New Phytologist

Research

Genome-wide patterns of differentiation and spatially varying selection between postglacial recolonization lineages of *Populus alba* (Salicaceae), a widespread forest tree

Kai N. Stölting[1], **Margot Paris**[1], **Cécile Meier**[1], **Berthold Heinze**[2], **Stefano Castiglione**[3], **Denes Bartha**[4] **and Christian Lexer**[1]

[1]Unit of Ecology and Evolution, Department of Biology, University of Fribourg, Chemin du Musee 10, 1700 Fribourg, Switzerland; [2]Department of Forest Genetics, Austrian Research Centre for Forests (BFW), Seckendorff-Gudent Weg 8, 1130 Vienna, Austria; [3]Department of Chemistry, University of Salerno, 84084 Fisciano, Italy; [4]Department of Botany, West-Hungarian University, 9400 Sopron, Hungary

Author for correspondence:
Christian Lexer
Tel: +41 26 300 8868
Email: christian.lexer@unifr.ch

Received: *22 December 2014*
Accepted: *25 February 2015*

New Phytologist (2015) **207:** 723–734
doi: 10.1111/nph.13392

Key words: adaptation, divergence continuum, divergent selection, population genomics, *Populus*, postglacial recolonization, selective sweep, whole-genome sequencing.

Summary

- Studying the divergence continuum in plants is relevant to fundamental and applied biology because of the potential to reveal functionally important genetic variation. In this context, whole-genome sequencing (WGS) provides the necessary rigour for uncovering footprints of selection.
- We resequenced populations of two divergent phylogeographic lineages of *Populus alba* ($n = 48$), thoroughly characterized by microsatellites ($n = 317$), and scanned their genomes for regions of unusually high allelic differentiation and reduced diversity using > 1.7 million single nucleotide polymorphisms (SNPs) from WGS. Results were confirmed by Sanger sequencing.
- On average, 9134 high-differentiation (≥ 4 standard deviations) outlier SNPs were uncovered between populations, 848 of which were shared by ≥ three replicate comparisons. Annotation revealed that 545 of these were located in 437 predicted genes. Twelve percent of differentiation outlier genome regions exhibited significantly reduced genetic diversity. Gene ontology (GO) searches were successful for 327 high-differentiation genes, and these were enriched for 63 GO terms.
- Our results provide a snapshot of the roles of 'hard selective sweeps' vs divergent selection of standing genetic variation in distinct postglacial recolonization lineages of *P. alba*. Thus, this study adds to our understanding of the mechanisms responsible for the origin of functionally relevant variation in temperate trees.

Introduction

Current progress in DNA sequencing technologies and computational biology facilitates fresh insights into genome-wide patterns of variation in wild species, and into the roles this variation may play in responses of organisms to environmental change (Anderson *et al.*, 2011; Orsini *et al.*, 2013). A useful family of concepts for approaching these issues revolves around the genomics of the divergence continuum, that is, the continuous gradient from differentiation between local populations to complete speciation (Schluter, 2000; Nosil *et al.*, 2009; Feder *et al.*, 2012).

Recently developed genotyping-by-sequencing approaches (Hohenlohe *et al.*, 2010; Elshire *et al.*, 2011) and whole-genome resequencing make it increasingly feasible to study the genomic architecture of divergence in many groups of animals and plants (Cao *et al.*, 2011; Ellegren *et al.*, 2012; Jones *et al.*, 2012; Ellegren, 2013; Renaut *et al.*, 2013). At the same time, new conceptual developments in population genetics (Hermisson & Pennings, 2005; Pritchard & Di Rienzo, 2010; Pritchard *et al.*, 2010), quantitative genetics (Le Corre & Kremer, 2012), and speciation biology (Smadja & Butlin, 2011; Seehausen *et al.*, 2014) have started to transform the way we think about the origin, dynamics, and fate of the genetic variation that forms the raw material for population divergence.

One set of questions of great current interest to the evolutionary genetics of diverging populations relates to the genomic architecture of differential adaptation, and of spatially varying selection more generally. Much has been learned about these topics in humans (Manolio *et al.*, 2009; Gibson, 2012), animals more generally (Barrett & Schluter, 2008; Rubin *et al.*, 2010), and plants (Fournier-Level *et al.*, 2011; Neale & Kremer, 2011) in recent years, based on a firm theoretical foundation (Orr, 1998; Barrett & Schluter, 2008; Pritchard *et al.*, 2010) and enabled by rapid progress in genomics. Recent studies of this topic suggest complex genomic architectures of adaptation – and complex histories of the underlying genetic variation – in species of animals and plants (Barrett & Schluter, 2008; Manolio *et al.*, 2009; de Carvalho *et al.*, 2010; Pritchard *et al.*, 2010; Neale & Kremer, 2011; Gibson, 2012). Understanding these issues is particularly relevant in temperate forest trees, where the functionally important, locally selected variation studied by evolutionary geneticists is potentially of direct relevance to breeding, forest management, and ecosystem restoration (Eckenwalder, 1996; Whitham *et al.*, 2006; Neale & Kremer, 2011; Slavov *et al.*, 2012). Important open research questions in this context revolve around genomic patterns of differentiation between divergent conspecific populations, the functional roles of genes or genetic elements affected by natural selection between populations (Neale & Kremer, 2011; Slavov *et al.*, 2012; Evans *et al.*, 2014), and the relative roles of adaptation from new mutations ('hard sweeps') vs standing genetic variation ('soft sweeps') during evolutionary responses to environmental change (Hermisson & Pennings, 2005; de Carvalho *et al.*, 2010; Pritchard *et al.*, 2010).

The 'model forest tree' genus *Populus* (poplars, aspens, cottonwoods) represents a prime example for the divergence continuum present in complexes of ecologically important and geographically widespread species (Eckenwalder, 1996; Whitham *et al.*, 2006; Jansson & Douglas, 2007). Genetic differentiation in *Populus* has been studied at several stages of the divergence continuum, ranging from locally adapted populations of the same species (Ingvarsson, 2005; de Carvalho *et al.*, 2010; Slavov *et al.*, 2012; Bernhardsson *et al.*, 2013) and taxa originating from recent (pleistocene) speciation events (Levsen *et al.*, 2012; Wang *et al.*, 2014) to much older species that appear to have diverged for millions of yr despite recurrent episodes of gene flow (Lindtke *et al.*, 2012; Stölting *et al.*, 2013).

Most available evolutionary genetic studies in *Populus* have not yet made use of the full power of whole genome scanning for resolving patterns of population divergence, with notable exceptions (Slavov *et al.*, 2012; Evans *et al.*, 2014; McKown *et al.*, 2014). The most extensive of these studies (Evans *et al.*, 2014) revealed clear associations between genome-wide patterns of spatially varying selection (inferred from five different types of selection scans) and the genomic architecture of phenotypic traits (phenology and growth) assayed in replicated common garden trials. This type of genomic association, long predicted by theory, lies at the heart of the population genomic approach of studying functionally relevant genetic variation (Luikart *et al.*, 2003), and it motivates the population genomic work reported for Eurasian species here.

Populus alba is a highly variable and patchily distributed taxon with a wide range of abiotic and biotic tolerances; it thrives in habitats as diverse as the deserts of northern Africa, European flood plain forests, and central Asian regions with highly continental climates and severe winter frosts (Dickmann & Kuzovkina, 2008). *P. alba* is known to hybridize occasionally with its related congener *Populus tremula*, but recent population genomic work indicates much stronger isolating barriers than previously assumed (Lexer *et al.*, 2010; Lindtke *et al.*, 2012, 2014). In line with its widespread, mosaic-like geographic distribution, significant nuclear genetic differentiation (e.g. in terms of F_{ST}) has been detected among local populations of *P. alba*, translating into estimates of only a few migrants per generation (Lexer *et al.*, 2005; Castiglione *et al.*, 2010; Lindtke *et al.*, 2012).

