ORIGINAL ARTICLE

A metapopulation model for the introgression from genetically modified plants into their wild relatives

Patrick G. Meirmans,[1,2] Jean Bousquet[3] and Nathalie Isabel[1,3]

1 Natural Resources Canada, Canadian Forest Service, Laurentian Forestry Centre, Québec, QC, Canada
2 Department of Ecology and Evolution, Université de Lausanne, UNIL, Lausanne, Switzerland
3 Canada Research Chair in Forest and Environmental Genomics, Pavillon Charles-Eugène-Marchand, Université Laval, Québec, QC, Canada

Abstract

Most models on introgression from genetically modified (GM) plants have focused on small spatial scales, modelling gene flow from a field containing GM plants into a single adjacent population of a wild relative. Here, we present a model to study the effect of introgression from multiple plantations into the whole metapopulation of the wild relative. The most important result of the model is that even very low levels of introgression and selection can lead to a high probability that the transgene goes to fixation in the metapopulation. Furthermore, the overall frequency of the transgene in the metapopulation, after a certain number of generations of introgression, depends on the population dynamics. If there is a high rate of migration or a high rate of population turnover, the overall transgene frequency is much higher than with lower rates. However, under an island model of population structure, this increased frequency has only a very small effect on the probability of fixation of the transgene. Considering these results, studies on the potential ecological risks of introgression from GM plants should look not only at the rate of introgression and selection acting on the transgene, but also at the metapopulation dynamics of the wild relative.

Introduction

The use of genetically modified (GM) plants in agriculture has triggered concern about adverse ecological and economic effects that may arise from gene flow from the transgenic crop into related species (Ellstrand 2003). One often-mentioned example is the creation of a herbicide-resistant weed through the introgression of a transgene conveying herbicide resistance from a crop species into a wild relative. A comprehensive assessment of the risks associated with the introduction of GM crops should combine empirical and modelling studies and consider all steps of the introgression process (Wilkinson et al. 2003a). The first step in this process consists of the formation of viable hybrids through pollination, seed production, and germination. The second step consists of the stabilization of the transgene in the recipient population by backcrossing of the F1 hybrids with the wild relative. The final step in the introgression process consists of the spread of the transgene over other populations of the wild

relative, influenced by its possible effects on fitness. The majority of both empirical and modelling studies have focused only on the first two steps and have concluded that for many crop species, transgene escape is inevitable (Jenczewski et al. 2003; Papa and Gepts 2003; Wilkinson et al. 2003b). Considering these results, the focus of risk assessment is now shifting to the more challenging task of predicting what will happen to the transgene once it has entered the population of the wild relative (Jenczewski et al. 2003; Haygood et al. 2004; Weis 2005).

One important factor determining the fate of an escaped transgene is the selection acting on it. Many transgenes that induce a commercially important trait to a crop species may actually also be selectively advantageous in the wild. Examples of such traits are, among others, herbicide resistance, resistance against herbivores, or cold hardiness. On the other hand, a transgene may also incur a fitness cost in the wild populations. Herbicide resistance is only advantageous if the wild populations are actually treated with the herbicide. If this is not the case, expressing the transgene

may actually present a waste of valuable resources. Many studies have therefore tested the fitness effects of the transgene when crossed into the wild relative (Burke and Rieseberg 2003; Snow et al. 2003).

To analyse the fate of the transgene within the wild relative, one must consider that those populations that are in contact with the transgenic crop are not isolated units. Instead, they are part of a dynamic metapopulation that consists of many separate populations that are connected by gene flow, and only a relatively small proportion of the populations may be in direct contact with the transgenic crop. In many aspects, the spread of a transgene therefore resembles the spread of a mutant allele over a metapopulation, which is a well-known topic in theoretical population genetics (Kimura 1962; Whitlock 2003). Indeed, several authors have used those results from theoretical population genetics to describe the spread of a transgene (Chapman and Burke 2006; Weis 2005).

The spread of a mutant allele is mostly seen in the light of the probability that the allele goes to fixation. In one of the most influential papers in the field, Kimura (1962) showed that, in an unstructured population, this probability is a function of the population size, the selection coefficient and the initial frequency of the allele. When the frequency of the allele is very low, the force of genetic drift is very strong and the probability of fixation is very small, even for alleles that are under strong positive selection. Surprisingly, when a simple island model of population structure (Wright 1930) is added, the probability of fixation does not change even though the population structure leads to an increase in the effective population size. This is because the increase in effective population size is exactly balanced by a decrease in the effectiveness of the selection process (Maruyama 1970; Cherry and Wakeley 2003). In more complex cases, population structure does affect the probability of fixation, usually by lowering it for beneficial alleles and increasing it for deleterious alleles (Whitlock 2003).

The use of standard population genetics models to describe the spread of a transgene is limited in that it ignores the influence of introgression itself on this process. The most important effect of introgression is that it may push the frequency of the transgene to such a level that selection gets stronger than drift, thus increasing the probability of fixation. For a comprehensive risk assessment, it is therefore important to know how the frequency of the transgene in the population of the wild relative is determined by continuous introgression during a given number of generations. Especially of interest are the fraction of natural populations that are in direct contact with the transgenic crop and the interplay among introgression, selection, population structure and population dynamics.

Here, we present a metapopulation model of transgene introgression based on standard population genetics models. We show that introgression for only a limited number of generations may lead to an appreciable overall frequency of the transgene in the metapopulation, with a high probability that the transgene goes to fixation in the whole metapopulation if it is under positive selection. We also show that there is an effect of the population structure and of the population dynamics on the overall frequency of the transgene; if there is a high migration rate between populations or a high rate of population turnover, the resulting transgene frequency becomes higher than when there is a low migration rate or low turnover. However, under the simple island model of population structure that we used, this increase in frequency has only a minor effect on the probability that the transgene goes to fixation.

The model

The model consists of a metapopulation of a wild species, where all populations are connected to each other by migration. Several of the populations are experiencing introgression from agricultural fields containing a GM crop of a related species. The transgene, which may be under selection in the wild populations, can therefore escape from the GM crop into the natural populations and may concurrently spread over the whole metapopulation (see Fig. 1). A program for Mac OS X to run the model, and the source code for this program, can be downloaded from: http://www.patrickmeirmans.com/software.

