Published in "Molecular Ecology 19(8): 1638-1650, 2010"
which should be cited to refer to this work.

Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula* L.), a widespread forest tree

DULCINEIA DE CARVALHO,*‡ PÄR K. INGVARSSON,† JEFFREY JOSEPH,‡ LEONIE SUTER,‡ CLAUDIO SEDIVY,‡ DAVID MACAYA-SANZ,‡§ JOAN COTTRELL,¶ BERTHOLD HEINZE,** IVAN SCHANZER†† and CHRISTIAN LEXER‡ ‡‡

**Universidade Federal de Lavras, Caixa Postal 3037 DCF/UFLA, 37200-000 Lavras, MG, Brazil, †Umeå Plant Science Centre, Department of Ecology and Environmental Science, Umeå University, Linneaus väg 6, SE-90187 Umeå, Sweden, ‡Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey TW 3DS, UK, §CIFOR-INIA, Ctra. de La Coruña, Km. 7.5, 28040 Madrid, Spain, ¶Forest Research, Northern Research Station, Roslin, Midlothian EH25 9SY, UK, **Federal Research Centre for Forests, Department of Genetics, Hauptstraße 7, A-1140 Vienna, Austria, ††Herbarium, Main Botanical Garden, Russian Academy of Sciences, Botanicheskaya Str. 4, 127276 Moscow, Russia, ‡‡Unit of Ecology and Evolution, Department of Biology, University of Fribourg, Chemin du Musée 10, CH-1700 Fribourg, Switzerland*

Abstract

Adaptation to new environments can start from new mutations or from standing variation already present in natural populations. Whether admixture constrains or facilitates adaptation from standing variation is largely unknown, especially in ecological keystone or foundation species. We examined patterns of neutral and adaptive population divergence in *Populus tremula* L., a widespread forest tree, using mapped molecular genetic markers. We detected the genetic signature of postglacial admixture between a Western and an Eastern lineage of *P. tremula* in Scandinavia, an area suspected to represent a zone of postglacial contact for many species of animals and plants. Stringent divergence-based neutrality tests provided clear indications for locally varying selection at the European scale. Six of 12 polymorphisms under selection were located less than 1 kb away from the nearest gene predicted by the *Populus trichocarpa* genome sequence. Few of these loci exhibited a signature of 'selective sweeps' in diversity-based tests, which is to be expected if adaptation occurs primarily from standing variation. In Scandinavia, admixture explained genomic patterns of ancestry and the nature of clinal variation and strength of selection for bud set, a phenological trait of great adaptive significance in temperate trees, measured in a common garden trial. Our data provide a hitherto missing direct link between past range shifts because of climatic oscillations, and levels of standing variation currently available for selection and adaptation in a terrestrial foundation species.

Keywords: adaptive divergence, admixture, genome scan, photoperiod, selective sweep, standing genetic variation

Introduction

The role of gene flow in adaptive evolution is a hotly debated topic in evolutionary biology. At the *within*-species level, gene flow has traditionally been seen as a homogenizing force that impedes local adaptation (Stearns & Hoekstra 2005), whereas others have stressed its role in contributing alleles to the standing variation available for local adaptation (Przeworski *et al.* 2005; Pennings & Hermisson 2006), or in facilitating species cohesion through the spread of beneficial alleles

Correspondence: Christian Lexer, Fax: +41 26 300 9698;
E-mail: christian.lexer@unifr.ch

(Rieseberg & Burke 2001). The role of gene flow at the *between*-species level is no less controversial, opinions ranging from 'evolutionary noise' to 'potent evolutionary force that creates opportunities for adaptive evolution and speciation' (Schemske 2000; Wu 2001; Rieseberg *et al.* 2003; Arnold 2006). Considering the intensity of the debate, surprisingly little attention has been paid to the *intermediate* scenario of gene flow between previously isolated conspecific populations, i.e. lineages that were isolated long enough for adaptive differentiation to arise but over too short time for the evolution of substantial reproductive isolation. This is expected to be the case for many species of the temperate zones, i.e. those regions of the world most strongly affected by the climatic shifts of the pleistocene and holocene (Hewitt 2000).

Cycles of genetic divergence and subsequent gene flow between temporarily isolated populations have been predicted to result in increased standing variation available for adaptive evolution (Hewitt 2000; Pennings & Hermisson 2006; Barrett & Schluter 2008), a hypothesis also supported by recent empirical work on fishes (Colosimo *et al.* 2005), mice (Mullen & Hoekstra 2008) and plants (Whibley *et al.* 2006). This scenario is particularly likely for long-lived organisms, such as trees; in temperate forest trees, rapid diversifying selection (facilitated by great environmental heterogeneity and large effective population size, N_e) and low nucleotide substitution rates per unit time suggest that amounts of standing variation available for local adaptation are often limiting (Petit & Hampe 2006; Savolainen & Pyhajarvi 2007). In contrast, admixture may also constrain adaptation via outbreeding depression or tradeoffs with negative effects on fitness (Stearns & Hoekstra 2005). Limited empirical evidence is currently available in terrestrial species for how admixture between divergent postglacial lineages affects adaptation from standing genetic variation.

The effects of gene flow between populations with varying degrees of isolation can be described in terms of geographic clines for phenotypes or allelic frequencies in space (Barton & Hewitt 1985), or in terms of single-locus clines against genome-wide admixture gradients, estimated by comparing the ancestry of individual loci to expectations obtained from many loci in the genome (Briscoe *et al.* 1994; Reich *et al.* 2005; Gompert & Buerkle 2009). More experience exists for geographic clines, but admixture-based analysis can help pinpoint the exact target loci and nature of selection acting in admixed populations (Lexer *et al.* 2007; Gompert & Buerkle 2009), a task that is not easily achieved by studying the shapes of geographic clines alone (Kruuk *et al.* 1999).

Populus tremula L. is a deciduous forest tree whose distribution ranges from the British Isles to the east tip of Asia and from Northern Africa to Northern Scandinavia. Its mating system is dioecious and both pollen and seeds are wind-dispersed, resulting in substantial levels of neutral variability (scaled synonymous mutation rate theta = 0.012; one single nucleotide polymorphism for every 100–150 bp; Ingvarsson 2008). The species can be used to study the link between biogeographic history and current standing variation available for adaptation, because a growing knowledge base exists for both.

With respect to biogeographic history, fossil pollen records suggest the presence of *Populus* in Central Europe from 9000 BC onwards (Huntley & Birks 1983). The debate over the precise locations of glacial refuges is ongoing (Petit *et al.* 2003), but some safe assumptions can be made. Most importantly, Scandinavian populations must have re-colonized after the last ice age, since Scandinavia was covered by ice during the glacial period up until 7000 years ago and large areas have only been colonized within the last 2000 years because of a continual land uplift of 90 cm per century (Ericson & Wallentinus 1979). With regard to standing variation for candidate adaptive traits and genes, strong genetic differentiation and clinal variation were observed for multiple fitness-related traits in populations from across a wide latitudinal gradient in Sweden (Hall *et al.* 2007). Clinal variation was also observed for DNA polymorphisms located within the phytochrome B2 locus, a key gene controlling adaptation to differences in daylength (Ingvarsson *et al.* 2006).

