

# Genomic evidence of an ancient inland temperate rainforest in the Pacific Northwest of North America

Megan Ruffley<sup>1,2</sup>  | Megan L. Smith<sup>3</sup> | Anahí Espíndola<sup>4</sup>  | Daniel F. Turck<sup>1,5</sup> | Niels Mitchell<sup>1</sup> | Bryan Carstens<sup>3</sup>  | Jack Sullivan<sup>1,2</sup>  | David C. Tank<sup>1,2,5</sup>

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

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

<sup>3</sup>Department of Evolution, Ecology, and Organismal Biology & Museum of Biological Diversity, The Ohio State University, Columbus, Ohio, USA

<sup>4</sup>Department of Entomology, University of Maryland, College Park, Maryland, USA

<sup>5</sup>Stillinger Herbarium, University of Idaho, Moscow, Idaho, USA

## Correspondence

Megan Ruffley, Department of Plant Biology, Carnegie Institution for Science, Stanford, CA 94305, USA.  
Email: [meganruffley@gmail.com](mailto:meganruffley@gmail.com)

## Present address

Megan Ruffley, Department of Plant Biology, Carnegie Institution for Science, Stanford, California, USA

Megan L. Smith, Department of Biology and Department of Computer Science, Indiana University, Bloomington, Indiana, USA

David C. Tank, Department of Botany & Rocky Mountain Herbarium, University of Wyoming, Laramie, Wyoming, USA

## Funding information

National Science Foundation, Grant/Award Number: DEB-1457726, NSF EPS-809935 and DEB-1457519; NIH, Grant/Award Number: NCR1 1P20RR016448-01 and NCR1 1P20RR016454-01

Handling Editor: Victoria Sork

## Abstract

The disjunct temperate rainforests of the Pacific Northwest of North America (PNW) are characterized by late-successional dominant tree species *Thuja plicata* (western redcedar) and *Tsuga heterophylla* (western hemlock). The demographic histories of these species, along with the PNW rainforest ecosystem in its entirety, have been heavily impacted by geological and climatic changes the PNW has experienced over the last 5 million years, including mountain orogeny and repeated Pleistocene glaciations. These environmental events have ultimately shaped the history of these species, with inland populations potentially being extirpated during the Pleistocene glaciations. Here, we collect genomic data for both species across their ranges to test multiple demographic models, each reflecting a different phylogeographical hypothesis on how the ecosystem-dominating species may have responded to dramatic climatic change. Our results indicate that inland and coastal populations in both species diverged ~2.5 million years ago in the early Pleistocene and experienced decreases in population size during glacial cycles, with subsequent population expansion. Importantly, we found evidence for gene flow between coastal and inland populations during the mid-Holocene. It is likely that intermittent migration in these species during this time has prevented allopatric speciation via genetic drift alone. In conclusion, our results from combining genomic data and demographic inference procedures establish that populations of the ecosystem dominants *Thuja plicata* and *Tsuga heterophylla* persisted in refugia located in both the coastal and inland regions of the PNW throughout the Pleistocene, with populations expanding and contracting in response to glacial cycles with occasional gene flow.

## KEYWORDS

demographic modelling, genomics, Pacific Northwest, phylogeography, site frequency spectrum, temperate rainforest

## 1 | INTRODUCTION

The old-growth cedar–hemlock forests of the Pacific Northwest of North America (PNW) characterize one of the most diverse temperate rainforests in the world (Newmaster et al., 2003). This ecosystem includes disjunct coastal and inland temperate rainforest (ITR) elements, with the latter located in the northern Rocky Mountains, completely separated from the larger coastal rainforest located in the Cascades ranges from northern California to Alaska. The coastal rainforest and the ITR are separated by ~300 km of xeric shrubland habitat that does not support range connectivity for species with conspecific populations across the disjunction. The mesic forest ecosystem sits between low-elevation xerophytic forests and subalpine, boreal forests, and harbours over 150 animal and plant lineages with disjunct populations (Brunsfeld et al., 2001; Nielson et al., 2001). The entire PNW region has been widely impacted by Pleistocene glacial/interglacial cycles (Waitt & Thorson, 1983), with its biota being strongly affected by these climatic changes. The ITR has been of particular interest because of the dramatic implications of the alternative hypotheses that have been proposed to explain its history during and after the Pleistocene. One, the Recent Dispersal (RD) hypothesis, posits the recent (<5000 years ago; ka) establishment of the ITR, invoking a post-Pleistocene colonization of the inland areas from coastal populations, and implying that the ITR is a recent derivative of the coastal forest with little evolutionary novelty (Brunsfeld et al., 2001; Daubenmire, 1975). The second hypothesis posits that the ITR represents an ancient disjunction between the inland and coastal forests (Brunsfeld et al., 2001) that occurred pre-Pleistocene (>2.5 million years ago; Ma). Under this Ancient Vicariance (AV) hypothesis, though the onset of the glaciers caused a massive contraction of the ITR populations, inland refugia persisted during cold periods and populations subsequently expanded *in situ* after the Pleistocene. The AV hypothesis predicts that allopatry may have led to speciation of some taxa in the ITR, and thus that the inland region harbours a unique endemic flora and fauna. These two hypotheses broadly encapsulate the proposed modes of the formation of the disjunction. They are critical to understanding general biogeographical processes associated with the ecosystem and the region, and they also have broad implications for conservation and management of this diversity hotspot.

The range of the PNW temperate rainforest is defined by the distributions of the two late-successional dominant species, *Tsuga heterophylla* Raf. (Sarg.) (western hemlock) and *Thuja plicata* Donn ex. D. Don. (western redcedar). Pollen records from the central and southern ITR suggest these species have only been present in the inland for <3000 years (Chase et al., 2008; Gavin et al., 2009; Mehringer, 1996; Rosenberg et al., 2003) and represents evidence for the RD hypothesis. Suitable inland habitat for *Tsuga heterophylla* is not completely occupied, suggesting the species range is still experiencing post-glacial expansion (Gavin & Hu, 2006). Similarly, Rosenberg et al. (2003) found no record of this species' pollen in southeastern British Columbia prior to ~3,500 years ago. Most pollen records concur that *Tsuga heterophylla* pre-dates evidence

of *Thuja plicata* (Mehringer, 1996; Whitlock, 1992). This coincides with microsatellite molecular evidence for *Thuja plicata* samples across the disjunct range (O'Connell et al., 2008), which supports one southern coastal refugium throughout the Pleistocene, with no evidence for ancient, disjunct inland refugia, nor for northern coastal refugia (e.g., the Haida Gwaii archipelago). O'Connell et al. (2008) further suggested that, given the lack of hierarchical structure in these coastal and inland population clusters, the divergence between them has been recent and rapid, which is congruent with post-glacial recolonization of the northern coast and ITR (O'Connell et al., 2008). While this inference was based on microsatellite loci, recent advances with reduced representation sequencing (Andrews et al., 2016; Peterson et al., 2012) provide enhanced power to test phylogeographical hypotheses regarding formation of disjunct populations (Carstens et al., 2012; Garrick et al., 2015). Indeed, a recent study used clustering analyses for population assignment based on a genome-wide panel of single nucleotide polymorphisms (SNPs) in *Thuja plicata* to support the presence of an ITR refugium (Fernandez et al., 2021). However, Fernandez et al. (2021) did not attempt to estimate population parameters, such as the timing of demographic events through model comparisons, which are critical to evaluating the relevant demographic hypotheses.

Though much of the available evidence supports the recent, post-glacial colonization of the ITR by these dominant tree species, a number of phylogeographical investigations have been conducted to evaluate the cause of the disjunction on other species in the PNW temperate rainforest (e.g., Brunsfeld et al., 2001; Soltis et al., 1997). To date, 13 other species complexes with disjunct ranges in the PNW have been investigated in a phylogeographical framework, including five amphibians (Carstens et al., 2004; Metzger et al., 2015; Nielson et al., 2001; Steele et al., 2005), one mammal (Carstens et al., 2005), two plants (Brunsfeld et al., 2007; Carstens et al., 2013; Ruffley et al., 2018), four molluscs (Smith et al., 2017, 2018; Wilke & Duncan, 2004) and an arthropod (Espíndola et al., 2016). These species span the tree of life and, based on analyses of their genetic variation, they also span the possible phylogeographical histories for the PNW temperate rainforest. Some species, such as the tailed frogs (*Ascaphus*; Nielson et al., 2001), giant salamanders (*Dicamptodon*, *Plethodon*; Carstens et al., 2004; Steele et al., 2005) and jumping slugs (*Hemphilia*; Rankin et al., 2019), show clear evidence of an ancient divergence between the ITR and coastal populations, indicating pre-Pleistocene divergence, and in some cases speciation. Conversely, other species, such as dusky willow trees (*Salix melanopsis*; Carstens et al., 2013), water voles (*Microtus richardsoni*; Carstens et al., 2005) and taildropper slugs (*Prophysaon andersoni*; Smith et al., 2018), show evidence of post-glacial recolonization of the inland from the coastal populations. Other phylogeographical models, such as pre-Pleistocene divergence with migration, have also been supported with genomic evidence (*Alnus rubra*; Ruffley et al. 2018). These results suggest that some ITR endemics might have been present before the ecosystem-dominant species were established if these ecosystem dominants colonized the ITR more

recently, as supported by pollen records and early molecular studies. Inferring the phylogeographical history of the ecosystem dominant species that establish the boundaries of the PNW temperate rainforest will provide critical insight for the availability of suitable habitat for refugial populations in the ITR, and will be a central contribution towards our understanding of the biogeographical history of the area.

The hypothesis that the ITR has an ancient (pre-Pleistocene) divergence from the coastal rainforest and has persisted throughout glacial cycles in refugia in the interior Northwest is compelling because its presence would support the habitat requirements of other species that show evidence of ancient vicariance. Additionally, palaeontologists have questioned the plausibility of the old-growth ITR becoming so established in less than 3500 years (Mehring, 1996). However, the persistence of the ITR throughout the Pleistocene has received no support from the pollen record (e.g., Chase et al., 2008; Gavin et al., 2009; Mehring, 1996). Whether the ITR persisted throughout the Pleistocene also has other implications for how the PNW disjunct community as a whole has adapted to the dramatic climatic changes, either in concert with one another or individualistically (Davis, 1981; Flessa et al., 2005; Habeck, 1987; Sullivan et al., 2000). Common insight from palaeoecology suggests that modern communities of PNW forest have assembled over a long history of individual responses to climate change (Davis, 1981; Flessa et al., 2005), and the hypothesis of a recently assembled, rapidly diverse ITR poses a challenge to this insight.

