MOLECULAR ECOLOGY

Molecular Ecology (2010) **19**, 1877–1897 doi: 10.1111/j.1365-294X.2010.04622.x

Phylogeography of Douglas-fir based on mitochondrial and chloroplast DNA sequences: testing hypotheses from the fossil record

PAUL F. GUGGER,* SHINYA SUGITA† and JEANNINE CAVENDER-BARES*

**Department of Ecology, Evolution and Behavior, University of Minnesota, Saint Paul, MN 55108, USA, †Institute of Ecology, Tallinn University, 10120 Tallinn, Estonia*

Abstract

The integration of fossil and molecular data can provide a synthetic understanding of the ecological and evolutionary history of an organism. We analysed range-wide maternally inherited mitochondrial DNA and paternally inherited chloroplast DNA sequence data with coalescent simulations and traditional population genetic methods to test hypotheses of population divergence generated from the fossil record of Douglas-fir (*Pseudotsuga menziesii*), an ecologically and economically important western North American conifer. Specifically, we tested (i) the hypothesis that the Pliocene orogeny of the Cascades and Sierra Nevada caused the divergence of coastal and Rocky Mountain Douglas-fir varieties; and (ii) the hypothesis that multiple glacial refugia existed on the coast and in the Rocky Mountains. We found that Douglas-fir varieties diverged about 2.11 Ma (4.37 Ma–755 ka), which could be consistent with a Pliocene divergence. Rocky Mountain Douglas-fir probably resided in three or more glacial refugia. More variable molecular markers would be required to detect the two coastal refugia suggested in the fossil record. Comparison of mitochondrial DNA and chloroplast DNA variation revealed that gene flow via pollen linked populations isolated from seed exchange. Postglacial colonization of Canada from coastal and Rocky Mountain refugia near the ice margin at the Last Glacial Maximum produced a wide hybrid zone among varieties that formed almost exclusively by pollen exchange and chloroplast DNA introgression, not seed exchange. Postglacial migration rates were 50–165 m/year, insufficient to track projected 21st century warming in some regions. Although fossil and genetic data largely agree, each provides unique insights.

Keywords: coalescent simulations, glacial refugia, introgression, isolation with migration, mutation rate, postglacial migration, *Pseudotsuga menziesii*, SAMOVA

Received 6 December 2009; Revision received 22 February 2010; accepted 3 March 2010

Introduction

The integration of fossil and molecular genetic data is critical to understanding the ecological and evolutionary history of an organism (Petit *et al.* 2008; Hu *et al.* 2009). Fossil data offer direct, dated evidence of species presence (and sometimes abundance), but suffer from lack of population (or even species) resolution and limited sampling. In contrast, inferences of population history from molecular data from modern forest trees can provide population and species level resolution from across an entire distribution, but yield estimates of population divergence that are imprecise. Hypotheses from the fossil record can be tested with molecular data, which in turn produce new insights that can be tested with fossil data as more sites are analysed. In this way, fossil and molecular data offer complementary information that can converge on the major geological, climatological or ecological causes of population divergence.

Range-wide syntheses of the fossil record in Europe and North America have revealed that temperate and

Correspondence: Paul F. Gugger, Fax: 612-624-6777;
E-mail: gugg0030@umn.edu

boreal forests responded to the climate warming after the Last Glacial Maximum (21–18 ka) by northward range expansion (postglacial migration) from glacial refugia south of the ice sheets (Davis 1976, 1981, 1983; Huntley & Birks 1983). This process had a profound effect on the population genetic structure of forests that can still be detected today (Petit *et al.* 1997, 2003; Hewitt 2000). Isolated from one another since at least the start of the Wisconsinan glaciation (115 ka), glacial refugia are characterized by within-refugium genetic diversity and among-refugium divergence due to the effects of isolation, mutation, drift and selection. This diversity is lost along postglacial migration routes due to the effects of successive dispersal bottlenecks and founding events (Petit *et al.* 1997; Bialozyt *et al.* 2006; Gugger *et al.* 2008). Although fossil and molecular studies largely corroborate one another (Petit *et al.* 2002a, 2003), molecular data have revealed cryptic northern refugia and identified migration pathways that suggest lower postglacial migration rates than the fossil record (McLachlan *et al.* 2005; Anderson *et al.* 2006; Magri *et al.* 2006; Godbout *et al.* 2008; Petit *et al.* 2008). Moreover, in some cases, deep genetic divergence among some populations observed in molecular phylogeographic studies has invited the reanalysis of the fossil record, suggesting pre-Wisconsinan population divergence (Magri *et al.* 2007).

The fossil record is a rich source of a priori hypotheses that can be tested using coalescent and other statistical phylogeographic methods that account for stochastic processes during population divergence (Knowles & Maddison 2002; Spellman & Klicka 2006). Despite the especially abundant fossil record for tree species, these methods have not commonly been used in plants. One persistent limitation has been the low sequence variation of plant mitochondrial (mtDNA) and chloroplast DNA (cpDNA) compared to animal mtDNA. Here, we formally integrate fossil and phylogeographic data, and attempt to overcome limitations that have prevented the use of coalescent methods in phylogeographic studies of plants, to study the evolutionary history of Douglas-fir (*Pseudotsuga menziesii*), an ecologically and economically important western North American tree.

Douglas-fir is well represented in the fossil record from the early Miocene (~23 Ma) to the late Holocene (Fig. 1; Hermann 1985). During the Miocene and Pliocene, *Pseudotsuga* fossils are found primarily from central British Columbia to southern California, from the coast to the interior as far as Idaho (Fig 1a). By the Pleistocene, *Pseudotsuga* fossils remain near the coast, disappear from the Columbia Plateau and western Great Basin, and appear in the central and southern Rocky Mountains. Coincident with this transition is the rise of the Cascade Range and Sierra Nevada primarily during the Pliocene, which is thought to have imposed a rain shadow that dried the Columbia Plateau and Great Basin (Brunsfeld *et al.* 2001). Orogeny and xerification could have caused the vicariant separation of Douglas-fir populations into what we now recognize as its two varieties: coastal (*P. menziesii* var. *menziesii*) and Rocky Mountain (*P. menziesii* var. *glauca*). These varieties are morphologically (Hermann & Lavender 1990), chemically (von Rudloff 1972; Zavarin & Snajberk 1973), and genetically (Li & Adams 1989; Aagaard *et al.* 1995) distinct with a transition zone in British Columbia thought to be restricted to the east slope of the Coast Range (Critchfield 1984; Li & Adams 1989; Hermann & Lavender 1990; Ponoy *et al.* 1994).

Subsequent population divergence within each variety probably occurred during Pleistocene glacial cycles (Fig. 1b). In a recent review of late Quaternary fossil pollen and packrat midden macrofossil data from over 500 sites in the USA and Canada, Gugger & Sugita (submitted) found that Douglas-fir resided in as many as two coastal refugia and three to four Rocky Mountain refugia. Refugia were defined broadly to mean regions or localities that were the sources of modern populations (Bennett & Provan 2008). Some of these refugia were uncertain because they were based on a limited number of fossil sites (southern Utah) or they had very low fossil Douglas-fir pollen abundances (Yellowstone National Park area). Moreover, because fossil data cannot definitively delimit the ranges of distinct populations, single-refugium hypotheses were also proposed for each variety. Canada was colonized by an unknown mixture of each variety from a coastal refugium in Washington near the ice margin at the Last Glacial Maximum and a Rocky Mountain refugium in the northern Rockies near Yellowstone. The southernmost populations in California and the Rockies retreated upslope 700–1000 m and probably did not contribute to the colonization of the northern part of the range.

Other hypotheses for the number of glacial refugia and postglacial migration routes have been proposed. Li & Adams (1989) observed differences in allozyme allele composition and frequency between the northern and southern Rockies and thus proposed two glacial refugia. St Clair *et al.* (2005) proposed that the postglacial migration of the Rocky Mountain variety southward along the east side of the Cascades via Canada to explain genecological dissimilarity of east and west Cascades populations and the apparent convergence of east Cascades populations with Rocky Mountain populations.

We used mtDNA and cpDNA sequence data to test (i) the hypothesis that the Pliocene orogeny of the Cascade Range and Sierra Nevada coincided with the divergence of Douglas-fir varieties; and (ii) hypotheses

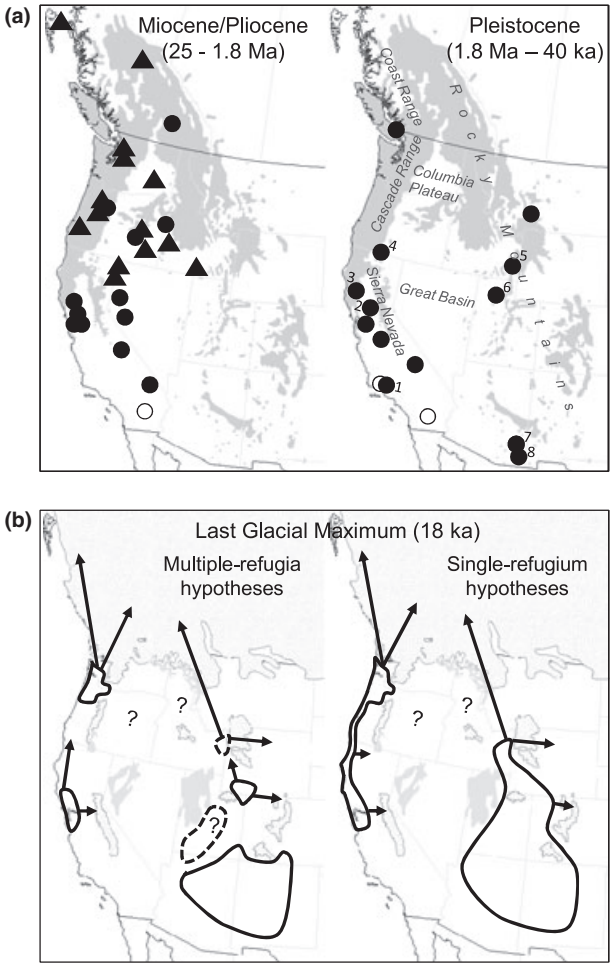

Fig. 1 Hypotheses from the fossil record. (a) *Pseudotsuga* fossil data are consistent with the Pliocene vicariance hypothesis. Miocene and Pliocene fossils were present near the coast and in the Great Basin and Columbia Plateau, but Pleistocene fossils were absent from the Great Basin and Columbia Plateau and present near the coast and in the Rockies. Miocene and Pliocene are grouped on one panel because when many of the fossil records were first published the geological timescale delineated those time periods differently with a boundary at 15 Ma. This means that many fossils originally attributed to the Pliocene were in fact from the Miocene. We marked fossils from 25 to 15 Ma as triangles and those from 15 to 1.8 Ma as circles. Both show that *Pseudotsuga* was present in the Great Basin. Filled shapes represent fossils from the genus *Pseudotsuga*, and empty circles indicate *Pseudotsuga macrocarpa* in particular. Modern range of Douglas-fir is in grey (Little 1971). This figure was modified from Hermann (1985) with our additions numbered: 1, Seacliff (Axelrod 1983); 2, Teichert Site (Rymer 1981); 3, Cache Formation (Ritter & Hatoff 1977); 4, Tule Lake (Adam & Vagenas 1990); 5, Grays Lake (Beiswenger 1991); 6, Great Salt Lake (Davis & Moutoux 1998); 7, Lake Cochise (Martin 1963); 8, Lehner Ranch Arroyo (Mehringer & Haynes 1965). (b) Based on a synthesis of late Quaternary fossil records, Gugger & Sugita (submitted) proposed multiple-refugia hypotheses for the coastal and Rocky Mountain varieties (left), which are shown with ice sheets (stippled) and glacial lakes (grey) at the Last Glacial Maximum. Their analysis suggested two coastal refugia and 3–4 Rocky Mountain refugia (encircled). Less certain refugia are marked with a dotted line and regions where late Quaternary fossil evidence is lacking are marked with question marks. The two-refugia hypothesis for the Rocky Mountains proposed by Li & Adams (1989) would approximately cluster the three southernmost refugia, but retain the northernmost refugium shown here. Because the fossil records do not offer population-level resolution, single-refugium hypotheses were also proposed for each variety (right). For either set of hypotheses, postglacial migration (arrows) was primarily northward into Canada with some eastward migration.

describing the number of glacial refugia and postglacial migration routes during the late Pleistocene and Holocene. Finally, we take advantage of the fact that in Douglas-fir mtDNA is maternally inherited and thus seed dispersed (Marshall & Neale 1992), whereas cpDNA is paternally inherited (Neale *et al.* 1986) and thus dispersed first in pollen and then as fertilized seed, to investigate differences in the history of seed and pollen dispersal.

Materials and methods

Sampling

Leaf or bud tissue was collected from 87 sites throughout the range of Douglas-fir in the USA and Canada (Fig. 2; Table S1, Supporting information). Seven sites were sampled from a common garden near Sooke, British Columbia (BC; 48.41667°N, −123.867°W, 140 m) and 17 from a common garden near Enderby, BC (50.5°N, −119°W, 600 m; Zhang *et al.* 1993), both under the direction of the British Columbia Ministry of Forests. Each provenance in the common gardens was seeded with multiple individuals from each of several maternal lines, thus we only used one individual to avoid the possibility of sampling the same tree (mother) twice. The remaining 63 sites were natural populations, where 1–11 individuals (mean = 3.1) separated from one another by at least 50 m were collected. We also collected four samples of *Pseudotsuga macrocarpa* and one of *Larix occidentalis* as outgroups.

