Theoretical and practical considerations of gene flow

Juan José Robledo-Arnuncio, Santiago Gonzaléz-Martínez, Peter E. Smouse

▶ **To cite this version:**

Juan José Robledo-Arnuncio, Santiago Gonzaléz-Martínez, Peter E. Smouse. Theoretical and practical considerations of gene flow. Forest and genetically modified trees, International Union of Forest Research Organisations (IUFRO). AUT., 2010, Rome, Italy. hal-02755450

HAL Id: hal-02755450
https://hal.inrae.fr/hal-02755450

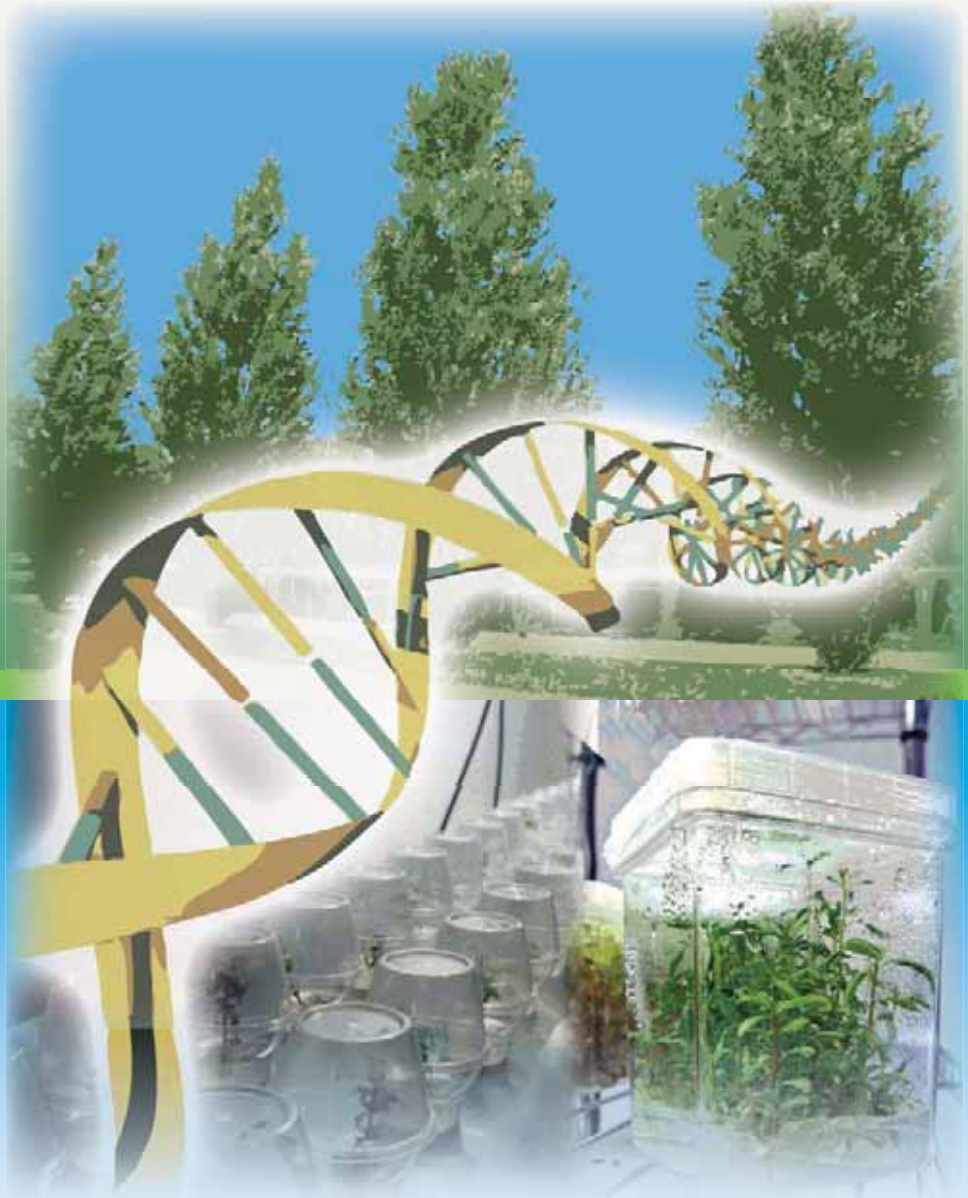
Submitted on 3 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution| 4.0 International License

FORESTS
and
GENETICALLY MODIFIED TREES



IUFRO

FAO
FIAT PANIS

FORESTS
and
GENETICALLY MODIFIED TREES

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS
Rome, 2010

The designations employed and the presentation of material in this information product do not imply the expression of any opinion whatsoever on the part of the Food and Agriculture Organization of the United Nations (FAO) concerning the legal or development status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. The mention of specific companies or products of manufacturers, whether or not these have been patented, does not imply that these have been endorsed or recommended by FAO in preference to others of a similar nature that are not mentioned.

The views expressed in this information product are those of the author(s) and do not necessarily reflect the views of FAO.

All rights reserved. FAO encourages the reproduction and dissemination of material in this information product. Non-commercial uses will be authorized free of charge, upon request. Reproduction for resale or other commercial purposes, including educational purposes, may incur fees. Applications for permission to reproduce or disseminate FAO copyright materials, and all queries concerning rights and licences, should be addressed by e-mail to copyright@fao.org or to the Chief, Publishing Policy and Support Branch, Office of Knowledge Exchange, Research and Extension, FAO, Viale delle Terme di Caracalla, 00153 Rome, Italy.

© FAO 2010

Contents

Part 2

ETHICAL AND SOCIO-ECONOMIC DIMENSIONS

7. Theoretical and practical considerations of gene flow

J.J. Robledo-Arnuncio, S.C. González-Martínez and P.E. Smouse

Gene flow, defined as the incorporation of genes from one gene pool into another, is at the core of the transgenic plant debate. In particular, a widespread societal perception of genetically modified plants is that of a hazardous material with high 'pollution' potential for the environment. The transfer of engineered genetic sequences (transgenes) from genetically modified plantations into natural populations of wild relatives via propagule dispersal is the natural vehicle for the feared 'pollution'. From a scientific risk assessment perspective, proper evaluation of the environmental implications of genetically modified plants involves both hazard and exposure assessments (Johnson *et al.*, 2006). Hazard assessment targets the identification and quantification of potential adverse effects of transgenic plants for the environment. Exposure assessment evaluates the probability of the environment being exposed to the hazards. Gauging the probability of transgenic incorporation into natural plant populations is the key step of exposure assessment.

