# CONGRUENT CLIMATE-RELATED GENECOLOGICAL RESPONSES FROM MOLECULAR MARKERS AND QUANTITATIVE TRAITS FOR WESTERN WHITE PINE (*PINUS MONTICOLA*)

Bryce A. Richardson,[1],* Gerald E. Rehfeldt,† and Mee-Sook Kim‡

*USDA Forest Service, Rocky Mountain Research Station, Provo, Utah 84606, U.S.A.; †USDA Forest Service, Rocky Mountain Research Station, Moscow, Idaho 83843, U.S.A.; and ‡Department of Forest Resources, Kookmin University, Seoul, Korea 136-702

Analyses of molecular and quantitative genetic data demonstrate the existence of congruent climate-related patterns in western white pine (*Pinus monticola*). Two independent studies allowed comparisons of amplified fragment length polymorphism (AFLP) markers with quantitative variation in adaptive traits. Principal component analyses were conducted on seedling traits in common gardens collected from 58 sites; principal coordinate analyses were conducted on AFLP data from 15 sites. Collection site eigenvector means were regressed on 35 climate variables to produce a genecological prediction for each data set. Both predictions explained a large percentage (>70%) of the genetic variation from the first eigenvector from similar predictors involving the interaction growing season precipitation and temperature. Both predictions were significant ($P < 0.01$), and projected genecological maps were largely congruent. The predominant pattern from the first eigenvector was two major groups, a northern and a southern, connected by a steep cline across the crest of the southern Cascade Range of Oregon. This cline was associated with growth potential in the adaptive traits. These results suggest that divergent climatic selection has influenced phenotypic traits associated with growth. Furthermore, the congruence of AFLPs suggests a linkage between some loci and genomic regions under selection and/or climatic influences on postmating reproductive success.

*Keywords:* adaptive traits, AFLP, divergent selection, ecological genetics, genome scan.

*Online enhancements:* appendix figures.

## Introduction

Widely distributed forest tree species are subjected to an array of environmental stresses that shape adaptation. Climatic adaptation was first demonstrated in forest trees by Langlet (1936), who showed that populations of forest trees exhibit clinal variation in growth and survival traits that parallel climate gradients. Since then, quantitative genecology—associating genetic responses to environment by measuring quantitative traits in common gardens—has been demonstrated in numerous broadly distributed plant species and is critical to understanding climatic adaptation (reviewed in Aitken et al. 2008).

Comparisons between adaptive traits and molecular markers have typically shown contrasting genetic patterns in forest trees (Karhu et al. 1996; Jaramillo-Correa et al. 2001; García-Gil et al. 2003) as well as other organisms (Reed and Frankham 2001). Because molecular markers, such as amplified fragment length polymorphisms (AFLPs) and microsatellites, are generally driven by nonselective forces associated with demographic processes, such disparities are not surprising. These neutral characteristics have made molecular markers valuable for inferring biogeographic and demographic processes caused by the impacts of glaciation and subsequent

postglacial colonization in forest tree species (Richardson et al. 2002; Petit et al. 2003; Godbout et al. 2005). However, increased resolution of molecular markers throughout the genome and new analytical approaches offer researchers better opportunities to identify loci with a putative signal of selection from the background of neutral variation shaped by demographic processes (reviewed in Storz 2005; Vasemagi and Primmer 2005; Biswas and Akey 2006).

Forest trees are amenable for using molecular markers to elucidate adaptive variation, largely because no domestication has occurred in many tree species and because numerous quantitative genetic studies of tree species have been performed (Neale and Savolainen 2004; González-Martínez et al. 2006). Furthermore, genetic variation in adaptive traits can be predicted from climate (Rehfeldt et al. 2001; St. Clair et al. 2005). These climate variables can be estimated from climate modeling at the spatial resolution and extent required for genetic analyses (McKenney et al. 2001; Hijmans et al. 2005; Rehfeldt 2006). Together, these approaches allow spatially explicit, climate-based predictions of adaptive genetic variability (Rehfeldt 2004; St. Clair et al. 2005).

Western white pine (*Pinus monticola* Dougl. ex. D. Don) is widely distributed over heterogeneous environments of western North America. In the Rocky Mountains, western white pine is a dominant seral species of mesic, montane forests. To the west in the Cascade Range, it is a minor, seral component

---

[1] Author for correspondence; e-mail: brichardson02@fs.fed.us.

of forests ranging from low elevation to subalpine. In these mountain ranges, western white pine is known for its rapid growth, which made it a desirable timber species. In the southern Cascade Range, Siskiyou Mountains, and Sierra Nevada, it is typically a minor component of upper montane and subalpine forests; however, in the Siskiyous, it also occupies the harsh sites on serpentine soils at lower elevations (Franklin and Dyrness 1973; Griffin and Critchfield 1972).

Genecology of western white pine is well documented with quantitative traits (Rehfeldt et al. 1984; Campbell and Sugano 1989; Meagher and Hunt 1998). This species has been described as a generalist with weak patterns of quantitative trait variation among seed sources from a broad range of elevations (Rehfeldt et al. 1984). However, two distinct geographical patterns in quantitative traits have been noted: seed sources north of 45°N latitude had high growth potential and low tolerance to cold, those sources south of 41°N had low growth potential and high cold tolerance, while those from a transitional region exhibited a steep cline (Rehfeldt et al. 1984). A subsequent common garden study focusing on this transitional area and the northern Cascade Range found similar geographic patterns, but growth potential patterns were apparently influenced by the crest of the Cascade Range in southern Oregon. Seed sources from west of the crest exhibited a high growth potential, while those from the east had a low growth potential (Campbell and Sugano 1989).

Our study uses climate variables in genecological models to compare molecular and quantitative genetic patterns. We use two independently sampled genetic data sets from western white pine, one of adaptive traits (Rehfeldt et al. 1984) and another of AFLP markers. The aim of this study is to elucidate climate-based genetic variation by the (1) development of genecological models from principal components and principal coordinates analyses that describe climate-related genetic variation across the species distribution, (2) identification of climate variables important in shaping western white pine genetic variation, (3) comparison of molecular and quantitative genecological predictions, and (4) evaluation of AFLP loci for a putative selective signal by using a genome scan followed by regression of locus frequencies with the genecological prediction. Finally, we discuss two hypotheses for the observed patterns in AFLPs that are directly or indirectly influenced by natural selection.

## Material and Methods

### Molecular Genetic Data

Bud and needle tissues were collected from typically old growth trees (i.e., >75 yr) in each of 15 natural collection sites of western white pine (fig. 1). Between 18 and 29 samples were analyzed per collection site. The fresh tissue was used to extract DNA using Qiagen plant DNA extraction kits. AFLP markers were generated following Kim et al. (2003). Briefly, EcoRI and MseI restriction enzymes were used to digest 500 ng of genomic DNA followed by ligation with adaptors and preamplification with +2 EcoRI and MseI primers (AC and CC, respectively). Selective amplification was performed with fluorescent dye-labeled +3 EcoRI primer set (ACG) and three unlabeled +4 MseI primers (CCAG, CCAT, and CCTA).

