4-1-2011

Low Levels of Population Genetic Structure in Pinus Contorta (Pinaceae) Across a Geographic Mosaic of Co-Evolution

T. L. Parchman

Craig Benkman
*University of Wyoming*, cbenkman@uwyo.edu

B. Jenkins

C. A. Buerkle

Follow this and additional works at: [https://repository.uwyo.edu/zoology_facpub](https://repository.uwyo.edu/zoology_facpub)

Part of the [Zoology Commons](https://repository.uwyo.edu/zoology_facpub)

Publication Information

LOW LEVELS OF POPULATION GENETIC STRUCTURE IN PINUS CONTORTA (PINACEAE) ACROSS A GEOGRAPHIC MOSAIC OF CO-EVOLUTION

THOMAS L. PARCHMAN, CRAIG W. BENKMAN, BRITTANY JENKINS, AND C. ALEX BUEKLE

The extent of gene flow among populations shapes the potential for independent population histories and adaptive evolution. Thus, knowledge of population genetic structure across the geographical range of a species provides information on the population context in which evolutionary processes operate. When combined with an understanding of neutral genetic differentiation, studies of molecular marker variation can thus improve our understanding of how adaptive geographic variation arises and is maintained in plant populations. Accounting for population structure is also a prerequisite for research on the genetic basis of phenotypic traits in natural populations. The geographic distribution of genetic variation affects the interpretation of traits that vary geographically by allowing an assessment of the extent to which divergence has been aided or impeded by gene flow. Similarly, knowledge of population genetic structure is used for genetic mapping in natural populations to distinguish true from spurious associations between genotype and phenotype (i.e., association genetics; Aranzana et al., 2005; Yu et al., 2006).

As a result of large population sizes, wind dispersal, and high levels of outcrossing, conifers typically harbor large amounts of genetic variation within populations and have low levels of genetic differentiation among populations (Hedrick, 2005). Despite this typical lack of neutral genetic differentiation among populations, many studies indicate that conifers are adapted to local environments and exhibit substantial geographic variation in adaptive traits (Morgenstern, 1996; Savolainen and Pyhajarvi, 2007a). Provenance trials demonstrate pronounced variation in adaptive phenotypes (Wheeler and Guries, 1982a; Xie and Ying, 1995), and geographic differentiation in morphology is often far greater than neutral genetic divergence (Wheeler, 1982b; Yang et al., 1996). High levels of phenotypic variation across populations that have limited neutral genetic differentiation indicate that conifers are readily capable of local adaptation, even in the face of extensive gene flow (Petit et al., 2004).

Lodgepole pine (Pinus contorta) is a dominant and widespread species occurring in the mountainous regions of western North America (Critchfield, 1980). Of its four recognized subspecies, the Rocky Mountain subspecies (P. contorta subsp. latifolia) has the most extensive range, growing throughout the Rocky Mountains and neighboring regions from the Yukon to Colorado (Fig. 1; Critchfield, 1980). Previous studies of Rocky Mountain lodgepole pine have highlighted substantial adaptive
mass (Smith, 1970; Benkman et al., 2001). However, in isolated ranges east and west of the Rocky Mountains where red squirrels are absent, lodgepole pine has evolved a higher ratio of seed mass to cone mass (Smith, 1970; Benkman et al., 2001). In addition, in these regions without squirrels, lodgepole pine has evolved larger cones with thicker distal scales in response to increases in seed predation and selection exerted by red crossbills (Loxia curvirostra) and crossbills have reciprocally evolved deeper bills (Benkman, 1999; Benkman et al., 2001, 2003). Divergence resulting from predator–prey arms races has occurred in a parallel
and replicated fashion in isolated ranges lacking squirrels to the east and west of the Rocky Mountains (CH and SH/AM in Fig. 1), leading to a geographic mosaic of coevolution (Thompson, 2005) that has driven geographic divergence in cone morphology (Benkman et al., 2001, 2003).

Populations of lodgepole pine across this region exhibit substantial geographic variation in the percentage of serotinous cones in a stand (Lotan, 1975; Tinker et al., 1994; Benkman and Siepielski, 2004). Individual trees largely produce either serotinous or nonserotinous cones (Teich, 1970; Lotan, 1975), and most stands of lodgepole pine contain a mix of such trees. Serotinous cones remain attached to branches and can hold viable seeds for decades (Critchfield, 1980), until fire heats and breaks resinous bonds between the cone scales and causes the release of seeds to the forest floor. Serotiny is favored by fire (Enright et al., 1998), and the percentage of serotinous trees within stands of several species of pines is positively correlated with the frequency of stand-replacing fires (Givnish, 1981; Muir and Lotan, 1985; Gauthier et al., 1996). However, predation by red squirrels reduces the accumulation of seeds in a canopy seed bank and thereby removes the benefit of serotiny (Benkman and Siepielski, 2004). The presence of red squirrels therefore should select against serotinous individuals. In areas with red squirrels, stands of lodgepole pine are generally less than 50% serotinous, whereas stands in ranges lacking squirrels are more than 90% serotinous (Fig. 1; Benkman and Siepielski, 2004). Thus, diverse selection pressures from the distribution of seed predators, fire, and perhaps other factors have given rise to substantial geographic variability in this trait.

