Theoretical and Applied Genetics 46, 109-115 (1975)

A Model for Genetic Relationship among Offspring from Open-pollinated Plant Populations

Hans-Rolf Gregorius
Lehrstuhl für Forstgenetik und Forstpflanzenzüchtung der Universität Göttingen (BRD)

Summary. Based on monoecious, diploid plant species, a model is constructed to determine genetic relationship within the seed production of an open-pollinated population, characterized by its rates of self-fertilization, population density and mode of pollen dispersal. Genetic relationship is measured by the coefficient of inbreeding of a seed produced by a mother plant located at a specified place, or by the coefficient of kinship between two seeds, produced from the same mother plant or produced from two different mother plants separated by a certain distance. The influence of the single parameters on these coefficients is demonstrated by some typical examples, which show that dimensionality of the habitat (one- or two-dimensional), as well as, range and type of pollen dispersal, has little influence on the relationship between seed produced from the same mother plant and, on the other hand, emphasize the important role of the rate of self-fertilization and population density. Some remarks on how to apply Wright's concept of neighbourhood to continuous plant populations close this paper.

Introduction

Many plant breeding activities require detailed knowledge about the genetic relationship among offspring (seed) from open-pollinated plant populations (see e.g. Squillace, 1975). Considering monoecious plant species, these offspring may be related in different ways: they may originate from the same mother plant and thus be half-sibs; full-sibs resulting from cross-fertilization or full-sibs resulting from self-fertilization; or they may originate from different mother plants and thus cannot be selfed full-sibs but may be related in any of the other ways including further possible relationships. The aim of this paper is to determine average coefficients of kinship among seed harvested from the same mother plant as well as from two different plants separated by specified distances. Special interest is paid to the influence of the rate of self-fertilization, population density and range of pollen dispersal on these parameters; the statement of the problem requires no assumptions concerning seed dispersal.

The consideration of pollen dispersal introduces migration of genetic material in a continuous way and relates the present problem to the field of continuous migration; models for this have been constructed from different points of view, mainly by Wright (1943, 1946) and Malécot (1948, 1950, 1967). Because of the difficulties which arise from these and also from more recent models for continuous migration, as e. g. given by Maruyama (1972) and Fleming and Su (1974), when trying to apply them to actual plant populations, Gregorius (1975) worked out a restatement of Malécot's model for this purpose. The fact that the concrete results derived from this model mainly describe aspects which become important after a long series of generations and thus primarily have evolutionary implications makes it necessary to develop a representation of the model which is appropriate for an application to the above case.

The Model

The following considerations are based on monoecious, diploid plant populations which are distributed over their habitats according to a specified population density. The seed production of each plant, in this context denoted as mother plant, is broken down into two parts, one resulting from self-fertilization and the other resulting from cross-fertilization. The cross-fertilized part of the seed production of a mother plant is accomplished under the assumption that each cross-pollinating pollen has an equal chance to come to fertilization. All members of the population flower at the same time and produce pollen to the same extent. Any form of gametic selection and mutation is excluded or at least regarded as negligible. With these assumptions, coefficients of inbreeding and kinship among the seed production of mother plants located at different places are computed; the coefficients refer to the one locus case.

The following notations will be used:

$\mu(x)$:= population density at place x

$q(x)$:= rate of self-fertilization of a plant located at place x, i.e. the probability that a seed produced by a mother plant at place x results from self-fertilization

$p(x|y)$:= probability density that a pollen which came to fertilization at place y originates from place x

$f(x)$:= coefficient of inbreeding of an individual from the parental generation located at place x

$\Phi(x,y)$:= coefficient of kinship between two individuals from the parental generation, one located at place x and the other at place y

$\Phi_0(x)$:= average coefficient of kinship of two seeds randomly drawn from the seed production of a plant located at place x

$\Phi_1(x,y)$:= average coefficient of kinship of two seeds randomly drawn from the seed production of two different plants, one located at place x and the other at place y

$f_1(x)$:= average coefficient of inbreeding of a seed randomly drawn from the seed production of a plant located at place x.

