




**ORIGINAL ARTICLE**

# Combining allele frequency and tree-based approaches improves phylogeographic inference from natural history collections

Megan Ruffley<sup>1,2,3</sup>  | Megan L. Smith<sup>4</sup> | Anahí Espíndola<sup>1,2</sup>  |  
 Bryan C. Carstens<sup>4</sup>  | Jack Sullivan<sup>1,2</sup> | David C. Tank<sup>1,2,3</sup>

<sup>1</sup>Department of Biological Sciences,  
University of Idaho, Moscow, ID, USA

<sup>2</sup>Institute for Bioinformatics and  
Evolutionary Studies (IBEST), Biological  
Sciences, Moscow, ID, USA

<sup>3</sup>Stillinger Herbarium, University of Idaho,  
Moscow, ID, USA

<sup>4</sup>Department of Evolution, Ecology, &  
Organismal Biology, The Ohio State  
University, Columbus, OH, USA

**Correspondence**

Megan Ruffley, Department of Biological  
Sciences, University of Idaho, Moscow, ID,  
USA.

Email: MeganRRuffley@gmail.com

**Funding information**

National Science Foundation, Grant/Award  
Number: DEB-1457519, DEB-1457726;  
Institute for Bioinformatics and Evolutionary  
Studies (IBEST) at the University of Idaho,  
Grant/Award Number: NIH NCR  
1P2ORR016454-01, NIH NCR  
1P2ORR016448-01, NSF EPS-809935

**Abstract**

Model selection approaches in phylogeography have allowed researchers to evaluate the support for competing demographic histories, which provides a mode of inference and a measure of uncertainty in understanding climatic and spatial influences on intraspecific diversity. Here, to rank all models in the comparison set and determine what proportion of the total support the top-ranked model garners, we conduct model selection using two analytical approaches—allele frequency-based, implemented in FASTSIMCOAL2, and gene tree-based, implemented in PHRAPL. We then expand this model selection framework by including an assessment of absolute fit of the models to the data. For this, we utilize DNA isolated from existing natural history collections that span the distribution of red alder (*Alnus rubra*) in the Pacific Northwest of North America to generate genomic data for the evaluation of 13 demographic scenarios. The quality of DNA recovered from herbarium specimen leaf tissue was assessed for its utility and effectiveness in demographic model selection, specifically in the two approaches mentioned. We present strong support for the use of herbarium tissue in the generation of genomic DNA, albeit with the inclusion of additional quality control checks prior to library preparation and analyses with multiple approaches that incorporate various data. Analyses with allele frequency spectra and gene trees predominantly support *A. rubra* having experienced an ancient vicariance event with intermittent and frequent gene flow between the disjunct populations. Additionally, the data consistently fit the most frequently selected model, corroborating the model selection techniques. Finally, these results suggest that the *A. rubra* disjunct populations do not represent separate species.

**KEYWORDS**

ddRAD sequencing, model selection, natural history collections, phylogeography

**1 | INTRODUCTION**

Understanding how conspecific populations evolve is central for identifying and quantifying diversity. Phylogeography aims to increase our understanding of these historical processes (e.g., Avise et al., 1987), and as the field has expanded, several approaches have

been used. While the earliest investigations derived their inferences from qualitative interpretations of patterns evident in the genetic and geographic data, later studies began to more explicitly test hypotheses (e.g., Knowles, 2001; Sullivan, Arellano, & Rogers, 2000) and estimate parameters under explicit analytical models, such as isolation with migration in IMA2 (Hey, 2010) and Migrate-n (Beerli &

Felsenstein, 2001). In hypothesis testing and model selection studies, models representing historical demographic scenarios are evaluated in a statistical framework, and the inferences are drawn from the results of the test (Knowles & Maddison, 2002). Early examples used parametric simulation and frequentist statistics (e.g., DeChaine & Martin, 2005), whereas later examples utilized Bayesian (e.g., Fagundes et al., 2007) or information-theoretic (Carstens, Stoute, & Reid, 2009) approaches to consider and rank multiple models. Such approaches allow historical knowledge of the species or complex of study to be incorporated into the models that are assessed (Gutenkunst, Hernandez, Williamson, & Bustamante, 2009a). Phylogeographic model selection can be implemented through a variety of approaches and software, such as approximate Bayesian computation (ABC; Csilléry, François, & Blum, 2012),  $\partial a \partial i$  (Gutenkunst, Hernandez, Williamson, & Bustamante, 2009b), *FASTSIMCOAL2* (Excoffier, Dupanloup, Huerta-Sánchez, Sousa, & Foll, 2013; Excoffier & Foll, 2011) and *PHRAPL* (Jackson, Morales, Carstens, & O'Meara, 2015). All incorporate coalescent theory (Kingman, 1982) to model evolutionary processes that occur at the population level, such as genetic drift, migration and population expansion and/or contraction over time. As opposed to hypothesis testing approaches that reject or fail to reject individual models (and thus experience difficulties with multiple comparisons), model selection frameworks can be designed to rank all models in the comparison set, and thus provide one measure of confidence in the form of what proportion of the total support is garnered by the top-ranked model. However, a potential shortcoming of such a framework is that there is no guarantee that a model that represents the true evolutionary history is included in the comparison set (Templeton, 2008).

Phylogeographic inference is ideally drawn from multiple sources, including geographic information (e.g., Hugall, Moritz, Moussalli, & Stanisc, 2002) and descriptive summaries of the data (e.g., Petit & Grivet, 2002). Analytical models that incorporate the coalescent process are valuable, particularly when they have a demonstrably good fit to the empirical data. While assessments of model adequacy and model fit have generally been lacking in phylogeographic research, they are vital components of inferences that are derived from statistical analysis (Gelman & Shalizi, 2013). Here, we expand a model selection framework such that it includes an assessment of model fit. We first conduct model selection using two analytical approaches—allele frequency-based, implemented in *FASTSIMCOAL2* (Excoffier et al., 2013), and gene tree-based, implemented in *PHRAPL* (Jackson et al., 2015)—and then assess the absolute fit of the models to the data.

Phylogeographic analysis depends on comprehensive sampling across the geographic range of a species or complex (Knowles & Maddison, 2002; Pincheel, Jordaens, Pfenninger, & Backeljau, 2005). Herbarium and other natural history museum specimens are important sources for such sampling when specimens are available for DNA extraction. Plant tissue dried and preserved in silica gel can be used to recover high-quality genomic data (Varma, Padh, & Shrivastava, 2007), even several years after collection (e.g., Eaton & Ree, 2013). However, there still remain a large number of herbarium

specimens without associated silica-dried tissue, which results in one having to use tissue directly from the herbarium specimen sheet that was not dried strategically for DNA preservation. The use of genomic data in phylogeography has increased the resolution at which we can discern competing hypotheses, and thus improved our overall understanding of phylogeographic processes (Carstens, Lemmon, & Lemmon, 2012). However, it is unclear whether the DNA that can be extracted from herbarium specimens is sufficiently intact to serve as the source material for generating genome-scale data sets, particularly when systematically distributed missing data can result in implicitly biased inferences (Andrews, Good, Miller, Luikart, & Hohenlohe, 2016).

In this work, in addition to extending model-based phylogeographic inference to incorporate model fit, we also aim to understand the quality of DNA needed to recover useful genomic data from herbarium-sampled leaf tissue. We further ask whether such genomic data are plagued with biased, or nonuniform, missing data. Finally, using genomic data from herbarium specimens, descriptive analyses and two model selection approaches, we aim to understand the phylogeographic history of *Alnus rubra* Bong. in the disjunct mesic forests of the Pacific Northwest of North America.

## 2 | MATERIALS AND METHODS

### 2.1 | Study system

The Pacific Northwest (PNW) temperate rainforests form a disjunct ecosystem that ranges from the Cascade Mountain Range to the Pacific coast, extending from northern California to southern Alaska, and exists along the northern Rocky Mountains (NRM) in central to northern Idaho. The range of mesic forests in the region was likely continuous prior to the uplift of the Cascades (c. 5 MYA; Waring & Franklin, 1979; Priest, 1990), which generated a rain shadow cast across the Columbia Basin and forced inland forests to retreat to suitable, wet habitat along the NRM. Because of this, the coastal and inland NRM forests became isolated by ~300 km of unsuitable habitat. The later onset of Pleistocene glaciations (c. 2.5 MYA) led to the expansion of Cordilleran ice sheets, which covered much of the inland rainforests, further reducing the available habitat for rainforest species. As a consequence of these events, at least some rainforest species were unable to persist in the inland NRM forest throughout the Pleistocene (reviewed in Brunfeldt, Sullivan, Soltis, & Soltis, 2001).

