GENE FLOW IN FOREST TREES: HOW FAR DO GENES REALLY TRAVEL? [1]

J. Burczyk [2], S. P. DiFazio [3] & W. T. Adams [4]

[2)] Department of Genetics, Institute of Biology and Environmental Protection, Bydgoszcz University, 85-064 Bydgoszcz, Poland
[3)] Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, U.S.A.
[4)] Department of Forest Science, Oregon State University, Corvallis, OR 97331, U.S.A.

Corresponding author: Jaroslaw Burczyk, Bydgoszcz University, Institute of Biology and Environmental Protection, Department of Genetics, ul. Chodkiewicza 30, 85-064 Bydgoszcz, Poland, phone: (+48-52) 3608396, 3419290, fax: (+48-52) 3400734, e-mail: burczyk@ab-byd.edu.pl

ABSTRACT

Gene flow is one of the most important factors shaping the genetic structure of populations. In recent decades, a number of studies have addressed issues of contemporary gene flow in forest trees, including pollen and seed dispersal, and gene immigration into natural and breeding populations (primarily seed orchards). Gene flow might be considered either beneficial or deleterious from the point of view of a conservation geneticist or a tree breeder. Extensive gene dispersal within local populations promotes panmixis and reduces family structuring in natural regeneration, thus reducing the potential for inbreeding. However, gene flow may reduce fitness of offspring if genes come from populations maladapted to the habitat of offspring establishment. Furthermore, substantial gene flow limits divergence among populations that might otherwise occur because of genetic drift and directional selection, and may enhance genetic diversity within populations.

The robustness and discriminatory power of parentage analysis have been significantly improved in recent years due to advances in molecular marker technology and analytical techniques. However, knowledge of gene flow in forest trees is still unsatisfactory due to continued shortcomings of available markers, inherent limitations of statistical models, and the anecdotal nature of many gene flow studies, which are typically limited in scope. Another limitation is that pollen gene flow is usually estimated by sampling seeds from a number of mother trees. Restricting sampling to seeds seems adequate when evaluating gene flow in seed collections to be used for artificial reforestation. However, in naturally regenerating populations, the most important parameter is the proportion of immigrant alleles in established seedlings. If natural selection favours local genotypes, this should be reflected in the difference between potential gene flow measured at the seed stage and realized gene flow measured in established seedlings. Realized gene flow is a complex phenomenon that depends on a large number of deterministic and stochastic variables. Profound understanding of effective gene flow observed at the landscape level may require joint efforts in population genetics, ecology, advanced multivariate statistics, and spatial simulation modelling.

INTRODUCTION

Evolutionary biologists have long been interested in gene flow[2] because it is one of the main factors determining the genetic architecture of populations, along with mutation, drift, and selection. For example, gene flow is crucial to all phases of Wright's *Shifting Balance Theory*, with restricted gene flow providing the settings for differentiation of demes by genetic drift and subsequent increase in frequency of favourable gene combinations, followed by spread of favourable combinations by interdemic gene flow (WRIGHT 1988). In fact, a major point of contention between the theories of Wright and Fisher revolved around gene flow and the effective size of populations, with Wright contending that gene flow was sufficiently restricted in wild populations that differentiation through genetic drift was a common occurrence

[1] This paper has been presented at the IUFRO Symposium on Population and Evolutionary Genetics of Forest Trees held in Stará Lesná, Slovakia, on August 25–29, 2002.

[2] By gene flow we are referring generically to transfer of genes between locations due to the dispersal of sexual and/or asexual propagules. By our definition, this en-compasses movement within populations as well as immigration and emigration.

(WADE & GOODNIGHT 1998). The *Biological Species Concept* is also primarily a question of gene flow, and disruption of gene flow between populations is thought to play a major role in speciation (LEWONTIN 1997; NIKLAS 1997). Finally, accurate estimation of gene dispersal is essential for assessing correlates of fitness in natural populations and clarifying the roles of sexual selection and ecological factors in the evolution of breeding systems (CHARNOV 1982; MORGAN & CONNER 2001).

Gene flow is also important in conservation biology, primarily because of its influence on the effective population size of threatened species (LANDE 1988; SCHEMSKE *et al.* 1994). For example, the metapopulation concept hinges upon gene flow among disjunct populations, some of which are subject to extinction without input of propagules from external populations (HANSKI & GILPIN 1997). Gene flow can also be the bane of conservation efforts, as rare species may be effectively lost through hybridization with common congeners (ELLSTRAND 1992b; FERDY & AUSTERLITZ 2002), and outcrossing depression can occur when historically separate populations are brought into contact (Templeton 1986). Even for widespread species like many forest trees, there is concern about the effects of silviculture and climate change on the genetic composition and future adaptability of forest trees (MILLAR & LIBBY 1991; SAVOLAINEN & KARKKAINEN 1992), and gene flow has a major impact on the influence of these effects. Accurate estimation of contemporary levels of realized gene flow is essential for designing reserves and conservation strategies that maximize interconnectedness of populations while minimizing unwanted gene flow (ADAMS & BURCZYK 2000).

From an ecological standpoint, gene flow is one of the key factors determining composition of ecosystems and responses to disturbance (CONNELL & SLATYER 1977). For example, recruitment into forest gaps is often determined by propagule availability, and gene flow by seed dispersal helps determine the alpha- and beta-diversity of temperate and tropical forests (EHRLEN & ERIKSSON 2000; CONDIT *et al.* 2002). Propagule dispersal can also modulate competitive interactions, allowing inferior competitors to preemptively occupy a site due to propagule pressure, thus causing an adaptive disequilibrium (LENORMAND 2002).

In addition to its prominent role in evolutionary biology and ecology, gene flow has garnered substantial interest in applied fields like crop improvement and ecological risk assessment. BATEMAN (1947) devised methods to measure and prevent the deleterious effects of pollen contamination of seed crops. This remains a major issue today, particularly for forest tree seed orchards, where skewed mating success and contamination from external pollen sources can cause substantial reductions in realized gains (DI-GIOVANNI & KEVAN 1991, ADAMS & BURCZYK 2000). Risk assessment for the introduction of exotic organisms also relies on accurate estimates of gene flow and migration (RUESINK *et al.* 1995; HIGGINS *et al.* 2000). Finally, with the introduction of genetically engineered organisms to the environment, there is enhanced interest in measuring prevailing gene flow from agricultural fields and tree plantations into wild populations, and applying this information to assessing risks of deleterious agronomic and ecological effects (SNOW & MORÁN-PALMA 1997; STRAUSS *et al.* 2001).

Here we review some of the approaches commonly used to study gene flow in forest trees, summarizing their strengths and weaknesses. In addition we review the gene flow literature, focusing mostly on pollen flow among tree populations. We also indicate gaps in our knowledge and suggest future research directions.

METHODS OF ESTIMATING GENE FLOW

One of the main objectives of population geneticists is to describe patterns of gene flow in terms of distances and rates, although accurate estimates of gene flow among populations are difficult to obtain. A number of methods have been developed to trace gene flow, ranging from tracking propagule movement (including vector behavior) to applying sophisticated statistical models to genetic marker data (see reviews in: ADAMS 1992; ELLSTRAND 1992a; BOSSART & PROWELL 1998; OUBORG *et al.* 1999; SORK *et al.* 1999; CAIN *et al.* 2000; HAMRICK & NASON 2000; DIFAZIO *et al.* 2004).

It is generally acknowledged that one of the most reliable approaches for studying effective gene flow is the application of selectively neutral genetic markers (NEIGEL 1997; SORK *et al.* 1999). Genetic markers have been employed to investigate gene flow both indirectly and directly. Indirect methods measure the level of genetic differentiation between populations (*e.g.*, F_{ST}) and infer the number of migrants that have historically been exchanged per generation (N_em). However, these methods suffer from well-known shortcomings. For example, the correlation between F_{ST} and migration rate may be quite tenuous because the assumptions of the underlying models are rarely met in real populations (*i.e.*, equilibrium between genetic drift and migration, negligible selection and mu-

tation, equal contributions of migrants from all populations; BOSSART & PROWELL 1998; WHITLOCK & MCCAULEY 1999). Another shortcoming of indirect measures is that they are not spatially explicit. At best, genetic structure-based approaches represent space in a hierarchical fashion, lacking the flexibility to incorporate factors such as spatial aggregation, anisotropy, and geographic barriers (but see SLATKIN (1993) for a method using pairwise comparisons among populations). All of these factors have been demonstrated to influence gene flow in comparative studies (MANASSE 1992; KLINGER *et al.* 1992; BERGELSON *et al.* 1993). Therefore, indirect gene flow estimates should generally be taken as long-term averages estimated over a large number of populations, and are generally not appropriate for inferring contemporaneous gene flow in specific settings (SORK *et al.* 1999).

