MOLECULAR ECOLOGY

Molecular Ecology (2013) **22**, 2640–2652 doi: 10.1111/mec.12280

Development of genetic diversity, differentiation and structure over 500 years in four ponderosa pine populations

M. R. LESSER,* T. L. PARCHMAN† and S. T. JACKSON‡§

*Department of Biology, Syracuse University, Syracuse, NY 13244, USA, †Department of Botany, University of Wyoming, Laramie, WY 82072, USA, ‡Program in Ecology, Department of Botany, University of Wyoming, Laramie, WY 82072, USA

Abstract

Population history plays an important role in shaping contemporary levels of genetic variation and geographic structure. This is especially true in small, isolated range-margin populations, where effects of inbreeding, genetic drift and gene flow may be more pronounced than in large continuous populations. Effects of landscape fragmentation and isolation distance may have implications for persistence of range-margin populations if they are demographic sinks. We studied four small, disjunct populations of ponderosa pine over a 500-year period. We coupled demographic data obtained through dendroecological methods with microsatellite data to discern how and when contemporary levels of allelic diversity, among and within-population levels of differentiation, and geographic structure, arose. Alleles accumulated rapidly following initial colonization, demonstrating proportionally high levels of gene flow into the populations. At population sizes of approximately 100 individuals, allele accumulation saturated. Levels of genetic differentiation among populations (F_{ST} and Jost's D_{est}) and diversity within populations (F_{IS}) remained stable through time. There was no evidence of geographic genetic structure at any time in the populations' history. Proportionally, high gene flow in the early stages of population growth resulted in rapid accumulation of alleles and quickly created relatively homogenous genetic patterns among populations. Our study demonstrates that contemporary levels of genetic diversity were formed quickly and early in population development. How contemporary genetic diversity accumulates over time is a key facet of understanding population growth and development. This is especially relevant given the extent and speed at which species ranges are predicted to shift in the coming century.

Keywords: allele accumulation, gene flow, genetic structure, *Pinus ponderosa*, population history, SSR, temporal development

Received 29 October 2012; revision received 24 January 2013; accepted 29 January 2013

Introduction

Population history, in conjunction with the life history traits of a species, can play a major role in observed levels of genetic variation and differentiation among populations (Hamrick & Nason 2000). Distance from other populations, number of founding individuals, temporally variable levels of migration and gene flow, and past bottlenecks interact with species' reproductive strategies, generation times, and ability and modes of dispersal to form distinct patterns of genetic variation and structure (Hamrick & Nason 2000; Conner & Hartl 2004). The relative importance and various roles of these mechanisms in shaping population-genetic structure can be assessed by empirically tracking temporal change in genetic diversity and variation (e.g. Hadly *et al*. 2004).

Many studies estimate historical gene flow through indirect measures. Gene flow can be indirectly estimated based on F_{ST} (Wright 1922; Rousset 1997). More advanced modelling methods (e.g. *Migrate*), utilize

Correspondence: M. R. Lesser, Fax: 315-443-2012;
E-mail: mrlesser@syr.edu
§Current address: DOI Southwest Climate Science Center, U.S. Geological Survey, 1955 E. Sixth Street, Ithaca, Tucson AZ, 85719

coalescence frameworks (Beerli & Felsenstein 2001; Beerli & Palczewski 2010) to assess gene flow between populations; and Hey (2010) uses isolation with migration (IM) models to assess population divergence. However, such techniques require assumptions that are difficult to test (Hamrick & Nason 2000) and may leave many historic events unaccounted for (Ramakrishnan & Hadly 2009).

Direct assessment of temporal dynamics is challenging, because mortality and decomposition typically erase the record of change. Temporal genetic data may sometimes be available over long time periods through fossil evidence, including ancient DNA (aDNA) (Ramakrishnan & Hadly 2009). Data from aDNA and other paleontological sources can provide valuable information on phylogenetic structure (Orlando *et al.* 2002), past bottlenecks (Ramakrishnan *et al.* 2005) and population isolation (Hadly *et al.* 1998). However, ancient DNA data are typically scarce and may include poorly constrained taphonomic distortions, a property shared with all fossil evidence (Jackson 2012). Data from aDNA are also typically too coarsely resolved to infer dynamics within and among populations at timescales of decades to centuries.

The development of genetic structure at these timescales is especially relevant for understanding expansion and colonization dynamics of native and non-native species (Slatkin 1987; Anderson *et al.* 2010). For tree populations, a solution to this challenge can come from combining dendroecology and population genetics. Through dendroecological methods, individual trees can be aged with high temporal precision (Fritts & Swetnam 1989). By combining these data with population-genetic data from the same individuals, genetic structure and diversity can be assessed with temporal precision throughout a population's history.

Range expansions into new territory often involve establishment, growth and coalescence of small, isolated populations at the range-margins of species (Gaston 2003; Lyford *et al.* 2003, Bridle & Vines 2006). In such populations, impacts of isolation, number of founders and immigration rates can be more pronounced than in large continuous range-centre populations (Ellstrand & Elam 1993). Populations at range-margins are often demographic sinks, with more central populations serving as sources (Pulliam 1988). Under these circumstances, the expectation is that if gene flow is weak, range-margin populations may become genetically distinct through inbreeding, genetic drift and local adaptation (Mimura & Aitken 2007). However, an alternative hypothesis is that gene flow from source populations may homogenize allelic frequencies in range-margin populations (García-Ramos & Kirkpatrick 1997; Aitken *et al.* 2008). Knowledge of how fine-scale genetic structure and variation underlie migration and expansion processes is crucial for understanding how, and at what time-scales, species will move into and occupy new territory (Caplat *et al.* 2008; Jackson *et al.* 2009). This is especially important given that an expected outcome of 21st century climate change is that many plant species will need to shift their distribution to remain in suitable niche space (Rehfeldt *et al.* 2006; Morin & Thuiller 2009).

Ponderosa pine occurs in small disjunct populations throughout much of its range, providing opportunities to study genetic variation in small isolated populations. While range-wide and landscape-level patterns of variation in ponderosa pine have been documented extensively (Conkle & Critchfield 1988; Latta & Mitton 1999; Epperson *et al.* 2001; Johansen & Latta 2003), few studies have focused on within-population diversity (however, see Linhart *et al.* 1981). Overall, populations of coniferous trees generally show high levels of within-population variation, but usually much lower differentiation among populations due to large population sizes, wind pollination and high levels of out-crossing (Hamrick *et al.* 1992, Savolainen *et al.* 2007). Furthermore, genetic structure in many conifers, including ponderosa pine is a result of their relative longevity, overlapping generations and long juvenile phases combined with northward migration patterns following the last glacial maximum (Austerlitz *et al.* 2000; Norris 2006).