AFLP profiles were generated on an ABI 3700 automated sequencer using Geneflo-625 size standard (Chimerx, Madison, WI). AFLP profiles were visually checked for polymorphism using Genescan software (ver. 3.7; Applied Biosystems, Foster City, CA). To insure reproducibility, this procedure was replicated for several samples. Two or more samples showing peaks greater than a threshold value of 150 units and ranging between 70 and 300 bp were scored as a locus. Out of a total of 102 markers, 66 were determined polymorphic from 357 samples. Polymorphic markers were used to create a distance matrix among collection sites (Cavalli-Sforza and Edwards 1967) followed by principal coordinate analysis using FAMD software (ver 1.1$\beta$; Schlüter and Harris 2006) to calculate mean eigenvectors for each collection site.

An $F_{ST}$-based simulation program, Dfdist (ver. D), was used to identify outlier loci, the loci considered to be candidates for directional selection (Beaumont and Nichols 1996; Beaumont and Balding 2004). Dfdist uses a Bayesian estimate of locus frequency for dominant markers (Zhivotovsky 1999). Locus $F_{ST}$ values were estimated from a simulated distribution of 50,000 iterations using an infinite alleles model. This distribution was used to determine outliers. Outlier loci were considered putatively under directional selection above the 0.995 quantile. These locus frequencies were regressed with the genecological prediction.

### Quantitative Genetic Data

Rehfeldt et al. (1984) described a common garden study of seedlings derived from wind pollinations in 58 natural collection sites (fig. 1). The present study uses data from 3-yr measurements at three disparate field tests. Traits measured included four variables describing growth and development in the field and injury from laboratory freezing tests. In our analysis, genetic variation among collection sites was assessed from eigenvector means from principal component analysis.

### Climate Estimates

To obtain climate estimates for all collection sites, we used climate surfaces for western North America developed from the thin plate splines of Hutchinson (1991, 2000) according to the procedures of Rehfeldt (2006; http://forest.moscowfsl.wsu.edu/climate/). The surfaces were developed from 1961–1990 monthly climate data from ~6,000 weather stations in the western United States and Canada. The surfaces allow monthly climate normals to be predicted for point locations, identified by latitude, longitude, and elevation. From these predictions, derived variables were calculated for as many as 35 climate variables (see Rehfeldt et al. 2006) that include simple expressions such as annual precipitation, complex temperature sums such as growing degree days (degree days >5°C), and temperature-precipitation interactions such as annual dryness index, the ratio of (growing degree days)$^{1/2}$ to mean annual precipitation (table 1).

To facilitate the mapping of our results, monthly normals were predicted for each cell of a digitized elevation model gridded at 0.5 min (~1 km$^2$). Derived variables were then cal-
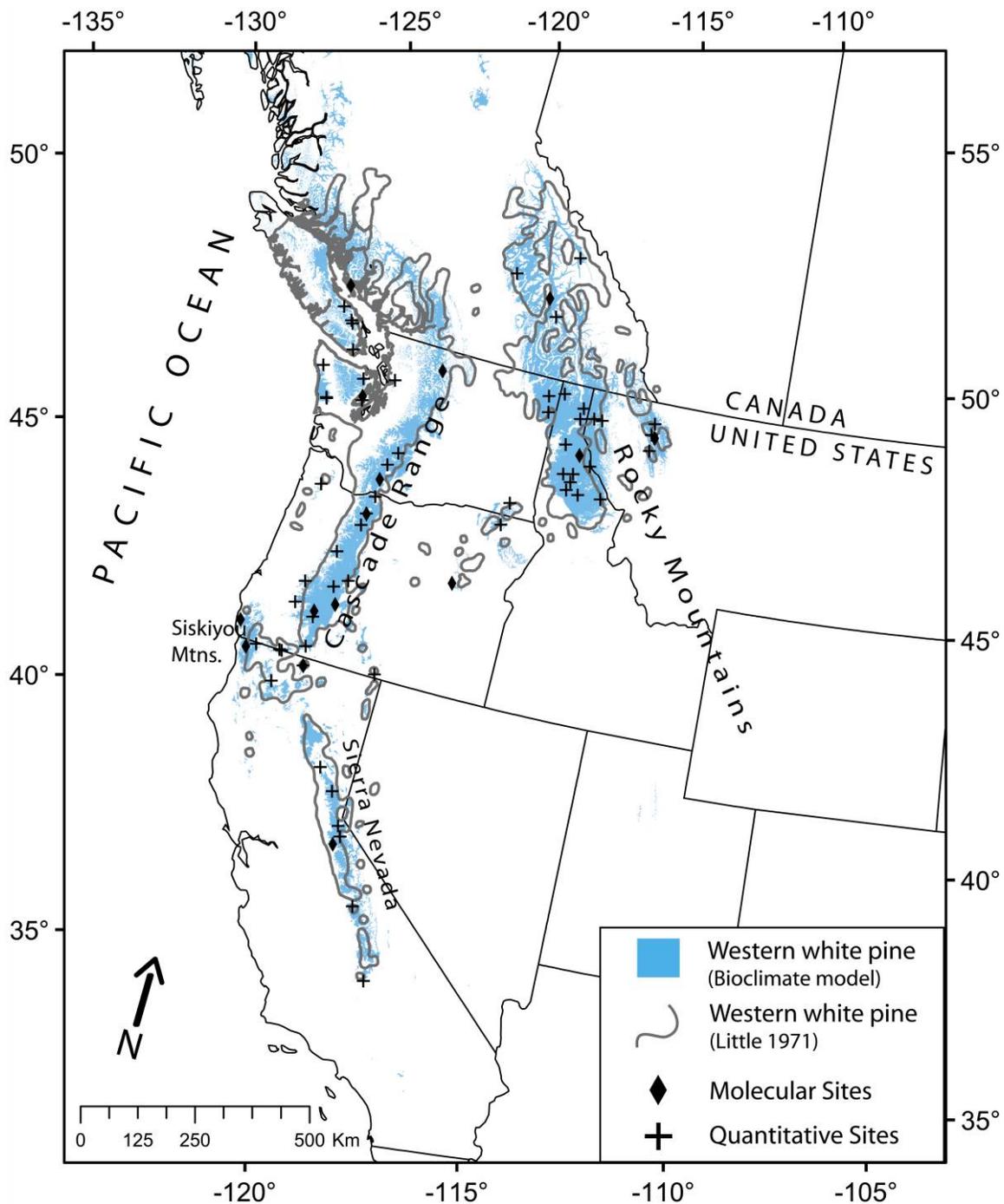
**Fig. 1** Predicted distribution of climates suitable for western white pine (*Pinus monticola*) according to a bioclimate model (Rehfeldt et al. 2006) plotted in reference to the range map of Little (1971). The climate variables used in this model are shown in table 1. Crosses indicate collection sites for the quantitative genetic data; diamonds indicate collection sites for the molecular (amplified fragment length polymorphism) data.

culated for each grid cell (see Rehfeldt et al. 2006). Thus, climate surfaces served three purposes: (1) to predict site-specific climate variables for the collection sites, (2) to map the predicted species distribution of western white pine (see Rehfeldt et al. 2006), and (3) to map the genecological predictions for both molecular and quantitative data.

