

Testing for the presence of cryptic diversity in tail-dropper slugs (*Prophysaon*) using molecular data

MEGAN L. SMITH^{1*}, MEGAN RUFFLEY^{2,3}, ANDREW M. RANKIN^{2,3}, ANAHÍ ESPÍNDOLA^{2,3}, DAVID C. TANK^{2,3}, JACK SULLIVAN^{2,3} and BRYAN C. CARSTENS¹

¹Department of Evolution, Ecology and Organismal Biology, The Ohio State University, 318 W. 12th Avenue, 300 Aronoff Labs, Columbus, OH 43210-1293, USA

²Department of Biological Sciences, University of Idaho, 875 Perimeter Drive MS 3051, Moscow, ID 83844-3051, USA

³Institute for Bioinformatics and Evolutionary Studies (IBEST), Biological Sciences, University of Idaho, 875 Perimeter Drive MS 3051, Moscow, ID 83844-3051, USA

Received 3 April 2018; revised 3 May 2018; accepted for publication 3 May 2018

The Pacific Northwest of North America contains two disjunct temperate rainforests, one in the Coastal and Cascades Ranges and another in the Northern Rocky Mountains. These rainforests harbour > 200 disjunct and endemic taxa, with coastal and inland populations separated by the Columbia Basin. For several taxa, molecular data have revealed cryptic diversity structured across the Columbia Basin. Here, we use information from previously studied taxa and a machine-learning framework to predict that tail-dropper slugs in the genus *Prophysaon* (*Prophysaon andersoni*, *Prophysaon coeruleum*, *Prophysaon dubium* and the *Prophysaon vanattae*/*Prophysaon humile* complex) should lack cryptic diversity. This prediction is supported by results from species distribution models (SDMs), which suggest that all taxa lacked suitable habitat in the inland rainforests during the Last Glacial Maximum. We collected *COI* data and tested these predictions using approximate Bayesian computation and found that models of recent dispersal between inland and coastal populations received strong support. Finally, we used posterior predictive simulations to show that the best model was a reasonable fit to the data for all taxa. Our study highlights the utility of predictive modelling in a comparative phylogeographical framework and illustrates how posterior assessments of model fit can improve confidence in model-based phylogeographical analysis.

ADDITIONAL KEYWORDS: approximate Bayesian computation – cryptic diversity – phylogeography – posterior predictive simulations.

INTRODUCTION

The Pacific Northwest of North America (PNW) supports two disjunct temperate rain forests, namely the Cascades and Coastal ranges in the west and the Northern Rocky Mountains in the east (Fig. 1). The inland and coastal rainforests were continuous before the orogeny of the Cascades range (2–5 Mya; Graham, 1999), when the elevation of the Cascades produced a rain shadow that led to the xerification of the Columbia Basin. This basin is now characterized by a shrub-steppe ecosystem and effectively separates inland

and coastal rainforests by > 200 km of habitat that is unsuitable for rainforest endemic species. The isolation of these two rainforests is somewhat diminished to the south by the Central Oregon highlands and by the Okanogan highlands in the north, but the basin has still acted as a barrier to dispersal for several rainforest endemics (Carstens *et al.*, 2005). In addition to the orogeny of the Cascades in the Pliocene, Pleistocene climatic fluctuations have influenced the distributions of rainforest endemics in the PNW (Pielou, 2008). Specifically, glaciers intermittently covered large parts of species' contemporary ranges, and rainforest endemics may have been eliminated from northern parts of their ranges completely or may have survived in isolated refugia (Brunsfield & Sullivan, 2005).

*Corresponding author. E-mail: megansmth67@gmail.com

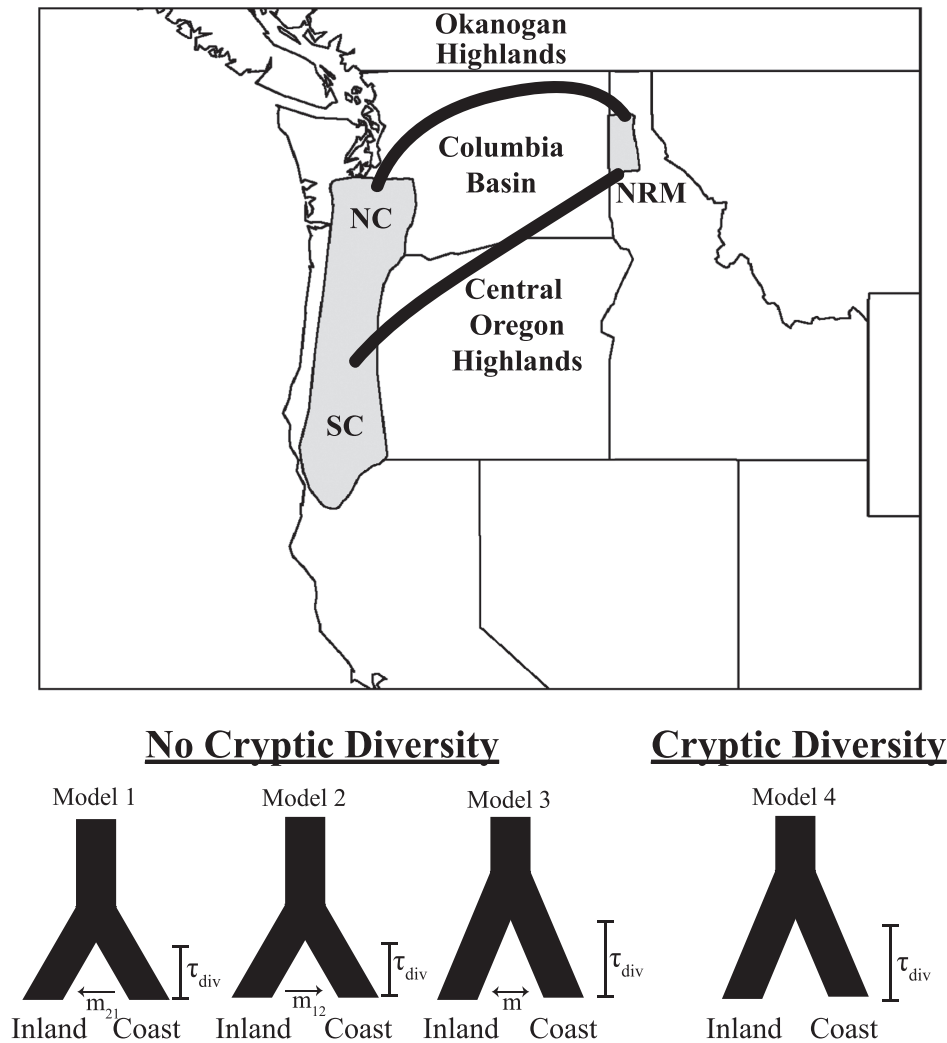


Figure 1. Phylogeographical models compared in the present study. Abbreviations: m_{12} , migration from population 1 into population 2; NC, North Cascades; NRM, Northern Rocky Mountains; SC, South Cascades; τ_{div} , divergence time. The grey areas on the map mark the distribution of *Prophysaon coeruleum*, adapted from [Burke \(2013\)](#). Black lines indicate northern and southern dispersal routes.

Several species with this disjunct distribution also have populations in the Blue and Willowa Mountains of southeastern Washington and northeastern Oregon. Some studies have found that these populations are closely related to populations in the Northern Rockies ([Nielson, Lohman & Sullivan, 2001](#)), whereas in other species, such as the polydesmid millipede *Chonaphe armata* (Harger, 1872; [Espíndola et al., 2016](#)) and *Prophysaon vanattae* ([Pilsbry, 1948](#)), populations in the Blue or Willowa Mountains are phenotypically more similar to populations in the Cascades.