Here, we focus on four small disjunct populations of ponderosa pine in the Bighorn Basin of north-central Wyoming. A near-complete demographic history of these populations has been documented, starting with the first colonizing individuals in the seventeenth and eighteenth centuries and continuing uninterrupted to the early 21st century (Lesser & Jackson 2012a). We combine this demographic history with data from nine nuclear microsatellite loci to analyse patterns and levels of genetic variation as the populations developed over the course of four to five centuries. Microsatellite markers are ideal for this type of analysis owing to high levels of allelic variability. Understanding rates and controls of allele accumulation in natural populations and determining the time spans required for the development of genetic differentiation within and among populations are critical for understanding evolutionary processes in range-margin populations (Jay *et al.* 2012). Our primary objectives were to assess the rate at which genetic diversity accumulated and determine the relationship between population sizes and diversity. Furthermore, we examine the alternate hypotheses of range-margin populations being genetically distinct populations with low levels of diversity and hence high genetic structure, or being homogenized through gene flow. An understanding of genetic structure is

important for informing conservation practices faced with rapid climate change and increasingly fragmented landscapes (Hannah *et al.* 2002; Bacles & Jump 2010; Loss *et al.* 2011).

Methods

Study sites and data collection

We studied four disjunct populations of ponderosa pine in the Bighorn Basin in north-central Wyoming (Fig. 1, Table 1). Populations were initially colonized between 1531 and 1655 (Table 1, Fig. 2). The number of individuals in each population ranged from 82 to over 900 (Table 1). All four populations are separated from other ponderosa pine populations by >10 km (Fig. 1). There is high certainty that the initial colonizers are still present and that trees are not missing from the record due to decomposition or disturbance (Lesser & Jackson 2012a). Evidence for the initial colonizers being present comes from the presence of dead wood and snags of *Pinus flexilis* and *Juniperus* spp. at all of the sites, while none of *Pinus ponderosa* was observed. Dead wood persists for centuries in semi-arid environments like the Bighorn Basin (Gray *et al.* 2004; Millar *et al.* 2007). The complete absence of fire-scarred trees along with the absence of surface or soil charcoal coupled with the scarcity of surface fuels suggest fire has not destroyed evidence of past establishment events. Thus, we were able to document the almost complete history of population development for the four sites. Lesser and Jackson (2012a, 2012b) provide detailed site descriptions.

An almost complete age census of the four populations was obtained using dendroecological techniques.

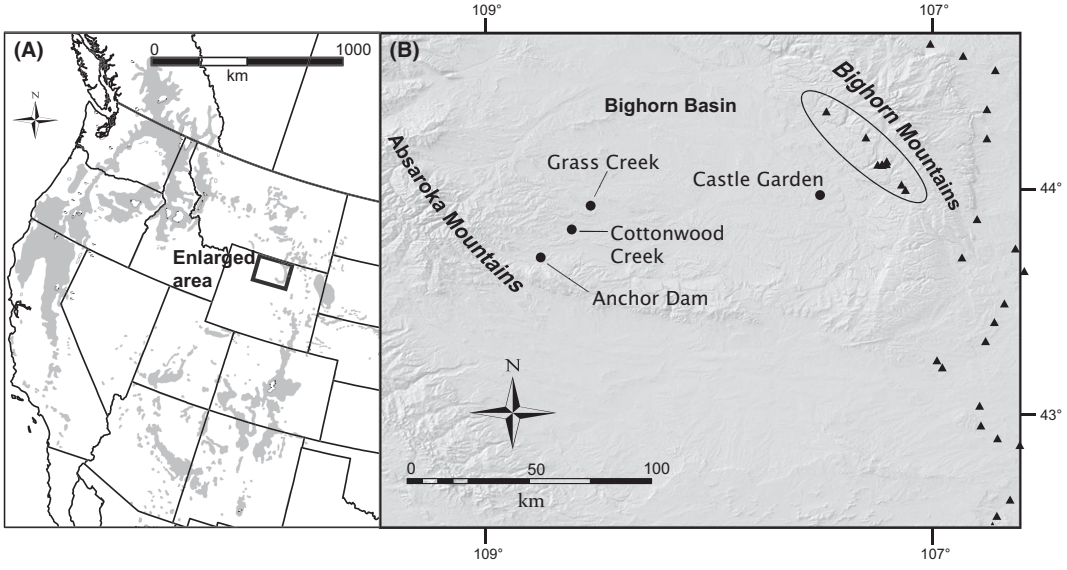

Fig. 1 (A) Geographic distribution of ponderosa pine (US Geological Survey 1999). (B) Inset portrays location of four study sites. The potential source area sampled along the western slope of the Bighorn Mountains is circled. Triangles designate ponderosa pine occurrence based on Rocky Mountain Herbarium Specimen Database records (2011).

Table 1 Study site locations, year of colonization and number of individuals sampled

Population	Latitude (N)	Longitude (W)	Year of first establishment	No. of individuals	No. genotyped and aged trees
Castle Garden	43.96	107.52	1592	177	144
Grass Creek	43.88	108.63	1605	82	62
Cottonwood Creek	43.80	108.70	1531	927	797
Anchor Dam	43.67	108.83	1655	177	122
Bighorn Mountains western slope	44.00	107.37	—	—	56*

*Trees only genotyped, not aged.

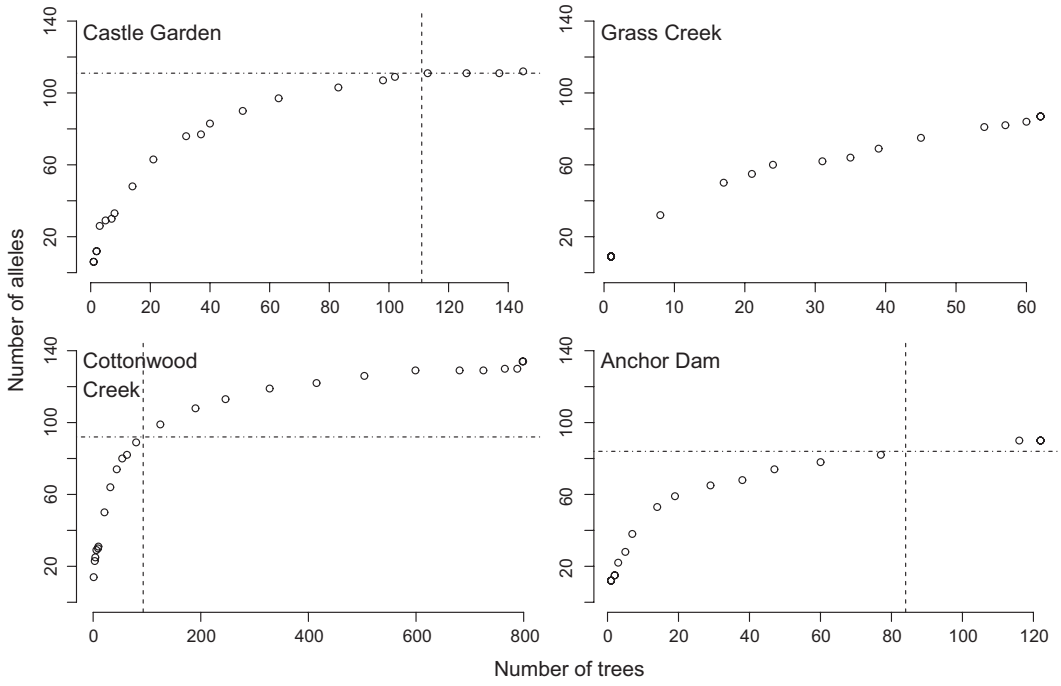

Fig. 2 Allele accumulation curves. The number of alleles is plotted against the number of individuals for each of the study populations, from the first individual to arrive to the present population. Dashed lines denote the number of alleles and the number of trees at the point where the ratio of alleles to trees falls below one. At Grass Creek, this saturation point is not reached and therefore no line is shown. Note different *x*-axis scales for each population.

