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

C. Lexer,* M. F. Fay,* J. A. Joseph,* M.-S. Nicola† and B. Heinze†

* Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3DS, UK, †Department of Genetics, Federal Office and Research Centre for Forests, Hauptstrasse 7, A-1140 Vienna, Austria

Abstract

The renewed interest in the use of hybrid zones for studying speciation calls for the identification and study of hybrid zones across a wide range of organisms, especially in long-lived taxa for which it is often difficult to generate interpopulation variation through controlled crosses. Here, we report on the extent and direction of introgression between two members of the 'model tree' genus *Populus*: *Populus alba* (white poplar) and *Populus tremula* (European aspen), across a large zone of sympatry located in the Danube valley. We genotyped 93 hybrid morphotypes and samples from four parental reference populations from within and outside the zone of sympathy for a genome-wide set of 20 nuclear microsatellites and eight plastid DNA restriction site polymorphisms. Our results indicate that introgression occurs preferentially from *P. tremula* to *P. alba* via *P. tremula* pollen. This unidirectional pattern is facilitated by high levels of pollen vs. seed dispersal in *P. tremula* (pollen/seed flow = 23.9) and by great ecological opportunity in the lowland floodplain forest in proximity to *P. alba* seed parents, which maintains gene flow in the direction of *P. alba* despite smaller effective population sizes (*Nₐ*) in this species (*P. alba* *Nₐ* c. 500–550; *P. tremula* *Nₐ* c. 550–700). Our results indicate that hybrid zones will be valuable tools for studying the genetic architecture of the barrier to gene flow between these two ecologically divergent *Populus* species.

Keywords: admixture, dispersal, ecological divergence, hybridization, microsatellites, plastid DNA

Received 30 September 2004; revision received 10 December 2004; accepted 10 December 2004

Introduction

The genetic architecture of barriers to gene flow plays a central role in molecular ecology and evolutionary genetics. It holds the key to reconstructing the sequence of genetic changes that accompany or facilitate speciation (Bradshaw *et al.* 1995; Rieseberg *et al.* 1999; Orr 2001). Genetic architecture also allows assessment of the role of hybrid zones as evolutionary filters: hybrid zones may permit introgression of beneficial genes into a new genetic background, while preventing the movement of detrimental genes across species boundaries (Barton & Hewitt 1985; Barton & Gale 1993; Martinsen *et al.* 2001). At the within-species level, studying genetic architecture provides information about the potential of gene flow and selection to facilitate the spread of advantageous alleles across populations and thereby to maintain species as cohesive units (Rieseberg & Morjan 2004), and about the role of natural selection in promoting population divergence (Wilding *et al.* 2001; Luikart *et al.* 2003; Campbell and Bernatchez 2004).

Hybrid zones have long been viewed as ‘natural laboratories’ for studying barriers to gene flow (Barton & Hewitt 1985; Harrison 1990; Barton & Gale 1993). The fascination of molecular ecologists for hybrid zones stems, in part, from the fact that hybrid zones are often composed of a wide variety of genotypes resulting from many generations of recombination. Moreover, linkage disequilibrium (LD) induced by hybridization is thought to facilitate studies of microevolutionary processes (Rieseberg *et al.* 1999; Rogers
et al. 2001; Rieseberg & Buerkle 2002), reminiscent of genetic mapping studies in admixed human populations (Briscoe et al. 1994). An important aspect is that hybrid zones potentially allow the fitness of adult hybrid genotypes to be evaluated under natural conditions, which is often difficult to achieve otherwise in long-lived or experimentally less tractable species (Rieseberg & Buerkle 2002; Lexer et al. 2003, 2004). Also, hybrid zones represent promising venues for studying interactions between heritable genetic variation and diversity at the community or ecosystem level (Whitham 1989; Whitham et al. 1999).

The potential usefulness of natural hybrid zones for addressing important evolutionary questions calls for the identification and study of hybrid populations across a wide range of animals and plants including different life histories and mating systems. A well-developed conceptual framework exists for characterizing newly described hybrid zones. Particularly important questions include: (i) What is the genomic composition of hybrids, i.e. is the hybrid zone dominated by F₁s (e.g. Milne et al. 2003), or are later generation hybrids frequent (e.g. Rieseberg et al. 1999; cases reviewed by Arnold et al. 2004)? (ii) Is the hybrid zone maintained by a balance between dispersal and selection against hybrids (Barton & Hewitt 1985; Barton & Gale 1993), or does its structure also depend on adaptation to the external environment (Moore 1977; Harrison 1990)? (iii) How symmetric or asymmetric are patterns of gene exchange, and to what extent are they caused by intrinsic genetic vs. ecological factors (Arnold 1993; Bacilieri et al. 1996)? Answering these questions informs molecular ecologists about which components of the barrier to gene flow are likely to be most important in maintaining the hybrid zone, and which analytical tools are most appropriate for studying its evolution.

The identification and study of hybrid zones is a particularly timely topic in long-lived forest trees. Although conventional quantitative trait locus (QTL) mapping has facilitated studies of the genetic architecture of interspecific differences in some taxa (e.g. Wu et al. 1997; Frewen et al. 2000; Saintagne et al. 2004), for most species the necessary advanced generation crosses are not available. This is unfortunate, because trees exhibit numerous features that distinguish them from annual model plants and likely affect the way in which their populations evolve, e.g. control of leaf and flower phenology, seasonal reallocation of nutrients, extreme plasticity in stress response-related characters, juvenile–mature phase change, and different dynamics of colonization into open habitats (Austerlitz et al. 2000; Bradshaw et al. 2000). Also, there is considerable interest in the question concerning the conditions under which trees are most likely to be able to adapt in situ to the expected rate of climate change (Davis & Shaw 2001). Natural hybrid zones may hold important clues regarding the genetic architecture of adaptively important traits in trees, in particular in cases in which species barriers are at least partly ecological, i.e. composed of genes or chromosomal segments conferring differential adaptation. Numerous well-defined examples exist for ecological species barriers among temperate trees (e.g. Van Valen 1976; Eckenwalder 1996; McKinnon et al. 2001; reviewed by Lexer et al. 2004), and no group of taxa may be better suited for studying them than the genus Populus (poplars, aspens, cottonwoods).

