The wave of advance of introduced genes in natural populations of plants.

Jarle Tufto,

Department of Mathematical Sciences - Lade, Norwegian University of Science and Technology, 7034 Trondheim, Norway. E-mail: jarlet@math.ntnu.no

Introduction

In plants, gene flow is limited to relatively short distances, implying that dynamic models which aim at predicting the fate of genes introduced by genetic engineering need to take into account the geographic location of the genes present in the population at any particular point in time. The frequency p of some allele is then a function not only of time t but also of the geographic location x. Several different formulations of such spatially explicit models are given in the literature. Here I give a brief overview of these different models and some of the predictions which may be relevant when discussing the long term spread of introduced genes. I also present a numerical solution for the resulting dynamics in one particular situation and discuss some implications of these results for the design of monitoring programs.

The model

We consider a population distributed continuously in one single dimension *x*. Local population density is constant and large, such that local genetic drift can be ignored, and generations are discrete. Let $p_t(x)$ denote the frequency of the transgene in generation *t* at geographic location *x*. We now assume that gene displacements, through seed and pollen dispersal, around every plant in the population follow some probability distribution *f(x)*. The frequency of the transgene also changes locally due to selection. With no dominance this gene frequency change is equal to $sp_t(x)[1- p_t(x)]$. With these assumption, the frequency of the transgene in generation $t+1$ at each geographic location *x* becomes

$$p_{t+1}(x) = \int f(y-x)\, p_t(y) dy + sp_t(x)[1- p_t(x)]\,. \qquad (1)$$

This equation defines the dynamics of a spatially explicit model with selection and dispersal acting every generation on local gene frequencies. The frequency of the transgene as a function of geographic location and time, is completely determined by (1) and the initial state of the

population, that is $p_t(x)$ at t=0. It should be noted that this model also can be applied to population distributed in two dimensions, as long as there is no dependencies on the second dimension y in the initial state of $p_t(x,y)$, and as long as the selection coefficient s is independent of y.

Analytic results

It is clear that if a selectively favoured gene is released at some geographic location, it will at first increase in frequency locally, and, at the same time, disperse into new parts of the population. After some initial generations, a steady state wave front moving will be formed. By doing certain approximations, Fisher and Kolmogorov [1,2] independently showed that this wave front will move with a constant speed of

$$v = \sqrt{2s}\,\sigma. \qquad (2)$$

Note that the dispersal probability distribution *f(x)* in the first term of the right hand side of (1) only appear through a single parameter single parameter σ^2 in (2), defined as the variance of the gene displacements, that is, $\sigma^2 = \int x^2 f(x)dx$. This parameter, the dispersal variance (or it's square root, the dispersal standard deviation) is the relevant quantity to estimate, and should not be confused with another frequently used measure of dispersal, the mean dispersal distance $\int |x| f(x)dx$, which in general have no simple relationship to σ^2. The dispersal variance can be estimated from different forms of data, either directly by using pollen traps [3] or genetic markers [4,5], or indirectly from the amount of genetic differentiation in subdivided populations [6]. In general, unless a random set of say, pollen grains can be followed, assumptions are made about the distributional form of *f(x)* are required. Some estimates of σ and some other measures of dispersal are given in Table 1.

If we use σ=40m, assume a selection coefficient of say, s=0.1, and also rely on Fisher's approximation, then this imply that the transgene will move with a speed of only v=18m/generation. For organisms with long generation times, such as many species of trees, the distance by which the wave front will move each year may become very small. Norway spruce, for example, have a generation time of about 20 years, which means that we should expect the wave front to move only about 1m/year. Although these calculations can only be taken as a rough approximation, they may still give us an idea about the relevant temporal and geographic scale of the problem.

Numerical analysis

The analytic results presented above refer to the asymtotic behaviour of the wave front, that is, the speed at which the wave front moves when stationarity has been attained many generations after the introduction. During the first few initial generations, however, the behaviour of the wave front may be quite different. The gene frequency dynamics will still be completely determined by (1) and the initial state of the system, but need to be analysed using numerical techniques. This can be done by slightly modifying the model, by assuming that the population is distributed in many small discrete subunits, instead of assuming that space is continuous. Using discrete coordinates, model (1) can be rewritten to

$$p_{t+1,i} = \sum_j m_{ij} p_{t,j} + s p_{t,i}[1 - p_{t,i}], \qquad (3)$$

where m_{ij}, determined by $f(x)$, represents the probability of dispersal to subunit i from some geographic location in subunit j. The dynamics of the model can then be simulated by setting all $p_{0,i}$ to some initial value, and then iterate (3) for the desired number of generations.

