ORIGINAL ARTICLE

Continental-scale assessment of genetic diversity and population structure in quaking aspen (*Populus tremuloides*)

Colin M. Callahan[1], Carol A. Rowe[1], Ronald J. Ryel[1], John D. Shaw[2], Michael D. Madritch[3] and Karen E. Mock[1]*

[1]*Department of Wildland Resources and the Ecology Center, Utah State University, Logan, UT, 84322-5230, USA,* [2]*USDA Forest Service, Rocky Mountain Research Station, Ogden, UT, 84401, USA,* [3]*Department of Biology, Appalachian State University, Boone, NC, 28608, USA*

ABSTRACT

Aim Quaking aspen (*Populus tremuloides*) has the largest natural distribution of any tree native to North America. The primary objectives of this study were to characterize range-wide genetic diversity and genetic structuring in quaking aspen, and to assess the influence of glacial history and rear-edge dynamics.

Location North America.

Methods Using a sample set representing the full longitudinal and latitudinal extent of the species' distribution, we examined geographical patterns of genetic diversity and structuring using 8 nuclear microsatellite loci in 794 individuals from 30 sampling sites.

Results Two major genetic clusters were identified across the range: a south-western cluster and a northern cluster. The south-western cluster, which included two subclusters, was bounded approximately by the Continental Divide to the east and the southern extent of the ice sheet at the Last Glacial Maximum to the north. Subclusters were not detected in the northern cluster, despite its continent-wide distribution. Genetic distance was significantly correlated with geographical distance in the south-western but not the northern cluster, and allelic richness was significantly lower in south-western sampling sites compared with northern sampling sites. Population structuring was low overall, but elevated in the south-western cluster.

Main conclusions Aspen populations in the south-western portion of the range are consistent with expectations for a historically stable edge, with low within-population diversity, significant geographical population structuring, and little evidence of northward expansion. Structuring within the south-western cluster may result from distinct gene pools separated during the Pleistocene and reunited following glacial retreat, similar to patterns found in other forest tree species in the western USA. In aspen, populations in the south-western portion of the species range are thought to be at particularly high risk of mortality with climate change. Our findings suggest that these same populations may be disproportionately valuable in terms of both evolutionary potential and conservation value.

*Correspondence: Karen E. Mock, Department of Wildland Resources, Utah State University, Logan, UT 84322-5230, USA.
E-mail: karen.mock@usu.edu

INTRODUCTION

Quaking aspen (*Populus tremuloides* Michx.; hereafter '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, occupying a broad range of ecosystems and elevations (Little, 1971; Perala, 1990). In the North American boreal forest, aspen is the dominant deciduous tree species by biomass (Hogg *et al.*, 2002) and is a commercially important source of wood fibre.

Aspen is prized for its aesthetic and cultural value, its value as forage for wildlife and livestock, and also for its importance to biodiversity, as many species rely on aspen habitat. In the Intermountain West of North America, aspen stands are associated with high levels of biodiversity for many taxonomic groups, including plants, birds and butterflies (Stohlgren *et al.*, 1997a,b; Mills *et al.*, 2000; Rumble *et al.*, 2001; Simonson *et al.*, 2001). In the western USA, aspen also functions as a firebreak (Fechner & Barrows, 1976).

Aspen is an attractive candidate as a model system for molecular adaptation (Ouborg *et al.*, 2010) because of its broad geographical range and ecological amplitude, along with the availability of a reference genome in the congener *Populus trichocarpa* (Tuskan *et al.*, 2006). Inferences of molecular adaptation, however, are limited by a lack of information about range-wide neutral variation in this species, because neutral variation can serve as a framework against which adaptive variation can be detected. Phylogeographical patterns have been described for other *Populus* species, namely *P. alba* and *P. tremula* in Europe (Brundu *et al.*, 2008; Fussi *et al.*, 2010) and *P. balsamifera* in North America (Breen *et al.*, 2012; Keller *et al.*, 2012), but a range-wide phylogeography of *P. tremuloides* has never been undertaken.

Aspen reproduces both sexually, through seed and pollen adapted for long-distance wind dispersal, and asexually through root sprouts (Barnes, 1966; Mitton & Grant, 1996). The species is notable for its ability to form large clones (genets), particularly in landscapes of the western USA where seed reproduction is considered rare and episodic due to the dry climate (Kemperman & Barnes, 1976; Grant *et al.*, 1992; Elliott & Baker, 2004). In the eastern USA and Canada, seed dispersal is thought to be a major mechanism for persistence and range expansion (Landhäusser & Wein, 1993; Landhäusser *et al.*, 2010).

Aspen has undergone a dramatic range expansion following the Last Glacial Maximum (LGM) 26–19 ka (Fig. 1). The majority of the species' current range was covered by the Cordilleran and Laurentide ice sheets during the Wisconsinan glacial episode (110–10 ka). Post-glacial dynamics at the southern edge of its range are unclear.

When species experience latitudinal shifts during climate oscillations, the rear range edge in a directional expansion may behave as a 'stable edge' or a 'trailing edge' (Hampe & Petit, 2005). Under the stable-edge scenario, populations persist throughout climate oscillations, usually in areas of varied topography, presumably matching suitable climatic profiles over time through elevational shifts (Comps *et al.*, 2001; Tzedakis *et al.*, 2002). Stable-edge populations may or may not contribute to range expansions. Alternatively, under a trailing-edge scenario, the species' entire range will have shifted in response to climate oscillations, with southern populations disappearing during northward shifts and being re-established from northern populations during southward shifts. For aspen, under either scenario, reduced within-population genetic diversity is expected due to genetic drift

and founder effects following isolation and/or shrinking sizes of populations over time in high-elevation habitats (Castric & Bernatchez, 2003; Petit *et al.*, 2003; Chang *et al.*, 2004). Under a stable-edge scenario, increased regional genetic diversity and significant divergence among southern populations are predicted (Comps *et al.*, 2001; Castric & Bernatchez, 2003; Hampe *et al.*, 2003; Petit *et al.*, 2003; Martin & McKay, 2004). In contrast, under a trailing-edge scenario, southern populations are expected to lack deep divergence among populations compared to core populations because they would not have been separated for substantially longer periods of time than populations in the remainder of the range (Hampe & Petit, 2005).