This compelling geological history has inspired a great deal of phylogeographical work, with several hypotheses proposed to explain the disjunct distributions of mesic forest endemics (reviewed by

[Brunsfield et al., 2001](#)). The ‘ancient vicariance’ hypothesis posits that populations survived in both inland and coastal rainforests throughout the Pleistocene climatic fluctuations and that no gene flow has occurred between inland and coastal populations since the Pliocene. A variation on this hypothesis considered by [Espíndola et al. \(2016\)](#) posits pre-Pleistocene divergence between inland and coastal populations but allows for subsequent intermittent gene flow between inland and coastal populations along ephemeral habitat corridors at the margins of retreating glaciers. Another class of hypotheses reviewed by [Brunsfield et al. \(2001\)](#), referred to as ‘recent dispersal’ hypotheses, posits that either inland or on the coast, no populations survived Pleistocene

climate fluctuations. These models predict post-Pleistocene divergence with subsequent dispersal either from coastal to inland or from inland to coastal populations. These models have received mixed support; data from several amphibian species suggest Pliocene divergence between inland and coastal populations and support the ancient vicariance model (Nielson *et al.*, 2001; Carstens *et al.*, 2004; Steele *et al.*, 2005), whereas data from dusky willows and robust lancetooth snails support a model of recent dispersal from coastal to inland rainforests (Carstens *et al.*, 2013; Smith *et al.*, 2017), and data from red alder support a more nuanced model of ancient vicariance, with intermittent gene flow in one or both directions (Ruffley *et al.*, 2018).

The ancient vicariance hypothesis predicts the presence of cryptic diversity structured across the Columbia Basin, owing to deep divergence between inland and coastal populations, and some species (e.g. the tailed frog, *Ascaphus montanus*; Nielson *et al.*, 2001) have been recognized after genetic data were collected to test this hypothesis. Recently, Espíndola *et al.* (2016) developed an approach to predict the presence or absence of such cryptic diversity that uses a machine-learning algorithm and genetic, taxonomic and environmental data from previously studied taxa to construct a classifier that attempts to predict the presence or absence of cryptic diversity in unsampled species. This method allows researchers to make predictions about which unstudied taxa are likely to harbour cryptic diversity, and may prove useful when limited resources force researchers to focus on only one or a few taxa. As part of the ongoing evaluation of this predictive framework for phylogeography, we apply random forest (RF) classification to terrestrial slugs endemic to the PNW and test the resulting predictions using species distribution models (SDMs) and genetic data.

MATERIAL AND METHODS

STUDY SYSTEM AND SAMPLING

Slugs of the genus *Prophysaon* are endemic to the mesic forests of the PNW, with several species exhibiting the characteristic mesic forest disjunct distribution on either side of the Columbia Basin. Here, we focus on three disjunct species: *Prophysaon andersoni* (J.G. Cooper, 1872), *Prophysaon coeruleum* (Cockerell, 1890) and *Prophysaon dubium* (Cockerell, 1890). Additionally, we include two sister species: *Prophysaon vanattae* (Pilsbry, 1948) and *Prophysaon humile* (Cockerell, 1948). *Prophysaon vanattae* occurs only in the Cascades, whereas *P. humile* is endemic to the inland rainforest. *Prophysaon vanattae* and *P. humile* were classified as belonging to the subgenus

Mimetarion (Pilsbry, 1948), along with *Prophysaon obscurum* (Cockerell, 1893) and *Prophysaon fasciatum* (Cockerell, 1890). *Prophysaon fasciatum* does not seem to be a distinct species (Pilsbry & Vanatta, 1898), and *P. obscurum* (Cockerell, 1893) is a narrow endemic found only south of the South Puget Sound and into western Washington and in the Columbia Gorge along the Washington–Oregon border (Burke, 2013). Molecular data indicate that *P. obscurum* is not distinct from *P. vanattae* (see [Supporting Information: Gene Tree of this article](#); Wilke & Duncan, 2004), suggesting that it is appropriate to consider *P. vanattae* and *P. humile* as sister taxa that diverged across the Columbia Basin. Although this may seem to suggest deep divergence between the inland and coastal sister taxa, we chose to include these taxa in the present study because there has been no study of the genetic divergence between these two taxa, and it is unknown whether the phenotypic divergence used to delineate the two species corresponds to deep genetic divergence. We collected 223 samples from throughout the ranges of the focal taxa and other species of *Prophysaon* [*Prophysaon foliolatum* (Gould, 1851) and *P. obscurum*] from field collections and from museums (Fig. 2; [Supporting Information, Table S1](#)). Sites were visited during the autumn of 2016, and slugs were stored in 95% ethanol immediately after collection. Additional samples were requested from museum collections (Carnegie Museum of Natural History, Royal British Columbia Museum and the California Academy of Sciences).

PREDICTING CRYPTIC DIVERSITY

We applied the approach developed by Espíndola *et al.* (2016) to predict whether the focal taxa harbour cryptic diversity structured across the Columbia Basin. Specifically, we trained an RF classification function on a set of taxa that had already been studied [the water vole *Microtus richardsoni* (J.E. Dekay, 1842), the tree *Salix melanopsis* (Nutt), the millipede *C. armata*, the frog *Ascaphus montanus/truei*, the salamander *Dicamptodon atterimus/copei* (Cope, 1867; Nussbaum, 1970) and the salamander *Plethodon idahoensis/vandykei* (Slater & Slipp, 1940; Van Denburgh, 1906)]. These reference taxa were classified as ‘cryptic’ or ‘non-cryptic’ based on genetic data provided by Espíndola *et al.* (2016), where ‘cryptic’ species had cryptic diversity structured across the Columbia Basin, and non-cryptic species did not. In addition to the data used by Espíndola *et al.* (2016) to train their predictive function, we included data from *Haplotrema vancouverense* (I. Lea, 1839), a snail endemic to the PNW for which genomic data have demonstrated a lack of cryptic diversity structured

