8-2012

Continental-scale characterization of molecular variation in quaking aspen

Colin M. Callahan
Utah State University

CONTINENTAL-SCALE CHARACTERIZATION OF MOLECULAR VARIATION IN QUAKING ASPEN

by

Colin M. Callahan

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Ecology

Approved:

_____	_____
Karen E. Mock Major Professor	James N. Long Committee Member
_____	_____
Paul G. Wolf Committee Member	Mark R. McLellan Vice President for Research and Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2012

ABSTRACT

Continental-scale Characterization of Molecular Variation in Quaking Aspen

by

Colin M. Callahan, Master of Science

Utah State University, 2012

Major Professor: Dr. Karen E. Mock
Department: Wildland Resources

Quaking aspen (*Populus tremuloides*) has the largest natural distribution of any tree native to North America, ranging from Alaska through the breadth of Canada and south to mid-Mexico. The Laurentide ice sheet occupied most of the current range of *P. tremuloides* until the late Pleistocene epoch, so this species has undergone a significant, geologically recent range expansion. Surprisingly, range-wide patterns of genetic variation in *P. tremuloides* have never been described. Using a sample set representing the full longitudinal and latitudinal extent of the species distribution, I have conducted a phylogeographic analysis for *P. tremuloides.* Preliminary results comparing both nuclear and chloroplast DNA sequences revealed surprisingly low levels of divergence across the range. Because of this remarkably shallow genetic divergence among aspen populations, I used a set of rapidly-evolving molecular markers (microsatellites) to describe patterns of gene flow and diversity and to correlate those patterns

with landscape features and histories. I analyzed eight microsatellite loci in 794 individuals from 30 sampling sites. From this multilocus data set, I identified pronounced genetic structuring across the range. Strikingly, sampling sites representing the southwestern portion of the range, the western United States and Mexico, form a distinct cluster. Sites within this southwestern cluster display dramatically reduced within-site genetic diversity but elevated regional genetic diversity, which suggests that populations in the southwestern portion of the range make up a stable edge persisting through multiple climate oscillations. Based on the uniqueness of the southwestern cluster and the climatic differences between the southwest and northern portions of the range, I propose that the southwestern cluster may represent a distinct ecotype. I also identified hotspots of diversity that correspond with potential refugia during the last glacial maximum but additional work is needed to refine these patterns. Further, my findings provide a solid foundation for a range of future studies on adaptive genetic and trait variation in this species.

(61 pages)

PUBLIC ABSTRACT

Continental-scale Characterization of Molecular Variation in Quaking Aspen

by

Colin Callahan

Utah State University, 2012

Quaking aspen has the largest natural distribution of any tree native to North America, ranging from Alaska through the breadth of Canada and south to mid-Mexico. Recent studies suggest a general decline of aspen throughout much of the range since at least the mid-20th century, though these findings remain inconclusive. Regardless, factors such as climate change, periods of drought, soil nutrient deficiencies, pathogens, insects, and encroachment by other tree species all pose risks to the health and maintenance of aspen across the continent. This situation is exemplified in the western United States where climate change is forecasted to have an extreme impact on the availability of suitable habitat for aspen. Facing all of these potential challenges, it is important for aspen to be able to adapt to changes in order to persist on the landscape. A critical factor in aspen's ability to adapt to change is genetic diversity. Surprisingly, range-wide patterns of genetic variation in aspen have never been described. Using a sample set representing the full longitudinal and latitudinal extent of aspen's distribution, I have assessed levels and patterns of genetic diversity in the species. To do this, I used a set of highly variable molecular

markers, known as microsatellites, to identify individuals based on unique genotypes. I identified two genetically distinct clusters. One cluster is made up entirely of individuals from the western United States and Mexico, the southwestern portion of the range. This cluster also displays significantly less within-site genetic diversity than is seen in the rest of the range but an increase in regional diversity for the Southwest as a whole. This pattern of genetic diversity coupled with the detection of a distinct genetic cluster in the Southwest suggests that not only do southwestern populations represent a stable edge that has persisted through multiple climate oscillations, but also the presence of a distinct aspen ecotype. I was also able to identify “hot spots” of genetic diversity which, combined with knowledge of glacial history, sheds light on potential glacial refugia and routes of postglacial recolonization. This information will be important as we move toward understanding how aspen responded to dramatic climate fluctuations in the past, and to predict how the species will respond to climate change in the future.

The results of this study are important in designing long-term conservation strategies for aspen. My results suggest that a high conservation value can be placed on the “hot spots” of diversity in order to preserve diversity for the range as a whole. Also, the finding of a stable edge and distinct ecotype in the Southwest highlights the need for unique conservation strategies for populations in this region.

ACKNOWLEDGMENTS

First, I would like to thank my academic advisor, Karen Mock, for her amazing guidance and mentorship throughout my time at Utah State. I would also like to thank my thesis committee members, James Long and Paul Wolf, for sharing their extensive knowledge and insight. Further, I would like to thank John Shaw, Ronald Ryel, and Michael Madritch for assistance with sample acquisition. Many thanks to E.C. Packee, Sr., M.L. Fairweather, Richard S. Gardner, R. Daigle, L. Kennedy, S. Wilson, R.J. DeRose, USFS FIA Program, L. Ballard, D. Keefe, E. Hurd, L. Nagel, F. Baker, E.F. Martínez Hernández, J. Higginson, H.G. Shaw, B. Pitman, S. Landhäusser, R. Magelssen, and J. DeWoody for collecting samples. I am particularly grateful to the various funding sources that have supported this work: the USDA Forest Inventory and Analysis Program, the USU Cedar Mountain Initiative, the NASA Biodiversity Program, the USDA National Research Initiative, the USU Research Catalyst Program, and the USU Advance Program.

Finally, I am especially thankful to my family for their unwavering support through all of my endeavors. Special thanks to the friends I have made while at Utah State for always being supportive and making this experience so enjoyable. Most importantly, I want to thank Meredith Berry for all of her love and support.

Colin M. Callahan

CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Quaking aspen (*Populus tremuloides*) has the largest natural distribution of any tree native to North America, ranging from Alaska through the breadth of Canada and south to mid-Mexico (Little, 1971; Perala, 1990). Within this distribution, aspen persists in a broad range of ecosystems and elevations. This species is capable of reproducing both sexually, through seed and pollen adapted for long-distance wind dispersal, and asexually through root sprouts which form extensive clonal colonies (Kemperman & Barnes, 1976). Quaking aspen is prized not only for its aesthetic and cultural value but also for its importance to biodiversity as many species rely on aspen habitat. In the Intermountain West of North America, where aspen is frequently the dominant deciduous tree species, aspen stands are associated with high levels of biodiversity of many taxonomic groups including plants, birds, and butterflies (Stohlgren *et al.*, 1997a; Stohlgren *et al.*, 1997b; Mills *et al.*, 2000; Rumble *et al.*, 2001; Simonson *et al.*, 2001). Further, in the western United States, aspen also functions as a firebreak (Fechner & Barrows, 1976).

Populus species have long served as important models in genetic studies (Wullschleger *et al.*, 2002). *Populus trichocarpa*, a closely related species, was the first tree genome sequenced (Tuskan *et al.*, 2006), providing a springboard for the study of genetic variation in other *Populus* species. Due to the importance of *P. tremuloides*, both ecologically and economically (Mitton & Grant, 1996) in

much of its range, an enhanced understanding of its range-wide genetic variation is particularly critical. In the fragmented southwestern portion of aspen's range, a study of ponderosa pine (*P. ponderosa*) identified two distinct gene pools (subsequently designated "varieties") representing historically separated lineages (Latta & Mitton, 1999). Discovery of these distinct ponderosa pine varieties in the western United States suggests the potential for locally adapted lineages, and distinct varieties, in other species in this region. *P. tremuloides* is a particularly attractive candidate as a model system for molecular adaptation due to the geographic extent of its range. However, inferences about molecular adaptation are limited by a lack of information about range-wide neutral variation in this species. Surprisingly, a range-wide phylogeography of *P. tremuloides* has never been undertaken.

Phylogeography is a relatively new field of science that integrates genetic and geographic information to study the distribution of genealogical lineages particularly within species (Avise, 2000). Surveys of geographic variation in mitochondrial DNA (mtDNA) taking place in the late 1970s and into the 80s led to the discovery of striking patterns in the spatial arrangement of mtDNA lineages and the coining of the term in 1987 (Avise *et al.*, 1987). The study of phylogeography has since grown and expanded to new areas, including biodiversity conservation, microbiology, parasitology, and virology (Beheregaray, 2008). There have been a multitude of phylogeographic assessments of both North American plants and trees (Marr *et al.*, 2008; Tomimatsu *et al.*, 2009;

Beatty & Provan, 2010; de Lafontaine *et al.*, 2010; Rebernig *et al.*, 2010). Phylogeographic patterns have also been described for other *Populus* species; *P. alba* and *P. tremula* in Europe (Brundu *et al.*, 2008; Fussi *et al.*, 2010) and *P. balsamifera* in North America (Keller *et al.*, 2010). Phylogeography has proven especially effective in answering questions about postglacial recolonization patterns (Hewitt, 2000). These are of particular interest because a plethora of species now inhabit areas once covered by ice during the last glacial maximum (LGM) and because of the profound impact of postglacial history on contemporary genetic diversity, patterns of adaptation, and reproductive congruity (Hewitt, 2004). In *P. tremuloides*, post-glacial history could be particularly important, since the Laurentide ice sheet occupied most of its current range until the late Pleistocene epoch (Dyke *et al.*, 2002).

Dramatic climate fluctuations during the Pleistocene epoch (2.5 – 0.01 MYA) were manifest as shifts between glacial and interglacial cycles. The Cordilleran and Laurentide ice sheets covered much of North America during the Wisconsinan glacial episode (110 – 10 KYA), reaching their maximum extent 26 – 19 KYA. Also during the LGM, sea levels dropped as much as 20 meters, exposing extended coastlines previously and currently under water (Clark & Mix, 2002). Phylogeographic surveys of several taxa in North America have revealed much about postglacial recolonization and the existence of multiple glacial refugia across the continent (Pielou, 1991). See Beatty and Provan (2010) for a review of putative glacial refugia for both plants and animals in North America.

The current distribution of *P. tremuloides* overlaps with multiple putative glacial refugia, both north and south of the ice sheets (Fig. 1). Based on examination of the fossil record, a likely precursor of *P. tremuloides* (*P. pliotremuloides)* is also known to have existed far south of the boundaries of the LGM prior to the Pleistocene, specifically as far south as southern California in the west during the Pliocene (Axelrod, 1941; Millar, 1996). The persistence of *P. tremuloides* in the unglaciated Beringia refugium of northern Alaska during the LGM has previously been investigated through mapped pollen data and fossil records (Brubaker *et al.*, 2005) but remains uncertain due to poor preservation of pollen in peat and lake sediments and the difficulty distinguishing wood from that of *P. balsamifera*.