In this study, we asked how admixture between divergent lineages affects adaptation from standing variation in the European aspen (*P. tremula* L.), a widespread forest tree. We studied a set of seven populations of this species, widely spaced across Europe, for 70 mapped microsatellite loci. Our goal was to infer patterns of neutral and non-neutral population divergence at the European scale and to test for admixture in Northern Europe. We then used phenotypic data from a common garden trial involving 12 populations from across Sweden to test whether admixture explains cline shape, variances and selection differentials for timing of bud set, a phenological trait of great adaptive significance. We first show that locally varying selection is detectable for gene-linked markers in this outcrossing forest tree at the subcontinental scale, contrary to expectations from previous studies with smaller geographic coverage (Hall *et al.* 2007; Ingvarsson 2008). We then demonstrate that admixture between divergent lineages has left its traces in phenotypic clines across the Northern European contact zone, leading to increased variance in and selection on adaptive traits. Our results are relevant to the ongoing discussion regarding the role of intraspecific admixture in adaptive evolution.

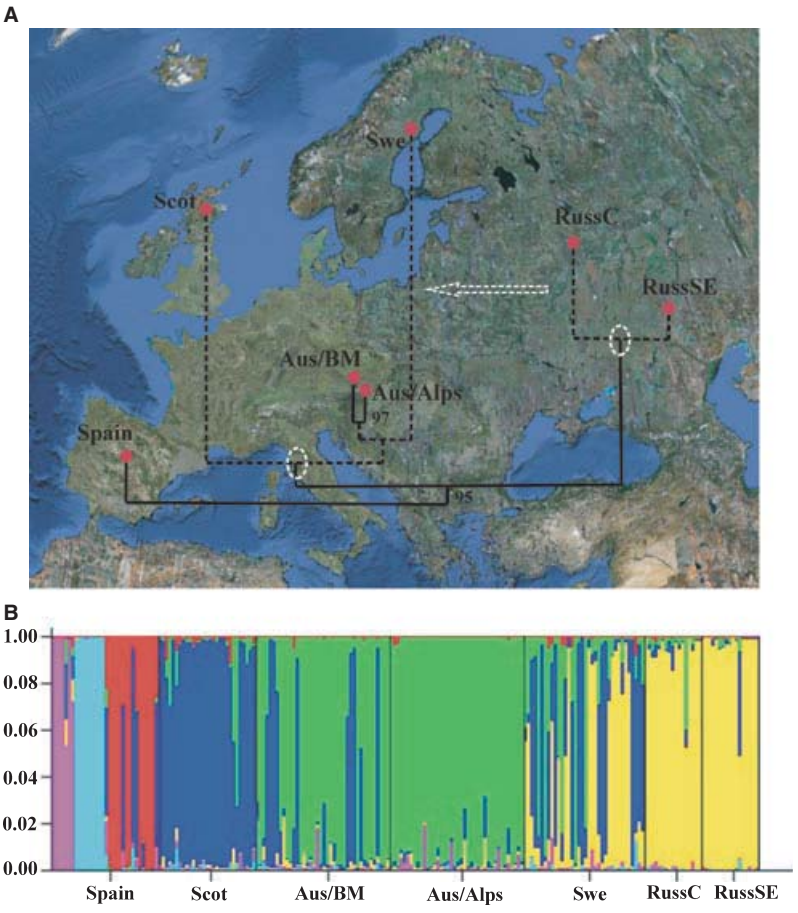

Fig. 1 Conventional (a) and Bayesian (b) analysis of population structure in *Populus tremula*, the European aspen, based on 70 microsatellite loci. (a) UPGMA tree based on Nei's standard genetic distance superimposed on geographic map showing locations of sampling sites. The dendrogram is informative regarding branching patterns but not branch lengths. Bootstrap percentages >60% are indicated beside each node, and nodes with <60% bootstrap support are indicated by dashed branches. Dashed white circles indicate nodes affected by the removal of 12 consistent non-neutral 'outlier' loci (Table 3) from the data set; removal of these loci results in a swap of population RussC to the clade containing populations Swe, Aus/BM, Aus/Alps and Scot, indicated by the white arrow. (b) Bayesian admixture coefficients (*Q*) for individual trees estimated within a linkage model with $K = 6$ gene pools; *Q* estimates for each tree are shown in the form of coloured, vertical bars. For population abbreviations and descriptions, see Table 1.

Materials and methods

Population samples and common garden trial

Population samples for molecular genetic work were collected in seven localities in Europe (Fig. 1a; Table 1) to facilitate analysis of population structure and the detection of non-neutral loci in pairwise comparisons at three different spatial scales: 'regional' (Austria/BM vs. Austria/Alps, Russia/C vs. Russia/SE), 'European' (10 pairwise comparisons) and 'sampling range-wide', i.e. all sampled European *Populus tremula* against Central European material of the closely related species *Populus alba* described in Lexer *et al.* (2005) (Table 2). The sample of *P. alba* was used as a reference to detect locus-specific reductions in genetic diversity because of selective sweeps across all studied populations of *P. tremula*; studying interspecific divergence (F_{ST}) was not a focus of this study. Each population was sampled over several dozens of square kilometres as in previous microsatellite studies of these species (Lexer *et al.* 2005, 2007), to avoid sampling clones and to account for the known great dispersal capacities of these trees. Field collections of this species must weigh the risk of slight population subdivision (Wahlund effects) because of wide local sampling against the risk of sampling multiple ramets of the same clone; cryptic subdivision will convert some of the among-population variation into within-population variation, which renders divergence (F_{ST})-based tests for locally varying selection conservative.

The common garden trial used to study patterns of phenotypic differentiation was previously described by Ingvarsson *et al.* (2006), Hall *et al.* (2007) and Luquez *et al.* (2008). In this study, extensive, previously unpublished multi-year data from this garden trial were analysed to examine adaptive trait differentiation across a Northern European admixture zone identified by molecular markers (below). Data from both the Ekebo garden (55.9°N) and the Sävar garden (63.4°N) described in Luquez *et al.* (2008) were used for this study. Each trial comprises 116 trees from 12 populations across a latitudinal gradient in Sweden (56–66°N), including up to eight replicate clones per tree planted in a randomized block design (Hall *et al.* 2007). Of a

Table 1 Populations of *Populus tremula* sampled for molecular genetic work, including population abbreviations, full names, geographic coordinates, no. of chromosomes sampled (*N*), allelic richness adjusted to the smallest number of individuals sampled by rarefaction, expected (H_E) and observed (H_O) heterozygosities, and within-population inbreeding coefficients (F_{IS})

Population abbreviation	Full name*	Latitude	Longitude	*N*	Allelic richness	H_E	H_O	F_{IS}†
Swe	Sweden/Umeå	63.829N	20.261E	72	4.897	0.549	0.473	0.134
Spain‡	Central Spain	40.017N	3.573W	64	3.497	0.474	0.468	0.022
Scot	Scotland	56.579N	3.606W	58	4.733	0.548	0.456	0.163
Aus/BM	Austria/Bohemian Massif	48.706N	16.097E	80	4.687	0.554	0.482	0.111
Aus/Alps	Austria/Eastern Alps	47.608N	16.099E	80	4.797	0.543	0.468	0.123
Russ/C	European Russia/Central	55.733N	37.674E	34	4.695	0.553	0.481	0.117
Russ/SE	European Russia/Southeast	48.677N	44.452E	34	4.708	0.545	0.457	0.156

*See Fig. 1 for location on a geographic map of Europe.
†Overall F_{IS} was significantly positive, with 95% bootstrap intervals ranging from 0.083 to 0.173.
‡A genetic bottleneck was detected for this population, based on both the sign test and Wilcoxon sign-rank test described by Piry *et al.* (1999).

