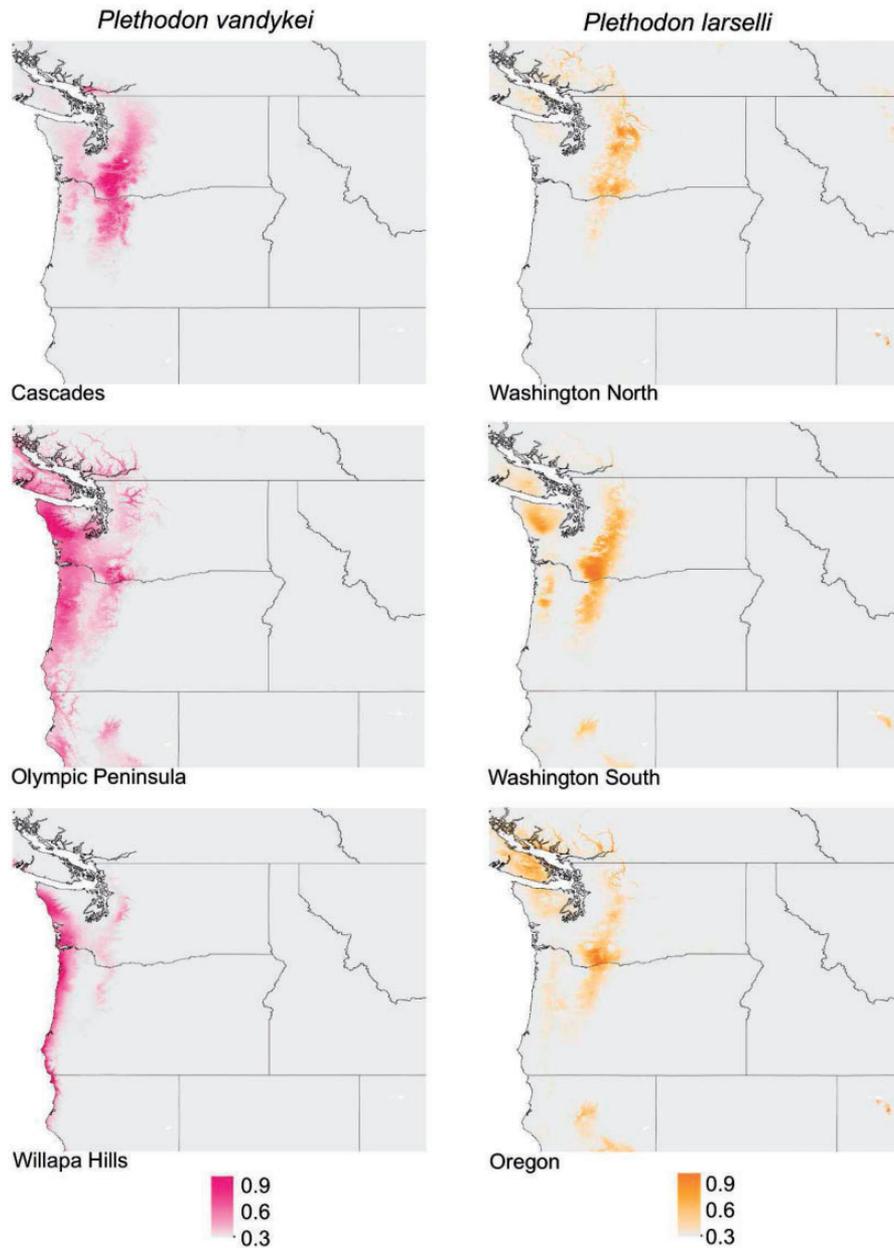


FIGURE 4. Current RDMs for *P. vandykei* and *P. larselli*. These models are based on the data set with correlated environmental variables removed (Supplementary Material, S2 available on Dryad at <http://dx.doi.org/10.5061/dryad.55137>).

these results could be explained by speculating that these regions have not been isolated for time sufficient to remove shared genetic polymorphism from the populations, or by the possibility that populations are currently exchanging alleles. Gene flow is unlikely due to the limited dispersal capabilities of *Plethodon* salamanders and the potential barriers to gene flow (e.g., valleys, rivers, and mountains) intermediate to the three disjunct regions. Furthermore, the reconstruction of the historical range of *P. vandykei* (Fig. 3a) demonstrates that there were large amounts of suitable habitat intermediate to the regions currently occupied by the

species at the close of the Pleistocene and suggests that the current populations are derived from a common ancestral population. This model is supported by the ecological similarity of the three regions (as evidenced by the RDMs), as well as the presence of suitable habitat intermediate to the current disjunct populations of *P. vandykei*. This ecological similarity suggests that additional factors (i.e., microhabitat features, dispersal limitations, and/or competition from other *Plethodon* salamanders present in these regions) prevent *P. vandykei* from expanding its range and bringing the three allopatric regions back into contact.

