

DNA from bird-dispersed seed and wind-disseminated pollen provides insights into postglacial colonization and population genetic structure of whitebark pine (*Pinus albicaulis*)

B. A. RICHARDSON,*† S. J. BRUNSFELD* and N. B. KLOPFENSTEIN†

*Department of Forest Resources, University of Idaho, Moscow, ID 83844, USA †USDA Forest Service, RMRS, 1221 S. Main, Moscow, ID 83843 USA

Abstract

Uniparentally inherited mitochondrial (mt)DNA and chloroplast (cp)DNA microsatellites (cpSSRs) were used to examine population genetic structure and biogeographic patterns of bird-dispersed seed and wind-disseminated pollen of whitebark pine (*Pinus albicaulis* Engelm.). Sampling was conducted from 41 populations throughout the range of the species. Analyses provide evidence for an ancestral haplotype and two derived mtDNA haplotypes with distinct regional distributions. An abrupt contact zone between mtDNA haplotypes in the Cascade Range suggests postglacial biogeographic movements. Among three cpSSR loci, 42 haplotypes were detected within 28 cpSSR sample populations that were aggregated into six regions. Analysis of molecular variance (AMOVA) was used to determine the hierarchical genetic structure of cpSSRs. AMOVA and population pairwise comparisons (F_{ST}) of cpSSR, and geographical distribution of mtDNA haplotypes provide insights into historical changes in biogeography. The genetic data suggest that whitebark pine has been intimately tied to climatic change and associated glaciation, which has led to range movements facilitated by seed dispersal by Clark's nutcracker (*Nucifraga columbiana* Wilson). The two hypotheses proposed to explain the genetic structure are: (i) a northward expansion into Canada and the northern Cascades in the early Holocene; and (ii) historical gene flow between Idaho and the Oregon Cascades when more continuous habitat existed in Central Oregon during the late Pleistocene. Genetic structure and insights gained from historical seed movements provide a basis on which to develop recovery plans for a species that is at risk from multiple threats.

Keywords: biogeography, chloroplast microsatellites, gene flow, mitochondrial DNA, *Pinus albicaulis*, postglacial colonization

Received 18 June 2001; revision received 25 October 2001; accepted 25 October 2001

Introduction

Biogeographic patterns are the result of the dispersal of organisms in response to historical climatic fluctuations and the distribution of suitable habitats. The oscillating climate associated with Pleistocene glaciation profoundly influenced the distribution of temperate organisms. During glacial cycles, species distributions changed in latitude

and elevation, and populations expanded and contracted (Delcourt & Delcourt 1991; Webb 1995). For plant populations, changes in distribution and genetic architecture are the ultimate result of the independent movement of seed and pollen. Thus, a robust understanding of biogeographical patterns in plants requires knowledge of gene flow (i.e. seed and pollen movement) as influenced by physical and biological constraints.

Tree species, such as whitebark pine (*Pinus albicaulis* Engelm.), that are restricted to the timberline and adjacent subalpine forest habitats would be particularly affected by

Correspondence: Steve Brunsfeld. Fax: 208 885 6226; E-mail: sbruns@uidaho.edu

the temperature variation associated with glacial and interglacial episodes. The climate-driven extinction and re-establishment of populations of whitebark pine on mountaintop islands would produce a genetic structure that is heavily influenced by gene-flow patterns. In this species, gene flow is a result of wind-dispersed pollen and seed dispersal facilitated by Clark's nutcracker (*Nucifraga columbiana* Wilson). The seed caching activities of the nutcracker have been shown to be integral to whitebark pine establishment and regeneration (Tomback 2001). Studies focusing on Clark's nutcracker seed-collection behaviour have demonstrated that the multistemmed growth form exhibited by most whitebark pine is attributable to the germination of multiple seeds from Clark's nutcracker seed caches (Tomback & Linhart 1990). These data have been confirmed by allozyme analysis indicating that the multiple stems are often different genotypes (Furnier *et al.* 1987; Rogers *et al.* 1999). Less is known about how long-distance transport of seed by nutcrackers influences genetic structure among populations. Observational data have reported bird-mediated seed transport up to 22 km (Vander Wall & Balda 1981). At the regional and range-wide scale, biparentally inherited allozymes, which reflect both seed and pollen gene flow, have been used to analyse genetic structure in whitebark pine. These studies revealed low genetic differentiation among populations, with most diversity found within populations (Yandell 1992; Jorgensen & Hamrick 1997; Bruederle *et al.* 1998), as has been shown for most other species in the Pinaceae (Hamrick & Godt 1996).

Whitebark pine is the focus of increased attention because of its ecological importance and declining abundance over much of its range. This decline is mainly attributable to blister-rust disease (caused by *Cronartium ribicola*, an exotic fungal pathogen) and encroachment of shade-tolerant species [e.g. *Abies lasiocarpa* (Hook.) Nutt., and *Tsuga mertensiana* (Bong.) Carrière] associated with fire exclusion (Kendall & Arno 1990). As a result, most stands of whitebark pine have been decimated, and thus produce low seed yields. Because whitebark pine seed is a major late summer food source for grizzly bears (*Ursus arctos horribilis*), red squirrels (*Tamiasciurus hudsonicus*), and for many birds besides Clark's nutcracker, the decline of whitebark pine has had an adverse impact on subalpine ecosystems (Mattson *et al.* 2001).

As genetic diversity becomes depleted from blister-rust disease and succession, successful conservation and restoration programmes require an understanding of the biological and physical mechanisms that have shaped population genetic structure. Mitochondrial (mt) and chloroplast (cp) DNA provide tools for the elucidation of population structure and gene flow in plants. In the Pinaceae, uniparental inheritance of organelles allows an independent assessment of gene flow from paternally transmitted cpDNA via pollen and maternally transmitted mtDNA via

seed (Wagner 1992). This unusual inheritance of organelles has been a basis for studies of population genetic structure and gene flow in limber pine (*Pinus flexilis* James; Latta & Mitton 1997; Mitton *et al.* 2000) and ponderosa pine (*Pinus ponderosa* Dougl.; Latta & Mitton 1999). Both studies found greater differentiation of mtDNA than cpDNA haplotypes among populations. Thus, the ability to separate the components of gene flow in this plant family offers the potential for understanding the influences of Clark's nutcracker-mediated seed and wind-disseminated pollen movement on the genetic structure and biogeographic history of whitebark pine.

Information on the biogeographic history of whitebark pine is limited. Macrofossil records greater than 50 000 years before present (BP) have only been detected in Yellowstone National Park, where the species was evidently present during the Wisconsin glaciation (Baker 1990). During the late Pleistocene (20 000–18 000 years BP), whitebark pine was probably displaced southward and lower in elevation. In the southern reaches of the species range, isolated populations in central Oregon and northern Nevada are thought to be the remnants of a much larger subalpine woodland that existed during glaciation (Thompson 1990; Charlet 1991). During the late glacial period (15 000–10 000 years BP) Haploxylon pine pollen, probably from whitebark pine, occurred as far north as northern Idaho and northwestern Montana (Mack *et al.* 1978, 1983). By the early Holocene, whitebark pine macrofossils appear in sediment from the central Canadian Rocky Mountains (Beaudoin & King 1989) and the east slope of the Coast Mountains, indicating an invasion of formerly glaciated mountains as early as 10 000 years ago (Clague & Mathewes 1989).

