Molecular Ecology (2009) **18**, 357–373 doi: 10.1111/j.1365-294X.2008.04016.x

Extensive pollen flow in two ecologically contrasting populations of *Populus trichocarpa*

G. T. SLAVOV,*† S. LEONARDI,‡ J. BURCZYK,§ W. T. ADAMS,¶ S. H. STRAUSS¶ and S. P. DIFAZIO*

Department of Biology, West Virginia University, Morgantown, WV 26506-6057, USA, †Department of Dendrology, University of Forestry, Sofia 1756, Bulgaria, ‡Dipartimento di Scienze Ambientali, Università di Parma, 43100 Parma, Italy, §Department of Genetics, Bydgoszcz University, Bydgoszcz, 85064, Poland, ¶Department of Forest Ecosystems and Society, Oregon State University, Corvallis, OR 97331-5752, USA

Abstract

Pollen-mediated gene flow was measured in two populations of black cottonwood using direct (paternity analysis) and indirect (correlated paternity) methods. The Marchel site was an area with an approximate radius of 250 m in a large continuous stand growing in a mesic habitat in western Oregon. In contrast, the Vinson site was an area with a radius of approximately 10 km and consisted of small, disjunct and isolated stands in the high desert of eastern Oregon. Pollen immigration was extensive in both populations, and was higher in the Marchel site (0.54 ± 0.02) than in the substantially larger and more isolated Vinson site (0.32 ± 0.02). Pollen pool differentiation among mothers was approximately five times stronger in the Vinson population (Φ_{FT} = 0.253, N = 27 mothers) than in the Marchel population (Φ_{FT} = 0.052, N = 5 mothers). Pollen dispersal was modelled using a mixed dispersal curve that incorporated pollen immigration. Predicted pollination frequencies generated based on this curve were substantially more accurate than those based on the widely used exponential power dispersal curve. Male neighbourhood sizes (*sensu* Wright 1946) estimated using paternity analysis and pollen pool differentiation were remarkably similar. They were three to five times smaller in the Vinson population, which reflected the substantial ecological and demographic differences between the two populations. When the same mathematical function was used, applying direct and indirect methods resulted in similar pollen dispersal curves, thus confirming the value of indirect methods as a viable lower-cost alternative to paternity analysis.

Keywords: black cottonwood, correlated paternity, gene flow, microsatellites, parentage analysis, *Populus*

Received 18 August 2008; revision received 8 October 2008; accepted 17 October 2008

Introduction

As a result of increasing habitat fragmentation, predicted climate change and concerns about the biosafety of transgenic organisms, interest in measuring contemporary levels of gene flow and modelling seed and pollen dispersal in forest trees has increased dramatically in recent years (Burczyk *et al*. 2004; DiFazio *et al*. 2004; Smouse & Sork 2004; van Frankenhuyzen & Beardmore 2004; Katul *et al*. 2005; Williams *et al*. 2006; Brunner *et al*. 2007; Farnum *et al*. 2007; Smouse *et al*. 2007). This interest, catalysed by the availability of highly informative microsatellite markers for many ecologically and economically important forest species (Brondani *et al*. 1998; Elsik *et al*. 2000; Slavov *et al*. 2004b; Tuskan *et al*. 2004), has spurred a number of empirical studies of contemporary gene flow (Burczyk *et al*. 2004; Smouse & Sork 2004; Robledo-Arnuncio & Gil 2005; Gérard *et al*. 2006; Goto *et al*. 2006; Bittencourt & Sebbenn 2007) and several significant analytical developments (Smouse *et al*. 2001; Burczyk *et al*. 2002, 2006; Austerlitz *et al*. 2004; Oddou-Muratorio *et al*. 2005; Robledo-Arnuncio *et al*. 2006).

Contemporary gene flow can be measured either directly or indirectly based on molecular marker data. Direct methods include parentage analysis (Jones & Ardren 2003) and the use of mating models incorporating genetic transition probabilities that are calculated by comparing the genotypes

Correspondence: G. T. Slavov, Fax: +1304 2936363; E-mail: gancho.slavov@mail.wvu.edu
S. H. Strauss and S. P. Difazio contributed equally to this work.

of parents and offspring (Adams & Birkes 1991; Burczyk *et al*. 2002, 2006). These methods provide a means to estimate pollen immigration, correlates of reproductive success (e.g. intermate distance, phenological overlap, fecundity, positions relative to predominant wind direction) and the effective number of male parents in the genetic neighbourhood of a female tree. Indirect methods combine features of traditional population structure analysis and paternity analysis to measure the degree of differentiation among the pollen pools of females sampled across the landscape (Smouse *et al*. 2001). This measure can then be converted into an estimate of genetic neighbourhood size (Smouse *et al*. 2001) and, when calculated for pairs of females separated by different distances, used to estimate a pollen dispersal curve (i.e. a mathematical relationship describing the probability of pollination as a function of distance between mates; Austerlitz *et al*. 2004; Robledo-Arnuncio *et al*. 2006).

Each of these groups of methods has certain advantages and shortcomings (reviewed by DiFazio *et al*. 2004; Smouse & Sork 2004). Briefly, direct methods require genotyping a large number of parents and offspring for at least half a dozen highly variable markers. Thus, these methods tend to be expensive and logistically difficult, and inferences are typically limited to the short-distance component of the pollen dispersal curve in individual populations and years. In contrast, indirect methods are time- and cost-efficient but have their own shortcomings. Pollen pool differentiation could potentially be confounded with population structure of adults and/or their relatedness (Austerlitz & Smouse 2001b), and the precision of dispersal-curve parameters estimated using these approaches is not high (Robledo-Arnuncio *et al*. 2006). Empirical studies employing both approaches could help determine what can and what cannot be learned using the more efficient indirect methods.

The genus *Populus* is distributed throughout the northern hemisphere and consists of approximately 30 species of forest trees that are divided into six sections (Eckenwalder 1996). Most species are primarily dioecious, wind-pollinated and wind-dispersed and under favourable conditions reach reproductive maturity within 5–10 years. *Populus* is facultatively clonal, but the relative importance of vegetative propagation and sexual reproduction varies considerably among species (Braatne *et al*. 1996; Schweitzer *et al*. 2002; Rood *et al*. 2003, 2007). A number of species and hybrid cultivars are economically important for production of lumber and pulp and for applications in horticulture and phytoremediation (Balatinecz & Kretschmann 2001; Doty 2008). There is also substantial interest in using *Populus* for production of biofuel feedstocks (Rubin 2008). Furthermore, some *Populus* species are ecologically dominant members of riparian communities, providing essential habitat for a wide range of organisms in arid regions (Whitham *et al*. 1999). *Populus* has been demonstrated to be a primary driver of the diversity and structure of a number of dependent communities (Whitham *et al*. 2006) and therefore can be considered a 'foundation species' with a disproportionately strong influence on associated ecosystems (Ellison *et al*. 2005).

Populus has many characteristics that make it an excellent system for molecular genetic investigations, including convenient vegetative propagation from root and stem cuttings, amenability to tissue culture, rapid growth and a relatively short juvenile period (Bradshaw *et al*. 2000; Taylor 2002). Furthermore, a well-developed molecular toolbox is available for *Populus*, including an efficient transformation system (Busov *et al*. 2005), dense genetic maps (Cervera *et al*. 2001; Yin *et al*. 2004), abundant expressed sequence tags (ESTs) (Sterky *et al*. 2004), BAC libraries (Stirling *et al*. 2003; Lescot *et al*. 2004), physical maps (Kelleher *et al*. 2007) and a whole genome sequence (Tuskan *et al*. 2006). Therefore, *Populus* is now widely recognised as one of the preeminent model trees for basic molecular biology investigations and a primary target for accelerated domestication through association studies and genetic engineering (Bhalerao *et al*. 2003; Boerjan 2005).

Extensive information about contemporary pollen flow in *Populus* would help manage and conserve the genetic resources of the genus, perform science-based risk assessment of transgene flow and design optimal genetic association studies. This information, however, is currently limited. Early studies focussed on characterising the population genetic structure of *Populus* species using allozyme markers. The weak inter-population differentiation revealed in these studies provided indirect evidence for substantial levels of gene flow (e.g. Weber & Stettler 1981; Jelinski & Cheliak 1992; Legionnet & Lefèvre 1996). More recently, allele frequencies for both allozyme and DNA-based markers were used to directly study natural hybridisation between *Populus fremontii* and *P. angustifolia* (Martinsen *et al*. 2001) and *P. alba* and *P. tremula* (Lexer *et al*. 2005), as well as to estimate levels of gene flow in populations of *P. nigra* in Europe, with particular interest in quantifying the introgression of alleles from the North-American *P. deltoides* and its hybrids introduced into Europe (Heinze 1998; Benetka *et al*. 1999; Fossati *et al*. 2003; Imbert & Lefèvre 2003; Vanden Broeck *et al*. 2004). Although these studies revealed important patterns of hybridisation and introgression, they provided little information about contemporary levels of pollen flow in *Populus*. To our knowledge, there have been three paternity analysis studies in *Populus*, all of which were designed to provide direct estimates of introgression of *P. deltoides* alleles into populations of *P. nigra* (Tabbener & Cottrell 2003; Pošpísková & Šálková 2006; Vanden Broeck *et al*. 2006) rather than to explicitly measure and model effective pollen dispersal.

In this study, we used both direct and indirect methods to characterise contemporary pollen-mediated gene flow in two ecologically and demographically contrasting populations of black cottonwood (*Populus trichocarpa* Torr.

& Gray). We performed paternity analysis at a large spatial scale (> 300 km²) and developed and applied a new maximum-likelihood approach for modelling pollen dispersal. This approach incorporates the estimated level of pollen immigration into the pollen dispersal curve and generates more realistic predictions than those commonly used in the literature. In addition, we compared pollen flow parameters calculated based on paternity analyses to those obtained through analyses of pollen pool differentiation and discussed the practical implications of our findings.