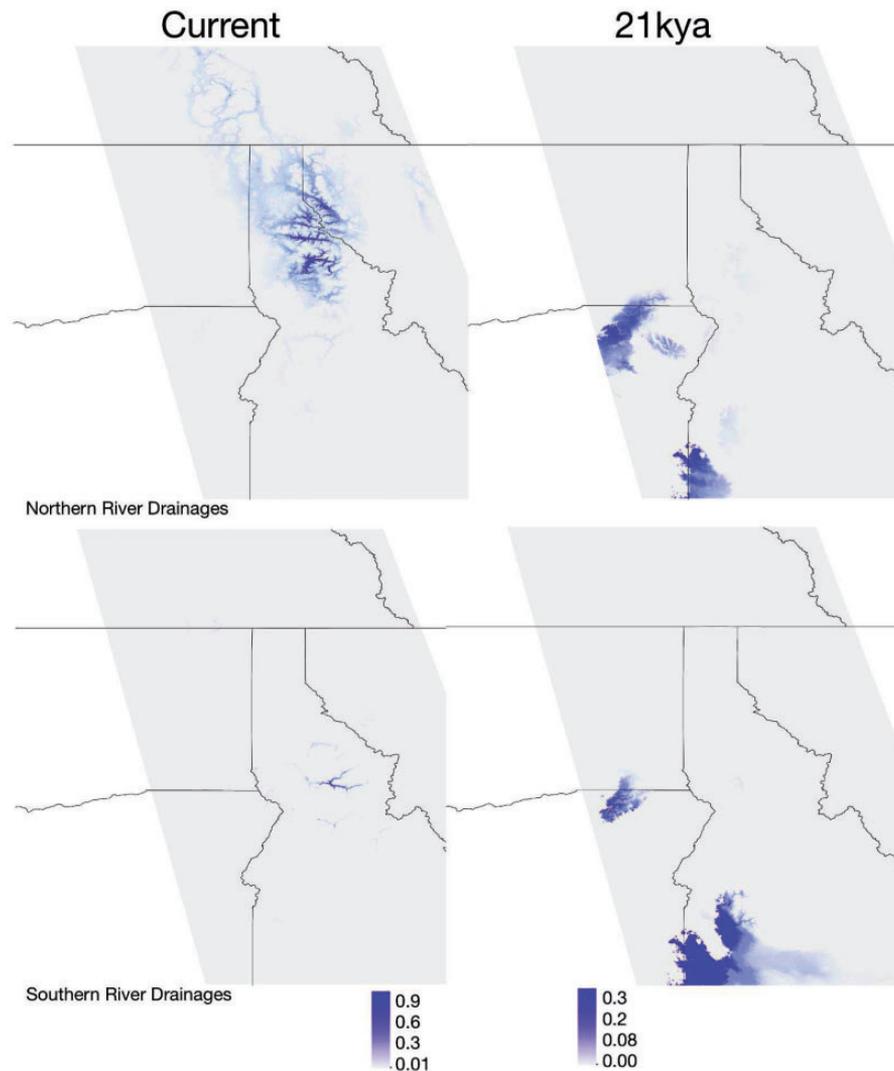


FIGURE 5. Current and historical (21 kya) RDMs for *P. idahoensis*. These models are based on the reduced data set with correlated environmental variables removed (Supplementary Material, S2 available on Dryad at <http://dx.doi.org/10.5061/dryad.55137>) and constructed on a reduced geographic area to prevent overprediction into the far southwest (Supplementary Material, S5 available on Dryad at <http://dx.doi.org/10.5061/dryad.55137>). Note the difference in probability score between the current and historical models for *P. idahoensis*.

Plethodon larselli

Like *P. vandykei*, the current distribution of *P. larselli* also includes populations that are apparently isolated from each other. *Plethodon larselli* occurs on either side of the Columbia River and in Washington is separated by high elevation mountains with a patchy distribution; this salamander has specific microhabitat requirements of forested rocky talus slopes and are only found early in the season during snow melt, restricting them to very cool moist areas (Aubry et al. 1987; TAP personal observation). The estimate of the current range appears to contain a much smaller habitable area than the LGM but larger than that observed in the field (TAP personal observation), whereas the historical range indicates that there was a great deal of suitable habitat throughout the Cascade Range during the LGM (Fig. 3b). Taken

together, these results also indicate that the current patchy distribution of *P. larselli* is a recent development, and probably much smaller than its historical range. Although ongoing gene flow may prevent genetic structure from forming within the Washington distribution of this species, results from a Migrate-n analysis (Supplementary Material, S4 available on Dryad at <http://dx.doi.org/10.5061/dryad.55137>) suggest that alleles are being exchanged across the broad Columbia River. Planned investigations into congeners *P. dunni* and *P. vehiculum* will more rigorously explore the potential of this river to inhibit gene flow in western *Plethodon*. In the meantime, our results support the assessment by the Department of Wildlife in Washington State that this species is threatened largely due to its strict habitat requirements (<http://wdfw.wa.gov/publications/01519>).

