Lineage Divergence in Coast Redwood (*Sequoia sempervirens*), Detected by a New Set of Nuclear Microsatellite Loci

Author(s): Vladimir Douhovnikoff and Richard S. Dodd

Lineage Divergence in Coast Redwood (*Sequoia sempervirens*), Detected by a New Set of Nuclear Microsatellite Loci

VLADIMIR DOUHOVNIKOFF

Department of Biology, Simmons College, Boston, Massachusetts 02115

AND

RICHARD S. DODD[1]

Environmental Science, Policy and Management, University of California, Berkeley 94720

ABSTRACT.—We developed a suite of tetranucleotide microsatellite loci and applied them to a study of genetic variation across the geographic range of coast redwood [*Sequoia sempervirens* (D. Don) Endl.]. The objectives of the study were to determine if the microsatellite loci could provide useful information on genetic diversity in this hexaploid species and to investigate earlier reports of divergent lineages within coast redwood. The microsatellite loci were highly variable, yielding a total of 142 alleles. Up to six alleles were detected in each individual consistent with the ploidy of coast redwood and suggesting that parental genomes must have been at least partially homologous. This does not rule out autoployploidy in the evolution of the redwood genome. We treated the microsatellite alleles as presence-absence data and we also estimated full genotypes assuming peak intensities varied with allele dosage. Both types of analyses revealed similar trends. Variation within the 17 watersheds sampled, explained most of the genetic diversity, with less than 4% of the variation attributable to watersheds. Our data showed a weak divergence between more or less continuous populations north of 36.8°N (the Sonoma- Mendocino county border) and disjunct populations south of this latitude. This is further north than indicated from earlier studies of marker systems that would be under selection and may reflect a demographic break. In view of the importance of clonal growth, we suggest that redwood may have difficulty adapting to new climatic conditions or of migrating into displaced habitats with anticipated climate change. Furthermore, the southern lineage of populations is likely to be at greatest risk and is therefore of conservation priority.

INTRODUCTION

Climate change scenarios predict modest to major changes in distributions of species over the next century. The magnitude of these changes depends on circulation model and assumed emissions scenarios, on the ability of species to disperse and on their altitudinal and latitudinal distribution. A recent study of the endemic flora of California suggests that species diversity will shift towards coastal areas, particularly in north-western California (Loarie *et al.*, 2008) and under extreme and moderate climate change, the majority of coastal species are predicted to experience northward range shifts. Although niche modeling studies provide valuable indications of potential effects of climate change on species' distributions, they rarely take account of genetic divergence among populations, or regional groups of populations, that might result in differential success in the new localities even if climatic conditions are the same. Restriction and isolation of populations during past environmental changes have commonly resulted in diverse lineages. These divergent lineages may have gene complexes that are well-suited to a suite of local environmental variables. This is particularly true for adaptations that require response to environmental

---

[1] Corresponding author: e-mail: dodd@berkeley.edu

cues correlated with latitude (*e.g.,* daylength). Unfortunately, information for such adaptations is rarely available and requires long-term studies, particularly for tree species. However, data on genetic variation of neutral markers is relatively easily attainable and can provide baseline information on the degree of structure among populations and whether this structure is discrete, or continuous in a spatial context.

The complex geological and climatic history of California has resulted in a number of vicariance events that have produced some common phylogeographic breaks among animal taxa (Rissler *et al.*, 2006; Chatzimanolis and Caterino, 2007). Among these, the Central Valley is the most important and in coastal California, prominent disjunctions occur at the Transverse Ranges, the San Francisco Bay and Russian River. These coastal disjunctions are probably a result of population divergence throughout long periods of isolation during the Pleistocene (Rissler *et al.*, 2006). Although more commonly detectable in animal taxa, phylogeographic breaks have also been detected in plants (Calsbeek *et al.*, 2003; Nettel *et al.*, 2009). Significant coastal disjunctions were detected at the San Francisco Bay and Mad River near Arcata for the tree *Notholithocarpus densiflorus* (syn. *Lithocarpus densiflorus*), a California Floristic Province (CFP) endemic (Nettel *et al.*, 2009). We are investigating potential common phylogeographic patterns of coastal California species, and we report here on genetic diversity in coast redwood [*Sequoia sempervirens* (D. Don) Endl.], one of the most characteristic components of north coastal forests.

Coast redwood is a California Floristic Province endemic, currently restricted to the coastal belt from central California to southern Oregon. It is unusual among conifers in having a hexaploid genome (Stebbins, 1948; Fozdar and Libby, 1968; Saylor and Simons, 1970; Ahuja and Neale, 2002; Ahuja, 2009). The high proportion of bivalents at meiosis with the presence of some multivalents led Stebbins (1948) to propose an autoallopolyploid or segmental allopolyploid origin. Karyotype analysis supported autoallopolyploidy (Saylor and Simons, 1970) and allozyme inheritance patterns provided some evidence for chromosomal duplication that would be consistent with autopolyploidy in the evolution of the genome of redwood (Rogers, 1997). Ahuja and Neale (2002) caution that chromosomal studies have been limited in numbers of individuals studied and that the evolution of redwood may include several episodes of polyploidization, not all of which were necessarily derived in the same manner.

Genetic diversity of growth traits and isozymes are reported to be high compared with other conifers (Anekonda, 1992; Rogers, 1997, 2000), but surprisingly little is known of the structure of genetic diversity throughout the range of the species. Two lines of evidence suggest that populations may be divided into southern and northern lineages with a break at the San Francisco Bay. First, growth performance in a common garden close to the San Francisco Bay was reported to be greater in more southerly provenances (Anekonda, 1992) and support for this has been found in provenance tests planted around the world (Kuser *et al.*, 1995). Secondly, Hall and Langenheim (1987) detected a break in foliar monoterpene concentration north and south of the San Francisco Bay, with a second possible disjunction in northern Mendocino County. However, neither single common garden growth performance, nor monoterpene profiles can distinguish local or regional adaptation from historical demographic processes. To detect lineages resulting from historical demographic processes, genetic variations at neutral markers are preferable. A suite of microsatellite markers was developed for redwood by Bruno and Brinegar (2004). However, of the five loci reported, only two showed moderate levels of polymorphism. Therefore, we undertook development of additional loci that would provide a set of hypervariable markers for a range of possible studies in redwood.
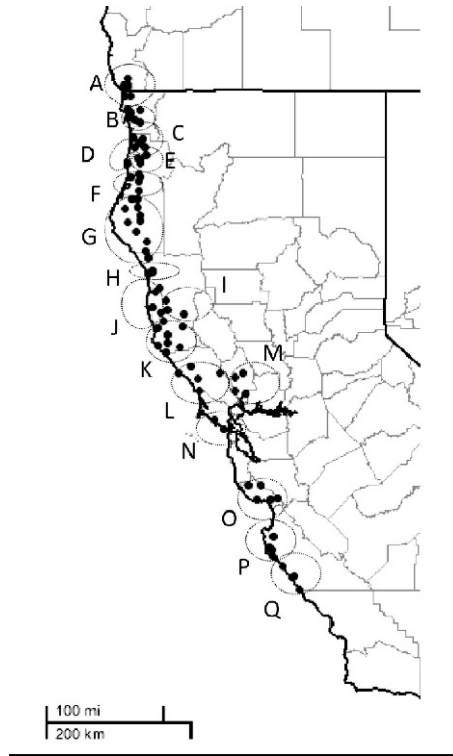
Fig. 1.—Map of collection sites and watershed groups. Sampling locations represented by black dots can include multiple samples. Watershed groups delineated by hatched ellipses are identified by letters