Materials and methods

Study sites and plant materials

Our two study sites, Marchel and Vinson, differ substantially in their ecological and demographic characteristics (Fig. 1). The Marchel site is located near Corvallis, Oregon, in the Willamette Valley, west of the Cascade Mountains. The climate at this site is characterised by dry summers and mild, wet winters, with annual precipitation of 1085 mm, and 1122 degree-days with temperatures of at least 10 °C. The study site was an area with an approximate radius of 250 m and included a large, continuous *Populus trichocarpa* stand along the Willamette River, as well as scattered patches of trees in an abandoned gravel mine (Fig. 2). The study site also included a 2.5-ha plantation, which consisted primarily of *P. trichocarpa* × *P. deltoides* hybrids and was established seven years before seeds were collected for this study. The plantation contained 15 different flowering male genotypes, each represented by four contiguous ramets, and one male genotype represented by a total of 74 ramets in four separate locations, all at 2.1 m × 3 m spacing (Hibbs *et al.* 2003; Fig. 2). We collected leaf tissue from 285 reproductively

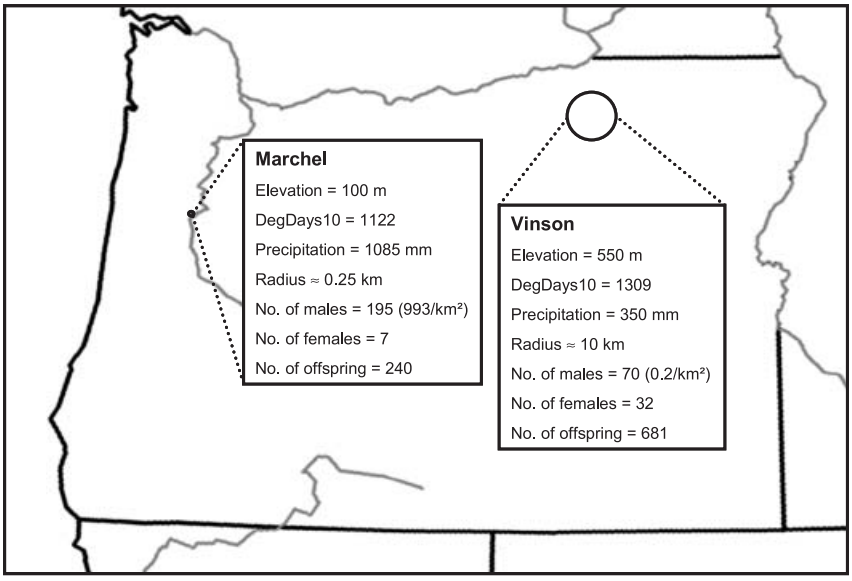

Fig. 1 Study sites: two *Populus trichocarpa* populations with contrasting ecological and demographic characteristics (not to scale). Major rivers are shown in light shading. See Figs 2–3 for detailed schematic diagrams of the two sites.

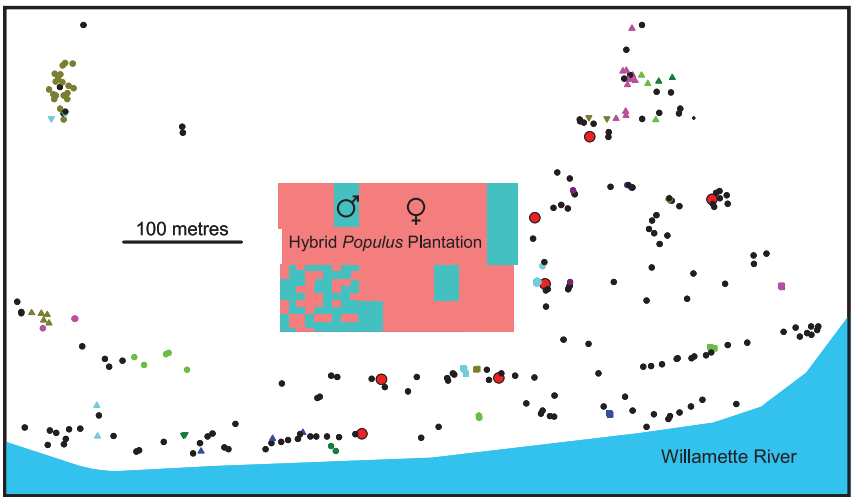

Fig. 2 Marchel site. Locations of plantation trees and wild male *Populus trichocarpa* trees. Black circles are male genets with single ramets. Coloured symbols represent male genets with multiple ramets. Large red circles are locations of sampled mother trees. Blue areas in the plantation are blocks of male trees, and pink areas are blocks of female trees.

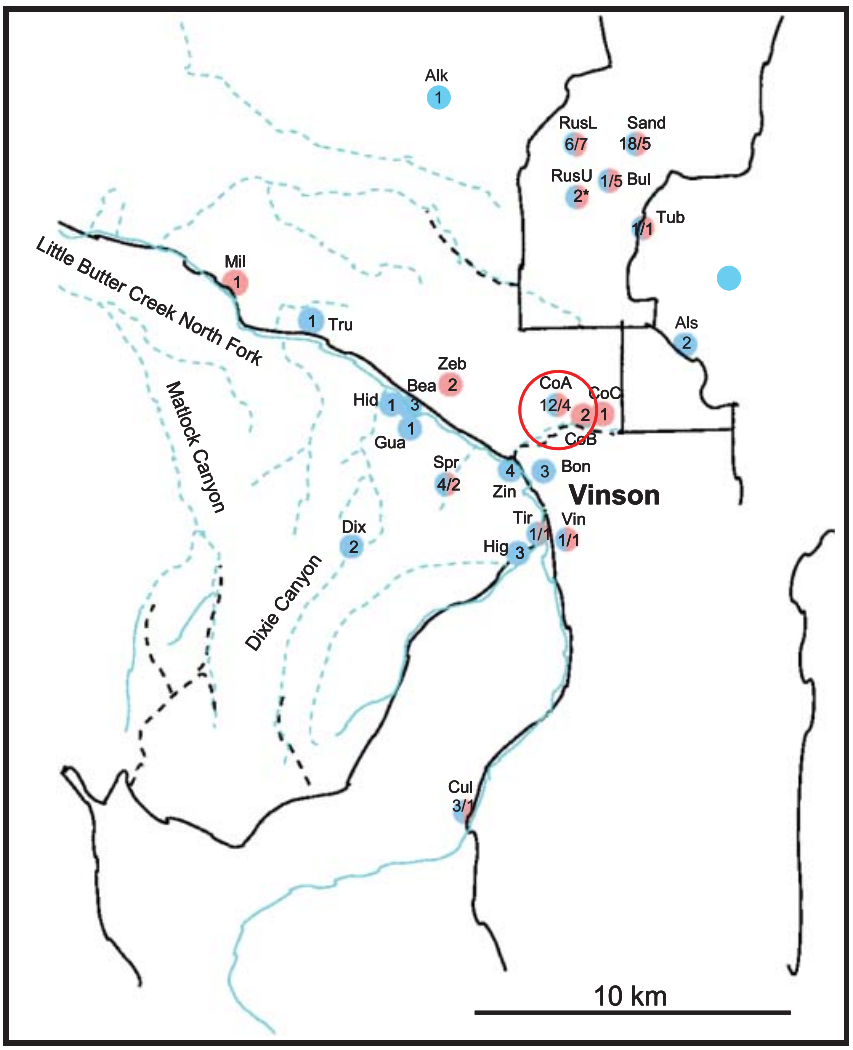

Fig. 3 Vinson site. Locations of groups of male (blue) and female (pink) *Populus trichocarpa* trees and numbers of genets sampled in each group (numbers in coloured ovals). Data from mother trees in groups CoA and CoB (delimited by a red circle) were used for analysis of pollen pool differentiation at a smaller spatial scale (see text).

mature male cottonwood trees in the natural stands, as well as from all male genotypes in the plantation. After accounting for clonality based on microsatellite genotypes (see below), gender and phenology, the number of unique male genets (excluding hybrids) was determined to be 195. Although the high density of this population made it unfeasible to sample every single stem, we are confident that we sampled all major sources of pollen (i.e. unsampled male trees were most likely ramets of genets that were sampled). We also sampled spatially interspersed female mother trees ($N = 7$), with an average pairwise distance among mothers of 173 m. We germinated and harvested young leaves from 240 seeds collected from the selected mother trees (mean = 34.3, range = 4–48 seeds per mother; Table S1, Supporting information).

The Vinson site is located approximately 50 km southwest of Pendleton, Oregon, in the high desert east of the Cascade Mountains. The annual precipitation at this site is 350 mm (more than three times lower than at the Marchel site), and the number of degree-days with temperatures of at least 10 °C is 1309 (approximately 17% higher than in the Marchel site). The study site was an area with a radius of approximately 10 km and consisted of small, disjunct stands, which occurred primarily along active and ephemeral creek beds, as is typical of this region (Fig. 3). Inspection of aerial photographs (1 : 72 000, Farm Service Agency, Heppner, Oregon) of the entire region indicated that no sources of *P. trichocarpa* pollen were located within 5 km of the borders of the sampled area. The arid landscape of eastern Oregon allowed us to successfully identify even single cottonwood trees on these aerial photographs. We collected leaf tissue from all 222 reproductively mature male ramets and one hermaphroditic clone with five mature ramets. After accounting for clonality, the number of male or hermaphroditic genets was 70. We sampled 32 female mother trees, with an average pairwise distance among sampled females of 6 km. We germinated and harvested young leaves from 681 seeds collected from the

Table 1 Number of alleles, heterozygosities and frequencies of null alleles for 10 microsatellite loci in two *Populus trichocarpa* populations. All statistics were calculated using version 3.0 of the Cervus program (Kalinowski *et al*. 2007)

	Population									
	Marchel					Vinson				
Locus*	*k*†	*N*‡	H_O§	H_E¶	f_0**	*k*	*N*	H_O	H_E	f_0
TGRC_AG1	26	172	0.669	0.868	0.132	25	84	0.786	0.905	0.068
PMGC_14	12	197	0.797	0.772	N/A††	10	99	0.515	0.586	0.079
PMGC_2011	17	164	0.835	0.863	0.017	12	88	0.682	0.753	0.049
PMGC_2156	21	183	0.672	0.856	0.121	16	93	0.624	0.835	0.144
PMGC_2235	24	197	0.848	0.917	0.037	17	96	0.885	0.829	N/A
PMGC_420	17	200	0.825	0.871	0.027	13	98	0.806	0.819	0.002
PMGC_433	17	190	0.953	0.911	N/A	17	83	0.928	0.922	N/A
PMGC_576	36	197	0.904	0.914	0.004	20	91	0.912	0.912	N/A
PMGC_684	21	198	0.803	0.910	0.063	16	86	0.477	0.650	0.153
TGRC_WOU1	7	200	0.080	0.078	N/A	5	97	0.134	0.251	0.323
Mean	**20**	**190**	**0.739**	**0.796**	**0.040**	**15**	**92**	**0.675**	**0.746**	**0.082**

*Primer sequences are available at http://www.ornl.gov/sci/ipgc/ssr_resource.htm.
†*k* is the observed number of alleles.
‡*N* is the number of genets genotyped.
§H_O is the observed heterozygosity.
¶H_E is the unbiased estimate of gene diversity (Nei & Roychoudhury 1974).
**f_0 is the frequency of null alleles estimated based on deviations from Hardy–Weinberg Equilibrium (Summers & Amos 1997).
††f_0 was not estimated because $H_O > H_E$.

selected mother trees (mean = 21, range = 1–49 seeds per mother; Table S1, Supporting information).

Relative tree locations at both sites were mapped using an ultrasonic device (Sonin 250, Sonin Inc., Charlotte, North Carolina) and a liquid-filled precision compass.

DNA isolation and microsatellite genotyping

DNA was extracted from leaf samples using a streamlined Sodium Dodecyl Sulfate (SDS) protocol. Briefly, frozen leaf tissue in liquid nitrogen was mechanically disrupted in microcentrifuge tubes using a drill bit mortar, treated with RNAase A, incubated in SDS, and finally extracted with phenol:chloroform:IAA, followed by ethanol precipitation (see http://www.as.wvu.edu/~sdifazio/protocols/streamlined_dna_extract.pdf for details).

