Estimation of population structure in coastal Douglas-fir [*Pseudotsuga menziesii* (Mirb.) Franco var. *menziesii*] using allozyme and microsatellite markers

Konstantin V. Krutovsky · John Bradley St. Clair ·
Robert Saich · Valerie D. Hipkins · David B. Neale

Abstract Characterizing population structure using neutral markers is an important first step in association genetic studies in order to avoid false associations between phenotypes and genotypes that may arise from non-selective demographic factors. Population structure was studied in a wide sample of ~1,300 coastal Douglas-fir [*Pseudotsuga menziesii* (Mirb.) Franco var. *menziesii*] trees from Washington and Oregon. This sample is being used for association mapping between cold hardiness and phenology phenotypes and single-nucleotide polymorphisms in adaptive-trait candidate genes. All trees were genotyped for 25 allozyme and six simple sequence repeat (SSR) markers using individual megagametophytes. Population structure analysis was done separately for allozyme and SSR markers, as well as for both datasets combined. The parameter of genetic differentiation (θ or F_{ST}) was standardized to take into account high within-population variation in the SSR loci and to allow comparison with allozyme loci. Genetic distance between populations was positively and significantly correlated with geographic distance, and weak but significant clinal variation was found for a few alleles. Although the STRUCTURE simulation analysis inferred the same number of populations as used in this study and as based on previous analysis of quantitative adaptive trait variation, clustering among populations was not significant. In general, results indicated weak differentiation among populations for both allozyme and SSR loci (θ_s=0.006–0.059). The lack of pronounced population subdivision in the studied area should facilitate association mapping in this experimental population, but we recommend taking the STRUCTURE analysis and population assignments for individual trees (Q-matrix) into account in association mapping.

Keywords Allozyme · Genetic markers · Microsatellites · Population structure · *Pseudotsuga menziesii* var. *menziesii* · SSR · Simple sequence repeats

K. V. Krutovsky
Institute of Forest Genetics, Pacific Southwest Research Station,
USDA Forest Service,
Davis, CA 95616, USA

K. V. Krutovsky
Department of Ecosystem Science and Management,
Texas A&M University,
College Station, TX 77843-2138, USA

J. B. St. Clair
Pacific Northwest Research Station, USDA Forest Service,
Corvallis, OR 97331-4401, USA

R. Saich · V. D. Hipkins
National Forest Genetics Laboratory,
Pacific Southwest Research Station, USDA Forest Service,
Placerville, CA 95667, USA

D. B. Neale (✉)
Department of Plant Sciences, University of California,
One Shields Avenue,
Davis, CA 95616, USA
e-mail: dbneale@ucdavis.edu

Introduction

Coastal Douglas-fir [*Pseudotsuga menziesii* (Mirb.) Franco var. *menziesii*] is the most economically important tree in the Pacific Northwest and is the major object of our association mapping study. Coastal Douglas-fir has a high

level of phenotypic variation that is strongly associated with geographic, topographic, and climatic features (Campbell and Sugano 1975; Campbell and Sorensen 1978; Steiner 1979; Li and Adams 1993; Aitken and Adams 1997; Anekonda et al. 2000; St. Clair et al. 2005; St. Clair 2006). Clinal patterns of variation in growth and phenological traits have been observed over latitudinal, longitudinal, and elevational transects despite the appreciable gene flow expected in this wind-pollinated species. Common garden studies have found that adaptive patterns of variation can occur among Douglas-fir populations within relatively small local regions (Campbell 1979; St. Clair et al. 2005). Populations can differ considerably for adaptive traits, in particular for bud phenology and growth. Variation in bud set timing and growth was strongly related to provenance elevation and cool season temperatures. Variation in bud burst and partitioning to stem diameter versus height was related to provenance latitude and summer drought. Clinal variation in response to microenvironmental heterogeneity has also been found in the central part of the Oregon Cascades (Campbell 1979).

The genes controlling adaptive traits in Douglas-fir are, however, still largely unknown. Quantitative trait loci (QTL) mapping studies have allowed us to begin dissecting these complex traits (Jermstad et al. 2001a, b, 2003; Wheeler et al. 2005; Carlson et al. 2007). Several genomic regions responsible for genetic control of growth rhythm and cold hardiness traits were found, but QTL mapping does not reveal which individual genes are responsible for these effects. Association mapping is a more powerful population genomic approach that can identify individual genes and alleles that are responsible for phenotypic differences in adaptive traits (Krutovsky and Neale 2005a; González-Martínez et al. 2006, 2007). Limited genetic resources and the large genome of Douglas-fir, however, prevent a full genome scan. Instead, we plan to carry out a candidate gene-based association mapping study using single-nucleotide polymorphisms (SNPs). To achieve this goal, we developed SNP markers in numerous candidate genes (Krutovsky and Neale 2005b; Pande et al. 2007). However, unchecked population structure can result in false positive, spurious associations and constrain the use of association studies (see, for instance, Aranzana et al. 2005 for *Arabidopsis* and Yu and Buckler 2006 for maize). Such false positives arise when testing random genetic markers with different frequencies in subpopulations for a trait with parallel phenotypic differences. To reduce this risk, estimates of population structure must be included in association analysis.

To address the issue of population structure in population-based samples, the two most common methods, genomic control (GC) and structured association (SA), are utilized in both human and plant studies (e.g., Devlin and Roeder 1999; Devlin et al. 2001; Yu and Buckler 2006). With GC, a set of random markers is used to estimate the degree of inflation of the test statistics generated by population structure, assuming such structure has a similar effect on all loci (Devlin and Roeder 1999). By contrast, SA analysis first uses a set of random markers to estimate individual population subdivision (Q-matrix) and then incorporates this estimate into further statistical analysis (Pritchard and Rosenberg 1999; Pritchard et al. 2000; Falush et al. 2003). Genetic association models that account for the different levels of relatedness found in natural populations have recently been developed (Pritchard et al. 2000; Falush et al. 2003; Thornsberry et al. 2001; Yu et al. 2006; Camus-Kulandaivelua et al. 2007). For instance, in several association studies in plants (Thornsberry et al. 2001; Liu et al. 2003; Wilson et al. 2004), population genetic structure has been assessed using the software STRUCTURE (Pritchard et al. 2000; Falush et al. 2003).

If, however, the distribution of functional alleles is highly correlated with population structure, statistical control for population structure can result in false negatives, particularly for small sample sizes. Two recent studies in maize illustrate the above scenario. In an attempt to validate the function of the *Dwarf8* (*D8*) locus, 71 elite European inbred lines were genotyped for *D8* polymorphism and phenotyped for flowering time (Andersen et al. 2005). Although significant association was detected without controlling for population structure, no association resulted when the population structure was controlled. By contrast, the association of *D8* polymorphism with flowering time has been validated in a large association mapping population of 375 maize inbred lines (Camus-Kulandaivelu et al. 2006). Association studies, therefore, are best carried out in independent populations with a large sample size (see Ball 2007 for review).

The objectives of this study were to explore population structure in coastal Douglas-fir populations using molecular markers and trees sampled from these populations and to justify further use of these samples as a large association mapping population with Q-matrix corrections if needed.

Materials and methods

Population sampling

Wind-pollinated seeds were collected from 1,283 parent trees of coastal Douglas-fir [*P. menziesii* (Mirb.) Franco var. *menziesii*] in naturally regenerated stands at 1,048 locations in western Oregon and Washington (Fig. 1). Most of the seeds were obtained from previous collections of the USDA Forest Service, USDI Bureau of Land Management, Oregon Department of Forestry, and Northwest Tree
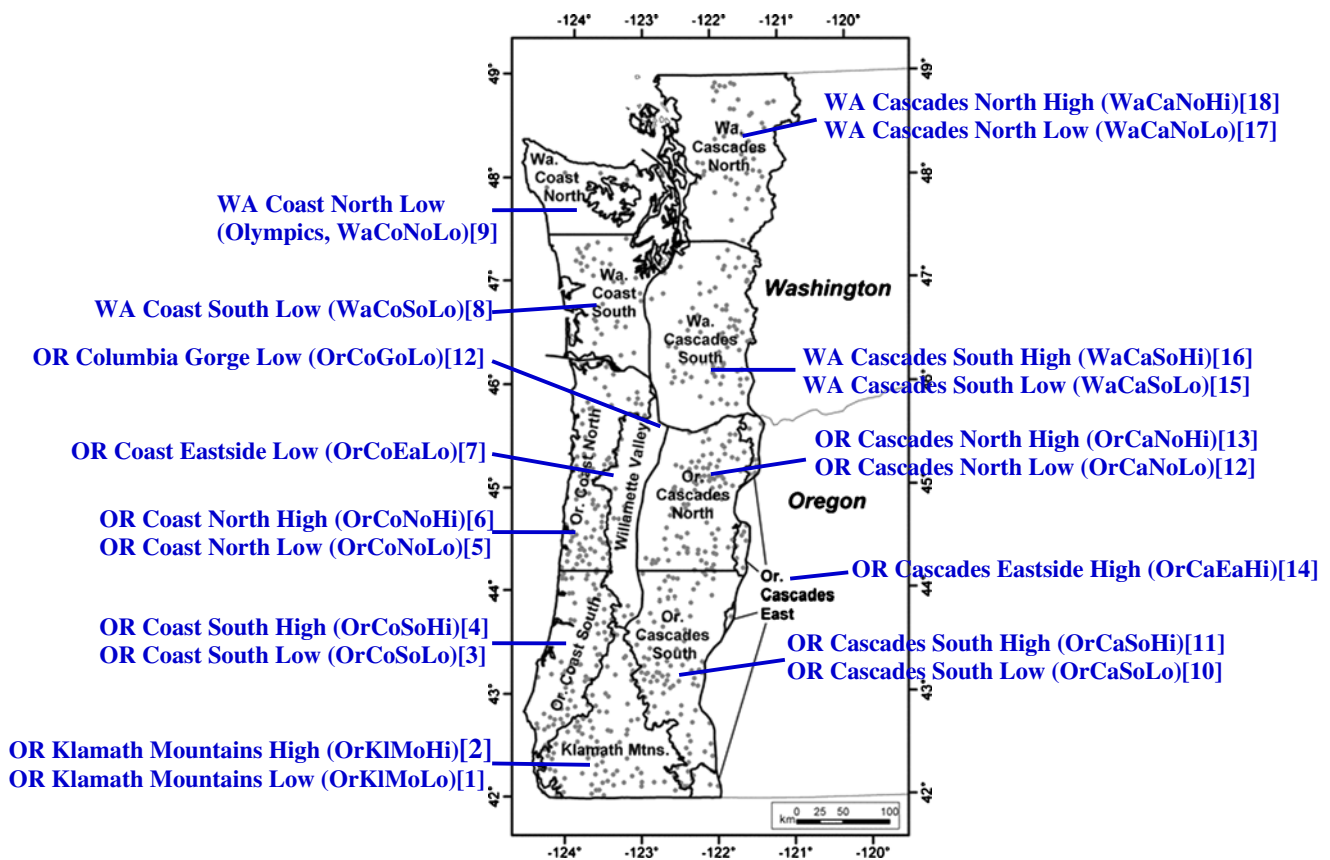
Fig. 1 Coastal Douglas-fir (*P. menziesii* var. *menziesii*) source locations and regional classification used in analysis. Most regions were divided into high and low elevations at 650 m (St. Clair 2006)

Improvement Cooperative made in the early 1990s. The range of coastal Douglas-fir in western Oregon and Washington was well sampled, although sampling intensity was lower along the Washington coast and in urban and agricultural areas around Puget Sound and the Willamette Valley (see St. Clair et al. 2005 for details).