This question is not novel to the PNW temperate rainforest, as the idea of the species responding individualistically, rather than in-concert, in response to climatic changes is an ecological hypothesis dating back to the early 20th century (Gleason, 1926), and has been demonstrated empirically (Burbrink et al., 2016). However, Kirkwood (1922), one of the first to characterize the ecology of species in the northern Rocky Mountains in general, emphasized how understanding of the ITR would be dramatically improved when “the individualities of the constituent species were understood.” Alternatively, there is evidence in other communities that species do, in fact, respond to climatic changes in concert (Chan et al., 2014; Gehera et al., 2017). For plant communities specifically, the idea of community-wide concerted response to climatic change can be traced back to early 20th century plant ecologist Clements (1916), and his idea that communities are “super-organisms” whose interactions are interwoven and dependent on one another. Regardless of the organism or ecosystem, researchers have long been fascinated with the question of whether species in the same environment respond asynchronously or synchronously to climatic changes (Carstens et al., 2005; Hickerson et al., 2006; Sullivan et al., 2000).

In this study, we test predictions about the phylogeographical history of these two species, specifically with respect to whether they harbour cryptic diversity across the disjunction; that is, they show evidence of pre-Pleistocene divergence and no subsequent migration. These predictions serve as a test of

the phylogeographical predictive framework that was originally developed by Espíndola et al. (2016) and recently updated with life history traits by Sullivan et al. (2019). We then validate these predictions, and ultimately test whether the ITR persisted throughout the Pleistocene (Brunsfield et al., 2001), by generating genomic data for individuals from these species throughout their ranges. For this, we rely on coalescent simulations, the joint site frequency spectrum (jSFS), and machine learning inference procedures to develop and test our phylogeographical hypotheses. We also validate the power of predictive phylogeography (Espíndola et al., 2016; Sullivan et al., 2019) in detecting the presence and absence of cryptic diversity. Finally, we evaluate the potential role of genomic data in uncovering the history of the past and explore how our inferences can be influenced by various data types and perspectives in genomics and palaeontology.

## 2 | MATERIALS AND METHODS

### 2.1 | Field sampling and sequencing

Field collections of plant material were made throughout the coastal and inland PNW temperate rainforest for western redcedar, *Thuja plicata*, and western hemlock, *Tsuga heterophylla*, between April and June of 2016 and 2018. Fresh tissue from specimens was dried and stored in silica gel. Voucher specimens of collections were preserved in the Stillinger herbarium and can be located on the Consortium of Pacific Northwest Herbaria data portal (<http://pnwherbaria.org>). Collections were also made for select Canadian localities from a clone bank of wild stand trees at the Cowichan Lake Research Station on Vancouver Island. In total, leaf tissue from 137 *Thuja plicata* individuals (Figure 1; Table S1) and 48 *Tsuga heterophylla* individuals (Figure 1; Table S2) were used to extract DNA using a modified CTAB protocol (Doyle & Doyle, 1987), purified using Sera-Mag SpeedBeads (Thermo Fisher Scientific; Rohland & Reich, 2012), and their DNA quantified using a Qubit 2.0 Fluorometer (Life Technologies).

Three double-digest restriction site-associated DNA sequencing (ddRADseq) libraries (Peterson et al., 2012) were prepared. Two of these were for *Thuja plicata*, assigning half of the samples to one or the other, and one for *Tsuga heterophylla*. For the *Thuja plicata* libraries, the restriction enzymes used were *EcoRI* and *SbfI*, while they were *SbfI* and *MspI* (New England Biolabs) for *Tsuga heterophylla*. All libraries were size selected using a size window of 200–500 bp using a BluePippin (Sage Science). All digestion, ligation and PCR (polymerase chain reaction) products were purified using the Agencourt AMPure XP purification system (Beckman Coulter). For *Thuja plicata*, sequences were generated as 50-bp single-end reads using an Illumina HiSeq 2500 at the Berkeley sequencing facility. For *Tsuga heterophylla*, sequences were generated as 150-bp paired-end reads using an Illumina HiSeq 4000 at The Ohio State University Wexner Medical Center. Raw sequences were processed using IPYRAD (Eaton, 2014; Eaton & Overcast, 2020) with a minimum coverage of

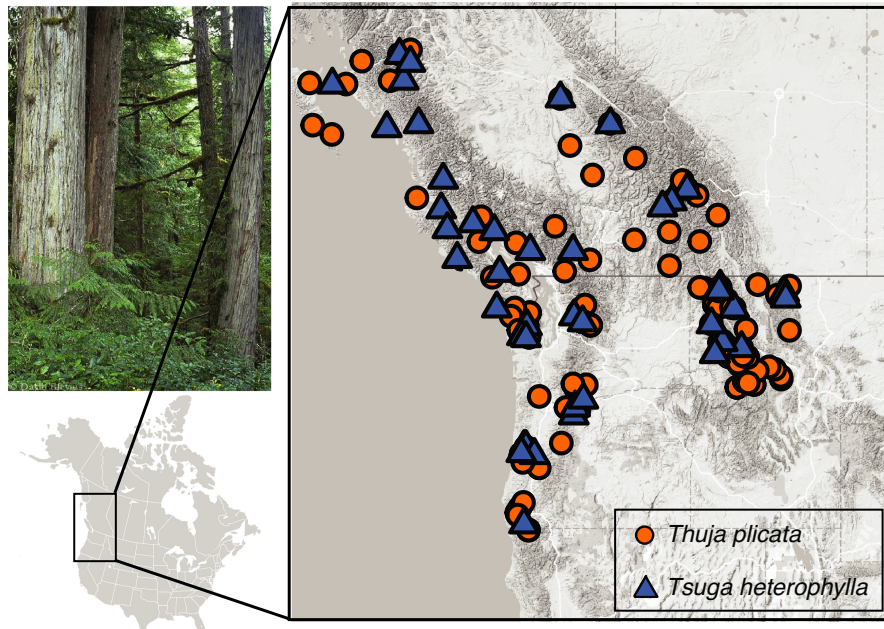


FIGURE 1 Localities sampled for *Thuja plicata* (western red cedar) and *Tsuga heterophylla* (western hemlock). Locality information for each collection can be found in Tables S1 and S2

10 and clustering threshold of 0.80. IPYRAD includes VSEARCH (Rognes et al., 2016) and MUSCLE (Edgar 2010) for sequence clustering. Though we had overlapping reads for *Tsuga heterophylla*, we opted to not merge them and only used single-end reads. Complete assembly procedures were performed and documented in JUPYTER notebooks (<https://github.com/ruffleymr/ThujaTsugaAnalyses>).

## 2.2 | Predictive phylogeography

Prior to analysing genomic data, we made predictions about whether *Thuja plicata* and *Tsuga heterophylla* are expected to harbour cryptic diversity using the random forest (RF) classifier developed for this system by Espindola et al. (2016) and Sullivan et al. (2019). For the predictor variables, we gathered occurrence data previously used for predictive phylogeography of species in the PNW (Espindola et al., 2016; Sullivan et al., 2019) and occurrence data from recently investigated species (Ruffley et al., 2018; Smith et al., 2017, 2018). These occurrence data are a combination of GBIF (Global Biodiversity Information Facility) records and field collections, and we used them to extract bioclimatic variables from WOLRDCLIM version 2 (Fick & Hijmans, 2017). Along with these bioclimatic variables, taxonomic rank and discrete trait variables, such as life stage at dispersal, out-croser or selfer, dispersal mechanism and trophic level (Table S3; Sullivan et al., 2019), were used as the predictor variables in the RF classifier.

The predicted trait (response variable) was harbouring or not harbouring cryptic diversity ("cryptic" vs. "noncryptic"). To date, 12 species complexes with disjunct ranges in the PNW have been investigated in a phylogeographical framework (Avisé et al., 1987): *Ascapus truei/A. montanus* (Metzger et al., 2015; Nielson et al., 2001), *Plethodon idahoensis/P. vandykei* (Carstens et al., 2004), *Prophyaon coeruleum* (Wilke & Duncan, 2004), *Microtus richardsoni*

(Carstens et al., 2005), *Dicamptodon aterrimus* and complex, and *D. tenebrosus* (Steele et al., 2005), *Salix melanopsis* (Brunsfeld et al., 2007; Carstens et al., 2013), *Conaphe armata* (Espindola et al., 2016), *Haplotrema vancouverense* (Smith et al., 2017), *Alnus rubra* (Ruffley et al., 2018), *Prophyaon dubium* and *P. andersoni* (Smith et al. 2018), and *Hemphillia* sp. complex (Rankin et al., 2019). Of these, eight species were classified as noncryptic, according to their respective study, meaning the species does not harbour a deep divergence between populations and has also not experienced significant gene flow between populations (Table S3). The remaining five species/complexes were classified as cryptic because the coastal and inland populations are deeply diverged and, in some cases, have been described as different species.

We constructed four different RF classifiers using different combinations of the predictor variables we had available: (i) bioclimatic variables only, (ii) bioclimatic variables and taxonomy, (iii) bioclimatic variables and life history traits, and (iv) bioclimatic variables, taxonomy and life history traits. In all of these, we are predicting the probability of a species being cryptic. We reported the average out-of-bag error rates for these classifiers. Since each decision tree composing the RF classifier is constructed with reference to only a subset of the training data, out-of-bag error rates can be calculated by considering only the trees constructed without reference to a particular observation, providing a convenient estimate of error rates.

With each of these classifiers, we predicted the presence or absence of cryptic diversity for *Thuja plicata* and *Tsuga heterophylla*, separately. To do this, we gathered occurrence records for the species in question: *Thuja plicata* (791; 569 GBIF records and 222 field collections; Table S4) and *Tsuga heterophylla* (468; 346 GBIF records and 111 field collections; Table S5). We excluded all occurrence records from GBIF that fell outside of the PNW temperate rainforest (35°–65°N, 160°–100°W). We used these locality coordinates



to extract values from 19 bioclimatic variables from WOLRDCLIM version 2 on February 5, 2019 (Fick & Hijmans, 2017) at a resolution of 30 arc-sec (~1 km<sup>2</sup>). We also assembled trait data to coincide with the trait data collected for PNW taxa for predictive phylogeography as in Sullivan et al. (2019). Using these data, we use the four classifiers and followed the procedure of Sullivan et al. (2019) to predict the probability of each species being cryptic. We ultimately aimed to validate these predictions using phylogeographical model testing as described below. After validation was successful, we included the newly classified species data gathered in this study to assess whether the classifier improved in overall accuracy with the addition of two plant species.