Due to its disjunct nature, the PNW rainforest has been the focus of several phylogeographic studies (Brunfeldt, Miller, & Carstens, 2007; Carstens, Stevenson, Degenhardt, & Sullivan, 2004; Metzger, Espindola, Waits, & Sullivan, 2015; Nielson, Lohman, & Sullivan, 2001; Steele, Carstens, Storfer, & Sullivan, 2005), indicating that the history of the species was tightly associated with that of the biome. Studies showed that while some species harbour cryptic diversity (i.e., pre-Pleistocene divergences) across the disjunction, others do not. This led to the definition of two principal phylogeographic hypotheses for the biome, which explain the presence or

absence of cryptic diversity along the disjunction. The first hypothesis explains the presence of cryptic diversity in a lineage and is known as pre-Pleistocene vicariance or ancient vicariance (AV) (Brunsfeld et al., 2001) (Figure 1a). The uplift of the Cascade Mountain Range has been implicated as causing the disjunction. The AV hypothesis posits that conspecific populations were continuously present along the coast and in the inland NRM forests, but that the two areas were genetically isolated from each other. Following the end of the Pleistocene glacial cycles (~13 KYA), conspecific populations locally recolonized newly freed suitable areas. Because this hypothesis posits that there has been no gene flow between the inland and coastal populations since the initial vicariance event, the lineages would have been evolving independently for c. 2 MY, leading to the presence of cryptic diversity.

The alternative hypothesis, post-Pleistocene dispersal or recent dispersal (RD) (Figure 1b), explains the absence of cryptic diversity in some taxa (Brunsfeld et al., 2001). This hypothesis posits local Pleistocene extinction in the NRM. The current species presence in the inland NRM forest is thus due to dispersal from coastal populations to the inland following glacial retreat, although there is also evidence of dispersal occurring from the inland NRM to the coast (e.g., Carstens et al., 2013). Ultimately, the RD hypothesis suggests absence of significant genetic differentiation between inland and coastal populations because dispersal to the inland happened after 13 KYA. However, RD is not the only phylogeographic scenario that could result in an absence of significant genetic differentiation between the inland and coastal populations. Due to the regions being exposed to cyclical glacial periods, there could also have been episodic, repeated migration since the ancient vicariance event (Figure 1c). In this case, there were still inland populations throughout the Pleistocene, as in the AV hypothesis, but persistent gene flow prohibited any deep divergence between the coastal and inland populations. Likewise, secondary contact, where the inland and coastal lineages underwent AV and only since the end of the Pleistocene came back in contact, could also result in the lack of cryptic diversity (Figure 1d).

The primary distribution of *Alnus rubra* is west of the Cascades in the coastal temperate rainforest from southeastern Alaska to central California, with disjunct populations in the inland NRM temperate rainforest of Idaho. Two studies have investigated the history of *A. rubra* using genetic data (Brumble, 2008; Strenge, 1994) and suggested recent dispersal to the inland rainforest. Strenge (1994; also see Soltis, Gitzendanner, Strenge, & Soltis, 1997) characterized two cpDNA genotypes, a coastal southern type and a coastal northern type, and the inland individuals included were of the southern coastal genotype. Brumble (2008) identified a 17-bp indel in the chloroplast *psbA-trnH* spacer that was also present in some, but not all, closely related *Alnus* sp. Thus, the genetic data in these two studies were limited, and the inferences were perhaps driven by a single ancestral polymorphism. Recently, a predictive framework was developed to detect the presence of cryptic diversity from locality data by assessing georeferenced climatic data and taxonomic ranks (Espindola et al., 2016). This predictive framework also predicts that *A. rubra* should not contain cryptic diversity.

## 2.2 | ddRAD Sequencing

Leaf tissue from 49 *A. rubra* herbarium specimens (18 localities; Table S1) were sampled from the Stillinger Herbarium at the University of Idaho. The sampled specimens were collected from the coastal and inland PNW rainforest between the years of 2000 and 2007 by various collectors (Table S1) and cover the complete range of the species. DNA was extracted using a modified CTAB protocol (Doyle & Doyle, 1987), purified using Sera-Mag Speed-Beads (Thermo Fisher Scientific; Faircloth & Glenn, 2012; Rohland & Reich, 2012) and quantified using a Qubit 2.0 Fluorometer (Life Technologies). Genomic data were generated using double-digest restriction site-associated DNA sequencing (ddRADseq) (Peterson, Weber, Kay, Fisher, & Hoekstra, 2012), with the restriction enzymes *EcoRI* and *SbfI* (New England Biolabs, USA), and size selection at a 650 ( $\pm 50$ )-bp window on a BluePippin (Sage Science). All digestion, ligation and PCR products were purified using Agencourt AMPure XP purification system (Beckman Coulter). Sequences were generated as 300-bp paired-end reads using an Illumina MiSeq in the Institute for Bioinformatics and Evolutionary Studies (IBEST) Genomics Resource Core at the University of Idaho. Raw sequences were processed using PyRAD (Eaton, 2014) under a minimum coverage of 7 and clustering threshold of 85% (see Dryad link for complete parameter file). PyRAD includes VSEARCH (Rognes, Flouri, Nichols, Quince, & Mahé, 2016) and Muscle (Edgar, 2004) for sequence clustering. To merge overlapping reads, Paired-End reAd mergeR (PEAR) (Zhang, Kobert, Flouri, & Stamatakis, 2014) was used, and only sequences that merged with their paired end were used in subsequent analyses.

## 2.3 | Data quality and effect on missing data

Before ddRADseq library preparation, DNA extracts from 13 of the 49 individuals were quantified using a Fragment Analyzer (Advanced Analytical), which describes the distribution of fragment sizes in a particular sample (Figure S1). The mode of the fragment size distribution, concentration of the fragments in the distribution, year of collection and each variable's potential interactions (Table S2) were used for linear regression to predict the total number of raw reads for the 13 samples. This allowed us to evaluate our ability to predict data quality based on DNA quality and/or quantity descriptors.

To confirm the presence of unbiased missing data across sampled localities, missing data were quantified across all 49 samples and organized by relatedness using population assignment probabilities from STRUCTURE at  $K = 3$  (Figure S2a). This ordered distribution was compared to a uniform distribution of the same size using a two-sided Kolmogorov–Smirnov (KS) test (Panchenko, 2006). The uniform distribution was simulated in R using the *runif* function to generate 49 random variables from a uniform distribution with a maximum and minimum bound corresponding to the maximum and minimum missing data value observed across all individuals (Figure S2b).

## 2.4 | Population structure

STRUCTURE 2.3.4 (Pritchard, 2010) was used to estimate population structure across all sampled localities. All unlinked SNPs from all samples were used in the analyses. Following Pritchard (2010), all parameters were kept as default, aside from the burn-in (set to 200,000 generations) and the MCMC length (set to 1,000,000 generations). The data were modelled assuming admixture and correlated allele frequencies between populations (Falush, Stephens, & Pritchard, 2003). We tested a range of  $K$  values from 1 to 10 and repeated each run 10 times to capture variation in the likelihood estimate of each  $K$  value. The individual- and population-level probabilities of belonging to a particular cluster  $K$  were visualized using STRUCTURE PLOT (Ramasamy, Ramasamy, Bindroo, & Naik, 2014).

## 2.5 | Species distribution models

To gather more information about the potential range extent of the species during the last glacial maximum (LGM; ~21 KYA), we used a species distribution modelling (SDM) approach (Peterson et al., 2011). To do this, we gathered 772 unique observations of *A. rubra* from the Global Biodiversity Information Facility (GBIF) and the Consortium of Pacific Northwest Herbaria. We selected eight of the least correlated bioclimatic variables from the 19 total WorldClim bioclimatic variables ( $r^2 < .7$ ; i.e., bio1, bio2, bio3, bio5, bio7, bio12, bio15 and bio17) at a resolution of 30 arc-seconds (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005) and used them to adjust SDMs. To do this, we used the package biomod2 (Thuiller, Lafourcade, Engler, & Araújo, 2009) and applied an ensemble approach in R. For this, we adjusted the final ensemble model using the AUCs (area under the curve) of nine modelling methods, including generalized linear model (GLM), generalized additive model (GAM), classification tree analysis (CTA), artificial neural network (ANN), surface range envelope (SRE), flexible discriminant analysis (FDA), multiple adaptive regression splines (MARS), random forest (RF) and Maxent as weighting units, and we selected 10,000 pseudo-absences from the background area (a polygon encompassing the entire range of the species and the totality of the PNW region). We then projected the ensemble model into geographic space, using both current and palaeoclimatic data obtained from WorldClim corresponding to current and LGM climatic conditions.