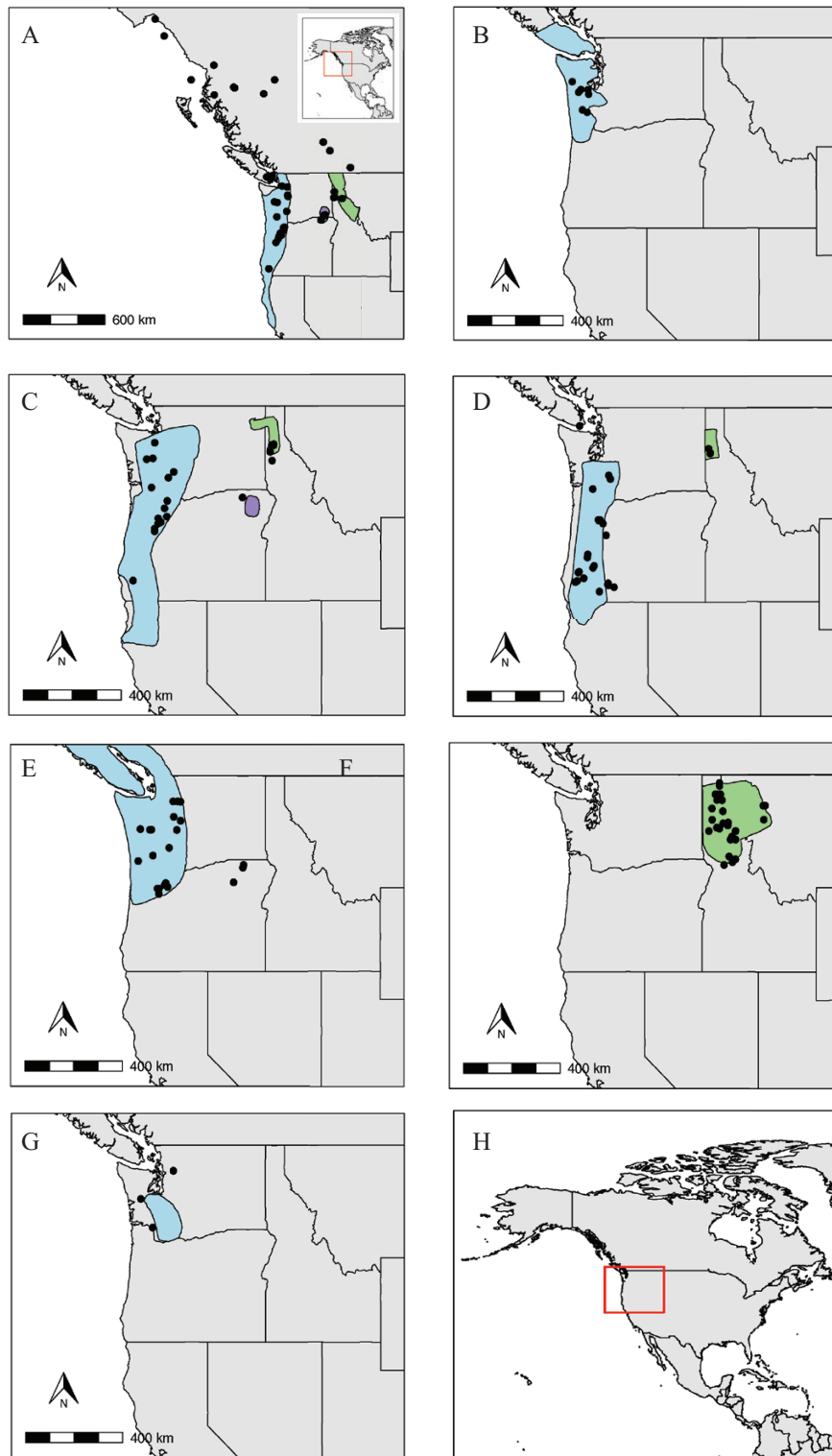


Figure 2. The distributions of all *Prophysaon* species considered in the present study, adapted from Burke (2013), with collection localities from the present study. Coastal ranges are shown in blue, inland ranges in green, and Blue and Wallowa ranges in purple. A, *Prophysaon andersoni*. B, *Prophysaon foliolatum*. C, *Prophysaon dubium*. D, *Prophysaon coeruleum*. E, *Prophysaon vanattae*. F, *Prophysaon humile*. G, *Prophysaon obscurum*. H, map showing context for maps B–G.

across the Columbia Basin (Smith *et al.*, 2017). We included this taxon because it is more closely related to *Prophysaon* than other species in the training set, and thus should improve the performance of the predictive function for the focal taxa. The occurrence data detailed by Espíndola *et al.* (2016) were also used here to extract eight environmental variables from the WorldClim database (Hijmans *et al.*, 2005): annual mean temperature, mean diurnal range, isothermality, maximum temperature of warmest month, temperature annual range, annual precipitation, precipitation seasonality and precipitation of driest quarter. In addition to environmental data, we used taxonomic information to train the classifier. Specifically, we classified each taxon as mollusc, arthropod, mammal, amphibian or plant. These broad taxonomic categories are intended to serve as a proxy for life-history characteristics that may influence traits such as dispersal ability. The dataset from Espíndola *et al.* (2016) combined with the data from *H. vancoverense* was used to train an RF classifier using the R package ‘randomForest’ (Liaw & Wiener, 2002). We used the same down-sampling strategy described by Espíndola *et al.* (2016) to account for differences in the number of cryptic and non-cryptic observations with 100 down-sampled replicates. We assessed the accuracy of these classifiers using cross-validation, where one taxon was omitted when building the RF classifier. The classifier was then applied to the omitted taxon, and the probability that the omitted taxon harboured or lacked cryptic diversity was calculated. We then applied the classifier to the four *Prophysaon* taxa with disjunct distributions and estimated the probability of each taxon harbouring or lacking cryptic diversity. Occurrence points for *Prophysaon* were from this study only, as the identification of occurrence data from data aggregators (e.g. the Global Biodiversity Information Facility, GBIF) could not be verified, and misidentifications are common in this group. Climate data were downloaded from WorldClim for these occurrence points for the eight bioclimatic variables listed above at a resolution of 30 arc s (Hijmans *et al.*, 2005). To evaluate the importance of the taxonomic predictors in driving the classifier, we also constructed and applied a classifier using no taxonomic variables and using a different set of classifications (Supporting Information: Random Forest Predictions) to evaluate the effects of including taxonomy in the classifier.

SPECIES DISTRIBUTION MODELS

In addition to the RF classifier, SDMs may help to predict whether taxa will harbour cryptic diversity structured across the Columbia Basin. Specifically in the PNW, where the presence or absence of cryptic

diversity is largely driven by the persistence of habitat through the Pleistocene glacial cycles, hindcast SDMs can be a powerful tool for predicting the presence of cryptic diversity (Richards, Carstens & Knowles, 2007). If no suitable habitat is predicted in the Northern Rocky Mountains during the LGM (assuming accurate SDMs and palaeoclimate reconstructions), the focal taxa are likely to have colonized the inland after glaciation, and we would not expect cryptic diversity across the Columbia Basin. On the contrary, if suitable habitat persisted in the inland rainforests during the LGM, there may have been refugia present at that time, and deep genetic divergence might exist between inland and coastal populations.

We built SDMs using the ensemble method implemented in the R package ‘biomod2’ (Thuiller, Georges & Engler, 2014) and the ten modelling approaches available in the package (Supporting Information: Species Distribution Modeling-Ensemble Approach). For the SDMs, we used only samples collected from the present study, as misidentifications are common in this group. We removed museum specimens with incorrect registers (e.g. registers that fell in the Pacific Ocean) and duplicate samples. This data curation resulted in 42 occurrence points for *P. andersoni*, 29 for *P. coeruleum*, 23 for *P. dubium*, and 52 for the *P. vanattae*/*P. humile* sister species pair (Supporting Information, Table S2). Climate data were downloaded from the WorldClim database (Hijmans *et al.*, 2005). Current climate data were at a resolution of 30 arc s, and LGM data were at a resolution of 2.5°. We selected uncorrelated bioclimatic variables for each species ($r < 0.7$) and chose among highly correlated variables by prioritizing variables that we thought would be most important for the focal taxa. The selected variables are reported in the Supporting Information (Table S3). We then cropped these layers to the extent of the focal region, which was defined as -150 to -100° longitude and 35 – 65° latitude. We chose an extent broader than the range of the focal taxa because our aim was to hindcast the SDMs. Given that suitable habitat in the past might have occurred outside current ranges, we included areas not currently occupied by the focal taxa, which is the usual approach in phylogeographical hindcasting studies (e.g. Espíndola *et al.*, 2012; Gavin *et al.*, 2014).

We used several modelling methods (Supporting Information: Species Distribution Modeling-Ensemble Approach) and ran five replicates per model. We randomly sampled 10 000 pseudoabsences from the entire background area (defined by the extent of the focal region) using the ‘random’ strategy available in BioMod. We chose this strategy because results have been equivocal on which sampling strategy should be preferred, with performance varying greatly across

datasets and methods, but with random sampling tending to perform best across regression methods, and with a relatively small decrease in performance with random sampling compared with other sampling methods in classification and machine-learning techniques (Barbet-Massin *et al.*, 2012).

To build models in Maxent, we used a maximum of 1000 iterations to reach convergence and included linear, quadratic, product, threshold and hinge features. For parameters not specified above, we used the default BioMod parameterization. We used 80% of our data for training models and 20% for testing, and five replicates for each model to evaluate model performance. We performed three replicates to determine variable importance and evaluated models using the area under the curve (AUC) of the receiver operating characteristic (ROC) curve. Models were rescaled, so that they could be combined later. We then built an ensemble model ignoring models with an ROC of < 0.85 and weighting all other models by ROC score. Finally, we forecast the ensemble model onto current and past (LGM) climate conditions. To evaluate the impacts of modelling choices on our SDM results, we also constructed SDMs using an alternative strategy and found that our results did not change substantially (Supporting Information: Species Distribution Modeling using Ecoregions and Maxent).