DNA preparation

Total genomic DNA was extracted from leaf or bud tissue using QIAGEN DNeasy Plant Mini Kit according to the manufacturer's instructions.

After surveying variation in several mtDNA and cpDNA sequences that were shown to vary neutrally

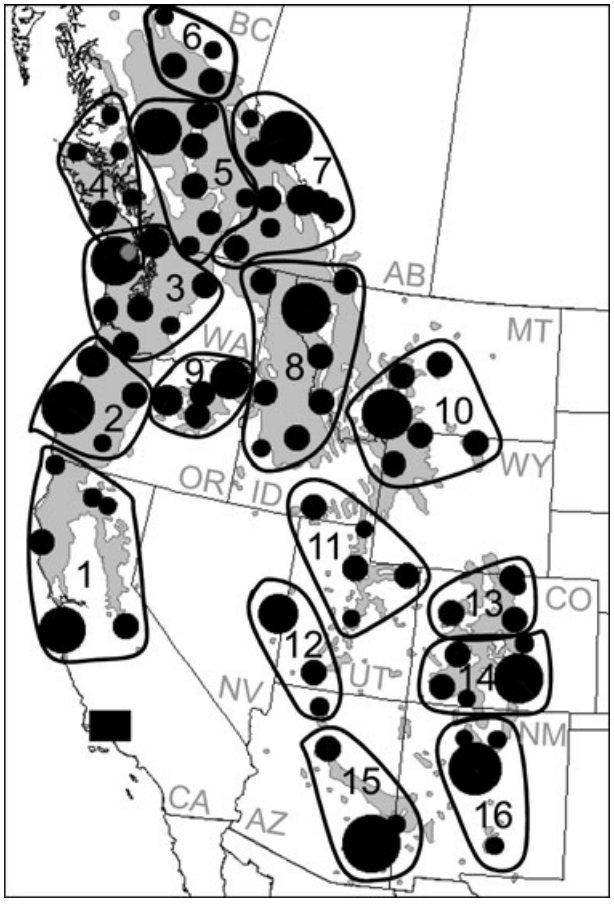

Fig. 2 Sampling strategy. We grouped 219 individuals from 87 sites (black/grey circles) into 16 populations (lines). These populations were defined primarily along major geographic breaks, such as the Klamath Mountains (populations 1/2 boundary), Columbia River Valley (2/3), Coast Range (4/5), Okanagan Valley (5/7), Deschutes River Valley (2/9), Hells Canyon (8/9), Snake River Plain (8/11), Great Divide Basin (10/11), Grand Canyon (12/15), other wide, arid valleys in the Southwest (e.g. 11/13, 15/16), and previously glaciated high peaks in Colorado (13/14; Fig. 1). Several boundaries were already recognized as probable varietal boundaries (2/9, 4/5, 5/7) and others were hypothesized subdivisions from the fossil record (e.g. 1/2, 10/11, ~11/12). Other groupings (not shown) based on geography were also evaluated, but did not lead to meaningful differences in our results or conclusions. Circles are scaled to the sampling intensity (1–11 individuals per site). One circle is grey to set it apart from overlapping points. Rectangle marks sampling location of four *Pseudotsuga macrocarpa*. The modern range of Douglas-fir is shown in grey (Little 1971). States and provinces are labelled with abbreviations: AB, Alberta; AZ, Arizona; BC, British Columbia; CA, California; CO, Colorado; ID, Idaho; MT, Montana; NM, New Mexico; NV, Nevada; OR, Oregon; UT, Utah; WA, Washington; WY, Wyoming.

in other species, we selected the two most variable segments from each genome for analyses. For cpDNA, we chose *rps7–trnL*, which contains intergenic spacers, the *ndhB* pseudogene and its intron; and *rps15–psaC*, which contains intergenic spacers and the ndh*H*, ndh*I*, and ndh*E* pseudogenes. ndh genes have been lost of are non-coding in Pinaceae (Wakasugi *et al.* 1994; Braukmann *et al.* 2009). For each, we designed novel primers (Table 1) based on the *Pinus thunbergii* complete chloroplast genome sequence (Wakasugi *et al.* 1994) using Primer3 (http://primer3.sourceforge.net/; Rozen & Skaletsky 2000).

The mtDNA segments we chose were variable region 7 (V7) of the small-subunit ribosomal RNA gene (*19S rDNA V7*; Duff & Nickrent 1997, 1999) and the first intron of the *nad7* gene (*nad7i1*; Jaramillo-Correa *et al.* 2004). Both were shown variable in *Picea* spp. (Jaramillo-Correa *et al.* 2003, 2004) and the latter in *Pinus* spp. (Godbout *et al.* 2005, 2008). Douglas-fir-specific primers (Table 1) were designed for each based on published sequences (Chaw *et al.* 2000; Jaramillo-Correa *et al.* 2004). For *V7*, care was taken to produce primers that would not amplify similar small-subunit rDNA sequences in the chloroplast (16S; Wakasugi *et al.* 1994) and nuclear (18S; Chaw *et al.* 1997, 2000) genomes.

Polymerase chain reactions (PCR; 25 μL) using QIAGEN *Taq* PCR Core Kit contained 2.5 μL 10 × PCR buffer, 5 μL Q solution, and final concentrations of 200 μM dNTPs, 2.5 mM $MgCl_2$, 4 μM of each primer, and 0.5 U *Taq* DNA polymerase. Except for annealing temperature (Table 1), amplification programs were the same: 4 min at 94 °C; 35 cycles of 1 min at 94 °C, 45 s at 55–58 °C, and 1.5 min at 72 °C; and 10 min 72 °C. PCR products were then cleaned with QIAquick PCR Purification Kit according to the manufacturer's instructions.

Purified products were sequenced on an ABI Prism 3730xl DNA Analyzer (PE Applied Biosystems). Each locus was sequenced using the forward primer yielding 300–700 bp (Table 1). Unique haplotypes were verified by resequencing with forward and reverse primers. Sequences were aligned and edited using Sequencher 4.8 (Gene Codes). Because organellar genomes in plants generally do not recombine and are uniparentally inherited (Reboud & Zeyl 1994), they each act as a single locus. Accordingly, we combined the two cpDNA sequences to produce a single cpDNA sequence and the two mtDNA sequences to produce a single mtDNA sequence. Here, their variants are referred to as chlorotypes and mitotypes, respectively.

Parsimony network and mutation rate

Parsimony networks were produced separately for mitotypes and chlorotypes using TCS 1.21 (Clement *et al.* 2000) with insertion/deletions coded as a fifth state.

Using DnaSP 5.0 (Librado & Rozas 2009), per base pair per year (μ_y) and per base pair per generation (μ_g)

Table 1 Primers developed for this study

Locus	Segment	Approximate amplified length (bp)	Primer name	Primer sequence (5′–3′)	Annealing temperature (°C)
mtDNA	*19S rDNA V7*	690	V7_3f	GAGCCAAGGAGGCAGATTG	58
			V7_3r	ATCCTTGGTCTGATGCTTCG	
	nad7i1	300	nad7i1_2f	ACCTAACAGAACGCACAAGG	55
			nad7i1_2r	TTCCAACCAAGAATTGATCC	
cpDNA	*rps7–trnL* (*ΨndhB*)	1810	rps7f	GGTTATTAGGGGCATCTCG	58
			trnLr	CGTGTCTACCATTTCACCATC	
	rps15–psaC (*ΨndhH/I/E*)	1630	rps15f	GGTATCCGTGGGCTAAAAAC	58
			psaCr	CAATACATCTGTGGGACAAGC	

substitution rates for each locus were calculated as the mean number of nucleotide substitutions per site (d_{XY}; Nei 1987) with Jukes & Cantor (1969) correction among all *Pseudotsuga* and *Larix* haplotypes divided by two times the divergence time in years or generations. We assumed the divergence of *Pseudotsuga* and *Larix* occurred about 50 Ma. The oldest *Larix* fossil dates to the Eocene (~45 Ma; Axelrod 1990; LePage & Basinger 1991; Schorn 1994) and oldest unambiguous *Pseudotsuga* fossil dates to the early Oligocene (~32 Ma; Lakhanpal 1958; Schorn 1994), though older fossils have been proposed (Penhallow 1902, 1907; Axelrod 1966; Hermann 1985). Consistent with our assumption, a molecular clock enforced on a *matK* gene phylogeny of Pinaceae suggests that *Larix* and *Pseudotsuga* split in the Palaeocene (~60 Ma; Wang *et al.* 2000). An approximate average generation time (T) can be calculated according to $T = \alpha + [s/(1–s)]$ (Lande *et al.* 2003; Spellman & Klicka 2006), where α is the time to maturity (~15 years; Herman & Lavender 1990) and s is the adult annual survival rate (~0.99–0.995; Franklin & DeBell 1988; Hann *et al.* 2003). Based on this, estimates for T range from 114 to 214 years. If we consider periodic major disturbances (e.g. fire or disease outbreaks) that were not observed in the studies estimating the above adult annual survival rates, a lower survival rate (Mathiasen *et al.* 1990) and therefore a lower average generation time would be more likely. We primarily assumed an average generation time of 100 years, but also explored the effects of generation times as high as 200 years.

Mutation rates calculated among genera (*Larix* and *Pseudotsuga*) should be valid within *Pseudotsuga menziesii* if the sequences have evolved under a molecular clock model. To test this null hypothesis, we compared the likelihood scores under a model with and without a molecular clock (Muse & Weir 1992) using PAUP* 4.0b10 (Swofford 2003). We conducted two tests per locus, one including one representative from each species and variety and one including only haplotypes observed in *P. menziesii* (all individuals, both varieties). If the molecular clock could not be rejected at either evolutionary scale, we inferred a constant rate of substitutions. These tests require that a particular nucleotide substitution model be chosen, which we chose using hierarchical likelihood ratio tests implemented in PAUP* and MODELTEST 3.7 (Posada & Crandall 1998).

Genetic diversity and neutrality

Because many of our sample sites only contained one individual, we quantified general patterns of genetic diversity by grouping data a priori from each sample site into 16 populations and four regions based on geography (Fig. 2). Using Arlequin 3.11 (Excoffier *et al.* 2005), we estimated the haplotype richness (h), number of segregating sites (S), haplotype diversity (H; Nei 1987), nucleotide diversity (π; Tajima 1983; Nei 1987), and D (Tajima 1989) for each population, for each region, and for the whole data set. We also estimated the haplotype richness after correcting for unequal sample sizes with rarefaction (h_r; Hurlbert 1971) using Contrib 1.01 (Petit *et al.* 1998), and R_2 (Ramos-Onsins & Rozas 2002) using DnaSP. Latitudinal trends in the diversity (H, π, h_r) of populations were investigated with linear regression. D and R_2 were used to test for departures from the neutral expectations of constant population size and selective neutrality, and their significance was assessed with coalescent simulations (Hudson 1990). More negative values of D suggest population expansion or purifying selection. The R_2 test statistic tests the null hypothesis of constant population size (Ramos-Onsins & Rozas 2002). We computed genetic differentiation among populations as G_{ST} (Nei 1973; Pons & Petit 1995).

Population structure

To identify clusters of genetically similar populations, we analysed our data with a spatial analysis of molecular variance (SAMOVA). Geographically close populations are grouped into a user-defined number of groups (K) using a simulated annealing approach to maximize the

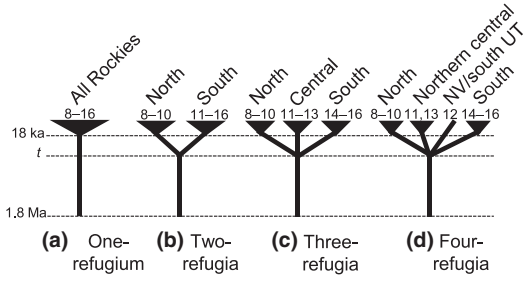

Fig. 3 (a) One-refugium (null), (b) two-refugia, (c) three-refugia and (d) four-refugia population models used to test hypotheses with coalescent simulations in Mesquite. Simulations were run so that a single ancestral population of constant size N_e split at time t into equal fractions of N_e. Dotted lines correspond to the Last Glacial Maximum (18 ka), the onset of recent glaciations (t; Wisconsinan = 115 ka, late Illinoisan = 190 ka), and the beginning of the Pleistocene (1.8 Ma). Population numbers matching those on Fig. 2 are shown above population groups.

variance (F_{CT}) among those groups (Dupanloup *et al.* 2002). We performed this analysis in SAMOVA 1.0 (Dupanloup *et al.* 2002) using K = 2–12 for each locus, and chose the number of groups that gave the highest F_{CT} or the number of groups for which F_{CT} began to plateau.

Testing refugia hypotheses with coalescent methods

Hypotheses for the number and extent of Rocky Mountain glacial refugia (Figs 1b and 3) were tested using coalescent simulations in Mesquite 2.6 (http://www.mesquiteproject.org). We did not use coalescent simulations to test hypotheses for the coastal variety due to lack of molecular variation (see Results). For the Rocky Mountain variety, we tested the following hypotheses: one-refugium, two-refugia, three-refugia and four-refugia, each with multiple Pleistocene divergence times among refugia: beginning of the Wisconsinan glaciation (115 ka) and beginning of the late Illinoisan glaciation (190 ka; Richmond & Fullerton 1986; Lisiecki & Raymo 2005).