In aspen, rear-edge populations may be disproportionately valuable in terms of evolutionary potential and conservation value, particularly if they represent a stable edge with substantial genetic diversity among populations. This issue is magnified for aspen west of the Continental Divide in the USA (a major north–south mountainous ridge separating Atlantic and Pacific watersheds in North America), where aspen habitat is projected to contract by up to 94% within a century, based on bioclimatic modelling (Rehfeldt *et al.*, 2009). Here, we use individual- and population-level clustering approaches to assess range-wide patterns of genetic divergence and diversity, and to assess rear-edge dynamics in aspen.

MATERIALS AND METHODS

Samples were collected from georeferenced sampling sites representing the full longitudinal and latitudinal extent of the range of aspen. A minimum of 20 ramets (trees) were sampled from each of 30 sampling sites (see Appendix S1 in Supporting Information). Within sampling sites, ramets were selected over a broad, relatively continuous area, in order to avoid resampling clones. Within each site, the maximum distance between ramets ranged from 1 km (UME) to 81 km (NMT). Leaves were stored in silica desiccant at ambient temperature. DNA was extracted from dried leaf tissue using a Qiagen DNEasy 96 Plant Kit (Qiagen, Chatsworth, CA, USA) following the manufacturer's protocol.

Nuclear microsatellites were used to conduct the range-wide phylogeography, based on a pilot study showing low range-wide divergence in a variety of nuclear and chloroplast sequences (Callahan, 2012). Microsatellites are a marker of choice in this situation, due to their rapid evolutionary rate, expected high population diversity, and broad genomic distribution. Eight microsatellite loci were amplified for all samples: WPMS14, WPMS15, WPMS17, WPMS20 (Smulders *et al.*, 2001), PMGC486, PMGC510, PMGC2571 and PMGC2658 (http://www.ornl.gov/sci/ipgc/ssr_resources.htm; Tuskan *et al.*, 2006). Reactions were prepared following Mock *et al.* (2008). Primer-specific annealing temperatures were 60 °C (WPMS14, WPMS15, WPMS20), 57 °C (PMGC486), 56 °C (WPMS17, PMGC510), and 55 °C (PMGC2571, PMGC2658). PCR products were analysed on
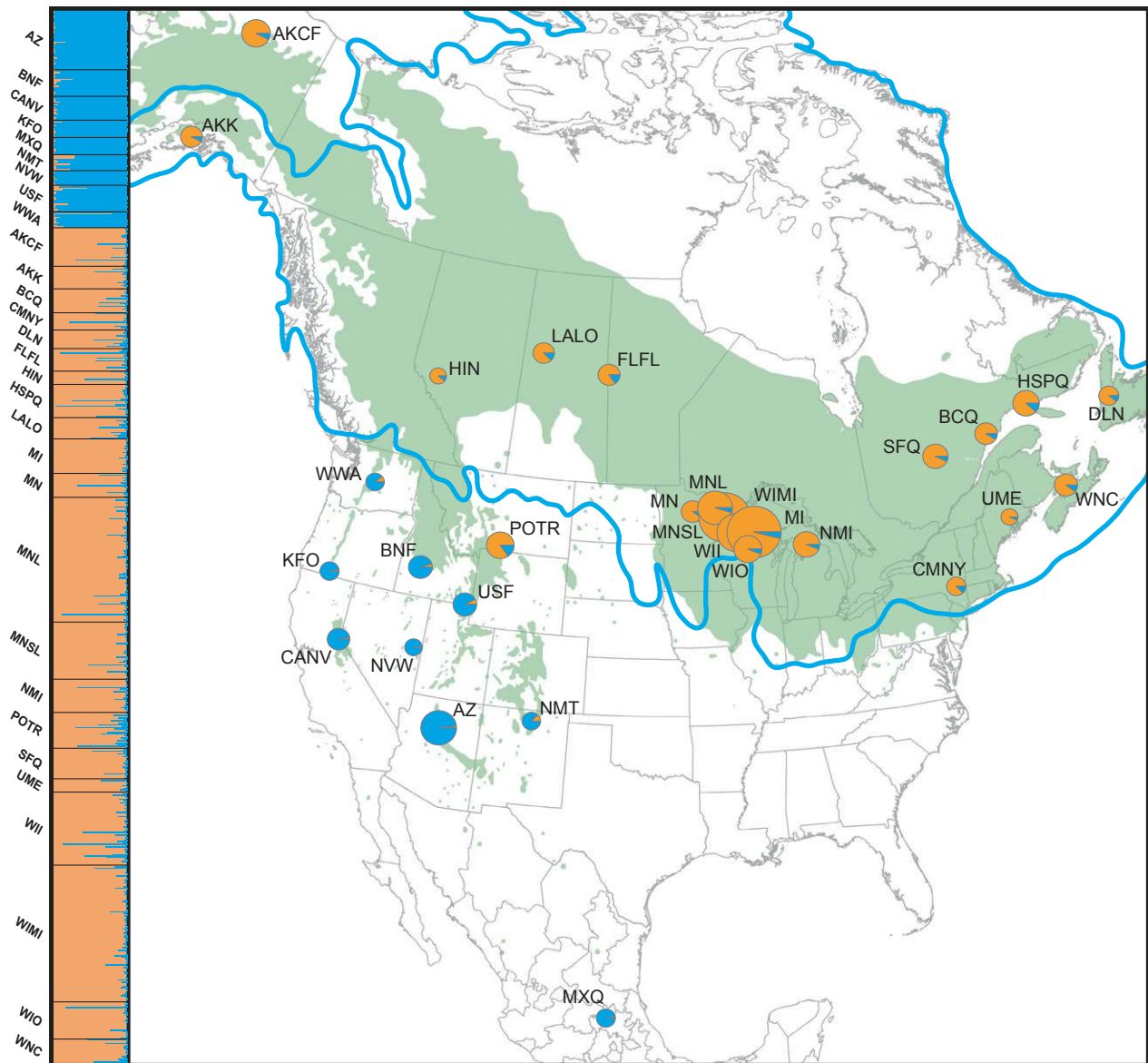
Figure 1 Sampling sites for aspen (*Populus tremuloides*) in North America. The distribution of aspen is shown in green (Little, 1971) and the outline of the Last Glacial Maximum is shown in blue (Ehlers *et al.*, 2011). Pie charts represent assignment probability of belonging to each of *K* = 2 clusters identified by STRUCTURE based on microsatellite allele frequencies, with probability values normalized using CLUMPP. Pie chart sizes are relative to sample size (*n*) at each sampling site. Individual membership coefficients are displayed in the bar plot: each individual is represented by a single horizontal bar, with sampling site labels shown on the left. The south-western cluster is shown in blue, and the northern cluster in orange. The figure was projected using the Albers North American equal-area conic projection.

an ABI 3100 or ABI 3730 sequencer using a LIZ500 (Applied Biosystems, Rotkreuz, Switzerland) size standard. Microsatellite chromatograms were scored using ABI GENEMAPPER 4 software (Applied Biosystems).