In this study, we use thoroughly validated whole-genome sequence data to address the genomic architecture of differentiation at a well-defined early stage of the divergence continuum present in this Eurasian *Populus* species, namely between phylogeographic lineages representing different postglacial recolonization routes of *P. alba* in Europe (Hewitt, 2000; Fussi *et al.*, 2010). We address three main questions of current interest to the evolutionary genetics of diverging populations in wild, perennial plants:

- What is the genomic architecture of population divergence at an early stage of the divergence continuum present in the 'model forest tree' *Populus*, that is, between well defined, distinct phylogeographic lineages?
- Is adaptation in these widespread forest trees more likely to occur from new mutations arising in large panmictic populations, or from standing genetic variation maintained by neutral processes in periods of glacial isolation?
- Which functional classes of genes are enriched among those affected by locally varying selection, and what might be the selective agents responsible for the unusually great allele frequency differences in these genes?

Materials and Methods

Characterization of *P. alba* populations with nuclear microsatellites

Two distinct recolonization lineages of *P. alba* L. in central Europe were identified in previous phylogeographic research based on plastid DNA (Fussi *et al.*, 2010). They comprise a southeastern recolonization route roughly following the Danube river valley upstream from Romania via Croatia, Serbia, and other countries into Hungary, and a southern route originating on the Italian peninsula. These two routes essentially follow major recolonization routes predicted by Hewitt (2000), Petit *et al.* (2003), and others. Populations of the two *P. alba* lineages are significantly differentiated for phenotypic (leaf and petiole) traits that appear to be under strong genetic control (several major quantitative trait loci detected; Lindtke *et al.*, 2013), and at least one of these (leaf size) is known to be linked with growth traits in poplar (Rae *et al.*, 2004). The populations are also differentiated for phytochemical defence traits (C. Caseys *et al.*, unpublished). The two lineages of *P. alba* meet at the northeastern end of the Alps in Austria, where a mixture of plastid DNA haplotypes from both lineages is found (Fussi *et al.*, 2010). The goal of the present study was to explore genomic differentiation between these two phylogeographic lineages, that is, across the intraspecific Austrian admixture zone (Fig. 1).

Population-level whole-genome resequencing represents a considerable investment in terms of sequencing effort and computation time, and thus we thoroughly characterized all study material with codominant nuclear genetic markers before sequencing. Microsatellite marker data for central European

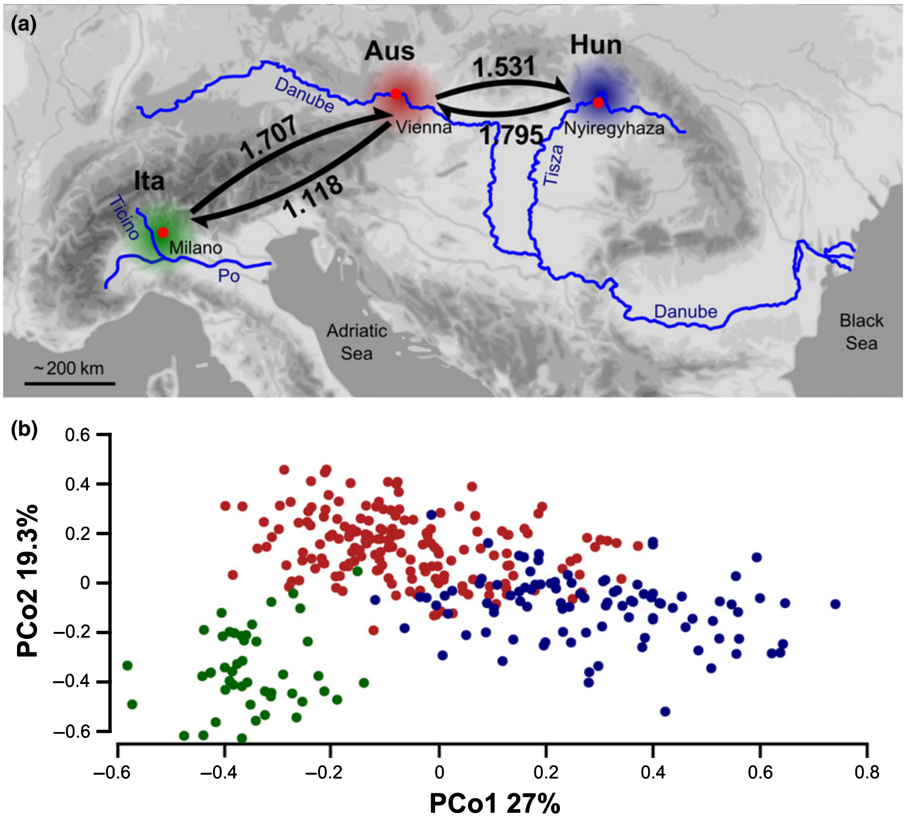

Fig. 1 Sampling locations and migration estimates for populations from two different central European phylogeographic lineages of *Populus alba*. (a) *P. alba* sampling sites, Italy (Ita, green), Austria (Aus, red), and Hungary (Hun, blue), with mutation-scaled migration parameter (*M*) estimates from Migrate for neighboring populations based on allelic data for codominant microsatellite loci. (b) Principal coordinate analysis (PCoA) of microsatellite data in Italian (green), Austrian (red), and Hungarian (blue) populations. The fraction of total variation explained by each principal coordinate axis is indicated.

populations of *P. alba* were available from previous studies (Lexer *et al.*, 2010; Lindtke *et al.*, 2012) and were comprehensively reanalyzed here. The dataset comprised 317 individuals, including 47 trees from the Ticino river drainage system in northern Italy (ITA), 173 from the Danube valley in northeastern Austria (AUT), and 97 from the Tisza river drainage in eastern Hungary (HUN) (Fig. 1). Characterization of these populations was based on 43 robust, polymorphic markers genotyped across all three localities with < 3% missing data; Supporting Information Table S1). The markers and genotyping procedures are described in detail in Lexer *et al.* (2010) and Lindtke *et al.* (2012).

We applied multivariate statistics and model-based analyses of migration rates to the nuclear microsatellite loci to complement the available plastid DNA data (Fussi *et al.*, 2010) for these localities. A principal coordinate analysis (PCoA) based on individual-level marker data was carried out using GenAlEx v6.5 (Peakall & Smouse, 2012) and the results were visualized in R 3.0.2 (R Core Team, 2013). Mutation-scaled migration parameters between populations were estimated using the coalescent theory and maximum-likelihood based approach implemented in Migrate v3.3.2 (Beerli & Felsenstein, 2001). Parameter estimation was based on a Brownian motion microsatellite data model, and genetic divergence between populations (F_{ST}) was used to obtain initial start values for the estimation of theta and migration rates. Migrate was run with 10 short chains (one heated) and a single long chain for 50 000 generations, discarding the first 10 000 generations as a burn-in.

Whole-genome sequencing (WGS) and single nucleotide polymorphism (SNP) detection

To represent the two phylogeographic lineages of interest, two subpopulation replicates from northern Italy (Ticino river valley) and two from eastern Hungary (Tisza river valley) were selected for pooled WGS (pool-seq WGS) using individual barcode tags (= identifier molecules) for each subpopulation replicate. The two subpopulation replicates for each locality were selected to be separated by 30–40 km of linear distance along their respective river valleys and were thus labeled Italy 'upstream' (sequencing pool A) and 'downstream' (pool B), and likewise for Hungary (sequencing pools C and D, respectively). Twelve individuals of *P. alba* previously characterized with microsatellites (Lexer *et al.*, 2010; Lindtke *et al.*, 2012) and genotyping-by-sequencing markers (Lindtke *et al.*, 2014) were chosen to represent each subpopulation replicate. The sequenced individuals were selected based on their geographic coordinates (minimum distance 50 m) and admixture coefficients (*Q*) from previous marker studies, which indicated they were free from signs of introgression from the related congener *P. tremula*. At distances of > 50 m, central European populations of *P. alba* are essentially panmictic and exhibit extremely weak spatial genetic structure (van Loo *et al.*, 2008). Our sampling design was chosen so as to yield sample sizes of $n = 24$ per population (northern Italy and eastern Hungary) upon combining data of subpopulation replicates for each locality, and to be able to check the robustness of the results at the individual subpopulation ($n = 12$) level. Sample sizes of this magnitude have previously been shown to yield useful and reliable results in pool-seq WGS (Rubin *et al.*, 2010).