First, 1000 coalescent genealogies were simulated under a hypothesized model of population divergence (Fig. 3) with effective population size, N_e (different numerical values were specified for different runs; more below). The fit of the simulated genealogies to the hypothesized model of population divergence was then assessed using the test statistic s (Slatkin & Maddison 1989), generating a null distribution for that hypothesis. The s value is the number of parsimony steps on a gene tree on which source population has been mapped as a multistate character. Typically, the s value for a single maximum likelihood tree from empirical data is compared to this null distribution to assess statistical significance of deviation from the hypothesized model (Knowles 2001; Spellman & Klicka 2006; Carstens & Richards 2007). However, as a way of dealing with the limited sequence variation and consequent poorly resolved tree, we computed s for 1000 trees drawn from the posterior distribution of a Bayesian Markov chain Monte Carlo analysis implemented in BEAST 1.4.8 (Drummond & Rambaut 2007). This generated an empirical distribution for s that accounted for the range of possible true trees. If the empirical s values were significantly higher (i.e. worse fit) than those from the simulated data, we rejected the hypothesis. Significance ($\alpha = 0.05$) was assessed conservatively by calculating the probability of observing an s value from the null distribution at least as high as the lowest empirical s. We could not assess the single-refugium hypothesis in this way because $s = 0$ for a single population calculated on any gene tree. Instead, we simulated 1000 genealogies under the single-refugium (null) hypothesis, and then measured their fit to each alternative hypothesis of two, three and four refugia, respectively (Knowles 2001). We rejected the null hypothesis in favour of the respective alternative if the highest empirical value of s was lower (i.e. better fit) than 95% of the simulated s values.

To estimate simultaneously the empirical trees and N_e, we ran BEAST for 20–30 million steps assuming a constant population size, a strict molecular clock, and a Hasegawa–Kishino–Yano substitution model (Hasegawa *et al.* 1985) with empirical base frequencies and no site heterogeneity. We repeated this twice for each locus to verify convergence upon the same values and combined replicate outputs. N_e was calculated for each locus from the median estimate of θ and the upper and lower bounds of its 95% highest posterior density (HPD) interval according to $\theta = 2N_e\mu_g$. We report these as the effective population sizes of females (mtDNA; N_{ef}) and males (cpDNA; N_{em}).

Coalescent simulations are sensitive to different N_e, and θ estimates from single loci have large errors (Edwards & Beerli 2000), so we tested each hypothesis in Mesquite using N_e calculated from the median point estimate of θ and the high and low values of its 95% HPD interval. We repeated this for N_e values calculated assuming a generation time of 200 years instead of 100 years, which had the effect of doubling μ_g and therefore halving N_e.

Divergence time among varieties

Divergence time among coastal (populations 1–3; Fig. 2) and Rocky Mountain (populations 8–16)

Douglas-fir varieties, excluding Canadian populations, was investigated using an isolation with migration (Neilson & Wakeley 2001) model as implemented in IMa (Hey & Neilson 2007). Under the full model, IMa simultaneously estimates six parameters (scaled by substitution rate): divergence time (t), migration from population one to two (m_1), migration from population two to one (m_2), effective population size of each population (N_1 and N_2) and effective population size of the ancestor (N_A). IMa assumes constant population size, neutral molecular markers, no recombination within loci, free recombination among loci, and a particular mutation model. We assessed selective neutrality and constant population size with D and R_2 as described earlier. Recombination is not common in organellar genomes (Chiu & Sears 1985; Reboud & Zeyl 1994; Birky 2001), but has been observed in *Picea* mtDNA (Jaramillo-Correa & Bousquet 2005). A four-gamete test (Hudson & Kaplan 1985) showed no evidence of mtDNA recombination in our data (not shown). We chose an infinite sites mutation model (Kimura 1969) for mtDNA and an Hasegawa–Kishino–Yano model (Hasegawa *et al.* 1985; Palsbøll *et al.* 2004) for cpDNA to accommodate homoplasy (see Results).

In IMa, 25 geometrically heated ($g_1 = 0.95$, $g_2 = 0.85$) Metropolis-coupled Markov chain Monte Carlo chains with 10 swaps per step and a maximum prior for $t = 5$ were run for 10–50 million steps beyond a 2 million step burn-in. The prior for t was chosen because values over five are biologically unreasonable, suggesting a divergence time of over 9 Ma. Maximum priors for m_1 and m_2 and scalars for the effective population size parameters (q_1, q_2, q_A) varied depending upon the comparison and were chosen based on preliminary runs of the program with larger parameter intervals (Won & Hey 2005). This analysis was performed for each locus and with both loci combined. To verify convergence upon the same parameter values, we ran this analysis three times for each comparison with different random seeds. Only estimates whose posterior distribution dropped to zero within the prior intervals investigated were trusted. To scale the outputs to demographic units, IMa uses the generation time and per locus per year substitution rate, which we calculated from μ_y by multiplying by the number of base pairs in the locus under consideration.

We also ran IMa in L-mode with 100 000–250 000 genealogies to test the null hypothesis that the likelihood of our data under the full model described above equals the likelihood under a simpler, nested model without migration (i.e. t, N_1, N_2, N_A, $m_1 = m_2 = 0$). This was conducted as a two part test, where first the full model was evaluated against a nested model in which $m_1 = m_2$. The test statistic is negative two times the natural log of the ratio of the estimated likelihoods of the nested model to the full model (-2Λ), which is X^2-distributed with a one degree of freedom (difference in number of parameters between models). If the nested model could not be rejected, the second test compared a model in which $m_1 = m_2$ against a yet simpler model, in which $m_1 = m_2 = 0$. Here, -2Λ has 0.5 probability of taking on a value of zero and 0.5 probability of taking on a value from a X^2 distribution with one degree of freedom (Chernoff 1954; Hey & Neilson 2007). If neither test was significant, the model with $m_1 = m_2 = 0$ was not rejected, so we ran the IMa analysis again with migration set to zero.

Results

We obtained sequences for both mtDNA segments (aligned, edited length of 943 bp) in 190 individuals from 82 sites and from both cpDNA segments (1537 bp) in 219 individuals from 87 sites (Table S1, Supporting Information). We observed two *nad7i1* haplotypes and six *V7* haplotypes, which combined, yielded seven mitotypes containing 10 base substitutions and five insertion/deletions (Fig. 4; GenBank accession nos in Table S2, Supporting Information). Of those, four were common (M1, M3, M4, M6) and three were rare (M2, M5, M7), occurring in four or fewer individuals. We also observed nine *rps7–trnL* haplotypes and 11 *rps15–psaC* haplotypes, which combined to 20 chlorotypes (Table S2, Supporting Information) containing 18 base substitutions and three insertion/deletions (Fig. 4). Two of these were common (C1, C5) and the rest were rare, occurring in one to nine individuals.

Visual inspection of the data suggested strong geographic structure in both mtDNA and cpDNA (Fig. 4). The two varieties were well delineated for both loci throughout much of the range, except in Canada, where cpDNA haplotypes were shared among varieties. The coastal variety lacked genetic structure for the sequences investigated here. Within the Rockies, additional population structure was clear in mtDNA data and suggestive in the cpDNA, where some mitotypes and chlorotypes had southern (e.g. M6, C20), central (e.g. M4, C16) or northern (e.g. M3, C7, C10) Rocky Mountain distributions.

Mutation rate

The per base pair per year substitution rate (μ_y) was 5.26×10^{-10} for mtDNA and 4.41×10^{-10} for cpDNA (Table 2). Sequence divergence and other mutation rates are reported in Table 2. For cpDNA, we rejected a

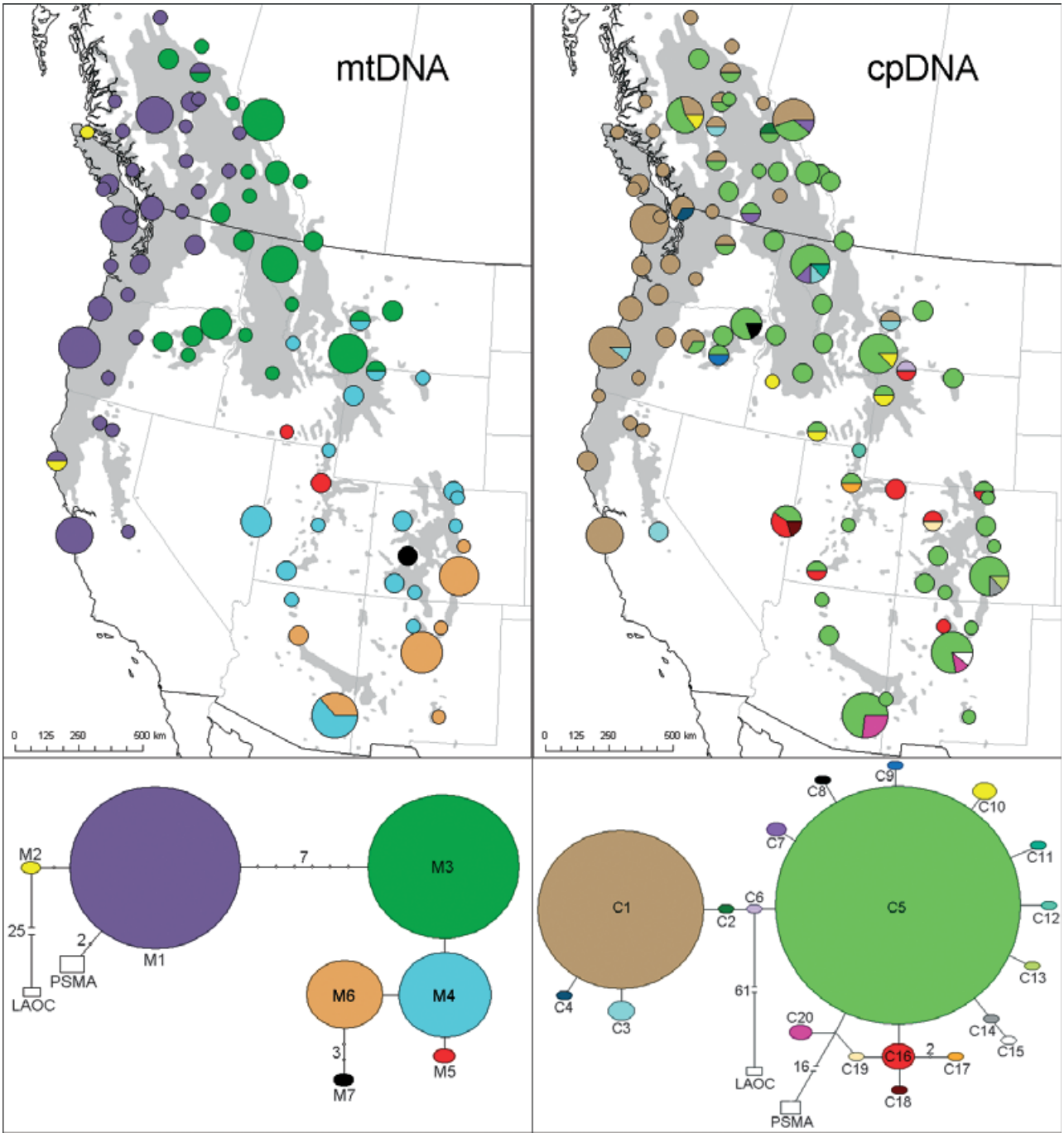

Fig. 4 Maps with sample sites coloured according to mitotype (left) and chlorotype (right) composition and proportioned according to sampling intensity. Below, parsimony networks show the relation of those mitotypes (left) and chlorotypes (right), with the size of the circles proportional to the frequency in the overall data set. Modern range of Douglas-fir is shown in grey (Little 1971).

molecular clock among species ($P = 0.034$; transversion model) and within *Pseudotsuga menziesii* chlorotypes ($P = 0.012$; F81; Felsenstein 1981). For mtDNA, a molecular clock was not rejected among species ($P = 0.25$; JC; Jukes & Cantor 1969) nor within *P. menziesii* ($P = 1$; F81; Table 2).

Genetic diversity and neutrality

Overall, haplotype diversity was similarly high for both mtDNA and cpDNA data ($H = 0.74$ and 0.64, respectively), and nucleotide diversity was low for both loci ($\pi = 0.00280$ and 0.00073; Table 3). Haplotype richness,

Table 2 Jukes–Cantor-corrected sequence divergence (d_{XY}) with standard deviation (SD), mutation rates among *Larix* and *Pseudotsuga*, and results of molecular clock tests among species and within *Pseudotsuga menziesii*

		mtDNA	cpDNA
d_{XY} (SD)		0.05260	0.04413
		(0.04332)	(0.03636)
μ_y*		5.26×10^{-10}	4.41×10^{-10}
μ_g†		5.26×10^{-8}	4.41×10^{-8}
μ_{IMa}‡		4.81×10^{-7}	6.66×10^{-7}
Insertion/deletion rate*		2.55×10^{-10}	1.49×10^{-10}
Model among species		JC	TVM
Molecular clock hypothesis	d.f.	2	2
	-2Λ	2.77	6.75
	P	0.25	0.034
Model within *P. menziesii*		F81	F81
Molecular clock hypothesis	d.f.	188	217
	-2Λ	64.60	266.63
	P	1	0.012

*Rate per site per year.
†Rate per site per generation (T = 100).
‡Rate per locus per year.
TVM, transversion model.

haplotype diversity and nucleotide diversity varied across populations and regions, with the coast having the lowest diversity by all measures (Table 3). However, no significant latitudinal trends were observed (mtDNA: $r_{hr\,|\,Lat} = -0.18$, $P = 0.49$; $r_{H\,|\,Lat} = -0.18$, $P = 0.50$; $r_{\pi\,|\,Lat} = 0.26$, $P = 0.34$; cpDNA: $r_{hr\,|\,Lat} = -0.10$, $P = 0.70$; $r_{H\,|\,Lat} = -0.01$, $P = 0.98$; $r_{\pi\,|\,Lat} = 0.14$, $P = 0.61$). Diversity was more strongly partitioned among populations in mtDNA ($G_{ST} = 0.73 \pm 0.07$) than in cpDNA ($G_{ST} = 0.35 \pm 0.09$).