Genetic markers

Allozyme loci Twenty-five allozyme loci (ACO-1, ACO-2, CAT, DIA, F-EST, G-6PD, GDH, GOT-1, GOT-2, GOT-3, GLYD, IDH, LAP-1, LAP-2, MDH-1, MDH-2, MDH-3, MDH-4, 6-PGD-1, PGM-1, PGM-2, PGI-1, PGI-2, PMI, and SOD) representing 16 different enzyme systems (Electronic supplementary materials, Table 1S) were genotyped in 1,236 trees (298 from Washington and 938 from Oregon). The genotypes of the trees were determined using eight megagametophytes per tree. The allozyme analysis was performed using starch gel electrophoresis and procedures described in Adams et al. (1990). The inheritance and linkage of these loci were studied earlier (El-Kassaby et al. 1982; Neale et al. 1984; Adams et al. 1990; Jermstad et al. 1994).

Microsatellite or SSR loci Six unlinked and highly polymorphic dinucleotide simple sequence repeat (SSR) loci (PmOSU_1C3, PmOSU_2G12, PmOSU_3B2, PmOSU_3F1, PmOSU_3G9, and PmOSU_4A7) were selected from a previous study (Slavov et al. 2004). For simplicity, the generic prefix "PmOSU" is omitted further in the text and tables. The 1,210 trees were genotyped using DNA isolated from ten pooled megagametophytes per tree (287 from Washington and 923 from Oregon). This allowed us to use the same seed collections that have been used for allozyme genotyping also for both SSR and SNP genotyping. Different seeds from the same open-pollinated seed collections were used for allozymes and for SSRs and SNPs genotyping, but the same DNA samples were used for SSRs and SNPs genotyping. A preliminary study pooling DNA from megagametophytes with known genotypes to make pools with ratios of different alleles 9:1, 8:2, 7:3, 3:7, 2:8, and 1:9 showed that the pooling had no effect on the SNP genotyping (results are available upon request). Pooling effects on SSR genotyping were not tested. Theoretically, smaller size alleles may have an amplification advantage over larger size alleles in uneven pools from heterozygous trees when ratio of both alleles is biased in

favor of a smaller size allele. It may potentially lead to the so-called large-allele dropout due to misgenotyping of heterozygotes for uneven size alleles as homozygotes for small alleles (DeWoody et al. 2006). In extreme cases, it may decrease observed heterozygosity and overestimate homozygosity. We addressed this potential problem by optimizing polymerase chain reaction (PCR) conditions and increasing PCR extension time to minimize competition between alleles, re-amplifying, re-genotyping, and carefully rescoring samples homozygous for short alleles. In addition, the MICRO-CHECKER software was used to test the genotyping of microsatellite (SSR) data and to identify various genotyping and typographic errors (Van Oosterhout et al. 2004). MICRO-CHECKER estimates the frequency of null alleles at a locus using a series of algorithms. Importantly, MICRO-CHECKER can also adjust the number of homozygote genotypes and allele and genotype frequencies of the amplified alleles, which allows the data to be used in further population genetic analysis, such as GENEPOP (Raymond and Rousset 1995a; Rousset 2008) and FSTAT (Goudet 1995, 2001).

Among 1,283 individual trees in this study, 1,236 trees were genotyped for allozyme loci, 1,210 trees for SSR loci, and 1,163 for both allozyme and SSR loci. Each set of markers was analyzed separately first, and then, both sets (allozyme and SSR loci) were analyzed jointly.

Population classification

Douglas-fir samples were assigned to specific eco-geographic groups to study their genetic differentiation. We used the 18 regions defined by St. Clair (2006) to classify populations for purposes of looking at population structure (Table 1 and Fig. 1). The regional classification is derived from the high-resolution genecological maps constructed for Douglas-fir in Oregon and Washington (St. Clair et al. 2005; St. Clair 2006). In the latter study, St. Clair (2006) defined regions based on ecoregions, latitudinal divisions within ecoregions, and elevational divisions within ecoregion and latitudinal strata. This population classification was designed to reflect known patterns of quantitative variation (St. Clair et al. 2005) and are similar to the breeding zones used in Douglas-fir breeding programs in Oregon and Washington (Howe et al. 2006). Regions were delineated using Omernik's level III ecoregions (Omernik 1995; Pater et al. 1998), but the Coast Range ecoregion was further divided at 44.2° N latitude, at the Columbia River (~46.2° N latitude), and at 47.8° N latitude. The Cascade ecoregion was further divided at 44.2° N latitude and at the Columbia River (~45.6° N latitude). All ecoregion latitudinal strata were further divided into low and high elevations at 650 m, although the Washington Coast Range and the Willamette Valley strata did not have high-elevation sources, and the Eastern

Cascades strata did not have low-elevation sources (Fig. 1). The entire sample was also studied to infer population structure without imposing any particular grouping a priori.

Statistical analysis

Descriptive gene diversity statistics, allelic richness, number of alleles, proportion of polymorphic loci, and observed (H_o) and expected (H_e) heterozygosities (Nei 1987) were calculated for each population using the FSTAT v. 2.9.3.2 software (Goudet 1995, 2001) and the Genetic Data Analysis software by Lewis and Zaykin (2001). Allelic richness is a measure of the number of alleles independent of sample size (Petit et al. 1998).

Pairwise difference between populations was evaluated using an unbiased estimate of the pairwise P values for allele distribution difference between populations (Fisher exact test) that was performed as described by Raymond and Rousset (1995b) using the GENEPOP v.4.0 software (Raymond and Rousset 1995a; Rousset 2008). In addition, for pairwise tests of differentiation, multi-loci genotypes were randomized between the two samples for each pair of samples, contingency tables of alleles by samples were generated, and the log-likelihood G statistics over all loci were used to estimate significance at 5%, 1%, or 0.1% levels using the FSTAT (Petit et al. 2001). The main advantage of this test compared to Fisher's procedure to combine individual loci P value is that loci are weighted according to their "information" content, that is, the P values for very polymorphic loci are weighted more than those for nearly monomorphic loci, unlike Fisher's procedure where each P value has the same weight. The standard (not sequential) Bonferroni corrections were used to adjust P value for each pair for multiple tests.

The genetic structure and differentiation between samples were estimated using the F statistics by Weir and Cockerham (1984), which measures the genetic variance among populations divided by the genetic variance of the total population. Significance of F (F_{IT}), θ (F_{ST}), and f (F_{IS}) was calculated using the FSTAT software and based on proportion of randomizations that gave larger F values than the observed. Testing for Hardy–Weinberg within samples was based on 1,000 randomizations of alleles within samples using f (F_{IS}) statistic.

The use of both allozyme and SSR markers increases the number of molecular markers and provides more data, but while the SSR variation located in non-coding regions is assumed to be selectively neutral, the allozyme variation can be more affected by natural selection (Altukhov 1991, 2006). Therefore, before using them together, it is important to compare results based on these two types of markers. However, comparison of population structure between allozyme and SSR loci can be difficult because the value

Table 1 Pacific Northwest ecoregions and coastal Douglas-fir regions and seed zones

Population	Region (St. Clair 2006)	Abbreviation	Number[a]	Ecoregion (US DOI BLM 2005)[b]
1	Oregon Klamath Mountains Low	OrKlMoLo	73	Klamath Mountains (78)
2	Oregon Klamath Mountains High	OrKlMoHi	96	
3	Oregon Coast South Low	OrCoSoLo	101	Coast Range (1)
4	Oregon Coast South High	OrCoSoHi	19	
5	Oregon Coast North Low	OrCoNoLo	128	
			2	Willamette Valley (3)
6	Oregon Coast North High	OrCoNoHi	16	Coast Range (1)
7	Oregon Coast Eastside Low	OrCoEaLo	1	
			38	Willamette Valley (3)
8	Washington Coast South Low	WaCoSoLo	42	Coast Range (1)
			13	Puget Lowland (2)
9	Washington Coast North Low (Olympics)	WaCoNoLo	8	Coast Range (1)
			6	Puget Lowland (2)
10	Oregon Cascades South Low	OrCaSoLo	59	Cascades (4)
			10	Willamette Valley (3)
11	Oregon Cascades South High	OrCaSoHi	168	Cascades (4)
12	Oregon Cascades North Low	OrCaNoLo	73	
			14	Willamette Valley (3)
			1	Puget Lowland (2)
	Oregon Columbia Gorge Low[c]	OrCoGoLo	4	Eastern Cascades Slopes and Foothills (9)
13	Oregon Cascades North High	OrCaNoHi	120	Cascades (4)
14	Oregon Cascades Eastside High	OrCaEaHi	37	Eastern Cascades Slopes and Foothills (9)
15	Washington Cascades South Low	WaCaSoLo	56	Cascades (4)
			3	Eastern Cascades Slopes and Foothills (9)
			9	Puget Lowland (2)
16	Washington Cascades South High	WaCaSoHi	79	Cascades (4)
			6	Eastern Cascades Slopes and Foothills (9)
17	Washington Cascades North Low	WaCaNoLo	12	Puget Lowland (2)
			45	North Cascades (77)
18	Washington Cascades North High	WaCaNoHi	44	
Total			1,283	

[a] Number of trees studied

[b] Ecoregion codes are in parentheses (see also EPA 2004). Based on Johnson et al. (2004), the ecoregions were divided into the following seed zones included in the study: North Cascades—207, 208, 210, 211; Puget Lowland—203–205, 208; Eastern Cascades Slopes and Foothills—100, 116, 211, 212; Cascades—100, 106, 108–116, 205, 206, 208, 209, 211, 212; Coast Range—101, 102, 104–108, 201, 203, 204, 205, 206; Willamette Valley—106, 108, 109, 114; Klamath Mountains—101, 102, 103, 106, 107, 110, 111, 116

[c] Due to limited sample size the Oregon Columbia Gorge Low (OrCoGoLo) population was combined with the neighboring Oregon Cascades North Low (OrCaNoLo) population

of the F_{ST} statistic depends on the amount of genetic variation within populations. A standardized measure of population genetic differentiation was developed recently that addressed this problem (Meirmans 2006). One of the advantages of the method is that it can be used to compare markers of different kinds and can be readily expanded to include different hierarchical levels in the tested population structure.