An understanding of population genetic structuring across this region is needed to determine whether the phenotypic differentiation described has arisen despite extensive gene flow and whether populations to the east and west of the Rocky Mountains have evolved convergent phenotypes independently. Previous studies of population genetic structuring of lodgepole pine in the Rocky Mountain region have focused mostly on the part of the range north of Montana (Wheeler and Guries, 1982a; Marshall et al., 2002; Godbout et al., 2008) and have included only one of the squirrel-less populations described (Cypress Hills). Much of the current distribution of lodgepole pine is thought to have resulted from expansion and northward migration of populations over the last 12,000 years following the last glacial retreat (MacDonald and Cwynar, 1985, 1991; Marshall et al., 2002; Godbout et al., 2008). Reflecting this migrational history, results of studies of genetic variation in P. c. latifolia based on allozymes (Wheeler and Guries, 1982a; Dancik and Yeh, 1983; Yeh et al., 1985; Epperson and Allard, 1989), mtDNA (Dong and Wagner, 1994; Godbout et al., 2008), and cpDNA SSR loci (Marshall et al., 2002) have shown high diversity within populations but minor, if any, differentiation among populations. However, these studies have used individual organellar markers or small sets of allozymes and have focused on the northern part of the range (mostly in Canada), where occupancy is recent (last 12,000 years). The portion of the range south of Canada (and where geographic variation in cone traits is evident) is characterized by a more fragmented distribution (Fig. 1) that has been less altered by glacial activity than the northern range (Dong and Wagner, 1994). In addition, some populations along the eastern and western periphery of the geographic range are especially isolated from the central Rocky Mountains (Fig. 1) and undergo divergent selection from different seed predators (Benkman et al., 2001, 2003; Benkman and Siepielski, 2004), and the level of population structure characterizing this region is unknown.

Here we surveyed genetic variation at nine nuclear simple sequence repeat (SSR) loci and 235 AFLP markers across 22 lodgepole pine populations that span the central and southern portion of the Rocky Mountains, including isolated populations occurring to the east and west that have diverged in cone structure and serotiny frequency. We used this combination of marker types because nuclear SSRs are typically highly variable and informative and because we could readily assay large numbers of AFLP markers and thereby base analyses of population structure on a larger portion of the genome. A population genetic survey based on large numbers of polymorphic markers should provide a context for understanding the adaptive phenotypic divergence in cone traits. A second motivation for this research is that a thorough understanding of population structure is a prerequisite for association genetics, in which future studies will map variation in cone serotiny and defense traits to variable genetic loci.

MATERIALS AND METHODS

Genetic resources and molecular markers—We sampled needles from a minimum of 18 individual trees from each of 22 populations ranging from southern Wyoming to southern Alberta (Fig. 1). Dissolved needles (50 mg) served as a source of DNA, which was extracted using a modification of the cetrimidemethyl ammonium bromide (CTAB) method (Doyle, 1991). We quantified the amount of extracted DNA using a Nanodrop spectrophotometer (Thermo Scientific, Waltham, Massachusetts, USA) and verified the presence of high molecular weight genomic DNA by electrophoresis in 1.5% agarose gels and staining and visualization with ethidium bromide. We used the polymerase chain reaction (PCR) to amplify a set of nine previously developed nuclear SSR loci (Liewlaksaneeyanawin et al., 2004) and determined the genotypes at these loci for 423 individuals from the 22 populations (Table 1). Each reaction consisted of 50–100 ng total genomic DNA, 2 pmol of each primer, 0.5 mM each of dATP, dCTP, dGTP, and dTTP, 1 µl of Taq polymerase. All PCR amplifications were performed with the following conditions: 94°C for 5 min, followed by 32 cycles of 94°C for 1 min, annealing temperature for 1 min (which varied, see Liewlaksaneeyanawin et al., 2004), and 72°C for 1 min, followed by a final extension step of 72°C for 3 min. Forward primers were labeled with various fluorescent dyes, and fragments were detected via capillary electrophoresis using an ABI 3130 genetic analyzer (ABI, Foster City, California, USA). We assessed the presence of fragments representing different alleles using the program Genemapper (ABI).