The mtDNA data showed little evidence of violating the assumptions of neutrality. Only for the combined coastal data was a significantly negative D observed (Table 3), suggesting possible recent population expansion (though R_2 was not significant). Overall, the cpDNA data had a significantly negative D value, but this was inconsistent across regions and populations. The coast had a significantly negative D value, Canada showed no evidence of deviating from neutrality, and the northern and southern US Rockies had significant R_2 values. Within the Rockies, only two D and one R_2 were significant, and two of these were in the northern Rockies. Population expansion in the northern Rockies and coast is consistent with the fossil record for Douglas-fir (Gugger & Sugita submitted), suggesting that selection was less likely a factor.

Population structure

SAMOVA of mtDNA data showed that F_{CT} began to plateau at 0.92 for five to six groups of populations (Table 4). The five groups corresponded to the coast (populations 1–5), northern Rockies (7–10), central Rockies (11–13), southern Rockies (14–16) and northern British Columbia (6). The six groups were the same except northern Utah (11) was split from the central Rockies. The separation of northern British Columbia as a distinct group in both cases probably occurred because that population is a mixture of mitotypes from neighbouring populations that are fixed for divergent coastal and Rocky Mountain mitotypes. Less than 1% of the variation was among populations within groups, and 7.8% was within populations.

In the SAMOVA of cpDNA, the highest F_{CT} (0.68) was for two groups, corresponding to the coast (1–4) and Rockies (5–16; Table 4). In this case, central British Columbia (5) was grouped with the Rockies instead of coast. For cpDNA, 4.1% of the variation was among populations within groups and 27.5% was within populations. SAMOVA of cpDNA compared to mtDNA showed that in both cases most of the variation was explained among groups, but that proportion was lower in the cpDNA than in the mtDNA data. Also, cpDNA had a much higher percentage of variation within populations than mtDNA.

Testing hypotheses with coalescent methods

The median estimates of θ were 0.000491 (95% HPD: 0.000104, 0.001060) for mtDNA and 0.001252 (95% HPD: 0.000585, 0.002086) for cpDNA. These corresponded to N_{ef} of 4667 (989, 10 076) and N_{em} of 14 185 (6628, 23 635), which we used in coalescent simulations. Because these estimates seemed low for a common and widespread tree, we also ran simulations with the median N_e estimate times 10.

Overall, simulation results did not clearly point to a single best hypothesis (Fig. 5; Table 5). For the mtDNA, simulations at the lowest N_{ef} (95% HPD low) for all divergence times rejected all multiple-refugia hypotheses but not the single-refugium hypothesis. For the median N_{ef} estimate, the single-refugium hypothesis was rejected in favour of the two-refugia hypothesis, and the two-refugia hypothesis could not be rejected for divergence dating to the beginning of the Wisconsinan glaciation (115 ka; Table 5; Fig. 5). For the 95% HPD high N_{ef}, the two-refugia hypothesis was not rejected at any divergence time, and the one-refugium hypothesis was rejected in favour of two- and three-refugia hypotheses. At $N_{ef} \times 10$, no multiple-refugia hypothesis was rejected, and all one-refugium hypotheses were rejected in favour of their alternative multiple-refugia hypotheses. For the cpDNA, no single-refugium hypothesis could be rejected for any N_{em}. In addition,

Table 3 Diversity and neutrality measures for 16 populations, four regions, and overall data set

		mtDNA								cpDNA							
US Coast	Population	*n*	*h*	h_r*	*S*	*H* (SD)	π (SD)	*D*†	R_2†	*n*	*h*	h_r‡	*S*	*H* (SD)	π (SD)	*D*†	R_2†
1	Northern CA	12	2	1.4	1	0.167 (0.134)	0.00018 (0.000293)	−1.14	0.28	14	2	1.6	1	0.264 (0.136)	0.000175 (0.000229)	−0.34	*0.13*
2	Western OR	14	1	1.0	0	0	0	0	0	15	2	1.3	1	0.133 (0.112)	0.000088 (0.000156)	−1.16	0.25
3	Western WA	16	1	1.0	0	0	0	0	0	20	3	1.5	3	0.195 (0.115)	0.000198 (0.000241)	*−1.72*	0.16
	Overall: US coast	**42**	**2**	**2.0**	**1**	**0.048 (0.045)**	**0.000052 (0.000142)**	***−1.12***	**0.15**	**49**	**4**	**4.0**	**4**	**0.193 (0.074)**	**0.000158 (0.000205)**	***−1.67***	**0.08**
Canada																	
4	Western BC	9	2	1.6	1	0.222 (0.166)	0.000241 (0.000354)	−1.09	0.31	9	1	1.0	0	0	0	0	0
5	Central BC	14	1	1.0	0	0	0	0	0	17	4	2.5	4	0.596 (0.099)	0.000797 (0.000598)	0.06	0.15
6	Northern BC	6	2	2.0	4	0.533 (0.172)	0.002272 (0.001688)	−1.18	0.27	6	2	2.0	2	0.6 (0.129)	0.000793 (0.000674)	1.75	0.3
7	Eastern BC	19	2	1.3	4	0.105 (0.92)	0.000448 (0.000476)	−1.86	0.22	24	4	2.4	3	0.591 (0.081)	0.000697 (0.000534)	0.79	0.18
	Overall: Canada	**48**	**3**	**2.9**	**5**	**0.529 (0.026)**	**0.002192 (0.001388)**	**2.05**	**0.2**	**56**	**6**	**5.6**	**5**	**0.592 (0.036)**	**0.000761 (0.000554)**	**0.14**	**0.11**
Northern US Rockies																	
8	Northern US Rockies	15	2	1.3	1	0.133 (0.112)	0.000143 (0.000252)	−1.16	0.25	21	5	2.0	6	0.352 (0.131)	0.000378 (0.000356)	*−2.06*	0.12
9	Eastern OR	10	1	1.0	0	0	0	0	0	12	4	2.5	3	0.561 (0.154)	0.000511 (0.000450)	−0.73	0.14
10	Yellowstone area	17	2	1.9	1	0.441 (0.098)	0.000472 (0.000496)	0.95	0.22	18	6	2.6	4	0.562 (0.134)	0.000424 (0.000386)	−1.35	*0.09*
	Overall: Northern US Rockies	**42**	**2**	**2.0**	**1**	**0.251 (0.078)**	**0.000269 (0.000343)**	**0.11**	**0.13**	**51**	**10**	**9.8**	**8**	**0.413 (0.086)**	**0.000508 (0.000420)**	**−1.54**	***0.05***
Southern US Rockies																	
11	Northern UT	5	2	2.0	1	0.600 (0.175)	0.000642 (0.000704)	1.22	0.3	8	5	3.8	5	0.857 (0.108)	0.001015 (0.000771)	−0.92	0.15
12	NV/southern UT	8	1	1.0	0	0	0	0	0	8	3	2.6	2	0.679 (0.122)	0.000543 (0.000491)	0.24	0.22
13	Northern CO	6	1	1.0	0	0	0	0	0	7	3	2.7	2	0.667 (0.16)	0.000567 (0.000516)	0.21	0.23
14	Southern CO	14	3	2.4	3	0.56 (0.125)	0.000953 (0.000793)	−0.17	0.15	14	3	1.7	2	0.275 (0.148)	0.000189 (0.000239)	−1.48	0.17
15	AZ	13	2	2.0	1	0.538 (0.06)	0.000577 (0.000572)	1.48	0.27	14	2	1.8	2	0.363 (0.13)	0.000479 (0.000426)	0.42	0.18
16	NM	12	2	1.4	1	0.167 (0.134)	0.000178 (0.000289)	−1.14	0.28	12	4	2.3	5	0.455 (0.17)	0.000551 (0.000474)	*−1.83*	0.15
	Overall: Southern US Rockies	**58**	**4**	**3.9**	**4**	**0.589 (.033)**	**0.000796 (0.000663)**	**−0.31**	**0.09**	**63**	**11**	**8.2**	**12**	**0.520 (0.073)**	**0.000851 (0.000600)**	**−1.41**	***0.04***
Overall		**190**	**7**	**–**	**10**	**0.741 (0.014)**	**0.002804 (0.001663)**	**1.27**	**0.13**	**219**	**20**	**–**	**18**	**0.637 (0.024)**	**0.000733 (0.000532)**	***−1.65***	**0.03**

*Rarefaction to 5 for populations and 42 for regions.
†Italicized values were significant under coalescent simulations.
‡Rarefaction to 5 for populations and 49 for regions.

Table 4 SAMOVA results that gave highest F_{CT} for each locus. All variance components were significant ($P < 0.001$)

	d.f.	Sum of squares	Variance components	Percentage of variation
mtDNA ($K = 5$)				
Among groups	4	343.18	2.474	91.5
Among populations Within groups	11	4.74	0.018	0.7
Within pops.	174	36.89	0.212	7.8
mtDNA ($K = 6$)				
Among groups	5	344.51	2.471	91.8
Among populations within groups	10	3.41	0.01	0.4
Within pops.	174	36.89	0.212	7.9
cpDNA ($K = 2$)				
Among groups	1	92.47	1.067	68.3
Among populations within groups	14	18.17	0.065	4.1
Within pops.	203	87.36	0.43	27.5

all multiple-refugia hypotheses were rejected at N_{em}, 95% HPD low N_{em} and 95% HPD high N_{em}. For $N_{em} \times 10$, two- and three-refugia hypotheses were not rejected at any divergence, and four-refugia hypotheses were not rejected only at 115 ka. These general patterns of significance upheld when considering a generation time of 200 years (not shown).

Divergence time among varieties

We ignored the possibility that some populations may have undergone recent expansions (Table 3) to avoid more complicated models (e.g. IM, Hey & Neilson 2004). Also, we excluded Canadian populations, which were largely a mixture of the two varieties (Fig. 4), and several individuals near Canada that were presumed recent migrants from the range of one variety to the other during postglacial contact (three C1s and two C3s from the northern US Rockies and one C5 from northern Washington).

Divergence time among coastal and Rocky Mountain varieties could not be estimated under the full IMa model for both loci combined nor for each locus independently (posterior distribution never dropped to zero). However, per generation migration rates were very low for both loci combined, mtDNA, and cpDNA, and we could not reject the null hypothesis that both migration parameters equalled zero (Table 6). Moreover, each variety in our analysis was fixed for divergent haplotypes, suggesting migration among varieties in the US has not been pronounced for a long time. Therefore, we set $m_1 = m_2 = 0$ and repeated the analyses. After this adjustment, we were able to estimate divergence time of about 2.11 Ma with the 90% HPD interval ranging 4.37 Ma–755 ka, which spans the early Pliocene through middle Pleistocene (Table 6). Divergence time estimates from mtDNA and cpDNA differed substantially (4.41 Ma and 491 ka, respectively).

Discussion

The distribution and divergence of Douglas-fir populations (Fig. 4) is well explained by both Pliocene geology and late Quaternary climate cycles. The divergence of coastal and Rocky Mountain varieties coincided with the uprising of the Cascades and Sierra Nevada (Fig. 1a; Table 6). Much of the variation within the Rocky Mountains is probably the product of isolation and recontact induced by glacial–interglacial cycles (Figs 1, 5 and 6; Tables 4 and 5). Phylogeographic patterns of mtDNA and cpDNA variation were broadly consistent (Fig. 4; Tables 3 and 4), but differed in ways suggesting that gene flow via pollen dispersal might have played an important role in connecting populations that were otherwise isolated from seed exchange. The most striking example of this is the ongoing cpDNA introgression observed primarily in Canadian populations east of the Coast Range that probably began in the early Holocene (Fig. 4).