Age could not be estimated for a small number of trees (10–16% of the populations) owing to heart-rot. For all trees, over 10 cm diameter, an increment core was collected. All cores were collected during the summer of 2007 and 2008. Age was estimated by whirl counts for trees smaller than 10 cm diameter. All cores were mounted, sanded, scanned and dated (Lesser & Jackson 2012a). Needle tissue was collected from every aged tree. Needle tissue was also collected from 56 trees across four locations along the western slope of the Bighorn Mountains (Fig. 1, Table 1). Populations on the western slope of the Bighorn Mountains are the most likely propagule source for the study populations because of their geographic proximity (Fig. 1).

DNA was extracted from 50 mg samples of desiccated needle tissue using a modified version of the cetyl-trimethyl ammonium bromide (CTAB) method (Doyle 1991). We verified the presence of high-molecular-weight DNA by electrophoresis in 1.5% agarose gels stained and visualized with ethidium bromide and quantified the amount of extracted DNA using a Nanodrop spectrophotometer (Thermo Scientific, Waltham, MA).

Polymerase chain reaction (PCR) was used to amplify nine nuclear microsatellite loci. Six of the loci, Pico2, Pico3, Pico31, Pico104, Pico109 and Pico138 were developed by Lesser *et al.* (2012). Two loci, Lop1 and Lop5 came from Liewlaksaneeyanawin *et al.* (2004). The final locus, NZPR4.6, was developed by Smith & Devey (1994) (Table 2). All reactions used 50–100 ng genomic DNA, 2 pmol of each primer, 0.5 mM each of dATP, dCTP, dGTP and dTTP, 1 × PCR buffer and 0.4 units of Taq polymerase. Thermocycling conditions followed Liewlaksaneeyanawin *et al.* (2004) for Lop1 and Lop5, Maherali *et al.* (2002) for NZPR4.6 and Lesser *et al.* (2012) for the remainder of the loci. To visualize genotypes, forward primers were labelled with fluorescent dyes, and PCR products were detected via capillary electrophoresis on an ABI 3130 genetic analyzer (ABI, Foster City, CA, USA). Alleles were scored using the GeneMapper software package (ABI, v 4.0).

Data analysis

For all analyses, we combined genotype data with the age data (Lesser & Jackson 2012a) for every tree. We excluded trees that were not genotyped at a minimum of four loci, which excluded a small percentage of individuals from the analysis (Table 1). Based on size and growth form, excluded trees, whether from lack of age data or failure to sufficiently amplify, were randomly distributed throughout the populations' history.

Expected and observed heterozygosity, number of alleles and allele size range were calculated for each

Table 2 Current characteristics of nine microsatellite loci used in the study across all sampled individuals. Expected heterozygosity is the estimated fraction of individuals that would be heterozygous under Hardy–Weinburg equilibrium. Estimated null allele frequencies are also given

Locus	No. of alleles	Min. allele size	Max. allele size	Het obs.	Het exp.	Est. null allele freq.
lop1	16	146	180	0.71	0.81	0.159[a]
lop5	13	165	210	0.73	0.79	0.047
NZPR4.6	21	195	238	0.44*	0.66	0.032
Pico2	15	183	211	0.65*	0.78	0.080[a]
Pico3	11	160	188	0.07*	0.30	0.290[a]
Pico138	22	126	170	0.43*	0.79	0.308[a]
Pico104	17	121	169	0.49*	0.89	0.270[a]
Pico109	25	122	170	0.46*	0.79	0.173[a]
Pico31	9	162	187	0.07*	0.28	0.302[a]

[a]loci with presence of null alleles based on van Oosterhout *et al.* (2004).
*Significant difference ($P < 0.05$) between expected and observed heterozygosities.

locus in each population using Microsatellite Analyzer (MSA v 4.05, Dieringer & Schlötterer 2003). Estimated null allele frequency was calculated using MICRO-CHECKER (van Oosterhout *et al.* 2004). We used these data to examine allele accumulation rates and frequencies through time. Allele saturation was calculated for each population as the point at which the ratio of alleles to trees fell below one, indicating that alleles were no longer accumulating more rapidly than trees. We plotted cumulative population growth curves for each population and calculated the time necessary to reach the population size corresponding with allele saturation. To deal with issues of sample size in the western Bighorn population, we performed rarefaction analysis using the PopGenkit package in R (Paquette 2012) to extrapolate the number of alleles at larger sample sizes.

To examine genetic differentiation among and within populations, we used the method of Chapuis & Estoup (2007) to calculate F_{ST}. This method adjusts F_{ST} values to correct for the presence of null alleles. The program SMOGD (Crawford 2010) was used to calculate Jost's D_{est}. Jost's D_{est} is a potentially more appropriate measure of genetic differentiation when using loci with high numbers of alleles, because it is not dependent on heterozygosity (Hedrick 2005; Jost 2008). We also examined genetic variation within populations using the F_{IS} statistic and observed heterozygosity (Raymond & Rousset 1995). We tested for significant differences in F_{IS} between populations using one-way analysis of variance. All measures of differentiation and heterozygosity were calculated from 1780 onwards. Calculation of these measures before 1780 was not possible because of increasingly small sample sizes.

The program STRUCTURE (v 2.2, Pritchard *et al.* 2000; Falush *et al.* 2003, 2007) was used to assess the number of populations (K), that genotyped samples grouped into assuming Hardy–Weinburg equilibrium and to assign individuals to each of K clusters. The most likely number of populations, given the data, was determined from posterior distributions generated with Monte Carlo Markov Chain (MCMC) sampling with a burn-in of 50 000 steps and an additional 150 000 steps for each round of simulation. We generated posterior distributions for 10 replicates of each K ranging from 1 to 8 and used the log-likelihood value from the MCMC to assess the most likely number of populations (Pritchard *et al.* 2000). K was also inferred based on the change in model likelihoods between successive values of K (ΔK) (Evanno *et al.* 2005).

Results

Eighty-five per cent of the individuals across the four populations were successfully genotyped and aged (Table 1). Most trees not genotyped and aged had heart-rot and could not be aged (Lesser & Jackson 2012a). Fifty-two aged individuals did not successfully genotype at a minimum of four loci and were therefore not included in the analysis.

All nine microsatellite loci were highly polymorphic, with the current number of alleles per locus ranging from 9 to 25. The total number of alleles across all loci and populations was 149, and allele sizes ranged from 121 to 238 bp (Table 2). Seven of the nine loci showed significant heterozygote deficiencies (Table 2). Estimated null allele frequencies ranged from 0.05 to 0.3; MICRO-CHECKER (van Oosterhout *et al.* 2004) designated all loci, except Lop5 and NZPR4.6, as having null alleles present (Table 2). Nonetheless, the presence of null alleles should not confound the analyses presented in this study (Chapuis & Estoup 2007; Carlsson 2008).

