Comparative Phylogeography of Pacific Northwest Gastropods

A Dissertation

Presented in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

with a

Major in Bioinformatics & Computational Biology

in the

College of Graduate Studies

University of Idaho

by

Andrew Michael Rankin

Major Professor: Jack Sullivan, Ph.D.

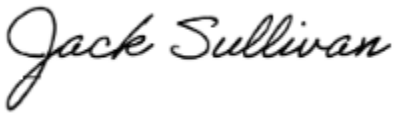
Committee Members: Paul Hohenlohe, Ph.D.; Lisette Waits, Ph.D.; Chris Remien, Ph.D.

Department Administrator: Dave Tank, Ph.D.

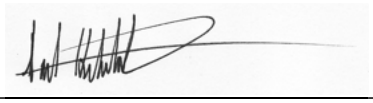
August 2021

Authorization to Submit Dissertation

This dissertation of Andrew Michael Rankin, submitted for the degree of Doctor of Philosophy with a Major in Bioinformatics & Computational Biology and titled "Comparative Phylogeography of Pacific Northwest Gastropods," has been reviewed in final form. Permission, as indicated by the signatures and dates below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor: 

_____ Date: 5/17/2021
Jack Sullivan, Ph.D.

Committee Members: 

_____ Date: 5/17/2021
Paul Hohenlohe, Ph.D.

_____ Date: 5/17/2021
Lisette Waits, Ph.D.

_____ Date: 5/17/2021
Chris Remien, Ph.D.

Department Administrator: 

_____ Date: 5/18/2021
David Tank, Ph.D.

Abstract

Comparative phylogeography is the study of the effects of biogeography and evolutionary history on the spatial distribution of genetic variation of codistributed species. This approach allows investigation of the links between population processes and regional patterns. Here, I present three comparative phylogeographic studies, each focused on different assemblages of gastropod species that are endemic to the temperate rainforests of the North American Pacific Coast and Northern Rocky Mountain interior regions. Slug and snail endemism in these areas is high and several factors are likely responsible for generating the high levels of diversity. The results suggest that codistributed species likely exhibit distinct patterns that reflect the unique aspects of species' phylogeographic histories, which emphasizes that phylogeographic structure is complex and is shaped by more than just a few abiotic factors. In addition, the results presented here significantly increase the taxonomic and natural history knowledge of several rare gastropod lineages, which is important because there is an increasing need to obtain data on the genetic and spatial structure of endemic taxa in areas such as the Pacific Northwest.

Acknowledgements

I thank my major professor, Jack Sullivan, for his patience and support throughout my graduate career. I also thank my committee members Paul Hohenlohe, Lisette Waits, Chris Remien, and my lab rotation supervisor, Christine Parent. I am also glad to have had the support and advice of Anahí Espíndola, Mason Linscott, and all the professors and students of the BCB program. I also want to mention my previous academic advisors, Kate Teeter, Kurt Galbreath, and Chris Floyd. Lastly, special thanks to Michael Lucid and all the employees and volunteers of The Multi Species Baseline Initiative.

Dedication

To my Father, Michael Steven Rankin.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

The geographic distribution of organisms has long been recognized as having implications for understanding evolution (Darwin, 1859; Wallace, 1876). This was eventually extended to the study of geographic distributions of intraspecific genetic lineages (i.e., phylogeography; Avise et al., 1987). Complementary to this is the comparison of geographic patterns of genetic variation in codistributed species to make regional assessments of biodiversity (comparative phylogeography; Zink, 1996; Bermingham & Moritz, 1998). Comparative phylogeographic analyses can determine whether species that currently overlap also exhibit congruent phylogeographic patterns, which would indicate that they differentiated in response to the same historical events. Incongruent patterns of differentiation are thus evidence of species having idiosyncratic histories, which could result from differences in migration patterns, molecular evolutionary rates, effective population sizes or generation times, as well as communities having broken up and reformed in different configurations time and again.

Three studies are presented here, with each being conducted in the Pacific Northwest but on different assemblages of endemic gastropods. In Chapter 1, I focus on the disjunct temperate rainforest ecosystem of the Pacific coast and Northern Rocky Mountain (NRM) interior regions and analyze the phylogenetic relationships and historical biogeogaphy of the endemic jumping slug genus, *Hemphillia*. Most phylogeographic studies focused on this disjunct ecosystem have been of single species (or species complexes) but few have analyzed patterns at higher taxonomic levels. The results suggest a complex interplay of ancient vicariance, and more recent speciation events that have shaped the biogeography of *Hemphillia*.

Although many interior rainforest species have disjunct relatives in the Pacific Coast, some have close affinities to taxa in the forests of eastern North America (Emberton, 1994). In Chapter 2, I focus on two NRM endemic snails, *Angusipira kochi occidentalis* and *Anguispira nimapuna*, that are currently separated by their congeners in eastern North American by ~2000 km. Several other taxonomic groups share this disjunct pattern of occurrence in both eastern and western North America (Vieites et al., 2007; Hendrixson & Bond, 2007), and these distributions hint at an originally transcontinental range with subsequent vicariance due to increasing aridity in the interior (Emberton & Roth, 1994). The results show that both *A. k. occidentalis* and *A. nimapuna* represent unique taxa that are both geographically and genetically distinct from their congeners and are part of a larger group of endemic gastropods that display contrasting biogeographic patterns in the NRM inland-rainforest: (1) widespread species with populations distributed allopatrically; (2) narrow-range endemics confined to discrete patches of suitable habitat.

In Chapter 4, I conduct a comparative phylogeographic analysis of three roundback slugs (Arionidae) that are endemic to the NRM inland-rainforest. The study uncovers idiosyncratic patterns of genetic differentiation among *Udosarx lyrata*, *Magnipelta mycophaga*, and *Kootenaia burkei*. Specifically, *U. lyrata* exhibits deep genetic structure across a small spatial area, *M. mycophaga* also exhibits genetic structure but across a much larger spatial area, and *K. burkei* exhibits shallow structure across a large spatial area. The results highlight that codistributed species likely exhibit unique patterns that are shaped by more than just a few abiotic historical factors.

Literature Cited

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, ... & Saunders NC. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annual review of ecology and systematics, 18(1), 489-522.

Bermingham E, & Moritz C. 1998. Comparative phylogeography: concepts and applications. Molecular Ecology, 7(4), 367-369.

Darwin C. 1859. The Origin of Species by Means of Natural Selection. Murray, London, U.K.

Emberton KC. 1994. Polygyrid land snail phylogeny: external sperm exchange, early North American biogeography, iterative shell evolution. Biological journal of the Linnean Society, 52(3), 241-271.

Emberton KC, Roth B. 1994. "Extralimital" Land Mollusks (Gastropoda) from the Deep River Formation, Montana: Evidence for Mesic Medial Tertiary Climate. Proceedings of the Academy of Natural Sciences of Philadelphia 145: 93 –106

Hendrixson BE, Bond JE. 2007. Molecular phylogeny and biogeography of an ancient Holarctic lineage of mygalomorph spiders (Araneae: Antrodiaetidae: Antrodiaetus). Molecular Phylogenetics and Evolution 42: 738 –755.

Vieites DR, Min MS, Wake DB. 2007. Rapid diversification and dispersal during periods of global warming by plethodontid salamanders. Proceedings of the National Academy of Sciences 104: 19903 –19907.

Wallace AR. 1876. The Geographical Distribution of Animals: With a Study of the Relations of Living and Extinct Faunas as Elucidating the Past Changes of the Earth's Surface. London: Macmillan

Zink RM. 1996. Comparative phylogeography in North American birds. Evolution, 50(1), 308-317.

.

Chapter 2: Complex interplay of ancient vicariance and recent patterns of geographical speciation in north-western North American temperate rainforests explains the phylogeny of jumping slugs (*Hemphillia spp.*)

Abstract

The history of the currently disjunct temperate rainforests of the Pacific Northwest of North America has shaped the evolution and diversity of endemics. This study focuses on how geological and climatic perturbations have driven speciation in the area by isolating lineages. We investigated the phylogenetic relationships and historical biogeography of the endemic jumping slugs (genus *Hemphillia*) using a multi-locus phylogeny. We evaluated the spatial distribution and divergence times of major lineages, generated ancestral area probabilities, and inferred the biogeographical history of the genus. Our study revealed eight genetic lineages that formed three clades: one clade consisting of two Coast/Cascade lineages, and two reciprocally monophyletic clades that each contain a Coast/Cascade and two Rocky Mountains taxa. The results of the biogeographical analysis suggest that the ancestral range of the genus occupied Coast/Cascade habitats and then spread across into Northern Rocky Mountain interior habitats with subsequent fragmentations isolating coastal and inland lineages. Finally, there have been more recent speciation events among three lineage pairs that have shaped shallow structures of all clades. We add to our knowledge of the biogeographical history of the region in that we discovered diversification and speciation events that have occurred in ways more complex than previously thought.

Introduction

North-western North America supports temperate rainforests in both its Pacific coast and northern Rocky Mountain interior regions. These forests are dominated by western hemlock (*Tsuga heterophylla*) and western red cedar (*Thuja plicata*) and contain many other endemic plant and animal taxa (e.g., DellaSala, 2011). These two disjunct rainforests are currently separated by xeric habitats of the Columbia Plateau (Graham, 1993, 1999), a flood basalt, shrub-steppe grassland spanning 300 km between the Coast/Cascade ranges and northern Rockies. This plateau formed from successive flows of basalt between 6 and 17 Mya (Tolan et al., 2009), and the Cascades Range uplift between 4 and 7 Mya (Priest, 1990). The latter generated the current rain shadow that further aridified the Columbia Plateau, which reached near-modern conditions by 4 Mya (Ashwill, 1983), and the transformation of the Plateau into a xeric sage-shrub habitat was complete by ~2 Mya (Brunsfeld et al., 2001). The

currently disjunct mesic forests of the Coast/Cascade Mountains and northern Rocky Mountains have thus been spatially isolated since xerification became complete. Subsequently, throughout the Pleistocene (2.6–0.012 Mya), the region was heavily affected by glacial cycles, and the Plateau experienced repeated flooding from proglacial Lake Missoula (Waitt, 1985; O'Connor & Baker, 1992; Booth et al., 2003). Most of the northern portions of the current range of the rainforest were covered by glaciers and smaller alpine glaciers formed in mountains during the Pleistocene.

These dramatic changes in the landscape have strongly affected the diversification patterns of rainforest endemics. Indeed, such taxa exhibit a phylogeographical break (sensu Swenson & Howard, 2005) between the Coast/Cascade and the northern Rockies (reviewed by Brunsfeld et al., 2001). This division has been explained by ancient vicariance caused by the aridification of the Columbia Plateau ecoregion, which has probably been unsuitable for dispersal between mountain ranges throughout the Pleistocene (Brunsfeld et al., 2001; Carstens & Richards, 2007) and into the Holocene. Phylogeographical studies of some regional endemics show Pacific coast and interior populations as being genetically differentiated (Nielson et al., 2001; Carstens et al., 2005; Steele et al., 2005), suggesting a pre-Pleistocene vicariance. Likewise, research on other non-endemic plant (Li & Adams, 1989; Albach et al., 2006) and animal (Demboski & Cook, 2001; Barrowclough et al., 2004; Galbreath et al., 2009; Kerhoulas et al., 2015) species from the region have uncovered phylogeographical divisions between the Coast/Cascade and northern Rockies phylogroups. Conversely, the phylogeography of other taxa suggests the presence of gene flow across the Columbia Basin (e.g., Ruffley et al., 2018), with the disjunction in some species having occurred via post-Pleistocene dispersal (e.g., Carstens et al., 2005; Smith et al., 2017).

For taxa with a long history (pre-Pleistocene) in inland mesic forests, the Pleistocene glacial cycles may have resulted in compartmentalized refugia in the northern Rockies (Brunsfeld et al., 2001). Specifically, isolation would have been promoted by montane glaciers, which were extensive in the Rocky Mountains north-west of the Wyoming Basin during the Last Glacial Maximum (LGM; Porter et al., 1983; Brouillet & Whetstone, 1993). In this scenario, populations of the northern Rockies would be expected to exhibit strong phylogeographical substructure and the presence of divergent lineages associated with different mountain systems. Such patterns have been identified in the inland endemics Constance's bittercress (*Cardamine constancei*; Brunsfeld & Sullivan, 2005), and Rocky Mountain tailed frogs (*Ascaphus montanus*; Nielson et al., 2006, Metzger et al., 2015), where phylogroups are consistent with the existence of a complex glacial refugium with multiple compartments.

To date, there has been considerable biogeographical research in the area, but the taxa previously examined typically have a high dispersal capacity. In contrast, there are many small,

endemic animals that live hidden and isolated, such as terrestrial invertebrates. Many of these organisms have an exceptionally low dispersal capacity, and therefore are good systems to provide additional insight into the biogeographical history of the Pacific Northwest (PNW). Among these invertebrates, mollusks exhibit substantial diversity and endemism within the PNW (Pilsbry, 1948; Frest & Johannes, 1995; Burke, 2013). In particular, many endemic gastropod taxa occur in the disjunct rainforest ecosystems of the PNW (Burke, 2013). Here, we conduct phylogenetic and biogeographical analyses on jumping slugs of the genus *Hemphillia*, a terrestrial gastropod genus that is endemic to temperate rainforests in the PNW. *Hemphillia* slugs are appropriate species for understanding the influence of the landscape on the diversification of species associated with the PNW disjunct rainforests because species are spatially structured and restricted to the region. We aim to unravel the complex interplay between environmental change and geographical speciation in PNW rainforests by using *Hemphillia* as model taxa. To do this, we investigate the origin of the major *Hemphillia* clades in space and time, test whether the timing of lineage diversification occurred in concert with the progression of the PNW ecosystem, and consider phylogeographical concordance with other co-distributed, endemic taxa.

Methods and Materials

Study system

Species of *Hemphillia* (jumping slugs) have traditionally been separated into two species-groups. One group, including *H. burringtoni* (Pilsbry, 1948) and *H. glandulosa* (Bland & Binney, 1872), is formed by smaller-bodied taxa known from western Washington and adjacent parts of western Oregon and western British Columbia (Burke, 2013). A third species within this group, *H. pantherina* (Branson, 1975), is of uncertain status as it was described from a single specimen and is now viewed as not warranting specific recognition (Burke 2015, T. E. Burke, personal communication, 2017). The second group is composed of larger-bodied taxa and includes *H. camelus* (Pilsbry & Vanatta, 1898), *H. danielsi* (Vanatta, 1914), *H. dromedarius* (Branson, 1972), *H. malonei* (Pilsbry, 1917), the newly described *H. skadei* (Lucid et al., 2018) and a suspected new species that closely resembles *H. danielsi* (Kelley et al., 1999; Burke, 2013), hereafter referred to as *H. sp.* Both *H. dromedarius* and *H. malonei* have Coast/Cascades Mountain distributions whereas the other species occur in the interior Rocky Mountain forests of south-eastern British Columbia, north-eastern Washington, western Montana, and northern Idaho.

Taxon sampling and molecular data

We obtained data from 200 *Hemphillia* specimens gathered from field or museum collections (Fig. 2.1; *H. burringtoni* = 20, *H. glandulosa* = 13, *H. camelus* = 43, *H. danielsi* = 18, *H. dromedarius* = 14, *H. malonei* = 14, *H. skadei* = 32 and *H. sp.* = 46). Field personnel preserved specimens in 70% ethanol and identified all ethanol-preserved specimens based on external morphological characters and geographical range following Burke (2013). Museum specimens were identified in a similar manner by their respective collectors. Total DNA was extracted from the foot of each specimen (N = 156; 10–15 mg) using the DNeasy Blood and Tissue Kit (Qiagen) per the manufacturer's protocols. Partial sequences of the mitochondrial cytochrome *c* subunit I (*COI*) gene, mitochondrial *16S* rRNA gene, nuclear internal transcribed spacer 1 (*ITS1*) marker and nuclear *actin* gene were amplified by PCR using the primers listed in Table 2.1. All PCRs were carried out in 50-µL reactions containing 3 µL DNA, 37.75 µL water, 5 µL buffer, 1 µL of 25 mm MgCl2, 1 µL of 10 mm dNTPs (Thermo Fisher Scientific), 1 µL of 10 mm forward and reverse primer, and 0.25 µL of 5 U/µL of Taq polymerase (New England Biolabs). PCR amplification consisted of an initial denaturation step at 95 °C for 2 min, followed by 30 cycles of a denaturation step at 95 °C for 35 s, an annealing step (Table 2.1) for 60 s and an elongation step at 72 °C for 45 s, and a final extension step at 72 °C for 5 min. Amplicons were electrophoresed in a 1% agarose gel to verify the amplifications and cleaned using the Qiaquick PCR cleanup kit (Qiagen). Bi-directional DNA sequencing was carried out by Eurofins (eurofinsgenomics.com) MWG Operon using the ABI Big Dye Terminator kit (v.3.1) and an automated DNA sequencer (model ABI 3730 XL). Sequences produced by Eurofins were visually examined and edited with Chromas v.2.6.2 (Technelysium, http://www.technelysium.com.au/chromas.html). Consensus sequences were then produced from both forward and reverse strands. To these data we added a set of mitochondrial (mtDNA) sequences from an additional 44 *Hemphillia* specimens generated according to the methods described by Wilke & Duncan (2004). Multiple sequence alignments were constructed for each locus separately using MAFFT online (http://www.ebi.ac.uk/Tools/msa/mafft/). In the *16S* and *ITS1* data sets, many regions were too divergent to be aligned across lineages, and therefore we used the Gblocks algorithm (Castresana, 2000; http://molevol.cmima.csic.es/castresana/Gblocks_server.html) to eliminate ambiguous regions and extract the conserved regions for subsequent analysis. No indels or premature stop codons were found in the *COI* and *actin* protein coding genes.

Phylogenetic inference

We conducted preliminary analyses to reconstruct gene trees for the four markers to assess potential incongruence. Each of the four separate data alignments was subjected to maximum-

likelihood (ML) phylogenetic estimation. We used the automodel command in PAUP* v.4.0a152 (preview release; Swofford, 2003) to select a model of nucleotide sequence evolution using the Bayesian Information Criterion and decision theory (Minin et al., 2003). The GTR+I+Γ model was specified for *COI*, GTR+Γ for *16S*, JC+I for *ITS1* and K2P+I+Γ for *actin*. ML analyses were performed in Garli (Zwickl, 2006) using default parameters, and each ML tree was first determined by conducting ten replicate runs with random starting trees. Node support was assessed using 100 bootstrap replicates with two tree searches per bootstrap. We used the resulting ML phylogenies to test the assumption that each data set has evolved in a clock-like fashion by testing for a global molecular clock in PAUP* v.4.0a152 using the likelihood-ratio test (LRT) of Felsenstein (1988). As the strict clock model was rejected, the relaxed clock model was used for subsequent analyses (Drummond et al., 2006).

Moreover, because nuclear data could not be obtained for specimens from all sites, we identified the major mitochondrial clades in the genus *Hemphillia*. We concatenated the mitochondrial *COI* and mitochondrial *16S* data sets (*COI+16S*) and used ML inference to identify the diversification branching pattern among *Hemphillia* species. Eight specimens representing the genus *Prophysaon*, another endemic slug of the PNW, were used as outgroups (*P. andersoni*: AY357610/AY357657, *P. coeruleum*: AY357617/AY357664, *P. dubium*: AY357611/AY357658, *P. foliolatum*: AY357612/AY357659, *P. humile*: AY357613/AY357660, *P. obscurum*: AY357614/AY357661, *P. sp.*: AY357616/AY357663 and *P. vanattae*: AY357615/AY357662) (Wilke & Duncan, 2004). ML analysis was conducted in Garli using the methods described above but with the data set partitioned by gene.

Species-tree inference

We performed species-tree inference under a multi-species coalescent model with *BEAST (Heled & Drummond, 2009) using the *COI, 16S, actin* and *ITS1* data sets (*COI+16S+actin+ITS1*), implemented in BEAST 2.4.4 (Bouckaert et al., 2014). The data matrix was partitioned by gene with unlinked substitution and clock models, unlinked *ITS1* and *actin* trees, but linked *COI* and *16S* trees. We used a relaxed lognormal molecular clock, a birth–death speciation tree prior, and a linear and constant root model for population size prior. One specimen each from the PNW endemic slug species *Zacoleus idahoensis* and *Magnipelta mycophaga* were used as outgroups. Because there are several published substitution rate estimates for both the *COI* and the *16S* gene in terrestrial gastropods and given a lack of time calibration for the taxa, we applied a range of mtDNA substitution rates to estimate divergence times. The rates for the *COI* gene in terrestrial gastropods are reported to vary between 2.8 × 10–8 and 1.3 × 10–7 substitutions/site/year (Van Riel et al., 2005), and

rates for *16S* are reported to vary between 1.6 × 10–8 and 1.29 × 10–7 substitutions/site/year (Thomaz et al., 1996; Chiba, 1999; Van Riel et al., 2005). Therefore, we estimated the timing of cladogenetic events by applying a normally distributed rate prior truncated to 0 and 0.2 for both *COI* and *16S*, with a mean *COI* site substitution rate of 0.08 per million years (SD: 0.03), and a mean *16S* site substitution rate of 0.07 per million years (SD: 0.04). The 95% interval of these distributions included all the values reported above. The analysis consisted of 500 million generations with a sampling interval of 50 000 and a burn-in of 25%. The BEAST output was analyzed using Tracer v.1.4 (Rambaut & Drummond, 2007) to verify an effective sample size exceeding 200 for all parameters being estimated. The BEAST tool TreeAnnotator was used to produce a median branch length maximum clade credibility tree from the post-burn-in trees.

Simultaneous divergence test

We tested a null hypothesis of simultaneous divergence under a hierarchical Approximate Bayesian Computation (hABC) approach as implemented by the PyMsBayes package (Oaks, 2014). Specifically, we tested whether divergence happened synchronously in four species pairs that represent phylogenetically related disjunct taxa (*H. danielsi*/*H. dromedarius*, *H. sp.*/*H. dromedarius*, *H. camelus*/*H. malonei* and *H. skadei*/*H. malonei*). The PyMsBayes program implements a modified version of msbayes (Huang et al., 2011) that specifies a Dirichlet-process prior (dpp-msbayes) over the hyperprior specifying the number of divergence events (Oaks et al., 2013). The dpp-msbayes model can use multiple loci to infer the temporal pattern of divergence across species pairs by comparing summary statistics among empirical and simulated data sets. We used both mitochondrial (*COI* and *16S*) and nuclear (*actin* and *ITS1*) loci together in the dpp-msbayes model to examine temporal congruence of divergence times between the four aforementioned species pairs. Our *BEAST divergence time estimates guided prior selection for dpp-msbayes as follows: concentration parameter of the Dirichlet process hyperprior~gamma [1000, 0.00055] such that there was an equal prior probability of one (TdivA = TdivB) or two (TdivA > TdivB or TdivA < TdivB) divergence events, population-scaled mutation rate (θ)~gamma [1, 0.0082], divergence times (τ)~gamma [1, 0.036], and the transition-to-transversion rate ratio (kappa) of the HKY substitution model, implemented in dpp-msbayes, was estimated for each alignment separately using PAUP*. We performed 1,000,000 simulations and retained the 1000 simulations with the best fit to the empirical data to estimate posterior parameter values.

Historical biogeographical analyses

We used BioGeoBEARS (Matzke, 2013) to conduct a historical biogeographical analysis in R (R Core Team, 2013). This package estimates ancestral geographical ranges using a time-calibrated phylogeny and current ranges, under an ML framework. Model testing was then performed to determine the fit of alternative biogeographical models. We used and compared two biogeographical models implemented in BioGeoBEARS to determine their fit to our data: (1) dispersal–extinction–cladogenesis (DEC; Ree et al., 2005; Ree & Smith, 2008) and (2) DEC+j. The DEC model focuses on vicariance, or allopatric speciation due to separation of the geographical range (when an ancestor with distribution ABC splits into two distributions A and BC) and allows for sympatric speciation (when an ancestor with a distribution ABC splits into two distributions A and ABC; Ronquist & Sanmartín, 2011). The DEC model has two free parameters that specify the rate of range expansion (d = dispersal) and range contraction (e = extinction), whereas the DEC+j model adds a third free parameter (j = jump dispersal) that corresponds to founder-event speciation (Matzke, 2014). The j parameter was initially implemented for island systems, in which new lineages may be established by colonization of a new island without a continuous ancestor (Clark et al., 2008). To infer ancestral ranges at internal nodes of the *Hemphillia* phylogeny, we used a pruned version of the multilocus species tree from *BEAST containing only in-group taxa. We coded each *Hemphillia* species as being present or absent in the PNW coastal and PNW interior region. The maximum range size, which limits the number of areas by tips and nodes, was set to two.

Species distribution modelling

To obtain an independent perspective on the distribution and divergences for *Hemphillia* species, we developed species distribution models (SDMs) for each species based on georeferenced locality data (Supplementary Information, Appendix S1) and current climate data. This allowed us to assess how closely the predicted habitat suitability of individual species reflected actual occurrence and to estimate the range of each species. Projections were based on a subset of seven uncorrelated (Pearson's correlation coefficient < |0.70|) standard bioclimatic parameters to include in models to describe the suitable climate of each species and to allow for comparisons across species. Variables included the following: BIO3 = Isothermality, BIO5 = Maximum Temperature of Warmest Month, BIO6 = Minimum Temperature of Coldest Month, BIO7 = Temperature Annual Range, BIO13 = Precipitation of Wettest Month, BIO14 = Precipitation of Driest Month and BIO15 = Precipitation Seasonality. We obtained these seven bioclimatic data layers for current (1950–2000) conditions from the WorldClim database at a resolution of 30″ (Hijmans et al., 2005) and cropped the study area to be between 100° and 180°W and 32° and 72°N. Setting a working area polygon to include only western

North America was to avoid sampling habitat greatly outside the species' known occurrences for the selection of background points, which are meant to be compared with the presence data and differentiate the environmental conditions which the species can potentially occur. Nine modelling methods were used to calculate separate SDMs: Generalized Linear Model (GLM), Boosted Regression Trees (GBM), Generalized Additive Model (GAM), Classification Tree Analysis (CTA), Artificial Neural Network (ANN), Surface Range Envelop (SRE), Flexible Discriminant Analysis (FDA), Multiple Adaptive Regression Splines (MARS) and Random Forest (RF). The relative contributions of these alternative models to the final combined model were weighted using the area under the receiver operating characteristic curve (AUC). All analyses were run using the R (R Core Team, 2013) package biomod2 (Thuiller et al. 2009), and for each method we used the default settings.

Results

Phylogenetic and species-tree estimation

Inference of individual gene trees indicated no conflict between mitochondrial gene trees and no strongly supported incongruence between mitochondrial and nuclear gene trees. Each genealogy revealed eight genetic clusters that correspond to species assignments. For ML analysis of the concatenated mitochondrial data set (Fig. 2.2), the deepest split within *Hemphillia* was inferred to be between the smaller bodied *H. burringtoni* (bootstrap support of 70) and *H. glandulosa* (97) (Clade I; 72) and the remaining larger bodied *Hemphillia* species (80). Within the larger bodied species-group, there were two clades (Clades II and III), each with three species. Within Clade II (93), the coastal species *H. dromedarius* (100) is sister to the reciprocally monophyletic *H. danielsi* and *H. sp.* (56), both inland species (100 and 100, respectively). Clade II is sister to Clade III (72), which contains the coastal species *H. malonei* (100) as sister to the reciprocally monophyletic *H. camelus* and *H. skadei* (88), both inland species (99 and 100, respectively). Relative to *H. burringtoni* and *H. glandulosa*, the six large-bodied species show similar, short intraspecific branch lengths. In contrast, branch lengths within *H. glandulosa* and *H. burringtoni* are half the total tree depth within the genus.

The *BEAST species-tree based on both mitochondrial and nuclear markers (Fig. 2.3) recovered branching patterns that were identical to those of the ML mtDNA analysis, although Bayesian nodal support was higher. One notable observation is that both reconstructions group *H. dromedarius*, *H. danielsi* and *H. sp.* together with high support (i.e., Clade II is well supported), but there is low support for the *H. danielsi*/*H. sp.* clade (bootstrap support of 56 and posterior probability of 0.71). Based on assumed substitution rates for *COI* and *16S* (see above), the analysis converged on a mean rate of 0.13 [95% highest posterior density (HPD): 0.084–0.17] substitutions/site/Myr for the

COI locus, 0.037 (95% HPD: 0.023–0.052) for *16S*, 0.014 (95% HPD: 0.0081–0.02) for *actin* and 0.012 (95% HPD: 0.007–0.017) for *ITS1*. The analysis suggests that *Hemphillia* is a relatively ancient lineage, with the deepest split between the large-bodied *Hemphillia* species (Clades II and III) and the small-bodied *H. burringtoni*/*H. glandulosa* (Clade I) placed at about 4.54 Mya (95% HPD: 3.05–6.19 Mya), and the split between *H. burringtoni* and *H. glandulosa* (Clade I) placed at 2.6 Mya (95% HPD: 1.79–3.56 Mya). Within the large-bodied species group, Clades II and III display a split dated at 3.2 Mya (95% HPD: 2.14–4.36 Mya). *Hemphillia dromedarius* split from *H. danielsi*/*H. sp.* 2.09 Mya (95% HPD: 1.27–3 Mya), and *H. danielsi*/*H. sp.* split 1.58 Mya (95% HPD: 0.9–2.33 Mya). *Hemphillia malonei* split from *H. camelus*/*H. skadei* around 2.44 Mya (95% HPD: 1.49–3.49 Mya), while the latter two split 1.5 Mya (95% HPD: 0.75–2.22 Mya).

Simultaneous divergence test

We used dpp-msbayes to test for simultaneous divergence across the Columbia Plateau by estimating divergence times for four species pairs (*H. danielsi*/*H. dromedarius*, *H. sp.*/*H. dromedarius*, *H. camelus*/*H. malonei* and *H. skadei*/*H. malonei*) and estimating the posterior probability of the number of divergence episodes. Results from analysis with dpp-msbayes support synchronous diversification (Fig. 2.4). The posterior probability for a single divergence was 0.59, whereas the next best supported scenario (two divergence events) received a posterior probability of 0.068, suggesting that Clades II and III began diversifying at approximately the same time. This inference is also supported by the broadly overlapping HPD intervals in the topology of the species tree (Fig. 2.3).

Historical biogeography

The species-tree inference from *BEAST yielded the overall highest likelihood for the DEC+j model in our BioGeoBEARS analysis [weighted corrected Akaike Information Criteria (AICc): 0.928; Table 2.2], with much less weight given to the DEC model (AICc: 0.0725; Table 2.2). However, Ree & Sanmartin (2018) stressed that the likelihoods of the DEC and DEC+j models are not statistically comparable, and because the latter model favors direct dispersal over widespread ranges owing to the assumption of extremely low extinction rates, the DEC+j model may not be adequate for reconstructing the history of older lineages (Sanmartin & Meseguer, 2016). We therefore consider the results from both the DEC and DEC+j models.

The branching pattern of the species-tree identifies one of the two primary branches leading exclusively to coastal species, whereas the second primary branch leads to both coastal and inland species. Given this, the area of the common ancestor of the genus *Hemphillia* is estimated as

primarily in the coast for both DEC and DEC+j (Fig. 2.3) inferences. The coastal region is also the most likely range for the ancestor of Clades II and III when considering the DEC+j model; however, DEC inference slightly favors a widespread range for this node, with the ancestral lineage having a distribution spanning the coast and inland (CI). Furthermore, in the DEC model, the ancestor of Clade II and the ancestor of Clade III show high support for a widespread range, which posits that the split between *H. dromedarius* and *H. danielsi*/*H. sp.* and the split between *H. malonei* and *H. camelus*/*H. skadei* occurred when an ancestral lineage with a distribution spanning the coast and inland (CI) split into two lineages with distribution C and I (widespread vicariance). However, the added jump-dispersal parameter emphasized by the DEC+j model suggests that long-distance dispersal/founder-event speciation may have had some important effects in obtaining the current disjunct distribution without the presence of widespread ancestors. That is, conditional on this model, there were two eastward colonization and founder effect speciation events driving the initial diversification of Clades II and III.

