Selection against recombinant hybrids maintains reproductive isolation in hybridizing *Populus* species despite F_1 fertility and recurrent gene flow

CAMILLE CHRISTE[1], KAI N. STÖLTING[1], LUISA BRESADOLA[1], BARBARA FUSSI[2], BERTHOLD HEINZE[3], DANIEL WEGMANN[1], CHRISTIAN LEXER[1,4]*

[1]University of Fribourg, Department of Biology, Chemin du Musée 10, CH-1700 Fribourg, Switzerland. [2]Bavarian Office for Forest Seeding and Planting, Applied Forest Genetics, Teisendorf, Germany. [3]Austrian Federal Research and Training Centre for Forests, Natural Hazards and Landscape, Department of Genetics, Seckendorff-Gudent-Weg 8, A-1130 Vienna, Austria. [4]University of Vienna, Department of Botany and Biodiversity Research, Rennweg 14, A-1030 Vienna, Austria

Keywords: Hybrid zones, admixture, speciation, RAD, genotyping by sequencing, local genetic ancestry, divergence, differentiation, *Populus*, common garden

[*] *Author for correspondence*: Christian Lexer, University of Vienna, Department of Botany and Biodiversity Research, Rennweg 14, A-1030 Vienna, Austria.

Tel: +43 1 4277 541 40. Email: christian.lexer@univie.ac.at

Running title: Genetic ancestry in poplar hybrid zones

Abstract

Natural hybrid zones have proven to be precious tools for understanding the origin and maintenance of reproductive isolation (RI) and therefore species. Most available genomic studies of hybrid zones using whole or partial genome resequencing approaches have focused on comparisons of the parental source populations involved in genome admixture, rather than exploring fine-scale patterns of chromosomal ancestry across the full admixture gradient present between hybridizing species. We have studied three well-known European 'replicate' hybrid zones of *Populus alba* and *P. tremula*, two wide-spread, ecologically divergent forest trees, using up to 432 505 Single Nucleotide Polymorphisms (SNPs) from Restriction site Associated DNA (RAD) sequencing. Estimates of fine-scale chromosomal ancestry, genomic divergence, and differentiation across all 19 poplar chromosomes revealed strikingly contrasting results, including an unexpected preponderance of F_1 hybrids in the centre of genomic clines on the one hand, and genomically localized, spatially variable shared variants consistent with ancient introgression between the parental species on the other. Genetic ancestry had a significant effect on survivorship of hybrid seedlings in a common garden trial, pointing to selection against early-generation recombinants. Our results indicate a role for selection against recombinant genotypes in maintaining RI in the face of apparent F_1 fertility, consistent with the intra-genomic 'coadaptation' model of barriers to introgression upon secondary contact. Whole genome resequencing of hybridizing populations will clarify the roles of specific genetic pathways in RI between these model forest trees and may reveal which loci are affected most strongly by its cyclic breakdown.

Introduction

When two differentiated populations or species come into contact, the evolutionary outcomes are highly variable and include the complete breakdown of reproductive barriers followed by the extinction of one or more taxa, the reinforcement and evolution of even stronger reproductive isolation (RI) between the parental taxa, the transfer of functionally important variants between divergent populations, and occasionally even hybrid speciation (Coyne & Orr 2004; The Heliconius Genome Consortium 2012; Hamilton & Miller 2015). Understanding the mechanisms that facilitate the maintenance of RI and trait differences in such contact zones is an important goal of molecular ecology and evolutionary biology (Coyne & Orr 2004; Arnold 2006). Genomic studies of hybrid zones have contributed greatly to our understanding of these mechanisms (Rieseberg *et al.* 1999; Whibley *et al.* 2006; Nolte *et al.* 2009; Teeter *et al.* 2010; Gompert *et al.* 2012; Abbott *et al.* 2013) and have recently experienced a remarkable renaissance, triggered in part by rapid advances in genomic technologies (e.g. The Heliconius Genome Consortium 2012; Ellegren 2014), the advancement of mathematical and verbal models for understanding barriers to gene flow (Barton& De Cara 2009; Feder *et al.* 2012; Uecker *et al.* 2015), and the development of novel analytical tools for studies of genome admixture (e.g. Price *et al.* 2009; Wegmann *et al.* 2011; Gompert & Buerkle 2013). Beyond biology, hybridization and introgression have recently attracted broad public interest because of the ongoing debate regarding social and genetic interactions among Neanderthals and modern humans (Racimo *et al.* 2015).

Considering that hybrid zones were first suggested as ‘natural labs’ for ecology, evolutionary and conservation biology already 30 years ago (Barton & Hewitt 1985; Hewitt 1988), astonishingly little is known about genomic patterns of ancestry along the chromosomes of animals and plants in contact zones (reviews by Feder *et al.* 2012; Ellegren 2014; Seehausen *et al.* 2014). The few available examples, reviewed by Gompert & Buerkle (2013), include dog-like canids (vonHoldt *et al.* 2011) and hominids (Racimo *et al.* 2015). More information on patterns of ancestry along admixture gradients would be highly desirable, because these patterns reveal the recent demographic and selective forces at play in such contact zones. As a direct result of genetic recombination, which will effectively re-shuffle the participating genomes in each generation, the mosaic of ancestries along hybrid chromosomes is particularly informative about (1) the demographic history of contact zones (Buerkle & Rieseberg 2008), (2) the strength of the barrier to gene flow (Coyne & Orr 2004; Barton & Hewitt 1985), and (3) the ecological impact (Whitham *et al.* 2006) and evolutionary outcomes (Abbott *et al.* 2013) of matings and admixture between highly divergent populations. Ideally, such studies of genomic ancestry in hybrids are combined with analyses of genomic differentiation between the parental populations, because these two sources of information cover complementary evolutionary time scales (Lexer *et al*. 2010; Lindtke *et al*. 2012; Gompert *et al*. 2012). This is especially relevant in organisms with a strong *a priori* hypothesis for past isolation-introgression cycles, such as Eurasian forest trees (Petit & Hampe 2006; Tzedakis *et al*. 2013).

Here, we studied patterns of ancestry in interspecific hybrid zones in the ‘model forest tree’ genus *Populus* (poplars, aspen, cottonwoods). This genus includes several pairs or groups of ecologically and economically important taxa with incomplete species barriers, i.e. taxa with ‘porous’ reproductive barriers (Martinsen *et al.* 2001; Lexer *et al.*2010; Geraldes *et al.* 2014). The Eurasian species complex of white poplars and aspens includes the wide-spread taxa *Populus alba* (White poplar) and *P. tremula* (European aspen) and represents a prime example for a group that would benefit greatly from knowledge of genetic ancestries in contact zones. The extensive ‘mosaic’ hybrid zones present in these forest foundation species contains a wealth of heritable phenotypic (e.g. morphological and chemical) variation (Lexer *et al.* 2009; Lindtke *et al.* 2013; Caseys *et al.* 2015), some of which is expected to affect entire associated communities of co-evolving herbivores and pathogens (Whitham *et al.* 2006; Bernhardsson *et al.* 2013; Caseys *et al.* 2015).

Low-resolution molecular ecology studies of these hybridizing species suggested a recombinant nature for genetically intermediate genotypes present in extensive, patchily distributed mosaic hybrid zones (Lexer *et al.* 2005; 2010; Lindtke *et al.* 2012). Further evidence for successful recombination and introgression came from population pair-wise genome scans (Lindtke *et al.* 2012; Stölting *et al.* 2013), and from molecular genetic analysis of experimental hybrids (Macaya-Sanz *et al.* 2011). Recent results from genotyping-by-sequencing (GBS) of trees from one hybrid zone locality, however, have cast doubts on the genomic make-up of hybrids from the genomic cline centre of these hybrid zones (Lindtke *et al.* 2014). That recent study focused on estimating genetic ancestries in open pollinated hybrid progeny, but it nevertheless also included 40 mature, genetically intermediate ‘hybrid references’ also studied here. Most (37) of these turned out to be F_1 hybrids based on GBS data (Lindtke et al. 2014), although previous microsatellite studies had suggested they were recombinants (Lindtke et al. 2012). Thus, these hybrid zones between two ecologically important tree species may serve as an illustrative example for some of the challenges encountered by present-day students of hybrid zones.

Here, we employ phased high density genetic marker data from Restriction site Associated DNA (RAD) sequencing to resolve some of the puzzles surrounding the genomic make-up and evolutionary dynamics of these poplar hybrid zones. We use our data to obtain novel insights into the genomic composition and fine-scale chromosomal ancestries (‘local ancestries’) of hybrids from three contact zones of these ecologically divergent species at a depth and level of precision rarely achieved in plant evolutionary genetics thus far. To make this possible, we adopt a modified approach previously used to study local ancestries in hominids (Price *et al*. 2009; Wegmann *et al*. 2011). We complement our analysis of local ancestry in hybrids by examining fine-scale patterns of genomic differentiation between the parental species, which may capture possible older gene flow episodes, and the effects of genetic ancestries on early fitness (survivorship) of hybrid seedlings in a common garden trial. In particular, we address the following specific questions:

(1) What are the relative proportions of F_1, early backcross, and later generation hybrids in three contact zone localities of these ecologically divergent forest trees?

(2) Based on genomic patterns of differentiation between the parental species, are there any indications for older admixture episodes?

(3) Can survivorship differences in a common garden trial explain the preponderance of F_1 genotypes found in natural populations?

Methods

Sampling design

This study was focused on two widespread European forest tree species, *P. alba* and *P. tremula*, and their hybrids (*P.* × *canescens*). These taxa were sampled in three natural mosaic hybrid zones (Lexer *et al.* 2010; Lindtke *et al.* 2012) situated along three major European river drainages: the Ticino river in Italy (45.28°N, 8.98°E), the Danube river in Austria (48.26°N, 16.23°E) and the Tisza river in Hungary (48.32°N, 22.26°E; coordinates are sampling mid points). In these three mosaic hybrid zone localities, *P. alba* and *P.* × *canescens* grow side by side in riverine lowland forest, while *P. tremula* grows on elevated sites along the river and in the surrounding hills. A total of 256 distinct individuals of these species and hybrids at their mature (reproductive) life stage, previously characterized with microsatellite markers (van Loo *et al.* 2008; Lexer *et al.* 2010; Lindtke *et al.*, 2012), were genotyped with RAD-seq in the present study.

RAD sequencing and bioinformatic data processing

DNA was extracted from silica-dried leaves with the Dneasy Plant Mini kit (Qiagen) following the manufacturer's instructions to obtain at least 1 ug of genomic DNA. RAD libraries were prepared following the protocol of Amores *et al.* (2011) using the restriction endonuclease PstI (New England Biolabs). This methylation-sensitive enzyme was selected based on a previous pilot study of these hybridizing species (Stölting *et al.* 2013). Eleven single-end libraries of maximum 24 individuals each barcoded with a set of unique tags comprising five to eight base pairs and differing by at least three nucleotides were sequenced on Illumina HiSeq 2000 devices.