Regression analysis and the Mantel test were used to analyze whether genetic differentiation between populations correlates with geographic distance between them using the SPAGEDI v.1.2 software (Hardy and Vekemans 2002) and the PGMAN software by Saúl Lozano-Fuentes (http://www.evolcafe.com/popgen/download.htm). Euclidian distance was calculated from the Universal Transverse Mercator coordinates X, Y, and Z using Franson CoordTrans v.2.2 software (http://franson.com/coordtrans) and was used as a measure of geographic distance between populations. Genetic differentiation between populations was calculated as $\theta/(1 - \theta)$ [$F_{ST}/(1 - F_{ST})$], which approximates relationships between genetic and geographic distance better than ordinary $\theta(F_{ST})$ (Rousset 1997).

To visualize phylogenetic relationships between populations, the consensus neighbor-joining trees based on Nei's (1972) genetic distance and 10,000 bootstraps were generated using the PHYLIP software package (http://evolution.genetics.washington.edu/phylip.html).

Coalescent simulations using the Island model (Strobeck 1987) to generate the theoretical distribution of F_{ST} and to reveal outlier loci with unusually high or low differentiation were performed following analyses described in Beaumont and Nichols (1996) and the FDIST2 program (Beaumont and Balding 2004). We used the LOSITAN program to visualize the expected confidence intervals for F_{ST} vs. H_e (expected heterozygosity) distribution computed by the FDIST2 program under an Island model of migration with neutral markers (Antao et al. 2008). LOSITAN was run first using 10,000 simulations and all loci to estimate the mean neutral F_{ST}. After the first run, all loci that were outside the 95% confidence intervals were removed, and the mean neutral F_{ST} was computed again using only putative neutral loci that were not removed. A second and final run of LOSITAN, using all loci, was then conducted using the last computed mean. This procedure lowers the bias on the estimation of the mean neutral F_{ST} by removing the most extreme loci from the estimation. All loci were present in the last run, and their estimated selection status was reported.

The population structure (number of potentially different clusters) and proportion of membership of each predefined population and each individual tree in each of the inferred clusters (Q-matrix) were inferred via a model-based clustering method implementing the Markov chain Monte Carlo (MCMC) algorithm and a Bayesian framework. The individuals were assigned to several genetic groups in a way to minimize within-group linkage disequilibrium and deviation from Hardy–Weinberg equilibrium using the STRUCTURE v2.1 software (Pritchard et al. 2000; Pritchard and Wen 2004). We set most of the parameters to their default values as advised in the user's manual (Pritchard and Wen 2004). We chose the admixture model and the option of correlated allele frequencies between populations, as this configuration is considered best by Falush et al. (2003) in cases of subtle population structure, and let the degree of admixture alpha be inferred from the data. Lambda, the parameter of the distribution of allelic frequencies, was set to 1, as the manual advises. We used long burn-in and MCMC of 50,000 and 100,000 cycles, respectively, although from preliminary tests, we found that 20,000 would be sufficient for both the burn-in and MCMC. Several different runs were done to test the consistency of the results. The number of possible clusters (K) was tested from 1 to 28. St. Clair (2006) concluded that quantitative variation best fits a subdivision into 18 populations. The most likely number of populations is usually identified using the maximal value of the posterior probability of the data among given K, Pr($X|K$) called $LnP(D)$ in STRUCTURE output (Pritchard et al. 2000). However, Evanno et al. (2005) found that in many cases, the estimated $LnP(D)$ does not help visualizing a correct number of clusters, K. They demonstrated that using an ad hoc statistic ΔK based on the rate of change in the log probability of data between successive K values evaluated by STRUCTURE more accurately detects the uppermost hierarchical level of structure. Although the different population structure and migration scenarios that they tested were more complex than in Douglas-fir, we also applied their approach and found that it indeed improved analysis of data generated by STRUCTURE in our study.

Clinal allelic variation was tested using Spearman rank correlation (STATISTICA software, v.6 StatSoft, Inc., Tulsa, OK, USA; www.statsoft.com) and logistic regression (JMP software; SAS Institute, Cary, NC, USA) with longitude, latitude, and elevation as the continuous independent variables and alleles as levels of the dependent variable.

Results

Genetic diversity

The observed (H_o) and expected (H_e) heterozygosities varied from 0.005 to 0.692 among 25 allozyme loci (Table 2) and from 0.187 to 0.223 between populations for all loci (Electronic supplementary materials, Appendix 1). The mean number of alleles per population (A=3.02; Electronic supplementary materials, Appendix 1) and per locus in the entire sample (A=4.96; Table 2) and heterozygosity for allozyme loci (H_o=0.203, H_e=0.206) were relatively high. The allelic richness and heterozygosity were the highest in the Oregon Cascades South High (OrCaSoHi) population and Washington Coast South Low (WaCoSoLo), respectively, and the lowest in Oregon Coast South High (OrCoSoHi) and Washington Coast North Low (Olympics; WaCoNoLo; Electronic supplementary materials, Appendix 1).

The mean number of alleles, as well as observed and expected heterozygosities, were also very high for SSR loci (A=24.3; H_o=0.639; H_e=0.936; Table 2 and Electronic supplementary materials, Appendix 1). The mean fixation indices, f (F_{IS}), were significantly different from Hardy–Weinberg expectation for some SSR loci in several populations, but not for allozyme loci. A high frequency of null alleles and allelic dropout in the SSR loci were the most likely reasons for false homozygote genotyping that inflated fixation indices. We minimized the problem by using the MICRO-CHECKER software to adjust the number of homozygous genotypes in each size class to

Table 2 Genetic variation and differentiation in coastal Douglas-fir

Locus	Alleles	Allelic richness	H_o	H_e	$f (F_{IS}) \pm SE$[a]	$\theta (F_{ST}) \pm SE$[a]	θ_S[b]	$P \pm SE$[c]
Allozyme loci								
ACO-1	3	1.67	0.053	0.055	0.036±0.035	0.002±0.003	0.002	**0.0374**±0.0028
ACO-2	5	1.49	0.032	0.032	0.046±0.040	0.002±0.003	0.002	0.7257±0.0094
PGM-1	6	2.86	0.280	0.329	0.125±0.058	0.026±0.023	0.037	**0.0000**±0.0000
PGM-2	7	2.38	0.113	0.121	0.049±0.036	0.001±0.002	0.001	0.0915±0.0079
PGI-1	3	1.19	0.012	0.013	0.225±0.115	−0.006±0.002	−0.005	0.9130±0.0033
PGI-2	5	2.36	0.128	0.127	−0.010±0.018	0.002±0.003	0.002	0.1156±0.0071
LAP-1	7	4.55	0.657	0.692	0.044±0.019	0.004±0.002	0.010	**0.0001**±0.0001
LAP-2	6	2.42	0.130	0.127	−0.001±0.021	0.003±0.004	0.003	0.0922±0.0063
SOD	5	2.05	0.100	0.109	0.062±0.032	0.004±0.004	0.004	**0.0080**±0.0018
PMI	3	1.19	0.010	0.012	0.224±0.137	0.005±0.005	0.005	**0.0007**±0.0002
GDH	4	1.10	0.005	0.005	−0.010±0.000	0.000±0.001	0.000	0.7166±0.0097
GOT-1	3	1.11	0.007	0.007	−0.002±0.000	−0.001±0.001	−0.001	0.7289±0.0061
GOT-2	6	2.56	0.225	0.222	0.005±0.029	0.004±0.003	0.004	**0.0014**±0.0006
GOT-3	5	2.28	0.127	0.138	0.052±0.024	−0.001±0.002	−0.001	0.0928±0.0081
G-6PD	7	3.15	0.575	0.545	−0.050±0.024	−0.002±0.001	−0.003	0.8060±0.0145
GLY-D	4	2.48	0.490	0.490	0.038±0.030	0.001±0.002	0.002	0.2240±0.0129
CAT	6	2.07	0.439	0.427	−0.025±0.027	0.005±0.004	0.008	**0.0408**±0.0120
F-EST	4	1.33	0.028	0.029	0.056±0.067	0.001±0.002	0.001	0.0619±0.0042
6-PGD	5	2.32	0.139	0.139	0.041±0.033	0.001±0.003	0.001	0.0603±0.0062
IDH	7	3.13	0.338	0.345	0.008±0.023	0.001±0.002	0.001	**0.0385**±0.0055
DIA	5	2.45	0.388	0.388	0.017±0.016	0.004±0.003	0.006	**0.0004**±0.0002
MDH-1	4	1.75	0.059	0.063	0.076±0.043	0.003±0.003	0.003	**0.0299**±0.0029
MDH-2	4	1.95	0.072	0.085	0.099±0.039	0.001±0.002	0.001	0.4779±0.0100
MDH-3	5	2.45	0.346	0.333	−0.019±0.024	0.000±0.002	0.000	0.2184±0.0137
MDH-4	5	2.48	0.333	0.328	0.007±0.040	−0.003±0.002	−0.004	0.7103±0.0153
Overall	4.96	2.19	0.203	0.206	0.020±0.012	0.003±0.002	0.004	Highly significant
95% CI					−0.002–0.044	0.001–0.007		
99% CI					−0.008–0.052	0.000–0.009		
SSR loci								
1C3	45	9.79	0.383	0.959	0.625±0.021	0.003±0.002	0.070	**0.0000**±0.0000
2G12	41	8.42	0.732	0.920	0.205±0.015	0.004±0.002	0.047	**0.0000**±0.0000
3B2	47	9.85	0.801	0.959	0.174±0.015	0.003±0.001	0.067	**0.0000**±0.0000
3F1	46	9.65	0.476	0.948	0.494±0.024	0.002±0.001	0.043	**0.0000**±0.0000
3G9	32	7.12	0.596	0.874	0.314±0.019	0.001±0.002	0.007	**0.0000**±0.0000
4A7	60	9.86	0.873	0.956	0.065±0.012	0.004±0.001	0.082	**0.0000**±0.0000
Overall	46.33	9.12	0.639	0.936	0.314±0.088	0.003±0.000	0.044	Highly significant
95% CI					0.164–0.479	0.002–0.004		
99% CI					0.129–0.527	0.002–0.004		
SSR loci after correction for null alleles[d]								
1C3	46	9.14	0.710	0.942	0.186±0.028	0.004±0.003	0.057	**0.0000**±0.0000
2G12	42	8.63	0.810	0.927	0.111±0.010	0.004±0.001	0.052	**0.0000**±0.0000
3B2	48	9.94	0.823	0.960	0.146±0.011	0.003±0.001	0.070	**0.0000**±0.0000
3F1	47	9.31	0.702	0.935	0.160±0.028	0.007±0.001	0.111	**0.0000**±0.0000
3G9	33	7.53	0.776	0.890	0.073±0.012	0.003±0.001	0.026	**0.0000**±0.0000
4A7	61	9.86	0.876	0.956	0.064±0.011	0.004±0.001	0.082	**0.0000**±0.0000
Overall	46.17	9.07	0.783	0.935	0.122±0.019	0.004±0.001	0.057	Highly significant
95% CI					0.089–0.156	0.003–0.005		
99% CI					0.079–0.165	0.003–0.006		