TABLE 6. Environmental variation between northern and southern lineages of *P. idahoensis*

Variable	Description	PC1	PC2	PC3	PC4
Bio 2	Mean diurnal temperature range	(0.38) 0.1698	(0.22) 0.4239	(0.11) 0.2453	(0.09) -0.1776
Bio 4	Temperature seasonality	0.0939	-0.0969	-0.4651	0.1898
Bio 5	Max temperature of warmest month	0.3964	0.0159	0.0488	-0.0038
Bio 6	Min temperature of coldest month	0.3279	-0.276	-0.0445	0.0828
Bio 7	Temperature annual range	0.144	0.4406	0.1454	-0.1306
Bio 8	Mean temperature of wettest quarter	0.3269	-0.0281	-0.1569	0.0376
Bio 9	Mean temperature of driest quarter	0.368	-0.1365	-0.0056	0.0862
Bio 14	Precipitation of driest month	-0.3255	-0.0136	-0.3685	0.2209
Bio 15	Precipitation seasonality	0.0154	-0.4524	0.3241	0.0168
Bio 17	Precipitation of driest quarter	-0.3646	-0.0068	-0.1112	0.2042
Bio 19	Precipitation of coldest quarter	-0.1734	-0.3507	0.374	0.1089
NDVI	Normalized Difference Vegetation Index (greenness)	0.107	0.1478	0.2263	0.5812
NDVISTD	Greenness seasonality (yearly standard deviation)	-0.1605	0.0776	0.3992	0.1183
QSCAT	Canopy or surface moisture and roughness	0.0307	0.3303	-0.1934	0.2208
SRTM	Elevation	-0.3507	0.1894	0.1	-0.1006
TREE	Percentage of tree cover	0.0663	0.1419	0.1496	0.6216
	Mean niche difference	2.97 * D	2.31 * C	1.77 * D	0.15
	Mean background difference	1.52	3.33	0.08	0.14
	98% CI background difference	(0.51–2.49)	(2.99–3.69)	(0.00–0.71)	(0.00–0.66)

Notes: For each of the four retained principal components axes, the loadings are shown for each environmental variable in addition to the proportion of variance explained (in parentheses). Mean niche differences for both the actual localities and random background localities are shown at the bottom of the table. Axes with significant divergence between lineages based on *t*-tests after Bonferroni correction are indicated by an asterisk. Axes that remain significantly divergent after comparison with the background distribution are labeled with a *D*, whereas axes that show significant conservatism relative to the background are labeled with a *C*.

Plethodon idahoensis

The current range of *P. idahoensis* is strikingly different from that of its close relatives in western Washington in that it is large and continuous throughout the northern Rocky Mountains. Such a range does not predict cryptic diversity, and this species was initially included in this study to provide a counter example to the ranges of the other taxa. However, much of its current range is north of the maximal extent of glaciation during the LGM, and the historical demographic model that has developed via phylogeographic work on this species is one of post glacial expansion from multiple glacial refugia in riverine canyons located to the south of the Pleistocene glaciers (Carstens et al. 2004; Carstens and Richards 2007). In contrast to *P. vandykei* and *P. larselli*, we find genetic evidence for cryptic independent lineages within *P. idahoensis*. First, the clustering level with the highest posterior probability from the Structurama analysis is $k=2$, and the division of samples into clusters largely corresponds to those sampled from northern and southern river drainages ($P=0.91, 0.93$). Although results from spedeSTEM do not support these regions as belonging to separate evolutionary lineages, the BPP analysis shows a high posterior probability (1.0) in support of these clusters as separate lineages. This discrepancy could be explained by the shallow divergence between these lineages, as spedeSTEM can fail to detect independent lineages when the loci that the gene trees are estimated from lack a sufficient number of SNPs, or from the fact that there might

be ongoing gene flow, despite divergence. Although BPP can be misled by inaccurate specification of the guide trees (Leaché and Fujita 2010; but see Zhang et al. 2014), this is not a potential problem in *P. idahoensis* because there are only two lineages (and thus one possible topology). The other source of error for BPP would be incorrect sample assignment to the putative lineages. To explore the probability for a false-positive result caused by error in sample assignment, we randomized assignment of *P. idahoensis* individuals and repeated the BPP analysis using priors (2, 10) (2, 10) and (2, 100) (2, 100). Results from 100 replicates indicate that population mis-assignment is not likely to mislead our analysis: when samples are randomized, the proportion of trees splitting *P. idahoensis* is 0.01 (mean PP=0.07, median PP=0) using priors (2, 10) (2, 10) and 0.3 (mean PP=0.34, median PP=0.01) using priors (2, 100) (2, 100). Taken in total, the genetic analyses support the division of *P. idahoensis* into a northern and southern lineage. This was surprising and is important because independent lineages can easily go undetected without the use of molecular approaches. Furthermore, the northern and southern lineages are statistically divergent in two PC niche axes, indicating that ecological niche differentiation appears to have occurred relatively rapidly for some environmental variables. These findings are similar to those of Raxworthy et al. (2007), who investigated geckos from Madagascar and found divergent ecological niches between closely related species. This strengthens our evidence for independent lineages in *P. idahoensis*,