## 2.6 | Demographic model selection

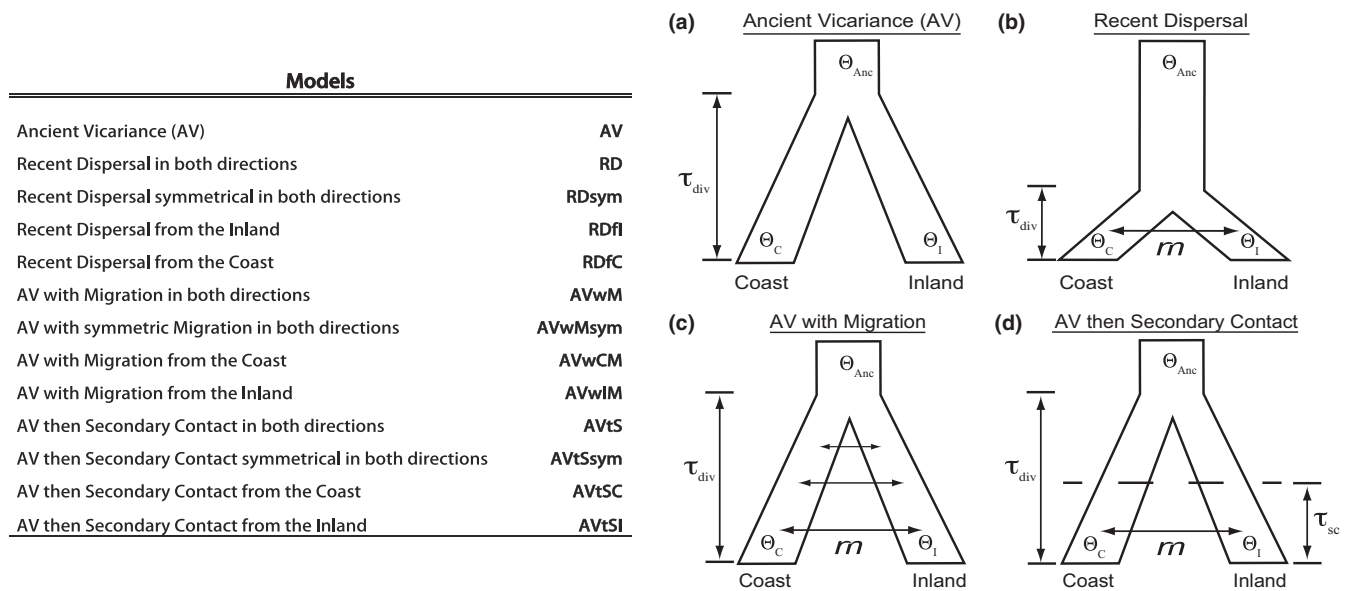
### 2.6.1 | Allele frequency approach

Alleles were grouped based on geography into two populations: a coastal population and an inland population. Folded joint allele frequency spectra (jAFS) were then constructed to summarize biallelic frequencies across both populations. AFS is a commonly used statistic for population genetic inference (Nielsen et al., 2009; Wakeley, 2008), and because of this, jAFS, as well as multidimensional AFS, have been increasingly used for demographic inference (Gutenkunst,

Hernandez, et al., 2009b; Keinan, Mullikin, Patterson, & Reich, 2007; Smith et al., 2017). An AFS cannot accommodate any missing data, and RADseq data are commonly plagued with missing data due to allelic dropout. Therefore, we constructed two sets of jAFS by subsampling SNPs at two different missing data thresholds: 20% and 30%. The threshold value indicates the percentage of individuals from each population that must contain a given SNP for it to be included in the jAFS. To account for variation in the subsampling technique, we constructed 20 jAFS in each subsampling category, for a total of 40 observed jAFS. All jAFS were made using custom Python scripts developed by J. Satler (<https://github.com/jordansatler/r/SNPtoAFS>). The first data set, subsampled at a 20% threshold, included jAFS from ten inland and nine coastal alleles ranging in 65–73 SNPs. The second data set, subsampled at a 30% threshold, included jAFS from 15 inland alleles and 14 coastal alleles ranging in 26–34 SNPs.

Model selection was performed on each observed jAFS using FASTSIMCOAL2 (Excoffier & Foll, 2011; Excoffier et al., 2013). Under this approach, we estimated the composite likelihood of a jAFS between two populations, given a particular demographic model, and for each model parameter. The optimization of each parameter and the composite likelihood was performed using the Expectation-Conditional Maximization (ECM) algorithm (Meng & Rubin, 1993). In ECM, the E-step consisted of 100,000 coalescent simulations to estimate the expected jAFS under the current demographic parameters to approximate the composite likelihood, as in Excoffier et al. (2013). The CM-step consisted of a series of conditional maximizations (Brent, 1974) corresponding to the number of parameters included in the model being investigated. The minimum and maximum number of ECM cycles were set to 10 and 30, respectively. The optimization process ended when the maximum number of cycles was complete, or when the difference in the composite likelihood under the current parameters compared to the likelihood under the proposed parameters was  $<0.001$ . Thirteen different demographic models were evaluated (Figure 1) with 20 independent optimizations from different starting parameters (Excoffier et al., 2013), and the maximum-likelihood parameter estimates resulting from each independent optimization were used as the starting parameters in a final maximization of the composite likelihood. We then calculated Akaike Information Criterion (AIC) values (Akaike, 1974) using the maximum composite likelihood estimated from this run and compared the models using Akaike weights, wAIC (Johnson & Omland, 2004). First, we compared just the four principal demographic models (AV, RD, AVwM, AVtS; Figure 1) using wAIC; then, we compared all 13 models using wAIC. We assume that because the collection of unlinked SNPs are randomly distributed across the genome (Excoffier et al., 2013), the composite likelihood is a good approximation of the true maximum likelihood and can thus be used in AIC calculations for model comparison.

For each model, we estimated  $\tau_{div}$  as divergence time in generations,  $\tau_{sc}$  as time of secondary contact in generations,  $m$  as various probabilities of migration to and from coastal and inland populations,  $\Theta_0$  and  $\Theta_1$ , as  $\Theta = 4N_e\mu$ , where  $N_e$  is the number of genes in each



**FIGURE 1** Four major demographic models representing the hypothesized phylogeographic history of *Alnus rubra* in the PNW temperate rainforest. In total, 13 demographic models were designed and evaluated for both data sets. Names and abbreviations of all models are listed in the left panel. Parameters estimated include population size,  $\theta_C$  and  $\theta_I$ ; migration probability,  $m$ ; divergence time,  $\tau_{div}$ ; and time of secondary contact,  $\tau_{sc}$ . The only additional parameter estimated but not included in the model design was mutation rate

deme and  $\mu$  is the mutation rate (the only parameter not shown in model design; Figure 1) in substitutions/site/generation. The initial values for parameters  $\theta_0$  and  $\theta_1$  were drawn from a log uniform distribution between 0.01 and 10, and the parameter space explored was constrained only by the minimum bound of the prior distribution. The mutation rate,  $\mu$ , was estimated from a minimum bounded log uniform distribution between  $1e^{-9}$  and  $1e^{-7}$ . Divergence time,  $\tau_{div}$ , was drawn from a minimum bounded uniform prior distribution with a minimum of 50,000 and maximum of 1,000,000 generations for all models involving an AV event, whereas  $\tau_{div}$  from recent dispersal models was drawn from a fully bounded, uniform distribution with a minimum of 500 and a maximum 50,000 generations. Divergence time estimates were converted from generations to years using a generation length of 6–8 years per generation (Orwa, Mutua, Kindt, Jamnadass, & Anthony, 2009). The secondary contact models included the time of the gene exchange event ( $\tau_{sc}$ ) as a parameter. The prior distribution for  $\tau_{sc}$  was uniform with a minimum in 500 and maximum in 50,000 generations. Migration was defined as the probability of a given lineage to move from one population to the other and was drawn from a log uniform prior distribution with min  $1e^{-10}$  and max 0.1. Migration was considered as either unidirectional (only from the inland, or only from the coast), bidirectional (a separate migration rate parameter was estimated for each direction), or symmetrically bidirectional (only one migration rate parameter is estimated; Figure 1).

To investigate model adequacy, we performed a goodness-of-fit test, which evaluates whether the observed data fit a particular model. The goodness-of-fit test is carried out using a likelihood ratio G-statistic,  $CLR = \log_{10}(CL_O/CL_E)$ , where  $CL_O$  is the observed maximum composite likelihood and  $CL_E$  is the estimated maximum

composite likelihood (Excoffier et al., 2013). To represent the expected distribution of data given the best model, we performed parametric bootstrapping with the maximum-likelihood (ML) parameter estimates of the selected model to generate 100 simulated jAFS that had an equal number of alleles per population as the empirical data. We then optimized the likelihood of each of these data sets given the model and used these maximum likelihoods to calculate the null distribution for the G-statistic. This process of parametric bootstrapping using the ML parameter estimates was carried out in FASTSIMCOAL2 and repeated for the three best models in each of the subsampling threshold categories, 20% and 30%. The  $p$ -value for each goodness-of-fit test was calculated as the proportion of simulated G-values that were greater than the observed test statistic over all the total number of G-values, in this case 100. These simulated data also permitted calculating 95% confidence intervals for parameters of interest under a model of interest.