Species distribution modelling

The projection of SDMs onto current climatic conditions shows that areas of high occurrence probability include moist forest stands, riparian and other wet, cool areas in the Cascade/Coastal Mountains and northern Rocky Mountain wet-belt ecosystem (Fig. 2.5). Sister species *H. burringtoni* and *H. glandulosa* exhibit a parapatric distribution with contact zones (Fig. 2.1) but more overlapping SDMs. For species of Clade II, the SDM for *H. dromedarius* closely matches its known distribution (Fig. 2.1), but sister species *H. danielsi* and *H. sp.* show SDMs that are broader than their current distributions (Fig. 2.1). The two species are known from the Clearwater and Salmon River watersheds, Idaho, and part of the Bitterroot Range, but show some occurrence probability across the Blue-Wallowa Mountains and Central Oregon Highland Mountains. In addition, the predicted SDM of *H. sp.* appears nested within that of *H. danielsi*. For species of Clade III, the SDM for *H. malonei* corresponds well with its known distribution (Fig. 2.1), whereas the SDMs of *H. camelus* and *H. skadei* show areas of high probability corresponding to their distribution as well as coastal/Cascade habitat. Lastly, the SDM of *H. skadei* is nested within that of *H. camelus*, similar to the observation between *H. danielsi* and *H. sp.*; however, evidence suggests that *H. camelus*/*H. skadei* are largely allopatric while *H. danielsi* and *H. sp* are largely sympatric (Fig. 2.1)

Discussion

Regional biogeography

The temperate rainforests of the Pacific coast and northern Rocky Mountain interior regions offer a compelling opportunity to study the impact of a disjunct ecosystem on structuring biotic diversity. An ancient vicariance hypothesis has been posited to explain the existence of extant, disjunct taxa in both the Coast/Cascade and northern Rocky Mountains components (Brunsfeld et al., 2001). Specifically, under this hypothesis, disjunct taxa are the result of formerly contiguous distributions that were split by the xerification of the Columbia Plateau associated with orogenesis of the Cascades. Here, we investigated how separation of the coastal and inland rainforests has shaped phylogenetic diversity for a genus of terrestrial slugs with putatively limited dispersal ability and a wide distribution in the PNW, to gain insights into the biogeographical history of the ecosystem.

Palaeontological data suggest that coniferous forests have been present in the northern Rocky Mountains since the mid-Eocene (Graham, 1993, 1999), or since the formation of the northern Rocky Mountains 45–36 Mya (English & Johnson, 2004). Thus, prior to uplift of the Cascades, a presumably continuous coniferous forest habitat stretched across the PNW. Establishment of the Columbia Plateau xeric habitat following uplift of the Cascades is thought to have fragmented this habitat, leading to the current disjunct range of the rainforest. Within this general historical framework, our results support the ancient vicariance hypothesis for the general coastal/inland diversification of the group in the sense that the divergences between coastal and inland types are older, substantially pre-dating possible post-Pleistocene dispersal (Brunsfeld et al., 2001). However, in contrast to other animals (e.g., *Ascaphus montanus*/*A. truei*, Nielson et al., 2001; *Plethodon idahoensis*/*P. vandykei*, Carstens et al., 2004), the structure in *Hemphillia* is more complex in that there are replicated ancient vicariance events in different lineages with subsequent inland speciation events within each lineage. The deep pre-Pleistocene structure characterized by the three major phylogenetic groups (Clades I, II and III), as well as shallow structure shaping those individual clades, indicates that there were multiple occurrences of range expansion/fragmentation or perhaps multiple long-distance dispersal/founder-event speciation events that structured diversity.

The two deepest phylogenetic breaks (between Clades I, II and III) are older than 2.14 Mya, and probably reflect Pliocene events. The most recent common ancestor of *Hemphillia*, which probably existed from 3 to 6 Mya, probably had a Pacific coast distribution (Fig. 3), and after the initial diversification of the group, the DEC model suggests that the ancestor of the larger-bodied species group spread across to the Northern Rocky Mountains (NRM) interior region such that there were contiguous populations from the Pacific coast to the interior during the Pliocene. The succeeding divergence of Clades II and III—the two reciprocally monophyletic large-bodied

Hemphillia species groups—is interesting because to our knowledge a similar phylogeographical pattern this old has not been observed in previously studied taxa from the region. However, a latitudinal split between northern populations still connected by the Okanogan Highlands (in north-central Washington) and southern populations connected via the Central Oregon Highlands may explain this pattern. Incidentally, the SDMs for both *H. danielsi* and *H. sp.* show areas of suitability in the Central Oregon Highlands (Fig. 2.5), although they are not currently known to occur in the region. The climatic influences of the early Cascades and associated loss of suitable habitat connecting the coastal and inland rainforest ecosystems probably split contiguous populations of Clades II and III into eastern and western groups, which would then have retracted to coastal and inland distributions.

The model with the highest likelihood in our BioGeoBEARS analysis was the DEC+j model, which received nearly all the weighted model support (as measured using AICc weights). Taken at face value, the resulting inference would suggest that long-distance dispersal/founder-event speciation played an important role in the invasions of the interior region, although this inference potentially contrasts with characteristics of *Hemphillia*, such as their sedentary nature and low vagility. However, despite their low vagility, there are reports of mollusc specimens or their eggs being dispersed by birds (e.g., Pearce et al., 2012; Shikov & Vinogradov, 2013), and rare colonization events such as being moved across the Columbia Basin by birds to the NRM interior forests may be possible, especially given the hermaphroditic nature of these organisms. Notably, other terrestrial gastropods (e.g., members of the tail-dropper genus *Prophysaon*) lack deep genetic structure across the Columbia basin (e.g., Wilke & Duncan, 2004; Smith et al., 2017, 2018), suggesting that there have been opportunities for gene flow across the disjunction. One possibility is that individuals dispersed along river corridors such as the Columbia River drainage basin, perhaps during large-scale flooding episodes that followed deglaciation, although such dispersal would have to have run against flood currents (generally NW to SE). Both river valleys and other events such as bird dispersal provide possible mechanisms for jump-dispersal/founder event speciation to have played an important role in colonization of the NRM interior forests by Clades II and III. However, given the broadly overlapping HPD intervals in the topology of the species tree (Fig. 2.3) and high posterior probability support for a single divergence event from the dpp-msbayes analysis (Fig. 2.4), two independent and simultaneous long-distance dispersal events seem an unlikely mechanism of diversification, and it is more plausible that simultaneous divergence was driven by an environmental mechanism, such as xerification of the Columbia Plateau. Thus, we favor the inferences derived under the DEC model over those from the DEC+j model, following the arguments of Ree & Sanmartin (2018). A similar reasoning has been invoked to explain comparable disjunction patterns in clades of amphibians (e.g., *Ascaphus montanus*/*A. truei*, Nielson et al., 2001; *Plethodon idahoensis*/*P. vandykei*, Carstens et al.,

2004; and *Dicamptodon copei*/*D. aterrimus*, Carstens et al., 2005). Carstens et al. (2005; using a coalescent-based method) estimated the mean time of divergence between the *Ascaphus*, *Plethodon* and *Dicamptodon* lineages to be 3.1, 4.1 and 1.2 Mya, respectively. Likewise, here (using a tree-based approach) we estimated the splits between coastal and interior sister groups of *Hemphillia* (*H. dromedarius* vs. *H. danielsi*/*H. sp.* and *H. malonei* vs. *H. camelus*/*H. skadei*) to be 2.09 Mya (95% HPD: 1.27–3 Mya) and 2.44 Mya (95% HPD: 1.49–3.49 Mya). These lineages may well have responded to Pliocene (5.33–2.49 Mya) drought in a similar manner. The xerification of the Columbian Plateau was probably not a discrete event but instead a long-term gradual drying, and therefore sister-groups may not have been strictly allopatric but perhaps parapatric with potential areas of contact. Ongoing hybrid zones may have slowed the process of lineage sorting to post-date the ecological separation of the Coast–Cascade mountains and Northern Rocky Mountains.

A corollary of the ancient vicariance hypothesis is that populations became isolated in both the Cascades and northern Rocky Mountains after uplift of the Cascades. This suggests that isolates persisted in coastal and inland refugia throughout the Pleistocene (2.6–0.012 Mya) glaciations, and until the present (Brunsfeld et al., 2001) (i.e., the most recent speciation events in Clades II and III probably reflect Pleistocene glacial events shaping the structure of those individual clades). During the Pleistocene glacial cycles, most of the PNW was subjected to repeated glaciation. Only the northernmost portions of the Cascades Range were affected, with montane glaciers in the Cascades probably pushing forest habitats to lower altitudes, and ice sheets covering a significant portion of the northern Rocky Mountains (Pielou, 1991; Delcourt & Delcourt, 1993). Nevertheless, the inland rainforest ecosystem contains a collection of pre-Pleistocene endemic species (i.e., old endemics), including Constance's bittercress (*Cardamine constancei*; Brunsfeld & Sullivan, 2005), Coeur d'Alene salamanders (*Plethodon idahoensis*; Carstens et al., 2004), Rocky Mountain tailed frogs (*Ascaphus montanus*; Nielson et al., 2006) and Idaho giant salamanders (*Dicamtodon aterrimus*; Steele et al., 2005), which are thought to have persisted through the Pleistocene glacial cycles in one or more inland refugia. In *Hemphillia*, the depth of phylogenetic divergence between the sister species *H. camelus* and *H. skadei*, as well as the divergence between *H. danielsi* and *H. sp.*, suggests that inland *Hemphillia* species have also persisted in the region throughout the Pleistocene climatic fluctuations. *Hemphillia danielsi* and *H. sp.* appear to be deeply divergent sister species; however, they do not appear to have differentiated spatially (Fig. 2.5) with their distributions centered around the Clearwater River drainage, a historically non-glaciated part of the inland ecosystem that has been a suspected refugium for many mesic forest endemics (Daubenmire, 1975; Carstens et al., 2004; Brunsfeld & Sullivan, 2005). Similarly, *H. camelus* and *H. skadei* are also well-separated sister species that nevertheless are predicted to have overlapping but not equal spatial distributions

(Supporting Information, Fig. S5). However, the distribution of the two species is predominately allopatric, suggesting some form of ecological exclusion. *Hemphillia skadei* has been found in the Coeur d'Alene, Saint Joe and Selkirk mountains in northern Idaho, and its range appears to be nested within that of *H. camelus* (Lucid et al., 2018). The latter occurs directly south of populations of *H. skadei*, but also in previously glaciated areas of the Selkirk and Purcell mountains of northern Idaho and the surrounding regions. Lucid et al. (2018) provide additional information on the geographical association among the subclades of *H. camelus* and *H. skadei*. It is likely that these northern areas were colonized through western Montana, surrounding the range of *H. skadei*, a pattern that has also been found for the Rocky Mountain tailed frogs (*Ascaphus montanus*; Metzger et al., 2015). The presence of more northerly, largely allopatric populations (Fig. 2.1) suggests the possibility of northern refugia in these species. It is possible that more northerly refugia occurred along other river canyons, such as the St. Joe or Coeur d'Alene rivers of northern Idaho, a hypothesis also supported by the phylogeographical structure identified in Constance's bittercress (*Cardamine constancei*; Brunsfeld & Sullivan, 2005). Our results thus provide more support for the presence of multiple, compartmentalized refugia within the northern Rockies (Brunsfeld et al., 2001), and further demonstrate the complexity of the biogeographical patterns and structure observed in northern Rocky Mountain endemics (reviewed by Shafer et al., 2010).

Within the Coast/Cascade mountains, there was a much greater extent of unglaciated habitat during the Pleistocene (Brunsfeld et al., 2001). However, *H. glandulosa* and *H. burringtoni* separate into two phylogroups that parse according to geography and appear to have separated during the late Pliocene/early Pleistocene. The current distribution of the two species is parapatric (Fig. 1), although they show very similar SDM projections (Fig. 2.5). It is possible that these two groups were isolated in separate Coast Mountain refugia and Cascades Mountain refugia during the Pleistocene and have only recently come into secondary contact. Individuals of the *H. burringtoni* clade occur on Vancouver Island, south throughout the Olympic Peninsula and western Washington, and along the north-eastern Oregon coast while individuals of the *H. glandulosa* clade occur from the central coast of Washington to the south-western Washington Cascades and north-western Oregon, and the two species come into contact in some areas of Washington and Oregon (Wilke & Ziegltrum, 2004; Burke et al., 2005). Further information on spatial population structure of *H. glandulosa* and *H. burringtoni* is *Discus*sed in Wilke & Ziegltrum (2004).

Lastly, members of the genus *Hemphillia* are often recognized as species of conservation importance in the states and provinces in which they occur (e.g., IDFG, 2017). This is due to their endemism and limited distribution, and a lack of taxonomic and natural history knowledge (IDFG, 2017). Our phylogenetic analyses help to clear some previous taxonomic difficulties of the group. For

example, our large-scale sampling shows that *H. glandulosa* and *H. burringtoni* separate into two phylogroups (see also Wilke & Ziegltrum, 2004), and therefore represent two genetically distinct species that parse according to geography. *Hemphillia* camelus and *H. skadei* appear to be deeply divergent sister species, even though *H. skadei* has long been treated as *H. camelus* (Burke, 2013) due to their highly similar external morphologies, and has only recently been described (Lucid et al., 2018). Our results also confirm that the suspected new taxon (*H. sp.*) is genetically distinct from the morphologically similar, sympatric species *H. danielsi* (Burke, 2013). However, there do appear to be morphological differences between the genitalia of the two species (Fig. 2.6). The biogeographical information we have detailed in this study will help guide managers to appropriately allocate resources for species conservation, and the support we document for previously suggested refugia will help guide land management to conserve putative biodiversity hotspots.

Conclusion

Many studies of phylogenetic concordance involving taxa endemic to the PNW show clear genetic breaks between coastal and inland populations, as well as evidence of multiple refugia in both coastal and inland regions throughout the Pleistocene glaciations. Our molecular data assembled for individuals of the endemic slug genus *Hemphillia* show elements of both ancient and shallow biogeographical patterns, suggesting that the biogeographical structure of taxa such as this one is more complex than seen in others from the region. For example, recent studies on other PNW invertebrates with limited dispersal abilities and wide distributions have shown either shallow divergence between coastal and inland populations (*Prophysaon* slugs; Wilke & Duncan, 2004; Smith et al., 2018), or genetic division between – but not within – coastal and inland populations (*Chonaphe millipedes*; Espíndola et al., 2016). Our data suggest that *Hemphillia* experienced replicated ancient vicariance events in two separate lineages, as well as more recent speciation events in both the coast and inland regions. We posit that late Pliocene and Pleistocene climatic oscillations, in conjunction with the geological and physiographic heterogeneity of the coastal and inland mesic forests, promoted allopatric diversification between the sister species pairs *H. burringtoni*/*H. glandulosa*, *H. camelus*/*H. skadei* and *H. danielsi*/*H. sp.* Today, two of the lineage pairs (*H. burringtoni*/*H. glandulosa* and *H. camelus*/*H. skadei*) are mostly non-overlapping, while *H. danielsi*/*H. sp.* are largely sympatric (Fig. 2.1). Our study demonstrates that a complex interplay of ancient vicariance and more recent speciation events has shaped the biogeography of *Hemphillia* in north-western North American temperate rainforests.

Literature Cited

Albach DC, Schoenswetter P, Tribsch A. 2006. Comparative phylogeography of the *Veronica alpina* complex in Europe and North America. Molecular Ecology 15: 3269 –3286.

Armbruster GF, van Moorsel CH, Gittenberger E. 2000. Conserved sequence patterns in the non-coding ribosomal ITS-1 of distantly related snail taxa. Journal of Molluscan Studies 66: 570 –573. Ashwill M. 1983. Seven fossil floras in the rain shadow of the Cascade Mountains, Oregon. Oregon Geology 45: 107 –111.

Barrowclough GF, Groth JG, Mertz LA, Gutiérrez RJ. 2004. Phylogeographic structure, gene flow and species status in blue grouse (*Dendragapus obscurus*). Molecular Ecology 13: 1911 –1922.

Bland T, Binney WG. 1874. XVI. —Description of *Hemphillia*, a new genus of terrestrial mollusks. Annals of the New York Academy of Sciences 10: 208 –211.

Booth DB, Troost KG, Clague JJ, Waitt RB. 2003. The Cordilleran ice sheet. Developments in Quaternary Sciences 1: 17 –43.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Computational Biology 10: e1003537.

Branson BA. 1972. *Hemphillia dromedarius*, a new Arionid slug from Washington. Nautilus 85: 100 –106.

Branson BA. 1975. *Hemphillia pantherina*, a new Arionid slug from Washington. The Veliger 18: 93 –94.

Brouillet L, Whetstone RD. 1993. Climate and physiography. In: FONAE Committee, ed. Flora of North America North of Mexico. New York: Oxford University Press, 15 –46.

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. Comparative phylogeography of northwestern North America: a synthesis. In: Silvertown J, Antonovics J, eds. Integrating ecological and evolutionary processes in a spatial context, Vol. 14. Oxford: Blackwell Science, 319 –339.

Burke T. 2015. Examination of *Hemphillia pantherina* type specimen photographs. Portland: US Forest Service Interagency Special Status/Sensitive Species Program.

Burke TE, Leonard WP. 2013. Land snails and slugs of the Pacific Northwest. Portland: Oregon State University Press.

Burke TE, Wainwright M, Duncan N. 2005. Conservation assessment for four species of the genus *Hemphillia*. Oregon and Washington: USDA Forest Service Region 6 and USDI Bureau of Land Management.

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, Richards CL. 2007. Integrating coalescent and ecological niche modeling in comparative phylogeography. Evolution 61: 1439 –1454.

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.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution 17: 540 –552.

Chiba S. 1999. Accelerated evolution of land snails Mandarina in the oceanic Bonin Islands: evidence from mitochondrial DNA sequences. Evolution 53: 460 –471.

Clark JR, Ree RH, Alfaro ME, King MG, Wagner WL, Roalson EH. 2008. A comparative study in ancestral range reconstruction methods: retracing the uncertain histories of insular lineages. Systematic Biology 57: 693 –707.

Daubenmire R. 1975. Floristic plant geography of eastern Washington and northern Idaho. Provo: Brigham Young University Press.

Delcourt PA, Delcourt HR. 1993. Paleoclimates, paleovegetation, and paleofloras during the late Quaternary. In: Flora North America Editorial Committee, ed. Flora of North America, Vol. 1. New York: Oxford University Press, 71 –94.

DellaSala D, Alaback P, Craighead L, Goward T, Pacquet P, Spribille P. 2011. Temperate and boreal rainforests of inland northwestern North America. In: DellaSalla DA, ed. Temperate and boreal rainforests of the world: ecology and conservation. Washington: Island Press, 82 –110.

Demboski JR, Cook JA. 2001. Phylogeography of the dusky shrew, *Sorex monticolus* (Insectivora, Soricidae): insight into deep and shallow history in northwestern North America. Molecular Ecology 10: 1227 –1240.

Drummond AJ, Ho SY, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biology 4: e88.

English JM, Johnston ST. 2004. The Laramide orogeny: what were the driving forces? International Geology Review 46: 833 –883.

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.

Felsenstein J. 1988. Phylogenies from molecular sequences: inference and reliability. Annual Review of Genetics 22: 521 –565.

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.

Frest TJ, Johannes EJ. 1995. Interior Columbia Basin mollusk species of special concern. Final Report Contract #43-0E00-4-9112. Washington: Interior Columbia Basin Ecosystem Management Project.

Galbreath KE, Hafner DJ, Zamudio KR. 2009. When cold is better: climate-driven elevation shifts yield complex patterns of diversification and demography in an alpine specialist (American pika, Ochotona princeps). Evolution 63: 2848 –2863.

Gavin DG, Hu FS, Walker IR, Westover K. 2009. The northern inland temperate rainforest of British Columbia: old forests with a young history? Northwest Science 83: 70 –78.

Graham A. 1993. History of the vegetation: Cretaceous–Tertiary. In: Flora North America Editorial Committee, eds. Flora of North America North of Mexico, Vol. 1. New York: Oxford University Press, 57 –70.

Graham A. 1999. Late Cretaceous and Cenozoic history of North American vegetation. New York: Oxford University Press.

Heled J, Drummond AJ. 2009. Bayesian inference of species trees from multilocus data. Molecular Biology and Evolution 27: 570 –580.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high-resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25: 1965 –1978.

Huang W, Takebayashi N, Qi Y, Hickerson MJ. 2011. MTML-msBayes: approximate Bayesian comparative phylogeographic inference from multiple taxa and multiple loci with rate heterogeneity. BMC bioinformatics 12: 1.

Idaho Department of Fish and Game. 2017. Idaho State Wildlife Action Plan, 2015. Boise: Idaho Department of Fish and Game. Grant No.: F14AF01068 Amendment #1. Available at: http://fishandgame.idaho.gov/. Sponsored by the US Fish and Wildlife Service, Wildlife and Sport Fish Restoration Program. Kelley R, Dowlan S, Duncan N, Burke T. 1999. Field guide to survey and manage terrestrial mollusk species from the Northwest Forest Plan. Portland: USDI Bureau of Land Management.

Kerhoulas NJ, Gunderson AM, Olson LE. 2015. Complex history of isolation and gene flow in hoary, Olympic, and endangered Vancouver Island marmots. Journal of Mammalogy 96: 810 –826.
Li P, Adams WT. 1989. Range-wide patterns of allozyme variation in Douglas-fir (*Pseudotsuga menziesii*). Canadian Journal of Forest Research 19: 149 –161.

Lucid MK, Rankin A, Espindola A, Chichester L, Ehlers S, Robinson L, Sullivan J. 2018. Taxonomy and biogeography of *Hemphillia* (Gastropoda: Pulmonata: Arionidae) in North American rainforests, with description of a new species (Skade's jumping slug, *Hemphillia skadei* sp. nov.). Canadian Journal of Zoology 96: 305 –316.

Matzke NJ. 2013. BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis in R Scripts. R package, version 0.2.1, 2013. http://CRAN.R-project.org/package=BioGeoBEARS. Accessed 5 November 2017.

Matzke NJ. 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. Systematic Biology 63: 951 –970.

Metzger G, Espíndola A, Waits LP, Sullivan J. 2015. Genetic structure across broad spatial and temporal scales: Rocky Mountain tailed frogs (*Ascaphus montanus*; Anura: Ascaphidae) in the inland temperate rainforest. Journal of Heredity 106: 700 –710.

Minin V, Abdo Z, Joyce P, Sullivan J. 2003. Performance-based selection of likelihood models for phylogeny estimation. Systematic Biology 52: 674 –683.

Mumladze L, Tarkhnishvili D, Murtskhvaladze M. 2013. Systematics and evolutionary history of large endemic snails from the Caucasus (*Helix buchii* and *H. goderdziana*) (Helicidae). American Malacological Bulletin 31: 225 –235.

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.

Nielson M, Lohman K, Daugherty CH, Allendorf FW, Knudsen KL, Sullivan J. 2006. Allozyme and mitochondrial DNA variation in the tailed frog (Anura: *Ascaphus*): the influence of geography and gene flow. Herpetologica 62: 235 –258.

O'Connor JE, Baker VR. 1992. Magnitudes and implications of peak discharges from glacial Lake Missoula. Geological Society of America Bulletin 104: 267 –279.

Oaks JR. 2014. An improved approximate-Bayesian model-choice method for estimating shared evolutionary history. BMC Evolutionary Biology 14: 150.

Oaks JR, Sukumaran J, Esselstyn JA, Linkem CW, Siler CD, Holder MT, Brown RM. 2013. Evidence for climate-driven diversification? A caution for interpreting ABC inferences of simultaneous historical events. Evolution: International Journal of Organic Evolution 67: 991 –1010.

Pearce TA, Mulvihill RS, Porter KA. 2012. Land slugs (Gastropoda: Pulmonata) on birds demonstrate dispersal potential. Nautilus 126: 38 –40.

Pielou EC. 1991. After the ice age: the return of life to glaciated North America. Chicago: University of Chicago Press.

Pilsbry HA. 1917. A new *Hemphillia* and other snails from near Mt. Hood, Oregon. The Nautilus 30: 117 –119.

Pilsbry H. 1948. Land Mollusca of North America (north of Mexico), Vol. 2. Philadelphia: The Academy of Natural Sciences of Philadelphia Monographs No. 3.

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.

Porter SC, Pierce KL, Hamilton TD. 1983. Late Wisconsin mountain glaciation in the western United States. Late Quaternary environments of the United States. In: Perter SC, ed. The late Pleistocene, Vol. 1. Minneapolis: University of Minnesota Press, 71 –111.

Priest GR. 1990. Volcanic and tectonic evolution of the Cascade Volcanic arc, central Oregon. Journal of Geophysical Research 95: 19583 –19600.

R Core Team. 2013. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available at: http://www.R-project.org/.

Rambaut A, Drummond AJ. 2007. Tracer v1.4. Available at: http://beast.bio.ed.ac.uk/Tracer.

Ree RH, Sanmartín I. 2018. Conceptual and statistical problems with the DEC+ J model of founder-event speciation and its comparison with DEC via model selection. Journal of Biogeography 45: 741 –749.

Ree RH, Smith SA. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Systematic Biology 57: 4 –14.

Ree RH, Moore BR, Webb CO, Donoghue MJ. 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. Evolution 59: 2299 –2311.

Ronquist F, Sanmartín I. 2011. Phylogenetic methods in biogeography. Annual Review of Ecology, Evolution, and Systematics 42: 441 –464.

Rowson B, Tattersfield P, Symondson WO. 2011. Phylogeny and biogeography of tropical carnivorous land-snails (Pulmonata: Streptaxoidea) with particular reference to East Africa and the Indian Ocean. Zoologica Scripta 40: 85 –98.

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.

Sanmartín I, Meseguer AS. 2016. Extinction in phylogenetics and biogeography: from timetrees to patterns of biotic assemblage. Frontiers in Genetics 7: 35.

Shafer A, Cullingham CI, Cote SD, Coltman DW. 2010. Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. Molecular Ecology 19: 4589 –4621.

Shikov EV, Vinogradov AA. 2013. Dispersal of terrestrial gastropods by birds during the nesting period. Folia Malacologica 21: 105 –110.

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.

Smith ML, Ruffley M, Rankin AM, Espíndola A, Tank DC, Sullivan J, Carstens BC. 2018. Testing for the presence of cryptic diversity in tail-dropper slugs (*Prophysaon*) using molecular data. Biological Journal of the Linnean Society 124: 518 –532.

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.

Swenson NG, Howard DJ. 2005. Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. The American Naturalist 166: 581 –591.

Swofford DL. 2003. PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4. Sunderland: Sinauer Associates.

Thomaz D, Guiller A, Clarke B. 1996. Extreme divergence of mitochondrial DNA within species of pulmonate land snails. Proceedings of the Royal Society of London B 263: 363 –368.

Thuiller W, Lafourcade B, Engler R, Araújo MB. 2009. BIOMOD–a platform for ensemble forecasting of species distributions. Ecography 32: 369 –373.

Tolan TL, Martin BS, Reidel SP, Anderson JL, Lindsey KA, Burt W. 2009. An introduction to stratigraphy, structural geology, and hydrogeology of the Columbia River Flood-Basalt Province—A primer for the GSA Columbia River Basalt Group field trips. In: O'Connor JE, Dorsey RJ, Madin IP, eds. Volcanoes to vineyards—Geologic field trips through the dynamic landscape of the Pacific Northwest: Geological Society of America Field Guide, Vol. 15, 599 –643.

Vanatta EG. 1914. Montana shells. Proceedings of the Academy of Natural Sciences of Philadelphia 66: 367 –371.

Van Riel P, Jordaens K, Van Houtte N, Martins AMF, Verhagen R, Backeljau T. 2005. Molecular systematics of the endemic *Leptaxini* (Gastropoda: Pulmonata) on the Azores islands. Molecular Phylogenetics and Evolution 37: 132 –143.

Waitt RB Jr. 1985. Case for periodic, colossal j ö kulhlaups from Pleistocene glacial Lake Missoula. Geological Society of America Bulletin 96: 1271 –1286.

Walsh HE, Kidd MG, Moum T, Friesen VL. 1999. Polytomies and the power of phylogenetic inference. Evolution 53: 932 –937.

Wilke T, Duncan N. 2004. Phylogeographical patterns in the American Pacific Northwest: lessons from the arionid slug *Prophysaon* coeruleum. Molecular Ecology 13: 2303 –2315.

Wilke T, Ziegltrum J. 2004. Genetic and anatomical analysis of the jumping slugs. USFS Contract Report #43-05g2-1-10086 Final Report. 2003

Zwickl DJ. 2006. GARLI: genetic algorithm for rapid likelihood inference. Available at: http://www.bio.utexas.edu/faculty/antisense/garli/Garli.html.

Table 2.1: Oligonucleotide sequences and annealing temperatures used for amplification of genetic markers for this study.

Locus	**Primer**	**Sequence (5′→3′)**	**Annealing temp. (°C)**	**Reference**
COI	LCO1490	TAAACTTCAGGGTGACCAAAAAATCA	52	Folmer et al., 1994
	HCO2198	GGTCAACAAATCATAAAGATATTGG		
16S	*16S*br-H	CCGGTCTGAACTCAGATCACGT	47.5	Lucid et al., 2018
	*16S*ar-L	CGCCTGTTTATCAAAAACAT		
ITS1	*ITS1*F	GCTGCGTTCTTCATCGATGC	52	Armbruster et al., 2000
	*ITS1*R	TAACAAGGTTTCCGTAGGTGAA		
actin	*Actin*A_S	ATGACATGGAGAAGATCTGGC	51.5	Rowson et al., 2011
	*Actin*BAS	TCCATACCAAGGAAAGATGGC		

Table 2.2: *BioGeoBEARS* results for each model implemented in the analysis: log-likelihood (LnL), number of parameters, dispersal (*d*), extinction (*e*), founder (*j*), corrected Akaike Information Criteria (AICc) and AICc weight.

Model	**LnL**	**Number of parameters**	**d**	**e**	**j**	**AICc**	**AICc weight**
DEC	−10.12	2	0.074	1.0×10^{-12}	0	24.24	0.019
DEC+j	−5.16	3	1.0×10^{-12}	1.0×10^{-12}	0.25	16.32	0.98

Figure 2.1: Area of study and *Hemphillia* sampling used in this study. Shaded areas represent distributions for *H. burringtoni*, *H. glandulosa*, *H. camelus*, *H. danielsi*, *H. dromedarius* and *H. malonei*, adapted from Burke (2013), and tentative distributions for *H. skadei* and *H. sp.*

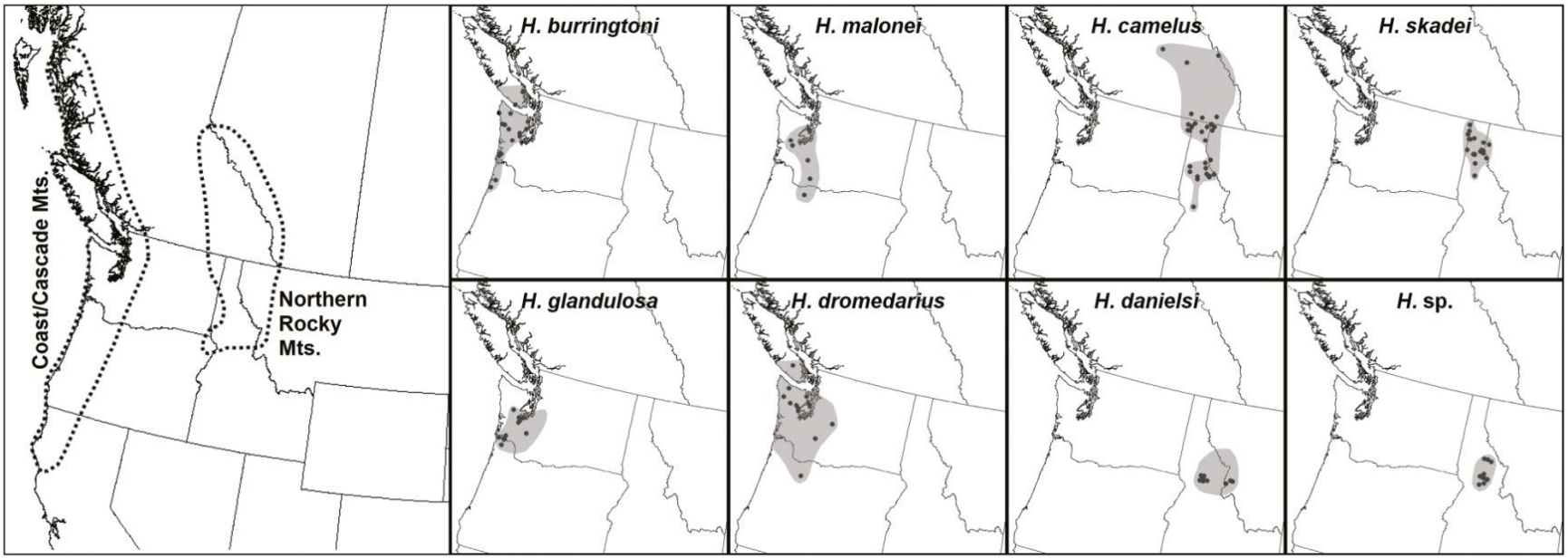


Figure 2.2: Best maximum-likelihood phylogeny for *Hemphillia* based on mtDNA haplotypes. The scale is given in substitutions per site, and node labels indicate maximum-likelihood bootstrap values for major groups.