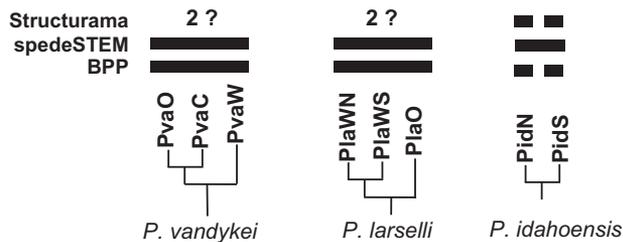


FIGURE 6. Species delimitation results from Structurama, spedeSTEM, and BPP (results plotted using guide tree estimated in *BEAST). Separate blocks represent separate species.

which may have easily been missed without the use of molecular species delimitation methods.

The historical RDMs encompass the current distributions at low probability (Fig. 5), whereas the historical SDM does not contain the current distribution of the population as in *P. larselli* and *P. vandykei* (Fig. 3). This result may suggest that *P. idahoensis* persisted throughout glacial cycles of the Pleistocene outside its current distribution and expansion played an important role in shaping the current genetic diversity, or more likely it was restricted to extremely small patches of habitat that are not easily detected using this approach (e.g., dual refuge model; see Carstens and Richards 2007).

Integration of Molecular and Environmental Data for Species Delimitation

Species delimitation should be conducted using data from a variety of sources (Sites and Marshall 2004; de Queiroz 2007; Knowles and Carstens 2007) and results are most meaningful when there are congruent signals across these data (Fujita et al. 2012). The widespread availability of climatic data, coupled with the clear relevance that estimates of the current and historical ranges of species have toward the question of species boundaries, make the integration of SDMs and molecular methods for species delimitation particularly useful. Estimates of the current and historical distributions are an asset when interpreting the results from genetic investigations, as demonstrated here in three *Plethodon* species with dissimilar distributions. In each case, we found that the historical distribution of the species was a much better predictor of the results of the genetic data than was the current range. Historic ranges may differ dramatically from current ranges (Kirkpatrick and Barton 1997; Tingley and Beissinger 2009), and evidence suggests that climate refugia may have been more common and widespread than previously thought (Hampe et al. 2013). Although our models of the historical range of these species are a simplification of a complex reality (like all models), they offer evidence external to that provided by the genetic diversity within these organisms, and enable a more nuanced interpretation of evolutionary history.

However, the climatic modeling used here does have shortcomings, and these should be acknowledged. First, the PNW has a complex history, and the range dynamics of species from this region will not be fully captured with binary models (i.e., the present and end-Pleistocene). Montane glaciers persisted in the region long after the close of the Pleistocene, and a dry period (the Hypsithermal Interval; Mathews and Heusser 1981) about 10,000–7500 ya likely initiated a major contraction in *Plethodon* ranges, factors not considered by our models. Additionally, as continental and montane glaciers retreated they left behind glacial till, and both the new substrates and potentially protracted plant successional communities would have likely created ephemeral conditions where patterns of gene flow may have been very different. Additional historical climate models may further elucidate range dynamics in this region, and therefore aid in developing hypotheses of species limits.

Gene Flow and Species Delimitation

Methods such as spedeSTEM and BPP incorporate a multispecies coalescent model that does not parameterize gene flow, and this complicates the interpretation of our results because we are forced to make assumptions about what type of analytical model is most appropriate for our data. For example, in species where there is no evidence of cryptic lineages, such as *P. vandykei* and *P. larselli*, an *n*-island model that does not parameterize temporal divergence among populations is likely justified. When we utilize such a model (Migrate-*n*; Beerli and Felsenstein 2001) to estimate gene flow, we find that estimates of gene flow among populations are non-zero in both *P. vandykei* and *P. larselli* (Supplementary Material, S4 available on Dryad at <http://dx.doi.org/10.5061/dryad.55137>). Model selection results from *P. vandykei* indicate that models of migration between the Cascades–Olympics and the Cascades–Willapa Hills ($P=0.53$) or symmetrical migration among all populations ($P=0.43$) are optimal. In *P. larselli*, the model with the highest probability is that of symmetrical migration among all populations ($P=0.98$), indicating that alleles are being exchanged across the Columbia River. These results indicate that ongoing gene flow may be inhibiting diversification in this species.

