Theor Appl Genet (1987) 73:724–736

A method for the estimation of gene flow parameters from a population structure caused by restricted gene flow and genetic drift *

H. Van Dijk

Department of Genetics, University of Groningen, Kerklaan 30, NL-9751 NN Haren, The Netherlands

Received September 5, 1986; Accepted November 22, 1986
Communicated by P. M. A. Tigerstedt

Summary. A method has been developed which enables the estimation of the plant gene flow parameters σ_p (pollen dispersal), σ_s (seed dispersal) and t (outcrossing rate) from a selection-free continuously structured population in equilibrium. The method uses Wright's F-coefficients and introduces a new F-function which describes the genetic similarity as a function of the spatial distance. The method has been elaborated for wind pollinated plant species but can be modified for insect pollination and for animal species. In practice allozymes will provide for the necessary neutral genetic variation. The more loci used and the more intermediate the gene frequencies, the more reliable the results. For the estimation of σ_p and t together (when the outcrossing rate is not known) at least two chromosomally unlinked loci are required. The method for estimating σ_s depends on whether the plant species is annual or perennial. The mechanism of selfing has been analysed by the explanation of the value of t by three components: population density (d), pollen flow (σ_p) and relative fertilization potential of own pollen (Z). The concepts of neighbourhood size and isolation by distance, developed by Wright, who used a single gene flow parameter σ, have been extended to the situation which is realistic for seed plants, using all three parameters σ_p, σ_s and t. When σ_p is large with respect to σ_s, σ_s largely determines the value of the neighbourhood size, whereas σ_p is the most dominating factor in isolation by distance. The use of "local effective population size" and "mean gene transport per generation" instead of "neighbourhood size" and "neighbourhood area", respectively, is proposed to avoid confusion. Computer simulations have been carried out to check the validity and the reliability of the method. Populations of 200 plants, using two or three loci with intermediate allele frequencies, gave good results in the calculation of σ_p with known value of t and of σ_s and N_e. With unknown t, especially with lower values of t, larger populations of at least 1,000 plants are necessary to obtain reasonably accurate results for σ_p and mean gene transport per generation M.

Key words: Gene flow – Genetic neighbourhood – Mathematical model – Computer simulation

* Grassland Species Research Group Publication No. 81

Introduction

A population is considered to be internally structured when the alleles of variable loci are not randomly distributed over the population area. Population structure may be brought about by local selectional differences, but also by restricted gene flow, allowing genetic drift to generate local differences in allele frequencies. In the latter case, the equilibrium state is only determined by the values of the gene flow parameters, which opens up the possibility of the estimation of these parameters from a selection-free population structure in equilibrium.

A lot of theoretical studies have been made on various forms of population structure (e.g. Wright 1951, 1965; Malécot 1969; Kimura and Weiss 1964). Analysis of populations which are subdivided into distinct subpopulations has been made possible by the use of F-coefficients (Wright 1951), especially the coefficient F_{ST}, which in its essence is connected with a discontinuous subdivision. A continuous situation with local differentiation, however, cannot be analysed according to a special 'recipe', in spite of theoretical achievements like 'neighbourhood size' (Wright 1943) and 'coefficient of kinship' (Malécot 1969).

The first important attempt of the estimation of neighbourhood size in a structured continuous population using allozyme variation was made by Schaal (1975). She divided the population into artificial squares which were thought to have the same area as the neighbourhood area. The latter was estimated by the use of directly determined dispersal parameters. The outcome of this experiment showed a much higher source of non-random mating within than between the squares, indicating a distinctly smaller neighbourhood size than assumed. Also, in these investigations population samples were used, whereas sampling just individuals would have been more appropriate.

In the present paper Wright's method of using F-coefficients is extended to a continuous situation in which it is possible to consider individuals themselves instead of samples. Since the need of such an analysis was generated during research on *Plantago* species, the method has been especially elaborated for wind-pollinating plant species, but can be modified for insect-pollinating plant species and for animal species.

The model requires an equilibrium situation without any effects of selection, which is assumed to be present in nature in populations growing under homogeneous environmental conditions, without any recent disturbances. The determination of the spatial distribution of allozyme genotypes will then enable the estimation of the gene flow parameters and the derived values for neighbourhood size and isolation by distance.

The problem

A path population of *Plantago major* was studied in detail with respect to the distribution of allozyme alleles over the population area. From several parts of the path, which is over 500 m long, samples were taken, and from one particular part (2 m × 2 m), all (adult) plants were scored for allozyme variation. The genotypes for the three variable enzyme loci and the position coordinates of those plants are represented in Fig. 1 (for methods: see Van Dijk et al., in prep.).

A high level of homozygosity can be noticed, far beyond the Hardy-Weinberg expectations, and even close to complete homozygosity, e.g. for *Est-4*. The average fixation index in this population appeared to be about 0.85.

In the second place the alleles are not at all equally distributed over the population area. The human eye can easily recognize certain patches, although this may be rather difficult to quantify without an objective method.

A third phenomenon is the linkage disequilibrium between the loci studied. These loci are chromosomally unlinked (Van Dijk 1981). Other combination of alleles are found at other sites of the path, indicating that there will be no distinct overall linkage disequilibrium in the population as a whole, only a local one.

Computer simulations were subsequently carried out, close to nature, using various gene flow levels, varying seed dispersal, pollen dispersal and degree of selfing. Some combinations of gene flow parameters led to homozygosities, degrees of patchiness and local linkage disequilibria which were comparable with the situation in Fig. 1. The question arising now is whether it is possible to estimate the gene flow parameters from the population structure found in nature.

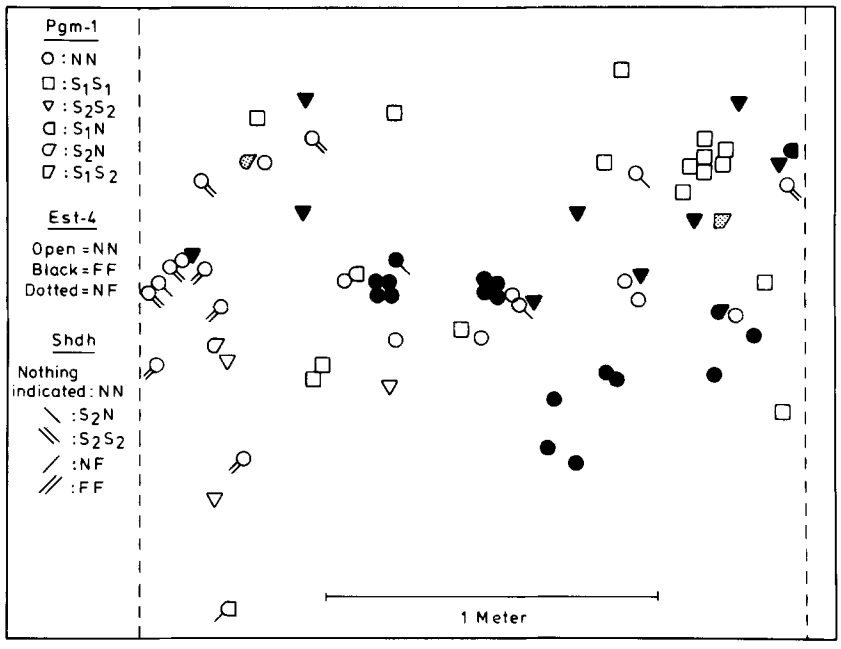

Fig. 1. The spatial distribution of allozyme genotypes in a part of a natural population of *Plantago major*. The (path-) population continues outside the broken lines, but is naturally limited in the vertical directions

Assuming no selection and an equilibrium situation, there is evidence for being able to estimate the three gene flow parameters because there are also three, partially independent, measurable phenomena (homozygosity, patchiness and local linkage disequilibrium). The first step in the analysis of population structure should therefore be the quantification of those three phenomena, and the next step the expression of the phenomena into the three gene flow parameters. The parameters will be chosen according to Wright (1943) and Crawford (1984): σ_p (the axial standard deviation of pollen dispersal), σ_s (the same for seed dispersal) and t (the fraction of cross-fertilizations).