Here, we report on the development of six hypervariable nuclear microsatellites. We apply these microsatellite loci to DNA from a collection of redwood clones covering the entire geographic range of coast redwood with the objectives: (1) to test their suitability to detect genetic variation in this hexaploid species, and (2) to assess genetic structure across the geographic range of the species and test for possible phylogeographic breaks. Detection of divergent lineages will be important for redwood management for commercial and conservation purposes.

METHODS

Range-wide collection at Russell Reserve.—In 1984 a range-wide collection of 180 *Sequoia sempervirens* clones was gathered for planting in an international common garden study. One trial was located at Russell Reserve with 175 samples. Original collections consisted of seedlings, saplings and sprouts from naturally occurring stands. Eighty-one stands within 12 counties from Curry County in Oregon south to Monterey County in California were sampled (Fig. 1). In most stands two or more samples were collected. Cuttings from collections were then vegetatively propagated to produce ramets for test plantations. Details of the original sampling, propagation and outplanting for the common garden are described in detail by Kuser *et al.* (1995). The number of samples from each population are

listed in Table 2. Over 20 y the plantation at Russell Reserve has been a common sample set for the study of redwood physiology (Anekonda *et al.*, 1993) and genetics (Hall and Langenheim, 1987; Anekonda, 1992). Each additional data set on these genotypes adds a new dimension to our knowledge and potential to leverage earlier work, which increases the scientific and conservation value of the collection itself.

The Russell Reserve is located approximately 15 km east of San Francisco Bay and 5 km from the nearest native redwood stand. On the eastern side of the Berkeley Hills the site generally has drier summers, colder winters and less fog than most native redwood stands. We sampled foliage from 135 genetically distinct ramets in the Russell Reserve collection and assigned them to one of 17 watersheds in coastal California (Fig. 1).

Development of microsatellite markers for redwood.—DNA was extracted from foliage of redwood following a simplified CTAB method (Cullings, 1992). Microsatellite clones were obtained and screened for biotin-enriched genomic DNA as described by Khasa *et al.* (2000). Briefly, five hundred microlitres of redwood DNA from a single tree was purified by cesium chloride density gradient ultracentrifugation (Sambrook *et al.*, 1989). Genomic DNA was restricted and ligated in a reaction mixture containing the restriction endonuclease *Hae*III (Pharmacia) and the oligonucleotide linkers M28, 5′-CTCTTGCTTGAATTCG-GACTA and M29, 5′-pTAGTCCGAATTCAAGCAAGAGCACA. A portion of the restriction-ligation reaction mixture was denatured and hybridized with 5′ biotin-labeled oligonucle-otide mix (TG_{12}, GA_{12}, AAG_8, AAC_8 and $GATA_6$). Biotin-captured DNA was collected with Dynal M280 streptavidin magnetic beads (Promega, Madison, Wisc.). Enriched DNA was recovered by polymerase chain reaction (PCR) and cloned into plasmid vectors. The purified amplification mixture was digested with *Eco*RI and ligated into *Eco*RI-digested pGEM3Z+ (Promega). Ligated DNAs were transformed into *E. coli* (strain SURE[TM], Stratagene, La Jolla, California) and screened and positive clones sequenced.

Of 52 clones sequenced, 43 were positive for the presence of a tetranucleotide repeat and 9 for a dinucleotide repeat. From these, 21 primer pairs were designed and tested. One primer from each pair was end-labeled with fluorescent dye (6-FAM, Operon). PCR included 5 ng DNA, 25 ng forward primer, 25 ng reverse primer, 200 µM of each dNTP (Promega), 0.5 U Taq DNA polymerase (GibcoBRL), 20 mM Tris–HCl pH 8.4, 50 mM KCl and 1.5 mM MgCl2. Amplification was performed in a Techne Genius thermocycler using an initial denaturation at 94 C for 2 min, followed by 35 cycles consisting of 40 s at 94 C, 1 min of annealing at 55 C, and 2 min extension at 72 C. A final extension was performed for 20 min at 72 C. Amplified DNA was analyzed according to PE Applied Biosystems protocols on an ABI Prism 3100 DNA Sequencing System. Band scoring was performed with Genemapper software (PE Applied Biosystems, Foster City, CA, USA). The 21 microsatellite loci were pre-screened in a test panel of eight redwood samples. Of those, six were identified as sufficiently polymorphic and then used in the present study of rangewide genetic variation (Table 1).

DNA analysis.—Analysis of co-dominant markers, such as microsatellites, in polyploids poses major difficulties because genotypes can only be assigned in the homozygous and fully heterozygous states. Two main approaches have been used to address these difficulties. One approach scores alleles as presence-absence data as if the alleles were dominant markers (Rodzen *et al.*, 2004). The second approach assigns genotypes by assessing dosage of each allele, to detect the presence of more than one copy in a sample, following the methods described by Esselink *et al.* (2004) in their MAC-PR method. Neither approach is ideal; the former fails to exploit the full information content in co-dominant marker data and the latter entails risks of errors in assignment, when two or more copies of an allele are present.