In this study, we use mtDNA and cpDNA microsatellite (cpSSR) haplotype data to assess genetic structure and gene flow via seed and pollen. Our objectives were to: (i) contrast the genetic structure of whitebark pine revealed by uniparentally inherited and separately disseminated mitochondrial and chloroplast genomes; (ii) use molecular and palaeoecological data to develop hypotheses with regard to the biogeographical distributions and re-colonization routes of whitebark pine; and (iii) obtain information on patterns of genetic differentiation and diversity that would be potentially useful in future conservation efforts.

Materials and methods

DNA extraction, and polymerase chain reaction (PCR)

Whitebark pine DNA was extracted from 75 to 100 mg of frozen needle or bud tissue using a rapid CTAB method (Stewart & Via 1993). This procedure was modified for use with an electric tissue homogenizer. PCR contained 0.2 mM dNTPs, 4 mM MgCl₂, 0.5 µM of primers (fluorescent

labelled for cpSSRs), and 1 U of Amplitaq DNA polymerase (Applied Biosystems) in a final reaction volume of 20 μ L. Amplifications were performed following the method of Vendramin *et al.* (1996) for cpSSRs and Wu *et al.* (1998) for mtDNA, with either a Perkin Elmer model 480 or a Techne Progene thermal cyclers.

mtDNA sequencing

Two geographically distant populations of whitebark pine, from Klamath Falls, OR and Salmon, ID, were used to detect polymorphisms. Five primer pairs that flank mtDNA introns (*atp6*, *cox1*, *cox3*, *nad5a*, and *nad5d*), were selected from a set developed for California closed-cone pines (Wu *et al.* 1998). PCR products ranged from 600 to 1500 base pairs (bp), and were sequenced with an ABI 377 DNA sequencer (Applied Biosystems) at the Laboratory of Ecological and Conservation Genetics (L.E.C.G.), University of Idaho, Moscow, ID. The sequences were edited and aligned with SEQUENCHER software (Gene Codes Corp.). Corrected sequences were then compiled for each locus to detect polymorphisms.

One transversion was detected in the *nad5a* (NADH-ubiquinone oxidoreductase subunit 5) intron, and the *nad5d* intron contained two adjacent-base transversions. Nucleotide Blast (<http://www.ncbi.nlm.nih.gov/BLAST/>) determined that the *nad5a* polymorphic sequence was recognized by the restriction enzyme *MseI*, which was subsequently used to screen all other samples. The two adjacent polymorphic bases in *nad5d* did not correspond to a restriction enzyme; therefore, detection was conducted by sequencing with nested primers (*nad5d*-short F: 5'-gca gtc cgg ctg ctg ccg, R: 5'-ggg cgc agg cag ccc tct acc). Both polymorphic loci were used to determine the haplotype of 195 samples in 37 populations throughout the range. In addition, limber pine, another member of section *Strobus*, was sequenced with the *nad5a* and *nad5d* primers to infer the likely polarity of the polymorphisms in whitebark pine.

cpDNA microsatellites

Three cpDNA microsatellite primer pairs (PT15169, PT30204, and PT71936), developed for *Pinus thunbergii* Franco (Vendramin *et al.* 1996), produced products that were found to be polymorphic in an initial screening of a total of 10 whitebark pine samples from Salmon, ID and Klamath Falls, OR. Each polymorphic locus was checked for homoplasmy caused by length variation in the flanking regions of the microsatellite by sequencing the PCR products of four samples from Oregon and Idaho.

The visualization of cpSSR loci was accomplished by diluting 1 μ L of fluorescent dye-labelled PCR products with 2 μ L of formamide, loading buffer, and dye standards (GS350 tamra, Applied Biosystems). This mixture was

denatured at 95 °C for 2 min and loaded on a 36-cm, 6% polyacrylamide gel. Products were multiplexed and run for 2 h on an ABI 377 DNA sequencer. GENESCAN and GENOTYPER v. 2.5 software packages (Applied Biosystems) were used to calculate cpSSR length.

Sampling

Final sampling for cpSSR haplotypes was conducted in 28 populations aggregated into six regions: northern Cascades, southern Cascades, Sierra Nevada, Yellowstone, central Idaho and northern Idaho. Each region was represented by at least two sampling populations. Sample sizes ranged from 82 in the northern Cascades to 17 in the Sierra Nevada.

Genetic analysis

Because cpDNA is nonrecombinant, the combination of allele sizes at the three loci was assigned as a haplotype. Gene diversity, equivalent to expected heterozygosity (H_E) for diploid data, was calculated as \hat{H} :

$$\hat{H} = \frac{n}{n-1} \left(1 - \sum_{i=1}^k p_i^2 \right)$$

where n is the number of samples in the population, k is the number of haplotypes, and p is the population frequency of the i th haplotype (Nei 1987).

Population pairwise comparisons were conducted using a genetic distance approach based on haplotype frequency variance under the infinite alleles model (IAM). This model assumes that genetic drift is the driving force behind population divergence (Weir & Cockerham 1984; Michalakis & Excoffier 1996). In a second approach, the stepwise mutation model (SMM) was used, which makes pairwise comparisons of allele length differences among haplotypes (Slatkin 1995). Pollen gene flow (Nm) was calculated from the private alleles method of Slatkin (1985). This estimate is based on a linear relationship that exists between the logarithm of Nm and the average frequency of the private alleles, corrected for differences in sample size (Barton & Slatkin 1986).

The IAM was used with analysis of molecular variance (AMOVA) to calculate genetic partitioning. This analysis included fixation indices at three hierarchical levels: Φ_{SC} among populations within groups; Φ_{ST} among all populations; and Φ_{CT} among groups of populations (Weir & Cockerham 1984; Excoffier *et al.* 1992). Statistical significance was tested for pairwise population comparisons and AMOVA Φ values by bootstrap analysis. These procedures were performed with ARLEQUIN software (v. 2.0; Schneider *et al.* 2000). Gene flow (Nm) from private alleles was calculated using GENEPOP software (v. 1.2; Raymond & Rousset 1995).

Table 1 List of whitebark pine (*Pinus albicaulis* Engelm.) populations, locations, sample size per population of mtDNA and cpSSR haplotypes, and latitude/longitude co-ordinates