Table 2 (continued)

Locus	Alleles	Allelic richness	H_o	H_e	$f(F_{IS})\pm SE$[a]	$\theta(F_{ST})\pm SE$[a]	θ_S[b]	$P\pm SE$[c]
Allozyme and SSR loci								
Overall	12.71	3.20	0.289	0.348	0.177±0.059	0.003±0.001	0.004	Highly significant
95% CI					0.062–0.281	0.002–0.005		
99% CI					0.035–0.316	0.001–0.006		
Allozyme and SSR loci after correction for null alleles[d]								
Overall	12.84	3.20	0.316	0.348	0.073±0.018	0.004±0.001	0.006	Highly significant
95% CI					0.037–0.103	0.002–0.006		
99% CI					0.026–0.112	0.001–0.008		

[a] Standard errors (SE) for f and θ for individual and all loci were obtained via jackknifing over populations and loci, respectively. Confidence intervals (CI) were obtained via bootstrapping over loci

[b] Standardized θ (Meirmans 2006)

[c] A log-likelihood ratio based exact G test (modified Fisher's exact probability test) for allelic differentiation (Goudet et al. 1996). Significant values are highlighted in bold

[d] Corrected using MICRO-CHECKER (van Oosterhout et al. 2004) and algorithm that has been developed to estimate null allele frequencies assuming that the population is in Hardy–Weinberg equilibrium (Chakraborty et al. 1992)

reflect the estimated "real" numbers of homozygotes. A zero value was entered to replace one of the alleles in some homozygous genotypes. Summary statistics of null allele estimates across all loci is presented for all 18 populations in Electronic supplementary materials, Appendix 2. Analysis of distribution of homozygote and allele frequencies for six SSR loci together with adjusted genotypes and corrected allele frequencies of amplified alleles based on the four methods of null allele estimation were presented only for the "OR Klamath Mountains Low" (OrKlMoLo) population as an example (Appendix 2). Similar analysis was done for all 18 populations (results are available upon request).

Pairwise differentiation between populations

All individual pairwise P values for allele distribution differences between populations are presented in Electronic supplementary materials, Appendix 3. We tested the null hypothesis of whether the allele distribution is identical across populations using the log-likelihood G statistics and Fisher exact test (see "Statistical analysis" in Materials and Methods). Electronic supplementary materials, Appendix 3 also presents pairwise values for standardized θ_S. They varied from 0 to 0.0506 (between OrCoEaLo and WaCo-SoLo) for the allozyme loci, from 0 to 0.2623 (between OrCoSoHi and WaCaNoLo) for the SSR loci, and from 0 to 0.0368 (between OrCoSoHi and WaCoSoLo) for the combined loci. The WaCoSoLo and Washington Cascade Mountain (WaCaSoLo, WaCaNoLo, and WaCaNoHi) populations were the most different from other populations, while the Oregon Coast North High (OrCoNoHi) population was similar to all populations. SSR alone and

combined with allozyme markers corroborated these observations, although more pairwise comparisons became significant based on the SSR or combined markers.

Allelic heterogeneity

Allelic differentiation was significant for the ACO-1, PGM-1, LAP-1, SOD, PMI, GOT-2, CAT, IDH, DIA, and MDH-1 loci among populations (see P values in Table 2). The allele frequencies were heterogeneous for all SSR loci, as well as globally for both marker sets separately and combined.

Genetic differentiation between populations

θ and θ_S were relatively low for both allozyme ($\theta=0.000–0.026$; $\theta_S=0.000–0.037$) and SSR ($\theta=0.001–0.004$; $\theta_S=0.007–0.082$) loci and combined ($\theta=0.003$; $\theta_S=0.004–0.044$), although values over all loci were statistically significantly different from zero in all cases based on confidence intervals (CI, 95%; Table 2). Coalescence simulation using overall $F_{ST}=0.005$ as an expected value for neutral markers and infinite allele model found the upper and the low F_{ST} limits at the 0.95 probability level (Fig. 2). The loci PGM-1 ($H_e=0.342$, $F_{ST}=0.038$, $P=0.9973$), 1C3 ($H_e=0.962$, $F_{ST}=0.018$, $P=0.9999$), and 3F1 ($H_e=0.955$, $F_{ST}=0.027$, $P=1.0$) were candidates for positive selection and locus MDH-4 ($H_e=0.327$, $F_{ST}=0.0$, $P=0.039$) for balancing selection (Fig. 2).

Genetic vs. geographic distance and clinal allelic variation

Neighbor-joining trees revealed no significant clustering except a few clusters with bootstrap values more than 50%,
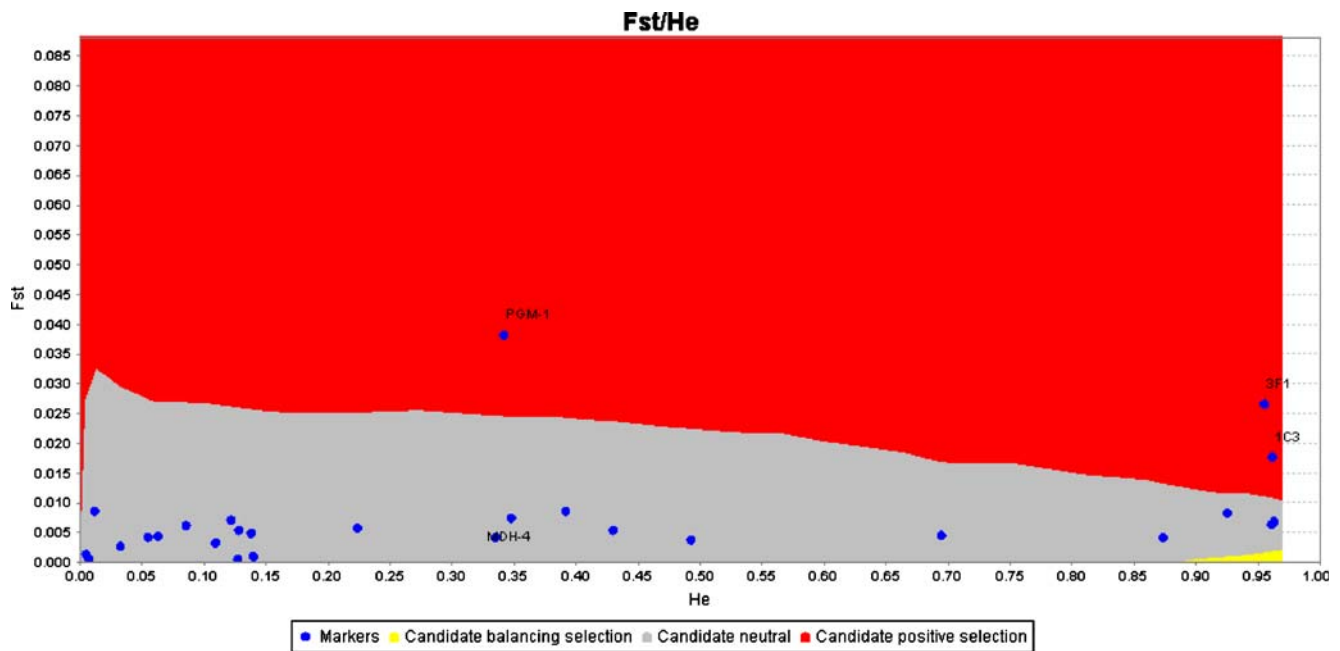
Fig. 2 F_{ST} and H_e distribution for 25 allozyme and six SSR loci based on differentiation among 18 coastal Douglas-fir populations. The LOSITAN (Antao et al. 2008) screen shot showing a graphical output with the simulated confidence area for neutral loci (gray color, middle band) with loci from the original empirical dataset represented as *dots* and computed using the FDIST2 simulation program (Beaumont and Nichols 1996; Beaumont and Balding 2004). Outliers are tagged with labels

but phylogenetic tree topography mostly followed geographic location of studied samples (Fig. 3), which was more pronounced when both types of markers were combined together (Fig. 3C). Elevation appeared to play a significant role in clustering. The populations from the same elevation had a tendency to cluster together, especially considering significant clusters. For instance, WaCaNoLo and WaCoNoLo, WaCoSoLo, and Oregon Coast North Low (OrCoNoLo) and OrCoSoHi and OrCoNoHi populations formed significant clusters (Fig. 3A, C).

The genetic distances were positively correlated with geographic distance for both types of markers (Fig. 4), but were statistically significant only for the SSR markers and combined data (Fig. 4B, C). Weak but significant clinal variation was found for several alleles (see Table 3 for examples). For instance, frequency of the CAT allele 1 significantly increases, while allele 3 decreases from south to north (Fig. 5A). The DIA alleles 2 and 3 gradually increase from east to west (Fig. 5B), and the PGM-1 allele 3 significantly increases, while alleles 1 and 2 decrease with elevation (Fig. 5C).

Population structure

Despite the non-random distribution of allelic variation in the studied area, the simulation study using the STRUCTURE software found very weak subdivision for the admixture ancestry and correlated allele frequency model tested under assumptions of $K=1, 2, 3,..., 28$ (e.g., Figs. 1S and 2S and Table 2S, Electronic supplementary materials; Q-matrix with inferred ancestry of individuals in each of the assigned 18 populations is available from the authors upon request). The differences in $LnP(D)$ were small for allozyme and SSR markers (Fig. 6), there were no clear geographic interpretation for the assignments, the assignments were roughly symmetric to all populations tested, and almost no individuals were strongly and unambiguously assigned. The $LnP(D)$ was almost unchanging for SSR markers for all K and for allozyme markers for K from 1 to 7. It dropped for allozyme markers for $K>7$. There was a pronounced drop for combined markers for $K=17$. Although the highest $LnP(D)$ for combined markers was for $K=15$, ΔK (Evanno et al. 2005) clearly demonstrated that the uppermost K equaled 18 (Fig. 6), which corresponded exactly to the same number of populations used in this study and suggested in previous analysis of quantitative adaptive trait variation (St. Clair 2006). There were a couple of ΔK peaks for allozyme loci at $K=13$ and 17, while ΔK was practically invariable for SSR markers.