DNA ISOLATION, SEQUENCING AND ANALYSES

DNA was extracted using DNeasy Blood and Tissue kits (Qiagen), following the manufacturer's standard protocol. A 710 bp portion of the *COI* gene was amplified using the primer pair LCO1490 and HCO2198 (described by Folmer *et al.*, 1994). Forward and reverse reads were assembled in Geneious v6.1.7 (Kearse *et al.*, 2012) and edited by eye when necessary. All available sequences (46 additional sequences) for *Prophyaon* were downloaded from GenBank in August 2017. The *muscle* (Edgar, 2004) algorithm available in Geneious v6.1.7 (Kearse *et al.*, 2012) was used to generate an alignment.

We used the AutoModel function in PAUP* v4.0a157 (Swofford, 2002) to select the best model of sequence evolution for the entire dataset. To select the best model of nucleotide substitution for our approximate Bayesian computation (ABC) analysis (see next subsection), we used the AutoModel function separately for datasets from each of the four disjunct taxa, excluding samples from the Blue and Wallowa Mountains (see next subsection). As a starting tree, we used a neighbor joining tree calculated from Jukes Cantor distances. We used the largest model set (11 substitution schemes), and considered models with and without gamma-distributed rate variation across sites and a proportion of invariable sites. We evaluated

models using the small-sample-size corrected version of the Akaike information criterion (AICc), Bayesian information criterion (BIC) and decision theory (DT; Minin *et al.*, 2003).

To obtain posterior distributions of the parameters for the model of sequence evolution for use in the ABC analyses (see next subsection), we used MrBayes v.3.2.6 (Ronquist *et al.*, 2012). For each of the four disjunct taxa, MrBayes was run under the model of sequence evolution selected by PAUP if the model could be implemented. If the model selected using the methods above could not be implemented in MrBayes, we reran the AutoModel function in PAUP* v4.0a157 (Swofford, 2002) using the reduced set of models (three substitution schemes). We conducted 400 independent runs, with 32 000 000 generations and four chains per run. For models that included gamma rate variation across sites, we used eight rate categories. We adjusted the temperature parameter to improve mixing, such that acceptance rates of initial runs were between 10 and 70%. We discarded 25% of runs as burn-in and combined the 100 independent runs for each of the focal taxa; the resulting posterior distributions of parameters were used in downstream analyses. We checked that the average deviation of split frequencies was < 0.01 for each run to assess convergence. Additionally, we estimated a maximum likelihood gene tree in GARLI v. 2.01 (Bazin, Zwickl & Cummings, 2014), with the slugs *Hemphillia malonei* and *Zacoleus idahoensis* as outgroups (Supporting Information: Gene Tree).

DEMOGRAPHIC MODEL SELECTION USING APPROXIMATE BAYESIAN COMPUTATION

To test whether the focal taxa harboured cryptic diversity structured across the Columbia Basin, we first evaluated recent dispersal models (Fig. 1). The first two models consisted of post-Pleistocene divergence with dispersal and subsequent gene flow either from coastal to inland rainforests or from inland to coastal rainforests. These models correspond to a lack of suitable habitat in either the inland or the coastal rainforests during the Pleistocene glacial cycles and subsequent post-Pleistocene colonization. The third recent dispersal model included pre-Pleistocene divergence with subsequent gene flow in both directions, approximating a scenario in which there was ancient vicariance followed by secondary contact (Ruffley *et al.*, 2018). After comparing the recent dispersal models, we calculated the posterior probability of the best dispersal model and a model of pre-Pleistocene divergence with no subsequent gene flow (ancient vicariance). The models tested were parameterized such that population sizes could vary for each of the two populations. All divergence time and

population size parameters were drawn from uniform priors, which were adjusted after initial runs to ensure that simulated summary statistics were in the range of observed summary statistics for each taxon and region. Full information on the priors for each parameter and each taxon is available in the [Supporting Information \(Table S4\)](#). For the purpose of the ABC analysis, we did not consider the Blue and Wallowa populations, because we had too few samples from these regions to model these populations separately.

To compare these models within each of the four focal taxa, we simulated 100 000 datasets under each of the three migration models in ms (Hudson, 2002). We then used the program Seq-Gen v.1.3.4 (Rambaut & Grassly, 1997) to simulate sequence data from the gene trees simulated in ms. We simulated 574 bp, to match the observed data. Parameters for the model of sequence evolution were drawn from the posterior distribution of parameters from the MrBayes analyses (see previous subsection). We drew the scaling parameter from a uniform prior distribution (Supporting Information, Table S4). We then used a custom python script (available at <https://github.com/meganlsmith>) to calculate six summary statistics: π ; the number of segregating sites (S); Watterson's θ (θ_w); π within each population; and the number of

nucleotide differences between populations (*nucdiv*).

We calculated π as $\pi = 2^* \sum_{i=2}^N \sum_{j=1}^{i-1} x_i x_j \pi_{ij}$, where x_i and

x_j are the frequencies of the i th and j th sequences, π_{ij} is the number of nucleotide differences per nucleotide site between sequences i and j , and N is the number of unique sequences in the sample (Nei, 1979). Watterson's θ was calculated as the number of segregating sites divided by the $(N - 1)$ th harmonic number (Watterson, 1975). The number of nucleotide differences between subpopulations was simply the sum of the number of nucleotide differences between sequences in each of the two populations. We calculated the same summary statistics from the observed data using a python script.

We evaluated a range of rejection methods (simple rejection and logistic regression), tolerances (0.001, 0.005, 0.01 and 0.05) and summary statistic combinations using a custom R script and the R package 'abc' (Csilléry, François & Blum, 2012). We performed ten cross-validation replicates for each combination of rejection method, tolerance and summary statistic combination. We did not use more replicates owing to the large number of combinations being evaluated and the computational requirements of each replicate. Each method was evaluated based on the mean posterior probability of the correct model across the three migration models. However, using the

logistic rejection method resulted in errors with many simulated and empirical datasets because there was too little variation in summary statistics in the selected region. Therefore, we considered only rejection methods for downstream analyses. The best three methods (based on the mean posterior probability of the best model across all three models) were evaluated further using 100 cross-validation replicates. When more than three methods had equivalent posterior probabilities across all models being compared, three methods were chosen at random, and 100 cross-validation replicates were run for these three methods. The best of these three methods was then selected based on the sum of the posterior probability of the correct model across all three models. We then used the winning method to calculate the posterior probabilities of each model for each taxon.

In a second step, we compared the recent dispersal model that had the highest posterior probability with the ancient vicariance model. The methods were the same as those in the first step of the ABC analysis. Cross-validation analyses were conducted in the same way as above, and the best model was selected based on the winning method.

Finally, we used posterior predictive distributions to assess the fit of the best model to the data directly. To generate the posterior predictive distribution, we simulated 100 datasets under each set of parameters from the posterior distribution of parameters under the best model. For each summary statistic used, we calculated the difference between the posterior predictive distribution of the summary statistic and the observed statistic, and evaluated whether the 95% highest density interval (HDI) included zero using the R package 'HDInterval' (Meredith & Kruschke, 2017). A 95% HDI that does not include zero indicates that the model is not a good fit to the data. In addition to the ABC analysis described here, we conducted an ABC analysis where we simulated only infinite sites data, rather than sequence data (Supporting Information: ABC with Infinite Sites Data).

RESULTS

STUDY SYSTEM AND SAMPLING

We collected samples or downloaded data from seven described species of *Prophyaon*: *P. andersoni*, *P. coeruleum*, *P. dubium*, *P. foliolatum*, *P. humile*, *P. obscurum* and *P. vanattae* (Fig. 2; Supporting Information, Table S1). For *P. andersoni*, we collected or downloaded sequence data from 11 inland samples, 66 coastal samples, and four samples from the Blue and Wallowa mountains. For *P. coeruleum*, we collected or downloaded data from eight inland and 39 coastal

samples. For *P. dubium*, we collected or downloaded data from eight inland samples, 27 coastal samples, and one sample from the Blue and Wallowa mountains. For *P. vannatae* and *P. humile*, we collected or downloaded data from 38 inland samples, 33 coastal samples, and five samples from the Blue and Wallowa mountains.