Region	Population	H	State/Province	<i>n</i> mtDNA	<i>n</i> cpSSR	Latitude (N)	Longitude (W)
Yellowstone	Hellroaring	•	MT	1	3	45°00'	110°27'
	Washburn	•	WY	4	7	44°48'	110°26'
	S. Baldy Mtn.	•	MT	1	2	45°28'	111°55'
	Black Butte	•	MT	2	2	44°56'	112°04'
	Little Bear	•	MT	2	3	45°35'	112°06'
	Sawtel Peak	•	ID	2	4	44°33'	111°26'
Central Idaho	Seven Devils	▲	ID	2	4	45°22'	116°29'
	Gospel Peak	▲	ID	1	5	45°37'	115°56'
	Eagle Cap	▲	OR	2	2	45°14'	117°33'
	Burnt Knob*	•▲	ID	2	2	45°42'	114°59'
	Lost Trail	•	ID	2	5	45°42'	113°58'
	Blacklead Mtn.	•	MT	2	2	46°38'	114°51'
	Pilot Knob*	•▲	ID	3	5	45°54'	115°42'
Northern Rockies	Gisborne	▲	ID	3	3	48°21'	116°45'
	Sheep Shed Mtn.	•	MT	1		47°31'	112°47'
	Sweet Grass	▲	MT	1		48°52'	111°09'
	Baree Mtn	•	MT	2	6	47°57'	115°32'
	Beaver Creek	•	ID	1		47°42'	115°46'
	Frozen Lake	▲	MT	1	1	48°59'	114°41'
	Farnham Peak		ID		4	48°51'	116°30'
	Mallard Larkin		ID		5	46°56'	115°31'
Salmo Lookout		WA		4	48°58'	117°06'	
Canadian Rockies	Paget Peak	▲	BC	2		51°26'	116°21'
	Jumbo Pass	▲	BC	2		50°20'	116°30'
	Scout Mtn.	▲	BC	2		49°04'	120°11'
Canadian Cascades	D'arcy	▲	BC	2		50°31'	122°34'
	Heckman Pass	▲	BC	4		52°32'	125°48'
Washington Cascades	Chinook Pass	□▲	WA	9		46°52'	121°31'
	Washington Pass	▲	WA	9		48°31'	120°39'
	Fox Mtn. Pass	▲	WA	15		48°05'	120°14'
	Rock Mtn	▲	WA	20		47°97'	120°58'
	Mission Ridge	▲	WA	12	28	47°18'	120°28'
	Manastash Ridge*	□▲	WA	27	28	47°04'	121°04'
	Ravens Roost	□	WA	27	26	47°01'	121°20'
Potato Hill	□	WA	8		46°19'	121°19'	
Oregon Cascades	Brown Mtn	□	OR	2	5	42°21'	122°23'
	Crater Lake		OR		3		
	Harriman	□	OR	2	3	42°21'	122°16'
	Pelican	□	OR	2	7	42°30'	122°08'
Sierras	Tioga Pass	□	CA	5	8	37°54'	119°15'
	Sonora Pass	□	CA	6	9	38°19'	119°38'

*Population polymorphic for mtDNA haplotypes (H), in a sample size of *n*. ▲, haplotype 1; •, haplotype 2; □, haplotype 3.

An exact test of population differentiation (Raymond & Rousset 1995) was used to assign cpSSR populations into AMOVA groups. This test assumes the null hypothesis of a random distribution of cpSSR haplotype frequencies among populations. Populations with significant differentiation from other populations ($P < 0.05$) were assigned to a separate AMOVA group. Populations that had no significant differentiation were assigned to the same AMOVA group.

Results

mtDNA haplotype distribution

Of the five mtDNA introns (*atp6*, *cox1*, *cox3*, *nad5a* and *nad5d*) sequenced, two (*nad5a* and *nad5d*) produced sequence polymorphisms yielding three haplotypes (GenBank accession no. AF434844) among 195 individuals surveyed (Table 1). Of the 37 populations sampled, 33 contained a single

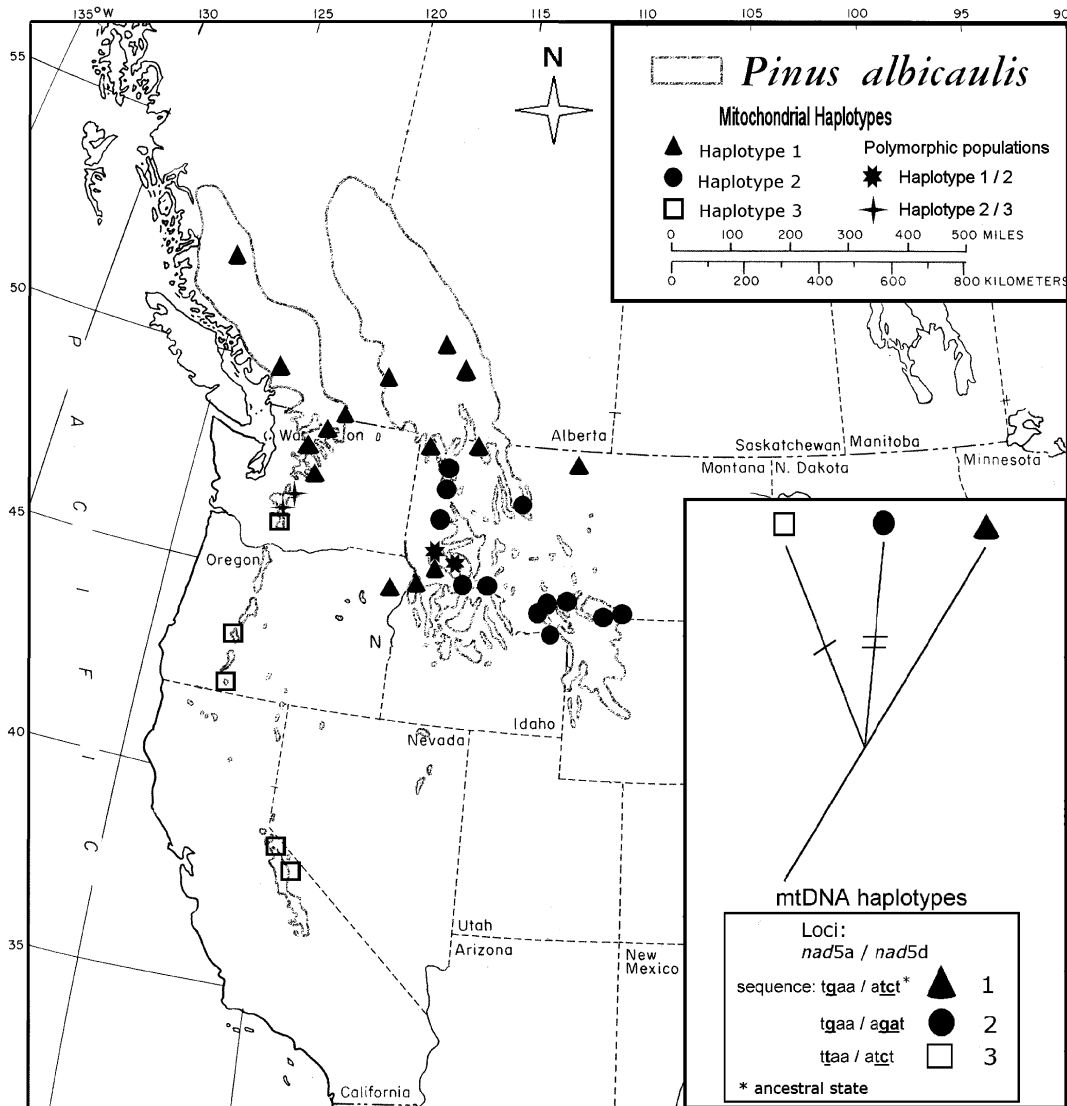


Fig. 1 Map of whitebark pine (*Pinus albicaulis* Engelm.) mtDNA haplotype distribution. Inset: a proposed phylogeny of mtDNA haplotype in whitebark pine based on polarity from *P. flexilis* James. Whitebark pine range derived from Little (1971).

haplotype, and four contained two haplotypes (Fig. 1). No individual exhibited signs of heteroplasmy, containing two or more biparentally inherited cytoplasmic haplotypes. The three mtDNA haplotypes differed in their geographical ranges. Haplotype 1 was the most wide ranging, occurring in the northern Cascades, central Idaho, eastern Oregon, and north into the Canadian Rockies and Coast Ranges. Haplotype 2 was found in the greater Yellowstone region and northward into the Bitterroot Range, where overlap occurred with haplotype 1. Haplotype 3 was exclusive to the Sierra Nevada and the central and southern Cascades. Contact of haplotypes 1 and 3 occurred in the vicinity of Mt. Rainier, WA. Haplotype 1 matched the *Pinus flexilis* outgroup at both polymorphic loci from which a simple phylogenetic tree for mtDNA was hand drawn (Fig. 1, inset).

cpSSR genetic diversity

A range-wide survey of 188 samples produced 42 cpSSR haplotypes in 28 populations that were aggregated into six regions for analysis. Unique haplotypes were found in 18 individuals, and at least one occurred in each region. The haplotypes were composed of three loci; locus pt15169 was the most variable with seven alleles, followed by pt30204 with six, and pt71936 with five. Microsatellite size variants typically conformed to a mono-repeat pattern (Fig. 2). No heteroplasmy was observed in the samples.