We generated AFLP markers based on a modified version of the Vos et al. (1995) protocol. Preselective and selective primers were modified and made more specific to accommodate the large genome size of lodgepole pine, following Remington et al. (1999). Restriction digestion and adaptor-ligation were carried out simultaneously on 0.5 µg of genomic DNA using the restriction

<table>
<thead>
<tr>
<th>Locus</th>
<th>H_o</th>
<th>H_e</th>
<th>No. of alleles</th>
<th>Average allele size</th>
<th>Maximum allele size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lop1</td>
<td>0.331</td>
<td>0.344</td>
<td>14</td>
<td>156</td>
<td>174</td>
</tr>
<tr>
<td>Lop5</td>
<td>0.91</td>
<td>0.838</td>
<td>28</td>
<td>180</td>
<td>211</td>
</tr>
<tr>
<td>PtTX2123</td>
<td>0.57</td>
<td>0.575</td>
<td>6</td>
<td>200</td>
<td>206</td>
</tr>
<tr>
<td>PtTX3011</td>
<td>0.956</td>
<td>0.72</td>
<td>48</td>
<td>182</td>
<td>226</td>
</tr>
<tr>
<td>PtTX3034</td>
<td>0.838</td>
<td>0.637</td>
<td>20</td>
<td>209</td>
<td>233</td>
</tr>
<tr>
<td>PtTX3049</td>
<td>0.901</td>
<td>0.604</td>
<td>30</td>
<td>315</td>
<td>340</td>
</tr>
<tr>
<td>PtTX3052</td>
<td>0.598</td>
<td>0.592</td>
<td>14</td>
<td>244</td>
<td>267</td>
</tr>
<tr>
<td>PtTX3107</td>
<td>0.797</td>
<td>0.509</td>
<td>21</td>
<td>167</td>
<td>199</td>
</tr>
<tr>
<td>PtTX3127</td>
<td>0.651</td>
<td>0.604</td>
<td>13</td>
<td>178</td>
<td>199</td>
</tr>
</tbody>
</table>

Note: H_o = observed heterozygosity
endonucleases EcoRI and Msel (NEB, Ipswich, Massachusetts). AFLP adaptor pairs were attached to digested fragments using T4 DNA ligase (NEB). We performed restriction and ligation reactions simultaneously in 11 µL volumes that were incubated for 18 h at 35°C. After incubation, we diluted these reactions with 170 µL 0.1X Tris-EDTA (TE) buffer. Preselective and selective primers were based on primer core sequences EcoRI 5’-GACTGGCTACCA-ATTCC-3’ and Msel 5’-GATGAGTCCGAGTAA-3’ (EcoRI and Msel hereafter). Preselective amplification reactions contained 4 µL of the diluted restriction-ligation products, 15 µL PCR core mix (385 µL H2O, 68 µL 10× PCR buffer, 41 µL 25 mM MgCl2, and 6.8 µL 40 mM dNTPs for a total volume of 500 µL), and 1 µL preselective primers, which each consisted of the adaptor primer sequences with one additional nucleotide at the 3′ ends (EcoRI-A and Msel-C). Preselective PCR conditions included 20 PCR cycles (94°C for 30 s, 56°C for 1 min, 72°C for 2 min), and a final extension at 60°C for 30 min. We diluted preselective amplification products with 170 µL 0.1X Tris-EDTA (TE). Selective amplification reactions contained 3 µL of diluted preselective amplification product, 15 µL standard PCR core mix, 1 µL of selective Msel primer, and 1 µL of the fluorescently labeled EcoRI selective primer. Both EcoRI and Msel selective amplification primers had three extra nucleotides at the 3′ ends to produce an appropriate number of amplified fragments.

We used three selective primer combinations (EcoRI-ACT and Msel-CTA; EcoRI-AAC and Msel-CCT; EcoRI-AAT and Msel-CAT) to generate AFLP fragments. We combined 1 µL of each selective amplification product with 8.75 µL formamide and 0.45 µL GeneScan 500 ROX labeled size standard (ABI) and resolved products on an ABI 3730 capillary sequencer. The presence or absence of AFLP fragments in each lane file was analyzed using the program Genemapper (ABI). We scored fragments generated by each selective primer combination by band intensity (i.e., presence or absence). We limited analyses to fragment sizes between 70 and 350 bp, only scored unambiguously discernible fragments, and genotyped a total of 536 individuals spanning the 22 populations (compared to 423 individuals with SSR data).

**Data analysis**—For the SSR loci, we calculated descriptive statistics for each population using microsatellite analyzer (MSA version 4.05; Dieringer and Schlötterer, 2003), which included observed and expected heterozygosities (Nei, 1987) and the mean number of alleles per locus. Prior to further analyses, we used the program Genepop (Raymond and Rousset, 1995) to check for violations of Hardy–Weinberg equilibrium for each SSR locus across all populations. To quantify genetic differentiation among populations based on the nine SSR loci, we used the program SMOGD (Crawford, 2010) to calculate GST, GSTt and Jost’s D for each population, and used MSA to calculate θ as in Weir and Cockerham (1984). GSTt and Jost’s D are not dependent on heterozygosity and are potentially more appropriate measures of genetic differentiation given loci with high diversity and numerous allelic states (Hedrick, 2005; Jost, 2008). We calculated basic descriptive statistics for the AFLP data, including heterozygosity and the proportion of polymorphic loci, for each population using the multiallelic SSR data (Hedrick, 2005; Jost, 2008).