It must be stressed that the detection of transgene flow into natural populations is not a demonstration of the risk of genetically modified organisms, which would require evidence of the transgenes being hazardous for the environment. This chapter deals solely with the role of gene flow in the genetically modified-plantation debate, without additional consideration of hazard assessment. The chapter is structured along four lines, describing the main contributions of gene flow researchers to exposure assessment of genetically modified trees:

- characterization of propagule dispersal patterns in non-genetically modified tree populations, which provides general insights into transgene flow potential and quantitative measurements for model parameterization;
- elaboration of theoretical models of gene flow from genetically modified tree plantations into natural populations, essential for predictive inference over large spatial and temporal scales;
- detection of transgene flow into natural populations, necessary for real-time monitoring, decision-making and management;
- formulation of transgene flow limitation practices.

There are several specific features of trees that are relevant in the genetically modified forest risk assessment context, which will be reiterated throughout this chapter. First, trees are long-lived perennials, a fact that has three important consequences:

- propagules will be dispersed from genetically modified plantations recurrently for many years before harvesting;

- it is very difficult to establish empirically the multiple-generation fate of these propagules in natural ecosystems;
- induced-sterility containment measures have increased chances of failing, due to temporal instability.

Second, trees disperse pollen and seed over broad spatial scales, increasing the probability of long-distance transgene movement and hampering its effective containment and accurate monitoring. Third, trees have typically very high fecundities, translating into large numbers of dispersed propagules, which are expected to increase the longest realized dispersal distance, particularly for fat-tailed dispersal distributions (Clark, Lewis and Horvath, 2001; Klein, Lavigne and Gouyon, 2006). Fourth, genotypes used for genetic modification are often taken from undomesticated tree stands and grown in similar locations, so cross-mating with natural populations of the same species (or close relatives) is likely to be common (González-Martínez, Robledo-Arnuncio and Smouse, 2005). Lastly, trees are the dominant life form of many terrestrial ecosystems, so introgression of transgenes into natural tree populations might have long-term and large-scale impacts on ecosystem function.

DISPERSAL PATTERNS IN NON-GENETICALLY MODIFIED TREE POPULATIONS

Given the absence of dispersal data for genetically modified trees and the legal and social restrictions on genetically modified-tree field trials, dispersal studies in non-genetically modified tree populations provide a necessary surrogate to investigate transgene flow potential. Assuming that no particular containment measures are taken and that genetic transformation for the target trait does not significantly alter the dispersal function, the available data on propagule dispersal patterns in natural tree populations, seed orchards and commercial plantations should reflect the potential scale of propagule flow from genetically modified plantations. Note that this section refers to the arrival of transgenes via pollen and seed dispersal into natural stands, and not to the long-term persistence of transgenes once they have arrived in the wild, which is discussed in the next section, on predictive models.

There are several statistical methods that have been developed for estimating gene movement within and among populations. Some of these methods provide historical estimates of gene flow, under various assumptions about evolutionary equilibrium, based on the spatial genetic structure of populations (Wright, 1931; Slatkin, 1985; Rousset, 1997; Beerli and Felsenstein, 1999) or individuals (Hardy and Vekemans, 1999; Rousset, 2000). Other methods yield contemporary gene flow estimates, inferred either from parentage analysis (Meagher, 1986; Devlin, Roeder and Ellstrand, 1988; Adams, Griffin and Moran, 1992; Smouse, Meagher and Kobak, 1999; Burczyk *et al.*, 2006) or from the spatial genetic structure of propagules (Smouse *et al.*, 2001; Austerlitz and Smouse, 2001; Robledo-Arnuncio, Austerlitz and Smouse, 2006). Several reviews on gene flow and transgenic trees have already extensively reported the main assumptions, statistical properties, pros and cons of each of these different estimation procedures (Ellstrand, 2003; Slavov,

DiFazio and Strauss, 2004; DiFazio *et al.*, 2004; Smouse, Robledo-Arnuncio and González-Martínez, 2007). The reader should refer to these previous works for detailed technical reference. Here, some results that are particularly relevant for genetically modified flow are summarized:

- Within-population mean dispersal distance estimates range from a few tens to several hundred metres (most frequently <1000 m in temperate forest trees), both for pollen (Dow and Ashley, 1998; Streiff *et al.*, 1999; Lian, Miwa and Hogetsu, 2001; Schuster and Mitton, 2000; Sork *et al.*, 2002; Robledo-Arnuncio and Gil, 2005; Goto *et al.*, 2006; Hardy *et al.*, 2006; Hardesty, Hubbell and Bermingham., 2006) and seeds (Clark *et al.*, 1999; Jones *et al.*, 2005; Goto *et al.*, 2006; Hardesty, Hubbell and Bermingham, 2006; González-Martínez *et al.*, 2006; Robledo-Arnuncio and García, 2007; Hardy *et al.*, 2006; Jordano *et al.*, 2007). Both insect- and wind-pollinated tree species show a similar range of mean dispersal distances in published studies, although there are large differences among species. It is noteworthy that estimates of the mean dispersal distance based on parentage analyses are likely to be downwardly biased, since the distribution of observed dispersal distances is usually truncated by the sampling plot boundaries, and propagules immigrating into the plot are usually discarded to compute this quantity.
- Yearly pollen immigration rates into forest fragments or stands are typically very high (>30%), and remain high (>5%) even with isolation distances of a few kilometres from the nearest conspecific stand (Kaufman, Smouse and Alvarez-Buylla, 1998; Adams and Burczyk, 2000; Schuster and Mitton 2000; Plomion *et al.*, 2001; Stoehr and Newton, 2002; Robledo-Arnuncio and Gil, 2005; Hanaoka *et al.*, 2007; O'Connell, Mosseler and Rajora, 2007).
- Seed immigration rates into sampling plots embedded within large forests (Jones *et al.*, 2005; González-Martínez *et al.*, 2006) and into isolated forest fragments (García, Jordano and Godoy, 2007) are both typically high (>10%). Secondary dispersal by fruit and seed predators, not always accounted for in seed migration estimates, is expected to increase the range of seed dispersal (Vander-Wall, 2001; Valbuena-Carabaña *et al.*, 2005).
- The estimated pattern of seed and pollen dispersal is very leptokurtic, i.e. there is a rapid decline in dispersal probability over short distances but non-negligible probability maintained beyond distances of several hundred metres (Clark *et al.*, 1999; Austerlitz *et al.*, 2004; Robledo-Arnuncio and Gil, 2005; Jones *et al.*, 2005; Goto *et al.*, 2006; Robledo-Arnuncio and García, 2007).
- Although empirical evidence for long-distance propagule dispersal in trees is abundant, its accurate probabilistic description remains a daunting challenge (Nathan, 2005). The usual procedure of parentage-based studies is to fit probability distributions to dispersal data collected on a small spatial scale and extrapolate the fit to the unobserved range of the distribution. Quantitative predictions established in this way should be considered with extreme caution, since functions with profoundly different tail-behaviour often fit observed data about equally well.