Discussion

The mean number of alleles ($A=3.02$) and heterozygosity for allozyme loci ($H_o=0.203$, $H_e=0.206$) were relatively high compared to conifers studied for a similar number of allozyme loci, but within the observed range ($A=1.5-2.5$, $H_e=0.120-0.220$; see El-Kassaby 1991; Hamrick et al.
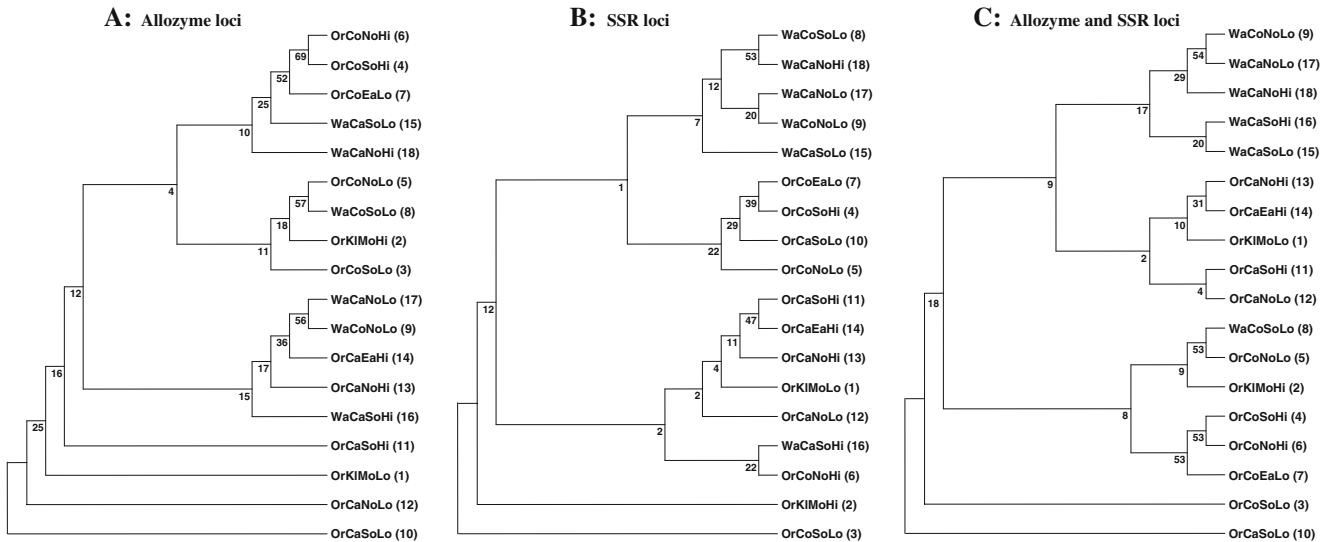
Fig. 3 Consensus neighbor-joining trees based on Nei's 1972 genetic distance and 10,000 bootstraps for 18 coastal Douglas-fir populations. The number near each node represents the percentage of times when the nodes occurred among 10,000 bootstraps
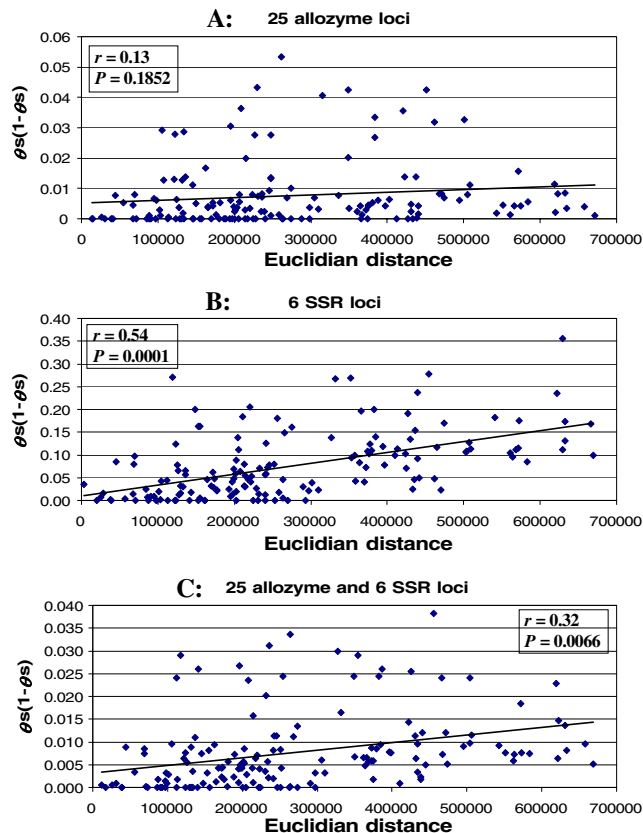
Fig. 4 Correlation between genetic vs. geographic distance for regions. Genetic distances were calculated as pairwise $\theta_S/(1 - \theta_S)$ values, where θ_S is a standardized θ (see "Material and Methods")

1992; Krutovskii et al. 1995; Ledig 1986, 1998). They were also higher than the mean expected heterozygosity (0.137) observed in an earlier range-wide study based on 20 enzyme loci and 104 geographical locations throughout the range of Douglas-fir (Li and Adams 1989), but also within the range (0.021–0.239) observed in that study. Relatively high genetic diversity within studied populations and low differentiation among populations (as measured by allozymes) were consistent with data for other conifers in general and for those long-living woody plant species that have large effective population size and are also wind-dispersed and wind-pollinated (Hamrick et al. 1992).

The mean number of alleles, observed and expected heterozygosities for SSR loci (A=11–36; H_o=0.549–0.718; H_e=0.877–0.944; see Electronic supplementary materials, Appendix 1) were also relatively higher than those estimated with the same type and number of markers in conifers with wide geographical distributions, such as *Larix occidentalis* (A=5.5, H_o=0.521, H_e=0.580; Khasa et al. 2006), *Pinus strobus* (A=9.6, H_o=0.522, H_e=0.607; Rajora et al. 2000; Marquardt and Epperson 2004), *Pinus contorta* (A=21.0, H_e=0.425; Thomas et al. 1999), *Picea glauca* (A=16.4, H_o=0.649, H_e=0.851; Rajora et al. 2005), and also *P. menziesii* (A=7.5, H_e=0.673; Amarasinghe and Carlson 2002) studied earlier. The difference can be explained by the fact that the SSR loci in our study were preselected as the most polymorphic.

The mean fixation indices f (F_{IS}) were significantly different from Hardy–Weinberg expectation for the SSR, but not allozyme loci. Both types of markers were genotyped in nearly the same samples, and this difference attests that a high frequency of null alleles and allelic dropout in the SSR loci were the major reasons for false

Table 3 Alleles with clinal variation in coastal Douglas-fir based on Spearman rank correlation and logistic regression

Locus	Latitude			Longitude			Elevation		
	Southward	Northward	P	Eastward	Westward	P	Upward	Downward	P
ACO-1			NS	1	3	0.0306	3		0.0682
ACO-2			NS		3	NS	2	4	0.0671
PGM-1		1	0.0546	2	3	<0.0001	3	1, 2	<0.0001
PGM-2		2	0.0684			NS			NS
LAP-1	2, 5	3, 4, 8	<0.0001	2	6, 7	0.0011	5, 7	2–4, 8	<0.0001
LAP-2	2, 3	5	0.0582			NS		5	NS
SOD	2	4	0.0002			NS			NS
PMI	2	1	<0.0001			NS	2	1	0.0768
GOT-3	2		0.0550	5	6	0.0782	2	5	0.0012
GLY-D	1		0.0194			NS			NS
CAT	3	1	NS			NS	1	3	0.0620
F-EST		1	NS		1	0.0079	NS		NS
6-PGD	4	1	0.0065			NS	4	1, 2	NS
IDH	2	5	0.0173			NS			NS
DIA	4	3	0.0083	4	2, 3	0.0013			NS
MDH-1	2	1	0.0004			NS	2	1	0.0398
MDH-2	3	4	0.0146			NS			NS
MDH-3			NS			NS		6	NS
1C3	13	20, 36, 41, 42	<0.0001	26, 40, 41	25	0.0005	22	26, 39, 41	<0.0001
2G12	12, 15, 31	13, 19	<0.0001	14	19	0.0050	12, 25	5	<0.0001
3B2	9, 24, 26, 29, 31, 41	2, 6, 12, 13, 14, 17, 18, 27, 35	<0.0001	7, 24	2, 13, 18, 20, 35	<0.0001	2, 20, 28	6, 7, 17, 29, 39	<0.0001
3F1	6, 7, 25, 26	9, 10, 16	<0.0001	20, 30, 39	1, 8, 45	<0.0001	1, 8, 14, 29, 43, 45	9, 15, 22, 25, 39	<0.0001
3G9	1, 8, 17	3, 18	0.0002	1, 7	18	<0.0001	11, 15, 17, 28	3, 4	0.0005
4A7	3, 12–14, 18, 25, 29, 49, 52	9, 33, 44	<0.0001	18, 21, 37, 50	9, 24, 26	0.0040	6, 9, 12, 26, 29, 49, 52	21, 30	0.0024

P—a whole model test probability for all alleles in an entire locus in logistic regression of latitude, longitude, and elevation on alleles. Alleles that had significant Spearman rank correlation are highlighted by italic, and alleles whose correlations were confirmed by both logistic regression and Spearman rank correlation are highlighted by bold