After sequencing, reads were trimmed to 90 bases and individual barcode tags were removed with Stacks (Catchen *et al.* 2011). Twelve individuals with less than 2 Mio reads were removed from the dataset. Reads not containing the expected RAD restriction site and library adapters were detected with BLASTN (Altschul *et al.* 1997) and excluded from subsequent steps.

We used the *Populus trichocarpa* genome assembly v2.1 (http://www.phytozome.net/poplar) for reference-mapping, which is justified by high levels of macrosynteny among different poplar species and sections indicated by genetic mapping studies (Macaya-Sanz *et al.* 2011; Wang *et al*. 2011; 2n=38 chromosomes throughout the genus), but we note that little information is currently available on microsynteny in poplars. Bowtie 2.1.0 (Langmead & Salzberg 2012) was used in the local alignment mode to map reads to the reference genome. The local mode allows some characters to be omitted ("soft clipped") from the ends of the alignment in order to achieve the greatest possible alignment score. SAMtools 0.1.18 (Li *et al*. 2009) and Picard (https://http://sourceforge.net/projects/ picard/) were used for sorting, filtering and indexing the reads, and GATK 2.5.2 UnifiedGenotyper with the option EMIT_VARIANTS_ONLY and default parameters (DePristo *et al.* 2011) was used for variant calling and genotyping. VCFtools v0.1.11 (Danecek *et al.* 2011) was used to apply validation criteria to the SNPs called by GATK 2.5.2. Biallelic positions covered by 8-250 reads in each individual and exhibiting a genotype quality (GQ in VCF v4.1) >20 were retained. All subsequent analyses were carried out with SNPs that mapped to the first 19 scaffolds (i.e. chromosomes) of the *P. trichocarpa* reference genome.

Different missing data thresholds were applied in our study, depending on the particular needs of each analytical method. The amounts of missing data across loci and individuals were both used as filtering criteria. First, a general threshold of loci with a maximum of 50% missing was applied. Then individuals with >30% overall missing data for the parental reference categories (defined by ADMIXTURE, below) and >50% for the hybrid category (defined by ADMIXTURE) were removed from the dataset. For estimates of diversity and differentiation between populations, only loci with a maximum of 10% missing data were accepted. For local ancestry analyses, we retained only SNPs that were scored for all parental individuals within each hybrid zone, as required by the software.

We used ADMIXTURE 1.23 (Alexander *et al.* 2009) to group the individuals into three categories: *P. alba*, *P. tremula*, and hybrids, based on genome-wide ancestry proportions for *K*=2 genetic clusters as previously inferred by Lindtke *et al.* (2012). Individuals with admixture proportions of the minor ancestry <10^{-5} were assigned to the parental reference categories *P. tremula* and *P. alba*, respectively, and individuals with larger admixture proportions to the *P.* × *canescens* hybrid category for further analysis. Monomorphic loci, loci with singletons and loci with private doubletons (SNPs where the minor allele occurred in a single individual in homozygous state) were removed for this analysis. We did this before and after splitting the dataset into the three spatially separated hybrid zone localities. This allowed us to retain similar numbers of high quality SNPs in each hybrid zone, while reflecting local patterns of genomic diversity in each dataset as closely as possible.

Since the estimation of absolute divergence (Dxy, below) between species required the full sequence data including invariant (monomorphic) sites, we also used GATK 2.5.2 UnifiedGenotyper (DePristo *et al.* 2011) with the option EMIT_ALL_SITES to produce the complete genotypes for the sampled individuals of the two parental species. In general, we applied the same validation criteria as for local ancestry analyses in hybrids, with the difference that genotype quality scores could not be estimated for the invariant loci.

We also used our RAD-seq data to infer the ancestries of plastid DNA haplotypes in all three hybrid zone localities by aligning all RAD sequence reads against the *P. trichocarpa* plastid genome (Tuskan *et al.* 2006) following the same steps as for the nuclear genome. We then visualized each SNP with IGV (Robinson *et al.* 2011; Thorvaldsdóttir *et al.* 2013) to check for the absence of heterozygotes, the high coverage expected for plastid DNA (>250x), and the presence of a PstI restriction site in the reference genome supporting the read harboring the polymorphism. The plastid DNA haplotypes identified by RAD-seq were validated by comparison to haplotypes identified by conventional PCR-RFLP fingerprinting of subsets of the samples, following procedures used in previous studies (Lexer *et al.* 2005; Fussi *et al.* 2010; Supporting information).

Population genetic analyses

Inter- and intraspecific patterns of genomic diversity

Global patterns of nuclear genomic variation were assessed using Principal Component Analysis (PCA) using the program SmartPCA from the software EIGENSOFT v5.0.1 (Price *et al.* 2006; Patterson *et al.* 2006) on all SNPs that passed the filtering criteria, and before splitting the dataset into three hybrid zone localities. Gene diversity for nuclear RAD-seq data was estimated in Arlequin 3.5.1.3 (Excoffier & Lischer 2010). Haplotypic data for plastid DNA were analyzed in the form of median-joining networks using the program Network 4.6.1.1 (Bandelt *et al.* 1999), and by plotting the inferred species origin of plastid haplotypes along nuclear admixture gradients for each hybrid zone locality.

Patterns of local genetic ancestry along hybrid chromosomes

We used the program RASPberry (Wegmann *et al.* 2011) to infer switches of local ancestry from *P. alba* to *P. tremula* or *vice versa* (resulting from 'junctions'; Fisher1954) and thus reconstruct local ancestries along the 19 poplar chromosomes in interspecific hybrids of *P. alba* and *P. tremula.* Full documentation for the software is available on https://bitbucket.org/phaentu/raspberry/. RASPberry uses a set of reference populations to infer the ancestry of chromosomal segments in admixed individuals using a Hidden Markov Model (HMM) previously developed by Price *et al.* (2009). Because the HMM-based inference used by RASPberry requires phased haplotypes as parental references, we used FastPHASE 1.4.0 (Scheet & Stephens 2006) to phase our reference panels identified by ADMIXTURE in Italy, Austria, and Hungary (N= 32, 40 and 30 for *P. alba* and 46, 30 and 26 for *P. tremula*, respectively). For the inference of local ancestry, we assumed a constant default recombination rate of 5 cM/Mb, following estimates for *P. trichocarpa* by Tuskan *et al.* (2006), but identified suitable values for the remaining model parameters by choosing the combination resulting in the highest likelihood. The chosen parameters were: (1) time since admixture = 1 generation, (2) the population recombination rate, which governs the recombination events among haplotypes from the same population as 1000 for *P. alba* and 2000 for *P. tremula,* (3) the rate at which mutations are allowed to occur in the reference panels to 0.1 for both species, (4) the miscopying rate set to 0.01 and (5) the miscopying mutation rate set to 0.01. The miscopying rate accounts for incomplete lineage sorting in the parental populations as well as finite reference panels. We set the window size to 3 cM (based on physical and genetic mapping information from *P. trichocarpa*; above) to guarantee that long-range LD information was used in low recombination regions. Values were stored during the forward runs every second SNP to reduce the memory load. RASPberry output was used to classify hybrids into F_1 (heterospecific ancestry throughout the genome), 'imperfect F_1' (heterospecific ancestry genomes with small homospecific segments representing <10% of each chromosome), backcrosses (large hetero- and homospecific ancestry segments), alba- or tremula-like (presence of small hetero-specific segments in individuals that did not pass our strict parental reference ADMIXTURE thresholds) and F_n (advanced recombinant hybrids).

Genome scans for differentiation and divergence between the parental species

Regions with low proportions of fixed differences can be candidate regions for introgression of alleles from one species into another. A custom R script was used to calculate the proportions of SNPs that were differentially fixed between the two parental species in each hybrid zone locality ('fixratio' from here onwards). This simple measure of differentiation has previously been shown to be significantly correlated with F_{ST} (Stölting *et al.* 2013). We also calculated the absolute measure of divergence between species (Dxy), which is independent of levels of diversity within populations (Cruickshank & Hahn 2014). Dxy was calculated within windows of 500kb, sliding by 250kb. A sliding window analysis (window size = 500 kb; step size = 250 kb) was used to report both fixratio and Dxy together with relative SNP density for all three hybrid zone localities.

Linkage disequilibrium (LD) as a potential indicator of recent introgression

We checked LD along all chromosomes including all candidate regions for introgression identified in genome scans for species differentiation (above), because LD can point to introgressed chromosome blocks stemming from recent gene flow (Barton & Hewitt 1985; Lexer *et al.* 2006; Stölting *et al.* 2013). Linkage disequilibrium (D') between pairs of markers in each parental species and hybrid zone locality was calculated using the program ldmax available within the GOLD 1.1.0 software package (Abecasis & Cookson 2000). The results were plotted with the R package LDheatmap (Shin *et al.* 2006).

Using genetic ancestries to predict survivorship in hybrid seedlings

Unusual patterns of local ancestry revealed by our hybrid zone RAD-seq data (discussed below) motivated us to explore whether patterns of genetic ancestry might be related to survivorship in hybrid seedlings in a common garden trial. The trial was established at the University of Fribourg, Switzerland, in 2011, consisting of 137 seedlings from 15 open pollinated families from the Italian Ticino river hybrid zone (Caseys *et al.* 2015). Seedlings were germinated in the greenhouse and were then planted in an unbalanced block design with randomization within blocks. Seedlings were allowed to grow without any interference, except for regular watering. The trial forms part of a larger, reciprocal common garden experiment to be RAD-genotyped and phenotyped for the purpose of admixture mapping at a later time. Survivorship was scored as binary data (dead / alive) after 4 years of growth.

Although no RAD-seq data is available for the common garden seedlings at this time, ancestry estimates were available to us from a previous genotyping-by-sequencing (GBS) study comprising close to 12 000 SNPs (Lindtke *et al.* 2014). We note that GBS (Elshire *et al*. 2011; Parchman *et al*. 2012) and RAD-seq (Baird *et al*. 2008; Amores *et al*. 2011) are closely related methods from the same family of genotyping-by-sequencing approaches, thus the two methods should yield comparable results given proper data curation and analysis; this was indeed confirmed by our data (below, Discussion). A difference between the previous GBS and the present RAD-seq study is that in the former, SNPs were not reference-mapped against a chromosome-level genome assembly, thus they did not allow the estimation of local ancestries along hybrid chromosomes. Nevertheless, they allowed the estimation of overall ancestry proportions for individual trees. Here we used admixture proportions (q, equivalent to a hybrid index) and inter-source (=heterospecific) ancestry (Q12) previously estimated for each individual seedling using a Bayesian framework (Lindtke *et al.* 2014). Because genome-wide heterozygosity will decrease with each generation of recombination (Buerkle & Lexer 2008), reduced Q12 allows the detection of recombinant hybrids among genetically variable individuals from natural populations. The biological principles behind this approach were outlined by Milne *&* Abbott (2008) and Lexer *et al*. (2010), among others, and the statistical approach was fully described by Lindtke *et al*. (2014). Here, we tested if q or Q12 predicted juvenile survivorship in the common garden trial using mixed effect models implemented in the lme4 and MuMIn packages in R. The models implemented binomial logistic regressions of survivorship with both linear and quadratic effects of q and Q12, including the family membership of each seedling as a random effect. An AIC difference (delta AIC) threshold of <7 to the best supported model was employed to identify the three top-ranking models.