The general pattern is that while a substantial proportion of dispersal events occur over short distances, the potential for long-distance gene movement among tree populations or stands is quite high, though difficult to predict. The probability of seed or pollen from genetically modified tree plantations effectively reaching natural populations located even a few kilometres away should be considered non-negligible, *a priori*, especially when dispersal episodes accumulate over several years or decades. For instance, a low (say $p = 0.01$) yearly probability of transgene dispersal from a genetically modified plantation can translate into a substantial ($1 - 0.99^{20} = 0.18$) probability over a period of 20 years (Haygood, Ives and Andow, 2004; Smouse, Robledo-Arnuncio and González-Martínez, 2007). Similarly, a low probability of escape from a single genetically modified stand can translate into substantial risk of spread if there are multiple genetically modified plantations.

Observed dispersal patterns in natural populations provide a rough idea of the rate and spatial scale of transgene dispersal. Obtaining more precise estimates of transgene escape rate by direct extrapolation of these patterns, however, may not be adequate: most empirical studies report seed or pollen immigration rates into small study plots or small populations, surrounded by widespread conspecific forests, while source genetically modified tree stands (especially experimental plantations) may be small relative to wild recipient populations. This demographic scenario would result in transgene escape being less frequent than observed migration rates among natural stands, since increasing population size is expected to decrease immigration rates (Ellstrand and Ellam, 1993). But even if probably lower than reported immigration rates into small natural stands, potential rates of gene movement from small genetically modified tree stands into large wild populations may still be significant, as suggested by the observed low levels of gene flow from hybrid poplar plantations into wild populations of interfertile congeneric species (reviewed in Slavov, DiFazio and Strauss, 2004), and by the available estimates of transgene spread from genetically modified agricultural crops (Rieger *et al.*, 2002; Beckie *et al.*, 2003; Watrud *et al.*, 2004). Moreover, a very low rate of gene flow may be sufficient for eventual transgene fixation in the wild if it occurs recurrently or if it confers a selective advantage over conventional trees (Haygood, Ives and Andow, 2004; see next section).

Overall, considering an appropriately large temporal scale, the available evidence strongly suggests that the efficient dispersal systems of trees render the movement of transgenes from genetically modified plantations into conventional forests highly probable. But although it is reasonable to assume a very high likelihood of occurrence of a certain amount of transgene flow, predicting the rate at which it will happen, especially over very long distances, requires further empirical and theoretical analysis.

Predictive models

Thoroughly assessing the long-term exposure of natural forests to genetically modified trees through gene flow can hardly be accomplished without theoretical

modelling. There are numerous challenging aspects of the problem for which field trials, though highly desirable, are not really feasible. The most difficult and critical factor is that the relevant spatial and temporal scales are very large, with serious implications for many aspects of the assessment of exposure through gene flow. One should be ready to imagine a mosaic landscape of genetically modified tree plantations and natural stands, in more or less close proximity, spreading over thousands of hectares of land, eventually across different properties or even national territories. One would like to be able to predict the expected rate of transgene movement into a particular natural population and the probability of long-term persistence and eventual fixation of the transgene in this population.

Long distance dispersal models

A first consequence of the large spatial scale of the problem is the need to quantify the frequency and range of long-distance transgene dispersal, so that one can make predictions about the expected rate of transgene dispersal in particular spatial and demographic scenarios. Measuring rare long-distance dispersal events is very difficult in practice and, as mentioned above, extrapolating phenomenological functions beyond the experimental range of real data does not constitute a reliable approach to predicting long-distance dispersal. As pointed out earlier, phenomenological model predictions are quite sensitive to model selection, which in turn is highly dependent on sampling scale (Kuparinen *et al.*, 2007a). Moreover, the dispersal process is expected to be highly dependent on environmental variation, and thus extrapolating case-specific dispersal patterns to different environments may lead to misleading predictions (Kuparinen, 2006). Mechanistic dispersal models, by quantitatively describing the relationship between dispersal and the underlying physical factors causing particle movement (mainly propagule terminal velocity, release height, canopy structure and air flow statistics), may be more adequate to infer solutions outside the spatial and environmental domain for which observed data are collected, providing a wider range of predictive relevance. It must be noted, however, that mechanistic models are not so easily applicable to animal-dispersed species.

Mechanistic wind dispersal models are especially suitable to model long-distance propagule transport because they can emulate stochastic turbulent transport processes, such as updrafts above the forest canopy, considered a major determinant of long-distance seed and pollen transport (see Kuparinen, 2006 for a review of mechanistic wind dispersal models). For instance, in a study involving laboratory and field experiments with five tree species in a deciduous forest in North America, Nathan *et al.* (2002) fitted a Eulerian-Lagrangian model that was able to predict the proportion (1–5%) of seeds collected at different heights of a 45-m tower, a proportion considered as an upper bound on the probability of their long-distance transport. Given the typical high seed fecundity of wind-dispersed trees (roughly 10^3–10^5 per tree per year; Clark *et al.*, 1999), this would represent substantial numbers of potential long-distance dispersal events. Using similar coupled Eulerian-Lagrangian simulations, parameterized for *Pinus taeda*, Williams

et al. (2006) predict 0.007% to 0.1% of seedlings establishing beyond 1 km from 16–25-year-old plantations, or about 40–60 seedlings per year, assuming a 10-ha genetically modified stand, a conservative annual fecundity of 10^5 seeds/ha, and a 6% germination rate.