The metapopulation of the wild species contains a certain number of populations (U), each of size N. The species is annual, diploid, and hermaphrodite, has no seedbank, and mates at random. A fraction k of the populations experiences introgression from the GM plants (these populations will be called 'directly introgressed populations'); the other populations $(1 - k)$ do not experience direct introgression. Introgression from the GM crop into the directly introgressed populations takes place through pollination of a fraction g of the individuals of the directly introgressed populations by GM plants. Only a single gene is considered, with two alleles: the transgene and the wildtype. Initially, the natural populations are fixed for the wildtype allele, and the GM crop is fixed for the transgene. It is assumed that the GM crop is harvested every year and regrown the next growing season from an independent seed stock. Therefore, the frequency of the transgene in the crop is not influenced by introgression from the natural populations and the transgene remains fixed in the GM crop.

A standard island model of population structure (Wright 1930) is assumed where every population in the metapopulation is linked to all other populations by migration. This
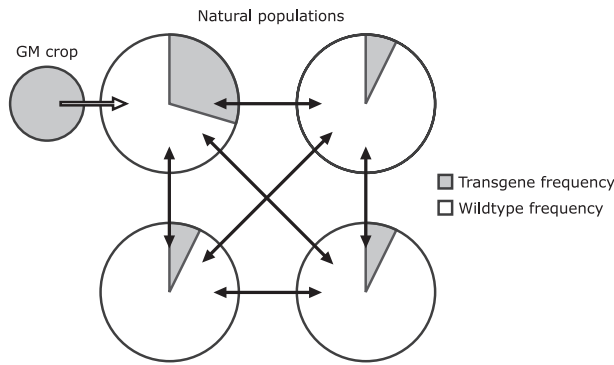
Natural populations

GM crop

☐ Transgene frequency
☐ Wildtype frequency

Figure 1 Schematic outline of the model. The four big circles represent the natural populations, which are connected to each other by migration (black arrows). One of the populations is in contact with a genetically modified (GM) crop field (smaller circle) and experiences unidirectional introgression (white arrowhead). The introgression leads to the presence of the transgene in the natural populations (grey shading), the frequency of which is highest in the directly introgressed population.

migration takes place through seeds: a fraction m of the seeds that are used to build a new generation comes from a 'seed pool' that is shared between all populations. Pollen flow between populations is not implicitly considered in the model, though the effects of this can be approximated by scaling up the migration rate.

Positive or negative selection acts on the transgene with strength s, with the transgene working additively. Selection takes place during or immediately after germination, when the population is replenished using the seeds produced in the previous generation. Every generation thus consists of the following steps: (i) germination of seeds, (ii) selection and establishment of individuals, (iii) production of seeds, and (iv) migration of seeds.

We made two versions of the model: stochastic and deterministic. The deterministic version makes the assumption that all subpopulations are of infinite size ($n = \infty$), while the stochastic one includes the effects of genetic drift due to finite population sizes.

Deterministic version

The absence of genetic drift in the deterministic model means that within each of the two groups of populations,

the directly introgressed populations and the remaining populations, the populations all have the same frequency of the transgene. Therefore, the metapopulation dynamics are only defined by k and m, and the actual number of populations U is not of interest. This is because the results for a set of 100 populations, of which 10 are receiving GM pollen, are actually the same as the results for a set of 10 populations of which one is receiving GM pollen.

The frequency of the transgene in the seeds that are produced by the directly introgressed populations ($P'_{sp(t+1)}$) is a function of both the current frequency of the transgene within these populations ($P_{p(t)}$) and the rate of introgression (g) from the GM field. For the other populations, the frequency of the transgene in the seeds ($P'_{sn(t+1)}$) simply equals the frequency of the transgene in the populations ($P_{n(t)}$). The frequency of the transgene in the shared seed pool ($P'_{s(t+1)}$) is then a weighted combination of the frequencies in the two types of populations:

$$
\begin{aligned}
P'_{sp(t+1)} &= P_{p(t)} + \tfrac{1}{2}g \cdot (1 - P_{p(t)}) \\
P'_{sn(t+1)} &= P_{n(t)} \\
P'_{s(t+1)} &= k \cdot P'_{sp(t+1)} + (1-k) \cdot P'_{sn(t+1)}
\end{aligned} \tag{1}
$$

At the beginning of the next generation, a fraction m of the seeds that are available for germination comes from the shared seed pool, and the remaining fraction $(1 - m)$ comes from the local population. Therefore, the frequency of the transgene in the seeds that are available for germination equals:

$$
\begin{aligned}
P'_{p(t+1)} &= (1-m) \cdot \left(P_{p(t)} + \tfrac{1}{2}g \cdot (1 - P_{p(t)})\right) + m \cdot P'_{s(t+1)} \\
P'_{n(t+1)} &= (1-m) \cdot P_{n(t)} + m \cdot P'_{s(t+1)}
\end{aligned} \tag{2}
$$

The frequency of the transgene in generation $t + 1$ is then determined by the selection acting on the transgene. Genic selection is applied using the standard method of equating the fitness of the homozygote for the wildtype allele (w_{22}) to 1, the fitness of the homozygote for the transgene (w_{11}) to $1 + s$ and the fitness of the heterozygote (w_{12}) to $1 + 0.5 \times s$ (Maynard Smith 1998). The frequency of the transgene after selection then equals:

$$
P_{p(t+1)} = \frac{w_{11} \cdot {P'_{p(t+1)}}^2 + w_{12} \cdot P'_{p(t+1)} \cdot (1 - P'_{p(t+1)})}{w_{11} \cdot {P'_{p(t+1)}}^2 + 2 \cdot w_{12} \cdot P'_{p(t+1)} \cdot (1 - P'_{p(t+1)}) + w_{22} \cdot (1 - P'_{p(t+1)})^2} \tag{3a}
$$

$$
P_{n(t+1)} = \frac{w_{11} \cdot {P'_{n(t+1)}}^2 + w_{12} \cdot P'_{n(t+1)} \cdot (1 - P'_{n(t+1)})}{w_{11} \cdot {P'_{n(t+1)}}^2 + 2 \cdot w_{12} \cdot P'_{n(t+1)} \cdot (1 - P'_{n(t+1)}) + w_{22} \cdot (1 - P'_{n(t+1)})^2} \tag{3b}
$$

Stochastic version

The stochastic version is mostly identical to the deterministic version, with the exception that the population sizes are finite, leading to genetic drift. Following the simulation model of Cherry and Wakeley (2003), the metapopulation is represented by an array of U integers, each bound between 0 and $2N$. These integers represent the number of copies of the transgene that are present in each population. Every generation, the new value of the integer is drawn from a binomial distribution, with an index parameter of $2N$. The probability parameter for the binomial distribution is formed by the expected frequency of the transgene in the population, which is a result of introgression, migration and selection. The expected frequency is calculated in a way similar to that of the deterministic model (Equations 1–3), with the exception that it has to be calculated for each population separately.