NS non-significant
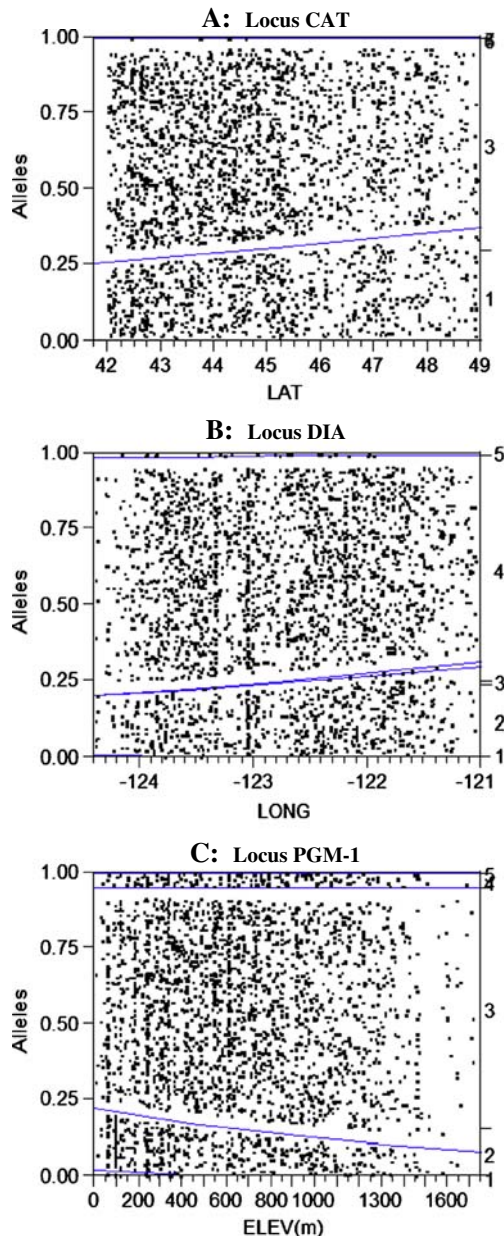
Fig. 5 Examples of logistic fit of the allozyme alleles by latitude (LAT), longitude (LONG), and elevation (ELEV). The left y-axis represents probability, and the proportions shown on the right y-axis represent the relative sizes of each allele group in the total sample. At each x value, the probability scale in the y direction is divided up (partitioned) into probabilities for each response allele category. The probabilities are measured as the vertical distance between the curves, with the total across all Y category probabilities sum to 1. Dots represent individual alleles that are drawn at their x-coordinate, with the y position jittered randomly within the range corresponding to the response allele category for that row. One can see that the dots tend to push the lines apart and make vertical space where they occur in numbers and allow the curves to get close together where there is no data. If the x variable has no effect on the response, then the fitted lines are horizontal and the probabilities are constant for each response across the continuous factor range

homozygote genotyping that inflated fixation indices in our study (e.g., Pompanon et al. 2005). Allelic dropout is the most prevalent source of genotyping error in general (Miller et al. 2002) and is the most likely case in our study due to using the DNA isolated from pooled megagametophytes. If the segregation ratio deviates much from 1:1 in a pooled individual sample, then a prevalent allele can over-compete an underrepresented allele during PCR, especially if a former one represents a shorter DNA fragment (aka "short allele dominance"). Several solutions were offered recently to deal with this problem (see Dewoody et al. 2006 for review), but most of them require multiple repeats or at least duplicate microsatellite genotypes (e.g., Miller et al. 2002; Johnson and Haydon 2007). We used the MICRO-CHECKER software that proved to be very efficient to minimize the problem. The observed heterozygosity based on the SSR genotypes corrected for null alleles increased, and discrepancy between observed and expected heterozygosity greatly decreased, respectively (Table 2). The fixation index f (F_{IS}) also decreased several folds, while estimates of differentiation [$\theta(F_{ST})$ and θ_S] remained almost the same (Table 2).

Pairwise comparisons demonstrated that the OrCoNoHi population was the most genetically similar to all populations. This population and the WaCoSoLo population were also the most polymorphic for SSR loci. These data support the hypothesis that southwestern Washington and western Oregon could have been glacial age refugia for coastal Douglas-fir (see papers discussed in St. Clair et al. 2005, p. 1212).

The values of standardized genetic differentiation (θ_S) for the SSR loci were several folds higher than for the allozyme loci (Table 2). One may suggest that the allozyme loci represent important primary metabolism and housekeeping genes that are likely under stabilizing selection that generally unifies allele frequencies across the entire area, while the SSR variation is apparently more selectively neutral and more affected by genetic drift that increases allelic heterogeneity. However, the distribution of individual values was relatively broad among the allozyme loci ($\theta=0.000–0.026$, $\theta_S=0.000–0.037$). Coalescence simulation study suggested that the PGM-1 locus with an especially high level of differentiation could represent the outlier locus that is probably under divergent or positive selection, while the MDH-4 locus with relatively high heterozygosity but lack of differentiation can be a candidate for the outlier locus that is probably under balancing selection (Fig. 2). The SSR loci 1C3 and 3F1 had not only a high level of heterozygosity, but unexpectedly also a relatively high level of differentiation. Additional analysis and data are needed to determine whether their high differentiation is a result of selection acting on a gene or genes closely linked to these markers or is a sign of recent demographic events that

Fig. 6 Estimates of the posterior probability of the data for a given *K*, *LnP(D)* (Pritchard et al. 2000) obtained in STRUCTURE analysis and an ad hoc quantity Δ*K* based on the second-order rate of change of the likelihood function with respect to *K* (Evanno et al. 2005) computed separately for allozyme and SSR loci and for all loci combined
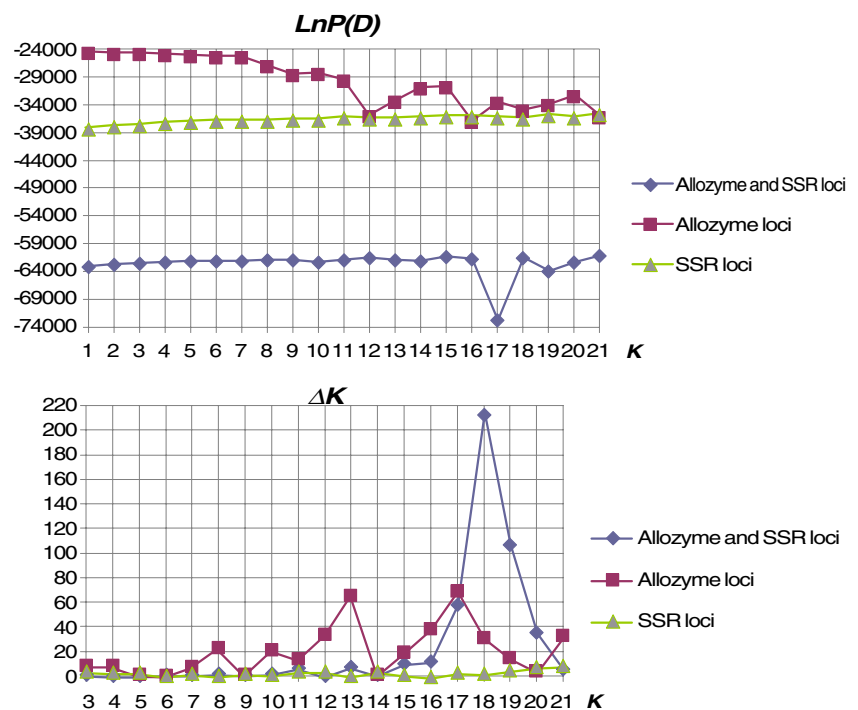
could lead to the violation of the Island model assumptions. The observed variation was rather neutral in most loci under the applied model of molecular evolution.

The relatively low genetic differentiation that was observed between populations in the study is typical for conifers for allozyme and SSR loci (see discussion and references above), but direct comparison between allozyme and SSR loci would be inappropriate without taking within-population variation into account and calculating standardized measures (Hedrick 2005; Meirmans 2006). The ordinary G_{ST} and F_{ST} estimates dependent on the amount of within-population genetic variation and high levels of genetic variation therefore generally lead to lower F_{ST} variation. This dependency makes it difficult to compare genetic markers with different mutation rates, such as microsatellites vs. allozymes. Standardized genetic differentiation measures based on G_{ST} (Hedrick 2005) and F_{ST} (Meirmans 2006) were suggested relatively recently. They are independent of the amount of genetic variation and therefore suitable for comparisons between studies that employ different genetic markers. Although these measures are relatively new, they were already used in a few recently published studies that, together with our study, validated their use (e.g., Criscione et al. 2007; Kettle et al. 2007; Langergraber et al. 2007; Ryynänen et al. 2007; Leinonen et al. 2008). Standardized F_{ST} was usually several times higher than unstandardized F_{ST} for microsatellite loci (e.g., Criscione et al. 2007; Jordan and Snell 2008; González-Pérez et al. 2009). Standardizing of F_{ST} became now a required procedure when different markers are compared or combined

for analysis (e.g., Langergraber et al. 2007; Ryynänen et al. 2007). Therefore, our study as well as other published studies confirmed a necessity of using a standardized measure of genetic differentiation when measures are based on different genetic markers.

The elevation played a significant role in clustering. Similar elevations from different populations tended to cluster together in the dendrogram that was based on allozyme loci, which are probably under selection (Fig. 3A). This interesting observation confirmed results from the common garden study that indicated considerable differentiation in bud set, emergence, and growth associated with elevation (e.g., St. Clair et al. 2005). On the other hand, the dendrogram based on the neutral SSR markers followed mostly geographic distribution so that high and low elevations within populations tended to cluster together (Fig. 3b). It appears that natural selection for adaptive traits and some allozyme loci has been strong despite high levels of gene flow between elevations within populations. Differentiation of adaptive variation as found in common garden studies and differentiation based on some supposedly selective allozyme loci compared to the lack of differentiation based on neutral molecular markers point to the importance of adaptive trait related genetic markers and common garden studies for delineating seed and breeding zones. For example, population differentiation at quantitative traits (Q_{ST}) related to adaptive traits, such as cold hardiness, bud burst, and bud set were from threefold to 15-fold greater than differentiation at anonymous and presumably neutral SSR markers (F_{ST}), suggesting the

action of natural selection acting upon these traits (Q_{ST}= 0.15–0.66, see Table 2 in St. Clair 2006, vs. standardized F_{ST}, θ_S=0.044, Table 2).

Considering that different evolutionary factors may affect variation in allozyme and SSR markers, it is also important to note here that one should be cautious when combining allozyme and SSR genotypic data for phylogenetic analysis. We would recommend doing a separate analysis first for each type of marker.

The positive and significant correlation between genetic and geographic distances for the SSR markers, but not for the allozyme loci (Fig. 4), suggests that isolation by distance may play an important (if not the most important) role in the coastal Douglas-fir differentiation in neutral SSR markers, while allozyme variations are under different (mostly stabilizing) forms of selection. Weak but significant clinal variation observed in several alleles may also suggest a subtle gradual genetic adaptation along ecological gradients (Table 3). It would be interesting to study whether any particular climatic factors are associated with observed clinal variation, but ecological data are very dynamic and much more complex than geographic variables and would need a separate analysis, which is beyond the scope of this paper. However, the geographic gradients are likely to couple with ecological ones in this range-wide study. Anyway, these observations corroborate results of the exact allelic differentiation tests that revealed a significant allelic differentiation ($P<<0$) among populations for many loci (Table 2). The genetic differentiation between populations (θ), although very low, was also significant in most cases (Appendix 3, Electronic supplementary materials). All these observations suggest that despite relatively low differentiation, genetic variation is not absolutely randomly distributed across the studied area. However, the simulation study using the STRUCTURE program failed to infer population substructure for supposedly selectively neutral SSR markers. Adding allozyme loci changed analysis, and the STRUCTURE program suggested that the uppermost number of clusters equaled the same number of populations used in this study and suggested in previous analysis of quantitative adaptive trait variation (St. Clair 2006). It demonstrated that the non-random distribution of allelic variation in the studied area was also due to supposedly selective allozyme variation, but in general was apparently insufficient to generate significant population subdivision, especially among neutral markers. In the association study of the same samples genotyped for 384 SNPs representing 117 genes, only seven SNPs demonstrated a significant differentiation (A. J. Eckert et al., unpublished).