Results

Polymorphisms detected by RAD sequencing of poplar hybrid zones

RAD sequencing in hybrid zones of *P. alba* and *P. tremula* yielded on average 4.4 x 10^6 high-quality sequence reads per individual (**Table S1**). On average 85% of reads were successfully reference-mapped against the *P. trichocarpa* genome assembly, and 47% of these represented unique hits to the reference genome. Sequencing coverage was 39.95 ± 15.37 std. dev. reads per individual for a total number of 432 505 SNPs after quality filtering. After splitting the dataset into three hybrid zones, applying the filtering for monomorphic loci, singletons, private doubletons and removing loci with missing data in the parental species, the total numbers of SNPs retained were 73 410, 70689 and 98 152 in Italy, Austria, and Hungary, respectively. The differences in numbers of recovered SNPs between the three hybrid zones may be due to different amounts of missing data and different numbers of individuals present in the parental references. Based on ADMIXTURE analysis, our final sample sizes (no. of haploid chromosomes after quality filtering) in Italy, Austria, and Hungary, were 32, 40 and 30 for *P. alba*, 46, 30 and 26 for *P. tremula*, and 172, 32 and 68 for *P.* × *canescens* hybrids, respectively.

Intra- and interspecific patterns of genomic diversity

Principal Component Analysis (**Fig. 1**) identified one major axis explaining 35.38% of the nuclear genomic variation revealed by the RAD-seq data. This axis clearly separated the two species from each other and from genetically intermediate hybrids. Potentially introgressed, parental-like genotypes are visible based on their deviations from the parental extremes on the left and right ends of the graph. The eigenvalues for individuals along the first axis were strongly correlated with admixture proportions from ADMIXTURE 1.23 (r=0.993, p=2.2x10^{-16}). The second axis (PC2) explained 3.23 % of the genomic variation and revealed the geographic structure of the studied hybrid zone localities, with three clusters corresponding to the three localities in their longitudinal order (**Fig. 1**). *Populus alba* (left side of graph) displayed much stronger genetic differentiation between geographically separated localities than *P. tremula*. The same clear geographic structure along PC2 was also visible for genetically intermediate hybrids and backcrosses, thus indicating that the two species had been locally affected by genetic interactions, i.e. hybridization and interspecific gene flow.

In line with these results, genomic diversity was greater in *P. tremula* than in *P. alba*, and greatest in trees classified as hybrids (**Table 1**), as expected from a previous microsatellite study (Lindtke *et al.* 2012). Pairwise genetic differentiation (F_{ST}) between *P. alba* and *P. tremula* was similar across localities (average = 0.516; **Table S2**), and the same was the case for differentiation between hybrids and each parental species (average = 0.156 and 0.169 for *P. alba* and *P. tremula*, respectively; **Table S2**).

Plastid DNA haplotype diversity uncovered by RAD-seq

All high-quality, high-coverage reads that mapped to the *P. trichocarpa* plastid genome were found next to the expected PstI restriction site. In total, we identified seven informative plastid SNPs that were present in five unique haplotypes (**Tables S1 and S3**). While two and three of these haplotypes were found among the pure members (with respect to their nuclear genomes) of *P. alba* and *P. tremula,* none of them was shared between the two species (**Table S1**). The plastid DNA differentiation between the two species was broadly congruent with results from classical PCR-RFLP based fingerprinting of plastid genomes for subsets of the samples (**Table S1**), as determined by checking their positions in a median-joining network (**Fig. S1**).

Based on our plastid haplotype data defined by RAD-seq, F_1 hybrids had diagnostic haplotypes from *P. alba* and *P. tremula* with a ratio of 70% to 30% (**Table S1**, **Fig. 2, Figs. S2 and S3**), roughly as expected from the abundances of the two species in these localities (Lexer *et al.* 2005; van Loo *et al.* 2008; Lexer *et al.* 2010). For backcrosses to *P. alba* this ratio was 58% to 42% (**Table S1**, **Fig. 2, Figs. S2 and S3**), respectively. A *P. alba* plastid haplotype was also present among the very few backcrosses to *P. tremula* detected in our study (below) (**Table S1, Fig. S2**).

Local ancestry along hybrid chromosomes

We inferred the genetic ancestry along the chromosomes of all hybrid individuals (**Figs 2 and 3**, **Figs. S2 and S3**) and discovered four distinct and characteristic patterns: (1) In all three hybrid zone localities, the majority of hybrid individuals (85%, 56%, 74% in

Italy, Austria and Hungary respectively) showed genomes that were almost completely of heterospecific ancestry, consistent with them being newly formed hybrids of the first generation (F_1) (**Figs. 2 and 3, Table 2 and 3, Figs. S2 and S3**). (2) Sixteen individuals (11%) were identified as backcrosses (**Table 2, Fig. 3, Figs. S2 and S3**), with their genome composed of chromosome segments of heterospecific ancestry and all remaining segments homozygous for the same ancestry. Among these, only three were backcrosses to *P. tremula*, and the remaining 13 were backcrosses to *P. alba* **(Tables 2 and 3, Fig. 2, Figs. S2 and S3)**. The number of inferred ancestry switch points in these backcrosses was on average 22.67 ± 4.11 across all 19 chromosomes (**Table 3**), and hence in line with one or two recombination events per chromosome per generation. (3) Parental-like patterns with almost all chromosomes of the same ancestry were found for individuals with genome-wide admixture proportions close to the pure parental threshold (**Fig. 2**, **Figs. S2 and S3, Tables 2 and3**; **Table S1**). (4) A single individual exhibited a balanced contribution of homozygous ancestry segments from both species, an elevated number of inferred switch points (48) and generally smaller segments, all consistent with it being an advanced product of matings among hybrids (IT077, **Figs. 2 and 3**, **Tables 2 and 3**). Importantly, this individual was not distinguishable genetically from trees that were intermediate between the two parental species in our PCA analysis (**Fig. 1**).

We also detected small segments homozygous for one or the other species that were spread throughout the whole spectra of samples and that appeared sporadically (**Figs. 2 and 3, Figs. S2** and **S3**). Their presence may indicate artefacts specific to our analytical method or sample set, or traces of older introgression events. Although present in all categories of hybrid individuals, we report them only for the F_1 category as 'imperfect F_1', because this category was large enough to examine differences among the three hybrid zone localities (**Tables 2 and 3**, **Table S1**). These unexpected segments covered less than 10% of the chromosomes in which they were found and were very small compared to the other segments. In imperfect F_1, their length was on average 0.36±0.23 Mio bp and 0.62±0.59 Mio bp for segments homozygous for *P. alba* and *P. tremula*, respectively. (**Table 3**, **Fig. S4**).

Genome scans for differentiation and divergence between the parental species

The proportion of fixed SNPs between the two parental species was highly variable along the 19 chromosomes and across hybrid zones. Hungary was the locality with the lowest overall proportion of fixed differences (2.9%). Austria had the highest value (7.9%), while Italy was intermediate (5.8%). We considered genome regions with zero fixratio (i.e. an absence of fixed SNPs) in only a single hybrid zone locality as plausible candidates for local introgression. In contrast, genome regions with very low or zero fixed differences across localities can plausibly be explained by older introgression events or shared ancestral polymorphism. Many examples for both spatial patterns were found in windowed analyses (**Fig. 4**; **Fig. S5**). For example, regions with zero fixation exclusively in the Italian hybrid zone were found in chromosomes I, III, IV, VI, and X (**Fig. 4**; **Fig. S5**), and regions with zero fixed differences across all three hybrid zones were found in chromosomes III, V, XIII, XV, XVII, and XIX (**Fig. 4**; **Fig. S5**). Globally and accounting for overlapping windows, 8.8%, 15.5%, and 25.8% of the genome exhibited zero fixation in Austria, Italy, and Hungary, respectively. At the level of particular localities, these percentages were 10.3%, 1.4% and 16.4% for Italy, Austria, and Hungary, respectively, and 5.2% in the case of zero fixation regions shared among all three hybrid zones. In most genome regions, absolute divergence Dxy did not exhibit noticeable fluctuations in regions with low fixratio compared to their flanking regions (**Fig. 4**; **Fig. S5**).

Linkage disequilibrium

Linkage disequilibrium (LD) in the two species generally decayed within several hundred base pairs or few kilo bases (results for exemplary chromosomes in **Figs S6 and S7)**, consistent with expectations for *Populus* spp. (Ingvarsson *et al.* 2008, Slavov *et al.* 2012). The three localities were very similar for their levels of LD (**Figs S6 and S7)**. The shared low fixation region in chromosome III (a candidate region for ancient introgression or shared ancestral polymorphism; **Fig. 4**) did not exhibit increased LD compared to its flanking regions (**Fig. S6**). In contrast, LD was slightly but noticeably increased in the locality-specific low fixation region in Italy on chromosome I (**Fig. 4**) in one parental species, *P. tremula*. Here, the region of interest was distinguishable from its flanking regions in terms of LD (**Fig. S7**).

Predicting survivorship in hybrid seedlings from genetic ancestries

After four years of growth, 120 of the 137 seedlings grown in the common garden trial were recorded as alive, and 17 (12.4%) as dead. The top three mixed effect models predicting survivorship included different combinations of both linear and quadratic effects of q (admixture proportion) and Q12 (interspecific ancestry) (**Table S4**). The top-ranked model contained both linear and quadratic effects of q and the linear effect of Q12, and all three were significant (**Table S5;** p-value < 0.005). The positive quadratic effect of q indicates that mortality was highest in genetically intermediate hybrids, as also visible in **Fig. 5**. The positive effect of Q12 reflects the increased survivorship of hybrids with higher Q12, including F1 hybrids with maximum Q12 (**Fig. 5)**. Graphical inspection further indicated a tendency of increased survivorship for individuals with high q (**Fig. S8),** consistent with the preferential backcrossing towards *P. alba* seen in population genetic studies of these species (**Fig. 2**; **Figs S2** and **S3**; Lexer *et al*. 2005; Lindtke *et al*. 2014). The plotted models also confirmed the quasi-linear increase in survivorship for the biologically most relevant portion of parameter space from intermediate to high values of Q12 (**Fig. S9),** i.e. towards F_1 hybrids.