Homozygosity is defined simply as the fraction of homozygotes (of whatever kind) in the population. In a Hardy-Weinberg population homozygosity is equal to $\sum p_i^2$. A measure of the degree of homozygosity which is independent of the allele frequencies is the fixation index F, which indicates how much a population is more homozygous than expected under Hardy-Weinberg conditions.

Patchiness in this context is a matter of how genetically equal neighbouring plants are in comparison to randomly chosen plants. Patchiness will be quantified in the next section. The local linkage disequilibrium, finally, can be expressed by the probability of the production of the same multi-locus-gamete by neighbouring plants above the expected level in Hardy-Weinberg conditions, and will be discussed later.

Spatial distance and genetic similarity

Population structure in a continuum caused by restricted gene flow and genetic drift is characterized by an increasing degree of genetic similarity between two individuals with decreasing spatial distance (local differentiation). Therefore, determination of the relationship between genetic similarity and spatial distance will describe this local differentiation in an essential way.

The genetic similarity I between two individuals on a particular locus is defined as the probability of producing the same gamete relative to that locus. In this definition I is independent of the ploidy level. For diploids I is 1, $\frac{1}{2}$, $\frac{1}{4}$ or 0 as illustrated below:

Genotype individual 1	Genotype individual 2	$I_{1,2}$
AA	AA	1
AA	AB	$\frac{1}{2}$
AA	BB	0
AB	AB	$\frac{1}{2}$
AB	AC	$\frac{1}{4}$
AB	CD	0

Notice that $I_{1,2}$ is also the expectation of the homozygosity in the progeny of a cross between individuals 1 and 2. Similar alleles as defined here are scored as being the same, but may differ in an undetectable way. This measure of genetic relatedness is different from the 'identity by descent' (Wright 1943; Malécot 1969), which is a measure based on the probability of two alleles being duplicates of one original gene.

In a population without any local differentiation, therefore with randomly distributed alleles, the expectation of I is $\sum p_i^2$ when the alleles have frequencies of $p_1, p_2, \ldots, p_i, \ldots$ etc. In a locally differentiated population I exceeds $\sum p_i^2$ when neighbouring plants are compared, to a maximum value of $I = 1$. The mean I for plants with distance r between them is now defined as I_r. This leads to the desired relationship: the function $I_r(r)$ describes the genetic similarity at any distance.

To make I_r values independent of allele frequencies, the scale ranging from $\sum p_i^2$ to 1 is converted into a scale from 0 to 1. The genetic similarity is now expressed by a new function $F_r(r)$, which is:

$$F_r(r) = \frac{I_r(r) - \sum p_i^2}{1 - \sum p_i^2}. \quad (1)$$

The change-over from I-scale to F-scale and the shape of the $F_r(r)$ function are illustrated by Fig. 2.

The symbol F was chosen in analogy to Wright's fixation index F, which is defined as the relative shortage of heterozygotes with regard to a population in Hardy-Weinberg equilibrium:

$$F = \frac{H_{exp} - H_{obs}}{H_{exp}} = \frac{Hom_{obs} - \sum p_i^2}{1 - \sum p_i^2} \quad (2)$$

with H_{exp}, H_{obs} and Hom_{obs} denoting respectively the heterozygosity expected, the heterozygosity observed and the homozygosity observed.

The difference between equations (1) and (2) is brought about by the terms $I_r(r)$ (the probability of finding the same allele in two individuals at distance r) and Hom_{obs} (the probability of finding the same allele on both chromosomes of a diploid individual).

In a practical situation the $F_r(r)$ function can easily be mapped when a computer is used in which the position coordinates and the genotype of each reproducing individual is put in. For each distance class, e.g. 0–5 cm, 5–10 cm, etc., the mean I_r is now calculated by comparing each individual to each later one. With the aid of the allele frequencies observed, the I_r values are transformed into F_r values. Also, a value N_r is noted for each distance class which expresses the mean number of individuals occurring in that distance class. The $N_r(r)$ function describes the population density by its height and the population dimensionality by its

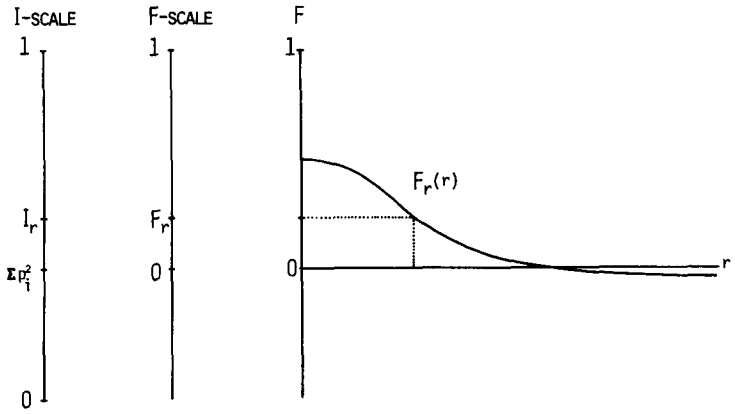

Fig. 2. The change-over from I-scale to F-scale; the shape of the $F_r(r)$ curve

shape. A 1-dimensional situation leads to a constant N_r independent of r, whereas in a 2-dimensional situation, N_r increases linearly with r. The most accurate determination of the $N_r(r)$ function is obtained when all individuals of the population are used, whether or not genetically examined.

The F_r values determined have equal expectations for different loci, which allows the use of the more accurate $\bar{F}_r$ values, obtained by averaging the F_r values over all loci. The reliability of the F_r values, however, is not independent of the allele frequencies. The larger the F-scale (the lower the $\sum p_i^2$ value), the more reliable are the F_r values. For this reason loci with one dominating allele and one or more rare alleles are not suitable for describing local differentiation. The calculation of any mean F value (F_{IT} or F_r) is always performed by weighing the respective F values by their denominators (H_{exp} values, see equations 1 and 2).

Homozygosity and the estimation of pollen flow

In a Hardy-Weinberg population homozygosity equals $\sum p_i^2$. Under less ideal conditions homozygosity may exceed $\sum p_i^2$ by different reasons of which selfing and local differentiation are the most important when selection is ignored.