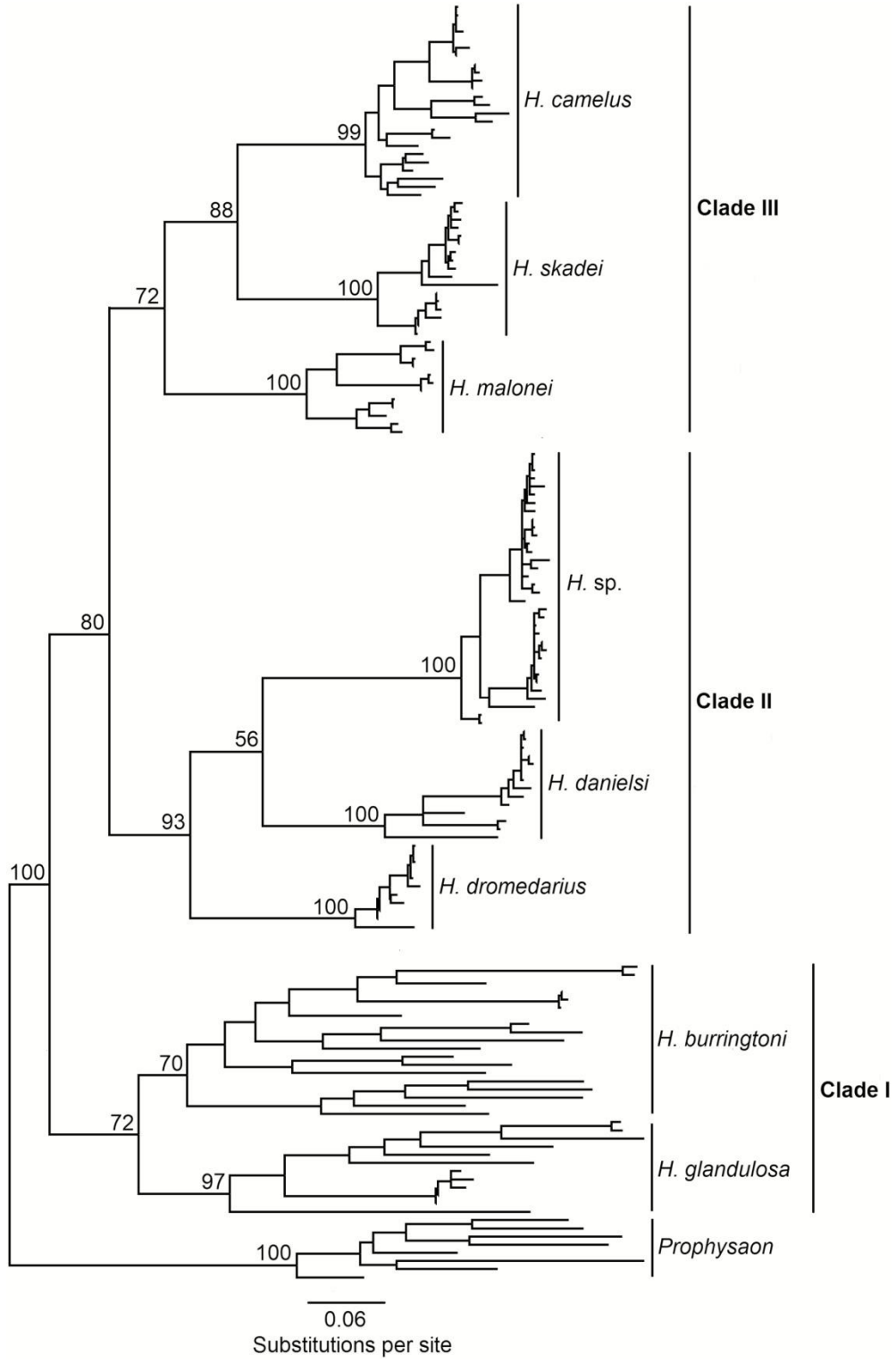


Figure 2.3: Maximum clade credibility species tree from the *BEAST analysis estimated from both mitochondrial and nuclear markers. Branch lengths are shown in millions of years and represent the median values of those present in the sampled trees. Node labels indicate Bayesian posterior probabilities for major groups and horizontal bars are the 95% highest posterior density (HPD) for the node estimates. Pie charts at nodes indicate the most probable ancestral range location(s) inferred from *BioGeoBEARS*, for both the DEC (top) and the DEC+j (bottom) models. Photo is of *Hemphillia camelus* (photo credit: Jack M. Sullivan).

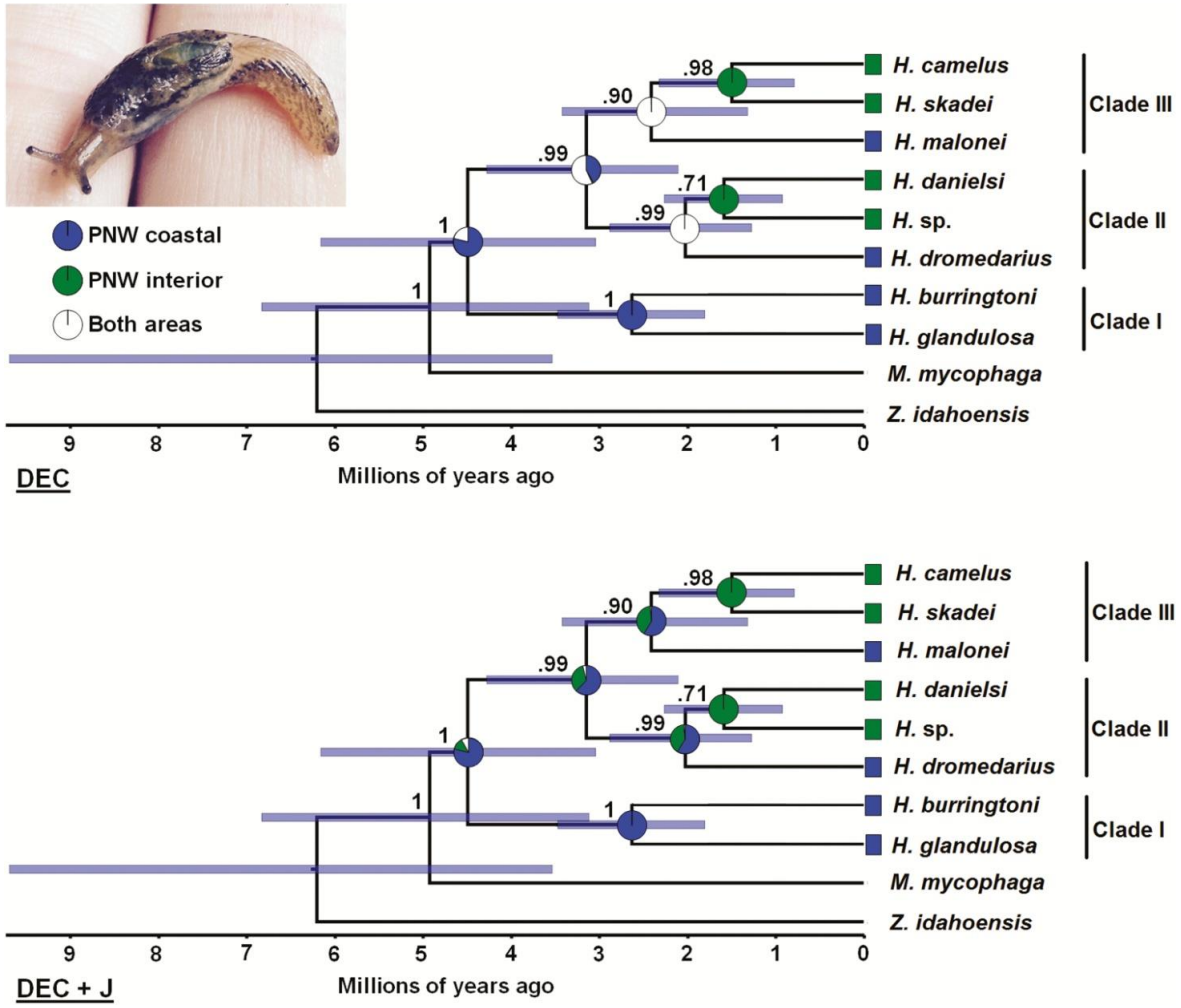


Figure 2.4: Results from analysis with *dpp-msbayes*. Top: posterior probability was highest for a single divergence event (0.59), whereas the next best supported scenario (two divergence events; bottom) received a posterior probability of 0.068.

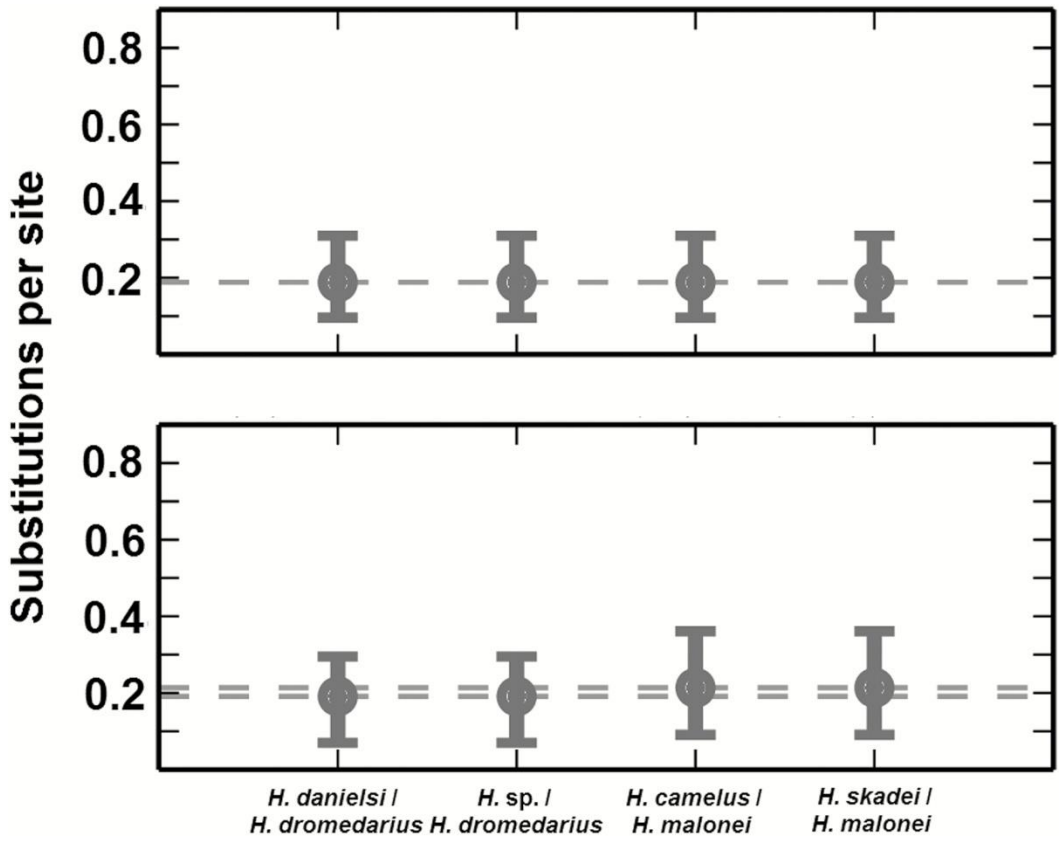


Figure 2.5: Species distribution models (SDMs) for each *Hemphillia* species. Scales indicate suitability values.

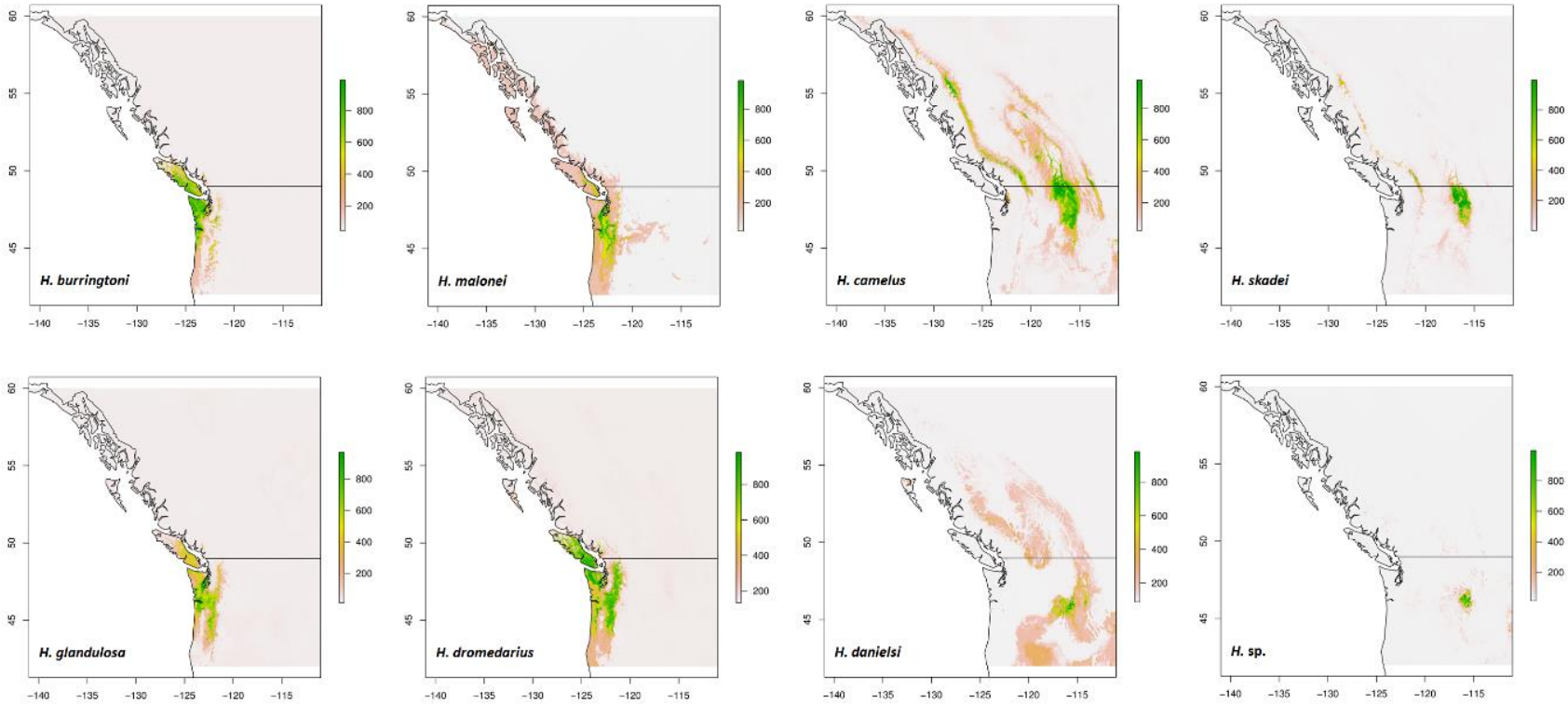
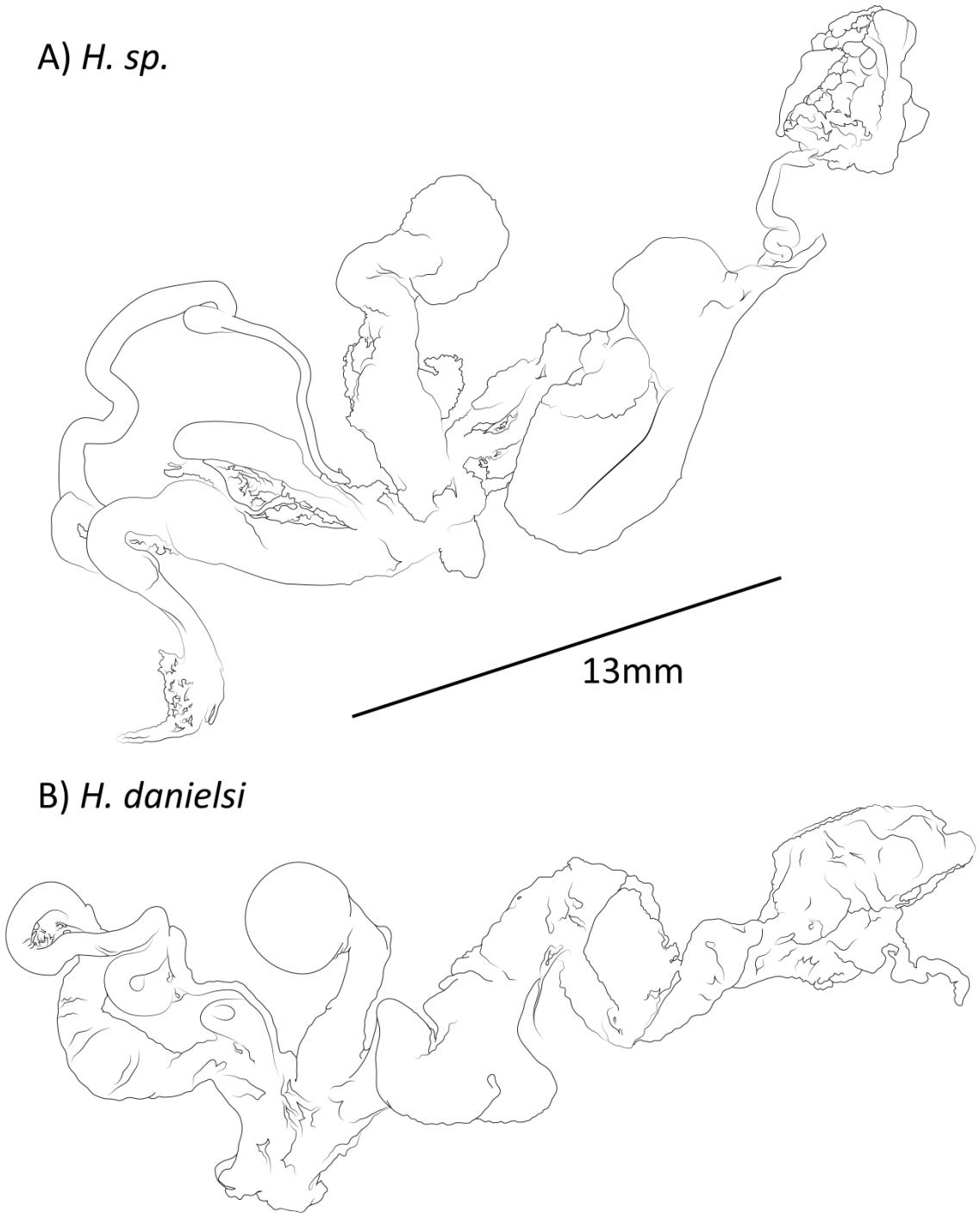


Figure 2.6: Distal genitalia of (A) undescribed jumping-slug (*Hemphillia sp*.) and (B) marbled Jumping-slug (*Hemphillia danielsi*)



Chapter 3: Comparative phylogeography of two Northern Rocky Mountain endemics: the widespread *Anguispira kochi occidentalis* and the narrow-range *Anguispira nimapuna* (Gastropoda: Discidae)

Abstract

The Northern Rocky Mountain ecosystem supports rich biological diversity with many endemic and rare species. Extant endemics display two biogeographic patterns: widespread species with fragmented populations, and narrow-range endemics. These distributions are shown by the congeneric snails *Anguispira kochi occidentalis* and *Anguispira nimapuna*. These two taxa are disjunct from the remaining species of the genus, which achieves its greatest diversity in eastern North America. Given the disjunct nature of *A. k. occidentalis* and *A. nimapuna*, we here present a mtDNA phylogeny of the genus that includes both eastern and western species to assess the phylogenetic position of *A. k. occidentalis* and *A. nimapuna*. We then reconstruct the demographic history of *A. k. occidentalis* and *A. nimapuna* by *analyzing*current patterns of genetic variation and interpreting the results considering the historical biogeography of the region. Both *A. k. occidentalis* and *A. nimapuna* represent unique taxa that are genetically and geographically distinct from their congeners. The current distribution and genetic structure of *A. k. occidentalis* has been shaped by both historical isolation in refugia and more recent northward shifts, whereas *A. nimapuna* is represented by two populations with shallow divergence in an area of long-term habitat stability.

Introduction

The Northern Rocky Mountain (NRM) mesic forest ecosystem is a relatively small biogeographic region that ranges from south-eastern British Columbia south through the Idaho Panhandle, into north-central Idaho and the north-eastern corners of Washington and Oregon, and west of the continental divide in Montana. Scattered throughout are discontinuous bands of mesic forest patches [sometimes referred to as inland rainforest (DellaSala, 2011)] that support a rich biodiversity of vascular and non-vascular plants, small-bodied vertebrates, and litter-dwelling organisms. Many of these appear to represent relicts, as many NRM species have disjunct relatives in the forests of the Pacific Coast (e.g., *Ascaphus*, *Prophysaon* and *Hemphillia*) and eastern North America (e.g., the Polygyridae and Plethodontidae). The current NRM biota includes two contrasting biogeographic patterns: (1) widespread species with populations distributed allopatrically in mid-elevation bands separated by stretches of dryer, lower-elevation habitat; (2) regional endemics confined to discrete patches of suitable habitat (Brunsfeld et al., 2001).

For widespread NRM taxa, it has been hypothesized that recurring glacial cycles in concert with altitudinal heterogeneity had a significant effect on the demographic history of regional populations (Brunsfeld et al., 2001). During glacial periods, multiple populations were fragmented in allopatric refugia across a topographically complex area [i.e., multiple inland refugia hypothesis; see Brunsfeld et al. (2001)]. The current distribution of widespread taxa would therefore have been achieved by subsequent post-glacial secondary contact among lineages, with the northernmost populations reflecting recent postglacial expansion. Phylogeographic studies of several NRM mesic species have indeed indicated populations survived in isolation long enough for genetic differences to be preserved among separate refugia [i.e., there was more than a single Pleistocene refugium; tailed frogs (Nielson et al., 2001); Constance's bittercress (Brunsfeld & Sullivan, 2005)]. Also present in this ecosystem are narrow-range endemics currently confined to discrete patches of suitable habitat. Numerous rare insects, gastropods, aquatic invertebrates, and wetland angiosperms are only known to occur in certain regions that were safe from glaciation during the Pleistocene such as within the Clearwater and Lolo National Forests (Fend & Gustafson, 2001; Stark & Gustafson, 2004; Stagliano et al., 2007; Newell et al., 2008; Stagliano, 2016). Specifically, the Clearwater drainage—a low-elevation tributary canyon of the Snake River in north-central Idaho, rich with moist, old-growth forest floors and higher-order tributaries—harbours many narrow-range endemics that appear to represent relicts that survived regional climatic changes in this protected canyon (Lichthardt & Moseley, 1994) but did not subsequently recolonize other parts of the ecosystem following glacial retreat.

Tiger snails of the genus *Anguispira* (family Discidae) are common, terrestrial snails that occupy moist, forested areas across portions of eastern and western North America. They achieve their greatest diversity in the east, where 12 of the 13 recognized species occur. In the west, only two forms occur: the subspecific *A. kochi occidentalis* and *A. nimapuna*. Specifically, *A. k. occidentalis* (Von Martens, 1882) is widespread in the moist, mixed coniferous forest of the NRM, whereas the type *Anguispira kochi kochi* (Pfeiffer, 1845) is a species of eastern North America. However, the original taxonomic description and subspecific designation of *A. k. occidentalis* was based primarily on subjective descriptions of shell characters that may be prone to homoplasy (e.g., Emberton, 1991a, 1995). Conversely, *A. nimapuna* (Baker, 1932), a narrow-range endemic known only from a few watersheds in Idaho County, Idaho, has a G1 global conservation status rank [Critically Imperiled (NatureServe, 2017)] and an S3 state rank [Species of Greatest Conservation Need Tier 3 (idfg.idaho.gov, retrieved Jan 15, 2021)] due to its rarity. The distribution of *A. nimapuna* is restricted to a small portion of the Clearwater drainage at the confluence of the Lochsa and Selway rivers, a V-shaped unglaciated valley where elevations rise abruptly from the river bottom to form steep slopes

(Lichthardt & Moseley, 1994). *A. nimapuna* may represent a relictual population that has persisted in an ancient refugium but has not expanded, perhaps due to the surrounding topological barriers.

Given the disjunction of *A. k. occidentalis* and *A. nimapuna* from their congeners—separated by ~2000 km from eastern populations—and the importance of our understanding of hierarchical evolutionary relationships, determining the phylogenetic position of these two species with respect to other members of *Anguispira* is critical. However, the only phylogenetic study on *Anguispira* published to date (Haskell & Pan, 2013) was limited to species/populations in the Cumberland Plateau region (southern Tennessee and northern Alabama) and did not include western forms. Here we present a mitochondrial DNA phylogeny of the genus with both eastern and western species to determine the phylogenetic position of *A. k. occidentalis* and *A. nimapuna*. After inferring this phylogeny, we focus on reconstructing the demographic history of *A. k. occidentalis* and *A. nimapuna* by *analyzing* current patterns of genetic variation and diversity of intraspecific lineages and interpreting the results considering the historical biogeography of the NRM region. Because contemporary patterns of genetic variation and distribution of populations are often the result of historical biogeographic processes, comparing the phylogeographic histories of species may highlight the different or similar effects that the biogeography of a region may have for species with contrasting distribution patterns. We first apply a traditional phylogeographic approach—reconstructing the history of the two species by assessing intraspecific patterns of diversity—then use coalescent simulations to compare alternative models of population history, and lastly evaluate our interpretations considering Species Distribution Models (SDM) produced for both species under current and Last Glacial Maximum (LGM) climate conditions using survey-presence records.

Methods and Materials

Sampling and tissue acquisition

We obtained sequence data from 120 *A. k. occidentalis* and 60 *A. nimapuna* specimens, the majority of which were obtained by field collection of live snails. To these, we added tissue samples from the California Academy of Sciences, Carnegie Museum of Natural History, Delaware Museum of Natural History, Field Museum of Natural History, Montana Natural Heritage Program, Royal British Columbia Museum and University of Florida Museum. Additional sequence data from 176 individuals representing nine of the remaining 11 currently recognized *Anguispira* species [missing are *Anguispira knoxensis* and *Anguispira rugoderma* (Integrated Taxonomic Information System; http://www.itis.gov; retrieved June 18, 2020)] were obtained from GenBank, including both unpublished and published (Haskell & Pan, 2013; accession numbers: JN544577 - JN544696) data, as well as data from an unpublished thesis (Clutts, 2008). To serve as confamilial phylogenetic

comparison, sequence data from 96 individuals representing seven of the 12 currently recognized *Discus* species was obtained from museums, field collections and GenBank (accession numbers: AF063140, AF063141, AF064406 - AF064409, AF064415 - AF064438, FJ917285, FJ969586 - FJ969703, JX911298, KM611953, KM612165, MF544381, MF544595, MF544774, MF544853, MF545101, MG421413, MG421515, MG421676, MG422014, MG422282, MG422422, MG422625, MG422626, MG422738, MG422943, MG423043, MG423120, MG423227, MG423525). Following Nordsieck (1986) and Emberton (1991b), we included the Oreohelicidae (*Oreohelix* and *Radiocentrum*) as a possible sister group as well as the Megomphicidae (*Polygyrella* and *Megomphix*) as a presumably more distantly related outgroup to root the discid phylogeny.

Laboratory work and data generation

Genomic DNA was extracted from tissue using either an E.Z.N.A. Blood and Tissue Kit (Omega Bio-tek, Doraville, CA, USA; this study) or a DNeasy Extraction Kit (Qiagen, Valencia, CA, USA; Clutts, 2008) as per the manufacturer's protocols. Regions of three mitochondrial genes—cytochrome c oxidase subunit I (*COI*), cytochrome b (*cytb*) and *16S* ribosomal RNA (*16S*)—were amplified via polymerase chain reaction (PCR) with universal metazoan and molluscan primer pairs: *COI* (Folmer et al., 1994), *16S* [Lucid et al. (2018) for this study and Geller et al. (1997) for Clutts (2008)] and *cytb* (Merritt et al., 1998). These genes were chosen to match data available from Clutts (2008) and in GenBank. Amplifications for newly generated data (this study) were performed in 25 μL reactions containing 2 μL DNA extract, 18 μL H2O, 2.5 μL buffer, 0.75 μL 50 mM MgCl2, 0.5 μL 10 mM dNTPs, 0.5 μL 10 mM forward primer, 0.5 μL 10 mM reverse primer and 1.25 U New England Biolabs Taq polymerase, while Clutts (2008) used HotStart Master Mix (Qiagen) following the manufacturer's protocols (half-volume reactions). Thermal cycling profiles included an initial denaturation step at 95o C for 2 min, followed by 30–35 cycles of denaturation (95o C for 45 s), primer annealing (45 s), extension (72o C for 60 s), followed by a terminal extension cycle of 72o C for 7 min. Primer annealing temperatures were: *COI* (52 °C); *16S* (50 °C); *cytb* (48 °C). Amplicons were then purified using the QiaQuick PCR Cleanup Kit (Qiagen; this study) or were gel-purified from 1% agarose gels using the QiaQuick Gel Extraction Kit (Qiagen; Clutts, 2008). Bi-directional DNA Sanger sequencing was outsourced to Eurofins MWG Operon, Louisvillle, KY, USA (http://www.eurofins.fr; this study) or cycle-sequenced using the BigDye Terminator Cycle Sequencing Ready Reaction Kit v.3.0 (Applied Biosystems, CA) and cleaned on Sephadex spin columns (Princeton Separations, Adelphia, NJ) and run out on an ABI Prism 377 (Applied Biosystems) automated DNA sequencer at Southern Illinois University (Clutts, 2008). Chromatograms in both directions were compared and consensus sequences were assembled using

either Chromas v.2.6.2 [Technelysium, http://www.technelysium.com.au/chromas.html (this study)] or Sequencher v.3.0 [Gene Codes, Ann Arbor, http://www.genecodes.com/ (Clutts, 2008)]. Because *COI*, *cytb* and *16S* genes each reside on the same mitochondrial DNA molecule and are inherited as one linkage group, they share a genealogical history and may share a single gene tree. Therefore, the three mitochondrial gene segments were concatenated into a single data set. We also explored partitioning sites into seven partitions, *COI* codon positions+*cytb* codon positions+*16S*; however, this partitioning scheme did not influence the results (e.g., poorly supported nodes were still poorly supported and highly supported nodes were still highly supported), so we focused our Discussion to results of a single partition. From these data, we produced three separate data matrices, including intraspecific sets for both *A. k. occidentalis* and *A. nimapuna*, as well as an interspecific set that contained all unique haplotypes (across the concatenated *COI-cytb-16S* matrix) of the remaining *Anguispira* species, *Discus* representatives and other outgroups. For each data matrix, multiple sequence alignments were constructed using MAFFT online with the default FFT-NS-2 option (http://www.ebi.ac.uk/Tools/msa/mafft/). For the interspecific data matrix only, several alignment-ambiguous loop regions of the *16S* gene were too divergent to be aligned across lineages and we therefore used the Gblocks Server v.0.91b algorithm with the least restrictive settings available [(Castresana, 2000); http://molevol.cmima.csic.es/castresana/Gblocks_server.html] to exclude these regions from phylogenetic inference.

Interspecific phylogeny

We analyzed the interspecific data matrix under maximum parsimony (MP), maximum likelihood (ML) and Bayesian frameworks. An MP analysis was performed in PAUP* (Swofford, 2003) using the heuristic search algorithm with ten random-addition replicates, tree bisection and reconnection (TBR) swapping, and all nucleotide substitutions treated as equal weight. Nodal support was estimated with 500 parsimony bootstrap replicates (Felsenstein, 1985). For the model-based ML analysis, we first selected a model of nucleotide sequence evolution via the automodel command in PAUP* under the Bayesian Information Criterion (BIC) and decision theory [DT (Minin et al., 2003)]. ML analyses were performed in Garli (Zwickl, 2006) using the GTR+I+Γ model (determined to be best fitting by PAUP*; with parameters estimated in Garli), conducting 100 replicate runs with random starting trees. Nodal support was assessed using 200 bootstrap replicates with two tree searches per bootstrap. We used the resulting phylogeny to test the assumption that the data set has evolved in a clock-like fashion by testing for a global molecular clock in PAUP* using the likelihood-ratio test of Felsenstein (1988). Then, a Bayesian analysis was performed with BEAST v.2.4.4 (Bouckaert et al., 2014) using the same GTR+I+Γ model determined earlier (with parameters

estimated in BEAST) as well as a relaxed lognormal molecular clock (based on results of clock analysis; Likelihood ratio test (LRT) = 822.3, d.f. = 328, P << 0.01) and a birth-death speciation tree prior [the birth-death prior has been shown to outperform the Yule prior when analyzing mixed inter- and intraspecies data sets (Ritchie et al., 2017)]. Posterior probabilities were estimated and used to assess support for each branch in the inferred phylogeny. Because we lacked fossils for calibration and recognizing that molecular clock estimates are often dubious (Hillis et al., 1996), we fixed the mean substitution rate to a value of 1 so that branch lengths would be reported in units of substitutions per site. Results were viewed in Tracer v.1.7 (Rambaut & Drummond, 2007) to ensure all parameters had converged and Effective Sample Size (ESS) values for all parameters exceeded 200, and maximum clade credibility trees were produced with the BEAST application TreeAnnotator.