For several practical and theoretical reasons it is interesting to study contemporary gene flow (SORK *et al.* 1999), which provides insights into prevailing population dynamics. Alterations of gene flow patterns detected through this type of analysis may give an early indication of future changes of genetic structure of populations. Several authors have attempted to measure gene dispersal from point sources in forest trees (ERICKSON & ADAMS 1989; YAZDANI *et al.* 1989). However this method almost always truncates the dispersal curve due to the difficulties of detecting propagules at low densities (ELLSTRAND 1992a; WHEELER *et al.* 1993), making such measures inadequate for extrapolation to gene flow levels between populations. The recently developed 'TwoGener' approach (SMOUSE *et al.* 2001) infers pollen dispersal distance from genetic differentiation among pollen pools of offspring samples from individual mother trees spatially dispersed on the landscape. In this approach, however, dispersal estimates are confounded with adult population structure, and the current formulations of the model are based on normal and exponential (*i.e.*, short-tailed) dispersal functions that may inadequately represent long-distance gene flow (LEWIS 1997; HIGGINS & RICHARDSON 1999).

Our current knowledge of contemporary gene flow between forest tree populations comes mainly from measures of pollen immigration into isolated (local) populations, usually seed orchards, or even to individual trees growing at low densities (receptor-oriented approach) (ELLSTRAND 1992a; WHEELER *et al.* 1993). Pollen immigration is primarily determined by genetic exclusion. In its simplest form, multilocus genotypes of offspring (or gametes) and putative parents are compared, and if an offspring (or gamete) genotype is not compatible with any adult within the local population it is considered an immigrant (apparent gene flow). This method requires that genotypes of all mature males within the local population are known, and its efficiency depends on the ability to discriminate between local and background pollen sources. If discrimination is not complete, some immigrants will not be detected, thus the estimates of background pollination will be downwardly biased. Application of simple genetic exclusion with microsatellites that have very high exclusion probabilities, has been effectively used to determine the level of pollen immigration in various oak populations (DOW & ASHLEY 1998; STREIFF *et al.* 1999; BUITJEVALD *et al.* 2001). These markers have also been used to reveal long distance pollen dispersal in other tree species (CHASE *et al.* 1996; KONUMA *et al.* 2000; KAMEYAMA *et al.* 2001; WHITE *et al.* 2002). One problem with using microsatellites in gene flow studies is a relatively high probability of genotyping errors, which we will discuss later.

When genetic markers of low exclusion probability (such as isozymes) are used for pollen gene flow studies, the proportion of 'apparent' immigrants (b) may be distinctly less than the true level of immigration (m) because at least some foreign gametes will likely have multilocus genotypes that could also be generated by local fathers (*i.e.*, cryptic gene flow). When it is possible to determine the haploid genotype of the pollen gamete in a seed (which can readily be done in conifers; MÜLLER 1976) the actual level of pollen gene flow can be estimated by $m = b/d$, where d is the probability of being able to distinguish immigrating pollen gametes (SMITH & ADAMS 1983; FRIEDMAN & ADAMS 1985; ADAMS *et al.* 1997). The parameter d is calculated as $d = 1-\Sigma g_i$, where g_i is the probability in the background pollen pool of the i-th pollen gamete genotype that can be produced by males in the local population. The g_i are estimated from allele frequencies of adult trees in background stands. While the 'conifer method' allows a straightforward correction to adjust for cryptic gene flow, this is not possible in angiosperms, or when diploid tissues of conifers are sampled. In these cases, the pollen allele at any particular locus can be determined unambiguously only if the maternal genotype is known and if the offspring is not heterozygous for the same alleles as the mother plant. Estimating m from diploid offspring genotypes requires a series of simulations to generate a data set which would have a similar proportion of 'apparent' immigrants given a particular level of true immigration (DEVLIN & ELLSTRAND 1990). These simulations require estimates of both b and allele frequencies in the background pollen pool. The above-mentioned paternity exclusion

methods have the advantage of making no assumptions about mating patterns within local populations, so the level of selfing and fertility variation among local pollen donors do not influence estimates of gene flow.

ADAMS and BIRKES (1989, 1991) developed a *neighborhood model* (probability mating model) designed to simultaneously account for the level of selfing, patterns of outcross-mating within a local population (neighborhood), and the level of pollen immigration from outside the neighborhood (gene flow). The amount of gene flow is estimated by fitting the model to the available data (local adult genotypes, offspring genotypes, allelic frequencies in the background pollen pool) based on maximum-likelihood procedures (see also BURCZYK *et al.* 1996). Other authors have since proposed similar models for conifers (XIE *et al.* 1991; STEWART 1994). While the neighborhood model was originally developed for conifers (*i.e.*, for haploid pollen gamete data) it was recently extended to utilize diploid offspring (angiosperm) data sets as well (BURCZYK *et al.* 2002).

There have been several attempts to study pollen dispersal within populations by assigning paternity to offspring based on genetic compatibility with known adult male genotypes (MEAGHER 1991; SCHNABEL & HAMRICK 1995). These studies, however, are often of little help in understanding long-distance gene flow. Though the amount of pollen generally decreases with distance from point sources (*i.e.*, male mating success decreases as the distance to females increases, see BURCZYK *et al.* 1996), a considerable proportion of pollen comes from outside local populations, at least in wind-pollinated species, and is typically not accounted for in the parentage assignment method.

ESTIMATES OF POLLEN GENE FLOW INTO FOREST TREE POPULATIONS

Pollen dispersal has long been considered a major avenue of gene flow in plant populations (HAMRICK & NASON 2000). While early studies of propagule dispersal from point sources suggested that gene flow might be quite restricted in forest trees (SILEN 1962; BRAMLETT 1981), more recent reports on the distribution of genetic variation in natural populations and on pollen contamination in seed orchards have demonstrated that gene flow can be extensive (ADAMS & BURCZYK 2000; HAMRICK & NASON 2000). There is also evidence that pollen of widely distributed forest tree species can disperse over large distances, up to tens or hundreds kilometers (HJELMROOS 1991; LINDGREN *et al.* 1995; DI-GIOVANNI *et al.* 1996; ROGERS & LEVETIN 1998). However, the effectiveness of such distant pollination may be questioned as successful fertilization, seedling establishment and survival depend on a number of environmental and genetic factors which may disfavor non-local genotypes (DYER & SORK 2001). Here we review some examples of estimates of pollen gene flow into forest tree populations. Further references can be found in earlier reviews (*e.g.* ADAMS & BURCZYK 2000; HAMRICK & NASON 2000).

Wind- pollinated trees

Effective pollen immigration into natural populations of wind-pollinated species is usually substantial, especially when studied in sample plots located within larger continuous forest stands (Table 1). In a mixed stand of oak species (*Quercus petraea* and *Q. robur*) in France, STREIFF *et al.* (1999) observed that over 65% of pollen fertilizing viable embryos originated from outside of a sample plot (5.8 ha). BURCZYK *et al.* (1996) found that nearly 56% of pollen immigrated from outside the neighborhoods (radius 11 m) of *Pinus attenuata*. Based on genetic exclusion only (*i.e.*, no correction for cryptic gene flow), apparent pollen immigration into two old growth stands of Douglas-fir (*Pseudotsuga menziesii*) was as high as 20 and 27 %, respectively (ADAMS 1992). Actual pollen gene flow, however, was probably 2–3 times greater, based on comparisons of apparent to actual gene flow estimates in other studies utilizing isozyme markers (NASON & HAMRICK 1997; ADAMS & BURCZYK 2000). In contrast to most examples of wind-pollinated trees in Table 1, SCHUSTER and MITTON (2000) estimated only 6.5 % pollen immigration into a population of *Pinus flexilis* well isolated from other scattered stands (>2 km) and a large continuous population (>100 km) of the same species. Low pollen gene flow (16 %) was also observed in an exotic plantation of Norway spruce in Ontario, Canada, isolated from another large plantation by about 4 km (XIE *et al.* 1994).

Wind-pollinated seed orchards are well known for experiencing high pollen immigration levels (pollen contamination) (SAVOLAINEN 1991; ADAMS & BURCZYK 2000). Contamination estimates often approach or exceed 40 % (FRIEDMAN & ADAMS 1985; WANG *et al.* 1991; PAKKANEN *et al.* 2000), and the estimates are only weakly related to the age, size and pollen production of seed orchards (PAKKANEN & PULKKINEN 1991; ADAMS *et al.* 1997; PAKKANEN *et al.* 2000). Isolating seed orchards from external pollen sources may be a hopeless task for broadly distributed, mass-pollinating forest tree species. For

example, HARJU and NIKKANEN (1996) estimated 48 % pollen contamination in a Scots pine seed orchard isolated at least 2 km from nearby native stands. The isolation zone contained Norway spruce stands, but scattered Scots pine individuals were also present. Similarly, BUITEVELD *et al.* (2001) detected 70 % contaminant pollen in a seed orchard of *Quercus robur*, which is comparable to the gene flow levels observed in natural oak stands (DOW & ASHLEY 1998; STREIFF *et al.* 1999). There were no oak trees present within 400 m of the *Q. robur* orchard, and only scattered individuals were growing within a 5 km radius. This led the authors to conclude that at least some of the foreign pollen could have originated from a large forest area located as far as 10 km from the seed orchard.