Wright (1951) calculated the effect of selfing (the outcrossing rate being t) as $F_{IS} = (1-t)/(1+t)$, occurring in each subpopulation, and the effect of combining subpopulations having different allele frequencies as F_{ST}. The overall F of the total population, F_{IT}, is related to both other F's by

$$1 - F_{IT} = (1 - F_{IS})\,(1 - F_{ST})\,. \qquad (3)$$

F_{ST} is determined in practice by measuring the allele frequencies in the various subpopulations. At this point the use of the known F-coefficients fails to analyse a continuous population structure, because no subpopulations can be noticed. By the use of the $F_r(r)$ function, however, this problem will be solved.

Homozygosity reflects the genetic similarity of male and female parent. For a given female the expected homozygosity of her progeny depends on her genetic similarity to all possible male parents and their respective probabilities of becoming the male parent. In considering plants, the male and female plant will be the same individual with a probability $1 - t$. The progeny is then homozygous for a fraction $\frac{1}{2} + \frac{1}{2}\mathrm{Hom}_{obs}$. The mean result of outcrossing is given by the weighted average of I_r over all distances:

$$I_x = \int_0^\infty I_r\, P_r\, N_r\, dr$$

or, because I_r and N_r are determined for distance classes:

$$I_x = \sum_r I_r\, P_r\, N_r$$

with P_r = the probability of a plant at distance r to become the male parent $\left(\sum_r P_r\, N_r = 1\right)$. In an equilibrium situation the homozygosity observed is now given by

$$\mathrm{Hom}_{obs} = (1-t)\,(\tfrac{1}{2} + \tfrac{1}{2}\mathrm{Hom}_{obs}) + t \sum_r I_r\, P_r\, N_r\,. \qquad (4)$$

Transformed to the F-scale, which allows the use of the average F_r and F_{IT} values:

$$\bar{F}_{IT} = (1-t)\,(\tfrac{1}{2} + \tfrac{1}{2}\bar{F}_{IT}) + t \sum_r \bar{F}_r\, P_r\, N_r\,. \qquad (5)$$

Substitution of F_{IS} for $(1-t)/(1+t)$ and F_x for $\sum \bar{F}_r\, P_r\, N_r$ changes this into

$$1 - F_{IT} = (1 - F_{IS})\,(1 - F_x) \qquad (6)$$

which shows that F_x is the analogon of F_{ST} for a continuous situation. When a population is divided into distinct subpopulations, each of which are internally

differentiated, the subdivision of F_{IT} can be extended:

$$1 - F_{IT} = (1 - F_{IS})\,(1 - F_x)\,(1 - F_{ST})\,.$$

Besides t, the only unknown variable in equation (5) is P_r, expressing the probability of a plant at distance r of becoming the male parent. This probability is proportional to the pollen density of such a plant at a distance r. To find out this pollen density, the distribution of pollen over the 2-dimensional space has to be known. It is believed that for wind-pollinators pollen density follows a leptokurtic distribution function in all directions (see for a review on pollen and seed dispersal Levin and Kerster 1974). Insect-pollinated species have a more complicated pollen dispersal which depends on the insect species and also on population density.

In this paper wind-pollination will be further elaborated, using a normal distribution function characterized by σ_p. It is also possible, however, to use any other function, on condition that it has only one unknown parameter. For insect-pollinated plant species and animal species the same requirement has to be made.

For a wind-pollinated plant species, using a normal distribution function, pollen density at distance r is given by (C being a constant)

$$\psi_r(r) = C \cdot e^{-r^2/2\sigma_p^2}. \tag{7}$$

The total density of "alien" pollen at the female parent's place equals $\sum_r \psi_r N_r$. The probability P_r is now equal to

$$P_r = \frac{\psi_r}{\sum_r \psi_r N_r}. \tag{8}$$

For a series of σ_p values $\sum_r F_r P_r N_r$ can now be calculated. When t = 1, or in the general case when t is known, the particular σ_p value which corresponds to the expected F_x from equation (6) will be the best estimate of σ_p for this population.

When t is unknown a second relationship between t and σ_p is necessary to calculate both parameters. That relationship will be the subject of the next section.

Determination of the degree of selfing

To separate the effects on population homozygosity of both pollen flow factors – pollen dispersal and outcrossing rate – the occurrence of two or more chromosomally non-linked loci with intermediate allele frequencies is required. An apparent linkage will be observed, brought about by both local differentiation and selfing. Local differentiation acts independently on each locus; the apparent linkage by this origin is the result of the simultaneous local similarities on each separate locus. Selfing, on the contrary, is always acting on all loci together, and is by this reason an additional source of correlation between loci.

The apparent linkage is manifested by an increased number of individuals who are simultaneously homozygous for all loci considered. In a Hardy-Weinberg population their frequency would be $\prod_j \sum_i p_{ij}^2$, p_{ij} being the frequency of the i-th allele of the j-th locus.

To set up an equation analogous to (4) a few additional population parameters have to be measured. These are the fractions of individuals who are homozygous for all m loci, for m − 1 loci, etc. When selfed their progeny will be homozygous for a fraction 1, $\frac{1}{2}$, $\frac{1}{4}$, etc. Furthermore, the relation between the spatial distance of two plants and their probability of producing the same multi-locus gamete must be known, ${}^mI_r$ being the analogon of I_r. The ${}^mI$ of each pair of individuals is calculated as the product of the I's of the separate loci for that pair.

With ${}^j\mathrm{Hom}$ being the fraction of plants which are homozygous for j loci when m loci are used,

$${}^m\mathrm{Hom} = (1 - t)\left(\sum_{j=0}^{m} {}^j\mathrm{Hom} \cdot \tfrac{1}{2}^{m-j}\right) + t \sum_r {}^mI_r P_r N_r. \tag{9}$$

Using $M = {}^m\mathrm{Hom}$, $A = \sum_{j=0}^{m} {}^j\mathrm{Hom} \cdot \frac{1}{2}^{m-j}$ and ${}^mI_x = \sum_r {}^mI_r \cdot P_r N_r$, (9) will change into

$$M = (1 - t)\,A + t\,{}^mI_x \quad \text{or} \quad t = \frac{A - M}{A - {}^mI_x}. \tag{10}$$

For separate loci the analogon was

$$t = \frac{1 - F_{IT}}{1 + F_{IT} - 2F_x}. \tag{11}$$

To solve both t and σ_p from (10) and (11), σ_p is plotted against t according to both equations. The intersection point represents the values of both parameters for the population examined (see Fig. 3). Not considering any chance-fluctuations, both curves will be permanently declining, and the curve for separate loci will decline faster than the curve for combined loci. This allows only one point of intersection.