Intraspecific phylogeny

We used a phylogeny-based framework to identify possible phylogeographic structure within *A. k. occidentalis* and *A. nimapuna*. For each data set, we first selected an appropriate model of nucleotide substitution using the automodel command in PAUP* under the BIC and DT (the best fitting models were HKY+I+Γ for *A. k. occidentalis* and HKY+I for *A. nimapuna*). Then, a ML tree was determined (from 100 replicate runs) for only unique haplotypes (across the concatenated *COI-cytb-16S* matrix; N = 64 for *A. k. occidentalis* and N = 40 for *A. nimapuna*) using Garli (Zwickl, 2006), and nodal support was assessed based on 200 bootstrap replicates (five tree searches per bootstrap). Next, we estimated trees under Bayesian inference with BEAST 2.4.4. Using all alleles (N = 120 for *A. k. occidentalis* and N = 60 for *A. nimapuna*), we specified both a coalescent constant population size tree prior and a Bayesian skyline tree prior with the same substitution models mentioned above. The skyline analysis attempts to assess putative fluctuations in effective population size over time, which is done by estimating the posterior distribution for effective population size at intervals along the phylogeny (see Ho & Shapiro, 2011). Each BEAST analysis was run for 100 million generations and samples were drawn every 10 000 generations, and we discarded 2500 samples from each run as burn-in. Results from each run were viewed in Tracer v.1.7 (Rambaut & Drummond, 2007) to ensure all parameters had converged and Effective Sample Size (ESS) values for all parameters exceeded 200, and maximum clade credibility trees were produced with the BEAST application Tree Annotator. Bayesian skyline plots (BSP) were also reconstructed using Tracer.

Intraspecific polymorphism

To estimate the extent of mtDNA variation, we calculated two intraspecific diversity statistics using MEGA6 (Tamura et al., 2013): the Watterson estimator [θ, estimated from the number of segregating sites (Watterson, 1975)] and nucleotide diversity [π, mean number of pairwise differences per site (Nei & Li, 1979)]. We then examined the behavior of these estimates as a function of sample size by calculating them for each n, where n is the number of sequences, and plotting the results of ten replicates. We randomly selected n samples from a data set, calculated θ and π, and repeated this ten times for each n. Because both θ and π are unbiased estimates of neutral diversity, the neutral model predicts that $\theta \approx \pi$; however, this assumption fails by violation of the infinite-sites model and under various influences of selection and demographic history (Tajima, 1989a, b). Therefore, θ and π can be compared to infer the types of variants present in each sample (e. g., when there are a lot of singletons, the latter underestimates polymorphisms compared to the former, whereas when there are many alleles at intermediate frequencies, pairwise differences are inflated when compared to segregating sites). To do so, we performed Tajima's *D* intraspecific polymorphism statistical test for neutrality to test the prediction that $\theta \approx \pi$ using Arlequin v.3.5 (Excoffier et al., 2007), running 1000 coalescent simulations under the infinite-sites model to test the significance of *D*.

Nucleotide diversity (π) is a useful measure of the extent of DNA polymorphism. However, the variance ($\sigma 2$) of an estimate of π has historically not been well studied (Nei & Jin, 1989) despite $\sigma 2$ being a potential source of information of demographic history (Wakeley, 1996a, 1996b). For example, consider two species having the same mean π but one species is panmictic whereas the other species is divided into several subpopulations between which there may or may not be genetic exchange. This may cause the two species to have different variances of π. Therefore, we quantified the variance of π for both *A. k. occidentalis* and *A. nimapuna* and compared those values to simulated data sets that fitted a single, neutrally evolving population of constant size, i.e., we compared the variance of π to the variance of a simulated population having the same mean of π. We quantified variance by calculating π for 200 randomly generated subsamples of size n = 20, plotted the distribution of values, and qualitatively compared the actual and simulated distributions against each other.

To estimate the most likely number of genetically differentiated clusters present in each species, a Bayesian Analyses of Population Structure (BAPS) was performed in BAPS 6.0 (Corander et al., 2008). BAPS performs a clustering with linked loci (codon linkage model) analysis that takes into consideration the non-independence of linked loci and attempts to infer the most likely number of putative genetic clusters (K) by maximizing the Hardy-Weinberg equilibrium amongst clusters [i.e., minimizing the Walhund effect (1928)] through a stochastic learning algorithm. We estimated the

probability of a different number of genetic clusters (K=1 to 15) with ten runs for each K, and the clustering of groups with the lowest log likelihood was selected.

Support for evolutionary scenarios estimated using abc

Although descriptive analyses of genetic variation and phylogenies are useful to identify patterns and compare hypotheses, Approximate Bayesian Computation (ABC) allows for the quantitative comparisons of alternative scenarios via simulation and estimation of the posterior distributions of important parameters. Such simulation approaches, which rely on implied assumptions of the many parameters, are therefore valuable when used in conjunction with other methods that do not rely on highly parameterized models. Thus, to better understand the evolutionary history of *A. k. occidentalis* and *A. nimapuna*, we performed coalescent simulations in an ABC framework using DIYABC v.1.0.4.46 (Cornuet et al., 2008), simulating many thousands of genealogies and retaining those simulations that produced genetic variation patterns close to the empirical data which were then used to discriminate among a set of alternative historical scenarios.

For *A. k. occidentalis*, we partitioned the data into five clusters that were recovered by the BAPS analysis and compared six alternative scenarios that considered divergence and population size variation (Fig. 3.1A). The *A. nimapuna* data were partitioned into two clusters that were recovered by BAPS and we compared five alternative scenarios (Fig. 3.1B) that considered population size variation. The similarity between the simulations and the empirical data was measured using both within- and between-population summary statistics, including number of segregating sites, mean pairwise differences, Tajima's D, and private segregating sites for a single population, and number of segregating sites, mean pairwise differences within, and mean pairwise differences between pairs of populations. For each species, 100 000 simulated data sets were generated for each scenario to build a reference under a mutation model with mean rate ranging from 1.00 × 10-09 to 1.00 × 10-07 and uniform prior distribution. A pre-evaluation step based on a principal component analysis (PCA) was performed to ensure scenarios and priors produced simulated data sets similar enough to the empirical data. The relative posterior probabilities of the competing scenarios were estimated via logistic regression on the 10% of simulated data sets closest to the empirical (Cornuet et al., 2010). The model with the highest posterior probability was considered the best model.

Results

Interspecific phylogeny

The interspecific data matrix consisted of 1374 bp and, excluding outgroups, included 665 variable sites, 617 of which were parsimony informative and 48 were singletons. The maximum likelihood (Fig. 3.2), maximum parsimony (Fig. 3.3), and Bayesian (Fig. 3.4) trees resulted in slightly different topologies; however, three general concordant results were revealed: the Discidae is monophyletic (MP bootstrap 91; ML bootstrap 98; Posterior Probability 0.96); the eastern *Anguispira* species (apart from *A. k. kochi*) form a strongly supported monophyletic group (MP 87; ML 83; PP 0.97) and *A. k. kochi+A. k. occidentalis* grouped together, though the support is moderate (MP < 70; ML < 70; PP 0.98).

The parsimony analysis produced a single most parsimonious tree. *A. k. kochi* and *A. k. occidentalis* group together and are distantly related to the eastern *Anguispira* species, and *A. nimapuna* clusters with *Discus* species (with < 70 bootstrap support). In the likelihood analysis, all *Angusipira* species cluster together with < 70 bootstrap support. *Anguispira nimapuna* is sister to a clade comprising the *A. kochi* clade and a well-supported eastern *Anguispira* clade. In the Bayesian analysis, the *A. kochi* samples are monophyletic, and, along with *A. nimapuna*, cluster with *Discus* species with < 0.90 posterior probability. Thus, the main difference among analyses is whether *Anguispira* as a whole is monophyletic.

A. k. occidentalis phylogeny and polymorphism

The *A. k. occidentalis* data set consisted of 120 sequences of 1486 bp. There were 250 variables sites, 189 of which were parsimony informative and 61 singletons, and 64 unique haplotypes, and for the protein-coding sections (*COI* and *cytb*) substitutions occurred primarily at third positions and synonymous substitutions outnumbered non-synonymous. Phylogenetic analyses were comparable between maximum likelihood (Fig. 3.2) and Bayesian trees (Fig. 3.4). There are two well-supported clades: a small clade composed of only Clearwater drainage individuals and a much larger clade composed of all remaining *A. k. occidentalis* individuals. The latter group of individuals are partitioned into subgroups in both trees. However, the deepest splits among these clusters are (generally) not well supported. Highly supported nodes include a clade of individuals collected from the southern tip of the Selkirk Mountains, three distantly related clades distributed allopatrically across the Coeur d'Alene (CDA) Mountains of Idaho and Montana (hereafter referred to as western, central and eastern CDA [WCDA], CCDA and ECDA), a pair of localities in the Bitterroot Mountains (BTRT), two distantly related clades distributed on either side of the Salmon River in west-central Idaho (near Riggins) [hereafter referred to as west Salmon River (WSR) and east Salmon

River (ESR)], a clade of Oregon individuals, and a genetically depauperate group of individuals collected from northern and eastern sampling locations (NE) (see Figs 3.5, 3.6).

Uncorrected nucleotide diversity (π) was 0.019 while the Watterson estimator (θ) was higher at 0.031 (Fig. 3.7), although this difference is not significant: Tajima's D = -1.28 ($P > 0.10$). However, we repeated Tajima's D for a single phylogroup (the NE clade), and this clade had Tajima's D = -2.44 ($P < 0.05$), in which case we conclude there is a significant excess of low-frequency polymorphisms in that clade. We compared the distribution curves of π (Fig. 3.8) for the *A. k. occidentalis* data set to that of a simulated data set (of a single, neutrally evolving population) produced to have an identical mean. The two distribution curves displayed identical means but qualitatively different variabilities. That is, the variance of the *A. k. occidentalis* is greater, suggesting that *A. k. occidentalis* is likely structured as a group of populations spatially separated across its range.

BAPS clustering analysis found the best partition of data to be K=5 populations (Fig. 3.6). Three clusters are composed of samples from small geographic areas and correspond to highly supported clades: Clearwater (N = 15), Selkirk (N = 5) and WSR individuals (N = 9). A fourth cluster (N = 28) includes individuals sampled across the central part of Idaho's panhandle (hereafter referred to as the Central cluster) and a couple of locations in Montana, and includes the WCDA, CCDA, ECDA, BTRT and ESR clades. The fifth cluster (N = 63) comprises individuals collected near the edges of the range, including north Idaho, Montana, British Columbia, and Oregon (hereafter referred to as the expanding margins cluster).

The Bayesian skyline plot revealed a historical signal of stability, followed by a decline and then increase during the most recent interval (Fig. 3.8). Due to the putative effects of lumping populations, we repeated the skyline analysis for three of the BAPS clusters (the Selkirk and WSR clusters were excluded due to low sample size). The expanding margins cluster shows a shallow history and a recent, precipitous increase. On the other hand, both the Central cluster and Clearwater cluster show a deeper history (relative to the expanding margins cluster) and a wavering but relatively stable trend. Lastly, the best supported scenario evaluated using DIYABC was scenario 2 (logistic approach PP = 0.39; Fig. 3.1). In this scenario, the oldest event (t3) was a simultaneous divergence of the five BAPS clusters with an increase in Ne for the expanding margins cluster in the most recent (t1 – 0) interval. The posterior probabilities of remaining scenarios were as follows: scenario 1, PP = 0.13; scenario 3, PP = 0.19; scenario 4, PP = 0.24; scenario 5, PP = 0.01; and scenario 6, PP = 0.04.

A. nimapuna phylogeny and polymorphism

The *A. nimapuna* data set consisted of 60 sequences of 1419 bp. There were 78 variable sites, 49 of which were parsimony informative and 29 singletons, and 40 unique haplotypes, and for the protein-coding sections (*COI* and *cytb*) substitutions occurred primarily at third positions and synonymous substitutions outnumbered non-synonymous. Results of phylogenetic analyses were nearly identical between the best maximum likelihood (Fig. 3.10) and Bayesian trees (Fig. 3.11). There are two well-supported clades that correspond to an eastern and western group (Fig. 3.11).

Uncorrected nucleotide diversity (π) was 0.0076 while the Watterson estimator (θ) was higher at 0.012 (Fig. 3.12), although this difference is not significant: Tajima's D = -1.22 ($P > 0.10$). We repeated Tajima's D for both western and eastern clades which showed D = -0.95 ($P > 0.10$) and D = -1.01 ($P > 0.10$), respectively. We compared the distribution curves of π (Fig. 3.13) for the *A. nimapuna* data set to that of a simulated data set (of a single, neutrally evolving population) produced to have an identical mean. The two distribution curves displayed identical means and qualitatively similar variabilities, suggesting that *A. nimapuna* is composed of a single population. However, BAPS clustering analysis returned the best partition of the data as two clusters (K=2) that correspond to the eastern and western phylogroups (Fig. 3.11).

The Bayesian skyline plot for all *A. nimapuna* individuals indicated a steadily increasing trend (Fig. 3.14). When we repeated the skyline analysis for the two BAPS clusters, the eastern group showed a similarly increasing trend; however, the trend for the western clade is more or less stable without abrupt change. Lastly, the best supported scenario evaluated using DIYABC was scenario 2 (logistic approach PP = 0.80; Fig. 3.1) in which there was a signal of population increase for both eastern and western clusters during the t2 – 0 interval. The posterior probabilities of remaining scenarios were as follows: scenario 1, PP = 0.04; scenario 3, PP = 0.01; scenario 4, PP = 0.03; and scenario 5, PP = 0.12.

Discussion

The inland temperate rainforests of Idaho, Montana and British Columbia are home to a large array of endemic taxa (Brunsfeld et al., 2001; DellaSala, 2011). Some of these have phylogenetic connections to Pacific coastal species (e.g., tailed frogs, *Ascaphus montanus*), some are relictual species that only occur in the inland rainforest (e.g., pygmy slugs, *Kootenaia burkei*), yet others have more complex biogeographic histories. Here, we examined two taxa in the tiger snail genus *Anguispira*, one of which (*A. kochi occidentalis*) has biogeographic affinities to eastern congeners and the other (*A. nimapuna*) exhibits a classic narrow endemic pattern associated with the single drainage of the Clearwater River.

Interspecific phylogeny

Phylogenetic analysis of the mtDNA indicates that *Anguispira* may not be monophyletic, and that *A. nimapuna* and *A. k. kochi+A. k. occidentalis* are differentiated from the remaining *Anguispira* species, the latter of which form a statistically supported clade in all three phylogenetic reconstruction methodologies (Figs. 3.2, 3.3, 3.4). Indeed, the most consistent result from these analyses is that eastern *Anguispira* species (apart from subspecific *A. k. kochi*) form a strongly supported clade. Only in the ML (Fig. 3.2) reconstruction do all *Anguispira* species form a monophyletic group, with *A. kochi* and then *A. nimapuna* splitting off deeply; however, monophyly is not statistically supported (< 70 bootstrap support). In the other two analyses, *A. nimapuna* and *A. kochi* cluster with *Discus* species (Figs. 3.3, 3.4). This low phylogenetic resolution is likely explained by a relatively short period of rapid evolution deep in the tree, resulting in short internal branches relative to the longer branches that lead to terminal taxa.

The distribution of *A. kochi* is disjunct, consisting of two components that extend from southern Ontario to Tennessee in the east and from southern British Columbia to Oregon in the west. The western form, *A. k. occidentalis* (Von Martens, 1882), was described after the eastern type based on shell morphology (coloration, thickness and size), and Pilsbry (1948) noted that while some western specimens are not discernible from eastern specimens, most could be easily separated. Although the western and eastern forms display superficially indistinguishable shell characteristics and morphology, Pilsbry (1948) maintained the subspecific status of *A. k. occidentalis* because of a separation of over 2000 km with no connections. These two subspecies are sisters in all three trees presented in this study, but with varied statistical support. That is, Bayesian posterior probability support is high (0.98; Fig. 3.4) but bootstrap support is < 70 in both the MP (Fig. 3.3) and ML (Fig. 3.2) reconstructions. This relationship between *A. k. kochi* and *A. k. occidentalis* is of particular interest given the large midcontinental gap between the two distributions. Several other taxonomic groups (with similarly low dispersal capacities) share this disjunct pattern of occurrence in both eastern and western North America, including salamanders (Vieites et al., 2007) and spiders (Hendrixson & Bond, 2007), as well as other snails (Emberton, 1994). These distributions hint at an originally continuous, transcontinental range with subsequent vicariance due to increasing aridity in the interior (Emberton & Roth, 1994). Indeed, a major vicariant event associated with the Western Interior Seaway (Cretaceous period; 145–66 Mya) followed by the decline of forest biomes and spread of grasslands (Roberts & Kirschbaum, 1995) may be the most parsimonious scenario for diversification at higher ranks, although divergence at lower levels may be more likely explained by more recent transcontinental movements. However, the former explanation is plausible for the *A. k. kochi*/*A. k. occidentalis* split given it is about as deep as the basal divergence for the clade comprising

all the other eastern species. Regardless, *A. k. kochi* and *A. k. occidentalis* have been in isolation long enough to have achieved reciprocal monophyly which takes, on average, xNe generations [where x is the inheritance scaler (Moritz, 1994; Allendorf & Luikart, 2009)]. Given that these two groups are independently evolving lineages that are geographically isolated, we believe that elevating *A. k. occidentalis* to species status is warranted in order to facilitate more efficient conservation planning or management.

A. nimapuna is the only *Anguispira* species other than *A. k. occidentalis* that is uniquely distributed in western North America and was described in a paper alongside *Discus marmorensis*, an extremely narrow endemic known only from the confluence of John Day Creek and the Salmon River in Idaho (Baker, 1932). Interestingly, Umiński (1963) suggested the possibility of a closer relationship between *A. nimapuna* and *D. marmorensis* (and proposed changing *D. marmorensis* to *A. marmorensis*) given similarities in the shape, aperture, and sculpture of the shell, as well as geographical distribution; both species inhabit mountainous regions of north-central Idaho. Although *A. nimapuna* and *D. marmorensis* cluster together in our Bayesian analysis (Fig. 3.3), this relationship has < 0.90 PP support. In the other analyses, *A. nimapuna* clusters with other *Discus* species in the MP reconstruction (Fig. 3.3) and is sister to all other *Anguispira* species in the ML reconstruction (Fig. 3.2). That *A. nimapuna* is a member of the Discidae is supported; however, like *A. kochi*, the taxon appears to be only distantly related to other *Anguispira* species. That is, if it is not sister to the rest of *Anguispira*, it is a probable *Discus* representative. However, this is not comprehensive because there are several *Discus* species that are not represented in our data set, including two taxa native to western North America—*D. brunsoni* [lake disc (Berry, 1955)] and *Discus shimekii* [striate disc (Pilsbry, 1890)]—as well as up to three other named *Discus* species in North America and Eurasia.

Results regarding only eastern *Anguispira* (apart from *A. k. kochi*) generally agreed with that of Haskell & Pan (2013) in the grouping patterns and their levels of support. There is a well-supported clade of carinate limestone specialists as follows: {[(*Anguispira alabama*, *Anguispira picta*), Dry Creek], *A. cumberlandiana*}. Specimens of *Anguispira alabama* were collected from Alabama and Tennessee whereas specimens of the other three clades were all collected in Tennessee. The Dry Creek Population originates from a rare population of snails recently discovered (2008) in Tennessee in an area north-west of the *A. picta* range (see Haskell & Pan, 2013). *A. cumberlandiana* is recovered as two additional clades, one of which contains *A. cumberlandiana* individuals while the other contains the subspecific *A. c. columba*. All three *A. cumberlandiana* clades are from Tennessee but do not appear to be closely related. Specimens of *A. fergusoni*, an Atlantic coastal plain species, form a strongly supported clade. *Anguispira jessica* are recovered as three distantly related clades, one including only *A. jessica* individuals (from North Carolina) whereas the other two each contain

Anguispira jessica individuals are recovered and *A. alternata* individuals, one from Virginia and the other from West Virginia. Additional specimens identified as *A. alternata* are recovered as members of several different clusters. The largest *A. alternata* clade contains representatives broadly distributed from the central and eastern United States and Ontario. A second *Anguispira alternata* clade is sister to *Anguispira columbiana columba*, and these two clades are sister to an *Anguispira strongylodes* clade, all three of which were collected from Tennessee and form a highly supported monophyletic group. A third well-supported cluster contains several *A. alternata* specimens as well as *A. macneilli*, *A. mordax*, and additional *A. strongylodes*. Collection localities from this group are broadly distributed across the southern United States. Given identifications were based on shell morphology, it is likely that some distantly related clades have converged on a similar external shell morphology.

A. k. occidentalis phylogeography

The current distribution of *A. k. occidentalis* spans both unglaciated and glaciated regions of the NRM [the Last Glacial Period occurred from ~115 000–11 700 years ago (Pielou, 1991)], and our data suggest that both historical isolation (e.g., the deep split between the Clearwater clade and all remaining clades) and more recent expansions (e.g., the NE clade) have shaped the spatial distribution and genetic structure of this species. Genetic diversity is structured at both deep and shallow scales, with there being several deeply divergent mtDNA lineages that, in turn, exhibit shallower coalescent events among samples. The deepest split separates the Clearwater clade from all other clades. Despite being from a small geographic area (samples collected around the confluence of the Lochsa, Clearwater and Selway rivers), the Clearwater clade has a relatively high level of diversity (13 haplotypes identified from 15 samples). The depth of coalescence within the Clearwater clade is also comparable to the interallelic divergences between the remaining phylogroups (Figs 3.5, 3.6) which, coupled with the predicted demographic stability from the Skyline analysis (Fig. 3.9), suggest long-term stability of both population and habitat. In addition to the Clearwater clade, there is a highly supported ‘expanding margins’ clade that includes a single haplotype from Idaho’s panhandle (C881), a highly supported Oregon clade, and a multitude of individuals from northern Idaho, Montana, and British Columbia, the lattermost of which exhibits a collapsed, comb-like structure indicative of a recent demographic expansion (Figs 3.5, 3.6). BAPS inferred an additional three clusters: Selkirk, WSR and Central (Fig. 3.6). The Selkirk clade is geographically separated from neighboring lineages by the Priest River basin whereas the WSR group is separated from ESR individuals by the Salmon River. Finally, within the Central BAPS cluster, three clades (WCDA, CCDA and ECDA) are

distributed allopatrically across the Coeur d'Alene Mountains of Idaho and Montana but any putative physiographic barriers affecting them is not immediately apparent (Fig. 3.6).

The overall phylogeographic pattern in *A. k. occidentalis*—considerable differentiation across a small geographic range, as well as mixing of phylogroup lineages—is often reported for other terrestrial snails (Thomaz et al., 1996), including another western United States snail group, *Oreohelix* (Dempsey et al., 2019, 2020; Linscott et al., 2020). Compared to other taxa, land snails exhibit high mtDNA substitution rates (Thomaz et al., 1996), and highly divergent lineages separated by short spatiotemporal events are often found (Chiba, 1999; Pinceel et al., 2005). We suggest that the mtDNA patterns observed in *A. k. occidentalis* are likely the result of a combination of high mtDNA substitution rates and prolonged isolation and *in situ* diversification of refugial populations and subsequent expansion from multiple refugia. Indeed, the lack of resolution of relationships among deeper branches of the *A. k. occidentalis* phylogeny, and the short internodes separating those splits, is consistent with a near-simultaneous fragmentation of a widespread ancestor. Our ABC analyses provide additional support for a near-simultaneous origin for the BAPS clusters, as well as the subsequent increase of Ne of the 'expanding margin' clade.

Taken together, these results strongly support, at a minimum, a NRM dual refugia hypothesis (Brunsfeld et al., 2001). That is, advancing glaciers likely forced the mesic ecosystem into refugia located to the south or in canyons deep enough to offer climatic insulation (Daubenmire, 1975), thereby fragmenting a previously contiguous distribution of populations into multiple compartmentalized refugia and fostering divergence among groups. Indeed, several studies have shown that during past glaciations species' distributions were not uniformly shifted to the south, but that multiple refugia occurred proximal to the ice sheets (Stewart & Lister, 2001). Genetic data from *Hemphillia* slugs (Rankin et al., 2019), Rocky Mountain tailed frog [*Ascaphus montanus* (Nielson et al., 2001; Metzger et al., 2015)] and Constance's bittercress [*Cardamine constancei* (Brunsfeld & Sullivan, 2005)] are all consistent with expansion from multiple refugia; however, there are other species that exhibit evidence of expansion from a single refugium [e.g., Idaho giant salamanders, *Dicamptodon aterrimus* (Carstens et al., 2005)].

A. nimapuna phylogeography

A. nimapuna is a narrow-range endemic species that occurs only around the watershed of the South Fork of the Clearwater River and upstream into the Lochsa and Selway drainages, Idaho County, Idaho (Fig. 3.11). Despite this relatively small area (our sampling spans ~60 km east to west and ~30 km north to south) phylogenetic analyses revealed two geographically structured and highly supported mtDNA lineages. However, the net nucleotide diversity between these two phylogroups

(0.00432 subs/site) is on par with intralineage nucleotide diversity (0.00495^{West} subs/site and 0.00594^{East} subs/site), in which case we assume the two groups have been diverging for only ~Ne generations (Moritz, 1994; Allendorf & Luikart, 2009). Indeed, the variance of π for all samples is qualitatively similar to that expected for a single, neutrally evolving population (Fig. 3.13). If the western and eastern phylogroups are indeed real, then it is likely they have only very recently diverged.

BAPS also recovered the western and eastern clusters as the best partition of the data, and subsequent DIYABC analysis of those clusters indicated this species has experienced an increase in population size, which is perplexing given its narrowly restricted range. Although this result was only partially supported by the skyline analyses in that the western group did not exhibit a strong signal of increasing Ne (Fig. 3.14). A stable and perhaps increasing demographic history would be predicted for a species inhabiting an area of habitat stability through fluctuating climates. Several additional lines of evidence suggest the Clearwater served as a refugium during past glacials (Daubenmire, 1952, 1978; Detling, 1968; Shafer et al., 2010). That is, the river canyons of the Clearwater basin are some of the northernmost canyons that were free of glacial ice and probably served as a refugium for many extant species with high moisture-heat requirements. This moist-canyon ecosystem contains many plants and animals that are endemic to the NRMs, as well as regional endemics whose distributions are closely tied to the Clearwater drainage such as the bank monkeyflower [*Mimulus clivicola* (Lorain, 1992)], several species of byrrhid beetles (Johnson, 1987) and the snail *Allogona lombardi* (Burke, 2013). It is also interesting to note that a widespread species like *A. k. occidentalis* contains a distinct (all private haplotypes) Clearwater clade with relatively high diversity, suggesting this is a long-established population as well.

High concentrations of narrow endemics can be linked to the accumulation of distinct evolutionary units within long-term stable environments (Molina-Venegas et al., 2017) and are sometimes indicative of refugia that have experienced long-term stability of climate and habitats (e.g., Jetz et al., 2004; Ohlemüller et al., 2008; Harrison & Noss, 2017). However, although this may explain why the Clearwater refugium is an endemic hotspot, it does not explain why some species with highly restricted ranges (*A. nimapuna*) have close relatives with widespread distributions (*A. k. occidentalis*). Given the distinctive conditions of the Clearwater drainage (a unique climate resulting from a combination of a low-elevation canyon, mountainous terrain, and high precipitation and moderate temperatures similar to Pacific coastal climates), it may be that *A. nimapuna* is adapted to the particular conditions of the Clearwater drainage and has not colonized new adjacent sites with more stressful conditions. However, the observation that *A. k. occidentalis* contains a unique and diverse Clearwater clade may also suggest colonization to and from this immediate area is difficult.

Indeed, the refugium ecosystem occupies a V-shaped valley and slopes rise abruptly from the valley to form steep breaklands and upstream and downstream migration pathways may prove to be too challenging for organisms with low dispersal ability (Lichthardt & Moseley, 1994). In any case, *A. nimapuna* is an exceptionally rare species that occupies habitat that is itself rare and sensitive. Species with these attributes should be allotted conservation priority, which is amongst the reasons this species is classified as a Species of Greatest Conservation Need by the U.S. state of Idaho (IDFG, 2017).

Conclusion

Both *A. k. occidentalis* and *A. nimapuna* represent unique taxa that are genetically and/or geographically distinct from their congeners and are part of a larger group of endemic mollusks that display restricted geographic distributions in the NRM. Several of these are highly localized in small areas such as the Clearwater drainage of north-central Idaho. A combination of intrinsic and extrinsic factors, such as mollusks having poor dispersal ability or limited accidental dispersal opportunities, geologic and historic factors, and long-term permanency of some NRM habitats likely have contributed to the high diversity and regional endemism in the area. There is an increasing need to obtain data on the genetic and spatial structure of endemic taxa in areas such as the NRM due to the threat of habitat fragmentation (both natural and induced by human activity), which may be exponentially greater to minute organisms that maintain small ranges and have little public appeal.

Literature Cited

Allendorf FW, Luikart G. 2009. Conservation and the genetics of populations. Hoboken, New Jersey: John Wiley & Sons.

Baker HB. 1932. New land snails from Idaho and eastern Oregon. The Nautilus 45: 82 –87.
Berry SS. 1955. An important new land-snail from the Mission Range, Montana. Bulletin of the Southern California Academy of Sciences 54: 17 –19.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Computational Biology 10: e1003537.

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. Comparative phylogeography of northwestern North America: a synthesis. Special Publication-British Ecological Society 14: 319 –340.

Burke TE. 2013. Land snails and slugs of the Pacific Northwest. Corvalis, Oregon: Oregon State University Press.

Carstens BC, Degenhardt JD, Stevenson AL, Sullivan J. 2005. Accounting for coalescent stochasticity in testing phylogeographical hypotheses: modelling Pleistocene population structure in the Idaho giant salamander *Dicamptodon aterrimus*. Molecular Ecology 14: 255 –265.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution 17: 540 –552.

Chiba S. 1999. Accelerated evolution of land snails Mandarina in the oceanic Bonin Islands: evidence from mitochondrial DNA sequences. Evolution 53: 460 –471.

Clutts SA. 2008. Systematics of *Anguispira* (Pulmonata: Discidae) based on molecular and morphometric data. MS Dissertation, Southern Illinois University.

Collins WD. 2006. The community climate system model version 3 (CCSM3). Journal of Climate 19: 2122 –2143.

Corander J, Marttinen P, Sirén J, Tang J. 2008. Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. BMC Bioinformatics 9: 1 –14.

Cornuet JM, Ravigné V, Estoup A. 2010. Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1. 0). BMC Bioinformatics 11: 401.

Cornuet JM, Santos F, Beaumont MA, Robert CP, Marin JM, Balding DJ, Guillemaud T, Estoup A. 2008. Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. Bioinformatics 24: 2713 –2719.

Daubenmire R. 1952. Plant geography of Idaho. In: Davis RJ, ed. Flora of Idaho. Provo: Brigham Young University Press, 1 –17.

Daubenmire R. 1975. Floristic plant geography of eastern Washington and northern Idaho. Journal of Biogeography 2: 1 –18.

Daubenmire R. 1978. Plant geography with special reference to North America. New York: Academic Press.

DellaSala DA, ed. 2011. Temperate and boreal rainforests of the world: ecology and conservation. Washington, D.C.: Island Press.

Dempsey ZW, Burg TM, Goater CP. 2019. Found, forgotten, and found again: systematics and distribution of Cooper's Rocky Mountain snail (*Oreohelix cooperi*) on a sky island in the Canadian Prairies. Canadian Journal of Zoology 97: 833 –840.

Dempsey ZW, Goater CP, Burg TM. 2020. Living on the edge: comparative phylogeography and phylogenetics of Oreohelix land snails at their range edge in Western Canada. BMC Evolutionary Biology 20: 1 –13.

Detling LE. 1968. Historical background of the flora of the Pacific Northwest. Bulletin No. 13. Eugene: Museum of Natural History, University of Oregon.

Emberton KC. 1991a. Ecology of a shell convergence between subfamilies of polygyrid land snails. Biological Journal of the Linnean Society
44: 105 –120.

Emberton KC. 1991b. Polygyrid relations: a phylogenetic analysis of 17 subfamilies of land snails (Mollusca: Gastropoda: Stylommatophora). Zoological Journal of the Linnean Society
103: 207 –224.

Emberton KC. 1994. Polygyrid land snail phylogeny: external sperm exchange, early North American biogeography, iterative shell evolution. Biological Journal of the Linnean Society
52: 241 –271.

Emberton KC. 1995. Sympatric convergence and environmental correlation between two land-snail species. Evolution 49: 469 –475.

Emberton KC, Roth B. 1994. "Extralimital" Land Mollusks (Gastropoda) from the Deep River Formation, Montana: Evidence for Mesic Medial Tertiary Climate. Proceedings of the Academy of Natural Sciences of Philadelphia 145: 93 –106.

Excoffier L, Laval G, Schneider S. 2007. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1: 47 –50.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39: 783 –791.

Felsenstei n J. 1988. Phylogenies from molecular sequences: inference and reliability. Annual Review of Genetics 22: 521 –565.

Fend SV, Gustafson DL. 2001. A new lumbriculid genus and species from North America (Clitellata, Lumbriculidae). In: Pilar Rodriguez and Piet F. M. Verdonschot, eds. Aquatic Oligochaete Biology. Vol. 8. Dordrecht: Springer, 13 –22.

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.

Geller JB, Walton ED, Grosholz ED, Ruiz GM. 1997. Cryptic invasions of the crab *Carcinus* detected by molecular phylogeography. Molecular Ecology 6: 901 –906.

Harrison S, Noss R. 2017. Endemism hotspots are linked to stable climatic refugia. Annals of Botany 119: 207 –214.

Haskell DG, Pan JW. 2013. Phylogenetic analysis of threatened and range-restricted limestone specialists in the land snail genus *Anguispira*. Conservation Genetics 14: 671 –682.