An original set of 1120 samples was reduced to a final set of 794 genets representing 30 sampling sites (Fig. 1, Appendix S1) by culling duplicate genets, putative triploids, putative hybrids, and any individuals that failed to amplify at a minimum of seven of eight microsatellite loci. Replicate samples from identical genets were excluded. Given the power of our microsatellite loci to distinguish individuals (average

probability of identity of 1.72×10^{-7} for random individuals and 8.3×10^{-4} for random siblings; Waits *et al.*, 2000) and the spatial proximity of identical genets, we were confident that the 67 duplicate genotypes we encountered were resampled genets. Samples with three alleles at any locus (*n* = 118) were presumed to be triploids (Mock *et al.*, 2012) and were removed from the data set, because loci with one or two alleles in triploids could not be genotyped accurately (due to dominance) and because triploids are expected to have very low fertility. Putative hybrids were identified as a cluster of outliers using an exploratory principal coordinates

analysis (PCoA) implemented in GenAlEx 6.3 (Peakall & Smouse, 2006) (data not shown). Based on this result, we identified and removed 37 putative hybrid genets – 35 from the Great Lakes area and two from Nova Scotia, both areas where the distribution of aspen overlaps that of *P. grandidentata*, which is known to produce at least first-generation hybrids with aspen (Barnes, 1961). Finally, 104 individuals that did not amplify at a minimum of seven of eight microsatellite loci were removed. Subsequent genetic analyses were conducted on the final set of 794 unique genets (Appendix S1).

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 3.5.1.3 (Excoffier & Lischer, 2010). Values of observed (H_O) and expected (H_E) heterozygosity were estimated for each sampling site, averaging across all eight loci. Inbreeding coefficients (F_{IS}) were calculated using FSTAT 2.9.3.2 (Goudet, 2001). One-sample *t*-tests were used to identify sites with significant F_{IS} values. To describe the genetic diversity within sampling areas, average allelic richness across all eight loci was calculated using FSTAT, with rarefaction employed to account for differences in sample size. The significance of differences in H_E and allelic richness among clusters and subclusters was determined using the Student's *t*-test.

To identify genetically distinct clusters of genets in our data set, we used an individual-based assignment approach, assuming correlated allele frequencies and admixed ancestry, as implemented in STRUCTURE 2.3 (Pritchard *et al.*, 2000). In general, admixture models are considered favourable to, and more robust than, models lacking admixture when local populations are likely to share migrants (François & Durand, 2010). For all STRUCTURE runs, the length of burn-in was set to 20,000 followed by 50,000 Markov chain Monte Carlo (MCMC) iterations. We tested *K*-values from 1 to 10 with 10 iterations completed for each *K* value. The most likely *K* value was determined by calculating Δ*K* following Evanno *et al.* (2005). All Δ*K* calculations were performed using the online version of STRUCTURE HARVESTER 0.6.91 (Earl & von-Holdt, 2012). We used CLUMPP 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 1.1 (Rosenberg, 2004).

Based on clusters identified by STRUCTURE, an analysis of molecular variance (AMOVA) was carried out using Arlequin 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 geographical distances between sites within clusters identified by STRUCTURE. Range-wide structure was also explored using BAPS 5.3 (Corander *et al.*, 2008b), and PCoA of individuals and populations in GenAlEx. We used the admixture model implemented in BAPS for analysis of range-wide structure and comparison with results generated

by STRUCTURE. The fixed-*K* analysis in BAPS was used for direct comparison to STRUCTURE results. All BAPS analyses were completed using 100 iterations, 50–150 reference individuals (depending on fixed-*K* value), and 20 iterations for each reference individual. We also constructed an unrooted neighbour-joining tree of populations using a Cavalli-Sforza distance matrix (Cavalli-Sforza & Edwards, 1967) and bootstrapping 1000 times over loci, implemented in the program POPULATIONS 1.2.31 (Langella, 2000).

RESULTS

Genetic structuring

Out of 240 locus–site combinations, and using Bonferroni-corrected significance levels, we found deviations from the genotypic proportions expected under Hardy–Weinberg equilibrium only at the AZ (WPMS17 and PMGC2658), MN (PMGC2571) and USF (WPMS20 and PMGC2658) sampling sites. No significant genotypic disequilibrium was detected following Bonferroni correction.

We identified an optimal solution of *K* = 2 clusters using the Bayesian algorithm implemented in STRUCTURE followed by calculation of Δ*K* (Fig. 1, Appendix S2). One of the clusters (the 'south-western cluster') is represented by individuals from the nine sampling sites in the south-western portion of aspen's range (AZ, BNF, CANV, KFO, MXQ, NMT, NVW, USF and WWA; Fig. 1). The second cluster (the 'northern cluster') is represented by the remaining sampling sites. In both cases, individuals with equivocal cluster assignments were infrequent.

Further STRUCTURE analyses were performed on each of the major clusters. Within the south-western cluster (*n* = 164 individuals), *K* = 2 optimal subclusters were identified, hereafter referred to as the south-west–south (SWS) and south-west–north (SWN) clusters (Fig. 2). Within the northern cluster (*n* = 630 individuals), although *K* = 2 was also the optimal solution, all individual membership assignments were divided between the two clusters, indicating no geographical structure of individuals. The STRUCTURE solutions for *K* = 3 on the whole data set also identified the northern cluster and the two south-western subclusters (not shown).