Animal pollinated trees

Insects are apparently very effective at promoting pollen dispersal within populations (BURCZYK *et al.* 2002; ISAGI & KANAZASHI 2002). However, gene flow in insect pollinated tree species is generally expected to be more restricted than for wind- pollinated trees because of inherent limitations on insect flight. The movement of most pollen vectors is primarily determined by foraging economy, such that dispersal distances will be limited if pollen (nectar) sources are

Table. 1. Estimates of pollen-mediated gene flow immgration into populations of forest trees based on genetic exclusion analyses (Type of population: N – natural population, SO – seed orchard, P – plantation); Marker: SSR – nuclear microsatellite(s), cpSSR – chloroplast microsatellite(s); Isoz – isozyme(s).

Species	Type of population (size)	Isolation (m)[a]	Marker (no of loci)	Immigration	Reference
Wind-pollinated					
Quercus macrocarpa	N (5 ha)	>100	SSR (4)	62	DOW & ASHLEY 1998
Quercus petraea	N (5.8 ha)	None	SSR (6)	69	STREIFF *et al.* 1999
Quercus robur	N (5.8 ha)	None	SSR (6)	65	STREIFF *et al.* 1999
Quercus robur	SO (4.5 ha)	>400	SSR (6)	70	BUITEVELD *et al.* 2001
Picea abies	SO (13.2 ha)	None	Isoz (11)	70[b]	PAKKANEN *et al.* 2000
Picea abies	P (1 ha)	>4000	Isoz (6)	16[b]	XIE & KNOWLES 1994
Pinus attenuata	N (0.04 ha)	>11	Isoz (11)	56[b]	BURCZYK *et al.* 1996
Pinus densiflora	N	>100	SSR (3)	31	LIAN *et al.* 2001
Pinus flexilis	N (15 ha)	>2000	Isoz (10)	6.5[b]	SCHUSTER & MITTON 2000
Pinus sylvestris	SO (22.9 ha)	>2000	Isoz (10)	48[b]	HARJU & NIKKANEN 1996
Pseudotsuga menziesii	N (2.4 ha)	None	Isoz (13)	20–27	ADAMS 1992
Pseudotsuga menziesii	SO (2 ha)	None	Isoz (11)	49[b]	ADAMS *et al.* 1997
Animal-pollinated					
Cordia alliodora	N (5.9 ha)	>175	Isoz (1)	~30	BOSHIER *et al.* 1995
Calophyllum longifolium	N (84 ha)	>210	Isoz (7)	62	STACY *et al.* 1996
Eucalyptus regnans	SO (0.5 ha)	>40	Isoz (10)	49–51[b]	BURCZYK *et al.* 2002
Gleditsia triacanthos	N (3 ha)	>140	Isoz (16)	17–28	SCHNABEL & HAMRICK 1995
	N (2.2 ha)	>85	Isoz (16)	19–30	
Magnolia obovata	N (69 ha)	>500	SSR (8)	56[c]	ISAGI & KANAZASHI 2002
Rhododendron metternichii	N (1 ha)	>20	SSR (11)	50[c]	KAMEYAMA *et al.* 2001
Spondias mombin	N (84 ha)	>300	Isoz (5)	2.5–5.2	STACY *et al.* 1996
Spondias mombin	N (3 ha)	>200	Isoz (11)	44[b]	NASON & HAMRICK 1997
Tachigali versicolor	N (5 ha)	>500	Isoz (1)	21[b]	LOVELESS *et al.* 1998
Turpinia occidentalis	N (50 ha)	>130	Isoz (9)	1–27	STACY *et al.* 1996
Yucca filamentosa	N (~2 ha)	>200	Isoz (10)	10	MASSEY & HAMRICK 1999

[a] distance of study plot from nearest trees of the same species,
[b] estimates adjusted for cryptic gene flow,
[c] estimate based on a sample of seedlings on the ground with diploid genotypes compatible with only one parent (presumably the mother) from within the plot, and an unknown parent outside the plot (presumably the father). All other estimates are based on seed samples from standing mother trees.

available at short distances, thus maximizing the net energy gained by the insect (LEVIN & KERSTER 1974). Nevertheless, impressive levels of pollen immigration have been observed for insect-pollinated trees (Table 1). For example, STACY *et al.* (1996) found that at least 62 % of effective pollen gametes of *Calophyllum longifolium* traveled distances greater than 200 m. Similarly, 40 % of effective pollen gametes originated from greater than 200 m for a *Spondias mombin* population (NASON & HAMRICK 1997), and 56 % of pollen gametes traveled at least 500 m for *Magnolia obovata* (ISAGI & KANAZASHI 2002). The longest effective pollen dispersal documented so far was reported for tropical figs (*Ficus* sp.), for which pollination is mediated by small, species-specific, short-lived wasps. It was found that effective pollination occurs between individuals separated by as much as 5–14 kilometers (NASON *et al.* 1996; 1998). One caveat is that the above studies all took place in tropical forests where tree densities of individual species are typically quite low, and it is unclear if these findings can be extrapolated to more continuously, and densely distributed temperate trees. For example, WHITE *et al.* (2002) demonstrated that habitat fragmentation may promote longer pollinator flights, thus diminishing potential threats (*e.g.*, loss of genetic diversity) of population fragmentation.

While high levels of pollen immigration are evident for many tree species, the origin of the foreign pollen remains unclear. Evolutionary and ecological consequences of gene flow from nearby stands is potentially quite different from the effects of gene flow from distant populations that are adapted to different environments (TEMPLETON 1986; ELLSTRAND 1992a).

CAUTIONS ON GENE FLOW ESTIMATES

Accurate estimates of pollen immigration are essential for deciding on management practices both in seed orchards and in gene conservation programs (ADAMS & BURCZYK 2000). Actions to reduce pollen contamination in seed orchards are costly, and decisions aimed at isolating plantations from outside pollen sources should be based only on sound and unbiased information. However, in reviewing available estimates of gene flow, one may raise questions about the precision of those measures, and the factors that introduce bias. Some sources of bias are obvious and have been known for decades. Non-genotyped individuals within the local population (or study plot) or genotyping errors (including both mutations and scoring mistakes) will inflate the number of 'apparent' immigrants (SAVOLAINEN 1991). Genotyping errors are particularly important when highly polymorphic markers such as microsatellites are used (WORTHINGTON WILMER *et al.* 1999). Although mutations in microsatellites are usually rare (usually less than 0.001), rates of mutation may vary among loci. RAHMAN and RAJORA (2001) found evidence of somaclonal variation at some microsatellite loci in *Populus tremuloides* clones. Null alleles of microsatellite loci can also be a significant source of error in paternity analyses (PEMBERTON *et al.* 1995) and mistyping of alleles is a common problem (MARIETTE *et al.* 2002). Thus, whenever possible, mutation rates and the presence of null alleles should be determined for newly developed microsatellites.

The probability of genotyping errors increases with the number of loci scored. When coupled with the fact that errors can occur in genotyping either parents or offspring (MARSHALL *et al.* 1998), the presence of errors may become quite meaningful, most likely leading to overestimates of gene flow. It is conceivable that a genotyping error could result in identifying an offspring as local, even though it is an immigrant, although it is more likely to be the other way around.

The probability of genotyping errors can be estimated. If ε is the probability of misidentifying an allele, with the probability of correctly genotyping an allele is $1-\varepsilon$, then the probability of correctly genotyping both alleles in a diploid individual is $(1-\varepsilon)^2$. Assuming that genotyping errors are independent over alleles, loci and individuals, the total probability of error (E) across loci and genotyped individuals is:

$$E = 1 - \prod_{i=1}^{n} \prod_{i=1}^{x} (1 - \varepsilon_{ij})^2$$

where ε_{ij} is the probability of error at the j-th locus of the i-th individual, n is the number of genotyped individuals (n = 2 for a pair (*i.e.*, offspring – single parent), or n = 3 for triplet (offspring – male parent – female parent) data), and x is the number of loci scored. Assuming error rates to be constant over loci and individuals, the above equation may be simplified:

$$E = 1 - (1 - \varepsilon)^{2nx}$$

For example, consider a seed from a known mother (genotype known) and unknown father, so that genotypes of pair of individuals, the seed and a putative male, must be compared to infer paternity. With 5 microsatellite loci and a mean error per allele as high as 0.01, which is considered to be realistic for

microsatellites (MARSHALL *et al.* 1998), the expected rate of error for any seed-male parent pair is $E = 1-(1-0.01)^{20} = 0.18$. Given that nearly all errors will lead to incorrect inference of parentage, paternity assignments in this example would be wrong in 18 % of the cases, resulting in substantial overestimation of gene flow. If pollen gametes, rather than diploid genotypes of offspring are assayed in the above example, n is 1.5, rather than 2, and $E = 0.14$; not much less than when diploid offspring must be genotyped. In the case where seed or seedling offspring on the ground are assayed and both parents must be inferred (*i.e.*, triplet data), the potential for misidentification of parentage (again assuming 5 loci) is increased by nearly 50 % ($n = 3$, $E = 0.26$). Even with an error rate as high as 0.01 for identifying an allele at a single locus of a haploid gamete (or $1-(1-0.01)^2 = 0.02$ for one or both alleles at a diploid locus), the probability of simultaneous error at two or more loci is extremely low; *i.e.*, less than $(0.01)^2 = 0.0001$ (or < 0.0004). Thus a potential approach to limiting mismatch errors and reducing the upward bias of gene flow estimates is to only infer a mismatch unless it occurs at more than one or two loci (depending on the number of loci scored and their error rates). With highly variable SSRs it is unlikely that a true match at one locus would not be detected at other loci as well (MARSHALL *et al.* 1998; WORTHINGTON WILMER *et al.* 1999). Surprisingly, there have been few studies investigating the effect of genotyping errors on the efficiency of gene flow estimation (MARSHALL *et al.* 1998).