The reliability of the estimation of the two pollen flow parameters depends, except for the examined plant numbers and the various allele frequencies, also on the position of the point of intersection. With lower values of t, both curves of Fig. 3 intersect at a point where they are almost horizontal. Small deviations brought about by sampling variance will have large consequencies for the estimated σ_p value. The reason is that differences in σ_p hardly reflect any differences in homozygosity. The reliabilities of these and other

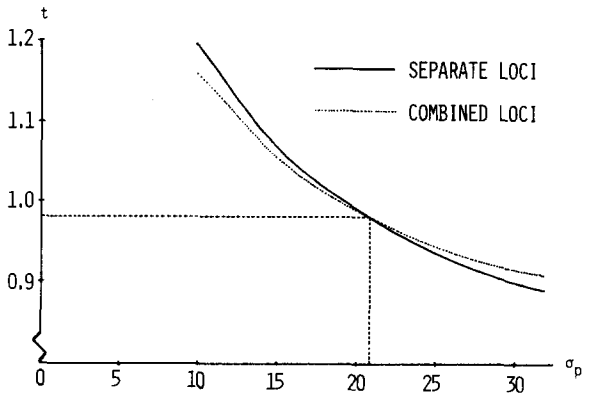

Fig. 3. The determination of σ_p and t simultaneously by using more than one locus

parameters estimated are discussed together with the computer simulation results.

When only one (sub)population is examined a transition to the F-scale is not necessary. With distinct subpopulations, or with a number of similar populations, combined results can be obtained by averaging ${}^{m}F_r$ and F_r values over (sub)populations and converting M and A into F_M and F_A using $\prod_j \sum_i p_{ij}^2$ instead of $\sum p_i^2$. One objection to the use of small (sub)populations is the possible occurrence of a net linkage disequilibrium. $\prod_j \sum_i p_{ij}^2$ is not expressing the real expectation of the multi-locus homozygote frequency in a random situation when the linkage disequilibria are substantial.

A closer examination of selfing

The selfing-outcrossing ratio is determined by the mean effective densities of own and alien pollen on the styles, taking into account any effects of incompatibility. When pollen dispersal follows a function like a normal distribution function, the density of own pollen at the plant's place is determined by the σ_p value using $r = 0$. This density will henceforth be called the 'macro-density' of own pollen. The density of alien pollen depends on both σ_p and the population plant density d. Although the local densities of own and alien pollen at the place of the styles will be proportional to the macro-densities, a lot of factors may modify it. The following are important:

– the extent of protogyny
– the extent in which flowers of the same plant flower simultaneously and, if so, their arrangement in space (e.g. spikes)
– the pollination mechanism (especially important in the case of insect pollination)
– the spread in time of plant flowering in the population.

Factors like these, together with incompatibility, determine the fertilization potential, Z, of own pollen relative to alien pollen. So defined, $Z = 1$ when self- and cross-fertilization occur in amounts proportional to the macro-densities of own and alien pollen. $Z = 0$ when plants are self-incompatible, and Z exceeds 1 when the local density of own pollen at the styles exceeds the macro-density.

When pollen density at distance r is given by ψ_r, the total density of alien pollen ψ_a at a plant's place equals $\sum_r \psi_r N_r$, whereas the macrodensity of own pollen ψ_o is $\psi_r(r = 0)$. Described by these parameters t will be

$$t = \frac{\psi_a}{Z\psi_o + \psi_a} \quad \text{and} \quad Z = \frac{\psi_a}{\psi_o} \cdot \frac{1 - t}{t}. \tag{12}$$

In a 2-dimensional situation

$$\psi_a = \int_0^\infty \psi_r N_r \, dr = 2\pi d C \sigma_p^2 \quad \text{and} \quad \psi_o = C$$

which gives

$$t = \frac{2\pi d\sigma_p^2}{Z + 2\pi d\sigma_p^2} \quad \text{and} \quad Z = 2\pi d\sigma_p^2 \cdot \frac{1 - t}{t}. \tag{13}$$

Equation (13) shows clearly the three constituent groups of factors by which t is determined: they work by way of σ_p, d or Z. With the aid of the estimated values of σ_p and t, and the observed density of reproducing plants d, Z can be calculated using (12) or (13).

The factors mentioned by which t is influenced are sometimes largely determined genetically and in other cases largely determined by environmental factors. Clearly t will be quite variable; its value will depend considerably on the situation in which it is measured. Because Z is density-independent and also free from environmental influences working on σ_p, this measure of selfing is a more constant one than t. In other words: Z is more than t a character of the species.

In the computer simulations mentioned later on, Z is one of the input parameters. It is not possible to use t as an input parameter because t is dependent on the particular density which is experienced by each individual plant.

Local differentiation and neighbourhood size

The point of intersection of the $F_r(r)$ curve with the F-axis, i.e. the point $F_r(0)$, (see Fig. 1), is a measure of the degree of local differentiation, ranging from 0 to 1.

In an equilibrium situation this local differentiation is maintained by the counterbalancing forces of gene flow and genetic drift. The role of gene flow is discussed in the next section. Genetic drift, bringing about the local decay of genetic variation which underlies local differentiation, is expressed by the probability that two individuals share one or more parents. In considering two plants at the same place, with a mean genetic similarity on the F-scale of $F_r(0)$, this probability of sharing parents depends on the "local effective population size", N_e, which is called the neighbourhood size. When two plants do not share any parents, their genetic similarity is on the average lower than $F_r(0)$, for their respective pairs of parents are growing at a larger than zero distance from each other. When they do share one or more parents their mean genetic similarity will exceed $F_r(0)$.

The probability of having, for instance, the same female parent, considering a 2-dimensional situation and a normal seed distribution, is given by

$$P_{ff} = \int_0^\infty P_r^2 N_r \, dr \tag{14}$$

with

$$P_r = \frac{C \cdot e^{-r^2/2\sigma_s^2}}{\int_0^\infty C \cdot e^{-r^2/2\sigma_s^2} N_r \, dr}.$$

Working with distance classes and an arbitrary dimensionality, the integrals have to be changed into summations. Elaboration of (14) gives

$$P_{ff} = \frac{\pi d C^2 \sigma_s^2}{(2\pi d C \sigma_s^2)^2} = \frac{1}{4\pi d \sigma_s^2}. \tag{15}$$

This outcome equals $1/N_{e,ff}$ = the probability of sharing the female parent in a population with $N_{e,ff}$ female individuals with equal probabilities of being the female parent.

Thus, $N_{e,ff} = 4\pi d \sigma_s^2$, which is identical to the formula $N_e = 4\pi d \sigma^2$ of Wright (1943). This equation does not hold for very low values of d or σ, because $1/N_e$ is a probability and is always between 0 and 1 so that N_e cannot be lower than 1. In considering (14), this discrepancy between N_e and lower values of $4\pi d \sigma^2$ is caused by the dependency of the total seed (or pollen) amount ($\int C e^{-r^2/2\sigma^2} N_r \, dr$) from whether or not a plant is standing nearby. The variance in the total seed amount also plays a part in this discrepancy. By computer simulation, the N_e for lower values of $4\pi d \sigma^2$ could be established as described in Table 1.

Analogous to equations (14) and (15) the probability P_{mm} that two outcrossing plants which are established at the same place have the same male parent is $1/4\pi d(\sigma_s^2 + \sigma_p^2)$, and $N_{e,mm}$ is, therefore, $4\pi d(\sigma_s^2 + \sigma_p^2)$. The probability P_{mf} that the female parent of the one plant is the male parent of the other is in the situation of complete outcrossing: $1/4\pi d(\sigma_s^2 + \frac{1}{2}\sigma_p^2)$ and $N_{e,mf}$ is $4\pi d(\sigma_s^2 + \frac{1}{2}\sigma_p^2)$.