PREDICTING CRYPTIC DIVERSITY

The cross-validation analysis indicated that the RF classifier performed well for most species. For all species except *C. armata*, the mean posterior probability of the correct classification was > 0.90 (Fig. 3), whereas for *C. armata*, the mean posterior probability of a

cryptic classification was 0.0125. This species was also identified as problematic by Espíndola *et al.* (2016), and these issues were attributed to difficulty in classifying *C. armata* as cryptic or non-cryptic using molecular data. All *Prophysaon* species groups were predicted to lack cryptic diversity, with a posterior probability > 0.98 (Fig. 3). However, when taxonomy was omitted or reclassified, these predictions changed. Specifically, without taxonomy, all focal taxa were predicted to harbour cryptic diversity, but with a low probability (Supporting Information, Fig. S1A), and when taxonomy was reclassified as vertebrate, invertebrate or plant, all taxa were predicted to harbour cryptic diversity with a high probability (Supporting Information, Fig. S1B).

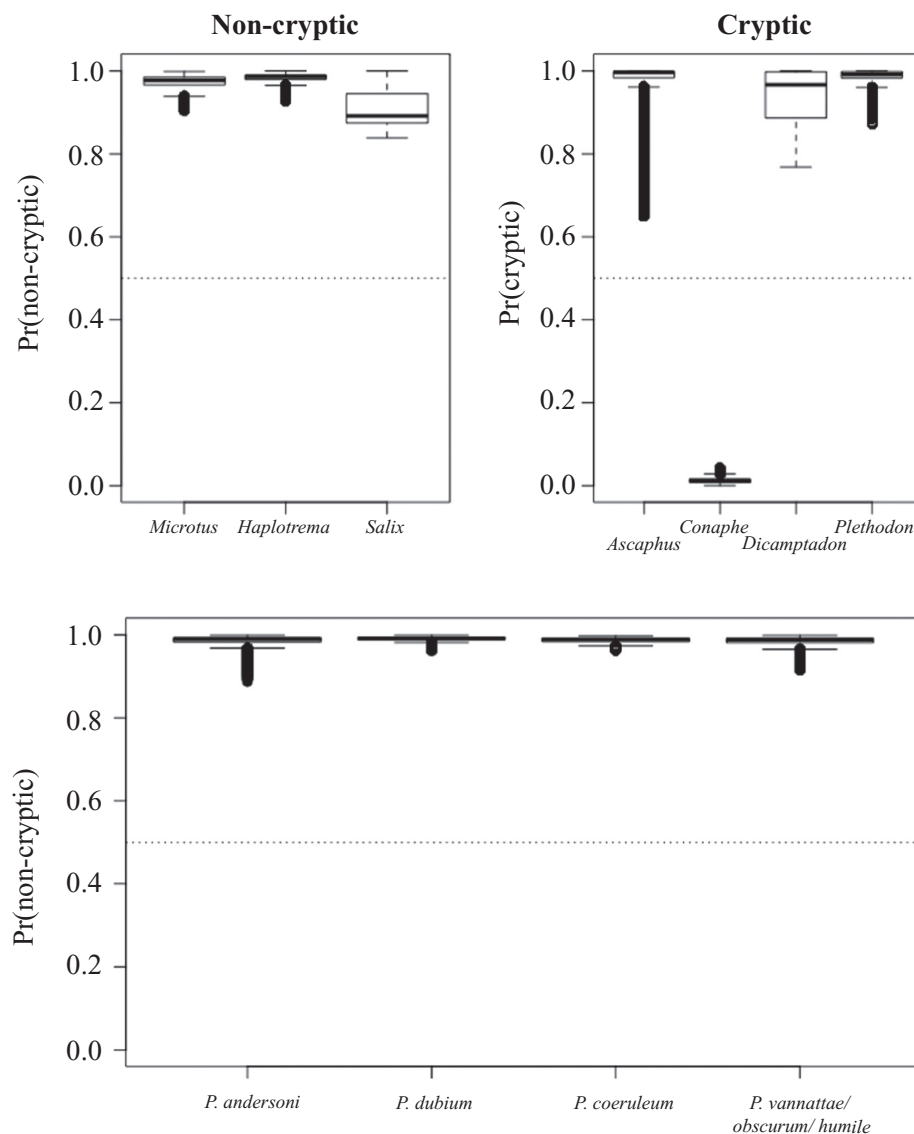


Figure 3. Results from the random forest analysis. Top: results from the cross-validation analysis. Bottom: classification results for *Prophysaon*.