Populus species are diploids (2n = 38) with relatively small genome sizes (550 Mb; 2C = 1.1 pg), and the numerous favourable attributes of this genus have made Populus a ‘model forest tree’ (Bradshaw et al. 2000) and led to an international effort to sequence its complete genome (Tuskan et al. 2004a; see http://www.ornl.gov/sci/ipgc). Species barriers in Populus are often porous, allowing for widespread hybridization, and genomic segments carrying neutral or favourable genes appear to be able to cross the species boundaries in many cases (Rajora & Dancik 1992; Eckenwalder 1996; Martinsen et al. 2001; Floate 2004). Comparative linkage mapping in three different species (Populus deltoides, Populus trichocarpa, and Populus nigra) revealed widespread homologies for most of the 19 linkage groups of the Populus genome (Cervera et al. 2001), suggesting that species barriers in Populus are genic rather than chromosomal. Further evidence for partially conserved gene order stems from cytogenetic data for ribosomal sequences (Prado et al. 1996). Taken together, this suggests that barriers are composed primarily of single genes and their interactions, although it is too early to rule out a role for chromosomal rearrangements in species isolation.

Here, we study the species barrier between Populus alba (white poplar) and Populus tremula (European aspen), two ecologically divergent species that hybridize frequently in Europe. We ask the following questions regarding hybrid zones of P. alba and P. tremula, using data from a wide zone of sympatry along the Danube as well as parental reference samples from outside the sympatric zone: (i) What is the genomic composition of hybrids when estimated with a genome-wide set of codominant nuclear microsatellites, and what do these estimates tell us about the extent and direction of introgression? (ii) How do patterns of nuclear gene exchange relate to the species origin of the plastid DNA molecule in the same set of individuals? In addition, we assess the role of effective population size (Nₑ) and pollen/seed flow in gene introgression and comment on the potential usefulness of hybrid zones among P. alba and P. tremula for genomic studies of microevolutionary processes.

Materials and methods

Hybrid zones between Populus alba and Populus tremula

The existence of hybrids between Populus alba and Populus tremula in the lowland floodplain forests of the Danube
near Vienna, Austria, has been suspected for several decades (Krembs 1956; Adler et al. 1994). Based on morphological evidence, some authors have argued that the hybrid form, known as Populus × canescens or grey poplar, may be even more common than pure P. alba (Lazowski 1997). To our knowledge, only two molecular studies exist on hybrids among the two species (Rajora & Dancik 1992; Fossati et al. 2004). Although neither of them was carried out within a population genetic framework (between one and 10 trees were sampled for each of several populations and compared to controlled crosses), both studies indicate that P. alba and P. tremula are well differentiated for allele frequencies at dominant and codominant genetic markers and that P. × canescens hybrids from different European river valleys appear to be a mixture of early generation hybrids and backcrosses (Rajora & Dancik 1992; Fossati et al. 2004).

P. alba and P. tremula differ strongly in their habitat requirements. In the Danube valley, P. alba favours flooded areas adjacent to the river and often extends to the active zone of the floodplain, where it plays a key role during early stages of succession (Lazowski 1997). Tolerance to flooding in P. alba may be mediated by a number of different characters, including life history traits (e.g. lack of dormancy, rapid germination and seedling growth, rapid root growth; Karrenberg et al. 2002), and physiological adaptations related to the prevention of cell damage caused by ethanol, a by-product of anaerobic processes in flooded plant tissue (Kreuzwieser et al. 1999). The preferred habitats of P. × canescens hybrids are located in close proximity to P. alba in the floodplain forest (Krembs 1956; Adler et al. 1994; Lazowski 1997), although fine-scale spatial patterns and niche widths of the two taxa have not yet been compared.

The second parental species, P. tremula, does not grow in flooded areas, or in close proximity of the Danube or its tributaries. Instead, its preferred habitats are located several kilometres away, reaching into the submontane and montane zones of the Alps (Adler et al. 1994). In addition to these divergent habitat requirements, numerous morphological traits distinguish P. alba and P. tremula (Adler et al. 1994), but their adaptive significance is unknown. It is also an open question whether the Danube valley hybrid zone follows a tension zone model with a balance of dispersal and selection against hybrids, or whether ecology plays a role in maintaining the zone (Barton & Hewitt 1985; Harrison 1990; Arnold 1997). One of the primary pieces of evidence for or against a role for ecological selection, the geographical width of the zone of hybridization, is difficult to assess because present-day floodplain forests represent only fragments of their historical distribution.