As it remains unclear whether *P. tremuloides* persisted in northern and southern glacial refugia during the LGM, it also remains unclear whether the rear-edge populations in the southern portion of the current range represent a stable or trailing edge (Hampe & Petit, 2005) throughout past climate oscillations. Under a stable edge scenario, southern populations will have persisted through multiple climate oscillations. Alternatively, under a trailing edge scenario, the whole species range will have shifted in response to climate oscillations, and present-day southern populations may have disappeared during northward shifts and been re-established from northern populations during southward shifts. If the populations representing the southern portion of aspen's range indeed represent a stable edge, one would expect to see reduced within-population genetic

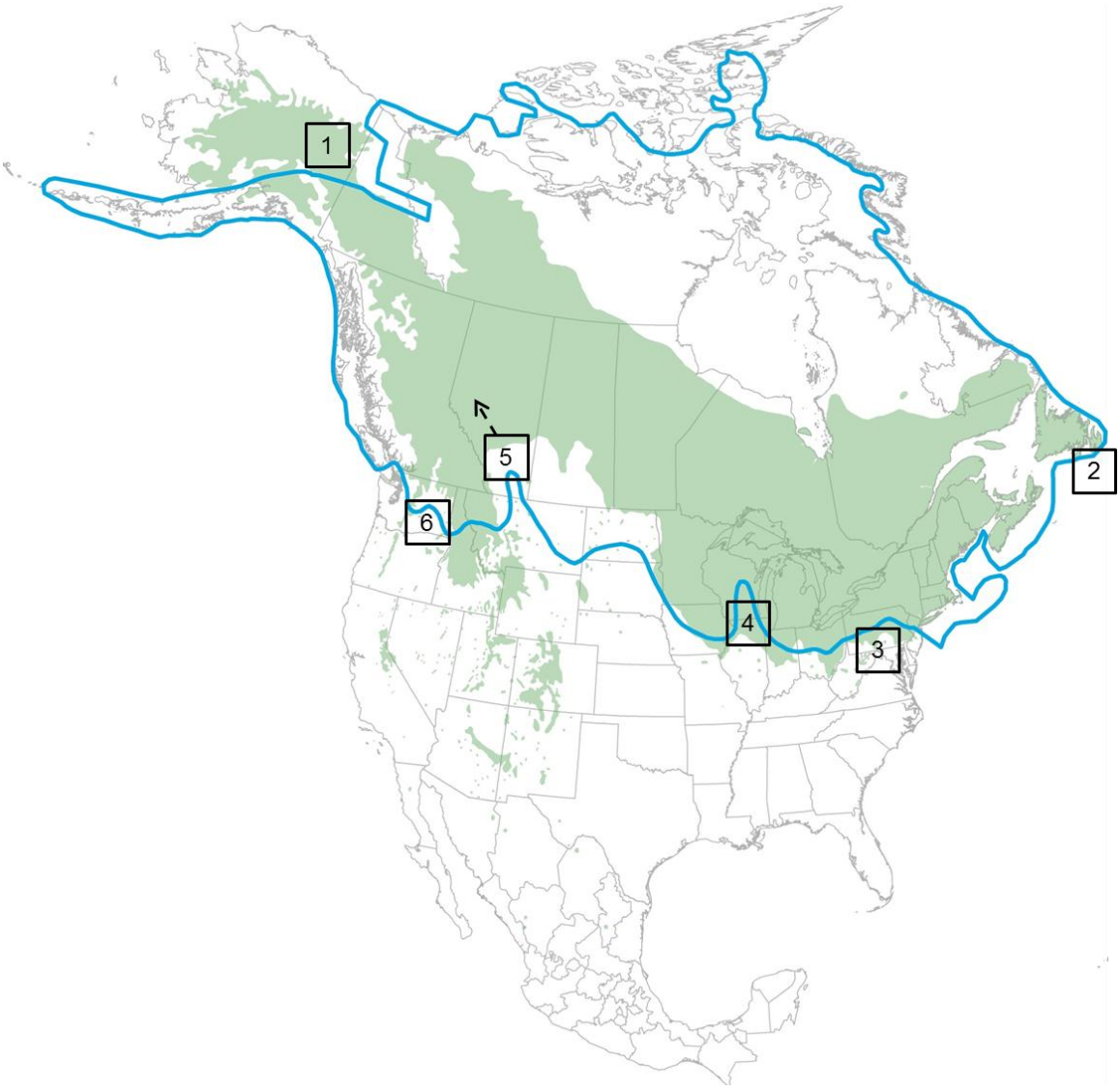

Figure 1 Potential refugial sites for *P. tremuloides* during the LGM: 1 – Beringia (Holder *et al.*, 1999; Tremblay & Schoen, 1999; Brubaker *et al.*, 2005; Anderson *et al.*, 2006; Loehr *et al.*, 2006); 2 – Grand Banks (Holder *et al.*, 1999); 3 – Northeastern United States (Tremblay & Schoen, 1999); 4 – "Driftless Area" (Rowe *et al.*, 2004); 5 – "Ice-free Corridor" (Mandryk *et al.*, 2001); 6 – Clearwater Refugium (Brunsfeld & Sullivan, 2005). Distribution of *P. tremuloides* is shown in green, outline of last glacial maximum shown in blue (Ray, 2001). The dashed arrow represents the uncertain extent of the ice-free corridor. Figure adapted from Beatty and Provan (2010).

diversity (Castric & Bernatchez, 2003; Petit *et al.*, 2003; Chang *et al.*, 2004) but increased regional genetic diversity due to markedly high levels of genetic differentiation among populations (Comps *et al.*, 2001; Castric & Bernatchez, 2003; Hampe *et al.*, 2003; Petit *et al.*, 2003; Martin & McKay, 2004). Conversely, if these populations represent a trailing edge, one would expect genetic diversity similar to that in the rest of the range or perhaps lower due to founder effects and/or increasing isolation and shrinking population sizes (Hampe & Petit, 2005). Identification of a stable versus trailing edge for *P. tremuloides* is critical as rear edge populations may be disproportionately valuable in terms of evolutionary potential and conservation value (Hampe & Petit, 2005). This issue is magnified for *P. tremuloides* in the western U.S., where the area currently occupied by aspen is projected to contract by up to 94% within a century, based on climate modeling (Rehfeldt *et al.*, 2009). Here, I use a phylogeographic approach to assess range-wide genetic structure, characterize genetic diversity across the range, and elucidate patterns of postglacial recolonization by *P. tremuloides*.

MATERIALS AND METHODS

Samples were collected from 39 georeferenced sampling sites representing the full longitudinal and latitudinal extent of aspen's range. A minimum of 10 ramets were sampled, by collecting whole leaves, from each site resulting in a total of 1221 individuals sampled across the 39 sampling sites. Samples from each site were collected in a dispersed fashion within a maximum area of 81 km^2 to avoid resampling of clones as some clones have been shown to be quite large (Mock *et al.*, 2008). Leaves from each individual were placed in separate envelopes which were then submerged in silica desiccant within air-tight containers stored at ambient temperature. DNA was extracted from dried leaf tissue using a Qiagen DNEasy 96 Plant Kit following manufacturer's protocol.

In a preliminary pilot study, I compared nuclear and chloroplast sequences across individuals from four geographically dispersed sampling sites (BNF, MI, MXH, and NH) (Fig. 2). The purpose of this work was to identify loci which were variable across the species range and would therefore be useful for assessment of rangewide phylogeographic patterns. Chloroplast loci to be sequenced were chosen based on previously demonstrated high levels of variation in other angiosperms; specific loci used in my study were *psbD* – $trnT^{(GGU)}$ and *petL* – *psbE* (Shaw *et al.*, 2007). Four nuclear exon sequences were chosen based on results of a previous study in *P. tremula* that reported an excess of polymorphism

Figure 2 Map of sampling sites. Distribution of *P. tremuloides* is shown in green, outline of last glacial maximum shown in blue (Ray, 2001).

across samples from a larger set of nuclear loci (Ingvarsson, 2008). Results of my exploratory sequencing revealed surprisingly low levels of divergence across the four sampling sites, with strong evidence of unsorted lineages and shared alleles between sites. For the first chloroplast locus *psbD* – *trnT*$^{(GGU)}$, the total fragment length was 394bp which yielded no variable sites among the four sampling sites sampled. The second chloroplast locus, *petL* – *psbE*, resulted in a 625bp fragment with four variable sites found among sampling sites (0.0064% divergence). The four nuclear coding regions resulted in a total alignment length of 2,975bp with only one variable site among sampling sites (0.00034%). Due to the shallow divergence among sites and apparent lack of lineage sorting revealed by preliminary sequencing, I chose to use microsatellites to construct the range-wide phylogeography. These more rapidly evolving markers serve as powerful tools to describe patterns of gene flow and diversity. Eight microsatellite loci were amplified for all samples: WPMS14, WPMS15, WPMS 17,and WPMS20 (Smulders *et al.*, 2001), and PMGC486, PMGC510, PMGC2571, PMGC2658 (http://www.ornl.gov/sci/ipgc/ssr_resources.htm); (Tuskan *et al.*, 2006). Reactions were prepared using 1.0 µL of template with final concentrations of 0.20 mM each dNTP, 1X PCR buffer, 0.25 µM each primer, and 0.01 U *Taq* polymerase for a final reaction volume of 10 µL. The following thermocycling conditions were used: 95° for 2 minutes followed by 30 cycles of 94° for 35 seconds, primer-specific annealing temperature for 40 seconds, 72° for 45 seconds followed by a final 10 minute extension at 72°. Primer-specific

annealing temperatures were 60° (WPMS14, WPMS15, WPMS20), 57° (PMGC486), 56° (WPMS17, PMGC510), and 55° (PMGC2571, PMGC2658). Following PCR, products were analyzed on either an ABI 3100 or ABI 3730 sequencer using a LIZ500 size standard. Microsatellite peaks were scored using ABI GeneMapper software.

After scoring all microsatellite loci, samples that failed to amplify for at least seven of the eight loci were removed from the data set. Any samples with three alleles at any loci (putative triploids) were also removed from the data set to allow for downstream analysis. Finally, any samples from the same sampling area that were identical at all loci were identified using the “multilocus matches” function in GenAlEx software (Peakall & Smouse, 2006). Duplicates were removed, leaving one representative of each unique genotype. Following the removal of these samples, I was left with a data set comprised of 842 individuals from 39 sampling areas. From this, any sampling areas represented by fewer than 10 individuals were removed in order to maximize the statistical power in downstream analyses. The final data set included 794 individuals from 30 sampling areas (Fig. 2; Table 1). Assessments of Hardy-Weinberg equilibrium (HWE) for all loci in all sampling sites and genotypic disequilibrium for all pairs of loci in all sampling sites were carried out in Arlequin version 3.5.1.3 (Excoffier & Lischer, 2010).

To identify genetically distinct clusters from my multi-locus data set, I used a Bayesian algorithm and clustering model implemented in the software program

Table 1 List of sampling sites. Longitude and latitude given in decimal degrees represent a central point at each location. The number of individuals from each site (N) is the final number included in the data set following removal of triploids, identical matches, and any individuals that failed to amplify at a minimum of seven of eight microsatellite loci. Sample collectors: [a]EC Packee, Sr.; [b]ML Fairweather; [c]R Daigle; [d]L Kennedy; [e]S Wilson; [f]RJ DeRose; [g]L Ballard; [h]E Hurd; [i]F Baker; [j]EF Martínez Hernández, J Higginson; [k]USFS FIA Program; [l]B Pitman; [m]R Magelssen.