### 2.3 | Population structure

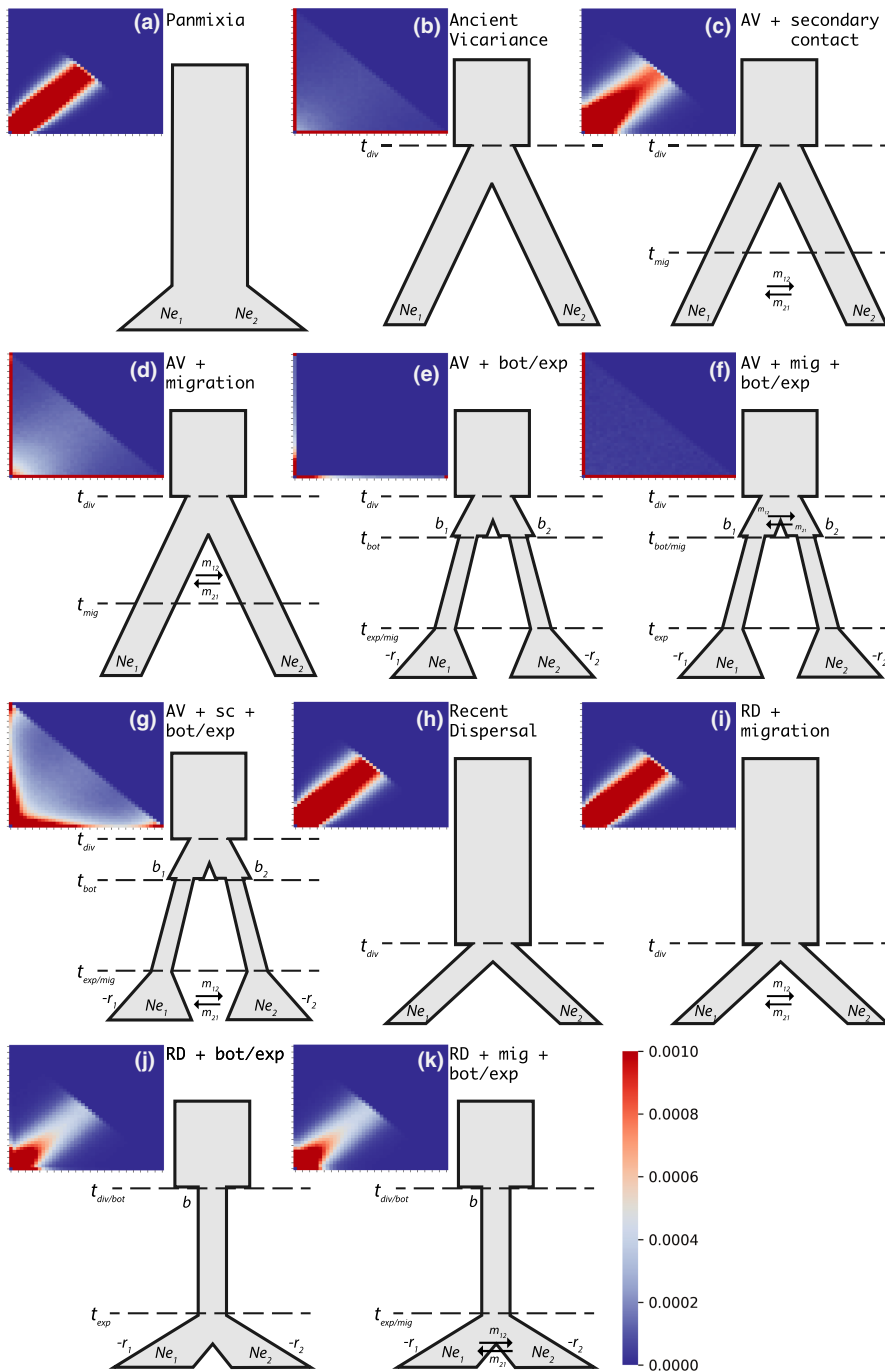
To assess population structure in each species, we explored the genome-wide SNP data using STRUCTURE version 2.3.4 (Pritchard et al., 2000). We ran STRUCTURE for  $K$  values of 1 to 10 with five replicates per  $K$ , where each replicate is a different sample of unlinked SNPs, subsampled from the same complete SNP data set. We ran STRUCTURE for 500,000 generations and discarded the first 10% as burn-in. The data were modelled assuming admixture and correlated allele frequencies between populations (Falush et al., 2003), while all other parameters were kept as their default. For the sake of comparison with results reported by Fernandez et al. (2021), STRUCTURE HARVESTER (Earl & vonHoldt, 2012) was then used to assess  $K$  using the Evanno method (Evanno et al., 2005). We acknowledge that many empirical studies have applied Evanno's  $K$  and interpreted the results as a measure of model fit, but the popularity of a given method should not be mistaken for its appropriateness. Evanno's  $K$  lacks properties that a parameter in evolutionary genetics should ideally possess; for example, it is extremely susceptible to uneven sampling (Puechmaile, 2016) and reproducible inference is challenging (Novembre, 2016). We therefore follow Pritchard et al. (2000) in treating STRUCTURE as a tool for data exploration rather than phylogeographical inference. Our inferences are derived from the results of the model-based analyses described below.

### 2.4 | Joint site frequency spectra

In the remainder of our analyses, we study our data using the jSFS because it summarizes much of the genomic data into one statistic that can be used for inference (Gutenkunst et al., 2009; Xue & Hickerson, 2015). The jSFS is essentially the overlap in the distribution in frequency of alleles between two populations and the pattern of overlap can tell us a lot about demographic processes in the past, both analytically (Excoffier & Foll, 2011; Gutenkunst et al., 2009) and visually (Figure 2). More specifically, a single SFS represents the distribution of the number of sites that are present at each of the  $N$  allele frequencies in the population, where  $N$  is equal to the number

of total chromosomes present in the population. For a diploid organism, this is twice the number of individuals. A jSFS is then the combination of two SFS, one for each population, as a matrix that is  $(N_{\text{pop1}} + 1)$  by  $(N_{\text{pop2}} + 1)$  cells. Each row is one of the allele frequencies in the first population, beginning with 0 and then ranging from  $1/N_{\text{pop1}}$  to  $N_{\text{pop1}}$  and each column is the allele frequencies in the second population, again beginning with 0 and ranging from  $1/N_{\text{pop2}}$  to  $N_{\text{pop2}}$ . Each cell then indicates the number of sites at that corresponding allele frequency in both populations. If the entire jSFS is standardized by the total number of sites, each cell indicates the proportion of sites at the corresponding population allele frequencies. The first row and column correspond to the sites that are at given frequencies in one population while not present at all in the other population, referred to hereafter as the "0" rows and columns. Again, these indicate the variants present in one population and not the other, and thus the cell at row "0" and column "0" indicates the sites that do not vary in either population. With SNP data and for demographic model selection, this cell is not typically considered because it is only relevant for scaling the proportion of invariant sites for parameter estimates. Thus, when estimating demographic parameters from these models, the monomorphic cell along with linked SNPs is needed to inform the composite likelihood of the models (Excoffier et al., 2013).

There is a trade-off between the number of chromosomes that can be included from each population and the number of unlinked SNPs included in the jSFS because the jSFS cannot accommodate missing data. The missing data are generally due to random missing data and allelic dropout from reduced representation sequencing (Andrews et al., 2016), where loci are not represented across all or even a majority of individuals in the population. Thus, the more samples per population are included, the fewer SNPs there are to sample from to construct the jSFS. We therefore down-sampled the number of SNPs and alleles (chromosomes in the population) to construct three different jSFS data sets for each species using custom python scripts ([github.com/isaacovercast/easySFS](https://github.com/isaacovercast/easySFS)). We enforced a different number of alleles to be included per population, which resulted in a different number of unlinked SNPs being sampled in each data set (Table 1). These data sets thus represent a spectrum of genomic information ranging from more individuals in the population but fewer SNPs to fewer individuals represented from the populations with many more SNPs included. We used unlinked SNPs for model selection (see below) to satisfy the assumption that SNPs are independent. We subsampled 100 different observed jSFS for each of the sample sizes for each of the species (600 observed jSFS in total) and masked monomorphic sites in all jSFS. For parameter estimation using the jSFS, we use the full SNP data set, meaning we included linked SNPs in the construction of the jSFS. We also considered the monomorphic cell in the jSFS when estimating parameters because this cell provides information important to scale the invariant sites in the genome. To calculate the monomorphic cell, we measured the ratio of monomorphic sites and polymorphic sites in our entire data sets for each species and then used those ratios, multiplied by the total number of biallelic SNPs used in the empirical jSFS.



**FIGURE 2** Summarized folded jSFS ( $43 \times 43$  cells) for 10,000 simulations under each associated demographic model, A to K. Scale indicates the proportion of loci in each cell, with 0.001 being the maximum, meaning if the proportion is higher than this value, the colour is the same as the maximum. Dashed lines represent the times of all events that can occur in a given model, including divergence (*div*), bottleneck (*bot*), expansion (*exp*) and migration initiation (*mig*) events. Migration arrows indicate asymmetrical migration between populations, *b* is the magnitude of a bottleneck and *r* is the population growth rate during expansion for a given population

## 2.5 | Demographic modelling

One of the most important recent advances in phylogeography is the development of model selection as a framework for inference of evolutionary processes (e.g., Carstens et al., 2013). After exploring our data using *STRUCTURE*, we constructed 11 demographic scenarios (Figure 2) to test using a machine-learning model-selection framework (Smith & Carstens, 2020). For each focal species, these alternative demographic hypotheses include divergence between the coastal and inland populations that occurred either before or after the Pleistocene glaciations. The pre-Pleistocene divergence scenarios (models B–G in Figure 2) model the populations diverging at the

time of the xerification of the Columbia Basin (Waring & Franklin, 1979), which followed the uplift in the Cascades Mountains (Priest, 1990). In the recent dispersal models (models H–K in Figure 2), post-Pleistocene divergence between the populations posits the ITR populations arising from the coastal populations only via dispersal of coastal migrants; these models imply a recent time of divergence, as the coastal migrants could only have recolonized the inland region after the last glacial retreat, ~10 ka (Waite & Thorson, 1983). The varying migration scenarios include divergence with migration, where migration eventually ends between the coast and ITR populations a substantial time after divergence. In divergence with secondary contact models, migration begins again between

**TABLE 1** The average number of unlinked SNPs used in the 100 empirical data sets, with the corresponding number of samples from the coastal and inland populations, where each sample represents an allele for an individual; most often both alleles are included, but in some cases only one allele from an individual is included in the construction of the jSFS. The error rate for all models represents the average error rate for all model classifications for the classifier constructed with the corresponding data size. The error rate with RD collapsed corresponds to the overall error rate for the classifier when the four Recent Dispersal models are collapsed into a single model, RD

Predictor variables	Error rate (%)	<i>Thuja plicata</i>		<i>Tsuga heterophylla</i>		Updated error rate (%)
		Noncryptic PP	Cryptic PP	Noncryptic PP	Cryptic PP	
Bioclim	0.205	0.713	0.287	0.669	0.331	0.155
Bioclim, Taxonomy	0	1	0	1	0	0
Bioclim, Traits	0	1	0	1	0	0
Bioclim, Taxonomy, Traits	0	1	0	1	0	0

the coast and ITR populations, at the very earliest, after the retreat of the Cordilleran ice sheet, ~10 ka (Waitt & Thorson, 1983). Thus, these models of ancient vicariance with secondary contact (models C and G) encompass both the older “ancient vicariance” and “recent dispersal” scenarios of Brunfeldt et al. (2001). The bottleneck events that are modelled are those that hypothetically occurred in the populations at the onset and for the duration of the Pleistocene and the subsequent population expansion events occur after the retreat of the glaciers, probably as recently as 3.5 ka (Mehringer, 1996; Whitlock, 1992).