Alleles accumulated rapidly following initial colonization at each site (Fig. 2). The total number of alleles ranged from 87 at Grass Creek to 134 at Cottonwood

Creek. All four populations show a saturating curve, with the rate of allele accumulation decreasing as population size increased (Fig. 2). Castle Garden, Cottonwood Creek and Anchor Dam reached saturation between 84 and 111 alleles. Population size at the time of allele saturation ranged from 84 trees at Anchor Dam to 111 trees at Castle Garden (Fig. 2). Time to saturation from population initiation ranged from 234 to 364 years at Cottonwood Creek and Castle Garden respectively (Fig. 3). Grass Creek never crossed the allele saturation point because the total number of trees never exceeded the number of alleles.

Alleles that arrived in the early stages of population growth tended to remain at high frequencies throughout the rest of population development (Fig. 4). The NZPR4.6 locus is representative of the patterns observed for the other eight loci. In all four populations, two or three alleles occur at relatively high frequencies (>0.1), while many alleles (between 5 and 22) occur at low frequencies. Furthermore, only five alleles were common (>0.1) and four of these were shared by two or more populations (Fig. 3).

Measures of differentiation among populations were generally low and remained constant through time (Fig. 5). One exception to this is the D_{est} value of 0.44 for the Castle Garden-Anchor Dam population pair in 1780. However, after 1840, D_{est} values became more similar between all population pairs and remained similar to present. An exception is the Castle Garden-Grass Creek pair, which remained consistently lower than the others. Current D_{est} (Jost 2008) ranged from 0.07 to 0.31 between the Castle Garden-Grass Creek and Grass Creek-Anchor Dam population pairs, respectively (Fig. 5A).

Adjusted F_{ST} values (Chapuis & Estoup 2007) showed a similar trend to $D_{est,}$ but were consistently lower. However, F_{ST} may not be an adequate measure of differentiation for highly diverse microsatellites (Jost 2008). Beginning in 1780, F_{ST} values ranged from 0.01 between the Castle Garden-Grass Creek pair to >0.2 for the Castle Garden-Anchor Dam and Grass Creek-Cottonwood Creek pairs (Fig. 5B). Following 1780, F_{ST} decreased rapidly in these two population pairs, becoming similar to other population pairs by 1840. From 1840 to the present, F_{ST} remained constant with all populations showing similar values. The exception to this is, as with D_{est}, the Castle Garden-Grass Creek pair, where F_{ST} was considerably lower than in the other population pairs. Current F_{ST} ranged from 0.01 to 0.1 in the Castle Garden-Grass Creek and Anchor Dam-Cottonwood Creek pairs, respectively (Fig. 5B).

F_{IS}, or the inbreeding coefficient, ranged from 0.27 at Cottonwood Creek to 0.44 at Anchor Dam in 1780 (Fig. 5C). F_{IS} values remained similar from 1780 to the present, with current F_{IS} ranging from 0.26 at Cottonwood Creek to 0.39 at Grass Creek (Fig. 5C). F_{is} values

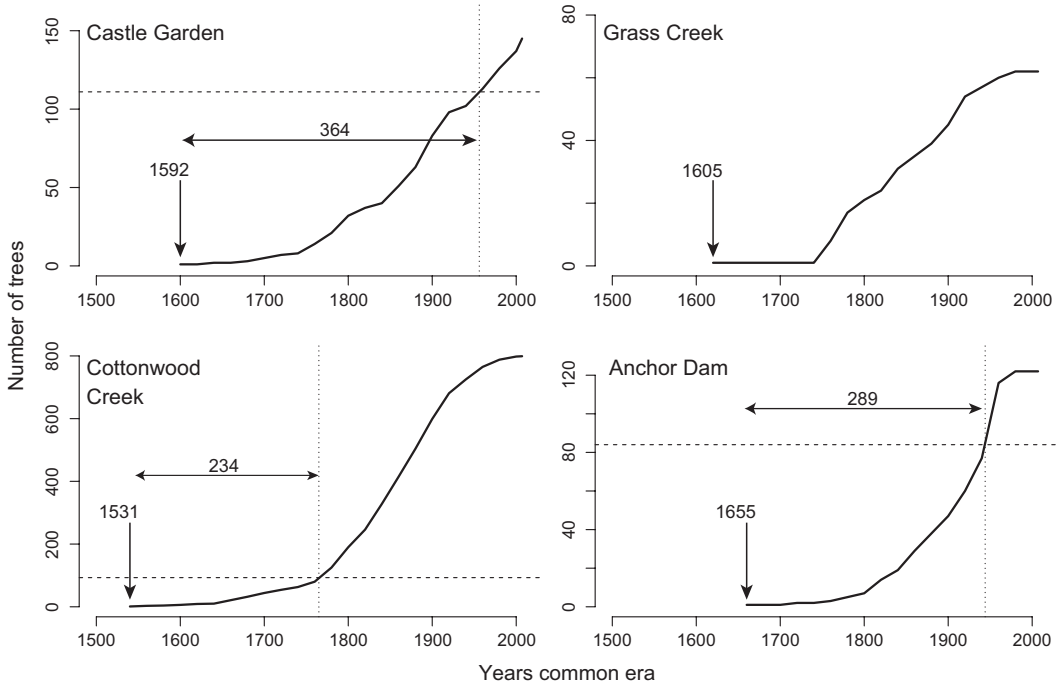

Fig. 3 Cumulative population growth for the four study populations. Dashed lines denote the number of trees and the year at which allele saturation is reached (from Fig 2). Vertical arrow indicates year of population initiation. Horizontal arrow indicates time span from year of initiation to allele saturation in years. Note different *y*-axis scales for each population.

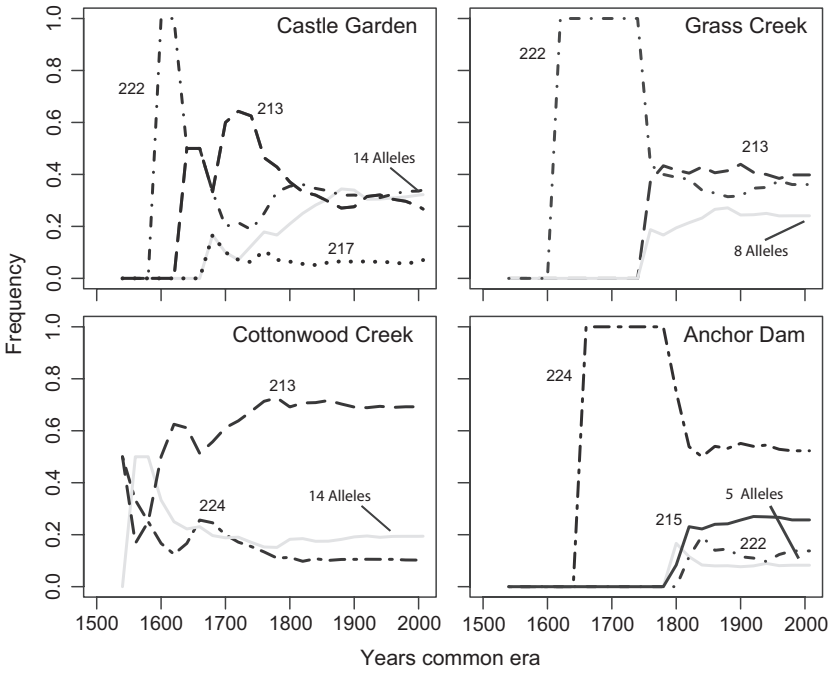

Fig. 4 Allele frequencies through time at the NZPR4.6 locus for the four study populations. For each population, the highest-frequency alleles are shown individually (alleles that are currently over 0.1). The frequency of all other alleles is combined (grey solid line), and the number of alleles represented by that line is shown. The NZPR4.6 locus is shown as a representative for all the loci studied. The same overall trend is seen for the eight other loci used in the study.

between populations were not significantly different. Observed heterozygosity (H_o) also remained relatively constant throughout time in all four populations, ranging from 0.3 to 0.5 (Fig. 5D).