Divergence of coastal and Rocky Mountain varieties

The divergence of the coastal and Rocky Mountain Douglas-fir varieties as estimated by IMa (2.11 Ma; Table 6) is consistent with the Pliocene divergence hypothesis and the fossil record (Fig. 1b), but it does not exclude the possibility of divergence times well into the Pleistocene (90% HPD interval, 4.37 Ma–755 ka). Part of the reason for the large interval is that the mtDNA and cpDNA data when run separately gave quite different estimates (4.41 Ma and 491 ka, respectively; Table 6), and the mtDNA estimate itself has a very broad HPD interval. This discrepancy could be understood in two ways: (i) the estimates reflect the variability expected among different loci evolving and sorting under stochastic processes (Edwards & Beerli 2000; Arbogast *et al.* 2002; Carstens & Knowles 2007);

Table 5 *P* values less than 0.05 indicate empirical and simulated distributions of *s* differed significantly, rejecting the null hypothesis. The rejection of a null one-refugium hypothesis suggests the multiple-refugia alternative in parentheses, whereas the rejection of a null multiple-refugia hypothesis with divergence time in parentheses suggests no particular alternative. Tests suggesting multiple refugia are in bold

	mtDNA				cpDNA			
Null hypothesis	N_{ef} low	N_{ef}	N_{ef} high	$N_{ef} \times 10$	N_{em} low	N_{em}	N_{em} high	$N_{em} \times 10$
One (H_A: two)	1	**0.034**	**<0.001**	**<0.001**	1	1	1	0.985
One (H_A: three)	1	0.900	**0.023**	**<0.001**	1	1	1	0.942
One (H_A: four)	1	0.999	0.337	**<0.001**	1	1	1	0.995
Two (115 ka)	<0.001	**0.157**	**0.897**	**1**	<0.001	<0.001	<0.001	**0.947**
Two (190 ka)	<0.001	0.006	**0.386**	**1**	<0.001	<0.001	<0.001	**0.582**
Three (115 ka)	<0.001	<0.001	0.020	**1**	<0.001	<0.001	<0.001	**0.734**
Three (190 ka)	<0.001	<0.001	<0.001	**0.990**	<0.001	<0.001	<0.001	**0.131**
Four (115 ka)	<0.001	<0.001	<0.001	**0.996**	<0.001	<0.001	<0.001	**0.420**
Four (190 ka)	<0.001	<0.001	<0.001	**0.838**	<0.001	<0.001	<0.001	0.016

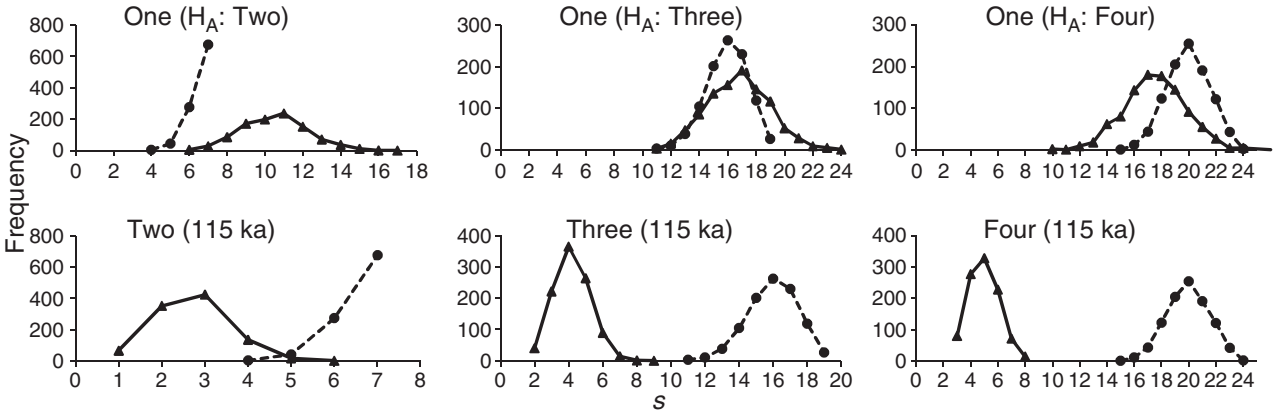

Fig. 5 Examples of distributions of *s* for simulated (solid line) and empirical (dashed line) trees for each refugial hypothesis for mtDNA data with N_{ef} = 4667 and a splitting time of 115 ka. Three- and four-refugia hypotheses were rejected, but the two-refugia hypothesis was not rejected. The one-refugium hypothesis was rejected in favour of the two-refugia hypothesis but was not rejected in favour of two- and three-refugia alternatives. *P* values are reported in Table 5.

Table 6 IMa results showing t, m_1 and m_2 for a run with the full model, followed by results of the likelihood ratio tests, and then t with its 90% HPD interval for a run with $m_1 = m_2 = 0$. In each case, the values for the run with the highest effective sample size (Hey & Neilson 2007) are shown

	Full model				$m_1 = m_2$		$m_1 = m_2 = 0$		Model without migration ($m_1 = m_2 = 0$)		
	Priors (m_1, m_2, q_1, q_2, q_A)	t (ka)	m_1	m_2	-2Λ	P	-2Λ	P	t (ka)	90% HPD low	90% HPD high
Both loci	**2.5. 0.5, 10, 10, 10**	**6580***	**0.0000055**	**0**	**$-0.1101 \approx 0$**	**1**	**0.001**	**0.49**	**2113**	**755**	**4372**
mtDNA	2.5, 1.5, 10, 10, 10	4360*	0.0000001	0	0.1846	0.67	$-0.0014 \approx 0$	≈ 0.5	4410	1813	8826
cpDNA	6, 2, 8, 30, 8	413*	0.0000002	0.0000001	0.4594	0.5	0.002	0.48	491	176	1832

*Posterior distribution for this parameter never dropped to zero, suggesting unreliable estimates.

or (ii) the estimates for each locus reflect different processes. Without data from more loci, it is conservative to assume the former; however, the latter has interesting implications, if true. For example, the mtDNA divergence time estimate of 4.41 Ma could reflect the time at which gene flow via seed dispersal was cut off,

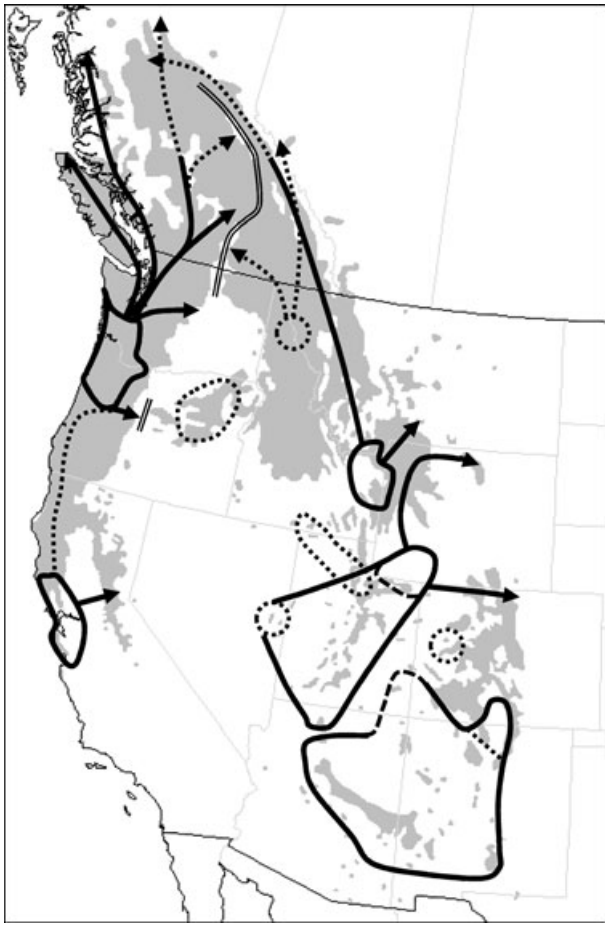

Fig. 6 Summary of glacial refugia and postglacial migration routes based on fossil and genetic data. Solid lines indicate refugia and postglacial migration routes supported by fossil and molecular data, dashed lines indicate those supported by fossil data only, and dotted lines indicate those supported by molecular data only. Double line marks some barriers to seed dispersal as suggested by mtDNA.

and the later cpDNA estimate of 491 ka could indicate that substantial gene flow via pollen continued until the middle Pleistocene. Under this scenario, pollen exchange could have been continuous over that interval or could have been in one large homogenizing burst just before final divergence.

However, additional uncertainties should be considered for all estimates. First, mtDNA mutations among varieties seemed to come in clusters that could have occurred as one event, and therefore, might not have fit the substitution model employed by IMa. The outgroup, *Pseudotsuga macrocarpa*, is positioned two mutational steps from the coastal variety (M1), but the coastal and Rocky Mountain varieties are seven mutational steps apart (Fig. 4). Under a molecular clock assumption, this implies that the varietal divergence preceded the species divergence, which is not supported by cpDNA (Fig. 4), morphology (Hermann & Lavender 1990) or the fossil record (Fig. 1; Hermann 1985). Overall, this could result in an overestimate of the divergence time among varieties based on mtDNA and both loci. Second and more generally, our mutation rate estimates may not be accurate. In calculating the per locus per year substitution rate used in IMa, we assumed a divergence time for *Pseudotsuga* and *Larix* of 50 Ma and that the mutation rate estimated among genera is applicable within species. The latter was assumed despite the fact that molecular clock model was rejected within and among species for the cpDNA (Table 2), and its effect on our conclusions is unknown. Otherwise, we expect the divergence time to be reasonably accurate based on the fossil record (LePage & Basinger 1991; Schorn 1994; Gernandt & Liston 1999) and other molecular data sets (Wang *et al.* 2000). However, if the true divergence time among genera was actually 100 Ma, for example, then our estimated mutation rates would have been halved and all our varietal divergence time estimates would have been doubled.

Pushing back the divergence date from 2.11 to 4.22 Ma for the analysis with both loci combined would not affect our main conclusion of divergence associated with the Cascade and Sierra Nevada orogeny and subsequent xerification of the Columbia Plateau and Great Basin between the present ranges of each variety. Moreover, our mutation rate estimates are consistent with the most recent fossil-calibrated neutral substitution rates for cpDNA observed in *Pinus* ($\mu_y = 2.2–4.2 \times 10^{-10}$; Willyard *et al.* 2007). In the Pacific Northwest, Douglas-fir shares this history with a number of similarly distributed taxa (Carstens *et al.* 2005), including tailed frogs (*Ascaphus* spp.; Neilson *et al.* 2001), Van Dyke's salamanders (*Plethodon* spp.; Carstens *et al.* 2004), the giant Pacific salamander (*Dicamptodon* spp.; Steele *et al.* 2005) and Constance's bittercress (*Cardamine constancei*; Brunsfeld & Sullivan 2005).

Glacial refugia

Glacial refugia for the coastal and Rocky Mountain varieties were distinct (Figs 1b, 4 and 6; Table 6). One chlorotype and one mitotype dominated the coastal region, and the few rare haplotypes were wide-ranging across the region. This low diversity and lack of substructure within the coastal variety is consistent with nuclear microsatellite and allozyme variation (Li & Adams 1989; Krutovsky *et al.* 2009) and suggests we cannot reject the hypothesis of one coastal refugium spanning central California to unglaciated western Washington. One refugium is also consistent with models that used modern climatic tolerances to predict Douglas-fir's distribution during the Last Glacial Maxi-

mum (Bartlein *et al.* 1998). Even so, the lack of structure is surprising considering the strong suggestion of two refugia in the fossil record (Fig. 1b; Gugger & Sugita submitted) and that leaf terpene chemistry differs in California compared to Oregon/Washington (von Rudloff 1972; Zavarin & Snajberk 1973; von Rudloff & Rehfeldt 1980; Critchfield 1984). Moreover, a diverse array of organisms form contact zones near the California/Oregon border (Soltis *et al.* 1997; Swenson & Howard 2005). However, given the very limited number of haplotypes observed in the region, we cannot exclude the possibility of multiple coastal refugia that more variable molecular markers with the signature of more recent events might reveal.

For the Rocky Mountains, in contrast, multiple glacial refugia were likely (Fig. 6). Coalescent tests for mtDNA and high N_{ef} estimates (95% HPD high and $N_{ef} \times 10$) tended to reject the single-refugium hypotheses in favour of multiple-refugia alternatives (Fig. 5; Table 5). Some two-refugia hypotheses were not rejected, and at $N_{ef} \times 10$ no multiple-refugia hypothesis was rejected. The tendency to not reject two-refugia hypotheses lends some support to the two refugia proposed by Li & Adams (1989) based on allozyme data. Nonetheless, there was limited power to distinguish among multiple-refugia hypotheses with the mtDNA coalescent tests, especially considering that most multiple-refugia hypotheses were not rejected at lower N_{ef}.

Aside from some coalescent tests, other evidence supports at least three glacial refugia. For example, mtDNA SAMOVA defines three to four Rocky Mountain population clusters: northern, southern, and one to two central (Table 4). The southern and northern mtDNA SAMOVA groups conform well to the refugia identified in the fossil record, and fall along topographic barriers such as the Grand Canyon and the high peaks of the Colorado Rockies and Yellowstone region that harboured extensive mountain glaciers at the Last Glacial Maximum and would have formed formidable barriers to seed dispersal (Figs 1b and 6; Gugger & Sugita submitted).

The broadly defined northern refugium indentified by mtDNA SAMOVA is further supported by, and perhaps further subdivided by, uniquely northern chlorotypes (C6–C11). In particular, C10 and C6 (private to the Yellowstone region) support the contention that low fossil *Pseudotsuga* pollen abundances during the full glacial (Baker 1983; Whitlock 1993) could represent a refugium in the region (Gugger & Sugita submitted). Furthermore, the distribution of C7, which spans northern Montana to east-central British Columbia and C11, which is private to northern Montana, could reflect a small refugium even farther north, with which fossil data are not available to compare (Fig. 6). Finally, two private chlorotypes (C8, C9) in eastern Oregon suggest that this geographically distinct region could have also been a distinct glacial refugium.

