Admixture mapping of quantitative traits in *Populus* hybrid zones: power and limitations

D Lindtke[1], SC González-Martínez[2], D Macaya-Sanz[2,3] and C Lexer[1,4]

Uncovering the genetic architecture of species differences is of central importance for understanding the origin and maintenance of biological diversity. Admixture mapping can be used to identify the number and effect sizes of genes that contribute to the divergence of ecologically important traits, even in taxa that are not amenable to laboratory crosses because of their long generation time or other limitations. Here, we apply admixture mapping to naturally occurring hybrids between two ecologically divergent *Populus* species. We map quantitative trait loci for eight leaf morphological traits using 77 mapped microsatellite markers from all 19 chromosomes of *Populus*. We apply multivariate linear regression analysis allowing the modeling of additive and non-additive gene action and identify several candidate genomic regions associated with leaf morphology using an information-theoretic approach. We perform simulation studies to assess the power and limitations of admixture mapping of quantitative traits in natural hybrid populations for a variety of genetic architectures and modes of gene action. Our results indicate that (1) admixture mapping has considerable power to identify the genetic architecture of species differences if sample sizes and marker densities are sufficiently high, (2) modeling of non-additive gene action can help to elucidate the discrepancy between genotype and phenotype sometimes seen in interspecific hybrids, and (3) the genetic architecture of leaf morphological traits in the studied *Populus* species involves complementary and overdominant gene action, providing the basis for rapid adaptation of these ecologically important forest trees.

INTRODUCTION

Understanding the origin and maintenance of phenotypic variation within and between species is of central importance in studies of adaptation and speciation. Within-species variation has been recognized as an important source for rapid adaptation (Barrett and Schluter, 2008), whereas between-species differences are important for understanding speciation driven through the accumulation of reproductive isolation and increasing biological divergence (Orr, 2001; Barton and Keightley, 2002; Funk *et al.*, 2006). In both cases, elucidating the genetic architecture of phenotypic variation is a critical step for understanding its origin and maintenance.

With the emergence of quantitative trait locus (QTL) mapping, the association of quantitative phenotypic variation with genetic polymorphisms and its applications in human medicine, agricultural genetics and evolutionary biology have received much attention (Tanksley, 1993; Lynch and Walsh, 1998; Mackay, 2001; Barton and Keightley, 2002). QTL mapping utilizes recombinant individuals from intra- or interspecific crosses that are obtained in the laboratory (reviewed by Tanksley, 1993; Orr, 2001; Lexer and Widmer, 2008), and has essentially contributed to advance our understanding of trait evolution, selection and speciation (e.g. Bradshaw *et al.*, 1998; Grandillo *et al.*, 1999; Zeng *et al.*, 2000). Despite the successful identification of QTLs in some organisms, no prevailing conclusions about the genetic architecture of phenotypic variation have been

reached to date because of a shortage of data and experimental limitations (Kearsey and Farquhar, 1998; Orr, 2001; Lexer and Widmer, 2008). In particular, classical QTL mapping studies have two main shortcomings. First, the use of artificial crosses restricts the method to a handful of species with short generation times, a large number of progeny and other characteristics that allow laboratory rearing. Second, mapping is typically carried out on F_2 or BC_1 individuals because of time constraints, which implies that, unless a large number of individuals are investigated, only coarse-scale genetic mapping can be achieved.

Admixture mapping (Chakraborty and Weiss, 1988; Briscoe *et al.*, 1994; McKeigue, 1998) is a promising alternative to classical QTL mapping studies that can overcome both of the above-mentioned limitations by using naturally occurring recombinant individuals from admixed populations or hybrid zones rather than laboratory crosses for gene mapping (reviewed by Rieseberg and Buerkle, 2002; Buerkle and Lexer, 2008). In contrast to association mapping studies that target variation present *within* populations or species, admixture mapping can be used to map phenotypic variation residing *between* divergent taxa (Buerkle and Lexer, 2008). Repeated recombination between parental chromosome blocks in hybrids breaks up the combinations of parental genotypes and thus reduces admixture linkage disequilibrium (LD) between markers. The resulting local chromosomal ancestry blocks are then statistically tested for

[1]Unit of Ecology and Evolution, Department of Biology, University of Fribourg, Fribourg, Switzerland; [2]INIA, Forest Research Centre, Department of Forest Ecology and Genetics, Madrid, Spain; [3]Technical University of Madrid, ETS Forestry Engineering, Department of Silviculture, Madrid, Spain and [4]Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey, UK
Correspondence: Dr D Lindtke, Unit of Ecology and Evolution, Department of Biology, University of Fribourg, Chemin du Musée 10, Fribourg 1700, Switzerland.
E-mail: doro.lindtke@gmail.com

association with phenotypic variation (McKeigue, 2005; Buerkle and Lexer, 2008). This approach is particularly promising as the chromosome block sizes that can be generated in the laboratory in reasonable time by conventional QTL studies (Erickson *et al.*, 2004) cannot keep pace with the recent advancements in marker technology, whereas naturally admixed populations can make the appropriate chromosome block sizes readily available. In addition, the use of natural populations potentially allows characterization of the genetic architecture of adaptation and phenotypic divergence during speciation *in situ* (Coyne and Orr, 2004; Buerkle and Lexer, 2008) and enables the study of these differences under a wide range of environmental conditions.

Despite the known utility of admixture mapping in human populations (Winkler *et al.*, 2010), the great promise it holds for evolutionary studies and the abundance of natural hybrid zones (Rieseberg and Buerkle, 2002), applications of admixture mapping to non-human organisms are extremely rare. Although the few existing studies on annual sunflowers (Rieseberg *et al.*, 1999b), sticklebacks (Malek *et al.*, 2012) and *Lycaeides* butterflies (Gompert *et al.*, 2013) have demonstrated the utility of hybrid zones for gene mapping, more empirical work and simulation studies are now needed to evaluate the power and limitations of admixture mapping in non-human organisms. In this study, we assess the performance of admixture mapping by applying the approach to naturally occurring hybrids between *Populus alba* (white poplar) and *P. tremula* (European aspen), and by performing simulation studies.

Populus alba and *P. tremula* are widely distributed, ecologically and morphologically well-differentiated species that often form large 'mosaic' hybrid zones along European river systems. *P. alba* is mainly restricted to lowland flood–plain forests, whereas *P. tremula* preferentially occurs in mixed upland habitats (Adler *et al.*, 1994). Species can be readily identified by leaf morphological characters (Lexer *et al.*, 2009), which are often associated with biomass accumulation and growth (Rae *et al.*, 2006).

Hybrids between *P. alba* and *P. tremula* fulfill many requirements for admixture mapping in an almost ideal way: they show admixture proportions that are intermediate between the two parental species and are highly recombinant with low but existent levels of LD, which will allow fine-scale mapping (Lexer *et al.*, 2007, 2010; Buerkle and Lexer, 2008); genetic maps among the parental species are thought to be collinear (Pakull *et al.*, 2009; Wang *et al.*, 2011), and the genome sequence of the closely related *P. trichocarpa* (Tuskan *et al.*, 2006) can be used as reference map; background LD in the parental populations is low to moderate (Lexer *et al.*, 2007); the allele frequency differential between the parental populations is high (Lexer *et al.*, 2007, 2010); and at least some of the phenotypic traits of hybrids are intermediate between the parental species (Rieseberg and Buerkle, 2002; Lexer *et al.*, 2009).

In this study, our primary aims were (1) to assess the power and limitations of admixture mapping of quantitative phenotypic trait differences in natural hybrid zones, and (2) to explore how the approach may be used to infer the genetic architecture of species differences as an alternative to laboratory crosses. By applying admixture mapping to both empirical and simulated data, we demonstrate that utilizing naturally occurring hybrids for genetic mapping is a promising approach to investigate the evolution of species differences.

MATERIALS AND METHODS
Sampling, genotyping and genetic classification
Eight hundred and thirty individuals were sampled in 'mosaic' hybrid zones and adjacent parental populations of *P. alba* and *P. tremula* along four different river drainage systems, located in Spain (Duero; $n = 137$), Italy (Ticino; $n = 219$), Austria (Danube; $n = 288$) and Hungary (Tisza; $n = 186$; Supplementary Figure S1). Individuals were genotyped at 77 mapped microsatellite markers distributed across all 19 chromosomes of the *Populus* genome, with linkage information available from the *P. trichocarpa* genome assembly v. 2 (http://www.phytozome.net/poplar) and a controlled backcross to *P. alba* (Supplementary Table S1). Laboratory procedures for DNA extraction, PCR amplification and genotyping followed Lexer *et al.* (2005a) and Lindtke *et al.* (2012), with allele size polymorphisms resolved on a 3100 (Spanish and Austrian samples) or a 3130 (Italian and Hungarian samples) Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). In 10 out of the 77 loci, amplification fragments were not consistent with a simple codominant inheritance pattern and these loci were thus scored conservatively as dominant markers. As genotypes for two microsatellites (G430 and O276) were not available for the Spanish population, 75 loci were used for analyses that combined all four hybrid zone localities.

Separately for each hybrid zone, individuals were assigned to three genotypic classes (*P. alba*, *P. tremula*, admixed) according to admixture proportions (*Q*) obtained by *structure* 2.3.2.1 (Pritchard *et al.*, 2000), which classified 179 out of all 715 phenotyped individuals (see below) as potentially admixed ($0.05 \leqslant Q \leqslant 0.95$). As most of these samples are from the Italian hybrid zone ($n = 109$), we have used the Italian population as our primary mapping population, and the pooled data set consisting of all four hybrid zone localities ($n = 179$) for complementary analyses only. Various checks were performed to verify the utility and reliability of the 77 microsatellite markers for mapping, including allele frequency differential (δ; Zhu *et al.*, 2005), missing genotypes and correlations between genome-wide individual ancestry and locus-specific ancestry (LSA).