Population and within-genome sampling

The sampling scheme used here was chosen to best meet the trade-off between sampling gene copies in populations and sampling loci in the genome. Twenty independent nuclear microsatellite loci (chosen from different linkage groups) were assayed in a total of 93 P. × canescens hybrid morphotypes from the Danube valley hybrid zone and two reference populations of each parental species, P. alba and P. tremula. Reference populations had an average sample size of N = 20 and included, for each species, one population from ‘within’ and one adjacent parental population from ‘outside’ the zone of sympathy. In addition, eight plastid DNA restriction site polymorphisms were typed in 46 of the sampled hybrids and in an average of 19 plants from each of the reference populations. The 20 nuclear microsatellites (Table 1), developed by Tuskan et al. (2004b) and Van der Schoot et al. (2000), are available at the following web site: http://www.ornl.gov/sci/ipgc/ssr_resource.htm. The four plastid DNA markers (described in the succeeding section) were identified based on Grivet et al. (2001).

Based on the highly outcrossing breeding system of the study species (wind-dispersed pollen and seeds; Petit et al. 2003) and a lack of population differentiation in preliminary microsatellite studies, populations as sampled here covered large geographical areas of up to 60 km² (i.e. populations typically included numerous ‘forest stands’ as usually defined by silviculturists). The following populations were sampled: P. × canescens/vie (P. × canescens hybrid zone in the Danube valley near Vienna, Austria; sampling mid-point: 48.26°N, 16.23°E), P. alba/vie (reference sample of P. alba from the zone of sympathy in the Danube valley near Vienna, Austria; sampling mid-point: 48.26°N, 16.23°E), P. alba/rom (P. alba from outside the zone of sympathy in the Danube valley in Romania; sampling mid-point: 43.77°N, 23.96°E), P. tremula/vie (reference sample of P. tremula from the zone of sympathy in the Danube valley near Vienna, Austria; sampling mid-point: 48.28°N, 15.89°E), P. tremula/car (P. tremula from outside the sympatric zone in the Eastern Alps in Carinthia, Austria; sampling mid-point: 46.62°N, 13.85°E).

Molecular marker assays

Total genomic DNA was extracted using either the DNeasy Plant Mini Kit (QIAGEN) or a modified approach based on Doyle & Doyle (1987), and DNA was quantified using either a TKO-100 fluorometer (Hoefer Scientific Instruments) or an Eppendorf BIO photometer. The 20 nuclear microsatellites were polymerase chain reaction (PCR)-amplified following methods described previously by Burke et al. (2002), making use of a standard touchdown cycling program with an annealing temperature (Tₘ) of 48 °C and a three-primer protocol including unlabelled M13-tagged forward and unlabelled/untagged reverse primers for each marker, and a third ‘universal’ M13-primer labelled...
analysed using Biosystems), and result files from the sequencers were mined using either the GENESIZE-500 ROX (GENPAK) multiplexing. Molecular sizes in base pairs were determined using either the GENSIZE-500 ROX (GENPAK) or the GENESCAN-500 ROX size standard (Applied Biosystems). Microsatellite genotypes were resolved either on an AB3700 or AB3100 automated sequencer (Applied Biosystems), making use of the different fluorescent dyes and size differences between loci for each locus.

Data analysis

Genetic diversity of sampled loci and populations. In order to characterize the microsatellite loci in the two study species and their hybrids, the number of alleles (A), variance in allele size (Var), expected heterozygosity (H_E), and observed heterozygosity (H_O) were calculated for each locus using the program MSA (Dieringer & Schloetterer 2003). For ease of presentation, populations of each parental species were combined in the analysis, which is justified by low levels of genetic divergence between them (succeeding discussion). All genetic diversity estimates obtained in MSA were corrected for differences in sample size. In addition, departures from Hardy–Weinberg equilibrium (HWE) in each studied taxon were computed using exact tests in genepop (Raymond & Rousset 1995), in order to test the possibility that particular loci may deviate from HWE because of null alleles (= allele nonamplification) stemming from cross-species amplification of the microsatellites. Subsequently, each population studied was characterized using the variance in allele size, H_E, and H_O calculated by MSA, allelic richness corrected for sample size in fstat (Goudet 1995), and the within-population inbreeding coefficient f estimated for

with one of the fluorescent dyes, FAM, HEX, NED or JOE (Applied Biosystems). Microsatellite genotypes were resolved either on an AB3700 or AB3100 automated sequencer (Applied Biosystems), making use of the different fluorescent dyes and size differences between loci for multiplexing. Molecular sizes in base pairs were determined using either the GENESIZE-500 ROX (GENPAK) or the GENESCAN-500 ROX size standard (Applied Biosystems), and result files from the sequencers were analysed using GENESCAN and GENOTyper software (Applied Biosystems).

Four plastid DNA fragments were screened for restriction site polymorphisms, yielding a total of eight polymorphisms that were used for identifying and characterizing plastid DNA haplotypes. The PCR primers/restriction enzymes used were: ccmp10R-trnHM/MspI, atpBSAM-rbcLSAM/EcoRI, trnTP-trnDP/HinI, ccmp10R-rpl16R1516/EcoRI-HhaI. The nomenclature of the primers follows Grivet et al. (2001) and the web-based resource at http://bfw.ac.at/200/1859.html, where the reaction conditions can also be found. The restriction fragments were resolved on 2% agarose gels and visualized by UV/ethidium bromide staining.