The same pattern of regional structuring (northern and south-western clusters with substructuring in the south-western cluster) was detected using BAPS when *K* = 2 or *K* = 3. The optimal solution identified by BAPS suggested *K* = 5, including the northern cluster and the two south-western clusters SWS and SWN, but with the Arizona population (AZ) and Mexico (MXQ) populations identified as separate clusters (Appendix S3). However, the algorithm implemented in BAPS is known to result in oversplitting (Latch *et al.*, 2006; Corander *et al.*, 2008a), and hence the finding of *K* = 5 may not be the most biologically realistic result. The neighbour-joining tree of populations also supported the STRUCTURE solutions, showing little support for structuring among the northern populations, but strong
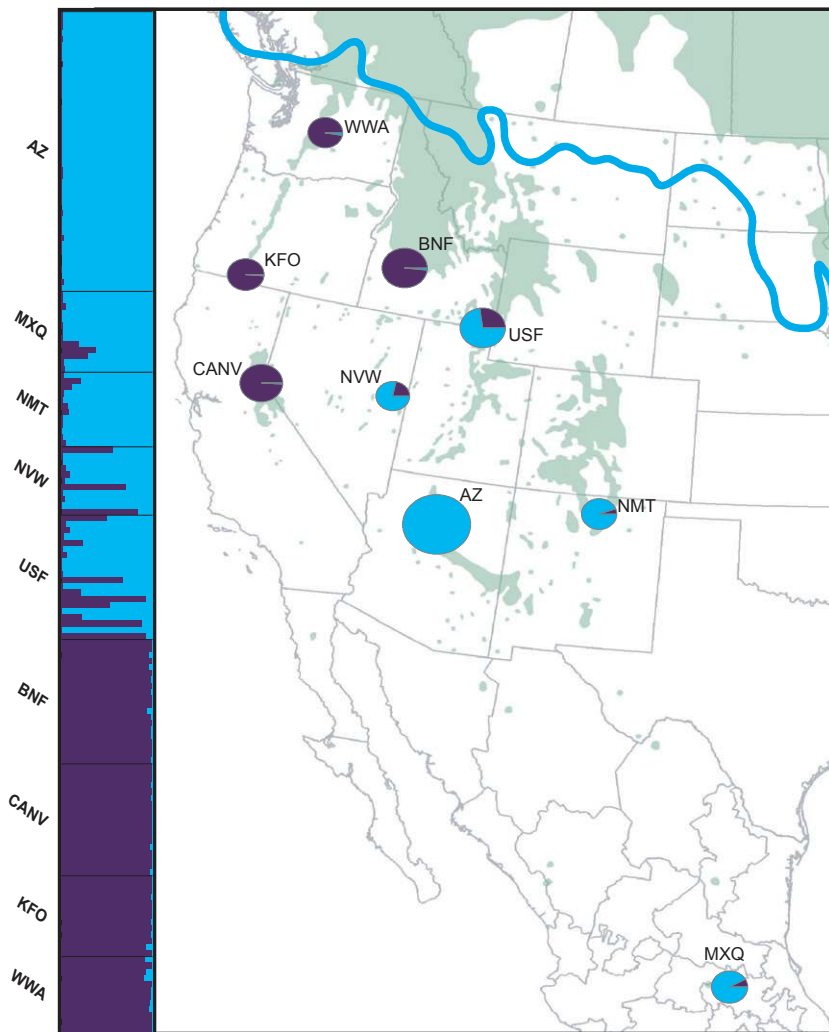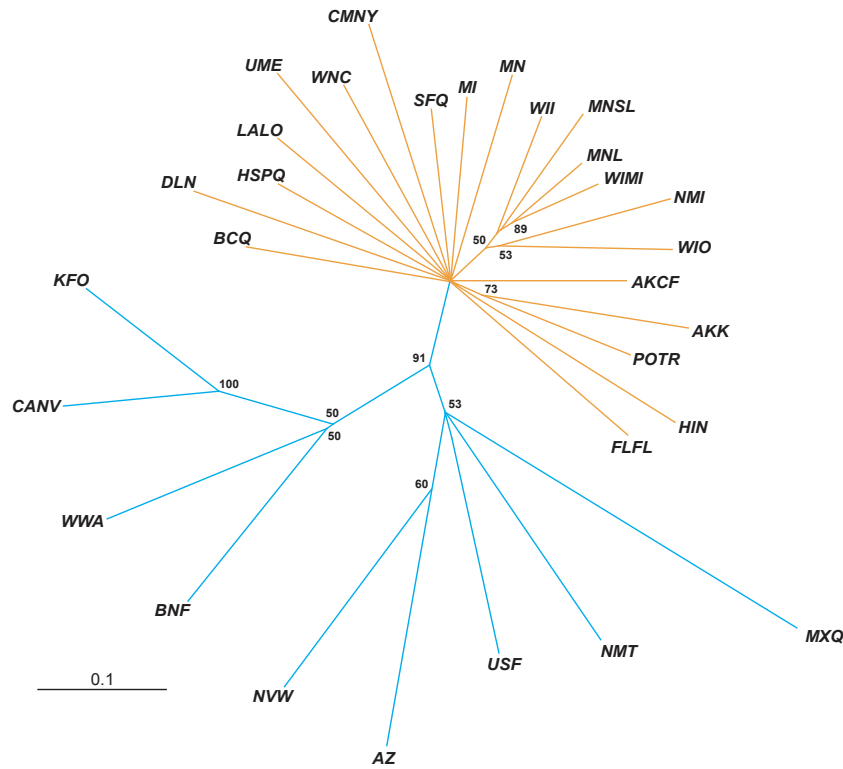
Figure 2 Assignment of North American aspen (*Populus tremuloides*) sampling sites from the south-western cluster to $K = 2$ clusters identified by STRUCTURE based on microsatellite allele frequencies. Pie charts represent probability of each sampling site belonging to each of the two clusters, with probability values normalized using CLUMPP. Pie chart sizes are relative to sample size (n) at each sampling site. Individual membership coefficients are displayed in the bar plot: each individual is represented by a single bar, with sampling site labels displayed on the left. SWS cluster shown in dark purple, SWN cluster shown in light blue. The figure was projected using the Albers North American equal-area conic projection.

bootstrap support for the south-western cluster and bootstrap support > 50% for the SWN and SWS subclusters (Fig. 3).