Total genomic DNA for pool-seq WGS was extracted for all plants from silica-preserved (lyophilized) leaf tissue using Qiagen's DNeasy Plant Mini Kit. DNA concentrations and purity were determined using a Nanodrop 1000 system (Thermo Fisher Scientific, Waltham, MA, USA). The four pools (Italy A, Italy B, Hungary C, and Hungary D) were prepared by combining equal quantities of DNA for each of the 12 individuals per subpopulation replicate. DNA pools (= subpopulation replicates) were individually barcoded and jointly sequenced on the SOLiD4 system (Applied Biosystems, Thermo Fisher Scientific) at the Functional Genomics Center Zurich (FGCZ) following the manufacturer's instructions. SOLiD4 sequencing produced paired-end reads of 50 + 35 bp in length. Our target sequencing depth was 24× per sequenced pool, based on recommendations from empirical and simulation studies (Rubin *et al.*, 2012; Rellstab *et al.*, 2013). Reference mapping, quality filtering, and SNP detection from pool-seq WGS data were carried out as described in Methods S1. We only retained high-quality bi-allelic SNPs covered by eight to 250 reads in each of the four subpopulation replicates. The dataset was further quality-filtered by removing sites with a minimum read count of 3 for the minor allele to avoid spurious SNP calls (Table S1).

SNP validation by Sanger sequencing

To validate the accuracy of SNP calling and allele frequency estimates from SOLiD pool-seq WGS (Table S2), a subset of SNP loci were sequenced using the conventional Sanger method for each of the 48 individuals used in pool-seq WGS (24 Italian and 24 Hungarian samples). Five genes were selected for fluorescent dideoxy Sanger resequencing (Table S3). These SNP validation genes were selected to maximize the total number of SNPs within PCR fragments < 1.2 kb in length. Three of the genes (Potri.014G162900, Potri.007G055500, Potri.014G164400) were selected because they exhibited several high-differentiation outlier SNPs per gene in pool-seq WGS, and two genes were picked from putatively neutral regions without such outlier SNPs (Potri.019G002600, Potri.014G068400). For detailed laboratory protocols for Sanger sequencing, see Methods S2.

After sequencing, raw data were proofread and manually edited to retain information on heterozygote allele states using BioEdit (Hall, 1999). Polymorphism calls from pool-seq WGS and Sanger sequencing were then validated through Pearson correlations for minor allele frequencies and allele frequency differentials (AFDs) between sequencing methods using R (R Core Team, 2013) (Table S4). For each Sanger-sequenced gene, genetic diversity and differentiation between phylogeographic lineages (Italy vs Hungary) were analyzed using DNASP 5.0 (Librado & Rozas, 2009) after haplotype phasing. Deviations from neutral equilibrium expectations were tested with Tajima's *D* (Tajima, 1989).

Detection of differentiation and diversity outlier polymorphisms

Upon successful SNP validation, an outlier approach was applied to the pool-seq WGS data to identify candidate SNPs or genomic regions with unusually high degrees of allele frequency differentiation, potentially indicating divergent (=spatially varying) natural selection between the two *P. alba* phylogeographic lineages. For this purpose, AFDs (Shriver *et al.*, 1997; Turner *et al.*, 2010) were calculated at the level of individual SNPs from pool-seq WGS, using read counts of each SNP variant to estimate allele frequencies (Table S5). Throughout this paper, AFDs should not be mistaken for genetic divergence parameters (e.g. numbers of polymorphisms within and between populations) used in classical tests for selection (Hudson *et al.*, 1987). Windowed measures of AFD and genetic diversity (pooled heterozygosity; Rubin *et al.*, 2010) were also calculated to allow for the detection of selective sweep regions, under the assumption of synteny between *P. alba* and *Populus trichocarpa*. The 'hard sweep' genetic diversity test statistic lnRH ($\log_e$RH) (Schlötterer & Dieringer, 2005) was estimated as the ratio of pooled heterozygosity between the Italian and Hungarian populations at the window level. lnRH measures the ratio of gene diversity (H_e) between pairs of populations. Window sizes in our study were set to 8 kb with a step size of 2 kb, to reflect known linkage disequilibrium (LD) distances in *Populus* spp. with similar life-history traits (Slavov *et al.*, 2012). Only windows with a minimum of 10 SNPs and a maximum of 200 SNPs were used for the analyses. For plotting purposes, we smoothed over multiple windows to reduce the complexity of graphical representation (see figure legends).

Outlier analyses were first carried out globally among phylogeographic lineages by comparing the Italian population (sequencing pools A and B combined) with the Hungarian population (pools C and D combined). Outlier analyses were then repeated at the subpopulation level by computing all four pairwise comparisons between the Italian and the Hungarian subpopulation replicates. In each analysis, SNPs were considered unusually divergent if AFDs were ≥ 4 SDs higher than the genome-wide average. Polymorphisms that were particularly differentiated at both the global and the subpopulation scales were considered as candidate loci potentially affected by divergent (= locally varying) selection between phylogeographic lineages. Only SNPs detected in at least three pairwise subpopulation comparisons (including two independent comparisons) were retained as consistent candidate loci for locally varying selection between phylogeographic lineages. Polymorphisms detected in only one or two comparisons were not retained, as they were regarded as being more likely due to subpopulation-specific adaptation or drift.

Functional characterization of highly differentiated candidate genes and polymorphisms

Regions containing outlier SNPs and windows with outlier status were parsed for the presence of gene sequences. All outlier genes were gene ontology (GO)-annotated using the Blast2GO-PRO v.

2.7.1 pipeline under default settings (Conesa *et al.*, 2005) based on Blastx searches against a local copy of the nr database (as of June 2014). GO term enrichment analyses were performed on GOs associated with at least two outlier genes using Fisher's exact tests as implemented in the Bioconductor package topGO v2.16 using a minimum node size of five as recommended by Alexa & Rahnenfuhrer (2010). Fisher's exact test *P*-values were corrected for multiple testing using the Benjamini & Hochberg (1995) false discovery rate and were retained and ranked when below a threshold of 15% (Table S6). Polymorphisms located within intergenic regions were checked for potential mutations in *cis*-regulatory elements using the SOGO New PLACE software tool (Higo *et al.*, 1999).

Results

Characterization of *P. alba* populations with nuclear microsatellites

Multivariate analysis of microsatellite marker data via PCoA indicated clear nuclear genetic differentiation between the northern Italian and eastern Hungarian populations in the wind-pollinated forest tree *P. alba*, and weaker differentiation between both of these localities and northeastern Austria; Austrian trees overlapped genetically with Hungarian and, to a lesser extent, Italian trees (Fig. 1). Mutation-scaled migration rates (*M*) estimated by Migrate were generally low and showed a tendency for asymmetric gene flow from both Italy and Hungary into Austria, as expected from previously hypothesized patterns and directions of postglacial recolonization of this species (Fussi *et al.*, 2010). Thus, the biparentally inherited nuclear microsatellite data were largely congruent with the presence of two genetically differentiated lineages with admixture in northeastern Austria (Fig. 1), as previously hypothesized based on maternally transmitted plastid DNA (Fussi *et al.*, 2010). Based on the congruence between spatial genetic patterns for plastid and nuclear markers, the Italian and Hungarian populations were regarded as appropriate for a whole-genome analysis of differentiation between these two intraspecific phylogeographic lineages.

WGS and SNP detection

Whole-genome resequencing resulted in a total of 596 689 236 paired-end reads of 50 and 35 bp, respectively (SRA accession number: SRP053219). A total of 49% of the longer 50 bp reads and 25% of the shorter 35 bp reads were accurately reference-mapped against the *P. trichocarpa* reference genome. We detected a total of 1775 768 SNPs in the studied populations (Table S2), covering, on average, 23.9×, 24.4×, 21.2× and 31.6× in pools A, B, C and D, respectively (Fig. S1). Given a known assembly size of 378 545 895 base pairs, the average distance among SNP loci was thus 213 bp.

The density of SNPs varied among chromosomes (Tables S2, S7). Among all detected SNPs, 66% were located within gene boundaries (including exons and introns) predicted from the *P. trichocarpa* genome assembly, and 41% were located in exonic regions. This value is higher than expected, as only 29 and 17% of the genome are composed of genic and exonic regions, respectively. This result probably stems from elevated degrees of divergence between *P. alba* and *P. trichocarpa* in intergenic regions compared with genes, resulting in a lower mapping efficiency in intergenic regions (Table S2). Thus, our results illustrate the benefits and pitfalls of using a heterologous reference genome in resequencing studies, as observed by others (Schlötterer *et al.*, 2014).