Phenotypic trait measurements
To apply admixture mapping to ecologically important species differences, leaf morphological traits were chosen as model traits, as they differ considerably between *P. alba* and *P. tremula* (Lexer *et al.*, 2009), are associated with biomass accumulation and growth, and have been used in previous QTL mapping studies of a related species pair (e.g. Rae *et al.*, 2006). We obtained phenotypic data for 13, 68, 20 and 55 *P. tremula*; 97, 41, 149 and 93 *P. alba*; as well as 15, 109, 19 and 36 admixed individuals in Spain, Italy, Austria and Hungary, respectively (715 individuals altogether; herbarium specimens suitable for measurements were not available for all 830 genotyped individuals), with species classified according to *structure Q* (above). Important aspects of leaf blade and petiole morphology were assessed by taking 11 raw measurements, complemented by 11 additional variables and aspect ratios that were computed from those values (Table 1 and Supplementary Table S2), using measurements from five leaves averaged per tree. All measurements and calculations were standardized, but not transformed to normality, as this is not necessarily

Table 1 Traits mapped and measurements/calculations taken

Trait	Description	Measurement/calculation
PETL	Petiole length	Raw measurement (mm)
LFAREA	Leaf area	Model based[a] (mm²)
PTRAT	Petiole ratio	Petiole length/lamina length
PETXM	Petiole flatness at petiole middle	Petiole cross-section length/width[b] at petiole middle
PETXB	Petiole flatness at petiole base	Petiole cross-section length/width[b] at petiole base
PETXA	Petiole flatness at petiole apex	Petiole cross-section length/width[b] at petiole apex
LFSHAP	Leaf shape	Lamina length/lamina width
LDRAT	Lobe depth	(Lobe width − sinus width)/lobe width

For additional traits see Supplementary Table S2.
[a]Leaf area = 28.5966 + 0.6168 × lamina length^0.8326 × lamina width^1.2065; best fitting allometric model based on 1288 observations from Spanish hybrid zone.
[b]The raw measurements of the petiole cross-sections were taken in a way that length is always the larger value of length and width.

2

correct in the presence of a QTL effect (Churchill and Doerge, 1994). To reduce the set of 22 mapping traits to a meaningful subset, a factor analysis using the promax rotation was carried out with the *factanal* function in R (R Development Core Team, 2011).

Trait mapping

QTLs were mapped by using (1) available admixture mapping software, and (2) linear regression analysis in R to overcome current limitations of existing software (see below).

ADMIXMAP software. The admixture mapping software *ADMIXMAP* (Hoggart *et al.*, 2004; http://homepages.ed.ac.uk/pmckeigu/admixmap/) is suitable for QTL mapping using microsatellite markers. The software applies a hidden Markov chain Monte Carlo model in combination with regression tests (Winkler *et al.*, 2010), and uses an algorithm similar to the *structure* software to model admixture and LSA, but cannot handle dominant markers. Thus, *ADMIXMAP* was run for 67 codominant markers only. We did not pool hybrid zones and considered the Italian mapping population only when using *ADMIXMAP*, as microsatellite fragment sizes (which we suspect to differ between localities because of processes unrelated to phenotype) are used as direct input.

Ancestry association tests were carried out separately for each trait, using the original allele counts (after addition of 0.5) from the parental reference populations to specify the prior distributions of ancestral allele frequencies. Samples were taken from 50 000 iterations with a thinning interval of 10 after removing 500 iterations as burn-in. During the burn-in, 10 annealing runs were performed. Default priors were applied except for the average number of generations back to the un-admixed ancestral population, where we used a gamma distribution with shape and rate parameters of 15 and 0.8, respectively (corresponding roughly to a mean of 19 and a variance of 22), to reflect the highly recombinant nature of the investigated hybrids. The performance of the sampler was evaluated through visual inspection of diagnostic graphs generated by the software.

Linear regression analysis. Although *ADMIXMAP* has the advantage that it computes LSA and performs association tests within a single step, it is not suitable for dominant markers, cannot deal with null alleles (non-amplifying alleles) or multiple populations, and can only perform single-locus regression analysis. Thus, in our second approach, we first estimated LSA and subsequently modeled the phenotype–genotype association in R (R Development Core Team, 2011) by applying generalized linear models. Separately for each hybrid zone locality, the software *structure* 2.3.2.1 (Pritchard *et al.*, 2000; Falush *et al.*, 2003, 2007) was used to estimate LSA for all 77 markers (the site-by-site output), using the recessive alleles option to account for the potential presence of null alleles, following steps and settings as in Lindtke *et al.* (2012). Using LSA rather than raw microsatellite fragment sizes as input for subsequent analyses will remove locality-specific differences unrelated to chromosome-block ancestry.

Structure. LSAs provide probability estimates for the ancestry of the observed genotype per locus and individual, that is, the probability that both alleles originate from *P. alba* (p_{aa}), both alleles originate from *P. tremula* (p_{tt}) or one allele is derived from each parent (p_{at} and p_{ta}). These values are calculated for each individual i from $\{1,\dots, N\}$ and each locus l from $\{1,\dots, L\}$, where $p_{aa}^{(i,l)} + p_{at}^{(i,l)} + p_{ta}^{(i,l)} + p_{tt}^{(i,l)} = 1$. Because our data are not phased ($p_{at}^{(i,l)} \approx p_{ta}^{(i,l)}$), knowing $p_{aa}^{(i,l)}$ and $p_{at}^{(i,l)}$ is already sufficient to describe the *locus-specific* individual admixture proportion as

$$S^{(i,l)} = p_{aa}^{(i,l)} + p_{at}^{(i,l)} \tag{1}$$

with $S^{(i,l)}$ ranging from 0 to 1 (from pure *P. tremula* to pure *P. alba* ancestry), analogous to the *genome-wide structure* Q.

The *genome-wide* individual admixture proportion can be calculated by averaging over the locus-specific estimates as

$$G^{(i)} = \overline{p_{aa}^{(i)}} + \overline{p_{at}^{(i)}} = \frac{1}{L}\sum_{l=1}^{L} p_{aa}^{(i,l)} + \frac{1}{L}\sum_{l=1}^{L} p_{at}^{(i,l)} \tag{2}$$

which is highly correlated to the *structure* admixture proportion Q (Italian data set: $r = 0.9999975$, P-value $< 2.2e^{-16}$).

We modeled a *single-locus* association between the phenotype y of individual i for each locus l by using generalized linear models as

$$y^{(i)} = \mu^{(l)} + \alpha^{(l)}G^{(i)} + \beta^{(l)}S^{(i,l)} + \varepsilon^{(i,l)} \tag{3}$$

where $\mu^{(l)}$ is the intercept, $\varepsilon^{(i,l)}$ the error term (Gaussian), $S^{(i,l)}$ is the locus-specific admixture estimate and $G^{(i)}$ provides the adjustment of the model for the confounding effect of genome-wide admixture (e.g. Hoggart *et al.*, 2003).

However, gene action is often not strictly additive, but may range continuously from dominance to recessiveness, or from over- to under-dominance (particularly in hybrids, overdominant gene action might be common; Tanksley, 1993; Charlesworth and Willis, 2009). In quantitative genetics, a dominance deviation term can be used to describe the degree to which the heterozygous genotype deviates from a phenotype expression that is exactly intermediate between the two homozygotes (Tanksley, 1993). Thus, we added the covariates $G'^{(i)} = \overline{p_{at}^{(i)}}$ and $S'^{(i,l)} = p_{at}^{(i,l)}$ to equation (3) to account for the potential non-additivity of the genome-wide and the locus-specific component, respectively, and modeled y as

$$y^{(i)} = \mu^{(l)} + \alpha_1^{(l)}G^{(i)} + \alpha_2^{(l)}G'^{(i)} + \beta_1^{(l)}S^{(i,l)} + \beta_2^{(l)}S'^{(i,l)} + \varepsilon^{(i,l)} \tag{4}$$

The ratio of the parameter estimates for each locus, $\beta_2^{(l)}/\beta_1^{(l)}$, can be used analogous to the dominance/additivity statistics in QTL mapping (Tanksley, 1993) to describe the effects of *P. alba* alleles as strictly additive for $\beta_2/\beta_1 = 0$, as dominant for $\beta_2/\beta_1 = 1$, as recessive for $\beta_2/\beta_1 = -1$, as overdominant for $\beta_2/\beta_1 > 1$ or as underdominant for $\beta_2/\beta_1 < -1$ (note that in case $\beta_1 < 0$, overdominant gene action will be defined as a decrease in the trait value for heterozygotes, whereas underdominance will indicate an increase in the trait value—the reversals of the definitions that hold for $\beta_1 > 0$). The same applies to the genome-wide parameters $\alpha_1^{(l)}$ and $\alpha_2^{(l)}$.

To assess if a focal locus explained a relevant proportion of the phenotype, an information theoretic approach was used to compare the model that includes the locus-specific effect to a model that contains the genome-wide effect only. In particular, we compared equation (3) to

$$y^{(i)} = \mu + \alpha G^{(i)} + \varepsilon^{(i)} \tag{5}$$

and equation (4) to

$$y^{(i)} = \mu + \alpha_1 G^{(i)} + \alpha_2 G'^{(i)} + \varepsilon^{(i)} \tag{6}$$

by calculating the Akaike information criterion (AIC; Akaike, 1973) using the *aictab* function (AICcmodavg package) in R. A locus was kept as a candidate QTL if $AIC_{model\ (5)} - AIC_{model\ (3)} \geqslant 4$ (additive model) or $AIC_{model\ (6)} - AIC_{model\ (4)} \geqslant 4$ (dominant/recessive model), denoting that explanation of the phenotype was considerably improved by incorporating a locus-specific effect in the model.