Based on other phylogeographic data, the northern Rockies are thought to have been a refugium for diverse mesic plant (Richardson *et al.* 2002; Brunsfeld & Sullivan 2005; Brunsfeld *et al.* 2007; Godbout *et al.* 2008) and animal taxa (Neilson *et al.* 2001; Carstens *et al.* 2004; Steele *et al.* 2005). Moreover, this is consistent with fossil evidence for a Douglas-fir refugium close to the ice margin in western Washington (Barnosky 1981, 1985; Gugger & Sugita submitted). Molecular phylogeographic studies and reinterpretations of the fossil record for a variety of North American and European taxa have revealed cryptic northern refugia (McLachlan *et al.* 2005; Anderson *et al.* 2006; Petit *et al.* 2008; Hu *et al.* 2009; Gugger & Sugita submitted). Douglas-fir could have survived on the unglaciated portions at the base of mountain ranges and ice-free ridges in the northern Rockies (Thompson *et al.* 1993; Whitlock 1995).

Similarly, the southern region defined by mtDNA SAMOVA and the distribution of some chlorotypes (C13–C15, C20) fit with expectations of a southern refugium from the fossil record (Fig. 1b). Again, chlorotypes suggest that the southern Rocky Mountain region may have been subdivided into multiple small refugia, consistent with isolation on distinct mountain chains.

Defining the number and extent of refugia in the central Rockies is less clear. Although at least one to two central Rocky Mountain refugia are observed in both fossil data (Fig. 1b) and the mtDNA SAMOVA groupings, their boundaries do not align as expected (Fig. 6). For example, the northern Utah group appears to have a distribution associated with the north and east shores of glacial Lake Lahontan (present-day Great Salt Lake), and the other central group populations are less well defined, spanning Nevada and the North Rim of the Grand Canyon to northern Colorado and Wyoming. In contrast, the limited fossil record in the region suggested that southern Utah was distinct from the remaining central Rocky Mountain populations (Gugger & Sugita submitted). The primarily central Rocky Mountain cpDNA clade (C16–C19) might also support a single central refugium. However, private chlorotypes in isolated mountain chains in Nevada (C18), northern Utah (C12, C17) and western Colorado (C19), and a private, genetically divergent mitotype (M7) in western Colorado suggest a more complicated scenario with multiple small refugia scattered throughout the region.

Overall, the absence of southern and central mitotypes and chlorotypes in the north supports the hypothesis that migration was primarily elevational (Gugger & Sugita submitted), not latitudinal, in the Southwest, similar to white fir (*Abies concolor*), bristlecone pine

(*Pinus longaeva*), limber pine (*Pinus flexilis*, Mitton *et al.* 2000) and many other taxa in the region (Maher 1961, 1963; Spaulding *et al.* 1983; Hall 1985).

Differences in mitotype and chlorotype distributions that may reflect differences in seed and pollen dispersal are apparent. Coalescent tests with mtDNA rejected one-refugium hypotheses in many cases, whereas tests with cpDNA never rejected a one-refugium hypothesis (Table 5). Similarly, mtDNA SAMOVA split the Rockies into northern, central and southern groups, whereas cpDNA SAMOVA did not split the Rockies at all, even though some chlorotypes have northern, central or southern Rocky Mountain distributions. The sharing of chlorotypes across boundaries defined by mitotypes could reflect exchange of pollen, but not seed, among regions. This pattern is widely observed in studies of conifers that compare maternally and paternally inherited organellar DNA data (Dong & Wagner 1994; Tsumura & Suyama 1998; Latta & Mitton 1999; Liepelt *et al.* 2002; Burban & Petit 2003; Petit *et al.* 2005; Jaramillo-Correa *et al.* 2006; Du *et al.* 2009).

Among other reasons (Knowles & Maddison 2002), this cautions against overinterpreting the location of refugia from haplotype distributions alone. However, our analysis poses hypotheses that could be tested with new fossil data from those regions (Fig. 6).

Finally, we caution that much of the observed sequence variation could predate the last glacial cycle. The mutation rates for cpDNA and mtDNA suggest one base substitution per locus per 800 000 to 1 million years (Table 2). If so, it is striking that any regional structure does exist, given the repeated range expansions and contractions expected across multiple glacial cycles. That it does suggests that regions may have repeatedly served as refugia across multiple climate cycles.

Postglacial migration, Holocene secondary contact and introgression

The postglacial colonization of Canada is characterized by the northward expansion and secondary contact of Douglas-fir's two long-isolated varieties (Figs 4 and 6; Table 3). Maternally inherited (i.e. seed-dispersed) mtDNA indicates that the Rocky Mountain variety migrated primarily along the west slope of the Canadian Rockies, while the coastal variety spread north along the coast and northeast into the interior of British Columbia, east of the Coast Range. Both varieties seem to have stopped at the same barrier, which from north to south, begins as the Rocky Mountain Trench, then follows the Columbia River downstream until Revelstoke, BC, where it veers west (possibly along the route of the Trans-Canada Highway), until regaining a southward trajectory along the Okanagan Valley into central Washington (Figs 4 and 6). Much of this narrow barrier is set apart from the surrounding areas by its low elevation, long north-south lakes, and relatively dry conditions favouring grass communities. There is some evidence of intermixing in the northernmost part of the range in British Columbia, where the Trench is less sharply cut into the landscape, the terrain settles to the more uniform Interior Plateau, and Douglas-fir inhabits lower elevations.

Pollen appears to have readily dispersed across the seed dispersal barrier and spread throughout British Columbia east of the Coast Range, generating a broad hybrid zone 450 km wide. Chloroplast DNA introgression was bidirectional among varieties, despite the prevailing westerly wind. Evidence for this effect is clear in Fig. 4, where coastal and Rocky Mountain chlorotypes are common on both sides of the barrier to seed dispersal as defined by mtDNA. Additional support comes from the higher genetic diversity in the contact zone (central, northern and eastern BC) compared to surrounding 'pure' regions (western BC, northern US Rockies; Table 3), and the fact that SAMOVA of mtDNA grouped the central British Columbian population with the coast, whereas the cpDNA SAMOVA grouped it with the northern Rockies. Overall, this process could not have begun prior to the start of the Holocene (11.7 ka) because ice still covered much of British Columbia, but the precise contact time remains unclear for lack of fossil sites in the area (Gugger & Sugita submitted).

Some have argued that the hybrid zone extends down the east slope of the Cascades, possibly due to the southward migration of the Rocky Mountain variety (Sorensen 1979; St Clair *et al.* 2005). The existence of coastal, but no Rocky Mountain, mitotypes in the east Cascades of Washington and Oregon argues against southward migration and in favour of migration from the coast (Fig. 4). However, the east Cascades site in Washington contains a mixture of chlorotypes from both varieties, suggesting hybridization by pollen dispersal from the Rockies. This could explain the similarity of phenology and growth traits of those populations to the Rocky Mountain variety (St Clair *et al.* 2005). Interestingly, allozymes (Li & Adams 1989) and putatively neutral nuclear single-nucleotide polymorphisms (SNPs) display weak differentiation across the crest of Cascades in Washington, but some SNPs associated with cold-hardiness traits are strongly differentiated (Eckert *et al.* 2009). This might suggest only limited gene flow from the Rocky Mountain variety to the east Cascades and subsequent strong selection for cold-adapted Rocky Mountain alleles. Douglas-fir did not become abundant in the east Cascades until relatively recently (early to mid-Holocene; Whitlock & Bartlein 1997), consistent with the idea that opportunities for gene flow among varieties may have been limited. In

Oregon, adaptive traits among varieties gradually transition from the Oregon Cascades to the Blue Mountains in eastern Oregon (Sorensen 1979), consistent with allozyme variation (Li & Adams 1989). However, studies of mitochondrial randomly amplified polymorphic DNA showed a sharp division among varieties in the region (Aagaard *et al.* 1995). Our data reconcile this discrepancy by showing that western and eastern Oregon have distinct maternal origins (mtDNA), but that the coastal variety in the Cascades made paternal contributions (cpDNA) to the Rocky Mountain variety in the Blue Mountains (Fig. 4).

Loss of diversity along migration routes is thought to depend on the width of the migration front, migration distance, the frequency of long-distance dispersal, and the presence of topographic obstacles (Davies *et al.* 2004; Bialozyt *et al.* 2006). European trees show severe losses in diversity at high latitudes (Petit *et al.* 2003) due to the action of founder effects from rare long-distance dispersal (Petit *et al.* 1997) and topographic barriers (Mátyás & Sperisen 2001) along narrow migration corridors. Those patterns have not been observed in eastern North America, where broad migration fronts lacking major topographic obstacles may have preserved diversity at northern range limits (Magni *et al.* 2005; McLachlan *et al.* 2005; Gugger *et al.* 2008). Given the relatively narrow, topographically complex corridors along the west coast and northern Rocky Mountains, it is somewhat surprising that Douglas-fir shows no significant loss of diversity with latitude (Fig. 4; Table 3). This pattern cannot be explained alone by the admixture of the two varieties (Walter & Epperson 2001; Petit *et al.* 2002b) because almost all haplotypes from northern refugia of each variety are found in Canada. Instead, either dispersal and topographic bottlenecks were not important, or low diversity in the region precluded our ability to observe this effect.

With the postglacial migration pathways defined by mtDNA and cpDNA, we are able to revise postglacial migration rate estimates previously based solely on fossil evidence. Radiocarbon-dated fossil records suggest that Douglas-fir migration tracked the ice sheets as they began to recede starting about 18 ka (Gugger & Sugita submitted). The precise timing of northernmost colonization is not known for lack of fossil sites, but a site whose dating was approximated based on correlations with other nearby radiocarbon dated sites suggests the coastal variety could have reached the present-day northern range limit as early as 9 ka (Heusser 1960). Dividing migration distance from refugium to northernmost site by time, we estimate a mean migration rate from coastal sources near the ice margin to central British Columbia of about 55–110 m/year, depending on the time of northernmost colonization (present-day or 9 ka). From northern Rocky Mountain sources to central British Columbia, the rate was about 60–165 m/year, where the lower bound is estimated from a present-day colonization of the northern limit from a refugium near the ice margin (limited cpDNA evidence, no fossil evidence) and the upper bound is estimated from colonization of the northern limit 9 ka from a source near Yellowstone (mtDNA, cpDNA and limited fossil evidence). These agree with recent estimates from the fossil record (50–220 m/year; Gugger & Sugita submitted) and are consistent with migration rates estimated from modern seed dispersal models (80 m/year; Thompson & Schorn 1998). However, they are substantially less than earlier estimates (450 m/year; Tsukada 1982), and more importantly, they are far lower than migration rates required to keep pace with projected 21st century climate change in some regions. For example, Douglas-fir is expected to disappear from much of its present range in Mexico, parts of the southern Rockies and coast, and central British Columbia, but could survive in parts of the Great Basin and northern British Columbia/southeast Alaska (Shafer *et al.* 2001). To colonize those regions in the next 100 years could require a migration rate greater than 1000 m/year, far higher any estimate considered probably naturally. This suggests human-assisted migration could be considered in some regions to help Douglas-fir reach some potential future habitats (McLachlan *et al.* 2007).

Synergy of fossil and molecular data

Combining fossil and molecular data provides insight into the geological and climatological causes of and patterns of population divergence that neither can provide independently (McLachlan *et al.* 2005; Magri *et al.* 2006; Hu *et al.* 2009). Here, we demonstrated that molecular data can be used to test alternative hypotheses suggested by the fossil record, which poorly resolves populations. When molecular and fossil data sets agree, we can attribute patterns of modern population structure to the timing of geologic and climatologic processes implied by fossils. When they disagree, we are forced to reassess our understanding of the causes or consequences of population divergence. Discordance among data sets was limited. However, lack of molecular variation on the coast prevented formal tests of expected population structure there. We identified concordance of molecular and fossil data sets for the Pliocene divergence of varieties, some of the multiple Pleistocene refugia in the Rockies, and the general postglacial migration patterns (Fig. 6). Molecular data added value by constraining the range of postglacial migration rate estimates and generating new hypotheses for possible additional glacial refugia in eastern Oregon, the central

Rockies and near the ice margin at the Last Glacial Maximum in the northern US Rockies. Future fossil pollen studies in each region could test those hypotheses. Finally, molecular data exposed an interesting hybrid zone in Canada driven by paternal gene flow, a pattern that could not have been shown from fossil data alone.

Acknowledgements

We thank B.C. Jaquish, M.U. Stoehr, W.M. Illick, J.S. McLachlan, and S.P. Tarsitano for assistance collecting samples; K. Koehler, and J. Vevea for laboratory support and assistance; A. Dyke and L. Robertson for ice sheet data used in Fig. 1; F.K. Barker and R.M. Zink for help with analyses; M.B. Bowen and the Cornell University Computational Biology Service Unit for computing resources; and F.S. Hu, G. Quiram, J.A. Savage, R.G. Shaw, P. Tiffin, R.M. Zink, and one anonymous reviewer for providing constructive comments that improved this manuscript. This work was supported by the National Science Foundation Graduate Research Fellowship, University of Minnesota Doctoral Dissertation Fellowship, and the following grants from the University of Minnesota: Dayton and Wilkie Natural History Research Grants, Florence Rothman Fellowship, Carolyn M. Crosby Fellowship, Elmer C. Birney Fellowship, Center for Community Genetics Research Grant, Thesis Research Grant, and Block Grants. Computational Biology Service Unit is partially funded by Microsoft Corporation.