Isozymes are probably less vulnerable to genotyping errors compared to microsatellites, but as mentioned earlier, their low levels of polymorphism necessitate correcting the observed proportion of immigrants for cryptic gene flow. One weakness of this approach is that in order to account for 'cryptic' gene flow estimates of allelic frequencies in background pollen pools are required (see *Methods of estimating gene flow*). Those estimates can be obtained by randomly sampling genotypes in surrounding adult populations (SMITH & ADAMS 1983; FRIEDMAN & ADAMS 1985). However, since individual variation in pollen fecundity is extensive and sources of background pollen are difficult to define explicitly, such estimates may not accurately account for allele frequencies in the immigrating pollen pool, thus leading to biased gene flow estimates (XIE *et al.* 1991). The methods of SMITH and ADAMS (1983) and DEVLIN and ELLSTRAND (1990) are quite commonly used for the estimation of gene flow in forest trees, especially when coupled with isozymes as genetic markers (see references in Table 1, and also cited in ADAMS and BURCZYK 2000; HAMRICK and NASON 2000). However, we suspect that at least some of these estimates could be biased if the presumed allele frequencies in the background pollen pool differed from the actual ones.

Mating models allow for simultaneous estimation of background pollination (m) and other mating parameters (ADAMS & BIRKES 1991), as well as background allele frequencies. This ensures that the estimate of m accounts for total gene flow (*i.e.*, apparent and cryptic) (XIE *et al.* 1991, BURCZYK *et al.* 2002). This approach not only eliminates the necessity of additional sampling in surrounding populations, thus reducing the costs of sampling and laboratory analyses, but also gives more accurate pollen gene flow estimates, especially if the true immigration levels are large and polymorphism is low (BURCZYK *et al.* 2002). Such mating models, however, are more complicated to apply than the genetic exclusion methods described earlier. Also, mating patterns within the local population must be assumed. The mating model of XIE *et al.* (1991) assumes that all mating within the local population is either selfing or random outcrossing, but several studies have indicated that random mating within populations is rarely met. The neighborhood model (ADAMS & BIRKES 1989, 1991; BURCZYK *et al.* 2002), however, can utilize additional information on the potential mating success of males within the local population, including distance of males from females, relative pollen fertility, etc., so that mating patterns among local trees can be more precisely accounted for.

REALIZED GENE FLOW – THE SEEDLING PERSPECTIVE

The usual way to estimate pollen gene flow based on genetic markers is to use progeny arrays (seeds) sampled from individual mother trees. This procedure provides the estimates of effective pollen gene flow observed at the seed stage only, and such information is quite appropriate if seeds collected from individual trees are to be used for artificial reforestation. Nevertheless, realized pollen gene flow (MEAGHER & THOMPSON 1987) observed at the stage of naturally regenerated seedlings, may substantially differ from effective pollen gene flow detected from seed samples. Several factors such as seed dispersal mechanisms, seed predation, and subsequent selection may influence the genetic composition of naturally regenerated seedlings. For example, if immigrating pollen comes from populations adapted

to substantially different environmental conditions (*e.g.*, colder or dryer climates), the resulting offspring may fail in competition with offspring derived from local matings.

Gene flow and parentage analyses of naturally regenerated seedling cohorts are among the greatest challenges facing plant population geneticists (MEAGHER & THOMPSON 1987). The results of such analyses may provide important insights into the dynamics of natural populations. However, because of the high amount of genetic exclusion power required for such studies (MARSHALL *et al.* 1998), there have been few attempts to study parentage of seedlings in forest tree species (but see YAZDANI *et al.* 1989; DOW & ASHLEY 1996; SCHNABEL *et al.* 1998; KAMEYAMA *et al.* 2001; GONZALEZ-MARTINEZ 2002; ISAGI & KANAZASHI 2002).

In an attempt to compensate for low exclusion power available in most studies, we have developed the *Seedling Neighborhood Model*, a probability model that attempts to describe the genetic composition of naturally regenerating seedling cohorts (Burczyk J., Adams W.T. & Birkes D.S., in preparation). The model is based on the earlier neighborhood models developed for pollen dispersal (ADAMS & BIRKES 1989, 1991; BURCZYK *et al.* 2002) and requires that the location of seedlings and all potential male and female parents within the local population be known, as well as the multilocus genotypes of all seedlings and potential parents and background allele frequencies. The model allows simultaneous estimation of levels of seed and pollen immigration, along with female and male reproductive success parameters. Additionally, it accounts for the proportion of selfing among the fraction of seedlings originating from local mothers. The procedure can be readily applied to seedlings but also to seeds (embryos) sampled on the ground or from seed traps, when seed dispersal parameters are investigated.

The model assumes that each seedling is mothered by (i) a female located outside of the arbitrarily defined local population (the seedling's neighborhood) due to seed immigration (with probability m_s) or (ii) by a local female growing within the seedling's neighborhood (with probability $1-m_s$). The neighborhood is defined as a circular area surrounding the location of a seedling. For each seedling with a local mother it is assumed that the paternal gamete could have come from one of three sources: (i) self-fertilization (with probability s), (ii) migrant pollen from outside of the mother's neighborhood (with probability m_p), or (iii) outcross fertilization with males located within the neighborhood (with probability $1-s-m_p$), much like earlier neighborhood models for pollen dispersal (ADAMS & BIRKES 1991; BURCZYK *et al.* 2002). The probability of observing a multilocus diploid genotype G_i among the seedlings, therefore, is:

$$P(G_i) = m_s \cdot P(G_i|B_j) + (1-m_s) \cdot \sum \psi_{ij} [s \cdot P(G_i|M_{ij},M_{ij}) + m_p P(G_i|M_{ij},B_p) + (1-s-m_p) \cdot \sum \varphi_{ijk} s P(G_i,M_{ij},F_{ijk})]$$

where $P(G_i|M_{ij},M_{ij})$, $P(G_i|M_{ij},B_s)$, $P(G_i|M_{ij},F_{ijk})$ are the genetic segregation or *transition* probabilities (DEVLIN *et al.* 1988); *i.e.*, the probabilities that a seedling has diploid genotype G_i when a mother plant of genotype M_{ij} is, respectively, self-pollinated, pollinated by a distant unknown background male, or pollinated by a neighboring plant having genotype F_{ijk}. $P(G_i|B_p)$ is the transition probability that a seedling immigrating from mother trees located outside of a seedling's neighborhood has genotype G_i. ψ_{ij} is the relative reproductive success of the j-th female in the neighborhood of the i-th seedling, and φ_{ijk} is the relative reproductive success of the k-th male within the neighborhood of the ij-th female.

By fitting the model to the multilocus genotypic data of seedlings and adult trees in the local population, the model parameters, including m_s, m_p, s and various additional terms relating female (ψ_{ij}) and male (φ_{ijk}) reproductive success to various factors likely to influence reproductive success (called selection gradients by MORGAN & CONNER 2001) can be estimated (see ADAMS 1992, BURCZYK *et al.* 1996, BURCZYK *et al.* 2002, for explanations). Hence, the unique feature of the *seedling neighborhood model* is that it can be applied to simultaneously estimating selection gradients of a given phenotypic trait (*e.g.*, tree size, age, flower fecundity) for both male and female functions, making it possible to evaluate the importance of each trait to male and female reproductive success. Despite the complexity of the model, it is possible to reliably estimate a relatively large number of model parameters when data sets are appropriate (*i.e.*, large sample sizes and high exclusion probability).