We used the program STRUCTURE (version 2.2; Pritchard et al., 2000; Falush et al., 2003, 2007) to assess whether the sampled genotypes were consistent with a single or multiple (K) populations, each in Hardy–Weinberg equilibrium. For these analyses, we used the combined data set of nine SSR and 235 AFLP loci. Log-likelihoods from Markov chain Monte Carlo (MCMC) sampling provide the basis for evaluating the number of clusters that best fit the data. We also used STRUCTURE to assign genotypic proportions of each individual to the K clusters based on admixture coefficients. We used the admixture model and ran simulations for 12 replicates for each value of K ranging from one to 15. We used a burn-in of 50000 steps and ran the MCMC for 150000 steps for each round of simulations. The number of populations (K) that was most likely given the data were determined using the method of Pritchard et al. (2000). We also calculated the change in model likelihoods between successive values of K (ΔK) and used the method of Evanno et al. (2005) to infer K.

We obtained 1000 bootstrapped matrices of estimates of θ and K for pairs of populations for the SSR data using the program FSTAT (Goudet, 1995). Similarly, we created 1000 bootstrapped matrices of θ and K for the AFLP data using AFLP SURV. Neighbor-joining trees were then constructed separately for the SSR and AFLP data sets based on the 1000 bootstrapped distance matrices using the neighbor routine in the program PHYLIP (version 3.6; Felsenstein, 2005). We used PHYLIP to construct consensus trees based on the 1000 trees generated as described, and the number of trees out of 1000 sharing nodes were used as bootstrap support values. We used correlations and Mantel tests to examine the relationship between genetic differentiation and geographic distance for both the SSR and AFLP data. Geographic coordinates of all populations were recorded and pairwise geographic distances between all sets of populations were obtained using the point-distance tool in the program ArcMap version 9.3 (RockWare, Golden, Colorado, USA). We compared geographic distances to estimates of θ(1 – θ) for the SSR data, and θFST/(1 – FST) for the AFLP data because these adjusted values are expected to vary linearly with geographic distances and are more appropriate for comparison (Rousset, 1997).

**RESULTS**

All nine of the nuclear SSR loci were polymorphic and highly heterozygous, with the number of alleles per locus ranging from 6 to 48 (Table 1). Pooled across all populations, some loci deviated from Hardy–Weinberg equilibrium; however, such departures were uncommon for individual SSR loci within populations and no tests were globally significant across individual populations. The mean number of alleles per locus was similar across populations, with no populations having significantly lower allelic diversity than the average (Table 2). Of the 235 AFLP loci we scored, 96% were also polymorphic. As indicated by estimates of heterozygosity, genetic diversity within populations was generally high and similar across populations for both the SSR and AFLP data sets (Table 2). As expected, per locus heterozygosity estimates were higher for the multiallelic SSR data (ranging from 0.543 to 0.708) than for the dominant, biallelic AFLP data (ranging from 0.264 to 0.306; Table 2).

Clustering analyses run in STRUCTURE based on the combined SSR and AFLP data sets found the highest log-likelihood support for three genotypic clusters in the population genetic data (K = 3; Fig. 2A), and the ΔK method of Evanno et al. (2005) resulted in support for the same value of K (Fig. 2B). Nonetheless, overall FST among the clusters in the K = 3 model was very low (0.013), and assignment of individuals to these three genotypic clusters was not consistent with population or geographic origin. Nearly all individuals had partial assignments to all three clusters, and the credible intervals on assignments (admixture coefficients) were very large (i.e., many ranged from 0 to 1). These analyses suggest an absence of pronounced population structure.

Estimates of genetic differentiation among populations were low for both the SSR and AFLP data sets. However, differences in the nature of these markers led to different information content and values of differentiation estimates, and there was substantial variation among loci and populations in the level of differentiation (Figs. 3, 4). Population pairwise and overall estimates of genetic differentiation for the SSR data were low, and many were statistically indistinguishable from zero (θ, mean = 0.016, range = 0.00–0.058; GSTt, mean = 0.111, range = 0.01–0.29; and Jost’s D, mean = 0.104, range = 0.01–0.26; Appendix S1, see Supplemental Data with online version of this article). Estimates of GSTt and Jost’s D were considerably higher than θ and GSTt, perhaps suggesting that estimates based on heterozygosity (θ and GSTt) do not adequately measure differentiation for highly diverse SSR loci (Hedrick, 2005; Jost, 2008). Nonetheless, the two standardized estimates of genetic differentiation among populations for the SSR data (GSTt and Jost’s D) were strongly correlated with θ and result in similar relative rankings for population differentiation. Analyses of the AFLP
data based on the method of Lynch and Milligan (1994) yielded an overall $F_{ST}$ of 0.01, and the Bayesian method of Holsinger et al. (2002) led to $F_{ST}$ of 0.02. Pairwise $F_{ST}$ estimates from the AFLP data for the 22 populations were low, ranging from 0 to 0.07, but many were statistically distinguishable from zero (online Appendix S2). Although overall estimates of genetic differentiation were low, estimates for individual loci varied substantially for both the SSR and AFLP data sets, with some loci being characterized by far higher levels of genetic differentiation (Fig. 4). A wide range of differentiation estimates also characterized the pairwise comparisons between populations (Fig. 3).