### Developing Genecological Predictions

Modern data mining tools with origins in machine learning are of increasing appeal to ecological modelers (Cutler et al. 2007) for applications such as predicting spatial distributions (Thuiller et al. 2003; Elith et al. 2006; Guisan and Zimmerman 2000) or predicting forest characteristics (Moisen and

## Table 1

### Climate Variables and Their Definitions Used in the Bioclimate and Genecological Analyses

| Acronym | Definition |
| --- | --- |
| MAT | Mean annual temperature |
| MTCM | Mean temperature in the coldest month |
| MMIN | Minimum temperature in the coldest month |
| MTWM | Mean temperature in the warmest month |
| MMAX | Maximum temperature in the warmest month |
| MAP[a] | Mean annual precipitation |
| GSP | Growing season precipitation, April–September |
| TD | Summer-winter temperature differential, MTWM − MTCM |
| DD5 | Degree days >5°C |
| DD0 | Degree days <0°C |
| MINDD0[a] | Minimum degree days <0°C |
| SDAY | Julian date of the last freezing date of spring |
| FDAY | Julian date of the first freezing date of autumn |
| FFP | Length of the frost-free period |
| GSDD5 | Degree days >5°C accumulating within the frost-free period |
| D100 | Julian date the sum of degree days >5°C reaches 100 |
| ADI | Annual dryness index, $(DD5)^{1/2}$/MAP |
| SDI[a] | Summer dryness index, $(GSDD5)^{1/2}$/GSP |
| PRATIO (PR)[a] | Ratio of summer precipitation to total precipitation, GSP/MAP |

Note. Interactions used in the analyses are MAP × DD5, MAP × MTCM, GSP × MTCM, GSP[a] × DD5, DD5[a] × MTCM, MAP × TD, GSP[a] × TD, MTCM/MAP, MTCM/GSP, DD5/GSP, ADI × MTCM, SDI × MTCM, TD/MAP, TDIFF/GSP, PRATIO × MTCM, and PRATIO × DD5.

[a] Variables used in the bioclimatic species model (fig. 1).

Frescino 2002). Of these tools, regression and classification trees are demonstrating a higher accuracy and better ability to characterize complex interactions among variables than more traditional parametric techniques (Cutler et al. 2007). We now use the Random Forests multiple regression tree of Breiman (2001) to predict genetic variation in an ecological genetic framework.

The eigenvector means for each collection site from both genetic data sets were regressed separately on the climate variables as independent variables using the Random Forests multiple regression tree (Breiman 2001), available in the R statistical package (Liaw and Wiener 2002; R Development Core Team 2006). The Random Forests algorithm reserves a random sample of one-third of the data for evaluating the training sets constructed from the other two-thirds of the data. Regression trees are built by bootstrapping (i.e., 1000 bootstrap replicates) from the training sets, while the omitted data are used to estimate prediction error and variable importance. Because errors converge to a limit, effects of overfitting are nil. Root mean squared errors averaged across all trees were used to calculate variance explained, a statistic used as a guide in choosing a reasonably parsimonious model, obtained by stepwise elimination of variables until the variance explained by the model began decreasing. Variable importance was judged from the increase in the mean square error after predictor variables were permuted. To obtain a statistic reflecting the fit of the chosen model, simple correlations were made between observed and predicted eigenvector means.

### Mapping Genetic Variation

The chosen genecological models were used to predict eigenvectors from the array of climate variables estimated for

each cell of the digitized elevation model. The results of the genecological models are herein referred to as genecological predictions. Genetic variation was then mapped using Arc Map software (ver. 9.1; ESRI 2005). Because our genecological predictions are designed to describe genetic variation along climatic gradients, we mapped genetic variation as a continuous variable across the landscape; however, discussion is made easier by referring to distinct landscape patterns as climatic ecotypes.

### Cross Validation of Molecular and Quantitative Predictions

In addition to a visual comparison of maps, we used the molecular prediction to plot values of eigenvector 1 for all locations in the quantitative data set and also used the quantitative prediction to plot values of eigenvector 1 for the molecular data set. Predicted and observed values were scaled between 0 and 1, and the relationships between them were assessed with Pearson's product-moment correlation coefficient.

### Bioclimate Species Model

We mapped genetic variation within a predicted geographic distribution of western white pine. The distribution was estimated by using the Random Forests classification tree of Breiman (2001) to predict presence or absence from permanent plots maintained by Forest Inventory and Analysis, U.S. Forest Service (Alerich et al. 2004). The data consist of ~117,000 plots in the western United States, of which western white pine occurs in 2160 plots. Statistical procedures follow those of Rehfeldt et al. (2006). A stepwise elimination proce-

dure was used to find a reasonably parsimonious model. Predictor variables were eliminated using variable importance scores based on the mean decrease in accuracy. The final model was chosen from the results of 10 forests each with 100 trees, using out-of-bag errors to judge the relative quality of the models. Mapping followed the procedures outlined above. All mapped predictions for Canada are extrapolations.

### Evaluation of Individual AFLP Loci

Because principal coordinates analyses does not allow for an evaluation of eigenvector loadings from AFLP loci, simple correlations were made between genecological predictions of the molecular data set. We assessed outlier AFLP loci identified by Dfdist, discussed above. Correlations of locus frequencies (Lynch and Milligan 1994; Zhivotovsky 1999), calculated using AFLP-SURV (Vekemans 2002) and the genecological prediction, were conducted with the R statistical package using Pearson's product-moment correlation coefficient. Scatter plots were constructed for the statistically significant correlations ($P < 0.01$). These plots were used to evaluate whether specific correlations correspond to the genecological prediction. Because of multiple tests, a false discovery rate (Benjamini and Hochberg 1995) was applied to the $P$ values from correlation.

## Results

### Principal Coordinates and Principal Components Analyses

For the molecular data, the first and second eigenvectors of principal coordinates analysis explained 43.8% and 18.4% of the genetic variation, respectively. For the quantitative data, the first two eigenvectors of principal components analysis explained 52.6% and 11.3% of the genetic variation, respectively. The first eigenvector of the quantitative data was associated primarily with growth traits (e.g., shoot elongation and growth phenology), while the second eigenvector primarily reflected cold tolerance (Rehfeldt et al. 1984). In both analyses, eigenvectors other than the first and second accounted for little variation ($<9\%$), and for the quantitative data, these eigenvectors showed no correspondence to quantitative traits. Therefore, the remaining eigenvectors are ignored hereafter.

### Bioclimate Species Model

Errors of prediction averaged 4.2% across the 10 forests for the models as variables of least importance were eliminated in a stepwise fashion. This error remained relatively constant throughout the process until eight variables remained. Thereafter, the errors increased slowly as variables were removed, reaching 4.7% with a four-variable model, 7.6% with three, 7.6% with two, and 10.3% with a one-variable model. We chose the seven-variable model (table 1) as being of reasonable parsimony while providing a buffer against relying on single variables, a potentially serious problem when mapping projections. Nonetheless, as demonstrated by the classification errors, the three variables of the three-variable model were the most effective. These three, in order of their importance values, were the ratio of summer to total precipitation (PRATIO), the summer dryness index (SDI), and the product of degree days $>5\ °C$ and summer precipitation (DD5 × GSP). Note that all involve summer precipitation, and the latter two are expressions of the summer balance between temperatures and precipitation.

### Genecological Predictions

The genecological predictions we chose as being reasonably parsimonious included two climate variables for both quantitative and molecular data sets. These two predictions explained 74% and 70% of the variance among populations for the first eigenvector of the quantitative and molecular data, respectively (table 2). In both analyses, the most important climate variables, as judged by the mean decrease in accuracy, involved interactions of growing season precipitation and temperature (GSP × TD; table 2). Significant relationships were found in plotting the observed eigenvectors and the genecological predictions ($r > 0.85$, $P < 0.001$; fig. 2).