Before using our data to assess these models, we first assessed our statistical power to make these inferences using simulations. Specifically, we simulated genomic data similar to the empirical genomic data we generated for *Thuja plicata* and *Tsuga heterophylla* under each of the 11 models. Specifically, we used the R package delimitR (Smith & Carstens, 2020), which relies on the multidimensional SFS (mSFS) and the machine learning algorithm abc-randomForests (Pudlo et al., 2015) for model selection. For this, we simulated jSFS under 11 demographic scenarios we deem plausible for both species (Figure 2) using FASTSIMCOAL2 (Excoffier & Foll 2011; Excoffier et al., 2013). We generated 10,000 simulated jSFS for each model. We then converted our jSFS into mSFS by flattening the matrix and binning the cells into a coarser representation of itself. In delimitR (Smith & Carstens, 2020), we then used the mSFS of the simulated data sets as the predictor variables to train an RF (Breiman, 2001; Liaw & Weirner, 2002; Pudlo et al., 2016) classifier to distinguish among the 11 demographic models. This allowed us to assess the limits of the inference we could make with respect to phylogeographical model selection and construct a confusion matrix to summarize this differentiability among models. It also generated the classifier used in model selection for our empirical data sets.

Following the constructing of a demographic model RF classifier, we assessed the demographic models (Figure 2) for the empirical data sets for each of *Thuja plicata* and *Tsuga heterophylla*. These classifiers simultaneously self-cross-validate by testing the accuracy of the decision trees being constructed using out-of-bag error rates. Again, each decision tree in the classifier is constructed from a random subset of the training data, and thus out-of-bag error rates can be calculated by considering only the trees constructed without reference to a particular observation. This results in overall error rates

for the classifier, as well as specific model misclassification rates. This is an error rate specific to the classifier and represents how often a model class is incorrectly identified, as well as the incorrect model choice.

We constructed six different classifiers to mimic the six empirical jSFS, with differing coastal and inland sample sizes and unlinked SNPs (Table 1). We then used the appropriate classifier to make predictions for the 100 corresponding subsampled jSFS. We summarized the support for each model and each data set by the number of votes for each model. We then estimated the posterior probability for the best model.

## 2.6 | Parameter estimation

Once the best model for each species was identified, we used FASTSIMCOAL2 to estimate its demographic parameters and their 95% confidence intervals (CI) using the full, linked SNP data sets for each species and the monomorphic cell of the jSFS. We also estimated an additional parameter not included in the prior modelling: the mutation rate ( $\mu$ ) in substitutions per site per million years. FASTSIMCOAL2 uses a modified expectation-maximization algorithm, known as a conditional expectation maximization (CEM; Brent, 1974; Meng & Rubin, 1993) for maximum-likelihood (ML) optimization, which can get trapped in local optima of the likelihood surface. Therefore, we performed 100 independent parameter optimizations with different initial values, 100,000 simulations to estimate the expected folded jSFS and 40 conditional CEM cycles per optimization. Following the first optimization, we identified the global ML and parameter estimates and performed an additional 100 independent optimizations using these ML parameter estimates as the starting values.

To estimate confidence intervals, we simulated 100 parametric bootstrap replicates using the ML parameter estimates from the final optimizations of the empirical data sets. We then re-optimized parameters of the simulated data sets, initiating the parameters at the ML estimates from the original optimization. We used these parameter estimates to generate 95% high-density CIs for all parameters (Kruschke, 2011).

All computational analyses were done using servers at the IBEST Computational Resources Core at the University of Idaho.

### 3 | RESULTS

#### 3.1 | Sequencing and jSFS

Following assembly of the ddRADseq data, we recovered 124,484 loci (214,183 SNPs) for *Thuja plicata* and 142,804 loci (893,487 SNPs) for *Tsuga heterophylla*, all of which were shared across a minimum of four individuals per species (Ruffley et al., 2022). To construct the jSFS for these species, the data were downsampled such that each SNP was represented in all individuals included in the jSFS (Table 1). In using the jSFS to make our inference about demographic histories, we excluded a considerable amount of sequence data, although the models are nevertheless distinguishable (Table 1).

#### 3.2 | Predictive phylogeography

Before assessing whether these species harbour cryptic diversity, we made phylogeographical predictions of cryptic and noncryptic for both species following the procedure introduced by Espíndola et al. (2016) using RF with bioclimatic variables associated with sample localities and taxonomic ranks. Following Sullivan et al. (2019), we also included trait values along with the bioclimatic and taxonomic variables. The error rates we recovered were congruent with those found by Sullivan et al. (2019) and thus these classifiers were used to make predictions about *Thuja plicata* and *Tsuga heterophylla*. Each classifier predicted neither species to harbour cryptic diversity (Table 2), with the only variation in the prediction being the less accurate classifier (bioclimatic data only).

#### 3.3 | Population structure

We explored the population structure for both species using STRUCTURE (Pritchard et al., 2000) for possible  $K$  of 1 to 10. For *Thuja plicata*, we inferred three clusters (Figure 3; Figure S1). While two of the clusters are geographically restricted to the coast or inland, the third has no clear geographical structure or primary location. At

$K = 2$  (Figure 3), we obtained the geographically structured pattern observed in the first two clusters of  $K = 3$  (although some coastal samples were sometimes present in the inland cluster).

For *Tsuga heterophylla*, we inferred  $K = 2$  as the optimal  $K$  value (Figure S2). We do not see a geographical association between the coastal and inland samples with the two clusters (Figure 3). However, when we investigate  $K = 3$ , we do observe a strong geographical structure, with one cluster mostly restricted to the inland forests and the other present along the coast. Because the justification of delta  $K$  is based on simulations with no hierarchical population structure (which is almost always present in nature), interpretation of multiple clustering scenarios is critical, regardless of the optimal  $K$  inferred (Janes et al., 2017).

#### 3.4 | Demographic modelling

The STRUCTURE results provided the guidance for deciding how many populations to model in the demographic models investigated. Since the primary goal of this study is to make inferences regarding the evolutionary history of the ecosystem dominants, we decided to model two populations on the basis of geography (i.e., samples from either the inland or coastal forests), which largely corresponds to the division between the two coastal and one inland population in the  $K = 3$  STRUCTURE analysis. We combine the clusters that contain coastal samples because these are not structured in a geographically meaningful manner (Figure 3), so organizing them by cluster would combine genetically similar populations that are not necessarily geographically close.

We developed 11 demographic models (Figure 2) to assess the phylogeographical history of each species and the origin of the genetic clusters. In these models, we considered both ancient and recent divergence events, and various migration scenarios, including divergence with and without migration and secondary contact. We also model possible bottleneck events associated with the onset of the Pleistocene (~2.5 Ma). We modelled population expansion to be associated with population regrowth after glacial retreat ~10 ka (Figure 2). We used FASTSIMCOAL2 to simulate DNA sequence data,

	SNPs	Coastal	Inland	Error rate (% all models)	Error rate (% RD collapsed)
<i>Thuja plicata</i>					
1	9,036	11	10	27.90	12.93
2	2,484	26	24	30.89	15.83
3	1,041	42	42	33.87	19.23
<i>Tsuga heterophylla</i>					
1	7,929	13	8	27.10	12.26
2	2,698	19	12	31.21	15.94
3	1,195	27	16	34.78	19.88

TABLE 2 Phylogeographical predictions of cryptic and noncryptic nature for *Thuja plicata* and *Tsuga heterophylla* using random forest (RF) with specified predictor variables

Note: Error rate indicates the error rate of the RF classifier used to make the predictions. PP: prediction probability. The updated error rate is the error rate of the new classifier constructed with the new data from *Thuja plicata* and *Tsuga heterophylla*.



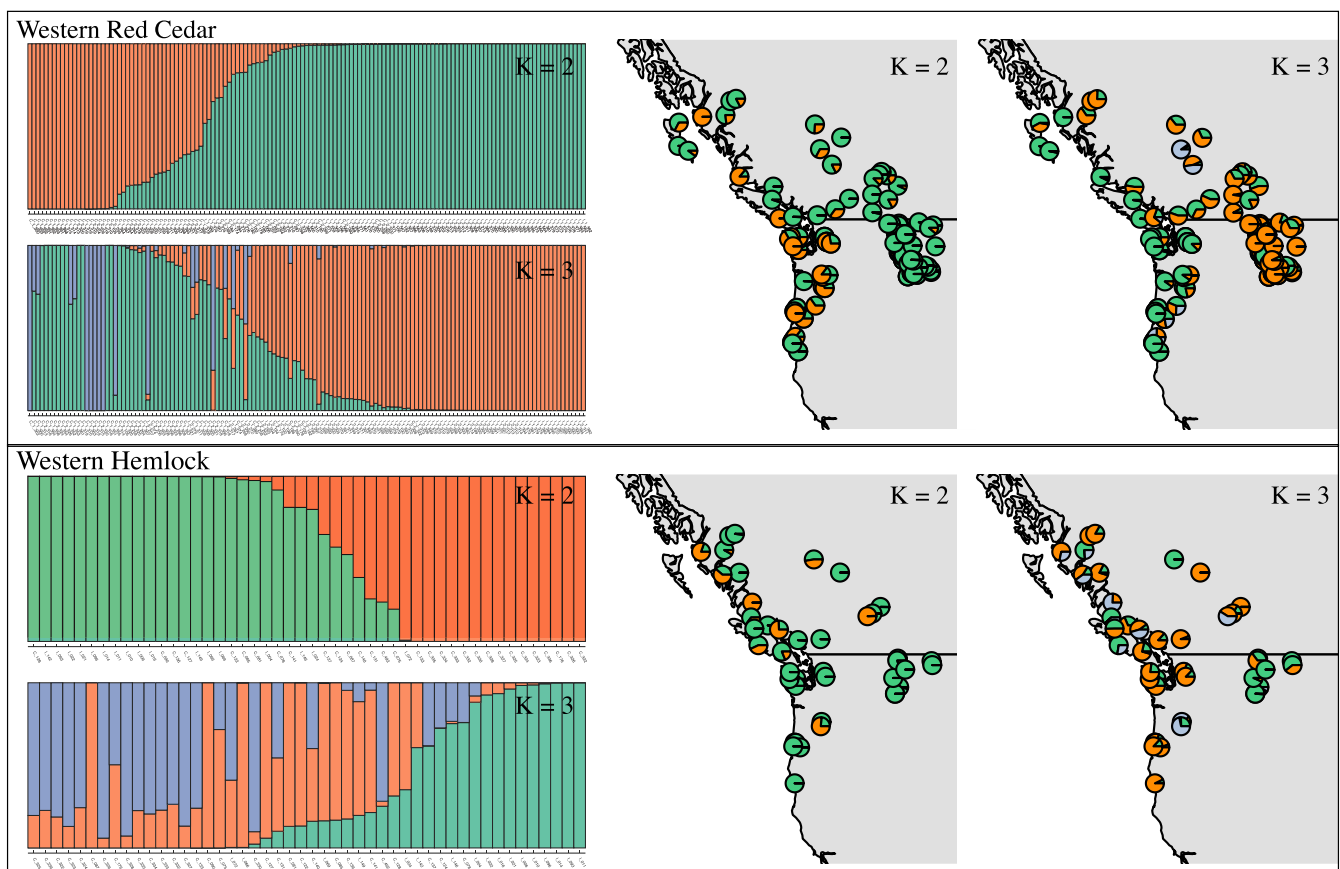
setting the number of loci and variable sites to match the empirical data, and then summarize that data using jSFS. We simulated 100 data sets for three different data set sizes and for each species. No missing data can be included in the calculation of the jSFS, so for a particular locus to be included, it must be present in all individuals. Thus, there is a trade-off between numbers of individuals and SNPs considered in the analysis.