Microsatellite loci with di- and tri-nucleotide repeat motifs were identified from published genomic and cDNA sequences downloaded from GenBank and sequences derived in-house in a separate project (Brunner *et al*. 2000). In addition, we sequenced clones from a genomic library developed at the University of Washington by the Poplar Molecular Genetics Cooperative (PMGC) and designed microsatellite primers using the Primer software (version 0.5, Whitehead Institute, Cambridge, Massachusetts). We amplified microsatellites using the following reaction conditions: 6 ng genomic DNA, 1 × Gibco BRL (currently, Invitrogen, Carlsbad, California) polymerase chain reaction (PCR) buffer, 1.3 mg/ml bovine serum albumin, 0.1 mM dNTP, 0.25 μM forward and reverse primers, a locus-specific concentration of $MgCl_2$ ranging from 2.0 to 3.5 mM, and 0.5 units of *Taq* DNA Polymarase (Invitrogen). Our PCR temperature profile began with 94 °C for 4 min; proceeded with 39 cycles of melting at 94 °C for 15 s, annealing at a locus-specific temperature ranging from 50 to 56 °C for 15 s, and extension at 72 °C for 15 s; and finished with an extension at 72 °C for 10 min.

For loci showing clear, consistent amplification of a single polymorphic locus, we purchased primers end-labelled with a FAM, HEX or TAMRA fluorescent dye [Applied Biosystems (ABI), Foster City, California]. Because there was good separation between the size ranges of some of the loci, we were able to multiplex five reactions per gel lane. Fragments were analysed using an ABI 377 automated sequencer (ABI) with a GeneScan ROX 500 internal size standard (ABI), or a standard derived from phage PhiX174 and labelled with Texas Red fluorescently-labelled nucleotides (http://www.as.wvu.edu/~sdifazio/protocols/Fluor_std.htm).

Alleles were sized using the GeneScan software (ABI), scored using the Genotyper software (ABI) and binned using custom PERL scripts, followed by manual adjustment of bin boundaries. We selected the ten most polymorphic and consistently amplifying loci (Table 1) for use in paternity analysis.

Data analysis

Paternity analysis. By comparing the genotypes of offspring to those of their respective mothers, we obtained an overall estimate of mistyping (i.e. including null alleles and genotyping errors) of 5.6% (DiFazio 2002). Paternities were therefore assigned and pollen immigration was estimated using genotypic exclusion allowing mismatching loci between offspring and candidate fathers in order to avoid false exclusions caused by mistyping (Wilmer *et al*. 1999; Cercueil *et al*. 2002; Slavov *et al*. 2005a, b; Vandeputte *et al*. 2006). Although other, potentially more efficient analytical approaches have been developed (e.g. mating models; Adams & Birkes 1991; Burczyk *et al*. 2002, 2006), the performance of these methods when rates of mistyping are non-negligible has not been tested extensively. We are currently developing extensions of existing mating models that will allow pollen flow and correlates of mating success to be estimated reliably even when rates of genotyping error are moderately high (Burczyk *et al.* unpublished data). For the purposes of this study, however, we chose to use paternity exclusion because with as few as 7–8 highly variable microsatellite markers, this approach: (i) allows one to effectively compensate for null alleles and genotyping error, while preserving reasonably high paternity assignment power (Slavov *et al*. 2005a; Vandeputte *et al*. 2006); and (ii) generates estimates of pollen immigration with little or no bias (Slavov *et al*. 2005b).

To increase the power of paternity exclusion, we allowed mismatching loci between offspring and candidate fathers only under certain conditions. A 'conditional mismatch' at a given locus was allowed when: (i) an offspring and a candidate father had alleles that differed by one microsatellite repeat unit (i.e. 2 or 3 bp); (ii) both the offspring and the candidate father appeared homozygous; or (iii) either one was homozygous and the other had an allele with a size that exceeded a locus-specific threshold. The first type of mismatch can occur as a result of single-step mutations or inaccurate allele sizing, the second type is expected to occur because of null alleles, and the third can result from allele drop-outs (i.e. preferential allele amplification caused by mutations in the priming site or sub-optimal DNA template quality), a problem that is exacerbated when there is a large discrepancy in allele sizes (Ewen *et al*. 2000). Paternity was assigned to a candidate father if its genotype: (i) had no more than three conditional mismatches with the genotype of the offspring analysed; (ii) had no other types of mismatches with the offspring genotype; (iii) was compared to the offspring genotype for at least five loci; and (iv) was the only paternal genotype for which the previous three criteria were met. Allowing up to four or five conditional mismatches had little effect on the final estimates of pollen immigration but increased the number of ambiguous paternities (i.e. offspring with two or more compatible fathers). Therefore, assignments based on up to three conditional mismatches were used in further analyses.

Following the approach developed by Smith & Adams (1983), pollen immigration (m) was estimated by dividing the observed proportion of immigrants (b, the proportion of offspring for which no compatible father is identified, which does not include offspring for which multiple compatible fathers are identified) by the detection probability (d, the probability of distinguishing an immigrant pollen gamete from those that could be produced by any of the candidate fathers). Detection probabilities were estimated using the Monte Carlo approach described by Slavov *et al*. (2005b). Briefly, a large number (i.e. > 10 000) of offspring resulting from pollen immigration were simulated based on the genotypes of the mother trees and allele frequencies estimated from the paternal haplotypes of offspring identified as pollen immigrants. The frequency and pattern of missing data in these simulated genotypes were identical to those in the observed offspring genotypes. The simulated genotypes were then analysed using paternity exclusion and detection probabilities estimated as the proportion of offspring classified as immigrants based on allowing the respective number of conditional mismatches. The standard error of pollen immigration estimates was calculated as:

$$\mathrm{SE}(m) \approx \frac{1}{d}\sqrt{\frac{b(1-b)}{n}} \qquad \text{(eqn 1)}$$

where n is the number of offspring analysed (Smith & Adams 1983). A simulation study, which assessed a wide range of parameter combinations, including scenarios similar to those in the Marchel and Vinson sites, revealed that this analytical approach results in estimates of pollen immigration with little or no bias and low variance when detection probabilities are relatively high (i.e. > 0.8) and a sufficient number of mismatching loci (2–3) has been allowed (Slavov *et al*. 2005b).

Pollen pool differentiation. The degree of genetic heterogeneity among the pollen clouds sampled by females in each of the two populations was assessed using the TwoGener approach (Smouse *et al*. 2001). This approach combines features of analysis of molecular variance (AMOVA) and paternity analysis to estimate a single parameter, Φ_{FT}, the proportion of variation in the global pollen pool, which is accounted for by differences among the local pollen pools sampled by individual mothers. Similar to other *F*-statistics, Φ_{FT} can be defined as: (i) a correlation coefficient: the intraclass correlation of pollen gametes drawn from a single mother; and (ii) a probability of identity by descent (IBD): the probability of IBD for two gametes drawn at random from the pollen pool of a single mother relative to the probability of IBD for two gametes drawn at random from the global pollen pool of all mothers (Austerlitz &

Smouse 2001a). Global estimates of Φ_{FT} for each population and pairwise Φ_{FT} estimates for each pair of mothers from the same population were obtained using version 6.0 of the GenAlEx software (Peakall & Smouse 2006) and were based only on data from mothers for which at least 10 offspring had been sampled. The statistical significance of global Φ_{FT} estimates was tested based on 999 permutations of pollen gametes among females using GenAlEx. To account for adult population structure, we used version 3.0 of the Arlequin program (Excoffier *et al*. 2005) to estimate the fixation index F_{IS} (Wright 1965) based on the multilocus genotypes of adult male trees and then divided Φ_{FT} estimates by $(1 + F_{IS})$, following equation 11 of Austerlitz & Smouse (2001b).

Pollen dispersal curves estimated based on paternity analyses. Under random mating, the frequency of mating events among trees separated by a certain distance is expected to be equal to the proportion of potential mating pairs separated by that distance. We evaluated the extent to which mating in each of the two study populations followed this expectation using goodness-of-fit chi-square tests (Wackerly *et al*. 2002) after grouping observed mating events and all pairwise distances among female and male trees into discrete distance classes. Distance classes were formed so that the expected number of mating events in each class was at least five.

In our preliminary analyses, fitting exponential, exponential power and Weibull functions resulted in poor prediction of either the local component (i.e. up to 500–600 m) or the long-distance component of the dispersal curve. We therefore chose to model effective pollen dispersal as a two-component process, with local dispersal described by an exponential distribution and long-distance dispersal described by a uniform distribution, which incorporates the estimated level of pollen immigration. Extending the approach of Streiff *et al*. (1999), who fitted an exponential distribution to the pollination data and then determined the parameters of the uniform distribution by truncating the exponential distribution near the edge of the sampled area, we estimated all parameters of the mixed distribution jointly using maximum likelihood (see below). Unlike Streiff *et al*. (1999), we did not truncate the exponential distribution and modelled long-distance dispersal using a uniform distribution with a lower bound of zero and an upper bound estimated from the empirical data. The biological basis for this approach is that locally dispersed pollen grains are subject to local air flows and eddies and follow predictable patterns of dispersal in which probability of deposition declines exponentially with distance from the source (e.g. Levin & Kerster 1974; Di Giovanni & Beckett 1990), whereas a portion of the pollen is caught in updrafts, potentially travelling great distances (e.g. Lanner 1965; Di Giovanni *et al*. 1996).

The one-dimensional pollen dispersal function is:

$$f(x) = \alpha\left(\frac{1}{\beta}\exp\left(-\frac{x}{\beta}\right)\right) + (1-\alpha)\left(\frac{1}{\gamma}\right)I_{[0,\gamma]} \qquad \text{(eqn 2)}^1$$

where x (m) is the intermate distance, α is a mixing parameter ($0 \le \alpha \le 1$), β (m) is the parameter defining the exponential distribution ($\beta > 0$), γ (m) is the upper-bound parameter of the uniform distribution defined from zero to γ ($\gamma > 0$), and $I_{[0,\gamma]}$ is an indicator variable, which takes a value of one for $0 \le x \le \gamma$, and zero otherwise. The mean dispersal distance based on this function is:

$$\delta = \alpha\beta + (1-\alpha)\frac{\gamma}{2} \qquad \text{(eqn 3)}$$

If sampled female trees are treated as pollen sinks, the joint likelihood of observing $y_1, y_2, \ldots, y_k$ pollinations in distance classes 1, 2, 3, ... , k can be expressed using the multinomial probability distribution:

$$L(\alpha,\beta,\lambda) = \frac{n!}{y_1!y_2!y_3!\ldots y_k!}p_1^{y_1}p_2^{y_2}p_3^{y_3}\ldots p_k^{y_k} \qquad \text{(eqn 4)}$$

where n is the total number of offspring analysed and $p_1, p_2, p_3 \ldots, p_k$ are the definite integrals of the dispersal function (eqn 2) between the lower and upper bounds of the respective distance classes. The last distance class was allocated to pollinations by unsampled male trees located beyond the borders of the study sites (i.e. pollen immigration). The unknown upper bound of this distance class (γ) was estimated along with the other two function parameters (α and β) by maximising the logarithm of the joint likelihood (eqn 4) using the Mead–Nelder simplex algorithm and quasi-Newton algorithms as implemented in the optimisation toolbox of Matlab (version 7.4, MathWorks Inc., Natick, Massachusetts). The number of distance classes and their bounds were varied in order to asses the extent to which this factor influenced the estimated function parameters (Table S2, Supporting information). Predicted pollination frequencies based on the estimated parameters were generated by integrating the dispersal function (eqn 2) between the lower and upper bound of each distance class and goodness-of-fit tests were performed as described above.