Genecological maps from the molecular and quantitative predictions for eigenvector 1 are illustrated (fig. 3). Mapped variation from the two predictions is strikingly similar: higher predicted eigenvectors, represented by dark shades of brown, were generally distributed north of 44°N in the Rocky Mountains and Cascade Range. Lower predicted eigenvectors, represented by shades of yellow, were distributed mostly south of 44°N, including Sierra Nevada, Siskiyous Mountains, and on the eastern slopes of the southern Cascade Range. Under the quantitative prediction, darker shades represent populations with higher growth potential and lighter shades represent those of lower growth potential. Also striking are the patterns of variation that cross the crest of the southern Cascade Range. Both predictions suggest steep genetic clines across the crest, with darker shades distributed on the windward, western slopes of the southern Cascade Range and lighter shades on the leeward, eastern slopes (fig. 3A, 3B, inset). The geographic distance between molecular and quantitative sites on either side of the

### Table 2

**Predictor Climate Variables, Proportion of Variance Explained, and Correlation Coefficients of the Observed with the Genecological Predictions Based on Random Forests Regression of Molecular and Quantitative Eigenvectors (Vectors 1 and 2)**

| | Quantitative model | | | | Molecular model | | | |
|---|---|---|---|---|---|---|---|---|
| | Climate variables | Variance (%) | $r$ | $P$ | Climate variables | Variance (%) | $r$ | $P$ |
| Vector 1 | GSP × DD5, GSP × TD | 74.3 | .86 | <.001 | GSP × TD, PRATIO | 70.1 | .85 | <.001 |
| Vector 2 | DDO, PR × MTCM | 59.2 | .77 | <.001 | MTCM × GSP | 34.6 | .60 | <.02 |

Note. Relationships of the eigenvectors and genecological predictions are shown in figure 2. Acronyms for climate variables are listed in table 1.
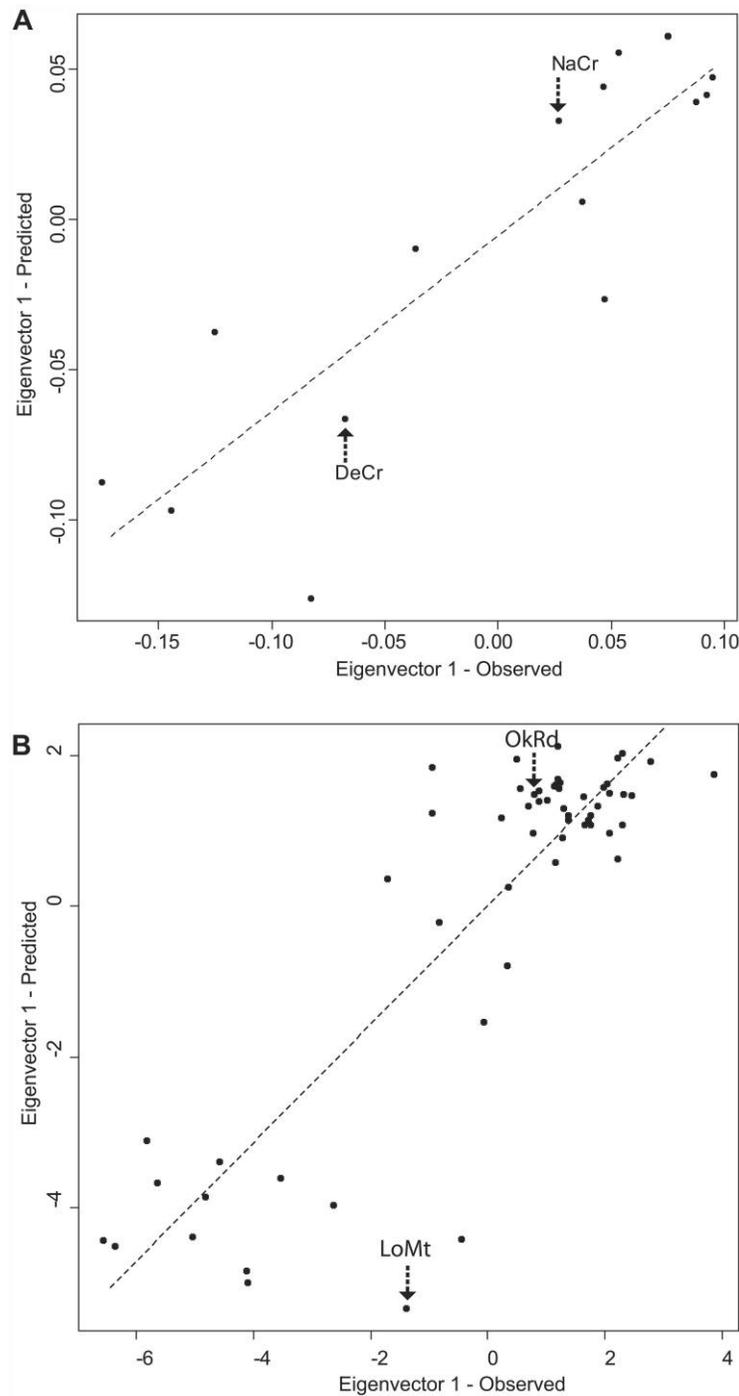
**Fig. 2** Observed values of the first eigenvector for collection sites (*X*-axis) plotted against values predicted from the Random Forest regression of climate data (*Y*-axis) for amplified fragment length polymorphism (*A*) and quantitative (*B*) data. The correlation coefficients and variables of importance are listed in table 2. Arrows indicate the proximal collection sites shown in the inset of fig. 3. The dashed line is the fitted linear regression model for the combined scatter plots (function not presented).

Cascade crest is only ~40 km, and the distribution of western white pine between sites is relatively continuous (B. A. Richardson, G. E. Rehfeldt, and M.-S. Kim, personal observation), but the genetic differences, reflected in the observed eigenvectors, are nonetheless pronounced (fig. 2).