Table 2 Number of candidate loci for divergent natural selection at different spatial scales out of a total of 70 marker loci analysed, including outlier loci detected by divergence- and diversity-based tests

	Regional scale		European scale*										Sampling range†
Test type	Aus BM-Alps‡	Russ C-SE‡	Swe-Spain	Swe-Scot	Swe-Aus	Swe-Russ‡	Spain-Scot	Spain-Aus	Spain-Russ	Scot-Aus	Scot-Russ	Aus-Russ	Trem-Alba
Divergence§	0	3	7	8	2	2	1	16	9	6	9	5	n/a
Diversity¶	1	4	2	1	3	3	3	3	1	2	1	3	0
Shared	0	2	0	0	1	0	0	0	0	0	0	0	n/a

*The following population pairs were combined at the European scale: Aus/BM and Aus/Alps, and Russia/C and Russia/SE.
†Reduced diversity across the sampled range of *Populus tremula* was tested by using Central European material of *Populus alba* as a reference.
‡Loci under balancing selection within populations were detected in three comparisons: Aus/BM-Aus/Alps, 1 locus; Russ/C-Russ/SE, 4 loci; Swe-Russ, 1 locus. For details about all non-neutral 'outlier' loci see Table S2 (Supporting information).
§'Divergence' outlier loci were significant in both the drift based and the F_{ST} based approach with a FDR of 10%.
¶'Diversity' oulier loci were significant for either one of the two diversity ratios, lnRH or lnRV, at a FDR of 10%.
For expansions of abbreviations in the column head, see Table 1.

number of correlated traits for which measurements were available, timing of bud set was chosen for linear and nonlinear cline fitting in this study. Bud set reveals information about the timing of growth cessation, a trait of known functional significance connected to photoperiod adaptation along latitudinal gradients in temperate trees (Ingvarsson *et al.* 2006; Hall *et al.* 2007). Timing of bud set was measured repeatedly from spring to summer in 2005–2008, as outlined in Luquez *et al.* 2008. Relative growthrate was used as a fitness proxy and was measured as the cumulative growth over the entire period from 2004 to 2008. The cumulative growth includes the effects of frost damage suffered by some of the trees during winters. Frost damage sometimes results in reductions in height (negative growth) as the top (annual) shoots are most likely to be killed by frost.

Microsatellite genotyping

Seventy microsatellite loci were used for population genetic analysis, including 62 markers for which repeat and primer information are available at http://www.ornl.gov/sci/ipgc/ssr_resource.htm and eight additional marker loci. For detailed information about all loci, see Tables S1 and S2, Supporting information and van der Schoot *et al.* (2000), Smulders *et al.* (2001), Tuskan *et al.* (2004) and Yin *et al.* (2009).

All markers were prescreened for single-locus amplification, polymorphism and codominance in the course of a related project on the genetics of species isolation in *P. alba* and *P. tremula*, i.e. the markers were thoroughly tested for robustness prior to this study. Linkage relationships (for graphical representation and genetic

structure analysis; below) were obtained from available *Populus trichocarpa* maps and in a few instances were complemented using recombination data from a controlled cross *P. alba* × *P. tremula*. Genomic DNA was extracted from silica-dried leaves using the Dneasy Plant Mini Kit (QIAGEN), and all markers were polymerase chain reaction (PCR) amplified using the reaction conditions described previously by Lexer *et al.* (2005) and precisely sized using an Applied Biosystems (AB) 3100 Genetic Analyzer, making use of the florescent dyes FAM and JOE as well as size differences among loci for multiplexing. For four loci with particularly low levels of polymorphism in *P. tremula*, the presence of the microsatellite repeat was confirmed by direct sequencing of PCR products, following the protocols of Joseph & Lexer (2008).

Patterns of genomic diversity

All populations of *P. tremula* were characterized for their genetic diversity and structure, including allelic richness (corrected for the smallest sample size by rarefaction), expected (H_E) and observed (H_O) heterozygosity, and the inbreeding coefficient F_{IS} with 95% confidence intervals obtained by bootstrapping over loci (Table S1, Supporting information). Characterization of each locus also included the variance in allele size as a simple diversity parameter under a stepwise mutation model, and single-locus estimates for F_{ST} and Hedrick's (2005) G'_{ST} (Table S1, Supporting information). All *P*-values were adjusted for multiple tests using the Bonferroni method. Population bottlenecks were tested for by the sign test and Wilcoxon sign-rank test described by Piry *et al.* (1999), two standard tests that compare allele numbers with equilibrium gene diversities, and isolation-by-distance (IBD) was tested by comparing the matrices of genetic and geographic distances with a Mantel test. This test allows detecting departures from migration–drift equilibrium which may arise for a variety of reasons, e.g. because of geographic barriers to gene flow, or because of artificial founding of local populations from nonlocal source material. Linkage disequilibrium (LD) among loci was not a major focus of this study and information on this is already available for *P. tremula* (Lexer *et al.* 2007; Ingvarsson 2008). Nevertheless, we were interested in confirming far-ranging LD on the proximal end of chromosome 19 as predicted by Yin *et al.* (2008), which we did using exact tests following Rousset (2008) based on marker order inferred by genetic mapping (Table S1, Supporting information) and assuming *P. trichocarpa* physical map distances.

Population genetic structure was analysed using both conventional and Bayesian methods. Conventional analysis comprised UPGMA cluster analysis in PHYLIP based on Nei's standard genetic distance, using 1000 bootstrap replicates to determine branch support (Fig. 1a). Bayesian analysis was carried out with Structure 2.2 using two different approaches described by Falush *et al.* (2003), namely a standard admixture model assuming correlated allele frequencies and a linkage model which incorporated linkage information from *P. trichocarpa*. The latter assumes good synteny in *Populus*, which we expect at this relatively coarse genomic scale (Cervera *et al.* 2001). Multiple programme runs were conducted for each approach to confirm stabilization of the summary statistics. The two approaches yielded highly similar results, indicating $K = 6$ as the most likely number of gene pools or 'genetic units' based on the model likelihoods and their variances, and the results of the '$k = 6$' linkage model were chosen for presentation and discussion (Fig. 1b), based on a run with 50 000 burn-in followed by 100 000 iterations. Based on their low genetic differentiation in the Bayesian analysis and their close proximity, the two neighbouring populations Aus/BM and Aus/Alps were combined in 'European' scale neutrality tests, as were populations Russia/C and Russia/SE.

Genomic footprint of selection

Neutrality tests were based either on diversity or on divergence. Genetic divergence based tests utilized the migration and drift (F_{ST})-based approach discussed by Beaumont & Balding (2004) and the alternative drift-based approach of Vitalis *et al.* (2001). The former is known to be robust across a wide range of demographic scenarios, and the latter is useful when departures from equilibrium conditions (e.g. bottlenecks) are expected. The migration and drift-based approach was carried out with one initial round of coalescent simulations to estimate 'neutral' F_{ST} and a second round to identify candidate loci subject to selection. We preferred the frequentist approach described by Beaumont & Balding (2004) over the Bayesian method available because more experience exists with it. Parameter settings for the alternative pure drift-based approach were chosen based on available knowledge on the biogeographic history of European forest trees (Huntley & Birks 1983), and the results remained stable across a wide range of parameter settings. The results shown (Table 2; Table S2, Supporting information) are from 100 000 simulations with mutation rate (μ) = 0.01, 0.001, 0.0001; population size before bottleneck (N_e) = 500, 1000, 5000; time since bottleneck (T_0) = 50, 500, 1000 generations; population size before split (N_0) = 50, 500; time (t) since split: 50, 500.