The frequency of the transgene in the shared seed cloud (cf. Equation 1) must be calculated by averaging over all populations:

$$P'_{s(t+1)} = \frac{\sum_{i=1}^{U} \left[P_{i(t)} + r_i \cdot \frac{1}{2} g \cdot \left(1 - P_{i(t)} \right) \right]}{U} \qquad (4)$$

Here, $P_{i(t)}$ is the frequency of the transgene in population i at generation t; $r_i = 1$ when population i is one of the directly introgressed populations, otherwise $r_i = 0$. The expected frequency of the transgene in the seeds that are available for germination for population i then equals (cf. Equation 2):

$$P'_{i(t+1)} = (1 - m) \cdot \left(P_{i(t)} + r_i \cdot \frac{1}{2} g \cdot \left(1 - P_{i(t)} \right) \right) + m \cdot P'_{s(t+1)} \qquad (5)$$

After selection, the expected frequency of the transgene in population i equals (cf. Equations 3a and 3b):

$$P_{i(t+1)} = \frac{w_{11} \cdot P'^2_{i(t+1)} + w_{12} \cdot P'_{i(t+1)} \cdot \left(1 - P'_{i(t+1)} \right)}{w_{11} \cdot P'^2_{i(t+1)} + 2 \cdot w_{12} \cdot P'_{i(t+1)} \cdot \left(1 - P'_{i(t+1)} \right) + w_{22} \cdot \left(1 - P'_{i(t+1)} \right)^2} \qquad (6)$$

Extinction and recolonization

For the above description, we used the term 'metapopulation' in a broad sense, meaning a group of separate but interacting populations. However, many uses of the term explicitly include extinction and recolonization. These processes can be included in the model by adding a probability d that a given population goes extinct in a generation. After extinction, the empty populations are immediately re-established using a migrant-pool model of recolonization (Slatkin 1977; Pannell 2003; Vuilleumier et al. 2008). To this end, the empty populations are filled from seeds drawn from the shared seed pool, so with an expected frequency of $(P'_{s(t+1)})$. In accordance with the main life-history of the model, the recolonization is immediately followed by a round of selection.

For the deterministic model, adding such population dynamics requires an extra assumption of a very large number of populations, together with the incorporation of the extinction (d) in Equation 2 that now becomes:

$$P'_{p(t+1)} = (1 - d) \cdot (1 - m) \cdot \left(P_{p(t)} + \tfrac{1}{2} g \cdot \left(1 - P_{p(t)} \right) \right) + (1 - d) \cdot m \cdot P'_{s(t+1)} + d \cdot P'_{s(t+1)}$$
$$P'_{n(t+1)} = (1 - d) \cdot (1 - m) \cdot P_{n(t)} + (1 - d) \cdot m \cdot P'_{s(t+1)} + d \cdot P'_{s(t+1)} \qquad (7)$$

For the stochastic model, the extinction and recolonization were added by drawing a random number from a uniform distribution and letting the population go extinct if this number is smaller than d.

Fixation probabilities after introgression

In the absence of introgression and extinction, our model reduces to the standard island model. Following the results of Maruyama (1970), this means that we can use Kimura's (1962) Equation 8 to calculate the probability of fixation if the introgression stops after a certain number of generations. This equation shows how the probability of fixation $u(P)$ is determined by the selection coefficient, the population size and the allele frequency:

$$u(P) = \frac{1 - e^{-4UNsP}}{1 - e^{-4UNs}} \qquad (8)$$

Here, P is the overall frequency of the transgene. The equation was slightly changed from Kimura's notation to match the notation from our model. More specifically, we use the product of the local population size (N) and the number of populations (U) as we are interested in the fixation of the allele in the whole metapopulation. We can only use this equation for the stochastic version of our model, as the deterministic version assumes infinite population sizes, which for the introgression rates considered here causes the fixation probability to always equal 0 for the used negative selection coefficients and 1 for the positive selection coefficients. We did not calculate the

probability of fixation for the case where there is population extinction as this case is much more complex. The main work on this (Whitlock 2003) only describes the probability of fixation of a new mutation, and does not give any approximations for the fixation probability of alleles with a higher frequency.

Tested parameter ranges

Both the deterministic and the stochastic versions of the model were used to test a wide range of parameter combinations. The rates of introgression that we used ranged from 0 to 0.5, which includes rates that are typical for interspecific gene flow and rates that are typical for intraspecific gene flow. We think that it is important to test such a wide range as the relationships between crops and their wild relatives range from intergeneric (e.g. wheat and jointed goatgrass) to intraspecific (e.g. sugar beet and sea beet) levels (Ellstrand 2003). The range of selection coefficients that we used was also chosen to be very wide from −0.1 to +0.1 as the effects of transgenes in natural populations are also expected to vary. Genes for herbicide tolerance or pest resistance may incur a very high positive selection (Snow et al. 2003), whereas genes that do not give a direct advantage in the wild relative may actually be mildly deleterious because of the wasted resources that are used for expressing those genes (Heil and Baldwin 2002). Population dynamics were simulated with extinction rates ranging from 0 to 0.4, following the model of Vuilleumier et al. (2008). Below, we point out some of the most important observations and illustrate them with some more detailed analyses.