The genecological prediction developed for the second quantitative eigenvector explained 59% of the variation in cold tolerance among sites. This eigenvector, reflecting cold tolerance among populations, was associated with two climate variables involving minimum temperatures and pre-
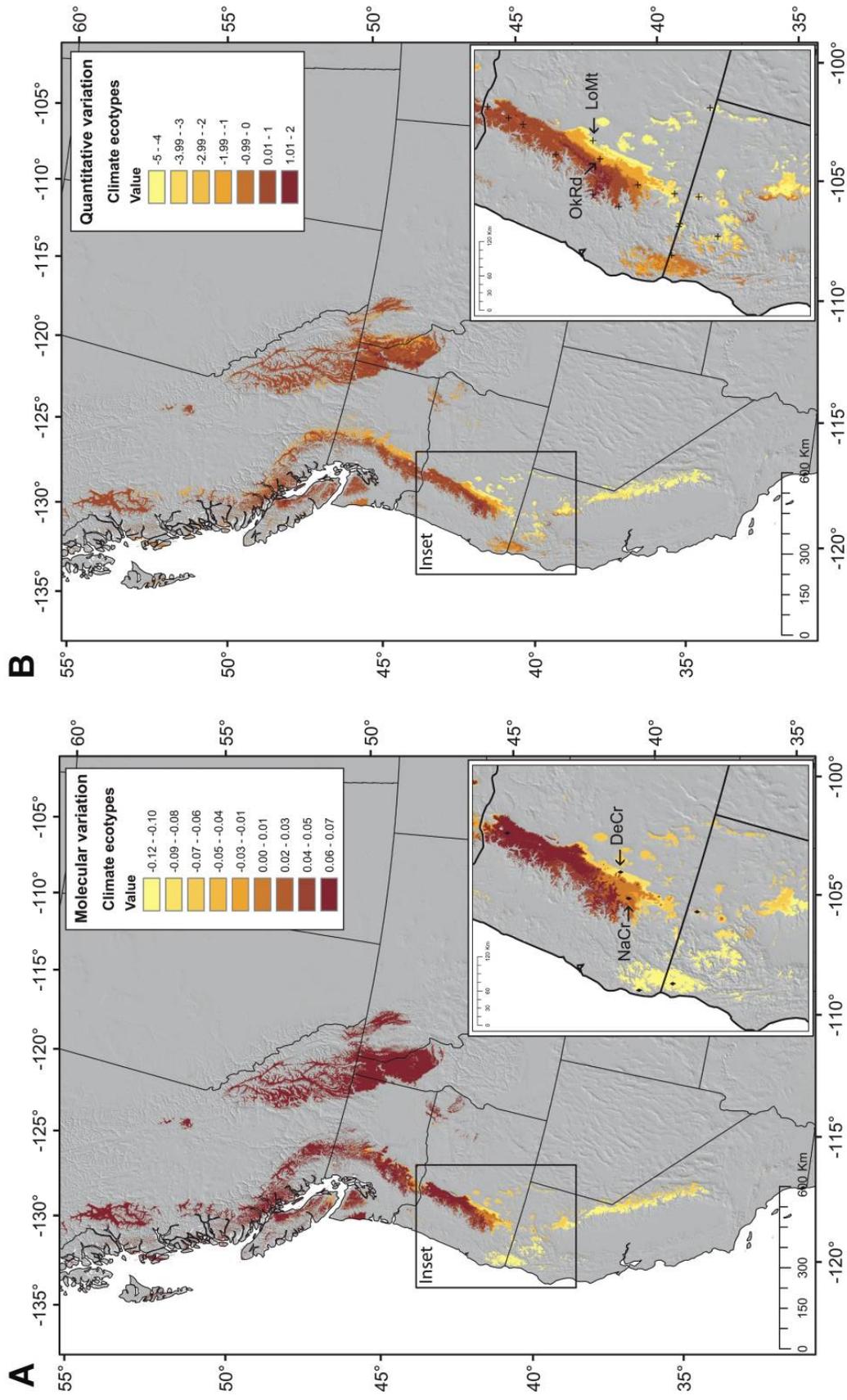
**Fig. 3** Mapped patterns of genetic variation using predicted values of eigenvectors (fig. 2) from molecular (*A*) and quantitative (*B*) genecological predictions. High to low values are represented by colors in a gradient from light yellow to dark brown. For the quantitative data, this color scheme is also directly related to predicted growth potential (yellow = low, dark brown = high). The insets illustrate a steep gradient across Cascade crest in southern Oregon, with identification of collection sites in opposing ecotypes: National Creek (*NaCr*), Deer Creek (*DeCr*), Oakridge (*OkRd*), and Lookout Mountain (*LoMt*).

cipitation (table 2). For the second molecular eigenvector, however, the Random Forests regression model accounted for only 35% of the variation among sites (table 2). After plotting the observed values against predicted values, it became obvious that the analyses were largely separating sites in the Siskiyou Mountains from the remaining 13 collection sites (data not shown). While this eigenvector represented significant population genetic structure, it showed no association to climate and no similarity to those of the second eigenvector of the quantitative prediction.

A genecological map for the second eigenvector of the quantitative data uses light shades to represent low values (i.e., low cold tolerance) and dark shades for high cold tolerance (fig. A1 in the online edition of the *International Journal of Plant Sciences*). These patterns are intuitively related to the coldness of the collection site: western white pine from the western slopes of the Cascade Range are exposed to milder climates moderated by the Pacific Ocean and consequently display lower cold tolerance. Western white pine from high elevations and regions influenced by a continental climate (i.e., east slopes of the northern Rocky Mountains) display a higher cold tolerance.

### Genecological Prediction Comparison

To demonstrate the similarities between quantitative and molecular genecological predictions, the quantitative prediction was plotted with the observed molecular eigenvectors of the first principal coordinates, and the molecular prediction was plotted with the observed quantitative eigenvectors of the first principal component. After scaling the predictions to values between 0 and 1, a simple correlation was calculated between observed and predicted values (fig. 4). A correlation of 0.84 ($P < 0.001$) was calculated between the observed quantitative eigenvectors and values for the molecular prediction, while a correlation of 0.68 ($P < 0.01$) was calculated between the observed molecular eigenvectors and values for the quantitative prediction. No significant relationships were found for the second eigenvectors of the two data sets (data not shown).

### Genome Scan and Regression with the Genecological Prediction

The genome scan revealed that eight of 66 loci (12%) were outliers exceeding neutral expectations (fig. A2 in the online edition of the *International Journal of Plant Sciences*). Regression of the molecular genecological prediction with the locus frequencies from outliers identified in the genome scan found that four of these eight were highly correlated ($P < 0.01$). When plotted with the prediction, two (TA103 and TA108) out of the four loci revealed disparate locus frequency differences between geographically neighboring sites of Deer Creek and National Creek, Oregon (fig. 5). These sites occupy opposite ends of the steep cline across the southern Cascade Range (see fig. 3A, inset). Likewise, these outliers were significantly correlated ($P < 0.01$), with individual climate variables included in the genecological prediction (data not shown). Further analysis showed that the removal of outlier loci reduced the variation explained in the first eigenvector and the similar-

ity between the molecular and quantitative genecological models (data not shown).

### Discussion

Western white pine genecological predictions of climate-related genetic variation in AFLP markers and quantitative traits were largely congruent. Similar climate variables accounted for the high proportion of variance in each prediction of the first eigenvectors (table 2). Mapped patterns of predicted genetic variation were nearly identical across the species contemporary climatic profile (fig. 3), and a cross validation exercise was strongly supportive (fig. 4). The patterns show that two major climatic ecotypes dominate both the molecular and quantitative data (fig. 3A, 3B). One climatic ecotype is predicted to occur in the Sierra Nevada and on eastern slopes of the southern Cascade Range. A second climatic ecotype occurs in the Rocky Mountains, northern Cascades, and along the western slopes of the southern Cascades. A steep cline between these groups occurs across the crest of the southern Cascade Range. This cline follows windward- to leeward-facing slopes of the Cascade Range and thus corresponds to a steep gradient in growing season precipitation and temperature. Support for these ecotypes is also found in relationships of pairwise genetic distances between collection sites and a Bayesian model-based clustering approach based on the AFLP data (M.-S. Kim, B. A. Richardson, G. I. McDonald, and N. B. Klopfenstein, unpublished manuscript). For the quantitative data, this climate gradient was associated with divergent selection in growth potential (fig. 3B, inset; Rehfeldt et al. 1984). The patterns, moreover, coincide with those demonstrated in an independent study of adaptive variation in western white pine within the Cascade Range of Oregon and Washington (Campbell and Sugano 1989). Unfortunately, the location data for this later study has been lost and cannot be analyzed anew.