TABLE 1.—Characteristics of microsatellite loci in *Sequoia sempervirens*

Locus	Repeat	GenBank accession no.	Primer sequences (5'-3')	T_a. (°C)	Size range (bp)	No. of alleles	H_E	Numbers of individuals with 1/2/3/4/5/6 alleles detected
RW30	(CTAT)$_{11}$	GU969043	F: GTTGCGTCAGACATTCAACG R: CGAGAGCTAGCCACTCAAGG	55	97–153	14	0.65	7/51/56/18/3/0
RW44	(CTAT)$_{21}$	GU969044	F: CCATCAAGCCTAATCCCTCA R: GGTAGGAGAGAGGGAGGGAGA	55	328–448	28	0.93	5/23/45/47/14/1
RW48	(CTAT)$_{12}$	GU969045	F: GGCCTACATAAGTGAACTAGAA R: TGAGGAGAACACAACCCTCA	55	182–246	16	0.87	14/38/50/31/1/1
RW39	(CTAT)$_{20}$	GU969046	F: CCATAAGGTTGAAATGAAGAAAAA R: GTTGATTGATCGTTGGTTGG	55	129–207	21	0.57	43/44/33/12/3/0
RW28	(CTAT)$_{11}$	GU969047	F: GATAGATAAATAGATGGATAG R: TTTTTAAGGTTTCATGGATAAGTACAA	55	186–240	25	0.93	5/30/29/47/43/10
RW33	(CTAT)$_{13}$	GU969048	F: GTCCCCCTTAACAAGCATCA R: CGCGGGTGAATAGATCAGAG	55	180–412	42	0.91	74/37/10/7/6/0

TABLE 2.—Watershed values of genetic diversity using both presence absence and allele frequency methods

Watershed	n	Mean latitude	Alleles no.	Presence-absence data		Allele frequency data		
				H_J	H_S	Allelic richness	H_E	Private allelic richness
A	9	42.08	56	0.13 (0.01)	0.12 (0.007)	7.31	0.69	0.27
B	11	41.76	68	0.15 (0.02)	0.12 (0.007)	8.62	0.80	0.62
C	9	41.40	70	0.16 (0.02)	0.13 (0.007)	9.15	0.84	0.05
D	4	41.28	44	0.14 (0.20)	0.13 (0.006)	7.33	0.77	0.34
E	5	41.22	53	0.17 (0.20)	0.13 (0.006)	8.31	0.83	0.21
F	9	40.98	63	0.15 (0.02)	0.13 (0.006)	8.22	0.78	0.04
G	21	40.40	80	0.13 (0.01)	0.13 (0.008)	8.73	0.80	0.41
H	4	39.84	55	0.17 (0.02)	0.13 (0.006)	9.17	0.82	0.56
I	7	39.37	61	0.16 (0.02)	0.13 (0.007)	8.60	0.84	0.31
J	8	39.43	64	0.15 (0.02)	0.13 (0.007)	8.51	0.79	0.37
K	13	38.97	72	0.14 (0.02)	0.12 (0.007)	8.43	0.77	0.28
L	6	38.58	58	0.15 (0.02)	0.12 (0.006)	8.48	0.78	0.72
M	7	38.49	54	0.13 (0.02)	0.12 (0.006)	7.72	0.77	0.43
N	4	37.96	41	0.13 (0.02)	0.12 (0.007)	6.83	0.71	0.20
O	8	37.05	64	0.13 (0.02)	0.12 (0.007)	8.33	0.75	0.93
P	4	36.35	45	0.18 (0.02)	0.13 (0.007)	9.00	0.78	0.33
Q	6	36.00	56	0.15 (0.02)	0.13 (0.007)	8.37	0.82	0.53

Here, we analyzed our co-dominant microsatellite data for the hexaploid redwood using these two parallel methods.

Presence-absence format.—To test for global genetic structure and potential inbreeding, we analyzed the data using a Bayesian method performed by Hickory v1.1 (Holsinger *et al.*, 2002; Holsinger and Wallace, 2004) that makes no prior assumptions of Hardy-Weinberg equilibrium. Hickory is model-based, and allows the user to test how well data fit to a full model that estimates f (the inbreeding coefficient) and θ (an analogue of F_{ST}), to a model in which f = 0, or a model in which θ = 0. For dominant markers, the authors caution on the interpretation of estimates of f, and recommend running a fourth model (f-free) that does not estimate f, but selects values of f from its prior distribution while estimating θ. Deviance information criterion (DIC) values are output for each model, which estimate how well the model fits the data, taking into account the number of parameters being estimated (Spiegelhalter *et al.*, 2002). Default settings were used for burnin (50,000), sampling (250,000) and thinning (50).

Genetic structure and gene diversity among the watersheds were analyzed using the fragment frequency method in AFLP-SURV v1.0 (Vekemans *et al.*, 2002, available at http://www.ulb.ac.be/sciences/lagev). Amplified fragments (alleles) that occurred only once in the total data set were removed. The fragment frequency method assumes fixed homozygosity at each fragment, or haploid individuals. We chose this approach rather than the square root or Bayesian methods because estimations are based on the frequencies of the amplified and non-amplified (null) products, rather than the square root of the null allele, as in typical dominant marker analyses. Estimates of diversity (H_j) and structure (F_{ST}) among watersheds were performed according to Lynch and Milligan (1994). Pairwise F_{ST} estimates were bootstrapped over fragments 1000 times. Bootstrapped matrices of pairwise
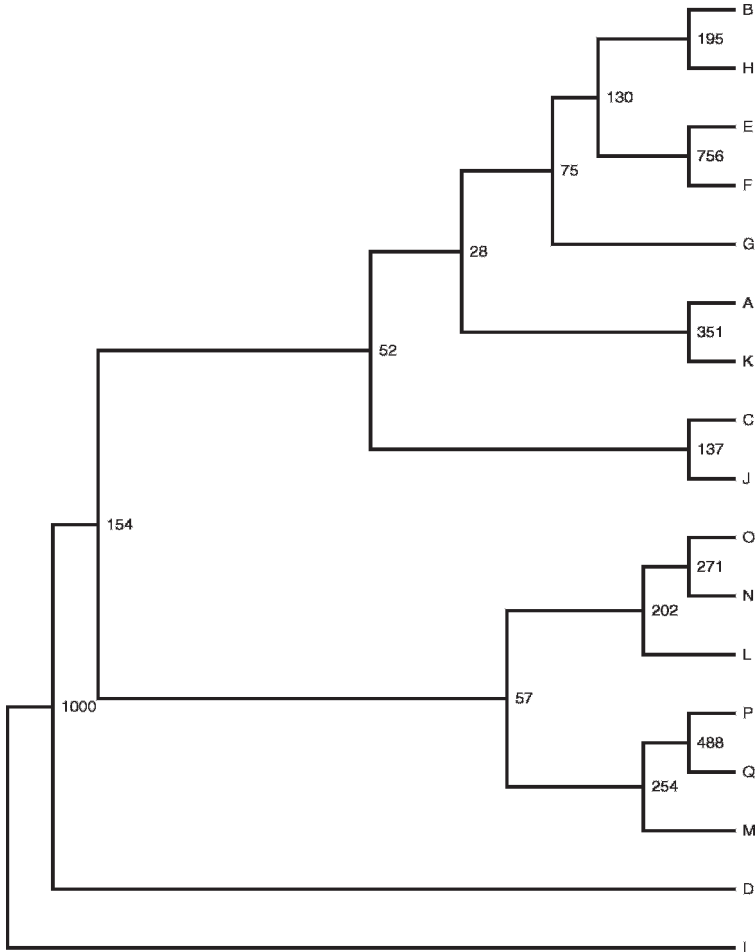
Fig. 2.—Unrooted concensus tree of pairwise F_{ST} among watersheds. Microsatellite alleles were treated as presence-absence data. Pairwise F_{ST} were estimated using AFLPSURV (Vekemans *et al.*, 2002). Neighbor joining trees were computed in PHYLLIP (Felsenstein, 1993). Bootstrap values (1000) shown at nodes