A common problem in multiple regression analysis is collinearity arising from high correlation between explanatory variables. LSA estimates for loci with low information content will mainly reflect their Markov chain Monte Carlo priors that are given by genome-wide admixture proportions, and will consequently show high correlation between locus-specific and genome-wide effects. Five markers (O30_1, G1416, G430, O14 and O206) with small allele frequency differential ($\delta < 0.1$; with $\delta = \sum |f_{i1} - f_{i2}|/2$, where f_{i1} and f_{i2} represent the ith allele frequencies in the two parental populations, respectively; Zhu *et al.*, 2005), and consequently high correlation between locus-specific and genome-wide admixture proportions ($r > 0.98$) were therefore excluded from the marker panel for all regression analyses.

To account for more complex genetic architectures of phenotypic traits controlled by multiple QTLs, and to reduce false-positive results due to ghost peaks, the identified candidate QTLs were subjected to a *stepwise model selection* approach. Forward selection was followed by backward elimination using the *stepAIC* function (MASS package), with the scope of the searched model space being the genome-wide effects ($G^{(i)}$ and $G'^{(i)}$) and candidate QTLs for both additive and dominant-recessive components ($S^{(i,l)}$ and $S'^{(i,l)}$). The Bayesian information criterion (BIC; Schwarz, 1978) was used for model selection, as it tends to favor more parsimonious models (Grueber *et al.*, 2011).

Forward selection followed by backward elimination has been shown to perform well for QTL mapping (Broman and Speed, 2002).

Although stepwise model selection is computationally fast, it cannot make use of the full power associated with the information theoretic approach, as it compares only two models at a time and ignores model uncertainty (Whittingham *et al.*, 2006). Therefore, we additionally used an *exhaustive search* over all possible models with the same scope as above, using the BIC (*dredge* and *model.avg* functions in MuMIn package). The set of best models with $\Delta BIC < 4$ from the top model was kept and full-model averaged parameter estimates were obtained. The coefficient of determination R^2 was computed to estimate the contribution of each marker (additive and dominant/recessive components) in explaining phenotype by controlling for all other variables included in the model, and to estimate the proportion of variance explained by all markers kept in the full model by controlling for genome-wide admixture (*r.squaredLR* function in MuMIn package).

We applied the linear regression analysis to both the Italian hybrid zone samples and the pooled data set. To combine samples, we used residual trait values obtained by a one-way analysis of variance, with locality as cofactor.

Simulations

Simulations were performed to assess the power of admixture mapping using *structure* LSA estimates and linear regression analysis. The performance of *ADMIXMAP* has been studied elsewhere (Hoggart *et al.*, 2003, 2004).

Admixed populations were simulated using the software *quantiNemo* 1.0.4 (Neuenschwander *et al.*, 2008), and were designed to match the real data as in Lindtke *et al.* (2012). Briefly, 77 microsatellite markers on 19 chromosomes were generated by using the map positions and original allele frequencies from *P. alba* and *P. tremula* calculated from the Italian population. Four QTLs were inserted into this genetic map and subsequently used to simulate quantitative trait values. Ten replicate admixed populations, consisting of different hybrid generations, backcrosses and immigrants, were sampled at generations 7, 12 and 22. Eight different genetic architectures for quantitative traits were modeled (Supplementary Table S3), with a proportion of phenotypic variation explained by genotype (total genetic effect) of 0.2 or 0.5, and sample sizes of 100 or 500 individuals. Candidate QTLs were identified and subjected to the stepwise model selection procedure as described above. Power (proportion of successfully detected QTL-linked markers) and false-positive rate (FPR; number of false positives divided by the number of final candidate QTLs) were determined for each generation, genetic architecture, genetic effect and

replicate. The full description of the simulation approach is available in Supplementary Information.

Linkage disequilibrium

The size of ancestral chromosome blocks generated by interpopulation recombination is critical for admixture mapping of phenotypic trait differences (Buerkle and Lexer, 2008) and is reflected by the extent of LD spanning along ancestral chromosome blocks. We therefore estimated correlations in LSA between adjacent genetic regions to evaluate the amount of non-random associations of ancestral alleles along chromosomes. Pairwise correlations between all locus pairs were calculated by resampling genotypes for each locus (0, 1 or 2 alleles originating from *P. alba*), with sampling probabilities defined by *structure* LSA estimates (p_{tt}, $p_{at} + p_{ta}$ or p_{aa}). This resampling approach allowed us to collapse LSA probabilities into discrete genotypes. Residuals from a binomial generalized linear model with Q as explanatory variable were calculated from the resampled genotypes to control for potential population structure due to different levels of admixture. After 500 replications, the median correlation coefficients between residuals were recorded for the Italian population, and two simulated data sets (generation 22 and 32; 109 randomly chosen individuals).

RESULTS
Phenotypic traits in hybrid zones

Phenotypic measurements in four hybrid zone localities, ranging from Southwestern to Central Europe (Supplementary Figure S1), yielded information on patterns of variation for 22 leaf morphological traits (Table 1 and Supplementary Table S2) in two *Populus* species and their hybrids. Factor analysis revealed that most variation within the 22 traits was explainable by 13 factors, with the first seven factors showing eigenvalues >1 (Supplementary Table S4). One representative trait for each of these first seven factors (absolute value of the loading >0.995) was chosen to be included in further analyses. This reduced the set of mapping traits to LFAREA, PTRAT, LFSHAP, PETXB, PETXM, PETXA and LDRAT. In addition, PETL was included in the set of mapping traits as it has already been examined in a previous QTL study (Rae *et al.*, 2006). Thus, eight leaf blade and petiole traits were included in the final analysis, of which we have
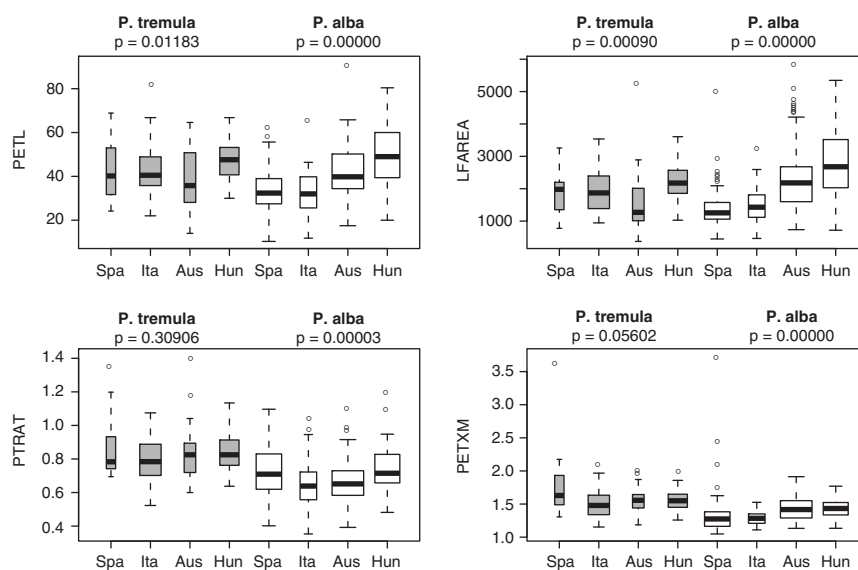

Figure 1 Boxplots showing untransformed phenotypic measurements for *P. tremula* (gray) and *P. alba* (white) at all four localities (Spa, Spain; Ita, Italy; Aus, Austria; Hun, Hungary). Indicated are *P*-values of Kruskal–Wallis rank-sum tests for the null hypothesis that within a species, all four localities have the same median. Four representative traits are shown. For additional traits see Supplementary Figure S2.

chosen PETL, LFAREA, PTRAT and PETXM as representative traits to be shown in more detail in the main article.

Considerable intraspecific phenotypic variation was identified for all eight mapping traits (Figure 1 and Supplementary Figure S2). The

four examined localities differed significantly in their median trait values for all traits in *P. alba*, and most traits in *P. tremula* (Figure 1 and Supplementary Figure S2). *P. alba* showed a clear southwest–northeast gradient in PETL and LFAREA, with the north–easternmost population having the largest values. Interspecific phenotypic differentiation between *P. alba* and *P. tremula* showed different magnitudes dependent on the trait studied. Owing to the southwest–northeast variation present in *P. alba*, this species exhibited shorter petioles and smaller leaves than *P. tremula* in Spain and Italy, but the opposite pattern in Austria and Hungary. PETL, LFAREA and PTRAT were highly correlated among admixed individuals for the pooled and the Italian data set (Spearman's rank correlation coefficients ρ for the correlations PETL–LFAREA, PETL–PTRAT and LFAREA–PTRAT were 0.85, 0.76 and 0.42 for pooled data, and 0.86, 0.81 and 0.50 for Italy, respectively; all *P*-values were $<4e^{-8}$; all other correlations had absolute values of $\rho < 0.45$).