No similarities were found in the comparison of the two genecological predictions developed from the second principal component/coordinate eigenvectors. However, the second mean eigenvectors of the quantitative data, which were associated with cold tolerance from laboratory freezing tests (Rehfeldt et al. 1984), were closely correlated with an interaction of minimum temperatures and precipitation of the seed source (table 2). When mapped, genecological predictions for these eigenvectors were assorted into a pattern of greater cold tolerance with continental climates and higher elevation (fig. A1). Regression of the second principal coordinate eigenvectors of the molecular data set on climatic variables revealed a significant ($P < 0.02$) linear relationship; however, this relationship is caused by two clusters of points separating collection sites in the Siskiyou Mountains from all other sites (data not shown). Incongruity between predicted effects for the second eigenvectors suggests that the molecular genetic variance associated with this second eigenvector is caused by the demographic effects. The Siskiyou collection sites are geographically relatively isolated, and variance along this axis is not likely to be directly climate related.

The central issue raised by the congruent patterns of adaptive traits and molecular markers in western white pine is the underlying evolutionary process. These congruent genetic patterns could be coincidental, derived independently from divergent climatic selection in growth potential and historical
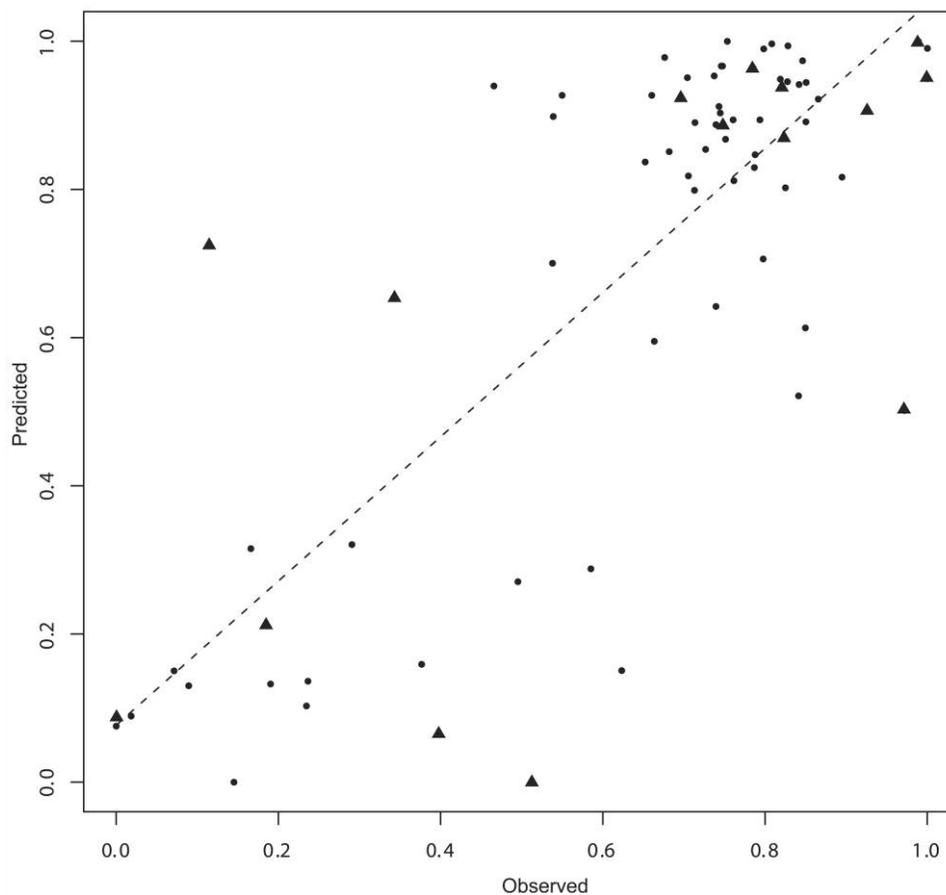
**Fig. 4** Cross validation scatter plots of observed means for the first eigenvector (*X*-axis) plotted against values of the genecological prediction (*Y*-axis). Circles are eigenvectors derived from quantitative data plotted with the values of the molecular genecological prediction; triangles are molecular eigenvectors plotted with the quantitative genecological prediction. Simple correlation coefficients were 0.84 and 0.68 ($P > 0.01$), respectively. The dashed line is the fitted linear regression model for the combined scatter plots (function not presented).

demography of molecular markers. Narrow zones of contact from expansion of postglacial refugia are common in temperate and boreal conifer species (Richardson et al. 2002; Godbout et al. 2005; Walter and Epperson 2005). These contact zones have been identified through nuclear and organellar DNA and paleoecological data. However, for western white pine, paleoecological records suggest that this species has been present on the western slopes of the southern Cascades and in northern California at least since the last glacial maximum (~18,000 yr before present; Grigg and Whitlock 1998; Mohr et al. 2000). Genetic structure between collection sites at higher latitudes (i.e., near the U.S. and Canadian border), are more likely caused by demographic processes resulting from glaciation (M.-S. Kim, B. A. Richardson, G. I. McDonald, and N. B. Klopfenstein, unpublished manuscript). Furthermore, both maternally inherited, mitochondrial (mt)DNA haplotypes (B. A. Richardson, unpublished data) and major-gene resistance, a hypersensitive reaction to a fungal disease caused by *Cronartium ribicola* (Kinloch et al. 2003), are shared in populations in either ecotype. The paleoecological data and sharing of mtDNA haplotypes and a resistance trait suggest that gene flow has occurred over time and the southern Cascade Range is not a recent contact zone. Thus, a

purely demographic hypothesis to explain a coincidental congruence in molecular markers and adaptive traits is inadequate.

In western white pine, divergent climatic selection is the most reasonable explanation for the congruent genetic patterns we describe. These patterns associated with growing season precipitation and temperature (table 2) could impose a reproductive barrier across these parapatric ecotypes. While the western white pine from these opposing ecotypes are reproductively compatible in breeding programs, no long-term, reciprocal transplant tests have been conducted to evaluate fitness of trees or ecotypic hybrids in these opposing environments. It is conceivable that immigrants from the opposing ecotypes or hybrids between ecotypes would have reduced fitness due to either competition by faster growing trees or damage/mortality caused by insufficient moisture. Thus, this steep environmental cline could act indirectly as a limited reproductive barrier. This type of reproductive isolation caused by natural selection is defined as postmating, hybrid inviability (Nosil et al. 2005). Such a process would likely have a widespread impact on the genome, detectable with AFLPs.