Results

Both the deterministic and the stochastic versions of the model show that there is an effect of population structure on the overall frequency of the transgene in the metapopulation. When there is a high rate of migration between populations, the overall transgene frequency after 15 generations of introgression can be up to twice as high as when there is no migration between populations (Fig. 2). This effect is due to the fact that introgression is most effective when the frequency of the transgene in the directly introgressed populations is low. With an increasing transgene frequency, there is also an increasing chance that an individual pollinated by a GM plant already possesses one or two copies of the transgene. Migration increases the effectiveness of introgression by lowering the frequency of the transgene in the directly introgressed populations. Migration has this effect as the frequency of the transgene in the shared seed pool is almost always lower than the frequency in the directly introgressed pop-
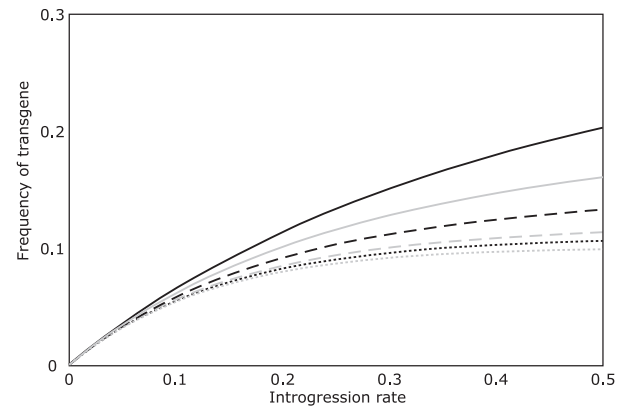

Figure 2 Effect of introgression and migration rates on the overall frequency of the transgene in the metapopulation after 15 generations. Deterministic version of the model, fraction of directly introgressed populations (k): 0.1, selection (s): 0.001, extinction (d): 0. The lines represent different migration rates (m), from top to bottom: 0.2, 0.1, 0.05, 0.02, 0.01, and 0.

ulations. The effect of population structure is already present at low introgression rates and for a small number of generations, though it is more pronounced when the rate of introgression is high.

The effect of population extinction and recolonization is comparable to that of migration (Fig. 3). With a high rate of population turnover, introgression is more effective, leading to higher frequencies of the transgene in the metapopulation. This result is not surprising as recolonization with a migrant-pool model (Slatkin 1977) can be seen as an extreme form of migration, where the migrants replace all the residents of the population. Like with
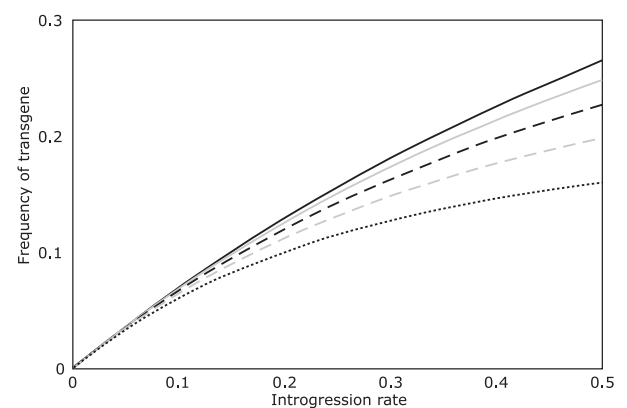

Figure 3 Effect of population extinction and recolonization on the overall frequency of the transgene in the metapopulation after 15 generations. Deterministic version of the model, fraction of directly introgressed populations (k): 0.1, selection (s): 0.001, migration (m): 0.1. The lines represent different extinction rates (d), from top to bottom: 0.4, 0.3, 0.2, 0.1, and 0.

migration, the effect is most pronounced for high levels of introgression, though it can also be observed for lower levels.

For the levels of selection that we analysed in our model, the effect of selection is small during the first generations of introgression when the transgene frequency is generally rather low (Fig. 4). This result is in line with expectations, as the effectiveness of selection is highest at intermediate allele frequencies (Fisher 1958). We only see a notable selection effect, when the selection coefficient is strongly positive or negative. A small interaction effect is found between selection and migration. For positive and slightly negative values of the selection coefficient, the above results hold where increasing the migration between population results in a higher overall frequency of the transgene. However, when the selection coefficient becomes strongly negative, high rates of migration lead to overall transgene frequencies that are actually slightly lower than those for lower rates of migration (Fig. 4). This interaction is due to the fact that higher migration rates lead to a higher overall allele frequency. As selection is rather ineffective at lower allele frequencies, an increased migration rate will also lead to an increased efficiency of selection.

The single difference between the stochastic and deterministic versions of the model is the population size, which is finite in the stochastic and infinite in the deterministic model. However, the differences in the results obtained between the two versions are slight. We ran the stochastic model with different population sizes and compared the results with that of the deterministic model with the same settings. The results show that there is very
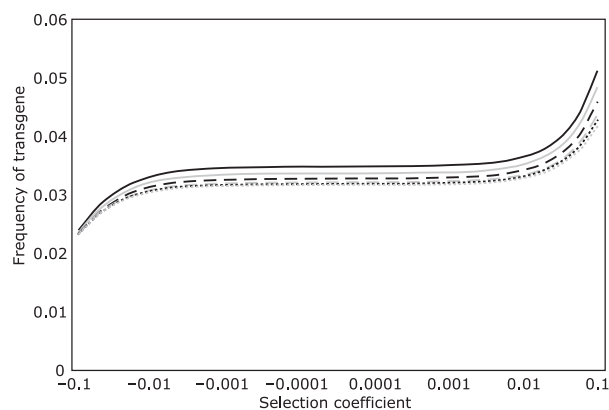

Figure 4 Effect of the selection coefficient and the migration rate on the overall frequency of the transgene in the metapopulation after 15 generations. Deterministic version of the model, fraction of directly introgressed populations (k): 0.1, rate of introgression (g): 0.05, extinction (d): 0. The lines represent different migration rates (m), from top to bottom: 0.2, 0.1, 0.05, 0.02, 0.01, and 0.

little effect of population size on the overall frequency of the transgene, except that for small populations, the overall frequency is somewhat lower than for large or infinite populations (see Supporting Information). This trend shows that in most cases, the deterministic model provides a good approximation.

Varying the fraction of populations that are in direct contact with the GM crop has a very straightforward effect on the overall transgene frequency: more directly introgressed populations simply lead to more introgression. More interesting are the consequences this has for the effectiveness of introgression as a result of migration and population turnover: the greater the fraction of directly introgressed populations, the smaller the impact of those processes on the overall frequency of the transgene (results not shown). If almost all populations experience introgression from the GM crop, the frequency of the transgene in the shared seed pool will almost equal the frequency in the directly introgressed populations, so migration and population turnover hardly lower the transgene frequency in these populations.