For both measures of differentiation among populations, the values between Castle Garden and Grass Creek with the western Bighorn population were similar to those between study populations. Anchor Dam and Cottonwood Creek had higher pairwise differentiation values from the western Bighorn population ($D_{est} > 0.4$ and $F_{ST} > 0.1$) than the other populations (Fig. 5E). The western Bighorn population also showed a similar F_{IS} value to all of the other populations ($F_{IS} = 0.38$) (Fig. 5D). Observed heterozygosity was lower for the western Bighorn population (0.38) than for the four study populations at present (Fig. 5E). The recorded number of alleles in the western Bighorn population (67) was lower than what was found in the study populations (149). Rarefaction analysis showed that even with increased sample size the estimated number of alleles in the western Bighorn population would remain below 80.

Clustering analysis in STRUCTURE showed no clear pattern of any geographic structure through time. Both likelihood and ΔK (Evanno *et al.* 2005) approaches showed the highest support for three populations. Of the time steps analysed, the present-day populations revealed the highest amount of visible structure, especially between Cottonwood Creek and the other populations (Fig. S1). However, most individuals were partially assigned to all three populations with credible assignments ranging from zero to one, giving the result no clear interpretation. This result was consistent through time, from 1780 to the present, and indicates a lack of pronounced population-genetic structuring from the time of stand establishment to the present.

Discussion

Tandem application of dendroecological data with genetic data allowed us to examine the temporal progression of population diversity and differentiation over the course of five centuries. Combining historic demographic data with genetic data overcomes many of the difficulties associated with directly measuring historical population-genetic parameters (Hamrick & Nason 2000). The utility of combining genetic data with paleoecological data has been shown for coarse-scale studies (Cwynar & MacDonald 1987; Hadly *et al.* 1998; Orlando *et al.* 2002; Petit *et al.* 2002; Magri *et al.* 2006; Gugger *et al.* 2010). However, the scale of paleoecological data does not allow for detailed population-level analysis (Jackson & Overpeck 2000), and few studies have integrated fine-scale demographic and genetic data. Steinitz *et al.* (2011) used a combination of dendroecology and microsatellites to examine parentage and Janzen-Connell effects in populations of Aleppo pine (*Pinus halepensis*) over the past century. Other cross-generational studies have only utilized cohort data, or at most two generations to examine parentage and fine-scale structure (e.g. González-Martinez *et al.* 2002).

Our results highlight the utility of using long-term demographic data at fine temporal and spatial scales.

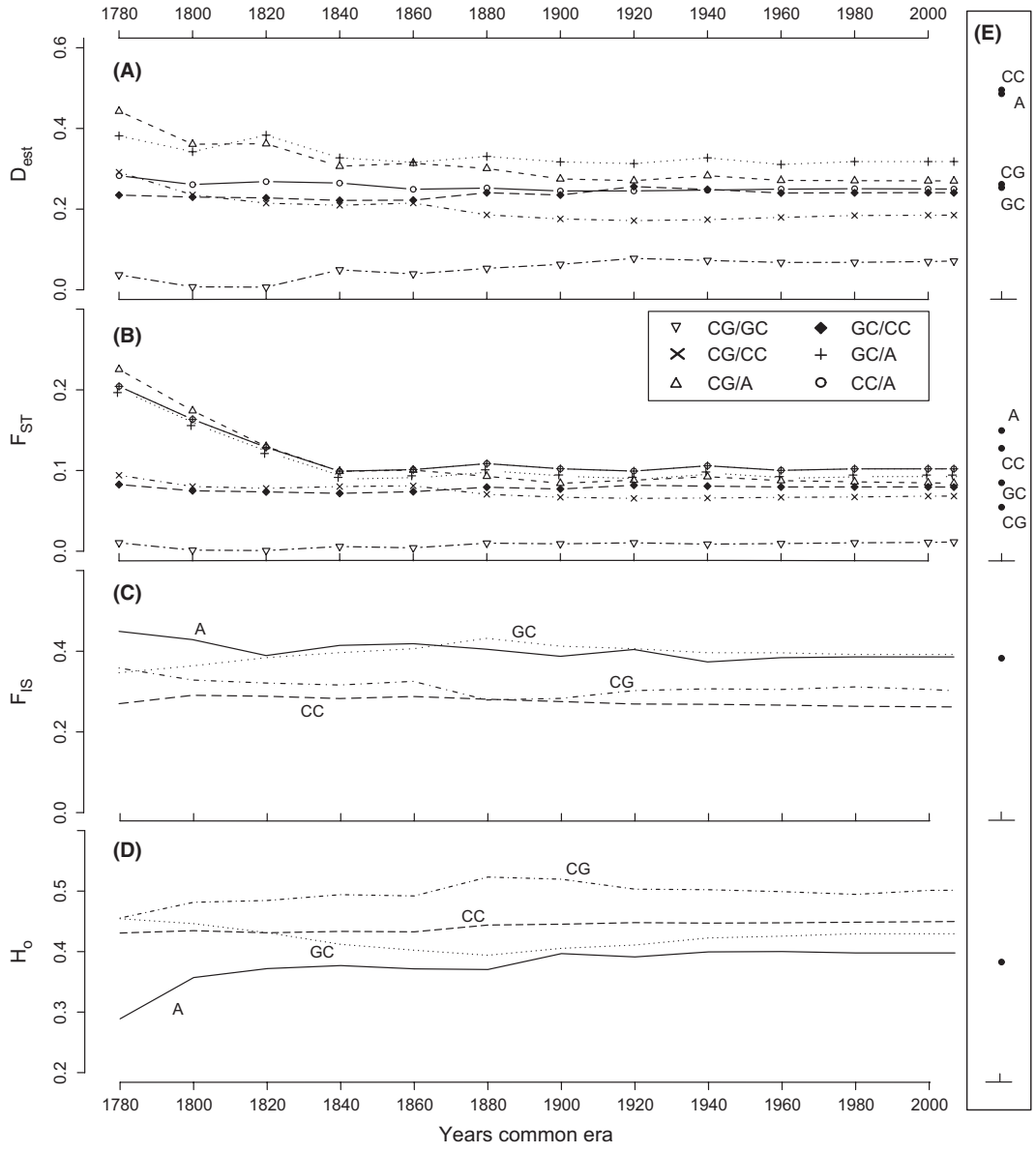

Fig. 5 Estimates of (A) Jost's D_{est}, and (B) F_{ST} for all population pairs. (C) F_{IS} estimates for all study populations. (D) Estimates of observed heterozygosity for all study populations. Estimates are shown from 1780 to 2007 AD (E) Jost's D and F_{ST} estimates between the western Bighorn Mountain population sample and each of the study populations, and F_{IS} and observed heterozygosity estimate of the Bighorn Mountain population sample. Populations are denoted as CG, GC, CC and A for Castle Garden, Grass Creek, Cottonwood Creek and Anchor Dam, respectively.