References

Aagaard JE, Volmer SS, Sorensen FC, Strauss SH (1995) Mitochondrial DNA products among RAPD profiles are frequent and strongly differentiated between races of Douglas-fir. *Molecular Ecology*, **4**, 441–447.

Adam DP, Vagenas GL (1990) Pollen data for a 3-m.y. core record from Tulelake, Siskiyou County, California. *United States Geological Survey Open-File Report*, **90–65**, 1–307.

Anderson LL, Hu F-S, Nelson DM, Petit RJ, Paige KN (2006) Ice-age endurance: DNA evidence of a white spruce refugium in Alaska. *Proceedings of the National Academy of Sciences, USA*, **103**, 12447–12450.

Arbogast BS, Edwards SV, Wakeley J, Beerli P, Slowinski JB (2002) Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annual Review of Ecology and Systematics*, **33**, 707–740.

Axelrod DI (1966) The Eocene Copper Basin flora of northeastern Nevada. *University of California Publications in Geological Science*, **59**, 1–83.

Axelrod DI (1983) New Pleistocene conifer records, coastal California. *University of California Publications in Geological Science*, **107**, 1–108.

Axelrod DI (1990) Environment of the Middle Eocene (45 Ma) Thunder Mountain flora, central Idaho. *National Geographic Research*, **6**, 355–361.

Baker RG (1983) Holocene vegetational history of the western United States. In: *Late-Quaternary Environments of the United States: The Holocene* (ed. Wright HE Jr), pp. 109–127. University of Minnesota Press, Minneapolis, MN.

Barnosky CW (1981) A record of late Quaternary vegetation from Davis Lake, southern Puget lowland, Washington. *Quaternary Research*, **16**, 221–239.

Barnosky CW (1985) Late Quaternary vegetation near Battle Ground Lake, southern Puget Trough, Washington. *Geological Society of America Bulletin*, **96**, 263–271.

Bartlein PJ, Anderson KH, Anderson PM *et al.* (1998) Paleoclimate simulations for North America over the past 21,000 years: features of the simulated climate and comparisons with paleoenvironmental data. *Quaternary Science Reviews*, **17**, 549–585.

Beiswenger JM (1991) Late Quaternary vegetational history of Grays Lake, Idaho. *Ecological Monographs*, **61**, 165–182.

Bennett KD, Provan J (2008) What do we mean by 'refugia'? *Quaternary Science Reviews*, **27**, 2449–2455.

Bialozyt R, Ziegenhagen B, Petit RJ (2006) Contrasting effects of long distance seed dispersal on genetic diversity during range expansion. *Journal of Evolutionary Biology*, **19**, 12–20.

Birky WC Jr (2001) The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annual Review of Genetics*, **35**, 125–148.

Braukmann TWA, Kuzmina M, Stefanović S (2009) Loss of all plastid *ndh* genes in Gnetales and conifers: extent and evolutionary significance for the seed plant phylogeny. *Current Genetics*, **55**, 323–337.

Brunsfeld SJ, Sullivan J (2005) A multi-compartmented glacial refugium in the northern Rocky Mountains: evidence from the phylogeography of *Cardamine constancei* (Brassicaceae). *Conservation Genetics*, **6**, 895–904.

Brunsfeld SJ, Sullivan J, Soltis DE, Soltis PS (2001) Comparative phylogeography of northwestern North America: a synthesis. In: *Integrating Ecology and Evolution in a Spatial Context* (eds Silvertown J, Antonovics J), pp. 319–339. Blackwell Publishing, Williston, Vermont.

Brunsfeld SJ, Miller TR, Carstens BC (2007) Insights into the biogeography of the Pacific Northwest of North America: evidence from the phylogeography of *Salix melanopsis*. *Systematic Botany*, **32**, 129–139.

Burban C, Petit RJ (2003) Phylogeography of maritime pine inferred with organelle markers having contrasted inheritance. *Molecular Ecology*, **12**, 1487–1495.

Carstens BC, Knowles LL (2007) Shifting distributions and speciation: species divergence during rapid climate change. *Molecular Ecology*, **16**, 619–627.

Carstens BC, Richards CL (2007) Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution*, **61**, 1439–1454.

Carstens BC, Stevenson AL, Degenhardt JD, Sullivan J (2004) Testing nested phylogenetic and phylogeographic hypotheses in the *Plethodon vandykei* species group. *Systematic Biology*, **53**, 781–792.

Carstens BC, Brunsfeld SJ, Dembroski JR, Good JM, Sullivan J (2005) Investigating the evolutionary history of the Pacific Northwest mesic forest ecosystem: hypothesis testing within a comparative phylogeographic framework. *Evolution*, **59**, 1639–1652.

Chaw S-M, Zharkikh A, Sung H-M, Lau T-C, Li W-H (1997) Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNA sequences. *Molecular Biology and Evolution*, **14**, 56–68.

Chaw S-M, Parkinson CL, Cheng Y, Vincent TM, Palmer JD (2000) Seed plant phylogeny inferred from all three plant genomes: monophyly of extant gymnosperms and origin of Gnetales from conifers. *Proceedings of the National Academy of Sciences, USA*, **97**, 4086–4091.

Chernoff H (1954) On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics*, **25**, 573–578.

Chiu W-L, Sears BB (1985) Recombination between chloroplast DNAs does not occur in sexual crosses of *Oenothera*. *Molecular and General Genetics*, **198**, 525–528.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1660.

Critchfield WB (1984) Impact of the Pleistocene on the genetic structure of North American conifers. In: *Proceedings of the Eighth North American Forest Biology Worksho* (ed. Lanner RM). pp. 70–118. Logan, Utah, USA.

Davies S, White A, Lowe A (2004) An investigation into effects of long-distance seed dispersal on organelle population genetic structure and colonization rate: a model synthesis. *Heredity*, **93**, 566–576.

Davis MB (1976) Pleistocene biogeography of temperate deciduous forests. *Geoscience and Man*, **13**, 13–26.

Davis MB (1981) Quaternary history and the stability of forest communities. In: *Forest Succession: Concepts and Application* (eds West DC, Shugart HH, Botkin DB), pp. 132–153. Springer-Verlag, New York.

Davis MB (1983) Holocene vegetational history of the eastern United States. In: *Late-Quaternary Environments of the United States: The Holocene* (ed. Wright HE Jr), pp. 166–181. University of Minnesota Press, Minneapolis, Minnesota.

Davis OK, Moutoux TE (1998) Tertiary and quaternary vegetation history of the Great Salt Lake, Utah, USA. *Journal of Paleolimnology*, **19**, 417–427.

Dong J, Wagner DB (1994) Paternally inherited chloroplast DNA polymorphism in *Pinus*: estimation of diversity and population subdivision, and tests of disequilibrium with a maternally inherited mitochondrial polymorphism. *Genetics*, **136**, 1187–1194.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214–221.

Du FK, Petit RJ, Liu J-Q (2009) More introgression with less gene flow: chloroplast vs. mitochondrial DNA in the *Picea asperata* complex in China, and a comparison with other Conifers. *Molecular Ecology*, **18**, 1396–1407.

Duff RJ, Nickrent DL (1997) Characterization of mitochondrial small-subunit ribosomal RNAs from holoparasitic plants. *Journal of Molecular Evolution*, **45**, 631–639.

Duff RJ, Nickrent DL (1999) Phylogenetic relationships of land plants using mitochondrial small-subunit rDNA sequences. *American Journal of Botany*, **86**, 372–386.

Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology*, **11**, 2571–2581.

Eckert AJ, Bower AD, Wegrzyn JL *et al.* (2009) Association genetics of coastal Douglas fir (*Pseudotsuga menziesii* var. *menziesii*, Pinaceae). *I. Cold-hardiness related traits*. *Genetics*, **182**, 1289–1302.

Edwards SV, Beerli P (2000) Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution*, **54**, 1839–1854.

Excoffier L, Laval G, Schneider S (2005) Arlequin version 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution*, **17**, 368–376.

Franklin JF, DeBell DS (1988) Thirty-six years of tree population change in an old-growth *Pseudotsuga-Tsuga* forest. *Canadian Journal of Forest Research*, **18**, 633–639.

Gernandt DS, Liston A (1999) Internal transcribed spacer region evolution in *Larix* and *Pseudotsuga* (Pinaceae). *American Journal of Botany*, **86**, 711–723.

Godbout J, Jaramillo-Correa JP, Beaulieu J, Bousquet J (2005) A mitochondrial DNA minisatellite reveals the postglacial history of jack pine (*Pinus banksiana*), a broad-range North American conifer. *Molecular Ecology*, **14**, 3497–3512.

Godbout J, Fazekas A, Newton CH, Yeh FC (2008) Glacial vicariance in the Pacific Northwest: evidence from a lodgepole pine mitochondrial DNA minisatellite for multiple genetically distinct and widely separated refugia. *Molecular Ecology*, **17**, 2463–2475.

Gugger PF, Sugita S (submitted) Glacial refugia and postglacial migration of Douglas-fir based on fossil pollen and macrofossil evidence. *Quaternary Science Reviews*.

Gugger PF, McLachlan JS, Manos PS, Clark JS (2008) Inferring long-distance dispersal and topographic barriers during post-glacial colonization from the genetic structure of red maple (*Acer rubrum* L.) in New England. *Journal of Biogeography*, **35**, 1665–1673.

Hall SA (1985) Quaternary pollen analysis and vegetational history of the Southwest. In: *Pollen Records of Late-Quaternary North American Sediment* (eds Bryant VM Jr, Holloway RG). pp. 95–124. American Association of Stratigraphic Palynologists Foundation, Dallas, Texas.

Hann DW, Marshall DD, Hanus ML (2003) *Equations for predicting height-to-crown-base, 5-year diameter-growth rate, 5-year height-growth rate, 5-year mortality rate, and maximum size-density trajectory for Douglas-fir and western hemlock in the coastal region of the Pacific Northwest*. Research Contribution 40. Forest Research Laboratory, Oregon State University, Corvallis.

Hasegawa MH, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, **22**, 160–174.

Hermann RK (1985) *The Genus Pseudotsuga: Ancestral History and Past Distribution*. Forest Research Laboratory, Oregon State University, Corvallis, Oregon.

Hermann RK, Lavender DP (1990) Douglas-fir. In: *Silvics of North America: Conifers* (eds Burns RM, Honkala BH), pp. 1080–1108. USDA Forest Service, Washington, District of Columbia.

Heusser CJ (1960) Late-Pleistocene environments of north Pacific North America: an elaboration of late-glacial and postglacial climatic, physiographic, and biotic changes. *American Geographical Society Special Publication*, **35**, 1–308.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hey J, Neilson R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, **167**, 747–760.

Hey J, Neilson R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences, USA*, **104**, 2785–2790.

Hu F-S, Hampe A, Petit RJ (2009) Paleoecology meets genetics: deciphering past vegetational dynamics. *Frontiers in Ecology and the Environment*, **7**, 371–379.

Hudson RR (1990) Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology*, **7**, 1–44.

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*, **111**, 147–164.

Huntley B, Birks HJB (1983) *An Atlas of Past and Present Pollen Maps for Europe: 0–13000 Years Ago*. Cambridge University Press, Cambridge.

Hurlbert SH (1971) The nonconcept of species diversity: a critique and alternative parameters. *Ecology*, **52**, 577–586.

Jaramillo-Correa JP, Bousquet J (2005) Mitochondrial genome recombination in the zone of contact between two hybridizing conifers. *Genetics*, **171**, 1951–1962.

Jaramillo-Correa JP, Bousquet J, Beaulieu J, Isabel N, Perron M, Bouillé M (2003) Cross-species amplification of mitochondrial DNA sequence-tagged-site markers in conifers: the nature of polymorphism and variation within and among species in *Picea*. *Theoretical and Applied Genetics*, **106**, 1353–1367.

Jaramillo-Correa JP, Beaulieu J, Bousquet J (2004) Variation in mitochondrial DNA reveals multiple distant glacial refugia in black spruce (*Picea mariana*), a transcontinental North American conifer. *Molecular Ecology*, **13**, 2735–2747.

Jaramillo-Correa JP, Beaulieu J, Ledig FT, Bousqet J (2006) Decoupled mitochondrial and chloroplast DNA population structure reveals Holocene collapse and population isolation in a threatened Mexican-endemic conifer. *Molecular Ecology*, **15**, 2787–2800.

Jukes TH, Cantor C (1969) Evolution of protein molecules. In: *Mammalian Protein Metabolism* (ed. Munro MN), pp. 21–132. Academic Press, New York.

Kimura M (1969) The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics*, **61**, 893–903.

Knowles LL (2001) Did Pleistocene glaciations promote divergence? Tests of explicit refugial models in montane grasshoppers. *Molecular Ecology*, **10**, 691–701.

Knowles LL, Maddison WP (2002) Statistical phylogeography. *Molecular Ecology*, **11**, 2623–2635.

Krutovsky KV, St Clair JB, Saich R, Hipkins VD, Neale DB (2009) Estimation of population structure in coastal Douglas-fir [*Pseudotsuga menziesii* (Mirb.) Franco var. *menziesii*] using allozyme and microsatellite markers. *Tree Genetics and Genomes*, **5**, 641–658.

Lakhanpal RE (1958) The Rujuda flora of west central Oregon. *University of California Publications in Geological Science*, **35**, 1–66.

Latta RG, Mitton JB (1999) Historical separation and present gene flow through a zone of secondary contact in ponderosa pine. *Evolution*, **53**, 769–776.