## 2.6.2 | Gene tree approach

Phylogeographic inference using approximate likelihoods, or PHRAPL (Jackson et al., 2015), is conducted using gene tree topologies without branch lengths as input. While the gene trees can be constructed using either linked or unlinked SNPs, we opted for the former because we were unable to use linked SNPs in the analysis with allele frequency spectra. For this, all loci that were present in at least four individuals from the coast and four individuals from the inland were used to construct a total of 63 gene trees in PAUP\* (Swofford, 2003). Before constructing trees, we used DT-ModSel (Minin, Abdo, Joyce, & Sullivan, 2003) to select an appropriate model of sequence evolution for each locus. A total of 42 models were



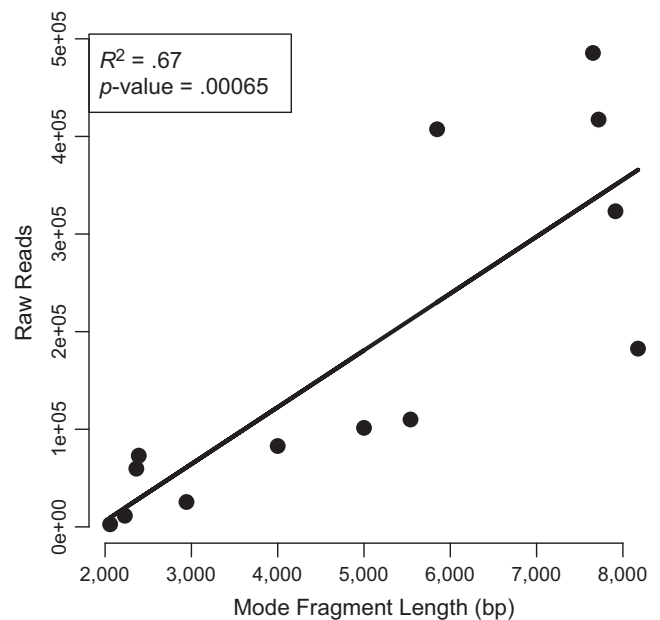
evaluated, and the best model was selected using decision theory. The model and corresponding parameter values were used to construct a maximum-likelihood tree in PAUP\*. A heuristic search started with a neighbour-joining tree and performed tree-bisection–reconnection (TBR) as the branch swapping mechanism for a maximum time limit of 5 min, at which time only the most optimal tree was saved. Two data sets were assembled; in the first data set, only gene trees produced with 20 or more SNPs were included, which resulted in seven trees total. In the second data set, only gene trees made with two or more SNPs were included, which resulted in 46 trees total. We performed the same model selection procedure that is described below, on both data sets.

All 13 models that were designed in FASTSIMCOAL2 were also designed and evaluated in PHRAPL. All the PHRAPL models have one less parameter than the models in FASTSIMCOAL2, because PHRAPL does not estimate mutation rate. PHRAPL uses a grid search to investigate parameter space and optimize approximate likelihoods under user-specified “grid values”. Runs begin with a broad range of grid values for all parameters, and move towards specific values once likelihood peaks are identified. We ran a total of four grid searches on each data set of trees, with 6–8 grid values investigated for the coalescent time parameter and the migration parameter(s) each, on every grid search. The first two grid searches investigated broad ranges in the coalescent time parameter, while the final two grid searches narrowed these values considerably (Table S3). All secondary contact models included an additional parameter representing the timing of the secondary contact, and this event time was set to occur prior to the coalescent event, at a relative time of 0.25 for all runs. We first compared only the four core models (Figure 1) using wAIC and then compared all 13 models at once using wAIC. All computational analyses were carried out using servers at the IBEST Computational Resources Core at the University of Idaho.

### 3 | RESULTS

#### 3.1 | ddRAD sequencing, data quality and effect on missing data

We recovered 648 loci with 614 unlinked biallelic SNPs, 5,494 total variable sites and 79% missing data. We expected to recover c. 6,000 loci, following approximate calculations (Peterson et al., 2012) given a genome size of roughly 5 Mbp (Benenst, Cox, & Leitch, 2012), 8-cutter and 6-cutter restriction enzymes, 70% of a half MiSeq lane (c. 6 million reads), and expected coverage of 20×. Recovery of fewer loci could be due to many reasons, a few of which include protocol modifications, restriction enzyme selection, suboptimal size selection window, or not enough sequencing power (Peterson et al., 2012). Here, the quality of the genomic DNA, or average fragment size, is a primary factor in explaining the variation in the number of reads recovered (Figure 2). Our linear regression analysis showed that the only significant predictor variable was the mode of the fragment size distribution (Figure 2), explaining around



**FIGURE 2** Number of raw reads recovered from ddRADseq experiments in relation to the mode fragment length found in the DNA extracts from 13 *Alnus rubra* individuals

60% of the variation observed in the total number of raw sequence reads.

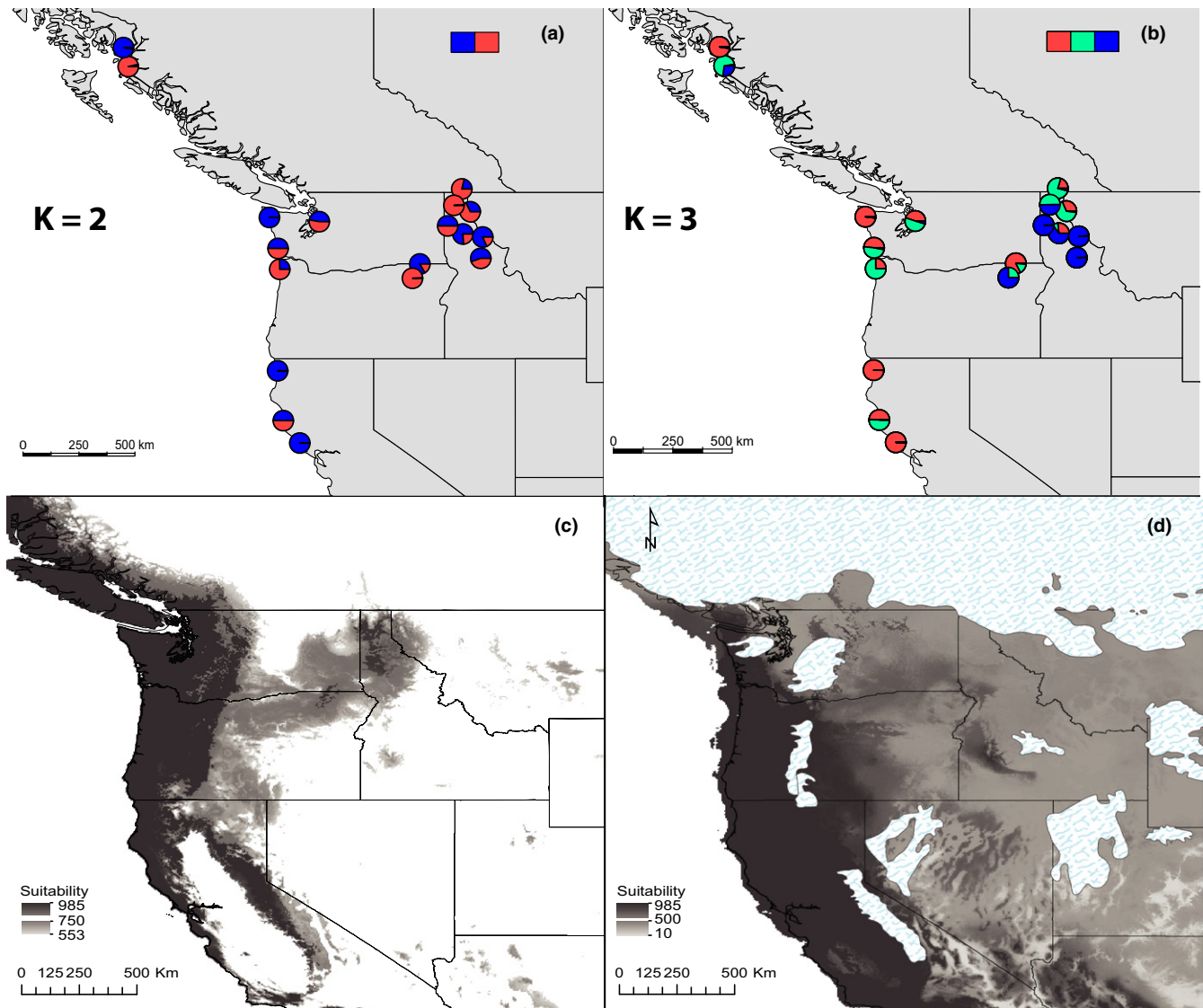
The KS test reported a  $p$ -value of .465 at  $\alpha = 0.05$ , indicating there is not a significant difference between the simulated uniform distribution and the observed distribution of missing data (Figure S2). Thus, we can conclude that the missing data are uniformly distributed across all individuals; that is, there are no individuals that have an extremely high amount of missing data relative to any other individuals, which indicates the data can be subsampled (i.e., missing data discarded) without biasing estimates (Wiuf, 2006).