F_{ST} estimates were clustered using the neighbor-joining method in Phylip 3.65 (Felsenstein, 1993) and an unrooted consensus tree was found using the extended majority rule in Consense (Phylip, 3.65). The consensus tree was drawn with TreeView 1.6.6 (Page, 1996).

We also carried out a principal coordinates analysis on a matrix of Dice similarity coefficients performed in NTSYS 2.1 (Applied Biostatistics). The use of Dice, that is equivalent to the coefficient of Nei and Li, is more suitable for codominant markers when they are scored dominantly because it does not treat shared absences as similarity. An analysis of variance was computed to test the effect of northern and southern groupings, with a divide including southern populations (L–Q) and northern populations (A–K) following the concensus tree in Figure 2. To test the significance of this grouping we permuted watersheds into two random groups (with a minimum of five watersheds in a

group, to minimize effects of outlier populations) 100 times and compared the *F* value for our north/south comparison with the distribution of permuted values. Analyses of variance were carried out with the ANOVA procedure in SAS (SAS, Inc.).

Allele assignment method.—The microsatellite DNA allele counting – peak ratios method (MAC-PR) described by Esselink *et al.* (2004) was used to assign precise allelic configurations (the actual genotype) based on quantitative values for peak areas from electropherograms. For each locus, all alleles were analyzed in pairwise combinations in order to determine their copy number in the individual samples. This was accomplished by calculating ratios between the peak areas for two alleles in all samples in which these two alleles occurred together.

Genetic diversities within watersheds were estimated using the two-phase mutation model for microsatellites scaled by ploidy level (Bruvo *et al.*, 2004), in GenoDive vers. 2.0b4 (Meirmans and van Tienderen, 2004). Allelic richness and private allelic richness were estimated using rarefaction to compensate for small sample size in ADZE (Szpiech *et al.*, 2008). Standardized sample sizes were set from 2 to 21 (the largest sample size in our data).

Analysis of molecular variance was performed using SPAGeDi ver. 1.2 (Hardy and Vekemans, 2002). We estimated global and pairwise F_{ST} based on allele identity (infinite alleles model; IAM) and R_{ST} based on allele size under a stepwise mutation model (stepwise mutation model; SMM).

We also performed a Bayesian analysis to evaluate genetic structure among individuals with Structure ver. 2.2 (Pritchard *et al.*, 2000; Falush *et al.*, 2003). This software infers groups of genotypes (K) that as far as possible satisfy Hardy Weinberg equilibrium and linkage equilibrium within groups. The admixture model with correlated frequencies was used. Preliminary analyses had indicated weak genetic structure, we allowed prior population information. We inferred population division (number of K populations) by performing 20 independent runs of each K ($K = 1$ to $K = 17$) with a burn-in of 100,000 iterations, and 1,000,000 iterations of the Gibbs sampler. Log-likelihood of the data was recorded for each run and the ad hoc statistic ΔK was calculated following Evanno *et al.* (2005).

To investigate whether the underlying genetic structure of redwood conformed to a continuous distribution model, rather than discrete populations we tested for isolation by distance using Mantel tests. We first compared a matrix of individual pairwise Dice coefficients with pairwise Euclidean spatial distances computed from latitude and longitude coordinates. We then performed matrix comparisons of pairwise watershed F_{ST} and R_{ST} (estimated in SPAGeDi) with Euclidean spatial distances obtained from the mean latitude and longitude coordinates for each watershed. Euclidean spatial distances and matrix comparisons with 10,000 permutations were performed using NTSYS-pc v2.0.

<div align="center">RESULTS</div>

The six microsatellite loci yielded a total of 146 alleles, ranging from 14–42 per locus. Numbers of alleles per sample ranged from 1–6, with high frequencies of 2 to 4 alleles per sample for 3 loci (RW30, RW44, RW48), 2–5 alleles for 1 locus (RW28), 1–3 alleles for one locus (RW39) and 1–2 alleles for one locus (RW33) (Table 1). Thus, allelic configurations were consistent with a hexaploid genome.

Genetic diversity.—Genetic diversity estimated from presence-absence data was in the range 0.12–0.17, in contrast to the higher estimates based on allelic data (0.69–0.84) (Table 2). This is to be expected, as estimated heterozygosity is a function of numbers of alleles that are maximally two for the presence-absence data. Despite the magnitude of differences, trends among populations can be compared using the different methods. All measures of

diversity were significantly correlated with one another after Bonferroni corrections, except private allele richness. Although populations varied in genetic diversity, these variations were not correlated with latitude of origin and had no obvious spatial pattern. For example highest allelic richness was detected in the northern watershed C, the central watershed H and the southern watershed P (Table 2).