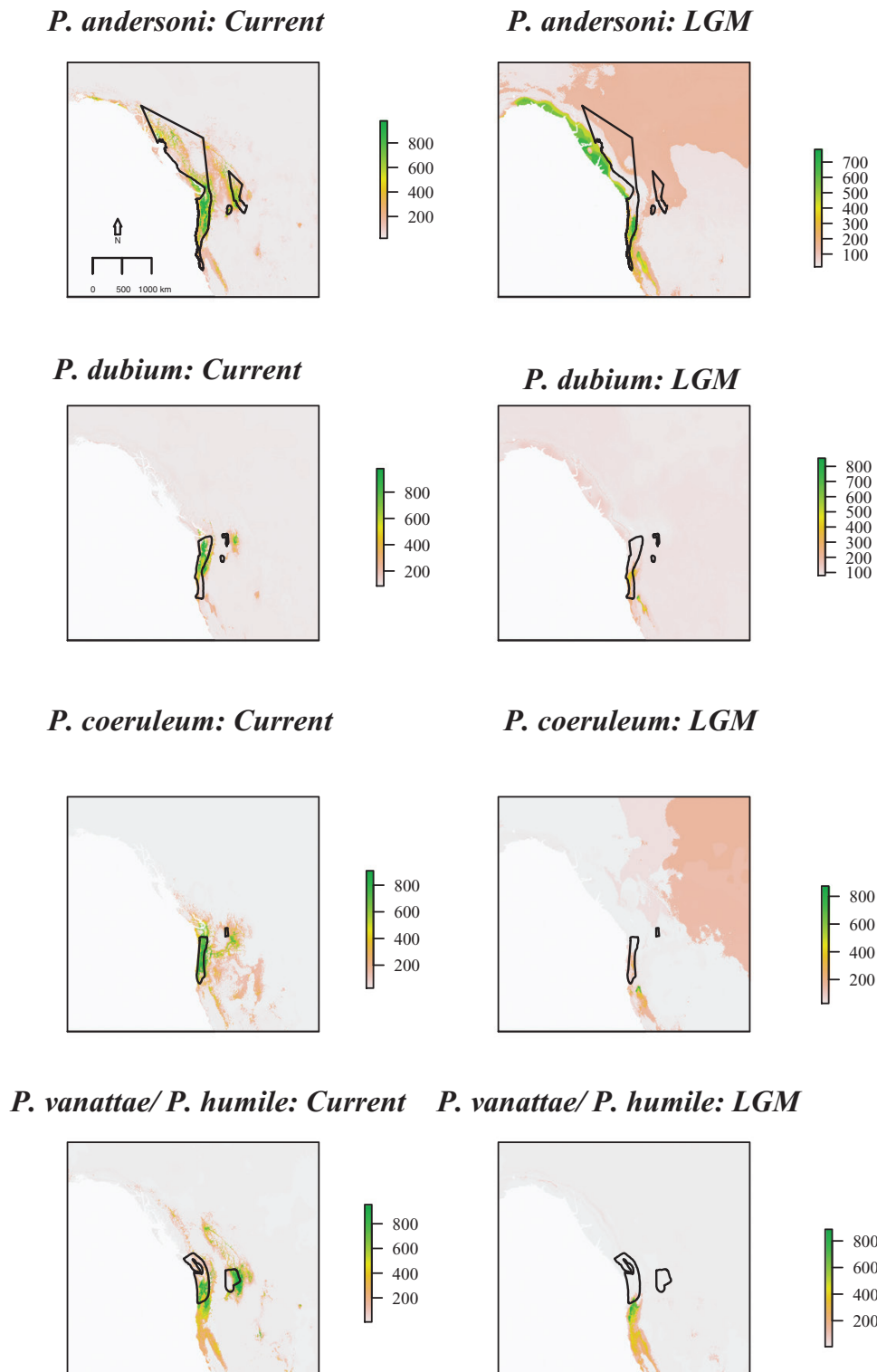


Figure 4. Species distribution models for the present and the Last Glacial Maximum (LGM). A, *Prophysaon andersoni*, current. B, *Prophysaon coeruleum*, current. C, *Prophysaon dubium*, current. D, *Prophysaon vanattae/Prophysaon humile*, current. E, *P. andersoni*, LGM. F, *P. coeruleum*, LGM. G, *P. dubium*, LGM. H, *P. vanattae/P. humile*, LGM. Current ranges, following Burke (2013), are outlined in black.

SPECIES DISTRIBUTION MODELS

The mean ROC AUC scores with and without models with an AUC < 0.85, respectively, were: 0.910 and 0.948 for *P. andersoni*, 0.925 and 0.975 for *P. dubium*, 0.923 and 0.963 for *P. coeruleum*, and 0.969 and 0.972 for *P. vanattae/P. humile*. Current distribution models indicated suitable habitat inland and on the coast (Fig. 4) and recovered the known range of the different taxa (Fig. 4). When SDMs were hindcast to conditions at the LGM, suitability values were extremely low in the inland portions of the species ranges (Fig. 4).

DNA ISOLATION AND SEQUENCING

The final alignment included 574 bases of the mitochondrial gene *COI*, with no stop codons in the reading frame. There was an AT bias in base composition (A, 0.321; C, 0.112; G, 0.108; T, 0.459), comparable to that found in other invertebrate mitochondrial genes (Lin & Danforth, 2004; Liu *et al.*, 2012; Hilgers *et al.*, 2016). Maximum within-species maximum likelihood distances were as follows: *P. andersoni*, 0.447; *P. foliolatum*, 0.153; *P. dubium*, 0.147; *P. coeruleum*, 1.080; *P. vanattae*, 0.654; *P. humile*, 0.117; and *P. obscurum*, 0.083. The results of model selection in PAUP for the full dataset were concordant across AICc, BIC and DT methods (TPM2uf + I + Γ), and all analyses on the full dataset were conducted using this model. We also used PAUP to determine the best model for each disjunct taxon. For *P. andersoni*, *P. coeruleum* and *P. vanattae/P. humile*, the results supported an HKY + I + Γ model regardless of the criteria (i.e. AICc, BIC, DT), whereas a TPM3uf + I model was supported for *P. dubium*. However, as this model cannot be implemented in MrBayes, we reran model selection for *P. dubium* using a reduced model set that considered only the three substitution schemes that can be implemented in MrBayes and chose an HKY + I model. Parameter estimates from PAUP are reported in the Supporting Information (Table S5). For the MrBayes analyses there was no evidence of a lack of convergence, and all runs had a standard deviation of split frequencies < 0.01.

DEMOGRAPHIC MODEL SELECTION USING ABC

Recent dispersal models

The summary statistics for the observed data are reported in Table 1. Cross-validation analyses indicated moderate to low ability to distinguish among the three dispersal models. Across all species, the posterior

probability of the correct model ranged from 0.361 to 0.763 in the cross-validation analysis (Table 2). For *P. andersoni*, *P. dubium* and the *P. vanattae/P. humile* sister species pair, model 1 had the highest posterior probability. For *P. coeruleum*, model 3 had the highest posterior probability (Table 3).

Recent dispersal vs. ancient vicariance

Cross-validation analyses indicated high power to distinguish between the recent dispersal and ancient vicariance models; the posterior probability of the correct model ranged from 0.973 to 0.999 (Table 4). For all species, the recent dispersal model received the highest posterior probability, and the posterior probability of the recent dispersal model was always one (Table 5). Across all species, the 95% highest density interval of the difference between the posterior predictive distribution and the observed data contained zero, meaning there was no indication of poor model fit. Plots of the posterior and posterior predictive distributions are available in the Supporting Information (Fig. S2). The results did not differ substantially when only infinite sites data, rather than sequence data, were simulated (Supporting Information: ABC with Infinite Sites Data).

DISCUSSION

PREDICTING CRYPTIC DIVERSITY

The predictive model correctly predicted that all disjunct taxa included in the present study lacked cryptic diversity structured across the Columbia Basin. However, this finding might be driven by the relative lack of taxonomic breadth of the groups included in our model, because the only other mollusc included was *H. vancouverense*, which also lacks cryptic diversity. To gain a better understanding of the role of taxonomy in our predictions, we omitted taxonomy and found that we predicted that all *Prophysaon* species would harbour cryptic diversity with low to moderate posterior probabilities. We also ran the predictive model with coarsened taxonomic classification by recognizing only vertebrates, invertebrates and plants. In this case, there were two invertebrates in the model, the land snail *H. vancouverense* and the millipede *C. armata*, and *C. armata* was classified as cryptic. This again resulted in predictions that all species of *Prophysaon* would harbour cryptic diversity with high posterior probabilities. These results indicate that our RF predictions might be overly reliant on taxonomy, most probably because taxonomy serves as a proxy for important traits, including dispersal ability, that are

Table 1. Summary statistics calculated for the observed data for use in the approximate Bayesian computation (ABC) analyses

Species	π	S	Tajima's D	θ_H	H	π_{inland}	π_{coast}	$nucdiv$	θ_w
<i>Prophyaon andersoni</i>	0.0475	144	-0.452	10.0	0.772	0.00824	0.0509	23.2	29.3
<i>Prophyaon coeruleum</i>	0.0950	180	0.233	18.6	0.639	0	0.09819	55.1	40.8
<i>Prophyaon dubium</i>	0.0503	90	0.832	16.5	0.564	0.00859	0.0510	31.3	21.9
<i>Prophyaon vanattae</i> / <i>Prophyaon humile</i>	0.0860	172	-0.0692	15.5	0.437	0.04910	0.0801	62.7	35.6

Fay and Wu's H (H), Fay's θ_H statistic (θ_H), nucleotide divergence between inland and coastal populations ($nucdiv$), nucleotide diversity (π), nucleotide diversity within the coastal population (π_{coast}), nucleotide diversity within the inland population (π_{inland}), number of segregating sites (S), and Watterson's theta (θ_w).

Table 2. Cross-validation results for the three recent dispersal models

Species	Method	Tolerance	Sumstats	pp(M1)	pp(M2)	pp(M3)
<i>Prophyaon andersoni</i>	Rejection	0.001	π (inland), π (coast), π_{12}	0.468	0.520	0.385
<i>Prophyaon dubium</i>	Rejection	0.001	π , S , π (inland), π (coast), π_{12}	0.465	0.464	0.361
<i>Prophyaon vanattae</i> / <i>Prophyaon humile</i>	Rejection	0.001	π , π (inland), π (coast), π_{12} , θ_w	0.456	0.456	0.414
<i>Prophyaon coeruleum</i>	Rejection	0.001	π , S , π (inland), π_{12}	0.530	0.763	0.493

Models correspond to those shown in Figure 1, and the posterior probability (pp) of a model (M) is the mean posterior probability across the cross-validation replicates in which the data were simulated under that model.