Two such simulations are shown in Figure 1, the first with selective advantage (s=+0.1), and the second with selective disadvantage (s=-0.1). In both cases, dispersal of genes follow a double exponential (Laplace) distribution with a standard deviation of 40m. The simulated populations consists of two regions; one transgenic 'cultivated' part at x<0 in which the frequency of the transgene is maintained at $p_{t,i}$=1, say by replantation of the crop every year, and a natural part at x>0 in which the trangene, in the initial generation, is assumed to be absent. The resulting gene frequency clines at every 5th generation are shown.

There are some interesting points to notice. Even when there is selection against the transgene it still become established several dispersal standard deviations into the natural part of the population. After about 30 generations, the effect of dispersal becomes balanced by the effect of selection, and the frequency of the transgene is then about 10% at a distance of 200m from the boundary. With selective advantage, after some initial generations, a steady state wave front is formed moving with a speed approximately equal to the prediction of Fisher's analysis.

Another result, and perhaps the most important one, is the small difference between the two scenarios in the predicted frequency of the transgene during the first few generations. After 5 generations the difference between the position of the gene frequency clines for selective advantage and disadvantage (represented by the two leftmost curves in Figure 1a and 1b) is very small and only about 10m. Only much later does the effect of selection become apparent. The main reason for this is that when $p_{t,i}$ is kept at 1 in the transgenic part of the population, dispersal will effectivly occur in only one direction across the boundary into the the

natural part of the population. This form of one-way migration will dominate the dynamics of the gene frequency cline during the first generations, regardless of whether the selection coefficient is positive or negative. Only serveral generation later, either an equilibrium is attained between dispersal and selection, or a steady state wave front moving 'with it's own help' is established.

Discussion

According to theory and existing estimates of some of the important parameters, the spread of transgenes into populations of wild relatives is likely to occur rather slowly. This has to be kept in mind when designing monitoring programmes and when evaluating the information that such monitoring programmes produce. Taken together with the fact that genetically engineering plants has been in use for at most two decades, that is, just a few generations, the absence of studies documenting the spread into populations of wild relatives and related environmental effects is perhaps not very surprising. It is well known, however, that at least in panmictic populations, even though the rate of increase in gene frequency may be slow, the probability of fixation of a gene rapidly tend to one for only slightly selectivly favoured genes [8], p. 425.

Even though the spread through large geographic regions may be slow, we should still be concerned with possible long-term effects, if the use of genetically modified organisms is to be sustainable, and the most critical parameter determining the long-term spread of a transgene is it's selection coefficient *s*. The result that the gene frequency cline during the first generations may depend only weakly on the coefficient of selection, suggests that a monitoring programme, sampling near a permanent transgenic plant field, will produce little information about the coefficent of selection. The models discussed in this paper are also fully deterministic, whereas a gene frequency cline, in a real population, will be influenced also by local genetic drift which is likely to mask any pattern that may be present in the data obtained.

Empirical evidence [9] and theoretical models for some modes of dispersal [3], suggest that the distributional form of seed and pollen displacements is quite leptokurtic, that is, a large part of seeds and pollen grains are deposited at either relatively short or long distances. Since dispersal of, at least, airborne pollen essentially is a stochastic non-deterministic process, this means that dispersal over any distance always is possible, and this has led some authors to the conclusion that gene flow by, for example, air borne pollen, is very effective also over long distances. Long distance dispersal will typically occur with a low probability, however, and because of competition with large amounts of local pollen, long distance dispersal from some

foreign population will in general not be as important as it may seem, because the contribution to the next generation made up by the foreign pollen in such cases will be very slight only. The actual shape of the dispersal distribution can be important, however, in estimation of the dispersal variance σ^2, because the distribution in most cases is observed over a limited study area only.