This approach allowed us to examine how levels of genetic diversity changed at decadal to centennial timescales and from the initial colonizers of the population through several generations to current population sizes. Results demonstrate that contemporary structure and diversity developed quickly following population initiation and have remained stable over the majority of population development. This has consequences for understanding the time frames and populations sizes required for range expansion of tree species. The rapid pace of current climate change (IPCC 2007) coupled with increasing occurrence and pace of plant invasions (Bradley *et al.*2010) makes understanding the fine-scale mechanisms of how plant species occupy new territory and develop populations increasingly important.

Differentiation and structure through time

Levels of differentiation between the study populations remained relatively constant through time. Levels of

differentiation are consistent with studies from other conifers (Hamrick *et al.* 1992; Parchman *et al.* 2011), which also typically show F_{ST} values below 0.1, little differentiation among populations and high levels of diversity within populations. In general, this is consistent with the large population sizes and high levels of outcrossing typical of conifers (Hamrick *et al.* 1992).

Previous studies of genetic variation in ponderosa pine have shown evidence for large-scale geographic separation into varieties and races across the range (Wells 1963; Conkle & Critchfield 1988; Latta & Mitton 1999; Epperson *et al.* 2001; Johansen & Latta 2003). Variation at finer geographic scales has been less studied, although Linhart *et al.* (1981) reported that differentiation was much greater spatially than temporally within populations. Our results also show little temporal change, with F_{IS}, H_o, and allele frequencies remaining relatively stable through time (Figs 4 and 5).

Peripheral populations are often demographic sinks (Pulliam 1988). Sink populations that experience low levels of immigration from outside sources are prone to become genetically distinct through inbreeding, genetic drift, and local adaptation (Mimura & Aitken 2007). Many examples of local adaptation in forest trees exist, primarily through common garden experiments (Savolainen *et al.* 2007). Furthermore, modelling results have shown that, in the absence of gene flow, peripheral populations can experience rapid evolution (García-Ramos & Kirkpatrick 1997). However, an alternative scenario is that continuous immigration may not only sustain peripheral populations, but homogenize genetic variation (Ellstrand & Elam 1993; Sork & Smouse 2006; Mimura & Aitken 2007). Further modelling results from Garcio-Ramos & Kirkpatrick (1997) showed that gene flow from central populations can override local adaptation. Savolainen *et al.* (2007) reported that northern range-margin populations of *Pinus sylvestris* were not locally adapted, due to extensive pollen flow from more southern sources. Our results also suggest that gene flow is homogenizing genetic variation in our four study populations. The similarity in F_{IS} between our four study populations and the western Bighorn population indicates that high levels of gene flow have overridden inbreeding and genetic drift in the peripheral populations. Parentage analysis of these data support high levels of gene flow, with nearly every individual establishing in the first century following site colonization having extra-local parents (Lesser & Jackson 2012b).

Diversity accumulation through time

Allele frequencies for all nine loci were highly variable in the early stages of population growth following initial colonization (Fig. 4). This variability is most likely the result of proportionally high levels of gene flow from dispersal events into the populations' during this time. High gene flow, in conjunction with small population size, can cause allele frequencies to change drastically (Ellstrand & Elam 1993; Sork & Smouse 2006). As population size increases, gene flow has less of an effect on allele frequency dynamics. Furthermore, as population size increases, allelic diversity can become saturated and it becomes increasingly unlikely that immigration will introduce new alleles into the population.

Our results show that alleles accumulated rapidly following initial colonization (Fig. 2). Castle Garden, Cottonwood Creek, and Anchor Dam all showed saturation of alleles. Population size at the saturation point ranged between 113 trees at Castle Garden and 125 trees at Cottonwood Creek. The number of alleles at saturation is also similar across the three populations, ranging from 84 at Anchor Dam to 111 at Castle Garden (Fig. 2). This similarity in population size and number of alleles suggests that not only are immigration rates similar across populations, but new alleles are introduced through immigration at a similar rate. Furthermore, the time it takes each population to reach allele saturation is similar, ranging from 234 years at Cottonwood Creek to 364 years at Castle Garden (Fig. 3).

Gene flow has the potential to erode differentiation between populations (F_{ST}) and increase variation within small populations (Ellstrand & Elam 1993). In our study populations, gene flow appears to increase variation. The western Bighorn population sample contained only 67 alleles across the nine loci, compared with the 149 alleles found in the study populations. The random sampling strategy utilized for the western Bighorn population may have missed rare alleles. However, rarefaction analysis showed that the number of alleles would still be <80 even at a sample size of 1200 individuals (results not shown). An alternative explanation is that not all gene flow is from that area, and long-distance dispersal has originated from multiple sources. In either case, the disjunct populations harbour as much, if not more, genetic diversity as the large, continuous forest population.

This result suggests that isolated populations do not necessarily have low levels of variation and that the number of seed sources may be higher than other studies have shown (Sork & Smouse 2006). Sezen *et al.* (2005) found that over half of the seedlings at second-growth sites were the offspring of only two parents. Aldrich & Hamrick (1998) found similar results in a study of isolated pasture trees. Our results, however, based on the number and rate of allele accumulation indicate a much higher number of source individuals

contributing to the populations. Furthermore, parentage analysis showed that almost every individual establishing for >100 years following initial colonization was the result of long-distance seed dispersal (Lesser & Jackson 2012b). Long-lived tree species, such as ponderosa pine, have the capacity to persist on the landscape for long periods of time without reproducing. During this time, population growth may occur only through long-distance seed-dispersal events, thus increasing the number of founders and decreasing founder effects (Austerlitz *et al.* 2000). These results suggest that gene flow has the potential to override expectations of low diversity and high structure in range-margin populations (Mimura & Aitken 2007). While range-margin populations may remain demographic sinks, this does not necessarily mean that they will harbour less genetic diversity given that the time frames for development, and the magnitude of gene flow that we show here, are realized.

Implications for conservation

Genetic diversity within populations is an important parameter for conservation (Frankel 1974; Kramer & Havens 2009), and many conservation efforts seek to maximize this. However, in the context of species migration potential, it is also important to understand the time frames and population sizes necessary for newly colonized populations to acquire threshold levels of genetic diversity. Our data suggest that for natural populations of conifers this process may require more than two centuries, with approximately 100 individuals established before the rate of increase in genetic diversity levels off. Allelic diversity may continue to increase after this point as populations continue to grow; however, it does so at an ever decreasing rate. Assisted migration has been proposed as a potential strategy for mitigating species responses to climate change (McLachlan *et al.* 2007; Schwartz *et al.* 2012). Understanding genetic diversity and how it accumulates over time can inform assisted migration strategies on where propagules should be obtained and how many individuals are needed to form, or supplement, populations to support a given level of genetic diversity (Frankham 2010).