Hendrixson BE, Bond JE. 2007. Molecular phylogeny and biogeography of an ancient Holarctic lineage of mygalomorph spiders (Araneae: Antrodiaetidae: Antrodiaetus). Molecular Phylogenetics and Evolution 42: 738 –755.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high-resolution interpolated climate surfaces for global land areas. International Journal of Climatology: A Journal of the Royal Meteorological Society 25: 1965 –1978.

Hillis DM, Mable BK, Moritz C. 1996. Applications of molecular systematics: the state of the field and a look to the future. In: Molecular systematics. Sunderland: Sinauer Associates, Inc., 515 –543.

Ho SY, Shapiro B. 2011. Skyline-plot methods for estimating demographic history from nucleotide sequences. Molecular Ecology Resources 11: 423 –434.

Idaho Department of Fish and Game. 2017. Idaho state wildlife action plan. In: Dixon R, ed. Boise.

Jetz W, Rahbek C, Colwell RK. 2004. The *COI*ncidence of rarity and richness and the potential signature of history in centres of endemism. Ecology Letters 7: 1180 –1191.

Johnson PJ. 1987. Larval taxonomy, biology, and biogeographyof the genera of North American Byrrhidae. M.S. thesis, University of Idaho, Moscow, ID.

Lichthardt J, Moseley RK. 1994. Ecosystem analysis and conservation planning for the Clearwater refugium, Clearwater and Nez Perce National Forests. Unpublished report for the Nez Perce and Clearwater National Forests. Boise: Idaho Department of Fish and Game, Conservation Data Center.

Linscott TM, Weaver K, Morales V, Parent CE. 2020. Assessing species number and genetic diversity of the Mountainsnails (Oreohelicidae). Conservation Genetics 21: 971 –985.

Lorain C. 1992. Conservation strategy for *Mimulus clivicola* (bank monkeyflower). Unpublished report. Boise: Idaho Department of Fish and Game, Conservation Data Center.

Lucid MK, Ehlers S, Robinson L, Cushman SA. 2018. Beer, brains, and brawn as tools to describe terrestrial gastropod species richness on a montane landscape. Ecosphere 9: e02535.

Lucid MK, Rankin A, Espindola A, Chichester L, Ehlers S, Robinson L, Sullivan J. 2018. Taxonomy and biogeography of *Hemphillia* (Gastropoda: Pulmonata: Arionidae) in North American rainforests, with description of a new species (Skade's jumping-slug, *Hemphillia skadei* sp. nov.). Canadian Journal of Zoology 96: 305 –316.

von Martens E. 1882. Ueber von den Gebrüdern Krause in Amerika gesammelte Conchylien. Sitzungsberichte der Gesellschaft Naturforschender Freunde zu Berlin (November): 138–143.
Merritt TJS, Shi L, Chase MC, Rex MA, Etter RJ, Quattro JM. 1998. Universal cytochrome b primers facilitate intraspecific studies in molluscan taxa. Molecular Marine Biology and Biotechnology 7: 7 – 11.

Metzger G, Espindola A, Waits LP, Sullivan J. 2015. Genetic structure across broad spatial and temporal scales: Rocky Mountain tailed frogs (*Ascaphus montanus*; Anura: Ascaphidae) in the inland temperate rainforest. Journal of Heredity 106: 700 –710.

Minin V, Abdo Z, Joyce P, Sullivan J. 2003. Performance-based selection of likelihood models for phylogeny estimation. Systematic Biology 52: 674 –683.

Molina-Venegas R, Aparicio A, Lavergne S, Arroyo J. 2017. Climatic and topographical correlates of plant palaeo-and neoendemism in a Mediterranean biodiversity hotspot. Annals of Botany 119: 229 – 238.

Moritz C. 1994. Applications of mitochondrial DNA analysis in conservation: a critical review. Molecular Ecology 3: 401 –411.
NatureServe. 2017. NatureServe Explorer: An online encyclopedia of life [web application]. Version 7.1. NatureServe, Arlington, Virginia. Available at: http://explorer.natureserve.org.

Nei M, Jin L. 1989. Variances of the average numbers of nucleotide substitutions within and between populations. Molecular Biology and Evolution 6: 290 –300.

Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Sciences 76: 5269 –5273.

Newell RL, Baumann RW, Stanford JA. 2008. Stoneflies of Glacier National Park and Flathead River basin, Montana. In: Hauer FR, Stanford JA, Newell RL, eds. International advances in the ecology, zoogeography and systematics of mayflies and stoneflies. Entomology Vol. 128, 173 –186. Berkeley, California: University of California Press.

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.

Nordsieck H. 1986. The system of the Stylommatophora (Gastropoda), with special regard to the systematic position of the Clausiliidae, II: Importance of the shell and distribution. Archiv für Molluskenkunde der Senckenbergischen Naturforschenden Gesellschaft 117: 93 –116.

Ohlemüller R, Anderson BJ, Araújo MB, Butchart SH, Kudrna O, Ridgely RS, Thomas CD. 2008. The *COI*ncidence of climatic and species rarity: high risk to small-range species from climate change.

Biology Letters 4: 568 –572.

Peterson AT. 2011. Ecological niches and geographic distributions (MPB-49). New Jersey: Princeton University Press.

Pfeiffer L. 1845 (1846). Description of thirty-six new species of Helix, belonging to the collection of H. Cuming, Esq. Proceedings of the Zoological Society of London 1845: 126 –133.

Pielou EC. 1991. After the ice age. The return of life to deglaciated North America. Chicago: University of Chicago Press.

Pilsbry HA. 1890. Nautilus, 4:3. Proceedings of the Academy of Natural Sciences of Philadelphia. 42: 297, pl. 5, figs. 9–11; reprinted in Nautilus, 5:39.

Pilsbry HA. 1948. Land Mollusca of North America. Academy of Natural Sciences of Philadelphia, Philadelphia Monograph 2: 521 –1113.

Pinceel J, Jordaens K, Backeljau T. 2005. Extreme mtDNA divergences in a terrestrial slug (Gastropoda, Pulmonata, Arionidae): accelerated evolution, allopatric divergence and secondary contact. Journal of Evolutionary Biology 18: 1264 –1280.

Rankin AM, Wilke T, Lucid M, Leonard W, Espíndola A, Smith ML,... & Sullivan J. 2019. Complex interplay of ancient vicariance and recent patterns of geographical speciation in north-western North American temperate rainforests explains the phylogeny of jumping slugs (*Hemphillia* spp.). Biological Journal of the Linnean Society 127: 876 –889.

R Core Team. 2014. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available at: http://www.R-project.org/

Rambaut A, Drummond AJ. 2007. Tracer v1.4. Available at: http://beast.bio.ed.ac.uk/Tracer

Ritchie AM, Lo N, Ho SY. 2017. The impact of the tree prior on molecular dating of data sets containing a mixture of inter-and intraspecies sampling. Systematic Biology 66: 413 –425.

Roberts LNR, Kirschbaum MA. 1995. Paleogeography and the Late Cretaceous of the Western Interior of middle North America; coal distribution and sediment accumulation. U.S. Geological Survey Professional Paper 1561: 115. Shafer AB, Cullingham CI, Cote SD, Coltman DW. 2010. Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. Molecular Ecology 19: 4589 –4621.

Stagliano DM. 2016. Mayflies (Insecta: Ephemeroptera) of conservation concern in Montana: status updates and management needs. Western North American Naturalist 76: 441 –451.

Stagliano DM, Stephens GM, Bosworth WR. 2007. Aquatic invertebrate species of concern on USFS Northern Region lands. Report prepared for USDA Forest Service, Northern Region, Missoula, MT. Boise and Helena: Montana Natural Heritage Program and Idaho Conservation Data Center. Agreement number 05-CS-11015600-036. 95 pp. + appendixes.

Stark BP, Gustafson DL. 2004. New species and records of Soliperla Ricker, 1952 from western North America (Insecta, Plecoptera, Peltoperlidae). SPIXIANA 27: 97 –105.

Stewart JR, Lister AM. 2001. Cryptic northern refugia and the origins of the modern biota. Trends in Ecology & Evolution 16: 608 –613.

Swofford DL. 2003. PAUP: phylogenetic analysis using parsimony. Mac Version 3. 1. 1. Available at: https://paup.phylosolutions.com/.

Tajima F. 1989a. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585 –595.

Tajima F. 1989b. The effect of change in population size on DNA polymorphism. Genetics 123: 597 –601.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution 30: 2725 –2729.

Thomaz D, Guiller A, Clarke B. 1996. Extreme divergence of mitochondrial DNA within species of pulmonate land snails. Proceedings of the Royal Society of London. Series B 263: 363 –368.

Thuiller W, Lafourcade B, Engler R, Araújo MB. 2009. BIOMOD–a platform for ensemble forecasting of species distributions. Ecography 32: 369 –373.

Uminski T. 1963. Taxonomy of *Anguispira* marmorensis (H.B. Baker, 1932) and notes on the taxonomy of the genera *Anguispira* (Morse) and *Discus* (Fitzinger) (Gastropoda, Endodontidae). Annales Zoologici Musei Polonici Historiae Naturalis 21: 81 –91.

Vieites DR, Min MS, Wake DB. 2007. Rapid diversification and dispersal during periods of global warming by plethodontid salamanders. Proceedings of the National Academy of Sciences 104: 19903 –19907.

Wahlund S. 1928. Zusammensetzung von Populationen und Korrelationserscheinungen vom Standpunkt der Vererbungslehre aus betrachtet. Hereditas 11: 65 –106.

Wakeley J. 1996a. The variance of pairwise nucleotide differences in two populations with migration. Theoretical Population Biology 49: 39 –57.

Wakeley J. 1996b. Distinguishing migration from isolation using the variance of pairwise differences. Theoretical Population Biology 49: 369 –386.

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. Theoretical Population Biology 7: 256 –276.

Zwickl DJ. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph. D. Thesis, The University of Texas, Austin.

Figure 3.1: Simulated historical scenarios tested in DIYABC for (A) five *A. k. occidentalis* BAPS clusters and (B) two *A. nimapuna* BAPS clusters. In these scenarios, t represents timescale in terms of the number of generations and width of the graph represents relative effective population size during the time period (e.g., 0 – t1).

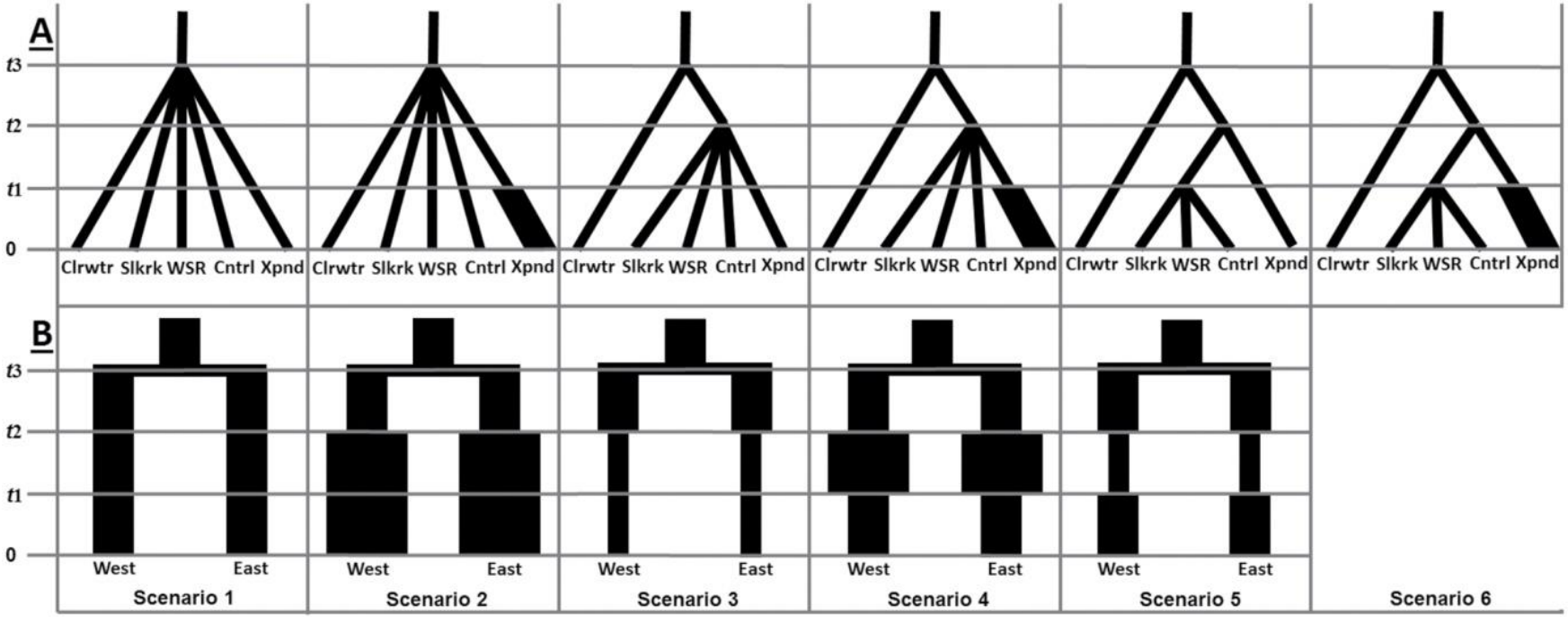


Figure 3.2: Best maximum likelihood phylogeny for *Anguispira* and *Discus* species based on concatenated *COI*, *cytb* and *16S* mtDNA sequences. Bootstrap values on nodes indicate relationships that are well supported (≥ 70).

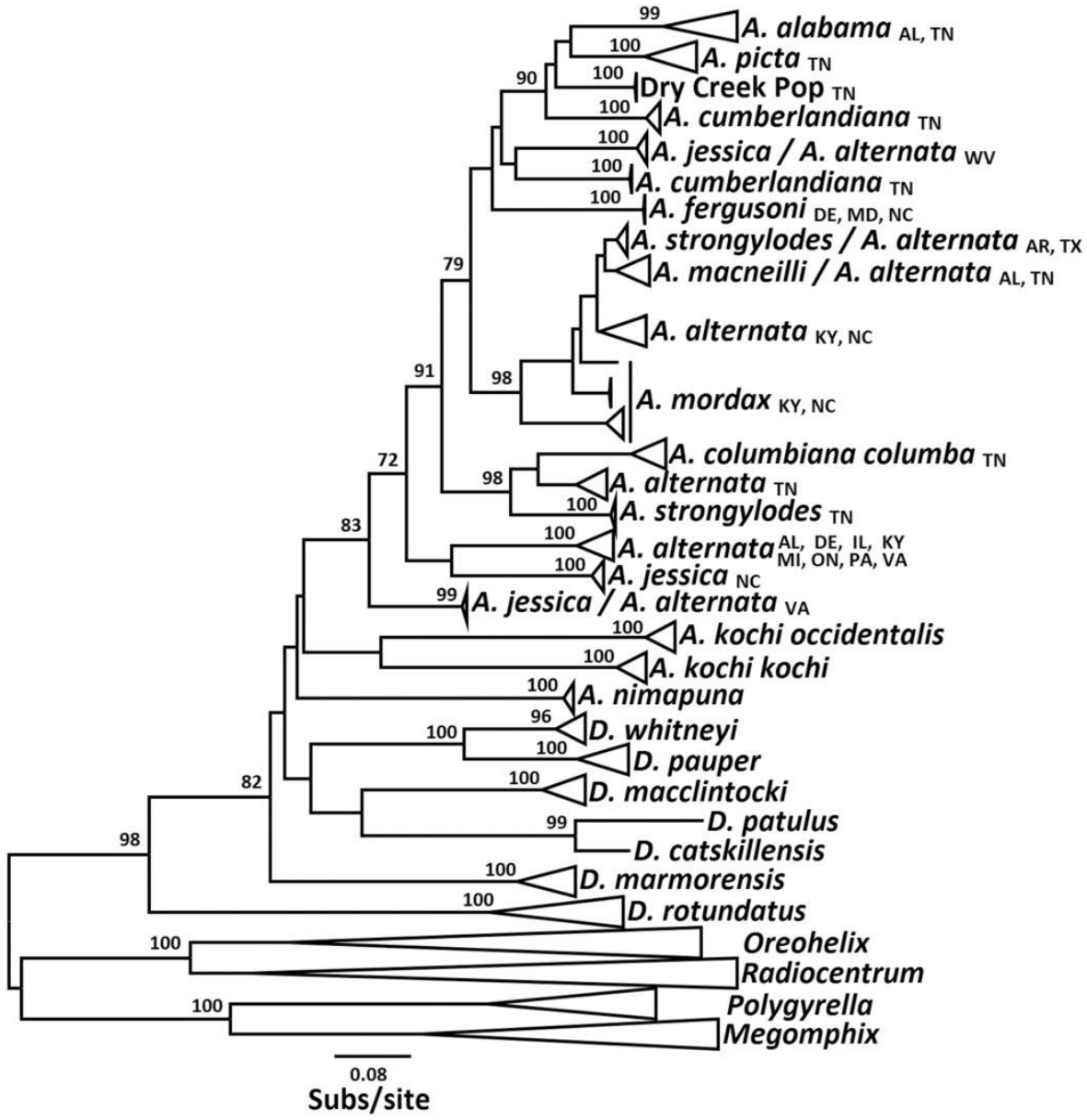


Figure 3.3: Maximum parsimony phylogeny of the interspecific *COI-cytb-16S* haplotypes data set. Numbers above the branches represent bootstrap values (500 replicates).

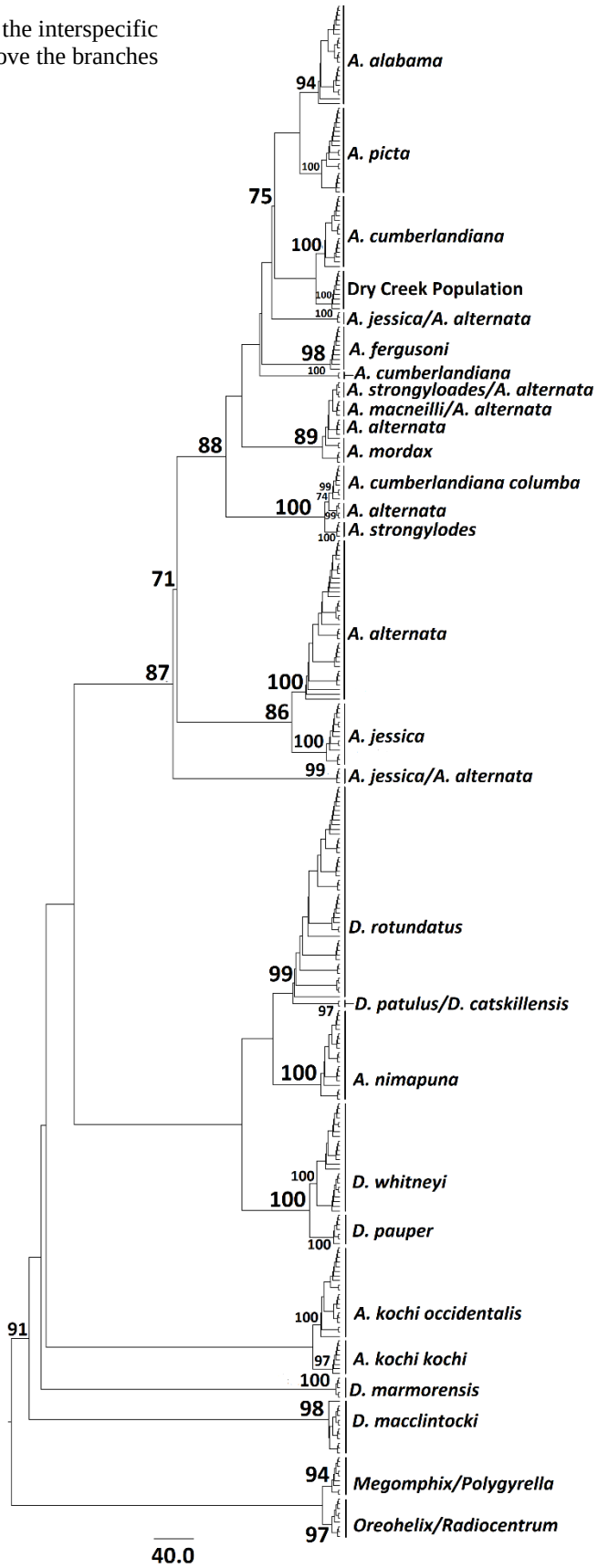


Figure 3.4: Bayesian phylogeny of the interspecific *COI-cytb-16S* haplotypes data set. Numbers above the branches represent posterior probabilities.

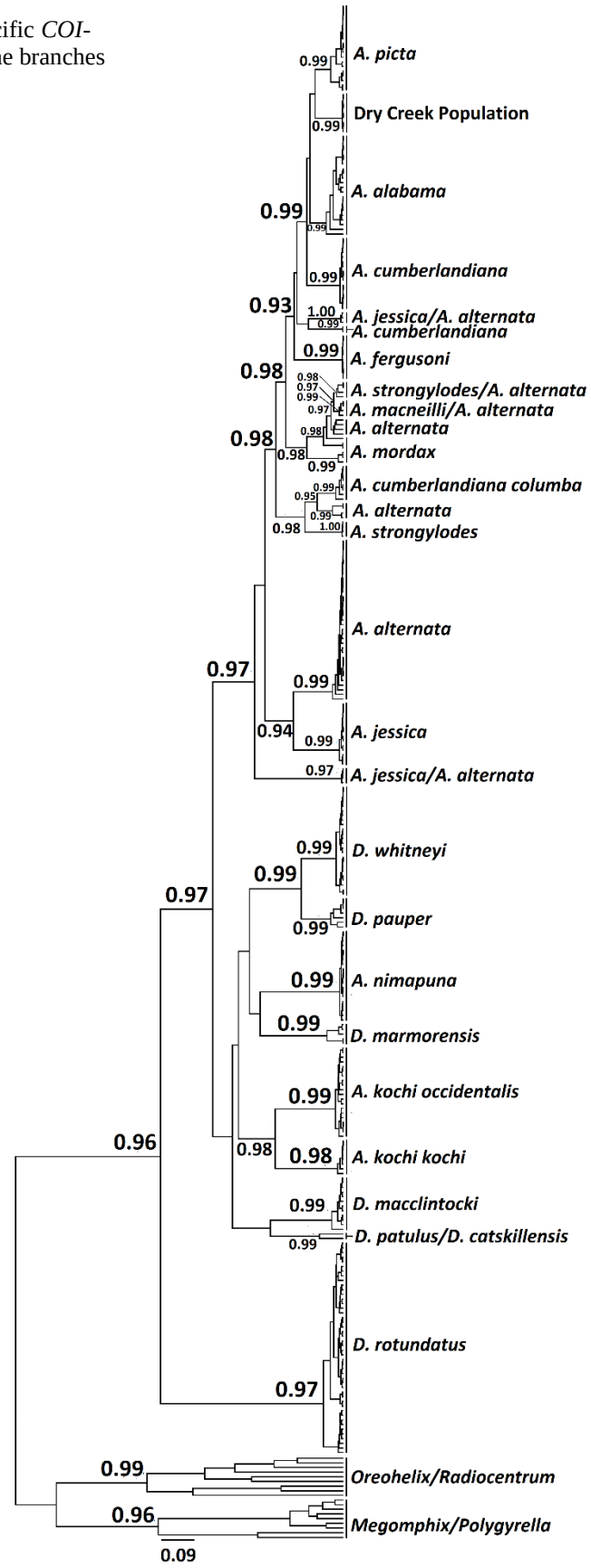


Figure 3.5: Best maximum likelihood phylogeny for *A. k. occidentalis* based on mtDNA haplotypes. Bootstrap values on nodes indicate relationships that are well supported (≥ 70). BAPS clusters that correspond to the best partition of the data are indicated by different colored bars.

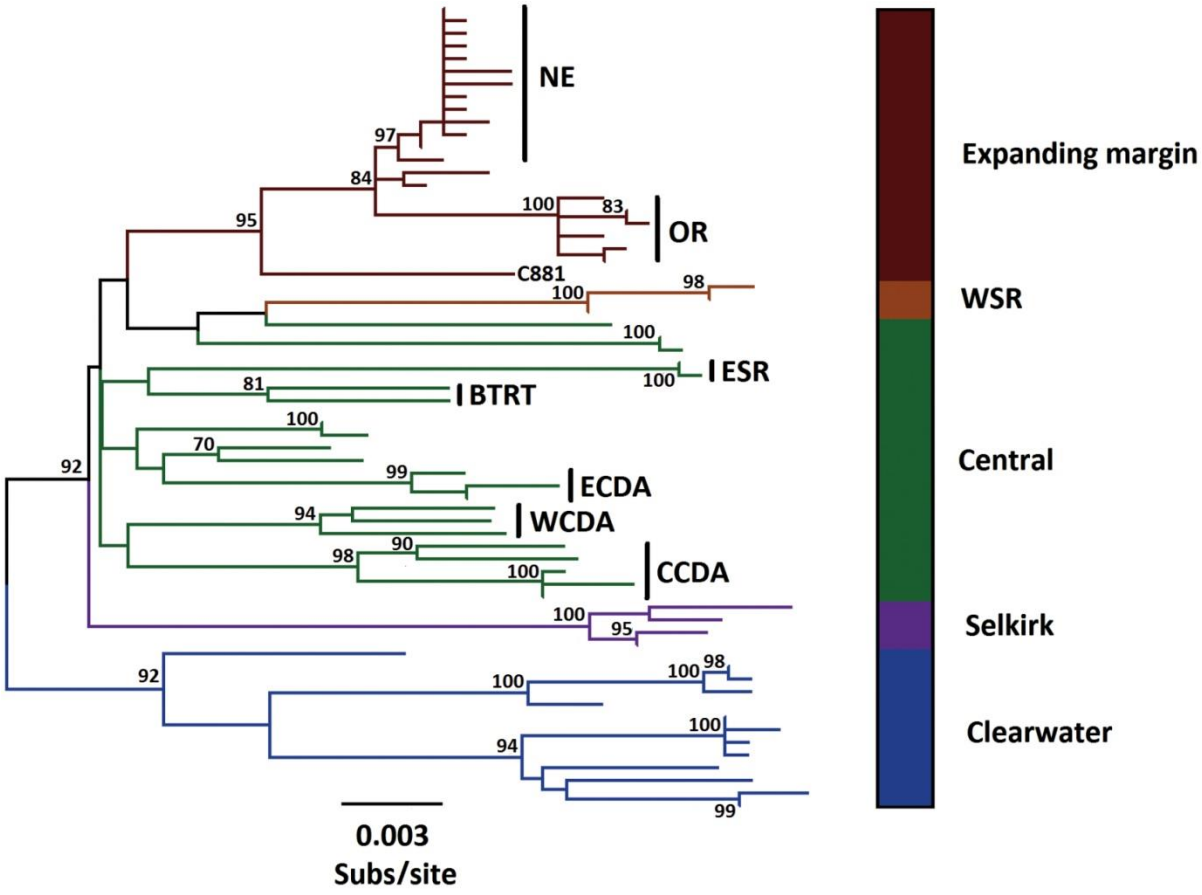


Figure 3.6: Bayesian phylogeny (constant population size tree prior) based on the mtDNA *A. k. occidentalis* all alleles data set. Circles on branches indicate Bayesian posterior probability ≥ 0.90. BAPS clusters that correspond to the best partition of the data are indicated by different colored bars. Map showing the locations of sampled individuals. State and province names are abbreviated as follows: AB, Alberta; BC, British Columbia; ID, Idaho; MT, Montana; OR, Oregon; WA, Washington.

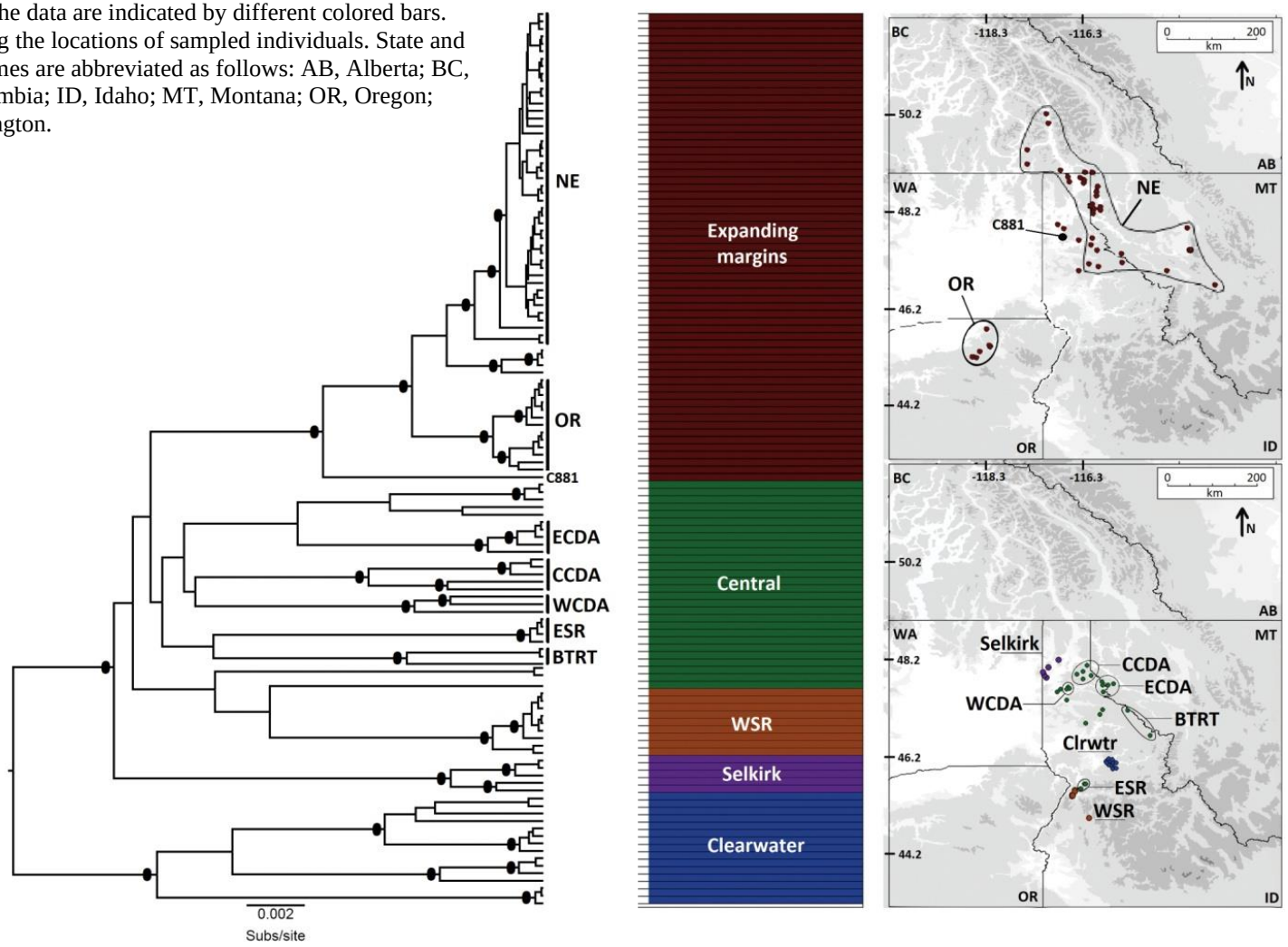


Figure 3.7: Uncorrected nucleotide diversity (π) and Watterson estimator (θ) for the *COI-cytb-16S A. k. occidentalis* data set plotted against sample size.