SNP validation by Sanger sequencing

A total of 89 SNPs predicted by pool-seq WGS were validated through direct Sanger sequencing of five nuclear loci (Tables S3, S4). Sanger sequencing yielded a total of 3639 sequenced bases in, on average, 44 of the 48 studied individuals (range 35–48, SD = 3.65). Allelic frequencies were significantly correlated between Sanger sequencing and pool-seq WGS with Pearson's *r*-values of 0.92 ($P < 0.001$) and 0.41 ($P < 0.001$) for the Italian and Hungarian populations, respectively.

The lower correlation in Hungary was traced back to putative paralogous reads mapped onto gene Potri.014G068400 in that population. This became apparent from a bias towards intermediate allele frequencies for this particular gene in Hungary, and even more so from information on sequence coverage: mean coverage was significantly higher for this gene in the Hungarian population (108.9×) compared with the Italian one (50.3×; $t = 8.181$, df = 35.98, $P < 0.001$), and also compared with the remaining four genes sequenced in this population (53.9×; $t = 7.923$, $P < 0.001$). The observed pattern is consistent with local duplication of gene Potri.014G068400 in the Hungarian population and mapping of paralogous reads against the same locus in the *P. trichocarpa* reference genome. Our results point to often underappreciated issues of resequencing studies arising from the dynamic nature and 'fluidity' of plant genomes. The correlation between allelic frequencies from pool-seq WGS and Sanger sequencing increased substantially for the Hungarian population after exclusion of this gene, resulting in a Pearson's *r* of 0.85 ($P < 0.001$). Correlations were also very high between AFDs obtained by pool-seq WGS and Sanger sequencing with *r*-values of 0.92 ($P < 0.001$, excluding the Potri.014G068400 gene).

Detection of differentiation and diversity outlier polymorphisms

On the level of individual SNPs from pool-seq WGS, outlier analysis on the global phylogeographic scale detected 11 099 highly differentiated SNPs between the Italian and Hungarian populations at a threshold of ≥ 4 SD compared with genome-wide expectations. This value is 198 times larger than expected from a normal distribution of the data at the given 4 SD cut-off, effectively translating into an expected 0.51% falsely discovered outliers. On average, 9134 high-differentiation outlier SNPs (range 8490–9833) were detected in pairwise comparisons among individual subpopulation replicates of the two lineages (sequencing pools A–D; see the Materials and Methods section) (Table S5). A total of 848 SNPs were detected as high-differentiation outliers in at least three out of four pairwise comparisons. All of these were also detected at the global phylogeographic scale and were thus retained as consistently strong SNP outliers for further analysis (Table 1).

On the level of genomic windows (window size = 8 kb, step size = 2 kb; see the Materials and Methods section), analysis of 159 026 sequence windows covered by ≥10 and < 200 SNPs revealed 519 particularly differentiated outlier windows (AFDs ≥ 4 SD) localized in 252 independent genome regions. This is *c.* 104 times more than expected from a normal distribution, equivalent to a false discovery rate of 0.96%. Outlier window size was rather small (median 4 kb, mean 6.2 kb, maximum 18 kb). Outlier regions for AFDs were observed on all 19 chromosomes of the *P. alba* genome. No window with particularly low genetic differentiation between phylogeographic lineages was detected at a cutoff of 4SD compared with genome-wide expectations (Fig. 2).

Measures of genetic diversity from pool-seq WGS differed slightly but significantly between the two studied populations, with higher pooled heterozygosity values (± SD) in Italy (0.271 ± 0.051) than in Hungary (0.264 ± 0.054; $t = 32.36$; $P < 0.001$). The 'hard sweep' test statistic lnRH detected 246 outlier windows with reduced diversity in Italy and 171 such windows in Hungary. Overall, this corresponds to 0.4% of all genomic windows and to 12% of all outlier windows detected by our scan for increased allele frequency differentiation (Figs 2, S2; Table S8). The results are informative regarding our discussion of hard selective sweeps vs more subtle allele frequency shifts during local adaptation in *P. alba* at this broad spatial scale (see later).

Individual Sanger sequencing of three highly differentiated candidate genes (Table 2) and two putatively neutral genes confirmed the results of pool-seq WGS. First, candidate genes showed high F_{ST} values between the Italian and the Hungarian lineages (range: 0.32–0.63, Table 2) compared with the putatively neutral controls (0.04 and 0.06), as predicted by pool-seq WGS. Second, Tajima's D was highly negative (= excess of low-frequency variants) for all candidate genes in the Italian population (Table 2), whereas it was nonsignificant for the two controls. This pattern across genes suggests positive selection on the three candidate genes, rather than population expansion, as the likely explanation for the differences in allele frequency spectra (Luikart *et al.*, 2003; Nielsen *et al.*, 2007).

Table 1 Summary of results on genetic differentiation at single nucleotide polymorphism (SNP) loci between Italian and Hungarian populations of *Populus alba*, including comparison of individual SNPs along the genome (in all windows) with SNPs observed in highly differentiated outlier windows

	All windows	Outlier windows
SNPs		
N	1775 768	12 244
Mean AFD	0.105	0.246
Differentiation outlier SNPs		
N	848	188
Mean outlier AFD	0.714	0.714
% in genic regions	64.3	75.0
Candidate outlier genes		
N	437	78
Mean AFD	0.181	0.248

Outlier, ≥ 4 SDs different from genome-wide averages of genetic differentiation. *N*, absolute number; AFD, allele frequency differential; candidate outlier genes include at least one SNP locus with outlier status.

Functional characterization of highly differentiated polymorphisms and genes

Annotation of the 848 consistently detected high-differentiation outlier SNPs (see earlier) indicated that 303 (35.7%) were located in nongenic regions and 545 (64.3%) were located in 437 different gene loci predicted from the *P. trichocarpa* genome assembly (22.2% in exons, 35.4% in introns, and 6.7% in untranslated regions). These candidate genes contained between one and 10 outlier SNPs (Table S8). Twenty-four of these 437 outlier genes were also outliers in screens for reduced genetic diversity lnRH, including 21 in Italy and three in Hungary, thus providing evidence for local hard selective sweeps in coding regions of *P. alba*. GO annotations identified molecular functions, biological processes, and/or the cellular components for 19 814 of the genes analyzed in this study, including 327 outlier genes. Among the 303 SNPs identified in intergenic regions, 89 were located in *cis*-regulatory elements.

Table 3 lists the top 29 candidate gene loci (genes containing a minimum of two high-differentiation outlier SNPs with at least one located in exonic regions), including descriptions of their likely functions. These top candidate genes were distributed along 14 of the 19 chromosomes of the *Populus* v2 assembly and included six genes with a diversity lnRH footprint of 'hard selective sweeps' (Table 3). GO term enrichment analyses of the 327 outlier genes detected 63 significantly overrepresented GO terms (58 molecular functions and five biological processes), and several of these were also overrepresented among differentiation outliers in recent studies of *P. trichocarpa* and *Arabidopsis lyrata*; Tables S6, S8). All 58 molecular functions were linked to transmembrane transport, catalytic activities, metal binding, and nucleotide binding (Table 4). The five significant biological processes were linked to ion transport and regulation of anatomical structure size, including clear candidate genes for adaptation to the divergent soil substrates found in the two studied river drainage systems (see later).