In *P. idahoensis*, where results from both Structurama and BPP suggest that the species is diverging into northern and southern lineages, the *n*-island model of gene flow implemented in Migrate-*n* is likely not appropriate because it does not parameterize population divergence. When we use it to analyze our data, the model with north to south migration has the highest probability (0.97), but we attribute this result to erroneous signal from shared ancestral polymorphism. Rather than the *n*-island model, a model that estimates both population divergence and migration is required. Carstens et al. (2009) analyzed sequence data (three loci)

from *P. idahoensis* using IMA (Hey and Nielson 2004) and found that the summed model probability of isolation-only models ($w_i = 0.546$) was slightly higher than that of isolation with migration models ($w_i = 0.454$). However, model choice experiments using approximate Bayesian computation (Pelletier and Carstens 2014; with five loci) indicate that the isolation with migration model offers a better fit to the data than isolation only. How do these results influence our interpretation of those from BPP, which does not estimate gene flow? Since gene flow homogenizes allele frequencies across the putative lineages, BPP fails to detect actual independent lineages more often in the presence of gene flow (Camargo et al. 2012). Given that we delimit the northern and southern lineages using BPP, even while there is some evidence of ongoing gene flow, indicates to us that the signal of lineage divergence remains strong even in the presence of some gene flow. Consequently, we are more inclined to accept the results of the BPP analysis and consider the northern and southern lineages of *P. idahoensis* to be evolutionarily distinct.

Conclusions

Given the variety of threats to global biodiversity (Daszak et al. 2000; McKee et al. 2004; Butchart et al. 2010; Vörösmarty et al. 2010) and the difficulties inherent in discovering cryptic lineages (e.g., Hoagland et al. 1995; Beheregaray and Caccone 2007), it is clear that more efficient discovery and characterization of biodiversity is needed (Maddison et al. 2012). The technique described above integrates current and historical SDMs into the species discovery process and can strengthen any investigation that seeks to quickly and accurately identify species boundaries. Notably, the addition of SDMs is accessible, and can be conducted anywhere with a computer without extensive investment in time, computational machinery, or the expense required for the generation of genetic data. The SDMs serve two purposes: they act as an independent line of evidence in support of results from genetic data, but they can also generate species hypotheses that can then be explored in greater detail with additional fieldwork and the molecular approaches essential in documenting biodiversity.

SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.55137>.

FUNDING

This work was supported by the National Science Foundation to BCC (DEB-0918212), an ASIH Gaige Fund Award and AMNH Theodore Roosevelt Grant to TAP, and the U.S. Forest Service, Pacific Northwest Research Station to CMC.

ACKNOWLEDGEMENTS

The authors thank Paul Chippindale, Tom Devitt, and John Wiens for primer sequences, as well as Sarah Hird and Melissa DeBiasse for helpful manuscript comments. Patricia A. Garvey-Darda provided useful information and locality data concerning *P. larselli*. Paul Pelletier, Jason Bazzano, Lindsay Henderson, and Wendy Lee helped with field collections. Many analyses were carried out using the Ohio Biodiversity Conservation Partnership (OBCP) Informatics Infrastructure housed within the C.A. Triplehorn Insect Collection Computing Rack at The Ohio State University with help from Joe Cora. The authors thank Jean-François Flot, Richard Glor, and several anonymous reviewers who provided comments that improved this manuscript.

REFERENCES

- Anderson D.R. 2008. Model based inference in the life sciences: a primer on evidence. New York: Springer-Verlag.
- Aubry K.B., Senger C.M., Crawford R.L. 1987. Discovery of Larch Mountain salamanders *Plethodon larselli* in the central cascade range of Washington. *Biol. Conserv.* 42:147–152.
- Austin M. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecol. Model.* 157:101–118.
- Berli P., Felsenstein J. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proc. Natl Acad. Sci. U. S. A.* 98:4563–4568.
- Berli P. 2009. How to use MIGRATE or why are Markov chain Monte Carlo programs difficult to use. *Popul. Genet. Anim. Conserv.* 17:42–79.
- Beheregaray L.B., Caccone A. 2007. Cryptic biodiversity in a changing world. *J. Biol.* 6:1–5.
- Brown J.H., Stevens G.C., Kaufman D.M. 1996. The geographic range: size, shape, boundaries and internal structures. *Ann. Rev. Ecol. Syst.* 27:597–623.
- Brunsfeld S., Sullivan J., Soltis D., Soltis P. 2001. Comparative phylogeography of northwestern North America: a synthesis. *Spec. Publ. Br. Ecol. Soc.* 14:319–340.
- Burbrink F.T., Yao H., Ingrasci M., Bryson R.W., Guiher T.J., Ruane S. 2011. Speciation at the mogollon rim in the Arizona mountain kingsnake (*Lampropeltis pyromelana*). *Mol. Phylogenet. Evol.* 60: 445–454.
- Butchart S.H., Walpole M., Collen B., van Strien A., Scharlemann J.P., Almond R.E., Baillie, B. Bomhard J.E., Brown C., Bruno J. 2010. Global biodiversity: indicators of recent declines. *Science* 328:1164–1168.
- Camargo A., Morando M., Avila L.J., Sites J.W. 2012. Species delimitation with ABC and other coalescent based methods: a test of accuracy with simulations and an empirical example with lizards of the *Liolaemus darwini* complex (squamata: Liolaemidae). *Evolution* 66:2834–2849.
- Carstens B.C., Brunsfeld S.J., Demboski J.R., Good J.M., Sullivan J. 2005. Investigating the evolutionary history of the Pacific Northwest mesic forest ecosystem: hypothesis testing within a comparative phylogeographic framework. *Evolution* 59: 1639–1652.
- 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., Richards C.L. 2007. Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution* 61: 1439–1454.