#### 3.2 | Population structure

Because the analyses of data with missing data can be suspected to involve an overestimation of  $K$  (Pritchard, 2010), we visualized all STRUCTURE results for  $K = 2$ –8. When  $K = 2$  (Figure 3a), there was apparent spatial genetic structure separating coastal from inland populations. This result agrees with the expectations under an AV scenario. When  $K = 3$  (Figure 3b), two of the clusters were restricted to either inland or coastal populations, and the third included individuals from both areas, suggesting gene flow between the disjunct populations. Results for  $K = 4, 5$  and 8 showed no geographic population structure (Figures S3 and S4), additionally suggesting the presence of gene flow between coastal and inland populations.

#### 3.3 | Species distribution models

Our ensemble SDM could successfully recover the current range of the species (Figure 3c). Part of the projected range of the species at



**FIGURE 3** STRUCTURE results for  $K = 2$  (a) and  $K = 3$  (b) for 18 populations of *Alnus rubra*. Species distribution models for *A. rubra* in the PNW rainforest under current (c) and last glacial maximum (LGM; d) conditions. The stippled area in the bottom right panel shows the extent of the Cordilleran ice sheet at the LGM

the LGM (Figure 3d) is substantially different from the current range. Specifically, the coastal area appeared to display high habitat suitability, while the suitable inland areas are more restricted than the current inland range. This suggests that during the LGM, the coastal area may have harboured large extents of continuous habitat for the species, whereas inland populations were likely more restricted.

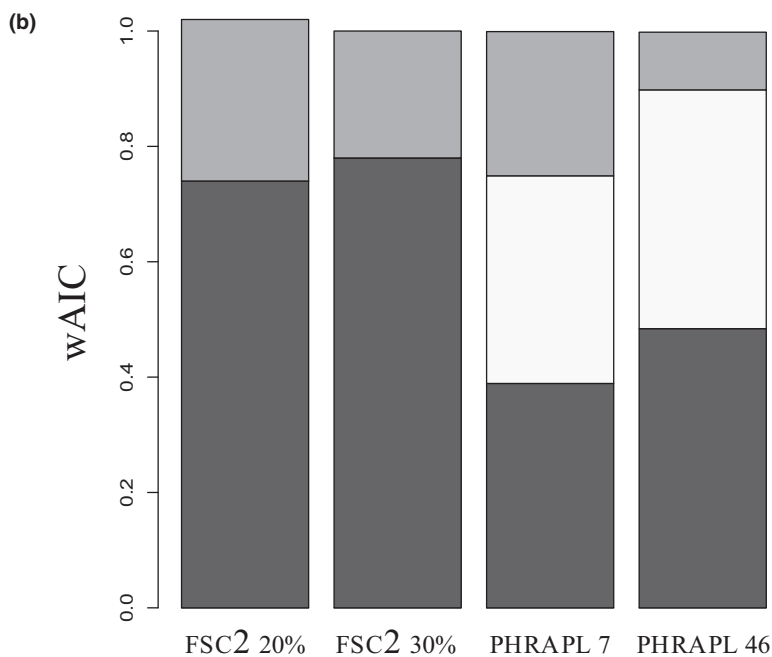
### 3.4 | Demographic model selection

When the four core models were compared, AVwM always had the highest wAIC, on average, regardless of the downsampling technique or whether FASTSIMCOAL2 or PHRAPL was used (Figure 4). AVwM was consistently selected using FASTSIMCOAL2, with fairly high average AIC values (Figure 4). In PHRAPL, with seven trees produced from loci with 20 or more SNPs, the AVwM model was selected 75% of the time

(Table S4) with an average wAIC of 0.389, and the RD model was selected 25% of the time with an average wAIC also of 0.362 (Figure 4). In PHRAPL, with 46 trees produced from loci with two or more SNPs, AVwM and RD were both selected 50% of the time (Table S4) with an average wAIC of 0.484 and 0.414, respectively (Figure 4).

When all 13 models were compared using wAIC from FASTSIMCOAL2 estimated likelihoods, two models were selected consistently, AVwCM and AVwIM. In the first data set (20% subsample threshold), AVwCM was selected around 50% of the time, while AVwIM was selected 43% of the time (Table S5). The remainder of selected models includes AVwM, AVtSC and AVtSI, all of which were selected c. 2.5% of the time. In the second FASTSIMCOAL2 data set (30% subsample threshold), AVwIM was selected 56% of the time, and AVwCM was selected 42% of the time. Only two other models

(a)		Ancient Vicariance	Recent Dispersal	AV with Migration	AV then Secondary Contact
Program	Data				
FASTSIMCOAL2	20%	0.000	0.000	<b>0.745</b>	0.276
	30%	0.000	0.000	<b>0.784</b>	0.215
PHRAPL	7 trees	0.001	0.295	<b>0.316</b>	0.400
	46 trees	0.002	0.513	<b>0.360</b>	0.123



**FIGURE 4** (a) Model selection results between the four major demographic models. (b) Averaged wAIC values across runs in both FASTSIMCOAL2 (FSC2) and PHRAPL data sets. White: Recent Dispersal; dark grey: Ancient Vicariance (AV) with Migration; light grey: AV then Secondary Contact. Notes: AV not shown; average wAIC of 0. Because these are averaged wAIC values, they do not sum to 1

(AVwMsym and AVtSI) were selected, each at very low rates (1% of the time). In both data sets, the three models with the highest average wAIC were AVwIM, AVwCM and AVwM (Figure 5a–b) and for all three models, the mode divergence time estimate was between 5.8 and 6.9 MYA (mean 6.3 MYA) (Table S6).

Model adequacy was evaluated for the three best models for both data sets assessed in FASTSIMCOAL2 (Figure 5c–d). In both data sets, the *p*-values were nonsignificant, indicating that the data fit all three models. In the first data set, the *p*-values for AVwM, AVwCM and AVwIM were .69, .77 and .79, respectively (Figure 5c). In the second data set, the *p*-values for AV with asymmetrical migration, AV with Coastal Migration and AV with Inland Migration were .93 for each (Figure 5d).

In PHRAPL, there were a handful of models amongst the 13 compared that carried a majority of wAIC support in any particular run. In all of the runs, the average wAIC for the best model was 0.16 – 0.26 (Tables S7 and S8), indicating not particularly strong support for any one model. In the data set with seven trees, the AVwIM model was selected 100% of the time; however, the average wAIC for the