However, only seven populations could be unambiguously assigned to a particular cluster each. Another five and six populations could be likely assigned to more than one or two clusters, respectively (Table 3S, Electronic supplementary

materials). Many individuals from the same population also had often memberships in different clusters. This indicates that the 18 genetic groups assigned by STRUCTURE did not necessarily correspond to those 18 populations delineated by St. Clair (2006). This can be explained by the sampling strategy used in this continuously distributed species with gradual isolation by distance. There is apparently no "true number" of populations and clear-cut divisions within the continuously distributed species range that more or less effectively partition variation and thus are more likely or less likely to be appropriate. The sampling strategy used in this study makes this more likely the case, unlike most other population genetic studies that used sampling of individuals at fewer geographic locations more isolated from each other and then studied if individuals are reassigned to their populations at those discrete locations.

Earlier published tests and comparative studies using empirical datasets have proven STRUCTURE's ability in assigning individuals to their known cluster of origin for highly differentiated populations (e.g., Pritchard and Donnelly 2001; Rosenberg et al. 2001; Manel et al. 2002; Turakulov and Easteal 2003), especially for populations that are relatively evenly distributed in space and with migration patterns not departing much from Wright's island model such as in Douglas-fir. Although little is known on the discriminating power of STRUCTURE to detect the real number of clusters (K) which composes a dataset (Evanno et al. 2005), it is unlikely that STRUCTURE would fail to detect clusters in Douglas-fir, if they would exist in reality. However, Evanno et al. (2005) found that in many cases, the estimated posterior probability of the data for a given K, $Pr(X|K)$ that is called $LnP(D)$ in STRUCTURE output and based on the log-likelihood of the data (Pritchard et al. 2000) does not help visualizing a correct number of clusters, K. They suggested an ad hoc statistic ΔK based on the rate of change in the log probability of data between successive K values evaluated by STRUCTURE and demonstrated that it accurately detects the uppermost hierarchical level of structure for population structure and migration scenarios that were more complex than Douglas-fir. Therefore, we analyzed the STRUCTURE results using also an ad hoc statistic ΔK. This method showed a clear peak at the "true" value of K=18 when both marker sets were combined (Fig. 6).

Conclusion

Our study demonstrated that the allozyme and SSR genotypic data can be combined for joint analysis of population structure and can improve results, but first, they should always be analyzed also separately due to their different nature. The SSR markers should be checked for

the allelic dropout and presence of null alleles, and specific corrections should be made if needed to mitigate their effects. To make comparisons between two types of markers legitimate, the standardized measures of differentiation should be calculated first.

Although genetic distance between populations positively and significantly correlated with geographic distance, clustering made geographic sense, and a weak but significant clinal variation was found for a few alleles, the results demonstrated a subtle differentiation (θ_s=0.003–0.044) for both allozyme and SSR loci in general. The STRUCTURE simulation analysis inferred a weak population subdivision in the studied area that should not interfere much with association mapping but has to be taken into account. Results confirmed previous observations that coastal Douglas-fir has relatively little within-population substructure based on allozyme loci (Merkle and Adams 1987; Moran and Adams 1989; Aagaard et al. 1998a, b; Viard et al. 2001). The studied coastal Douglas-fir populations apparently represent large and rather continuous populations with practically unobstructed gene flow that prevent strong genetic differentiation due to genetic drift alone. More importantly, our study suggests that the samples used in the analysis can be used together efficiently as a single association mapping population. Douglas-fir is an excellent perennial plant species for studying adaptive phenotypic traits and genetic adaptation using association mapping. It is evolutionarily old; phenotypically and genetically highly diverse; and distributed in large, outcrossed, natural populations with high gene flow. All these should facilitate the association mapping in this experimental population, which is currently in progress.

References

Aagaard JE, Krutovskii KV, Strauss SH (1998a) RAPDs and allozymes exhibit similar levels of diversity and differentiation among populations and races of Douglas-fir. Heredity 81:69–78

Aagaard JE, Krutovskii KV, Strauss SH (1998b) RAPD markers of mitochondrial origin exhibit lower population diversity and higher differentiation than RAPDs of nuclear origin in Douglas-fir. Mol Ecol 7:801–812

Adams WT, Neale DB, Doerksen AH, Smith DB (1990) Inheritance and linkage of isozyme variants from seed and bud tissues in coastal Douglas-fir [Pseudotsuga menziesii var. menziesii (Mirb.) Franco]. Silvae Genet 39:153–167

Aitken SN, Adams WT (1997) Spring cold hardiness under strong genetic control in Oregon populations of coastal Douglas-fir. Can J For Res 27:1773–1778

Altukhov YP (1991) The role of balancing selection and overdominance in maintaining allozyme polymorphism. Genetica 85:79–90

Altukhov YP (2006) Intraspecific genetic diversity. Springer, Berlin

Amarasinghe V, Carlson JE (2002) The development of microsatellite DNA markers for genetic analysis in Douglas-fir. Can J For Res 32:1904–1915

Andersen JR, Schrag T, Melchinger AE, Zein I, Lubberstedt T (2005) Validation of Dwarf8 polymorphisms associated with flowering time in elite European inbred lines of maize (Zea mays L.). Theor Appl Genet 111:206–217

Anekonda TS, Adams WT, Aitken SN, Neale DB, Jermstad KD, Wheeler NC (2000) Genetics of cold-hardiness in a cloned full-sib family of coastal Douglas-fir. Can J For Res 30:837–840

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a F_{st}-outlier method. BMC Bioinformatics 9:323

Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M, Jakob K, Lister C, Molitor J, Shindo C, Tang C, Toomajian C, Traw B, Zheng H, Bergelson J, Dean C, Marjoram P, Nordborg M (2005) Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes. PLoS Genet 1(5):e60. doi:10.1371

Ball RD (2007) Statistical analysis and experimental design, Ch8. In: Oraguzie NC, Rikkerink EHA, Gardiner SE, de Silva HN (eds) Association mapping in plants. Springer, New York

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. Mol Ecol 13:969–980

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. Proc R Soc Lond B Biol Sci 263:1619–1626

Campbell RK (1979) Genecology of Douglas-fir in a watershed in the Oregon Cascades. Ecology 60(5):1036–1050

Campbell RK, Sorensen FC (1978) Effect of test environment on expression of clines and on delimitation of seed zones in Douglas-fir. Theor Appl Genet 51:233–246

Campbell RK, Sugano AI (1975) Phenology of bud burst in Douglas-fir related to provenance, photoperiod, chilling and flushing temperature. Bot Gaz 136:290–298

Camus-Kulandaivelu L, Veyrieras J-B, Madur D, Combes V, Fourmann M, Barraud S, Dubreuil P, Gouesnard B, Manicacci D, Charcosset A (2006) Maize adaptation to temperate climate: relationship with population structure and polymorphism in the Dwarf8 gene. Genetics 172:2449–2463

Camus-Kulandaivelua L, Veyrierasa J-B, Gouesnardb B, Charcosseta A, Manicacciaa D (2007) Evaluating the reliability of structure outputs in case of relatedness between individuals. Crop Sci 47:887–890

Carlson JE, Traore A, Agrama HA, Krutovsky KV (2007) Douglas-fir. In: Kole C (ed) Genome mapping and molecular breeding in plants, vol. 7: forest trees. Springer, Berlin, pp 199–210

Chakraborty R, De Andrade M, Daiger SP, Budowle B (1992) Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. Ann Hum Genet 56:45–57

Criscione CD, Anderson JD, Sudimack D, Peng W, Jha B, Williams-Blangero S, Anderson TJC (2007) Disentangling hybridization and host colonization in parasitic roundworms of humans and pigs. Proc R Soc B 274:2669–2677

Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55:997–1004

Devlin B, Roeder K, Wasserman L (2001) Genomic control, a new approach to genetic-based association studies. Theor Popul Biol 60:155–166

DeWoody J, Nason JD, Hipkins VD (2006) Mitigating scoring errors in microsatellite data from wild populations. Mol Ecol Notes 6 (4):951–957

El-Kassaby YA (1991) Genetic variation within and among conifer populations: review and evaluation of methods. In: Fineschi S, Malvolti ME, Cannata F, Hattemer HH (eds) Biochemical markers in the population genetics of forest trees. Proceedings of the International Workshop, Porano-Orvieto, Italy, October 1988. SPB Academic Publishing B. V., The Hague, pp 59–74

El-Kassaby YA, Sziklai O, Yeh FC (1982) Linkage relationships among 19 polymorphic allozyme loci in coastal Douglas-fir (Pseudotsuga menziesii var. 'menziesii'). Can J Genet Cytol 124:101–108

EPA (2004) Level III Ecoregions. Western Ecology Division, Corvalis, Oregon. Available at http://www.epa.gov/wed/pages/ecoregions/level_iii.htm. Revised March 2007

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–262

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

González-Martínez SC, Krutovsky KV, Neale DB (2006) Forest tree population genomics and adaptive evolution. New Phytol 170 (2):227–238

González-Martínez SC, Wheeler NC, Ersoz E, Nelson CD, Neale DB (2007) Association genetics in Pinus taeda L. I. Wood property traits. Genetics 175(1):399–409

González-Pérez MA, Sosa PA, Rivero E, González-González EA, Naranjo A (2009) Molecular markers reveal no genetic differentiation between Myrica rivas-martinezii and M. faya (Myricaceae). Ann Bot 103:79–86

Goudet J (1995) FSTAT (vers. 1.2): a computer program to calculate F-statistics. J Heredity 86:485–486

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available at http://www.unil.ch/izea/softwares/fstat.html

Goudet J, Raymond M, de Meeüs T, Rousset F (1996) Testing differentiation in diploid populations. Genetics 144:1931–1938

Hamrick JL, Godt MJ, Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plant species. New Forest 6:95–124

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol Ecol Notes 2:618–620

Hedrick PW (2005) A standardized genetic differentiation measure. Evolution 59:1633–1638

Howe GT, Jayawickrama KJ, Cherry ML, Johnson GR, Wheeler NC (2006) Breeding Douglas-fir. In: Janick J (ed) Plant breeding reviews, vol 27. Wiley, New York, pp 245–353

Jermstad KD, Reem AM, Wheeler NC, Neale DB (1994) Inheritance of restriction fragment length polymorphisms, random amplified polymorphic DNAs and isozymes in coastal Douglas-fir. Theor Appl Genet 89:758–766