Sampling Site	Location	Longitude	Latitude	N	Max. Dist. (km) among Ramets
AKCF[a]	Coldfoot, AK	-150.114	67.425	29	79
AKK[a]	Kenai, AK	-149.724	60.480	17	24
AZ[b]	Coconino Co., AZ	-111.846	35.448	45	7
BCQ[c]	Baie Comeau, Quebec, CANADA	-68.469	49.395	18	24
BNF[d]	Boise National Forest, ID	-115.399	43.502	20	30
CANV[e]	South Lake Tahoe, CA and NV	-120.064	38.929	18	25
CMNY[f]	Catskill Mtns., NY	-74.087	42.119	13	8
DLN[c]	Deer Lake, Newfoundland, CANADA	-57.445	49.160	14	21
FLFL[g]	Flin Flan, Manitoba, CANADA	-101.705	54.730	17	10
HIN[g]	Hinton, Alberta, CANADA	-117.531	53.429	10	34
HSPQ[c]	Haure St. Pierre, Quebec, CANADA	-64.343	50.289	25	23
KFO[h]	Klamath Falls, OR	-122.075	42.222	13	22
LALO[g]	La Loche, Saskatchewan, CANADA	-108.125	55.568	16	18
MI	Ontonagon Co., MI - Ottawa N.F.	-89.178	46.643	26	17
MN[i]	Itasco County, MN	-94.056	47.636	18	8
MNL	Lake Co., MN	-91.441	47.280	94	15
MNSL	St. Louis Co., MN	-92.169	47.802	45	41
MXQ[j]	Queretaro, MEXICO	-99.593	20.830	13	4
NMI	Emmet Co., MI	-84.981	45.544	49	27
NMT[k]	Taos Co., NM	-105.573	36.407	12	81
NVW[k]	White Pine Co., NV	-114.647	39.350	11	44
POTR[l]	Yellowstone Park, MT	-109.524	45.306	27	30
SFQ[c]	St. Felicien, Quebec, CANADA	-73.179	48.919	23	26
UME[f]	Univ. of Maine	-68.662	44.910	10	1
USF	Swan Flats - UT	-111.503	41.996	20	2
WII	Iron Co., WI	-90.520	46.460	56	10
WIMI	Wisconsin/Michigan Border	-89.097	46.435	103	42
WIO	Oakland Co., MI	-89.689	45.573	37	3
WNC[c]	Wentworth, Nova Scotia, CANADA	-63.561	45.608	21	23
WWA[m]	Kittitas Co., WA	-120.545	47.277	12	16

Structure version 2.3 (Pritchard *et al.*, 2000). Because I know that *P. tremuloides* has undergone a recent and dramatic range expansion, and that both pollen and seed are capable of long distance wind dispersal, I chose the admixture ancestry model in Structure. In general, admixture models are considered favorable and more robust than models lacking admixture when local populations are likely to share migrants (Francois & Durand, 2010). I also chose the correlated allele frequencies model (*F* model) which assumes a unique rate of change in allele frequencies due to genetic drift, conversely related to effective population size, in subpopulations simultaneously diverging from a parental population (Falush *et al.*, 2003). For all Structure runs, the length of the burnin period was set to 20,000 with 50,000 MCMC iterations after burnin. Structure assigns each individual probabilistically to any number of clusters (*K*) where the range of *K* values to be tested is determined *a priori*. Each individual is assigned membership coefficients for each cluster which sum to 1 across the number of *K* clusters identified. The range of possible *K values* tested was 1 – 10 with 10 iterations completed for each *K* value. The most likely *K* was determined by calculating Δ*K* following Evanno (2005). All calculations of Δ*K* were performed using the online version of Structure Harvester v0.6.91 (Earl & vonHoldt, 2012). I used *CLUMPP* version 1.1.2 (Jakobsson & Rosenberg, 2007) to account for problems with multimodality and label switching between iterations of Structure runs. All Structure results were plotted using *Distruct* version 1.1 (Rosenberg, 2004).

Using clusters identified by Structure, an analysis of molecular variance (AMOVA) was carried out, implemented in Arlequin version 3.5.1.3 (Excoffier & Lischer, 2010). Mantel tests, implemented in GenAlEx (Peakall & Smouse, 2006), were used to test for significant correlations between genetic (linearized Φ_{ST}) and geographic distances between sites within clusters identified by Structure. To describe genetic diversity within sampling areas across the range, average allelic richness values, across all eight loci, were calculated using FSTAT version 2.9.3.2 (Goudet, 2002). To visualize allelic richness across the range, the kriging interpolation tool was implemented in ArcGIS version 10 (ESRI, Redlands, CA, USA) to construct an interpolated surface based on average allelic richness values input for each sampling area. Observed (H_0) and expected (H_E) heterozygosity values were calculated for each sampling site, averaged across all eight loci, in Arlequin. Inbreeding coefficients (F_{IS}) were calculated using FSTAT.

RESULTS

From my multilocus data set of 794 individuals from 30 sampling sites, I identified *K*=2 clusters using the Bayesian algorithm implemented in Structure followed by calculation of Δ*K* (Fig. 3). One of the clusters (hereafter the 'southwestern cluster') is represented by individuals from the nine sampling sites in the southwestern portion of aspen's range (AZ, BNF, CANV, KFO, MXQ, NMT, NVW, USF, and WWA) (Fig. 3). The alternate cluster is represented by the rest of the sampling sites in the data set, all of which are north of the southwestern cluster and will hereafter be referred to as the 'northern cluster'. Next, I tested for further structuring within each of these two clusters. I isolated the southwestern cluster, 164 individuals from nine sampling sites, and repeated the Structure analysis. Once again, following calculation of Δ*K*, *K*=2 optimal clusters were identified, referred to as the southwest-south (SWS) and southwest-north (SWN) clusters from here on (Fig. 4). When the same analysis was performed for the north cluster, 630 individuals from 21 sampling sites, *K*=2 clusters were again identified. This time, however, membership coefficients revealed weaker structuring as many individuals had strong probabilities of assigning to either one of the two clusters (Fig. 5).

Out of 240 locus/site combinations, and using Bonferroni-corrected significance levels, we found deviations from Hardy-Weinberg expected genotypic proportions at the AZ site (WPMS17 and PMGC2658), the MN site

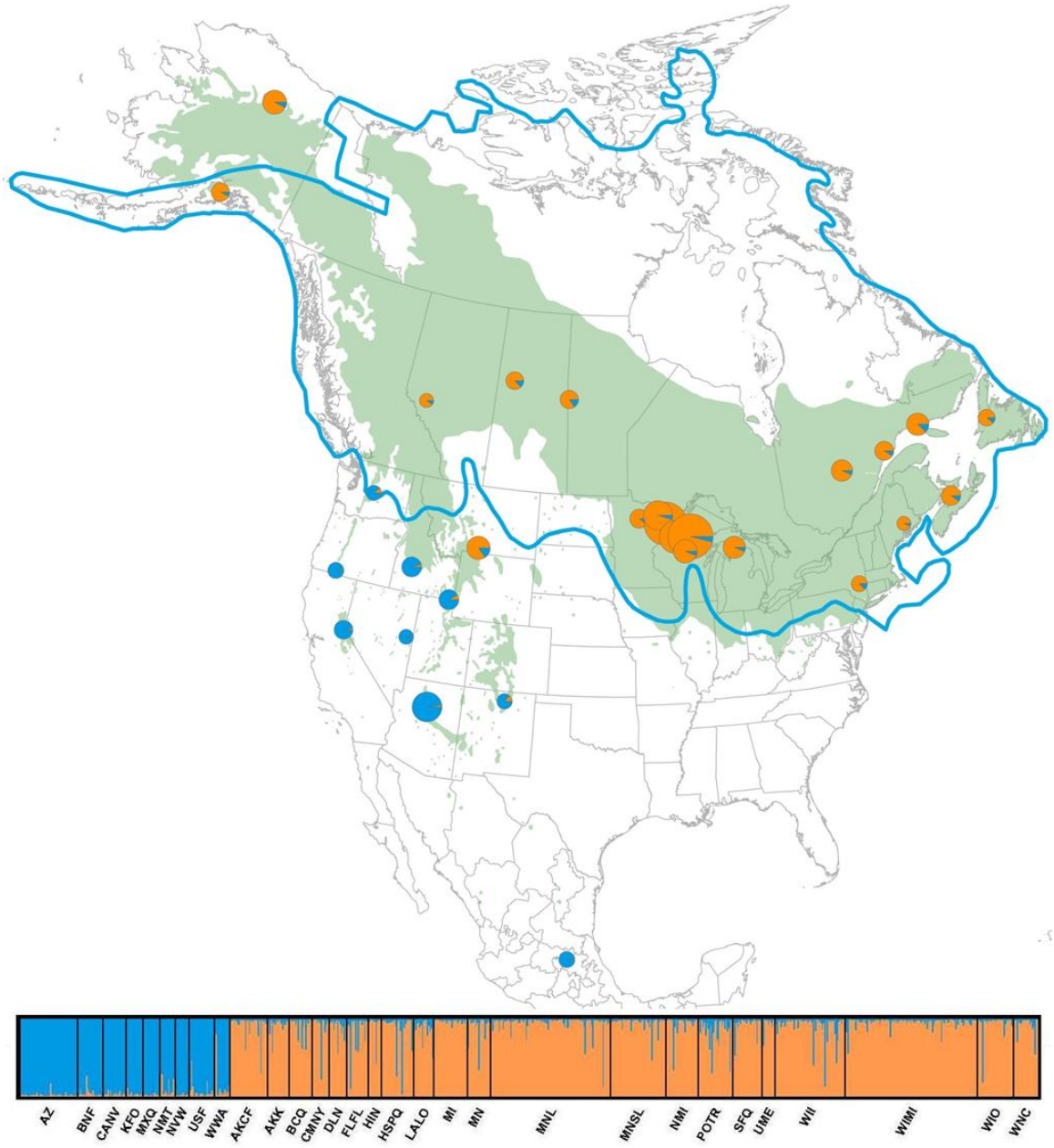

Figure 3 Assignment of sampling sites to *K*=2 clusters identified by Structure based on microsatellite allele frequencies. Pie charts represent probability of each sampling site belonging to each of the two clusters, probability values normalized using CLUMPP. Pie chart sizes are relative sample size (N) at each sampling site. Individual membership coefficients are displayed in bar plot. Each individual is represented by a single bar, sampling site labels displayed below. Southwestern cluster shown in blue, northern cluster shown in orange.