Despite the low average level of genetic differentiation, there was a tendency for geographically proximate populations to be less differentiated than were more distant populations, consistent with a pattern of isolation by distance (Fig. 5). Geographic and genetic distances were correlated for both the SSR ($\rho = 0.28$; Mantel test, $P < 0.001$) and AFLP data sets ($\rho = 0.31$, Mantel test, $P < 0.001$). In addition, neighbor-joining dendrograms constructed from pairwise estimates of Nei’s $D$ between populations generally clustered geographically proximate populations together (Fig. 6). Pairwise comparisons of populations at the opposite extremes of the sampled geographic distribution exhibited among the highest differentiation estimates (e.g., CH vs. SH, CH vs. SU; Fig. 5). For example, the average differentiation was higher for comparisons between isolated populations at the northeastern (CH, BP, LR, SG) and southwestern margins of the range (SH, AM, DC, SB) (mean $G_{ST}' = 0.17$).

Table 2. The number of individuals genotyped in each population for nine SSR loci ($n$), expected ($H_e$) and observed heterozygosities ($H_o$), and the mean number of SSR alleles per locus. Following is the number of individuals genotyped across 235 AFLP loci ($N$), expected heterozygosity, and the percentage of polymorphic loci (PLP). Abbreviations after location names correspond to the labels in Fig. 1.

<table>
<thead>
<tr>
<th>Population, State/Province (Abbreviation)</th>
<th>SSR Mean no. alleles</th>
<th>AFLP $N$</th>
<th>$H_e$</th>
<th>$H_o$</th>
<th>Mean no. alleles</th>
<th>$n$</th>
<th>$H_e$</th>
<th>PLP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absaroka Mts., WY (AB)</td>
<td>21</td>
<td>0.702</td>
<td>0.616</td>
<td>8.78</td>
<td>42</td>
<td>0.281</td>
<td>84.3</td>
<td></td>
</tr>
<tr>
<td>Albion Mts., ID (AM)</td>
<td>14</td>
<td>0.709</td>
<td>0.648</td>
<td>6.67</td>
<td>36</td>
<td>0.268</td>
<td>83</td>
<td></td>
</tr>
<tr>
<td>Bears Paw Mts., MT (BP)</td>
<td>24</td>
<td>0.723</td>
<td>0.638</td>
<td>11.33</td>
<td>24</td>
<td>0.288</td>
<td>87.7</td>
<td></td>
</tr>
<tr>
<td>Big Horn Mts., WY (BH)</td>
<td>14</td>
<td>0.712</td>
<td>0.618</td>
<td>7</td>
<td>19</td>
<td>0.277</td>
<td>82.1</td>
<td></td>
</tr>
<tr>
<td>Crow’s Nest Pass, AB (CN)</td>
<td>21</td>
<td>0.716</td>
<td>0.566</td>
<td>9.56</td>
<td>25</td>
<td>0.28</td>
<td>82.1</td>
<td></td>
</tr>
<tr>
<td>Cypress Hills, AB (CH)</td>
<td>20</td>
<td>0.746</td>
<td>0.609</td>
<td>9.11</td>
<td>20</td>
<td>0.285</td>
<td>87.2</td>
<td></td>
</tr>
<tr>
<td>Deep Creek Range, ID (DC)</td>
<td>23</td>
<td>0.695</td>
<td>0.624</td>
<td>7.78</td>
<td>22</td>
<td>0.265</td>
<td>84.3</td>
<td></td>
</tr>
<tr>
<td>East Glacier, MT (EG)</td>
<td>19</td>
<td>0.746</td>
<td>0.652</td>
<td>8.56</td>
<td>19</td>
<td>0.294</td>
<td>85.5</td>
<td></td>
</tr>
<tr>
<td>Highwood Mts., MT (HI)</td>
<td>21</td>
<td>0.732</td>
<td>0.622</td>
<td>9.11</td>
<td>19</td>
<td>0.3</td>
<td>85.1</td>
<td></td>
</tr>
<tr>
<td>Judith Mts., MT (JU)</td>
<td>20</td>
<td>0.744</td>
<td>0.577</td>
<td>8.89</td>
<td>20</td>
<td>0.306</td>
<td>88.5</td>
<td></td>
</tr>
<tr>
<td>Little Belt Mts., MT (LB)</td>
<td>25</td>
<td>0.731</td>
<td>0.6</td>
<td>9</td>
<td>17</td>
<td>0.305</td>
<td>85.5</td>
<td></td>
</tr>
<tr>
<td>Little Rocky Mts., MT (LR)</td>
<td>21</td>
<td>0.74</td>
<td>0.655</td>
<td>9</td>
<td>24</td>
<td>0.277</td>
<td>89.4</td>
<td></td>
</tr>
<tr>
<td>Pine Creek Pass, ID (PC)</td>
<td>17</td>
<td>0.706</td>
<td>0.543</td>
<td>8.11</td>
<td>17</td>
<td>0.293</td>
<td>83</td>
<td></td>
</tr>
<tr>
<td>Roger’s Pass, MT (RP)</td>
<td>16</td>
<td>0.753</td>
<td>0.708</td>
<td>9.11</td>
<td>23</td>
<td>0.291</td>
<td>87.7</td>
<td></td>
</tr>
<tr>
<td>Snowy Range, WY (SR)</td>
<td>16</td>
<td>0.707</td>
<td>0.594</td>
<td>7.33</td>
<td>17</td>
<td>0.282</td>
<td>81.3</td>
<td></td>
</tr>
<tr>
<td>South Hills, ID (SH)</td>
<td>24</td>
<td>0.652</td>
<td>0.57</td>
<td>8.11</td>
<td>56</td>
<td>0.264</td>
<td>81.7</td>
<td></td>
</tr>
<tr>
<td>Sublett Range, ID (SU)</td>
<td>21</td>
<td>0.668</td>
<td>0.565</td>
<td>7.44</td>
<td>29</td>
<td>0.264</td>
<td>79.6</td>
<td></td>
</tr>
<tr>
<td>Sweetgrass Hills, MT (SG)</td>
<td>10</td>
<td>0.685</td>
<td>0.607</td>
<td>6.56</td>
<td>9</td>
<td>0.303</td>
<td>84.7</td>
<td></td>
</tr>
<tr>
<td>Wasatch Mts., UT (WA)</td>
<td>18</td>
<td>0.695</td>
<td>0.577</td>
<td>8</td>
<td>19</td>
<td>0.288</td>
<td>85.5</td>
<td></td>
</tr>
<tr>
<td>West Yellowstone, WY (WY)</td>
<td>19</td>
<td>0.754</td>
<td>0.631</td>
<td>9.44</td>
<td>18</td>
<td>0.269</td>
<td>77.4</td>
<td></td>
</tr>
<tr>
<td>Wind River Mts., WY (WR)</td>
<td>20</td>
<td>0.719</td>
<td>0.654</td>
<td>8.56</td>
<td>43</td>
<td>0.267</td>
<td>82.1</td>
<td></td>
</tr>
<tr>
<td>Vedauwoo, WY (VE)</td>
<td>19</td>
<td>0.671</td>
<td>0.355</td>
<td>8</td>
<td>19</td>
<td>0.29</td>
<td>83</td>
<td></td>
</tr>
</tbody>
</table>