Discussion

Genomic patterns of differentiation between phylogeographic lineages in a widespread forest tree: implications for adaptive evolution

Classical population genetic theory has had a strong focus on adaptation from new mutations that become fixed by divergent natural selection in populations colonizing new environments (Orr, 1998; Barrett & Schluter, 2008). This traditional model was accompanied by the long-held view that adaptation may often involve relatively few genes with alleles of large effect

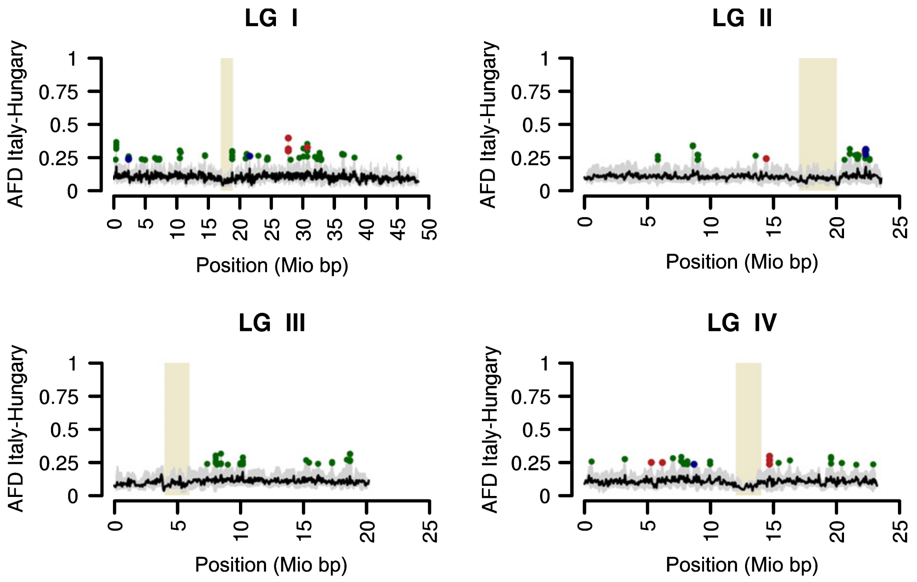

Fig. 2 Genomic patterns of diversity and differentiation in populations from two different central European phylogeographic lineages of *Populus alba*. Sliding windows (size = 8000 bp, step size = 2000 bp) summarize results for 10–200 single nucleotide polymorphisms per window on four exemplary poplar linkage groups (LGs) identified by roman numerals. Beige rectangles highlight putative centromere locations (Slavov *et al.*, 2012) on each LG. A Loess smoother summarizes the region-wide degrees of differentiation (window size 250 kb, sliding by 100 kb). Positive allele frequency differential (AFD) outliers (≥ 4 SD different from genome-wide expectations) are highlighted in green for the comparison of Italian and Hungarian populations. AFD outlier windows, which are also of reduced diversity (lnRH ≤ −4 SD different from genome-wide expectations) in Italian populations are indicated in red, whereas those of reduced diversity in Hungary (lnRH ≥ 4 SD) are indicated in blue.

Table 2 Estimates of genetic diversity and differentiation and neutrality tests for three candidate genes and two negative controls individually sequenced by the Sanger method

			Diversity						Differentiation	Neutrality test
Genes	Pop.	Fragment size	# Seq.	S	Si	Pi	# Hap.	Hd	F_{ST}	Tajima's *D*
Candidate genes										
Potri.014G162900	Italy	707	48	14	0	0.0018	4	0.233	0.634	−**1.841***
	Hungary		48	4	0	0.0006	4	0.233		−1.175
Potri.007G055500	Italy	898	46	18	7	0.0019	9	0.561	0.317	−**1.838***
	Hungary		28	14	2	0.0055	11	0.899		1.273
Potri.014G164400	Italy	903	46	25	11	0.0027	6	0.492	0.428	−**1.887***
	Hungary		44	14	0	0.0034	14	0.874		−0.13
Putatively neutral genes										
Potri.019G002600	Italy	694	46	24	4	0.0092	12	0.86	0.056	0.132
	Hungary		48	30	3	0.0092	17	0.792		−0.483
Potri.014G068400	Italy	879	44	8	3	0.0024	11	0.794	0.042	0.39
	Hungary		34	15	4	0.0041	21	0.964		−0.26

Significant Tajima's *D*-values are presented in bold.

*Significant at $P < 0.05$; Pop., population; # Seq., number of sequences; S, polymorphic sites; Si, singletons; Pi, nucleotide diversity; # Hap., number of haplotypes; Hd, haplotype diversity.

(reviewed by Orr, 1998). A growing number of human genetics studies indicates that this view is overly simplistic, as functionally and adaptively important traits often have a complex basis, entailing either many common genetic variants of small effect and/or a number of rare variants of large effect (Manolio *et al.*, 2009; Gibson, 2012). Similar evidence has started to emerge in plants, and particularly so in forest trees (reviewed by Neale & Kremer, 2011; Le Corre & Kremer, 2012; Evans *et al.*, 2014; McKown *et al.*, 2014; Zhou *et al.*, 2014). To date, most studies of these issues in wild, outcrossing plants were based on limited numbers of polymorphic markers and genes, typically assayed via SNP arrays (Neale & Kremer, 2011). In the present study, we brought > 1.7 million SNPs from population-level genome resequencing in *P. alba* to bear on these issues.

Our genomic data on two genetically (Fig. 1) and phenotypically (Lindtke *et al.*, 2013) divergent intraspecific lineages of *P. alba* from different river drainage systems with divergent soil properties (see later) and local climates are consistent with the prediction from earlier population and quantitative genetic work on forest trees that adaptive divergence has a complex, polygenic basis: on average, > 9000 high-differentiation outlier SNPs were detected between subpopulations of the studied intraspecific

Table 3 Twenty-nine genes with multiple outlier single nucleotide polymorphisms (SNPs) for genetic differentiation between Italian and Hungarian populations of *Populus alba*

Outlier genes	# out	Description
DNA/RNA binding – transcription factors		
Potri.014G166300	**4**	**transcriptional adapter ada2-like isoform x4**
Potri.006G221800	**3**	**myb transcription factor r2r3-like protein**
Potri.005G183700	**3**	**phd finger family protein**
Potri.014G164400*	2	scarecrow-like transcription factor pat1-like
Potri.009G093000	2	homeodomain-like protein with ring fyve phd-type zinc finger isoform 1
Potri.001G101800	2	at-rich interactive domain-containing protein 2-like
Potri.001G220100	2	la-related protein 1-like
Potri.005G112300	2	proline-rich family protein
Potri.012G086800	2	50s ribosomal protein
Potri.002G017500	2	thioredoxin superfamily protein
Potri.012G082600	2	u3 small nucleolar ribonucleoprotein mpp10 isoform 2
Ion transport		
Potri.001G123700	10	potassium transporter 11 family protein
Potri.007G055500*	**3**	**calcium-transporting atpase plasma membrane-type-like**
Potri.010G045900	3	abc transporter b family member chloroplastic-like
Response to stress – defence		
Potri.013G058100	4	chaperone protein
Potri.014G195900	3	phytochelatin synthase
Potri.017G074000	3	abc transporter b family member 15-like
Potri.003G066400	2	flavonoid 3-monooxygenase-like
Potri.016G050300	2	protein polychome-like
Potri.014G153400	2	aconitate hydratase mitochondrial-like
Structural		
Poptr.0005s20630	4	caffeic acid methyltransferase
Potri.010G098800	3	ubx domain-containing protein
Potri.010G153000	3	phragmoplast orienting kinesin 1-like
Potri.014G162900*	2	fasciclin-like arabinogalactan protein 7 isoform 1
Potri.001G148800	2	pleckstrin homology domain-containing protein
Other		
Potri.013G008500	8	uncharacterized loc101202927
Potri.004G067200	**3**	**upf0415 protein c7orf25 homolog**
Potri.004G067400	**3**	**endosomal targeting bro1-like domain-containing protein isoform 2**
Potri.010G170900	2	btb poz domain-containing protein at3 g19850-like

Candidate genes in bold were located in genomic regions with a significant reduction of genetic diversity (lnRH) in the Italian population.
*Candidate genes tested for traces of selection using Sanger sequencing; # out, number of SNP loci with outlier status.

lineages using stringent (4 SD) detection thresholds, including 848 SNPs within 437 different genes revealed by replicated contrasts (Table 1). These results from single SNPs were complemented by those from window-averaged analyses of differentiation across the genome (Figs 2, S2). Thus, to the extent that allele frequency differentiation is informative regarding the nature of population divergence (Le Corre & Kremer, 2012), our results are consistent with a polygenic architecture of locally varying selection (cf. polygenic adaptation) in *P. alba* at the level of divergent postglacial recolonization lineages in Europe. Our study provides an informative 'snapshot' of this early stage of the divergence continuum (Nosil *et al.*, 2009; Feder *et al.*, 2012) between well-defined postglacial recolonization lineages in a widespread, ecologically important forest tree.