Discussion

The availability of novel tools from evolutionary genomics makes this the perfect time to realize the long-standing and often-cited prediction that hybrid zones may serve as ‘natural laboratories’ for studying the evolutionary process (Barton & Hewitt 1985; Hewitt 1988), for example by tracing the origin and fate of the ancestry segments affected by selection and drift during adaptation and speciation. Here, we applied high-density SNP genotyping using RAD-seq and a suite of bioinformatic and analytical tools to three *Populus* mosaic hybrid zones. The results provide a ‘snapshot’ of the likely biogeographic history of the genomic variation present in these ecologically divergent, hybridizing species at an unprecedented depth. They also reveal previously unknown aspects of the genomic make-up of these well-known hybrid zone localities, including fine-scale patterns of ancestry that we would not have expected based on genetic tools available years ago (Lexer *et al.* 2005; van Loo *et al.* 2008; Lindtke *et al.* 2012), and glimpses of genetic interactions in the ancient history of these highly divergent species that call for further research. Combining genomic data with a common garden trial, we also tested the hypothesis that the unusual genomic make-up of these hybrid zones may be caused by selection during the juvenile life stages of these long-lived forest trees.

From genome-wide ancestry to local ancestries along the chromosomes of hybrids

Our exploratory analysis of overall genomic ancestries by PCA (**Fig. 1**) recovered an "all-in-one" picture of numerous biogeographic patterns predicted by earlier studies of these hybridizing tree species: (1) strong genomic differentiation between species, a preponderance of genetically intermediate hybrids, the presence of backcrosses, and signs of parental-like introgressants (Lexer *et al.* 2010; Lindtke *et al.* 2012) along the first (horizontal) PCA axis; (2) stronger geographic differentiation in *P. alba* than in *P. tremula* along the second (vertical) axis, consistent with a model of geographically separate ice age refugia in southern European for the former and interconnected refugia for the latter species in central Europe (Fussi *et al.* 2010; Tzedakis *et al.* 2013); (3) a "sandwich"-like arrangement of genotypes from along the entire inter-specific axis (PC1) within three layers along PC2 corresponding to the three sampled geographic localities (Italy, Hungary, and Austria), thus reinforcing the notion of localized genetic contact after the divergence of geographic lineages within each species. The unusually divergent hybrid genotypes in Italy (blue outliers towards the top) may stem from hybridization with long-distance dispersers or with divergent genotypes from planted *P. alba* in the area. While overall genomic ancestries from PCA were thus broadly consistent with expectations from earlier work, the same cannot be said about local ancestries along hybrid chromosomes.

Estimates of local ancestries along individual hybrid chromosomes indicated that many of the genetically intermediate early-generation hybrids suspected to be recombinants in previous studies (Lexer *et al.* 2010; Lindtke *et al.* 2012) were in fact F_1's (**Figs. 2 and 3**; **Figs. S2 and S3**). Indeed, F_1 hybrids were the most dominant class of hybrids in all localities, encompassing 56% to 85% of all admixed individuals in the studied zones (**Table 2**). This included 37 genetically intermediate adult trees included as 'hybrid references' in a recent GBS study of seedling families from the Italian hybrid zone (Lindtke *et al.* 2014). All 37 trees had maximum genome-wide heterospecific ('intersource') ancestries inferred from unmapped SNP data in that earlier study, and all exhibited genomic ancestries consistent with F_1 status in the present study, which effectively cross-validates these two statistical approaches. If our estimates of local ancestry in three spatially separated hybrid zones are correct, then contact zone localities of *Populus alba* and *P. tremula* may be seen as 'F_1-dominated hybrid zones', in which later generation hybrids are rare despite apparent F_1 fertility (Milne *et al.* 2003; Milne & Abbott 2008; Lindtke *et al*. 2014). Nevertheless, the localities studied here do not fit well with the expectation that F_1-dominated hybrid zones should form preferentially in undisturbed settings (Milne *et al.* 2003), since riverine flood plain forests are characterized by strong natural disturbance regimes (Karrenberg *et al.* 2002). So other causes must be responsible for the preponderance of F_1 hybrids in these tree hybrid zones, and we shall re-visit these below.

Traces of recombination in F_1-dominated hybrid zones inferred from homospecific ancestry segments of different lengths

Our analysis of local ancestry also revealed the presence of early-generation backcrosses (mainly to *P. alba*), characterized by large homo- and heterospecific ancestry segments (**Table 3**). Backcrosses made up between 8% and 37% of hybrids in each locality (**Fig. 2**; **Figs. S2 and S3; Table 2**), which is likely too high to prevent the introgression and spread of neutral or advantageous genetic variants in the long term (Slatkin 1976; Barton 2001; Roux *et al.* 2013). This is consistent with results from previous genomic cline-based studies of fewer markers (Lexer *et al.* 2010) and qualitative patterns seen along the first (horizontal) genomic PCA axis in our present RAD-seq study (**Fig. 1**). Further, the genomes of many of our inferred F_1 were interrupted by small segments of homospecific ancestry (**Tables 2 and 3**; **Fig. S4**). These 'imperfect F_1' may point to methodological limitations of our analysis, e.g. fine-scale departures from synteny with our reference genome, or limits to the precise estimation of parental haplotype frequencies with our reference samples. Nevertheless, it seems unlikely that all of these homospecific ancestry segments are fully explained by statistical artefacts alone. Their small sizes (**Table 3;** see Results) may also point to admixture in the more distant past. Given the high recombination rates in poplar (5 cM/MB, five times higher than in humans; Tuskan *et al*. 2006) and assuming that the size of ancestry segments follows an exponential distribution (Price *et al*. 2009), small homospecific segments such as those seen in 'imperfect F1' (on average 0.5 megabases; **Table 3**) may have arisen within as few as 10 generations (i.e. 200 years assuming a generation time of 20 years).Thus, detecting older admixture pulses (e.g. those pre-dating the last glaciation) from local ancestry patterns would benefit from the much greater marker densities afforded by whole genome resequencing. Within the present RAD-seq study, more direct indications for ancient gene flow were gleaned by analyzing genomic patterns of differentiation between the parental species.

Genomic and spatial patterns of species differentiation: footprints of past gene flow episodes

Whereas many speciation genomic studies have focused on especially divergent genome regions, i.e. the so-called genomic 'islands' or 'continents' of speciation (Nosil *et al.* 2009; Feder *et al.* 2012), highly divergent taxa at a late stage of speciation such as ours offer the opportunity to do just the opposite: to search for low-differentiation genomic 'hotspots' of introgression across strong and genomically wide-spread species barriers (Stölting *et al.* 2013; Roux *et al.* 2013). This is equivalent to detecting 'pores' in the genome through which genetic variants may pass (Wu 2001), thus contributing genetic variation to the recipient gene pools and potentially affecting the fate of speciating pairs or groups of taxa (Gavrilets & Vose 2005; Abbott *et al.* 2013). In the absence of multiple taxa allowing polarized comparisons (e.g. four-taxon ABBA / BABA tests for introgression; The Heliconius Genome Consortium 2012), insights into the potential causes of unusually strong allele sharing in low-differentiation 'hotspots' may be gleaned by analyzing genomic variation in a spatial, biogeographic context (Muir & Schlötterer 2005), a genomic-map based context (Stölting *et al.* 2013), or both.

Our RAD-based scan for genome regions with particularly low proportions of fixed SNPs revealed low-differentiation ‘hotspots’ that were highly localized in the genome (**Fig. 4; Fig. S5**). A previous study has confirmed the statistical significance of this phenomenon using computationally intensive genome-wide autocorrelation analysis in one of these hybrid zone localities (Stölting *et al.* 2013). Many of these low-differentiation genome regions were shared between our three spatially separated hybrid zone localities, thus pointing to shared ancestral variation, shared selection pressures, or ancient introgression events, pre-dating geographic divergence of populations in each species (e.g. chromosome III, **Fig. 4**). Some low-differentiation regions, on the other hand, were found in only one or two localities, consistent with more recent introgression after geographic divergence of populations (e.g. chromosome I, **Fig. 4**). Although these patterns are suggestive of past gene flow, we must note that we can currently not rule out shared ancestral polymorphism and drift as potential causes for spatially uniform and variable patterns of differentiation. Linkage disequilibrium (LD) (**Fig. S7**) was of limited help in defining regions stemming from recently introgressed haplotypes in our data, presumably due to the known, rapid decay of LD in aspens (Ingvarsson 2008) and the relatively low marker densities afforded by RAD-seq. Nevertheless, our example genome region for recent, geographically localized introgression on chromosome I (**Fig. 4**) did exhibit increased LD in one of the two species (**Fig. S7**), consistent with recent introgression.

Selection against recombinants in a common garden trial

A particular life history characteristic of trees is their long juvenile period with very high rates of juvenile mortality, which implies very strong selection at early life stages (Petit & Hampe 2006; Larcombe *et al.* 2015). Based on a comparison of genomic ancestries in 1st year seedlings and mature trees from the same hybrid zone locality of *P. alba* and *P. tremula*, Lindtke *et al.* (2014) hypothesized the presence of strong postzygotic selection against recombinant hybrid genotypes (F_n and backcrosses) acting between the very early seedling stage and maturity. Indeed, a broad range of recombinant genotypes was present among 1st year seedlings, whereas mainly genetically intermediate hybrids with high heterospecific ancestry (confirmed to indicate F_1 hybrid status above) and parental-like genotypes were found in mature trees (Lindtke *et al.* 2014). Since the seedlings studied by Lindtke *et al.* (2014) were planted in a common garden trial (Caseys *et al.* 2015), we had the opportunity to test whether selection at the early juvenile stage (the first four years) may indeed contribute to the dominance of such F_1-like genotypes among mature trees in hybrid zones (**Fig. 2**; **Figs S2 and S3**).

Together, admixture proportions (q) and interspecific ancestry (Q12), the proportion of the genome with ancestry in different source species (Lindtke *et al*. 2014), were able to explain 4thyear survivorship in our common garden trial. In effect, selection was primarily against genetically intermediate hybrids (intermediate q) with lower-than-maximum Q12 (**Fig. 5**), i.e. recombinant hybrids (Buerkle & Lexer 2008; Milne *et al*. 2008; Lindtke *et al*. 2014). Thus, genomic patterns of survival and mortality in our common garden trial (**Fig. 5**) suggest that selection against recombinant hybrids contributes to the high proportion of F_1 hybrids and strong RI seen in poplar hybrid zones (**Fig.2; Figs S2 and S3**). The increased survivorship of individuals with high Q12 (F_1 hybrids) and high q (including *P. alba*-like backcrosses) (**Figs. S8 and S9**) also indicates a potential for the occasional, partial breakdown of RI due to heterosis and subsequent introgression towards *P. alba*. We note the caveat that hybrid survivorship was assayed in only one environment using germplasm from a single hybrid zone locality. Testing more localities and environments may yield a more nuanced picture of hybrid survivorship and fitness, and these experiments should ideally be carried out by comparing larger cohorts of experimental F_1 and recombinant hybrids.

Genomics of incomplete barriers to gene flow in secondary contact

Our results point to a 'paradox' (Roux *et al.* 2013; Stölting *et al.* 2013) that we expect will be encountered fairly often in highly divergent, hybridizing taxa once studied with genomic tools: even strong and seemingly 'genome-wide' barriers to gene flow, such as those suggested by our local ancestry analysis of European poplar hybrid zones (above), are unlikely to prevent introgression in the long-term (**Fig. 4; Fig. S5**). Although long expected from population genetic theory (Slatkin 1976; Barton 2001), this contrast is not always immediately intuitive to empirical students and researchers in ecology and biogeography. Results such as these raise the question regarding the build-up and maintenance of genomically wide-spread RI in the face of recurrent gene flow in broad zones of sympatry, especially when secondary contact is established repeatedly following the onset of speciation.