As all transgenic crops will eventually be replaced by new varieties, introgression of a certain transgene will only take place for a limited number of generations. Therefore, it is of interest to see what happens to the transgene after introgression stops, given the frequency it reached due to the introgression. We applied Kimura's (1962) equation to the results of our model after 15 generations of introgression, for various introgression rates and selection coefficients, but without population extinction. The obtained fixation probabilities show that even when there is a low rate of introgression and a moderately low positive selection coefficient, there can be a high probability that the transgene goes to fixation (Fig. 5). Apparently, a very low rate of introgression suffices to push the overall frequency of the transgene to such a level that the force of genetic drift becomes unimportant relative to the force of selection.

When fixation probabilities are compared between simulations with and without migration, we see that the increased transgene frequency that results from the migration has only a very small effect on the fixation probability. In fact, plotting the fixation probabilities for a migration rate of 0.2 produces a graph that is indistinguishable from the ones without migration (not shown). In a similar vein, graphs with the probability of fixation for negative selection coefficients were very uninformative as the probability was indistinguishable from zero (not shown), except for cases with a very small population size combined with a high level of introgression and weak selection (for more on this subject, see Kimura 1962; Whitlock 2003; Whitlock and Gomulkiewicz 2005). This result shows that after introgression has stopped,
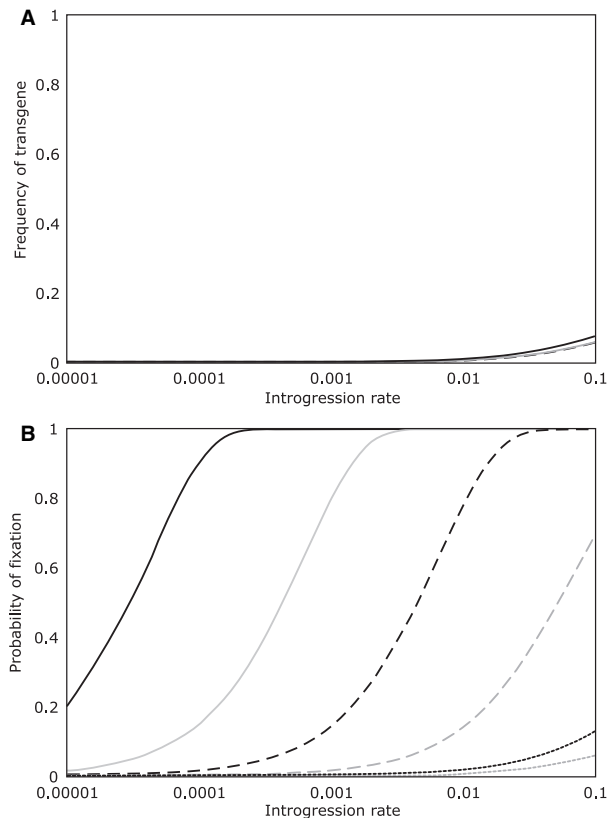
Figure 5 Fate of the transgene when introgression stops after 15 generations, such as for a genetically modified (GM) variety replacement in agriculture. Stochastic version of the model, which reduces to a pure island model after introgression stops and there is no extinction; 1000 runs, number of populations (U): 100, population size (N): 1000, fraction of directly introgressed populations (k): 0.1, migration rate (m): 0.05, extinction (d): 0. The lines represent different selection coefficients (s), from top to bottom: 0.1, 0.01, 0.001, 0.0001, 0.00001 and 0. (A) The overall transgene frequency after 15 generations, as a function of the introgression rate, for several values of the selection coefficient. Note that the curves overlap for all s < 0.1. (B) Probability that the transgene goes to fixation in the metapopulation, following Kimura (1962), using the frequencies in graph A.

population size plays a more important role than during introgression.

Discussion

Our results show that very small levels of introgression, combined with a very small selective advantage for the transgene and a moderately large population size, can already lead to a high probability that the transgene goes to fixation in the wild relative. A small number of generations of introgression is enough to push the transgene frequency to such a level that the force of selection becomes more important than the force of genetic drift. Figure 5

illustrates this point well. For instance, let us consider an introgression rate of 0.001 and a selection coefficient of 0.01. Both values are already so small that they become hard to measure experimentally with a reasonable degree of accuracy. Figure 5 indicates that if introgression lasts for no more than 15 generations at this rate, the probability that the transgene will go to fixation in the wild relative is close to 80%. Even though the selection coefficient used here is small, this does not necessarily mean that the likelihood of economic damage is small. For example, if only one in a hundred populations are treated with herbicides, a transgene for herbicide resistance approximately has a selective advantage of about 1%. However, it may in total result in considerable economical damage if the species is widespread. In most cases where the transgene is under negative selection in natural populations, the probability of fixation after introgression becomes very close to zero, even though the transgene may reach appreciable frequencies during introgression. The only exceptions to this are the cases where selection is very weak or population sizes are very small, indicating that negative selection on the transgene may only present a problem at the metapopulation level when the wild relative is a rare species. However, when a species is common but highly fragmented, negative selection can also affect small isolated populations.

Several modelling studies that have investigated introgression from GM plants into natural populations have focused on a single population of the wild relative (Haygood et al. 2004; Walklate et al. 2004; Damgaard and Kjellsson 2005; Kuparinen and Schurr 2007). These studies have mostly found that even with moderate levels of introgression, the transgene can reach appreciable frequencies of up to 5% in the wild population already during the introgression phase itself (Kuparinen and Schurr 2007). Like the present study, Haygood et al. (2004) looked at the fixation probability of a transgene under positive selection, though they did this in a single-population system. They found that the probability of fixation can be high even for low rates and short periods of introgression. Their results correspond to those of our study, though in our case the fixation probabilities are higher as the total effective size of a metapopulation, consisting of many demes that all have size N, can be several orders of magnitude larger than that of a single deme of size N. These single-population models have also shown that the realized introgression depends on the distance from individual plants to the crop field (Kuparinen and Schurr 2007; Watrud et al. 2004). Therefore, gradients in transgene frequency may develop within a natural population, which in itself may influence introgression. However, the main interest of this paper is the effect of metapopulation dynamics on the introgression process and the

within-population dynamics that have been analysed in these other studies have a negligible effect on the between-population dynamics. Therefore, we feel that it is justified to use a simplified introgression process at the local population level, even though we are aware that the within-population dynamics are in fact much more complex.