To allow comparisons with dispersal curves fitted using patterns of correlated paternity (see below), a similar procedure was used to fit a dispersal curve from the exponential power family:

[1]Correction added after online publication 24 December 2008: eqn 2 was changed from $f(x) = \alpha\left(\frac{1}{\beta}\exp\left(\frac{x}{\beta}\right)\right) + (1-\alpha)\left(\frac{1}{\gamma}\right)I_{[0,\gamma]}$ to $f(x) = \alpha\left(\frac{1}{\beta}\exp\left(-\frac{x}{\beta}\right)\right) + (1-\alpha)\left(\frac{1}{\gamma}\right)I_{[0,\gamma]}$.

$$f(x,y)=\frac{b}{2\pi a^2\Gamma(2/b)}\exp\left[-\left(\frac{\sqrt{x^2+y^2}}{a}\right)^b\right] \quad \text{(eqn 5)}$$

where $\sqrt{x^2+y^2}$ is the dispersal distance, a is a scale parameter, b is a shape parameter, Γ is the gamma function (Wackerly *et al*. 2002) and the mean dispersal distance is:

$$\delta=\frac{a\Gamma(3/b)}{\Gamma(2/b)} \quad \text{(eqn 6)}$$

Because of its shape flexibility (fat-tailed when $b < 1$ and thin-tailed when $b > 1$), this curve has been used in several recent studies of seed and pollen dispersal (Clark *et al*. 1998; Austerlitz *et al*. 2004; Oddou-Muratorio *et al*. 2005; Robledo-Arnuncio *et al*. 2006; Robledo-Arnuncio & García 2007). Substituting $r=\sqrt{x^2+y^2}$, assuming isotropic pollen dispersal, and using equation A.5 derived by Tufto *et al*. (1997), we transformed the bivariate dispersal curve given in eqn 5 to a one-dimensional dispersal function:

$$f(r)=\frac{br}{a^2\Gamma(2/b)}\exp\left[-\left(\frac{r}{a}\right)^b\right] \quad \text{(eqn 7)}$$

This function was used to estimate the distribution parameters a and b by maximizing the likelihood function (eqn 4) as described above. The probability of observing intermate distances in the last distance class (i.e. pollen immigration) was expressed as:

$$p_k=1-\sum_{i=1}^{k-1}p_i \quad \text{(eqn 8)}$$

Predicted pollination frequencies based on the estimated parameters of the exponential power dispersal function were generated using Monte Carlo integration (Ross 1996) based on 10^6 pseudorandom numbers per interval and goodness-of-fit tests were performed as described above.

Pollen dispersal curves estimated from correlated paternity. The scale (a) and shape (b) parameters of the exponential power dispersal curve given in eqn 5 were estimated based on measures of correlated paternity among the progeny arrays of pairs of sampled female trees (Robledo-Arnuncio *et al*. 2006). This approach is conceptually similar to using pairwise Φ_{FT} values among females to fit pollen dispersal functions (Austerlitz *et al*. 2004) but is superior because: (i) unlike the TwoGener-based approach, it does not depend on the effective density of males; and (ii) it yields more accurate and precise estimates of the parameters of dispersal functions from the exponential power family (Robledo-Arnuncio *et al*. 2006). Dispersal function parameters were estimated based on data for females from which at least 10 offspring had been genotyped using the Kindist program within the Poldisp software package (Robledo-Arnuncio *et al*. 2007).

Table 2 Estimates of pollen immigration in two *Populus trichocarpa* populations with contrasting ecological and demographic characteristics

Population	*R**	*n*†	*b*‡	*d*§	*m*¶	SE(*m*)**
Marchel	0.25	240	0.46	0.85	0.54	0.02
Vinson	10	681	0.27	0.84	0.32	0.02

**R* is the approximate radius (km) of the sampled area.
†*n* is the number of offspring analysed.
‡*b* is the observed proportion of pollen immigrants based on exclusion allowing up to three conditional mismatches (see Materials and methods).
§*d* is the detection probability based on exclusion allowing up to three conditional mismatches (see Materials and methods).
¶*m* is the estimate of pollen immigration adjusted for cryptic gene flow (*b*/*d*).
**SE(*m*) is the standard error of pollen immigration estimates (eqn 1).

Number of breeding males and neighbourhood size. The size of the breeding male population (N_{bmp}) was estimated following the fractional paternity assignment approach of Nielsen *et al*. (2001). A maximum likelihood estimate of N_{bmp} was obtained by maximising the likelihood given in equation 7 of Nielsen *et al*. (2001), and 95% likelihood-based confidence intervals (Neale & Miller 1997) were calculated using Matlab. The effective number of fathers in the neighbourhood of each female (N_{ep}), which is analogous to Wright's neighbourhood size (Wright 1946), was estimated directly from the paternity analysis data [i.e. as the inverse of the squared relative reproductive success summed over all fathers (Burczyk *et al*. 2002)] and from the analysis of pollen pool differentiation (i.e. as $1/(2\Phi_{FT})$; Austerlitz & Smouse 2001a; Smouse *et al*. 2001).

Results

Pollen immigration

The number and variability of the microsatellite loci we used (Table 1) allowed us to apply a conservative adjustment for biases introduced by null alleles and genotyping error, while preserving relatively high detection probabilities (i.e. limiting the extent of cryptic gene flow) in both populations (Table 2). As expected, pollen immigration was extensive in both populations, and was higher in the Marchel population in which a substantially smaller area (and therefore a smaller proportion of the potential fathers) was sampled (Table 2). Allowing more than three mismatching loci between offspring and candidate fathers did not result in substantial changes of pollen immigration estimates.

Table 3 Pollen pool differentiation among females in two *Populus trichocarpa* populations with contrasting ecological and demographic characteristics

Population	n*	N†	D_{am}‡	Φ_{FT}§	F_{IS}¶	Φ'_{FT}**
Marchel (whole area)	230	5	180	0.052 (< 0.001)	0.031	0.050
Vinson (whole area)	661	27	5514	0.253 (< 0.001)	0.084	0.233
Vinson (small scale)[2]	231	6	518	0.111 (< 0.001)	0.084	0.102

*n is the number of offspring analysed.
†N is the number of mother trees from which offspring included in the analysis had been sampled (see Materials and methods).
‡D_{am} is the average pairwise distance among mother trees (m).
§Φ_{FT} is the degree of pollen pool differentiation among mothers (Smouse *et al*. 2001). One-sided P-values from permutation tests of the significance of Φ_{FT} are shown in parentheses.
¶F_{IS} is the fixation index (Wright 1965) estimated from the genotypes of adult male trees in the respective population.
**Φ'_{FT} is the degree of pollen pool differentiation among mothers after accounting for population structure among adult male trees (see Materials and methods).

When up to five conditional mismatches were allowed, for example, pollen immigration estimates were $m = 0.51 \pm 0.02$ for the Marchel population and $m = 0.31 \pm 0.02$ for the Vinson population (i.e. within 3–5% of the estimates reported in Table 2). Observed pollen immigration, calculated for individual mothers for which more than 10 seedlings were analysed, was negatively correlated with the number of reproductively mature male trees located in close proximity to these mothers. In the Vinson population, this correlation was moderate and statistically significant for the number of male trees within 100 m of sampled mother trees ($r = -0.45$, $P = 0.018$, $N = 27$) and somewhat stronger for the number of male trees within 500 m ($r = -0.62$, $P = 0.0005$, $N = 27$). In the Marchel population, the correlation for the number of male trees within 100 m of sampled mother trees was strong, but not statistically significant ($r = -0.82$, $P = 0.091$, $N = 5$). Because of the small size of the area sampled and the high density of males in the Marchel population, the number of males within circles with radii greater than 100 m did not vary substantially across mothers. Therefore, we were not able to determine whether male density at larger spatial scales was correlated with pollen immigration in mother trees sampled from this population.

Pollen pool differentiation

In both populations, the differentiation among the pollen pools of female trees was statistically significant (Table 3). The extent of pollen pool differentiation among mothers was approximately five times stronger in the Vinson population ($\Phi_{FT} = 0.253$) than in the Marchel population ($\Phi_{FT} = 0.052$). Analysis of offspring from mother trees located within a more restricted area near the centre of the Vinson population (Fig. 3) still revealed pollen pool differentiation approximately two times stronger ($\Phi_{FT} = 0.111$) than in the Marchel population (Table 3). Accounting for population structure among males did not result in substantial changes of Φ_{FT} estimates in either population (Table 3). We hypothesised that differentiation among pollen pools was caused by excessively frequent pollination by a small number of males in close proximity to the sampled mother trees. We therefore recalculated pollen pool differentiation after excluding all offspring that had paternity assigned to males within 500 meters. Pollen pool differentiation decreased more than twelve-fold in the small-scale Vinson sample ($\Phi_{FT} = 0.009$, $P = 0.050$) but not in the Marchel sample ($\Phi_{FT} = 0.064$, $P < 0.001$).

Pollen dispersal curve

Paternity analysis. The proportion of pairwise distances among males and females in a given distance class was a good predictor of within-population pollination frequency in the continuous, high-density Marchel population (Fig. 4A) but not in the disjunct Vinson population (Fig. 4B). Fitting exponential, exponential power and Weibull functions resulted in poor prediction of the local component (i.e. up to 500–600 m) and particularly of the long-distance component of the dispersal curve (see below). In contrast, the three-parameter mixed dispersal function we used generated reasonable predictions for both local and long-distance pollinations, including pollen immigration (Fig. 5A, B). The dispersal function parameters ($\alpha = 0.536$, $\beta = 195$ m, $\gamma = 32\,306$ m) could only be estimated for the Vinson population for which we had extensive long-distance pollination data (Fig. 5B). Based on these parameter values, the estimated mean pollination distance was $\delta = 7599$ m. Remarkably, when these parameter estimates were used for the Marchel population, the dispersal curve predictions were generally concordant with the observed frequencies of local mating and pollen immigration (Fig. 5A). Varying the number of distance classes (from as few as nine to as many as 37) did not result in substantial changes in

[2]Correction added after online publication 24 December 2008: Vinson (small scale) was moved to the third row in Table 3.