- Carstens B.C., Stevenson A.L., Degenhardt J.D., Sullivan J. 2004. Testing nested phylogenetic and phylogeographic hypotheses in the *Plethodon vandykei* species group. *Syst. Biol.* 53:781–792.
- Carstens B.C., Stoute H.N., Reid N.M. 2009. An information-theoretical approach to phylogeography. *Mol. Ecol.* 18:4270–4282.
- Collevatti R.G., Terribile L.C., Lima Ribeiro M.S., Nabout J.C., Oliveira G., Rangel T.F., Rabelo S.G., Diniz Filho J.A. 2012. A coupled phylogeographical and species distribution modelling approach recovers the demographical history of a Neotropical seasonally dry forest tree species. *Mol. Ecol.* 21:5845–5863.
- Coyne J.A., Orr A.H. 2004. *Speciation*. Sunderland (MA): Sinauer Associates.
- Daszak P., Cunningham A.A., Hyatt A.D. 2000. Emerging infectious diseases of wildlife—threats to biodiversity and human health. *Science* 287:443–449.
- Dolman G., Phillips B. 2004. Single copy nuclear DNA markers characterized for comparative phylogeography in Australian wet tropics rainforest skinks. *Mol. Ecol. Notes* 4:185–187.
- de Queiroz K. 2007. Species concepts and species delimitation. *Syst. Biol.* 56:879–886.
- Drummond A., Ashton B., Buxton S., Cheung M., Cooper A., Duran C., Field M., Heled J., Kearse M., Markowitz S. 2011. Geneious v5. 4. Available from <http://www.geneious.com/>.
- Drummond A.J., Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214.
- Ence D.D., Carstens B.C. 2011. SpedeSTEM: a rapid and accurate method for species delimitation. *Mol. Ecol. Resour.* 11:473–480.
- ESRI. 2011. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute.
- Florio A., Ingram C., Rakotondravony H., Louis E., Raxworthy C. 2012. Detecting cryptic speciation in the widespread and morphologically conservative carpet chameleon (*Furcifer lateralis*) of Madagascar. *J. Evol. Biol.* 25:1399–1414.
- Flot J.F. 2007. Champuru 1.0: a computer software for unraveling mixtures of two DNA sequences of unequal lengths. *Mol. Ecol. Notes* 7:974–977.
- Flot J.F. 2010. SeqPHASE: a web tool for interconverting PHASE input/output files and FASTA sequence alignments. *Mol. Ecol. Resour.* 10:162–166.
- Fu Y.-X., Li W.-H. 1993. Statistical tests of neutrality of mutations. *Genetics* 133:693–709.
- Fujita M.K., Leaché A.D., Burbrink F.T., McGuire J.A., Moritz C. 2012. Coalescent-based species delimitation in an integrative taxonomy. *Trends Ecol. Evol.* 9:480–488.
- Guisan A., Thuiller W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* 8:993–1009.
- Hampe A., Rodríguez Sánchez F., Dobrowski S., Hu F.S., Gavin D.G. 2013. Climate refugia: from the Last Glacial Maximum to the twenty-first century. *New Phytol.* 197:16–18.
- Hey J., Nielson R. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* 167:747–760.
- Highton R. 1995. Speciation in eastern North American salamanders of the genus *Plethodon*. *Annu. Rev. Ecol. Syst.* 26:579–600.
- Highton R., Hastings A.P., Palmer C., Watts R., Hass C.A., Culver M., Arnold S.J. 2012. Concurrent speciation in the Eastern Woodland Salamanders (Genus *Plethodon*): DNA sequences of the complete albumin nuclear and partial mitochondrial 12s genes. *Mol. Phylogenet. Evol.* 63:278–290.
- Highton R., Larson A. 1979. The genetic relationships of the salamanders of the genus *Plethodon*. *Syst. Biol.* 28:579–599.
- Hillis D.M., Dixon M.T. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. *Q. Rev. Biol.* 66:411–453.
- Hoagland P., Kaoru Y., Broadus J.M. 1995. A methodological review of net benefit evaluation for marine reserves. Environment Department, World Bank.
- Howard J.H., Seeb L.W., Wallace R. 1993. Genetic variation and population divergence in the *Plethodon vandykei* species group (Caudata: Plethodontidae). *Herpetologica* 49:238–247.
- Huelsenbeck J.P., Andolfatto P., Huelsenbeck E.T. 2011. Structurama: Bayesian inference of population structure. *Evol. Bioinformatics Online* 7:55.