More recently, Kuparinen *et al.* (2007b) have developed a specific mechanistic approach to airborne dispersal of propagules in forested areas that explicitly addresses long-distance transport by modelling complex turbulent flows in upper parts of the atmospheric boundary layer. Consistent with previous studies, their simulations suggest that large amounts of light pollen, and small but significant proportions of heavier particles like seeds, may easily disperse over several kilometres. Lower propagule terminal velocities, higher release heights and changing wind conditions significantly increased the predicted rate and range of long-distance transport. They also point out, however, that further work is needed for better understanding of implementing release and deposition processes and within-canopy turbulences, which are critical for effective seed and pollen dispersal.

Population dynamics models

Once estimates of the frequency and spatial range of transgene escape are available, the next step is to investigate the long-term demographic dynamics of immigrant transgenes in natural populations, in competition with wild genotypes, and under a range of environmental conditions, including the presence or absence of the agent that the transgene may have been engineered to mitigate (Farnum, Lucier and Meilan, 2007). Only in this way will it be possible to predict the degree and duration of the exposure of natural forests to transgenes, which will range between fixation of the transgene in the recipient natural population or its quick elimination by natural selection. Given the long lifespan of trees, and taking into account that the relative fitness of transgenes may have multiple components expressed at different life stages, the necessity of theoretical models to examine multiple-generation transgene population dynamics becomes evident.

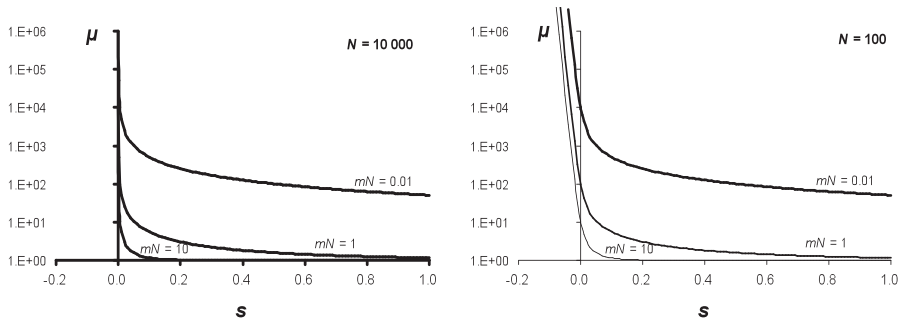
Simple demographically and spatially unstructured models can provide a first insight of transgene population dynamics. As an example, Williams and Davis (2005) use deterministic phenomenological simulations to investigate the fate of transgenes in a small escaped genetically modified-tree colony, with assumed initial transgene frequency of 50%, under different selective and demographic scenarios. Although their quantitative predictions are not easily interpretable, because of the absence of stochastic drift in the model and because of the artificial assumption that any immigrants arriving into the colony after its foundation had the same transgene frequency as colony residents, they illustrate the intuitive idea that both the relative abundance and relative fitness of the escaped genetically modified individuals are critical for transgene spread. Specifically, transgenic alleles in the escaped genetically modified-tree colony will tend to fixation if the transgene confers a net fitness advantage relative to wild-type alleles, but this process may be retarded if the genetically modified colony is embedded within a relatively large natural forest.

A more straightforward and formal description of the probability of transgene escape in a spatially and demographically unstructured model is provided by the analytical treatment of Haygood, Ives and Andow *(*2004). They define transgene escape into a wild population as the arrival of a transgene whose descendants will eventually take over the population, i.e. the descendant lineage of which will be destined for fixation, showing that the probability distribution of escape time (not time to fixation), defined in this way, is approximately geometric, with mean equal to the inverse of the probability of transgene escape.

Here, we derive the probability of transgene escape in a similar fashion to Haygood, Ives and Andow *(*2004), but considering a diploid transgenic locus and allowing for negative selective coefficients for the transgene, in order to illustrate the interplay between the probability of transgene escape, the transgene migration rate, the recipient population size and the adaptive value of the transgene. Let N be the number of mature individuals in the wild population, m the fraction of gametes in the wild population that flow from the genetically modified plantation per generation ($m < 0 < 1$), and s the selection coefficient for the transgene under wild conditions. From standard population genetics theory (e.g. eqs. 3.31 and 5.47 in Ewens, 2004), the probability that a newly arrived transgenic lineage is destined for fixation is approximately $\pi = (1 - e^{-s})/(1 - e^{-2Ns})$, assuming there is no dominance. The first generation after gene flow begins, we have Nm transgenes in the wild population, and the probability that at least one of them is destined for fixation, i.e. the probability of transgene escape, is given by $p = 1 - (1 - \pi)^{Nm}$. If the transgene does not escape in the first generation, we assume (following Haygood, Ives and Andow, 2004) that the situation is essentially the same in subsequent generations, until the transgene escapes or gene flow ends. That is, we assume that in these subsequent generations the amount of transgenes produced in the wild population and the number of individuals in transgenic lineages destined for extinction are small enough that p remains approximately the same until a transgene escape event occurs. That leads to a probability distribution of escape time, in generations, that is approximately geometric with mean $\mu = 1/p$ (Haygood, Ives and Andow, 2004).

Using this model, we examined (Figure 7-1) the estimated value of the mean escape time (μ), in generations, for different values of wild population size (N = 100 and 10 000), transgene selective value (s = -0.1 to 1.0) and number of transgene migrants per generation (Nm = 0.01, 1 and 10). A first interesting result is that the escape time becomes virtually independent of N as soon as the transgene has a relatively small selective advantage ($s > 0.01$ in our examples). If the transgene is neutral ($s = 0$), by contrast, the mean escape time is greatly reduced for small population sizes, assuming a fixed number of transgene migrants per generation (Nm). For instance, if $Nm = 10$, we have $\mu \approx 10$ for $N = 100$ and $\mu \approx 1000$ for $N = 10\,000$ (since, as expected under our model assumptions, the fixation probability of the transgene becomes approximately m for $s = 0$). Now, if the transgene is maladaptive in the wild ($s < 0$), escape time becomes enormous for large populations (the probability of transgene escape becomes negligible),

FIGURE 7-1
Expected mean transgene escape time in generations (*μ*) into a wild population of size *N* as a function of the selective coefficient of the transgene under wild conditions (*s*) and for different numbers of transgene immigrants per generation (*mN*)



irrespective of the number of migrants, while it can be relatively short if the wild population is small and the number of migrants relatively large (e.g. $\mu \approx = 100$ for $N = 100$, $s = -0.02$, and $Nm = 10$) (Figure 7-1). This is because stochastic drift reduces the efficiency of selection in small populations. Finally, for any given value of the selective coefficient (with $s \geq 0$), escape time increases as the number of transgene migrants decreases. Interestingly, however, escape time becomes fairly short ($\mu < 100$) as soon as the number of migrants per generation is not too small ($Nm \geq 1$) and the selection coefficient of the transgene is very slightly positive ($s > 0.001$). We believe that the arrival of at least one transgene migrant per generation ($Nm \geq 1$) can be considered a minimal working rate for genetically modified tree populations (given that this is a per generation rate and that trees may have a generation time of several decades), and thus that the probability of transgene escape will be generally very large for any transgene that is even slightly favoured by selection.