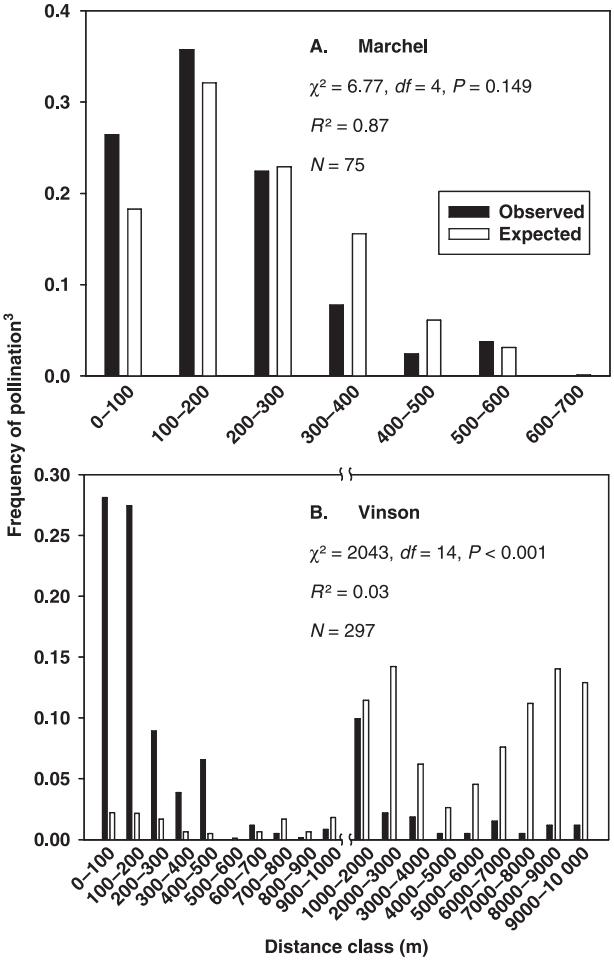

Fig. 4 Observed vs. expected frequency of within-population pollinations in different distance classes in two *Populus trichocarpa* populations. Expected frequencies were based on the proportion of potential intermate distances in each class. Goodness-of-fit tests were based on distance classes pooled to ensure that the expected number of pollinations in each class was at least five. *N* is the number of offspring for which unique paternity was assigned.

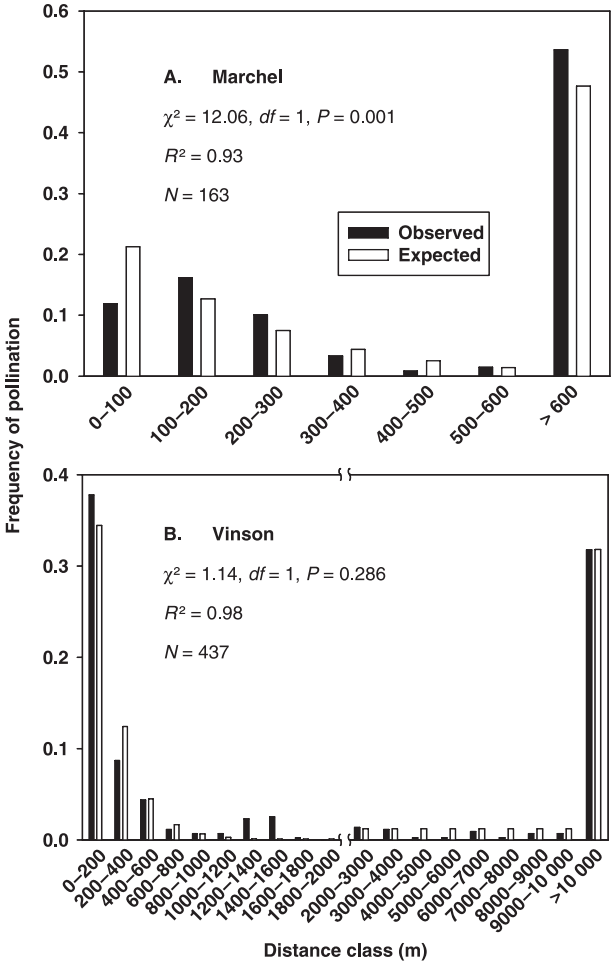

Fig. 5 Observed vs. expected pollination frequencies based on predictions from a mixed probability density function, whose parameters were estimated based on paternity analysis data. Parameter estimates used to generate predictions for both populations ($\alpha = 0.536$, $\beta = 195$ m, $\gamma = 32\,306$ m) were obtained based on the data from the Vinson population (see text). The mean dispersal distance based on these parameter estimates was $\delta = 7599$ m. Goodness-of-fit tests were based on distance classes pooled to ensure that the expected number of pollinations in each class was at least five. *N* is the number of offspring for which unique paternity was assigned plus the number offspring resulting from pollen immigration (calculated proportionally based on the estimates in Table 2).

the estimated function parameters (Table S2, Supporting information).

Correlated paternity. Pairwise correlated paternity did not appear to decline with distance between mothers in the Marchel population. Thus, realistic parameter estimates could not be obtained for this population. In contrast, there was a clear negative relationship between the extent of correlated paternity among mothers and the physical distance among mothers in the Vinson population (Fig. S1, Supporting information). Setting the threshold distance separating uncorrelated pollen pools at 1300 m (i.e. based on visual assessment of the scatter plot of correlated paternity vs. distance between pairs of mothers; Fig. S1, Supporting information), the estimated parameters of the exponential power dispersal function were $a = 0.115$ and $b = 0.244$. Varying the threshold distance between 500 and 1500 m resulted in estimates of a ranging from 0.089 to 0.126, estimates of b ranging from 0.238 to 0.246, and mean dispersal distances ranging from 1205 to 1372 m. Predictions generated based on these parameter estimates were roughly compatible with the shape of the observed frequency distribution of mating events in different distance classes,

[3]Correction added after online publication 24 December 2008: the y axis label 'Frequency of pollinations' in Fig 4 was corrected to 'Frequency of pollination'.

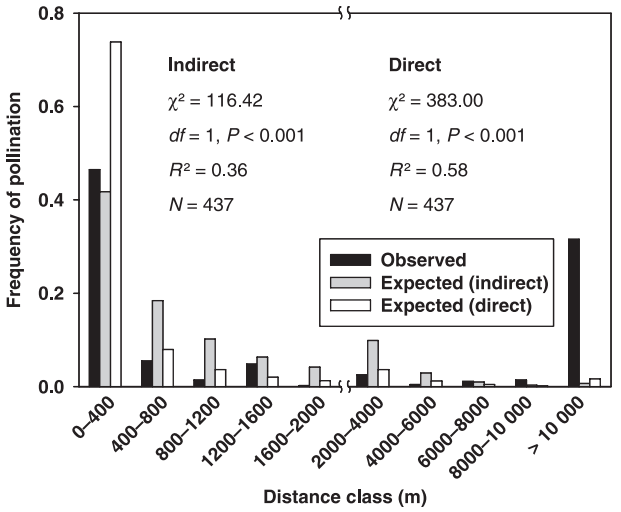

Fig. 6 Observed vs. expected pollination frequencies based on predictions from an exponential power dispersal function. Indirect estimates of the function parameters ($a = 0.115$, $b = 0.244$) were obtained based correlated paternity for pairs of females from the Vinson population. The mean dispersal distance based on these parameter estimates was $\delta = 1249$ m. Direct estimates ($a = 8.18 \times 10^{-17}$, $b = 0.078$) were obtained using maximum likelihood and paternity analysis data for the Vinson population. The mean dispersal distance based on direct parameter estimates was $\delta = 1283$ m. Goodness-of-fit test were based on distance classes pooled to assure that the expected number of pollinations in each class was at least five. *N* is defined in Fig. 5.

although they severely under-predicted long-distance dispersal (Fig. 6). The mean pollination distance calculated based on these estimates ($\delta = 1249$ m) was approximately six times smaller than that estimated based on paternity analysis and the three-parameter mixed dispersal function, but remarkably similar to that estimated based on fitting an exponential power dispersal curve to the paternity analysis data ($\delta = 1283$ m; Fig. 6).

Number of breeding males and neighbourhood size

The number of breeding males (N_{bmp}) was an order of magnitude higher in the Marchel population than in the Vinson population (Table 4). The neighbourhood sizes (N_{ep}) estimated directly from the paternity analysis data and from the degree of pollen pool differentiation were remarkably similar, and estimates for the Marchel population were 3.3 to 4.9 times higher than those for the Vinson population (Table 4).

Discussion

Pollen immigration

Although the spatial scales, the ecological characteristics and the demographic structures of the two study populations differed dramatically (Figs 1–3), pollen immigration was consistently high. This is not surprising in the light of results from paternity analysis studies in other wind-pollinated forest trees (Burczyk *et al*. 2004; Slavov *et al*. 2004a; Gérard *et al*. 2006; Goto *et al*. 2006), including other *Populus* species (Tabbener & Cottrell 2003; Pošpísková & Šálková 2006; Vanden Broeck *et al*. 2006). As expected, the rate of pollen immigration was nearly 70% higher in the Marchel population, which had a substantially smaller area sampled (Fig. 1) and an order of magnitude higher of unsampled breeding males (Table 4). Given the fact that we sampled all males located within 6–16 km from mother trees in the Vinson population, however, it is remarkable that approximately one-third of the offspring analysed in this population resulted from pollen immigration. While it is possible that we have overestimated pollen immigration in the Vinson population as a result of false paternity exclusions caused by null alleles and genotyping errors, or by failing to sample highly fertile male trees within the study area, both of these scenarios appear unlikely. First, the pollen immigration estimate for the Vinson population did not change substantially even after performing paternity exclusion based on up to five conditional mismatches between the genotypes of offspring and candidate fathers. Second, the Vinson site was searched extensively and great care was taken to sample all reproductively mature male trees in the study area. Because the landscape of eastern Oregon is arid and depauperate of vegetation, the aerial photographs we used as an independent check provided excellent capability to identify *Populus* stands, and even individual trees. We are therefore confident that we have sampled all significant sources of pollen within the study area.

Pollen immigration calculated for individual mothers was negatively correlated with the number of reproductively mature male trees located in close proximity. In the Vinson

Table 4 Number of breeding males and male neighbourhood sizes in two *Populus trichocarpa* populations with contrasting ecological and demographic characteristics

Population	N_{bmp} (95% CI)*	N_{ep} (direct)†	N_{ep} (indirect)‡
Marchel	3467 (1802–9337)	8.2	9.7
Vinson	280 (233–344)	2.5	2.0

*N_{bmp} is the number of breeding male trees (Nielsen *et al*. 2001). Likelihood-based 95% confidence intervals (Neale & Miller 1997) are shown in parenthesis.
†N_{ep} (direct) is the male genetic neighbourhood size estimated based on paternity analysis data.
‡N_{ep} (indirect) is the male genetic neighbourhood size estimated based on the degree of pollen pool differentiation (Φ_{FT}).

population, the strength of this correlation increased as the number of males within 100, 200, 300, 400 and 500 m was considered. This result is consistent with our two-component model of pollen dispersal, which predicts higher rate of pollinations through long-distance dispersal and pollen immigration (i.e. sampling from the uniform tail of the pollen dispersal function) when there are fewer phenologically compatible and fertile male trees in close proximity (i.e. within the exponential component of the dispersal curve). The small size of the area sampled, the small number of mother trees for which at least 10 seeds were analysed and the high density of males in the Marchel population did not allow us to fully evaluate whether pollen immigration in this population followed the same pattern. The consistent sign and the strength of the correlation between pollen immigration and the number of males within 100 m of mother trees sampled in the Marchel population, however, suggested that this is likely to be the case.