Our results also allow us to touch on the enigmatic history of the potentially adaptive genetic variation present between divergent phylogeographic lineages of this widespread forest tree. As shown by Hermisson & Pennings (2005) and illustrated by Pritchard *et al.* (2010), adaptation is very likely to occur from standing genetic variation rather than from new mutations when the mutational target size is large, as is the case for polygenic adaptation. In effect, adaptive population divergence in this case will make use of moderate allele frequency changes at many loci of small effect ('soft selective sweeps'; Hermisson & Pennings, 2005; Pritchard *et al.*, 2010; Messer & Petrov, 2013). It appears that this subtle genomic signature is more easily identified by pairwise tests for allelic differentiation than by tests for hard selective sweeps based on locally reduced diversity in single populations (Figs 2, S2).

The greater number of differentiation than diversity outliers under adaptation from standing variation (de Carvalho *et al.*, 2010; our study) may arise from a variety of biological factors. Most importantly, selective allele frequency shifts from standing variation ('soft sweeps') are expected to involve multiple old alleles that may have arisen independently on different genomic backgrounds (Barrett & Schluter, 2008). This will result in a much more subtle reduction of local genetic diversity, compared with expectations for 'hard sweeps' starting from new mutations (Pritchard *et al.*, 2010). Spatially varying selection of this type can easily lead to shifts in allele frequencies, resulting in greatly increased differentiation at particular loci, but small or zero net effects on local diversity (lnRH), for example when alternative alleles at bi-allelic SNPs are driven to high frequency in different localities. We acknowledge that the power of the lnRH statistic to detect selective sweeps (Schlötterer & Dieringer, 2005) has been more thoroughly evaluated than that of AFD.

The putative older age of alleles contributing to the standing genetic variation (Barrett & Schluter, 2008) also represents a plausible explanation for the limited overlap between single outlier SNPs vs entire outlier windows for AFD (Table 1; see the Results section). The latter point to more recent and/or stronger local selection, which will affect broader LD regions along chromosomes as a result of processes related to genetic hitchhiking. Future studies of hard vs soft selective sweeps should make more explicit use of information on the length of selected haplotypes (Chen *et al.*, 2010), an approach that would benefit greatly from the further development of haplotyping methods for pool-seq data (Schlötterer *et al.*, 2014).

Despite the apparent prevalence of soft selective sweeps in the studied polymorphisms and populations, it is noteworthy that 12% of all genomic outlier windows detected in our study exhibited the local reduction of genetic diversity (lnRH) expected for classical, hard selective sweeps. This suggests that effective population sizes (N_e) and/or mutation rates in this widespread,

Table 4 Major gene ontology (GO) terms with significant enrichment among highly differentiated outlier genes compared with all genes covered by this whole-genome resequencing effort

Significantly enriched GO term in outlier genes	Sig. child terms	# genes	# outliers genes
(a) Molecular function			
Transmembrane transporter activity	22	1205	34**
Active transmembrane transporter activity	7	649	19*
Cation transmembrane transporter activity	11	568	18**
Potassium ion transmembrane transporter activity	3	81	6**
Calcium-transporting ATPase activity	0	35	3*
Hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances	5	285	12**
Xenobiotic transporter activity	1	35	3*
Catalytic activity	33	11309	204*
Hydrolase activity, acting on ester bonds	8	1182	29*
Lyase activity	3	437	13*
Oxidoreductase activity, acting on a heme group of donors	3	10	2*
Binding			
Metal cluster binding	1	124	6*
Cyclic nucleotide binding	0	20	3**
(b) Biological process			
Ion transport	2	1304	36***
Regulation of anatomical structure size	1	150	9***

Sig. child terms, significantly enriched GO child terms; # genes, total number of annotated genes; # outlier genes, number of genes containing outlier (high genetic differentiation) SNP loci. Molecular functions (a) and biological processes (b) of genes with significantly enriched GO terms. Significance of Fisher exact tests: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. All false discovery rates are < 0.15.

perennial plant species are sufficiently large to give rise to new genetic variants that make an important contribution to the genetic variation used by spatially varying selection (Pritchard *et al.*, 2010). Our results are consistent with selection acting on both new mutations and standing variation after the onset of postglacial recolonization of these river drainages, that is, within the last few hundreds of tree generations (de Carvalho *et al.*, 2010; Fussi *et al.*, 2010). We note that a previous study has estimated migration rates and effective population sizes in central European populations of this species (Lexer *et al.*, 2005), and that a full demographic analysis based on the joint site frequency spectrum (SFS) from many sequenced individuals of these populations is forthcoming (C. Christe *et al.*, unpublished).

At the present time, we cannot exclude with certainty the possibility that some of the differentially selected variants in *P. alba* might stem from local interspecific introgression from the related congener *P. tremula*. Nevertheless, we consider it unlikely that our results are strongly affected by introgression, as these two parapatric species are isolated by much stronger postzygotic barriers than traditionally thought (Lexer *et al.*, 2010; Lindtke *et al.*, 2012, 2014). Also, all of our sequenced trees were previously characterized as pure *P. alba* based on large numbers of genetic markers (Lexer *et al.*, 2010; Lindtke *et al.*, 2012, 2014). Extension of WGS to the related *P. tremula* might reveal whether any of the unusually divergent genome regions discovered in this study have a heterospecific origin.

Functional roles and potential selective agents of highly differentiated candidate genes

A focused inspection of functional (GO) categories with significant enrichment among our spatially varying selection candidate loci (Table 4) and our top list of highly differentiated candidate genes (Tables 3, S8) reveals plausible links with the perhaps most striking environmental difference between the two studied river drainages, soil substrate texture and nutrient content. The soils of both river drainages, Ticino in northern Italy and Tisza in eastern Hungary, are classified as typical flood-plain humaquepts and dystrodepts according to established soil taxonomy (Soil Survey Staff, 2010), but the soils at the Tisza are much more loamy and clayey with high water retention, whereas those at the Ticino exhibit a strong gravel and sand component with low water retention and low nutrient availability in terms of cation exchange capacity (CEC) and pH (European soil portal: http://eu-soils.jrc.ec.europa.eu/; Solaro, 2006). Notwithstanding the spatial mosaics in soil nutrient availability (CEC and pH) generally expected in river flood plains, the difference in soil substrate (loam and clay vs gravel and sand) and nutrient status between these river drainages represents a plausible potential agent of selection on genetic variation present in flood plain forest trees at this spatial scale.

Consistent with this hypothesis, many of our enriched GO categories were also identified by population genomic scans of other taxa with pronounced intraspecific variation in soil substrates and nutrient content, for example, *P. trichocarpa* (Evans *et al.*, 2014) and *A. lyrata* (Turner *et al.*, 2010) (Table S6). Also, our enriched GO terms and most highly differentiated selection outlier candidate loci (Tables 3, 4) include multiple genes involved in cation transport, root development, and water stress response. For example, the most highly divergent candidate gene in our study, Potri.001G123700, with 10 highly differentiated SNPs (Table 3), codes for a potassium (K^+) transporter from a highly enriched GO molecular function (transmembrane transporter activity; Tables 4, S6). A total of six highly differentiated candidate genes in this GO category code for K^+ uptake permeases, K^+ transporters (ATK1), and K^+ channels (Table S8). In *A. thaliana*, ATK1 is expressed predominantly in root hairs and root endodermis where K^+ is taken up from the soil (Cao *et al.*, 1995). This high-affinity transporter operates at intermediate and low external K^+ concentrations (Nieves-Cordones *et al.*, 2014). Eleven additional, highly differentiated candidate genes involved in transport of ions such as calcium, magnesium, nitrate, phosphate, ammonium, boron, and sulfur were identified by our study (Table S8).

In addition to ion uptake and transport, 11 of our highly differented candidate genes (Tables 3, S8) were linked to root development. The architecture of a plant's root system is central to adaptation to the soil substrate, and optimal development of the primary and lateral roots, including root hairs, is essential for exploitation of the soil. Thus, increased allelic differentiation of these genes is consistent with the hypothesis of spatially varying selection as a result of the pronounced differences in soil substrate texture and grain among these river drainages (see earlier). Among the highly differentiated candidate genes identified, Potri.010G162500 is known to affect root meristem size and growth (Zhou *et al.*, 2011). Four additional genes in this group are linked to response and/or transport of auxin, which mediates root cell elongation, cell cycle progression, and tissue differentiation (Table S8). For example, Potri.005G112300, one of our top candidates (Table 3), is involved in cell differentiation and elongation in the root tip through the modulation of auxin transporters under limiting conditions (Gonzalez-Mendoza *et al.*, 2013).