Spatial distribution, demographic structure and environmental variation may influence population dynamics in real systems, interacting with population genetic processes. Therefore, predictive models for the spatial and temporal dynamics of escaped transgenes need to be spatially, ecologically and demographically realistic. An early example of long-term spatial simulation modelling of transgene spread is provided by the STEVE model (DiFazio *et al.*, 2004), aimed at investigating potential invasiveness of transgenic poplars in the northwest United States of America. This stochastic model tracks transgenic and conventional genotypes in a virtual landscape that includes topographical and ecological information, with population dynamics being governed by modules simulating growth, reproduction, seed and pollen dispersal, and competition. The authors performed sensitivity analyses to study the consequences of different dispersal and selective conditions, several deployment and flowering control scenarios, and contrasting

selective values for the transgene. The main results highlighted by the authors (Slavov, DiFazio and Strauss, 2004) are:

- transgenic introgression into conventional stands was insensitive to the slope of local dispersal curves, but highly sensitive to changes in the proportion of long-distance dispersal;
- an imperfect, but tightly linked, sterility gene could dramatically slow the spread of a transgene that provided even a strong selective advantage;
- the spread of neutral transgenes could be greatly reduced by sterility levels whose effectiveness was of the order of 95%.

Perhaps the most elaborated and realistic spatial simulation model of transgene escape to date is that of Kuparinen and Schurr (2007). The model, which can be run in deterministic or stochastic form, includes: modules for seed dormancy; seedling establishment; tree growth; individual mortality; ovule, pollen and seed production; and pollen and seed dispersal. Many of the relevant demographic and reproductive processes are density-, genotype- and size-dependent. Seed and pollen dispersal are simulated using a mechanistic Lagrangian stochastic model especially configured to account for long-distance dispersal events. As an application, the authors examined the sensitivity of transgene escape to demographic differences between genetically modified and conventional trees, to the expression (dominant or recessive) of the transgene, and to the initial genotype of the genetically modified plants at the engineered loci (homozygous or heterozygous). After 100 years, a neutral transgene had diffused through short distance dispersal from the plantation into a contiguous conventional stand, with declining frequency with distance. Additionally, small numbers of transgenes had escaped the plantation via long-distance dispersal to distances beyond 1000 m. Decreased density-dependent mortality and increased growth, relative to conventional trees, were the two demographic factors of transgenes that resulted in a higher increase of escape rate into natural populations. The expression of transgenes only affected the probability of escape when they had demographic effects, with markedly reduced escape for recessive transgenes. Escape rate was also reduced for dominant transgenes if the initial genetically modified population consisted of heterozygous individuals.

Despite the utility of modelling, it must be kept in mind that theoretical models lacking realistic calibration will only provide qualitative insights on the sensitivity of transgene escape to particular factor effects. Quantitative predictions will require adequate parameterization, requiring experimental data, which should be pursued to the extent that model factors are legally amenable to empirical testing. That necessary caveat translates into a pair of serious challenges facing forest geneticists. One is the need to validate long-distance dispersal models empirically, including mechanistic models. The other, most critical, is to quantify the relative fitness of transgenes under different ecological conditions. As has been pointed out (Lee and Natesan, 2006), predictive models will not be really useful for transgene risk assessment if the uncertainty surrounding transgenic fitness impacts is not reduced.

REAL-TIME TRANSGENE FLOW ASSESSMENT

What do we need?

Another front where gene flow researchers can contribute to exposure assessment of genetically modified trees is the development of methods for real-time detection of transgenes. Although field release of genetically modified trees is still uncommon (Van Frankenhuyzen and Beardmore, 2004), there will soon be high demand for tools for field assessment for transgenic presence in natural forests. Many of the available methods for gene flow analysis are not adequate for this purpose. Genetic methods for assigning individuals to populations (Manel, Gaggiotti and Waples, 2005), for instance, require a thorough characterization of the recipient and the genetically modified donor populations, and will be of little help unless there is very strong divergence within the allele frequency spectra of the populations, since otherwise assignment error rates are likely to be larger than the presumably very low transgenic frequency to be estimated. Parentage assignment, in contrast, requires exhaustive genotyping of all potential parents within the study area, which becomes unfeasible over the spatial scales that are relevant for transgene flow detection, being moreover subject to a level of statistical uncertainty that may be unacceptable for decision-making. In fact, parental designation is not necessary for detecting transgenes, which only requires a categorical diagnostic criterion to conclude whether an individual is carrying the engineered sequence or not, for which several more powerful monitoring methods are available (Stewart, 2005).

Transgene monitoring methodologies

The most straightforward detection method is laboratory screening of the transgenic sequence directly. This will require tissue collection and DNA extraction from potentially escaped genetically modified individuals in conventional populations for the examination of a diagnostic DNA segment at the modified region. This can typically be achieved by means of PCR amplification, followed by automated sequencing or by single nucleotide polymorphism (SNP) analysis. European regulatory schemes are already demanding all sequence information of transgenes in applications for authorization for release of genetically modified organisms, including the location of primers used for detection (EFSA, 2006). Ideally, the proposed 'biobarcode' technology (Gressel and Ehrlich, 2002) would permit a standardized procedure for transgene detection. This technology would consist of the inclusion of a non-coding DNA segment in the engineered DNA sequence, flanked by universal PCR primers, which would contain a variable region encoding information on transgene identity and origin.

More elaborated screening procedures, using nanotechnologies, would allow faster and *in vivo* monitoring of transgenes in the field. These techniques, still not implemented for commercial transgene detection in plants, involve developing nucleic acids that are complementary to the target transgenic transcript and that carry a fluorescent label that can be seen by shining an ultraviolet light on the plant (see Stewart [2005] for a detailed description of different methods). There are, however, several barriers to the use of this kind of approach, that may

eventually prevent its implementation, such as safety concerns about fluorescence-based technologies, the additional investment for genetically modified tree re-engineering, and, most importantly, legal restrictions on and social rejection of further transgenic engineering (Stewart, 2005).