A preliminary application of this model to microsatellite data of naturally regenerated *Pinus pinaster* seedlings is promising (GONZALEZ-MARTINEZ *et al.* 2001). The authors attempted to estimate several parameters, including the levels of seed and pollen immigration, as well as gradients relating female reproductive success to the distance between seedlings and mother trees, and male reproductive success to distance between males and females. The estimates reasonably met expectations. The level of seed immigration from outside a 50 m neighborhood

(*i.e.*, from greater than 50 m away from the location of the seedling) was estimated to be m_s=0.56, while pollen immigration observed for the fraction of the seedlings originating from local mothers (*i.e.*, effective pollen coming from greater than 50 m from the mother tree) was as high as 85 %. The probability of maternity increased with decreasing distance between seedlings and potential females. Additionally, stem diameter appeared to be an important determinant of female but not male reproductive success (Gonzalez-Martinez, S.C., personal communication). Computer simulations indicated that the seedling model gives reasonably robust estimates of reproductive parameters when genetic markers with at least moderate exclusion probabilities (EP>0.8) are available (BURCZYK *et al.*, in preparation). However, it is necessary to further explore the efficiency of the model under a wide variety of mating systems and reproductive success gradients.

GENE FLOW AT A LANDSCAPE LEVEL

The methods currently available can provide intricately detailed pictures of gene dispersal and reproductive success within small populations or within portions of larger populations. They also make it possible to estimate gene flow into local populations from background sources. However, current methods provide little information on dynamics of gene flow on a landscape scale, *i.e.*, the extent and distribution of effective gene dispersal among populations. Direct observation of long-distance gene flow is made particularly difficult, because relevant events are usually infrequent and mostly stochastic. Furthermore, the process of gene flow is extremely complex, involving the full range of life history stages from seed bank through seedlings to mature individuals, and incorporating all of the biotic and abiotic interactions that affect mating, propagule production, dispersal, establishment, survival, and maturation. All of these factors must be taken into consideration when extrapolating gene flow through space or time.

Spatial simulation modeling provides one means of accomplishing such extrapolation (KING 1991; DUNNING *et al.* 1995). Models provide an extensible framework for integrating data from disparate demographic and genetic field studies with landscape-scale analyses of ecosystem dynamics (SORK *et al.* 1999; HIGGINS *et al.* 2000). In addition, such models allow 'virtual experiments' through sensitivity analyses in which selected components of the system are manipulated to determine their importance in determining long-term outcomes (TURNER *et al.* 2001).

Many different types of models have been used for simulating gene flow across a landscape. One approach is to devise mechanistic models of pollen and seed dispersal based on the physical properties of the propagules and the environment (GREENE & JOHNSON 2000; NATHAN *et al.* 2001). Such models have a distinct advantage in that they are easily parameterized for a large number of species because flight characteristics of pollen and seeds are readily measured, detailed microclimatic data can be obtained for many sites, and the physics of dispersal by abiotic agents are fairly well-characterized (NATHAN *et al.* 2001). Disadvantages include the large number of parameters that require estimation (particularly if realized gene flow is to be modeled), and the high computational requirements that limit the extent of the area and time frame that can be modeled.

An alternative approach is to model gene flow phenomenologically based on field observations of dispersal and demographic processes. A common method is to use reaction-diffusion models to depict the movement of an 'invasion front' using a diffusion approximation and logistic growth models (FISHER 1937; ANDOW *et al.* 1990). Alternatively, probability density functions of propagule movement and/or reproductive success can be used to determine the probability of dispersal between points on a lattice of habitat cells (HIGGINS & RICHARDSON 1996; LAVOREL *et al.* 1999). This approach has the advantage of being easily parameterized from historical data (*e.g.*, a chronosequence of air photos, or survey data), and readily integrated with geographic information systems (GIS). A major disadvantage is the difficulty of measuring contemporaneous realized gene flow on appropriate space and time scales to parameterize the models, as described above.

As an example of the latter approach, we developed a spatially-explicit model of gene flow from hybrid poplar plantations based on observations of realized gene flow in wild populations (DIFAZIO 2002). The model, called STEVE (Simulation of Transgene Effects in a Variable Environment), was applied to a landscape grid in northwest Oregon (23 km × 37 km, 100 m^2 cells) containing information about elevation, habitat type, and poplar populations. The simulation has an annual time step, with modules to simulate creation and conversion of poplar patches, growth, reproduction, dispersal, and competition within poplar cohorts. Our field studies of gene flow indicated that long-distance dispersal is considerable for *Populus*, with an apparently panmictic regional pollen cloud accounting for nearly half of the successful pollinations. We therefore chose to model gene dispersal as a two-stage process, with local dis-

persal modeled explicitly by a negative exponential distribution, and long-distance dispersal modeled as a uniform distribution with probability of pollination independent of distance between mates. This is analogous to a 'mixed model' approach, as advocated by CLARK *et al.* (1998) and HIGGINS and RICHARDSON (1999).

The STEVE model has proved useful for exploring experimentally intractable aspects of potential escape of transgenes (*e.g.*, for herbicide tolerance or insect resistance) from genetically engineered trees in hybrid poplar plantations into wild populations of native poplar. For example, using sensitivity analyses, we investigated landscape-level implications of factors such as fertility and relative competitiveness of transgenic trees, variable rates of disturbance and habitat creation, silvicultural options such as rotation length, gender of cultivated clones, and area and placement of transgenic plantations, and case studies for traits such as herbicide resistance and insect resistance. In so doing, we confirmed the previously suspected importance of rare, long-distance dispersal events in determining impacts of transgenic plantations, and the strong influence of transgenic competitiveness and habitat availability in determining extent of transgene movement (DIFAZIO 2002; SLAVOV *et al.* 2002; DIFAZIO *et al.* 2004). The model also yielded some surprising results, including:

Herbicide tolerance had little effect on gene flow from hybrid plantations, and only marginal effects on transgenic poplar establishment in agricultural fields. In contrast, insect resistance had the potential to greatly enhance gene flow when insect pressure was high.

Reducing the relative fertility of transgenic trees by moderate amounts (*e.g.*, 90 %), greatly reduced gene flow, even with substantial transgenic competitive advantages.

Reductions in transgenic fertility below about 1 % had little effect on transgene flow. This was due to a lack of pollen competition from wild trees in the interiors of large plantations containing transgenic males and conventional females. Even though the total amount of pollen reaching conventional plantation females was very low (due to low transgenic male fertility), a large number of transgenic seeds was produced because the model assumed no pollen limitation on seed production.

These findings have provided insights into the processes of gene flow from plantations and should help to guide further risk assessment efforts. However, models are only as accurate as their underlying hypotheses and the quality of the data used for parameterization. Therefore, further large-scale experimentation is required to validate model predictions and refine gene flow estimates. Verification of large-scale dynamics can be accomplished by tracking spread of seedlings from exotic plantations (HIGGINS & RICHARDSON 1998), or tracking the spread of transgenic trees containing easily assayed marker genes. If verification is possible, large-scale spatial simulation models can provide a much-needed tool to explore the multi-faceted processes of gene flow on scales that are relevant for evolutionary biologists, ecologists, land managers and policy makers.

CONCLUSIONS

Gene flow is considered one of the major factors maintaining the high levels of genetic diversity found within forest tree populations. Its importance is well acknowledged both in natural and breeding populations, and the extent of gene flow within and among populations has attracted the attention of many researchers in recent decades. However, despite great progress in genetic marker technology and estimation procedures, several problems remain and some new questions have arisen.

One consideration is the amount of gene flow from plantations to natural stands. In some cases, extensive plantations are cultivated in close proximity to stands of interfertile native trees. If the plantations exhibit lower genetic diversity (as in the case of clonally propagated species like *Populus*), gene flow from such populations may have negative impacts on natural stands, potentially reducing diversity and adaptability of later generations (ADAMS & BURCZYK 2000). Unfortunately, there are no simple ways to measure gene flow from plantations to natural populations (DIFAZIO *et al.* 2004), so this issue poses a major challenge to forest geneticists and conservationists.

Understanding realized gene flow at the stage of naturally regenerated seedlings should receive high priority. Revealing patterns and determinants of gene dispersal and reproductive success is fundamental to understanding the genetic aspects of natural regeneration processes in plant populations (MEAGHER & THOMPSON 1987). This will lead to enhanced recognition of the interactions among genetics, ecology and demography at the ecosystem level. Such information is of primary interest in genetic conservation programs, where natural regeneration is the major mode of reproduction.

The relationship between distant gene flow and adaptation needs to be investigated in more detail (*e.g.* GARCIA-RAMOS & KIRKPATRICK 1997). For example,

level. Such information is of primary interest in genetic conservation programs, where natural regeneration is the major mode of reproduction.