The cpSSR haplotypes 1 and 2 were the most common, occurring in 25 (13.4%) and 27 (14.5%) samples, respectively. Haplotype 1 was present in all regions except for the Sierra Nevada, and haplotype 2 was only absent in the

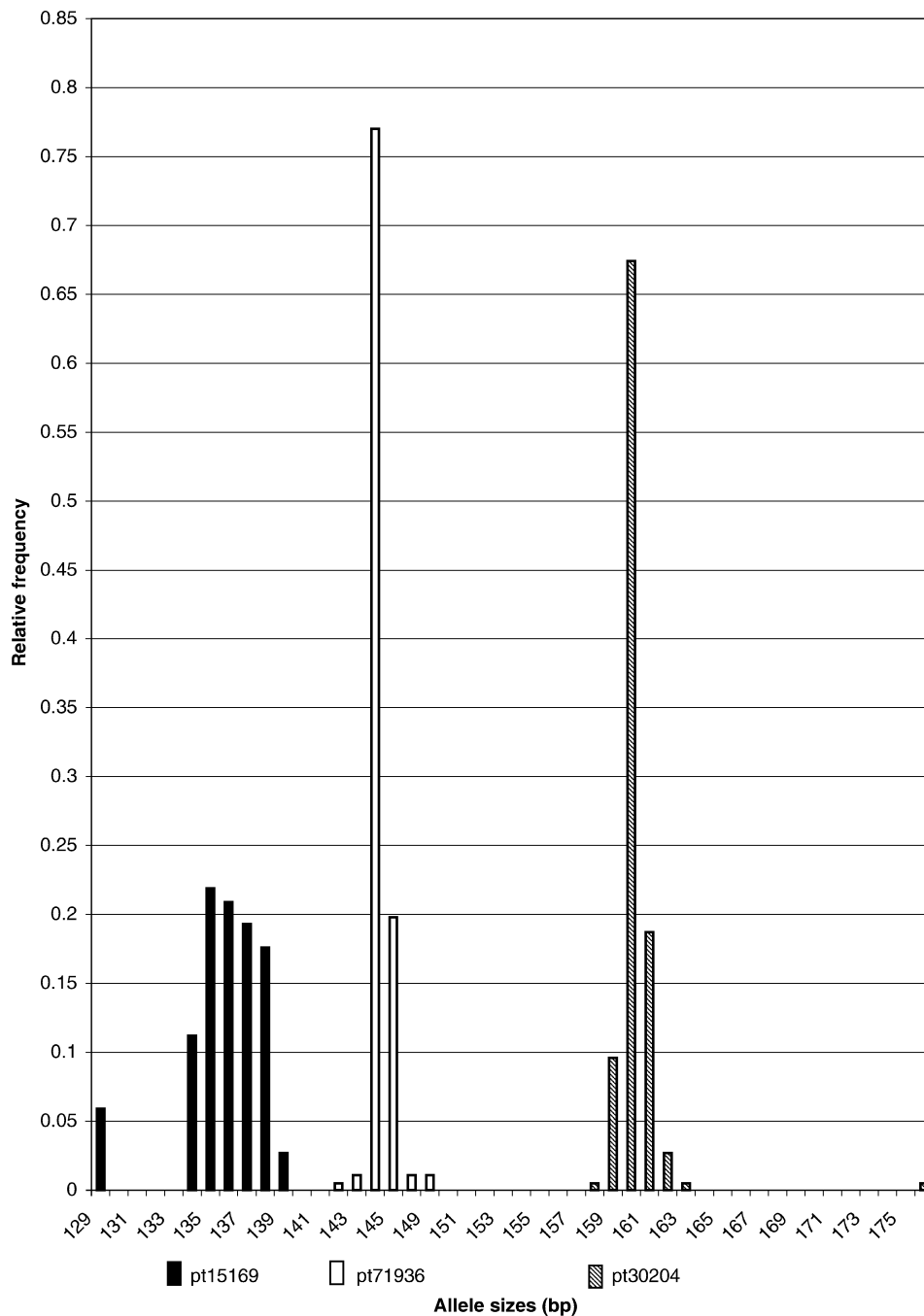


Fig. 2 Frequency distribution of allele sizes (base pairs) for the three mono-repeat cpSSR loci in whitebark pine (*Pinus albicaulis* Engelm.): pt15169, pt71936, pt30204.

Yellowstone region. Haplotype 7 had a total of 11 samples (5.9%), but was present in all regions. Private haplotypes within regions usually consisted of a single or low frequency occurrence. A notable exception was haplotype 23, which had a high frequency of 0.41 in the Sierra Nevada. Gene diversity (H) was high with a regional average of 0.91. Most H -values ranged above 0.90, except for the Sierra Nevada, which had an estimate of 0.79 (Table 2).

Pairwise population/region distances, gene flow and hierarchical genetic analyses

Using the IAM, pairwise genetic distances between regions based on cpSSRs resulted in low and insignificant genetic structure ($F_{ST} = 0-0.034$, $P > 0.05$) among regions from the northern Cascades (A), southern Oregon (B), central Idaho (E) and northern Idaho (F). Greater distances and significant

Table 2 The cpSSR haplotype gene diversity and sample size (*n*) in whitebark pine (*Pinus albicaulis* Engelm)

Region	<i>n</i>	Gene diversity
(A) Northern Cascades, WA	82	0.928
(B) Southern Oregon	18	0.915
(C) Sierra Nevada, CA	17	0.794
(D) Yellowstone, WY	24	0.938
(E) Central Idaho	25	0.940
(F) Northern Idaho	22	0.926
Average		0.907

genetic structure ($F_{ST} = 0.04-0.12$, $P < 0.05$) were found in most regional comparisons involving the Sierra Nevada (C). One exception was the F_{ST} value between the Sierra Nevada and those from southern Oregon ($F_{ST} = 0.053$, $P > 0.05$). This is not unexpected based on their proximity, but the lack of significance could also be due to the small sample sizes for these regions ($n < 20$). The Yellowstone region also showed significant genetic structure among all other regions except the Oregon Cascades. The most divergence was observed between the Sierra Nevada and Yellowstone regions ($F_{ST} = 0.127$, $P < 0.05$).

The private alleles estimations (Slatkin 1985) of Nm were generally consistent with the F_{ST} estimates (Table 3), which is attributable to the number of private haplotypes found in each region. Again, Sierra Nevada showed consistently low gene flow with other regions ($Nm = 0.41-1.36$). Using the SMM, most R_{ST} estimates produced negative values, while five comparisons had positive values and only one was statistically significant (data not shown). These negative values are due to the higher variance in allele size within, compared to between regions.