An alternative approach for transgene screening is testing for diagnostic phenotypic traits expressed by the transgene, such as herbicide and pest resistance, or some easily detectable protein. This procedure can allow an intensive, low-cost screening, prior to more direct assessment using DNA-PCR analysis. Watrud *et al. (*2004), for instance, used two cycles of herbicide spraying to detect the presence of herbicide-resistance transgenes in progenies collected from conventional populations of creeping bentgrass. Survivors of the second cycle were then tested for the presence of a transgene-encoded protein using commercial test strips. Finally, DNA from herbicide resistant and protein-positive plants was extracted and sequenced for final confirmation of transgene presence. Similar screening protocols might also prove useful for forest trees, as long as the engineered traits are expressed at an early life stage (Smouse, Robledo-Arnuncio and González-Martínez, 2007), which may not be the case for altered fibre quality or growth. Testing for herbicide or pest resistance by spraying progenies collected from seed trees could be feasible for detection of transgene flow via pollen, but similar tests on naturally regenerated seedlings in the wild might be ecologically unacceptable.

Challenges related to sampling

The challenge of categorically detecting the early stages of transgene spread in the wild can be intimidating. Assume that a transgene is present in the natural regeneration of a conventional forest at a frequency of $q = 10^{-3}$. Then, if we wanted to reduce the probability of not detecting the transgene below $\alpha = 0.01$, we would need to screen at least $n = 4600$ seedlings (ensuring that $\alpha = (1 - q)^n < 0.01$). If the introgression rate were as low as $q = 10^{-4}$, we would then need over 46 000 samples to ensure $\alpha = <0.01$. Given the additional advisability of sampling over large spatial and temporal scales, the problem becomes such that some have simply concluded the impossibility of proving that transgenes are absent from a given region (Ortiz-García *et al.*, 2005). Of course, if a transgene were ultimately to reach fixation, its frequency would eventually have to reach levels much easier to detect, but this may only happen after a minimal initial frequency (as low as $1/2N$, N being the recipient population size) and several generations of random drift or positive selection, which probably means several centuries for forest tree species. Nevertheless, early detection is critical if we are to intervene. That being the case, strongly replicated sampling over large spatial scales seems unavoidable, which, if legally enforced, might have implications for the economic payoff associated with genetically modified tree plantations.

The intricacy of accurate early detection of transgenes is illustrated by the intense and publicized scientific debate about the presence or absence of transgenic flow into maize landraces in Mexico, with more than ten studies conducted since

2001 and several replies and counter-replies disputing statistical and sampling issues (see Mercer and Wainwright, 2008 for review and discussion). In fact, it has been argued that too much emphasis is being placed on the rate of transgene flow, when the parameter of greater concern should be the relative fitness of the transgenes (Hails and Morley, 2005; Lee and Natesan, 2006; Chapman and Burke 2006). The reasons for this argument can be summarized as follows:

- it is reasonable to assume that occasional transgene flow into natural populations is unavoidable in practice, even if at very low rates;
- the magnitude of the transgene migration rate may be very difficult to estimate;
- the relative fitness of transgenes is the primary force governing their spread.

One agrees with this view, and stresses the need for a shift towards further empirical research on life-time fitness costs and benefits of transgenes under contrasting ecological conditions, a challenging task for long-living forest trees. It is also likely, however, that any scientific risk assessment protocol and, perhaps more importantly, any political or social debate on the risks of genetically modified trees, will hardly pass without convincing transgene flow estimates.

Transgene flow avoidance

The exposure of natural ecosystems to genetically modified trees could be essentially avoided if effective gene flow from transgenic plantations were interrupted. Since, as discussed above, spatial isolation does not provide an efficient barrier to transgene flow, alternative transgenic containment and mitigation strategies are being developed. Specifically, containment methods use different forms of genetic engineering to prevent transgenes from leaving genetically modified plants, either by inducing sterility or by removing the transgene from gametes before their release (excision techniques). Mitigation procedures intend to reduce the fitness of transgenes by tightly linking it to an engineered gene that is maladaptive in the wild, hence providing a useful complement to the expected leakages in containment strategies. Technical and practical details concerning the development, implementation and efficiency of different containment and mitigation strategies were extensively dealt with in an earlier chapter. Here, it is simply asserted that fully safe transgene containment methods are yet to be developed and thoroughly tested on a case-by-case basis. A recent study reports promising results along this line, with some excision techniques achieving 100% deletion of functional transgenes from pollen and/or seed, as tested on more than 25 000 progeny of tobacco plants for each transgenic event (Luo *et al.*, 2007). Further research is needed, however, to test the temporal and environmental stability of this technique for tree species and different transgenes. Due to the long life cycle of forest trees and the diverse ecological conditions they experience, the stability of any genetically engineered transgene containment strategy remains a matter of concern. It must be kept in mind that containment failure rates much lower than 10^{-3} may be necessary to reduce transgene escape probabilities to acceptable levels (Haygood, Ives and Andow, 2004).

REFERENCES

Adams, W.T., Griffin, A.R. & Moran, G.F. 1992. Using paternity analysis to measure effective pollen dispersal in plant populations. *American Naturalist,* 140: 762–780.

Adams, W.T. & Burczyk, J. 2000. Magnitude and implications of gene flow in conservation reserves. pp. 215–224, *in:* A. Young, D. Boshier & T. Boyle (editors). *Forest conservation genetics, principles and practice*. CSIRO, Collingwood, Australia.

Austerlitz, F. & Smouse, P. 2001. 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., Dick, C.W., Dutech, C., Klein, E.K., Oddou-Muratorio, S., Smouse, P.E. &, Sork, V.L. 2004. Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology*, 13(4): 937–954.

Beckie, H.J., Warwick, S.I., Nair, H. & Séguin-Swartz, G. 2003. Gene flow in commercial fields of herbicide-resistant canola (*Brassica napus*). *Ecological Applications,* 13(5): 1276–1294.

Beerli, P. & Felsenstein, J. 1999. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics,* 152: 763–773.