Among other functionally interesting groups of genes, our highly differentiated candidates for spatially varying selection also include two genes involved in response to water deprivation (Potri.001G229100 and Potri.005G219000). Thus, it would appear that our replicated WGS scan in *P. alba* has revealed multiple spatially varying selection candidate genes with plausible links to edaphic differences between these river drainages. This includes many genes for which local adaptation has made use of standing genetic variation, and several (e.g. Potri.007G055500; Tables 2, 3) for which it has proceeded by classical hard selective sweeps.

Implications for adaptation genomics, breeding, and applied plant science

Taken at face value, our results suggest that adaptive differentiation at this stage of the divergence continuum in Eurasian *Populus* – between divergent postglacial recolonization lineages at the subcontinental scale – is based on both new mutations and standing genetic variation, consistent with recent results in the North American *P. trichocarpa* (Evans *et al.*, 2014). The relative prevalence of spatially varying selection from standing variation implies the potential for rapid evolution of tree populations in the face of environmental (e.g. climate) change (Barrett & Schluter, 2008). Indeed, common garden trials in the related *P. tremula* show that genomic admixture between divergent postglacial lineages at a continent-wide scale contributes to variation for adaptively important quantitative traits (e.g. bud phenology; de Carvalho *et al.*, 2010). More rigorous tests of the genomic architecture of adaptive and functional divergence in Eurasian *Populus* species should now entail a combination of genome-scale population genetics, association genetics, and experimental common gardens (Evans *et al.*, 2014), best including estimation of the covariance of allelic effects (Le Corre & Kremer, 2012). Populations with admixture between adaptively differentiated intraspecific lineages such as those identified here (Figs 1, 2) may be useful in this context, as the genomic variation in ancestry induced by admixture can be utilized in association scans (Buerkle & Lexer, 2008). The role of structural variation and copy number variants (Mills *et al.*, 2011) also deserves special attention in studies of adaptive divergence in this system, as exemplified by an apparent structural variation identified by Sanger-based SNP validation in *P. alba* in our study (gene Potri.014G068400; see the Results section). The increased use of homospecific reference genomes in future WGS efforts is expected to further increase the fraction of the genome that can be interrogated for polymorphisms in *Populus* and many other plant taxa. Fortunately, the potential for all these research topics in *Populus* and other forest trees is excellent (Neale & Kremer, 2011; Slavov *et al.*, 2012; Evans *et al.*, 2014; McKown *et al.*, 2014). In *Populus* spp., the results may reveal which genome regions are of special interest to adaptation genomics, breeding for bioenergy feedstock development, and restoration efforts in riverine habitats.

Acknowledgements

We thank Hans Herz, Wilfried Nebenführ, Stefano Gomarasca, István Asztalos, and other colleagues for help during field work; Thelma Barbará and Alexa Oppliger for help in the laboratory; and Rémy Bruggmann and staff of the Functional Genomics Centre Zurich (FGCZ) for advice regarding pool-seq WGS sampling design and bioinformatics work flow, and for DNA sequencing. Thanks also go to Myriam Heuertz, Antonello Bonfante, Joachim Hermisson, and Camille Christe for helpful discussions, and to three anonymous reviewers for valuable comments on the manuscript. Financial support came from grant award nos 31003A_127059 and 31003A_149306 of the Swiss National Science Foundation (SNSF) to C.L.

References

Alexa A, Rahnenfuhrer J. 2010. topGO: enrichment analysis for gene ontology. R package version2.16.0.

Anderson JT, Willis JH, Mitchell-Olds T. 2011. Evolutionary genetics of plant adaptation. *Trends in Genetics* **27**: 258–266.

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

Beerli P, Felsenstein J. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences, USA* **98**: 4563–4568.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate – a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**: 289–300.

Bernhardsson C, Robinson KM, Abreu IN, Jansson S, Albrectsen BR, Ingvarsson PK. 2013. Geographic structure in metabolome and herbivore community co-occurs with genetic structure in plant defence genes. *Ecology Letters* **16**: 791–798.

Buerkle CA, Lexer C. 2008. Admixture as the basis for genetic mapping. *Trends in Ecology and Evolution* **23**: 686–694.

Cao J, Schneeberger K, Ossowski S, Gunther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C *et al.* 2011. Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nature Genetics* **43**: 956–963.

Cao YW, Ward JM, Kelly WB, Ichida AM, Gaber RF, Anderson JA, Uozumi N, Schroeder JI, Crawford NM. 1995. Multiple genes, tissue-specificity, and expression-dependent modulation contribute to the functional diversity of potassium channels in *Arabidopsis thaliana*. *Plant Physiology* **109**: 1093–1106.

de Carvalho D, Ingvarsson PK, Joseph J, Suter L, Sedivy C, Macaya-Sanz D, Cottrell J, Heinze B, Schanzer I, Lexer C. 2010. Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula* L.), a widespread forest tree. *Molecular Ecology* **19**: 1638–1650.

Castiglione S, Cicatelli A, Lupi R, Patrignani G, Fossati T, Brundu G, Sabatti M, van Loo M, Lexer C. 2010. Genetic structure and introgression in riparian populations of *Populus alba* L. *Plant Biosystems* **144**: 656–668.

Chen H, Patterson N, Reich D. 2010. Population differentiation as a test for selective sweeps. *Genome Research* **20**: 393–402.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**: 3674–3676.

Dickmann D, Kuzovkina YA. 2008. *Poplars and willows in the world.* Working Paper IPC/9-2. Rome, Italy.

Eckenwalder JE. 1996. Systematics and evolution of *Populus*. In: Stettler RF, Bradshaw HD, Heilman PE, Hinckley TM, eds. *Biology of Populus, and its implications for management and conservation*. Ottawa, Canada: NRC Research Press, 7–32.

Ellegren H. 2013. Genome sequencing and population genomics in non-model organisms. *Trends in Ecology and Evolution* **29**: 51–63.

Ellegren H, Smeds L, Burri R, Olason PI, Backstrom N, Kawakami T, Kunstner A, Makinen H, Nadachowska-Brzyska K, Qvarnstrom A *et al.* 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* **491**: 756–760.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**: e19379.

Evans LM, Slavov GT, Rodgers-Melnick E, Martin J, Ranjan P, Muchero W, Brunner AM, Schackwitz W, Gunter L, Chen J-G *et al.* 2014. Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. *Nature Genetics* **46**: 1089–1096.

Feder JL, Egan SP, Nosil P. 2012. The genomics of speciation-with-gene-flow. *Trends in Genetics* **28**: 342–350.

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM. 2011. A map of local adaptation in *Arabidopsis thaliana*. *Science* **334**: 86–89.

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.

Gibson G. 2012. Rare and common variants: twenty arguments. *Nature Reviews Genetics* **13**: 135–145.

Gonzalez-Mendoza V, Zurita-Silva A, Sanchez-Calderon L, Sanchez-Sandoval ME, Oropeza-Aburto A, Gutierrez-Alanis D, Alatorre-Cobos F, Herrera-Estrella L. 2013. APSR1, a novel gene required for meristem maintenance, is negatively regulated by low phosphate availability. *Plant Science* **205**: 2–12.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**: 95–98.

Hermisson J, Pennings PS. 2005. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**: 2335–2352.

Hewitt G. 2000. The genetic legacy of the Quaternary ice ages. *Nature* **405**: 907–913.

Higo K, Ugawa Y, Iwamoto M, Korenaga T. 1999. Plant *cis*-acting regulatory DNA elements (PLACE) database. *Nucleic Acids Research* **27**: 297–300.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* **6**: e1000862.

Hudson RR, Kreitman M, Aguadé M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**: 153–159.

Ingvarsson PK. 2005. Molecular population genetics of herbivore-induced protease inhibitor genes in European aspen (*Populus tremula* L., Salicaceae). *Molecular Biology and Evolution* **22**: 1802–1812.

Jansson S, Douglas CJ. 2007. *Populus*: a model system for plant biology. *Annual Reviews Plant Biology* **58**: 435–458.

Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S *et al.* 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**: 55–61.

Le Corre V, Kremer A. 2012. The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology* **21**: 1548–1566.

Levsen ND, Tiffin P, Olson MS. 2012. Pleistocene speciation in the genus *Populus* (Salicaceae). *Systematic Biology* **61**: 401–412.