Table 1 Genetic variability at 20 nuclear microsatellite loci in Populus alba, Populus tremula, and their hybrids (P. × canescens), including locus name, repeat type, number of alleles (A), variance in allele size (Var), expected heterozygosity (H_E), and observed heterozygosity (H_O) for each locus.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Repeat type</th>
<th>P. alba</th>
<th>P. × canescens</th>
<th>P. tremula</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>A</td>
<td>Var</td>
<td>H_E</td>
</tr>
<tr>
<td>PMGC 2852‡</td>
<td>di</td>
<td>10</td>
<td>24.2</td>
<td>0.715</td>
</tr>
<tr>
<td>WMPS 15§</td>
<td>tri</td>
<td>7</td>
<td>20.5</td>
<td>0.773</td>
</tr>
<tr>
<td>ORPM 312‡</td>
<td>tri</td>
<td>9</td>
<td>28.3</td>
<td>0.789</td>
</tr>
<tr>
<td>ORPM 344‡</td>
<td>di</td>
<td>4</td>
<td>0.8</td>
<td>0.320</td>
</tr>
<tr>
<td>ORPM 206‡</td>
<td>tri</td>
<td>2</td>
<td>3.7</td>
<td>0.023</td>
</tr>
<tr>
<td>ORPM 127‡</td>
<td>di</td>
<td>3</td>
<td>2.1</td>
<td>0.118</td>
</tr>
<tr>
<td>ORPM 202‡</td>
<td>tri</td>
<td>3</td>
<td>1.5</td>
<td>0.571</td>
</tr>
<tr>
<td>ORPM 29‡</td>
<td>di</td>
<td>2</td>
<td>1.4</td>
<td>0.028</td>
</tr>
<tr>
<td>ORPM 30–1‡</td>
<td>di</td>
<td>2</td>
<td>7.4</td>
<td>0.088</td>
</tr>
<tr>
<td>ORPM 30–2‡</td>
<td>di</td>
<td>14</td>
<td>17.7</td>
<td>0.828</td>
</tr>
<tr>
<td>ORPM 220‡</td>
<td>tetra</td>
<td>1</td>
<td>0.0</td>
<td>0.000</td>
</tr>
<tr>
<td>ORPM 28‡</td>
<td>di</td>
<td>3</td>
<td>1.3</td>
<td>0.514</td>
</tr>
<tr>
<td>ORPM 137‡</td>
<td>di</td>
<td>7</td>
<td>13.4</td>
<td>0.630</td>
</tr>
<tr>
<td>ORPM 14‡</td>
<td>tetra</td>
<td>1</td>
<td>0.0</td>
<td>0.000</td>
</tr>
<tr>
<td>ORPM 21‡</td>
<td>di</td>
<td>2</td>
<td>0.1</td>
<td>0.204</td>
</tr>
<tr>
<td>ORPM 60‡</td>
<td>tri</td>
<td>7</td>
<td>28.5</td>
<td>0.675</td>
</tr>
<tr>
<td>ORPM 149‡</td>
<td>di</td>
<td>4</td>
<td>12.6</td>
<td>0.600</td>
</tr>
<tr>
<td>ORPM 167‡</td>
<td>di</td>
<td>2</td>
<td>0.2</td>
<td>0.024</td>
</tr>
<tr>
<td>ORPM 214‡</td>
<td>di</td>
<td>3</td>
<td>1.9</td>
<td>0.371</td>
</tr>
<tr>
<td>WPMS 5§</td>
<td>di</td>
<td>7</td>
<td>31.7</td>
<td>0.830</td>
</tr>
</tbody>
</table>

†Markers developed by Tuskan et al. (2004b); †markers from http://www.ornl.gov/sci/ipgc/ssr_resource.htm. (2000); §markers by Van der Schoot et al. 2000; ‡markers from http://www.ornl.gov/sci/ipgc/ssr_resource.htm. (2000); ‡markers from http://www.ornl.gov/sci/ipgc/ssr_resource.htm. (2000); ‡markers by Van der Schoot et al. 2000; ‡two independent loci with nonoverlapping allele size ranges amplified; *P < 0.05, **P < 0.005, exact tests for departure from Hardy–Weinberg equilibrium.
Effective population sizes and migration rates. Theta (4Nₑµ, with Nₑ = effective population size and µ = mutation rate) for populations of *P. alba* and *P. tremula,* and the effective number of migrants (Nₑm) between populations of each species, were estimated following a coalescent theory and maximum-likelihood-based approach using MIGRATE 1.5 (Beerli & Felsenstein 1999). Genetic divergence between populations of each species (Fₛₜ; Weir & Cockerham 1984) was used to obtain initial start values for the estimation of theta and Nₑm. The computations were carried out under both the infinite allele model (IAM; Kimura & Crow 1964) and the stepwise mutation model (SMM; Kimura & Ohta 1978), and effective population sizes were estimated from theta values by assuming a microsatellite mutation rate of 10⁻³ per gamete per generation.

Nuclear admixture analysis and assignment tests. Nuclear admixture proportions (Q) were estimated using the Bayesian approach implemented in STRUCTURE version 2 (Pritchard et al. 2000). The calculations were carried out under the admixture model allowing for correlated allele frequencies, using the burn-in and simulation settings recommended by the user manual. We did not estimate the number of populations K from the data because this would assume that populations are in HWE, which is not expected in hybrid zones. Instead, populations of each parental species were combined in accordance with high levels of gene flow (Nₑm) between them, and admixture proportions Q for *P. × canescens* hybrids were estimated based on pooled allele frequency data from each parental species. Prior population information was used for all parental genotypes originating from outside the zone of sympatry (plants from populations *P. alba/rom* and *P. tremula/car*), which allowed us to make optimal use of the allopatric reference samples, and to estimate the probability that any of these plants had received genes from the other species within the last three generations.

For ease of presentation, patterns of admixture between *P. alba* and *P. tremula* were depicted in the form of twodimensional population assignments rather than one-dimensional admixture proportions. Simple likelihood-based assignment tests (sensu Rannala & Mountain 1997) were carried out in ARLEQUIN, and the results were verified using a genetic distance-based method which requires even fewer assumptions about population structure (Cornuet et al. 1999).