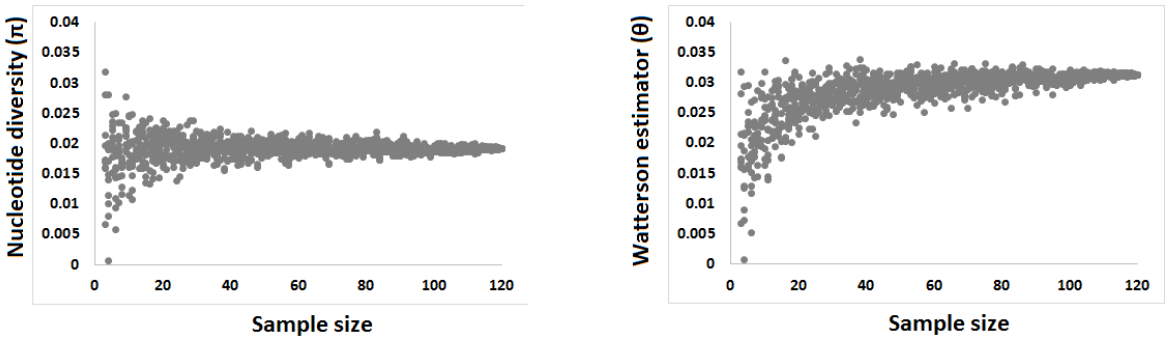
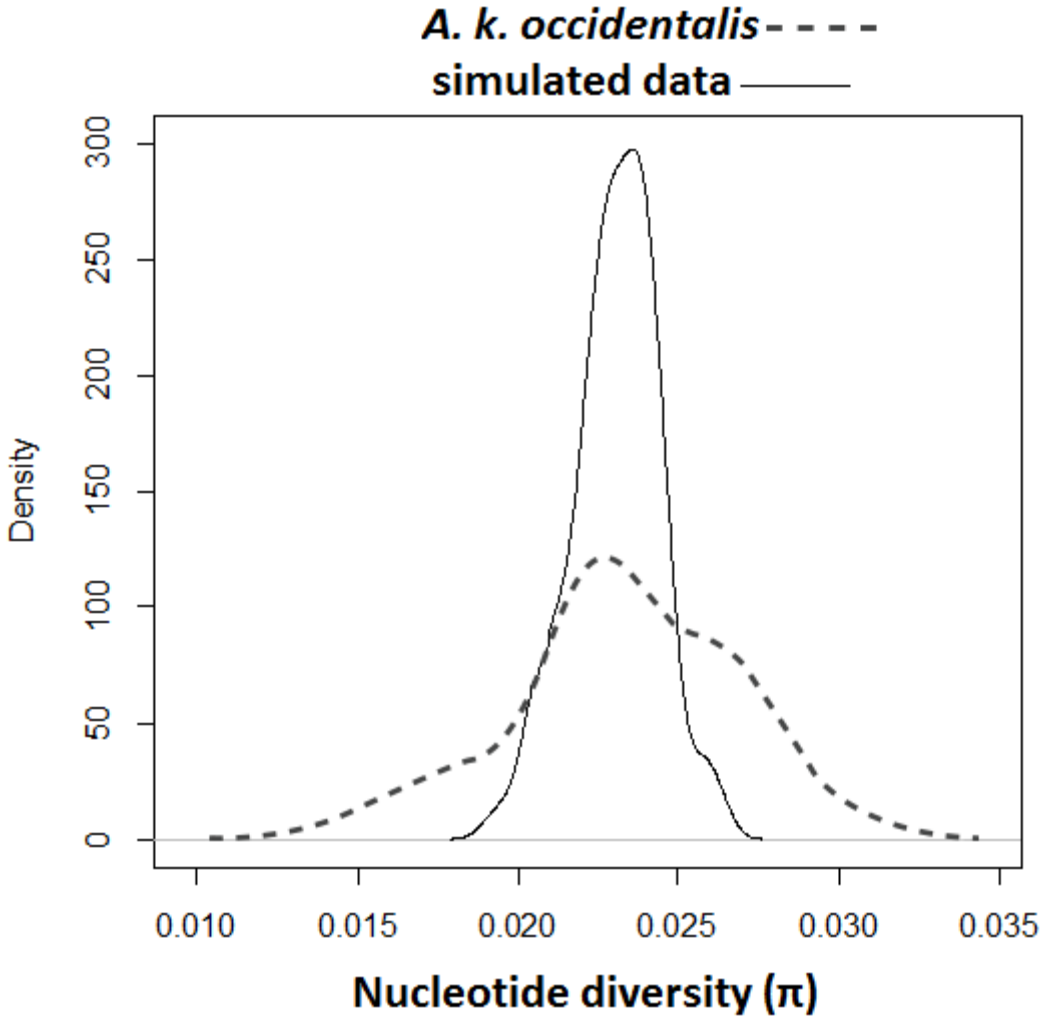


Figure 3.8: The distribution curve of π for the *COI-cytb-16S A. k. occidentalis* data set compared to that of a simulated data set (of a single, neutrally evolving population) produced to have an identical mean.



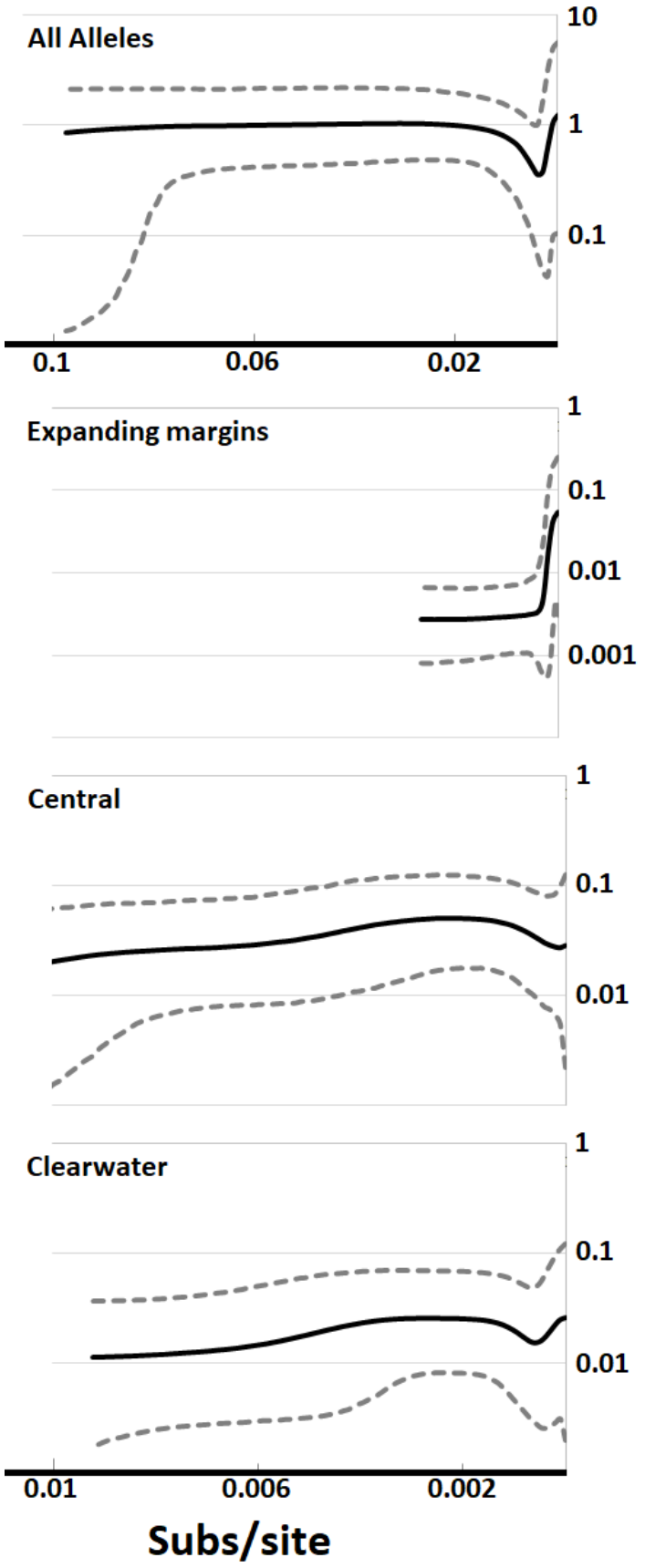
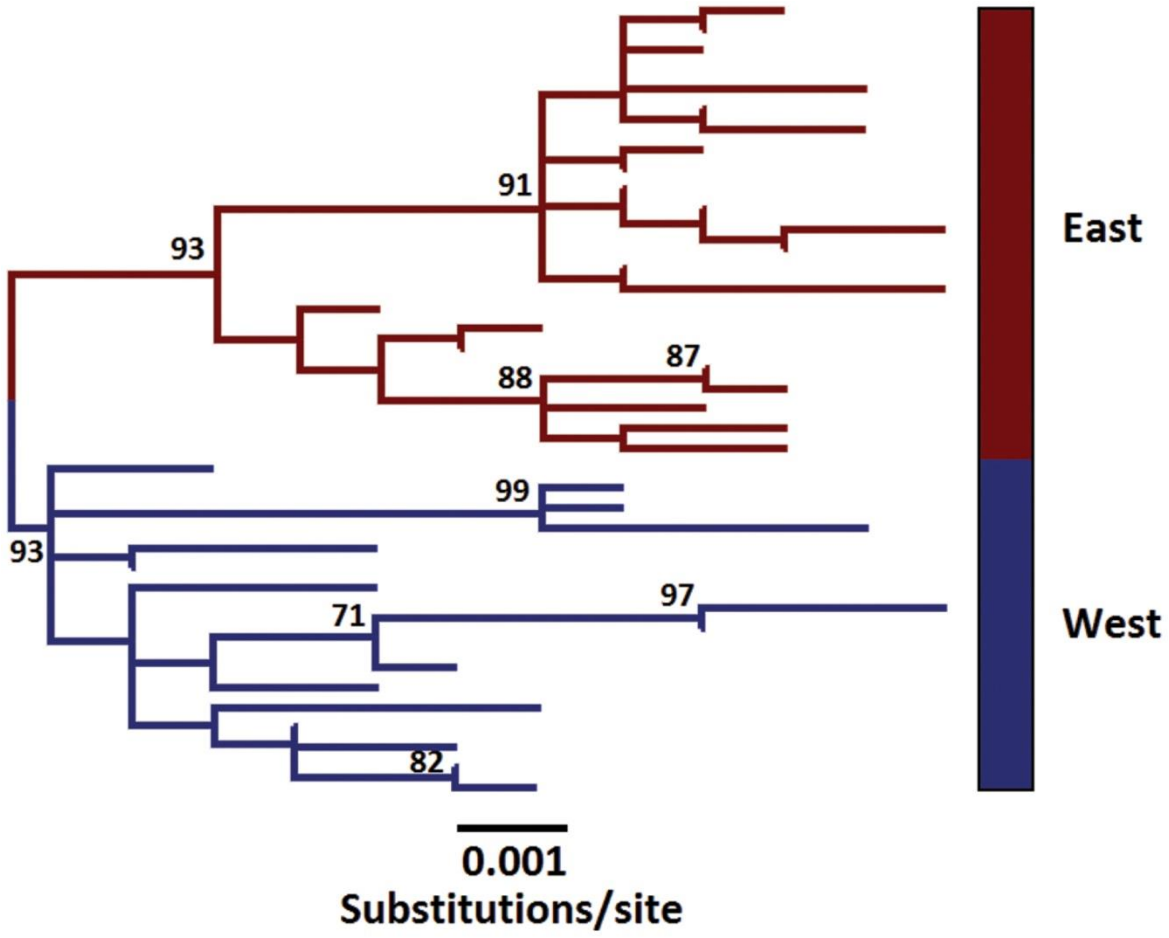


Figure 3.9: Bayesian skyline plots for all *A. k. occidentalis* alleles (N = 120) and BAPS clusters (Expanding margins, N = 63; Central, N = 28; and Clearwater, N = 15) showing effective population size (scaled by mutation rate) through time. Note the time scale is given in substitution per site, which can be converted to units of time using a molecular clock calibration. Black lines indicate the median value while grey dashed lines denote the 0.95 highest posterior probability intervals.

Figure 3.10: Best maximum likelihood phylogeny for *A. nimapuna* based on mtDNA haplotypes. Bootstrap values on nodes indicate relationships that are well-supported (≥ 70). BAPS clusters that correspond to the best partition of the data are indicated by different colored bars.



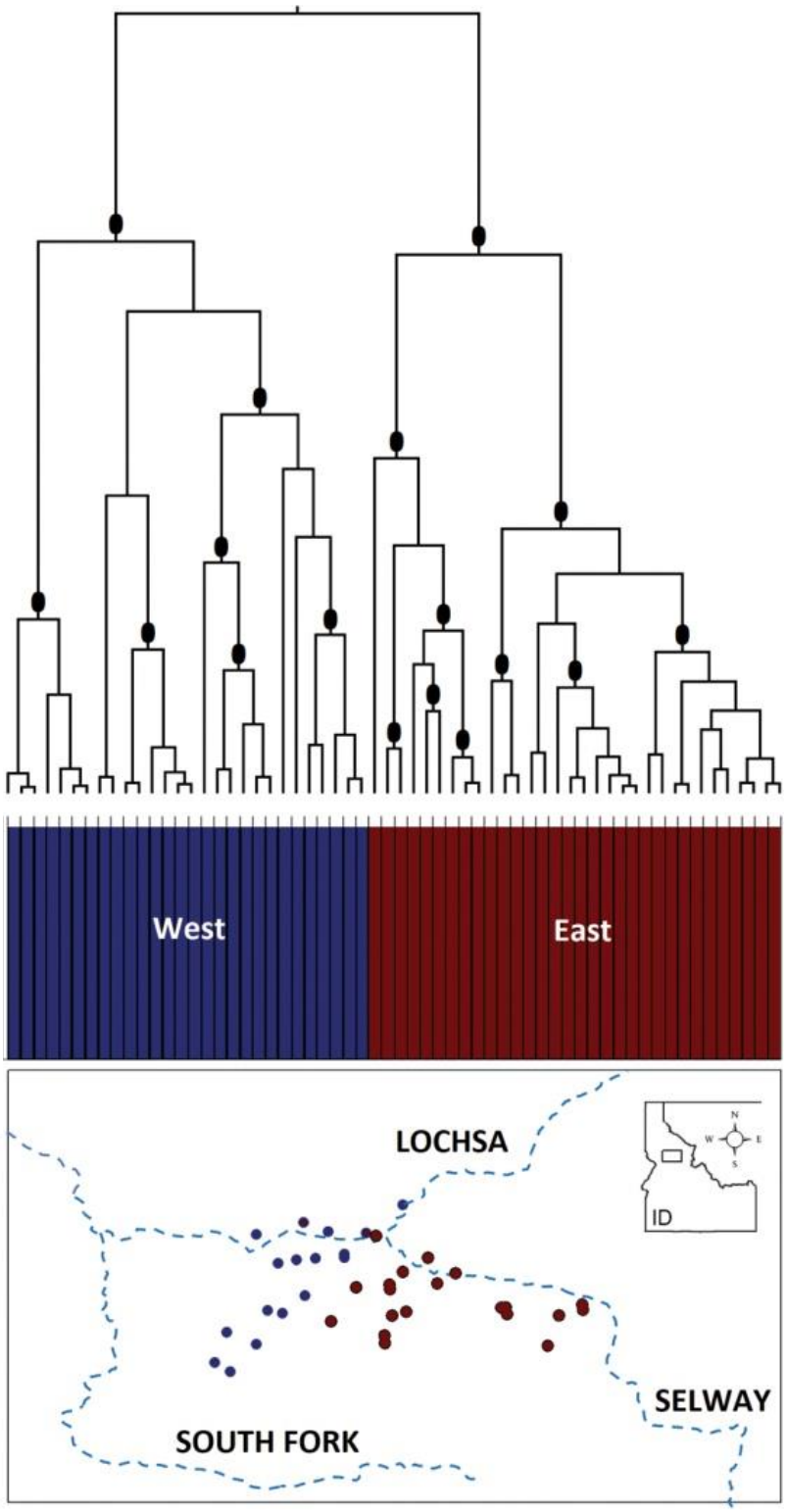
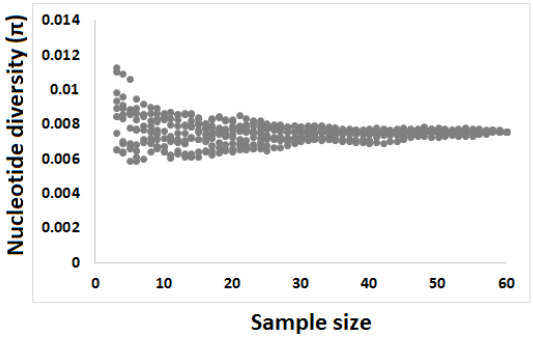


Figure 3.11: Bayesian phylogeny (constant population size tree prior) based on the mtDNA *A. nimapuna* all alleles data set. Circles on branches indicate Bayesian posterior probability ≥ 0.90. BAPS clusters that correspond to the best partition of the data are indicated by different colored bars. Map showing the distribution of the *A. nimapuna* sampling localities and two mtDNA clades along the Lochsa, Selway and South Fork rivers in the Clearwater River Drainage of Idaho County, ID.

Figure 3.12: Uncorrected nucleotide diversity (π) and Watterson estimator (θ) for the *COI-cytb-16S A. nimapuna* data set plotted against sample size.



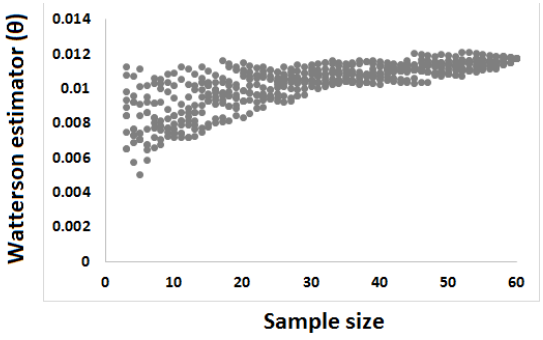


Figure 3.13: The distribution curves of π for the *COI-cytb-16S A. nimapuna* data set compared to that of a simulated data set (of a single, neutrally evolving population) produced to have an identical mean.

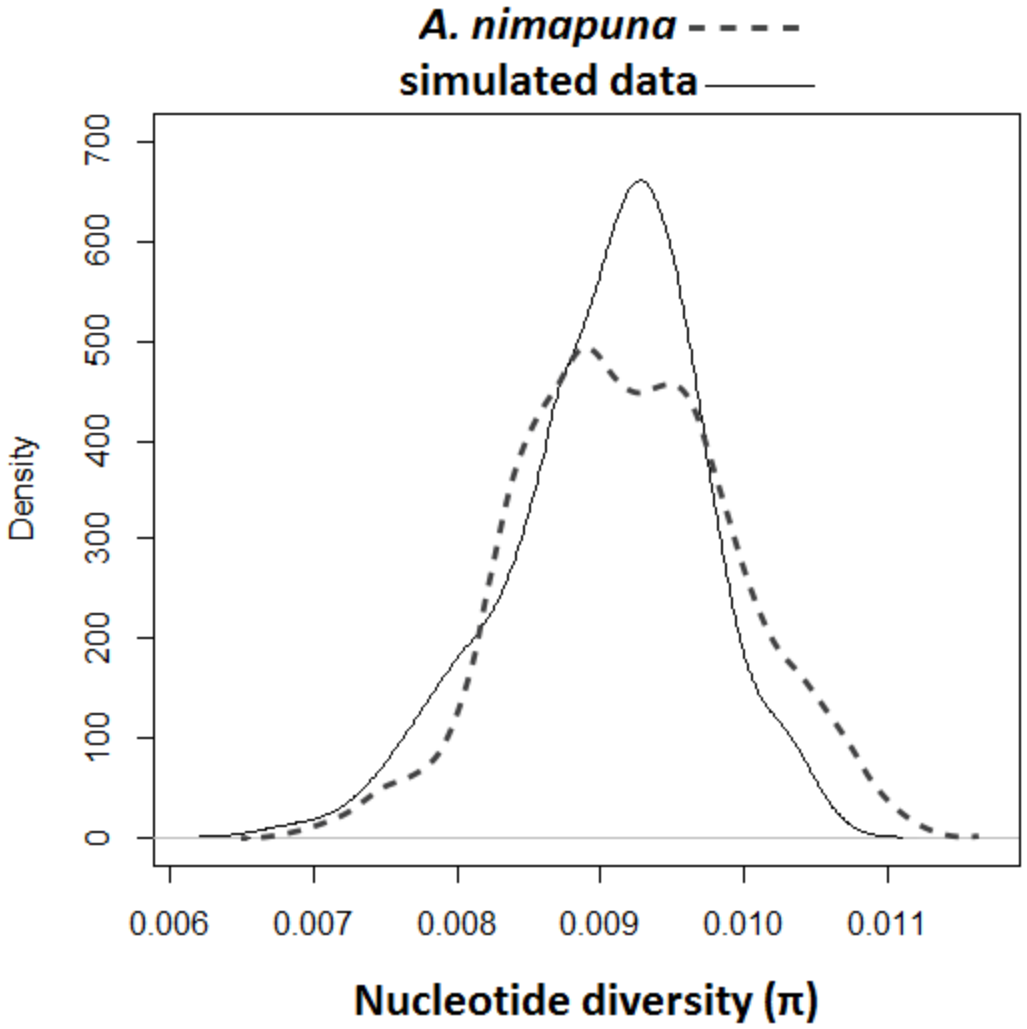
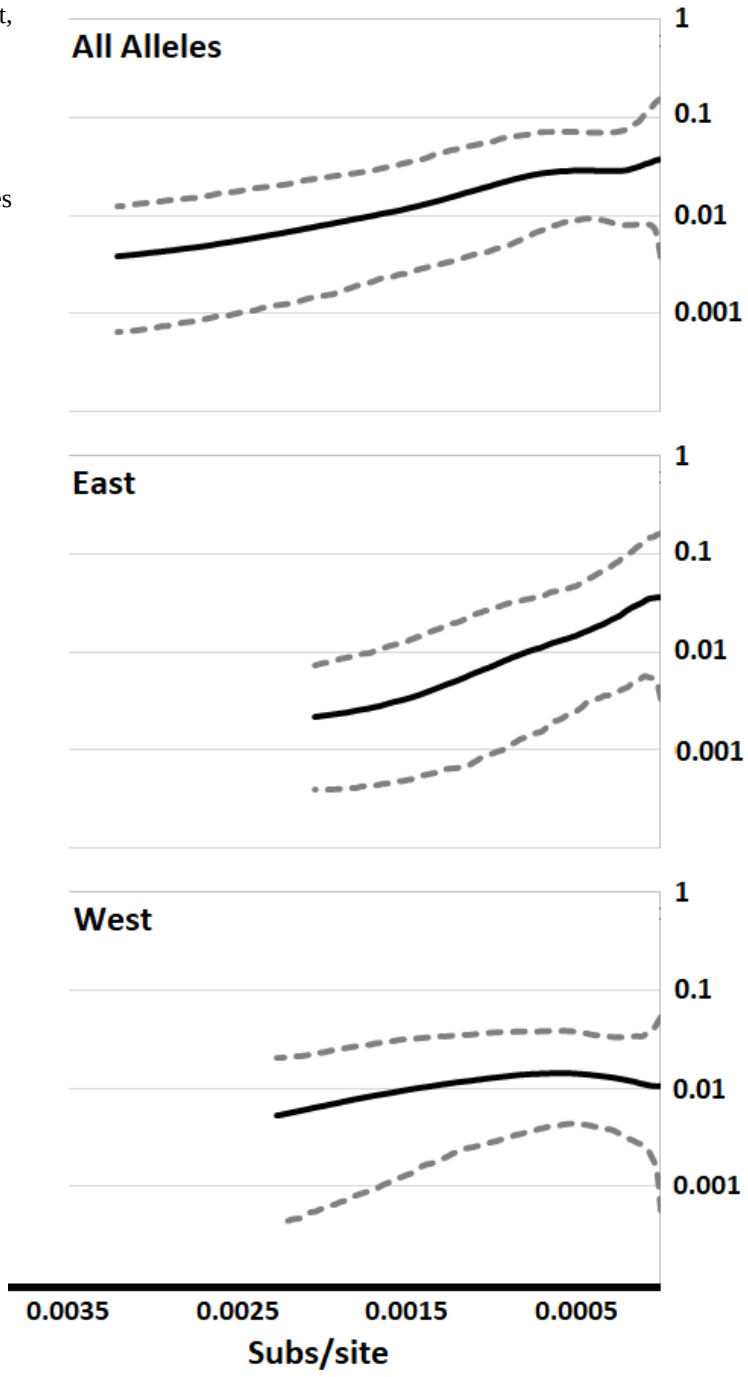


Figure 3.14: Bayesian skyline plots for all *A. nimapuna* alleles (N = 60) and BAPS clusters (East, N = 32; and West, N = 28) showing effective population size (scaled by mutation rate) through time. Note the time scale is given in substitution per site, which can be converted to units of time using a molecular clock calibration. Black lines indicate the median value while grey dashed lines denote the 0.95 highest posterior probability interval.



Chapter 4: Comparative phylogeographic analysis of three endemic roundback slugs (Arionidae) in the North American inland temperate rainforest

Abstract

The Northern Rocky Mountain region was heavily affected by Pleistocene glacial-interglacial successions that impacted population-level processes due to the creation and erosion of geographic and habitat barriers. Current day forest patches harbor impressive numbers of endemics, and within this framework, populations of co-distributed species might be expected to have differentiated in response to the same historical events and thus show similar patterns of genetic divergence. Here, we conduct a comparative analysis of three species of regional endemic slugs with overlapping distributions. Results suggest largely non-congruent histories, and highlights that codistributed species likely exhibit distinct patterns that reflect unique phylogeographic histories.

Introduction

Phylogeographic comparisons among co-distributed species shed light on the generality or specificity of species' responses to biogeographical processes and can improve predictions of responses to future environmental changes. Because multiple species within communities share a common history, we might expect co-distributed species to show similar phylogeographic patterns, especially among closely related taxa that might also maintain similar population sizes, generation times, and molecular evolutionary rates. Indeed, common structure and divergence times have been documented among co-occurring, closely related taxa such as some mammal communities (Arbogast and Kenagy, 2001) and some bird communities (Ralston et al., 2021). However, comparative studies oftentimes show that taxa vary in their patterns of genetic differentiation (e.g., Ritter et al., 2021). In any case, understanding how biogeographical forces have either similarly or differentially impacted the evolution of biotic assemblages allows us to distill a regional biogeographic perspective and may help to focus conservation efforts on regions containing endemic species.

The North American inland temperate rainforest, composed of discontinuous bands of mesic patches across the Northern Rocky Mountains (NRM), is a unique ecosystem (DellaSala, 2011) that is home to a rich biodiversity of organisms, including many endemics. This region was heavily impacted by Quaternary climatic oscillations (Richmond et al., 1965; Waitt & Thorson, 1983; Delcourt & Delcourt, 1993) which led to episodic contractions and expansions of available habitats, and ultimately impacted the geographic distribution of genetic diversity of regional taxa (e.g., Good

& Sullivan, 2001, Carstens et al., 2005, Metzger et al., 2015; Rankin et al., 2021). Many single-taxon studies have provided novel insights into regional patterns, such as putative refugia spots and post-glacial colonization routes, which has provided a rich setting for comparative studies. Here, we analyze patterns of geographic genetic structure of three codistributed endemic slug species to assess whether taxa with overlapping distributions share congruent phylogeographic structure. This will enable determination of the concerted or idiosyncratic responses of independent lineages to biogeographic processes in this region.

Slug and snail endemism in the NRM is high, and new species are continuing to be discovered and described (Leonard et al., 2011; Burke, 2013; Lucid et al., 2018a). In northern Idaho and adjacent areas, gastropod biodiversity has been extensively surveyed (Lucid et al., 2018b) and the number of genetic studies on these organisms is beginning to increase (e.g., Smith et al., 2018; Rankin et al., 2019). Here, we extend our knowledge to include three roundback slug species, *Udosarx lyrata*, *Magnipelta mycophaga*, and *Kootenaia burkei*, using both mtDNA and genomic SNP datasets. *U. lyrata* is found in scattered patches within the Bitterroots and north into the Coeur d'Alene Mts and parts of the Clearwater and Lolo National Forests. *M. mycophaga* has a more widespread but patchy distribution, and is found in the Selkirk, Purcell, Cabinet, Saint Joe, and Bitterroot Mts. *K. burkei* has a similar distributional range as *M. mycophaga* but does not extend further south than the Saint Joe Mts and occurs at higher abundances (Burke, 2013; Lucid et al., 2018b). Each species is the only representative of their respective genera, but all belong to the same family (Arionidae). In this study, we evaluate the genetic diversity and phylogeographic structure of these species to determine to what extent the diversity and structure of each species is congruent, which will allow for incorporating information on genetic diversity into regional evaluations of biodiversity.

Methods and Materials

Sampling and laboratory methods

Most samples were collected from 2010 to 2014 during the Multi-species Baseline Initiative (https://idfg.idaho.gov/baseline) survey for terrestrial gastropods in the northern regions of Idaho and adjacent portions of Northeast Washington and Northwest Montana, USA (see Lucid et al., 2018b). Additional samples were provided by the Montana Natural Heritage Program (Missoula, MT). Genomic DNA was extracted from ethanol-preserved tissue with an EZNA Blood and Tissue Kit (Omega Bio-tek, Doraville, CA) following the manufacturer's protocol. Total genomic DNA was qualitatively assessed for each sample on a 1.5% agarose gel and quantified using a Qubit fluorometer.

Regions of two mitochondrial genes—cytochrome *c* oxidase subunit I (*COI*) and *16S* ribosomal RNA—were amplified via PCR and Sanger sequenced following Rankin et al. (2021). Resulting *COI* and *16S* sequences were concatenated for each individual to produce a mtDNA dataset. In total, mtDNA data was obtained from 197 individuals (*U. lyrata* = 57, *M. mycophaga* = 70, and *K. burkei* = 70).

A genotyping by sequencing (GBS) library was prepared for 174 individuals (*U. lyrata = 40*, *M. mycophaga* = 43, and *K. burkei* = 91) using a *PstI* cutting enzyme (New England Biolabs, MA, USA) following the protocol of Elshire et al. (2011) with some modifications suggested by Smith and Carstens (2020). Briefly, digestion and individual-barcode-adapter ligation were performed on 200 ng of DNA from each sample. Samples were then pooled, purified, and amplified via PCR (see Elshire et al., (2011) for primer sequences). Products were separated on a 1.5% agarose gel to ensure fragments within 400—500 bp size range were recovered. The resulting libraries were size selected and then sequenced on an Illumina Hi-Seq 4000 at QB3 genomics (Berkeley, CA), generating single-end reads of 50 bp.

We processed each species' raw Illumina reads separately in ipyrad (Eaton and Overcast, 2020) to demultiplex samples (using their unique barcodes), remove adapter sequences and low-quality bases, and align reads. Default ipyrad parameters were used except that we specified a *clustering threshold* of 0.91 and the *minimum number of samples for a specific locus* set to 33%. These parameters were chosen to obtain a moderate amount of missing data with a reasonable number of shared loci.

Interspecific mtDNA phylogeny

To place our results into a broader evolutionary context, we reconstructed a mtDNA phylogeny to include data from the three focal species as well as 19 additional arionid slug species that are also endemic to the western United States, including two species of *Ariolimax*, two species of *Zacoleus*, eight species of *Hemphillia*, and seven species of *Prophysaon*. An interspecific *COI-16S* data matrix was assembled, pruned to include ~3 individuals per species, and aligned using the MAFFT online server (http://www.ebi.ac.uk/Tools/msa/mafft/). The final matrix consisted of 75 individuals and 1093 bp with 526 parsimony informative sites.

We employed the automodel command in PAUP* (Swofford, 2003) to select a model of nucleotide sequence evolution under the Bayesian Information Criterion and decision theory (Minin et al., 2003). The GTR+I+Γ model was determined to be best fitting. Maximum likelihood (ML) analyses were performed in Garli (Zwickl 2006), conducting 100 replicate runs with random starting trees. Nodal support was then assessed with 200 bootstrap replicates.

A Bayesian analysis was performed with BEAST v2.4.4 (Bouckaert et al., 2014) using the GTR+I+Γ model, a relaxed lognormal molecular clock, and a birth-death speciation tree prior. In addition, although molecular clocks are oftentimes erratic, evolutionary time frames can sometimes provide useful insights, so we applied rate calibrations for both *COI* and *16S*. That is, we specified a normally distributed molecular clock rate prior, with a mean *COI* site substitution rate of 0.08 per million years (SD: 0.03), and a mean *16S* site substitution rate of 0.07 per million years (SD: 0.04). The 95% interval of these distributions was set to include previously reported clock rates in gastropods (Thomaz et al., 1996; Chiba, 1999; Van Riel et al., 2005). The analysis consisted of 100 million generations with a sampling interval of 10,000 and a burn-in of 25%. The resulting tree was used as input for the tree-based species delimitation method, the Generalized Mixed Yule Coalescent (GMYC; Pons et al. 2006), which is part of the R package splits (R Core Team, 2013).

Intraspecific mtDNA and GBS phylogeography

We first identified major mtDNA clades within each species by performing ML analyses in Garli (as above) but using the GTR+I+G model for *U. lyrata*, HKY+G model for *M. mycophaga*, and HKY+I model for *K. burkei* (selected using PAUP* as above). Then, to estimate the most likely number of mtDNA clusters present within each species, we used the Bayesian assignment analysis program BAPS 6.0 (Corander et al., 2008), choosing the options *clustering with linked loci* and the *codon linkage model*. We tested K=1—12, performing ten replicates for each K.

With each species' full SNP dataset, we produced an SVD-quartets phylogeny (Chifman and Kubatko, 2015) implemented in PAUP*. All possible quartets were sampled, and 100 bootstrap replicates were carried out. We then exported uncorrected P-distance matrices with the *savedists* command, and distance matrices were then used to estimate a phylogenetic network with the NeighborNet algorithm implemented in SplitsTree v5 (Huson & Bryant, 2005).

We applied two methods to infer overall patterns of population genetic structure based only on unlinked SNPs (from the *ustr* output file from ipyrad). First, we used the model-based approach implemented in ADMIXTURE v1.3.0 (Alexander, Novembre, & Lange, 2009) to infer individual ancestries in a ML framework. We estimated ancestry coefficients in 10 replicate runs for each K=1—12, and performed two replicate runs of 10-fold cross-validation error for each K. We then plotted the log-likelihood and cross-validation errors and identified the K-value having the lowest cross-validation error as best fitting. Lastly, a two-dimensional Principal Coordinate Analysis (PCoA) was performed on P-distance matrices (produced from unlinked SNPs) using the *cmdscale* command of the R statistical program (R Development Core Team).