Table 1. The correction of lower neighbourhood sizes (results of 20,000 simulations)

$4\pi d\sigma^2$	N_e estimated by computer simulation	$4\pi d\sigma^2$	N_e estimated by computer simulation
0	1	5	4.99
0.5	1.28	5.5	5.47
1	1.60	6	5.96
1.5	1.95	7	6.94
2	2.33	8	7.93
2.5	2.73	9	8.92
3	3.15	10	9.92
3.5	3.59	11	10.92
4	4.05	12	11.92
4.5	4.51		

The overall probability of sharing parents is expressed by

$$1/N_e = \tfrac{1}{4}(P_{ff} + 2P_{mf} + P_{mm}). \tag{16}$$

This is a general method for the calculation of all kinds of effective population sizes, for instance that of N_m male and N_f female individuals in an animal population, P_{mf} being zero for separate sexes.

When selfing does occur (outcrossing rate t) the terms of (16) are given by

$$P_{ff} = \frac{1}{N_{e,ff}} \tag{17}$$

$$P_{mf} = \frac{1-t}{N_{e,ff}} + \frac{t}{N_{e,mf}} \tag{18}$$

$$P_{mm} = \frac{(1-t)^2}{N_{e,ff}} + \frac{2t(1-t)}{N_{e,mf}} + \frac{t^2}{N_{e,mm}}. \tag{19}$$

The overall probability of sharing parents is

$$1/N_e = \frac{(1-\frac{1}{2}t)^2}{N_{e,ff}} + \frac{t(1-\frac{1}{2}t)}{N_{e,mf}} + \frac{\frac{1}{4}t^2}{N_{e,mm}}. \tag{20}$$

With both $t = 0$ or $\sigma_p = 0$, N_e will reduce to $N_{e,ff}$. When σ_p is very large with respect to σ_s, the overall N_e will become $N_{e,ff}/(1-\frac{1}{2}t)^2$. This is an interesting case, firstly because it is realistic and secondly because it shows that for $t = 1$, $N_e = 4N_{e,ff} = 16\pi d\sigma_s^2$, which illustrates that the smallest parameter dominates. This is contradictory to the N_e derived by Levin and Kerster (1974) in which the largest dispersal parameter is dominating. The overall σ^2 they use is not the variance relative to effective population size, but the one relative to isolation by distance, as will be discussed later.

In fact, domination of the smallest dispersal parameter appears to be logical: even when pollen may come from everywhere, two plants will often have a common female parent when seed dispersal is very low.

Seed dispersal

The counterbalancing forces, gene flow and genetic drift, which determine together the height of $F_r(0)$, both depend on the three gene flow parameters σ_p, σ_s and t. Knowing σ_p and t, σ_s can be calculated from the height of $F_r(0)$.

Actually, two models will be considered, depending on the turnover value of the population. In the first model with distinct generations (annual species) two plants at the same place will have one or more parents in common with a probability of $1/N_e$. The other model describes the situation of a perennial species with a very low turnover value. A plant germinating at a particular place will sometimes find an older individual (which has already reproduced) nearby. The question is no longer whether they have one or more parents in common but whether the new plant is a descendant of the older one.

Model I

Two new plants, which are established at the same place, have parents which are either the same or different. Different parents have I values according to their spatial distance. The probability distribution of their spatial distance depends on the dispersal mechanism. The function $P_r(r)$ as defined in (8) for pollen dispersal must be modified for seed dispersal (female parents), seed + pollen dispersal (male parents) and for parents of the different sex. The P_r's will be denoted as $P_{r,ff}$, $P_{r,mm}$ and $P_{r,mf}$. The appropriate variances are $2\sigma_s^2$, $2(\sigma_s^2+\sigma_p^2)$ and $2\sigma_s^2+\sigma_p^2$, respectively. The following equation will now describe the forces which determine $F_r(0)$:

$$F_r(0)=\frac{1}{N_e}\left(\tfrac{1}{2}+\tfrac{1}{2}F_{IT}\right)+\tfrac{1}{4}(1-P_{ff})\sum_r F_r P_{r,ff} N_r \qquad (21)$$
$$+\tfrac{1}{2}(1-P_{mf})\sum_r F_r P_{r,mf} N_r+\tfrac{1}{4}(1-P_{mm})\sum_r F_r P_{r,mm} N_r.$$

By substituting a range of σ_s values, the value which gives an $F_r(0)$ value corresponding with the $F_r(0)$ value calculated from the data is the best estimate of σ_s. $N_{e,ff}$, $N_{e,mf}$, $N_{e,mm}$ (together determining N_e), P_{ff}, P_{mf} and P_{mm} have to be corrected according to Table 1 if necessary.

Model II

The probability that the new individual is a descendant of the older one depends largely on seed dispersal. Neglecting the probability that the older one is the male parent of the new one but not the female parent, and defining the fractions of own and alien seed with respect to the older plant as φ_o and φ_a, then $\varphi_o/(\varphi_o+\varphi_a)$ is the probability that the older one is the female parent. The genetic similarity then equals $(\frac{1}{4}+\frac{3}{4}F_{IT})$, which includes the possibility of selfing. This leads to the following equation:

$$F_r(0)=\frac{\varphi_o}{\varphi_a+\varphi_o}\left(\tfrac{1}{4}+\tfrac{3}{4}F_{IT}\right)+\frac{\varphi_a}{\varphi_a+\varphi_o}$$
$$\cdot\left\{(1-\tfrac{1}{2}t)\sum_r F_r P_{r,f} N_r+\tfrac{1}{2}t\sum_r F_r P_{r,m} N_r\right\}. \qquad (22)$$

Because φ_a can be calculated as $\sum_r \varphi_r N_r$ and φ_o as $\varphi_r(r=0)$, φ_r being $C\cdot e^{-r^2/2\sigma_s^2}$, the substitution of a range of σ_s values indicates again the best estimate of σ_s. The variances used in $P_{r,f}$ and $P_{r,m}$ are σ_s^2 and $\sigma_s^2+\sigma_p^2$ respectively.

The determination of the value of $F_r(0)$ requires special treatment. The numbers of data in the smallest distance classes are relatively low, so that direct reading from the plotted curve is very inaccurate. The best estimation of $F_r(0)$ from the data is obtained by calculating the best fitting curve of known mathematical form, based on a sufficient distance range to also include the more accurate F_r values of the higher distance classes. This best fitting curve is calculated by the method of least squares; the empirically found mathematical function is

$$F_r(r)=a\,e^{-r^2/b}+c$$

valid for distances within the area which is 2-dimensional in practice.