Genetic structure.—The DIC criterion for model choice favoured a no-inbreeding model over all other models tested. Smaller DIC values are preferred in model choice and differences of less than 5–6 units are not considered justification for selecting one model over another (Spiegelhalter *et al.*, 2002). The DIC value for the full model was 4310, compared with 4613 for the model of no population structure ($\theta = 0$), providing strong support for population structure in the presence/absence data. The DIC value for the no-inbreeding model ($f = 0$) was 4323, so that we could not rule out that some inbreeding occurred in these populations. According to the f – free model, $\theta^{II} = 0.029$ (0.022–0.038, 95% credible region), which is a measure of contemporary population structure equivalent to Nei's G_{ST}.

Global population structure from AFLPSurv was 0.041 (0.014–0.043, 95% confidence limits), similar to the Bayesian estimate above. A neighbor-joining tree of pairwise F_{ST} among watersheds is shown in Figure 2. Two major clades are suggested, although bootstrap values are low. One clade is a southern group extending from the southern most populations in Monterey County to Sonoma and Napa counties. The second clade includes most of the populations north of the Sonoma – Mendocino county border ($\sim$38.6°N).

We examined this north-south divergence further using principal coordinates analysis. The high dimensionality of the data was only moderately reduced, with the first three axes accounting for 5.0%, 4.4%, 4.1% of the data respectively. Two-dimensional plots of the first three principal coordinates dimensions indicated a weak separation of northern and southern populations (Fig. 3). Analysis of variance for this divergence was significant for Pcoord1 but not for the other two dimensions (Table 3). A permutation test showed that the F-statistic for the north-south break was in the upper 2% of the distribution of F-statistics from random permutations of populations into two groups.

Global F_{ST} and R_{ST} estimated from allele frequency data in SPAGeDi were not significantly different from zero ($F_{ST} = 0.0031$, 95% confidence limits -0.006–0.006, $R_{ST} = 0.024$, 95% confidence limits -0.04–0.04).

The Bayesian analysis of population structure carried out in STRUCTURE yielded Ln P(D) probabilities that increased from K = 1 (-1130) to K = 10 (-9986), after which probabilities reached a plateau and fluctuated. The optimal value of K, using the procedure recommended by Evanno *et al.* (2005), was K = 3 groups. We therefore evaluated groupings of 2, 3 and 4 for optimal biological relevance. In Table 4, we record populations that had more than 50% of individuals assigned to a STRUCTURE group, assuming K = 2, more than 40% assuming K = 3 and more than 30% assuming K = 4. Southern populations from L–Q were assigned to the same cluster for K = 2 and K = 3 but were broken up for K = 4. For K = 2 and 3, population H was included with the southern cluster. There was no obvious geographic pattern for the segregation between clusters 2 and 3, for the K = 3 model. The K = 4 model made less sense in a spatial context than either the K = 2, or K = 3 models.

Mantel tests revealed a significant association between geographic distance and genetic distances for individual pairwise matrices using the complement of Dice's similarity coefficients as a measure of genetic distance (Mantel r = 0.09; P = 0.005). At a watershed level, genetic distances were significantly associated with geographical distance for F_{ST}
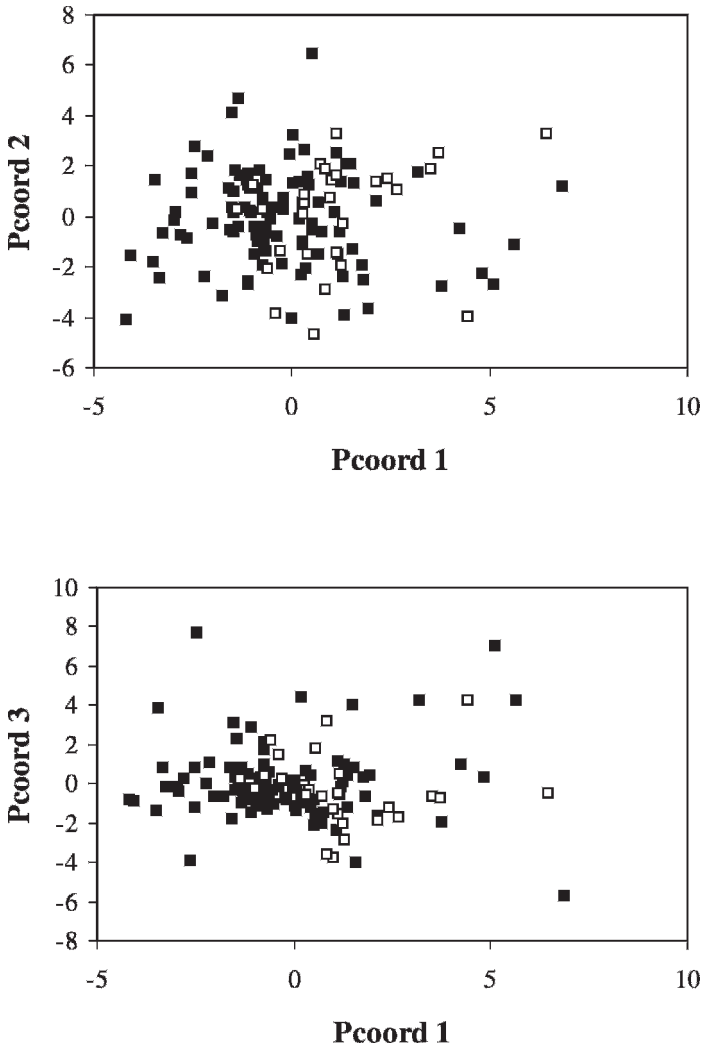
FIG. 3.—Plot of the first three principal coordinates axes of microsatellite alleles treated as presence-absence data. Filled rectangles – individuals from north of Sonoma County, California ~36.8°N; open rectangles – individuals from Sonoma county south

TABLE 3.—Analysis of variance of effects of a north – south divergence at the Sonoma-Mendocino border on individual tree scores for the first three principal coordinates axes of microsatellite alleles scored as presence-absence data

Principal coordinates axes	MSQ	F-ratio	P
Pcoord 1	42.46	12.57	0.0005
Pcoord 2	4.75	1.37	0.24
Pcoord 3	1.75	0.44	0.51

TABLE 4.—Population membership in STRUCTURE clusters of K = 2, 3 and 4, where populations are assigned to a cluster if individual assignments to the cluster are: >50% for K = 2, >40% for K = 3 and >30% for K = 4

Number of groups K	Inferred cluster			
	1	2	3	4
2	C, H, L, M, N, O, P, Q	A, B, D, E, F, G, I, J, K		
3	H, L, M, N, O, P, Q	A, D, E, F, I, J	B, C, G, K	
4	D, J, N, Q	A, B, E, F, G, L	C, H, M, P	I, K, O

following an IAM (Mantel r = 0.37; P = 0.003) but not with R_{ST} following a SMM (Mantel r = −0.05; P = 0.29).