AMOVA results (Table 1a) indicated that 5.4% of the genetic variation was partitioned among the two major clusters (northern and south-western), 3.2% among sampling sites within the two major clusters, and 91.4% within sampling sites, with an overall F_{ST} of 0.086 ($P < 0.001$). When AMOVA was performed by pooling sampling sites within each cluster (i.e. treating clusters as populations), F_{ST} was 0.058 ($P < 0.001$). AMOVA was also used to assess the degree of structuring among the two south-western subclusters (SWN and SWS) and among sampling sites within each of these subclusters (Table 1b). AMOVA results indicated that 11.4% of the genetic variation in the south-western clus-

ter 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 0.177 ($P < 0.001$). In the northern cluster, AMOVA indicated that 1.3% of the genetic variation (F_{ST} of 0.013) was partitioned among sampling sites and 98.7% within sampling sites ($P < 0.001$).

Isolation by distance, a pattern produced by the interaction of gene flow and genetic drift (Hutchison & Templeton, 1999), was strikingly different between the northern and south-western clusters. Genetic distance was significantly correlated with geographical distance in the south-western cluster ($r^2 = 0.202$, $P < 0.01$) but not the northern cluster ($r^2 = 0.003$, $P = 0.311$). Sites within the south-western cluster displayed a much broader range of linearized pairwise Φ_{ST} than did those within the northern cluster (Fig. 4).

Figure 3 Unrooted neighbour-joining tree of North American aspen (*Populus tremuloides*) based on a Cavalli-Sforza (Cavalli-Sforza & Edwards, 1967) distance matrix. The proportion of 1000 bootstrap replicates supporting nodes are shown when > 50%; nodes with < 50% support are collapsed.

Table 1 Results of analysis of molecular variance (AMOVA) for *Populus tremuloides* in North America ($n = 794$ genets), based on microsatellite allele frequencies for (a) south-western and northern clusters, and (b) subclusters within the south-western clusters: SWS and SWN. All AMOVA results were significant ($P < 0.001$).

| Source of variation	Sum of squares	Variance components	% variation
(a)			
Among south-western and northern clusters	101.59	0.18	5.45
Among sites within clusters	232.27	0.10	3.17
Within sites	4644.44	2.98	91.38
Total	4978.30	3.26	
(b)			
Among SWS and SWN clusters	70.43	0.38	11.41
Among sites within clusters	68.01	0.20	6.13
Within sites	877.29	2.75	82.46
Total	1015.73	3.34	

Genetic diversity

Comparison of genetic diversity between the northern and south-western clusters and among sampling sites was assessed using allelic richness and expected heterozygosity (H_E).
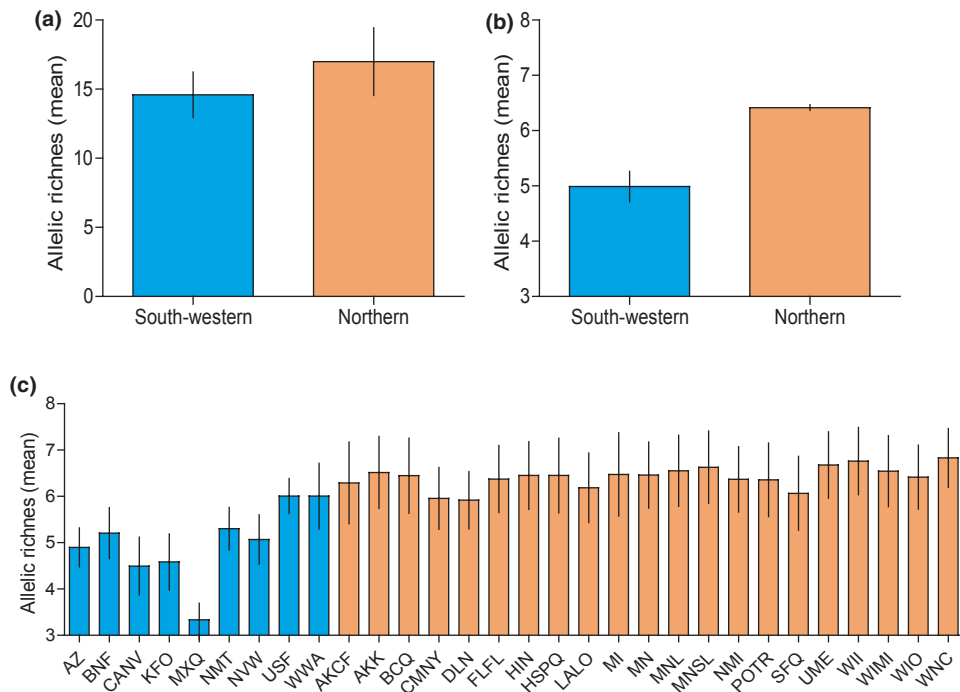
Among sampling sites, allelic richness varied from 3.337 (MXQ) to 6.832 (WNC) (Fig. 5c). Sampling sites within the south-western cluster had an average allelic richness of 4.989, significantly lower ($P < 0.001$) than sites within the northern cluster, which had an average allelic richness of 6.415 (Fig. 5b, Table 2). Comparing the two south-western subclusters, sampling sites contained similar levels of average allelic richness: 4.923 and 5.073 ($P = 0.797$) for SWS and SWN, respectively.

Pooling all samples within clusters, average allelic richness was marginally lower in the south-western cluster (14.592) than in the northern cluster (16.995), but not significantly different ($P = 0.078$) (Fig. 5a, Table 2). When samples were pooled within each of the south-western subclusters, we found similar levels of allelic richness (SWS, 11.42; SWN, 10.45; $P = 0.373$; Table 2).