The congruence of AFLP and adaptive trait genetic variance in western white pine could involve a second hypothesis
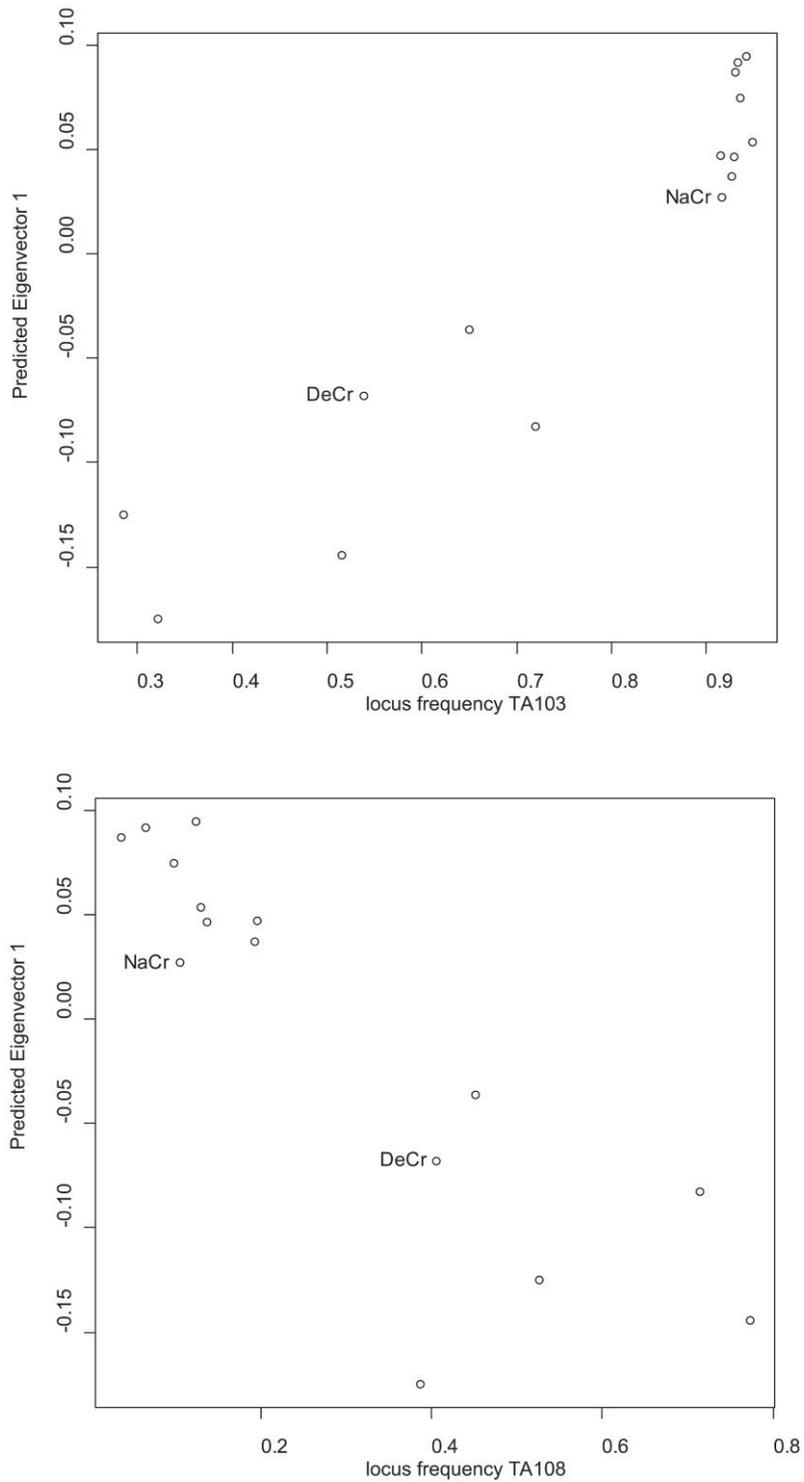
**Fig. 5** Plotted relationships of the molecular genecological prediction (*Y*-axis) and the plus-allele frequency of amplified fragment length polymorphism loci TA103 and TA108 and (*X*-axis). These loci were identified as outliers by a genome scan. The correlation coefficient was −0.87, and $P > 0.0145$ corrected for multiple testing using a false discovery rate. The collection sites Deer Creek (*DeCr*) and National Creek (*NaCr*) are indicated to show frequency differences from adjacent collection sites in opposing ecotypes.

in which regions of the genome under divergent selection are linked to some AFLP loci. This hypothesis is not mutually exclusive of reproductive isolation caused by climatic selection. However, support for AFLP linkage to regions of the genome under divergent selection is found in the differential $F_{ST}$ values from individual AFLP loci (fig. A2) and strong correlation of some genome scan outliers with genecological predictions (fig. 5). The proportion of loci under putative selection reported in this study of 12% is comparable to outlier detection in other conifers (Acheré et al. 2005; Namroud et al. 2008). The discovery of putative genetic hitchhiking loci caused by selection has been increasingly reported across ecological gradients in diverse taxa (Jump et al. 2006; Rogers and Bernatchez 2007; review in Nosil et al. 2009). Further study is needed to characterize associations between growth traits and molecular markers in western white pine and to test the fitness of western white pine ecotypes through reciprocal transplant experiments.

Global climate change poses an enormous challenge to those developing conservation strategies for diverse forest species (Aitken et al. 2008). Here, we delineate the climate ecotypes using climate surfaces (see Rehfeldt 2006). These surfaces can be readily updated to incorporate global warming scenarios from general circulation models to predict the future distribution of climates expected to be suited to each ecotype throughout this century. Such analyses on quantitative traits have been performed for Engelman spruce (*Picea engelmannii*; Rehfeldt 2004) and for Douglas fir (*Pseudotsuga menziesii*; St. Clair et al. 2005; St. Clair and Howe 2007). Information on the future distribution of climates suited to contemporary populations is critical for breeding, conservation, and long-term restoration programs (Tchebakova et al. 2005). As genomic techniques become more refined in the methods to detect an adaptive signature, this approach may be essential for time-dependent decisions regarding conservation strategies in response to climate change. While quantita-

tive genetic approaches are ideal for genecological studies, the length of time to acquire data (i.e., typically 5–10 yr) is an impediment to genetic conservation decisions for tree species under rapid climate change.

## Conclusion

This genecological study demonstrates that AFLP loci reflect climate-driven divergent selection in western white pine. We show that geographic patterns developed from molecular predictions are congruent to those developed from quantitative data obtained in independent studies conducted in common gardens. The proposed evolutionary process for the observed AFLP patterns is loosely linked loci to genomic regions under selection and/or postmating reproductive isolation cause by climatic selection. Our approach requires knowledge of the range-wide genetic variation in adaptive traits, a climate model capable of estimating site-specific climate variables, and a versatile regression algorithm, such as Random Forests, capable of handling large data sets. This study provides a basis for evaluating molecular markers for climatic selection, identifying geographic regions for further research in genecology, and prioritizing populations for conservation under current and future climates.

## Acknowledgments

## Literature Cited

Acheré V, JM Favre, G Besnard, S Jeandroz 2005 Genome organization of molecular differentiation in Norway spruce (*Picea abies*). Mol Ecol 14:3191–3201.

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

Alerich CA, L Klevgard, C Liff, PD Miles 2004 The forest inventory and analysis database: database description and users guide. Version 1.7. http://www.ncrs2.fs.fed.us/4801/fiadb/fiadb_documentation/FIADB_v17_122104.pdf.

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

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

Benjamini Y, Y Hochberg 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc 57:289–300.

Biswas S, JM Akey 2006 Genomic insights into positive selection. Trends Genet 22:437–446.

Breiman L 2001 Random Forests. Mach Learn 45:5–32.

Campbell RK, AI Sugano 1989 Seed zones and breeding zones for white pine in the Cascade Range of Washington and Oregon. USDA

Forest Service Research Paper PNW-RP-407. Pacific Northwest Research Station, Portland, OR.

Cavalli-Sforza LL, AWF Edwards 1967 Phylogenetic analysis: models and estimation procedures. Evolution 21:550–570.

Cutler RD, TC Edwards Jr, KH Beard, A Cutler, KT Hess, J Gibson, JJ Lawler 2007 Random Forests for classification in Ecology. Ecology 88:2783–2792.

Elith J, CH Graham, RP Anderson, M Dudik, S Ferrier, A Guisan, RJ Hijmans, et al 2006 Novel methods improve prediction of species' distributions from occurrence data. Ecography 29:129–151.

ESRI 2005 ArcGIS 9.1. ESRI, Redlands, CA.

Franklin JF, CT Dyrness 1973 Natural vegetation of Oregon and Washington. USDA Forest Service General Technical Report PNW-8 417. Washington, DC.