Notes: AB = Alberta, Canada; ID = Idaho, MT = Montana, UT = Utah, WY = Wyoming, USA.
light of substantial phenotypic divergence in cone traits and its implications for future research aimed at detecting the genetic architecture of such traits.

Population genetic structure in lodgepole pine—The SSR and AFLP data sets indicate high levels of genetic diversity within and subtle levels of differentiation between lodgepole populations in our study region. Overall estimates of population genetic differentiation for the AFLPs and the SSRs were low ($F_{ST} = 0.01-0.02$, $\theta = 0.016$), indicating that the majority of genetic variation is partitioned within populations. Despite utilizing a large number of polymorphic markers, analyses conducted using the Bayesian clustering method of Pritchard et al. (2000, with extensions from Falush et al., 2003) similarly produced little evidence of clear or pronounced population
structure (Fig. 2). Given the average low levels of differentiation estimates ($F_{ST} < 0.02$), it is not surprising that STRUCTURE did not detect evidence of true genotypic clustering, as analyses of simulated data sets failed to detect the true $K$ when overall $F_{ST}$ was less than 0.02 (Latch et al., 2006). These results are similar to those documented in previous population genetic surveys of lodgepole pine based on the northern part of the distribution and smaller sets of markers (Wheeler and Guries, 1982a; Yeh et al., 1985; Dong and Wagner, 1994; Marshall et al., 2002) and are consistent with other population genetic surveys of conifers, where per locus estimates of $F_{ST}$ are typically less than 0.1 (Hamrick et al., 1992). High levels of outcrossing, wind dispersal, and large population sizes predispose these populations to harbor large amounts of genetic diversity and to exhibit low levels of geographic differentiation and structure.

**Fig. 5.** Plots of geographic and genetic distances between all pairs of populations sampled and the lines of best fit. Genetic differentiation for the SSR data are based on $\theta$ and differentiation for the AFLP data are based on $F_{ST}$. Several pairwise comparisons exhibiting noteworthy patterns of variation are labeled with abbreviations that match those in Fig. 1. Populations indicated in bold face are relatively isolated and occur at the periphery of the distribution.