Plastid DNA haplotype network. A median-joining (MJ) network (Bandelt et al. 1999) was constructed based on plastid DNA haplotypes using the program NETWORK (www.fluxus-engineering.com). This method is capable of resolving even complex haplotype phylogenies (Posada & Crandall 2001). It uses parsimony criteria to identify median vectors, i.e. consensus sequences of mutually close sequences of markers which are biologically equivalent to possible unsampled or extinct ancestral haplotypes. Initial runs using all haplotypes detected by the eight restriction-site polymorphisms yielded a highly complex structure with numerous reticulations. Hence, following the user manual, we excluded singleton haplotypes and downweighted rapidly mutating characters in subsequent runs. The default settings were used for all other parameters. To test for a possible role of cyto-nuclear incompatibilities in determining patterns of introgression across the hybrid zone (Arnold 1993), nuclear admixture proportions and plastid DNA haplotype data were compared in a simple way: nonparametric Spearman rank correlations were used to test for a possible association between nuclear and plastid genomic composition using SPSS (SPSS Inc.).

Results

Variability at microsatellite loci

All 20 nuclear microsatellites were variable with up to 20 alleles segregating in each taxon, expected heterozygosities (Hₑ) of up to 0.903, and observed heterozygosities (Hₒ) as high as 0.947 (Table 1). Variance in allele size, a simple diversity parameter under the SMM, followed a similar trend as diversity estimates under the IAM. One locus displayed significant deviations from HWE in *Populus alba* and another did so in *Populus tremula,* possibly because
of population subdivision. In contrast, four loci deviated from HWE in \( P. \times \) canescens, consistent with more frequent departures from random mating in the hybrid population. No single locus deviated from HWE across all three taxa (Table 1). Allele frequency tables for each locus/taxon in Table 1 are available from the corresponding author on request.

**Genetic diversity, effective population size, and gene flow**

Microsatellite-based diversity estimates for each of the five populations studied (Table 2) revealed higher \( H_E, H_O \), and variance in allele size for populations of \( P. \) tremula compared to \( P. \) alba, while \( P. \) × canescens hybrids were generally intermediate between the two parental species, and \( H_O \) for plastid DNA markers followed a similar pattern across populations (Table 2). In contrast, allelic richness at microsatellite loci, the most sensitive diversity parameter used in this study, revealed increased variability in \( P. \times \) canescens, probably because of the admixture of alleles from each parental species (Table 2).

In accordance with higher levels of diversity, effective population size was higher in \( P. \) tremula than in \( P. \) alba (Fig. 1). The estimates reported here were obtained under the SMM, although the IAM estimates followed the same pattern across populations and species. Estimates of \( N_m \) obtained from the same analysis (Fig. 1) revealed high levels of gene flow and connectedness between populations of each parental species (\( P. \) alba: \( N_m = 3.1 \) and 3.4 depending on direction of gene exchange, \( P. \) tremula: \( N_m = 14.7 \) and 6.9, respectively). The between-species differences in \( N_m \) likely reflect different geographical distances between populations – c. 780 km for the reference populations of \( P. \) alba and 240 km for populations of \( P. \) tremula. Based on high levels of gene exchange, populations from each parental species were pooled for their use as reference samples in admixture analyses.

**Nuclear admixture analysis**

Bayesian admixture analysis indicated that the sympatric populations of \( P. \) alba and \( P. \) tremula were pure (admixture coefficients for each individual \( Q > 0.950 \) or \( Q < 0.050 \), respectively), with the exception of two individuals of \( P. \) alba which were reclassified as \( P. \times \) canescens based on their \( Q \) values (all results reported here are for populations after reclassification). The probability that any of the allopatric populations (\( P. \) alba/vie and \( P. \) tremula/car) had received genes by interspecific introgression within the last three generations before sampling did not reach the \( P = 0.05 \) level for any of the individuals. In effect, Bayesian admixture analysis suggested that the error rate of parental species identification in the field was low. In contrast, a wide range of admixture proportions was observed for \( P. \times \) canescens hybrid morphotypes (95% confidence interval of \( Q = 0.468–0.998 \), range = 0.072–0.998), indicating the presence of a wide range of hybrid generations and preferential backcrossing to \( P. \) alba. Hybrid morphotypes with \( P. \) alba-like genotypes (\( Q > 0.950 \)) were kept in

**Table 2 Genetic variability in populations of Populus alba, Populus tremula, and \( P. \times \) canescens, including the variance in allele size (Var), allelic richness corrected for sample size, expected heterozygosity (\( H_E \)), observed heterozygosity (\( H_O \)), and the within-population inbreeding coefficient \( f \) for microsatellites, the number of plastid DNA haplotypes found in each sample, and \( H_E \) for plastid DNA markers.**

<table>
<thead>
<tr>
<th>Species/population</th>
<th>Microsatellites</th>
<th>Plastid DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Var</td>
<td>Allelic richness</td>
</tr>
<tr>
<td>( P. ) alba/vie</td>
<td>12.2</td>
<td>3.23</td>
</tr>
<tr>
<td>( P. ) alba/rom</td>
<td>8.3</td>
<td>3.27</td>
</tr>
<tr>
<td>( P. \times ) canescens/vie</td>
<td>152.2</td>
<td>3.97</td>
</tr>
<tr>
<td>( P. ) tremula/vie</td>
<td>23.4</td>
<td>3.86</td>
</tr>
<tr>
<td>( P. ) tremula/car</td>
<td>20.5</td>
<td>3.89</td>
</tr>
</tbody>
</table>

*Exact tests for departures from Hardy–Weinberg equilibrium \( P < 0.005 \).
the P. × canescens category in this study, because, by necessity, admixture analysis underestimates allelic contributions from the donor species to highly backcrossed individuals unless many loci in the genome have been sampled.