Identifying the most realistic population model for divergence-based outlier detection is subject of intense

ongoing research (Foll & Gaggiotti 2008; Excoffier *et al.* 2009). Here, a combination of the Vitalis *et al.* (2001) and Beaumont & Balding (2004) methods were chosen because extensive experience exists for them. A stringent set of criteria was used to identify loci potentially subject to selection: each candidate locus had to be significant in the purely drift based *and* the F_{ST} based test at a false discovery rate (FDR) of 10% (Benjamini & Hochberg 2000). For completeness, candidate loci for balancing selection were recorded in the form of outliers in the lower tail of the F_{ST} distribution (alpha = 5%) for each pair of populations, but these were not a major focus of this study.

Diversity-based tests for selective sweeps made use of two different diversity ratios, namely the ratio of gene diversity (lnRH) and the ratio of the variance in repeat number (lnRV) for pairs of populations. Significance was assigned using the neutrality tests of Schlötterer (2002) and false positives were accounted for by using a FDR of 10%. We did not use a dynamically adjusted number of linked microsatellites to account for false positives (Wiehe *et al.* 2007) because the number of linked markers available was low. The two ratios, lnRH and lnRV, were interpreted separately because they may reflect different aspects of the genealogical history of a locus.

The divergence- and diversity-based tests were applied to pairwise population comparisons at three different spatial scales as outlined above (Population samples and common garden trial). Our rationale was to examine the interplay of gene flow and selection at these different spatial scales. With respect to timescales under investigation, assuming a generation time of 20 years for poplars yields no more than 600–700 generations since the end of the last ice age 13 000 years ago when most European forest trees were restricted to spatially disconnected refugia. Thus, the relatively long generation time of poplars (compared with annual plants) should render our data ideal for detecting post-glacial, locally varying selection with rapidly evolving microsatellites (Schlötterer 2003).

For each candidate locus found to be under locally varying selection at the European scale in at least three pairwise population comparisons, the nearest gene along the chromosome was identified using *P. trichocarpa* genome assembly version 1; version 2 of the genome assembly became available while this study was in review, but its browsing functions at that time were limited. Thus, we preferred the more evolved version 1, with the exception of one locus for which the chromosomal location in version 1 was unclear (footnote to Table 3).

Table 3 Twelve candidate loci for divergent selection identified in three or more pairwise population comparisons in *Populus tremula*, including marker name, linkage group (Lg), closest *Populus trichocarpa* gene model, approximate distance to the nearest gene in kilobases (kb), putative gene function, gene expression information and genomic space of the microsatellite repeat

Marker*	Lg	Closest *Populus* gene model†	Distance to gene (kb)†	Putative gene function	Expression information‡	Genomic space of repeat
G1376	2	fgenesh4_pg.C_LG_II002615	0.4	Transcriptional regulator	No	Intergenic
G1416	3	eugene3.00030612	2.1	26S proteasome regulatory subunit	Yes	Intergenic
ASP322	6	estExt_fgenesh1_pg_v1.C_LG_VI0144	0.0	Adenine phospho- ribosyltransferase	Yes	UTR
ASP933	6	eugene3.00061049	0.0	Amino oxidase	Yes	UTR
G1485	6	grail3.0023026901	1.4	Anion transporter	No	Intergenic
O268	8	fgenesh4_pg.C_LG_VIII001710	0.8	NADH kinase	No	Intergenic
G1353	13	fgenesh1_pg.C_LG_XIII000001	1.0	Pectinesterase inhibitor	No	Intergenic
G1608	15	grail3.0082003901	7.6	NADH2 dehydrogenase	Yes	Intergenic
O14	16	gw1.XVI.1041.1	3.0	Alpha-L-Arabinofuranosidase	Yes	Intergenic
O276	19	gw1.117.48.1	0.1	NBS-LRR–type disease resistance gene	No	Intergenic
Con58.1§	19	grail3.0117003601	3.3	Small auxin-upregulated RNA	Yes	Intergenic
Yin2	19	gw1.XIX.66.1	0.0	Oxysterol-binding protein	Yes	Exon

*Markers with G refer to GCPM, markers with O refer to ORPM microsatellite loci on the *Populus* genome web page (http:// ipgc/ ssr_resource.htm). For complete information about all marker loci including detailed references see Table S1 (Supporting information).

†Gene models and approximate kilobase (kb) distances refer to *P. trichocarpa* genome assembly version 1, except for G1608. This lorus did not map clearly in assembly version 1 but was found on chromosome 15 in *P. trichocarpa* genome assembly version 2. See Materials and methods for the use of the two genome assemblies.

‡Indicates whether or not expressed sequence tag (EST) information is available in JGI *Populus* genome browser version 1.1.

§Locus located on scaffold 117, known to be homologous to *P. trichocarpa* chromosome 19.

Phenotypic clines along a latitudinal gradient

Data on bud set and growth were taken from Luquez *et al.* (2008), where the SwAsp collection and common garden experiments are also described in detail. Clone-specific breeding values for bud set and growth were estimated using linear mixed-model Best Linear Unbiased Predictors (BLUPs) as described by Luquez *et al.* (2008). Within-population variation in bud set was then estimated from variation among clonal BLUPs within the 12 original populations of the SwAsp collection. The relationship between bud set and latitude was studied using a simple linear model ($y = a + b*x$) and a more complex, nonlinear model which allowed for a steeper slope in the centre of the cline ($y = a + b/[1 + \exp[c*x]]$). The two models were fitted to the data using linear and nonlinear regression using the *lm* and *nls* functions in *R*. The two models are not nested and the best-fitting model was therefore selected based on Akaike's Information Criteria (AIC), with $AIC = 2*\log(L) + 2*n$, where L is the likelihood and n is the number of free parameters of a given model (two for the linear model and three for the non-linear model). Selection differentials were calculated within each population in the form of covariances between bud set and fitness (Lande & Arnold 1983), using relative growthrate as a fitness proxy. These analyses were first carried out for the Sävar garden only (the trial with the slightly greater level of replication), then for both gardens combined. Since the result remained the same, the results of the larger, combined data set are presented and discussed below. Note that growth rate is just one of many components of fitness and may thus not fully capture fitness tradeoffs and lifetime reproductive output; the use of partial fitness proxies is common in evolutionary ecology because lifetime fitness is often difficult to measure (Lande & Arnold 1983).

Locus-specific ancestries in the Swedish admixture zone

Genomic admixture and marker locus-specific ancestries were estimated for the Northern European admixture zone based on microsatellite genotype data for trees from Scotland, Sweden and Russia, using the R-script INTROGRESS (Z. Gompert and C. A. Buerkle) which analyses admixture gradients and genomic clines (Gompert & Buerkle 2009). Overall admixture was estimated in the form of a maximum-likelihood (ML) hybrid index and 95% confidence intervals. The ancestry of each locus in each individual was assessed in terms of homo- or heterozygosity for Scottish and Russian alleles, with multi-allelic microsatellites binned into two informative allelic classes per locus as described in Lexer *et al.* (2007). Four loci were removed as they yielded no information about the population origin of alleles in admixed individuals. Scotland and Russia were used to estimate parental allele frequencies for admixture analysis, because the Structure results pointed to these two as predominant source populations contributing to the admixed Swedish sample.