Jermstad KD, Bassoni DL, Jech KS, Wheeler NC, Neale DB (2001a) Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. I. Timing of vegetative bud flush. Theor Appl Genet 102:1142–1151

Jermstad KD, Bassoni DL, Wheeler NC, Anekonda TS, Aitken SN (2001b) Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. II. Spring and fall cold-hardiness. Theor Appl Genet 102:1152–1158

Jermstad KD, Bassoni DL, Jech KS, Ritchie GA, Wheeler NC (2003) Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. III. QTL by environment interactions. Genetics 165:1489–1506

Johnson PCD, Haydon DT (2007) Maximum-likelihood estimation of allelic dropout and false allele error rates from microsatellite genotypes in the absence of reference data. Genetics 175:827–842

Johnson GR, Sorenson FC, St Clair JB, Cronn RC (2004) Pacific Northwest forest tree seed zones: a template for native plants? Nativ Plants J 5:131–140

Jordan MA, Snell HL (2008) Historical fragmentation of islands and genetic drift in populations of Galápagos lava lizards (Microlophus albemarlensis complex). Mol Ecol 17:1224–1237

Kettle CJ, Hollingsworth PM, Jaffré T, Moran B, Ennos RA (2007) Identifying the early genetic consequences of habitat degradation in a highly threatened tropical conifer, Araucaria nemorosa Laubenfels. Mol Ecol 16:3581–3591

Khasa DP, Jaramillo-Correa JP, Jaquish B, Bousquet J (2006) Contrasting microsatellite variation between subalpine and western larch, two closely related species with different distribution patterns. Mol Ecol 15:3907–3918

Krutovskii KV, Politov DV, Altukhov YP (1995) Isozyme study of population genetic structure, mating system and phylogenetic relationships of the five stone pine species (subsection Cembrae, section Strobus, subgenus Strobus). In: Baradat P, Adams WT, Mller-Starck G (eds) Population genetics and genetic conservation of forest trees. Proceedings of the International Symposium, Carcans Maubuisson, France, 24–28 August 1992. SPB Academic Publishing, Amsterdam, The Netherlands, pp 279–304

Krutovsky KV, Neale DB (2005a) Forest genomics and new molecular genetic approaches to measuring and conserving adaptive genetic diversity in forest trees. In: Geburek T, Turok J (eds) Conservation and management of forest genetic resources in Europe. Arbora, Zvolen, Slovakia, pp 369–390 Available online at http://www.fao.org/DOCREP/003/X6884E/X6884E00.HTM

Krutovsky KV, Neale DB (2005b) Nucleotide diversity and linkage disequilibrium in cold hardiness and wood quality related candidate genes in Douglas-fir. Genetics 171:2029–2041

Langergraber KE, Siedel H, Mitani JC, Wrangham RW, Reynolds V, Hunt K, Vigilant L (2007) The genetic signature of sex-biased migration in patrilocal chimpanzees and humans. PLoS ONE 2 (10):e973. doi:10.1371/journal.pone.0000973

Ledig FT (1986) Heterozygosity, heterosis and fitness in outbreeding plants. In: Soule ME (ed) Conservation biology: the science of scarcity and diversity. Sinauer, Sunderland, Massachusetts, pp 77–104

Ledig FT (1998) Genetic variation in Pinus. In: Richardson DM (ed) Ecology and biogeography of Pinus. Cambridge University Press, Cambridge, UK, pp 251–280

Leinonen T, O'Hara RB, Cano JM, Merila J (2008) Comparative studies of quantitative trait and neutral marker divergence: a meta-analysis. J Evol Biol 21:1–17

Lewis PO, Zaykin D (2001) Genetic data analysis: computer program for the analysis of allelic data. Version 1.0 (d16c). Free program distributed by the authors over the Internet from http://lewis.eeb.uconn.edu/lewishome/software.html

Li P, Adams WT (1989) Range-wide patterns of allozyme variation in Douglas-fir (Pseudotsuga menziesii). Can J For Res 19:149–161

Li P, Adams WT (1993) Genetic control of bud phenology in pole-size trees and seedlings of coastal Douglas-fir. Can J For Res 23:1043–1051

Liu KJ, Goodman M, Muse S, Smith JS, Buckler E, Doebley J (2003) Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. Genetics 165:2117–2128

Manel S, Brthier P, Luikart G (2002) Detecting wildlife poaching: identifying the origin of individuals with Bayesian assignment tests and multilocus genotypes. Conserv Biol 16:650–659

Marquardt PE, Epperson BK (2004) Spatial and population genetic structure of microsatellites in white pine. Mol Ecol 13:3305–3315

Meirmans PG (2006) Using the AMOVA framework to estimate a standardized genetic differentiation measure. Evolution 60:2399–2402

Merkle SA, Adams WT (1987) Pattern of allozyme variation within and among Douglas-fir breeding zones in southwest Oregon. Can J For Res 17:402–407

Miller CR, Joyce P, Waits LP (2002) Assessing allelic dropout and genotype reliability using maximum likelihood. Genetics 160:357–366

Moran GF, Adams WT (1989) Microgeographical patterns of allozyme differentiation in Douglas-fir from southwest Oregon. For Sci 35:3–15

Neale DB, Weber JC, Adams WT (1984) Inheritance of needle tissue isozymes in Douglas-fir. Can J Genet Cytol 26:459–468

Nei M (1972) Genetic distance between populations. Am Nat 106:283–292

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Omernik JM (1995) Ecoregions: a spatial framework for environmental management. In: Davis WS, Simon TP (eds) Biological assessment and criteria: tools for water resource planning and decision making. Lewis, Boca Raton, FL, pp 49–62

Pande B, Krutovsky KV, Jermstad KD, Howe GT, St. Clair JB, Wheeler NC, Neale DB (2007) Candidate gene-based association genetics in Douglas-fir (Poster #502). In: Plant & Animal Genome XV. The International Conference on the Status of Plant and Animal Genome Research, Final Program and Abstracts Guide, January 13–17, 2007. San Diego, CA, USA, p 227. http://www.intl-pag.org/15/abstracts/PAG15_P05j_502.html

Pater DE, Bryce SE, Thorson TD, Kagan J, Chappel C, Omernik JM, Azevedo SH, Woods AJ (1998) Ecoregions of western Washington and Oregon. Map scale 1:1,350,000. US Geological Survey, Reston, VI

Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. Conserv Biol 12:844–855

Petit E, Balloux F, Goudet J (2001) Sex biased dispersal in a migratory bat: a characterization using sex-specific demographic parameters. Evolution 55:635–640

Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences and solutions. Nat Rev Genet 6:847–859

Pritchard JK, Donnelly P (2001) Case–control studies of association in structured or admixed populations. Theor Popul Biol 60:227–237

Pritchard JK, Rosenberg NA (1999) Use of unlinked genetic markers to detect population stratification in association studies. Am J Hum Genet 65:220–228

Pritchard JK, Wen W (2004) Documentation for structure software: version 2. http://pritch.bsd.uchicago.edu/software/readme_structure2_1.pdf

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

Rajora OH, Rahman MH, Buchert GP, Dancik BP (2000) Microsatellite DNA analysis of genetic effects of harvesting in old-growth eastern white pine (Pinus strobus) in Ontario, Canada. Mol Ecol 9:339–348

Rajora OP, Mann IK, Shi Y-Z (2005) Genetic diversity and population structure of boreal white spruce (Picea glauca) in pristine conifer-dominated and mixedwood forest stands. Can J Bot 83:1096–1105

Raymond M, Rousset F (1995a) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Heredity 86:248–249

Raymond M, Rousset F (1995b) An exact test for population differentiation. Evolution 49:1280–1283

Rosenberg NA, Burke T, Elo K, Feldman MW, Freidlin PJ, Groenen MAM, Hillel J, Maki-Tanila A, Tixier-Boichard M, Vignal A, Wimmers K, Weigend S (2001) Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. Genetics 159:699–713

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

Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol Ecol Res 8 (1):103–106

Ryynänen HJ, Tonteri A, Vasemägi A, Primmer CR (2007) A comparison of biallelic markers and microsatellites for the estimation of population and conservation genetic parameters in Atlantic salmon (Salmo salar). J Heredity 98:692–704

Slavov GT, Howe GT, Yakovlev I, Edwards KJ, Krutovskii KV, Tuskan GA, Carlson JE, Strauss SH, Adams WT (2004) Highly variable SSR markers in Douglas-fir: Mendelian inheritance and map locations. Theor Appl Genet 108:873–880

St. Clair JB (2006) Genetic variation in fall cold hardiness in coastal Douglas-fir in western Oregon and Washington. Can J Bot 84:1110–1121

St. Clair JB, Mandel NL, Vance-Borland KW (2005) Genecology of Douglas-fir in western Oregon and Washington. Ann Bot 96:1199–1214

Steiner KC (1979) Variation in bud-burst timing among populations of interior Douglas-fir. Silvae Genet 28:76–79

Strobeck C (1987) Average number of nucleotide differences in a sample from a single subpopulation: a test for population subdivision. Genetics 117(149):153

Thomas BR, MacDonald SE, Hick M, Adams DL, Hodgetts RB (1999) Effects of reforestation methods on genetic diversity of lodgepole pine: an assessment using microsatellite and randomly amplified polymorphic DNA markers. Theor Appl Genet 98:793–801

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D et al (2001) Dwarf8 polymorphisms associate with variation in flowering time. Nat Genet 28:286–289

Turakulov R, Easteal S (2003) Number of SNPS loci needed to detect population structure. Hum Hered 55:37–45

US DOI BLM (2005) Final Programmatic Environmental Impact Statement (PEIS) on wind energy development on Bureau of Land Management (BLM)-administered lands in the Western United States. Appendix F: Ecoregions of the 11 Western States and Distribution by Ecoregion of Wind Energy Resources on BLM-Administered Lands Within Each State. http://windeis.anl.gov/documents/fpeis/maintext/Vol2/appendices/appendix_f/Vol2AppF_1.pdf

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4:535–538 http://www.microchecker.hull.ac.uk/index.jsp

Viard F, El-Kassaby YA, Ritland K (2001) Diversity and genetic structure in populations of Pseudotsuga menziesii (Pinaceae) at chloroplast microsatellite loci. Genome 44:336–344

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

Wheeler NC, Jermstad KD, Krutovsky KV, Aitken SN, Howe GT, Krakowski J, Neale DB (2005) Mapping of quantitative trait loci

controlling adaptive traits in coastal Douglas-fir. IV. Cold-hardiness QTL verification and candidate gene mapping. Mol Breed 15:145–156

Wilson LM, Whitt SR, Ibanez AM, Rocheford TR, Goodman M, Buckler E (2004) Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell 16:2719–2733

Yu J, Buckler ES IV (2006) Genetic association mapping and genome organization of maize. Curr Opin Biotechnol 17:155–160

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208