Burczyk, J., Adams, W.T., Birkes, D.S. & Chybicki, I.J. 2006. Using genetic markers to directly estimate gene flow and reproductive success parameters in plants based on naturally regenerated seedlings. *Genetics,* 173: 363–372.

Chapman, M.A. & Burke, J.M. 2006. Letting the gene out of the bottle: the population genetics of genetically modified crops. *New Phytologist*, 170: 429–443.

Clark, J.S., Silman, M., Kern, R., Macklin, E. & HilleRisLambers, J. 1999. Seed dispersal near and far: patterns across temperate and tropical forests. *Ecology,* 80: 1475–1494.

Clark, J.S., Lewis, M. & Horvath, L. 2001. Invasion by extremes: population spread with variation in dispersal and reproduction. *American Naturalist,* 157: 537–554.

Devlin, B., Roeder, K. & Ellstrand, N.C. 1988. Fractional paternity assignment: theoretical development and comparison to other methods. *Theoretical and Applied Genetics,* 76: 369–380.

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. pp. 405–422, *in:* C. Walter & M. Carson (editors). *Plantation forest biotechnology for the 21st century*. Research Signpost, Trivandrum, India.

Dow, B.D. & Ashley, M.V. 1998. Factors influencing male mating success in bur oak, *Quercus macrocarpa*. *New Forest,* 15: 161–180.

Ellstrand, N.C. 2003. Current knowledge of gene flow in plants: implications for transgene flow. *Philosophical Transactions of the Royal Society of London, Series B*, 358: 1163–1170.

Ellstrand, N.C. & Elam, D.R. 1993. Population genetic consequences of small population size: implications for plant conservation. *Annual Review of Ecology and Systematics,* 24: 217–242.

EFSA [European Food Safety Authority]. 2006. Guidance document of the Scientific Panel on Genetically Modified Organisms for the risk assessment of genetically modified plants and derived food and feed. *The EFSA Journal,* 99: 1–100.

Ewens, W.J. 2004. *Mathematical population genetics. I. Theoretical introduction.* 2nd edition. Springer, New York, USA.

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

García, C., Jordano, P. & Godoy, J.A. 2007. Contemporary pollen and seed dispersal in a *Prunus mahaleb* population: patterns in distance and direction. *Molecular Ecology,* 16: 1947–1955.

González-Martínez, S.C., Burczyk, J., Nathan, R., Nanos, N., Gil, L. & Alía, R. 2006. Effective gene dispersal and female reproductive success in Mediterranean maritime pine (*Pinus pinaster* Aiton). *Molecular Ecology,* 15: 4577–4588.

González-Martínez, S.C., Robledo-Arnuncio, J.J. & Smouse, P.E. 2005. The consequences and implications of introgression in the conservation of forest trees. pp. 17–23, *in:* M.C. de Vicente (editor). *Gene flow and germplasm management.* Topical Reviews in Agricultural Biodiversity. International Plant Genetic Resources Institute, Rome, Italy.

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.

Gressel, J. & Ehrlich, G. 2005. Universal inheritable barcodes for identifying organisms. *Trends in Plant Science,* 7: 542–544.

Hails, R.S. & Morley, K. 2005. Genes invading new populations: a risk assessment perspective. *Trends in Ecology & Evolution,* 20: 245–252.

Hanaoka, S., Yuzurihara, J., Asuka, Y., Tomaru, N., Tsumura, Y., Kabukari, Y. & Mukai, Y. 2007. Pollen-mediated gene flow in a small, fragmented natural population of *Fagus crenata*. *Canadian Journal of Botany,* 85: 404–413.

Hardesty, B.D., Hubbell, S.P. & Bermingham, E. 2006. Genetic evidence of frequent long-distance recruitment in a vertebrate-dispersed tree. *Ecology Letters,* 9: 516–525.

Hardy, O.J. & Vekemans, X. 1999. Isolation by distance in a continuous population: reconciliation between spatial autocorrelation and population genetics models. *Heredity,* 83: 145–154.

Hardy, O.J., Maggia, L., Bandou, E., Breyne, P., Caron, H., Chevallier, M.H., Doligez, A., Dutech, C., Kremer, A., Latouche-Hallé, C., Troispoux, V., Veron, V. & Degen, B. 2006. Fine-scale genetic structure and gene dispersal inferences in 10 Neotropical tree species. *Molecular Ecology,* 15: 559–571.

Haygood, R., Ives, A.R. & Andow, D.A. 2004. Population genetics of transgene containment. *Ecology Letters*, 7: 213–220.

Johnson, K.L., Raybould, A.F., Hudson, M.D. & Poppy, G.M. 2006. How does scientific risk assessment of GM crops fit within the wider risk analysis? *Trends in Plant Science,* 12: 1–5.

Jones, F.A., Chen, J., Weng, G.J. & Hubbell, S.P. 2005. A genetic evaluation of seed dispersal in the neotropical tree *Jacaranda copaia* (Bignoniaceae). *American Naturalist,* 166: 543–555.

Jordano, P., García, C., Godoy, J.A. & García-Castaño, J.L. 2007. Differential contribution of frugivores to complex seed dispersal patterns. *Proceedings of the National Academy of Sciences of the United States of America,* 104: 3278–3282.

Kaufman, S.R., Smouse, P.E. & Alvarez-Buylla, E.R. 1998. Pollen-mediated gene flow and differential male reproductive success in a tropical pioneer tree, *Cecropia obtusifolia* Bertol. (Moraceae): a paternity analysis. *Heredity,* 81: 164–173.

Klein, E.K., Lavigne, C. & Gouyon, P.H. 2006. Mixing of propagules from discrete sources at long distances: comparing a dispersal tail to an exponential. *BMC Ecology*. 6: Art. no. 3. 12 p.

Kuparinen, A. 2006. Mechanistic models for wind dispersal. *Trends in Plant Science,* 11: 296–301.

Kuparinen, A. & Schurr, F.M. 2007. A flexible modelling framework linking the spatio-temporal dynamics of plant genotypes and populations: application to gene flow from transgenic forests. *Ecological Modelling,* 202: 476–486.

Kuparinen, A., Markkanen, T., Riikonen, H. & Vesala, T. 2007a. Modeling air-mediated dispersal of spores, pollen and seeds in forested areas. *Ecological Modelling*, 208: 177–188.

Kuparinen, A., Snäll, T., Vänskä, S. & O'Hara, R.B. 2007b. The role of model selection in describing stochastic ecological processes. *Oikos,* 116: 966–974.