Results

Population differentiation and admixture

Our analysis of molecular genetic variability for European aspen – the most comprehensive molecular data set available for this species to date in terms of geographic and genomic coverage – revealed distinct patterns of diversity and differentiation at the European scale. Whereas overall genomic divergence was low (average F_{ST} over loci = 0.051, 95% CI = 0.040–0.063; average G'_{ST} = 0.181, 95% CI = 0.152–0.211), significant IBD was detected among populations ($r = 0.601$; $P < 0.005$; $R^2 = 36\%$), thus indicating an important role for gene flow and drift in shaping patterns of molecular diversity. Accordingly, geographic structure was detectable by both conventional (Fig.1a; Table S3, Supporting information) and Bayesian (Fig. 1b) methods. Conventional cluster analysis revealed geographic structure but several nodes had <60% bootstrap support, affecting populations from Central, Northern and Eastern Europe (Fig. 1a). Bayesian analysis revealed the likely cause: extensive admixture between Eastern and Western European lineages in Scandinavia and, to a lesser extent, in Central Europe (Fig. 1b, populations Swe and Aus/BM). Note that *all* models from $K = 4$ to $K = 8$ yielded congruent results for Sweden, namely admixture between a Western and an Eastern European lineage (not shown). Bayesian analysis also revealed the presence of several groups of closely related genotypes in Spain (Fig. 1b). This population has undergone a recent genetic bottleneck, detectable by standard tests that compare allele numbers and equilibrium gene diversities (sign test: $P < 0.001$; Wilcoxon test: $P < 0.001$).

In general, microsatellite diversities were intermediate ($H_E = 0.563 \pm 0.030$ SE; $H_O = 0.468 \pm 0.029$ SE), i.e. low enough for meaningful F_{ST}-based analysis of locally varying selection using frequentist approaches. A slight homozygote surplus (positive inbreeding coefficient F_{IS}) was detectable for all populations, the 95% bootstrap interval for overall F_{IS} ranging from 0.083 to 0.173, and this was attributable primarily to cryptic population subdivision (see Materials and methods).

Genomic footprint of locally varying selection

Our stringent divergence-based tests for locally varying selection (a locus had to be significant in both the purely drift-based and the F_{ST}-based tests at a FDR of 10%; see Materials and methods) yielded between 1 and 16 candidate loci for divergent selection at the European scale (6 loci ± 1 SE across all 10 pairwise population comparisons at this geographic scale) and smaller numbers of candidate loci in regional comparisons (Table 2; Table S2, Supporting information).

Diversity-based tests for selective sweeps (at FDR = 10%) yielded between 1 and 3 sweep loci at the European scale, between 1 and 4 at the regional scale and no range-wide sweep was found when *Populus tremula* was compared with Central European populations of its closely related sister species *Populus alba*. Except for three 'sweep loci' with locally reduced diversity, there was no overlap between candidate 'outlier' loci detected by the divergence- and the diversity-based tests (Table 2; Table S2, Supporting information). Six loci were found to be under balancing selection within populations, five of them in regional comparisons.

The top 12 candidate loci for locally varying selection (loci that matched our stringent significance criteria in at least three pairwise population comparisons) comprised three loci located within expressed genes and four other loci that were located 1 kb or closer to the nearest gene model predicted by the *Populus trichocarpa* genome project (average distance to nearest gene: 1.6 ± 0.6 kb SE; Table 3). The gene list includes, among others, a transcription factor and genes involved in plant defence, such as a NBS-LRR-type resistance gene and a pectinesterase inhibitor. Two of the genes are located at the proximal end of chromosome 19, recently put forward as an incipient sex chromosome in *Populus* (Yin *et al.* 2008). Significant within-population LD between markers in this region extended over >400 kb in our study, assuming *P. trichocarpa* physical map distances (markers Con 58.1 and Yin2; details not shown), as opposed to the general prediction of a decay of LD within a few 100 bp in this species (Ingvarsson 2008). Population comparisons including the bottlenecked Spanish and the admixed Swedish population contributed rather moderately to the detection of replicated non-neutral outlier loci: except for loci O268, G1608 and Con58, all loci in Table 3 would have been detected as replicated outliers even without considering Spain, and except for G1608, Con58 and Yin2 all loci would have been detected without considering the admixed Swedish population (Table S2, Supporting information). The contribution of Spain and Sweden to all neutrality tests is visible in Table 2. Likewise, the commonly observed tendency across loci and populations for inbreeding coefficients (F_{IS}) to be positive (Wahlund effects; see Materials and methods; Table 1; Table S1, Supporting information) had little influence on our inference of selection, i.e. there was no significant difference between F_{IS} for the top 12 candidate outlier loci shown in Table 3 and the remaining loci (nonparametric Mann–Whitney *U*-test, $P = 0.35$).

The role of admixture in shaping phenotypic clines in Scandinavia

A nonlinear model provided a better fit for the phenotypic cline for bud set across a latitudinal gradient in Sweden compared with a linear model [Fig. 2a; linear: Akaike's Information Criterion (AIC) = 68.3; nonlinear: AIC = 49.3]. The difference in model fits was caused by a step in the cline between 58° and 63° of latitude (Fig. 2a). This is just south of the admixed Swedish material sampled in the microsatellite-based genome scan (Figs 1a,b). The variance for bud set was greatly increased in this area (Fig. 2b), and selection differentials (covariances between bud set and relative growth-rate as a fitness proxy) were elevated in the contact zone as well (Fig. 2c).

Considerable within-genome variation for marker ancestry was observed when Scotland and Russia were used as proxies for the source gene pools that contributed to the admixed Swedish population (Fig. 3a; choice of Scotland and Russia as source populations motivated by Fig. 1). Whereas a complete admixture gradient was recovered across these populations for the genome-wide panel of 70 microsatellites (Fig. 3b), the ancestries of individual loci varied greatly along the chromosomes (Fig. 3a). This can easily be illustrated by comparing chromosome 6, known to exhibit normal recombination rates (Yin *et al.* 2004), and chromosome 19, known to exhibit greatly reduced recombination for the region studied (Yin *et al.* 2008). Ancestry varied greatly across chromosome 6, its left end exhibiting a noticeable excess of Russian ancestry (light green) and the right half being predominantly Scottish (dark green). In contrast, locus-specific ancestries changed gradually from Scottish to Russian along chromosome 19 (Fig. 3c), consistent with a history of lower recombination.

Discussion

The role of standing variation in adaptive evolution is a topic of fundamental importance for evolutionary biology (Colosimo *et al.* 2005; Stearns & Hoekstra 2005; Pennings & Hermisson 2006; Barrett & Schluter 2008; Mullen & Hoekstra 2008). As beneficial alleles present as standing variation are older than new mutations,