A population differentiation test analogous to Fisher's exact test based on haplotype frequencies (Raymond & Rousset 1995) produced only one cpSSR population, Sierra Nevada (C), that was significantly different ($P < 0.05$) from all other regions. Yellowstone also showed significant divergence from central Idaho and Sierra Nevada. Based

on the results of the exact test, the Sierra Nevada and Yellowstone regions were assigned to their own AMOVA groups, while the other four regions (northern Cascades, northern Idaho, central Idaho and southern Oregon) were assigned to a third group for genetic hierarchical analysis (Fig. 3). The overwhelming majority of the genetic variance (95.4%) was found within regions (Table 4). Among hierarchical levels: divergence among the three groups (Φ_{CT}) and among populations within groups (Φ_{SC}) was low ($\Phi_{CT} = 0.039$, $P < 0.05$; $\Phi_{SC} = 0.008$, $P < 0.001$) and among populations (Φ_{ST}) was moderate ($\Phi_{ST} = 0.046$, $P < 0.05$).

A local genetic analysis of cpSSRs was also conducted on three populations in the northern Cascades region. This area was of particular interest due to the abrupt geographical contact of mtDNA haplotypes 1 and 3. A sample of 45 at Manastash Ridge population produced 21 trees with haplotype 1 and 24 with haplotype 3. The Ravens Roost population (22 km to the southwest) was monomorphic for haplotype 3, and the Mission Ridge population (52 km to the north of Manastash Ridge) was monomorphic for haplotype 1. Despite distinct structure with mtDNA haplotypes, there was little genetic structure (F_{ST}) in cpSSR data among these populations (Manastash Ridge: Mission Ridge = 0.0049, Manastash Ridge: Ravens Roost = 0.0058, Ravens Roost: Mission Ridge = 0).

Discussion

Inferences from seed and pollen gene flow

The geographical distribution of whitebark pine mtDNA haplotypes is consistent with that expected for a species dispersed by Clark's nutcracker, for which the maximum seed-caching flights are estimated to be 22 km (Vander Wall & Balda 1981). The existence of disjunct regions with distinct mtDNA haplotypes (e.g. Yellowstone and the Sierra Nevada) suggests that large areas of unsuitable habitat (> 100 km) impose a significant barrier to seed dispersal. In contrast, cpSSR genetic structure is low across mtDNA contact zones and among regions suggesting that

Table 3 The cpSSR genetic distances and gene flow among six whitebark pine (*Pinus albicaulis*, Engelm.) regions. Values in the upper diagonal are Nm estimates from private alleles method; F_{ST} distances appear in the lower diagonal

Region	Northern Cascades	Southern Oregon	Sierra Nevada	Yellowstone	Central Idaho	Northern Idaho
Northern Cascades		4.68	2.19	6.56	13.56	10.90
Southern Oregon	0.012		0.41	7.19	1.68	3.83
Sierra Nevada	0.046*	0.053		0.51	0.83	1.36
Yellowstone	0.034*	0.000	0.127*		1.17	4.27
Central Idaho	0.017	0.034	0.082*	0.07*		5.37
Northern Idaho	0.000	0.013	0.032	0.047*	0.008	

* $P < 0.05$.

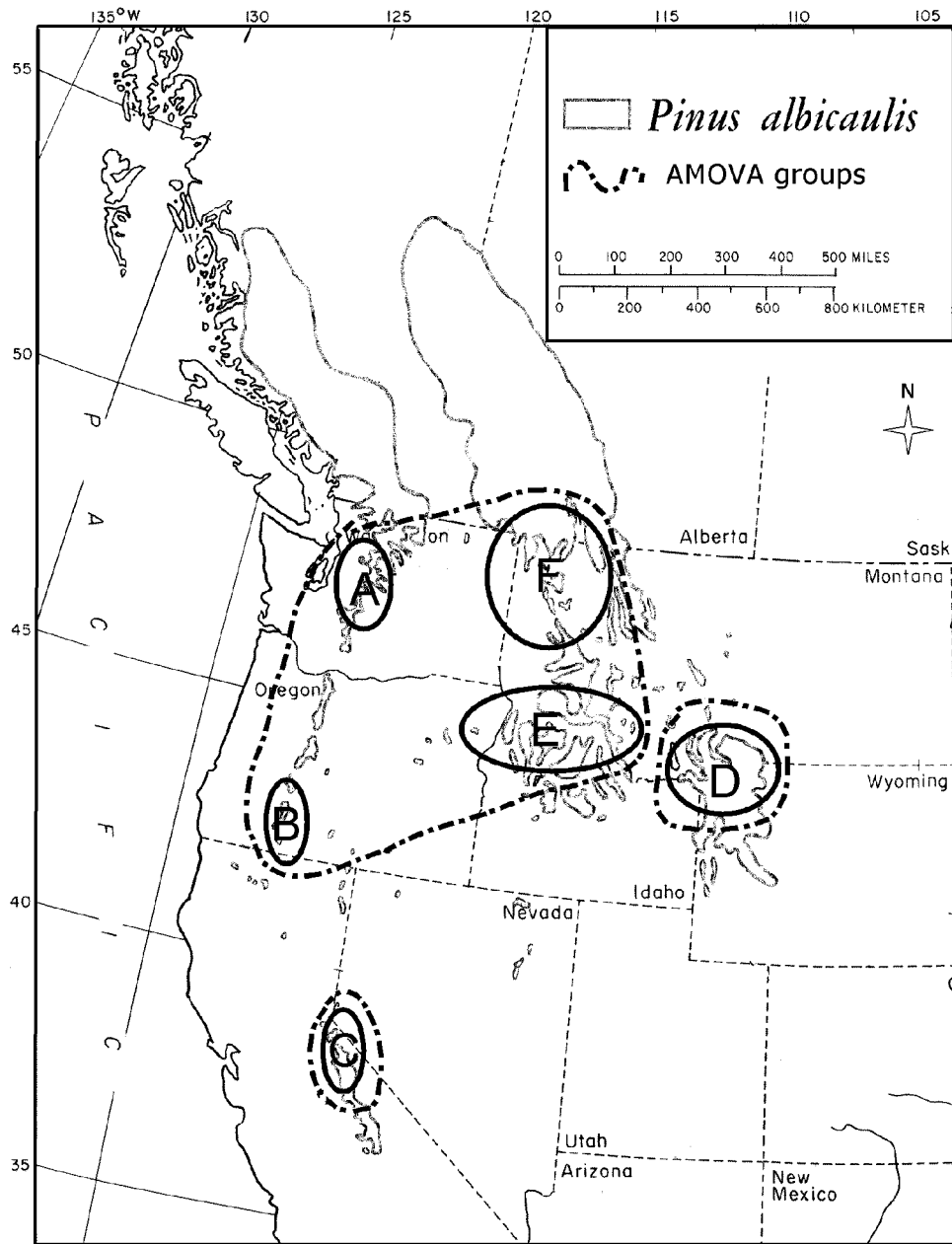


Fig. 3 Areas of sampling for whitebark pine (*Pinus albicaulis* Engelm.) cpSSR regions and AMOVA groups based on an exact test are circumscribed. Region A, northern Cascades; B, southern Oregon; C, Sierra Nevada Mountains; D, Yellowstone; E, central Idaho; F, northern Idaho. Whitebark pine range derived from Little (1971).

the majority of gene flow is via pollen. The narrowness of contact zones and general lack of mixing among mtDNA haplotypes suggest that nutcrackers play a specific role in the overall gene flow. These data and those collected from seed-caching behaviour (e.g. Tomback 1978; Vander Wall & Balda 1981) indicate that nutcrackers typically deposit seed within the same population; however, rare interpopulation seed caching flights may establish new populations, and over longer time scales are responsible for range expansions (see below). Genetic heterozygosity

of these expanding populations is apparently maintained by paternal gene flow.