- Hugall A., Moritz C., Moussalli A., Stanisci J. 2002. Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosiphia bellendenkerensis* (Brazier 1875). *Proc. Natl. Acad. Sci. U. S. A.* 99:6112–6117.
- Kirkpatrick M., Barton N.H. 1997. Evolution of a species' range. *Am. Nat.* 150:1–23.
- Knowles L.L., Carstens B.C. 2007. Delimiting species without monophyletic gene trees. *Syst. Biol.* 56:887–895.
- Kozak K.H., Wiens J.J. 2006. Does niche conservatism promote speciation? A case study in North American salamanders. *Evolution* 60:2604–2621.
- Kozak K.H., Mendyk R.W., Wiens J.J. 2009. Can parallel diversification occur in sympatry? Repeated patterns of body size evolution in coexisting clades of North American salamanders. *Evolution* 63:1769–1784.
- Kozak K.H., Weisrock D.W., Larson A. 2006. Rapid lineage accumulation in a non-adaptive radiation: phylogenetic analysis of diversification rates in eastern North American woodland salamanders (Plethodontidae: *Plethodon*). *Proc. R. Soc. Lond. B Biol.* 273:539–546.
- Kubatko L.S., Carstens B.C., Knowles L.L. 2009. STEM: species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics* 25:971–973.
- Leaché A.D., Fujita M.K. 2010. Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proc. R. Soc. Lond. B Biol.* 277:3071–3077.
- Liu L., Yu L., Kubatko L., Pearl D.K., Edwards S.V. 2009. Coalescent methods for estimating phylogenetic trees. *Mol. Phylogenet. Evol.* 53:320.
- Lozier J., Aniello P., Hickerson M. 2009. Predicting the distribution of Sasquatch in western North America: anything goes with ecological niche modelling. *J. Biogeogr.* 36:1623–1627.
- Maddison D.R., Guralnick R., Hill A., Reysenbach A.-L., McDade L.A. 2012. Ramping up biodiversity discovery via online quantum contributions. *Trends Ecol. Evol.* 27:72–77.
- Mathews R.W., Heusser L.E. 1981. A 12,000 year palynological record of temperature and precipitation trends in southwestern British Columbia. *Can. J. Bot.* 59:707–710.
- Mayr E. 1942. *Systematics and the origin of species*, from the viewpoint of a zoologist. Harvard University Press, Cambridge, MA.
- McCormack J.E., Zellmer A.J., Knowles L.L. 2010. Does niche divergence accompany allopatric divergence in *Aphelocoma* jays as predicted under ecological speciation?: insights from tests with models. *Evolution* 123:1244.
- McKee J.K., Sciuilli P.W., Foose C.D., Waite T.A. 2004. Forecasting global biodiversity threats associated with human population growth. *Biol. Conserv.* 115:161–164.
- Minin V., Abdo Z., Joyce P., Sullivan J. 2003. Performance-based selection of likelihood models for phylogeny estimation. *Syst. Biol.* 52:674–683.
- Mizuno S., Macgregor H.C. 1974. Chromosomes, DNA sequences, and evolution in salamanders of the genus *Plethodon*. *Chromosoma* 48:239–296.
- Mueller R.L., Macey J.R., Jaekel M., Wake D.B., Boore J.L. 2004. Morphological homoplasy, life history evolution, and historical biogeography of plethodontid salamanders inferred from complete mitochondrial genomes. *Proc. Natl. Acad. Sci. U. S. A.* 101:13820–13825.
- Otto-Bliesner B., Brady E.C., Clauzet G., Toma R., Levis S., Kothavala Z. 2006. Last glacial maximum and holocene climate in CCSM3. *J. Climate* 19:2526–2544.
- Papes M., Gaubert P. 2007. Modeling ecological niches from low numbers if occurrences: assesment of the conservation status of poorly known viverrids (Mammalia, Carnivora) across two continents. *Divers. Distrib.* 13:890–902.
- Pelletier T.A., Carstens B.C. 2014. Model choice for phylogeographic inference using a large set of models. *Mol. Ecol.* 23:3028–3043.
- Phillips S.J., Anderson R.P., Schapire R.E. 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190:231–259.
- Pond S.L.K., Frost S.D.W. 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21:2531–2533.