The analyses of model identifiability resulted in low error rates for classifying each of the 11 models (Table 1; Figures S3 and S4). Most models were classified correctly most of the time, with all of them having a classification accuracy above 0.72 (Figures S3 and S4), except for the models with a recent divergence event between coastal and inland populations. We therefore collapsed these models, which all varied in the presence/absence of migration and bottleneck and expansion events, into a single recent dispersal model (Figure 4; Figure S5). This decreased the overall error rate dramatically (Table 1); the identifiability of the recent dispersal class of models increased to 0.90.

The first classifier, with all 11 models, was used to make predictions using the observed jSFS for each species (Figure 5). For each data set size, we used 100 different jSFS that were constructed by

subsampling unlinked SNPs randomly. For *Thuja plicata*, all data sets had the highest prediction probability for the same model: ancient divergence between coastal and inland populations, followed by a bottleneck in both populations, and then population expansion contemporaneous with secondary contact between populations due to migration ("AV + sc + bot/exp," Figure 5; model G in Figure 2). On average, each *Thuja plicata* data set received 552/1000 votes for that model and had an average posterior probability of 0.72 (Figure 5). Therefore, some data sets, or certain subsampled jSFS, had higher and/or lower prediction probabilities for this model.

The results were different for *Tsuga heterophylla* in that each data set did not have the highest prediction probability for the same model. Those with most SNPs supported models similar to *Thuja plicata* ("AV + sc + bot/exp," model G in Figure 2). On average, this model received 564/1000 votes in the classifier for each observed jSFS that supported this model (Figure 5), and had an average estimated posterior probability of 0.83 (Figure 5). With fewer SNPs included in the jSFS, but more samples represented in the population, the model that had the highest prediction probability was model C, or ancient divergence between coastal and inland populations, followed by secondary contact between populations



**FIGURE 3** Left panels: STRUCTURE results for *Thuja plicata* (western redcedar) and *Tsuga heterophylla* (western hemlock), where each bar indicates an individual in the population and the colour indicates the proportion of genetic variation associated with a particular cluster. Clusters are indicated by  $K$  values in the top right corner. Coastal samples are denoted with a C in the label and inland samples with an I. Right panels: sampling localities plotted according to the proportion of genomic variation attributed to each cluster, with clusters at  $K = 2$  and  $K = 3$

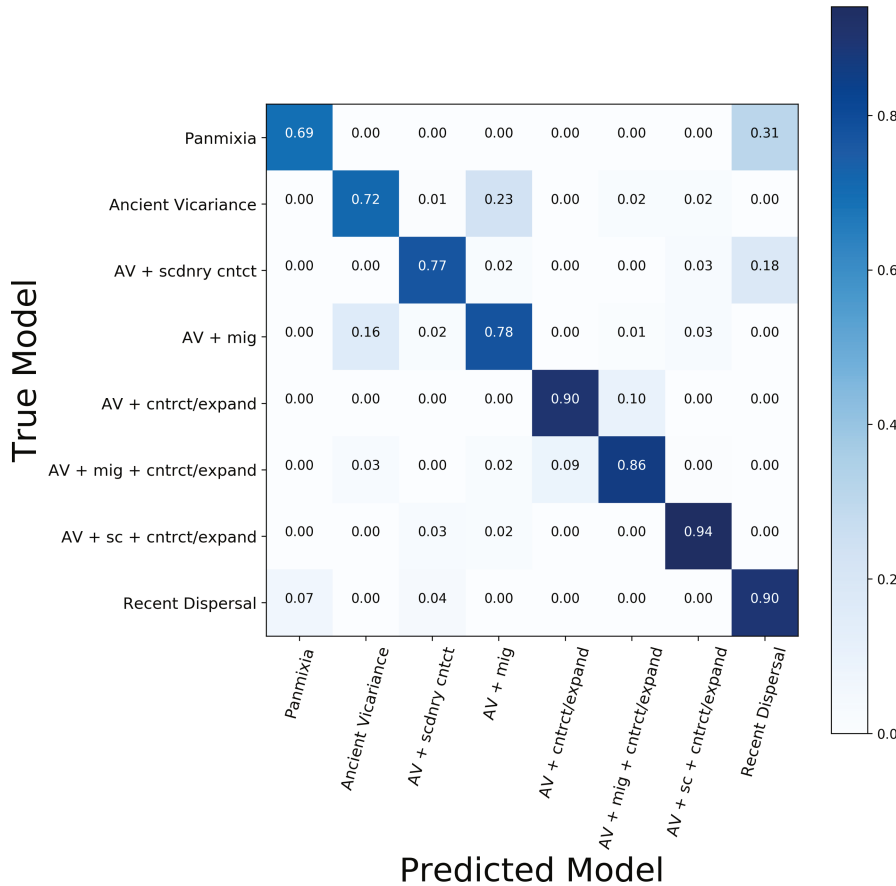


FIGURE 4 Confusion matrix depicting the prediction accuracies and inaccuracies for eight demographic models using delimitR for model selection, which involves the simulated mSFS and “abcrf.” Rows indicate the model the data were simulated under and columns indicate the model that was predicted; each cell then indicates the proportion of simulated data under the true model that is classified as the predicted model. Thus, the diagonal cells of the matrix depict the proportion of correct model classifications, as the predicted model aligns with the true model, whereas the off-diagonal cells depict the proportion of model simulations that are incorrectly classified, and specifically which model the simulations are incorrectly classified as

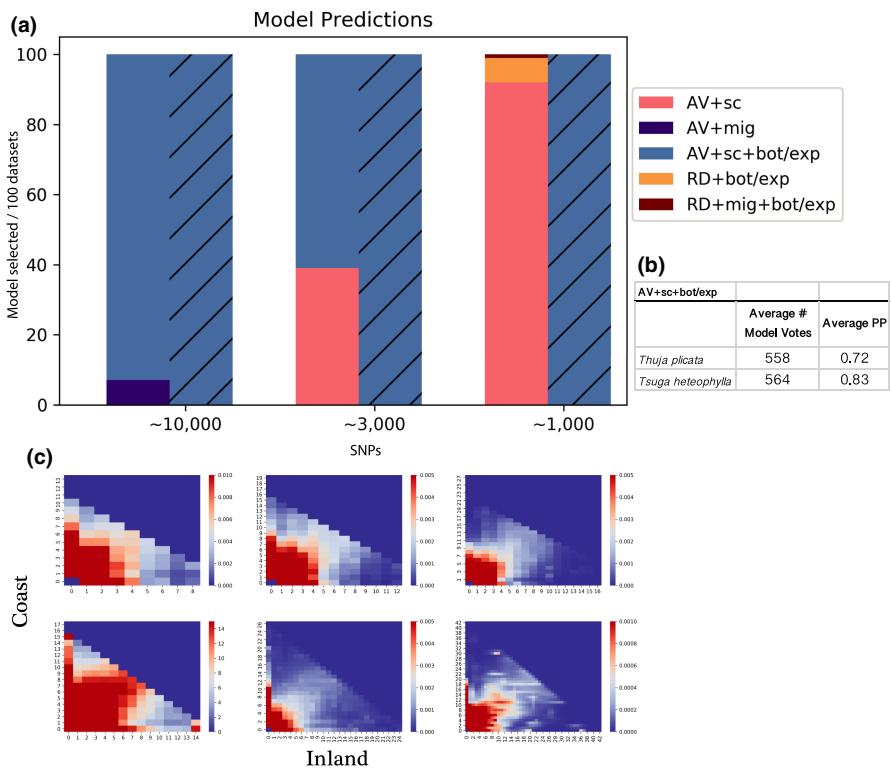


FIGURE 5 (a) Barplots representing the proportion of observed jSFS, at the corresponding average number of SNPs in the jSFS, that are classified as a given model, which is indicated by the colour of the barplot. Solid barplots (left side) represent *Tsuga heterophylla* predictions and diagonal striped barplots (right side) represent predictions for *Thuja plicata*. (b) Table indicating the average number of model votes for the most selected model, “AV + sc + bot/exp,” for both species, along with the average estimated posterior probability (PP) for the same model. (c) Corresponding observed jSFS at each SNP count (10,000, 3000 and 1000) for *Tsuga heterophylla* (top row) and *Thuja plicata* (bottom row)

due to migration (“AV + sc,” Figure 2), which is very similar to model G. The difference is that model G includes the population bottleneck and expansion that are not modelled in the former model C

(“AV + sc,” Figure 2). On average, this model received 532/1,000 votes in the classifier for each observed jSFS that supported this model, and had an average estimated posterior probability of 0.78.

TABLE 3 Maximum-likelihood (ML) parameter estimates for *Thuja plicata* and *Tsuga heterophylla* for the model selected more often for the data, "AV + sc + bot/exp"

	<i>Thuja plicata</i>			<i>Tsuga heterophylla</i>		
	ML estimate	Minimum 95% CI	Maximum 95% CI	ML estimate	Minimum 95% CI	Maximum 95% CI
N inland	1,317,431	1,118,973	1,611,659	1,574,048	1,165,534	1,938,174
N coast	1,514,468	1,301,011	1,673,645	3,361,225	2,865,576	3,708,440
$T_{div}$	252,814	231,341	295,009	252,285	223,890	295,967
$T_{bot}$	53,436	50,545	62,430	59,417	51,439	92,440
$T_{exp}$	1043	1010	1095	2021	1592	2290
BtnMag inland	0.5930	0.5391	0.6706	0.4792	0.4403	0.4992
BtnMag coast	0.4791	0.4336	0.4982	0.5888	0.5230	0.7201
Gro inland	-1.7E-04	-3.1E-04	1.9E-05	-4.4E-04	-5.0E-04	-3.9E-04
Gro2 coast	-6.8E-04	-8.0E-04	-5.9E-04	-1.7E-04	-2.1E-04	-1.1E-04
MIG inland → coast	1.4E-05	1.0E-05	1.9E-05	1.2E-05	6.2E-06	2.0E-05
MIG coast → inland	1.1E-04	9.5E-05	1.2E-04	4.5E-05	3.4E-05	5.7E-05
Mutation rate	5.7E-09	5.0E-09	6.1E-09	4.6E-09	4.2E-09	5.1E-09

Note: The population sizes,  $N$  inland and  $N$  coast, are in units of the number of alleles in the population. All of the events,  $T_{div}$ ,  $T_{bot}$  and  $T_{exp}$ , are in units of coalescent generations. The magnitude of the bottleneck, btnmag, indicates the instantaneous shrinkage of the population by that proportion. The growth rates indicate population size change, backward in time, as the number of alleles removed from the population per generation. Thus, a negative rate indicates population expansion forward in time. The migration rates indicate the proportion of alleles moving to the other population per generation. The mutation rate is in substitutions per site per generation.