LePage BA, Basinger JF (1991) A new species of *Larix* (Pinaceae) from the early Tertiary of Axel Heiberg Island, Arctic Canada. *Review of Palaeobotany and Palynology*, **70**, 89–111.

Li P, Adams WT (1989) Range-wide patterns of allozyme variation in Douglas-fir (*Pseudotsuga menziesii*). *Canadian Journal of Forest Research*, **19**, 149–161.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Liepelt S, Bialozyt R, Ziegenhagen B (2002) Wind-dispersed pollen mediates postglacial gene flow among refugia. *Proceedings of the National Academy of Sciences, USA*, **99**, 14590–14594.

Lisiecki LE, Raymo ME (2005) A Pliocene–Pleistocene stack of 57 globally distributed benthic $\delta^{18}O$ records. *Paleooceanography*, **20**, PA1003.

Little EL Jr (1971) *Atlas of United States Trees, Volume 1: Conifers and Important Hardwoods*. US Department of Agriculture, Washington, District of Columbia.

Magni CR, Ducousso A, Caron H, Petit RJ, Kremer A (2005) Chloroplast DNA variation of *Quercus rubra* L. in North America and comparison with other Fagaceae. *Molecular Ecology*, **14**, 2.

Magri D, Vendramin GG, Comps B *et al*. (2006) A new scenario for the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. *New Phytologist*, **171**, 199–221.

Magri D, Fineschi S, Bellarosa R *et al*. (2007) The distribution of *Quercus suber* chloroplast haplotypes matches palaeogeographical history of the western Mediterranean. *Molecular Ecology*, **16**, 5259–5266.

Maher Jr LJ (1961) *Pollen analysis and postglacial vegetation history in the Animas Valley Region, Southern San Juan Mountains, Colorado*. PhD Dissertation, University of Minnesota, Minneapolis, Minnesota.

Maher LJ Jr (1963) Pollen analysis of surface materials from the southern San Juan Mountains, Colorado. *Geological Society of America Bulletin*, **74**, 1485–1504.

Marshall KA, Neale DB (1992) The inheritance of mitochondrial DNA in Douglas-fir (*Pseudotsuga menziesii*). *Canadian Journal of Forest Research*, **22**, 73–75.

Martin PS (1963) Geochronology of pluvial Lake Cochise, southern Arizona. *II. Pollen analysis of a 42-meter core*. *Ecology*, **44**, 436–444.

Mathiasen RL, Hawksworth FG, Edminster CB (1990) Effects of dwarf mistletoe on growth and mortality of Douglas-fir in the Southwest. *Great Basin Naturalist*, **50**, 2, 173–179.

Mátyás G, Sperisen C (2001) Chloroplast DNA polymorphisms provide evidence for postglacial re-colonisation of oaks (*Quercus* spp.) across the Swiss Alps. *Theoretical and Applied Genetics*, **102**, 12–20.

McLachlan JS, Clark JS, Manos PS (2005) Molecular indicators of tree migration capacity under rapid climate change. *Ecology*, **86**, 2088–2098.

McLachlan JS, Hellman JJ, Schwartz MW (2007) A framework for debate of assisted migration in an era of climate change. *Conservation Biology*, **21**, 297–302.

Mehringer Jr PJ, Haynes CV Jr (1965) The pollen evidence for the environment of early man and extinct mammals at the Lehner Mammoth Site, southeastern Arizona. *American Antiquity*, **31**, 17–23.

Mitton JB, Kreiser BR, Latta RG (2000) Glacial refugia of limber pine (*Pinus flexilis* James) inferred from the population structure of mitochondrial DNA. *Molecular Ecology*, **9**, 91–97.

Muse S, Weir B (1992) Testing for equality of evolutionary rates. *Genetics*, **132**, 269–276.

Neale DB, Wheeler NC, Allard RW (1986) Paternal inheritance of chloroplast DNA in Douglas-fir. *Canadian Journal of Forest Research*, **16**, 1152–1154.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences, USA*, **70**, 3321–3323.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Neilson R, Wakeley JW (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885–896.

Neilson MK, Lohman K, Sullivan J (2001) Phylogeography of the tailed frog (*Ascaphus truei*): implications for the biogeography of the Pacific Northwest. *Evolution*, **55**, 147–160.

Palsbøll PJ, Bérubé M, Aguilar A, Notarbartolo-Di-Sciara G, Nielsen R (2004) Discerning between recurrent gene flow and recent divergence under a finite-site mutation model applied to North Atlantic and Mediterranean Sea fin whale (*Balaenoptera physalus*) populations. *Evolution*, **58**, 670–675.

Penhallow DP (1902) Notes on Cretaceous and Tertiary plants of Canada. *Royal Society of Canada Transactions, Series 2*, **8**, 31–91.

Penhallow DP (1907) *Manual of North American Gymnosperms*. The Athenaeum Press, Boston, Massachusetts.

Petit RJ, Pineau E, Demesure B, Bacilieri R, Ducousso A, Kremer A (1997) Chloroplast DNA footprints of postglacial recolonization by oaks. *Proceedings of the National Academy of Sciences, USA*, **94**, 9996–10001.

Petit RJ, Mousadik AE, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, **12**, 844–855.

Petit RJ, Brewer S, Bordács S *et al.* (2002a) Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecology and Management*, **156**, 49–74.

Petit RJ, Csaikl UM, Bordács S *et al.* (2002b) Chloroplast DNA variation in European white oaks: phylogeography and patterns of diversity based on data from over 2600 populations. *Forest Ecology and Management*, **156**, 1–3.

Petit RJ, Aguinagalde I, de Beaulieu JL *et al.* (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science*, **300**, 1563–1565.

Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D, Vendramin GG (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology*, **14**, 689–701.

Petit RJ, Hu F-S, Dick CW (2008) Forests of the past: a window to future changes. *Science*, **320**, 1450–1452.

Ponoy B, Hong Y-P, Woods J, Jaquish B, Carlson JE (1994) Chloroplast DNA diversity of Douglas-fir in British Columbia. *Canadian Journal of Forest Research*, **24**, 1824–1834.

Pons O, Petit RJ (1995) Estimation, variance and optimal sampling of gene diversity. I. Haploid locus. *Theoretical and Applied Genetics*, **90**, 462–470.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**, 2092–2100.

Reboud X, Zeyl C (1994) Organelle inheritance in plants. *Heredity*, **72**, 132–140.

Richardson BA, Brunsfeld SJ, Klopfenstein NB (2002) DNA from bird-dispersed seed and wind-disseminated pollen provides insights into postglacial colonization and population genetic structure of whitebark pine (*Pinus albicaulis*). *Molecular Ecology*, **11**, 215–227.

Richmond GM, Fullerton DS (1986) Introduction to Quaternary glaciations in the United States of America. *Quaternary Science Reviews*, **5**, 3–10.

Ritter EW, Hatoff BW (1977) Late Pleistocene pollen and sediments: and analysis of a central California locality. *Texas Journal of Science*, **29**, 195–207.

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds Krawetz S, Misener S), pp. 365–386. Humana Press, Totowa, New Jersey.

von Rudloff E (1972) Chemosystematic studies in the genus *Pseudotsuga*. I. Leaf oil analysis of the coastal and Rocky Mountain varieties of the Douglas-fir. *Canadian Journal of Botany*, **50**, 1025–1040.

von Rudloff E, Rehfeldt GE (1980) Chemosystematic studies in the genus *Pseudotsuga*. IV. Inheritance and geographical variation in the leaf oil terpenes of Douglas-fir from the Pacific Northwest. *Canadian Journal of Botany*, **58**, 546–556.

Rymer MJ (1981) Stratigraphic revision of the Cache Formation (Pliocene and Pleistocene), Lake County, California. *United States Geological Survey Bulletin*, **1502–C**, 1–35.

Schorn HE (1994) A preliminary discussion of fossil larches (*Larix*, Pinaceae) from the Arctic. *Quaternary International*, **22–23**, 173–183.

Shafer SL, Bartlein PJ, Thompson RS (2001) Potential changes in the distributions of western North America tree and shrub taxa under future climate scenarios. *Ecosystems*, **4**, 200–215.

Slatkin M, Maddison WP (1989) A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics*, **123**, 603–613.

Soltis DE, Gitzendanner MA, Strenge DD, Soltis PS (1997) Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. *Plant Systematics and Evolution*, **206**, 353–373.

Sorensen FC (1979) Provenance variation in *Pseudotsuga menziesii* seedlings from the var. *menziesii*-var. *glauca* transition zone in Oregon. *Silvae Genetica*, **28**, 37–119.

Spaulding WG, Leopold EB, Van Devender TR (1983) Late Wisconsin paleoecology of the American Southwest. In: *Late-Quaternary Environments of the United States: The Late Pleistocene* (eds Wright HE Jr, Porter SC), pp. 259–293. University of Minnesota Press, Minneapolis, Minnesota.

Spellman GM, Klicka J (2006) Testing hypotheses of Pleistocene population history using coalescent simulations:

phylogeography of the pygmy nuthatch (*Sitta pygmaea*). *Proceedings of the Royal Society B*, **273**, 3057–3063.

St Clair JB, Mandel NL, Vance-Boland KW (2005) Genecology of Douglas fir in western Oregon and Washington. *Annals of Botany*, **96**, 1199–1214.

Steele CA, Carstens BC, Storfer A, Sullivan J (2005) Testing hypotheses of speciation timing in *Dicamptodon copei* and *Dicamptodon aterrimus* (Caudata: Dicamptodontidae). *Molecular Phylogenetics and Evolution*, **36**, 90–100.

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

Swofford DL (2003) *paup*: Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Version 4.0b10. Sinauer Associates, Sunderland, Massachusetts.

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**, 437–460.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Thompson A, Schorn HE (1998) Mechanisms of migration in Douglas-fir [*Pseudotsuga menziesii* (Mirb.) Franco]: a model. *Integrative Paleontology and the Future*, University of California, Museum of Paleontology, Berkeley, California.

Thompson RS, Whitlock C, Bartlein PJ, Harrison SP, Spaulding WG (1993) Climatic changes in the western United States since 18,000 yr B.P. In: *Global Climates since the Last Glacial Maximum* (eds Wright HE Jr, Kutzbach JE, Webb T III *et al.*), pp. 468–513. University of Minnesota Press, Minneapolis, Minnesota.

Tsukada M (1982) *Pseudotsuga menziesii* (Mirb.) Franco: its pollen dispersal and late Quaternary history in the Pacific Northwest. *Japanese Journal of Ecology*, **32**, 159–187.

Tsumura Y, Suyama Y (1998) Differentiation of mitochondrial DNA polymorphisms in populations of five Japanese *Abies* species. *Evolution*, **52**, 1031–1042.

Wakasugi T, Tsudzuki J, Ito S, Nakashima T, Tsudzuki T, Sugiura M (1994) Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *Proceedings of the National Academy of Sciences, USA*, **91**, 9794–9798.

Walter R, Epperson BK (2001) Geographic pattern of gentic variation in *Pinus resinosa*: area of greatest diversity is not the origin of postglacial populations. *Molecular Ecology*, **10**, 103–111.

Wang X-Q, Tank DC, Sang T (2000) Phylogeny and divergence times in Pinaceae: evidence from three genomes. *Molecular Biology and Evolution*, **17**, 773–781.

Whitlock C (1993) Postglacial vegetation and climate of Grand Teton and southern Yellowstone National Parks. *Ecological Monographs*, **63**, 173–198.

Whitlock C (1995) The history of *Larix occidentalis* during the last 20,000 years of environmental change. In: *Ecology and Management of Larix forests: A Look Ahead, General Technical Report GTR-INT-319* (eds Schmidt WC, McDonald KJ), pp. 83–90. USDA Forest Service, Intermountain Research Station, Ogden, Utah.

Whitlock C, Bartlein PJ (1997) Vegetation and climate change in northwest America during the past 125 kyr. *Nature*, **388**, 57–61.

Willyard A, Syring J, Gernandt DS, Liston A, Cronn R (2007) Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for *Pinus*. *Molecular Biology and Evolution*, **24**, 90–101.

Won YJ, Hey J (2005) Divergence population genetics of chimpanzees. *Molecular Biology and Evolution*, **22**, 297–307.

Zavarin E, Snajberk K (1973) Geographic variability of monoterpenes from cortex of *Pseudotsuga menziesii*. *Pure and Applied Chemistry*, **34**, 411–434.

Zhang J, Marshall JD, Jaquish BC (1993) Genetic differentiation in carbon isotope discrimination and gas exchange in *Pseudotsuga menziesii*: a common garden experiment. *Oecologia*, **93**, 80–87.

P.G. is a PhD Candidate whose dissertation research combines fossil and molecular data to understand the ecological and evolutionary history of Douglas-fir and other western North American trees. S.S. is a palaeoecologist with expertise in fossil pollen analysis and theory-based quantitative reconstruction of past vegetation abundance using palaeorecords. J.C.-B. is an Assistant Professor interested in the processes that lead to species distributions and community assembly across a variety of spatial and temporal scales, especially in North American oaks.

Supporting Information

Additional supporting information may be found in the online version of this article.

Table S1 Population and sample site information with mitotype and chlorotype frequencies.

Table S2 Mitotype and chlorotype definitions in terms of *V7*, *nad7i1*, *rps7-trnL*, and *rps15-psaC* haplotypes reported to GenBank (accessions in parentheses).

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.