The relationship between distant gene flow and adaptation needs to be investigated in more detail (*e.g.* GARCIA-RAMOS & KIRKPATRICK 1997). For example, the effects of long distance gene flow on adaptation and competition of immigrants in a new habitat is poorly understood. MORGAN (2002) suggested that genome-wide deleterious mutations existing within populations may reduce fitness of local matings due to increased probability of homozygosity for deleterious alleles, thus favoring distant gene flow. The advantage of immigrants is expected to increase with the extent of local inbreeding depression (MORGAN 2002, see also references therein). It remains to be seen whether these ideas are relevant to forest trees, which are known to suffer from high inbreeding depression (WILLIAMS & SAVOLAINEN 1996), but which also harbor substantial genetic diversity within populations. More information is needed on other mechanisms favoring immigration, and the effects of anthropogenic habitat fragmentation on gene flow.

Gene flow in forest trees, and plants in general, is often considered idiosyncratic, varying among species, populations and even individuals (ELLSTRAND 1992a). Nearly two decades ago, SLATKIN (1981) wrote that "there is no way to determine the importance of gene flow in natural populations because there is no direct way to estimate levels of gene flow". Now, however, we anticipate that a more comprehensive picture of gene flow and the genetic systems of forest trees will emerge from the increasing number of studies on actual gene flow, new insights from mitochondrial and chloroplast DNA data, and advances in statistical modeling.

REFERENCES

ADAMS, W.T. 1992: Gene dispersal within forest tree populations. *New Forests* **6**:217–240.

ADAMS, W.T. & BIRKES, D.S. 1989: Mating patterns in seed orchards. *In*: Proceedings of the 20th Southern Forest Tree Improvement Conference, pp. 75–86, Charleston.

ADAMS, W.T. & BIRKES, D.S. 1991: Estimating mating patterns in forest tree populations. *In*: Biochemical Markers in the Population Genetics of Forest Trees (eds. Fineschi, S., Malvolti, M. E., Cannata, F. & Hattemer, H.H.), pp. 157–172, SPB Academic Publishing, The Hague, The Netherlands.

ADAMS, W.T. & BURCZYK, J. 2000: Magnitude and implications of gene flow in gene conservation reserves. *In*: Forest Conservation Genetics: Principles and Practice (eds. Young, A., Boshier, D. & Boyle, T.), pp. 215–224, CSIRO Publishing, Collingwood, Australia.

ADAMS, W.T., HIPKINS, V.D., BURCZYK, J. & RANDALL, W.K. 1997: Pollen contamination trends in a maturing Douglas-fir seed orchards. *Can. J. For. Res.* **27**: 131–134.

ANDOW, D., KAREIVA, P., LEVIN, S.A. & OKUBO, A. 1990: Spread of invading organisms. *Landsc. Ecol.* **4**: 177–188.

BATEMAN, A.J. 1947: Contamination of seed crops II. Wind pollination. *Heredity* **1**: 235–246.

BERGELSON, J., NEWMAN, J.A. & FLORESROUX, E.M. 1993: Rates of weed spread in spatially heterogeneous environments. *Ecology* **74**: 999–1011.

BOSHIER, D.H., CHASE, M.R. & BAWA, K.S. 1995: Population genetics of *Cordia alliodora* (Boraginaceae), a neotropical tree. 3. Gene flow, neighborhood, and population structure. *Am. J. Bot.* **82**: 484–490.

BOSSART, J.L. & PROWELL, D.P. 1998: Genetic estimates of population structure and gene flow: limitations, lessons, and new directions. *Trends in Ecol. Evol.* **13**: 202–206.

BRAMLETT, D.L. 1981: Effectiveness of wind pollination in seed orchards. *In*: Pollen Management Handbook (ed. Franklin, E. C.), pp. 10–14, USDA Forest Service, Agric. Handbook 587.

BUITEVELD, J., BAKKER, *E.G.*, BOVENSCHEN, J. & VRIES DE, S.M.G. 2001: Paternity analysis in a seed orchard of *Quercus robur* L. and estimation of the amount of background pollination using microsatellite markers. *For. Genet.* **8**: 331–337.

BURCZYK. J., ADAMS, W.T. & SHIMIZU, J.Y. 1996: Mating patterns and pollen dispersal in a natural knobcone pine (*Pinus attenuata* Lemmon) stand. *Heredity* **77**: 251–260.

BURCZYK, J., ADAMS, W.T., MORAN, G.F. & GRIFFIN, A.R. 2002: Complex patterns of mating revealed in a *Eucalyptus regnans* seed orchard using allozyme markers and the neighborhood model. *Mol. Ecol.* **11**: 2379–2391.

CAIN, M.L., MILLIGAN, B.G. & STRAND, A.E. 2000: Long-distance seed dispersal in plant populations. *Am. J. Bot.* **87**: 1217–1227.

CHASE, M., KESSELI, R. & BAWA, K. 1996: Microsatellite markers for population and conservation genetics of tropical trees. *Am. J. Bot.* **83**: 51–57.

CHARNOV, E. L. 1982: The Theory of Sex Allocation. Princeton University Press, Princeton, N.J.

CLARK, J. S., FASTIE, C., HURTT, G., JACKSON, S. T., JOHNSON, C., KING, G.A., LEWIS, M., LYNCH, J., PACALA, S., PRENTICE, C., SCHUPP, E.W., WEBB, T.I. & WYCKOFF, P. 1998: Reid's paradox of rapid plant migration: dispersal theory and interpretation of palaeoecological records. *BioScience* **48**:13–24.

CONDIT, R., PITMAN, N., LEIGH, E. G. JR., CHAVE, J., TERBORGH, J., FOSTER, R. B., NUNEZ, P., AGUILAR, S., VALENCIA, R., VILLA, G., MULLER-LANDAU, H. C., LOSOS, E. & HUBBELL, S. P. 2002: Beta-Diversity in Tropical Forest Trees. *Science* **295**: 666–669.

CONNELL, J.H. & SLATYER, R.O. 1977: Mechanisms of selection in natural communities and their role in community stability and organization. *Am. Natur.* **111**:

1119 –1144.

DEVLIN, B. & ELLSTRAND, N.C. 1990: The development and application of a refined method for estimating gene flow from angiosperm paternity analysis. *Evolution* **44**: 248–259.

DEVLIN, B., ROEDER, K. & ELLSTRAND, N.C. 1988: Fractional paternity assignment: theoretical development and comparison to other methods. *Theor. Appl. Genet.* **76**: 369–380.

DI-GIOVANNI, F. & KEVAN, P.G. 1991: Factors affecting pollen dynamics and its importance to pollen contamination: a review. *Can. J. For. Res.* **21**: 1155 –1170.

DI-GIOVANNI, F., KEVAN, P. G. & ARNOLD, J. 1996: Lower planetary boundary layer profiles of atmospheric conifer pollen above a seed orchard in northern Ontario, Canada. *For. Ecol. Manage.* **83**: 87–97.

DIFAZIO, S. P. 2002: Measuring and Modeling Gene Flow from Hybrid Poplar Plantations: Implications for Transgenic Risk Assessment. PhD Thesis, Oregon State University, Corvallis, OR. *http://www.fsl.orst.edu/tgerc/dif_thesis/difaz_thesis.pdf*

DIFAZIO, S. P., SLAVOV, G. T., BURCZYK, J., LEONARDI, S. & STRAUSS, S. H. 2004: Gene Flow From Tree Plantations and Implications for Transgenic Risk Assessment. *In*: Forest Biotechnology for the 21st Century (eds. Walter, C. & Carson, M.), Research Signpost, Kerala, India. pp. (in press).

DOW, B. D. & ASHLEY, M. V. 1996: Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa*. *Mol. Ecol.* **5**: 120–130.

DOW, B. D. & ASHLEY, M. V. 1998: High levels of gene flow in bur oak revealed by paternity analysis using microsatellites. *J. Hered.* **89**: 62–70.

DUNNING, J. B., STEWART, D. J., DANIELSON, B. J., NOON, B. R., ROOT, T. L., LAMBERSON, R. H. & STEVENS, E. E. 1995: Spatially explicit population models: Current forms and future uses. *Ecol. Applic.* **5**: 3–11.

DYER, R. J. & SORK, V. L. 2001: Pollen pool heterogeneity in shortleaf pine, *Pinus echinata* Mill. *Mol. Ecol.* **10**: 859 –866.

EHRLEN, J. & Eriksson, O. 2000: Dispersal limitation and patch occupancy in forest herbs 29. *Ecology* **81**: 1667 –1674.

ELLSTRAND, N. C. 1992a: Gene flow among seed plant populations. *New Forests* **6**: 241–256.

ELLSTRAND, N. C. 1992b: Gene flow by pollen: Implications for plant conservation genetics. *Oikos* **63**: 77–86.

ERICKSON, V. J. & ADAMS, W. T. 1989: Mating success in a coastal Douglas-fir seed orchard as affected by distance and floral phenology. *Can. J. For. Res.* **19**: 1248–1255.

FERDY, J. -B. & AUSTERLITZ, F. 2002: Extinction and introgression in a community of partially cross-fertile plant species. *Am. Natur.* **160**:74–86.