For the Italian hybrid zone (our primary mapping population), hybrids showed trait values that were either intermediate between the parental species or similar to one or the other parent (Table 2, Figure 2a and Supplementary Figure S3a). The parental populations were significantly differentiated for most of the studied traits, and admixed individuals showed a non-random relationship between genome-wide *structure Q* and phenotype for all traits except LDRAT (*P*-values of linear regression slopes <0.1), which suggests a heritable component for these traits (Table 2).
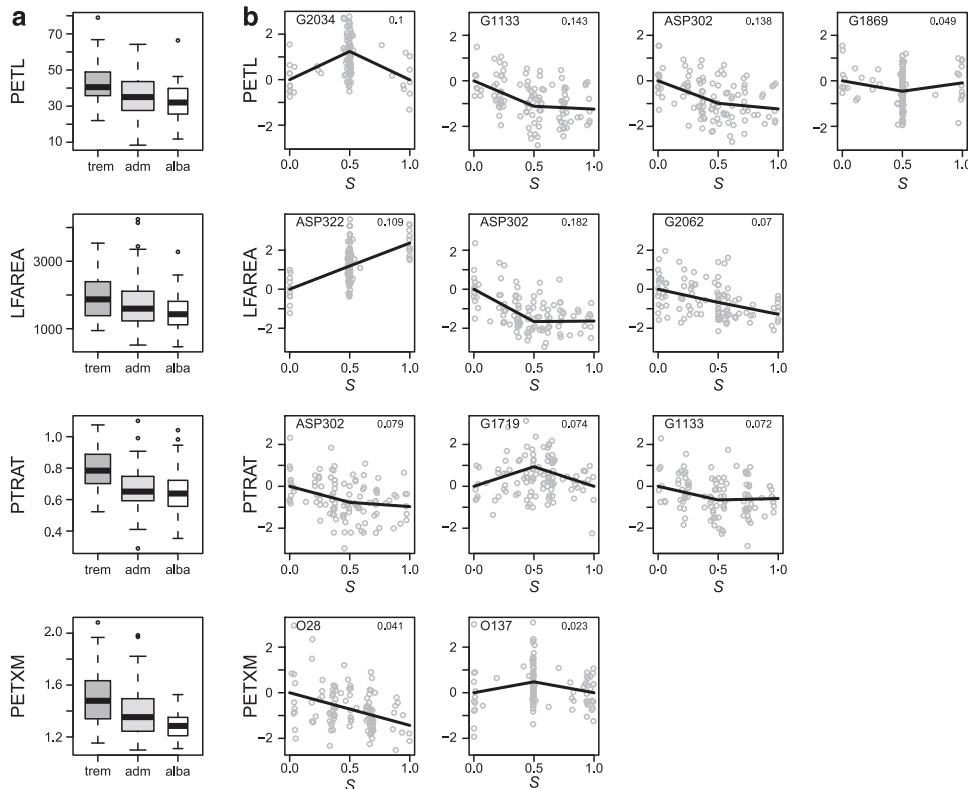
Table 2 Summary statistics of phenotypic traits for the Italian hybrid zone

Trait	P. tremula mean (s.d.)	Admixed mean (s.d.)	P. alba mean (s.d.)	P-value Wilcoxon test[a]	P-value trait~Q for admixed[b]
PETL	42.32 (10.28)	35.50 (10.78)	32.01 (9.89)	**0.000**	**0.018**
LFAREA	1940 (633)	1754 (743)	1476 (548)	**0.000**	0.062
PTRAT	0.796 (0.127)	0.665 (0.128)	0.659 (0.159)	**0.000**	**0.002**
PETXM	1.498 (0.225)	1.376 (0.175)	1.290 (0.103)	**0.000**	0.080
PETXB	1.407 (0.206)	1.419 (0.184)	1.303 (0.159)	**0.006**	0.075
PETXA	1.326 (0.175)	1.289 (0.118)	1.339 (0.156)	0.493	**0.018**
LFSHAP	1.098 (0.152)	1.205 (0.166)	1.194 (0.167)	**0.003**	**0.022**
LDRAT	0.189 (0.062)	0.175 (0.060)	0.197 (0.044)	0.160	0.999

Abbreviations: LDRAT, lobe depth; LFAREA, leaf area; LFSHAP, leaf shape; PETL, petiole length; PETXA, petiole flatness at petiole apex; PETXB, petiole flatness at petiole base; PETXM, petiole flatness at petiole middle; PTRAT, petiole ratio; s.d., standard deviation.
P-values <0.05 are marked in bold.
[a]*P*-values for Mann–Whitney–Wilcoxon test for the null hypothesis that trait values for parental populations are from the same distribution.
[b]*P*-values for the null hypothesis that linear regression slopes for phenotypic traits of admixed individuals as a function of *structure Q* are zero.

Figure 2 Raw phenotypes and locus-specific associations for four representative traits, Italian hybrid zone. (**a**) Boxplots of untransformed measurements for *P. tremula* (trem), admixed individuals (adm) and *P. alba* (alba), boxplot widths being proportional to sample size. (**b**) Genotype–phenotype associations of admixed individuals for each candidate locus with relative importance >0.5. Residual phenotypes (y axis) were calculated by using full-model averaged parameter estimates (Coef; Table 3), including all variables except the focal marker in the model, and are plotted against locus-specific admixture proportions *S* of the focal marker (x axis, with 0.0 representing *P. tremula* and 1.0 representing *P.alba* ancestry; marker names indicated at top left of each scatterplot). Regression lines were computed from Coef (additive and dominant/recessive components) of the focal marker. R^2 for focal markers are indicated at the top right of each scatterplot. For additional traits see Supplementary Figure S3.

Genetic architecture of phenotypic traits

Italian hybrid zone. By conducting single-locus tests for genotype–phenotype associations using an information theoretic approach, several candidate QTLs for all eight traits could be identified (Figure 3, Supplementary Figure S4 and Supplementary Table S5). Correlated traits such as PETL and LFAREA also showed overlap in their set of candidate loci. By incorporating dominant/recessive terms into the model, additional candidate loci were identified, most of which showed strong evidence for over- or underdominant gene action in subsequent multivariate regression analysis (Table 3). The software *ADMIXMAP* tended to identify a similar set of candidate QTLs as the single-locus additive model (Supplementary Table S5 and Supplementary Figure S5).

An exhaustive search algorithm and model selection were carried out to build the final multivariate regression model and to obtain full-model averaged parameter estimates. Only a subset of candidate QTLs showed a relative importance of >0.5 (Table 3). Full-model averaged parameter estimates are informative regarding the direction of QTL effects (negative signs indicate reduction of trait values by *P. alba* alleles), effect size (e.g. the parameter estimate can be transformed into QTL effect expressed in percent species difference) and gene action (Table 3). QTL effects were sometimes biased towards a single direction, as for PETL (of nine coefficients, eight were

negative), PETXM (of nine coefficients, seven were negative) and LFSHAP (all four coefficients were positive). Effect sizes in terms of percent species differences showed a wide range of magnitudes. For example, the final model for PETXM incorporated many QTLs of small effect, whereas large but complementary QTL effects were found for LFAREA.

Gene action was often not strictly additive, and QTLs showed evidence for over- or underdominance in many instances (Table 3, Figure 2b and Supplementary Figure S3b). Note that estimates for gene action may involve a large error, particularly if β_1 is close to zero. The estimates should thus be seen as a rough approximation, although the overall trend for gene action should be robust. In cases where the same candidate locus was identified for several traits, gene action was often consistent across traits (i.e. dominance of ASP302 for PETL, LFAREA and PTRAT, and of G1133 for PETL and PTRAT; Figure 2b and Table 3).

After controlling for the residual genome-wide effect, the final models explained 37.9%, 38.7%, 27.3% and 28.5% of phenotypic variation for PETL, LFAREA, PTRAT and PETXM, respectively (13.5–24.6% for the other four traits; Table 3). A single QTL explained on average 8.6% (range 2.3–18.2%) of phenotypic variation (only loci with relative importance >0.5, joint additive and recessive/dominant components, all other parameters in the model controlled for; Table 3).
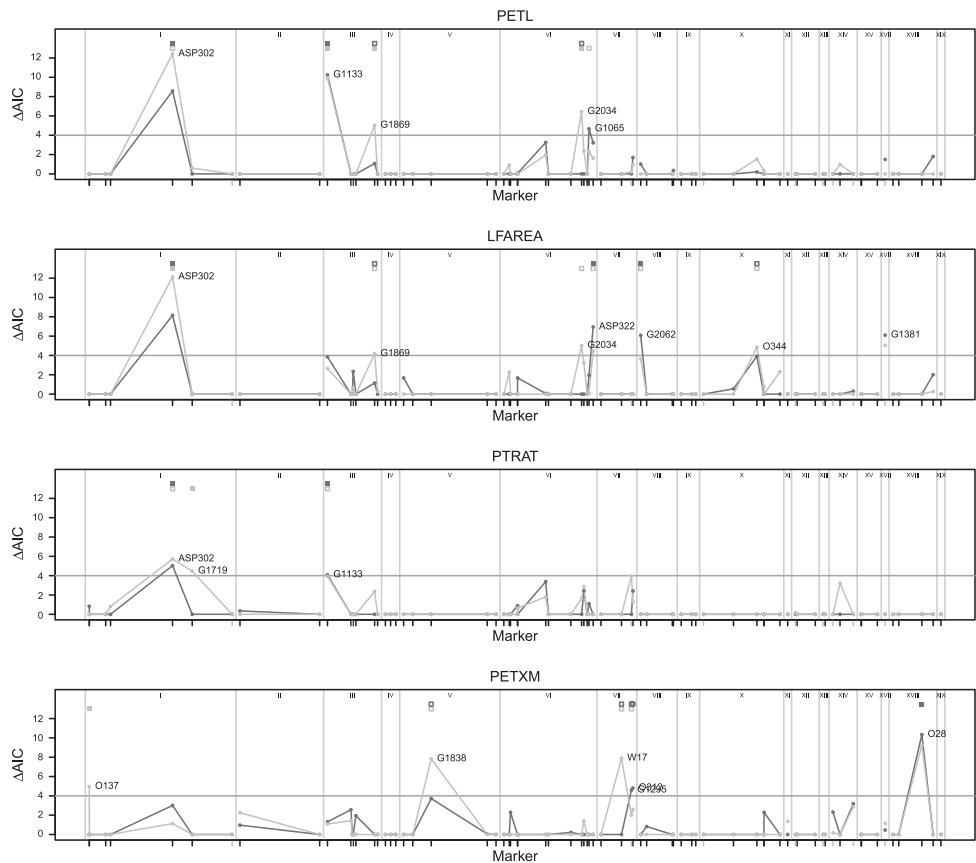