Acknowledgements

This work is in part based on financial support from the Norwegian Institute for Nature Research and the Directorate for Nature Management.

References

[1] Fisher, R. A. (1937) The wave of advance of advantegeous genes. *Ann Eugenics* 7: 355-369

[2] Kolmogorov, A., Petrovskii, I., Piskunov, N. (1937) In: F. Oliveira-Pinto, B. Connolly (eds): *Applicable Mathematics of Non-Physical Phenomena* John Wiley & Sons, New York.

[3] Tufto, J., S. Engen, K. Hindar (1997) Stochastic dispersal processes in plant populations. *Theor Pop Biol* 52: 16-26

[4] Kareiva, P., Morris, W., Jacobi, C. M. (1994) Studying and managing the risk of cross-fertilization between transgenic crops and wild relatives. *Mol Ecol* 3: 15-21.

[5] Nurminiemi, M., Tufto, J., Nilsson, N.-O., Rognli, O.-A. (1998) Spatial models of pollen dispersal in the forage grass meadow fescue. *Evol Ecol, In press.*

[6] Tufto, J., S. Engen, K. Hindar (1997) Inferring patterns of migration from gene frequencies under equilibrium conditions. *Genetics* 144: 1911-1921.

[7] Wright, J. W. (1953) Pollen dispersion studies: Some practical applications. *J Forestry* 51: 114-118

[8] Crow, J., Kimura, M. (1970) *An introduction to population genetics theory*. Harper & Row, New York.

[9] Levin, D. A., Kerster, H. W. (1974) Gene Flow in Seed Plants. *Evol Biol* 7: 139-220.

Table 1. Some estimates of σ and other measures of dispersal.

Species	Method	Estimate	Reference
Meadow fescue (*Festuca pratensis*)	Genetic markers	30m	[5]
Sea beet (*Beta vulgaris maritima*)	Indirect	74.6m	[6]
Norway spruce (Picea abias)		30.6m*	[7]
Pinus cembriodes		16.8m*	[7]

* Assuming an exponential dispersal distribution

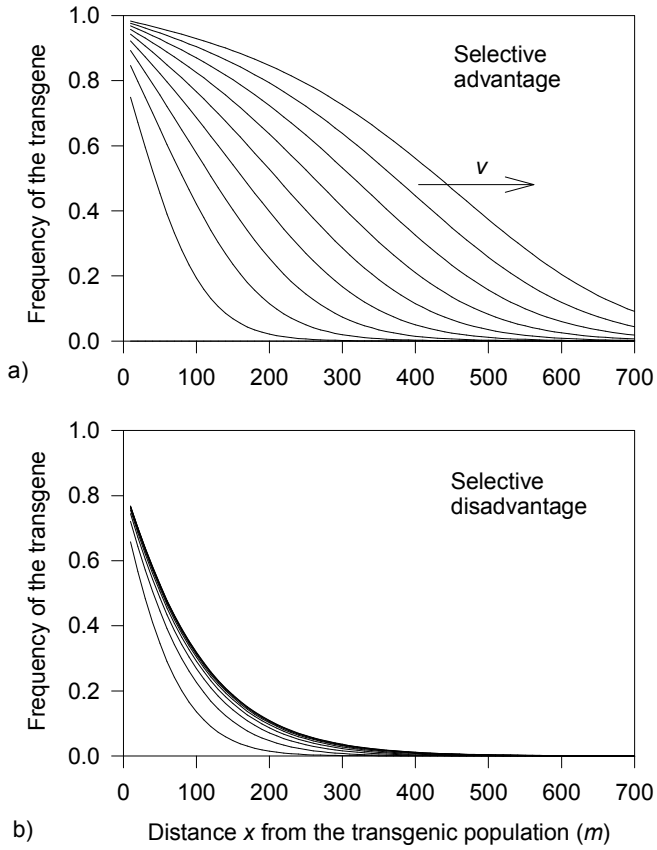


Figure 1. The frequency of the transgene as a function of distance from a permanent, transgenic population, with a) selective advantage, and b) selective disadvantage. The gene frequency cline at every 5th generation is shown. Further details are given in the main text.