In cases with apparent selection against recombinant hybrids as seen here (**Fig. 5**), wide-spread epistatic gene interactions are a highly plausible hypothesis for the maintenance of RI (Gavrilets *et al.* 2003; Coyne & Orr 2004). These can take the form of classical Dobzhansky-Muller incompatibilities (DMI's; reviewed by Gavrilets *et al.* 2003), or of intra-genomic coadaptation of loci within genetic or biochemical pathways (Johnson & Porter 2000; Edmands & Timmerman 2003). In fact, genetic evidence supporting the presence of epistasis in these poplar hybrid zones was put forward previously (Lexer *et al.* 2010), and simulation studies suggest that genotypic patterns in these poplar hybrids are compatible with the intra-genomic coadaptation model of RI (Lindtke & Buerkle 2015; Dorothea Lindtke, personal communication). This agrees well with the observation that F_1 hybrids of these species are at least partially fit and fertile (Macaya-Sanz *et al*. 2011; Lindtke *et al*. 2014), and with their preponderance in nature (this study; **Fig. 2**). Both observations would not necessarily be expected for classical DMI's (Barton & Bengtsson 1986; Gavrilets 1997). Available metabolomic data also point to coadaptation within genetic pathways. The flavonoid pathway (Rausher *et al.* 1999) appears to be a case in point, since its metabolites exhibit strong genetic control in poplars, many quantitative trait loci (QTL) for flavonoid abundances are segregating in hybrids between *P. alba* and *P. tremula*, and their allelic composition has apparently been shaped by heterosis (Caseys *et al.* 2015). Heterosis is also supported by the increased survivorship of seedlings with high intersource ancestry (Q12) in our common garden trial (**Fig. 5**; **Fig. S8**), in addition to the clear preponderance of F_1 hybrids in nature (**Fig. 2**; **Figs. S2** and **S3**). It is possible that slightly deleterious alleles segregating in these species trigger heterosis in F_1's (Bierne *et al*. 2002; Harris &

Nielsen 2015). This hypothesis is currently awaiting rigorous tests based on patterns of DNA sequence evolution in protein-coding genome regions of these species.

Many different mechanisms may cause the origin and maintenance of RI in the face of gene flow (Coyne & Orr 2004; Baack *et al.* 2015), and genomic studies of hybridizing species offer superb opportunities to address these. Mutation load in the donor species may predict selection against introgressed genome segments in the recipient species with or without epistasis (Harris & Nielsen 2015). Cyto-nuclear interactions may contribute to genomic isolation, although this seems less likely in these three hybrid zone localities, where cytoplasms of each species appear to combine freely with different nuclear genomes in hybrids (Results), with haplotype frequencies roughly matching the species' abundances in floodplain forests (Lexer *et al*. 2005; van Loo *et al*. 2008). In addition to post-zygotic mechanisms, pre-mating (e.g. phenology) and early post-mating barriers (e.g. pollen-stigma interactions and pollen competition) may also contribute to RI in plants (Baack *et al.* 2015) including poplars (Rajora 1989; Hall *et al.* 2007; Lindtke *et al.* 2014). Thus, selection against hybrids, as inferred here, may in principle also 'reinforce' existing pre-zygotic barriers stemming from sexual selection among gametes (Rajora 1989) or differences in phenology. Genomic studies of hybridizing species allow explicit tests of these hypotheses, especially if genomic patterns of ancestry and differentiation are examined for populations with different geographic distances to contact zones. In summary, selection against recombinant hybrids appears to maintain RI during secondary contact of ecologically divergent, hybridizing Eurasian poplar species, leading to F_1-dominated hybrid zones, but the precise genetic and molecular mechanisms responsible remain to be identified.

Acknowledgements

We thank Hans Herz, Wilfried Nebenführ, Stefano Castiglione, Stefano Gomarasca, Denes Bartha, and István Asztalos for help during fieldwork, Alexa Oppliger, Eligio Bossolini, and Thelma Barbará for help in the lab, the Lausanne and Bern genomics facilities for Illumina sequencing, Josef Schöpfer, David Frey and other colleagues for help with the common garden trial, Dorothea Lindtke, Celine Caseys, Alex Buerkle, Om Rajora, and Sven Bacher for helpful discussions, and three anonymous reviewers for insightful comments. This work was supported by grants PDFMP3_134660 and 31003A_149306 of the Swiss National Science Foundation (SNSF) to CL.

References

Abbott R, Albach D, Ansell S *et al.* (2013) Hybridization and speciation. *Journal of Evolutionary Biology*, **26**, 229–246.

Arnold ML (2006) Evolution through Genetic Exchange. In: Oxford University Press, Oxford, UK.

Abecasis GR, Cookson WO (2000) GOLD — Graphical overview of linkage disequilibrium. *Bioinformatics*, **16**, 182–183.

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, **19**, 1655–1664.

Altschul S, Madden T, Schaffer A *et al.* (1997) Gapped BLAST and PSI- BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389–3402.

Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH (2011) Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. *Genetics*, **188**, 799–808.

Baack E, Melo MC, Rieseberg LH, Ortiz-Barrientos D (2015) The origins of reproductive isolation in plants. *New Phytologist*, **207**, 968–984.

Babst BA, Chen HY, Wang HQ, Payyavula RS, Thomas TP, Harding SA, et al (2014) Stress-responsive hydroxycinnamate glycosyltransferase modulates phenylpropanoid metabolism in *Populus*. *Journal of Experimental Botany*, **65**, 4191–4200.

Baird NA, Etter PD, Atwood TS et al. (2008) Rapid SNP discoveryand genetic mapping using sequenced RAD markers.*PLoS ONE*, **3**, e3376.

Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Barton NH (2001) The role of hybridization in evolution. *Molecular Ecology*, **10**, 551–568.

Barton N, Bengtsson BO (1986) The barrier to genetic exchange between hybridising populations. *Heredity*, **57**, 357–376.

Barton NH, De Cara MAR (2009) The evolution of strong reproductive isolation. *Evolution*, **63**, 1171–1190.

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

Bernhardsson C, Robinson KM, Abreu IN *et al.*(2013) Geographic structure in metabolome and herbivore community co-occurs with genetic structure in plant defence genes. *Ecology Letters*, **16**, 791–798.

Bierne N, Lenormand T, Bonhomme F, David P (2002) Deleterious mutations in a hybrid zone: can mutational load decrease the barrier to gene flow?*Genetics Research*, **80**, 197-204.Buerkle CA, Lexer C (2008) Admixture as the basis for genetic mapping. *Trends in Ecology and Evolution*, **23**, 686–694.

Buerkle CA, Rieseberg LH (2008) The rate of genome stabilization in homoploid hybrid species. *Evolution*, **62**, 266–275.

Caseys C, Glauser G, Stölting KN *et al.* (2012) Effects of interspecific recombination on functional traits in trees revealed by metabolomics and genotyping-by-resequencing. *Plant Ecology & Diversity*, **5**, 457–471.

Caseys C, Stritt C, Glauser G, Blanchard T, Lexer C (2015) Effects of hybridization and evolutionary constraints on secondary metabolites: The genetic architecture of phenylpropanoids in European *Populus* species. *PLoS ONE*, **10**, e0128200.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics*, **1**, 171–182.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Sunderland, MA.

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, **23**, 3133–3157.

Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

DePristo MA, Banks E, Poplin R *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, **43**, 491–498.

Edmands S, Timmerman CC (2003) Modeling factors affecting the severity of outbreeding depression. *Conservation Biology*, **17**, 883–892.

Ellegren H (2014) Genome sequencing and population genomics in non-model organisms. *Trends in Ecology and Evolution*, **29**, 51–63.

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.

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

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

Fisher RA (1954) A fuller theory of “junctions” in inbreeding. *Heredity*, **8**, 187–197.

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.

Gavrilets S (1997) Hybrid zone with Dobzhansky-type epistatic selection. *Evolution*, **51**, 1027–1035.

Gavrilets S (2003) Perspective: models of speciation: what have we learned in 40 years? *Evolution*, **57**, 2197–2215.

Gavrilets S, Vose A (2005) Dynamic patterns of adaptive radiation. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 18040–18045.

Geraldes A, Farzaneh N, Grassa CJ *et al.* (2014) Landscape genomics of *Populus trichocarpa*: the role of hybridization, limited gene flow and natural selection in shaping patterns of population structure. *Evolution*, **68**, 3260–3280.

Gompert Z, Buerkle CA (2013) Analyses of genetic ancestry enable key insights for molecular ecology. *Molecular Ecology*, **22**, 5278–5294.

Gompert Z, Parchman TL, Buerkle CA (2012) Genomics of isolation in hybrids. *Philosophical Transactions of the Royal Society B*, **367**, 439–450.

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.

Hamilton JA, Miller JM (2015) Adaptive introgression: a resource for management and genetic conservation in a changing climate. *Conservation Biology*, doi: 10.1111/cobi.12574.

Harris K, Nielsen R (2015) The genetic cost of Neanderthal introgression. Bio-archive, doi: http://dx.doi.org/10.1101/030387

Hewitt GM (1988) Hybrid zones - natural laboratories for evolutionary studies. *Trends in Ecology and Evolution,***3**,158-167.

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

Johnson NA, Porter AH (2000) Rapid speciation via parallel, directional selection on regulatory genetic pathways. *Journal of Theoretical Biology*, **205**, 527–542.

Karrenberg S, Edwards PJ, Kollmann J (2002) The life history of Saliacaceae living in the active zone of floodplains. *Freshwater Biology*, **47**, 733–748.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–359.

Larcombe MJ, Holland B, Steane DA *et al.* (2015) Patterns of reproductive isolation in *Eucalyptus* – a phylogenetic perspective. *Molecular Biology and Evolution*, **32**, 1833–1846.

Lexer C, Fay MF, Joseph JA, 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, Joseph J, Loo M Van *et al.*(2009) The use of digital image-based morphometrics to study the phenotypic mosaic in taxa with porous genomes. *Taxon*, **58**, 5–20.

Lexer C, Joseph JA, van Loo M *et al.*(2010) Genomic admixture analysis in European *Populus* spp. reveals unexpected patterns of reproductive isolation and mating. *Genetics*, **186**, 699–712.

Lexer C, Kremer A, Petit RJ (2006) Shared alleles in sympatric oaks: recurrent gene flow is a more parsimonious explanation than ancestral polymorphism. *Molecular Ecology*, **15**, 2007–2012.

Li H, Handsaker B, Wysoker A *et al.* (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Lindtke D, Buerkle CA (2015) The genetic architecture of hybrid incompatibilities and their effect on barriers to introgression in secondary contact. *Evolution*, doi:10.1111/evo.12725.