Likelihood-based and genetic distance-based genotype assignments gave comparable results, and the former are plotted in Fig. 2. The tendency of P. × canescens hybrids to backcross towards P. alba is clearly visible in Fig. 2a. In addition, Fig. 2b reveals assignment probabilities among P. alba and P. × canescens that would not have been discernible in Fig. 2a because of the high differentiation of allele frequencies between the two parental species (scaling effects). Hybrid plants that are differentiated from both parental species in plots a and b are recombinant early generation hybrids, as revealed by an inspection of individual microsatellite genotypes, except for four plants which appear to be F1s.

Genetic structure at nuclear and plastid DNA markers

Genetic divergence statistics obtained by AMOVA were significant at all hierarchical levels, when only populations of the two parental species were considered or when the analysis was carried out with P. × canescens nested within P. alba. For simplicity, only the results obtained under the IAM are summarized here (Table 3). Most importantly, the proportion of variation among populations within each species was small for nuclear microsatellites (2%; $F_{SC} = 0.038$) and clearly larger for plastid DNA (15%; $F_{SC} = 0.246$), while the proportion of variation within populations followed a complementary pattern (Table 3). Our data inform us about the role of pollen vs. seed dispersal in maintaining patterns of gene introgression. As expected from gene flow and admixture analyses, populations of each parental species grouped together with high bootstrap support in the neighbour-joining tree, and P. × canescens had a stronger affinity to P. alba than to P. tremula (Fig. 3).
Median-joining analysis resulted in a haplotype network with two major groups, one of which contained haplotypes typical of *P. alba*, whereas the other one consisted of *P. tremula* haplotypes (Fig. 4). The complete matrix of haplotype frequencies across taxa is available from the authors on request. Among eight haplotypes found in *P. × canescens*, six were either typical of *P. alba* or were derived directly from *P. alba* types (Fig. 4; Table 4). One haplotype found in hybrids was typical of *P. tremula* (haplotype H9) and one was shared among all three taxa, possibly representing retained ancestral polymorphism (haplotype H16; Fig. 4; Table 4). Hybrids that carried *P. alba*-derived haplotypes were characterized by a wide range of nuclear admixture proportions, ranging from individuals with an affinity to *P. tremula* \( (Q = 0.072) \) via trees with intermediate \( Q \) values to *P. alba*-like plants \( (Q > 0.950; \) Table 4). A simple test for correlation between nuclear and plastid genomic composition in these plants yielded no significant result (Spearman’s \( r = -0.230, P = 0.197 \)). Our data suggest that introgression among the two species is primarily via *P. tremula* pollen and *P. alba*.
Table 4 Plastid DNA haplotypes found in *P. x canescens* hybrids after exclusion of three singletons, including indication about haplotype sharing among taxa, group position in the plastid DNA network, frequencies in both species and their hybrids, and range of nuclear admixture proportions (Q) for plants carrying each haplotype

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Taxon</th>
<th>Haplotype group</th>
<th>Frequencies</th>
<th>Nuclear admixture (Q)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>alba (N = 41)</td>
<td>can (N = 46)</td>
<td>trem (N = 33)</td>
</tr>
<tr>
<td>H8</td>
<td>alba/can</td>
<td>alba</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>H9</td>
<td>trem/can</td>
<td>trem</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>H12</td>
<td>can</td>
<td>alba</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>H13</td>
<td>can</td>
<td>alba</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>H15</td>
<td>alba/can</td>
<td>alba</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>H16</td>
<td>alba/can/trem</td>
<td>alba</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>H23</td>
<td>can</td>
<td>alba</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>H25</td>
<td>alba/can</td>
<td>alba</td>
<td>4</td>
<td>17</td>
</tr>
</tbody>
</table>

seed parents, and that this asymmetric pattern is shaped by factors other than intrinsic cytonuclear incompatibilities.

Discussion

Recently, the role of natural selection in population divergence and speciation has been hotly debated (Merila & Crnokrak 2001; Wilding *et al.* 2001; McKay & Latta 2002; Rieseberg *et al.* 2002; Campbell and Bernatchez 2004), and this has triggered renewed interest in hybrid zones as venues for studying the factors that create or maintain barriers to gene flow between divergent populations (Rieseberg *et al.* 1999; Martinsen *et al.* 2001; Rogers *et al.* 2001; Remington & Perugganan 2003; Vines *et al.* 2003). The identification and characterization of hybrid zones in diverse taxa is therefore an important task with numerous potential spin-offs in molecular ecology. This is particularly true for organisms with existing genome programs such as *Populus* (Tuskan *et al.* 2004a), because the generation of complete genomic sequences, expressed sequence tag (EST) libraries, or other sources of genetic markers make it possible to study hybrid zones from a genomic perspective. Our results on the extent and direction of introgression between the ecologically divergent species, *Populus alba* and *Populus tremula*, directly contribute to the development of this approach in members of a genus with considerable ecological importance (Whitham 1989; Eckenwalder 1996; Bradshaw *et al.* 2000).