**Fig. 6.** Neighbor-joining trees based on pairwise estimates of Nei’s (1972) genetic distance between all pairs of 22 populations for the (A) nine SSR loci, and (B) 235 AFLP loci. Values at the branches indicate the percentage of trees constructed from 1000 bootstrapped distance matrices that shared that node. Isolated populations that have diverged in cone morphology from populations in the central part of the range are indicated in bold, italicized font. Those west of the Rocky Mountains are also underlined.
Despite the lack of pronounced range-wide population structure, there was substantial variation in estimated differentiation among populations (Fig. 3) and among individual loci (Fig. 4) and marker types (Fig. 3). Pairwise population differentiation estimates spanned a wide range (Fig. 3), and those comparisons exhibiting higher levels of differentiation often involved peripheral populations (Fig. 5). Estimates of genetic differentiation based on heterozygosity ($\theta$ and $F_{ST}$) underestimate levels of differentiation for markers such as SSRs that have high allelic diversity (Hedrick, 2005; Jost, 2008). Our estimates of differentiation for the SSR data based on Jost’s $D$ and $G_{ST}$ are substantially higher than those based on $\theta$, suggesting these corrected estimates reveal higher levels of differentiation than those based on heterozygosity ($\theta$ and $F_{ST}$). With highly diverse SSR loci, the presence of rare alleles at low frequencies in different populations could contribute to this pattern. Because our sample sizes per population are small relative to the number of different alleles in populations, our estimates of Jost’s $D$ and $G_{ST}$ could be biased upwards. The strong correlation and similar rankings of Jost’s $D$ and $G_{ST}$ with $\theta$ suggest similar interpretations despite differences in the absolute numerical values of these estimates. Nonetheless, the nine SSR loci do give rise to slightly higher estimates of differentiation (based on $\theta$) than the 235 AFLP markers ($F_{ST}$; Fig. 3). Overall, the differentiation estimates and the large number of pairwise comparisons that were statistically distinguishable from zero indicate that there is subtle genetic divergence across this geographic region.

Although the average estimates of differentiation across all loci are low, some loci exhibit higher levels of differentiation (Fig. 4), which could reflect the influence of natural selection on these or closely linked loci. A common approach for detecting evidence of selection in such data sets involves simulating neutral distributions of $F_{ST}$ to detect markers that have higher than expected $F_{ST}$ values (Beaumont and Balding, 2004; Beaumont, 2005). Comparison of neutral distributions of $F_{ST}$ (based on simulations of the coalescent) with estimates for the SSR and AFLP loci (conducted with the programs FDist and DFDist; Beaumont and Balding, 2004) indicated that a small number ($<3\%$ for all comparisons) of these markers fall outside of the simulated neutral distribution and thus could be influenced by selection. Because the loci detected as possibly experiencing selection were different for separate independent pairwise population comparisons and exhibited no relationship with geographic or phenotypic patterns, we do not present these results in detail here. In addition, we have concerns about the utility of AFLPs for such single-locus analyses, due to their dominant nature, the possibility of AFLP products arising from endosymbiotic DNA (e.g., Gompert et al., 2010), and uncertainty about their molecular evolution (what proportion of variation corresponds to unrecognized length variation rather than the presence or absence of products?). Nonetheless, the large variance in differentiation estimates highlights that genetic differentiation is heterogeneous across the genome (Fig. 4), and suggests that genome scans based on higher quality marker or sequence data (SSRs or haplotypes) should be effective in detecting genetic regions involved in adaptive evolution (Beaumont, 2005; Nosil et al., 2009).

Pairwise geographic and genetic distances between populations for both sets of markers were correlated, indicating a pattern of isolation by distance (Fig. 5). This pattern was also evident in neighbor-joining phenograms where geographically proximate populations typically clustered together (Fig. 6). A similar relationship between genetic and geographic distances has been documented in previous research on lodgepole pine (Marshall et al., 2002) and other conifers in western North America (Krutovsky et al., 2009) and is likely influenced by a recent northward migration and expansion occurring over the past 10,000–12,000 years following the last glacial retreat (Wheeler and Guries, 1982a; MacDonald and Cwynar, 1991; Godbout et al., 2008). Although populations have not been isolated for long enough to accumulate substantial divergence at neutral markers, divergence, and expansion has apparently resulted in an association between geographical and genetic distance.