Expected heterozygosity (H_E), ranged from 0.613 (MXQ) to 0.801 (WNC) (Table 2). Average H_E among sampling sites was lower in the south-western cluster (0.711) than in the northern cluster (H_E, 0.779; $P = 0.008$). Average H_E among sampling sites was not significantly different between the SWS (0.715) and SWN (0.707) subclusters ($P = 0.850$). When sampling sites within major clusters were pooled and each cluster treated as a single group, cluster-level H_E values were 0.789 for the northern cluster and 0.805 for the south-western cluster and not significantly different ($P = 0.793$;
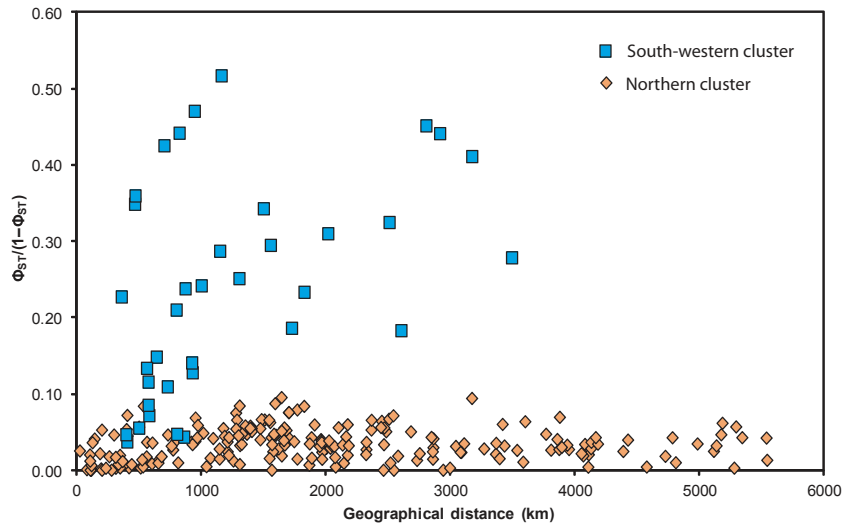
Figure 4 Isolation-by-distance patterns within the south-western (SW) and northern (N) clusters of North American aspen (*Populus tremuloides*). Each point represents the genetic distance between two sites as a function of geographical distance. Sites within the south-western cluster displayed a much broader range of linearized pairwise Φ_{ST} values than those within the northern cluster.

Figure 5 Measures of allelic richness in North American aspen (*Populus tremuloides*) populations, averaging across all eight loci: (a) mean cluster-level allelic richness values, pooling individuals within each cluster; (b) mean sampling site-level allelic richness for each cluster; (c) sampling site-level allelic richness for each of the 30 sampling sites, colours represent cluster assignment.

Table 2). Levels of inbreeding within sampling sites were low for all sampling sites, with an average F_{IS} of 0.019 ($P = 0.122$; Table 2).

DISCUSSION

Our primary findings were: (1) the existence of two geographical genetic clusters within the range of aspen, the northern and south-western clusters (Figs 1 & 3); (2) further subdivision of the south-western cluster (Figs 2 & 3); and (3) characteristics consistent with a historically stable rear edge in the south-western cluster.

Range-wide population structuring

The geographical locations of the south-western and northern clusters suggest two distinct varieties in the landscape, occupying, and possibly adapted to, regions with distinct cli-

Table 2 Summary statistics for North American *Populus tremuloides* clusters and sampling sites. Significant values are in italics. Cluster-specific values were calculated regionally rather than within-site averages.

Cluster/site	*n*	AR	H_E	F_{IS}
South-western cluster	164	14.59	0.81	
SWS cluster	101	11.42	0.76	
AZ	45	4.90	0.70	0.19
MXQ	13	3.34	0.61	−0.12
NMT	12	5.30	0.77	0.03
NVW	11	5.07	0.72	0.02
USF	20	6.01	0.77	0.16
SWN cluster	63	10.45	0.73	
BNF	20	5.21	0.72	0.01
CANV	18	4.50	0.68	−0.05
KFO	13	4.59	0.64	0.08
WWA	12	6.00	0.78	0.09
Northern cluster	630	16.99	0.79	
AKCF	29	6.29	0.76	0.04
AKK	17	6.52	0.80	0.06
BCQ	18	6.45	0.76	−0.02
CMNY	13	5.96	0.78	−0.03
DLN	14	5.92	0.77	0.04
FLFL	17	6.37	0.78	0.07
HIN	10	6.45	0.78	−0.06
HSPQ	25	6.45	0.78	0.03
LALO	16	6.19	0.78	0.01
MI	26	6.47	0.76	−0.02
MN	18	6.46	0.79	0.05
MNL	94	6.55	0.78	0.00
MNSL	43	6.63	0.78	*0.03*
NMI	25	6.37	0.79	0.00
POTR	27	6.36	0.77	−0.03
SFQ	23	6.07	0.76	0.04
UME	10	6.68	0.79	−0.08
WII	55	6.76	0.79	*0.05*
WIMI	103	6.54	0.78	0.01
WIO	28	6.42	0.78	−0.02
WNC	19	6.83	0.80	−0.02

n, the number of samples representing each cluster, subcluster, or site; AR, mean allelic richness; H_E, expected heterozygosity; F_{IS}, inbreeding coefficient.

mates: a semi-arid climate in the western USA and a mesic and largely continental climate in Canada and Alaska (Mock *et al.*, 2012). The geography of the south-western cluster generally corresponds with the range-wide incidence of triploidy recently described in aspen (Mock *et al.*, 2012), suggesting that aspen populations in this cluster have a tendency to form or retain triploids at a higher rate than those in the northern cluster. It is possible that differences in average generation time (due to the episodic nature of seedling recruitment in the landscape occupied by the south-western cluster) are contributing to differences in allele frequencies. This would be consistent with the finding of large clones and high rates of triploidy in this same landscape (Barnes, 1966; Mock *et al.*, 2012). This would not, however, preclude the possibility of adaptive differences between the major clusters.

The idea that south-western aspen may represent a distinct variety is not new. Axelrod (1941) compared the leaf morphology and habitat types of two fossil species of *Populus* existing during the Miocene and Pliocene, *P. lindgreni* and *P. pliotremuloides*, noting that leaves of *P. lindgreni* more closely resemble leaves of modern *P. tremuloides* growing in more mesic environments, and that the leaves of *P. pliotremuloides* are indistinguishable from leaves of modern *P. tremuloides* growing in the more arid south-western portion of the range. From this comparison of leaf morphology and habitat types of these two fossil species, Axelrod (1941) concluded that, given the enormous and varied distribution of contemporary aspen, it seems 'highly probable' that multiple ecotypes could be present. Further, Barnes (1975) described clinal variation in aspen leaf morphology from south to north in the western USA and Canada. Given that the populations comprising the south-western cluster have occupied the more arid environments in the species' range, they may well have adaptive traits not found in the rest of the range.