Pollen pool differentiation

The Φ_{FT} estimate we obtained based on data from the continuous Marchel population was similar to values observed in other wind-pollinated trees believed to have extensive pollen dispersal, such as species from the genera *Pinus*, *Picea* and *Quercus* (Smouse & Sork 2004; Fernández-Manjarrés & Sork 2005; Fernández-Manjarrés *et al*. 2006; O'Connell *et al*. 2006). In contrast, the estimate of pollen pool differentiation based on the entire data set from the Vinson population appears to be the highest reported to date for a temperate forest tree. One reason for this contrast was the different spatial scales at which mothers were sampled in the two populations (Fig 1–3; column D_{am} in Table 3). When we analysed a subset of the data including offspring from six of the Vinson mother trees that were located relatively close to one another (Table 3; Fig. 3), the Φ_{FT} estimate decreased by 56% but still remained nearly twice as high as the value we estimated in the Marchel population (Table 3). This is consistent with theoretical results and computer simulations (Austerlitz & Smouse 2001b; Austerlitz & Smouse 2002) showing that Φ_{FT} estimates are expected to increase as the average distance among sampled females increases up to approximately five times the mean pollen dispersal distance, which was estimated to be 7599 m based on fitting a mixed pollen dispersal function to paternity analysis data (discussed below). Two additional potential causes of inflated Φ_{FT} estimates are genetic structure and/or relatedness among males (Austerlitz & Smouse 2001b). Corrections for population structure among males resulted in relatively small changes of Φ_{FT} estimates in both populations (i.e. less than 10% of the estimated values). The strong spatial genetic structure we detected in the Vinson population (Slavov *et al.* unpublished data), on the other hand, suggests that relatedness among males that mate most frequently with a given female is a likely reason for the strong pollen pool differentiation. This effect is probably accentuated by the small number of males in close proximity to sampled female trees (mean number of males within 500 m = 3.8, range = 0–10) and the high occurrence of full sibs among the offspring of most mother trees sampled in the Vinson population. As an illustration of the latter factor, six of the mother trees sampled in the Vinson population had a single identified pollen donor, which in all cases was a male tree located within 500 m. Furthermore, when offspring for which a unique father was identified within 500 m of the mother tree were excluded from the TwoGener analysis, pollen pool differentiation among females in the Vinson population decreased by an order of magnitude. We therefore conclude that the remarkably high estimate of Φ_{FT} obtained for the Vinson population is most likely a result of: (i) the large average distance among mothers; (ii) relatedness among dominant pollen donors; and (iii) frequent mating with a small number of male trees located within 500 m of sampled mother trees.

Pollen dispersal curve

Paternity analysis. As expected under random mating, the frequency distribution of within-population pollinations in the Marchel population did not deviate significantly from the frequency distribution of potential intermate distances. In the Vinson population, however, the frequency distribution of potential intermate distances severely underpredicted local pollination frequencies and over-predicted long-distance pollination frequencies. The most likely explanations for this contrast are the dramatic differences between the spatial scales and the demographic characteristics of the two populations (Figs 1–3).

The mixed dispersal function we used generated realistic predictions of the frequencies of both local and long-distance pollinations in the Vinson population. The function parameters could not be estimated for the Marchel population because the small size of the area sampled in this population did not allow us to obtain enough information about the long-distance component of the pollen dispersal curve. *Ad hoc* computer simulations based on our data indicated that to obtain reliable estimates of the mixed dispersal function using our analytical approach, males have to be exhaustively sampled in an area with a radius of at least 3–4 km (Slavov, unpublished data). While the minimum area that needs to be sampled will vary in different situations, data on long-distance dispersal are indispensable for applying our maximum-likelihood approach. This is because for short pollination distances, the exponential distribution and the uniform distribution forming the mixed dispersal function (eqn 2) overlap extensively, which makes the estimation of their parameters using maximum-likelihood methods on

grouped data particularly challenging (Macdonald 1986; Macdonald & Green 1988; Clark 1998; Higgins & Richardson 1999). Our analyses of both simulated and real data indicated that estimates of function parameters are relatively insensitive to the number and width of the distance classes used (e.g. Table S2, Supporting information), as long as a substantial amount of data on long-distance dispersal is available.

Regardless of the dramatic ecological and demographic differences between the two study populations, the observed general patterns of pollen dispersal were remarkably consistent. In both populations, pollination frequencies declined exponentially for distances up to 500–600 m, and pollinations within this range of distances accounted for approximately half of the mating events. Furthermore, the mixed dispersal function parameterised based on data from the Vinson population generated predictions that were strongly correlated with pollen immigration and pollination frequencies in the six distance classes for which mating events were detected in the Marchel population (Fig. 5A).

Correlated paternity. The exponential power-dispersal function, whose parameters were estimated based on correlated paternity among mothers from the Vinson population, generated predictions that did not fit but were moderately correlated with the observed frequency distribution of pollinations (Fig. 6). The shape parameter estimates we obtained using paternity analysis data and patterns of correlated paternity (b = 0.078 and b = 0.238–0.246, respectively) were comparable to those obtained in a large-scale study in *Sorbus torminalis* (b = 0.17–0.38 based on paternity analysis and b = 0.091–0.565 based on correlated paternity; Austerlitz *et al*. 2004; Oddou-Muratorio *et al*. 2005). Furthermore, the mean dispersal distances obtained using the two approaches differed by only 34 m (i.e. 2.7%). Thus, our empirical results are consistent with the findings of previous simulation studies indicating that pollen dispersal parameters estimated from patterns of correlated paternity are not highly accurate, but can be used reliably to obtain a general idea about the shape (i.e. thin- vs. fat-tailed) of the pollen dispersal curve and the mean dispersal distance (Austerlitz *et al*. 2004; Burczyk & Koralewski 2005; Robledo-Arnuncio *et al*. 2006).

Choosing a dispersal function and an estimation approach. Weibull and exponential power-dispersal functions that were either fitted to paternity analysis results (e.g. Fig. 6) or indirectly estimated based on correlated paternity data from the Vinson population (Fig. 6) were invariably fat-tailed (b < 1). Their overall fit, however, was poor, which was mostly caused by substantial over-prediction of short-distance pollinations and under-prediction of long-distance pollinations and pollen immigration. A similar, although less pronounced, pattern was observed in two other studies in which pollen dispersal of forest trees was modelled using an exponential power function (Fig. 2 in Oddou-Muratorio *et al*. 2005; Fig. 3 in Robledo-Arnuncio & Gil 2005). It appears therefore that the exponential power family of distribution functions, which has become a popular choice for modelling propagule dispersal because of its shape flexibility (Clark 1998; Clark *et al*. 1998; Austerlitz *et al*. 2004; Oddou-Muratorio *et al*. 2005; Gérard *et al*. 2006; González-Martínez *et al*. 2006; Robledo-Arnuncio *et al*. 2006; Robledo-Arnuncio & García 2007) may not be universally appropriate. The need to try different types of dispersal functions was also illustrated in a recent landscape-level study of pollen flow in *Beta vulgaris*, which revealed extensive long-distance pollen dispersal (Fénart *et al*. 2007).

Modelling pollen dispersal as a two-component process can provide a substantially better fit to empirical data compared to any of the commonly used one- and two-parameter dispersal functions (Goto *et al*. 2006; this study). However, the two-component dispersal function we used is a crude and probably unrealistic model of effective pollen dispersal. Modelling long-distance pollination using a uniform distribution, for example, is probably too simplistic. The mean dispersal distance calculated based on our two-component dispersal function (δ = 7599 m) is likely to be an overestimate because we modelled long-distance dispersal using a uniform distribution, whereas using a decaying function (e.g. exponential power or Weibull) would have been more biologically realistic. Our approach could be refined by using a mixture of exponential power or Weibull functions to model short- and long-distance dispersal (Clark 1998; Higgins & Richardson 1999). However, this would increase the number of parameters that need to be estimated from three to five and would make the maximum-likelihood estimation more challenging (Macdonald 1986; Macdonald & Green 1988; Clark 1998; Higgins & Richardson 1999).

When the same function was used, pollen dispersal curves fitted based on direct (i.e. paternity analysis) and indirect (i.e. correlated paternity) methods had relatively similar shapes and very similar means (Fig. 6). It is possible, therefore, that indirect methods could be enhanced substantially by modelling pollen dispersal as a two-component process.

Number of breeding males and neighbourhood size

As expected, the estimated numbers of breeding males reflected the dramatic difference between the demographic characteristics of the two study populations. Values of N_{bmp} should be treated as upper limits because their estimation does not account for the occurrence of null alleles and genotyping errors, even low rates of which have

been demonstrated to result in inflated N_{bmp} estimates (Oddou-Muratorio *et al*. 2003).

Although patterns of pollen dispersal appeared very similar in the two study populations (discussed above), the male neighbourhood sizes for females in the Marchel population were 3–5 times higher than those for females in the Vinson population. One way to think about the N_{ep} values presented in Table 4 is that a progeny array from a mother tree in the Marchel population would have been generated by mating with an average of 8–10 'idealised' males (i.e. having equal reproductive success), whereas a progeny array from a mother tree in the Vinson population would have been generated by mating with only 2–3 idealised males.

In both populations, the average distance among sampled mothers was less than five times the mean pollination distance, which has been recommended as a minimum for the reliable estimation of Φ_{FT} and its derivative parameters (Austerlitz & Smouse 2001a). Nevertheless, neighbourhood sizes estimated directly from the paternity analysis data and from the degree of pollen pool differentiation were remarkably similar, reinforcing the value of indirect estimates of correlated paternity as a relatively inexpensive and rapid means of obtaining information of great importance for management and conservation (Smouse & Sork 2004; Burczyk & Koralewski 2005).

Practical implications

This study is part of a larger project aimed at assessing the potential introgression of transgenes into wild tree populations using spatial simulation modelling (Brunner *et al*. 2007). Poplars have been used more extensively for genetic engineering field studies than any other tree taxon (Bradshaw & Strauss 2001), and transgenic insect-resistant poplars are already being employed on a commercial scale in China (Wang *et al*. 2004), so there is a strong interest in assessing potential risks of transgenic *Populus* plantations. Our results indicate that long-distance dispersal is extensive and pollen dispersal curves are similar in populations with very different ecological and demographic characteristics. Pollen dispersal in simulation models for diverse sites can therefore be described using a two-component process, with local dispersal approximated by an exponential distribution, and long distance dispersal approximated by a uniform distribution (Brunner *et al*. 2007).

In addition, our results have several implications for gene conservation in *P. trichocarpa* and other forest trees. First, pollen immigration is expected to be high even with substantial spatial isolation of the populations of interest. Second, despite extensive long-distance dispersal, genetic neighbourhood sizes are likely to be very small in disjunct populations east of the Cascade Mountains. Finally, the effective number of male parents (N_{ep}) estimated directly from paternity analysis, which is used as a diversity index in certifying seeds for afforestation (Stoehr *et al*. 2004), can be reliably approximated using time- and cost-efficient indirect methods.

Conclusions

This study supports the emerging consensus that long-distance pollen dispersal can be very extensive in forest trees. As expected, levels of pollen immigration were inversely related to the size of the area in which all candidate fathers have been sampled and the number of candidate fathers in close proximity to mother trees. The relative degrees of pollen pool differentiation and the sizes of male neighbourhoods around sampled mother trees reflected the dramatic ecological and demographic differences between the two study sites.