Glacial and postglacial biogeography

Climate change between the late Pleistocene and early Holocene had a dramatic impact on the biogeography of whitebark pine. At lower latitudes, full-glacial climates apparently led to a sizeable expansion of subalpine and montane habitats, which were displaced as much as

Table 4 AMOVA table using genetic distance based on haplotype frequencies (F_{ST}) in whitebark pine (*Pinus albicaulis* Engelm)

Source of variation	Degrees of freedom	Sum of squares	Variance components	Percentage of variation
Among groups	2	4.212	0.03406	3.89
Among populations within groups	3	3.109	0.00653	0.075
Within populations	181	151.123	0.83493	95.36
Total	186	158.444	0.87552	

	Fixation indices	<i>P</i> -value
Φ_{SC}	0.00776	0.008
Φ_{ST}	0.04636	0.031
Φ_{CT}	0.0389	0.048

1000 m below the present limits (Thompson 1990; Charlet 1991). This climate probably favoured the expansion of whitebark pine into the northern Great Basin and central Oregon. As warming occurred, whitebark pine from these lower latitude regions were reduced to remnant populations on the highest peaks. Meanwhile, in the northern Rocky Mountains and Cascades, the retreating ice provided new expanses of subalpine habitat that were colonized by whitebark pine, as indicated by macrofossils found in early Holocene strata (Beaudoin & King 1989; Clague & Mathewes 1989).

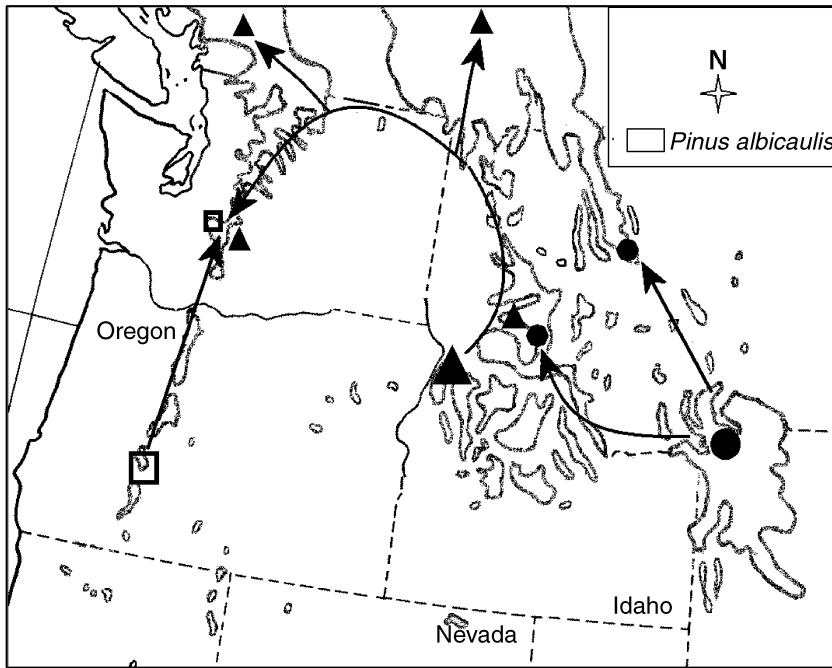
Whitebark pine genetic structure and diversity should reflect the patterns of expansion into formerly glaciated regions. The distribution of mtDNA haplotypes in the glaciated northern Cascades is not consistent with the hypothesis proposed by Jorgensen & Hamrick (1997) that a late-glacial refugium persisted near Mount Rainier, WA. Their hypothesis was based on higher heterozygosity found in allozymes at this site. The distribution of mtDNA haplotypes supports an alternative hypothesis: whitebark pine was absent in the northern Cascades during the late Pleistocene and was colonized in the Holocene from two geographically separate regions associated with mtDNA haplotypes 1 and 3. This hypothesis predicts that haplotype 1 expanded westward across subalpine habitat in north-central Washington, and then spread northward and southward into the Cascades. Meanwhile, haplotype 3 was proceeding northward into the Washington Cascades (Fig. 4). At present, the narrow contact zone of haplotypes 1 and 3 occurs along a 100-km area extending from Chinook Pass, WA to Snoqualmie Pass, WA. The high allozyme heterozygosity detected by Jorgensen and Hamrick in this area could be the result of the contact between two formerly isolated regions of whitebark pine. The westward colonization of the northern Cascades from Idaho is also consistent with the low pairwise F_{ST} and high gene flow estimates (Tables 3 and 4); this hypothesis is also supported by the lack of whitebark macrofossils from late

Pleistocene strata in this region (Dunwiddie 1986). It would be expected that a glacial refugium in the northern Cascades would produce a wider zone containing both mtDNA haplotypes 1 and 3 in the northern Cascades after Holocene expansion, and genetic differentiation would probably be higher between Idaho and this region due to isolation.

The Yellowstone region, which contains the oldest known whitebark pine macrofossils (50 000 years BP; Baker 1990), is likely to have supported populations containing mtDNA haplotype 2 during the Pleistocene. This population apparently expanded northward along the Continental Divide, where it mixed with haplotype 1 in the Bitterroot range and central Idaho (Fig. 4). Expansion of haplotype 2 could potentially have occurred earlier in the Pleistocene; however, under such a scenario haplotype 2 would have probably disseminated into Canada and the northern Cascades along with haplotype 1.

The cpSSR data, including relatively high gene flow and low genetic structure, suggest that considerable paternal gene flow has occurred between Idaho and southern Oregon, although it appears that seed dispersal (and associated mtDNA) did not span this gap. The climate of the late Pleistocene may have provided whitebark pine with suitable habitat to extend across central Oregon, providing a corridor for gene flow (Fig. 5). In addition, this hypothesis is supported by the exact test that showed insignificant differentiation between these regions. Additional sampling of mtDNA from mountaintop island populations of whitebark pine across central Oregon may provide additional insight into this hypothesis.

Although whitebark pine populations in the Cascades may have had historical ties to Idaho populations via two corridors (north-central Washington and central Oregon), the Sierra Nevada region apparently exhibited limited gene flow with other regions. Geographic isolation and a possible founder effect may have created comparatively low genetic diversity (Table 2) and high genetic divergence that



▲ = haplotype 1, ● = haplotype 2, □ = haplotype 3

Fig. 4 Hypothesized postglacial colonization of mtDNA haplotypes associated with bird-dispersed seed in whitebark pine (*Pinus albicaulis* Engelm.). Symbols represent mtDNA haplotypes, postglacial colonization routes (arrows) from three putative glacial refugia, and subsequent secondary contact in the Cascade Mountains.