Extent and direction of introgression

Our results confirm earlier reports of introgression between *P. tremula* and *P. alba* which were based on less extensive sampling of dominant and/or codominant nuclear markers (Rajora & Dancik 1992; Fossati *et al.* 2004), and they also allow us to suggest possible mechanisms responsible for the observed patterns of hybridization. In the present study, 20 microsatellite loci, representative of a large proportion of the nuclear genome of *Populus*, and eight plastid DNA restriction site polymorphisms allow us to estimate effective population size (*N*<sub>e</sub>) and gene flow (*N*<sub>ef</sub>) in the two parental species (Fig. 1), study nuclear admixture (Fig. 2) and the species origin of plastid DNA variants (Fig. 4), and assess genetic relationships and structure within and among populations from both within and outside a large zone of sympatry (Fig. 3; Table 3). Our plastid DNA data reveal that introgression between the two species occurs preferentially via *P. tremula* pollen and *P. alba* seed parents (Fig. 4; Table 4), and our nuclear microsatellite data indicate that almost all hybrids in the studied hybrid zone are recombinant backcrosses towards *P. alba* (Fig. 2). Our data not only inform us about the genetic architecture of the Danube valley hybrid zone, they also contain important information about the factors that may be responsible for maintaining the observed genetic structure.

The role of ecological and life history factors in gene introgression

Effective population sizes (*N*<sub>e</sub>) were large in both species and were larger in *P. tremula* than in *P. alba* (Fig. 1), in agreement with the community ecology of the species. In the Danube valley, *P. tremula* occurs in mixed upland forest communities covering huge geographical areas, whereas *P. alba* is mainly restricted to patchy, fragmented floodplain forests adjacent to the Danube (Krembs 1956; Adler *et al.* 1994; Lazowski 1997). This, together with the fact that *P. tremula* probably recolonized the area earlier than the warmth-loving *P. alba* (Huntley & Birks 1983), may account for larger *N*<sub>e</sub> in the former. Note that the differences in *N*<sub>e</sub> (95% confidence intervals do not overlap; legend of Fig. 1) are unlikely to be artefacts arising from differences in cross-transferability of the microsatellites, since genetic diversity was higher in *P. tremula* for both nuclear microsatellite and plastid DNA markers (Table 2). Large effective population sizes in the two species indicate that selection should be more important in promoting evolutionary...
change than drift. However, the direction of the interspecific difference in $N_e$ leaves us with a puzzle — why is introgression in the direction of the species with the smaller effective population size, $P. alba$? As outlined in the succeeding discussion, a joint genetic structure analysis of nuclear and plastid DNA may inform us about the likely causes for this pattern.

Analysis of molecular variance indicates that a much larger proportion of the plastid DNA variation resides among populations of each parental species compared to nuclear microsatellites (15% vs. 2%, respectively; Table 3). If, for example, only $P. tremula$ populations are considered in genetic divergence analyses, this results in $F_{ST}$ estimates of 0.310 for plastid DNA and 0.017 for nuclear microsatellites. In a finite island model under migration-drift equilibrium, this translates into a pollen/seed flow ratio of 23.9, following equation 5 in Ennos (1994). Hence, pollen/seed flow is similar to that found in other long-lived forest trees with wind-dispersed pollen and seeds (e.g. *Pinus muricata*; pollen/seed ratio c. 24; *Pinus radiata*, pollen/seed ratio c. 31; Strauss et al. 1993), but much lower than in trees with wind-dispersed pollen and heavy animal-dispersed fruits (e.g. *Quercus robur*/*petraea* complex, pollen/seed ratio c. 196; Kremer et al. 1991).

Our results confirm that pollen dispersal in $P. tremula$ is much more efficient than seed dispersal although both are moved by wind, and this has important implications for interpreting patterns of hybridization between the two species.

It has long been known that hybridization between $P. alba$ and $P. tremula$ is tightly connected to the disturbance regime of the floodplain forest (Krembs 1956; Adler et al. 1994; Lazowski 1997; Karrenberg et al. 2002), the habitat ‘mosaic’ in which $P. alba$ also occurs. This is where *P. × canescens* hybrid morphotypes were found and sampled in the present work. In contrast, the preferred upland habitats of $P. tremula$ are separated by at least several kilometres and extend far into the Alps. Hence, considering the low mobility of seeds compared to pollen in $P. tremula$, it makes sense that established hybrids on the floodplain preferentially carry the plastid DNA of their geographically closer parent, $P. alba$ (Fig. 4). Note also that poplars are dioecious (male and female sex expressed on different trees). Hence, chances for establishment of female $F_1$ hybrids with $P. tremula$-derived plastids, capable of passing $P. tremula$ plastid DNA on to $P. alba$ via backcrossing, are reduced even further (50% for a sex ratio of 1:1). The close geographical proximity of $P. alba$ and $P. × canescens$, well documented by ecologists (Krembs 1956; Adler et al. 1994; Lazowski 1997), also presents a simple and plausible explanation for unidirectional introgression as an alternative to ‘intrinsic’ genetic factors. Of course, differences in flowering phenology must be expected to contribute to the observed pattern as well — $P. tremula$ tends to start flowering several weeks earlier than $P. alba$ (Senghas & Seybold 1993), i.e. $P. tremula$ flowers may already be fertilized when $P. alba$ pollen is spread.

Although studying locus-specific effects was not the goal of this study, our nuclear and plastid DNA data lead us to speculate about the role of cytonuclear incompatibilities in gene introgression. In the Danube valley, *Populus × canescens* hybrids preferentially carry plastid DNA haplotypes from $P. alba$ and are characterized by a wide range of nuclear admixture proportions (Table 4). In effect, there is no apparent incompatibility between $P. alba$ plastid DNA and *P. tremula* nuclear genomes. We were not able to test for incompatibilities in the opposite direction, since asymmetric dispersal and location of ‘hybrid’ habitats act like a filter, effectively preventing $P. tremula$ plastid DNA from introgressing into $P. alba$ as outlined earlier. To our knowledge, historical crossing experiments (e.g. Wettstein 1929) do not contain sufficient information to solve this puzzle either, e.g. offspring from laboratory crosses need not be viable under field conditions. Likewise, the fact that introgression is in the same backcross direction (to $P. alba$) in other European river valleys (Rajora & Dancik 1992; Fossati et al. 2004) does not resolve the question either, because ecological and spatial patterns in these places are similar. Nevertheless, studies of North American *Populus* hybrid zones report a lack of cytonuclear incompatibilities in both crossing directions (Paige et al. 1991). This provides weak evidence against a role for cytonuclear incompatibilities in *Populus*, although we caution against premature generalizations.