We suspect this lower model support could be due to the fact that, in comparison to *Thuja plicata*, there were fewer SNPs to inform parameters associated with the additional process—population bottleneck and expansion—as well as fewer individuals sampled from each population.

### 3.5 | Parameter estimation

The parameter estimates obtained in our analyses were in units of coalescent generations and generally fit with most of our expectations based on the history of the bioregion. For both species, the population sizes estimated for the coastal population are slightly larger than those of the ITR (Table 3), consistent with their current distributions. For both species, the median divergence time estimates,  $T_{div}$ , between the coastal and inland rainforests were ~252,000 generations ago (Table 3). The timing of the population bottleneck event,  $T_{bot}$ , for both species was estimated to be between 50 and 90 kya (Table 3). The time of the population expansion,  $T_{exp}$ , for *Thuja plicata* was ~1050 generations ago, whereas  $T_{exp}$  for *Tsuga heterophylla* was nearly twice that value (2020 generations ago). The magnitude of the *Thuja plicata* coastal bottleneck was apparently slightly larger than that of the inland bottleneck. The opposite was true for *Tsuga heterophylla*, where the bottleneck in the inland forests was more severe than that on the coast (Table 3). Not surprisingly, the populations with the most severe bottleneck also had the largest population expansion rates (Table 3). Note that this expansion rate was modelled backwards in time; a negative rate indicates the population is getting smaller towards the past, thus

expanding forward in time. In both species, migration rates from the coast to the ITR were larger than migration rates from the ITR to the coast (Table 3).

To interpret our time estimates, which in coalescent analyses are expressed in units of generations, we need to consider the generation length of each species. The generation length is essentially the average amount of time between consecutive generations in a population. For *Thuja plicata*, estimates of trees reaching maturity typically range from 20 to 30 years (Turner, 1985), but trees can reach maturity as early as 10 years in some open grown areas (Minore, 1990). The same is true for *Tsuga heterophylla*, where most estimates suggest maturity is reached between 25 and 30 years (Owens & Molder 1984) but with trees reaching maturity earlier in some cases (Tesky 1992). To be conservative, we assumed a relatively short generation length of 10 years per generation and a longer generation length of 25 years per generation for both species. In doing this, we can convert our estimates of the timing of the events from generation into years while accounting for uncertainty in generation length amongst the populations (Table 4). From these generation lengths, we calculate median divergence times between inland and coastal populations of 6.3–2.5 Ma, where each estimate has an associated 95% CI (Table 4). This indicates the divergence between inland and coastal populations for both species probably began before the onset of the Pleistocene, continued into the Pliocene–Pleistocene boundary, and possibly even into the early onset of the Pleistocene, as the Pleistocene is recorded to have begun 2.588 Ma (Gibbard et al. 2010). Similarly, given the overlap in 95% CI for the estimates for both species, we cannot reject the hypothesis that these species' disjunct populations diverged in concordance.

TABLE 4 Divergence time estimates and time of population expansion and secondary contact estimates for both species

	ML estimate	Minimum 95% CI	Maximum 95% CI
<i>Thuja plicata</i>			
$T_{div}$	2,528,140	2,313,410	2,950,090
$T_{exp}$	10,430	10,100	10,950
<i>Tsuga heterophylla</i>			
$T_{div}$	2,522,850	2,238,900	2,959,670
$T_{exp}$	20,210	15,920	22,900

Note: Estimates are in years, calculated by multiplying the divergence time in generations by an estimated generation length of 10 years and 25 years for both species.

## 4 | DISCUSSION

### 4.1 | History of the ITR

The demographic modelling results suggest that the ITR represents expansion from pre-Pleistocene relictual ITR populations and that this forest periodically received migrants from coastal populations, presumably via wind dispersal. Genomic evidence from both *Thuja plicata* and *Tsuga heterophylla* supports this ancient divergence between the ITR and the coastal rainforest, with the evidence apparent in the statistical model selection as well as the observed allele frequencies. The genomic signature of refugia (an anciently diverged population) is an abundance of rare alleles not shared with other populations. O'Connell et al. (2008), while they acknowledge some genetic differentiation between interior and coastal populations, suggested the divergence was shallow enough to support recent divergence with an absence of subsequent migration (e.g., model H, Figure 2). However, none of the recent dispersal models were supported by our data (Figure 5). Moreover, in a recent genomic study of *Thuja plicata* (and mountain hemlock, *Tsuga mertensiana*, a sub-alpine congener of *Tsuga heterophylla* examined here), Fernandez et al. (2021) found that ITR populations of *Thuja plicata* are a mix of two genomic clusters, one restricted to the southern portion of the ITR and the other more prominent in the northern portion of the ITR and shared with the central Cascades. Thus, their genomic data are the first to suggest persistence of this late-successional dominant tree species in the inland region throughout the Pleistocene. Here, we have collected thousands of loci across individuals from both ITR and coastal populations for *Thuja plicata* (see also Fernandez et al., 2021), and the other late-successional dominant, *Tsuga heterophylla*. With these data, we have been able to examine the jSFS and observe the high frequency of rare alleles present in the coastal and ITR populations separately, which indicates their ancient divergence. More importantly, we have modelled coalescent processes to account for coalescent stochasticity and explicitly conducted statistical tests of multiple plausible evolutionary (phylogeographical) models. Our results have ultimately allowed us to infer the presence of ITR refugia for both species throughout the Pleistocene,

indicating a partial confirmation of the AV hypothesis. We were able to date the divergence times between inland and coastal populations for both species to just prior to or at the onset of the Pleistocene, assess demographic histories, and estimate migration patterns between coastal and inland populations of each of the two species. These inferences illustrate the analytical power of explicit statistical phylogeographical modelling. They also reiterate that there can be both evidence of ancient vicariance and recent dispersal patterns, although our approach explicitly questions whether inland populations that survived glaciation in the ITR are extant today.

These results have implications for how the fossil pollen record informs our understanding of the history of the PNW (Whitlock, 1992). Given the difficulty of identifying cedar pollen (Faegri & Iversen, 1992), the pollen classification of *Thuja plicata* is generally limited to being a member of the family Cupressaceae, while the identification of *Tsuga heterophylla* pollen is more reliable. Of this *Thuja-Tsuga* pollen record, the ITR is inferred to have been present at the southern and central ranges of its current distribution <6.3–4 ka (Chase et al., 2008; Herring & Gavin, 2015; Mehrlinger, 1996; Rosenberg et al., 2003), and not detected in high levels at the northern range extent until roughly 3–2 ka (Gavin et al., 2009). Before this, the *Thuja-Tsuga* pollen record is not recognized in the area prior to 100 ka, suggesting a rather recent population expansion of both species throughout the range. However, we know pollen records are subject to taphonomic biases, and just because pollen is not detected does not mean it was not in the vicinity or habitat-type. Given the genomic evidence showing an abundance of rare alleles in the ITR populations and the coalescent analysis, we ultimately infer that populations of *Thuja plicata* and *Tsuga heterophylla* must have been in the ITR during the Pleistocene earlier than 6 ka, though perhaps in small or isolated populations that limited detection.

### 4.2 | Analytical considerations

The limitations to the use of the jSFS to summarize genomic data should be recognized. As described above, when we summarized our data into a single jSFS, we downsampled data so that every SNP was included in each individual in the jSFS. We note that doing this required us to forfeit a considerable amount of data (Table 2). We performed a sensitivity analysis on the use of the jSFS by constructing three different data set sizes, of 100 jSFS each, for each species, *Thuja plicata* and *Tsuga heterophylla* (Table 2), which resulted in 100 model predictions per species, per data set (Figure 5). The largest discrepancy in the entire inference is within the *Tsuga heterophylla* prediction, where a different demographic history is supported by the jSFS that included more individuals and fewer SNPs, and the jSFS with the most SNPs and fewest individuals (Figure 5). While the two models supported are generally consistent with our overall inference of pre-Pleistocene divergence followed by secondary contact, they differ in the presence of a population bottleneck during the Pleistocene and subsequent population expansion after the last glacial retreat. The difference in the model support for *Tsuga*

*heterophylla* across downsampling regimes could be due to the data set with more SNPs being able to estimate the bottleneck and expansion parameters more effectively, and therefore showing strong support for that model. Conversely, the information in the data set with fewer SNPs may have just been insufficient to estimate those parameters. For the purposes of our study, more data are not necessary, but for future and more precise demographic parameter inference, this may be the case.

Our model-selection procedure supports, for both *Thuja plicata* and *Tsuga heterophylla*, a pre-Pleistocene divergence event, followed by secondary gene flow between the populations. The approach used here for model selection using RF and the jSFS (Smith & Carstens, 2020; Smith et al., 2017) had yet to be tested using plant species or demographic models of this complexity. This is a likelihood-free approach that is based on simulating allelic data while accounting for coalescent stochasticity and demographic processes. Model selection, both in general and when implemented with machine-learning as employed here, is as accurate as the data are distinct in model space. This means that we should be able to assess if the empirical data are insufficient to distinguish among these models, which is indicated by the classifier's error rates. Indeed, our simulations indicated that the genomic signatures of the class of four recent dispersal models (Models H–K, Figure 2) are not differentiable from each other (Figure S4). This is probably due to the recent divergence time between the coast and inland populations resulting in low resolution of the distinct migration patterns under which those models were simulated. However, all hypotheses positing post-Pleistocene dispersal, regardless of migration pattern, are well differentiated from those positing a pre-Pleistocene dispersal, or specifically the persistence of disjunct coastal and inland populations through the Pleistocene (Figure S4). Additionally, when we pool the recent dispersal models into a general recent dispersal model, our data show that the error rates in all of our classifiers are extremely low, indicating high confidence in our classifier and high information content in data with respect to distinguishing among the final eight demographic scenarios we propose (Figure 4). Again, the power of the machine learning classifier depends on how distinct the data are in model space and models can be very simple or very complex, all of which influences the power of the classifier.

The approach used here provides flexibility to the demographic model designs and simulation of data, as well as computational efficiency. As is true for all inferences based on model selection, it remains possible that some as yet unexamined model may be a better description of the true evolutionary history of these taxa, perhaps specifically those that model more than two populations and therefore more complex divergence and migration scenarios. Nevertheless, the approach to inference that we have adopted here (i.e., developing models that are derived from extrinsic information such as pollen records and climate data, collecting genomic data, ranking models, assessing their identifiability and making inferences) is an extremely powerful framework for phylogeographical research. In contrast to the approach that bases inference on methods designed for data exploration, our approach to inference utilizes

existing data to formulate hypotheses that can then be supported (or not) as new data are collected and analysed.