FISHER, R. A. 1937: The wave of advance of advantageous genes. *Ann. Eugen.* **7**: 355–369.

FRIEDMAN, S. T. & ADAMS, W. T. 1985: Estimation of gene flow into two seed orchards of loblolly pine (*Pinus taeda* L.). *Theor. Appl. Genet.* **69**: 609–615.

GARCIA-RAMOS, G. & KIRKPATRICK, M. 1997: Genetic models of adaptation and gene flow in peripheral populations. *Evolution* **51**:21–28.

GONZÁLEZ-MARTÍNEZ, S. C., BURCZYK, J., ROBLEDO, J. J., GIL, L. & R. ALÍA. 2001: Sistemas de reproducción y flujo genético: aplicación a la mejora genética de *Pinus pinaster* Ait. *In*: Proceedings of the III Spanish Forestry Conference. September 25–28, Granada. Volume II. Pp. 572–578.

GONZALEZ-MARTINEZ, S. C., GERBER, S., CERVERA, M. T., MARTINEZ-ZAPATER, J. M., GIL, L. AND ALIA, R. 2002: Seed gene flow and fine–scale structure in a Mediterranean pine (*Pinus pinaster* Ait.) using nuclear microsatellite markers. *Theor. Appl. Genet.* **104**:1290 –1297.

GREENE, D. F. & JOHNSON, E. A. 2000: Tree recruitment from burn edges. *Can. J. For. Res.* **30**: 1264–1274.

HAMRICK, J. L. & NASON, J. D. 2000: Gene flow in forest trees. *In*: Forest Conservation Genetics: Principles and Practice (eds. Young, A., Boshier, D. & Boyle, T.), pp. 81–90, CSIRO Publishing, Collingwood, Australia.

HANSKI, I. & GILPIN, M. 1997: Metapopulation Biology: Ecology, Genetics, and Evolution. Academic Press, London, UK.

HARJU, A. M. & NIKKANEN, T. 1996: Reproductive success of orchard and nonorchard pollens during different stages of pollen shedding in a Scots pine seed orchard. *Can. J. For. Res.* **26**: 1096–1102.

HIGGINS, S. I. & RICHARDSON, D. M. 1996: A review of models of alien plant spread. *Ecol. Modell.* **87**:249 –265.

HIGGINS, S. I. & RICHARDSON, D. M. 1998: Pine invasions in the southern hemisphere: modelling interactions between organism, environment and disturbance. *Plant Ecol.* **135**: 79–93.

HIGGINS, S. I. & RICHARDSON, D. M. 1999: Predicting plant migration rates in a changing world: The role of long-distance dispersal. *Am. Natur.***153**: 464–475.

HIGGINS, S.I., RICHARDSON, D.M. & COWLING, R.M. 2000: Using a dynamic landscape model for planning the management of alien plant invasions. *Ecol. Applic.* **10**: 1833–1848.

HJELMROOS, M. 1991: Evidence of long-distance transport of *Betula* pollen. *Grana* **30**: 215–228.

ISAGI, Y. & KANAZASHI, T. 2002: Gene flow analysis in *Magnolia obovata* Thunb. using highly variable microsatellite markers. *In*: Diversity and interaction in a temperate forest community: Ogawa Forest Reserve of Japan (ed. Matsumoto, N.), pp. 257–269, Springer-Verlag, Tokyo.

KAMEYAMA, Y., ISAGI, Y. & NAKAGOSHI, N. 2001: Patterns and levels of gene flow in *Rhododendron metternichii* var. *hondoense* revealed by microsatellite analysis. *Mol. Ecol.* **10**: 205–216.

KING, A.W. 1991: Translating models across scales in the landscape. *In*: Quantitative Methods in Landscape Ecology. Springer-Verlag, New York. pp. 479–517.

KLINGER, T., ARRIOLA, P.E. & ELLSTRAND, N.C. 1992: Crop-weed hybridization in radish (*Raphanus sativus*): Effects of distance and population size. *Am. J. Bot.* **79**: 1431–1435.

KONUMA, A., TSUMURA, Y., LEE, C.T., LEE, S.L. & OKUDA, T. 2000: Estimation of gene flow in the tropical-rainforest tree *Neobalanocarpus heimii* (Dipterocarpaceae),

inferred from paternity analysis. *Mol. Ecol.* **9**: 1843–1852.

LANDE, R. 1988: Genetics and demography in biological conservation. *Science* **241**: 1455–1460.

LAVOREL, S., SMITH, M.S. & REID, N. 1999: Spread of mistletoes (*Amyema preissii*) in fragmented Australian woodlands: a simulation study. *Landsc. Ecol.* **14**: 147–160.

LENORMAND, T. 2002: Gene flow and the limits to natural selection. *Trends Ecol. Evol.* **17**: 183.

LEVIN, D.A. & KERSTER, H.W. 1974: Gene flow in seed plants. *In*: Evolutionary Biology 7. (eds. Dobzhansky, T. M., Hecht, M. T. & Steere, W. C.), pp. 139–220, Plenum Press, New York.

LEWIS, M. 1997: Variability, patchiness, and jump dispersal in the spread of an invading organism. *In* Spatial Ecology: The role of space in population dynamics and interspecific interactions. Princeton University Press, Princeton. pp. 46–69.

LEWONTIN, R. 1997: Dobzhansky's *Genetics and the Origin of Species*: Is it still relevant? *Genetics* **147**: 351–355.

LIAN, C.L., MIWA, M. & HOGETSU, T. 2001: Outcrossing and paternity analysis of *Pinus densiflora* (Japanese red pine) by microsatellite polymorphism. *Heredity* **87**: 88–98.

LINDGREN, D., PAULE, L., SHEN, X.H., YAZDANI, R., SEGERSTROM, U., WALLIN, J.E. & LEJDEBRO, M.L. 1995: Can viable pollen carry Scots pine genes over long distances? *Grana* **34**: 64–69.

LOVELESS, M.D., HAMRICK, J.L. & FOSTER, R.B. 1998: Population structure and mating system in *Tachigali versicolor*, a monocarpic neotropical tree. *Heredity* **81**: 134–143.

MANASSE, R.S. 1992: Ecological risks of transgenic plants: effects of spatial dispersion on gene flow. *Ecol. Applic.* **2**: 431–438.

MARIETTE, S., COTTRELL, J., CSAIKL, U.M., GOIKOECHEA, P., KONIG, A., LOWE, A.J., VAN DAM, B.C., BARRENECHE, T., BODENES, C., STREIFF, R., BURG, K., GROPPE, K., MUNRO, R.C., TABBENER, H. & KREMER, A. 2002: Comparison of levels of genetic diversity detected with AFLP and microsatellite markers within and among mixed *Q. petraea* (Matt.) Liebl. and *Q. robur* L. stands. *Silvae Genet.* **52**: 72–79.

MARSHALL, T.C., SLATE, J., KRUUK, L.E. & PEMBERTON, J.M. 1998: Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* **7**: 639–655.

MASSEY, L.K. & HAMRICK, J.L. 1999: Breeding structure of a *Yucca filamentosa* (Agavaceae) population. *Evolution* **53:** 1293–1298.

MEAGHER, T.R. 1991: Analysis of paternity within a natural population of *Chamaelirium luteum*. II. Patterns of male reproductive success. *Am. Natur.* **137**: 738–752.

MEAGHER, T.R. & THOMPSON, E. 1987: Analysis of parentage for naturally established seedlings of *Chamaelirium luteum* (Liliaceae). *Ecology* **68:** 803–812.

MILLAR, C.I. & LIBBY, W.J. 1991: Strategies for conserving clonal, ecotypic, and disjunct population diversity in widespread species. *In*: Genetics and Conservation of Rare Plants. Oxford University Press, Oxford. pp. 149–170.

MORGAN, M.T. 2002: Genome-wide deleterious mutations favors dispersal and species integrity. *Heredity* **89**: 253–257.

MORGAN, M.T. & CONNER, J.K. 2001: Using genetic markers to directly estimate male selection gradients. *Evolution* **55**: 272–281.

MÜLLER, G. 1976: A simple method of estimating rates of self-fertilization by analyzing isozymes in trees seeds. *Silvae Genet.* **25**: 15–17.

NASON, J.D. & HAMRICK, J.L. 1997: Reproductive and genetic consequences of forest fragmentation: Two case studies of neotropical canopy trees. *J. Hered.* **88**: 264–276.

NASON, J.D., HERRE, E.A. & HAMRICK, J.L. 1996: Paternity analysis of the breeding structure of strangler fig populations: Evidence for substantial long-distance wasp dispersal. *J. Biogeogr.* **23**: 501–512.

NASON, J.D., HERRE, E.A. & HAMRICK, J.L. 1998: The breeding structure of a tropical keystone plant resource. *Nature* **391**: 685–687.

NATHAN, R., SAFRIEL, U.N. & NOY MEIR, I. 2001: Field validation and sensitivity analysis of a mechanistic model for tree seed dispersal by wind. *Ecology* **82**: 374–388.