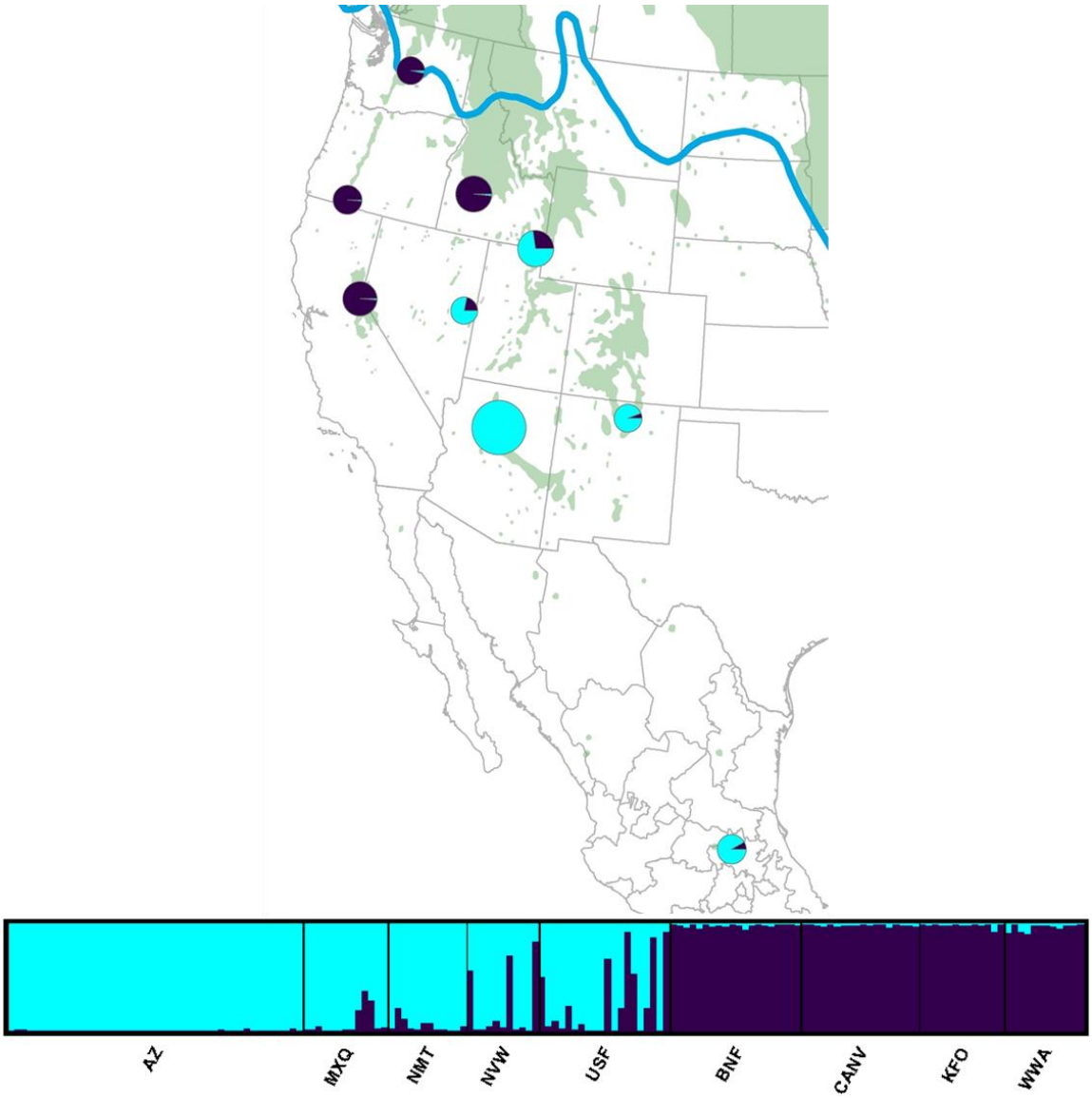

Figure 4 Assignment of sampling sites from the southwestern cluster to *K*=2 clusters identified by Structure based on microsatellite allele frequencies. Pie charts represent probability of each sampling site belonging to each of the two clusters, probability values normalized using CLUMPP. Pie chart sizes are relative to sample size (N) at each sampling site. Individual membership coefficients are displayed in bar plot. Each individual is represented by a single bar, sampling site labels displayed below. SWS cluster shown in dark purple, SWN cluster shown in light blue.

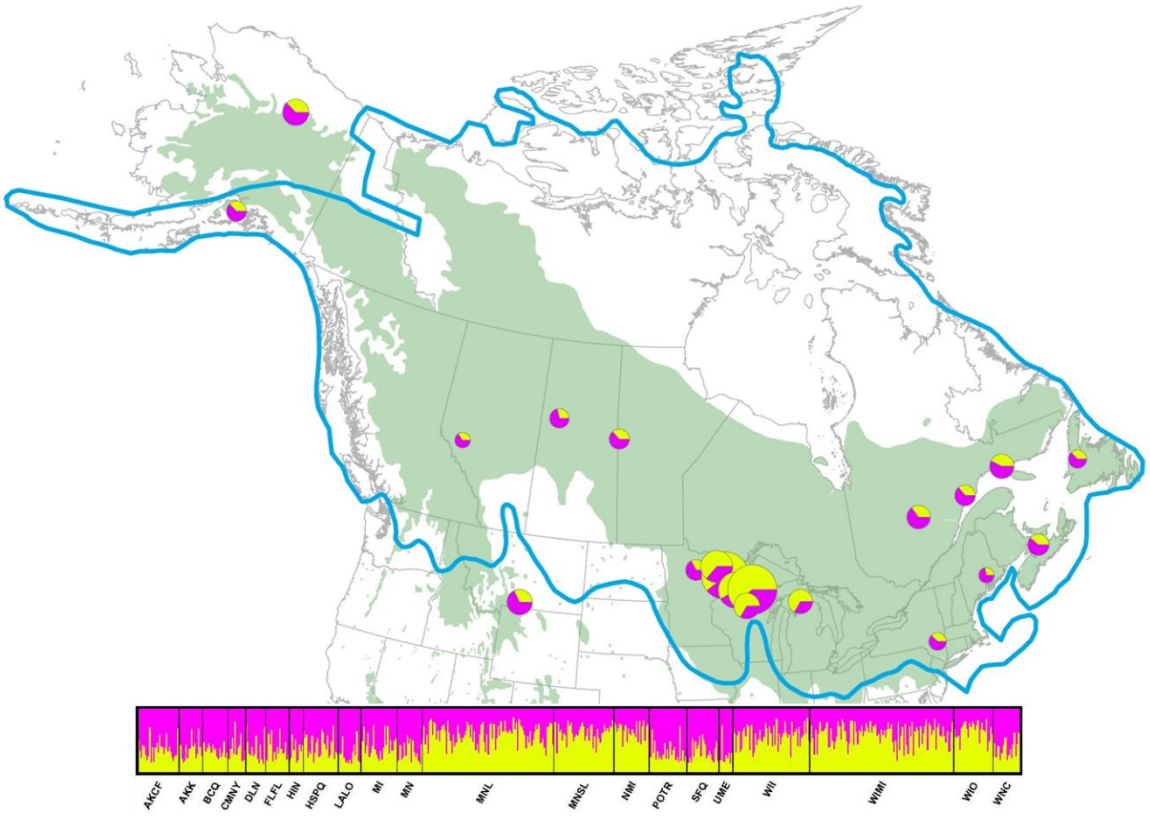

Figure 5 Assignment of sampling sites from the northern cluster to *K*=2 clusters identified by Structure based on microsatellite allele frequencies. Pie charts represent probability of each sampling site belonging to each of the two clusters, probability values normalized using CLUMPP. Pie chart sizes are relative to sample size (N) at each sampling site. Individual membership coefficients are displayed in bar plot. Each individual is represented by a single bar, sampling site labels displayed below.

(PMGC2571), and the USF site (WPMS20 and PMGC2658). No significant genotypic disequilibrium was detected following Bonferroni correction. I used AMOVA and F_{ST} to assess the degree of structuring among the two major clusters (southwestern and northern) and sampling sites within each of these clusters (Table 2a). AMOVA results revealed that 5.4% of the genetic variation was partitioned among the two major clusters, 3.2% among sampling sites within the two major clusters, and 91.4% within sampling sites with an F_{ST} of 0.086. The same AMOVA was run to assess the degree of structuring among the two subclusters identified within the southwestern cluster (SWN and SWS) and among sampling sites within each of these subclusters (Table 2b). AMOVA results revealed that 11.4% of the genetic variation in the southwestern cluster was partitioned among the two subclusters, 6.1% among sampling sites within the two subclusters, and 82.5% within sampling sites with an F_{ST} of 0.175.

Genetic distance was significantly correlated with geographic distance in the southwestern cluster (P<0.01) but not the northern cluster (P=0.311). Sites within the southwestern cluster displayed a much wider range of linearized pairwise Φ_{ST} than those within the northern cluster (Fig. 6). Allelic richness values, averaged across all eight loci for each of the two major clusters, did not differ significantly at 14.592 and 16.995 for the southwestern and northern clusters respectively (P=0.078) (Fig. 7a; Table 3). There was also no significant difference detected between each of the southwestern subclusters, with allelic richness values of 11.424 and 10.455 for the SWS and SWN clusters,

Table 2 Results of analysis of molecular variance (AMOVA) based on microsatellite allele frequencies for (a) southwestern and northern clusters and (b) subclusters within the southwestern cluster, SWS and SWN.

Source of Variation	Sum of squares	Variance Components	% Variation
(a)			
Among southwestern and northern clusters	101.594	0.17787	5.45
Among sites within clusters	232.269	0.10331	3.17
Within sites	4644.436	2.98102	91.38
Total	4978.299	3.26220	
(b)			
Among SWS and SWN clusters	70.425	0.38069	11.41
Among sites within clusters	68.012	0.20448	6.13
Within sites	877.292	2.75013	82.46
Total	1015.729	3.33530	

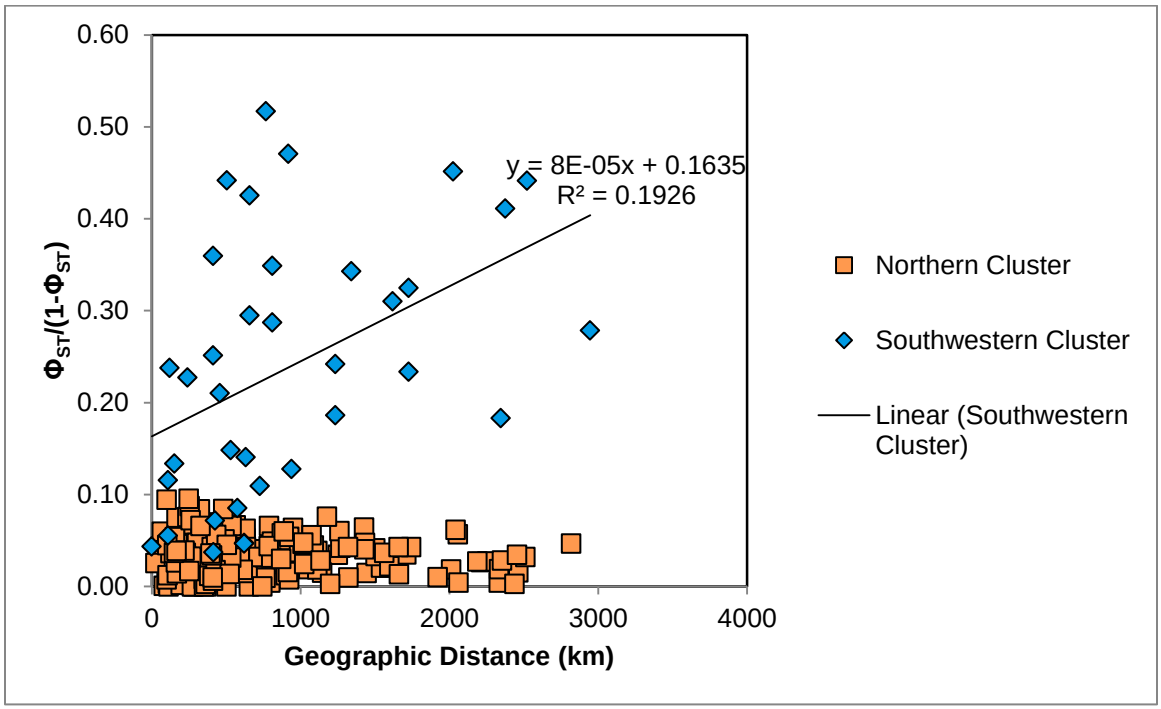

Figure 6 Geographic versus genetic distance for sites within the southwestern and northern clusters. Each point represents the genetic distance between two sites as a function of geographic distance.

respectively (P=0.373) (Table 3). Within sites, allelic richness across the range, varied from 3.337 (MXQ) to 6.832 (WNC) (Fig. 7b). Sampling sites within the southwestern cluster had an average allelic richness of 4.989, significantly less than sites within the northern cluster which had an average allelic richness of 6.415 (P<0.001) (Fig. 7c). Average allelic richness of sampling sites within the two subclusters was 4.923 and 5.073 for the SWS and SWN subclusters respectively. Elevated levels of allelic richness among sampling sites, along with interpolation of allelic richness across the range, identified Alaska, the Great Lakes area, and the northeastern U.S./Canada as areas of elevated within-site diversity and the southwestern portion of the range as an area of decreased within-site diversity (Fig. 8).