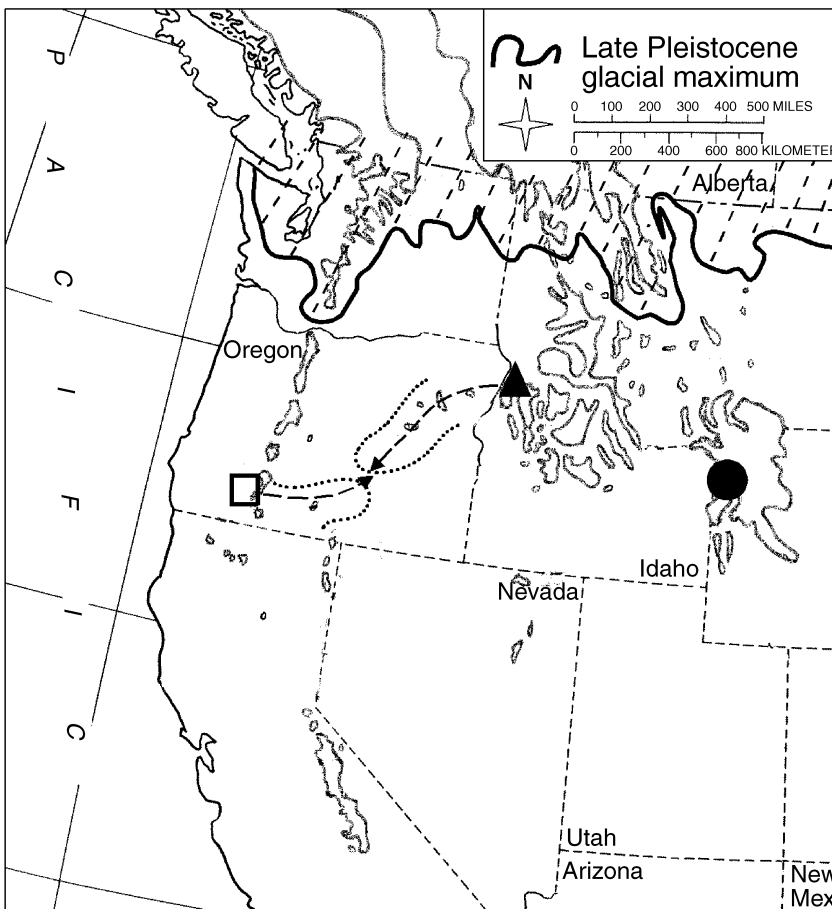


Fig. 5 Hypothesized distribution of mtDNA haplotypes at glacial maximum (Porter *et al.* 1983). Arrows represent hypothesized cpDNA (pollen) gene flow across suitable whitebark pine habitat (dotted line) that likely existed during this time.

distinguishes the Sierra Nevada (Table 3). Sierra Nevada whitebark pine and smaller islands in northern Nevada could be considered contemporary refugia that could be lost if climatic warming or other ecological threats persist.

Evaluation of infinite alleles and stepwise mutation models

This study compared the IAM and SMM and found that the former yielded results that are more biologically reasonable in light of whitebark pine biogeography and the known climatic history of the region. Genetic distances estimated using the R_{ST} under the SMM are based on pairwise, size-length differences in haplotype loci. Although the SMM might appear to be suitable given the stepwise pattern of microsatellite lengths (Fig. 2), no divergence among most regions was detected with this method, apparently due to the generally high allele size variance within regions that produces negative values (data not shown). Differences among most populations were detected using the IAM, which estimates genetic distances based on haplotype frequencies (F_{ST} ; Table 3). One caveat in the estimation of F_{ST} under the IAM is that low-level homoplasy could potentially create identical haplotypes in different populations, thereby causing an underestimation of the genetic structure.

The lack of resolution using R_{ST} may indicate that a number of assumptions are not met for cpSSRs in whitebark pine. For detection of population divergence using R_{ST} , Goldstein *et al.* (1995) reported that 500 generations are needed before the SMM is effective in detecting genetic differentiation at nuclear microsatellite loci. This estimate assumes a mutation rate of one per 10^4 gametes and moderately sized populations. At present, cpSSR mutation rates are unknown, although it has been estimated that cpDNA mutation rates are twice as slow as nuclear DNA (Wolfe *et al.* 1987; Wang *et al.* 2000). This apparently slower mutation rate in cpSSRs coupled with a long generation time (≈ 30 years; Krugman & Jenkinson 1974) necessitates an exceedingly long time frame before genetic differentiation is detected using R_{ST} . Other studies have reported low resolution of R_{ST} with populations that have more than one migrant per generation, suggesting that F_{ST} estimates are a more sensitive means by which to detect the genetic structure of allopatric or recently diverged populations (Forbes *et al.* 1995). Given the apparent changes in biogeography and high gene flow among populations, the conditions needed to fit SMM are unlikely to occur in existing whitebark pine populations.

Implications for conservation and management

The biogeography of whitebark pine has been profoundly shaped by geographical isolation and by large shifts in the

subalpine ecosystems associated with climate change and glaciation. These movements have been facilitated by nutcracker-mediated seed dispersal over thousands of years. Populations existing before Holocene warming, which resided south of the Cordilleran ice sheet (e.g. Sierra Nevada, Yellowstone, Idaho and northern Nevada), probably hold the majority of genetic resources. In contrast, populations north of the extent of glaciation (Fig. 5) have colonized these regions relatively recently. Cumulative founding events during this northward expansion are expected generally to reduce genetic diversity (Petit *et al.* 2001). From a gene conservation perspective, these populations would be likely to have lower priority.

Recovery plans in whitebark pine have emphasized conducting breeding programmes for blister-rust resistance to supplement populations with high mortality, and the creation of canopy openings through silvicultural treatments or prescribed fire to perpetuate natural regeneration by nutcrackers (Keane and Arno 2001; Hoff *et al.* 2001). However, it is critically important that these recovery plans consider the existing genetic structure (Hoff *et al.* 2001). Substantial population genetic structure from cpSSR data warrants caution in transfer of whitebark pine material among regions (e.g. Yellowstone to Idaho, or Sierra Nevada to others regions). Until ongoing whitebark pine provenance tests provide essential information on genetic and environmental interactions, discretion should be used in the movement of seed or the full-scale implementation of blister-rust resistance breeding programmes.

Acknowledgements

This project was funded by the Rocky Mountain Research Station, Moscow Idaho, Soil, Pathology, Microbe Unit (RM-4552), and a grant provided by Donna Dekker, Forest Ecosystem Processes Unit (RM-4155), Research Joint Venture Agreement No. INT-95123-RJVA. We would like to thank Dr Lisette Waits for her insights throughout this project. Special thanks to Dr Mary Frances Mahalovich and Jodie Krakowski for acquiring samples; Dr Mee-Sook Kim for laboratory expertise; Paul Joyce for statistical advice and Marcus Warwell, John Hanna, Jill Hoff and Terry Miller for assistance with collections. We would also like to thank three anonymous reviewers for their helpful comments.

References

- Baker RG (1990) Late Quaternary history of whitebark pine in the Rocky Mountains. In: *Proceedings – Symposium on Whitebark Pine Ecosystems: Ecology and Management of a High-Mountain Resource. General Technical Report. INT-270* (ed. Schmidt WC), pp. 40–48. USDA Forest Service, Washington DC.
- Barton NH, Slatkin M (1986) A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. *Heredity*, **56**, 409–415.
- Beaudoin AB, King RH (1989) *Late quaternary vegetational history of Wilcox Pass, Jasper National Park, Alberta*, unpublished paper. Geology Department, University of Iowa, Iowa City IA.