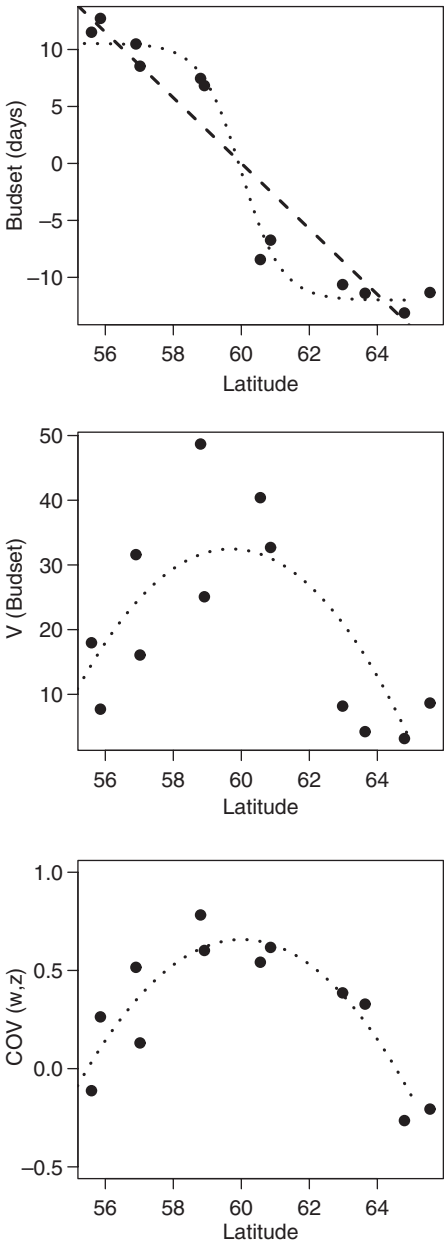

Fig. 2 Cline shapes, trait variances and selection differentials for timing of bud set, a trait of known adaptive significance in deciduous temperate trees, measured in 12 populations of *Populus tremula* from across a latitudinal gradient in Scandinavia grown in two common gardens. (a) Best Linear Unbiased Predictors (BLUPs, expressed in deviations from mean bud set date), with regression curves for linear and non-linear model fits indicated by dashed and dotted lines, respectively. (b) Trait variances, nonlinear fit indicated by dotted line. (c) Selection differentials, estimated as the covariance between bud set and relative growth rate as a fitness proxy, nonlinear fit indicated by dotted line.

their current spatial distributions will reflect key aspects of a species' biogeographic history, particularly in temperate taxa which must undergo regular cycles of range contractions and expansions to track Earth's climatic oscillations (Hewitt 2000). This implies that beneficial alleles have been pre-tested by selection in past environments or in other parts of a species' range (Kremer *et al.* 2002; Barrett & Schluter 2008). In many cases, this past 'selective filter' may even have operated in a different species, followed by interspecific gene flow and recombination in the form of adaptive introgression (Arnold 2006) or recombinational speciation (Rieseberg *et al.* 2003).

Our molecular marker-based scan indicates that locally varying selection is detectable in the wind-pollinated, dioecious *Populus tremula* at the European scale (Table 2; Table S2, Supporting information). This implies linkage of putatively 'neutral' marker alleles with selected coding or regulatory mutations or heritable epialleles, the presize nature of the causal molecular changes under selection being unknown at this stage of the work. At first sight, our detection of locally varying selection is at odds with the known high levels of inter-population gene flow in this species (N_em = up to 15 effective migrants per generation; Lexer *et al.* 2005) and the rapid decay of LD observed at the DNA sequence level (within just a few 100 bp, Ingvarsson 2008). The finding in this study of an average of 8% non-neutral outlier loci (±2% SE across all population comparisons) in our stringent divergence-based tests (Tables 2 and 3) indicates that selection (s) on particular genome regions in *P. tremula* is stronger *or* recombination (r) in these regions is lower than previously thought.

A role for low recombination is indicated by the fact that 3 of the 12 loci with consistent and replicated departures from neutrality in our study (Table 3) are located in the proximal end of chromosome 19, thought to be an incipient sex chromosome conserved across different sections of the genus *Populus* (Yin *et al.* 2008). Within-population LD in this region is greatly elevated in the studied populations (extending up to 400 kb, assuming *Populus trichocarpa* physical map distances). Elevated LD in this region is expected based on genomic sequencing, scaffold assembly and genetic mapping of this region in *P. trichocarpa* (Yin *et al.* 2008). An indication for selection, however, is that 10 of the 12 replicated outlier polymorphisms detected in this study were located less than 3 kb from the nearest gene, and three markers were even located within transcribed sequences (Table 3). This makes it plausible that genes or *cis*-regulatory elements adjacent to the studied polymorphisms were indeed the targets of selection. Of course, the causal molecular changes (mutations or heritable epigenetic modifications) under locally varying selection may also reside in a different gene or DNA region adjacent to each microsatellite. At the very least, our cross-check of the 12 top candidates against the

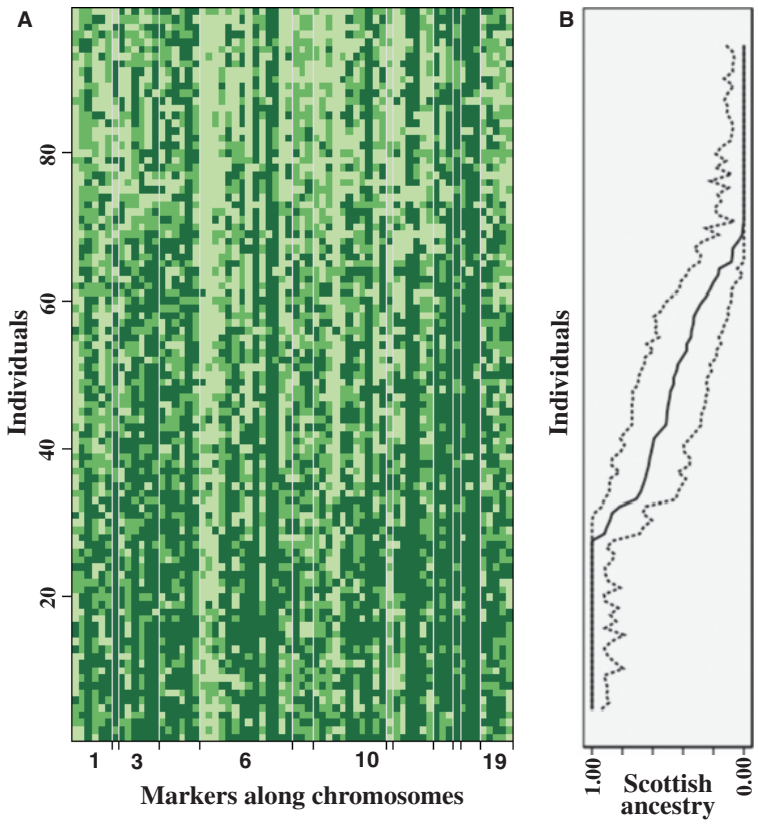

Fig. 3 Admixture and genomic variation for marker ancestry in trees from the Northern European admixture zone of *Populus tremula*. (a) Genomic variation for marker ancestry in terms of homo- or heterozygosity for Scottish and Russian alleles, respectively, with multi-allelic microsatellites binned into two informative allelic classes for each locus. Dark green, homozygotes for Scottish allele; light green, homozygotes for Russian allele; medium green, heterozygotes for Scottish and Russian alleles. Markers along chromosomes are shown in consecutive order on the horizontal axis, chromosomes separated by thin white lines, and individuals are shown on the vertical axis. The five chromosomes with the best genomic coverage are indicated by arabic numbers along the horizontal axis; for all other chromosomal locations, see Table S1 (Supporting information). (b) Overall admixture and 95% confidence intervals for the trees shown, corresponding exactly to the individuals seen along the vertical axis in (a).

Populus genome assembly (Table 3) suggests that all 12 are situated within gene-rich regions.

Two conspicuous findings of our population genomic work were the small number of non-neutral outlier loci detected by diversity based tests for 'selective sweeps' (on average just 3% of the studied loci ±0.4% SE), and the limited overlap between outliers found with divergence vs. diversity based tests (Table 2). This pattern was not caused by the presence of the admixed Swedish and the bottlenecked Spanish populations in the data set, e.g. population comparisons of Scotland–Austria, Scotland–Russia and Austria–Russia followed the same trend (Table 2). The pattern was also not caused by errors in the spatial scale chosen for analysis, as the number of selective sweeps was low at all three geographic scales examined (Table 2). On the contrary, this pattern is expected if local adaptation occurs primarily from standing variation (soft sweeps) rather than from new mutations (hard sweeps; Pennings & Hermisson 2006). Soft sweeps will result in a severe reduction of the chromosomal width of sweep regions (hence a reduction of statistical power to detect selection) compared with hard sweeps, and this is exactly what was observed in this study: locus-specific reductions in diversity were rare (Table 2), and they almost never affected pairs of adjacent loci, whereas the latter was commonly observed for divergence-based tests (Table S2, Supporting information). Note that low levels of neutral divergence as found in *P. tremula* (average F_{ST} = 0.051 ± 0.047 SD) greatly facilitate the detection of locally varying selection in divergence-based tests, because selected loci will emerge from the neutral distribution more clearly.