Isolation by distance

The original derivations by Wright of formulae describing local effective population size ($N_e=4\pi d\sigma^2$) and isolation by distance (the parent is situated with a probability of $1-e^{-2}=0.865$ within a circle of radius 2σ and area $4\pi\sigma^2$) are directly linked by the fact that N_e is the number of individuals within the circle. Thus, in this simple situation of only one dispersion mechanism, σ^2 describes N_e and isolation by distance directly and simultaneously. Taking into account the three dispersal parameters σ_p, σ_s and t, neighbourhood size N_e gives a much more complicated formula (20), in which the smallest dispersal component (in females; seed dispersal only) dominates. Isolation by distance, on the contrary, will be governed by the highest dispersal component (in males; pollen + seed dispersal). By this reason, even when a circle relevant for isolation by distance is constructed which is related with the combined dispersal components, no straightforward relationship will exist between N_e and the number of individuals within the circle. Also, the probability of finding the parent within the circle is no longer 0.865, but will be different for male and female parents.

The only rationale for the choice of $4\pi\sigma^2$, the "neighbourhood area", N_a, as the measure of isolation by distance has been the coincidence of the local effective population size N_e and the number of individuals within the area N_a. Because this elegance of the model is lost in the more complex situation of seed plants, and the relationship $N_e=N_a$. d does not exist anymore, scrupulously sticking to Wright's model will lead to confusion. For that reason a new approach of the conception of isolation by distance has to be made. A more straightforward measure of isolation by distance is "the mean gene transport per generation (M)". The mean distance of the female parent from her offspring is $\sqrt{\frac{1}{2}\pi\sigma_s^2}=1.2533\sigma_s$. The mean distance of a male

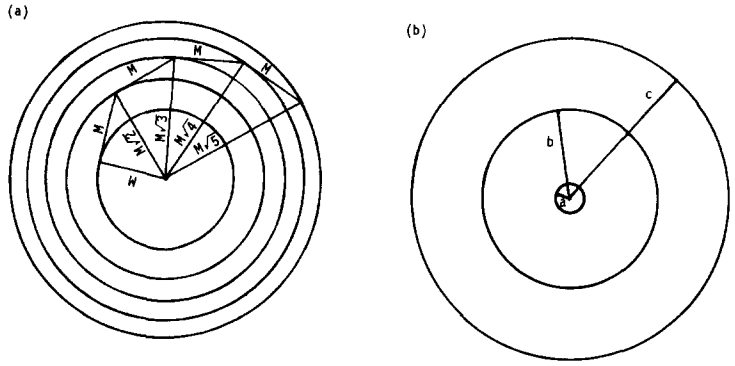

Fig. 4. **a** The mean gene transport in 1, 2, 3, 4 and 5 generations (the radii of the circles). **b** The mean distances of the grandparents (a: female line, b: mixed male and female line, c: male line). $\sigma_s = 1$; $\sigma_p = 10$; $t = 1$: a (♀♀) = 1.8; b (♀♂ and ♂♀) = 12.7 and c (♂♂) = 17.8

parent from his offspring is $\sqrt{\frac{1}{2}\pi(\sigma_s^2 + \sigma_p^2)}$ in the case of outcrossing and equals $\sqrt{\frac{1}{2}\pi\sigma_s^2}$ in the case of selfing, thus being $\sqrt{\frac{1}{2}\pi(\sigma_s^2 + t\sigma_p^2)}$ on average on the long-term. The total long-term mean gene transport will be brought about by equal numbers of male and female dispersion phases. The mean gene transport per generation is therefore:

$$M = \sqrt{\tfrac{1}{2}\pi(\sigma_s^2 + \tfrac{1}{2}t\sigma_p^2)} \; . \qquad (23)$$

In g generations the mean gene transport will be $M\sqrt{g}$ as illustrated in Fig. 4a. In a two-dimensional situation, the circle with radius $M\sqrt{g}$ has an area of $\pi g M^2$, increasing with πM^2 each generation. This "area per generation" equals $\frac{1}{2}\pi^2(\sigma_s^2 + \frac{1}{2}t\sigma_p^2)$ and is proportional to the neighbourhood area as calculated by Crawford (1984) which is $4\pi(\sigma_s^2 + \frac{1}{2}t\sigma_p^2)$, being the direct extension of Wright's N_a with respect to isolation by distance.

As mentioned, the mean gene transport per generation M is a long-term conception. Fig. 4b illustrates that in reality a binomial distribution exists of gene transport through male or female dispersion steps. This explains the existence of a certain degree of local differentiation and a restricted N_e in the case of a low σ_s and a very high σ_p: this local differentiation is brought about by the restricted gene transport in the pure female parental lines, and is a short-term effect only.

An additional advantage of using M instead of the neighbourhood area N_a is the possibility to calculate M from any pollen and seed distribution, which may be completely different from a normal distribution.

Computer simulations

In order to check the described methods for the estimation of the three gene flow parameters from a given internally structured population, a series of computer simulations have been carried out in which the natural process of the forming of local differentiation is imitated, using different gene flow regimes. After reaching equilibrium, the population structure developed is analysed by means of the "recipes" which are described in this paper. The recovery of the gene flow parameters will give information about the correctness and the reliability of the method.

The simulations are based on the following model which corresponds to a real situation of a population of *Plantago major* growing on a path: an area of 6 × 2 m is populated with plants, naturally limited at the long sides by decreasing density and abruptly ending at the short sides. The mean population density is about 34 plants per m^2. Each "simulation year" a certain percentage (25% in this simulations) of older plants, randomly chosen, is replaced by new plants at randomly chosen sites. For each new plant a female parent is selected from the neighbouring older plants, according to their expected local seed densities. From the female parent one set of alleles is chosen. A male parent is selected from the neighbouring plants of the female parent in an analogous way. The amount of selfing is determined by an introduced value of Z. Three loci with two alleles each are used; one locus may serves as a self-incompatibility (SI) locus (gametophytic system with 2N alleles) in some simulations. At the beginning of the simulations the population is in Hardy-Weinberg equilibrium without any linkage disequilibria. The starting allele frequencies are 0.5 in all cases, except for the SI-locus.

During the simulations the value of t is registered by counting the numbers of cross- and self-fertilizations. Also F_{IT} is followed, indicating equilibrium when its value does no longer change systematically. After 10 to 15 simulation years of constant F_{IT} the simulation is stopped. Actually a real equilibrium will not be reached, only a 'quasi-equilibrium', because long term changes will occur through genetic drift of the whole population. In order to consider real equilibrium, mutation and/or long-distance seed or pollen dispersal have to be included into the simulation model. In fact, incorporation of long distance seed transport (concerning 0.1% to 1% of all new plants) did not cause any essential difference in the simula-

Table 2. The gene flow regimes used for the computer sumulations (all distances in cm). d is about 34 plants per square meter

Regime no.	Input parameters				Observed	Calculated[b]		No. of simulation years
	σ_p	σ_s	Z	SI[a]	t	N_e	M	
1	20	5	0	+	1	4.2	18.8	30
2	20	5	0	–	1	4.3	18.8	30
3	20	5	1	–	0.856	3.7	17.6	30
4	20	5	5	–	0.564	2.6	14.7	35
5	20	5	50	–	0.125	1.8	8.9	45
6	5	15	1	–	0.263	8.5	18.9	50
7	100	20	0	+	1	33.8	92.1	25
8	100	20	100	–	0.437	19.6	63.7	35
9	∞[c]	1	1	–	0.997	4.0	∞	30

[a] + means that one locus has been used as a self-incompatibiltiy locus (gametophytic system) and two instead of three loci have been used for the population analysis