### 4.3 | Predictive comparative phylogeography of PNW rainforest taxa

Comparative phylogeographical studies have addressed disjunct mesic forest taxa in the PNW of North America for decades. These have largely focused on the rather coarse phylogeographical hypotheses synthesized by Brunsfeld et al. (2001). Several taxa, primarily amphibians (*Ascaphus truei*/*A. montanus*, Nielson et al., 2001; *Plethodon vandykei*/*P. idahoensis*, Carstens et al., 2005; *Dicamptodon copei*/*D. aterrimus*, Carstens et al., 2005), show evidence of an ancient vicariance and persistence of populations in inland refugia throughout the Pleistocene, with no evidence of gene flow. Red alder (*Alnus rubra*) genomics (Ruffley et al., 2018) show evidence of ancient vicariance but with consistent migration between coastal and inland populations through the Pleistocene. Still other disjunct rainforest endemics, including water voles (*Microtus richardsoni*, Carstens et al., 2005) and several disjunct tailedropper slugs (*Prophysaon andersoni*, *P. dubium*, *P. coeruleum* and *P. vanatta*/*P. humile*; Smith et al., 2018), show no evidence of Pleistocene persistence in ITR refugia, but rather have attained the disjunct distribution via post-Pleistocene dispersal. Likewise, other more broadly distributed tree species, not specific to the mesic forest but persistent in the PNW ecoregion, have also shown similar patterns of post-Pleistocene dispersal (Bagley et al., 2020; Callahan et al., 2013), albeit that the populations have connected ranges across the disjunction.

Because of this collection of complex evolutionary histories for mesic forest disjunct taxa, and the conservation and evolutionary implications, Espindola et al. (2016) and Sullivan et al. (2019) have developed a framework for predicting the presence or absence of cryptic, or ancient, divergence in this system. Our data on the two mesic forest climax community dominant tree species represent a critical refinement to this predictive framework. This is especially true because *Tsuga heterophylla* and *Thuja plicata* show strong evidence for a pre-Pleistocene divergence as well as evidence of post-Pleistocene gene flow through the nonzero estimation of migration rates between the populations (Table 3), a pattern thus far seen in only one other disjunct taxon (*Alnus rubra*). This should increase our ability to identify other inland rainforest taxa that may demonstrate a similar evolutionary history, especially plants with wind-dispersed pollen and/or seeds/spores.

In addition to *Alnus rubra* and *Salix melanopsis*, there are many other nonecosystem dominant plant species that exhibit the current mesic forest disjunct distribution (Brunsfeld et al., 2001). These species vary in life histories from the plants that have been studied, which are all long-lived tree species, and include annuals such as *Lysimachia latifolia* (Pacific starflower), and perennials such as a fern *Polystichum munitum* (western sword fern), an herbaceous flowering plant *Tellima grandiflora* (bigflower tellima), and a sedge *Carex hendersonii* (Henderson's sedge). It remains unclear whether these species



are a result of recent dispersal or persisted in inland throughout glaciation, but we speculate that the ancient ITR was probably more diverse than just *Tsuga heterophylla*, *Thuja plicata* and *Alnus rubra*. These additional disjunct plant species are prime targets for future studies to characterize plant phylogeography in the PNW and how biodiverse the ancient ITR was.

## 5 | CONCLUSION

Using genomic evidence and modern demographic inference procedures with machine learning, we have shown evidence for the persistence of ITR populations of both ecosystem dominant tree species *Thuja plicata* (western redcedar) and *Tsuga heterophylla* (western hemlock) throughout the Pleistocene. This is critical because both other mesic forest disjuncts and ITR endemics (e.g., endemic jumping slugs such as *Hemphilia camelus*, *H. skadei* and *H. danielsi*; Rankin et al., 2019) are rainforest-dependent. Our results, as well as the recent results for *Thuja plicata* of Fernandez et al. (2021), indicate that these late-successional dominant tree species persisted throughout the Pleistocene, providing habitat for rainforest dependents, including numerous understorey plant species (Björk, 2010) that form the ecosystem. Further, the refugial populations in the ITR were probably small, as we show support here for Pleistocene-related population bottleneck events in both species. This evidence does coincide with the palaeontological record, which suggests that the temperate rainforest did not dominate the PNW landscape until the last 5,000 years, and only in the last 3,500 years did the expansion of ITR begin. Coupled with the recent population expansion, we also show evidence for secondary contact at this time between the coastal and ITR populations for both species. This recent gene flow has probably muddled other genetic inferences made about *Thuja plicata* previously, which suggested that the ITR populations were a result of coastal recolonization. While our data indicate that coastal migrants contributed to the genetic architecture of the current ITR populations, they provide strong evidence that Pleistocene refugia contributed to that architecture as well. This is supported by the high proportion of rare alleles observed in the ITR populations for *Tsuga heterophylla* and *Thuja plicata*, rare alleles that could only be the result of an ancient vicariant event with the coastal population.

## ACKNOWLEDGEMENTS

Support for this work was from National Science Foundation grants no. DEB-1457519 and DEB-1457726, and the Institute for Bioinformatics and Evolutionary Studies (IBEST) at the University of Idaho, which is supported by NIH NCR 1P20RR016454-01, NIH NCR 1P20RR016448-01 and NSF EPS-809935. We also thank John H. Russell, Charlie Cartwright, Sally Aitken and Lise van der Merwe for their help in obtaining and sampling *Thuja plicata* and *Tsuga heterophylla* tissue samples from localities throughout Canada from the clone bank of wild stand trees at the Cowichan Lake Research Station. We also thank members of the Tank, Sullivan, and

Carstens laboratories for sharing their knowledge on phylogeography and this ecosystem.

## AUTHOR CONTRIBUTIONS

MR, DCT and JS developed the research concept; MR, DCT, JS, MLS, AE and BC contributed to study design and implementation. MR, MLS, AE, DT, MS and NM all collected samples in the field and performed laboratory work for sequencing. MR performed analysis and wrote the manuscript. All authors contributed to critiques of the analysis and subsequent revisions of the text.

## CONFLICT OF INTEREST

We report no conflicts of interest.

## BENEFIT SHARING STATEMENT

Benefits from this research accrue from the sharing of our data and results on public databases as described above.

## OPEN RESEARCH BADGES



This article has earned an Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at [10.5061/dryad.pkOp2ngq9](https://doi.org/10.5061/dryad.pkOp2ngq9).

## DATA AVAILABILITY STATEMENT

All raw Sequence reads and assembly data for each library were deposited in Dryad: <https://doi.org/10.5061/dryad.pkOp2ngq9>. JUPYTER notebooks corresponding to all analyses in this study, including sequence assembly from raw reads, for each species are available at: <https://github.com/ruffleymr/ThujaTsugaAnalyses>.

## ORCID

Megan Ruffley <https://orcid.org/0000-0003-1796-2719>  
 Anahí Espíndola <https://orcid.org/0000-0001-9128-8836>  
 Bryan Carstens <https://orcid.org/0000-0002-1552-227X>  
 Jack Sullivan <https://orcid.org/0000-0003-0216-6867>

## REFERENCES

- Andrews, K., Good, J., Miller, M., Gordon, L., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Review Genetist*, 17, 81–92. <https://doi.org/10.1038/nrg.2015.28>
- Avise, J. C., Arnold, J., Ball, R., Bermingham, E., Lamb, T., Neigel, J., Reeb, C., & Saunders, N. C. (1987). Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, 18, 489–522. <https://doi.org/10.1146/annurev.es.18.110187.002421>
- Bagley, J. C., Heming, N. M., Gutiérrez, E. E., Devisetty, U. K., Mock, K. E., Eckert, A. J., & Strauss, S. H. (2020). Genotyping-by-sequencing and ecological niche modeling illuminate phylogeography, admixture, and Pleistocene range dynamics in quaking aspen (*Populus tremuloides*). *Ecology and Evolution*, 10(11), 4609–4629.
- Björk, C. R. (2010). Distribution patterns of disjunct and endemic vascular plants in the interior wetbelt of northwest North America. *Botany-Botanique*, 88(4), 409–428. <https://doi.org/10.1139/B10-030>

- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
- Brent, R. P. (1974). Algorithms for minimization without derivatives. *IEEE Transactions on Automatic Control*, 19, 632–633. <https://doi.org/10.1109/TAC.1974.1100629>
- 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 northwestern North America: A synthesis. In: J. Silvertown, & J. Antonovics (Eds), *Integrating ecological and evolutionary processes in a spatial context* (pp. 319–339). Blackwell Science.
- Burbrink, F. T., Chan, Y. L., Myers, E. A., Ruane, S., Smith, B. T., & Hickerson, M. J. (2016). Asynchronous demographic responses to Pleistocene climate change in eastern Nearctic vertebrates. *Ecology Letters*, 19, 1457–1467. <https://doi.org/10.1111/ele.12695>
- Callahan, C. M., Rowe, C. A., Ryel, R. J., Shaw, J. D., Madritch, M. D., & Mock, K. E. (2013). Continental-scale assessment of genetic diversity and population structure in quaking aspen (*Populus tremuloides*). *Journal of Biogeography*, 40, 1780–1791.
- Carstens, B., Lemmon, A. R., & Lemmon, E. M. (2012). The Promises and Pitfalls of Next-Generation Sequencing Data in Phylogeography. *Systematic Biology*, 61(5), 713–715. <https://doi.org/10.1093/sysbio/sys050>
- Carstens, B. C., Brennan, R. S., Chua, V., Duffie, C. V., Harvey, M. G., Koch, R. A., McMahan, C. D., Nelson, B. J., Newman, C. E., Satler, J. D., Seeholzer, G., Posbic, K., Tank, D. C., & Sullivan, J. (2013). Model selection as a tool for phylogeographic inference: An example from the willow *Salix melanopsis*. *Molecular Ecology*, 22(15), 4014–4028.
- Carstens, B. C., Brunsfeld, S. J., Demboski, J. R., Good, J. M., & Sullivan, J. (2005). Investigating the evolutionary history of the Pacific Northwest mesic forest ecosystem: Hypothesis testing within a comparative phylogeographic framework. *Evolution*, 59, 1639–1652. <https://doi.org/10.1111/j.0014-3820.2005.tb01815.x>
- 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>
- Chan, Y. L., Schanzenbach, D., & Hickerson, M. J. (2014). Detecting concerted demographic response across community assemblages using hierarchical approximate Bayesian computation. *Molecular Biology and Evolution*, 31, 2501–2515. <https://doi.org/10.1093/molbev/msu187>
- Chase, M., Bleskie, C., Walker, I. R., Gavin, D. G., & Hu, F. S. (2008). Midge-inferred Holocene summer temperatures in Southeastern British Columbia, Canada. *Palaeogeography Palaeoclimatology Palaeoecology*, 257, 244–259. <https://doi.org/10.1016/j.palaeo.2007.10.020>
- Clements, F. E. (1916). Plant succession: An analysis of the development of vegetation. *Carnegie Institution of Washington Publication Sciences*, 242, 1–512.
- Daubenmire, R. (1975). Floristic plant geography of eastern Washington and northern Idaho. *Journal of Biogeography*, 2, 1–18. <https://doi.org/10.2307/3038197>
- Davis, M. B. (1981). Quaternary history and the stability of forest communities. In D. C. West, H. H. Shugart, & D. B. Botkin (Eds.), *Forest succession. Springer advanced texts in life sciences* (pp. 132–153). Springer.
- Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, 19, 11–15.
- Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359–361. <https://doi.org/10.1007/s12686-011-9548-7>
- 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., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, 3(8), 2592–2594. <https://doi.org/10.1093/bioinformatics/bt2966>
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19), 2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>
- Espindola, A., Ruffley, M., Smith, M. L., Tank, D. C., Carstens, B. C., & Sullivan, J. (2016). Identifying cryptic diversity with predictive phylogeography. *Proceedings of the Royal Society B: Biological Sciences*, 283(1841), 20161529. <https://doi.org/10.1098/rspb.2016.1529>
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14, 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics*, 9(10), 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>
- Faegri, K., & Iversen, J. (1992). *Textbook of pollen analysis*. Hafner.
- 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. <https://doi.org/10.1093/genetics/164.4.1567>
- Fernandez, M. C., Hu, F. S., Gavin, D. G., Lafontaine, G., Heath, K. D., & Vanderpoorten, A. (2021). A tale of two conifers: Migration across a dispersal barrier outpaced regional expansion from refugia. *Journal of Biogeography*, 00, 1–11. <https://doi.org/10.1111/jbi.14209>
- Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37, 4302–4315. <https://doi.org/10.1002/joc.5086>
- Flessa, K. W., Jackson, S. T., Aber, J. D., Arthur, M. A., Crane, P. R., Erwin, D. H., Graham, R. W., Jackson, J. B. C., Kidwell, S. M., Maples, C. G., Peterson, C. H., & Reichman, O. J. C. (2005). *The geological record of ecological dynamics: Understanding the biotic effects of future environmental change*. National Academies Press.
- Garrick, R. C., Bonatelli, I. A. S., Hyseni, C., Morales, A., Pelletier, T. A., Perez, M. F., Rice, E., Satler, J. D., Symula, R. E., Thomé, M. T. C., & Carstens, B. C. (2015). The evolution of phylogeographic datasets. *Molecular Ecology*, 24, 1164–1171. <https://doi.org/10.1111/mec.13108>
- Gavin, D. G., & Hu, F. S. (2006). Spatial variation of climatic and non-climatic controls on species distribution: The range limit of *Tsuga heterophylla*. *Journal of Biogeography*, 33, 1384–1396. <https://doi.org/10.1111/j.1365-2699.2006.01509.x>
- 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>
- Gehara, M., Garda, A. A., Werneck, F. P., Oliveira, E. F., da Fonseca, E. M., Camurugi, F., Magalhães, def. M., Lanna, F., Sites, Jr J. W., Marques, R., Silveira-Filho, R., São Pedro, V. A., Colli, G. R., Costa, G. C., & Burbrink, F. T. (2017). Estimating synchronous demographic changes across populations using hABC and its application for a herpetological community from northeastern Brazil. *Molecular Ecology*, 26, 4756–4771. <https://doi.org/10.1111/mec.14239>
- Gibbard, P. L., Head, M. J., Walker, M. J. & Subcommittee on Quaternary Stratigraphy. (2010). Formal ratification of the Quaternary System/Period and the Pleistocene Series/Epoch with a base at 2.58 Ma. *Journal of Quaternary Science*, 25(2), 96–102.