Observed site-level heterozygosity (H_0) values ranged from 0.572 (AZ) to 0.850 (UME) and expected site-level heterozygosity (H_E) values ranged from 0.613 (MXQ) to 0.801 (WNC) (Table 3). Average H_0 among sampling sites within the two major clusters was 0.677 and 0.773 for the southwestern and northern clusters respectively. Average H_E among sampling sites within the two major clusters was 0.711 and 0.779 for the southwestern and northern clusters respectively. Average H_0 among sampling sites was 0.670 and 0.685 for the SWS and SWN subclusters, respectively. Additionally, average H_E among sampling sites within each of the subclusters was 0.715 and 0.707 for the SWS and SWN subclusters respectively. When sampling sites within major clusters were pooled and each cluster treated as a single group, cluster-level H_E values

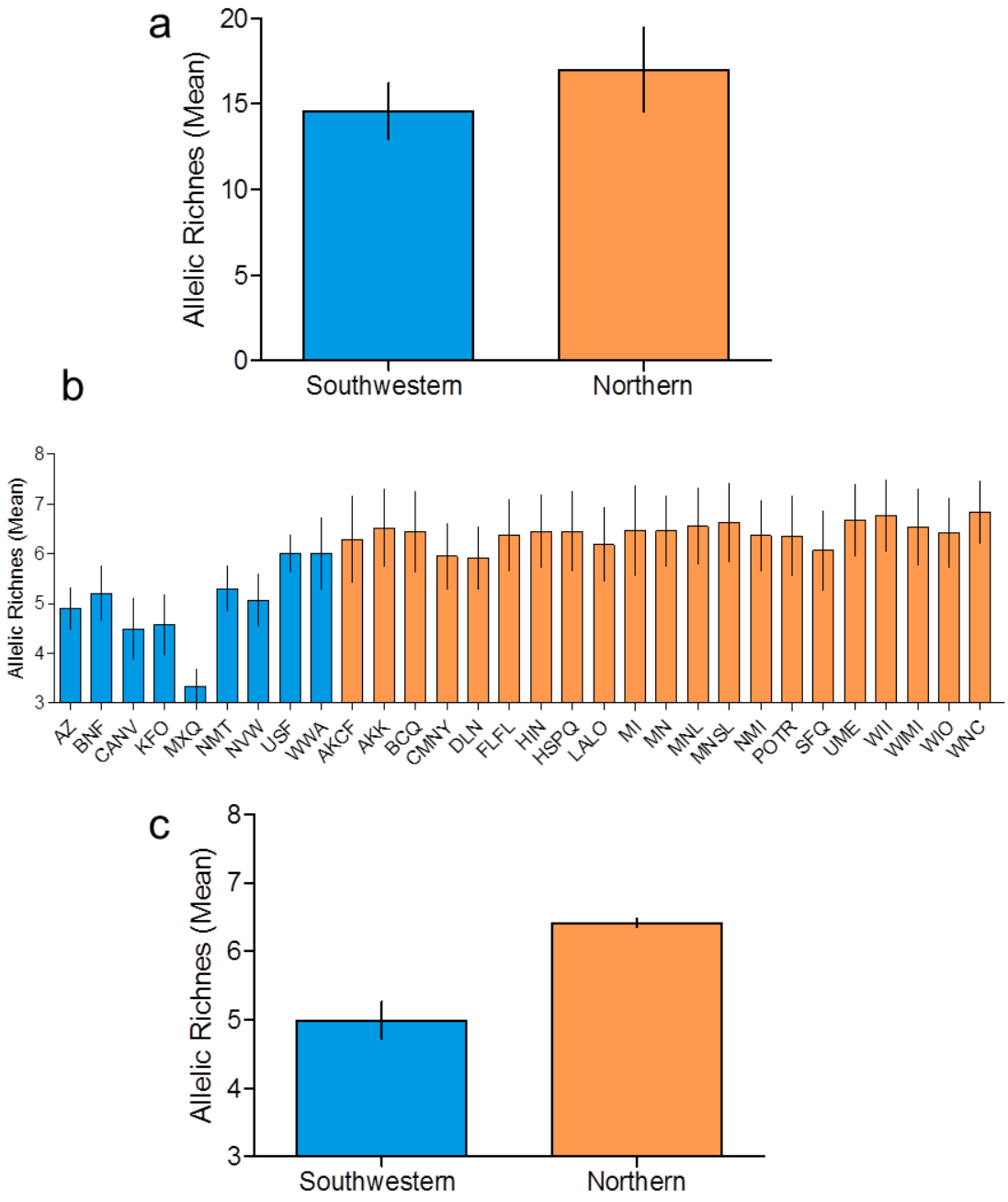

Figure 7 Measures of allelic richness. (a) Regional allelic richness values, averaged across all eight loci for each cluster. (b) Within-site allelic richness for each of the 30 sampling sites, colors represent cluster assignment. (c) Average within-site allelic richness for each cluster.

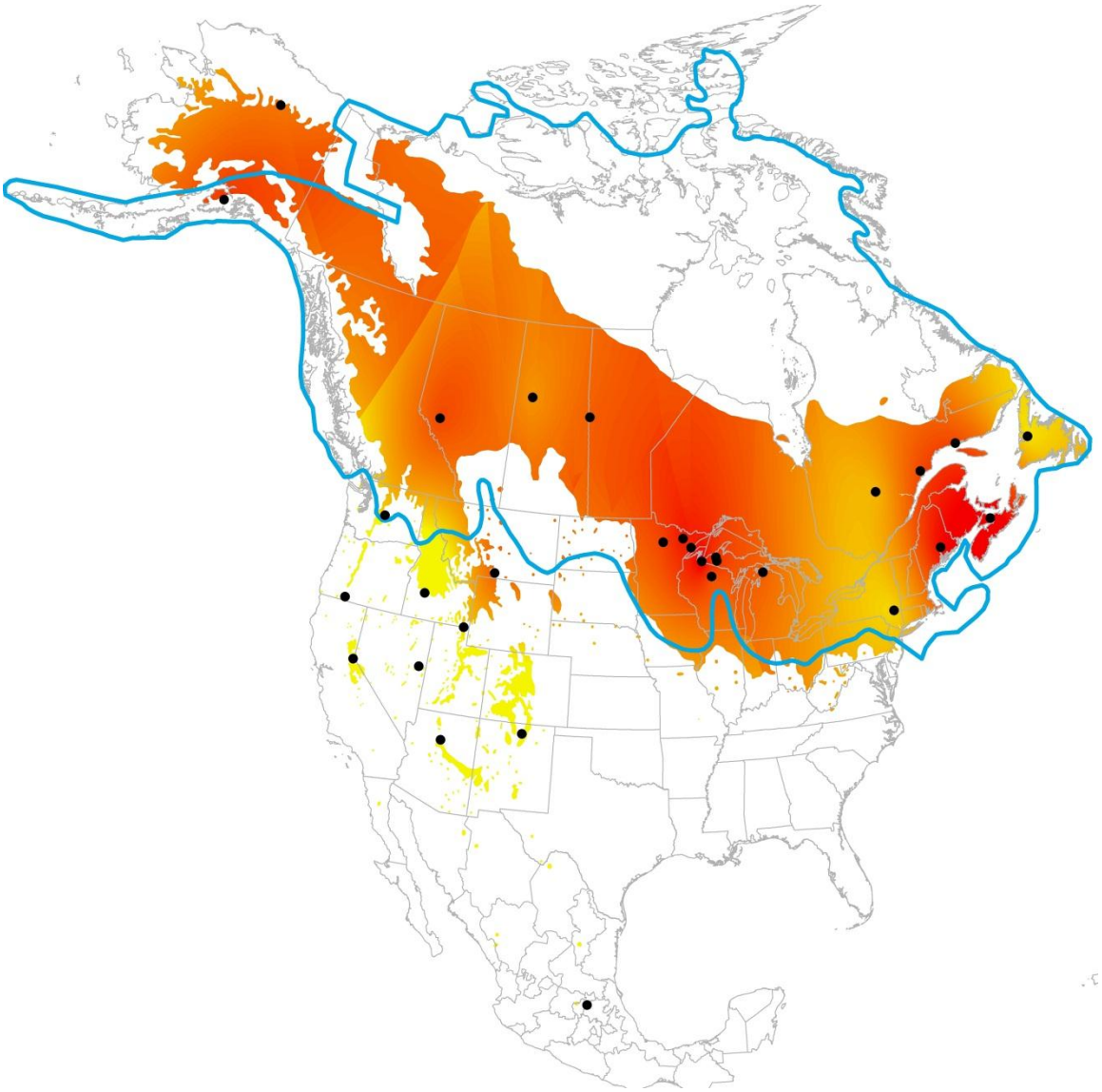

Figure 8 Interpolation of allelic richness across the range of *P. tremuloides* based on average allelic richness values at each sampling site. Allelic richness of sites across the range, varied from 3.337 (MXQ) to 6.832 (WNC). Red represents areas of higher allelic richness, yellow represents areas of lower allelic richness. Outline of last glacial maximum shown in blue (Ray, 2001).

Table 3 Summary statistics for each cluster and sampling site. Allelic richness (AR), observed heterozygosity (H_0), expected heterozygosity (H_E), inbreeding coefficients (F_{IS}). Cluster-specific values were calculated regionally rather than within-site averages.

Cluster/Site	AR	H_0	H_E	F_{IS}
Southwestern cluster	14.592250	0.65639	0.80502	0.185
SWS cluster	11.424375	0.63511	0.76374	0.169
AZ	4.897500	0.57174	0.70466	0.190
MXQ	3.337125	0.68437	0.61267	-0.123
NMT	5.303000	0.74602	0.77025	0.033
NVW	5.070000	0.70341	0.71807	0.022
USF	6.007375	0.64605	0.76814	0.163
SWN cluster	10.454875	0.69024	0.73235	0.058
BNF	5.205875	0.71875	0.72468	0.008
CANV	4.495500	0.71487	0.67939	-0.054
KFO	4.585125	0.59535	0.64378	0.078
WWA	6.004750	0.71117	0.77893	0.091
Northern cluster	16.994500	0.77097	0.78875	0.023
AKCF	6.291625	0.72891	0.75971	0.041
AKK	6.515250	0.75276	0.79715	0.057
BCQ	6.445875	0.77778	0.76241	-0.021
CMNY	5.956625	0.79728	0.77788	-0.027
DLN	5.917375	0.74367	0.77117	0.037
FLFL	6.374875	0.72326	0.77607	0.070
HIN	6.450500	0.82500	0.78341	-0.057
HSPQ	6.451000	0.75340	0.77774	0.032
LALO	6.185750	0.77007	0.77834	0.011
MI	6.473375	0.77538	0.76003	-0.021
MN	6.459500	0.74413	0.78564	0.054
MNL	6.550750	0.78590	0.78282	-0.004
MNSL	6.628250	0.75291	0.77644	0.031
NMI	6.367250	0.79000	0.79082	0.001
POTR	6.357750	0.79104	0.77016	-0.028
SFQ	6.065625	0.73098	0.75939	0.038
UME	6.676750	0.85040	0.79153	-0.079
WII	6.761750	0.75880	0.79497	0.046
WIMI	6.542250	0.77306	0.77840	0.007
WIO	6.416125	0.79018	0.77792	-0.016
WNC	6.831625	0.81798	0.80053	-0.022

ranged from 0.789 (northern cluster) to 0.805 (southwestern cluster) (Table 3). Lastly, levels of inbreeding within sampling sites were low with an average F_{IS} of 0.019 (Table 3).