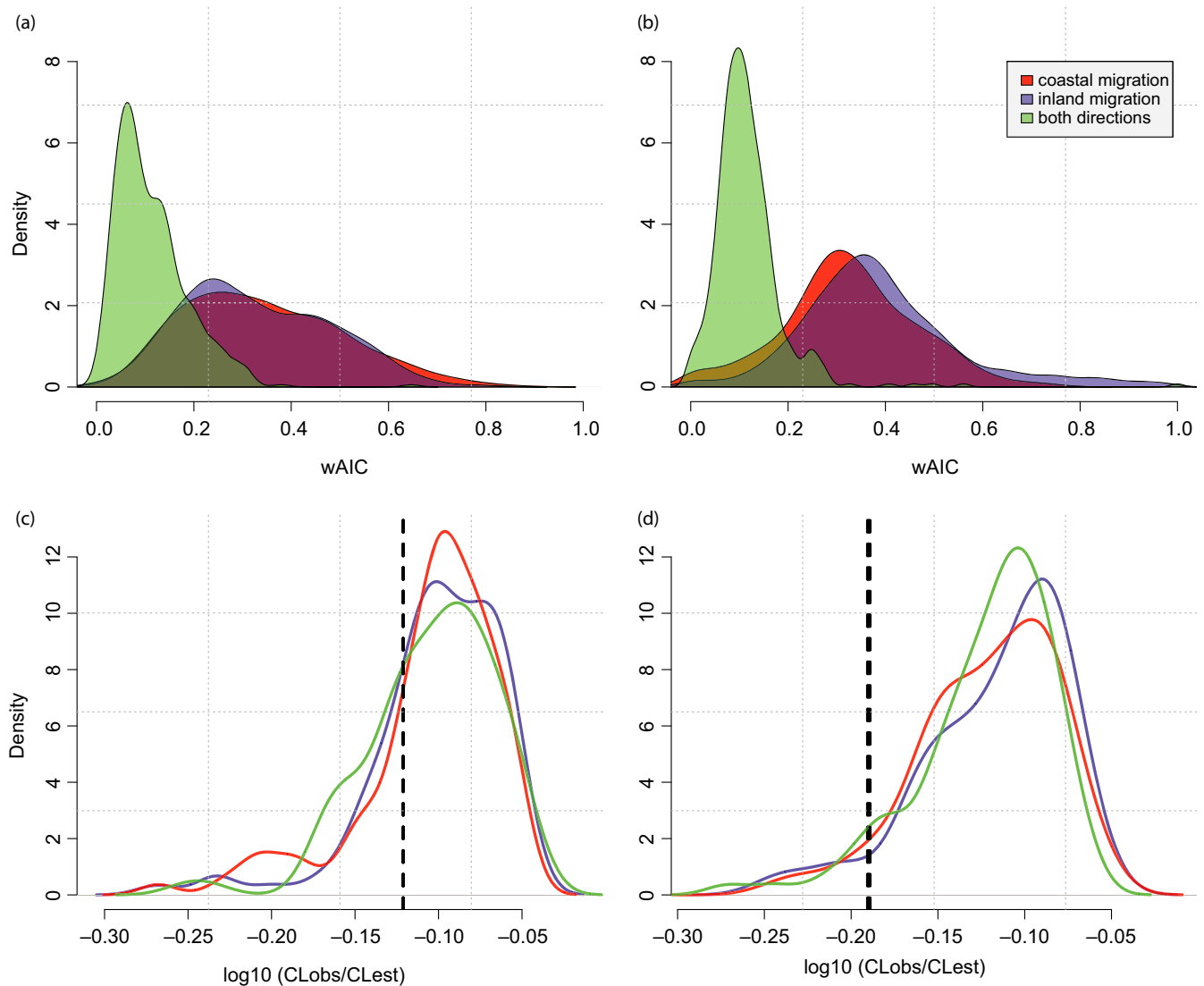
AVwIM model was only 0.167. In all four runs with the seven-tree data set, there were at least five models that carried more than 10% of the model weight (i.e., wAIC > 0.10) (Table S7). In the data set with 46 trees, AVwMsym and RDsym were both selected 50% of time with an average wAIC of 0.22 and 0.214 (Table S8). In all four runs with the 46-tree data set, there were at least four models with more than 10% of the model weight. In both PHRAPL data sets, the RD models collectively occupied over a third of the model weight in any given grid search.

## 4 | DISCUSSION

### 4.1 | Allele frequency and gene tree approaches

FASTSIMCOAL2 and PHRAPL each have unique properties that make them useful in performing model selection. While FASTSIMCOAL2 is appealing because it summarizes unlinked SNPs in an allele frequency spectrum, PHRAPL uses topologies that can be generated from SNP data or entire loci. In this study, the number of unlinked SNPs that could be





**FIGURE 5** Model selection results from *FASTSIMCOAL2* show three models with the highest wAIC densities out of all 13 demographic models evaluated across 400 independent optimizations of the likelihood for each model (20 subsampled jAFS at a subsampling threshold of 20% (a) and 30% (b), each optimized 20 times). The three best models represent continuous migration after the ancient vicariance event with varying migration patterns. (c, d) Model adequacy results for the three best models identified in both jAFS data sets. Note that in each data set, the test statistics for all three models are very similar making it appear as if there is only one test statistic (black dashed line); however, there are actually three overlapping values. In (d), the three test statistics are more variable, hence the appearance of a thicker line. All six evolutions of model adequacy were nonsignificant

used in *FASTSIMCOAL2* was limited because AFS does not accommodate missing data, and therefore, the unlinked SNPs must be downsampled. Alternatively, *PHRAPL* accommodates missing data because not all individuals need be present in each gene tree, only a minimum total number of individuals from each population need be present. Depending on the missing data and number of linked and unlinked SNPs present in one's data set, either approach could be viable.

*FASTSIMCOAL2* results were more consistent than *PHRAPL* results in selecting the same model with high support, especially when only comparing the four core models (Figure 4). However, even when comparing all 13 models in *FASTSIMCOAL2*, AVwCM and AVwIM maintained a majority of the wAIC support. Additionally, model adequacy

in *FASTSIMCOAL2* supported that the data fit the three best models, all of which support an ancient vicariance event with some intermittent gene flow. When comparing the four core models in *PHRAPL*, there was consistency in selecting the AVwM model, but with relatively low support from wAIC. Because no single model had an overwhelming majority of the model weight, there was overall less support in which model was selected when comparing all 13 models in *PHRAPL*. We attribute this identifiability issue to the lack of resolution in many of the gene trees.

As for the performance of each analytical approach, it is unreasonable to conclude on which is more accurate given this study, because each approach performs model selection differently, that is,

uses a different approximate likelihood to evaluate model parameters. Specifically, *FASTSIMCOAL2* uses an observed AFS to optimize an approximate likelihood based on an expected AFS under coalescent simulations (Excoffier et al., 2013), while *PHRAPL* uses gene tree topologies to optimize an approximate likelihood that is based on concordant gene tree topologies under coalescent simulations (Jackson et al., 2015). Both incorporate coalescent theory, and whether one approach is more accurate than the other may depend on whether information in the available data is better captured in an AFS or gene trees.

The 95% confidence intervals for parameter estimates for the three best models are not particularly narrow (Tables S9–14), which suggests that there is a lot of uncertainty associated with the estimation of the parameters, and that the data may not be sufficient for such estimation. However, here we are not specifically drawing inferences from the parameter estimates, but rather, the overall model selected. The power analysis performed (see supplemental methods) shows that the correct model is actually selected a majority of the time (86%, Table S15) when comparing the four main models. Given these results, we feel that regardless of the wide confidence intervals on parameter estimates, we still have considerable power to distinguish between the four main classes of models.

## 4.2 | Genomic data from herbarium specimens

While ddRADseq can be successful in recovering thousands of loci and SNPs (Peterson et al., 2012), this was not the case in our study. Our analyses (Figure 2) show that this is mostly due to lower quality (mode fragment size) DNA used in the protocol. No other variable, or interaction of variables, including year collected or concentration, recovered any significant relationship with the number of reads recovered. The ~10-year-old herbarium specimens used in this study were presumably dried and stored in various ways that resulted in some specimens having better-quality DNA. Our results indicate that as long as high-quality DNA can be identified (e.g., with a Fragment Analyzer or Bioanalyzer), the method used here is a cost-effective approach for generating genomic data. Specifically, our results strongly suggest that the mode of the distribution of fragment sizes is a strong predictor of the number of reads that will be recovered in the ddRAD approach. Examining the mode fragment length as an additional step for standardization can identify degraded samples, that is, samples with an average fragment length <5,000 bp, when concentration alone cannot. Practically, we recommend verifying the quality of individual samples with fragment length identification to ensure that no highly degraded samples are used in library preparation. In addition, characterizing missing data as uniformly missing was crucial for implementing our subsampling strategies for jAFS construction. If the distribution of missing data were systematically structured, subsampling could have drastically biased our likelihood estimates. Taken together, these results suggest that generating genomic data using DNA obtained from herbarium specimens is possible, but the average fragment size of resulting DNAs and the

distribution of missing data should be considered for both the experimental and analytical approaches employed here.

Though we have investigated the use of herbarium tissue and seen promising results, it remains unclear whether genomic data generated from herbarium specimens can be combined with genomic data generated from fresh tissue. Given what we have learned about the influence of fragment size on the number of loci recovered, we imagine that the disparity in allelic dropout between herbarium specimens and fresh tissue would be quite large, potentially making the data unworkable together. Thus, we emphasize to proceed with caution if combining fresh tissue with herbarium tissue for studies using RADseq family approaches for data collection.

## 4.3 | Phylogeographic history of red alder

Overall, the Ancient Vicariance with Migration (AVwM) model was the strongest and most consistently supported model across data sets and approaches. Although this model selection approach was not intended to make inferences from the parameter estimates, but rather from the overall model selected, the divergence times estimated seem to indicate congruence with the PNW history of pre-Pleistocene divergence. The divergence time estimates ranged 5.8–6.9 MYA and are older than the Pacific Northwest rainforest disjunction (~3–5 MYA; but see the 95% CI in Table S8). Although we cannot precisely determine divergence times, we can characterize the timing of divergence as pre-Pleistocene.

The AVwM model involves four possible migration scenarios. Two of these four scenarios are highly supported with allele frequency data and include migration in only one direction, AVwIM and AVwCM. When comparing all models, gene trees supported AVwIM as the best model. These results suggest that migration has been predominantly unidirectional, although the actual direction still remains unclear in light of these results. Disentangling which direction is the most likely requires investigating recolonization and/or specific migration route models, potentially, with the inclusion of more genetic data. Due to the lack of genetic structure within the coast or the inland, we do not think that expanding the geographic breadth of our samples is necessary (Figures S3 and S4). That said, model adequacy results suggest that more genetic data are required to distinguish between the three best migration scenarios for the AVwM model. These results also highlight the limit to the phylogeographic inference that we are able to make given these data, which is something that we feel should always be identified.