Lindtke D, Buerkle CA, Barbará T *et al.*(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, González-Martínez SC, Macaya-Sanz D, Lexer C (2013) Admixture mapping of quantitative traits in *Populus* hybrid zones: power and limitations. *Heredity*, **111**, 474–85.

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

Macaya-Sanz D, Suter L, Joseph J *et al.*(2011) Genetic analysis of post-mating reproductive barriers in hybridizing European *Populus* species. *Heredity*, **107**, 478–486.

Martinsen GD, Whitham TG, Turek RJ, Keim P (2001) Hybrid populations selectively filter gene introgression between species. *Evolution*, **55**, 1325–1335.

Milne RI, Abbott RJ (2008) Reproductive isolation among two interfertile *Rhododendron* species: low frequency of post-F1 hybrid genotypes in alpine hybrid zones. *Molecular Ecology*, **17**, 1108–1121.

Milne RI, Terzioglu S, Abbott RJ (2003) A hybrid zone dominated by fertile F1s: maintenance of species barriers in *Rhododendron*. *Molecular Ecology*, **12**, 2719–2729.

Muir G, Schlötterer C (2005) Evidence for shared ancestral polymorphism rather than recurrent gene flow at microsatellite loci differentiating two hybridizing oaks (*Quercus* spp.). *Molecular Ecology*, **14**, 549–561.

Nolte AW, Gompert Z, Buerkle CA (2009) Variable patterns of introgression in two sculpin hybrid zones suggest that genomic isolation differs among populations. *Molecular Ecology*, **18**, 2615–2627.

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

Parchman TL, Gompert Z, Mudge J, Schilkey F, Benkman CW, Buerkle CA (2012) Genome wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology*, **21**, 2991–3005.

Patterson N, Price AL, Reich D (2006) Population structure and eigen analysis. *PLoS Genetics*, **2**, e190.

Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 187–214.

Price AL, Patterson NJ, Plenge RM *et al.* (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, **38**, 904–909.

Price AL, Tandon A, Patterson N *et al.* (2009) Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genetics*, **5**, e1000519.

Racimo F, Sankararaman S, Nielsen R, Huerta-Sánchez E (2015) Evidence for archaic adaptive introgression in humans. *Nature Reviews Genetics*, **16**, 359–371.

Rajora OP (1989) Sexual plant reproduction pollen competition among *Populus deltoides* Marsh ., *P. nigra* L. and *P. maximowiczii* Henry in fertilizing *P. deltoides* ovules and siring its seed crop. *Sexual Plant Reproduction*, **2**, 90–96.

Rausher MD, Miller RE, Tiffin P (1999) Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. *Molecular Biology and Evolution*, **16**, 266–274.

Rieseberg LH, Whitton J, Gardner K (1999) Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics*, **152**, 713–727.

Robinson JT, Thorvaldsdóttir H, Winckler W *et al.* (2011) Integrative genomics viewer. *Nature Biotechnology*, **29**, 24–26.

Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species barrier: genomic hotspots of introgression between two highly divergent *Ciona intestinalis* species. *Molecular Biology and Evolution*, **30**, 1574–1587.

Scheet P, Stephens M (2006) A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *American Journal of Human Genetics*, **78**, 629–644.

Seehausen O, Butlin RK, Keller I *et al.*(2014) Genomics and the origin of species. *Nature Reviews Genetics*, **15**, 176–192.

Shin J, Blay S, McNeney B, Graham J (2006) LDheatmap : An R function for graphical display. *Journal of Statistical Software*, **16**, 1–10.

Slatkin M (1976) The rate of spead of an advantageous allele in a subdivided population. In: *Population Genetics and Ecology* (eds Karlin S, Nevo E), pp. 767-780. Academic Press, Inc., New York.

Slavov GT, DiFazio SP, Martin J *et al.* (2012) Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree *Populus trichocarpa*. *New Phytologist*, **196**, 713–725.

Stölting KN, Nipper R, Lindtke D *et al.*(2013) Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. *Molecular Ecology*, **22**, 842–855.

Stölting KN, Paris M, Heinze B *et al.* (2015) Genome-wide patterns of differentiation and spatially varying selection between postglacial recolonization lineages of *Populus alba* (Salicaceae), a widespread forest tree. *New Phytologist*, **207**, 723–734.

Teeter KC, Thibodeau LM, Gompert Z *et al.* (2010) The variable genomic architecture of isolation between hybridizing species of house mice. *Evolution*, **64**, 472–485.

The Heliconius Genome Consortium (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, **487**, 94–98.

Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, **14**, 178–192.

Tuskan GA, Difazio S, Jansson S *et al.* (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, **313**, 1596–1604.

Tzedakis PC, Emerson BC, Hewitt GM (2013) Cryptic or mystic? Glacial tree refugia in northern Europe. *Trends in Ecology & Evolution*, **28**, 696–704.

Uecker H, Setter D, Hermisson J (2015) Adaptive gene introgression after secondary contact. *Journal of Mathematical Biology*, **70**, 1523–1580.

vonHoldt BM, Pollinger JP, Earl DA *et al.* (2011) A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. *Genome Research*, **21**, 1294–1305.

Wang Y, Zhang B, Sun X, Tan B, Xu LA, Huang M, Wang M (2011) Comparative genome mapping among *Populus adenopoda*, *P. alba*, *P. deltoides*, *P. euramericana* and *P. trichocarpa*. *Genes and Genetic Systems,* **86**, 257-68.

Wegmann D, Kessner DE, Veeramah KR *et al.*(2011) Recombination rates in admixed individuals identified by ancestry-based inference. *Nature Genetics*, **43**, 847–853.

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

Whitham TG, Bailey JK, Schweitzer JA *et al.*(2006) A framework for community and ecosystem genetics: from genes to ecosystems. *Nature Reviews Genetics*, **7**, 510–523.

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

Data accessibility

Sequence raw data are compiled under NCBI-SRA accession ID SRP069231. VCF files with nuclear and plastid RAD-seq variant data for each hybrid zone locality are provided as DRYAD accession number doi:10.5061/dryad.pj44k.

Authors contributions

CC and CL conceived the study, CC, LB, BF, and BH gathered data, CC, DW, KNS, LB, and CL analyzed data, CC and CL wrote the paper with input and revisions from all coauthors.

Table 1. Gene diversity estimated for populations of *P. alba*, *P. tremula* and *P.* × *canescens* hybrids and each hybrid zone locality from RAD-seq data.

		Gene diversity per population	**Overall gene diversity per locality**
Italy	*P. alba*	0.048 ± 0.023	0.098 ± 0.046
	P. tremula	0.075 ± 0.036	
	Hybrids	0.097 ± 0.046	
Austria	*P. alba*	0.054 ± 0.026	0.099 ± 0.047
	P. tremula	0.076 ± 0.037	
	Hybrids	0.101 ± 0.049	
Hungary	*P. alba*	0.056 ± 0.027	0.098 ± 0.047
	P. tremula	0.077 ± 0.038	
	Hybrids	0.099 ± 0.047	

Table 2. Percentage and number of individuals (in brackets) of each hybrid class inferred based on RASPberry genetic ancestries for all three hybrid zone localities.

Hybrid class	**Italy**	**Austria**	**Hungary**
P. alba-like	4.7 (4)	0.0	8.8 (3)
Backcross to *P. alba* (BCA)	5.8 (5)	31.3 (5)	8.8 (3)
F_1	44.2 (35)	43.8 (2)	26.5(16)
Imperfect F_1	40.7 (38)	12.5 (7)	47.1 (9)
F_n	1.2 (1)	0.0	0.0
Backcross to *P. tremula* (BCT)	2.3 (2)	6.3 (1)	0.0
P. tremula-like	1.2 (1)	6.3 (1)	8.8 (3)

Table 3. Means and standard deviations of admixture proportions (q), total numbers of ancestry switch points, % of heterospecific ancestry, and sizes in Mio bp. of all three inferred types of ancestry segments (homozygous *P. alba*, homozygous *P. tremula*, heterospecific) inferred for different classes of hybrid genotypes from three localities.

Hybrid class	Admixture proportion q	Total nb. of switch points	% hetero ancestry	Segment size hom / alba	Segment size hom / tremula	Segment size hetero ancestry
P. alba-like	0.03 ± 0.01	0.57 ± 0.91	0.00 ± 0.00	19.67 ± 8.88	-	0.47 ± 0.39
BCA	0.29 ± 0.07	22.67 ± 4.51	0.62 ± 0.33	7.82 ± 7.84	0.51 ± 0.48	12.49 ± 7.95
F_1	0.50 ± 0.02	0	1.00 ± 0.01	-	-	18.29 ± 8.88
Imperfect F_1	0.49 ± 0.02	2.07 ± 1.12	1.00 ± 0.01	0.36 ± 0.23	0.62 ± 0.59	18.07 ± 9.07
F_n	0.46	48	0.58 ± 0.32	4.60 ± 5.19	3.95 ± 4.41	6.19 ± 5.71
BCT	0.67 ± 0.06	22.67 ± 3.71	0.62 ± 0.33	0.44 ± 0.49	8.30 ± 5.38	9.44 ± 6.50
P. tremula-like	0.94 ± 0.02	0.60 ± 1.21	0.04 ± 0.20	-	16.89 ± 9.44	0.17 ± 0.07

Figure legends

Figure 1. Principal Component Analysis (PCA) based on 432 505 SNPs illustrating the population structure of *P. alba* (rectangles), *P. tremula* (triangles), and their hybrids (circles) in the Italian (blue), Austrian (green), and Hungarian (orange) hybrid zone locality, respectively. Species and hybrids were delimited based on admixture proportions from ADMIXTURE 1.23 as described in text.

Figure 2. (A) Local genetic ancestry along all 19 poplar chromosomes of hybrids from the Austrian hybrid zone locality inferred from polymorphic markers using a Hidden Markov Model (HMM) of admixture, including genome segments homozygous for *P. alba* (blue), segments homozygous for *P. tremula* (orange), segments with heterospecific ancestry (green), and undefined segments (white). (B) Genome-wide admixture proportions from ADMIXTURE 1.23and the species origin of plastid DNA for each individual (*P. alba*, blue; *P. tremula*, orange; no data, white).

Figure 3. Local genetic ancestry along all 19 poplar chromosomes of four exemplary individuals, representing F_1, imperfect F_1, backcrosses to *P. alba* (BCA), and advanced recombinants (F_n). Shown are genome segments homozygous for *P. alba* (blue), segments homozygous for *P. tremula* (orange), segments with heterospecific ancestry (gray), and undefined segments (white).