García-Gil MR, M Mikkonen, O Savolainen 2003 Nucleotide diversity at two phytochrome loci along a latitudinal cline in *Pinus sylvestris*. Mol Ecol 12:1195–1206.

Godbout J, JP Jaramillo-Correa, J Beaulieu, J Bousquet 2005 A mitochondrial DNA minisatellite reveals the postglacial history of jack pine (*Pinus banksiana*), a broad-range North American conifer. Mol Ecol 14:3497–3512.

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

Griffin JR, WB Critchfield 1972 The distribution of forest trees in California. USDA Forest Service Research Paper PSW-82. Pacific Southwest Forest and Range Experiment Station, Berkeley, CA.

Grigg LD, C Whitlock 1998 Late-glacial vegetation and climate change in western Oregon. Quat Res 49:287–298.

Guisan A, NE Zimmerman 2000 Predictive habitat distribution models in ecology. Ecol Model 135:147–186.

Hijmans RJ, SE Cameron, JL Parra, PG Jones, A Jarvis 2005 Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 25:1965–1978.

Hutchinson MF 1991 Continent-wide data assimilation using thin plate smoothing splines. Pages 104–113 in JD Jasper, ed. Data assimilation systems. Bureau of Meteorology, Melbourne.

——— 2000 Anusplin user's guide. Version 4.1. Australian National University Centre for Resource and Environmental Studies, Canberra.

Jaramillo-Correa JP, J Beaulieu, J Bousquet 2001 Contrasting evolutionary forces driving population structure at expressed sequence tag polymorphisms, allozymes and quantitative traits in white spruce. Mol Ecol 10:2729–2740.

Jump AS, JM Hunt, JA Martinez-Izquierdo, J Penuelas 2006 Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in Fagus sylvatica. Mol Ecol 15:3469–3480.

Karhu A, P Hurme, M Karjalainen, P Karvonen, K Kärkkäinen, D Neale, O Savolainen 1996 Do molecular markers reflect patterns of differentiation in adaptive traits of conifers? Theor Appl Genet 93:215–221.

Kim M-S, SJ Brunsfeld, GI McDonald, NB Klopfenstein 2003 Effect of white pine blister rust (Cronartium ribicola) and rust-resistance breeding on genetic variation in western white pine (Pinus monticola). Theor Appl Genet 106:1004–1010.

Kinloch BB, Jr, RA Sniezko, GE Dupper 2003 Origin and distribution of Cr2, a gene for resistance to white pine blister rust in natural populations of western white pine. Phytopathology 93:691–694.

Langlet O 1936 Studier over tallens fysiologiska variabilitet och dess samband med klimatet. Medd Statens Skogsforsoksanstalt 29:219–470.

Liaw A, M Wiener 2002 Classification and regression by Random Forests. R News 2:18–22.

Little EL, Jr 1971 Atlas of United States trees. Vol 1. Conifers and important hardwoods. USDA Miscellaneous Publication 1146. Washington, DC.

Lynch M, BG Milligan 1994 Analysis of population genetic structure with RAPD markers. Mol Ecol 3:91–99.

McKenney DW, MF Hutchinson, JL Kesteven, LA Venier 2001 Canada's plant hardiness zones revisited using modern climate interpolation techniques. Can J Plant Sci 81:129–143.

Meagher MD, RS Hunt 1998 Early height growth of western white pine provenances in British Columbia plantations. West J Appl For 13:47–53.

Mohr JA, C Whitlock, CN Skinner 2000 Postglacial vegetation and fire history, eastern Klamath Mountains, California, USA. Holocene 10:587–601.

Moisen GA, TS Frescino 2002 Comparing five modeling techniques for predicting forest characteristics. Ecol Model 157:209–225.

Namroud M, J Beaulieu, N Juge, J Laroche, J Bousquet 2008 Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. Mol Ecol 17:3599–3613.

Neale DB, O Savolainen 2004 Association genetics of complex traits in conifers. Trends Plant Sci 9:325–330.

Nosil P, DJ Funk, D Ortiz-Barrientos 2009 Divergent selection and heterogeneous genomic divergence. Mol Ecol 18:375–402.

Nosil P, TH Vines, DJ Funk 2005 Reproductive isolation caused by natural selection against immigrants from divergent habitats. Evolution 59:705–719.

Petit RJ, I Aguinagalde, J-L de Beaulieu, C Bittkau, S Brewer, R Cheddadi, R Ennos, et al 2003 Glacial refugia: hotspots but not melting pots of genetic diversity. Science 6:1563–1565.

R Development Core Team 2006 R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. http://www.R-project.org.

Reed DH, R Frankham 2001 How closely correlated are molecular and quantitative measures of genetic variation? a meta-analysis. Evolution 55:1095–1103.

Rehfeldt GE 2004 Interspecific and intraspecific variation in Picea engelmannii and its congeneric cohorts: biosystematics, genecology, and climate change. USDA Forest Service General Technical Report RMRS-GTR-134. Fort Collins, CO.

——— 2006 A spline climate model for western United States. USDA Forest Service General Technical Report RMRS-GTR-165. Fort Collins, CO.

Rehfeldt GE, NL Crookston, MV Warwell, JS Evans 2006 Empirical analyses of plant-climate relationships for the western United States. Int J Plant Sci 167:1123–1150.

Rehfeldt GE, RJ Hoff, RJ Steinhoff 1984 Geographic patterns of genetic variation in Pinus monticola. Bot Gaz 145:229–239.

Rehfeldt GE, WR Wykoff, CC Ying 2001 Physiologic plasticity, evolution, and impacts of a changing climate on Pinus contorta. Clim Change 50:355–376.

Richardson BA, SJ Brunsfeld, NB Klopfenstein 2002 DNA from bird-dispersed seed and wind-disseminated pollen provides insights into postglacial colonization and population genetic structure of whitebark pine (Pinus albicaulis). Mol Ecol 11:215–227.

Rogers SM, L Bernatchez 2007 The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (Coregonus sp. Salmonidae). Mol Biol Evol 24:1423–1438.

Schlüter PM, SA Harris 2006 Analysis of multilocus fingerprinting data sets containing missing data. Mol Ecol Notes 6:569–572.

St. Clair JB, GT Howe 2007 Genetic maladaptation of coastal Douglas fir seedlings to future climates. Global Change Biol 13:1441–1454.

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

Storz J 2005 Using genome scans of DNA polymorphism to infer adaptive population divergence. Mol Ecol 14:671–688.

Tchebakova NM, GE Rehfeldt, EI Parfenova 2005 Impacts of climate change on the distribution of Larix spp. and Pinus sylvestris and their climatypes in Siberia. Mitigation Adapt Strategies Global Change 11:861–882.

Thuiller W, MB Araujo, S Lavorel 2003 Generalized models vs. classification tree analysis: predicting spatial distributions of plant species at different scales. J Veg Sci 14:669–680.

Vasemagi A, CR Primmer 2005 Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. Mol Ecol 14:3623–3642.

Vekemans X 2002 AFLP-SURV. Universite Libre de Bruxelles, Belguim. http://www.ulb.ac.be/sciences/lagev/aflp-surv.html.

Walter R, BK Epperson 2005 Geographic pattern of genetic diversity in Pinus resinosa: contact zone between descendants of glacial refugia. Am J Bot 92:92–100.

Zhivotovsky LA 1999 Estimating population structure in diploids with multilocus dominant DNA markers. Mol Ecol 8:907–913.