Whereas previous population genetic surveys of lodgepole pine focused mostly on the northern part of the range (Wheeler and Guries, 1982a; Yeh et al., 1985; Dong and Wagner, 1994; Marshall et al., 2002; Godbout et al., 2008), we sampled a portion of the range where lodgepole pine has been present for longer time periods and is characterized by a more fragmented distribution (Fig. 1). During the last glacial maximum, ice covered much of the current distribution north of the Canadian border, but lodgepole pine was present to the south in areas of Idaho, Montana, and Wyoming (Baker, 1976; Mehringer et al., 1977; Beiswenger, 1991). Perhaps surprisingly, the level of genetic differentiation among the populations in our study is not more pronounced than to the north. Nonetheless, the isolated peripheral populations have slightly higher levels of genetic differentiation from other populations than the populations in the core of the range (Fig. 5). Many of these peripheral populations are isolated from the central Rocky Mountain region by large expanses of grassland or sagebrush (Fig. 1, Benkman, 1999; Edelaar and Benkman, 2006), which likely reduces gene flow. Previous studies have not included these populations, with the exception of the Cypress Hills, which have higher genetic divergence than other populations in the Rocky Mountain region based on mtDNA (Godbout et al., 2008) and allozymes (Wheeler and Guries, 1982a; Dancik and Yeh, 1983). During a warming period that ended approximately 7000 years ago, lodgepole pine likely had a wider distribution that extended to the Cypress Hills and the isolated ranges in northern Montana, after which time these forests became isolated by large expanses of grassland (Anderson et al., 1989). In southeastern Idaho, lodgepole pine is known to have been in the Albion Mountains for at least 12,000 years (Davis et al., 1986). These forests may have been connected to those in the Rocky Mountains as recently as 10,000 years ago (Wells, 1983; Beiswenger, 1991) although they have been isolated since. That genetic differentiation of these populations is not greater is potentially the result of the relatively recent isolation of lodgepole pine in these mountain ranges rather than high levels of contemporary gene flow.

**The population genetic context for adaptive variation within Pinus contorta**—Together with previous studies of morphological variation, our results indicate that the substantial phenotypic divergence in cone traits across these populations (Benkman, 1999; Benkman et al., 2001, 2003; Benkman and Siepielski, 2004; Edelaar and Benkman, 2006) has occurred within a background of limited neutral genetic divergence. This suggests that adaptive divergence has arisen despite recent shared ancestry and has been relatively unimpeded by gene flow and agrees with other studies examining the relationship between phenotypic variation and genetic structure in lodgepole pine (Wheeler, 1982b; Xie and Ying, 1995; Yang et al., 1996). The ability of forest trees in general to evolve rapidly in response to changing biotic and abiotic conditions has been widely recognized (Petit et al., 2004; Savolainen and Pyhäjärvi, 2007a;
Savolainen et al., 2007b) and is likely enabled by large amounts of genetic variation resulting from high levels of outcrossing, high levels of gene flow, and large effective population sizes.

Despite subtle genetic differentiation among populations, there are patterns in these data that further bear on our understanding of geographic variation in cone traits. In particular, populations to the east and west of the Rocky Mountains that exhibit parallel divergence in cone traits are more genetically similar to populations in the middle of the range than they are to each other (Fig. 6). Trees in these populations, including the South Hills, Idaho and the Cypress Hills, Alberta have evolved larger cones with thicker distal scales as a result of coevolution with crossbills (Benkman, 1999; Benkman et al., 2001, 2003). These populations also exhibit the most extreme levels of cone serotiny, where the frequency of trees with serotinous cones approaches 100% (Fig. 1). This is in stark contrast to populations in the central part of the range, where the frequency of serotiny is consistently less than 50% (Benkman and Siepielski, 2004). Phenotypic plasticity is unlikely to contribute substantially to the divergence in cone traits among populations because cone traits in pines generally have highheritabilities (broad sense heritabilities of 0.6–0.8, Matziris, 1998; Sivacioglu and Ayan, 2010) and serotiny is thought to be under strong genetic control (Teich, 1970; Critchfield, 1980). Patterns of genetic variation presented here along with the likely biogeographical history of lodgepole pine in this region suggest that cone traits diverged in a parallel and independent fashion in the isolated ranges to the east and west of the Rocky Mountains (Benkman, 1999; Benkman et al., 2001, 2003).

Finally, an understanding of population structure and geographic variation in phenotypic traits is crucial for investigating the genetic control and architecture of traits in natural populations (González-Martínez et al., 2006; Zhao et al., 2007). If not accounted for, population structure can lead to spurious associations between genetic and phenotypic variation, particularly when populations differ in genotypic and phenotypic composition (Aranzana et al., 2005; Yu et al., 2006). The subtle levels of population differentiation across our study region mean that markers associated with phenotypic variation should be distinguishable from low levels of background neutral variation. Nonetheless, knowledge of this subtle genetic structure should still provide important information for accounting for population structure in association studies. Recent transcriptome sequencing for lodgepole pine (Parchman et al., 2010) and high-throughput sequencing assays (e.g., Hohenlohe et al., 2010) will facilitate future association and population genomic approaches involving these populations. The population genetic analyses presented here will provide a source of information for the planning and execution of such studies.

LITERATURE CITED


Felsenstein, J. 2005. PHYLIP (phylogeny inference package), version 3.6. Distributed by the author, Department of Genome Sciences, University of Washington, Seattle, Washington, USA.


Hedrick, P. 2005. Genetics of populations, 3rd ed. Jones and Bartlett, Sudbury, Massachusetts, USA.


Vekemans, X. 2002. AFLP–SURV version 1.0. Distributed by the author, Laboratoire de Génétique et Ecologie Végétales, Université Libre de Bruxelles, Brussels, Belgium.