Results

Interspecific mtDNA phylogeny and species delimitation

In both ML and Bayesian phylogenies (Figures 4.1 & 4.2), *U. lyrata* individuals form a highly supported monophyletic group (bootstrap=88 and Posterior Probability [PP]=0.99) that is recovered as sister to *Zacoleus* (bootstrap=99 and PP=1.00). The *Udosarx*/*Zacoleus* clade clusters with *Prophysaon* in the ML phylogeny (with <50 bootstrap support) but clusters with all other species in the Bayesian phylogeny (PP=0.88). There is a *M. mycophaga*, *K. burkei*, and *Ariolimax* clade in both trees (bootstrap=54 and PP=1.00) but these three groups are separated by long internal branches. This clade, in turn, clusters with *Hemphillia* in both trees (bootstrap=55 and PP=0.99).

Based on assumed substitution rate priors in our BEAST run, the analysis finished with a posterior mean rate of 0.115 substitutions/site/Myr for *COI* and 0.027 substitutions/site/Myr for *16S*. However, the 95% highest posterior density (HPD) intervals are large: 0.017—0.18 for *COI* and 0.01—0.045 for *16S*, indicating the time scale should be interpreted with caution (Fig. 4.2). Regardless, the Bayesian phylogeny shows *M. mycophaga* and *K. burkei* as having short intraspecific branch lengths, like the depth seen within other species, but branch lengths within *U. lyrata* are much longer and similar to the depth seen between species (Fig. 4.2). Indeed, the GMYC method delimited 28 species (Fig. 4.2) with several putative species noted for *U. lyrata*. Because there was evidence of multiple species represented in *U. lyrata*, we subsequently subjected the *U. lyrata* mtDNA dataset to the distance-based Automatic Barcode Gap Discovery method (ABGD; Puillandre et al. 2012) through the online server (https ://bioin fo.mnhn.fr/abi/public/abgd/abgdw eb.html) with the default settings.

Intraspecific mtDNA and GBS phylogeography

Udosarx lyrata—We identified 46 mtDNA haplotypes among 57 individuals. There is a total of 877 bp with 288 variable sites and 226 parsimony informative sites. The total number of retained GBS loci was 28,198 (with 50% missing data) and the mean number of variable sites per locus was 3.67.

Seven partitions were inferred by ABGD (sp1—7; Fig. 4.3). These seven partitions also appear in the best ML tree as highly supported clades, which also generally agree with the five clusters inferred by BAPS (Fig. 4.3). Specifically, the ABGD partitions sp1 and sp2 correspond to BAPS cluster 1. Likewise, the partitions sp3 and sp4 correspond to BAPS2. Partition sp5 includes two specimens, *C1545* & *MT1*, which were not included in the BAPS analysis due to only having *16S* (i.e., *COI* was not obtained for those two specimens). Partition sp6 corresponds to BAPS3. Lastly, partition sp7 corresponds to BAPS clusters 4 and 5, and includes 38 of the 57 individuals in the

dataset. Results of the NeighborNet and PCo are generally concordant with the mtDNA results (Fig. 4.3), except for one mtDNA phylogroup that exhibits a different pattern in the genomic data.

Within partition sp7 (BAPS clusters 4 and 5) there are four highly supported clades present in the mtDNA tree, hereafter referred to as lyr-a—d (Fig. 4.3). The lyr-d clade corresponds to BAPS5 and has a parapatric distribution with the lyr-c clade. The latter clade, along with lyr-a and lyr-b, all correspond to BAPS4 but form a paraphyletic group in the ML tree and are all geographically allopatric. In the NeighborNet and PCo results, the clades lyr-a, lyr-c, and lyr-d clump together but lyr-b clusters apart. Thus, lyr-b exhibits some mitonuclear discordance.

In the ADMIXTURE analysis, likelihood scores increased with increasing K and do not appear to plateau by K=12 (Fig. 4.4). Furthermore, no sharp decrease in cross-validation error estimates is observed but reach a low point at K=3. As such, we show results for K=3—5 and a corresponding map is shown depicting results for K=5 (Fig. 4.4). At K=3, the data are partitioned so that clades lyr-a, lyr-c, and lyr-d form a cluster, lyr-b forms a second cluster, and the remainder of individuals form a third cluster with some evidence of admixture with lyr-b. At K=4, individuals belonging to partition sp6 (BAPS3) separate. At K=5, individuals belonging to BAPS1 and BAPS2 separate. Results for K=5 is also depicted on the SVD-quartets tree (Fig. 4.5)

Magnipelta mycophaga—We identified 25 mtDNA haplotypes among 70 *M. mycophaga* sequences. There is a total of 917 bp with 72 variable sites and 58 parsimony informative sites. The total number of retained GBS loci was 49,869 (with 50% missing data) and the mean number of variable sites per locus was 1.1.

The best ML tree revealed three diverged, well-supported mtDNA clades, with intralineage diversity being relatively low compared to interlineage divergence (Fig. 4.6). BAPS results were concordant, identifying the same three clusters. The three clades roughly follow a north-to-south geographical pattern and are hereafter referred to as the South, Central, and North clades. The South and North clades are recovered as sister in the mtDNA tree (bootstrap=85). Individuals in the South clade were collected in parts of the Bitterroot, Saint Joe, and Coeur d'Alene Mts while representatives of the Central clade are from the Coeur d'Alene and Cabinet Mts. Finally, individuals in the North clade were collected from the Cabinet, Purcell, and Selkirk Mountains. Results of the NeighborNet and PCo analyses strongly support the mtDNA results (Fig. 4.6).

Cross-validation error estimates for the ADMIXTURE models decreased sharply from K=1 and reach a low point at K=4 (Fig. 4.7). Likewise, likelihood scores increased dramatically from K=1 and appear to plateau at K=4. We show results for K=2—4 and a corresponding map is shown depicting results for K=4 (Fig. 4.7). At K=2, the North clade separates from the remaining individuals (South+Central clades), while at K=2 the North clade is split into two groups with five individuals

showing admixture. Then, at K=4, the South and Central clades separate. Results for K=4 is also depicted on the SVD-quartets tree (Fig. 4.8). The SVD-quartets tree differs from the mtDNA tree in an important way: the South and Central clades are recovered as sister (bootstrap=100).

Kootenaia burkei—We identified 38 haplotypes among 70 *K. burkei* mtDNA sequences. There are a total of 1109 bp with 85 variable sites and 62 parsimony informative sites. The total number of retained GBS loci was 41,131 (with 35% missing data) and the mean number of variable sites per locus was 0.98.

BAPS recovered five mtDNA clusters that generally correspond to clades in the best ML tree, although bootstrap support for clades is very low (Fig. 4.9). Also, specimen *C580* is not part of any ML clade but clustered with BAPS1. Of note, the BAPS5 cluster corresponds to a large unresolved polytomy that includes individuals collected from non-overlapping geographic regions. There is a northeastern group (R1) primarily located in the Purcell, Cabinet, and Selkirk Mts and a second group (R2) with individuals from the Coeur d'Alene and Saint Joe Mts. Results of the NeighborNet and PCo analyses are similar to the mtDNA results but differ in a few ways (Fig. 4.9). Groups R1 and R2 separate from one another in the GBS data but the R2 group clusters closely together with group B1 (BAPS1). Groups R2 and B1 overlap partially in the Coeur d'Alene Mts but representatives of B1 were the only individuals collected in the Saint Joe Mts. In addition, two mtDNA clades (BAPS3 & BAPS4; both located in the middle of the sampling area) cluster closely together in the NeighborNet and PCo results. Finally, A group of individuals collected in the Selkirk Mts cluster together in all three analyses.

In the ADMIXTURE analysis, likelihood scores increased with increasing K and do not appear to plateau by K=12 (Fig. 4.10). However, cross-validation error estimates decrease from K=1 and reach a low point at K=4 and seem to stabilize from there. We show results for K=4—6 and a corresponding map is shown depicting results for K=6. At K=4, the data are roughly partitioned by region with there being a Selkirk group in the northwest, a northeastern group (R1), and then individuals collected from the middle and southern portions of the sampling area that show some admixture and includes groups B1 and R2 as well as BAPS3 and BAPS4. At K=5, a small group from the western Coeur d'Alene Mts splits (which also corresponds to BAPS4). At K=6, the B1 group splits apart from the R2 group but some individuals show admixture. Results for K=6 is also depicted on the SVD-quartets tree (Fig. 4.11). Results from all analyses show populations collected from closer to the periphery of the sampling area are most distinct, and there is some evidence of mitonuclear discordance.

Discussion

Several factors are likely responsible for generating the high levels of gastropod diversity observed in the North American inland temperate rainforest. Biotic factors such as dispersal abilities as well as spatiotemporal processes like vicariance through isolation in a heterogeneous landscape have likely been strong driving forces. For codistributed gastropod species with presumably similar life histories or ecological associations, congruent patterns of spatial genetic structure might be expected if species' historical responses were concerted. However, our findings mostly support a non-congruent phylogeographic history among the species we examined.

Both *M. mycophaga* and *K. burkei* show a similar depth of intraspecific mtDNA divergence (Fig. 4.2) and diversity (72 variable sites and 85 variable sites, respectively), whereas *U. lyrata* has relatively high mtDNA diversity (288 variable sites) and some terminal nodes within *U. lyrata* are separated by long branches (Fig. 4.2). In addition, both *M. mycophaga* and *K. burkei* have a similar mean number of SNPs per locus of ~1 (Figs 4.8 & 4.11), whereas it was 3.67 in *U. lyrata* (Fig. 4.5). These results strongly suggest that *Udosarx* is a polytypic genus.

Udosarx is narrowly distributed in north Idaho and adjacent parts of Montana but despite its small range, *Udosarx* exhibits deep genetic structure. Even within the lyr-a—d phylogroups there is genetic structure congruent with the geographical location of its different populations (Fig. 4.3). However, given that its current range is located south of glacial extent, it is not unreasonable to assume that long-term habitat stability (i.e., there has been more time for substitutions to accrue) and topographic complexity coupled with the sedentary nature of these organisms has led to the deep, hierarchical genetic structure observed in *Udosarx*. Unfortunately, our sampling did not expand into the furthest reaches of the eastern and southern parts of the range. Given that highly diverged genetic lineages were recovered in our data, it is likely that there remains unsampled diversity.

It was previously suggested that there may be two subspecies in Montana, *U. l. lyrata* (Webb 1959) and *U. l. russelli* (Russell and Webb 1980), but there have not been any follow-up evaluations on the validity of these subspecies, and this is the first study to include genetic data. Although there appears to be enough evidence to consider *Udosarx* a polytpic genus that is sister to the likewise polytypic *Zacoleus* genus, it is interesting to note that *Z. idahoensis* has a much broader range across the inland northwest, extending much further north into Canada, whereas *Z. Leonardi* is narrowly distributed in the Cascade Mts (Burke, 2013). This raises the question as to why *Udosarx* lineages have not similarly extended further north and tentatively suggests that *Z. Leonardi* reached its current distribution via dispersal from the inland.

If we compare results from all three focal species, then we see a contrast in that *Udosarx* exhibits deep genetic structure across a small spatial area, *M. mycophaga* also exhibits genetic

structure but across a much larger spatial area, and *K. burkei* exhibits shallow structure across a large spatial area. Indeed, *M. mycophaga* and *K. burkei* show similar levels of intraspecific genetic diversity but the two species have different patterns of genetic structure. *M. mycophaga* is divided into three distinct lineages with no evidence of admixture among those three lineages. Conversely, *K. burkei* shows more shallow structure and considerable admixture among lineages. Potential differences in habitat tolerance between the two species may explain these results. Specifically, *M. mycophaga* is mostly found in upper subalpine zones whereas *K. burkei* is commonly found in moist montane forests as well as riparian areas (Burke 2013; Lucid et al., 2018b). Thus, dispersal rates of the two species may differ through different habitat types (i.e., *K. burkei* may have higher rates of dispersal via riparian areas). However, the three *M. mycophaga* lineages are distributed widely across mountain systems and there is evidence that the species has (relatively) recently colonized the Purcell and Selkirk Mts via the Cabinet Mts, suggesting valley bottoms are not completely impermeable to the species.

The North *M. mucophaga* phylogroup is genetically depauperate, an expected result for a population following a retreating glacier. However, our ADMIXTURE result splits the North phylogroup, with one type predominately in the Cabinet Mts and the second type predominately in the Selkirk and Purcell Mts, with some admixed individuals from locations in-between (Fig. 4.7). Furthermore, two mtDNA haplotypes (*C147* and *FIA1234*) form a separate clade (with moderate support) in the ML tree, branching apart from the remaining North individuals (Fig. 4.6). *C147* is in the Selkirk Mts and *FIA1234* is in the Cabinet Mts. We suggest the Selkirk and Purcell populations have originated relatively recently from Cabinet immigrants.

Individuals of *K. burkei* show a pattern of high admixture at the center and more genetically distinct populations at the edges, with the edge populations being the Selkirk, R1, and B1 groups. This pattern is likely to occur under equilibrium conditions in which the spatial distribution of genetic variation within local areas and over larger distances is a trade-off between drift and gene flow. Furthermore, in the *K. burkei* mtDNA phylogeny, there is a large unresolved polytomy that resembles a comb-like structure indicative of and expanding population. However, the geographic distribution of individuals that belong to this mtDNA clade are non-overlapping (the R1 and R2 groups). It is thus interesting to note that the genomic data shows a closer relationship between the R2 and B1 groups with the R1 group being more distantly related. Both R2 and B1 groups are from the middle and southern parts of the sampling area and partially overlap, whereas our sampling does not show any overlap with the R2 group. It is possible these populations are connected through Montana and that the dispersal route into the Cabinet and Purcell Mts occurred in a counterclockwise fashion. A similar

pattern was observed in both *Ascaphus* frogs (Metzger et al., 2015) and *Hemphillia* slugs (Rankin et al., 2019)

Overall, the genomic data revealed partially congruent (with the mtDNA) but more distinct patterns of genetic structure and each species show some evidence of mitonuclear discordance, which can arise from incomplete lineage sorting, introgression, or mitochondrial capture. These results highlight the importance of using mtDNA in conjunction with nuclear data to improve the power of phylogeographic inference. Our study also emphasizes that (1) many understudied taxa groups in this region likely have cryptic lineages present, as was uncovered in the *Udosarx* data, and information on cryptic lineages is important to better understand the level of gastropod biodiversity in the region, (2) strong intraspecific genetic differentiation is potentially present within some taxa (e.g., *M. mycophaga*), which may have conservation implications. For example, our study provides important baseline data on the distribution of genetic diversity within *M. mycophaga*, which includes three phylogroups that could potentially be considered distinct evolutionarily significant units for management purposes (sensu Moritz, 1994). (3) Even closely related species may have unique histories owing to differences in life histories. For example, both *M. mycophaga* and *K. burkei* are part of the same phylogroup and show similar levels of intraspecific diversity (), but their spatial genetic structure is much different and likely caused by differences in migration patterns.

Literature Cited

Arbogast BS, & Kenagy GJ. 2001. Comparative phylogeography as an integrative approach to historical biogeography. Journal of Biogeography. 28: 819-825

Ralston J, FitzGerald AM, Burg TM, Starkloff NC, Warkentin IG, & Kirchman JJ. 2021. Comparative phylogeographic analysis suggests a shared history among eastern North American boreal forest birds. Ornithology.

Ritter CD, Coelho LA, Capurucho JM, Borges SH, Cornelius C, & Ribas CC. 2021. Sister species, different histories: comparative phylogeography of two bird species associated with Amazonian open vegetation. Biological Journal of the Linnean Society, 132: 161-173.

DellaSala DA. (Ed.). 2011. Temperate and boreal rainforests of the world: ecology and conservation. Island Press.

Richmond GM, Fryxell R, Neff GE, Weis PL. 1965. The Cordilleranice sheet of the northern Rocky Mountains and related Quaternary history. In: The Quarternary of the United States (eds Wright Jr HE, Frey DG), pp. 231–242. Princeton University Press, Princeton, NJ

Waitt Jr RB, Thorson RM. 1983. The Cordilleran ice sheet in Washington, Idaho, and Montana. In: Late Quaternary Environments of the United States, the Late Pleistocene (eds Wright HE Jr, Porter SC), pp. 53 – 70. University of Minnesota Press, Minneapolis, MN.

Delcourt PA, Delcourt HR. 1993. Paleoclimates, paleovegetation, and paleofloras during the late Quaternary. In: Flora of North America, Vol. 1, 1st edn. Editorial Committee, pp. 71– 94. Oxford University Press, New York.

Good JM, & Sullivan J. 2001. Phylogeography of the red-tailed chipmunk (*Tamias ruficaudus*), a northern Rocky Mountain endemic. Molecular Ecology, 10: 2683-2695.

Carstens BC, Degenhardt JD, Stevenson AL, & Sullivan J. 2005. Accounting for coalescent stochasticity in testing phylogeographical hypotheses: modelling Pleistocene population structure in the Idaho giant salamander *Dicamptodon aterrimus*. Molecular Ecology, 14: 255-265.

Metzger G, Espindola A, Waits LP, & Sullivan J. 2015. Genetic structure across broad spatial and temporal scales: Rocky Mountain tailed frogs (*Ascaphus montanus*; Anura: Ascaphidae) in the inland temperate rainforest. Journal of Heredity, 106: 700-710.

Rankin AM, Anderson FE, Clutts SA, Espíndola A, Carstens BC, Lucid M, & Sullivan J. 2021. Comparative phylogeography of two Northern Rocky Mountain endemics: the widespread *Anguispira kochi occidentalis* and the narrow-range *Anguispira nimapuna* (Gastropoda: Discidae). Biological Journal of the Linnean Society.

Leonard WP, Chichester L, Richart CH, & Young TA. 2011. *Securicauda hermani* and *Carinacauda stormi*, two new genera and species of slug from the Pacific Northwest of the United States (Gastropoda: Stylommatophora: Arionidae), with notes on *Gliabates oregonius* Webb 1959. Zootaxa, 2746: 43-56.

Burke TE. 2013. Land snails and slugs of the Pacific Northwest. Corvalis, Oregon: Oregon State University Press.

(A) Lucid MK, Rankin A, Espindola A, Chichester L, Ehlers S, Robinson L, & Sullivan J. 2018. Taxonomy and biogeography of *Hemphillia* (Gastropoda: Pulmonata: Arionidae) in North American rainforests, with description of a new species (Skade's jumping-slug, *Hemphillia skadei* sp. nov.). Canadian Journal of Zoology, 96: 305-316.

(B) Lucid MK, Ehlers S, Robinson L, & Cushman SA. 2018. Beer, brains, and brawn as tools to describe terrestrial gastropod species richness on a montane landscape. Ecosphere, 9: e02535.

Smith ML, Ruffley M, Rankin AM, Espíndola A, Tank DC, Sullivan J, & Carstens BC. 2018. Testing for the presence of cryptic diversity in tail-dropper slugs (*Prophysaon*) using molecular data. Biological Journal of the Linnean Society, 124: 518-532.

Rankin AM, Wilke T, Lucid M, Leonard W, Espíndola A, Smith ML, ... & Sullivan J. 2019. Complex interplay of ancient vicariance and recent patterns of geographical speciation in north-western North American temperate rainforests explains the phylogeny of jumping slugs (*Hemphillia* spp.). Biological Journal of the Linnean Society, 127: 876-889.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS one, 6: e19379.

Smith ML, & Carstens BC. 2020. Process-based species delimitation leads to identification of more biologically relevant species. Evolution, 74: 216-229.

Eaton DA, & Overcast, I. 2020. ipyrad: Interactive assembly and analysis of RADseq datasets. Bioinformatics, 36: 2592-2594.

Minin V, Abdo Z, Joyce P, Sullivan J. 2003. Performance-based selection of likelihood models for phylogeny estimation. Systematic Biology 52: 674 –683

Zwickl DJ. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion (Doctoral dissertation).

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, ... & Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol, 10: e1003537.

Swofford DL. 2003. PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4. Sunderland: Sinauer Associates.

Thomaz D, Guiller A, Clarke B. 1996. Extreme divergence of mitochondrial DNA within species of pulmonate land snails. Proceedings of the Royal Society of London B 263: 363 –368.

Chiba S. 1999. Accelerated evolution of land snails Mandarina in the oceanic Bonin Islands: evidence from mitochondrial DNA sequences. Evolution 53: 460 –471.

Van Riel P, Jordaens K, Van Houtte N, Martins AMF, Verhagen R, Backeljau T. 2005. Molecular systematics of the endemic *Leptaxini* (Gastropoda: Pulmonata) on the Azores islands. Molecular Phylogenetics and Evolution 37: 132 –143.

Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, ... & Vogler AP. 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Systematic

biology, 55: 595-609.

Corander J, Marttinen P, Sirén J, & Tang J. 2008. Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. BMC bioinformatics, 9: 1-14.

Chifman J, & Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics, 30: 3317-3324.

Huson DH, & Bryant D. 2005. Estimating phylogenetic trees and networks using SplitsTree 4. Manuscript in preparation, software available from www. splitstree. org.

Alexander DH, Novembre J, & Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome research, 19: 1655-1664.

Puillandre N, Lambert A, Brouillet S, & ACHAZ G. 2012. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. Molecular ecology, 21: 1864-1877.

Webb GR. 1959. Two new north-western slugs, *Udosarx lyrata*, and *Gliabates oregonia*. Gastropodia, 1: 22-23.

Russell RH, & Webb GR. 1980. The slug *Udosarx lyrata*: additional data on distribution, anatomy, and taxonomy. Gastropodia, 2(1), 3.

Moritz C. 1994. Defining ‘evolutionarily significant units’ for conservation. Trends in ecology & evolution, 9: 373-375.

Figure 4.1: Best maximum-likelihood phylogeny for Pacific Northwest arionid species based on mtDNA. The scale is given in substitutions per site, and node labels indicate maximum-likelihood bootstrap values.

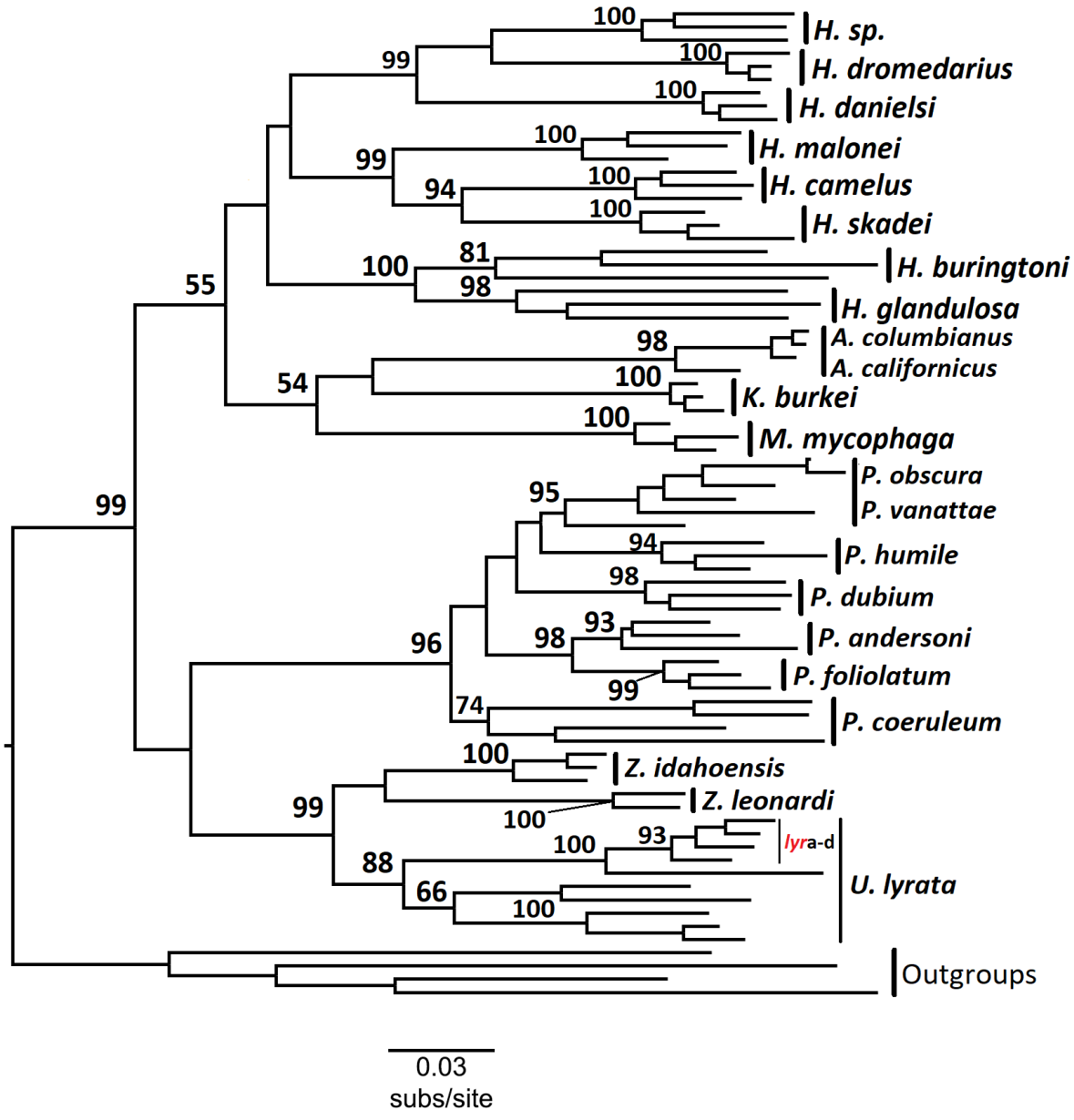


Figure 4.2: Bayesian phylogeny for Pacific Northwest arionid species based on mtDNA. Circles on branches indicate Bayesian posterior probabilities >95% from the BEAST analysis.

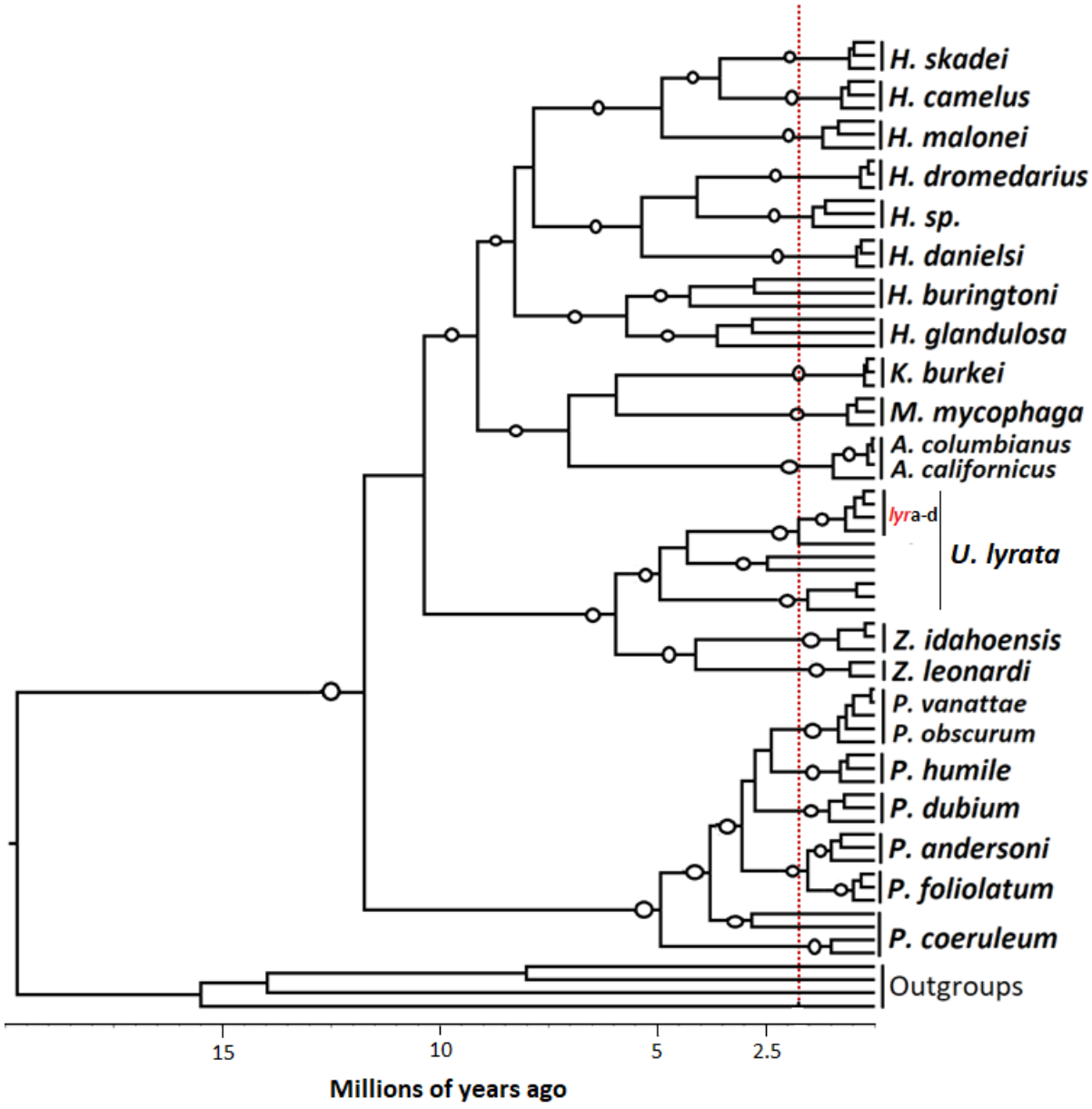


Figure 4.3: (1) Best maximum likelihood phylogeny based on the *U. lyrata* mtDNA data. Bootstrap support values ≥ 50 are given. BAPS clusters that correspond to the best partition of the data are indicated by different colored bars. Map showing the locations of sampled individuals. (2) NeighborNet phylogeny based on full SNP dataset, and (3) PCoA based on unlinked SNP datas

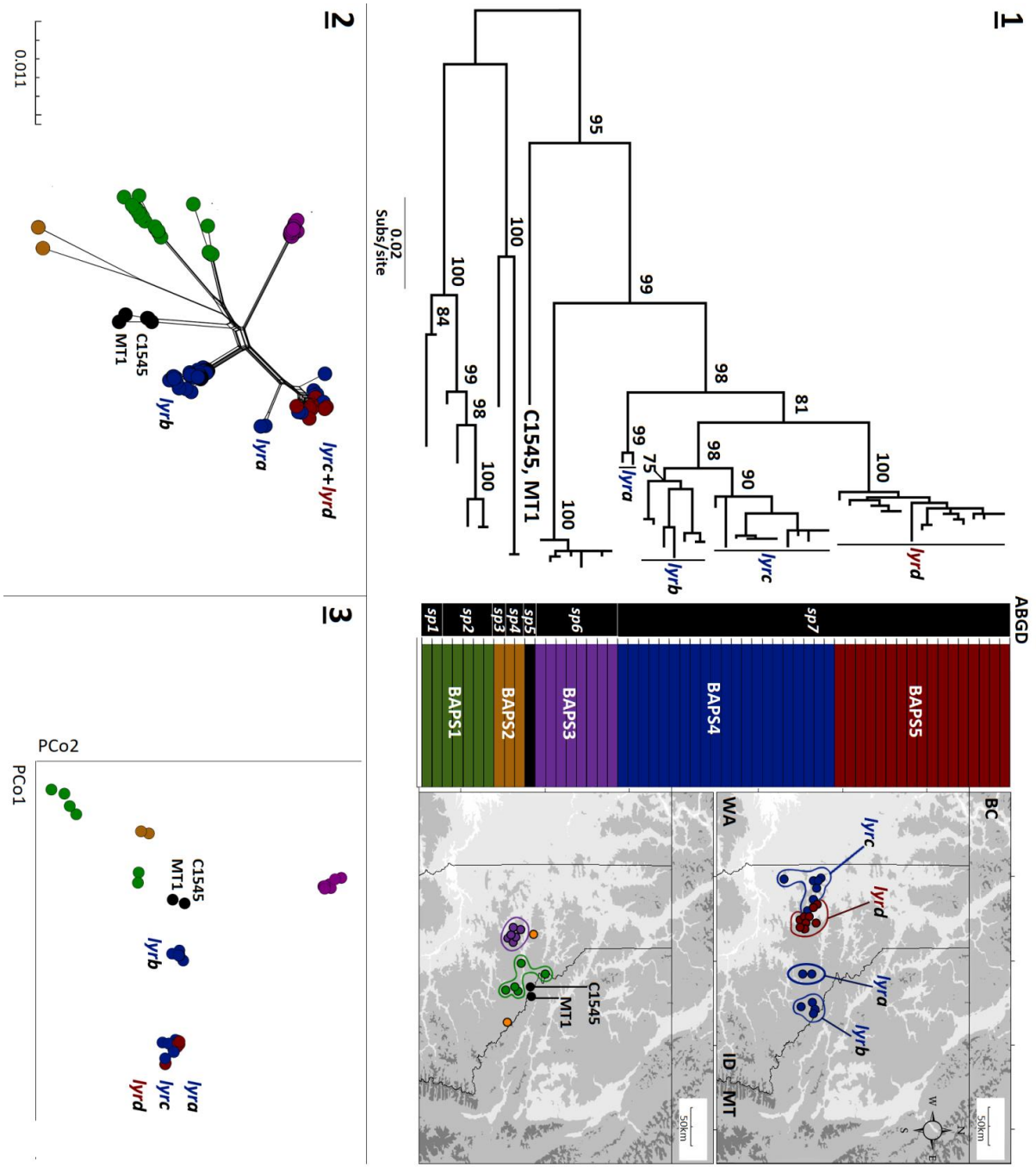


Figure 4.4: *U. lyrata* ADMIXTURE results for K=3—5 based on unlinked SNP data and log likelihood and cross-validation error estimates for each value of K. Results for K=5 are depicted as pie-charts on map

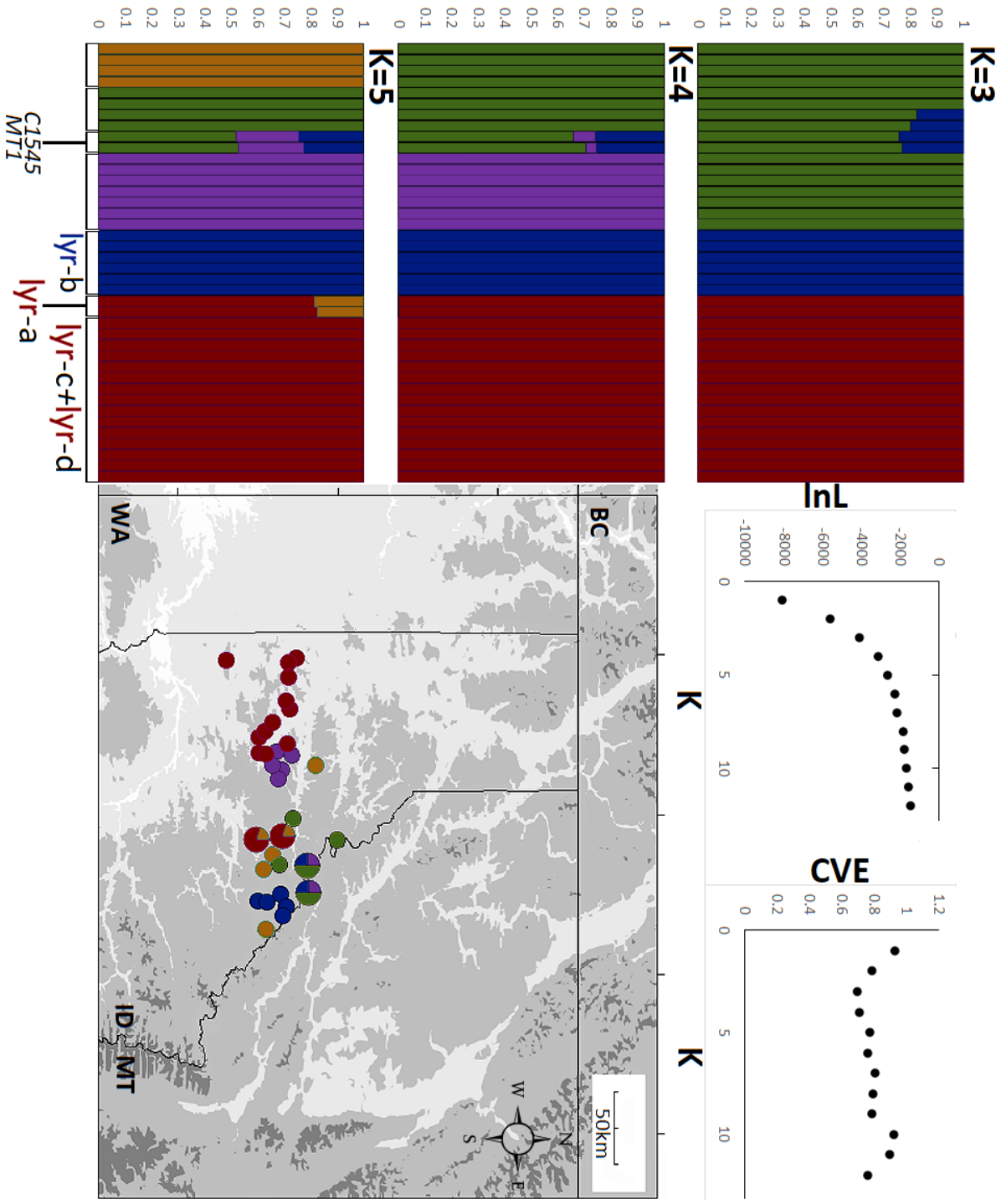


Figure 4.5: SVD-quartets phylogeny and bootstrap values for *U. lyrata* based on full SNP dataset, and the distribution of SNPs per locus. ADMIXTURE results for K=5 is depicted as pie-charts for each individual.