Figure 3 Genotype–phenotype associations of four representative traits mapped in the Italian hybrid zone. Depicted are ΔAIC from single-marker regression analyses, with values <0 truncated to zero. Dark gray lines, additive model; light gray lines, dominant/recessive model. Marker loci with ΔAIC values ⩾4 (marked by horizontal line) are labeled. Squares indicate markers that were included in final multivariate regression models (Table 3); dark gray, additive component; light gray, dominant/recessive component; filled squares, relative importance >0.5. Tick marks on x axis indicate markers according to their relative map position along chromosomes I to XIX (separated by gray vertical lines); black tick marks, codominant markers; gray tick marks, dominant markers. For additional traits see Supplementary Figure S4.

Table 3 Final models of multivariate regression analysis, exhaustive search, for the Italian hybrid zone

Trait	Marker	Chr.	Position (bp)	δ	Additive components				Dominant/recessive components				Gene action[a]	R² combined[b]
					Imp[c]	Coef[d] (β_1)	Effect[e]	R²[f]	Imp[c]	Coef[d] (β_2)	Effect[e]	R²[f]		
PETL	**ASP302**	1	30 444 558	0.58	**0.91**	−1.24	−129.19	0.114	0.47	−0.74	−77.17	0.031	0.60	**0.138**
	G1133	3	4 716 293	0.81	**0.93**	−1.24	−130.11	0.089	**0.51**	−0.99	−103.48	0.033	0.80	**0.143**
	G1869	3	16 805 774	1.00	0.09	−0.09	−9.31	0.014	**0.60**	−0.83	−86.47	0.028	9.28	**0.049**
	G2034	6	22 219 025	0.96	0.01	−0.01	−0.96	0.018	**1.00**	2.48	259.50	0.087	−271.04	**0.100**
	G1065	6	24 119 014	0.82					0.01	−0.01	−1.13	0.014	−Inf	0.014
	Q				**0.52**	0.96	100.01	0.066	0.04	−0.13	−13.74	0.004	−0.14	**0.076**
	Total						−169.57				−22.49			0.379[g]
LFAREA	**ASP302**	1	30 444 558	0.58	**1.00**	−1.63	−261.44	0.128	**0.88**	−1.69	−269.99	0.072	1.03	**0.182**
	G1869	3	16 805 774	1.00	0.25	−0.28	−45.14	0.028	0.04	−0.05	−7.64	0.010	0.17	0.045
	G2034	6	22 219 025	0.96					0.28	0.36	58.40	0.002	Inf	0.002
	ASP322	6	25 184 620	1.00	**1.00**	2.37	378.83	0.058	0.03	0.01	2.38	0.013	0.01	**0.109**
	G2062	8	5 051 212	0.86	**0.93**	−1.28	−205.52	0.027	0.07	−0.07	−10.70	0.011	0.05	**0.070**
	O344	10	14 738 667	0.93	0.13	0.11	17.71	0.037	0.17	0.18	29.21	0.022	1.65	0.040
	Q								0.16	0.39	62.35	0.011	Inf	0.011
	Total						−115.56				−135.98			0.387[g]
PTRAT	**ASP302**	1	30 444 558	0.58	**0.80**	−0.97	−90.31	0.056	0.37	−0.56	−52.53	0.025	0.58	**0.079**
	G1719	1	35 488 312	0.56					**0.79**	1.87	174.56	0.074	Inf	**0.074**
	G1133	3	4 716 293	0.81	**0.56**	−0.58	−54.43	0.034	0.40	−0.72	−67.00	0.025	1.23	**0.072**
	Q								0.08	−0.19	−17.30	0.000	-Inf	0.000
	Total						−144.74				37.73			0.273[g]
PETXM	**O137**	1	9 087 453	0.98					**0.78**	0.95	79.73	0.023	Inf	**0.023**
	G1838	5	8 802 231	0.81	0.28	−0.32	−27.05	0.031	0.22	−0.29	−24.63	0.012	0.91	0.039
	W17	7	8 696 038	0.53	0.05	−0.06	−5.25	0.024	0.22	−0.47	−39.40	0.037	7.50	0.039
	G1295*	7	11 243 952	0.65	0.25	−0.28	−23.50	0.013	0.25	−0.59	−49.57	0.000	2.11	0.037
	O312	7	11 625 195	0.89	0.40	0.45	37.42	0.010					0.00	0.010
	O28	18	11 993 250	0.49	**0.91**	−1.43	−119.66	0.041					0.00	**0.041**
	Q				0.31	0.67	56.35	0.044	0.05	−0.13	−11.00	0.012	−0.20	0.061
	Total						−81.69				−44.87			0.285[g]
PETXB	**O268**	8	13 427 006	0.83	**0.51**	0.74	130.94	0.049	0.07	−0.07	−13.10	0.008	−0.10	**0.054**
	G1949	9	1 444 490	0.34	0.10	−0.08	−14.30	0.002	0.44	−1.54	−273.11	0.052	19.10	0.058
	Q				**0.62**	−1.24	−220.54	0.075	0.22	0.72	127.77	0.040	−0.58	**0.089**
	Total						−103.90				−158.43			0.135[g]
PETXA	G1838	5	8 802 231	0.81	0.04	−0.01	−14.26	0.005	**0.53**	−0.74	−707.29	0.059	49.59	**0.062**
	G1831	6	3 671 141	0.72	**1.00**	1.38	1324.57	0.103	0.31	−0.37	−358.20	0.040	−0.27	**0.123**
	Q								0.13	0.30	292.52	0.043	Inf	0.043
	Total						1310.31				−772.97			0.178[g]
LFSHAP	**O220**	4	7 778 968	0.43	**0.79**	1.03	177.60	0.059					0.00	**0.059**
	O149	10	16 581 540	1.00					**0.89**	1.30	223.40	0.064	Inf	**0.064**
	G1381*	17	6 100 168	0.77	0.42	0.45	77.83	0.017	0.34	0.80	137.79	0.001	1.77	0.055
	Q				0.05	−0.07	−12.51	0.012					0.00	0.012
	Total						242.92				361.19			0.222[g]
LDRAT	**O167**	6	5 821 040	0.96	0.24	−0.26	−182.58	0.013	**0.64**	1.07	759.63	0.068	−4.16	**0.098**
	W12	6	19 471 676	0.92	**0.81**	−0.98	−693.38	0.022	0.33	0.63	443.18	0.041	−0.64	**0.084**
	G1894	15	809 326	0.97	**0.95**	1.61	1137.83	0.085	0.39	0.69	488.50	0.063	0.43	**0.105**
	Q				0.11	−0.27	−190.36	0.004	**0.77**	−4.00	−2835.28	0.140	14.89	**0.151**
	Total						71.51				−1143.97			0.246[g]

Abbreviations: Chr, chromosome; δ, allele frequency differential; LDRAT, lobe depth; LFAREA, leaf area; LFSHAP, leaf shape; Q, genome-wide effects G and G'; QTL, quantitative trait locus; PETL, petiole length; PETXA, petiole flatness at petiole apex; PETXB, petiole flatness at petiole base; PETXM, petiole flatness at petiole middle; PTRAT, petiole ratio.

Markers with a relative importance >0.5 are marked in bold and loci scored as dominant markers are indicated with an asterisk.

[a]Gene action of *P. alba* alleles, β_2/β_1, where values of 0 denote additivity, −1 recessiveness, 1 dominance, < −1 underdominance and >1 overdominance.

[b]R^2 for each marker (additive and dominant/recessive component), adjusted for all other parameters in the full model (*r.squaredLR* function MuMIn package).

[c]Relative importance of the predictor variables, calculated as a sum of the Akaike weights over all of the best models in which the parameter of interest appears (*model.avg* function MuMIn package).

[d]Full-model averaged parameter estimates (coefficient is zero if not included in model; *model.avg* function MuMIn package). Positive or negative signs indicate an increase or decrease of the trait value by *P. alba* alleles.

[e]QTL effect in percent species difference. *Total* indicates the sum over individual effects maintained in the full model. Positive or negative signs indicate an increase or decrease of the trait value by *P. alba* alleles.

[f]R^2 for each parameter, adjusted for all other parameters in the full model (*r.squaredLR* function MuMIn package).