Figure 4. Patterns of SNP density, interspecific divergence (Dxy) and differentiation (fixratio, proportion of fixed SNPs among all variable SNPs) between *Populus alba* and *P. tremula* for the Italian (blue), Austrian (green), and Hungarian (orange) hybrid zone locality, respectively. Shown are sliding window analyses (window size: 500 kb, step size: 250 kb) for relative SNP density (top), Dxy (middle) and fixratio (bottom) for (A) the first 15 million base pairs (bp) of chromosome I and (B) from 10 to 22.5 million bp of chromosome III. Results of all windowed analyses are plotted against window midpoints in million bp for windows with >20 SNPs. Genome regions with very low or zero fixation (discussed in text) are labeled by yellow rectangles. Chromosome I exemplifies spatially variable patterns (low differentiation in Italy only), whereas chromosome III exemplifies spatially uniform patterns (low differentiation in all three localities).

Figure 5. Survivorship of seedlings grown in a common garden trial. Symbols represent individuals (circle if alive, cross if dead in year 4). The position of each individual indicates its admixture proportion (*q*, x-axis) and heterospecific (= intersource) ancestry (Q12, y-axis), re-drawn from Lindtke *et al.* (2014). For pure *P. tremula* q = 0, for pure *P. alba* q = 1. Lines indicate the maximum value Q12 can assume for a given *q*. Some data points are overlapping. Colors represent putative genotypic classes, identified graphically based on discontinuities in two-dimensional parameter space.

Supporting information
Additional supporting information accompanies the online version of this article

Legends to Supporting information Figures S1 – S9

Figure S1 Median joining network for plastid DNA haplotypes visualised by the PCR-RFLP method.Haplotype names are congruent with the terminology used by Fussi *et al.* (2010). New haplotypes (newly found in the PCR-RFLP validation dataset) are labeled with an asterisk. Missing haplotypes are represented as black dots along lines connecting haplotypes. Pie sizes are proportional to haplotype frequencies. Haplotypes are colored according to their population origin: *P. alba* Italy, dark blue; *P. alba* Hungary, light blue; *P. tremula* Italy, dark green; *P. tremula* Hungary, light green; *P. x canescens* hybrids Italy, dark purple; *P. x canescens* hybrids Hungary, light purple.

Figure S2 (A) Local genetic ancestry along the 19 chromosomes of hybrids in the Italian hybrid zone, representing homozygous segments for *P. alba* (blue), homozygous segments for *P. tremula* (orange), heterospecific ancestry segments (green) and undefined segments (white). (B) Genomic admixture proportions and species origin of plastid DNA for each individual (*P. alba*, blue; *P. tremula*, orange; no data, white).

Figure S3 (A) Local genetic ancestry along the 19 chromosomes of hybrids in the Hungarian hybrid zone, representing homozygous segments for *P. alba* (blue), homozygous segments for *P. tremula* (orange), heterospecific ancestry segments (green) and undefined segments (white). (B) Genomic admixture proportions and species origin of plastid DNA for each individual (*P. alba*, blue; *P. tremula*, orange; no data, white).

Figure S4 Sizes of genomic segments (x-axis) found in backcrosses to *P. alba* (BCA, blue), backcrosses to *P. tremula* (BCT, orange), a single advanced recombinant hybrid (F_n, green), and 'imperfect F_1' hybrids (gray), plotted against the percentage they cover on each chromosome (y-axis). Segment sizes and percentages were extracted from RASPberry ancestry results.

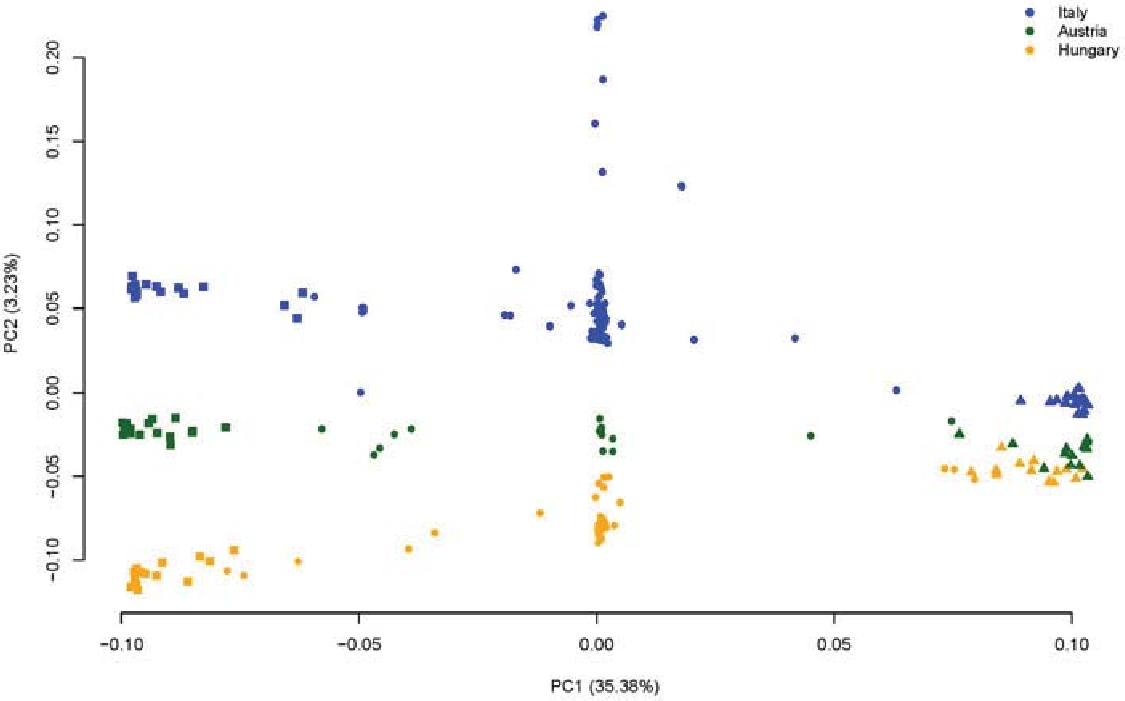
Figure S5 Patterns of SNP density, interspecific genomic divergence (D_{XY}) and differentiation (fixratio) between *Populus alba* and *P. tremula* for Italy (blue), Austria (green) and Hungary (orange). Shown are sliding window analyses (window size, 500 kb; step size, 250 kb) for relative SNP density (top), D_{XY} (middle) and proportion of fixed SNPs among all variable SNPs (fixratio, bottom). Results of all windowed analyses are plotted against window midpoints in million bp for windows with >20 SNPs. Zero fixation windows are highlighted in light yellow, thus facilitating inspection of low-differentiation regions with spatially variable and spatially uniform patterns. Approximate *P. trichocarpa* centromere positions are highlighted in gray.

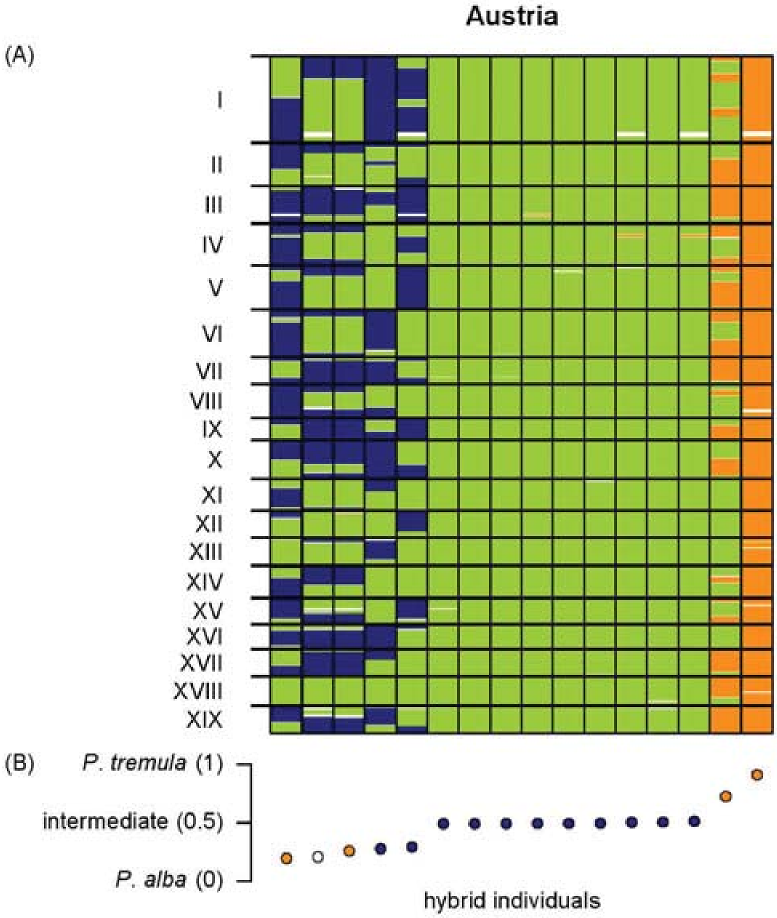
Figure S6 Heat map showing linkage disequilibrium (D') from 17.5 to 21.7 million base pairs (bp) of chromosome III for the parental populations of *P. alba* and *P. tremula* in the Italian, Austrian, and Hungarian hybrid zone, respectively. D' is depicted on a gray scale as indicated by the chart below the heat map.Positions of markers along the reference genome are indicated by black bars along the diagonale. The genome region of interest is framed by a bold black line.

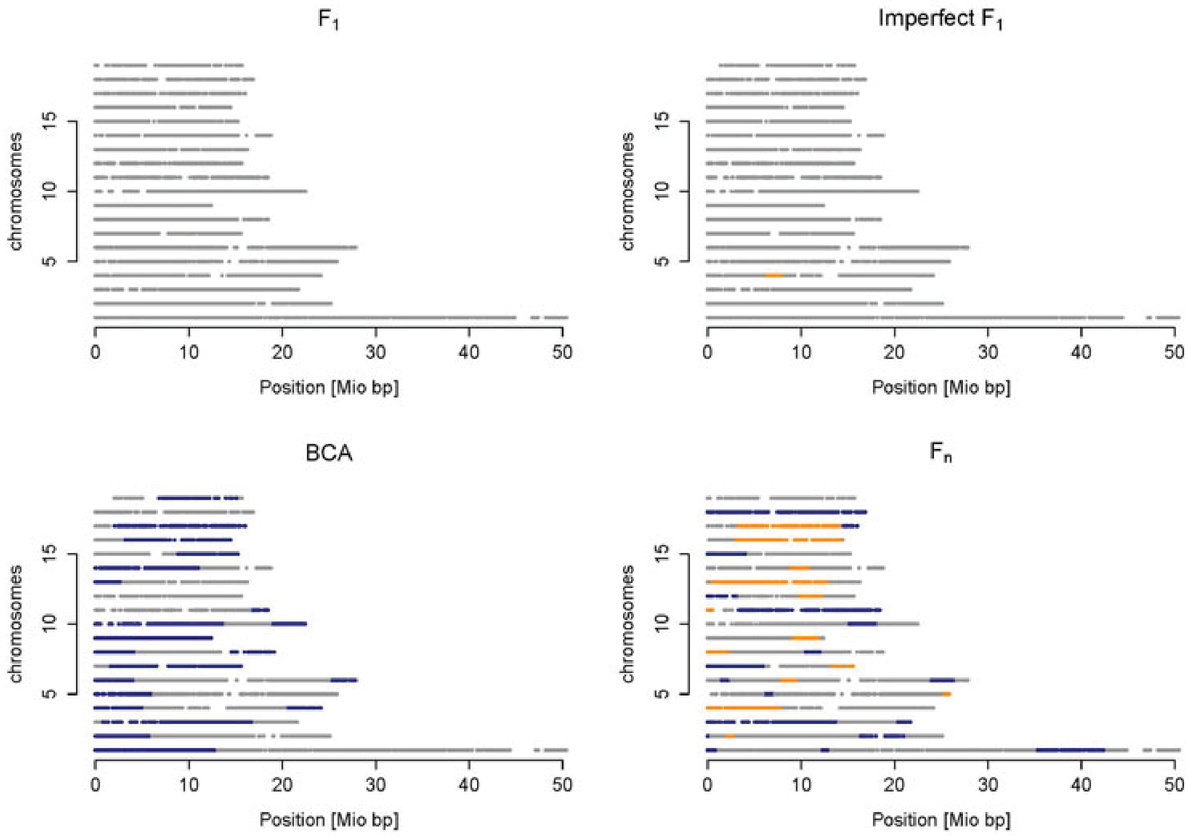
Figure S7 Heat map showing linkage disequilibrium (D') from 2.5 to 7.5 million base pairs (bp) of chromosome I for the parental populations of *P. alba* and *P. tremula* in the Italian, Austrian, and Hungarian hybrid zone, respectively. D' is depicted on a gray scale as indicated by the chart below the heat map.Positions of markers along the reference genome are indicated by black bars along the diagonale. The genome region of interest is framed by a bold black line.