Modelling pollen dispersal as a two-component process appears to be superior to trying to fit a single pollen dispersal curve. Regardless of whether a direct or an indirect approach is used, fitting a mixed dispersal curve should be attempted when long-distance pollen dispersal is extensive. Paternity analysis data and patterns of correlated paternity among mothers generated similar estimates of: (i) male neighbourhood sizes; (ii) the mean pollen dispersal distance; and (iii) the shape parameter of the pollen dispersal curve. Thus, studies designed to satisfy the critical assumptions of indirect approaches (i.e. sampling mothers at sufficiently large distances) can provide reliable information about these parameters at a relatively low cost, which would allow contemporary patterns of pollen-mediated gene flow to be characterised for populations growing under a variety of ecological conditions and over multiple years.

Acknowledgements

Assistance with field and lab work, as well as with data analysis, was provided by Shuping Cheng, Gokcin Temel, Jace Carson, Caprice Rosato, Eliza Walthers, Paul Rosenfeld, Rick Meilan, Toby Bradshaw, Brian Watson and David Birkes. This manuscript was greatly improved by the suggestions of Victoria Sork and two anonymous reviewers. Financial support was provided by an EPA STAR Fellowship to SPD, the Tree Biosafety and Genomics Research Cooperative, USDA Biotechnology Risk Assessment Grant 97-39210-5022, the DOE Oak Ridge National Lab Bioenergy Program, the DOE BioEnergy Science Center and NSF FIBR grant DEB-0425908.

References

Adams WT, Birkes DS (1991) Estimating mating patterns in forest tree populations. In: *Biochemical Markers in the Population Genetics of Forest Trees* (eds Fineschi S, Malvoti ME, Cannata F, Hattemer HH), pp. 157–172. SPB Academic Publishing, The Hague.

Austerlitz F, Smouse PE (2001a) Two-generation analysis of pollen flow across a landscape. II. Relation between Φ_{FT}, pollen dispersal and interfemale distance. *Genetics*, **157**, 851–857.

Austerlitz F, Smouse PE (2001b) Two-generation analysis of pollen flow across a landscape. III. Impact of adult population structure. *Genetical Research*, **78**, 271–280.

Austerlitz F, Smouse PE (2002) Two-generation analysis of pollen flow across a landscape. IV. Estimating the dispersal parameter. *Genetics*, **161**, 355–363.

Austerlitz F, Dick CW, Dutech C *et al*. (2004) Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology*, **13**, 937–954.

Balatinecz JJ, Kretschmann DE (2001) Properties and utilization of poplar wood. In: *Poplar Culture in North America* (eds Dickmann D, Isebrands JG, Eckenwalder JE, Richardson J), pp. 277–290. NRC Press, Ottawa, ON.

Benetka V, Mottl J, Vacková K *et al*. (1999) Estimation of the introgression level in *Populus nigra* L. populations by means of isozyme gene markers. *Silvae Genetica*, **48**, 218–223.

Bhalerao R, Nilsson O, Sandberg G (2003) Out of the woods: forest biotechnology enters the genomic era. *Current Opinion in Biotechnology*, **14**, 206–213.

Bittencourt JVM, Sebbenn AM (2007) Patterns of pollen and seed dispersal in a small, fragmented population of the wind-pollinated tree *Araucaria angustifolia* in southern Brazil. *Heredity*, **99**, 580–591.

Boerjan W (2005) Biotechnology and the domestication of forest trees. *Current Opinion in Biotechnology*, **16**, 159–166.

Braatne JH, Rood SB, Heilman PE (1996) Life history, ecology, and reproduction of riparian cottonwoods in North America. In: *Biology of Populus and its Implications for Management and Conservation* (eds Stettler RF, Bradshaw HD Jr, Heilman PE, Hinckley TM), pp. 57–85. NRC Research Press, Ottawa, ON.

Bradshaw HD, Strauss SH (2001) Breeding strategies for the 21st century: Domestication of poplar. In: *Poplar Culture in North America* (eds Dickmann DI, Isebrands JG, Eckenwalder JE, Richardson J), pp. 383–394. NRC Research Press, Ottawa, ON.

Bradshaw HD, Ceulemans RE, Davis J, Stettler RF (2000) Emerging model systems in plant biology: poplar (*Populus*) as a model forest tree. *Journal of Plant Growth Regulation*, **19**, 306–313.

Brondani RPV, Brondani C, Tarchini R, Grattapaglia D (1998) Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. *Theoretical and Applied Genetics*, **97**, 816–827.

Brunner AM, Rottmann WH, Sheppard LA *et al*. (2000) Structure and expression of duplicate AGAMOUS orthologues in poplar. *Plant Molecular Biology*, **44**, 619–634.

Brunner AM, Li JY, DiFazio SP *et al*. (2007) Genetic containment of forest plantations. *Tree Genetics and Genomes*, **3**, 75–100.

Burczyk J, Koralewski TE (2005) Parentage versus two-generation analyses for estimating pollen-mediated gene flow in plant populations. *Molecular Ecology*, **14**, 2525–2537.

Burczyk J, Adams WT, Moran GF, Griffin AR (2002) Complex patterns of mating revealed in a *Eucalyptus regnans* seed orchard using allozyme markers and the neighborhood model. *Molecular Ecology*, **11**, 2379–2391.

Burczyk J, DiFazio SP, Adams WT (2004) Gene flow in forest trees: How far do genes really travel? *Forest Genetics*, **11**, 179–192.

Burczyk J, Adams WT, Birkes DS, Chybicki IJ (2006) Using genetic markers to directly estimate gene flow and reproductive success parameters in plants on the basis of naturally regenerated seedlings. *Genetics*, **173**, 363–372.

Busov VB, Brunner AM, Meilan R *et al*. (2005) Genetic transformation: a powerful tool for dissection of adaptive traits in trees. *New Phytologist*, **167**, 9–18.

Cercueil A, Bellemain E, Manel S (2002) PARENTE: Computer program for parentage analysis. *Journal of Heredity*, **93**, 458–459.

Cervera MT, Storme V, Ivens B *et al*. (2001) Dense genetic linkage maps of three *Populus* species (*Populus deltoides*, *P. nigra* and *P. trichocarpa*) based on AFLP and microsatellite markers. *Genetics*, **158**, 787–809.

Clark JS (1998) Why trees migrate so fast: Confronting theory with dispersal biology and the paleorecord. *American Naturalist*, **152**, 204–224.

Clark JS, Macklin E, Wood L (1998) Stages and spatial scales of recruitment limitation in southern Appalachian forests. *Ecological Monographs*, **68**, 213–235.

Di Giovanni F, Beckett PM (1990) On the mathematical modeling of pollen dispersal and deposition. *Journal of Applied Meteorology*, **29**, 1352–1357.

Di Giovanni F, Kevan PG, Arnold J (1996) Lower planetary boundary layer profiles of atmospheric conifer pollen above a seed orchard in northern Ontario, Canada. *Forest Ecology and Management*, **83**, 87–97.

DiFazio SP (2002) *Measuring and Modeling Gene Flow from Hybrid Poplar Plantations: Implications for Transgenic Risk Assessment*. PhD Thesis. Oregon State University, Corvallis, OR.

DiFazio SP, Slavov GT, Burczyk J, Leonardi S, Strauss SH (2004) Gene Flow From Tree Plantations and Implications for Transgenic Risk Assessment. In: *Forest Biotechnology for the 21st Century* (eds Walter C, Carson M), pp. 405–422. Research Signpost, Kerala.

Doty SL (2008) Enhancing phytoremediation through the use of transgenics and endophytes. *New Phytologist*, **179**, 318–333.

Eckenwalder JE (1996) Systematics and evolution of *Populus*. In: *Biology of Populus and its Implications for Management and Conservation* (eds Stettler RF, Bradshaw HD Jr, Heilman PE, Hinckley TM), pp. 7–32. NRC Research Press, Ottawa, ON.

Ellison AM, Bank MS, Clinton BD *et al*. (2005) Loss of foundation species: consequences for the structure and dynamics of forested ecosystems. *Frontiers in Ecology and the Environment*, **3**, 479–486.

Elsik CG, Minihan VT, Hall SE, Scarpa AM, Williams CG (2000) Low-copy microsatellite markers for *Pinus taeda* L. *Genome*, **43**, 550–555.

Ewen KR, Bahlo M, Treloar SA *et al*. (2000) Identification and analysis of error types in high-throughput genotyping. *American Journal of Human Genetics*, **67**, 727–736.

Excoffier L, Laval G, Schneider S (2005) *Arlequin*, ver 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47–50.

Farnum P, Lucier A, Meilan R (2007) Ecological and population genetics research imperatives for transgenic trees. *Tree Genetics and Genomes*, **3**, 119–133.

Fénart S, Austerlitz F, Cuguen J, Arnaud JF (2007) Long distance pollen-mediated gene flow at a landscape level: the weed beet as a case study. *Molecular Ecology*, **16**, 3801–3813.

Fernández-Manjarrés JF, Sork VL (2005) Mating patterns of a subdivided population of the Andean oak (*Quercus humboldtii* Bonpl., Fagaceae). *Journal of Heredity*, **96**, 635–643.

Fernández-Manjarrés JF, Idol J, Sork VL (2006) Mating patterns of black oak *Quercus velutina* (Fagaceae) in a Missouri oak-hickory forest. *Journal of Heredity*, **97**, 451–455.

Fossati T, Grassi F, Sala F, Castiglione S (2003) Molecular analysis of natural populations of *Populus nigra* L. intermingled with cultivated hybrids. *Molecular Ecology*, **12**, 2033–2043.

Gérard PR, Klein EK, Austerlitz F, Fernández-Manjarrés JF, Frascaria-Lacoste N (2006) Assortative mating and differential male mating success in an ash hybrid zone population. *BMC Evolutionary Biology*, **6**, 96.

González-Martínez SC, Burczyk J, Nathan R *et al*. (2006) Effective gene dispersal and female reproductive success in Mediterranean maritime pine (*Pinus pinaster* Aiton). *Molecular Ecology*, **15**, 4577–4588.

Goto S, Shimatani K, Yoshimaru H, Takahashi Y (2006) Fat-tailed gene flow in the dioecious canopy tree species *Fraxinus mandshurica* var. *japonica* revealed by microsatellites. *Molecular Ecology*, **15**, 2985–2996.

Heinze B (1998) Molecular genetic differentiation and identification of black poplar and hybrid poplar clones. *FBVA Berichte*, **105**, 44 pp.

Hibbs D, Withrow-Robinson B, Brown D, Fletcher R (2003) Hybrid poplar in the Willamette Valley. *Western Journal of Applied Forestry*, **18**, 281–285.

Higgins SI, Richardson DM (1999) Predicting plant migration rates in a changing world: The role of long-distance dispersal. *American Naturalist*, **153**, 464–475.

Imbert E, Lefèvre F (2003) Dispersal and gene flow of *Populus nigra* (Salicaceae) along a dynamic river system. *Journal of Ecology*, **91**, 447–456.

Jelinski DE, Cheliak WM (1992) Genetic diversity and spatial subdivision of *Populus tremuloides* (Salicaceae) in a heterogeneous landscape. *American Journal of Botany*, **79**, 728–736.