[b] For the calculation of N_e and M, the input values of σ_p and σ_s and the observed value of t have been used

[c] Random choice of the male parent, irrespective of his distance to the female parent

Table 3. Results of the computer simulations: F values and plant numbers

Regime and simulation no.	F_{IT}	$F_r(0)$	N	Regime and simulation no.	F_{IT}	$F_r(0)$	N
1-1	0.170	0.254	200	6-1	0.689	0.267	203
-2	0.070	0.224	190	-2	0.750	0.328	205
-3	0.152	0.303	205	-3	0.801	0.461	194
-4	0.018	0.241	196	-4	0.660	0.323	209
-5	0.068	0.306	197	-5	0.837	0.501	210
-combined	0.097	0.270	988	-combined	0.747	0.366	1,004
2-1	0.130	0.331	201	7-1	0.055	0.031	206
-2	0.211	0.336	195	-2	0.034	0.167	200
-3	0.270	0.384	210	-3	0.006	0.034	195
-4	0.140	0.203	198	-4	0.019	0.051	204
-5	0.126	0.312	199	-5	– 0.051	0.037	197
-combined	0.176	0.316	1,003	-combined	0.013	0.036	1,004
3-1	0.261	0.304	209	8-1	0.358	0.098	201
-2	0.268	0.325	206	-2	0.484	0.302	189
-3	0.270	0.367	212	-3	0.385	0.191	206
-4	0.211	0.267	200	-4	0.436	0.181	205
-5	0.149	0.250	196	-5	0.432	0.106	199
-combined	0.235	0.307	1,023	-combined	0.418	0.168	1,000
4-1	0.358	0.312	198	9-1	– 0.013	0.178	212
-2	0.484	0.506	200	-2	– 0.021	0.194	199
-3	0.475	0.439	195	-3	0.062	0.242	213
-4	0.453	0.368	194	-4	– 0.022	0.267	194
-5	0.432	0.522	202	-5	– 0.054	0.177	203
-combined	0.440	0.414	989	-combined	– 0.009	0.209	1,021
5-1	0.886	0.755	203				
-2	0.826	0.925	204				
-3	0.868	0.740	199				
-4	0.875	0.912	195				
-5	0.849	0.806	210				
-combined	0.860	0.825	1,011				

Table 4. Results of the computer simulations: estimations of the gene flow parameters[a]

Regime and simulation no.	σ_p[b]	σ_p	t	σ_s[c]	Regime and simulation no.	σ_p[b]	σ_p	t	σ_s[c]
1-1	15.3	6.0	1.173	8.5	6-1	4.7	∞	0.160	13.8
-2	33.7	∞	0.865	5.9	-2	1.9	∞	0.143	15.3
-3	28.6	4.2	1.643	4.3	-3	4.2	15.9	0.178	12.7
-4	64	18.5	1.188	0	-4	46	∞	0.204	13.9
-5	40	∞	0.867	0	-5	5.1	1.6	1.000	12.9
-combined	29.3	36.5	0.964	3.8	-combined	4.0	90	0.160	13.3
(expected)	(20)	(20)	(1)	(5)	(expected)	(5)	(5)	(0.263)	(15)
2-1	23.1	15.0	1.122	0.8	7-1	0	–	–	31.0
-2	13.5	11.2	1.080	6.4	-2	13.8	–	–	6.1
-3	15.5	62	0.711	4.2	-3	140	–	–	27.3
-4	14.8	7.3	1.206	11.9	-4	122	–	–	21.9
-5	40.1	320	0.792	0.7	-5	∞	–	–	18.1
-combined	18.7	22.2	0.955	4.6	-combined	70.5	–	–	24.8
(expected)	(20)	(20)	(1)	(5)	(expected)	(100)	–	–	(20)
3-1	16.0	11.6	0.925	8.8	8-1	∞	–	–	22.0
-2	20.5	13.8	0.977	9.5	-2	47	–	–	11.0
-3	20.0	26.0	0.781	5.1	-3	800	–	–	12.5
-4	21.6	19.4	0.873	8.7	-4	35	–	–	14.2
-5	31.4	∞	0.737	5.3	-5	30	–	–	21.0
-combined	21.4	26.0	0.811	6.8	-combined	88	∞	0.408	14.9
(expected)	(20)	(20)	(0.856)	(5)	(expected)	(100)	(100)	(0.437)	(20)
4-1	23.2	7.0	0.782	9.7	9-1	∞	–	–	3.2
-2	15.8	4.3	0.778	7.2	-2	∞	–	–	1.6
-3	14.3	5.5	0.848	8.7	-3	17	–	–	2.9
-4	13.4	∞	0.375	7.9	-4	∞	–	–	0
-5	29.4	5.3	1.474	5.4	-5	∞	–	–	1.4
-combined	17.4	8.4	0.750	8.6	-combined	∞	–	–	1.0
(expected)	(20)	(20)	(0.564)	(5)	(expected)	(∞)	–	–	(1)
5-1	11.4	0.9	0.844	6.9					
-2	38.1	∞	0.095	0					
-3	26.5	3.3	0.492	8.2					
-4	22.1	∞	0.067	0					
-5	14.6	25	0.113	4.9					
-combined	20.1	8.5	0.260	5.6					
(expected)	(20)	(20)	(0.125)	(5)					

[a] Zero values of σ_p or σ_s are indications of a higher $F_r(0)$ than can be explained by any positive value of the appropriate σ value. t values larger than one are not real

[b] The estimation of σ_p if the value of t would have been known

[c] σ_s has been calculated by using the estimated values of σ_p and t. If this was impossible ($\sigma_p = 0$ or not determined), the appropriate input or observed value has been used

tion results: the quasi-equilibrium is reached in a much shorter time than a distinct overall effect of genetic drift would need.

The subsequent analysis of the population structure evolved has been made usable for any number of loci with any number of alleles and any dimensionality of the population. In the situation mentioned, only the inner 5×2 m of the 6×2 m area has been used for the analysis to avoid border effects. The equations (21) and (22), which describe the height of $F_r(0)$ according to the two different models for seed dispersal, are both used because the turnover value is 25%: $F_r(0)$ is explained by (21) for 25% and by (22) for 75%.

Nine gene flow regimes have been chosen for the simulations (see Table 2) and five independent simulations have been carried out with each regime. In Table 3 the respective F values (F_{IT} and $F_r(0)$) at the end of each of the simulations are shown. The combined parallel simulations of one regime are also analysed together, indicated as 'combined'. The estimation results of the gene flow parameters have been summarized in Table 4 in the same way: in the separate and the combined simulations. A comparison of the 'combined' results with the expected values show clearly that the recovery of σ_p when t is known and of σ_s are very accurate. When both σ_p and t are un-

known, the results are less reliable, especially for lower t values. Populations of about 1,000 individuals and three or two (in the case of simulations 1 and 7) loci with intermediate allele frequencies are apparently large enough for reasonable results. The separate simulations (about 200 plants each) provide for an indication of the error of the estimation method. In some of the regimes also this number of individuals leads to reliable results.