DISCUSSION

To predict how species may respond to rapid habitat displacements as a result of climate change has become an important area of inquiry. Low rates of vagility may leave some species stranded in less than optimal environments, resulting in population extinctions or the need for human-assisted movements. Past evolutionary events may have led to divergent lineages within species that may have special conservation importance and differential capacity to respond to environmental change (Moritz, 1994). This is particularly important for keystone taxa of species-rich, communities, where future community structure could depend strongly on the success of the keystone species. Here, we report for the first time population genetic structure of coast redwood that is the dominant species of the coastal redwood community in one of the world's hotspots of diversity in the California Floristic Province.

Microsatellite markers.—The six tetranucleotide repeat microsatellite loci that we developed and report here are highly variable. Analysis of 135 individuals detected 146 alleles, ranging from 14–42 alleles per locus. These loci are much more variable than the five microsatellite loci developed previously for redwood (Bruno and Brinegar, 2004), for which only two loci showed moderate variability. Highly polymorphic loci that are relatively easy to score are desirable for addressing ecological questions where high exclusion probabilities are needed such as parentage analysis, for clonal identification and for quantifying gene flow among populations. We observed up to six alleles per individual for three loci and up to five alleles per individual for the remaining three loci. This pattern would be expected in an autopolyploid and in an autoallopolyploid or segmental polyploid if parental genomes were at least partially homologous. Therefore, we cannot rule out that the hexaploid genome of coast redwood is of autopolyploid origin. In contrast to our markers, Brinegar *et al.* (2007) observed no more than four alleles per individual for their five nuclear microsatellite loci and suggested that this could be consistent with an autoallopolyploid, where the locus was not amplifying for one of the parental chromosome sets.

Interpretation of co-dominant genetic markers in a population genetic framework is not straightforward for polyploid species and only a few analytical approaches have addressed polyploidy directly. Here, we used both a presence/absence approach to interpretation of alleles and inferred genotypes according to allele dosage. Results from the two approaches were mostly concordant, and, although the allele dosage approach exploits all of the available data, the phenetic approach has proven valuable for detection of genetic diversity and genetic structure in wild plants (Hollingsworth and Ennos, 2004).

Population genetic structure.—Estimates of global population differentiation ranged from 3–4% for alleles treated phenetically, to less than 1% for F_{ST} and ~2.5% for R_{ST} estimated from allele frequency data. Although estimates from allele frequency data were not significantly different from zero, taken together the data indicate a consistent pattern of low genetic structure among populations. This is not surprising for a narrowly distributed endemic tree species with wind dispersed pollen (Hamrick and Godt, 1996) and is probably reinforced by the hexaploid genome in redwood. Effects of genetic drift will be slower for a polyploid with greater effective population size than for a diploid with the same census population size.

The consensus tree of pairwise phenetic F_{ST} indicated a partition of southern and northern populations with a break close to the Sonoma/Mendocino county border. This tree received only weak support, but was confirmed by analyses of variance of principal coordinates scores. Furthermore, Bayesian assignment of individuals suggested two or three groups as being the most likely structure in the microsatellite data treated as allele frequencies, with southern populations again tending to diverge from the north at the same location. Although our finding of a north-south divergence among populations of redwood is in line with earlier reports based on growth performance from a common gardens study (Anekonda, 1992) and from foliar monoterpene profiles (Hall and Langenheim, 1987), the break is further north than the San Francisco Bay that was indicated by these earlier studies. It may not be surprising that divergence indicated by growth performance and monoterpene composition is not fully concordant with our microsatellite data. Local adaptation for growth performance and herbivore selection pressures for monoterpene composition (Hall and Langenheim, 1987; Kulman, 1971; Sturgeon, 1979; Langenheim *et al.*, 1986) may buffer effects of genetic drift due to population isolation. The unrooted concensus tree showed population D as being divergent from either the northern or southern lineages. The sample size was small and alleleic richness very low and we suspect that this was a sampling artifact rather than a true difference.

Genetic distance increased with geographic distance at both an individual level (measured as Dice's pairwise distances) and at a population level (F_{ST}) suggesting that a clinal population structure probably best fits the pattern of genetic diversity in coast redwood. Interestingly, there was no significant association between geographic distance and pairwise population R_{ST}, so that when mutational information content of microsatellites is taken into account, an isolation by distance model does not apply. The most likely explanation for this lack of concordance between F_{ST} and R_{ST} is that the entire population of redwood is not in mutation/migration/drift equilibrium because of divergence between northern and southern groups of populations described above. The present distribution of coast redwood is more or less continuous from southern Oregon to northern Sonoma County and then becomes fragmented to its southern limit (Roy, 1966). This coincides with the two lineages that we have detected here and may reflect a long-term break in the distribution of coast redwood. The palynological record reports low levels of redwood pollen from the middle of the last glacial through the last glacial maximum from southern Oregon to central California (Heusser, 1998). Although populations were possibly small and isolated, the redwood pollen record suggests that the limits of its distribution have changed little since ~60 ka (1000 y before present). The rather constant latitudinal limits of coast redwood through time are likely due to limited southward movement of the polar front during the last glacial maximum (Heusser, 1998) and enhanced upwelling resulting in coastal fog (Pisias *et al.*, 2001). Increases in redwood pollen during the late glacial ~12 ka followed by a more rapid increase at ~5.2 ka are more or less coincident in sites close to the

California-Oregon border and in Sonoma County, northern California (Heusser, 1998). The more recent increase was probably due to a transition from a continental climate to a maritime climate with cool summers and mild winters (Barron *et al.*, 2003).