Our metapopulation model also shows that increased migration and population turnover increase the effectiveness of transgene introgression from a GM crop into natural populations of its wild relative. These effects are most pronounced when the rate of introgression is very high, but are also present for lower rates of introgression. The effect of population structure (migration) that we found only takes place during the actual introgression process. For the cases where there is no population extinction, the standard island model applies once the introgression stops and therefore, the results from Maruyama (1970) apply, which show that under the island model, the probability of fixation is independent of the amount of population structure. Calculating the fixation probability under different levels of migration shows that the increase in transgene frequency that results from migration does not translate in a drastic change in the probability of fixation.

We did not calculate the probability of fixation for the case where there is population extinction as this case is much more complex. The main work on this (Whitlock 2003) only describes the probability of fixation of a new mutation, and does not give any approximations for the fixation probability of alleles with a higher frequency. However, it is well known that extinction and recolonization can lead to drastic reductions in the effective population size (Slatkin 1977; Whitlock and Barton 1997; Whitlock 2003; Vuilleumier et al. 2008). Using a simulation approach, Vuilleumier et al. (2008) showed that this reduction in effective population size indeed leads to a reduction in the probability of fixation of a beneficial allele. So regarding the risk of transgene introgression, population extinction and recolonization work in both ways: on the one hand, the risk increases slightly because of the increased frequency of the transgene due to the more effective introgression. On the other hand, population turnover reduces the risk through a reduction of the probability of fixation. Of these two, the reduction of the probability of fixation through the smaller effective population size is probably most important.

We deliberately made our model of transgene introgression as simple as possible, so it would most closely resemble the classical island model, with obvious benefits. However, adopting this strategy means that a number of simplifying assumptions were made. Many important parameters that will influence the spread of a transgene over a wild population have been omitted, such as more realistic population dynamics, seed banks, long-distance pollen dispersal, and heterogeneity in the selection pressure. Of course, the validity of any model is determined by the assumptions it makes and the extent to which any simplifications of complex processes are defensible. Therefore, we will discuss the most important assumptions of the present model and how they affect our main conclusions.

Introgression

We simplified the process of introgression by using a fixed rate of pollination of individuals from the wild populations by individuals from the GM crop. Such a fixed rate is applicable, for instance, to cases where the crop and its wild relative are genetically closely related and crosses between them are equally successful as crosses within the two groups (Watrud et al. 2004). However, there are also cases where crops and their wild relatives are genetically differentiated, often belonging to different species or differing in ploidy level. Because of such differentiation, hybrid breakdown is often observed, with many possible outcomes such as the sterility of the F_1 generation, reduced growth and backcrossing with only a single parental species (Ellstrand 2003). Such hybrid breakdown can be quite severe, causing the actual introgression rate to be much lower than the hybridization rate (Ellstrand et al. 1999). Our model does not consider such complex effects of hybridization, but the fixed rate of introgression used can be seen as the 'realized' rate of introgression in which hybrid breakdown effects are already accounted for. A similar argument can be made for the transgene containment strategies that have often been suggested (see Brunner et al. 2007 for an overview). An example of such a containment strategy is the induction of male sterility in the transgenic crop, which directly prevents pollen flow from the crop to the wild population. However, as no such system will ever be completely failsafe, the risk of transgene escape will be determined by the amount of leakage of the system (Haygood et al. 2004). Like hybrid breakdown, such leakage can be incorporated in our model as a simple downscaling of the introgression rate. For example, Haygood et al. (2004) consider a hybridization rate of 0.1 and a leakage of 0.025 to be plausible values, together with a selection coefficient of 0.1. For our model, these values translate into an introgression rate of 0.0025, with a probability of fixation of 1 after 15 generations of introgression (Fig. 5). Therefore, adding a more complex system of hybridization and introgression will not qualitatively change the main outcomes of the present model, though they may have a large quantitative effect.

Another simplification that we made in the process of introgression is that we assumed only short-distance dispersal of the pollen from the crop-field into an adjacent natural population. This is realistic for some crops but certainly not for all: for example many wind-pollinated species such as trees or grasses can have pollen dispersal distances that equal or surpass that of their seeds (Watrud et al. 2004; Robledo-Arnuncio and Gil 2005; Fénart et al. 2007). Plantations of GM poplars have already been established in China, and their release in North America is expected in the near future (Strauss et al. 2001). The main reason for not including long-distance pollen dispersal in our model is that we were mostly interested in the effect of population structure on the introgression process, and we believe it is of relatively minor importance whether the connectivity between populations stems from seed dispersal or pollen dispersal. Furthermore, the impact of long distance pollen dispersal is expected to be small, as in the relatively rare event, that pollen from the GM crop reaches a distant population it may not necessarily be able to hybridize with the present wild relatives. In other words, the contribution of long-distance pollen dispersal is determined by the product of its rate and the introgression rate.

Many studies have analysed the possibility of hybridization between crop species and their wild relatives. As the observed rates of hybridization vary strongly between taxa, a 'typical' value, or even a 'typical range' for the rate of introgression cannot be given. One of the best studied systems is oilseed rape (*Brassica napus*), which can hybridize with both of its progenitor species, *Brassica rapa* and *Brassica oleracea*. In this system, the observed rates of spontaneous hybridization are quite low at around 1% (Scott and Wilkinson 1999). Furthermore, there is a difference in ploidy between the allotetraploid *B. napus* and the two diploid progenitor species, causing hybrid breakdown so that backcrossing the F1 to parental species may cause a reduction in fitness (Hauser et al. 1998). Effective rates of introgression from oilseed rape into natural populations of *B. rapa* and *B. oleracea* are therefore expected to be low. On the other hand, volunteers of oilseed rape itself are important weeds in many crop fields and many feral populations of oilseed rape have been established. The risks of transgene escape are much higher here, and indeed herbicide resistant feral populations, originating from transgenic oilseed rape, are frequently encountered and cause economic damage (Hall et al. 2000; Senior and Dale 2002; Warwick et al. 2003).

Selection

We modelled selection to be the same across all populations of the wild relative. Whether this assumption is

realistic depends to a large extent on the fitness effect that the transgene has on individuals and on the interaction between the transgene and the environment. This simplification may, for example, be realistic in cases where the transgene conveys resistance to herbivores, such as the often-used *Cry* genes from *Bacillus thuringiensis* (Snow et al. 2003). As the targeted herbivores are usually widespread, an escaped transgene is expected to provide a fitness advantage in nearly all populations of the wild relative. The simple selection model that we used is however less realistic for transgenes that provide herbicide resistance, another often used trait. Usually, only a small fraction of the natural populations of a species are treated with herbicides, namely those that exist in or close to agricultural fields to which the herbicide is applied. In that case, the transgene gives a large fitness advantage to introgressed individuals from such treated populations, but may be selectively neutral or provide a disadvantage to individuals from untreated populations (Meagher et al. 2003).