Regardless of the outcome of this debate, the wide distribution of backcrossed hybrid genotypes (Fig. 2) indicates that the species barrier between $P. alba$ and $P. tremula$ is extremely porous. Continued backcrossing and introgression create a wealth of interspecific recombination products, resulting in increased levels of genetic variability compared to pure $P. alba$ (Table 1) and notoriously large variances in phenotypic characters (Adler et al. 1994; Eckenwalder et al. 1996). Given sufficient heritability in the relevant traits, this may allow hybrids to take advantage of the highly dynamic setting provided by the floodplain forest. To put it in Edgar Anderson’s words, ‘around a thousand different kinds of habitat would be needed to permit the various recombinations to find a niche somewhere’ (Anderson 1948). We argue that the ‘mosaic’ of the floodplain forest provides these habitats, and therefore the ecological opportunity required for the establishment of recombinant hybrid backcrosses.

**Implications for studies of microevolutionary processes**

Hybrid zones are important tools for studying the evolution of reproductive barriers in sympathy, the role of selection in creating or maintaining population/species differences, the role of hybridization in evolution, or the genetic architecture of important phenotypic traits that differ between the hybridizing populations (Barton & Hewitt 1985; Mallet & Barton 1989; Nuernberger et al. 1995; Rieseberg & Buerkle...
2002; Vines et al. 2003). The last topic is particularly relevant for hybridizing taxa with divergent ecological preferences, such as the two Populus species studied here (floodplain pioneer vs. upland species). The studied hybrid zone should be suitable for QTL mapping-type studies of interspecific character differences (sensu Rieseberg & Buerkle 2002), because (i) numerous diagnostic phenotypic characters differ between P. alba and P. tremula (Adler et al. 1994), (ii) numerous ecological differences separate the two species (iii) the two species are strongly differentiated for nuclear microsatellite allele frequencies, and (iv) the proportion of recombinant genotypes in the hybrid zone is high.

Although the wide and continuous distribution of backcrossed genotypes (Fig. 2) speaks against a typical tension zone model (balance between dispersal and selection against hybrids), the question remains as to what extent ecological selection associated with the habitat ‘mosaic’ of the floodplain forest accounts for the persistence of the hybrid zone (Harrison 1990; Barton & Gale 1993; Vines et al. 2003). In the absence of reliable data on historical hybrid zone width, this question may be addressed by studying spatial patterns of hybridization, or associations between hybrid genonomic composition and environmental variables (Vines et al. 2003; Dodd & Afzal-Rafii 2004). Also, the role of clonal reproduction on hybrid zone persistence (Schweitzer et al. 2002) deserves special attention in these species, since clonal propagation may allow P. × canescens hybrids to persist in the hybrid zone even in the absence of sexual recombination.

Finally, an important aspect is that backcrossed (recombinant) hybrids between P. alba and P. tremula are found in several European river valleys (Rajora & Dancík 1992; Fossati et al. 2004). This, combined with the high potential of nuclear microsatellites indicated by the present study, suggests the possibility of comparative genomic studies among several hybrid zones in Europe, equivalent to ‘replicated natural experiments’ across largely independent evolutionary trajectories.

Acknowledgements

We wish to thank Loren Rieseberg and Keith Gardner for numerous helpful discussions during the early phases of the project, Christian Fraissl and Franz Kovacs of the Danube Floodplain National Park, Austria, the Forstverwaltung Lobau of the Vienna City Council, Herbert Tiefenbacher and several other private landowners in Austria for their continued support during sample collection, and Alex Buerkle for helpful discussions about data analyses in hybrid zones. This work was supported by award No. NE/C507037/1 of the Natural Environment Research Council (NERC) in the United Kingdom to CL.

References


© 2005 Blackwell Publishing Ltd, Molecular Ecology, 14, 1045–1057

Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. Genetics, 49, 725–738.
Rogers SM, Campbell D, Baird SJE, Dazmman RG, Bernatchez L (2001) Combining the analyses of introgressive hybridization and linkage mapping to investigate the genetic architecture of
population divergence in the lake whitefish \( (\text{Coregonus clupeaformis}, \text{Mitchell}) \). Genetica, 111, 25–41.


Strauss SH, Hong Y-P, Hipkins VD (1993) High levels of population differentiation for mitochondrial DNA haplotypes in \( \text{Pinus radiata}, \text{muricata}, \text{and attenuata} \). Theoretical and Applied Genetics, 86, 605–611.


This work was carried out within a larger collaboration between the Genetics section at Kew and the Federal Office and Research Centre for Forests in Vienna, Austria. Christian Lexer, the P. I. at Kew, is a population geneticist in Mike Fay’s Genetics section, and Jeffrey Joseph is a research assistant there. The whole Genetics group studies population genetic and cytogenetic aspects of plant evolution with a focus on conservation. Berthold Heinze’s group is interested in the application of genomic tools to the functional characterization, conservation, and breeding of forest trees, and Marius-Sorin Nica was a student in his lab.