Although distinguishing between young soft sweeps and old hard sweeps is not trivial, the presence of soft sweeps is expected from theory when the scaled mutation rate, $\Theta = 2N_e\,\mu$, exceeds 0.01 (Pennings & Hermisson 2006). This threshold is easily exceeded in *P. tremula* for a wide range of possible mutation rates (μ), because effective population size (N_e) is in the order of 10^5 (Ingvarsson 2008). Theory also predicts that migration from geographically disconnected source gene pools (e.g. glacial refugia) will contribute to the standing variation available for local adaptation, in a manner similar to recurrent mutation (Pennings & Hermisson 2006). This is what we find here for European populations of the widespread forest tree *P. tremula*.

Our Bayesian structure analysis based on 70 mapped loci clearly indicates genetic admixture between a Western and an Eastern European lineage of *P. tremula* in Sweden, with possible additional contributions from Central European populations (Fig. 1b). The known biogeographic history of Scandinavia (ice cover during last glaciation and recent recolonisation following land

uplift) allows us to exclude that this genetic pattern was produced by anything other than postglacial intraspecific admixture. The Northern European admixture zone of *P. tremula* reported here coincides with a postulated zone of postglacial contact affecting many species of animals and plants (Hewitt 2000; Tollefsrud *et al.* 2008), including bears, rodents, conifers and orchids.

Admixture of postglacial lineages of *P. tremula* in Sweden is the simplest explanation for the significant step in the geographic cline for bud set (Fig. 2a), a heritable trait of clear adaptive significance in temperate trees. Of course, clinal variation in Scandinavia cannot be attributed to admixture alone: the critical daylength for initiating growth cessation varies linearly across the range of the studied populations (Hall *et al.* 2007), and adaptation of forest trees to local photoperiod is known to generate clinal variation in important phenological traits (Howe *et al. 2003*). Accordingly, DNA polymorphisms within the phytochrome B2 locus exhibit clinal variation across the same geographic gradient (Ingvarsson *et al.* 2006), whereas multiple neutral polymorphisms located elsewhere in the genome do not (Hall *et al.* 2007). Admixture contributes to the variation available for adaptation along the gradient, as visible from the step in the cline (Fig. 2a) and from two other observations.

First, the variance of bud set is greatly elevated in the centre of the cline, consistent with increased genetic variation available for selection, stemming from admixture of differentiated gene pools (Fig. 2b). Second, selection differentials (estimated as the covariance between bud set and vegetative fitness) are also elevated in the centre of the cline (Fig. 2c). Thus, admixture between differentiated postglacial lineages contributes to the standing variation available for natural selection and adaptation, as predicted by theory (Arnold 2006; Pennings & Hermisson 2006; Barrett & Schluter 2008). It is unlikely that admixture constrains adaptation in *P. tremula*: no signs of outbreeding depression (=reduced vegetative fitness) were observed in central Sweden, and geographic clines for fitness-related growth traits are shallow and linear as expected from the gradual change in the length of the growing season (Hall *et al.* 2007; Luquez *et al.* 2008). Our data provide the necessary, hitherto missing direct link between range shifts because of past climatic oscillations on the one hand and current levels of standing variation available for adaptive evolution on the other. Our three-pronged approach (genomic scan for selectively differentiated DNA regions, common garden measurements of adaptive trait differentiation, map-based analysis of genomic ancestry) allowed us to demonstrate the effects of admixture on standing variation at a depth rarely (if ever) seen in previous studies of clinal variation in ecological keystone or foundation species.

The most reliable way to clarify the exact source of beneficial alleles in any species is to map and isolate the genes responsible for adaptation, estimate their fitness effects and establish their genealogical histories (Colosimo *et al.* 2005; Barrett & Schluter 2008). In long-lived organisms such as trees, such studies are greatly facilitated by emerging approaches to identify fitness-related genes in natural populations (Savolainen & Pyhajarvi 2007), particularly 'admixture mapping' in intraspecific admixture or contact zones (Buerkle & Lexer 2008). In the case of *P. tremula*, the chromosomal variation in marker ancestry required for admixture mapping appears to be present in the Scandinavian contact zone, as exemplified by the mosaic-like nature of chromosome 6, consisting of blocks of DNA of either 'Scottish' or 'Russian' ancestry (Fig. 3).

Understanding the role of standing variation in facilitating rapid adaptation to new environments is a crucial task for evolutionary biologists, because rapid evolution will be required for species' survival in the face of human-mediated environmental perturbations (Barrett & Schluter 2008). This is particularly relevant for long-lived organisms, such as trees (Davis & Shaw 2001; Petit & Hampe 2006; Savolainen & Pyhajarvi 2007). Detecting the missing links between biogeographic history and current levels of standing variation available for adaptation is a key to both science-based conservation and responsible exploitation of wild species in a changing world.

Acknowledgements

We thank M. Fay and M. van Loo for help in the laboratory, T. Barbará for help with data analysis; C. A. Buerkle, S. P. DiFazio, T.M. Yin, A. Widmer and S. Gonzalez-Martinez for helpful discussions; the *Populus* genome consortium for sharing information and four anonymous referees for reading the manuscript. The research was financially supported in part by grants from the Brazilian CNPq to DdC, from the Russian RFBR to IS, and from the British NERC to CL.

References

Arnold ML (2006) *Evolution Through Genetic Exchange*. Oxford University Press, Oxford.

Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology and Evolution*, **23**, 38–44.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology Evolution and Systematics*, **16**, 113–148.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

Benjamini Y, Hochberg Y (2000) On the adaptive control of the false discovery rate in multiple testing with independent