Studies on the actual fitness effect of transgenes in natural populations show that these are largely variable between species and transgenes (see Chapman and Burke 2006 for a review). However, assessing the effect experimentally is difficult as the actual fitness of individuals that carry the transgene may be strongly context dependent. For example, an insecticidal *Bt*-gene, when introgressed from GM crop sunflowers into wild sunflowers, resulted in an average fitness increase of 14% in Colorado and a 55% increase in Nebraska (Snow et al. 2003). Furthermore, as noted above, the selective pressure may not always be present; so it is also important to measure any possible negative fitness effects of the transgene in the absence of the selection pressure. An example of such a broad study is the case of a transgene conferring resistance to white mould in sunflower (Burke and Rieseberg 2003). In this study, the transgene was found to indeed provide resistance in crop-wild hybrids, with no negative fitness effect in the absence of infection. However, the effect of the disease varied between locations and as a result, the mould-resistance provided by the transgene did not translate into a higher reproductive fitness.

The study of the selection coefficient across different ecological contexts is especially important as such heterogeneity strongly impacts the spread of the transgene. Theoretical studies have shown that such spatially varying selection increases the probability of fixation of an allele under positive selection, compared to homogeneous selection with an equivalent average selection coefficient (Whitlock and Gomulkiewicz 2005; Vuilleumier et al. 2008). Populations where selection for the transgene is strong have a disproportional effect on the overall fixation probability as they act as buffer populations where

an allele is much less likely to become extinct due to drift.

Metapopulation dynamics

Perhaps the most stringent assumptions made by our model concern the metapopulation dynamics: the model is non-spatial, has fixed population sizes, and has a simple model of extinction and recolonization. In spite of these assumptions, the island model (Wright 1930), on which our model is based, is probably the most often used model in population genetics and is generally found to work surprisingly well (Felsenstein 2004). More realistic population models that do include spatial structure or different population sizes generally lead to a lower effective population size, and hence to a lower probability of fixation for beneficial alleles and a higher probability for deleterious alleles (Whitlock and Barton 1997; Whitlock 2003; Whitlock and Gomulkiewicz 2005). Below, we make some predictions about the effects of such complex metapopulation dynamics on the effectiveness of introgression, though there is a need to study these issues in more detail.

Adding an explicit spatial population structure will most likely not influence our conclusions very much. The effectiveness of introgression is influenced by the population structure as migration removes some transgenes from the directly introgressed populations and replaces them with wildtype alleles from the populations that are not in direct contact with the GM crop. This process will also take place when migration is spatially restricted, unless all directly introgressed populations are clustered together. A similar prediction can be made for adding pollen flow between populations. Our model assumes that all migration takes place through the movement of seeds, but of course there are many plant species in which long distance dispersal takes place through pollen. However, one of the main outcomes of our model is that migration enhances the effectiveness of introgression, and we believe that it is of minor importance whether the migration takes place though seeds or through pollen and the relative proportion of each. Thus, migration through pollen can simply be taken into account by scaling up, in our model, the migration rate parameter.

Relaxing the assumption that all populations have the same size will also have only a moderate impact on the effectiveness of the introgression as long as there is a fixed rate of introgression. In reality, however, the rate of introgression is expected to be linked to the ratio between the size of the crop field and the population size of the wild species directly affected by introgression: if the wild population is very small compared with the crop, the individuals are swamped by pollen and may be more

likely to hybridize with the crop. If the wild population is large, there may be enough conspecific pollen for an efficient dilution effect and hybridization may be less likely.

The lack of a seed bank in our model is not realistic for most wild relatives of crop species. As a majority of these species is annual, a seed bank is an important part of their life-history. The main effect of a seed bank is that it causes an increase in the species' effective population size. This increase is proportional to the average time to germination of seeds in the seed bank (Nunney 2002). Thus, the longer the seeds remain viable in the seed bank, the larger the effective population size. As only a part of this larger effective population is at one time directly exposed to introgression, this is equivalent to scaling the introgression rate down. On the other hand, a seed bank may cause a slight increase in the effectiveness of the introgression in a similar same way as migration does. When introgression starts, the seed bank is completely filled with wildtype alleles, but during introgression seeds containing the transgene are stored in the seed bank and the standing population is partly replenished with wildtype individuals from the seed bank. This keeps the transgene frequency in the standing population relatively low and therefore increases the effectiveness of the introgression. This latter effect of the seedbank will mostly be present when the duration of the introgression is relatively short compared to the average time to seed germination. Combined, the increase in the effective population size due to the seed bank is expected to have a much larger impact than the increase in the effectiveness of introgression.

Conclusions

Using a metapopulation model in which only a subset of populations is in direct contact with the GM crop, we found that population structure and population turnover impact the effectiveness of transgene introgression: more migration between populations or a higher extinction rate of populations leads to higher overall frequencies of the transgene. Furthermore, using analytical approximations under an island demographic model, we confirmed that earlier results showing that even very low levels of introgression and positive selection can lead to a high probability that the transgene goes to fixation in the metapopulation of the wild relative (Haygood et al. 2004). However, the increase in the transgene frequency that results from the population structure does not have a significant effect on the probability of fixation. Evidently, the island model of population structure that we used is an abstraction, and the results may be different for other, more realistic models of population structure. Nevertheless, we believe that relaxing the model's assumptions does not qualitatively change our main results about

the effect of metapopulation dynamics on the effectiveness of introgression. However, for quantitative predictions of the outcome of transgene introgression under specific scenarios, more precise metapopulation models of transgene introgression are needed. Furthermore, it is important to study in more detail the effect of population turnover and population structure on the probability that the transgene goes to fixation.

The results of our model have important implications for GM risk assessment. They show that a comprehensive risk assessment strategy should not only include analyses of the realized introgression rate, the survival of the hybrids, and the fitness effect of the transgene (Wilkinson et al. 2003a), but also an analysis of the effective size of the metapopulation of the wild relative, the rate of population turnover and the rate of migration between the populations.