- Pond S.L.K., Posada D., Gravenor M.B., Woelk C.H., Frost S.D.W. 2006. GARD: a genetic algorithm for recombination detection. *Bioinformatics* 22:3096–3098.
- Rambaut A., Drummond A. 2007. Tracer v1. 5. Computer program. Available from: URL <http://tree.bio.ed.ac.uk/software/tracer>.
- Raxworthy C.J., Ingram C.M., Rabibisoa N., Pearson R.G. 2007. Applications of ecological niche modeling for species delimitation: a review and empirical evaluation using day geckos (*Phelsuma*) from Madagascar. *Syst. Biol.* 56:907–923.
- Riley S.P.D., Shaffer H.B., Voss S.R., Fitzpatrick B.M. 2003. Hybridization between a rare, native tiger salamander (*Ambystoma californiense*) and its introduced congener. *Ecol. Appl.* 13:1263–1275.
- Rissler L.J., Apodaca J.J. 2007. Adding more ecology into species delimitation: ecological niche models and phylogeography help define cryptic species in the black salamander (*Aneides flavipunctatus*). *Syst. Biol.* 56:924–942.
- Rittmeyer E.N., Austin C.C. 2012. The effects of sampling on delimiting species from multi-locus sequence data. *Mol. Phylogenet. Evol.* 65:451–463.
- Rosell J.A., Olson M.E., Weeks A., De-Nova J.A., Lemos R.M., Camacho J.P., Feria T.P., Gómez-Bermejo R., Montero J.C., Eguiarte L.E. 2010. Diversification in species complexes: tests of species origin and delimitation in the *Bursera simaruba* clade of tropical trees (Burseraceae). *Mol. Phylogenet. Evol.* 57:798–811.
- Rozas J., Sánchez-DelBarrio J.C., Messeguer X., Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497.
- Sessions S.K., Larson A. 1987. Developmental correlates of genome size in plethodontid salamanders and their implications for genome evolution. *Evolution* 41:1239–1251.
- Shepard D.B., Burbrink F.T. 2009. Phylogeographic and demographic effects of Pleistocene climatic fluctuations in a montane salamander, *Plethodon fourchensis*. *Mol. Ecol.* 18:2243–2262.
- Sites J.W., Marshall J.C. 2004. Operational criteria for delimiting species. *Annu. Rev. Ecol. Syst.* 35:199–227.
- Soberon J. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodivers. Inform.* 2:1–10.
- StataCorp. 2009. Stata statistical software: release 11. College Station, TX: StataCorp LP.
- 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.
- Storfer A., Cross J., Rush V., Caruso J. 1999. Adaptive coloration and gene flow as a constraint in the streamside salamander, *Ambystoma barbouri*. *Evolution* 53:889–898.
- Swofford D. 2002. Paup*. Phylogenetic analysis using parsimony (* and other models). Version 4:b10. Illinois Natural History Survey, Champaign, IL.
- Swenson N.G., Howard D.J. 2005. Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *Am. Nat.* 166:581–591.
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
- Tingley M.W., Beissinger S.R. 2009. Detecting range shifts from historical species occurrences: new perspectives on old data. *Trends Ecol. Evol.* 24:625–633.
- Vörösmarty C.J., McIntyre P., Gessner M.O., Dudgeon D., Prusevich A., Green P., Glidden S., Bunn S.E., Sullivan C.A., Liermann C.R. 2010. Global threats to human water security and river biodiversity. *Nature* 467:555–561.
- Wagner R.S., Miller M.P., Crisafulli C.M., Haig S.M. 2005. Geographic variation, genetic structure, and conservation unit designation in the Larch Mountain salamander (*Plethodon larselli*). *Can. J. Zool.* 83:396–406.
- Wake D.B. 1997. Incipient species formation in the salamanders of the *Ensatina* complex. *Proc. Natl. Acad. Sci. U.S.A.* 94:7761–7767.
- Wake D.B. 2009. What salamanders have taught us about evolution. *Annu. Rev. Ecol. Syst.* 40:333–352.
- Weisrock D.W., Rasoloarison R.M., Fiorentino I., Ralison J.M., Goodman S.M., Kappeler P.M., Yoder A.D. 2010. Delimiting species without nuclear monophyly in Madagascar's mouse lemurs. *PLoS One* 5:e9883.
- Wiens J.J., Engstrom T.N., Chippindale P.T. 2006. Rapid diversification, incomplete isolation, and the “speciation clock” in North American salamanders (genus *Plethodon*): testing the hybrid swarm hypothesis of rapid radiation. *Evolution* 60:2585–2603.
- Yang Z., Rannala B. 2010. Bayesian species delimitation using multilocus sequence data. *Proc. Natl. Acad. Sci. U. S. A.* 107:9264–9269.
- Zhang C., Rannala B., Yang Z. 2014. Bayesian species delimitation can be robust to guide-tree inference errors. *Syst. Biol.* 63:993–1004.
- Zhou W.W., Wen Y., Fu J., Xu Y.B., Jin J.Q., Ding L., Min M.S., Che J., Zhang Y.P. 2012. Speciation in the *Rana chensinensis* species complex and its relationship to the uplift of the Qinghai–Tibetan Plateau. *Mol. Ecol.* 21:960–973.