The SDMs and population structure results provide further evidence for the AVwM scenario. The climatic niche projections using LGM conditions (Figure 3d) indicate that the expected inland range of *A. rubra* shifted into southern Idaho. This is consistent with the hypothesis that species ranges in this area were displaced south during the Pleistocene (Sullivan et al., 2000). Although not apparent in the SDM because of its later occurrence, pollen records indicate the presence of *Alnus* in the NRM of Canada throughout the Holocene (Gavin, Hu, Walker, & Westover, 2009). This could indicate surviving populations of *A. rubra* in nunatak refugia as far north as Canada, or

a rapid colonization of the area following glacial retreat. During the Holocene, the climate was considerably warmer than today (Wagner, Melles, Hahne, Niessen, & Hubberten, 2000), which could explain why the inland forest extended much farther North than it does currently. Ultimately, the persistence of *A. rubra* in the inland during the Pleistocene is supported by our results, although whether southern inland populations, or migrants from the northern Cascades, or both, colonized the inland forest remains unclear. The strong genetic divide between the inland and coastal populations (Figure 3a–b) also corroborate that inland populations of *A. rubra* likely persisted through the Pleistocene.

Given the prominent role that gene flow has played in the phylogeographic history of *A. rubra*, we conclude, in agreement with Espindola et al.'s (2016) prediction, that coastal and inland populations of red alder do not harbour cryptic diversity and thus do not represent incipient sister species. Previously, in this system, non-cryptic taxa were often considered to be the result of recent dispersal. However, we show here that an alternative phylogeographic hypothesis—specifically, ancient vicariance with periods of gene flow—can also explain why some lineages in the disjunct mesic forests of the PNW may not harbour cryptic diversity, despite evidence of ancient population structure. We also show that the inclusion of more intraspecific data, genetic and geographic, does in fact increase our phylogeographic understanding of *A. rubra*, specifically because the AVwM and RD models could not have been distinguished using the cpDNA from Strengé (1994; Soltis et al., 1997) or Brumble (2008). Finally, we acknowledge that the inclusion of more data would allow for the evaluation of more complex models; therefore, if in the future more *A. rubra* data are generated, we recommend the evaluation of more complex models that include population expansion and contraction.

## 5 | CONCLUSIONS

In this study, we compared two approaches in phylogeographic model selection, allele frequency-based (Excoffier et al., 2013) and gene tree-based (Jackson et al., 2015), and used the results from both to draw phylogeographic inference in an emerging model system in comparative phylogeography. Importantly, both approaches resulted in a ranking of models that was useful in gauging relative support for all competing models. The most overwhelming indicator of successful model selection comes from the review of model adequacy, where we see the data consistently fit the models that were most frequently selected. Because assessing model fit is a critical component of any statistical inference, we feel that future phylogeographic studies should include explicit tests of model adequacy, as performed here. Further, we also conclude that the mode fragment length is an effective measure of sample quality that will help in identifying samples that may be problematic for RAD-based genomic reduction sequencing strategies – samples where the concentration alone is not enough to indicate levels of degradation – early on in library preparation. We also demonstrate that genomic data obtained

from DNA isolated from herbarium specimens do not necessarily result in systematically missing amounts of data, which allows for downsampling without the fear of drastically biasing the data. Finally, we were successful in using DNA from herbarium specimens to gather the genomic data necessary to make inferences regarding the phylogeographic history of *A. rubra*, where the combination of descriptive and model selection based tools was invaluable in recovering a meaningful phylogeographic inference that is supported by multiple, independent lines of evidence.

## ACKNOWLEDGEMENTS

Support for this work comes from National Science Foundation grant nos. DEB-1457519 and DEB-1457726, and the Institute for Bioinformatics and Evolutionary Studies (IBEST) at the University of Idaho, which is supported by NIH NCRR 1P20RR016454-01, NIH NCRR 1P20RR016448-01 and NSF EPS-809935. We thank Paul Hohenlohe and Kim Andrews for advice on sequencing and library preparation, and the Tank Lab group for helpful reviews of the earlier versions of this manuscript.

## DATA ACCESSIBILITY

Parameter file for processing raw reads in PyRAD, along with the unprocessed sequence reads, the full SNP data set and 63 loci are available on Dryad (<https://doi.org/10.5061/dryad.288c2>).

## AUTHOR CONTRIBUTIONS

M.R. and A.E. generated and analysed the data. All authors contributed to the design of the study and drafting the article. All authors approved the final version of the manuscript.

## ORCID

Megan Ruffley  <http://orcid.org/0000-0003-1796-2719>

Anahí Espindola  <http://orcid.org/0000-0001-9128-8836>

Bryan C. Carstens  <http://orcid.org/0000-0002-1552-227X>

## REFERENCES

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, advance on, *17*, 81–92. <https://doi.org/10.1038/nrg.2015.28>
- Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., ... Saunders, N. C. (1987). Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, *18*, 489–522.
- Beerli, P., & Felsenstein, J. (2001). Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations

- by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 4563–4568. <https://doi.org/10.1073/pnas.081068098>
- Benenett, M. D., Cox, A. V., & Leitch, I. J. (2012). *Angiosperm DNA C-values database (release 8.0, Dec. 2012)*. Retrieved from <http://www.kew.org/cvalues/homepage.html>.
- Brent, R. P. (1974). Algorithms for minimization without derivatives. *IEEE Transactions on Automatic Control*, 19, 632–633.
- Brumble, A. J. (2008). *Phylogeography of Alnus rubra*. Master's Thesis. University of Idaho.
- Brunsfeld, S. J., Miller, T. A., & Carstens, B. C. (2007). Insights into the biogeography of the Pacific Northwest of North America: Evidence from the phylogeography of *Salix melanopsis*. *Systematic Biology*, 32, 129–139.
- Brunsfeld, S. J., Sullivan, J., Soltis, D. E., & Soltis, P. S. (2001). Comparative phylogeography of north-western North America: A synthesis. In: *Integrating ecological and evolutionary processes in a spatial context*, pp. 319–339.
- Carstens, B. C., Brennan, R. S., Chua, V., Duffie, C. V., Harvey, M. G., Koch, R. A., ... Seeholzer, G. (2013). Model selection as a tool for phylogeographic inference: An example from the willow *Salix melanopsis*. *Molecular Ecology*, 22, 4014–4028. <https://doi.org/10.1111/mec.12347>
- Carstens, B., Lemmon, A. R., & Lemmon, E. M. (2012). The promises and pitfalls of next-generation sequencing data in phylogeography. *Systematic Biology*, 61, 713–715. <https://doi.org/10.1093/sysbio/sys050>
- Carstens, B. C., Stevenson, A. L., Degenhardt, J. D., & Sullivan, J. (2004). Testing nested phylogenetic and phylogeographic hypotheses in the *Plethodon vandykei* species group. *Systematic Biology*, 53, 781–792. <https://doi.org/10.1080/10635150490522296>
- Carstens, B. C., Stoute, H. N., & Reid, N. M. (2009). An information-theoretical approach to phylogeography. *Molecular Ecology*, 18, 4270–4282. <https://doi.org/10.1111/j.1365-294X.2009.04327.x>
- Csilléry, K., François, O., & Blum, M. G. B. (2012). abc: An R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution*, 3, 475–479. <https://doi.org/10.1111/j.2041-210X.2011.00179.x>
- DeChaine, E. G., & Martin, A. P. (2005). Marked genetic divergence among sky island populations of *Sedum lanceolatum* (Crassulaceae) in the Rocky Mountains. *American Journal of Botany*, 92, 477–486. <https://doi.org/10.3732/ajb.92.3.477>
- Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, 19, 11–15.
- Eaton, D. A. R. (2014). PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, 30, 1844–1849. <https://doi.org/10.1093/bioinformatics/btu121>
- Eaton, D. A. R., & Ree, R. H. (2013). Inferring phylogeny and introgression using RADseq data: An example from flowering plants (*Pedicularis*: Orobanchaceae). *Systematic Biology*, 62, 689–706. <https://doi.org/10.1093/sysbio/syt032>
- Edgar, R. C. (2004). MUSCLE user guide. *Nucleic Acids Research*, 32, 1–15.
- Espíndola, A., Ruffley, M., Smith, M. L., Carstens, B. C., Tank, D. C., & Sullivan, J. (2016). Identifying cryptic diversity with predictive phylogeography. *Proceedings of the Royal Society B: Biological Sciences*, 283, 20161529.
- Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics*, 9, e1003905. <https://doi.org/10.1371/journal.pgen.1003905>
- Excoffier, L., & Foll, M. (2011). fastsimcoal: A continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*, 27, 1332–1334. <https://doi.org/10.1093/bioinformatics/btr124>
- Fagundes, N. J., Ray, N., Beaumont, M., Neuenschwander, S., Salzano, F. M., Bonatto, S. L., & Excoffier, L. (2007). Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences*, 104, 17614–17619.
- Faircloth, B., & Glenn, T. (2012). AMPure XP substitute cost savings protocol. *Genome Research*, 22, 939–946.
- Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, 164, 1567–1587.
- Gavin, D. G., Hu, F. S., Walker, I. R., & Westover, K. (2009). The Northern Inland temperate rainforest of British Columbia: Old forests with a young history? *Northwest Science*, 83, 70–78. <https://doi.org/10.3955/046.083.0107>
- Gelman, A., & Shalizi, C. R. (2013). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, 66, 8–38. <https://doi.org/10.1111/j.2044-8317.2011.02037.x>
- Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009a). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, 5, e1000695.
- Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009b). Diffusion Approximations for Demographic Inference: *∂a∂i*. *Nature Precedings*.
- Hey, J. (2010). Isolation with migration models for more than two populations. *Molecular Biology and Evolution*, 27, 905–920. <https://doi.org/10.1093/molbev/msp296>
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965–1978. [https://doi.org/10.1002/\(ISSN\)1097-0088](https://doi.org/10.1002/(ISSN)1097-0088)
- Hugall, A., Moritz, C., Moussalli, A., & Stanisic, J. (2002). Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosophia bellendenkerensis* (Brazier 1875). *Proceedings of the National Academy of Sciences*, 99, 6112–6117. <https://doi.org/10.1073/pnas.092538699>
- Jackson, N. D., Morales, A., Carstens, B. C., & O'Meara, B. C. (2015). Phylogeographic inference using approximate likelihoods. *Journal of Chemical Information and Modeling*, 53, 160.
- Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology and Evolution*, 19, 101–108. <https://doi.org/10.1016/j.tree.2003.10.013>
- Keinan, A., Mullikin, J. C., Patterson, N., & Reich, D. (2007). Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. *Nature Genetics*, 39, 1251–1255. <https://doi.org/10.1038/ng2116>
- Kingman, J. (1982). The coalescent. *Stochastic Processes and their Applications*, 13, 235–248. [https://doi.org/10.1016/0304-4149\(82\)90011-4](https://doi.org/10.1016/0304-4149(82)90011-4)
- Knowles, L. L. (2001). Genealogical portraits of speciation in montane grasshoppers (genus *Melanoplus*) from the sky islands of the Rocky Mountains. *Proceedings of the Royal Society B: Biological Sciences*, 268, 319–324. <https://doi.org/10.1098/rspb.2000.1364>
- Knowles, L. L., & Maddison, W. P. (2002). Statistical phylogeography. *Molecular Ecology*, 11, 2623–2635. <https://doi.org/10.1046/j.1365-294X.2002.01637.x>
- Meng, X.-L., & Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, 80, 267–278. <https://doi.org/10.1093/biomet/80.2.267>
- Metzger, G., Espíndola, A., Waits, L. P., & Sullivan, J. (2015). Genetic structure across broad spatial and temporal scales: Rocky mountain tailed frogs (*Ascaphus mantanus*; 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. <https://doi.org/10.1080/10635150390235494>
- Nielsen, R., Hubisz, M. J., Hellmann, I., Torgerson, D., Andrés, A. M., Albrechtsen, A., ... Indap, A. (2009). Darwinian and demographic forces affecting human protein coding genes. *Genome Research*, 19, 838–849.