[g]Total R^2 of the full model including all markers, adjusted for Q (*r.squaredLR* function MuMIn package).
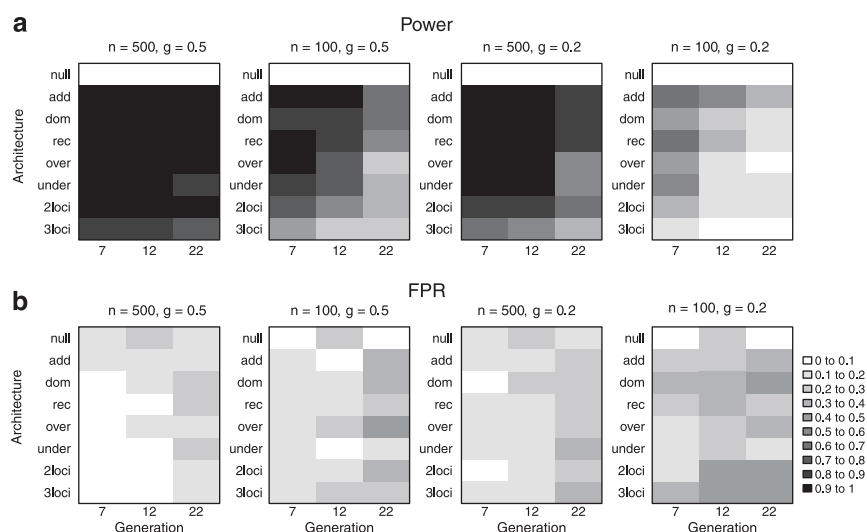
Figure 4 (**a**) Power and (**b**) FPR of multivariate regression analysis using a stepwise model selection approach. Shown are results from simulation studies for 7, 12 or 22 generations, since admixture, different sample sizes ($n = 100$ or 500), genetic effects ($g = 0.2$ or 0.5) and genetic architectures or gene action (null, no QTL; add, additive; dom, dominant; rec, recessive; over, overdominant; under, underdominant; 2loci, two additive loci; 3loci, three additive loci).
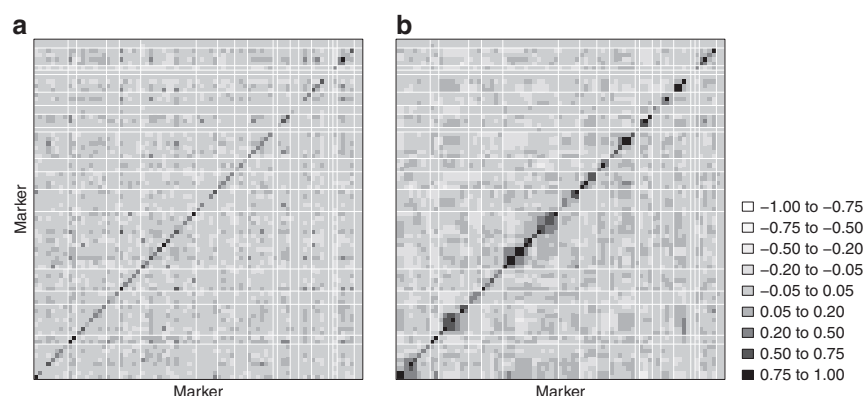

Figure 5 LD measured as correlation coefficients between all locus-pairs for admixed individuals for (**a**) the Italian hybrid zone locality; (**b**) simulated data after 22 generations (above diagonal) and 32 generations of admixture (below and on diagonal). Given are median correlation coefficients obtained from resampling *structure* LSAs in 500 replicates. Dark color indicates strong positive correlation between loci. White lines separate chromosomes.

Pooled data from four hybrid zones. By pooling hybrid zones, the set of identified candidate QTLs differed considerably from the set of QTLs identified in the Italian hybrid zone. Overlapping candidate loci from single-marker regression include G1133 (PETL and PTRAT), G1065 (PETL), O312 (PETXM) and O344 (LFAREA; Figure 3 and Supplementary Figure S6). Compared with the Italian data set, a substantially smaller amount of residual phenotypic variation could be explained by genotype in the final models (18.8%, 8.3%, 8.5% and 4.1% for PETL, LFAREA, PTRAT and PETXM, respectively).

Simulations
Simulation studies on the power of admixture mapping were performed for varying sample sizes, numbers of generations since admixture and genetic effects, by modeling eight different genetic architectures (Figure 4). Stepwise model selection on candidate QTLs reduced the number of ghost QTLs, even if those showed ΔAIC values higher than the target loci, and also allowed identification of physically linked QTLs as independent candidates (data not shown).

As expected, power increases significantly if 500 rather than 100 individuals are analyzed, or if the genetic effect is 0.5 rather than 0.2 (Figure 4). The total number of false positives does not increase by analyzing 100 individuals or traits with a small genetic effect (Supplementary Table S6); however, the FPR increases (Figure 4). An extended number of generations since admixture resulted in reduced power and higher FPR. More complex genetic architectures (two or three additive QTLs) tended to reduce power as well. This might be due to the smaller genetic effect of a single QTL on a trait for complex architectures, as the total genetic effect on a trait was held constant at 0.2 or 0.5, independent of the number of QTLs.

Linkage disequilibrium
For the Italian hybrid zone, an age of 32 generations since admixture was estimated by *ADMIXMAP* (95% confidence interval (CI): 22.9–42.6 generations). Correlation coefficients between all locus-pairs estimated by re-sampling genotypes showed weak and rather diffuse indications of LD irrespective of physical linkage (Figure 5a).

In contrast, simulated data after 22 and 32 generations of admixture with a similar distribution of Q and the same marker information content as the real data showed evidence for LD between physically linked loci (Figure 5b). The results indicate that the marker densities achieved by our microsatellite-based QTL scan are low compared with the number of recombination break points along hybrid chromosomes.

DISCUSSION
Power of admixture mapping in natural hybrid zones
Despite the frequent application of admixture mapping in human medicine (reviewed by Winkler *et al.*, 2010), very few studies have adopted this technique for phenotypic trait mapping in non-human organisms (Rieseberg *et al.*, 1999b; Malek *et al.*, 2012; Gompert *et al.*, 2013). Thus, the power and limitations of admixture mapping of quantitative traits in natural hybrid populations (which may differ in several aspects from human admixed populations) require exploration. We will discuss some of the obstacles and opportunities encountered by applying admixture mapping to interspecific hybrid zones of *Populus* spp. and simulated data. The pronounced differences between our study system and human admixed populations might be typical for other natural hybrid zones for which the application of admixture mapping would be desirable (e.g. reviewed by Buerkle and Lexer, 2008). These differences may include greater levels of divergence, limitations in sample size and genome coverage and the number of generations since admixture (Lexer *et al.*, 2010). Hybridizing populations of wild species may also display higher levels of intraspecific genetic and phenotypic variation, depending on their life-history characteristics (Lynch and Walsh, 1998).

Sample size. Knowledge from classical QTL studies (Tanksley, 1993) concerning the required sample size for mapping appears to be directly transferable to admixture mapping. For example, small sample size reduces power because QTLs of small effect will remain undetected (Bradshaw *et al.*, 1998; Zeng *et al.*, 2000; Orr, 2001). Studies comparing results for different sample sizes indicate that small sample size does not necessarily result in a higher absolute number of false positives, as QTLs found in smaller data sets were also detected after analyzing more individuals (Bradshaw *et al.*, 1998; Zeng *et al.*, 2000). Our simulation results for admixture mapping are perfectly in line with these findings (Figure 4 and Supplementary Table S6).

Genetic architecture. Our simulations indicate that over- or underdominant gene action can be detected by admixture mapping by including a dominant/recessive deviation term in regression models, similar to an approach commonly applied in classical QTL studies (Tanksley, 1993). This might improve QTL identification particularly in the presence of over- or underdominance, or pseudo-overdominance (Erickson *et al.*, 2004). More complex genetic architectures involving two or three additive loci resulted in lower detection power, probably because the total genetic effect on phenotype was partitioned across loci in the simulations (Figure 4 and Supplementary Table S6).

Results from the Italian mapping population indicate that the genetic architecture of the investigated quantitative traits is in fact rather complex. For the eight examined traits, on average 2.6 markers with a relative importance >0.5 remained in the final regression models, many of them with evidence for over- or underdominant gene action (Table 3; Figure 2b, Figure 3, Supplementary Figure S3b and Supplementary Figure S4).

LD and genome coverage. The estimated number of >30 generations since admixture in the Italian *Populus* hybrid zone suggests that recombination should have had sufficient time to break up physically linked loci. The estimated number of generations since admixture may be inaccurate depending on the hybrid zone history, the precision of the specified map and the amount of continuous gene flow from the parental taxa (Falush *et al.*, 2003; Hoggart *et al.*, 2004). Nevertheless, the existence of small ancestral chromosome blocks achieved by a considerable amount of recurrent recombination was also suggested by correlation coefficients between locus-pairs that did not differ between physically linked and unlinked loci (Figure 5a). In contrast, high correlation coefficients occasionally seen between physically unlinked loci suggest that the exact map location of identified candidate QTLs should be considered with caution. This LD pattern may indicate imperfectly shared synteny to the *P. trichocarpa* genome (Pakull *et al.*, 2009; Wang *et al.*, 2011), methodological issues in estimating LSA (e.g. due to errors during PCR amplification, or other problems associated to empirical rather than simulated data) or correlations of LSA between physically unlinked genetic regions (for example, because of shared selection pressures or epistasis). As our marker panel might thus not always cover the causal QTL regions, our results could be interpreted as a lower bound for the number of QTLs controlling the studied traits in natural interspecific hybrids.

The weak patterns of LD in the Italian mapping population (Figure 5a) highlight the importance to test for LD when attempting admixture mapping with a limited number of genetic markers. This becomes also evident from our simulation results where an increased number of generations since admixture reduced power and increased FPR (Figure 4), probably due to recombination between QTLs and genetic markers. Using an appropriate marker density can easily solve this issue in future studies. The small chromosome blocks identified in *Populus* hybrids indicate that fine-scale admixture mapping will be feasible, for example, by using data generated by genotyping-by-sequencing (e.g. Stölting *et al.*, 2013).