In Table 5 the values of N_e and M, determined after the estimations of the gene flow parameters, are shown. In this case the figures of 'combined' and 'expected' are also rather close to each other. Sometimes M (and also σ_p) are estimated as infinitely high. In fact 'large' (e.g. 100 cm, being large with respect to the plant density used) and infinitely high are very close: the scale is not linear at all. The estimations of M are very susceptible for deviations in σ_p, whereas σ_s and N_e are rather insensible for the value of σ_p.

It is not possible to derive formulae which describe the variance in the estimation of the various parameters, nor is it possible to calculate that variance from the separate simulation results because of the 'non-linearity' of the parameter estimations. The sampling variance in F is given by Rasmussen (1964). For allele frequencies of 0.5, this sampling variance in a population of N individuals is:

$$\mathrm{Var}(F) = \frac{(1-F)(1+F)}{N}. \qquad (24)$$

A comparison can be made with the variances in F_{IT} and $F_r(0)$ calculated from the simulations. The variance in F_{IT} between loci in the same simulation appeared on the average to be 1.45× the variance as predicted by (24). This factor was 1.82 for the same locus in different parallel simulations. The variance in $F_r(0)$, which is an estimated value instead of a directly measured one, was on average 4.29× the variance expected according to (24).

Table 5. Results of the computer simulations: estimations of N_e and M[a]

Regime and simulation no.	N_e	M	Regime and simulation no.	N_e	M
1-1	5.9	12.1	6-1	8.7	∞
-2	5.4	∞	-2	10.1	∞
-3	4.8	7.2	-3	6.5	17.0
-4	3.7*	17.9*	-4	8.7	∞
-5	3.6*	∞*	-5	6.8	16.2
-combined	4.3	32.1	-combined	7.9	36.0
(expected)	(4.2)	(18.8)	(expected)	(8.5)	(18.9)
2-1	3.1	14.1	7-1	27.4	38.9
-2	3.5	13.1	-2	4.1	14.4
-3	3.3	46.6	-3	49.5	128.7
-4	6.1	16.5	-4	39.8	111.5
-5	2.8	252.4	-5	33.9	∞
-combined	4.1	20.1	-combined	38.2	69.8
(expected)	(4.3)	(18.8)	(expected)	(33.8)	(92.1)
3-1	5.2	14.8	8-1	23.0*	–
-2	6.0	17.0	-2	7.3*	–
-3	3.8	21.3	-3	9.5*	–
-4	6.1	19.4	-4	11.8*	–
-5	4.1	∞	-5	21.5*	–
-combined	5.0	22.4	-combined	12.3	∞
(expected)	(3.7)	(17.6)	(expected)	(19.6)	(63.7)
4-1	4.3	13.3	9-1	5.3	∞
-2	2.6	9.6	-2	4.4	∞
-3	3.7	11.8	-3	3.7	15.5
-4	3.9	∞	-4	4.0*	∞*
-5	2.3	8.9	-5	4.2	∞
-combined	3.9	12.6	-combined	4.0	∞
(expected)	(2.6)	(14.7)	(expected)	(4.0)	(∞)
5-1	2.3	8.7			
-2	1.1*	∞*			
-3	2.9	10.5			
-4	1.1*	∞*			
-5	1.8	9.7			
-combined	2.1	8.0			
(expected)	(1.8)	(8.9)			

[a] The values of N_e and M have been calculated by using the estimated values of σ_p, σ_s and t. If this was impossible ($\sigma_s = 0$ or any gene flow parameter not determined) the appropriate input or observed value has been used. In that case a '*' is added

* See footnote a

Discussion

As the method developed appears to be correct and gives reliable results using reasonable numbers of individuals, the question remaining is how realistic are the necessary conditions.

The first condition is that selection does not have any influence. The existence of local fitness differences of the alleles of the loci used may especially severely disturb the results. The choice of a homogeneous population area would therefore be necessary, although this does not completely exclude selectional differences. The use of allozyme loci, which are usually selectively neutral, diminishes the probability of selection even more, but tight linkage of allozyme loci influencing fitness is not unlikely, as has been found in *Plantago major* (Van Dijk 1984). Also, linkage with a self-incompatibility locus (Cornish et al. 1980; Van Dijk 1985), or with male sterility loci (Van Dijk 1985) may occur. A necessary precaution prior to the use of allozyme loci should be to check that they all behave in the same way with respect to the F_r-curves, which makes selection less likely.

A second condition is equilibrium. To reach equilibrium in the simulation studies needed 25 to 50

"years", departing from total randomness and with a yearly turnover of 25%. To restore the equilibrium after a disturbance, or to reach equilibrium after (partial) colonization, will take less time: the acquirement of a departure from Hardy-Weinberg equilibrium will always be a slower process than going into the direction of that equilibrium. A period of constant management and known absence of severe disturbances during several years (annual species) up to a few decades (perennial species with a low turnover) seems to be sufficient to reach equilibrium.

A series of other effects may also interfere, such as mutation, long distance seed transport and the existence of a seed bank. Self-incompatibility lowers the F_r-curve by diminishing genetic drift, but does not influence the method itself, as is confirmed by the simulation results. Mutation and long distance seed transport have essentially the same influence on population structure: the introduction of alleles independent of local allele frequencies. They form a necessary condition for long-term equilibrium because they prevent the population from total fixation of one genotype. The effects of mutation and long distance seed transport are expressed in a higher value of σ_s than caused by normal "passive" seed dispersal only. The existence of a seed bank implies a higher effective density than is concluded from the observed number of reproducing individuals. All density dependent estimates: σ_s and N_e, have to be corrected for that larger effective density.

The method as it is described in this paper does not claim to be completely applicable to all kinds of populations, but will provide a framework which may be modified according to the special needs of a real situation. Minimally, it will produce an idea of the order of magnitude of gene flow. Together with the knowledge of selectional differences within or between populations or ecotypes, a notion of the strength of gene flow forms a necessary condition in order to explain the distribution and the abundance of a species in different habitats.

References

Cornish MA, Hayward MD, Lawrence MJ (1980) Self-incompatibility in Ryegrass. 3. The joint segregation of S and *Pgi-2* in *Lolium perenne* L. Heredity 44:55–62

Crawford TJ (1984) The estimation of neighbourhood parameters for plant populations. Heredity 52:273–283

Kimura M, Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics 49:561–576

Levin DA, Kerster HW (1971) Gene flow in seed plants. Evol Biol 7:139–220

Malécot G (1969) The mathematics of heredity. Freeman, San Francisco

Rasmussen DI (1964) Blood group polymorphism and inbreeding in natural populations of the deer mouse *Peromyscus maniculatus*. Evolution 18:219–229

Schaal BA (1975) Population structure and local differentiation in *Liatris cylindracea*. Am Nat 109:511–528

Van Dijk H (1984) Genetic variability in *Plantago* species in relation to their ecology. 2. Quantitative characters and allozyme loci in *P. major*. Theor Appl Genet 68:43–52

Van Dijk H (1985) Allozyme genetics, self-incompatibility and male sterility in *Plantago lanceolata*. Heredity 52:53–63

Wright S (1943) Isolation by distance. Genetics 28:114–138

Wright S (1951) The genetical structure of populations. Ann Eug 15:323–354

Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19:395–420