DISCUSSION

Range-wide Population Structure

My results indicate the existence of two genetically and geographically distinct clusters across the range of aspen. The geographic division of the southwestern and northern clusters reveals the potential for two distinct ecoptypes on the landscape; the southwestern cluster locally adapted to the semiarid climate of the western United States and the northern cluster adapted to the mesic and largely continental climate of Canada and Alaska. Given the extent of the Laurentide and Cordilleran ice sheets during the LGM, much of the current range of *P. tremuloides* was glaciated until at least the end of the Pleistocene epoch. Consequently, the vast majority of aspen's current distribution is the result of a geologically recent range expansion resulting from postglacial recolonization. The assignment of all of the previously glaciated sites to a single genetically diverse cluster suggests rapid post-glacial recolonization, potentially with widespread lineage mixing following the retreat of the ice sheets with populations in the Southwest contributing little, if at all, to recolonization. Also, the lack of clustering of sites near putative refugia reveals the possibility of recolonization from a single refugium.

Although the major genetic clusters also represented geographic clusters, the POTR sampling site within the greater Yellowstone area (POTR) was an anomaly. This site was consistently assigned to, and strongly affiliated with, the northern cluster in the individual-based assignment testing, contrary to my

expectations based on the geographical location of this site. Average allelic richness was also higher at this site in comparison to the rest of the southwestern sites, and more similar to those observed in northern cluster sites. I speculate that the affiliation of POTR with the northern cluster may seem geographically anomalous because I lack sampling sites throughout Montana and North Dakota that could serve as a logical connection to the northern cluster populations. The affiliation of POTR with the northern cluster is consistent with the finding that the northern and southwestern clusters are separated by the continental divide, at least in the U.S.

My results suggest that the southwestern cluster sites are more isolated than the northern cluster sites (Fig. 6). This finding is consistent with the general pattern of pronounced habitat fragmentation in the southwestern portion of the species range, and may reflect post-glacial habitat contraction. It appears that the northern cluster was established from glacial refugia both north and south of the ice sheets, or potentially a single refugium, but not from a northward expansion of the southwestern cluster. This lack of evidence of northward expansion of the southwestern cluster is also consistent with the finding that relictual populations may not significantly contribute to postglacial recolonizations, which acts to preserve their genetic distinctiveness (Bilton *et al.*, 1998; Petit *et al.*, 2003; Hampe & Petit, 2005). The lack of further structuring within the northern cluster, i.e. clustering of sites near putative glacial refugia, is likely due to the recentness of recolonization coupled with widespread lineage

mixing due to the largely continuous nature of the habitat following retreat of the ice sheets (Petit *et al.*, 2003).

Stable vs. Trailing Edge Dynamics

Species distributions expand, contract, and shift in response to changes in climate and other variables. A warming climate, for example, may result in the shift of a species distribution to higher latitudes, exemplified by some boreal species of North America (Williams *et al.*, 2004). This shifting distribution has both a leading edge, expanding to new territory, and a rear edge (Hampe & Petit, 2005). Two scenarios exist to describe populations at the rear edge of a distribution; these populations may be characterized as “stable” or “trailing” edges. Under the stable edge scenario, these rear edge populations persist throughout climate oscillations, usually in areas of disparate topography which increases the chances of matching suitable climatic profiles through slight altitudinal shifts (Comps *et al.*, 2001; Tzedakis *et al.*, 2002). As mentioned previously, stable edge populations are expected to have low levels of within-population genetic diversity (Castric & Bernatchez, 2003; Petit *et al.*, 2003; Chang *et al.*, 2004) but increased regional genetic diversity (Comps *et al.*, 2001; Castric & Bernatchez, 2003; Hampe *et al.*, 2003; Petit *et al.*, 2003; Martin & McKay, 2004). Conversely, under a trailing edge scenario, rear edge populations are remnants of the previous distribution left behind as the entire range shifts to more favorable conditions. In this case, rear edge populations are at risk of complete extirpation as conditions become less favorable through time and the

entire distribution shifts; examples in North America include certain spruce species as well as *Pinus banksiana* and *Pinus resinosa* (Jackson *et al.*, 1997; Jackson *et al.*, 2000; Williams *et al.*, 2004). Trailing edge populations are expected to have within-population genetic diversity similar to that observed in the rest of the range, or perhaps markedly lower due to founder effects and/or increasing isolation and shrinking population size. Additionally, trailing-edge populations are expected to lack deep divergence among populations due to these populations being only slightly older than those in the rest of the range (Hampe & Petit, 2005). A high conservation value would be placed on stable edge populations to preserve the elevated levels of genetic diversity and divergent lineages present (Petit *et al.*, 1998). Alternatively, trailing edge populations should be conserved for their inherent value on the landscape. In the case of *P. tremuloides*, such populations should be preserved for the biodiversity harbored in western aspen habitat (Stohlgren *et al.*, 1997a; Stohlgren *et al.*, 1997b; Mills *et al.*, 2000; Rumble *et al.*, 2001; Simonson *et al.*, 2001) as well as the cultural and aesthetic value placed on western aspen.

My results indicate that the southwestern cluster represents a distinct lineage, and a potentially distinct ecotype, which is consistent with these populations being part of a stable edge throughout multiple climate oscillations. This southwestern portion of the range is home to multiple mountain ranges and basins which would have provided ample opportunity for altitudinal shifts in response to changes in climate throughout past glacial and interglacial episodes.

Allelic richness within sites in the southwestern cluster were lower than within-site levels in the rest of the range but allelic richness of the southwestern cluster, as a whole, was near that of the northern cluster which reveals the low within-site genetic diversity and increased regional genetic diversity characteristic of a stable edge. I also found no evidence of northward expansion of the southwestern cluster which is common of relict stable edge populations (Bilton *et al.*, 1998; Petit *et al.*, 2003).

The idea that southwestern aspen may represent a distinct ecotype is not new. Axelrod (1941) compared leaf fossil morphology and habitat types of two potential precursor species of *P. tremuloides* existing during the Miocene and Pliocene, *P. lindgreni* and *P. pliotremuloides*. He suggested that, when compared to modern *P. tremuloides*, these two taxa should be recognized as distinct ecotypes. Leaves of *P. lindgreni* more closely resemble leaves of modern aspen growing in more mesic environments and leaves of *P. pliotremuloides* are indistinguishable from leaves of modern aspen growing in the more arid portion of the range (southwestern). From this comparison of leaf morphology and habitat types between these two pre-historic species, Axelrod (1941) concluded that, given the enormous and varied distribution of contemporary *P. tremuloides*, it seems “highly probable” that multiple ecotypes be present. Further, Barnes (1975) described clinal variation in aspen leaf morphology from south to north in the western United States and Canada. Throughout long-term persistence in the western United States and Mexico, selection would be expected to favor

adaptation to local conditions in these populations rather than increased dispersal and generalism (Dynesius & Jansson, 2000). This expectation of local adaptation, coupled with reduced gene flow due to isolation in the Southwest, can lead to the development of exceptionally well-defined ecotypes (Hampe & Bairlein, 2000; Castric & Bernatchez, 2003; Perez-Tris *et al.*, 2004; Hampe & Petit, 2005). Given that the populations that make up the southwestern cluster likely persisted throughout multiple climate oscillations in the western United States and Mexico, the more arid environment of this southwestern portion of the range, and the greater degree of isolation among populations in this cluster, I find it improbable that these populations do not represent a distinct ecotype.

Sub-structuring Within the Southwestern Cluster

Geographic and genetic subdivision within the southwestern cluster closely resembles that demonstrated for ponderosa pine (*Pinus ponderosa* Laws Pinaceae) (Latta & Mitton, 1999). Latta and Mitton (1999) describe two historically separate groups of ponderosa pine east and west of the Great Basin, currently meeting in a transition zone in west-central Montana, and further classify these two groups as distinct varieties of ponderosa pine (*P. p. ponderosa* and *P. p. scopulorum*). They hypothesized that these two groups were separated during the LGM and the current transition zone represents a point of secondary contact in which case unique haplotypes would be expected on either side of the transition zone with a potential mixture of haplotypes within the transition zone. These researchers demonstrated the existence of primarily east and west

haplotypes using length variation in mtDNA and cpDNA markers, with a mixture of haplotypes in the transition zone. My structuring results, based on microsatellite allele frequencies, show a similar result for *P. tremuloides* (Fig. 4). Due to a lower number of sampling sites in this area, though, it is difficult to clearly delineate a transition zone between the SWS and SWN clusters. The USF site does show a greater level of mixed assignment between the two subclusters than the other sites and is nearly longitudinally identical in location to the transition zone in west-central Montana for ponderosa pine. I propose that the SWS and SWN clusters represent unique gene pools but are of the same ecotype described above (southwestern cluster). These gene pools may have been separated and reunited, through a transition zone, multiple times through multiple glacial cycles. The varied topography that these populations inhabit could lead to subdivision, with gene flow between the two subclusters only occurring during glaciation or warming events that would open new habitable areas (Hewitt, 2000).

Centers for Diversity and Potential Glacial Refugia

Based on the current distribution of *P. tremuloides*, there are several potential refugial sites which may have harbored aspen during the LGM and served as points of origin for postglacial recolonization (Fig. 1). One would expect relatively higher levels of allelic richness at such refugial sites since they are sources for recolonization of the surrounding region following retreat of the ice sheets. One would also expect that populations near glacial refugia will form

clusters due to the presence of unique alleles representing divergent lineages. Due to the survival of *P. balsamifera* in the Beringia refugium during the LGM (Brubaker *et al.*, 2005), I identified this site as a likely refugium for *P. tremuloides*. I also expected to find an area of high diversity near the Great Lakes as this area is close to a putative refugium, the “driftless area” (Rowe *et al.*, 2004), at the southern edge of the range.