Variation between hybrid zones. Genetic data obtained from different hybrid zone localities are expected to differ because of processes unrelated to phenotype (e.g. genetic drift). With admixture mapping, this effect can be diminished by using local parental populations as reference for estimating LSA. Nevertheless, the possibilities of genetic drift of ancestry blocks within hybrid zones and local chromosome rearrangements represent a general issue for mapping when genetic data from different localities are pooled.

The substantial intraspecific variation for mappable (i.e. heritable) traits among hybrid zones, particularly for PETL and LFAREA, suggests the presence of considerable amounts of standing genetic variation (Barrett and Schluter, 2008) maintained by environment-dependent selection pressures on the investigated quantitative traits (Petit and Hampe, 2006; Figure 1 and Supplementary Figure S2), or QTL-by-environment interactions. On one hand, if QTL-by-environment interactions are present, admixture mapping can be an attractive approach for studying the genetics of species differences *in situ* (i.e. exactly where selection and speciation take place), and might provide more meaningful estimates of QTL effects than those obtained under unrealistic artificial growth conditions (Erickson *et al.*, 2004). On the other hand, environment-associated population structure combined with environmentally influenced plasticity of phenotypic traits may result in false-positive QTLs. Including appropriate environmental covariates into the linear model (e.g. see Hoggart *et al.*, 2003) or complementary multisite

common garden experiments would thus be desirable for future studies.

From a purely practical point of view, environment-dependent intraspecific variation in phenotypes causes statistical and conceptual issues when pooling data across hybrid zones, or across spatially separated localities of large 'mosaic' hybrid zones. For example, the southwest–northeast gradient for PETL and LFAREA in *P. alba* (Figure 1) results in a locality-dependent sign shift for regression slopes. This sign shift cannot be accounted for by the simple one-way analysis of variance approach that we have applied in this study. Using more sophisticated methods such as mixed-effect models would be desirable but would require larger sample sizes. Further, if phenotypes result from the combined effect of several genes and the environment, similar phenotypes may not necessarily result from similar genotypes (Tanksley, 1993; Kearsey and Farquhar, 1998). Thus, by pooling data across localities, power will only increase if the same QTL alleles are segregating in all pooled populations, but will decrease otherwise (Tanksley, 1993; Kearsey and Farquhar, 1998). This may account for the smaller amount of phenotypic variation explained by genotypic data from pooled populations compared with the Italian mapping population only. QTL studies indicate that environment dependence can result in few overlapping QTLs between localities, although major QTLs tend to be more strongly conserved (Tanksley, 1993). Genetic regions that were identified in both the pooled (Supplementary Figure S6) and the Italian data set (Figure 3 and Supplementary Figure S4) can be considered as good candidates for loci involved in species differences that are conserved among localities, and may thus represent important QTLs for species divergence. By contrast, geographically variable QTL expression, and thus little overlap between localities, is anticipated for this outcrossing forest tree, where large amounts of standing genetic variation can be the source for local adaptation (Petit and Hampe, 2006; de Carvalho *et al.*, 2010).

Genetic architecture of phenotypic traits

Direction and magnitude of effects. The genetic architecture of the studied traits appears to be complex and comprises mixtures of positive and negative effects (except LFSHAP; Table 3, Figure 2b and Supplementary Figure S3b). The bias of detected QTL effects into a single direction for PETL, LFSHAP and PETXM suggests that directional selection has been acting during species divergence for these traits (Orr, 1998), but would require a greater density of markers to be tested explicitly. For the remaining traits, QTLs with complementary positive and negative factors were detected, a feature expected for older taxa that are subjected to stabilizing selection over long divergence times (Orr, 2001). Complementary gene action will facilitate the maintenance of 'cryptic' variation, which can be the source for rapid adaptive response to changing environments (Barrett and Schluter, 2008). QTL effect sizes in terms of percent species difference were highly variable and often exceeded 100% if complementary factors were involved, consistent with previous findings in wild, annual sunflowers (Lexer *et al.*, 2005b).

The amounts of phenotypic variation explained by single QTL and by the final multivariate regression models, as well as the number of detected QTLs per trait, are in line with other QTL mapping studies in plants (Kearsey and Farquhar, 1998; Lynch and Walsh, 1998). Moreover, the total amount of phenotypic variation explained (13.5–38.7%) is similar to other admixture mapping studies (11.0–45.0%, Malek *et al.*, 2012; 4.9–24.1%, Gompert *et al.*, 2013; their 'admixed-naïve' (AN) analysis with individuals from five admixed populations). By considering the relatively low genomic coverage of our marker panel compared with the levels of LD (Figure 5a), the large effect sizes found by our study may be surprising but can be explained mainly by three factors. First, small sample sizes can lead to overestimation of QTL effects (see below). Second, effect sizes are likely to be larger in interspecific compared with intraspecific studies (Lexer *et al.*, 2005b). Third, the underlying genetic architecture of the studied traits might involve many QTLs (see above), of which only a subset could be identified with our low marker density and sample size. If these QTLs are genetically associated in hybrids (e.g. because of epistasis; Lexer *et al.*, 2010; Lindtke *et al.*, 2012), combined effect sizes from several QTLs in different parts of the genome might be detected by a single (surrogate) marker locus, even though block sizes are small. Our results suggest that much of the genetic contribution on phenotype could be explained by the set of identified candidate QTLs rather than a diffuse genome-wide effect, as the genome-wide control parameter Q did not reach high importance for most traits (Table 3).

Gene action. The identified QTLs showed varying modes of gene action, with a characteristic trend of *P. alba* alleles being dominant or overdominant ($\beta_2/\beta_1 > 0$; Table 3). These estimates for gene action may involve a large error, particularly if β_1 is close to zero. However, the observed trend still holds when considering calculations for $|\beta_1| > 0.3$ only. The finding is consistent with the observation that the investigated *Populus* hybrids tend to be more similar in morphology and habitat type to *P. alba* than to *P. tremula*. Floras and morphometric studies have often assigned hybrids as backcrosses towards *P. alba* (e.g. Adler *et al.*, 1994; Lexer *et al.*, 2009), whereas recent molecular genetic work has shown that hybrids are genetically intermediate between *P. alba* and *P. tremula* (Lexer *et al.*, 2010; Lindtke *et al.*, 2012). This discrepancy between phenotype and genotype might be partially explained by dominance effects of alleles originating from *P. alba*. Similar processes may be at work in interspecific hybrid zones of other groups of taxa with character conflicts involving phenotypic and genotypic data (Lexer *et al.*, 2009).

Many of the candidate QTLs appear to be over- or underdominantly expressed ($\beta_2/\beta_1 > 1$, or < -1; Table 3). Even though there is some evidence that the genetic basis actually involves pseudo-overdominance (Erickson *et al.*, 2004), this pattern is commonly observed for interspecific hybrids (Rieseberg *et al.*, 1999a), and represents a source for new genetic variation that can potentially lead to the evolution of novel lineages (Nolte and Tautz, 2010).

Colocalization of QTLs. For phenotypically correlated traits such as PETL, LFAREA and PTRAT, a similar set of candidate QTLs was detected for the Italian data set, although this association was less pronounced for the pooled data (Table 3, Figure 3 and Supplementary Figure S6). This colocalization of QTLs suggests the presence of pleiotropy (e.g. through involvement in plant growth or other developmental constraints) or tight linkage of causal QTLs for these traits, at least in Italy. Interestingly, colocalization of petiole length and leaf area has been detected by a classical QTL mapping study using a pair of related *Populus* species (Rae *et al.*, 2006). Pleiotropy or close linkage between QTLs for PETL, LFAREA and PTRAT is conceivable, as it would facilitate a coordinated response to selection on complex phenotypes (Lynch and Walsh, 1998). For example, it is very likely that an optimal fluttering physiology of leaves, a trait typical for *P. tremula*, will only be achieved by coordination between certain petiole and leaf blade characteristics.

Statistical issues. Admixture mapping likely suffers from similar statistical bias as classical QTL mapping studies, thus we might have

missed QTLs because of (1) low coverage resulting in insufficient linkage between causal QTLs and marker loci; (2) small sample size implying low resolution of QTLs of small effect, and overestimation of QTL effect sizes (Beavis effect; Beavis, 1998). By contrast, using natural populations rather than line crosses cultured under controlled environmental conditions will increase noise on phenotype expression and will thus diminish effect sizes.

The applied stepwise model selection approach appears to be a convenient procedure to distinguish distinct QTLs that are physically linked (Broman and Speed, 2002). Although we have not tested the exhaustive search algorithm on simulated data because of its high computational demand, its performance should exceed that of the stepwise selection procedure (Whittingham et al., 2006). For the Italian data set, results obtained by the stepwise and exhaustive search algorithm were very similar for some of the studied traits, for example, PETL, LFAREA and PETXA (Supplementary Table S5).

CONCLUSIONS AND FUTURE DIRECTIONS

By applying admixture mapping to natural hybrid zones of *Populus*, we have identified several candidate QTLs controlling ecologically important trait differences, including their effect sizes and modes of gene action. The genetic architecture of the investigated leaf traits appears to be complex, which might be attributed to the advanced stage of divergence of the examined species pair. High levels of intraspecific genetic and phenotypic variation inherent in natural populations as seen here in *Populus* spp., as well as QTL-by-environment interactions, represent a potential obstacle to admixture mapping. Improving sample sizes will facilitate the detection of genetic variants that represent important species differences, but this strongly relies on the availability of recombinant hybrids in nature. By jointly interpreting the results from four hybrid zones and numerical simulations, we have demonstrated the feasibility of admixture mapping of quantitative phenotypic trait differences in (non-human) natural hybrid zones. Future studies should consider enhancing genome coverage to increase mapping resolution, for example, by using genotyping-by-sequencing approaches. For *Populus*, this work is currently in progress and may enable validation of QTLs that were identified in this study. The results of this study indicate the potential to achieve high-resolution mapping even for species with long generation times. This demonstrates the potential of admixture mapping over classical QTL studies in unlocking the considerable amounts of variation present in long-lived, outcrossing species.