- 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. <https://doi.org/10.1111/j.0014-3820.2001.tb01280.x>
- Orwa, C., Mutua, A., Kindt, R., Jamnadass, R., & Anthony, S. (2009). *Agroforestry Database: a tree reference and selection guide version 4.0*. Retrieved from [www.worldagroforestry.org/sites/treedbs/treedatabases.asp](http://www.worldagroforestry.org/sites/treedbs/treedatabases.asp), 0, 1–5.
- Panchenko, P. (2006). Kolmogorov-Smirnov Test. In: *Statistics for Applications*, pp. 83–90.
- Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M., & Araújo, M. B. (2011). *Ecological niches and geographic distributions*. Princeton, NJ: Princeton University Press.
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, 7, e37135. <https://doi.org/10.1371/journal.pone.0037135>
- Petit, R. J., & Grivet, D. (2002). Optimal randomization strategies when testing the existence of a phylogeographic structure [1]. *Genetics*, 161, 469–471.
- Pinceel, J., Jordaens, K., Pfenninger, M., & Backeljau, T. (2005). Range-wide phylogeography of a terrestrial slug in Europe: Evidence for Alpine refugia and rapid colonization after the Pleistocene glaciations. *Molecular Ecology*, 14, 1133–1150. <https://doi.org/10.1111/j.1365-294X.2005.02479.x>
- Priest, G. R. (1990). Volcanic and tectonic evolution of the cascade volcanic arc, central Oregon. *Journal of Geophysical Research*, 95, 19583–19599. <https://doi.org/10.1029/JB095iB12p19583>
- Pritchard, J. K. (2010). Documentation for structure software: Version 2.3. *In Practice*, 6, 321–326.
- Ramasamy, R. K., Ramasamy, S., Bindroo, B. B., & Naik, V. G. (2014). STRUCTURE PLOT: A program for drawing elegant STRUCTURE bar plots in user friendly interface. *SpringerPlus*, 3, 431. <https://doi.org/10.1186/2193-1801-3-431>
- Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, 4, e2409v1.
- Rohland, N., & Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, 22, 939–946. <https://doi.org/10.1101/gr.128124.111>
- Smith, M. L., Ruffley, M., Espíndola, A., Tank, D. C., Sullivan, J., & Carstens, B. C. (2017). Demographic model selection using random forests and the site frequency spectrum. *Molecular Ecology*, 26, 4562–4573. <https://doi.org/10.1111/mec.14223>
- Soltis, D. E., Gitzendanner, M. A., Strenge, D. D., & Soltis, P. S. (1997). Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. *Plant Systematics and Evolution*, 206, 353–373. <https://doi.org/10.1007/BF00987957>
- Steele, C. A., Carstens, B. C., 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. <https://doi.org/10.1016/j.ympev.2004.12.001>
- Strenge, D. (1994). *The intraspecific Phylogeography of Polystichum munitum and Alnus rubra*. Master's Thesis. Washington State University.
- Sullivan, J., Arellano, E., & Rogers, D. S. (2000). Comparative phylogeography of mesoamerican highland rodents: Concerted versus independent response to past climatic fluctuations. *The American Naturalist*, 155, 755–768. <https://doi.org/10.1086/303362>
- Swofford, D. L. (2003). *PAUP\*: phylogenetic analysis using parsimony, version 4.0b10*. 21 Libro, 11 pp.
- Templeton, A. R. (2008). Nested clade analysis: An extensively validated method for strong phylogeographic inference. *Molecular Ecology*, 17, 1877–1880. <https://doi.org/10.1111/j.1365-294X.2008.03731.x>
- Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD – A platform for ensemble forecasting of species distributions. *Ecography*, 32, 369–373. <https://doi.org/10.1111/j.1600-0587.2008.05742.x>
- Varma, A., Padh, H., & Shrivastava, N. (2007). Plant genomic DNA isolation: An art or a science. *Biotechnology Journal*, 2, 386–392. [https://doi.org/10.1002/\(ISSN\)1860-7314](https://doi.org/10.1002/(ISSN)1860-7314)
- Wagner, B., Melles, M., Hahne, J., Niessen, F., & Hubberten, H. W. (2000). Holocene climate history of Geographical Society Ø, East Greenland — evidence from lake sediments. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 160, 45–68. [https://doi.org/10.1016/S0031-0182\(00\)00046-8](https://doi.org/10.1016/S0031-0182(00)00046-8)
- Wakeley, J. (2008). *Coalescent Theory: An Introduction*.
- Waring, R. H., & Franklin, J. F. (1979). Evergreen coniferous forests of the pacific northwest. *Science (New York, N.Y.)*, 204, 1380–1386. <https://doi.org/10.1126/science.204.4400.1380>
- Wiuf, C. (2006). Consistency of estimators of population scaled parameters using composite likelihood. *Journal of Mathematical Biology*, 53, 821–841. <https://doi.org/10.1007/s00285-006-0031-0>
- Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: A fast and accurate Illumina Paired-End reAd merger. *Bioinformatics*, 30, 614–620. <https://doi.org/10.1093/bioinformatics/btt593>

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Ruffley M, Smith ML, Espíndola A, Carstens BC, Sullivan J, Tank DC. Combining allele frequency and tree-based approaches improves phylogeographic inference from natural history collections. *Mol Ecol*. 2018;00:1–13. <https://doi.org/10.1111/mec.14491>