Jones AG, Ardren WR (2003) Methods of parentage analysis in natural populations. *Molecular Ecology*, **12**, 2511–2523.

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, **16**, 1099–1106.

Katul GG, Porporato A, Nathan R *et al*. (2005) Mechanistic analytical models for long-distance seed dispersal by wind. *American Naturalist*, **166**, 368–381.

Kelleher CT, Chiu R, Shin H *et al*. (2007) A physical map of the highly heterozygous *Populus* genome: integration with the genome sequence and genetic map and analysis of haplotype variation. *Plant Journal*, **50**, 1063–1078.

Lanner RM (1965) Needed: a new approach to the study of pollen dispersion. *Silvae Genetica*, **15**, 50–52.

Legionnet A, Lefèvre F (1996) Genetic variation of the riparian pioneer tree species *Populus nigra* L. I. Study of population structure based on isozymes. *Heredity*, **77**, 629–637.

Lescot M, Rombauts S, Zhang J *et al*. (2004) Annotation of a 95-kb *Populus deltoides* genomic sequence reveals a disease resistance gene cluster and novel class I and class II transposable elements. *Theoretical and Applied Genetics*, **109**, 10–22.

Levin DA, Kerster HW (1974) Gene flow in seed plants. In: *Evolutionary Biology 7* (eds Dobzhansky TM, Hecht MT, Steere WC), pp. 139–220. Plenum Press, New York.

Lexer C, Fay MF, Joseph JA *et al*. (2005) Barrier to gene flow between two ecologically divergent *Populus* species, *P. alba* (white poplar) and *P. tremula* (European aspen): the role of ecology and life history in gene introgression. *Molecular Ecology*, **14**, 1045–1057.

Macdonald PDM (1986) MIX: an interactive program for fitting mixtures of distributions. *American Statistician*, **40**, 53.

Macdonald PDM, Green PEJ (1988) *User's Guide to Program MIX: an Interactive Program for Fitting Mixtures of Distributions, Release 2.3*. Ichthus Data Systems, Hamilton, ON.

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

Neale MC, Miller MB (1997) The use of likelihood-based confidence intervals in genetic models. *Behavior Genetics*, **27**, 113–120.

Nei M, Roychoudhury AK (1974) Sampling variances of heterozygosity and genetic distance. *Genetics*, **76**, 379–390.

Nielsen R, Mattila DK, Clapham PJ, Palsbøll PJ (2001) Statistical approaches to paternity analysis in natural populations and applications to the Atlantic humpback whale. *Genetics*, **157**, 1673–1682.

O'Connell LM, Mosseler A, Rajora OP (2006) Impacts of forest fragmentation on the mating system and genetic diversity of white spruce (*Picea glauca*) at the landscape level. *Heredity*, **97**, 418–426.

Oddou-Muratorio S, Houot ML, Demesure-Musch B, Austerlitz F (2003) Pollen flow in the wildservice tree, *Sorbus torminalis* (L.) Crantz. I. Evaluating the paternity analysis procedure in continuous populations. *Molecular Ecology*, **12**, 3427–3439.

Oddou-Muratorio S, Klein EK, Austerlitz F (2005) Pollen flow in the wildservice tree, *Sorbus torminalis* (L.) Crantz. II. Pollen dispersal and heterogeneity in mating success inferred from parent-offspring analysis. *Molecular Ecology*, **14**, 4441–4452.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Pošpísková M, Šálková I (2006) Population structure and parentage analysis of black poplar along the Morava River. *Canadian Journal of Forest Research*, **36**, 1067–1076.

Robledo-Arnuncio JJ, García C (2007) Estimation of the seed dispersal kernel from exact identification of source plants. *Molecular Ecology*, **16**, 5098–5109.

Robledo-Arnuncio JJ, Gil L (2005) Patterns of pollen dispersal in a small population of *Pinus sylvestris* L. revealed by total-exclusion paternity analysis. *Heredity*, **94**, 13–22.

Robledo-Arnuncio JJ, Austerlitz F, Smouse PE (2006) A new method of estimating the pollen dispersal curve independently of effective density. *Genetics*, **173**, 1033–1045.

Robledo-Arnuncio JJ, Austerlitz F, Smouse PE (2007) POLDISP: a software package for indirect estimation of contemporary pollen dispersal. *Molecular Ecology Notes*, **7**, 763–766.

Rood SB, Kalischuk AR, Polzin ML, Braatne JH (2003) Branch propagation, not cladoptosis, permits dispersive, clonal reproduction of riparian cottonwoods. *Forest Ecology and Management*, **186**, 227–242.

Rood SB, Goater LA, Mahoney JM *et al*. (2007) Flood, fire, and ice: disturbance ecology of riparian cottonwoods. *Canadian Journal of Botany*, **85**, 1019–1032.

Ross SM (1996) *Simulation*, pp. 36–44. Academic Press, San Diego, CA.

Rubin EM (2008) Genomics of cellulosic biofuels. *Nature*, **454**, 841–845.

Schweitzer JA, Martinsen GD, Whitham TG (2002) Cottonwood hybrids gain fitness traits of both parents: a mechanism for their long-term persistence? *American Journal of Botany*, **89**, 981–990.

Slavov GT, DiFazio SP, Strauss SH (2004a) Gene flow in forest trees: Gene migration patterns and landscape modeling of transgene dispersion in hybrid poplar. In: *Introgression from Genetically Modified Plants Into Wild Relatives* (eds den Nijs JCM, Bartsch D, Sweet J), pp. 89–106. CAB International, Wallingford.

Slavov GT, Howe GT, Yakovlev I *et al*. (2004b) Highly variable SSR markers in Douglas-fir: Mendelian inheritance and map locations. *Theoretical and Applied Genetics*, **108**, 873–880.

Slavov GT, Howe GT, Adams WT (2005a) Pollen contamination and mating patterns in a Douglas-fir seed orchard as measured by simple sequence repeat markers. *Canadian Journal of Forest Research*, **35**, 1592–1603.

Slavov GT, Howe GT, Gyaourova AV *et al*. (2005b) Estimating pollen flow using SSR markers and paternity exclusion: accounting for mistyping. *Molecular Ecology*, **14**, 3109–3121.

Smith DB, Adams WT (1983) Measuring pollen contamination in clonal seed orchards with the aid of genetic markers. In: *Proceedings of the 17th Southern Forest Tree Improvement Conference*, pp. 64–73. Athens, GA, USA.

Smouse PE, Sork VL (2004) Measuring pollen flow in forest trees: an exposition of alternative approaches. *Forest Ecology and Management*, **197**, 21–38.

Smouse PE, Dyer RJ, Westfall RD, Sork VL (2001) Two-generation analysis of pollen flow across a landscape. I. Male gamete heterogeneity among females. *Evolution*, **55**, 260–271.

Smouse PE, Robledo-Arnuncio JJ, González-Martínez SC (2007) Implications of natural propagule flow for containment of genetically modified forest trees. *Tree Genetics and Genomes*, **3**, 141–152.

Sterky F, Bhalerao RR, Unneberg P *et al*. (2004) A *Populus* EST resource for plant functional genomics. *Proceedings of the National Academy of Sciences, USA*, **101**, 13951–13956.

Stirling B, Yang ZK, Gunter LE, Tuskan GA, Bradshaw HD (2003) Comparative sequence analysis between orthologous regions of the Arabidopsis and *Populus* genomes reveals substantial synteny and microcollinearity. *Canadian Journal of Forest Research*, **33**, 2245–2251.

Stoehr M, Webber J, Woods J (2004) Protocol for rating seed orchard seedlots in British Columbia: quantifying genetic gain and diversity. *Forestry*, **77**, 297–303.

Streiff R, Ducousso A, Lexer C *et al*. (1999) Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L. & *Q. petraea* (Matt.) Liebl. *Molecular Ecology*, **8**, 831–841.

Summers K, Amos W (1997) Behavioral, ecological, and molecular genetic analyses of reproductive strategies in the Amazonian dart-poison frog, *Dendrobates ventrimaculatus*. *Behavioral Ecology*, **8**, 260–267.

Tabbener HE, Cottrell JE (2003) The use of PCR based DNA markers to study the paternity of poplar seedlings. *Forest Ecology and Management*, **179**, 363–376.

Taylor G (2002) *Populus*: Arabidopsis for forestry. Do we need a model tree? *Annals of Botany*, **90**, 681–689.

Tufto J, Engen S, Hindar K (1997) Stochastic dispersal processes in plant populations. *Theoretical Population Biology*, **52**, 16–26.

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

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

Vanden Broeck A, Storme V, Cottrell JE *et al*. (2004) Gene flow between cultivated poplars and native black poplar (*Populus nigra* L.): a case study along the river Meuse on the Dutch-Belgian border. *Forest Ecology and Management*, **197**, 307–310.

Vanden Broeck A, Cottrell J, Quataert P *et al*. (2006) Paternity analysis of *Populus nigra* L. offspring in a Belgian plantation of native and exotic poplars. *Annals of Forest Science*, **63**, 783–790.

Vandeputte M, Mauger S, Dupont-Nivet M (2006) An evaluation of allowing for mismatches as a way to manage genotyping errors in parentage assignment by exclusion. *Molecular Ecology Notes*, **6**, 265–267.

van Frankenhuyzen K, Beardmore T (2004) Current status and environmental impact of transgenic forest trees. *Canadian Journal of Forest Research*, **34**, 1163–1180.

Wackerly DD, Mendenhall W, Scheaffer RL (2002) *Mathematical Statistics with Applications*. Duxburry, USA.

Wang L, Han YF, Hu J (2004) Transgenic forest trees for insect resistance. In: *Molecular Genetics and Breeding of Forest Trees* (eds Kumar S, Fladung M), pp. 243–262. Hawarth Press, New York.

Weber JC, Stettler RF (1981) Isoenzyme variation among ten populations of *Populus trichocarpa* Torr. et Gray in the Pacific Northwest. *Silvae Genetica*, **30**, 82–87.

Whitham TG, Martinsen GD, Floate KD *et al*. (1999) Plant hybrid zones affect biodiversity: Tools for a genetic-based understanding of community structure. *Ecology*, **80**, 416–428.

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

Williams CG, Ladeau SL, Oren R, Katul GG (2006) Modeling seed dispersal distances: Implications for transgenic *Pinus taeda*. *Ecological Applications*, **16**, 117–124.

Wilmer JW, Allen PJ, Pomeroy PP, Twiss SD, Amos W (1999) Where have all the fathers gone? An extensive microsatellite analysis of paternity in the grey seal (*Halichoerus grypus*). *Molecular Ecology*, **8**, 1417–1429.

Wright S (1946) Isolation by distance under diverse systems of mating. *Genetics*, **31**, 39–59.

Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*, **19**, 295–420.

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

Supporting Information

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

Fig. S1 Correlated paternity vs. distance between pairs of mothers. Only the range in which pairwise correlated paternity decays is shown.

Table S1 Relative positions of mother trees within sites and number of seeds analysed.

Table S2 Parameters of a mixed distribution function (see Materials and methods) estimated based on pollination distance data from the Vinson population grouped in different distance classes. Estimates shown in bold italics were reported in the text.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.