Future for redwood under climate change.—Most predictions suggest that over the next century, north temperate taxa will need to shift to higher latitudes or altitudes to match their current climatic habitats (Thuiller *et al.*, 2005; Iverson and Prasad, 2002; Kueppers *et al.*, 2005). Coastal California is expected to be a sink for diversity from more interior regions but will also lose taxa as temperatures rise (Loarie *et al.*, 2008). The narrow distribution of redwood suggests that it is unlikely to succeed well outside of its current climatic envelope. Despite high levels of genetic diversity that we have observed for microsatellite loci and that have been reported earlier for growth performance (Anekonda, 1992) and monoterpenes (Hall and Langenheim, 1987), the species appears not to have adapted to environmental gradients at the limits of its range. This is probably in part due to the predominant asexual reproductive mode; although large quantities of seed can be produced, their viability is very low (Olsen *et al.*, 1990). Furthermore, seed viability is of short duration (Fritz and Rydelius, 1966) and seedlings are extremely susceptible to damping off (Hepting, 1971). Stands are composed of a high proportion of clonal members (Douhovnikoff *et al.*, 2004; Rogers, 2000) that may have persisted over many generations. This persistence mode of reproduction can be advantageous in the face of disturbance events and also may have been of great benefit to redwoods during the cold periods of the Pleistocene. However, two major disadvantages to this mode of growth are (1) lack of recombination that would otherwise allow new adaptive gene combinations to be expressed, and (2) low vagility of the species. These two consequences of vegetative reproduction in long-lived organisms such as redwood are likely to mean that it will be unable to adapt to the new conditions or migrate into the displaced habitat. Assisted migration may be necessary to allow populations to match the new climates.

The southern range of redwood is of critical conservation concern. These populations that appear from our work here to be divergent from the northern range of the species are likely to be most adversely affected by climate change. Further work is needed to elucidate population structure in the south and to confirm the break between the northern and southern lineages. More intensive sampling is necessary that will likely give greater power to discriminatory tests. We also have a panel of additional microsatellite loci that could be used to provide wider genome sampling. Our results are an important step forward in our understanding of redwood genetics, probably the greatest contribution of these data is as a guide for future more intensive sampling. This would allow for analysis at a finer scale, the inclusion of more independent variables and the exploration of how biological properties such a hexaploidy and clonal growth influence genetic variation.

LITERATURE CITED

AHUJA, M. R. 2009. Genetic constitution and diversity in four narrow endemic redwoods from the family Cupressaceae. *Euphytica*, **165**:5–19.
——— AND D. B. NEALE. 2002. Origins of polyploidy in coast redwood (*Sequoia sempervirens* (D. Don) Endl.) and relationship of coast redwood to other genera of Taxodiaceae. *Silvae Genet.*, **51**:93–100.
ANEKONDA, T. S., R. S. CRIDDLE AND W. J. LIBBY. 1993. Spatial and temporal relationships between growth traits and metabolic heat rates in coast redwood. *Can J Forest Res.*, **23**:1793–1798.

———. 1992. A genetic architecture study of coast redwood. Dissertation Thesis. University of California, Berkeley.

BARRON, J. A., L. HEUSSER, T. HERBERT AND M. LYLE. 2003. High-resolution climatic evolution of coastal northern California during the past 16,000 years. *Paleoceanography*, **18**:1020. doi:10.1029/2002PA000768, 2003.

BRINEGAR, C., D. BRUNO, R. KIRKBRIDE, S. GLAVAS AND I. UDRANSZKY. 2007. Applications of Redwood Genotyping by Using Microsatellite Markers. *In:* R. B. Standiford, G. A. Giusti, Y. Valachovic, W. J. Zielinski and M. J., technical. Furniss (eds.). 2007. Proceedings of the redwood region forest science symposium: What does the future hold? Gen. Tech. Rep. PSW-GTR-194. Albany, California: Pacific Southwest Research Station, Forest Service, U.S. Department of Agriculture, p. 47–56.

BRUNO, D. AND C. BRINEGAR. 2004. Microsatellite markers in coast redwood (*Sequoia sempervirens*). *Mol. Ecol. Notes*, **4**:482–484.

BRUVO, R., N. K. MICHIELS, T. G. D'SOUZA AND H. SCHULENBURG. 2004. A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Mol. Ecol.*, **13**:2101–2106.

CALSBEEK, R., J. N. THOMPSON AND J. E. RICHARDSON. 2003. Patterns of molecular evolution and diversification in a biodiversity hotspot: the California Floristic Province. *Mol. Ecol.*, **12**:1021–1029.

CHATZIMANOLIS, S. AND M. S. CATERINO. 2007. Limited phylogeographic structure in the flightless ground beetle, *Calathus ruficollis*, in southern California. *Divers. Distrib.*, **13**:498–509.

CULLINGS, K. W. 1992. Design and testing of a plant-specific PCR primer for ecological and evolutionary studies. *Mol. Ecol.*, **1**:233–240.

DOUHOVNIKOFF, V., A. CHENG AND R. S. DODD. 2004. Incidence, size and spatial structure of clones in second growth stands of coast redwood, *Sequoia sempervirens* (Cupressaceae). *Am J Bot.*, **91**:1140–1146.

ESSELINK, G. D., H. NYBOM AND B. VOSMAN. 2004. Assignment of allelic configuration in polyploids using the MAC-PR (microsatellite DNA allele counting – peak ratios) method. *Theor Appl Genet.*, **109**:1–7.

EVANNO, G., S. REGNAUT AND J. GOUDET. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.*, **14**:2611–2620.

FALUSH, D., M. STEPHENS AND J. K. PRITCHARD. 2003. Inference of population structure: Extensions to linked loci and correlated allele frequencies. *Genetics*, **164**:1567–1587.

FELSENSTEIN, J. 1993. PHYLIP (Phylogeny Inference Package) version 3.65. Distributed by the author. Department of Genetics, University of Washington, Seattle, Washington.

FOZDAR, B. S. AND W. J. LIBBY. 1968. Chromosomes of *Sequoia sempervirens*; 8-hydroxy-quinoline-castor oil pretreatment for improving preparation. *Stain Technol.*, **43**:97–100.

FRITZ, E. AND J. A. RYDELIUS. 1966. Redwood reforestation problems: an experimental approach to their solution. Foundation for American Resource Management, San Francisco, California, U.S.A.

HAMRICK, J. L. AND M. J. W. GODT. 1996. Effects of life history traits on genetic diversity in plant species. *Philos T Roy Soc B*, **351**:1291–1298.

HALL, G. D. AND J. H. LANGENHEIM. 1987. Geographic variation in leaf monoterpenes of *Sequoia sempervirens*. *Biochem. Syst. Ecol.*, **15**:31–43.

HARDY, O. J. AND X. VEKEMANS. 2002. SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes*, **2**:618–620.