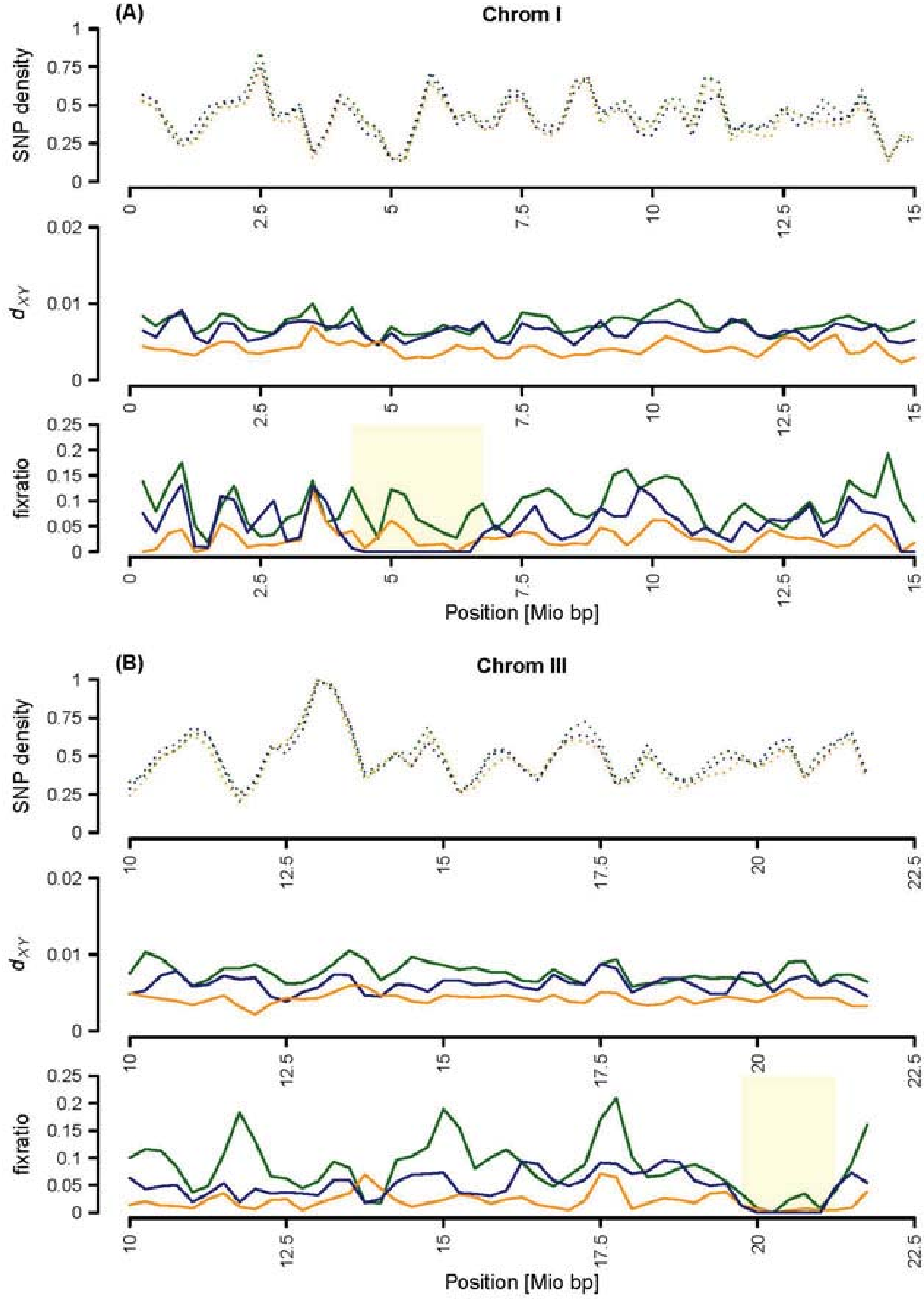
Figure S8 Graphical representation of the linear and quadratic effects of admixture proportions q on seedling survivorship in a common garden trial. Reduced survivorship of individuals with intermediate q (early generation hybrids) and increased survivorship of individuals with large q (*P. alba*-like plants including backcrosses) are visible from the plotted regression line (dashed lines are 95% confidence intervals).

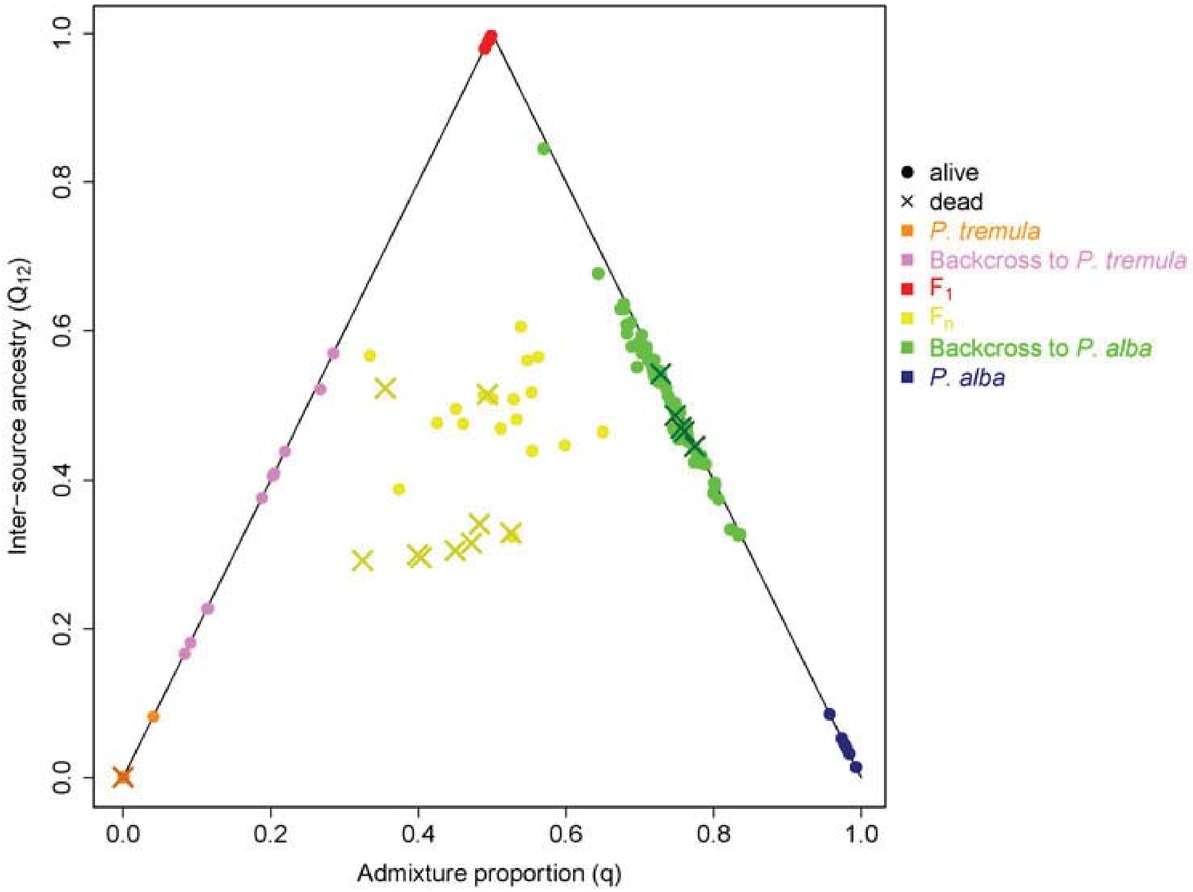
Figure S9 Graphical representation of the linear and quadratic effects of inter-source ancestry Q12 on seedling survivorship in a common garden trial. Reduced survivorship of individuals with intermediate Q12 (recombinant early generation hybrids) and increased survivorship of individuals with maximum Q12 (F_1 hybrids) is visible from the plotted regression line (dashed lines are 95% confidence intervals). See Milne & Abbott (2008) and Lindtke *et al.* (2012) for the expected biologically relevant parameter space of Q12 and related measures of interspecific heterozygosity.


● Italy
● Austria
● Hungary
PC2 (3.23%)
0.20
0.15
0.10
0.05
0.00
−0.05
−0.10
PC1 (35.38%)
−0.10
−0.05
0.00
0.05
0.10


Austria
(A)
I
II
III
IV
V
VI
VII
VIII
IX
X
XI
XII
XIII
XIV
XV
XVI
XVII
XVIII
XIX
(B)
P. tremula (1)
intermediate (0.5)
P. alba (0)
hybrid individuals


F_1
Imperfect F_1
BCA
F_n
chromosomes
5
10
15
Position [Mio bp]
0
10
20
30
40
50


(A)
Chrom I
SNP density
1
0.75
0.5
0.25
0
0
2.5
5
7.5
10
12.5
15
d_{XY}
0.02
0.01
0
fixratio
0.25
0.2
0.15
0.1
0.05
0
Position [Mio bp]
(B)
Chrom III
10
12.5
15
17.5
20
22.5
Position [Mio bp]


Inter-source ancestry (Q_{12})
1.0
0.8
0.6
0.4
0.2
0.0
Admixture proportion (q)
0.0
0.2
0.4
0.6
0.8
1.0
● alive
× dead
P. tremula
Backcross to P. tremula
F_1
F_n
Backcross to P. alba
P. alba