Another important issue in conservation genetics is how fragmentation affects gene flow and population structure (Sork & Smouse 2006; Bacles & Jump 2010). Our results suggest that populations that are far removed from other populations (>10 km) still experience levels of gene flow that are high enough to eliminate geographic structure and keep differentiation among populations low. However, our results may not be applicable to species that have lower pollen productivity and dispersibility, and lower seed-dispersal capabilities, than pines. Furthermore, as landscapes become more fragmented, populations may become even more isolated than our study sites, and move beyond threshold distances for gene flow. Artificial augmentation of gene flow is one possible solution to this issue (Frankham 2010). Again, data such as that we present here is highly informative for this, and other management strategies, on how systems have operated over the past several centuries, and what expectations of genetic diversity and structure should be at various stages of population development along range-margins. The applicability of this study design to other species and systems may be limited due to life history characteristics and disturbance regimes that do not allow long-term persistence of almost all individuals. However, where possible, there is a need for similar studies on species with different life history and dispersal strategies than what we present here, so that a more comprehensive understanding of plant migration potential can be reached.

Acknowledgements

Support was provided by a USDA McIntire-Stennis Competitive Grant and a NSF Doctoral Dissertation Improvement Grant (DEB-0910173). We thank C. Benkman, A. Buerkle, D. Doak, and D. Tinker for critical comments and statistical advice. We thank B. Jenkins, J. Theurer, H. Olsen, and S. Brewer for laboratory and field assistance. We thank J. F. Scheepens, M. Heuertz, and two anonymous reviewers for valuable comments and suggestions.

References

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S (2008) Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications*, **1**, 95–111.

Aldrich PR, Hamrick JL (1998) Reproductive dominance of pasture trees in fragmented tropical forest mosaic. *Science*, **281**, 103–105.

Anderson CD, Epperson BK, Fortin M *et al.* (2010) Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Molecular Ecology*, **19**, 3565–3575.

Austerlitz F, Mariette S, Machon N, Gouyon P, Godelle B (2000) Effects of colonization processes on genetic diversity: differences between annual plants and tree species. *Genetics*, **154**, 1309–1321.

Bacles CEF, Jump AS (2010) Taking a tree's perspective on forest fragmentation genetics. *Trends in Plant Science*, **16**, 13–18.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopualtions by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 4563–4568.

Beerli P, Palczewski M (2010) Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics*, **109**, 1125–1132.

Bradley BA, Blumenthal DM, Wilcove DS, Ziska LH (2010) Predicting plant invasions in an era of global change. *Trends in Ecology and Evolution*, **25**, 310–318.

Bridle JR, Vines TH (2006) Limits to evolution at range margins: when and why does adaptation fail. *Trends in Ecology and Evolution*, **22**, 140–147.

Caplat P, Anand M, Bauch C (2008) Interactions between climate change, competition, dispersal, and disturbances in a tree migration model. *Theoretical Ecology*, **1**, 209–220.

Carlsson J (2008) Effects of microsatellite null alleles on assignment testing. *Journal of Heredity*, **99**, 616–623.

Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, **24**, 621–631.

Conkle MT, Critchfield WB (1988) Genetic variation and hybridization of ponderosa pine. In: *Pondersa Pine: The Species and its Management* (eds Baumgartner DM, Lotan JE), pp. 27–43. Washington State Univ. Press, Pullman, Washington.

Conner JK, Hartl DL (2004) *A Primer of Ecological Genetics*. Sinuarer Associated Inc, Sunderland, Massachusetts. 304 pp.

Crawford NG (2010) SMOGD: software for the measurement of genetic diversity. *Molecular Ecology Resources*, **10**, 556–557.

Cwynar LC, MacDonald GM (1987) Geographical variation of lodgepole pine in relation to population history. *The American Naturalist*, **129**, 463–469.

Dieringer D, Schlötterer C (2003) Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes*, **3**, 167–169.

Dieringer D, Schlötterer C (2003) Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes*, **3**, 167–169.

Doyle J. (1991) DNA protocols for plants: a CTAB total DNA isolation In: *Molecular techniques in taxonomy*, (eds Hewitt GM, Johnson A) pp. 283–293. Springer-Verlag, Berlin.

Epperson BK, Telewski FW, Plovanich-Jones AE, Grimes JE (2001) Clinal differentiation and putative hybridization in a contact zone of *Pinus ponderosa* and *P. arizonica* (Pinaceae). *American Journal of Botany*, **88**, 1052–1057.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multiple locus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, **7**, 574–578.

Frankel OH (1974) Genetic conservation: our evolutionary responsibility. *Genetics*, **78**, 53–65.

Frankham R (2010) Challenges and opportunities of genetic approaches to biological conservation. *Biological Conservation*, **143**, 1919–1927.

Fritts HC, Swetnam TW (1989) Dendroecology: a tool for evaluating variations in past and present forest environments. In: *Advances in Ecological Research*, Vol. **19** (eds Begon M, MacFadyen A, Fitter AH, Ford ED), pp. 111–175. Academic Press Limited, San Diego, California.

Garcio-Ramos G, Kirkpatrick M (1997) Genetic models of adaptation and gene flow in peripheral populations. *Evolution*, **51**, 21–28.

Gaston KJ (2003) *The Structure and Dynamics of Geographic Ranges*. Oxford University Press, New York, New York. 267 pp.

González-Martinez SC, Gerber S, Cervera MT, Martinez-Zapater JM, Gil L, Alia R (2002) Seed gene flow and fine-scale structure In a Mediterranean pine (*Pinus pinaster* Ait.) using nuclear microsatellite markers. *Theoretical Applied Genetics*, **104**, 1290–1297.

Gray ST, Fastie CL, Jackson ST, Betancourt JL (2004) Tree-ring based reconstructions of precipitation in the Bighorn Basin, Wyoming since 1260 A.D. *Journal of Climate*, **17**, 3855–3865.

Gugger PF, Sugita S, Cavender-Bares J (2010) Phylogeography of Douglas-fir based on mitochondrial and chloroplast DNA sequences: testing hypothesis from the fossil record. *Molecular Ecology*, **19**, 1877–1897.

Hadly EA, Kohn MH, Leonard JA, Wayne RK (1998) A genetic record of population isolation in pocket gophers during Holocene climatic change. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 6893–6896.

Hadly EA, Ramakrishnan U, Chan YL *et al.* (2004) Genetic response to climate change: insights from Ancient DNA and palychronology. *Plos Biology*, **2**, 1600–1608.

Hamrick JL, Nason JD (2000) Gene flow in forest trees. In: *Forest Conservation Genetics* (eds Young A, Boshier D, Boyle T), pp. 81–90. CABI publishing, Wallingford, UK.

Hamrick JL, Godt MJW, Broyles MS (1992) Factors influencing levels of genetic diversity in woody plant species. *New Forests*, **6**, 95–124.

Hannah L, Midgley GF, Lovejoy T *et al.* (2002) Conservation of biodiversity in a changing climate. *Conservation Biology*, **16**, 264–268.

Hedrick P (2005) *Genetics of Populations*, 3rd edn. Jones and Bartlett Publishers, Boston, Massachusetts, 728 pp.

Hey J (2010) Isolation with migration models for more than two populations. *Molecular Biology and Evolution*, **27**, 905–920.