HEPTING, G. H. 1971. Diseases of forest and shade trees of the United States. *U.S. Department of Agriculture, Agricultural Handbook*, **386**:658.

HEUSSER, L. 1998. Direct correlation of millennial-scale changes in western North American vegetation and climate with changes in the California Current system over the past ~60 kyr. *Paleoceanography*, **13**:252–262.

HOLLINGSWORTH, P. M. AND R. A. ENNOS. 2004. Neighbour joining trees, dominant markers and population genetic structure. *Heredity*, **92**:490–498.

HOLSINGER, K. E. AND L. E. WALLACE. 2004. Bayesian approaches for the analysis of population genetic structure: an example from *Platanthera leucophaea* (Orchidaceae). *Mol. Ecol.*, **13**:887–894.

———, P. O. LEWIS AND D. K. DEY. 2002. A Bayesian method for analysis of genetic population structure with dominant marker data. *Mol. Ecol.*, **11**:1157–1164.

IVERSON, L. R. AND A. M. PRASAD. 1998. Predicting the abundance of 80 trees species following climate change in the eastern United States. *Ecol Monogr.*, **66**:465–485.

KHASA, P. D., C. H. NEWTON, M. H. RAHMAN, B. JAQUISH AND B. P. DANCIK. 2000. Isolation, characterization, and inheritance of microsatellite loci in alpine larch and western larch. *Genome*, **43**:439–448.

KUEPPERS, L. M., M. A. SNYDER, L. C. SLOAN, E. S. ZAVALETA AND B. FULFROST. 2005. Modeled regional climate change and California endemic oak ranges. *Proc. Natl. Acad. Sci. U.S.A.*, **102**: 16281–16286.

KULMAN, H. M. 1971. Effects of insect defoliation on growth and mortality of trees. *Ann. Rev. Entomol.*, **16**:289–324.

KUSER, J. E., A. BAILLY AND A. FRANCLET. 1995. Early results of a rangewide provenance test of *Sequoia sempervirens*. Recursos Geneticos Forestales (FAO), p. 21–25.

LANGENHEIM, J. H., C. A. MACEDO, M. K. ROSS AND H. STUBBLEBINE. 1986. Leaf development in the tropical leguminous tree *Copaifera* in relation to microlepidopteran herbivory. *Biochem. System. Ecol.*, **14**:51–60.

LOARIE, S. R., B. E. CARTER, K. HAYHOE, S. MCMAHON, R. MOE, C. A. KNIGHT AND D. D. ACKERLY. 2008. Climate change and the future of California's endemic flora. *Plos. One*, **3**:e2502 1–10.

LYNCH, M. AND B. G. MILLIGAN. 1994. Analysis of population genetic structure with RAPD markers. *Mol. Ecol.*, **3**:91–99.

MEIRMANS, P. AND P. VAN TIENDEREN. 2004. GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Mol. Ecol. Notes*, **4**:792–794.

MORITZ, C. 1994. Defining ''Evolutionarily Significant Units'' for conservation. *Trends in Ecol. Evol.*, **9**:373–3.

NETTEL, A., R. S. DODD AND Z. AFZAL-RAFII. 2009. Genetic diversity, structure, and demographic change in tanoak, *Lithocarpus densiflorus* (Fagaceae), the most susceptible host to the Sudden Oak Death disease in California. *Am. J. Bot.*, **96**:2224–2233.

OLSEN, D. F., D. F. ROY AND G. A. WALTERS. 1990. *Sequoia sempervirens* (D. Don) Endl. Redwoods. *In:* Silvics of North America. U.S. Dep Agric. Agric. Handb. 654, p. 541–551.

PAGE, R. D. M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.*, **12**:357–358.

PISIAS, N. G., A. C. MIX AND L. HEUSSER. 2001. Millenial scale climate variability of the northeast Pacific Ocean and northwest North America based on radiolaria and pollen. *Quaternary Sci. Rev.*, **20**:1561–1576.

PRITCHARD, J. K., M. STEPHENS AND P. J. DONNELLY. 2000. Inference of population structure using multilocus genotype data. *Genetics*, **155**:945–959.

RISSLER, L. J., R. J. HIJMANS, C. H. GRAHAM, C. MORITZ AND D. B. WAKE. 2006. Phylogeographic lineages and species comparisons in conservation analyses: a case study of California herpetofauna. *Am. Nat.*, **167**:655–666.

RODZEN, J. A., T. R. FAMULA AND B. MAY. 2004. Estimation of parentage and relatedness in the polyploid white sturgeon (Acipenser transmontanus) using a dominant marker approach for duplicated microsatellite loci. *Aquaculture*, **232**:165–182.

ROGERS, D. L. 1997. Inheritance of allozymes from seed tissues of the hexaploid gymnosperm, *Sequoia sempervirens* (D. Don) Endl. (Coast redwood). *Heredity*, **78**:166–175.

———. 2000. Genotypic diversity and clone size in old-growth populations of coast redwood (*Sequoia sempervirens*). *Can. J. Bot.*, **78**:1408–1419.

ROY, D. F. 1966. Silvical characteristics of redwood [*Sequoia sempervirens* (D. Don) Endl.]. Berkeley, California, Pacific Southwest Forest and Range Experiment Station PSW-28, 20 p.

SAMBROOK, J., E. F. FRITSCH AND T. MANIATIS. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

SAYLOR, L. C. AND H. A. SIMONS. 1970. Karyology of *Sequoia sempervirens*: karyotype and accessory chromosomes. *Cytologia*, **35**:294–303.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin and A. van der Linde. 2002. Bayesian measures of model complexity and fit. *J. Roy. Stat. Soc. B.*, **64**:483–689.

Stebbins, G. L. 1948. The chromosomes and relationships of *Metasequoia* and *Sequoia*. *Science*, **108**:95–98.

Sturgeon, K. B. 1979. Monoterpene variation in ponderosa pine xylem resin related to western white pine beetle predation. *Evolution*, **40**:290–302.

Szpiech, Z., M. Jakobsson and N. A. Rosenberg. 2008. ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*, **24**:2498–2504.

Thuiller, W., S. Lavorel, M. B. Araújo, M. T. Syles and I. C. Prentice. 2005. Climate change threats to plant diversity in Europe. *Proc. Natl. Acad. Sci. U.S.A.*, **102**:8245–8250.

Vekemans, X., T. Beauwens, M. Lemaire and I. Roldan-Ruiz. 2002. Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Mol. Ecol.*, **11**:139–151.