Table 3. Approximate Bayesian computation results for the three recent dispersal models

Species	M1	M2	M3	BF
<i>Prophyaon andersoni</i>	0.697	0.034	0.269	2.60
<i>Prophyaon dubium</i>	0.540	0.173	0.287	1.88
<i>Prophyaon vanattae</i> / <i>Prophyaon humile</i>	0.580	0.290	0.130	2.00
<i>Prophyaon coeruleum</i>	0.287	0.300	0.413	1.38

The models (M) correspond to those shown in Figure 1, and the posterior probabilities are those calculated using the best method selected by cross-validation. The Bayes factors (BF) shown are those comparing the model having the highest posterior probability with the model having the second highest posterior probability.

not quantified directly. Continuing to add more species to the dataset, as this work has done, and generating data on the variables we are attempting to summarize with taxonomy will improve future classifications.

Given that *P. vanattae* and *P. humile* are currently described as different species, our finding of recent dispersal between these two is surprising. Although researchers commonly discuss the lack of phenotypic divergence when genetic divergence is present (cryptic diversity), the inverse is also a common phenomenon. Even within *Prophyaon*, subspecies designations based on morphology have been questioned in the light

of molecular data (Wilke & Duncan, 2004). Within *P. vanattae*, there exists extensive phenotypic variation (Burke, 2013), which might exceed that between *P. vanattae* and *P. humile*. Given that taxonomic classifications drive how we group organisms, and thus our results in phylogeographic studies, this work highlights the need to revisit these classifications in light of phylogeographical evidence (such as that presented here) and genomic data. It is possible that our inference is driven by incorrectly treating all *P. vanattae* as a single species or by incorrectly splitting *P. vanattae* and *P. humile*, and future work should test these hypotheses using increased sampling, in terms of both individuals and loci. Furthermore, support for *P. vanattae* and *P. humile* as sister species is limited, and the gene tree estimated here does not support this relationship. It is possible that these species are not sister species, and future work including more loci should evaluate this relationship and reconsider the results reported here for these species. Careful species delimitation using multiple lines of evidence (e.g. morphological, ecological and genomic data) will be necessary to characterize these taxa accurately.

ASSESSING MODEL FIT

Given that the use of ABC for phylogeographical inference can tell us only which of the tested

Table 4. Cross-validation results for the step comparing the best recent dispersal model with the ancient vicariance model

Species	Recent dispersal model	Method	Tolerance	Sumstats	pp(recent dispersal)	pp(ancient vicariance)
<i>Prophysaon andersoni</i>	1	Rejection	0.001	π, S, π (inland), π (coast), θ_w	0.977	0.986
<i>Prophysaon dubium</i>	1	Rejection	0.001	S, π_{12}	0.977	0.981
<i>Prophysaon vanattae</i> / <i>Prophysaon humile</i>	1	Rejection	0.001	π, π (inland), π_{12}, θ_w	0.973	0.995
<i>Prophysaon coeruleum</i>	3	Rejection	0.001	π, S, π (coast)	0.984	0.999

Models correspond to those shown in Figure 1, and the posterior probability (pp) of a model is the mean posterior probability of the model across the cross-validation replicates in which the data were simulated under that model.

Table 5. Approximate Bayesian computation results for step comparing the best recent dispersal (RD) model with the ancient vicariance (AV) model

Species	RD model	p(RD)	p(AV)	BF
<i>Prophysaon andersoni</i>	1	1	0	Inf
<i>Prophysaon dubium</i>	1	1	0	Inf
<i>Prophysaon vanattae</i> / <i>Prophysaon humile</i>	1	1	0	Inf
<i>Prophysaon coeruleum</i>	3	1	0	Inf

The models correspond to those shown in Figure 1, and the posterior probabilities are those calculated using the best method selected by cross-validation. The Bayes factors (BF) shown are those comparing the model having the highest posterior probability with the model having the second highest posterior probability. Inf = infinite.

models is the best fit to the data, and not how well the models fit the data, it is essential to accompany model selection with tests of model fit. This need is particularly pronounced when analysing relatively limited datasets, such as the single mitochondrial locus analysed here. We addressed this issue with posterior predictive simulation tests of model fit, and we found no indication that the best model was a poor fit to the data for any of the focal taxa. This increases our confidence in the model selection results, because it suggests that not only was the selected model a better fit to the data than other models in the model set, but also the selected model was a reasonable fit to the data. This diminishes our concerns about how our choice of models might have influenced the results of model selection.

CONCLUSIONS

The work presented here represents a substantial expansion of the predictive framework described by Espíndola *et al.* (2016). It highlights the utility of this framework, while suggesting areas where it could be improved. The framework appears to make accurate predictions in all species analysed here, but these predictions are driven largely by taxonomy. Perhaps the most promising aspect of the predictive framework

for phylogeography is its capacity to integrate phylogeographical research conducted at different points in time and in different empirical systems. An attribute of this integration is that the framework should increase in its utility and accuracy as more diverse taxa are incorporated. Our study has nearly doubled the number of species complexes (from seven to 12) that can be included in the predictive framework for the PNW temperate rainforest, and should improve the accuracy of this framework for future research via both the expansion of the taxon set and the careful exploration of the ABC methodology.

Indeed, such an iterative approach has been central to taxonomic research since its inception; hypotheses are formed, tested and then modified. The predictive framework developed by Espíndola *et al.* (2016), and used and expanded here, allows phylogeographical research to proceed in the same way. Information about previously studied species is used to construct hypotheses about what factors affect the phylogeographical histories of different taxa. In the present study these factors were environmental and taxonomic, but they could also include trait and life-history data. These hypotheses are then used to generate predictions about unstudied taxa. After these predictions are tested, we can modify our hypotheses by considering environmental, ecological

and other traits of the newly added taxa. This leads to a cyclical approach, in which we constantly modify and improve our hypotheses about how environmental and intrinsic factors drive species' responses to climatic and geological events, and provides a novel means to integrate phylogeographical datasets. This is no trivial accomplishment, given the scale of phylogeographical datasets that have been collected to date: a Web of Science search of the term phylogeograph* returned 15 911 hits (7 December 2017). Integrating these data into a common framework that allows researchers to develop and test hypotheses on a large scale has been a difficult task, and the predictive framework used here is a step towards that goal.

ACKNOWLEDGEMENTS

We would like to acknowledge Bill Leonard, Ben Stone and Tara Pelletier for assistance in the field. Funding was provided by the US National Science Foundation (DEB 1457726/14575199). M.L.S. was supported by an National Science Foundation Graduate Research Fellowship (DG-1343012). We thank the Royal British Columbia Museum, the Carnegie Museum and the California Academy of Sciences for providing samples. We also thank Michael Lucid of Idaho Fish and Game for donations of samples and the Ohio Supercomputer Center for computing resources (allocation grant PAS1181-1).