IPCC (2007) Climate change 2007 (Intergovernmental Panel on Climate Change). In: *The Physical Science Basis*(eds Solomon S, Qin D, Manning M, Chen Z, Marquis MC, Averyt KB, Tignor M & Miller HL), pp. 996. Cambridge University Press, Cambridge.

Jackson ST (2012) Representation of flora and vegetation in Quaternary fossil assemblages: known and unknown knowns and unknowns. *Quaternary Science Reviews*, **49**, 1–15.

Jackson ST, Overpeck JT (2000) Responses of plant populations and communities to environmental changes of the Late Quaternary. *Paleobiology (supplement)*, **26**, 194–220.

Jackson ST, Betancourt JL, Booth RK, Gray ST (2009) Ecology and the ratchet of events: climate variability, niche dimensions, and species distributions. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 19685–19692.

Jay F, Manel S, Alvarez N *et al.* (2012) Forcasting changes in population genetic structure of alpine plants in response to global warming. *Molecular Ecology*, **21**, 2354–2368.

Johansen AD, Latta RG (2003) Mitochondrial haplotype distribution, seed-dispersal and patterns of postglacial expansion of ponderosa pine. *Molecular Ecology*, **12**, 293–298.

Jost L (2008) GST and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.

Kramer AT, Havens K (2009) Plant conservation genetics in a changing world. *Trends in Plant Sciences*, **14**, 599–607.

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

Lesser MR, Jackson ST (2012a) Making a stand: five centuries of population growth in colonizing populations of *Pinus ponderosa*. *Ecology*, **93**, 1071–1081.

Lesser MR, Jackson ST (2012b) Contributions of long-distance dispersal to population growth in colonizing *Pinus ponderosa* populations. *Ecology Letters*, **16**, 380–389 doi:10.1111/ele.12053.

Lesser MR, Parchman TL, Buerkle CA (2012) Cross-species transferability of SSR loci developed from transcriptome sequencing in lodgepole pine. *Molecular Ecology Resources*, **12**, 448–455.

Liewlaksaneeyanawin C, Ritland CE, El-Kassaby YA, Ritland K (2004) Single-copy, species-transferable microsatellite markers developed from loblolly pine ESTs. *Theoretical and Applied Genetics*, **109**, 361–369.

Linhart YB, Mitton JB, Sturgeon KB, Davis ML (1981) Genetic variation in space and time in a population of ponderosa pine. *Heredity*, **46**, 407–426.

Loss SR, Terwilliger LA, Peterson AC (2011) Assisted colonization: integrating conservation strategies in the face of climate change. *Biological Conservation*, **144**, 92–100.

Lyford ME, Jackson ST, Betancourt JL, Gray ST (2003) Influence of landscape structure and climate variability on a late Holocene plant migration. *Ecological Monographs*, **73**, 567–583.

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

Maherali H, Williams BL, Paige KN, Delucia EH (2002) Hydraulic differentiation of ponderosa pine populations along a climatic gradient with ecotypic divergence. *Functional Ecology*, **16**, 510–521.

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.

Millar CI, Westfall RD, Delany D (2007) Response of high-elevation limber pine (*Pinus flexilis*) to multiyear drought and 20th century warming, Sierra Nevada, California, USA. *Canadian Journal of Forest Research*, **37**, 2508–2520.

Mimura M, Aitken SN (2007) Increased selfing and decreased effective pollen donor number in peripheral relative to central populations in *Picea sitchensis* (Pinaceae). *American Journal of Botany*, **94**, 991–998.

Morin X, Thuiller W (2009) Comparing niche- and process-based models to reduce prediction uncertainty in species range shifts under climate change. *Ecology*, **90**, 1301–1313.

Norris JR (2006) Influence of climate on the modern and late Holocene biogeography of ponderosa pine in the Central Rockies. Dissertation, University of Wyoming, Laramie, Wyoming.

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Orlando L, Bonjean D, Bocherens H *et al.* (2002) Ancient DNA and the population genetics of cave bears (*Ursus spelaeus*) through time and space. *Molecular Biology and Evolution*, **19**, 1920–1933.

Paquette SR (2012) PopGenKit: Useful Functions for (batch) file conversion and data resampling in microsatellite datasets. R package version 1.0. http://CRAN.R-project.org/package=PopGenKit.

Parchman TL, Benkman CW, Jenkins B, Buerkle CA (2011) Low levels of population genetic structure in *Pinus contorta* (Pinaceae) across a geographic mosaic of co-evolution. *American Journal of Botany*, **98**, 669–679.

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

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Pulliam HR (1988) Sources, sinks, and population regulation. *The American Naturalist*, **132**, 652–661.

Ramakrishnan U, Hadly EA (2009) Using phylochronology to reveal cryptic population histories: review and synthesis of 29 ancient DNA studies. *Molecular Ecology*, **18**, 1310–1330.

Ramakrishnan U, Hadly E, Mountain JL (2005) Detecting part population bottle necks using temporal genetic data. *Molecular Ecology*, **14**, 2915–2922.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **82**, 648–649.

Rehfeldt GE, Crookston NL, Warwell MV, Evans JE (2006) Empirical analysis of plant-climate relationships for the western United States. *International Journal of Plant Sciences*, **167**, 1123–1150.

Rousset F (1997) Genetic differentiation and estimation of genetic flow from F-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.

Savolainen O, Pyhäjärvi T, Knürr T (2007) Gene flow and local adaptation in trees. *Annual Review of Ecology Evolution and Systematics*, **38**, 595–619.

Schwartz MW, Hellman JJ, McLachlan JM *et al.* (2012) Managed relocation: integrating the scientific, regulatory, and ethical challenges. *BioScience*, **62**, 732–743.

Sezen UU, Chazdon RL, Holsinger KE (2005) Genetic consequences of tropical second-growth forest regeneration. *Science*, **307**, 891.

Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science*, **236**, 787–792.

Smith DN, Devey ME (1994) Occurrence and inheritance of microsatellites in *Pinus radiata*. *Genome*, **37**, 977–983.

Sork VL, Smouse PE (2006) Genetic analysis of landscape connectivity in tree populations. *Landscape Ecology* **21**: 821–836.

Steinitz O, Troupin D, Vendramin GG, Nathan R (2011) Genetic evidence for a Janzen-Connell recruitment pattern in

reproductive offspring of *Pinus halepensis* trees. *Molecular Ecology*, **20**, 4152–4164.
Wells OO (1963) Geographic variation in ponderosa pine. *Silvae Genetica*, **13**, 89–103.
Wright S (1922) Coefficients of inbreeding and relationship. *The American Naturalist*, **56**, 330–338.

M.R.L. and S.T.J. conceived the study. M.R.L. and T.L.P. performed analysis. M.R.L. wrote the manuscript with contributions from S.T.J. and T.L.P.

Data accessibility

Microsatellite data along with tree ages and coordinates are available at DRYAD (doi:10.5061/dryad.pc683).

Supporting information

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

Fig. S1 Structure plot of genetic assignment for Pinus ponderosa individuals in the four study populations, where k=3. Populations are coded as Anchor Dam (A), Cottonwood Creek (C), Castle Garden (CG), and Grass Creek (G).