Lexer C, Fay MF, Joseph JA, Nica MS, 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, Joseph JA, van Loo M, Barbara T, Heinze B, Bartha D, Castiglione S, Fay MF, Buerkle CA. 2010. Genomic admixture analysis in European *Populus* spp. reveals unexpected patterns of reproductive isolation and mating. *Genetics* **186**: 699–712.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451–1452.

Lindtke D, Buerkle CA, Barbara T, Heinze B, Castiglione S, Bartha D, Lexer C. 2012. Recombinant hybrids retain heterozygosity at many loci: new insights into the genomics of reproductive isolation in *Populus*. *Molecular Ecology* **21**: 5042–5058.

Lindtke D, Gompert Z, Lexer C, Buerkle CA. 2014. Unexpected ancestry of *Populus* seedlings from a hybrid zone implies a large role for postzygotic selection in the maintenance of species. *Molecular Ecology* **23**: 4316–4330.

Lindtke D, Gonzalez-Martinez SC, Macaya-Sanz D, Lexer C. 2013. Admixture mapping of quantitative traits in *Populus* hybrid zones: power and limitations. *Heredity* **111**: 474–485.

van Loo M, Joseph JA, Heinze B, Fay MF, Lexer C. 2008. Clonality and spatial genetic structure in *Populus x canescens* and its sympatric backcross parent *P. alba* in a Central European hybrid zone. *New Phytologist* **177**: 506–516.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P. 2003. The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics* **4**: 981–994.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A *et al.* 2009. Finding the missing heritability of complex diseases. *Nature* **461**: 747–753.

McKown AD, Klapste J, Guy RD, Geraldes A, Porth I, Hannemann J, Friedmann M, Muchero W, Tuskan GA, Ehlting J *et al.* 2014. Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of *Populus trichocarpa*. *New Phytologist* **203**: 535–553.

Messer PW, Petrov DA. 2013. Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology and Evolution* **28**: 659–669.

Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye K, Cheetham RK *et al.* 2011. Mapping copy number variation by population-scale genome sequencing. *Nature* **470**: 59–65.

Neale DB, Kremer A. 2011. Forest tree genomics: growing resources and applications. *Nature Reviews Genetics* **12**: 111–122.

Nielsen R, Hellmann I, Hubisz M, Bustamante C, Clark AG. 2007. Recent and ongoing selection in the human genome. *Nature Reviews Genetics* **8**: 857–868.

Nieves-Cordones M, Alemán F, Martínez V, Rubio F. 2014. K^+ uptake in plant roots. The systems involved, their regulation and parallels in other organisms. *Journal of Plant Physiology* **171**: 688–695.

Nosil P, Harmon LJ, Seehausen O. 2009. Ecological explanations for (incomplete) speciation. *Trends in Ecology and Evolution* **24**: 145–156.

Orr HA. 1998. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* **52**: 935–949.

Orsini L, Andrew R, Eizaguirre C. 2013. Evolutionary ecological genomics. *Molecular Ecology* **22**: 527–531.

Peakall R, Smouse PE. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* **28**: 2537–2539.

Petit R, Aguinagalde I, de Beaulieu JL, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M *et al.* 2003. Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* **300**: 1563–1565.

Pritchard JK, Di Rienzo A. 2010. Adaptation – not by sweeps alone. *Nature Reviews Genetics* **11**: 665–667.

Pritchard JK, Pickrell JK, Coop G. 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology* **20**: R208–R215.

R Core Team. 2013. *R: a language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Rae AM, Robinson KM, Street NR, Taylor G. 2004. Morphological and physiological traits influencing biomass productivity in short-rotation coppice poplar. *Canadian Journal of Forest Research* **34**: 1488–1498.

Rellstab C, Zoller S, Tedder A, Gugerli F, Fischer MC. 2013. Validation of SNP allele frequencies determined by pooled next-generation sequencing in natural populations of a non-model plant species. *PLoS ONE* **8**: e80422.

Renaut S, Grassa CJ, Yeaman S, Moyers BT, Lai Z, Kane NC, Bowers JE, Burke JM, Rieseberg LH. 2013. Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Communications* **4**: 1827.

Rubin CJ, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT, Jiang L, Ingman M, Sharpe T, Ka S *et al.* 2010. Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* **464**: 587–591.

Schlötterer C, Dieringer D. 2005. A novel test statistic for the identification of local selective sweeps based on microsatellite gene diversity. In: Nurminsky D, ed. *Selective sweep*. New York, NY, USA: Landes Bioscience, 55–64.

Schlötterer S, Tobler R, Kofler R, Nolte V. 2014. Sequencing pools of individuals – mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics* **15**: 749–763.

Schluter D. 2000. *The ecology of adaptive radiation*. Oxford, UK: Oxford University Press.

Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, Peichel CL, Saetre GP, Bank C, Brannstrom A *et al.* 2014. Genomics and the origin of species. *Nature Reviews Genetics* **15**: 176–192.

Shriver MD, Smith MW, Jin L, Marcini A, Akey JM, Deka R, Ferrell RE. 1997. Ethnic-affiliation estimation by use of population-specific DNA markers. *American Journal of Human Genetics* **60**: 957–964.

Slavov GT, DiFazio SP, Martin J, Schackwitz W, Muchero W, Rodgers-Melnick E, Lipphardt MF, Pennacchio CP, Hellsten U, Pennacchio LA *et al.* 2012. Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree *Populus trichocarpa*. *New Phytologist* **196**: 713–725.

Smadja CM, Butlin RK. 2011. A framework for comparing processes of speciation in the presence of gene flow. *Molecular Ecology* **20**: 5123–5140.

Soil Survey Staff. 2010. *Keys to soil taxonomy*. Washington, DC: USDA-Natural Resources Conservation Service.

Solaro S. 2006. *Introduction to soil types of Ticino floodplain and pre-alps*. Ispra, Varese, Italy: ERSAF – Ente Regionale per i Servizi all'Agricoltura e alle Foreste Struttura Sviluppo Rurale, Suoli e Supporto alla Filiera Vitivinicola.

Stölting KN, Nipper R, Lindtke D, Caseys C, Waeber S, Castiglione S, Lexer C. 2013. Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. *Molecular Ecology* **22**: 842–855.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.

Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin SV. 2010. Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nature Genetics* **42**: 260–263.

Wang J, Kallman T, Liu J, Guo Q, Wu Y, Lin K, Lascoux M. 2014. Speciation of two desert poplar species triggered by Pleistocene climatic oscillations. *Heredity* **112**: 156–164.

Whitham TG, Bailey JK, Schweitzer JA, Shuster SM, Bangert RK, Leroy CJ, Lonsdorf EV, Allan GJ, DiFazio SP, Potts BM *et al.* 2006. A framework for community and ecosystem genetics: from genes to ecosystems. *Nature Reviews Genetics* **7**: 510–523.

Zhou L, Bawa R, Holliday JA. 2014. Exome resequencing reveals signatures of demographic and adaptive processes across the genome and range of black cottonwood (*Populus trichocarpa*). *Molecular Ecology* **23**: 2486–2499.

Zhou XJ, Li Q, Chen X, Liu JP, Zhang QH, Liu YJ, Liu KD, Xu J. 2011. The Arabidopsis RETARDED ROOT GROWTH gene encodes a mitochondria-localized protein that is required for cell division in the root meristem. *Plant Physiology* **157**: 1793–1804.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Histogram of read coverage per SNP and DNA pool.

Fig. S2 Genomic patterns of genetic diversity and differentiation in *P. alba* visualized by windowed analysis.

Table S1 Microsatellite genotypes

Table S2 Raw data for 1775 768 SNP loci from pool-seq WGS in *P. alba*

Table S3 Description and primer sequences for five Sanger-sequenced genes

Table S4 Number of haplotypes and reads recovered by Sanger sequencing and SOLiD4 pool-seq WGS

Table S5 SNP outlier detection and allele frequency differentials (AFDs) for comparisons among subpopulation replicates

Table S6 Number of genes, Fisher's exact tests and false discovery rates (FDRs) for 63 gene ontology (GO) terms with significant enrichment in high-differentiation outlier genes

Table S7 SNP distribution along chromosomes

Table S8 Description and gene ontology (GO) annotations of the 437 outlier genes

Methods S1 Reference-mapping and SNP detection from pool-seq WGS data.

Methods S2 SNP validation by Sanger sequencing.