Assessment of allelic richness across the range provided insight into potential glacial refugia for *P. tremuloides* and allows inference of postglacial recolonization patterns. The ten sites with the highest average allelic richness values (WNC, WII, UME, MNSL, MNL, WIMI, AKK, MI, MN, and HSPQ) lie in close proximity to proposed glacial refugia. Three of these sites (WNC, UME, and HSPQ) are in the northeastern United States and Canada near the proposed Grand Banks refugium (Holder *et al.*, 1999), an area off of the present day coast of Newfoundland which remained ice free and above sea level during the LGM. Interestingly, however, our Newfoundland sampling site (DLN) displayed below average allelic richness. A possible explanation for this may be shrinking population size coupled with a lack of gene flow due to isolation through time. Currently, Newfoundland is separated from the Labrador Peninsula by the Gulf of St. Lawrence and the Straight of Belle Isle which together may act as a barrier to gene flow, effectively isolating aspen populations on the island from those on the mainland. Another six sites (WII, MNSL, MNL, WIMI, MI, and MN) from the ten most diverse sites are from the Great Lakes area, near an area known as the

"driftless area" (Rowe *et al.*, 2004). The term "driftless area" comes from the lack of glacial drift and evidence of glaciation. This was a large area that remained ice free during the LGM despite the advancement of the Laurentide ice sheet. The only remaining site I have yet to mention from our ten most diverse sites is located in Kenai, Alaska (AKK), a peninsula on the southern portion of the state. This sampling site lies near the putative Beringia refugium, which has not only been shown to harbor other *Populus* species during the LGM (Brubaker *et al.*, 2005), but other taxa as well (Holder *et al.*, 1999; Tremblay & Schoen, 1999; Anderson *et al.*, 2006; Loehr *et al.*, 2006). Our other sampling site in Alaska (AKCF) displayed only slightly elevated allelic richness relative to other sites near refugia. I expected a higher level of allelic richness at this site due to its location near the Beringia refugium. This site lies at the northern limit of aspen's current range though, which may explain the slightly lower allelic richness versus the AKK site (i.e. possible founder effects at an expanding edge) (Hewitt, 2000).

It is also worth noting that another site displaying an elevated level of allelic richness (HIN) is in Alberta, Canada in close proximity to the controversial "ice-free corridor" (Mandryk *et al.*, 2001). It is currently generally accepted that the corridor did not extend to the southern glacial limit during the LGM due to coalescence of the ice sheets in southern Alberta but geological evidence lends support to an ice-free area north of southern glacial limit. There is also evidence of this area serving as a refugium for mountain sheep based on mtDNA (Loehr *et al.*, 2006). Increased diversity in this area lends support to the persistence of

aspen during the LGM but may also be explained by mixing of lineages during postglacial recolonization (Petit *et al.*, 2003). As the range expanded north from south of the glacial margins following retreat of the ice sheets, it is possible these southern groups coalesced with groups expanding southward from Beringia as the corridor opened and widened.

Interestingly, sites that are indicated as glacial refugia for *P. tremuloides* during the LGM, by elevated levels of allelic richness, did not form distinct clusters based on allele frequencies. A possible explanation for this may be the lack of divergent local adaptive differentiation among populations that survived the LGM in refugia both north and south of the ice sheets, followed by rapid recolonization and lineage mixing among several glacial refugial sources. If *P. tremuloides* expanded to areas of suitable climate and habitat following retreat of the ice sheets, populations that persisted in refugia would have been pre-adapted for such conditions (Davis & Shaw, 2001). Furthermore, Petit (2003) explains that populations near glacial refugia should be highly divergent due to prolonged periods of isolation, particularly in those populations that did not contribute to postglacial recolonization. In the case of *P. tremuloides*, the lack of evidence of deep divergence at refugial sites outside of the southwestern portion of the range may be explained by the rapid, multi-directional nature of the expansion and widespread lineage mixing. Finally, my sparse sampling strategy made it difficult to distinguish between hypotheses about various glacial refugia. A more powerful way to identify refugia and elucidate patterns of postglacial

recolonization would be to sample the entire range in a more dispersed fashion, rather than within pre-defined populations, in a way that includes the various scales being investigated (Guillot *et al.*, 2009; Schwartz & McKelvey, 2009).

Conclusions

From my multilocus data set of 794 individuals from 30 sampling sites across the range of *P. tremuloides*, I identified distinct population structuring at the continental scale. Two major clusters were discovered: a northern cluster composed almost entirely of sites within the boundaries of the LGM and a southwestern cluster composed of sites representing the fragmented southwestern portion of the range. Within this southwestern cluster, there is further population structure with two more subclusters identified, potentially representing two distinct gene pools within the southwestern cluster.

I identified the southwestern portion of aspen's range as a stable edge of genetic diversity with low levels of within-site genetic diversity but increased regional genetic diversity. I suggest that these southwestern populations represent a distinct ecotype based on the potential for local adaptation to different climatic regimes and minimal gene flow between the southwestern and northern clusters through time. The populations of this stable edge likely persisted through multiple climate oscillations by utilizing the varied topography of the southwestern portion of the range to match suitable climate through altitudinal shifts. With sudden dieback affecting stands throughout much of the species range (Frey *et al.*, 2004), aspen in the western United States may be

particularly vulnerable, with significant decline over a 48-year period documented (Di Orio *et al.*, 2005) along with the projected loss of up to 94% of area occupied by its contemporary climate profile within this century (Rehfeldt *et al.*, 2009). Although populations of the Southwest have survived past climate shifts, current rates of change may exceed the adaptive and migratory capacity of many species (Tett *et al.*, 1999). It is important to monitor these rear-edge populations as environmental conditions change to distinguish climate effects from other factors such as increasing fragmentation and competition with other species (Hampe & Petit, 2005). Also, though within-site genetic diversity is low in the Southwest, it is important to identify and preserve as many local populations as possible in order to maintain the divergent lineages that contribute to the regional genetic diversity (Petit *et al.*, 1998; Hampe & Petit, 2005). Restoration efforts such as controlled burning and management to limit ungulate browsing should be employed where necessary to maintain the health of these populations. Though assisted migration may represent a powerful conservation effort for aspen in the northern portion of the range in the face of a warming climate (Gray *et al.*, 2011), southwestern aspen are at the southern extent of the range with no "preadapted" populations to draw from (Davis & Shaw, 2001).

The identification of genetic diversity hot spots sheds light on potential glacial refugia as well as postglacial recolonization patterns. I identified three potential refugia as sources of postglacial recolonization for the northern clade, with a fourth area representing either a refugium or a postglacial mixing of

lineages, although my inferences about the geography of these potential refugia were limited by my sampling design. The three likely refugial sites are Beringia in Alaska, the “driftless region” near the Great Lakes, and the Grand Banks refugium off the present day coast of Newfoundland. The fourth area, in Alberta, Canada, may represent a site where *P. tremuloides* persisted during the LGM or a mixing of lineages from southern glacial margins moving north with lineages from Beringia moving south. In any case, these diversity hot spots are of high conservation value, because by concentrating on these sites it may be possible to conserve a large portion of the genetic diversity present throughout much of the range.

REFERENCES

Anderson, L.L., Hu, F.S., Nelson, D.M., Petit, R.J. & Paige, K.N. (2006) Ice-age endurance: DNA evidence of a white spruce refugium in Alaska. *Proceedings of the National Academy of Sciences USA*, **103**, 12447-12450.

Avise, J.C. (2000) *Phylogeography: The history and formation of species*. Harvard University Press, Cambridge, Massachusettes.

Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A. & Saunders, N.C. (1987) Intraspecific Phylogeography - the mitochondrial-DNA bridge between population-genetics and systematics. *Annual Review of Ecology and Systematics*, **18**, 489-522.

Axelrod, D.I. (1941) The concept of ecospecies in tertiary paleobotany. *Proceedings of the National Academy of Sciences USA*, **27**, 545-51.

Barnes, B.V. (1975) Phenotypic variation of trembling aspen in western North America. *Forest Science*, **21**, 319-328.

Beatty, G.E. & Provan, J. (2010) Refugial persistence and postglacial recolonization of North America by the cold-tolerant herbaceous plant Orthilia secunda. *Molecular Ecology*, **19**, 5009-5021.

Beheregaray, L.B. (2008) Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology*, **17**, 3754-3774.

Bilton, D.T., Mirol, P.M., Mascheretti, S., Fredga, K., Zima, J. & Searle, J.B. (1998) Mediterranean Europe as an area of endemism for small mammals rather than a source for northwards postglacial colonization. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **265**, 1219-1226.

Brubaker, L.B., Anderson, P.M., Edwards, M.E. & Lozhkin, A.V. (2005) Beringia as a glacial refugium for boreal trees and shrubs: new perspectives from mapped pollen data. *Journal of Biogeography*, **32**, 833-848.

Brundu, G., Lupi, R., Zapelli, I., Fossati, T., Patrignani, G., Camarda, I., Sala, F. & Castiglione, S. (2008) The origin of clonal diversity and structure of Populus alba in Sardinia: evidence from nuclear and plastid microsatellite markers. *Annals of Botany*, **102**, 997-1006.

Brunsfeld, S.J. & Sullivan, J. (2005) A multi-compartmented glacial refugium in the northern Rocky Mountains: evidence from the phylogeography of Cardamine constancei (Brassicaceae). *Conservation Genetics*, **6**, 895-904.

Castric, V. & Bernatchez, L. (2003) The rise and fall of isolation by distance in the anadromous brook charr (Salvelinus fontinalis Mitchill). *Genetics*, **163**, 983-996.

Chang, C.S., Kim, H., Park, T.Y. & Maunder, M. (2004) Low levels of genetic variation among southern peripheral populations of the threatened herb,

Leontice microrhyncha (Berberidaceae) in Korea. *Biological Conservation*, **119**, 387-396.

Clark, P.U. & Mix, A.C. (2002) Ice sheets and sea level of the Last Glacial Maximum. *Quaternary Science Reviews*, **21**, 1-7.

Comps, B., Gomory, D., Letouzey, J., Thiebaut, B. & Petit, R.J. (2001) Diverging trends between heterozygosity and allelic richness during postglacial colonization in the European beech. *Genetics*, **157**, 389-397.

Davis, M.B. & Shaw, R.G. (2001) Range shifts and adaptive responses to Quaternary climate change. *Science*, **292**, 673-679.

de Lafontaine, G., Turgeon, J. & Payette, S. (2010) Phylogeography of white spruce (Picea glauca) in eastern North America reveals contrasting ecological trajectories. *Journal of Biogeography*, **37**, 741-751.

Di Orio, A.P., Callas, R. & Schaefer, R.J. (2005) Forty-eight year decline and fragmentation of aspen (Populus tremuloides) in the South Warner Mountains of California. *Forest Ecology and Management*, **206**, 307-313.

Dyke, A.S., Andrews, J.T., Clark, P.U., England, J.H., Miller, G.H., Shaw, J. & Veillette, J.J. (2002) The Laurentide and Innuitian ice sheets during the Last Glacial Maximum. *Quaternary Science Reviews*, **21**, 9-31.

Dynesius, M. & Jansson, R. (2000) Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *Proc Natl Acad Sci U S A*, **97**, 9115-9120.

Earl, D.A. & vonHoldt, B.M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359-361.

Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611-2620.

Excoffier, L. & Lischer, H.E.L. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564-567.

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

Fechner, G.H. & Barrows, J.S. (1976) *Aspen stands as wildfire fuel breaks*. USDA Forest Service Rocky Mountain Forest and Range Experiment Station, Fort Collins, CO.

Francois, O. & Durand, E. (2010) Spatially explicit Bayesian clustering models in population genetics. *Molecular Ecology Resources*, **10**, 773-784.

Frey, B.R., Lieffers, V.J., Hogg, E.H. & Landhausser, S.M. (2004) Predicting landscape patterns of aspen dieback: mechanisms and knowledge gaps. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, **34**, 1379-1390.

Fussi, B., Lexer, C. & Heinze, B. (2010) Phylogeography of Populus alba (L.) and Populus tremula (L.) in Central Europe: secondary contact and hybridisation during recolonisation from disconnected refugia. *Tree Genetics & Genomes*, **6**, 439-450.