Lee, D. & Natesan, E. 2006. Evaluating genetic containment strategies for transgenic plants. *Trends in Biotechnology*, 24: 109–114.

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

Luo, K., Duan, H., Zhao, D., Zheng, X., Deng, W., Chen, Y., Stewart, C.N. Jr, McAvoy, R., Jiang, X., Wu, Y., He, A., Pei, Y. & Li, Y. 2007. 'GM-gene-deletor': fused loxP-FRT recognition sequences dramatically improve the efficiency of FLP or CRE recombinase on transgene excision from pollen and seed of tobacco plants. *Plant Biotechnology Journal,* 5: 263–274.

Manel, S., Gaggiotti, O.E. & Waples, R.S. 2005. Assignment methods: matching biological questions with appropriate techniques. *Trends in Ecology & Evolution,* 20: 136–142.

Meagher, T.R. 1986. Analysis of paternity within a natural population of *Chamaelirium luteum*. 1. Identification of most-likely male parents. *American Naturalist,* 128: 199–215.

Mercer, K.L. & Wainwright, J.D. 2008. Gene flow from transgenic maize to landraces in Mexico: an analysis. *Agriculture Ecosystems & Environment,* 123(1-3): 109–115. [See also *ibid.* 126(4): 293–293.]

Nathan, R. 2005. Long-distance dispersal research: building a network of yellow brick roads. *Diversity Distribution*, 11: 125–130.

Nathan, R., Katul, G., Horn, H.S., Thomas, S.M., Oren, R., Avissar, R., Pacala, S.W. & Levin, S.A. 2002. Mechanisms of long-distance dispersal of seeds by wind. *Nature,* 418: 409–413.

O'Connell, L.M., Mosseler, A. & Rajora, O.P. 2007. Extensive long-distance pollen dispersal in a fragmented landscape maintains genetic diversity in white spruce. *Journal of Heredity,* 98: 640–645.

Ortiz-García, S., Ezcurra, E., Schoel, B., Acevedo, F., Soberón, J. & Snow, A.A. 2005. Reply to Cleveland *et al.*'s critique of 'Absence of detectable transgenes in local landraces of maize in Oaxaca, Mexico (2003-2004)'. *Environmental Biosafety Research.* 4: 209–215.

Plomion, C., LeProvost, G., Pot, D., Vendramin, G., Gerber, S., Decroocq, S., Brach, J., Raffin, A. & Pastuszka, P. 2001. Pollen contamination in a maritime pine polycross seed orchard and certification of improved seeds using chloroplast microsatellites. *Canadian Journal of Forest Research,* 31: 1816–1825.

Rieger, M.A., Lamond, M., Preston, C., Powles, S.B. & Roush, R.T. 2002. Pollen-mediated movement of herbicide resistance between commercial canola fields. *Science,* 296: 2386–2388.

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

Robledo-Arnuncio, J.J. & 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, J.J., Austerlitz, F. & Smouse, P.E. 2006. A new indirect method of estimating the pollen dispersal curve, independently of effective density. *Genetics*, 173: 1–14.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, 145: 1219–1228.

Rousset, F. 2000. Genetic differentiation between individuals. *Journal of Evolutionary Biology,* 13: 58–62.

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

Slatkin, M. 1985. Gene flow in natural populations. *Annual Review of Ecology and Systematics*, 16: 393–430.

Slavov, G.T., DiFazio, S.P. & Strauss, S.H. 2004. Gene flow in forest trees: gene migration patterns and landscape modelling of transgene dispersal in hybrid poplar. pp. 89–106, *in:* H.C.M. den Nijs, D. Bartsch & J. Sweet (editors). *Introgression from genetically modified plants into wild relatives*. CABI, Wallingford, UK.

Smouse, P.E., Meagher, T.R. & Kobak, C.J. 1999. Parentage analysis in *Chamaelirium luteum* (L.) Gray (Liliaceae): why do some males have higher reproductive contributions? *Journal of Evolutionary Biology,* 12: 1069–1077.

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.

Smouse, P.E., Robledo-Arnuncio, J.J. & González-Martínez, S.C. 2007. Implications of natural propagule flow for containment of genetically modified forest trees. *Tree Genetics and Genomes*, 3: 141–152.

Sork, V.L., Davis, F.W., Smouse, P.E., Apsit, V.J., Dyer, R.J., Fernandez, J.F. & Kuhn, B. 2002. Pollen movement in declining populations of California valley oak, *Quercus lobata*: where have all the fathers gone? *Molecular Ecology,* 11: 1657–1668.

Stewart, C.N. 2005. Monitoring the presence and expression of transgenes in living plants. *Trends in Plant Science,* 10: 390–396.

Stoehr, M.U. & Newton, C.R. 2002. Evaluation of mating dynamics in a lodgepole pine seed orchard using chloroplast DNA markers. *Canadian Journal of Forest Research,* 32: 469–476.

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. *Molecular Ecology,* 8(5): 831–841.

Valbuena-Carabaña, M., González-Martínez, S.C., Sork, V., Collada, C., Soto, A., Goicoechea, P.G. & Gil, L. 2005. Gene flow and hybridization in a mixed oak forest (*Quercus pyrenaica* Willd. and *Q. petraea* (Matts.) Liebl.) in central Spain. *Heredity,* 95: 457–465.

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

Vander-Wall, S.B. 2001. The evolutionary ecology of nut dispersal. *Botanical Review,* 67: 74–117.

Watrud, L.S., Lee, E.H., Fairbrother, A., Burdick, C., Reichman, J.R., Bollman, M., Storm, M., King, G. & Van de Water, P.K. 2004. Evidence for landscape-level, pollen-mediated gene flow from genetically modified creeping bentgrass with *CP4* EPSPS as a marker. *Proceedings of the National Academy of Sciences of the United States of America,* 101: 14533–14538.

Williams, C.G. & Davis, B.H. 2005. Rate of transgene spread via long-distance seed dispersal in *Pinus taeda*. *Forest Ecology and Management,* 217: 95–102.

Williams, C.G., LaDeau, S.L., Oren, R. & Katul, G.G. 2006. Modeling seed dispersal distances: implications for transgenic *Pinus taeda*. *Ecological Applications*, 16(1): 117–124.

Wright, S. 1931. Evolution in Mendelian populations. *Genetics*, 6: 111–123.