- Bruederle LP, Tomback DF, Kelly KK, Hardwick RC (1998) Population genetic structure in bird-dispersed pine, *Pinus albicaulis* (Pinaceae). *Canadian Journal of Botany*, **76**, 83–90.
- Charlet DA (1991) *Relationships of the Great Basin alpine flora: a quantitative analysis*. Master's Thesis, University of Nevada, Reno.
- Clague JJ, Mathewes RW (1989) Early Holocene thermal maximum in western North America: new evidence from Castle Peak, British Columbia. *Geology*, **17**, 277–280.
- Delcourt HR, Delcourt PA (1991) Gradients, continua and ecotones. In: *Quaternary Ecology, a Paleocological Perspective* pp. 93–121. Chapman & Hall, New York.
- Dunwiddie PW (1986) A 6000-year record of forest history on Mount Rainier, Washington. *Ecology*, **67**, 51–68.
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: applications to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.
- Forbes SH, Hogg JT, Buchanan FC, Crawford AM, Allendorf FW (1995) Microsatellite evolution in congeneric mammals: domestic and bighorn sheep. *Molecular Biology and Evolution*, **12**, 1106–1113.
- Furnier GR, Knowles P, Clyde MA, Dancik BP (1987) Effects of avian seed dispersal on the genetic structure of whitebark pine populations. *Evolution*, **41**, 607–612.
- Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellites. *Genetics*, **139**, 463–471.
- Hamrick JL, Godt MJ (1996) Conservation genetics of endemic plant species. In: *Conservation Genetics: Case Histories from Nature* (eds Avise JC, Hamrick JL), pp. 281–333. Chapman & Hall, New York.
- Hoff RJ, Ferguson DE, McDonald GI, Keane RE (2001) Strategies for managing whitebark pine in the presence of white pine blister rust. In: *Whitebark Communities: Ecology and Restoration* (eds Tomback DF, Arno SF, Keane RE), pp. 346–366. Island Press, Washington DC.
- Jorgensen SM, Hamrick JL (1997) Biogeography and population genetics of whitebark pine, *Pinus albicaulis*. *Canadian Journal of Forestry Research*, **27**, 1574–1585.
- Keane RE, Arno SF (2001) Restoration concepts and techniques. In: *Whitebark Communities: Ecology and Restoration* (eds Tomback DF, Arno SF, Keane RE), pp. 367–400. Island Press, Washington DC, USA.
- Kendall KC, Arno SF (1990) Whitebark pine—an important but endangered wildlife resource. In: *Proceedings—Symposium on Whitebark Pine Ecosystems: Ecology and Management of a High-Mountain Resource. General Technical Report. INT-270* (ed. Schmidt WC), pp. 264–263. USDA Forest Service, Washington DC.
- Krugman SL, Jenkinson JL (1974) *Pinus L. Pine*. In: *Seeds of Woody Plants in the United States USDA Forest Service, Agriculture Handbook no. 450* (ed. Schopmeyer CS), pp. 598–638. USDA Forest Service, Washington DC.
- Latta RG, Mitton JB (1997) A comparison of population differentiation across four classes of gene markers in limber pine (*Pinus flexilis* James). *Genetics*, **146**, 1153–1163.
- Latta RG, Mitton JB (1999) Historical separation and present gene flow through a zone of secondary contact in ponderosa pine. *Evolution*, **53**, 769–776.
- Little EL (1971) *Atlas of United States Trees: Conifers and Important Hardwoods*. USDA Forest Service Publication no. 1146. USDA Forestry Service, Washington DC.
- Mack RN, Rutter NW, Bryant VM Jr, Valastro S (1978) Reexamination of postglacial vegetation history in northern Idaho: Hagar Pond, Bonner County. *Quaternary Research*, **10**, 241–255.
- Mack RN, Rutter NW, Valastro S (1983) Holocene vegetation history of the Kootenai River Valley, Montana. *Quaternary Research*, **20**, 177–193.
- Mattson DJ, Kendall KC, Reinhart DP (2001) Whitebark pine, grizzly bears, and red squirrels. In: *Whitebark Communities: Ecology and Restoration* (eds Tomback DF, Arno SF, Keane RE), pp. 121–136. Island Press, Washington DC, USA.
- Michalakis Y, Excoffier L (1996) A generic estimation of population subdivision using distances between alleles with special reference to microsatellite loci. *Genetics*, **142**, 1061–1064.
- 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.
- Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Petit RJ, Bialozyt R, Brewer S, Cheddadi R, Comps B (2001) From spatial patterns of genetic diversity to post-glacial migration processes in forest trees. In: *Integrating Ecological and Evolutionary Processes in a Spatial Context* (eds Silvertown J, Antonovics J), pp. 295–318. Blackwell Science, Oxford.
- Porter SC, Pierce KL, Hamilton TD (1983) Late Wisconsin mountain glaciation in the western United States. In: *Late-Quaternary Environments of the United States* (ed. Porter SC), pp. 71–111. University of Minnesota, Minneapolis.
- Raymond M, Rousset F (1995) GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.
- Rogers DL, Millar CI, Westfall RD (1999) Fine-scale genetic structure of whitebark pine (*Pinus albicaulis*): associations with watershed and growth form. *Evolution*, **53**, 74–90.
- Schneider S, Roessli D, Excoffier L (2000) *ARLEQUIN, Version 2.0: A software for population genetics data analysis*. Genetics and Biometry Laboratory, University of Geneva, Geneva Switzerland.
- Slatkin M (1985) Rare alleles as indicators of gene flow. *Evolution*, **39**, 53–65.
- Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.
- Stewart C Jr, Via L (1993) A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. *Biotechniques*, **14**, 748–749.
- Thompson RS (1990) Late Quaternary vegetation and climate in the Great Basin. In: *Packrat Middens: the Last 40,000 Years of Biotic Change* (eds Betancourt JL, Van Devender TR, Martin PS), pp. 200–239. University of Arizona Press, Tucson.
- Tomback DF (1978) Foraging strategies of Clark's nutcracker. *Living Bird*, **16**, 123–161.
- Tomback DF (2001) Clark's nutcracker: agents of regeneration. In: *Whitebark Communities: Ecology and Restoration* (eds Tomback DF, Arno SF, Keane RE), pp. 89–104. Island Press, Washington DC.
- Tomback DF, Linhart YB (1990) The evolution of bird-dispersed pines. *Evolutionary Ecology*, **4**, 185–219.
- Vander Wall SB, Balda RP (1981) Ecology and evolution of food-storage behavior in conifer-seed-caching corvids. *Zeitschrift Fur Tierpsychologie*, **56**, 217–242.
- Vendramin GG, Lelli L, Rossi P, Morgante M (1996) A set of primers for the amplification of 20 chloroplast microsatellites in Pinaceae. *Molecular Ecology*, **5**, 595–598.
- Wagner DB (1992) Nuclear, chloroplast and mitochondrial DNA polymorphisms as biochemical markers in population genetic analysis of forest trees. *New Forests*, **6**, 373–390.

Wang X, Tank DC, Sang T (2000) Phylogeny and divergence times in the Pinaceae: evidence from three genomes. *Molecular Biology and Evolution*, **17**, 772–781.

Webb T (1995) Pollen records of Late Quaternary vegetation change: Plant community rearrangements and evolutionary implications. In: *Studies in Geophysics: Effects of Past Global Change on Life*, pp. 221–232. National Academy Press, Washington DC.

Weir BS, Cockerham CC (1984) Estimating *F*-statistic for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Wolfe KH, Li WH, Sharpe PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondria, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the USA*, **84**, 9054–9058.

Wu J, Krutovskii KV, Strauss SH (1998) Abundant mitochondrial genome diversity, population differentiation and convergent evolution in pines. *Genetics*, **150**, 1605–1614.

Yandell UG (1992) *An allozyme analysis of whitebark pine*. Master's Thesis, University of Nevada, Reno.

Bryce Richardson, a biological technician with the USDA Forest Service, has research interests in population and ecological genetics of forest trees and associated pathogens. This research was part of his MS thesis at the University of Idaho, Department of Forest Resources. Steve Brunsfeld is an Associate Professor in the Department of Forest Resources with research interests in conservation genetics, phylogeography and molecular systematics, especially of woody plants. Ned B. Klopfenstein is a Research Plant Pathologist in the USDA Forest Service with research interests in the genetics of host–pathogen interactions within forest ecosystems.