- Gleason, H. A. (1926). The individualistic concept of the plant association. *Bulletin of the Torrey Botanical Club*, 53(1), 7–26. <https://doi.org/10.2307/2479933>
- Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, 5(10), e1000695. <https://doi.org/10.1371/journal.pgen.1000695>
- Habeck, J. R. (1987). Present-day vegetation in the northern rocky mountains. *Annals of the Missouri Botanical Garden*, 74(4), 804–840. <https://doi.org/10.2307/2399451>
- Herring, E. M., & Gavin, D. G. (2015). Climate and vegetation since the Last Interglacial (MIS 5e) in a putative glacial refugium, northern Idaho, USA. *Quaternary Science Reviews*, 117, 82–95. <https://doi.org/10.1016/j.quascirev.2015.03.028>
- Hickerson, M. J., Stahl, E. A., & Lessios, H. A. (2006). Test for simultaneous divergence using approximate Bayesian computation. *Evolution*, 60, 2435–2453. <https://doi.org/10.1111/j.0014-3820.2006.tb01880.x>
- Janes, J. K., Miller, J. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., Cullingham, C. I., & Andrew, R. L. (2017). The  $K = 2$  conundrum. *Molecular Ecology*, 26, 3594–3602. <https://doi.org/10.1111/mec.14187>
- Kirkwood, J. E. (1922). Forest distribution in the northern Rocky Mountains. *Univ. Montana Bulletin*, 247, 1–180.
- Kruschke, J. K. (2011). *Doing Bayesian data analysis: A tutorial with R and BUGS*. Academic Press.
- Liaw, A., & Wiener, M. (2002). Classification and Regression by random Forest. *R News*, 2(3), 18–22.
- Mehring, P. J., Jr. (1996). Columbia River Basin Ecosystems: Late Quaternary Environments. Contract Report. Interior Columbia Basin Ecosystem Management Project.
- 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., Espindola, A., Waits, L. P., & 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.
- Minore, D. (1990). *Thuja plicata* Donn ex D. Don western redcedar. In: R. M. Burns, & B. H. Honkala, technical coordinators. *Silvics of North America. Volume 1. Conifers. Agric. Handb. 654* (pp. 590–600). U.S. Department of Agriculture, Forest Service.
- Newmaster, S. G., Belland, R. J., Arsenaault, A., & Vitt, D. H. (2003). Patterns of bryophyte diversity in humid coastal and inland cedar-hemlock forests of British Columbia. *Environmental Reviews*, 11, S159–S185. <https://doi.org/10.1139/a03-016>
- 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>
- Novembre, J. (2016). Pritchard, Stephens, and Donnelly on population structure. *Genetics*, 204, 391–393. <https://doi.org/10.1534/genetics.116.195164>
- O'Connell, L. M., Ritland, K., & Thompson, S. L. (2008). Patterns of post-glacial colonization by western redcedar (*Thuja plicata*, Cupressaceae) as revealed by microsatellite markers. *Botany-Botanique*, 86, 194–203.
- Owens, J. N., & Molder, M. (1984). The reproductive cycles of western and mountain hemlock. Victoria, BC: Ministry of Forests. *Information Services Branch*, 32, p. [19144].
- 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(5), e37135. <https://doi.org/10.1371/journal.pone.0037135>
- 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., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959. <https://doi.org/10.1093/genetics/155.2.945>
- Pudlo, P., Marin, J. M., Estoup, A., Cornuet, J. M., Gautier, M., Robert, C. P. (2016). Reliable ABC model choice via random forests. *Bioinformatics*, 32(6), 859–866. <https://doi.org/10.1093/bioinformatics/btv684>
- Puechmaille, S. J. (2016). The program STRUCTURE does not reliably recover the correct population structure when sampling is uneven: Subsampling and new estimators alleviate the problem. *Molecular Ecology Resources*, 16, 608–627.
- Rankin, A. M., Wilke, T., Lucid, M., Leonard, W., Espindola, A. E., Smith, M. L., Carstens, B. C., & 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(4), 876–889. <https://doi.org/10.1093/biolinnean/blz040>
- Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, 4, e2584. <https://doi.org/10.7717/peerj.2584>
- 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>
- Rosenberg, S. M., Walker, I. R., & Mathewes, R. W. (2003). Postglacial spread of hemlock (*Tsuga*) and vegetation history in Mount Revelstoke National Park, British Columbia, Canada. *Canadian Journal of Botany*, 81, 139–151.
- Ruffley, M., Smith, M. L., Espindola, A., Carstens, B. C., Sullivan, J., & Tank, D. C. (2018). Combining allele frequency and tree-based approaches improves phylogeographic inference from natural history collections. *Molecular Ecology*, 27, 1012–1024. <https://doi.org/10.1111/mec.14491>
- Ruffley, M., Smith, M. L., & Turck, D. (2022). *Thuja plicata* and *Tsuga heterophylla* ddRADseq raw data. <https://doi.org/10.5061/dryad.pk0p2ngq9>
- Smith, M. L., & Carstens, B. C. (2020). Process-based species delimitation leads to identification of more biologically relevant species\*. *Evolution*, 74, 216–229. <https://doi.org/10.1111/evo.13878>
- Smith, M. L., Ruffley, M., Espindola, A., Tank, D. C., Sullivan, J., & Carstens, B. C. (2017). Demographic model selection using random forests and the site frequency spectrum. *Molecular Ecology*, 26(17), 4562–4573. <https://doi.org/10.1111/mec.14223>
- Smith, M. L., Ruffley, M., Rankin, A. M., Espindola, A., Tank, D. C., Sullivan, J., & Carstens, B. C. (2018). Testing for the presence of cryptic diversity in tail-dropper slugs (Prophyaon) using molecular data. *Biological Journal of the Linnean Society*, 124(3), 518–532. <https://doi.org/10.1093/biolinnean/bly067>
- 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.
- 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.
- 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.
- Sullivan, J., Smith, M. L., Espindola, A., Ruffley, M., Rankin, A., Tank, D. C., & Carstens, B. C. (2019). Integrating life history traits into predictive phylogeography. *Molecular Ecology*, 28, 2062–2073. <https://doi.org/10.1111/mec.15029>

- Tesky, J. L. (1992). *Tsuga heterophylla*. In: *Fire Effects Information System*, [Online]. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory (Producer). Retrieved from <https://www.fs.fed.us/database/feis/plants/tree/tsuhet/all.html> [2020, April 14]
- Turner, D. P. (1985). Successional relationships and a comparison of biological characteristics among six northwestern conifers. *Bulletin of the Torrey Botanical Club*, 112(4), 421–428.
- Waitt, R. B., & Thorson, R. M. (1983). The Cordilleran ice sheet in Washington, Idaho, and Montana. In S. C. Porter (Ed.), *Late-Quaternary environments of the United States* (pp. 53–70). University of Minnesota Press.
- Waring, R. H., & Franklin, J. F. (1979). Evergreen coniferous forests of the Pacific Northwest. *Science (New York, N.Y.)*, 204, 1380–1386.
- Whitlock, C. (1992). Vegetational and climatic history of the Pacific Northwest during the last 20,000 years: Implications for understanding present day biodiversity. *Northwest Environmental Journal*, 8, 5–28.
- Wilke, T., & Duncan, N. (2004). Phylogeographical patterns in the American Pacific Northwest: Lessons from the arionid slug *Prophysaon coeruleum*. *Molecular Ecology*, 13, 2303–2315. <https://doi.org/10.1111/j.1365-294X.2004.02234.x>
- Xue, A. T., & Hickerson, M. J. (2015). The aggregate site frequency spectrum for comparative population genomic inference. *Molecular Ecology*, 24, 6223–6240.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Ruffley, M., Smith, M. L., Espíndola, A., Turck, D. F., Mitchell, N., Carstens, B., Sullivan, J., & Tank, D. C. (2022). Genomic evidence of an ancient inland temperate rainforest in the Pacific Northwest of North America. *Molecular Ecology*, 31, 2985–3001. <https://doi.org/10.1111/mec.16431>