Acknowledgements

Literature cited

Brunner, A. M., J. Li, S. P. DiFazio, O. Shevchenko, B. E. Montgomery, R. Mohamed, H. Wei et al. 2007. Genetic containment of forest plantations. Tree Genetics & Genomes **3**:75–100.

Burke, J. M., and L. H. Rieseberg. 2003. Fitness effects of transgenic disease resistance in sunflowers. Science **300**:1250.

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

Cherry, J. L., and J. Wakeley. 2003. A diffusion approximation for selection and drift in a subdivided population. Genetics **163**:421–428.

Damgaard, C., and G. Kjellsson. 2005. Gene flow of oilseed rape (Brassica napus) according to isolation distance and buffer zone. Agriculture Ecosystems & Environment **108**:291–301.

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-Biological Sciences **358**:1163–1170.

Ellstrand, N. C., H. C. Prentice, and J. F. Hancock. 1999. Gene flow and introgression from domesticated plants into their wild relatives. Annual Review of Ecology and Systematics **30**:539–563.

Felsenstein, J. 2004. Inferring Phylogenies. Sunderland, Sinauer.

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

Fisher, R. A. 1958. The Genetical Theory of Natural Selection, 2nd edn. Dover, New York.

Hall, L., K. Topinka, J. Huffman, L. Davis, and A. Good. 2000. Pollen flow between herbicide-resistant Brassica napus is the cause of multiple-resistant B. napus volunteers. Weed Science **48**:688–694.

Hauser, T. P., R. B. Jørgensen, and H. Østergård. 1998. Fitness of backcross and F2 hybrids between weedy Brassica rapa and oilseed rape (B. napus). Heredity **81**:436–443.

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

Heil, M., and I. Baldwin. 2002. Fitness costs of induced resistance: emerging experimental support for a slippery concept. Trends in Plant Science **7**:61–67.

Jenczewski, E., J. Ronfort, and A.-M. Chèvre. 2003. Crop-to-wild gene flow, introgression and possible fitness effects of transgenes. Environmental Biosafety Research **2**:9–24.

Kimura, M. 1962. On the probability of fixation of mutant genes in a population. Genetics **47**:713–719.

Kuparinen, A., and F. M. Schurr. 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.

Maruyama, T. 1970. On the fixation probability of mutant genes in a subdivided population. Genetical Research **15**:221–225.

Maynard Smith, J. 1998. Evolutionary Genetics. Oxford University Press, Oxford.

Meagher, T. R., F. C. Belanger, and P. R. Day. 2003. Using empirical data to model transgene dispersal. Philosophical Transactions of the Royal Society B: Biological Sciences **358**:1157–1162.

Nunney, L. 2002. The effective size of annual plant populations: The interaction of a seed bank with fluctuating population size in maintaining genetic variation. American Naturalist **160**:195–204.

Pannell, J. 2003. Coalescence in a metapopulation with recurrent local extinction and recolonization. Evolution **57**:949–961.

Papa, R., and P. Gepts. 2003. Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (Phaseolus vulgaris L.) from Mesoamerica. Theoretical and Applied Genetics **106**:239–250.

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

Scott, S. E., and M. J. Wilkinson. 1999. Low probability of chloroplast movement from oilseed rape (Brassica napus) into wild Brassica rapa. Nature Biotechnology **17**:390–392.

Senior, I. J., and P. J. Dale. 2002. Herbicide-tolerant crops in agriculture: oilseed rape as a case study. Plant Breeding **121**:97–107.

Slatkin, M. 1977. Gene flow and genetic drift in a species subject to frequent local extinctions. Theoretical Population Biology **12**:253–262.

Snow, A. A., D. Pilson, L. H. Rieseberg, M. J. Paulsen, N. Pleskac, M. R. Reagon, D. E. Wolf *et al.* 2003. A Bt transgene reduces herbivory and enhances fecundity in wild sunflowers. Ecological Applications **13**:279–286.

Strauss, S. H., P. Coventry, M. M. Campbell, S. N. Pryor, and J. Burley. 2001. Certification of genetically modified forest plantations. International Forestry Review **3**:85–101.

Vuilleumier, S., J. M. Yearsley, and N. Perrin. 2008. The fixation of locally beneficial alleles in a metapopulation. Genetics **178**:467–475.

Walklate, P. J., J. C. R. Hunt, H. L. Higson, and J. B. Sweet. 2004. A model of pollen-mediated gene flow for oilseed rape. Proceedings of the Royal Society of London Series B-Biological Sciences **271**:441–449.

Warwick, S. I., M.-J. Simard, A. Légère, H. J. Beckie, L. Braun, B. Zhu, P. Mason *et al.* 2003. Hybridization between transgenic *Brassica napus* L. and its wild relatives: *Brassica rapa* L., *Raphanus raphanistrum* L., *Sinapis arvensis* L., and *Erucastrum gallicum* (Willd.) OE Schulz. Theoretical and Applied Genetics **107**:528–539.

Watrud, L. S., E. H. Lee, A. Fairbrother, C. Burdick, J. R. Reichman, M. Bollman, M. Storm *et al.* 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.

Weis, A. E. 2005. Assessing the ecological fitness of recipients. In G. M. Poppy, and M. J. Wilkinson, eds. Gene Flow From GM Plants, pp. 143–168. Blackwell, Oxford.

Whitlock, M. C. 2003. Fixation probability and time in subdivided populations. Genetics **164**:767–779.

Whitlock, M. C., and N. H. Barton. 1997. The effective size of a subdivided population. Genetics **146**:427–441.

Whitlock, M. C., and R. Gomulkiewicz. 2005. Probability of fixation in a heterogeneous environment. Genetics **171**:1407–1417.

Wilkinson, M. J., J. Sweet, and G. M. Poppy. 2003a. Risk assessment of GM plants: avoiding gridlock? Trends in Plant Science **8**:208–212.

Wilkinson, M. J., L. J. Elliott, J. Allainguillaume, M. W. Shaw, C. Norris, R. Welters, M. Alexander *et al.* 2003b. Hybridization between *Brassica napus* and *B. rapa* on a national scale in the United Kingdom. Science **302**:457–459.

Wright, S. 1930. Evolution in mendelian populations. Genetics **16**:97–159.

Supporting Information

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

Figure S1 The frequency of the transgene (P_t) after 25 generations of introgression, for several different values of the selection coefficient (s), rate of introgression (g), migration rate (m), and population size (N).

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