Goudet, J. (2002) *FSTAT, version 2.9.3.2, A program to estimate and test gene diversities and fixation indices.*

Gray, L.K., Gylander, T., Mbogga, M.S., Chen, P.Y. & Hamann, A. (2011) Assisted migration to address climate change: recommendations for aspen reforestation in western Canada. *Ecological Applications*, **21**, 1591-1603.

Guillot, G., Leblois, R., Coulon, A. & Frantz, A.C. (2009) Statistical methods in spatial genetics. *Molecular Ecology*, **18**, 4734-4756.

Hampe, A. & Bairlein, F. (2000) Modified dispersal-related traits in disjunct populations of bird-dispersed Frangula alnus (Rhamnaceae): a result of its Quaternary distribution shifts? *Ecography*, **23**, 603-613.

Hampe, A. & Petit, R.J. (2005) Conserving biodiversity under climate change: the rear edge matters. *Ecology Letters*, **8**, 461-467.

Hampe, A., Arroyo, J., Jordano, P. & Petit, R.J. (2003) Rangewide phylogeography of a bird-dispersed Eurasian shrub: contrasting Mediterranean and temperate glacial refugia. *Molecular Ecology*, **12**, 3415-3426.

Hewitt, G. (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907-913.

Hewitt, G.M. (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **359**, 183-195.

Holder, K., Montgomerie, R. & Friesen, V.L. (1999) A test of the glacial refugium hypothesis using patterns of mitochondrial and nuclear dna sequence variation in rock ptarmigan (Lagopus mutus). *Evolution*, **53**, 1936-1950.

Ingvarsson, P.K. (2008) Multilocus patterns of nucleotide polymorphism and the demographic history of Populus tremula. *Genetics*, **180**, 329-340.

Jackson, S.T., Overpeck, J.T., Webb, T., Keattch, S.E. & Anderson, K.H. (1997) Mapped plant-macrofossil and pollen records of late quaternary vegetation change in eastern North America. *Quaternary Science Reviews*, **16**, 1-70.

Jackson, S.T., Webb, R.S., Anderson, K.H., Overpeck, J.T., Webb, T., Williams, J.W. & Hansen, B.C.S. (2000) Vegetation and environment in Eastern North America during the Last Glacial Maximum. *Quaternary Science Reviews*, **19**, 489-508.

Jakobsson, M. & Rosenberg, N.A. (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801-1806.

Keller, S.R., Olson, M.S., Silim, S., Schroeder, W. & Tiffin, P. (2010) Genomic diversity, population structure, and migration following rapid range

expansion in the Balsam Poplar, Populus balsamifera. *Molecular Ecology*, **19**, 1212-1226.

Kemperman, J.A. & Barnes, B.V. (1976) Clone size in American aspens. *Canadian Journal of Botany-Revue Canadienne De Botanique*, **54**, 2603-2607.

Latta, R.G. & Mitton, J.B. (1999) Historical separation and present gene flow through a zone of secondary contact in ponderosa pine. *Evolution*, **53**, 769-776.

Little, E.L., Jr. (1971) *Checklist of United States trees (native and naturalized)*. US Department of Agriculture, Washington D.C.

Loehr, J., Worley, K., Grapputo, A., Carey, J., Veitch, A. & Coltman, D.W. (2006) Evidence for cryptic glacial refugia from North American mountain sheep mitochondrial DNA. *Journal of Evolutionary Biology*, **19**, 419-430.

Mandryk, C.A.S., Josenhans, H., Fedje, D.W. & Mathewes, R.W. (2001) Late Quaternary paleoenvironments of Northwestern North America: implications for inland versus coastal migration routes. *Quaternary Science Reviews*, **20**, 301-314.

Marr, K.L., Allen, G.A. & Hebda, R.J. (2008) Refugia in the Cordilleran ice sheet of western North America: chloroplast DNA diversity in the Arctic-alpine plant Oxyria digyna. *Journal of Biogeography*, **35**, 1323-1334.

Martin, P.R. & McKay, J.K. (2004) Latitudinal variation in genetic divergence of populations and the potential for future speciation. *Evolution*, **58**, 938-945.

Millar, C.I. (1996) *Tertiary vegetation history*. Institute of Forest Genetics, US Forest Service, Albany, California.

Mills, T.R., Rumble, M.A. & Flake, L.D. (2000) Habitat of birds in ponderosa pine and aspen/birch forest in the Black Hills, South Dakota. *Journal of Field Ornithology*, **71**, 187-206.

Mitton, J.B. & Grant, M.C. (1996) Genetic variation and the natural history of quaking aspen. *Bioscience*, **46**, 25-31.

Mock, K.E., Rowe, C.A., Hooten, M.B., Dewoody, J. & Hipkins, V.D. (2008) Clonal dynamics in western North American aspen (Populus tremuloides). *Molecular Ecology*, **17**, 4827-4844.

Peakall, R. & Smouse, P.E. (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288-295.

Perala, D.A. (1990) *Silvics of North America: Volume 2, Hardwoods. Agriculture Handbook 654*. US Department of Agriculture, Forest Service, Washington D.C.

Perez-Tris, J., Bensch, S., Carbonell, R., Helbig, A.J. & Telleria, J.L. (2004) Historical diversification of migration patterns in a passerine bird. *Evolution*, **58**, 1819-1832.

Petit, R.J., El Mousadik, A. & Pons, O. (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, **12**, 844-855.

Petit, R.J., Aguinagalde, I., de Beaulieu, J.L., Bittkau, C., Brewer, S., Cheddadi, R., Ennos, R., Fineschi, S., Grivet, D., Lascoux, M., Mohanty, A., Muller-Starck, G.M., Demesure-Musch, B., Palme, A., Martin, J.P., Rendell, S. & Vendramin, G.G. (2003) Glacial refugia: Hotspots but not melting pots of genetic diversity. *Science*, **300**, 1563-1565.

Pielou, E.C. (1991) *After the Ice Age: the return of life to glaciated North America*. University of Chicago Press, Chicago.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-959.

Ray, N., Adams, J. M., (2001) A GIS-based Vegetation Map of the World at the Last Glacial Maximum (25,000-15,000 BP). *Internet Archeology*, **11**

Rebernig, C.A., Schneeweiss, G.M., Bardy, K.E., Schonswetter, P., Villasenor, J.L., Obermayer, R., Stuessy, T.F. & Weiss-Schneeweiss, H. (2010) Multiple Pleistocene refugia and Holocene range expansion of an abundant southwestern American desert plant species (Melampodium leucanthum, Asteraceae). *Molecular Ecology*, **19**, 3421-3443.

Rehfeldt, G.E., Ferguson, D.E. & Crookston, N.L. (2009) Aspen, climate, and sudden decline in western USA. *Forest Ecology and Management*, **258**, 2353-2364.

Rosenberg, N.A. (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137-138.

Rowe, K.C., Heske, E.J., Brown, P.W. & Paige, K.N. (2004) Surviving the ice: Northern refugia and postglacial colonization. *Proceedings of the National Academy of Sciences USA*, **101**, 10355-10359.

Rumble, M.A., Flake, L.D., Mills, T.R. & Dykstra, B.L. (2001) Do pine trees in aspen stands increase bird diversity? *Sustaining Aspen in Western Landscapes: Symposium Proceedings: 13-15 June 2000, Grand Junction, Colorado (ed by Sheppard WD et al.)*, pp. 185-191.

Schwartz, M.K. & McKelvey, K.S. (2009) Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conservation Genetics*, **10**, 441-452.

Shaw, J., Lickey, E.B., Schilling, E.E. & Small, R.L. (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. *American Journal of Botany*, **94**, 275-288.

Simonson, S.E., Opler, P.A., Stohlgren, T.J. & Chong, G.W. (2001) Rapid assessment of butterfly diversity in a montane landscape. *Biodiversity and Conservation*, **10**, 1369-1386.

Smulders, M.J.M., Van Der Schoot, J., Arens, P. & Vosman, B. (2001) Trinucleotide repeat microsatellite markers for black poplar (Populus nigra L.). *Molecular Ecology Notes*, **1**, 188-190.

Stohlgren, T.J., Chong, G.W., Kalkhan, M.A. & Schell, L.D. (1997a) Multiscale sampling of plant diversity: Effects of minimum mapping unit size. *Ecological Applications*, **7**, 1064-1074.

Stohlgren, T.J., Coughenour, M.B., Chong, G.W., Binkley, D., Kalkhan, M.A., Schell, L.D., Buckley, D.J. & Berry, J.K. (1997b) Landscape analysis of plant diversity. *Landscape Ecology*, **12**, 155-170.

Tett, S.F.B., Stott, P.A., Allen, M.R., Ingram, W.J. & Mitchell, J.F.B. (1999) Causes of twentieth-century temperature change near the Earth's surface. *Nature*, **399**, 569-572.

Tomimatsu, H., Kephart, S.R. & Vellend, M. (2009) Phylogeography of Camassia quamash in western North America: postglacial colonization and transport by indigenous peoples. *Molecular Ecology*, **18**, 3918-3928.

Tremblay, N.O. & Schoen, D.J. (1999) Molecular phylogeography of Dryas integrifolia: glacial refugia and postglacial recolonization. *Molecular Ecology*, **8**, 1187-1198.

Tuskan, G.A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R.R., Bhalerao, R.P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapman, J., Chen, G.L., Cooper, D., Coutinho, P.M., Couturier, J., Covert, S., Cronk, Q., Cunningham, R., Davis, J., Degroeve, S., Dejardin, A., Depamphilis, C., Detter, J., Dirks, B., Dubchak, I., Duplessis, S.,

Ehlting, J., Ellis, B., Gendler, K., Goodstein, D., Gribskov, M., Grimwood, J., Groover, A., Gunter, L., Hamberger, B., Heinze, B., Helariutta, Y., Henrissat, B., Holligan, D., Holt, R., Huang, W., Islam-Faridi, N., Jones, S., Jones-Rhoades, M., Jorgensen, R., Joshi, C., Kangasjarvi, J., Karlsson, J., Kelleher, C., Kirkpatrick, R., Kirst, M., Kohler, A., Kalluri, U., Larimer, F., Leebens-Mack, J., Leple, J.C., Locascio, P., Lou, Y., Lucas, S., Martin, F., Montanini, B., Napoli, C., Nelson, D.R., Nelson, C., Nieminen, K., Nilsson, O., Pereda, V., Peter, G., Philippe, R., Pilate, G., Poliakov, A., Razumovskaya, J., Richardson, P., Rinaldi, C., Ritland, K., Rouze, P., Ryaboy, D., Schmutz, J., Schrader, J., Segerman, B., Shin, H., Siddiqui, A., Sterky, F., Terry, A., Tsai, C.J., Uberbacher, E., Unneberg, P., Vahala, J., Wall, K., Wessler, S., Yang, G., Yin, T., Douglas, C., Marra, M., Sandberg, G., Van de Peer, Y. & Rokhsar, D. (2006) The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). *Science*, **313**, 1596-1604.

Tzedakis, P.C., Lawson, I.T., Frogley, M.R., Hewitt, G.M. & Preece, R.C. (2002) Buffered tree population changes in a quaternary refugium: evolutionary implications. *Science*, **297**, 2044-2047.

Williams, J.W., Shuman, B.N., Webb, T., Bartlein, P.J. & Leduc, P.L. (2004) Late-quaternary vegetation dynamics in north america: scaling from taxa to biomes. *Ecological Monographs*, **74**, 309-334.

Wullschleger, S.D., Jansson, S. & Taylor, G. (2002) Genomics and forest biology: Populus emerges as the perennial favorite. *Plant Cell*, **14**, 2651-2655.