Although the major genetic clusters were generally consistent with geographical boundaries, the POTR sampling site within the greater Yellowstone area was an anomaly. This site was consistently assigned to, and strongly affiliated with, the northern cluster in both the individual-based assignment testing (Fig. 1) and the population-based neighbour-joining tree (Fig. 3). Average allelic richness was also higher at this site than in the rest of the south-western sites, and more similar to those observed in northern cluster sites. Further, the POTR site has been shown to have a particularly low rate of triploidy more typical of eastern and northern populations (Mock *et al.*, 2012). The affiliation of POTR with the northern cluster may seem geographically anomalous because we lack the sampling sites throughout Montana and North Dakota that would provide more obvious continuity with the northern cluster populations. The affiliation of POTR with the northern cluster is consistent with the finding that the northern and south-western clusters are generally separated by the Continental Divide, at least in the USA.

Substructuring within the south-western cluster

Geographical and genetic subdivision within the south-western cluster resembles that demonstrated for many other species (Brunsfeld *et al.*, 2001), including ponderosa pine (*Pinus ponderosa*; Latta & Mitton, 1999) and Douglas fir (*Pseudotsuga menziesii*; Gugger *et al.*, 2010), which commonly co-occur with aspen in western North America. These species are thought to have been subdivided into distinct refugia (Cascade versus Rocky Mountains) during the Wisconsin glaciation, with current contact zones in Montana and Idaho. The USF site shows a greater level of mixed assignment between the two subclusters than the other sites and nearly at the same longitude as the transition zone in west-central Montana for ponderosa pine (Latta & Mitton, 1999) (Fig. 2). Our results at both the individual level

(Fig. 2) and the population level (Fig. 3) suggest that aspen in the south-western cluster may have been subdivided into separate glacial refugia during the Pleistocene, subsequently establishing a contact zone along the eastern side of the Great Basin. However, the low sampling-site density in this region makes fine-scaled comparisons among species difficult.

Stable versus trailing-edge dynamics

Our results suggest that the south-western cluster represents a distinct lineage, and a potentially distinct variety, which is consistent with these populations being part of a stable edge during climate oscillations. This south-western portion of the range is home to multiple mountain ranges and basins, which would have provided ample opportunity for elevational shifts in response to changes in climate throughout past glacial and interglacial episodes, as has been suggested for other species (Tzedakis *et al.*, 2002). A stable-edge scenario for the south-western cluster is also supported by our observed patterns of population divergence and allelic richness (structured populations with relatively low diversity but regional allelic richness similar to that in the rest of the range). We also found little evidence of northward post-glacial expansion of the south-western cluster, which is characteristic of other relictual stable-edge populations (Bilton *et al.*, 1998; Petit *et al.*, 2003; Hampe & Petit, 2005).

By contrast, the northern clade showed reduced genetic structuring among populations; F_{ST} was an order of magnitude lower among northern clade populations than among south-western populations. Substructuring within the northern cluster was not detectable at the individual level, and only minor structuring was evident at the population level (Fig. 3), and there was no pronounced pattern of isolation by distance, even at great geographical distances (Fig. 4). These patterns indicate that aspen in the northern portion of its range has undergone, and may still be undergoing, rapid post-glacial expansion with high levels of gene flow. Testing specific hypotheses about glacial refugia within the northern cluster is likely to require finer-scale sampling of populations and the use of nuclear and/or organellar sequence data, which can be more reliably ordered into an evolutionary sequence than microsatellites (Ellegren, 2004).

CONCLUSIONS

The range-wide structuring results have implications for translocation programmes and seed-zone boundaries as well as for the design of ecological studies in aspen. We recommend that additional populations and perhaps a frequency-based sequencing approach (Breen *et al.*, 2012) or an approach based on single nucleotide polymorphisms (SNPs) (Keller *et al.*, 2010) be used to confirm and refine the boundaries between the genetic clusters described here, and to examine more closely the role of glacial refugia in the current distribution of genetic variation. We also recommend

that high-throughput sequencing approaches (e.g. restriction site associated DNA markers; Miller *et al.*, 2007) and common-garden studies (e.g. Gray *et al.*, 2011) be used to assess the extent and nature of adaptive variation among clusters and along environmental gradients and to improve our understanding of the ecological and demographic forces shaping the evolution of aspen.

The stable-edge dynamics in the south-western populations present a management challenge. These populations are more isolated and less genetically diverse than those in the northern cluster, and climate modelling suggests a future range contraction in this region (Rehfeldt *et al.*, 2009). However, the south-western populations collectively may represent a large portion of the species' genetic variation and evolutionary history, and aspen in the Intermountain West is particularly valuable as a foundation species, because it is the only common deciduous hardwood tree in these forests. Restoration and climate change mitigation efforts will be difficult in this region due to the rarity of conditions required for seedling establishment (McDonough, 1985) and the plethora of ecological challenges facing aspen (Frey *et al.*, 2004), but assisted migration along elevational gradients (Gray *et al.*, 2011) might be appropriate.

REFERENCES

Axelrod, D.I. (1941) The concept of ecospecies in Tertiary paleobotany. *Proceedings of the National Academy of Sciences USA*, **27**, 545–551.

Barnes, B.V. (1961) Hybrid aspens in the lower peninsula of Michigan. *Rhodora*, **63**, 311–324.

Barnes, B.V. (1966) The clonal growth habit of American aspens. *Ecology*, **47**, 439–447.

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

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 B: Biological Sciences*, **265**, 1219–1226.

Breen, A.L., Murray, D.F. & Olson, M.S. (2012) Genetic consequences of glacial survival: the late Quaternary history of balsam poplar (*Populus balsamifera* L.) in North America. *Journal of Biogeography*, **39**, 918–928.

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., Soltis, D.E. & Soltis, P.S. (2001) Comparative phylogeography of northwestern North America: a synthesis. *Integrating ecology and evolution in a spatial context* (ed. by J. Silvertown and J. Antonovics), pp. 319–339. Blackwell, Oxford.

Callahan, C.M. (2012) *Continental-scale characterization of molecular variation in quaking aspen*. MS Thesis, Utah State University, Logan, UT.