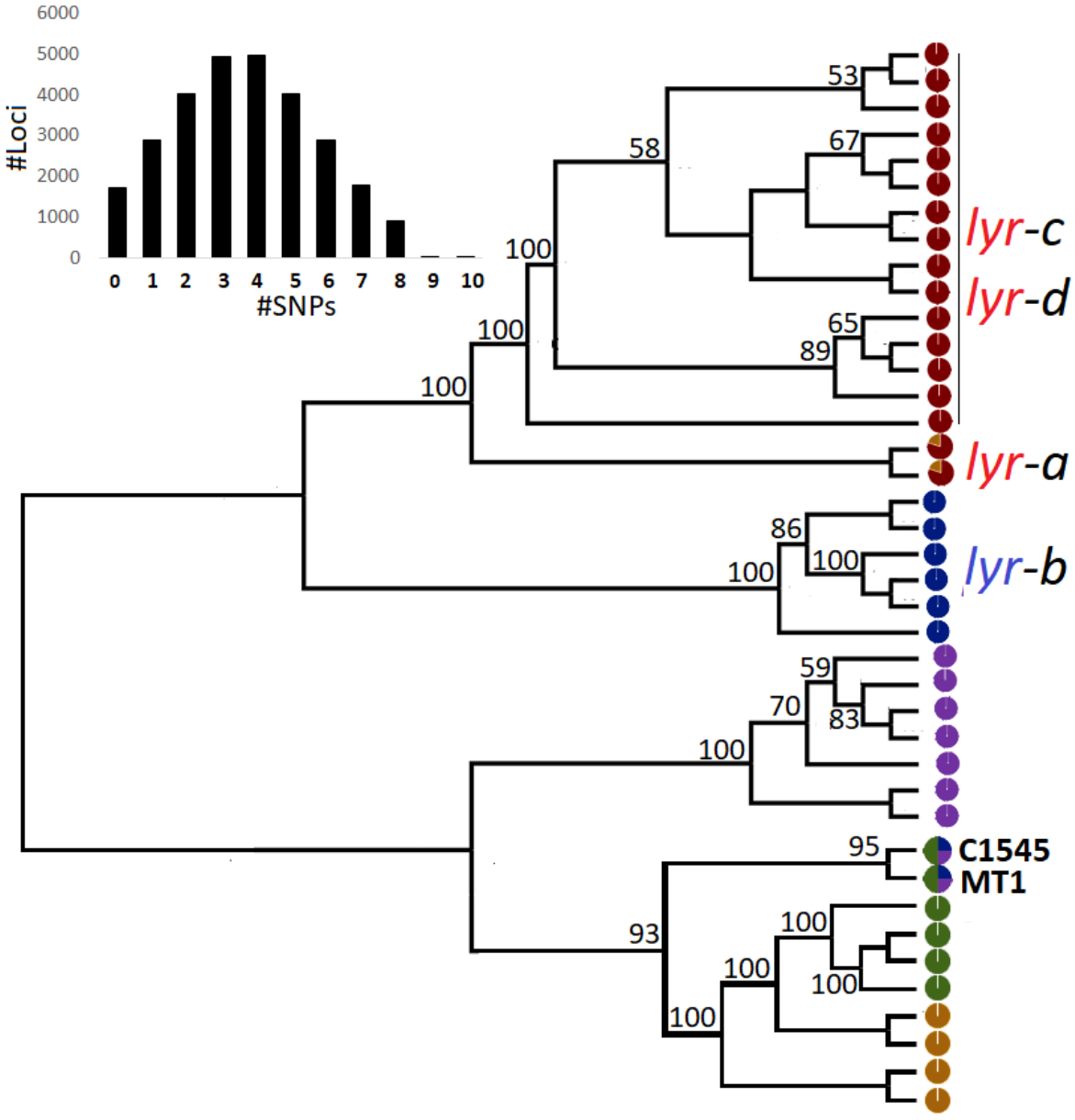


Figure 4.6: (1) Best maximum likelihood phylogeny based on the *M. mycophaga* mtDNA data. Bootstrap support values ≥ 50 are given. BAPS clusters that correspond to the best partition of the data are indicated by different colored bars. Map showing the locations of sampled individuals. (2) NeighborNet phylogeny based on full SNP dataset, and (3) PCoA based on unlinked SNP dataset.

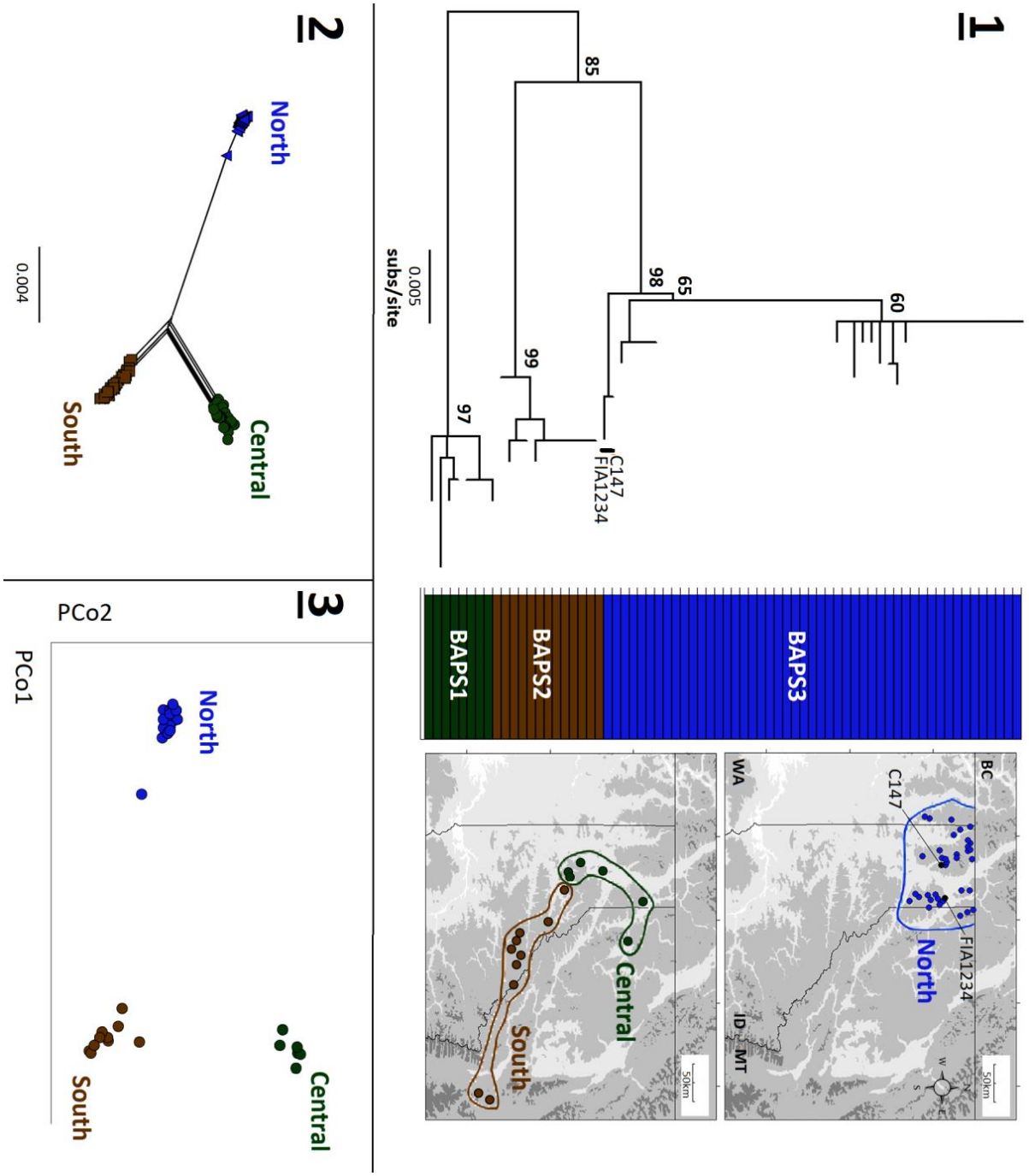


Figure 4.7: *M. mycophaga* ADMIXTURE results K=2—4 based on unlinked SNP data and log likelihood and cross-validation error estimates for each value of K. Results for K=5 is depicted as pie-charts on map.

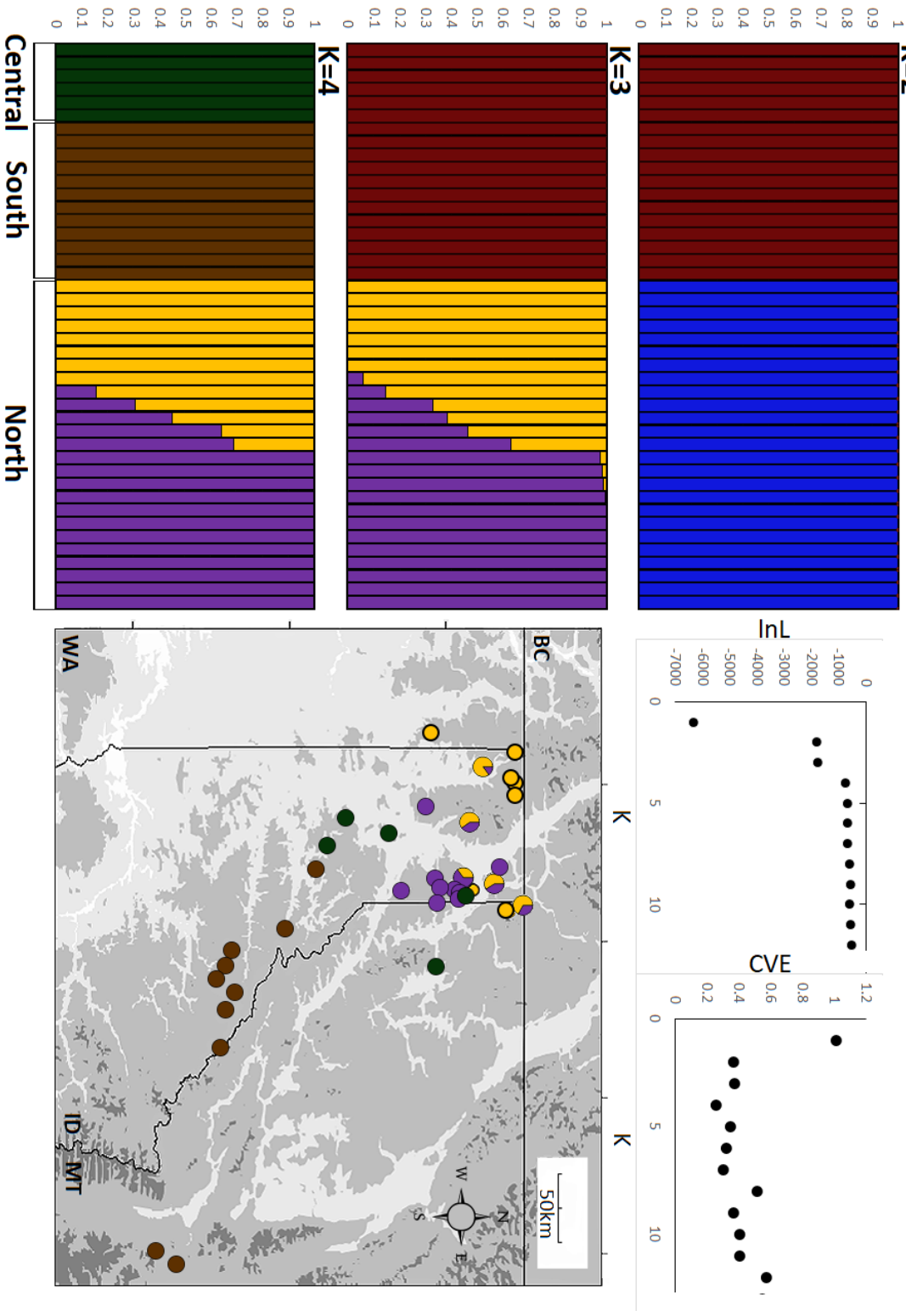
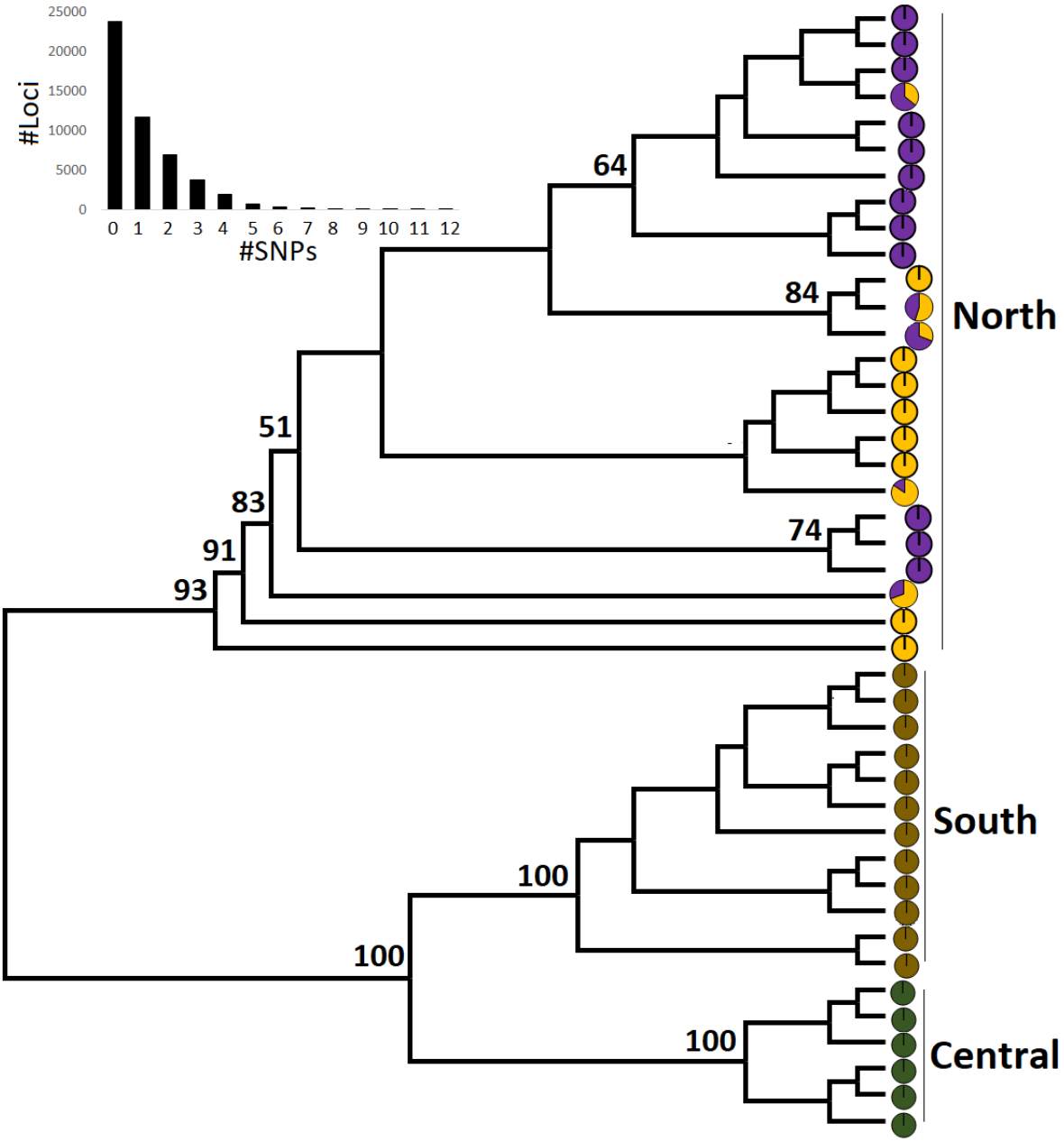


Figure 4.8: SVD-quartets phylogeny and bootstrap values for *M. mycophaga* based on full SNP dataset, and the distribution of SNPs per locus. ADMIXTURE results for K=5 is depicted as pie-charts for each individual.



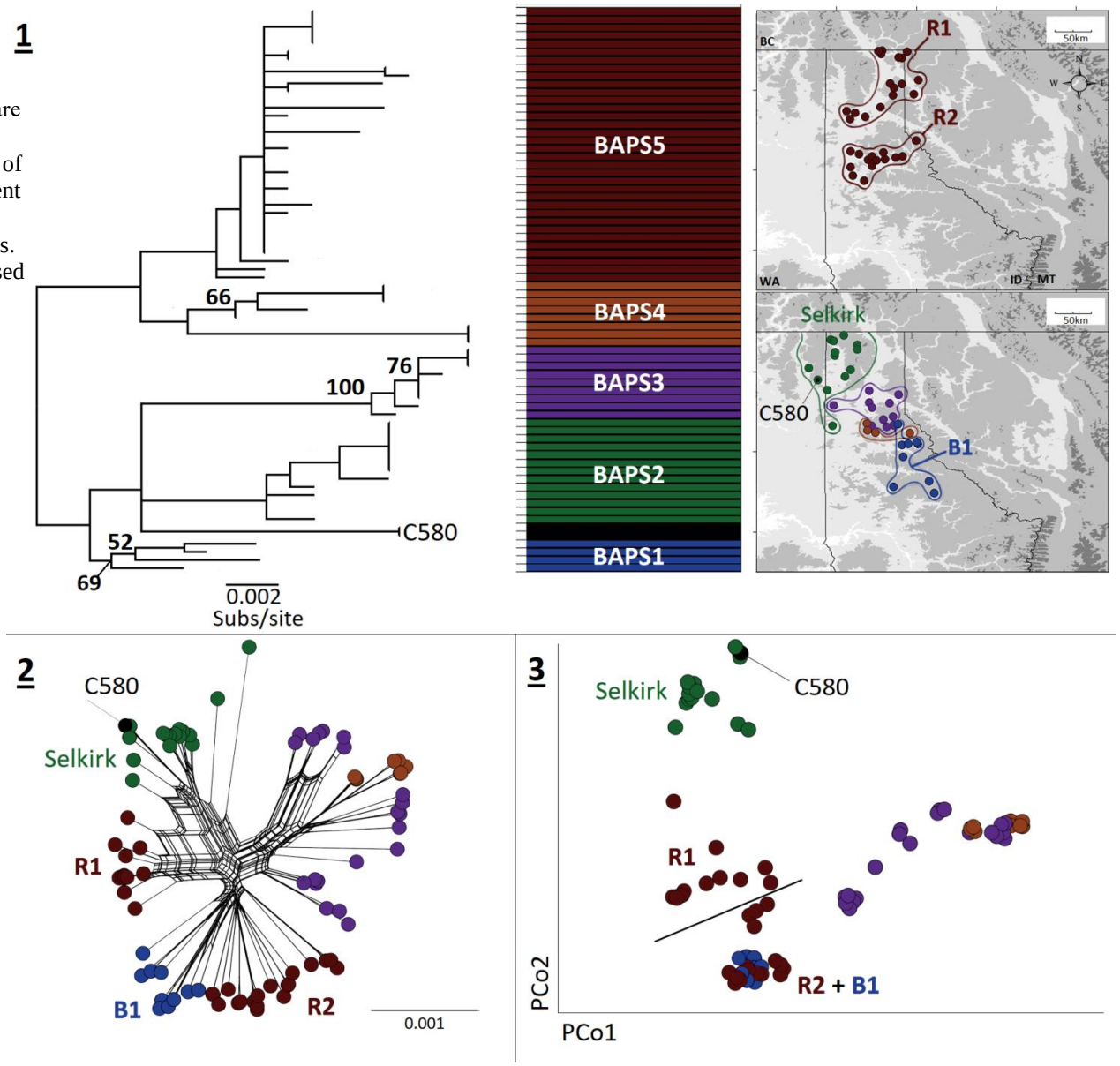


Figure 4.9: (1) Best maximum likelihood phylogeny based on the *K. burkei* mtDNA data. Bootstrap support values ≥ 50 are given. BAPS clusters that correspond to the best partition of the data are indicated by different colored bars. Map showing the locations of sampled individuals. (2) NeighborNet phylogeny based on full SNP dataset, and (3) PCoA based on unlinked SNP dataset.

Figure 4.10: *K. burkei* ADMIXTURE results K=2—4 based on unlinked SNP data and the log likelihood and cross-validation error estimates for each value of K. Results for K=5 is depicted as pie-charts on map.

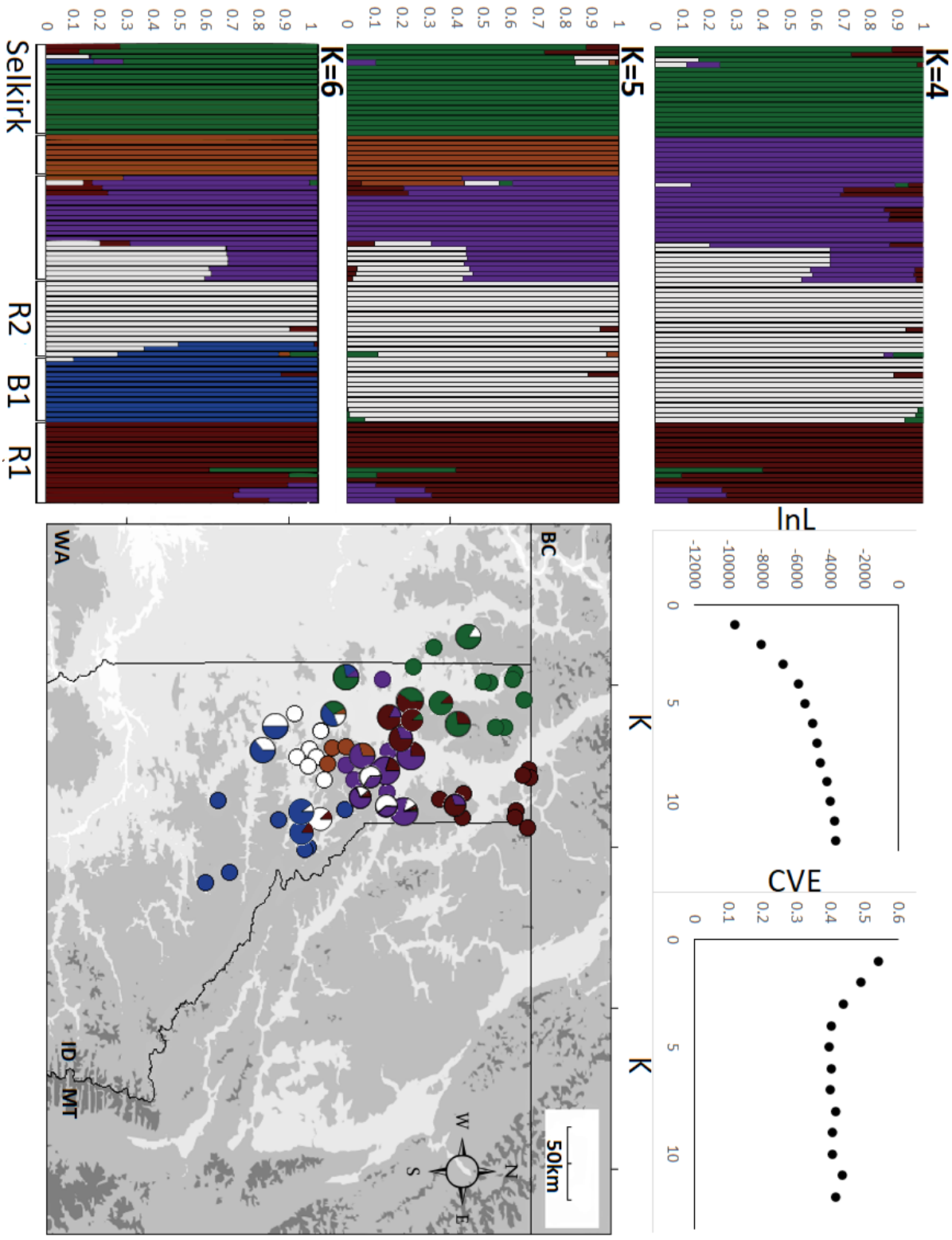
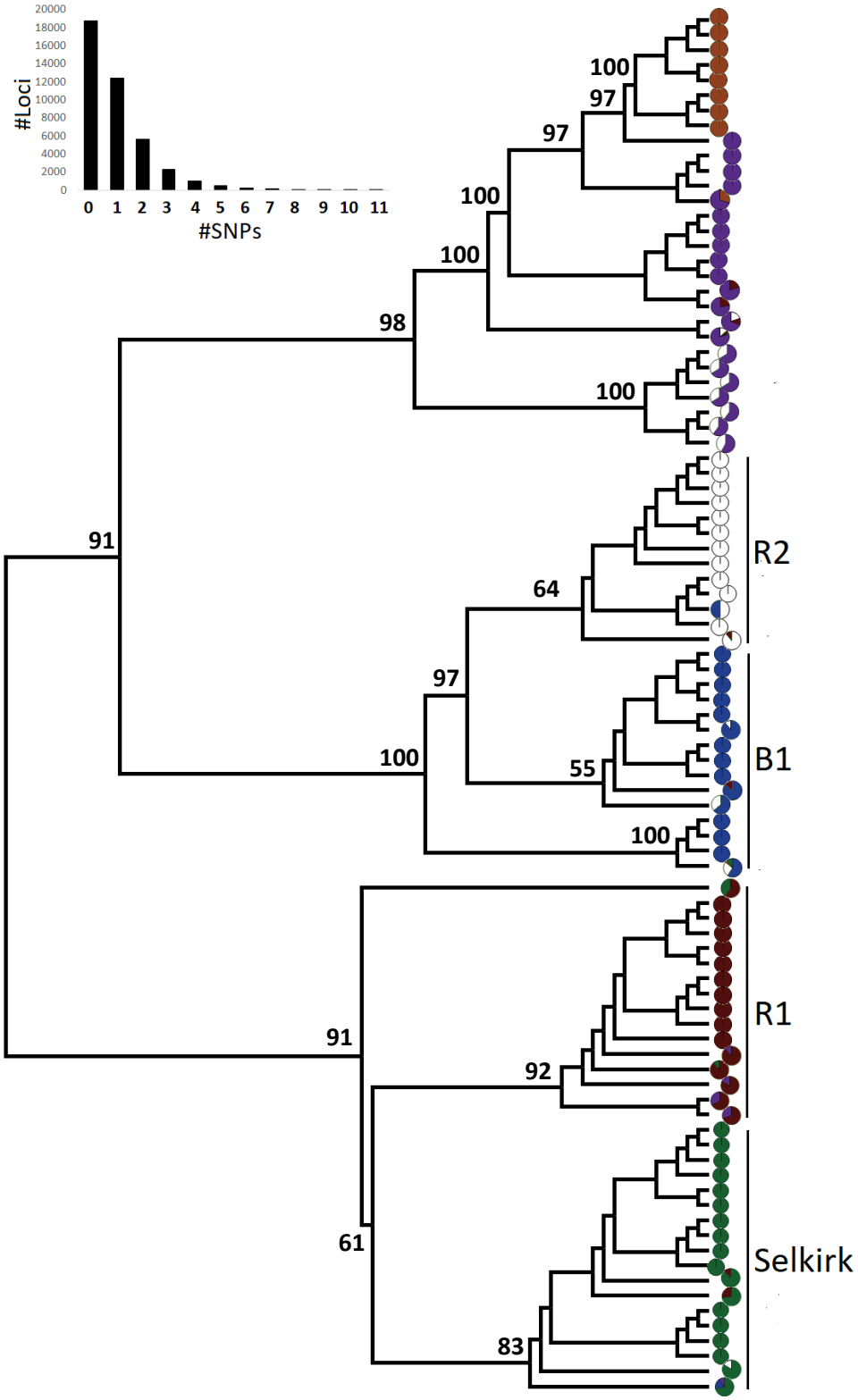


Figure 4.11: SVD-quartets phylogeny and bootstrap values for *K. burkei* based on full SNP dataset, and the distribution of SNPs per locus. ADMIXTURE results for K=5 is depicted as pie-charts for each individual.



Chapter 5: Conclusion

This dissertation adds to our knowledge of the biogeographical history of the Pacific Northwest region in that we show codistributed species are likely to exhibit distinct patterns that reflect the unique aspects of species' phylogeographic histories. The results presented here also significantly increase the taxonomic and natural history knowledge of several rare gastropod lineages, which is important because there is an increasing need to obtain data on the genetic and spatial structure of endemic taxa in areas such as the Pacific Northwest due to the threat of habitat fragmentation (both natural and induced by human activity), which may be exponentially greater to minute organisms that maintain small ranges and have little public appeal.