DATA ARCHIVING

Microsatellite genotypes, phenotype measurements for all 22 traits and R code for single-locus and multivariate regression analyses have been deposited at Dryad: doi:10.5061/dryad.12kh6.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

Adler W, Oswald K, Fischer R (1994). *Exkursionsflora von Oesterreich*. Verlag Eugen Ulmer: Stuttgart and Vienna, Germany and Austria.

Akaike H (1973). Information theory as an extension of the maximum likelihood principle. In: Petrov BN, Csaki F (eds) *Second International Symposium on Information Theory*. Akademiai Kiado: Budapest, pp 267–281.

Barrett RDH, Schluter D (2008). Adaptation from standing genetic variation. *Trends Ecol Evol* **23**: 38–44.

Barton NH, Keightley PD (2002). Understanding quantitative genetic variation. *Nat Rev Genet* **3**: 11–21.

Beavis WD (1998). QTL analyses: power, precision and accuracy. In: Paterson AH (ed) *Molecular Dissection of Complex Traits*. CRC Press: New York, NY, USA, pp 145–162.

Bradshaw HD, Otto KG, Frewen BE, Mckay JK, Schemske DW (1998). Quantitative trait loci affecting differences in floral morphology between two species of monkeyflower (*Mimulus*). *Genetics* **149**: 367–382.

Briscoe D, Stephens JC, Obrien SJ (1994). Linkage disequilibrium in admixed populations—applications in gene-mapping. *J Hered* **85**: 59–63.

Broman KW, Speed TR (2002). A model selection approach for the identification of quantitative trait loci in experimental crosses. *J R Stat Soc B* **64**: 641–656.

Buerkle CA, Lexer C (2008). Admixture as the basis for genetic mapping. *Trends Ecol Evol* **23**: 686–694.

Chakraborty R, Weiss KM (1988). Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proc Natl Acad Sci USA* **85**: 9119–9123.

Charlesworth D, Willis JH (2009). The genetics of inbreeding depression. *Nat Rev Genet* **10**: 783–796.

Churchill GA, Doerge RW (1994). Empirical threshold values for quantitative trait mapping. *Genetics* **138**: 963–971.

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

de Carvalho D, Ingvarsson PK, Joseph J, Suter L, Sedivy C, Macaya-Sanz D et al. (2010). Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula* L.), a widespread forest tree. *Mol Ecol* **19**: 1638–1650.

Erickson DL, Fenster CB, Stenoien HK, Price D (2004). Quantitative trait locus analyses and the study of evolutionary process. *Mol Ecol* **13**: 2505–2522.

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

Falush D, Stephens M, Pritchard JK (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes* **7**: 574–578.

Funk DJ, Nosil P, Etges WJ (2006). Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proc Natl Acad Sci USA* **103**: 3209–3213.

Gompert Z, Lucas LK, Nice CC, Buerkle CA (2013). Genome divergence and the genetic architecture of barriers to gene flow between *Lycaeides idas and L melissa*. *Evolution*; doi:10.1111/evo.12021 .

Grandillo S, Ku HM, Tanksley SD (1999). Identifying the loci responsible for natural variation in fruit size and shape in tomato. *Theor Appl Genet* **99**: 978–987.

Grueber CE, Nakagawa S, Laws RJ, Jamieson IG (2011). Multimodel inference in ecology and evolution: challenges and solutions. *J Evol Biol* **24**: 699–711.

Hoggart CJ, Parra EJ, Shriver MD, Bonilla C, Kittles RA, Clayton DG et al. (2003). Control of confounding of genetic associations in stratified populations. *Am J Hum Genet* **72**: 1492–1504.

Hoggart CJ, Shriver MD, Kittles RA, Clayton DG, McKeigue PM (2004). Design and analysis of admixture mapping studies. *Am J Hum Genet* **74**: 965–978.

Kearsey MJ, Farquhar AGL (1998). QTL analysis in plants; where are we now? *Heredity* **80**: 137–142.

Lexer C, Buerkle CA, Joseph JA, Heinze B, Fay MF (2007). Admixture in European *Populus* hybrid zones makes feasible the mapping of loci that contribute to reproductive isolation and trait differences. *Heredity* **98**: 74–84.

Lexer C, Fay MF, Joseph JA, Nica MS, Heinze B (2005a). 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. *Mol Ecol* **14**: 1045–1057.

Lexer C, Joseph J, Van Loo M, Prenner G, Heinze B, Chase MW et al. (2009). The use of digital image-based morphometrics to study the phenotypic mosaic in taxa with porous genomes. *Taxon* **58**: 349–364.

Lexer C, Joseph JA, Van Loo M, Barbara T, Heinze B, Bartha D et al. (2010). Genomic admixture analysis in European *Populus* spp. reveals unexpected patterns of reproductive isolation and mating. *Genetics* **186**: 699–U391.

Lexer C, Rosenthal DM, Raymond O, Donovan LA, Rieseberg LH (2005b). Genetics of species differences in the wild annual sunflowers, *Helianthus annuus* and *H. petiolaris*. *Genetics* **169**: 2225–2239.

Lexer C, Widmer A (2008). The genic view of plant speciation: Recent progress and emerging questions. *Philos Trans R Soc Ser B* **363**: 3023–3036.

Lindtke D, Buerkle CA, Barbara T, Heinze B, Castiglione S, Bartha D et al. (2012). Recombinant hybrids retain heterozygosity at many loci: new insights into the genomics of reproductive isolation in *Populus*. *Mol Ecol* **21**: 5042–5058.

Lynch M, Walsh B (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer: Sunderland, MA, USA.

Mackay TFC (2001). The genetic architecture of quantitative traits. *Annu Rev Genet* **35**: 303–339.

Malek TB, Boughman JW, Dworkin I, Peichel CL (2012). Admixture mapping of male nuptial colour and body shape in a recently formed hybrid population of threespine stickleback. *Mol Ecol* **21**: 5265–5279.

McKeigue PM (1998). Mapping genes that underlie ethnic differences in disease risk: methods for detecting linkage in admixed populations, by conditioning on parental admixture. *Am J Hum Genet* **63**: 241–251.

McKeigue PM (2005). Prospects for admixture mapping of complex traits. *Am J Hum Genet* **76**: 1–7.

Neuenschwander S, Hospital F, Guillaume F, Goudet J (2008). quantiNemo: an individual-based program to simulate quantitative traits with explicit genetic architecture in a dynamic metapopulation. *BioInformatics* **24**: 1552–1553.

Nolte AW, Tautz D (2010). Understanding the onset of hybrid speciation. *Trends Genet* **26**: 54–58.

Orr HA (1998). Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data. *Genetics* **149**: 2099–2104.

Orr HA (2001). The genetics of species differences. *Trends Ecol Evol* **16**: 343–350.

Pakull B, Groppe K, Meyer M, Markussen T, Fladung M (2009). Genetic linkage mapping in aspen (*Populus tremula* L. and *Populus tremuloides* Michx.). *Tree Genet Genom* **5**: 505–515.

Petit RJ, Hampe A (2006). Some evolutionary consequences of being a tree. *Annu Rev Ecol Evol Syst* **37**: 187–214.

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

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria, http://www.r-project.org/.

Rae AM, Ferris R, Tallis MJ, Taylor G (2006). Elucidating genomic regions determining enhanced leaf growth and delayed senescence in elevated CO_2. *Plant Cell Environ* **29**: 1730–1741.

Rieseberg LH, Archer MA, Wayne RK (1999a). Transgressive segregation, adaptation and speciation. *Heredity* **83**: 363–372.

Rieseberg LH, Buerkle CA (2002). Genetic mapping in hybrid zones. *Am Nat* **159**: S36–S50.

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

Schwarz G (1978). Estimating dimension of a model. *Ann Stat* **6**: 461–464.

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

Tanksley SD (1993). Mapping polygenes. *Annu Rev Genet* **27**: 205–233.

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U *et al.* (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**: 1596–1604.

Wang YX, Zhang B, Sun XY, Tan BY, Xu LA, Huang MR *et al.* (2011). Comparative genome mapping among *Populus adenopoda*, *P. alba*, *P. deltoides*, *P. euramericana* and *P. trichocarpa*. *Genes Genet Syst* **86**: 257–268.

Whittingham MJ, Stephens PA, Bradbury RB, Freckleton RP (2006). Why do we still use stepwise modelling in ecology and behaviour? *J Anim Ecol* **75**: 1182–1189.

Winkler CA, Nelson GW, Smith MW (2010). Admixture mapping comes of age. *Annu Rev Genom Hum Genom* **11**: 65–89.

Zeng ZB, Liu JJ, Stam LF, Kao CH, Mercer JM, Laurie CC (2000). Genetic architecture of a morphological shape difference between two *Drosophila* species. *Genetics* **154**: 299–310.

Zhu XF, Luke A, Cooper RS, Quertermous T, Hanis C, Mosley T *et al.* (2005). Admixture mapping for hypertension loci with genome-scan markers. *Nat Genet* **37**: 177–181.