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.

Cavalli-Sforza, L.L. & Edwards, A.W.F. (1967) Phylogenetic analysis: models and estimation procedures. *American Journal of Human Genetics*, **19**, 233–257.

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.

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

Corander, J., Sirén, J. & Arjas, E. (2008a) Bayesian spatial modeling of genetic population structure. *Computational Statistics*, **23**, 111–129.

Corander, J., Marttinen, P., Sirén, J. & Tang, J. (2008b) Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics*, **9**, 539.

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.

Ehlers, J., Gibbard, P.L. & Hughes, P.D. (eds) (2011) *Quaternary glaciations – extent and chronology*. A closer look, Elsevier Science, Amsterdam.

Ellegren, H. (2004) Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, **5**, 435–445.

Elliott, G.P. & Baker, W.L. (2004) Quaking aspen (*Populus tremuloides* Michx.) at treeline: a century of change in the San Juan Mountains, Colorado, USA. *Journal of Biogeography*, **31**, 733–745.

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.

Fechner, G.H. & Barrows, J.S. (1976) *Aspen stands as wildfire fuel breaks*. Eisenhower Consortium Bulletin 4. USDA Forest Service Rocky Mountain Forest and Range Experiment Station, Fort Collins, CO.

François, 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. & Landhäusser, S.M. (2004) Predicting landscape patterns of aspen dieback: mechanisms and knowledge gaps. *Canadian Journal of Forest Research*, **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. (2001) *FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3)*. Available at: http://www2.unil.ch/popgen/softwares/fstat.htm.

Grant, M.C., Mitton, J.B. & Linhart, Y.B. (1992) Even larger organisms. *Nature*, **360**, 216.

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.

Gugger, P.F., Sugita, S. & Cavender-Bares, J. (2010) Phylogeography of Douglas-fir based on mitochondrial and chloroplast DNA sequences: testing hypotheses from the fossil record. *Molecular Ecology*, **19**, 1877–1897.

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.

Hogg, E.H., Brandt, J.P. & Kochtubajda, B. (2002) Growth and dieback of aspen forests in northwestern Alberta, Canada, in relation to climate and insects. *Canadian Journal of Forest Research*, **32**, 823–832.

Hutchison, D.W. & Templeton, A.R. (1999) Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution*, **53**, 1898–1914.

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.

Keller, S.R., Levsen, N., Olson, M.S. & Tiffin, P. (2012) Local adaptation in the flowering-time gene network of balsam poplar, *Populus balsamifera* L. *Molecular Biology and Evolution*, **29**, 3143–3152.

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

Landhäusser, S.M. & Wein, R.W. (1993) Postfire vegetation recovery and tree establishment at the Arctic treeline: climate-change–vegetation-response hypotheses. *Journal of Ecology*, **81**, 665–672.

Landhäusser, S.M., Deshaies, D. & Lieffers, V.J. (2010) Disturbance facilitates rapid range expansion of aspen into higher elevations of the Rocky Mountains under a warming climate. *Journal of Biogeography*, **37**, 68–76.

Langella, O. (2000) *Populations 1.2*. Available at: http://bioinformatics.org/~tryphon/populations/.

Latch, E.K., Dharmarajan, G., Glaubitz, J.C. & Rhodes, O.E. (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics*, **7**, 295–302.

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) *Atlas of United States trees, Vol. 1, Conifers and important hardwoods*. Miscellaneous Publication 1146. USDA Forest Service, Washington, D.C.

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

McDonough, W.T. (1985) Sexual reproduction, seeds and seedlings. *Aspen: ecology and management in the western United States* (ed. by N.V. DeByle and R.P. Winokur), pp. 25–28. General Technical Report RM-119. USDA Forest Service, Fort Collins, CO.

Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A. & Johnson, E.A. (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, **17**, 240–248.

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.

Mock, K.E., Callahan, C.M., Islam-Faridi, M.N., Shaw, J.D., Rai, H.S., Sanderson, S.C., Rowe, C.A., Ryel, R.J., Madritch, M.D., Gardner, R.S. & Wolf, P.G. (2012) Widespread triploidy in western North American aspen (*Populus tremuloides*). *PLoS ONE*, **7**, e48406.

Ouborg, N.J., Pertoldi, C., Loeschcke, V., Bijlsma, R. & Hedrick, P.W. (2010) Conservation genetics in transition to conservation genomics. *Trends in Genetics*, **26**, 177–187.

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) *Populus tremuloides* Michx., quaking aspen. *Silvics of North America, Vol. 2, Hardwoods* (ed. by R.M. Burns and B.H. Honkala), pp. 555–569. Agriculture Handbook 654. USDA Forest Service, Washington, D.C.

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., Müller-Starck, G.M., Demesure-Musch, B., Palmé, A., Martín, J.P., Rendell, S. & Vendramin, G.G. (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science*, **300**, 1563–1565.

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

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.

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. June 13–15, 2000. Grand Junction, Colorado* (ed. by W.D. Sheppard, D. Binkley, D.L. Bartos, T.J. Stohlgren and L.G. Eskew), pp. 185–191. USDA Forest Service Proceedings RMRS-P-18. USDA Forest Service, Washington, D.C.

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.

Tuskan, G.A., DiFazio, S., Jansson, S. *et al.* (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.

Waits, L., Taberlet, P., Swenson, J.E., Sandegren, F. & Franzén, R. (2000) Nuclear DNA microsatellite analysis of genetic diversity and gene flow in the Scandinavian brown bear (*Ursus arctos*). *Molecular Ecology*, **9**, 421–431.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 List of sampling sites.
Appendix S2 Plot of ΔK used to assess the optimal numbers of cluster of *Populus tremuloides*.
Appendix S3 Assignment of *Populus tremuloides* sampling sites using BAPS.

BIOSKETCH

Colin M. Callahan is a graduate student at Utah State University interested in molecular ecology, phylogeography, and evolutionary biology. The focus of this research team is on conservation genetics and forest ecology.

Author contributions: C.C. and K.M conceived the ideas; C.C. and C.R. collected data; C.C. analysed the data; R.R., J.S. and M. M. provided samples and contributed to the writing; and C.C. and K.M. led the writing.

Editor: Pauline Ladiges