REFERENCES

- Barbet-Massin M, Jiguet F, Albert CH, Thuiller W. 2012. Modelling species distributions to map the road towards carnivore conservation in the tropics. *Methods in Ecology and Evolution* **3**: 327–338.
- Bazinot AL, Zwickl DJ, Cummings MP. 2014. A gateway for phylogenetic analysis powered by grid computing featuring GARLI 2.0. *Systematic Biology* **63**: 812–818.
- Brunsfeld SJ, Sullivan J. 2005. A multi-compartmented glacial refugium in the northern Rocky Mountains: evidence from the phylogeography of *Cardamine constancei* (Brassicaceae). *Conservation Genetics* **6**: 895–904.
- Brunsfeld SJ, Sullivan J, Soltis DE, Soltis PS. 2001. Chapter 15 Comparative phylogeography of north-western North America: a synthesis. Special Publication-British Ecological Society **14**: 319–340.
- Burke TE. 2013. *Land snails and slugs of the Pacific Northwest*. Corvallis, OR: Oregon State University Press.
- Carstens BC, Brennan RS, Chua V, Duffie CV, Harvey MG, Koch RA, McMahan CD, Nelson BJ, Newman CE, Satler JD, Seeholzer G, Posbic K, Tank DC, Sullivan J. 2013. Model selection as a tool for phylogeographic inference: an example from the willow *Salix melanopsis*. *Molecular Ecology* **22**: 4014–4028.
- Carstens BC, Brunsfeld SJ, Demboski JR, Good JM, 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 BC, Stevenson AL, Degenhardt JD, Sullivan J. 2004. Testing nested phylogenetic and phylogeographic hypotheses in the *Plethodon vandykei* species group. *Systematic Biology* **53**: 781–792.
- Csilléry K, François O, Blum MGB. 2012. abc: an R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution* **3**: 475–479.
- Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**: 113.
- Espíndola A, Pellissier L, Maiorano L, Hordijk W, Guisan A, Alvarez N. 2012. Predicting present and future intra-specific genetic structure through niche hindcasting across 24 millennia. *Ecology Letters* **15**: 649–657.
- Espíndola A, Ruffley M, Smith ML, Carstens BC, Tank DC, Sullivan J. 2016. Identifying cryptic diversity with predictive phylogeography. *Proceedings of the Royal Society B: Biological Sciences* **283**: 20161529.
- Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* **3**: 294–299.
- Gavin DG, Fitzpatrick MC, Gugger PF, Heath KD, Rodríguez-Sánchez F, Dobrowski SZ, Hampe A, Hu FS, Ashcroft MB, Bartlein PJ, Blois JL, Carstens BC, Davis EB, de Lafontaine G, Edwards ME, Fernandez M, Henne PD, Herring EM, Holden ZA, Kong WS, Liu J, Magri D, Matzke NJ, McGlone MS, Saltré F, Stigall AL, Tsai YH, Williams JW. 2014. Climate refugia: joint inference from fossil records, species distribution models and phylogeography. *New Phytologist* **204**: 37–54.
- Graham A. 1999. *Late Cretaceous and Cenozoic history of North American vegetation: north of Mexico*. New York: Oxford University Press on Demand.
- Hijmans RJ, Cameron S, Parra J, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**: 1965–1978.
- Hilgers L, Grau JH, Pfaender J, von Rintelen T. 2016. The complete mitochondrial genome of the viviparous freshwater snail *Tylomelania sarasinorum* (Caenogastropoda: Cerithioidea). *Mitochondrial DNA Part B: Resources* **1**: 330–331.
- Hudson R. 2002. ms - a program for generating samples under neutral models. *Bioinformatics* **18**: 337–338.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–1649.
- Liaw A, Wiener M. 2002. Classification and regression by randomForest. *R News* **2**: 18–22.

- Lin CP, Danforth BN. 2004.** How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. *Molecular Phylogenetics and Evolution* **30**: 686–702.
- Liu GH, Wang SY, Huang WY, Zhao GH, Wei SJ, Song HQ, Xu MJ, Lin RQ, Zhou DH, Zhu XQ. 2012.** The complete mitochondrial genome of *Galba pervia* (Gastropoda: Mollusca), an intermediate host snail of *Fasciola* spp. *PLoS ONE* **7**.
- Meredith M, Kruschke J. 2017.** HDInterval: highest (posterior) density intervals. R package version 0.1.3.
- Minin V, Abdo Z, Joyce P, Sullivan J. 2003.** Performance-based selection of likelihood models for phylogeny estimation. *Systematic Biology* **52**: 674–683.
- Nei M, Li WH. 1979.** Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America* **76**: 5269–5273.
- Nielson M, Lohman K, Sullivan J. 2001.** Phylogeography of the tailed frog (*Ascaphus truei*): implications for the biogeography of the Pacific Northwest. *Evolution* **55**: 147–160.
- Pielou EC. 2008.** *After the ice age: the return of life to glaciated North America*. Chicago, IL: University of Chicago Press.
- Pilsbry HA. 1948.** *Land Mollusca of North America. Vol. 2, part 2*. Philadelphia, PA: The Academy of Natural Sciences of Philadelphia.
- Pilsbry HA, Vanatta EG. 1898.** Revision of the North American Slugs: Binneya, Hemphillia, Hesperarion, Prophysaon and Anadenulus. *Proceedings of the Academy of Natural Sciences of Philadelphia* **50**: 219–261.
- Rambaut A, Grassly NC. 1997.** Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Computer Applications in the Biosciences* **13**: 235–238.
- Richards CL, Carstens BC, Knowles LL. 2007.** Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. *Journal of Biogeography* **34**: 1833–1845.
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012.** MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**: 539–542.
- Ruffley M, Smith ML, Espíndola A, Carstens BC, Sullivan J, Tank DC. 2018.** Combining allele frequency and tree-based approaches improves phylogeographic inference from natural history collections. *Molecular Ecology* **27**: 1012–1024.
- Smith ML, Ruffley M, Espíndola A, Tank DC, Sullivan J, Carstens BC. 2017.** Demographic model selection using random forests and the site frequency spectrum. *Molecular Ecology* **26**: 4562–4573.
- Steele CA, Carstens BC, Storfer A, Sullivan J. 2005.** Testing hypotheses of speciation timing in *Dicamptodon copei* and *Dicamptodon aterrimus* (Caudata: Dicamptodontidae). *Molecular Phylogenetics and Evolution* **36**: 90–100.
- Swofford DL. 2002.** *PAUP* version 4.0 a157*. Sunderland, MA: Sinauer.
- Thuiller W, Georges D, Engler R. 2014.** *biomod2: Ensemble platform for species distribution modeling. R package version 3.1–64*. Available at: <http://CRAN.R-project.org/package=biomod2>
- Watterson GA. 1975.** On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* **7**: 256–276.
- Wilke T, Duncan N. 2004.** Phylogeographical patterns in the American Pacific Northwest: lessons from the arionid slug *Prophysaon coeruleum*. *Molecular Ecology* **13**: 2303–2315.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. Sampling localities from this study, including samples from GenBank.

Table S2. Occurrence points used to build species distribution models (SDMs).

Table S3. Bioclimatic variables used in species distribution models.

Table S4. Priors for the approximate Bayesian computation (ABC) analysis.

Table S5. Parameters from PAUP* model selection results.

Table S6. Cross-validation results for the three recent dispersal models with infinite sites data.

Table S7. Approximate Bayesian computation (ABC) results for the three recent dispersal models with infinite sites data.

Table S8. Cross-validation results for the recent dispersal vs. ancient vicariance step with infinite sites data.

Table S9. Approximate Bayesian computation (ABC) results for the recent dispersal vs. ancient vicariance step with infinite sites data.

Figure S1. Results of predictions from the RF classifier when (A) no taxonomy and (B) a revised classification was used. The revised classification included three ranks: vertebrate, invertebrate and plant.

Figure S2. Posterior and posterior predictive distributions are shown compared with observed summary statistics when sequence data were simulated. Posterior distributions shown in blue, posterior predictive distributions are shown in red, and observed data are indicated by black dotted lines.

Figure S3. Species distribution models constructed in Maxent. The average across five replicates is shown both for the present and for the Last Glacial Maximum (LGM) for each species. A, *Prophysaon andersoni*, current. B,

P. andersoni, LGM. C, *Prophysaon dubium*, current. D, *P. dubium*, LGM. E, *Prophysaon coeruleum*, current. F, *P. coeruleum*, LGM. G, *Prophysaon vanattae* and *Prophysaon humile*, current. H, *P. vanattae* and *P. humile*, LGM.

Figure S4. Gene tree estimated in GARLIv2.0, with bootstrap support for major groups.

Figure S5. Majority rule consensus tree including all compatible groups. Bootstrap replicates were constructed in GARLIv2.0, and the consensus tree was computed in PAUP* v.4.0.

SHARED DATA

All sequences are available on GenBank (accession numbers: MH324506–MH324729). Scripts are available from github (<https://github.com/meganlsmith/>).