statistics. *Journal of Educational and Behavioral Statistics*, **25**, 60–83.
Briscoe D, Stephens J, O'Brien S (1994) Linkage disequilibrium in admixed populations: applications in gene mapping. *Journal of Heredity*, **85**, 59–63.
Buerkle CA, Lexer C (2008) Admixture as the basis for genetic mapping. *Trends in Ecology and Evolution*, **23**, 686–694.
Cervera MT, Storme V, Ivens B *et al.* (2001) Dense genetic linkage maps of three *Populus* species (*Populus deltoides*, *P. nigra* and *P. trichocarpa*) based on AFLP and microsatellite markers. *Genetics*, **158**, 787–809.
Colosimo PF, Hosemann KE, Balabhadra S *et al.* (2005) Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*, **307**, 1928–1933.
Davis MB, Shaw RG (2001) Range shifts and adaptive responses to quaternary climate change. *Science*, **292**, 673–679.
Ericson L, Wallentinus HG (1979) Sea-shore vegetation around the Gulf of Bothnia. *Wahlenbergia*, **5**, 1–142.
Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**, 285–298.
Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.
Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–993.
Gompert Z, Buerkle CA (2009) A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Molecular Ecology*, **18**, 1207–1224.
Hall D, Luquez V, Garcia VM *et al.* (2007) Adaptive population differentiation in phenology across a latitudinal gradient in European aspen (*Populus tremula*, L.): a comparison of neutral markers, candidate genes and phenotypic traits. *Evolution*, **61**, 2849–2860.
Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution*, **5**, 1633–1638.
Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.
Howe GT, Aitken SN, Neale DB *et al.* (2003) From genotype to phenotype: unraveling the complexities of cold adaptation in forest trees. *Canadian Journal of Botany*, **81**, 1247–1266.
Huntley B, Birks H (1983) *An Atlas of Past and Present Pollen Maps of Europe: 0–13 000 Years ago*. Cambridge University Press, Cambridge.
Ingvarsson PK (2008) Multilocus patterns of nucleotide polymorphism and the demographic history of *Populus tremula*. *Genetics*, **180**, 329–340.
Ingvarsson PK, Garcia MV, Hall D, Luquez V, Jansson S (2006) Clinal variation in phyB2, a candidate gene for daylength-induced growth cessation and bud set, across a latitudinal gradient in European aspen (*Populus tremula*). *Genetics*, **172**, 1845–1853.
Joseph JA, Lexer C (2008) A set of novel DNA polymorphisms within candidate genes potentially involved in ecological divergence between *Populus alba* and *P. tremula*, two hybridizing European forest trees. *Molecular Ecology Resources*, **8**, 188–192.
Kremer A, Kleinschmit J, Cottrell J *et al.* (2002) Is there a correlation between chloroplastic and nuclear divergence, or what are the roles of history and selection on genetic diversity in European oaks? *Forest Ecology and Management*, **156**, 75–87.
Kruuk LEB, Baird SJE, Gale KS, Barton NH (1999) A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids. *Genetics*, **153**, 1959–1971.
Lande R, Arnold SJ (1983) The measurement of selection on correlated characters. *Evolution*, **37**, 1210–1226.
Lexer C, Fay M, Joseph J, Nica M-S, Heinze B (2005) Barrier to gene flow between two ecologically divergent *Populus* species, *P. alba* (white poplar) and *P. tremula* (European aspen): the role of ecology and life history in gene introgression. *Molecular Ecology*, **14**, 1045–1057.
Lexer C, Buerkle CA, Joseph JA, Heinze B, Fay MF (2007) Admixture in European *Populus* hybrid zones makes feasible the mapping of loci that contribute to reproductive isolation and trait differences. *Heredity*, **98**, 183–184.
Luquez V, Hall D, Albrectsen BR *et al.* (2008) Natural phenological variation in aspen (*Populus tremula*): the SwAsp collection. *Tree Genetics and Genomes*, **4**, 279–292.
Mullen LM, Hoekstra HE (2008) Natural selection along an environmental gradient: a classic cline in mouse pigmentation. *Evolution*, **62**, 1555–1569.
Pennings PS, Hermisson J (2006) Soft sweeps II-molecular population genetics of adaptation from recurrent mutation or migration. *Molecular Biology and Evolution*, **23**, 1076–1084.
Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. *Annual Review of Ecology Evolution and Systematics*, **37**, 187–214.
Petit RJ, Aguinagalde I, de Beaulieu JL *et al.* (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science*, **300**, 1563–1565.
Piry S, Luikart G, Cornuet JM (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, **90**, 502–503.
Przeworski M, Coop G, Wall JD (2005) The signature of positive selection on standing genetic variation. *Evolution*, **59**, 2312–2323.
Reich D, Patterson N, De Jager PL *et al.* (2005) A whole-genome admixture scan finds a candidate locus for multiple sclerosis susceptibility. *Nature Genetics*, **37**, 1113–1118.
Rieseberg LH, Burke JM (2001) The biological reality of species: gene flow, selection, and collective evolution. *Taxon*, **50**, 47–67.
Rieseberg LH, Raymond O, Rosenthal DM *et al.* (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science*, **301**, 1211–1216.
Rousset F (2008) GENEPOP '007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.
Savolainen O, Pyhajarvi T (2007) Genomic diversity in forest trees. *Current Opinion in Plant Biology*, **10**, 162–167.
Schemske DW (2000) Understanding the origin of species. *Evolution*, **54**, 1069–1073.
Schlötterer C (2002) A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics*, **160**, 753–763.
Schlötterer C (2003) Hitchhiking mapping-functional genomics from the population genetics perspective. *Trends in Genetics*, **19**, 32–38.

van der Schoot J, Pospiskova M, Vosman B, Smulders MJM (2000) Development and characterization of microsatellite markers in black poplar (*Populus nigra* L.). *Theoretical and Applied Genetics*, **101**, 317–322.

Smulders MJM, 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.

Stearns SC, Hoekstra RF (2005) *Evolution: An Introduction*. Oxford University Press, Oxford.

Tollefsrud MM, Kissling R, Gugerli F *et al.* (2008) Genetic consequences of glacial survival and postglacial colonization in Norway spruce: combined analysis of mitochondrial DNA and fossil pollen. *Molecular Ecology*, **17**, 4134–4150.

Tuskan GA, Gunter LE, Yang ZK *et al.* (2004) Characterization of microsatellites revealed by genomic sequencing of *Populus trichocarpa*. *Canadian Journal of Forest Research*, **34**, 85–93.

Vitalis R, Dawson K, Boursot P (2001) Interpretation of variation across marker loci as evidence of selection. *Genetics*, **158**, 1811–1823.

Whibley AC, Langlade NB, Andalo C *et al.* (2006) Evolutionary paths underlying flower color variation in *Antirrhinum*. *Science*, **313**, 963–966.

Wiehe T, Nolte V, Zivkovic D, Schlotterer C (2007) Identification of selective sweeps using a dynamically adjusted number of linked microsatellites. *Genetics*, **175**, 207–218.

Wu CI (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865.

Yin TM, DiFazio SP, Gunter LE *et al.* (2004) Large-scale heterospecific segregation distortion in *Populus* revealed by a dense genetic map. *Theoretical and Applied Genetics*, **109**, 451–463.

Yin T, DiFazio SP, Gunter LE *et al.* (2008) Genome structure and emerging evidence of an incipient sex chromosome in Populus. *Genome Research*, **18**, 422–430.

Yin TM, Zhang XY, Gunter LE *et al.* (2009) Microsatellite primer resource for *Populus* developed from the mapped sequence scaffolds of the Nisqually-1 genome. *New Phytologist*, **181**, 498–503.

Christian Lexer's team is interested in the evolutionary genomics of adaption and speciation, using Populus and other plants as study systems. Dulcineia de Carvalho is a group leader in Conservation Genetics in Brazil, she carried out this study as a visiting professor in C.L.'s lab. Jeffrey Joseph, Leonie Suter, and Claudia Sedivy were members of C.L.'s group. Pär Ingvarsson, Joan Cottrell, Berthold Heinze, and Ivan Schanzer are group leaders with broad interests in plant evolutionary and conservation genetics; they contributed to this work via sample and data collection, analysis, and interpretation of results.

Supporting Information

Table S1 Microsatellite marker loci used in this study, including locus names, linkage groups (Lg), allele numbers (A), expected (H_E) and observed (H_O) heterozygosities, inbreeding coefficients F_{IS}, and genetic divergence in the form of F_{ST} and Hedrick's (2005) G'_{ST}

Table S2 Results of marker-based genome scans for non-neutral population divergence based on divergence and diversity based tests, pairwise population comparisons arranged side-by-side on separate pages to facilitate inspection of locus-specific outlier patterns across populations and to allow printing of particular population comparisons if desired

Table S3 Matrix of F_{ST} values for pairs of populations used in divergence- and diversity based neutrality tests at the European scale