NEIGEL, J.E. 1997: A comparison of alternative strategies for estimating gene flow from genetic markers. *Ann. Rev. Ecol. System.* **28**: 105–128.

NIKLAS, K.J. 1997: The Evolutionary Biology of Plants. University of Chicago Press, Chicago.

OUBORG, N.J., PIQUOT, Y. & GROENENDAEL, J.M. 1999: Population genetics, molecular markers and the study of dispersal in plants. *J. Ecol.* **87**: 551–568.

PAKKANEN, A. & PULKKINEN, P. 1991: Pollen production and background pollination levels in Scots pine seed orchards of northern Finnish origin. *In*: Pollen Contamination in seed orchards (ed. Lindgren, D.), Proceedings of the Meeting of the Nordic Group for Tree Breeding 1991, pp. 14–21, Swedish University of Agricultural Sciences, Department of Forest Genetics and Plant Physiology, Umeå.

PAKKANEN, A., NIKKANEN, T. & PULKKINEN, P. 2000: Annual variation in pollen contamination and outcrossing in a *Picea abies* seed orchard. *Scand. J. For. Res.* **15:** 399–404.

PEMBERTON, J.M., SLATE, J., BANCROFT, D.R. & BARRETT, J.A. 1995: Nonamplifying alleles at microsatellite loci: A caution for parentage and population studies. *Mol. Ecol.* **4**: 249–252.

RAHMAN, M.H. & RAJORA, O.P. 2001: Microsatellite DNA somaclonal variation in micropropagated trembling aspen (*Populus tremuloides*). *Plant Cell Rep.* **20**: 531–536.

ROGERS, C.A. & LEVETIN, E. 1998: Evidence of long-distance tranport of mountain cedar pollen into Tulsa, Oklahoma. *Intern. J. Biomet.* **42**: 65–72.

RUESINK, J.L., PARKER, I.M., GROOM, M.J. & KAREIVA, P.M. 1995: Reducing the risks of nonindigenous species introductions. *BioScience* **45**: 465–477.

SAVOLAINEN, O. 1991: Background pollination in seed orchards. *In*: Pollen Contamination in Seed Orchards (ed. Lindgren, D.), Proceedings of the Meeting of the Nord-

ic Group for Tree Breeding 1991, pp. 6–13, Swedish University of Agricultural Sciences, Department of Forest Genetics and Plant Physiology, Umeå,.

SAVOLAINEN, O. & KÄRKKAINEN, K. 1992: Effect of forest management on gene pools. *New Forests* **6**: 329–345.

SCHEMSKE, D.W., HUSBAND, B.C., RUCKELSHAUS, M. H., GOODWILLIE, C., PARKER, I.M. & BISHOP, J.G. 1994: Evaluating approaches to the conservation of rare and endangered plants. *Ecology* **75**: 584–606.

SCHNABEL, A. & HAMRICK, J.L. 1995: Understanding the population genetic structure of *Gleditsia triacanthos* L: The scale and pattern of pollen gene flow. *Evolution* **49**: 921–931.

SCHNABEL, A., NASON, J.D. & HAMRICK, J.L. 1998: Understanding the population genetic structure of *Gleditsia triacanthos* L.: seed dispersal and variation in female reproductive success. *Mol. Ecol.* **7**: 819–832.

SCHUSTER, W.S.F. & MITTON, J.B. 2000: Paternity and gene dispersal in limber pine (*Pinus flexilis* James). *Heredity* **84**: 348–361.

SILEN, R.R. 1962: Pollen dispersal considerations for Douglas-fir. *Journal of Forestry* **60**:790–795.

SLATKIN, M. 1981: Estimating levels of gene flow in natural populations. *Genetics* **75**: 735–756.

SLATKIN, M. 1993: Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47**: 264–279.

SLAVOV, G. T., DIFAZIO, S. P. & STRAUSS, S. H. 2002: Gene flow in forest trees: From empirical estimates to transgenic risk assessment. *In*: Ecological and Agronomic Consequences of Gene Flow from Transgenic Crops to Wild Relatives. Ohio State University, Columbus, pp. 106–119 http://www.biosci.ohio-state.edu/~lspencer/Proceedings .pdf.

SMITH, D. B. & ADAMS, W. T. 1983: Measuring pollen contamination in clonal seed orchards with the aid of genetic markers. *In*: Proceedings of the 20th Southern Forest Tree Improvement Conference, pp. 69–77, Athens.

SMOUSE, P. E., DYER, R. J., WESTFALL, R. D. & SORK, V. L. 2001: Two-generation analysis of pollen flow across a landscape. I. Male gamete heterogeneity among females. *Evolution* **55**: 260–271.

SNOW, A. A. & MORÁN-PALMA. 1997: Commercialization of transgenic plants: potential ecological risks. *BioScience* **47**: 86–97.

SORK, V. L., NASON, J., CAMPBELL, D. R. & FERNANDEZ, J. 1999: Landscape approaches to contemporary gene flow in plants. *Trends Ecol. Evol.* **14**: 219 –224.

STACY, E. A., HAMRICK, J. L., NASON, J. D., HUBBELL, S. P., FOSTER, R. B. & CONDIT, R. 1996: Pollen dispersal in low-density populations of three neotropical tree species. *Amer. Natur.* **148**: 275–298.

STEWART, S. C. 1994: Simultaneous estimation of pollen contamination and pollen fertilities of individual trees in conifer seed orchards using multilocus genetic data. *Theor. Appl. Genet.* **88**: 593–596.

STRAUSS, S. H., DIFAZIO, S. P. & MEILAN, R. 2001: Genetically modified poplars in context. *For. Chron.* **77**: 271 –279.

STREIFF, R., DUCOUSSO, A., LEXER, C., STEINKELLNER, H., GLOESSL, J. & KREMER, A. 1999: Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L. and *Q. petraea* (Matt.) Liebl. *Mol. Ecol.* **8**: 831–841.

TEMPLETON, A. R. 1986: Coadaptation and outbreeding depression. *In*: Conservation Biology: The Science of Scarcity and Diversity. Sinauer, Sunderland, MA. pp. 105–116.

TURNER, M. G., GARDNER, R. & O'NEILL. 2001: Landscape Ecology in Theory and Practice: Pattern and Process. Springer-Verlag, New York.

WADE, M.J. & GOODNIGHT, C.J. 1998: Perspecitve: The theories of Fisher and Wright in the context of metapopulations: When nature does many small experiments. *Evolution* **52**:1537–1553.

WANG, X., LINDGREN, D., SZMIDT, A. E. & YAZDANI, R. 1991: Pollen migration into a seed orchard of *Pinus sylvestris* L. and the methods of its estimation using allozyme markers. *Scand. J. For. Res.* **6**: 379 –385.

WHEELER, N. C., ADAMS, W. T. & HAMRICK, J. L. 1993: Pollen distribution in wind-pollinated seed orchards. *In*: Advances in Pollen Management (eds. Bramlett, D. L., Askew, G. R., Blush, T. D., Bridgwater, F. E. & Jett, J. B.), pp. 25–31, U.S. Department of Agriculture Forest Service, Washington, DC.

WHITE, G. M., BOSHIER, D. H. & POWELL, W. 2002: Increased pollen flow counteracts fragmentation in a tropical dry forest: An example from *Swietenia humilis* Zuccarini. *Proc. Natl. Acad. Sci. USA* **99**: 2038 –2042.

WHITLOCK, M. C. & MCCAULEY, D. E. 1999: Indirect measures of gene flow and migration: F_{ST} ne 1/(4Nm + 1). *Heredity* **82**: 117–125.

WILLIAMS, C. G. & SAVOLAINEN, O. 1996: Inbreeding depression in conifers: Implications for breeding strategy. *For. Sci.* **42**: 102–117.

WORTHINGTON WILMER, J., ALLEN, P. J., POMEROY, P. P., TWISS, S. D. & AMOS, W. 1999: Where have all the fathers gone ? An extensive microsatellite analysis of paternity in the grey seal (*Halichoreus grypus*). *Mol. Ecol.* **8**: 1417–1429.

WRIGHT, S. 1988: Surfaces of selective value revisited. *Amer. Natur.* **131**: 115–123.

XIE, C. Y. & KNOWLES, P. 1994: Mating system and effective pollen immigration in a Norway spruce (*Picea abies* (L.) Karst) plantation. *Silvae Genet.* **43**: 48 –52.

XIE, C. Y., YEH, F. C., DANCIK, B. P. & STROBECK, C. 1991: Joint estimation of immigration and mating system parameters in gymnosperms using the EM algorithm. *Theor. Appl. Genet.* **83**: 137–140.

YAZDANI, R., LINDGREN, D. & STEWART, S. 1989: Gene dispersion within a population of *Pinus sylvestris*. *Scand. J. For. Res.* **4**: 295–306.