Computation of $\Phi_0(x)$:

By ♀, $♂_s$, $♂_c$ we denote the three different types of gametes that can be contained in a seed produced from a plant located at a certain place; ♀ is the female gamete, $♂_s$ is the male gamete which comes from the same plant (self-fertilization) and $♂_c$ is the male gamete coming from a different plant (cross-fertilization). Thus two types of seed, viz. $♂_s♀$, $♂_c♀$ exist, which can be combined into three different pairs, i.e. $(♂_s♀, ♂_s♀)$, $(♂_c♀, ♂_s♀)$ and $(♂_c♀, ♂_c♀)$, having probabilities $q(x)^2$, $2q(x)(1-q(x))$ and $(1-q(x))^2$ to be drawn from the seed production of a plant, located at place x. The coefficients of kinship for these three pairs may each be obtained by determining the identity relations between the four different combinations of two gametes, one drawn from the first and the other from the second seed:

$$(♂_s♀, ♂_s♀) : \frac{1}{2}(1+f(x))$$

$$(♂_s♀, ♂_c♀) : \frac{1}{4}(1+f(x)) + \frac{1}{2}\cdot\int\Phi(x,y)p(y|x)\,dy$$

$$(♂_c♀, ♂_c♀) : \frac{1}{8}(1+f(x)) + \frac{1}{2}\cdot\int\Phi(x,y)p(y|x)\,dy + \frac{1}{4}\cdot\int\Phi(y,z)p(y|x)p(z|x)\,dydz + \frac{1}{4}\cdot\int(\frac{1}{2}(1+f(y)) - \Phi(y,y))\frac{p(y|x)^2}{\mu(y)}\,dy.$$

The last integral contributes to the fact that the two cross-fertilizing gametes $♂_c$ can originate from the same plant (the derivation method is analogous to that given by Malécot (e.g. 1967) and Gregorius (1975)). Summing up all the terms we are led to

$$\Phi_0(x) = \frac{1}{2}(1+f(x))\frac{1}{4}(1+q(x))^2 + \frac{1}{2}(1-q(x)^2)\times \int\Phi(x,y)p(y|x)\,dy + \frac{1}{4}(1-q(x))^2\Big[\int\Big(\frac{1}{2}(1+f(y)) - \Phi(y,y)\Big)\frac{p(y|x)^2}{\mu(y)}\,dy + \int\Phi(y,z)p(y|x)p(z|x)\,dydz\Big]. \quad (1)$$

Computation of $\Phi_1(x,y)$:

Again, choosing the above way of proceeding and considering that we now are concerned with pairs of seeds produced from two different plants located at place x and y, the four types $(♂_s♀_x, ♂_s♀_y)$, $(♂_s♀_x, ♂_c♀_y)$, $(♂_c♀_x, ♂_s♀_y)$ and $(♂_c♀_x, ♂_c♀_y)$ of pairs of seed can be distinguished and have the respective probabilities $q(x)q(y)$, $q(x)(1-q(y))$, $(1-q(x))q(y)$ and $(1-q(x))(1-q(y))$ to be drawn (note that e.g. $♂_s♀_x$ now describes a seed resulting from self-fertilization at place x). The coefficients of kinship for these four pairs are:

$$(♂_s♀_x, ♂_s♀_y) : \Phi(x,y)$$

$$(♂_s♀_x, ♂_c♀_y) : \frac{1}{2}\Phi(x,y) + \frac{1}{2}\cdot\int\Phi(z,x)p(z|y)\,dz + \frac{1}{2}\cdot\Big[\frac{1}{2}(1+f(x)) - \Phi(x,x)\Big]\frac{p(x|y)}{\mu(x)}$$

$$(♂_c♀_x, ♂_c♀_y) : \frac{1}{2}\Phi(x,y) + \frac{1}{2}\cdot\int\Phi(z,y)p(z|x)\,dz + \frac{1}{2}\cdot\Big[\frac{1}{2}(1+f(y)) - \Phi(y,y)\Big]\frac{p(y|x)}{\mu(y)}$$

$$\begin{aligned}(♂_c♀_x, ♂_c♀_y) : {} & \frac{1}{4}\Phi(x,y) + \frac{1}{4}\cdot\int\Phi(z,y)p(z|x)\,dz + \\ & + \frac{1}{4}\cdot\Big[\frac{1}{2}(1+f(y)) - \Phi(y,y)\Big]\frac{p(y\ x)}{\mu(y)} + \\ & + \frac{1}{4}\cdot\int\Phi(z,x)p(z|y)\,dz + \\ & + \frac{1}{4}\cdot\Big[\frac{1}{2}(1+f(x)) - \Phi(x,x)\Big]\frac{p(x|y)}{\mu(x)} + \\ & + \frac{1}{4}\cdot\int\Phi(u,v)p(u|x)p(v|y)\,dudv + \\ & + \frac{1}{4}\cdot\int\Big(\frac{1}{2}(1+f(z)) - \Phi(z,z)\Big)\frac{p(z|x)p(z|y)}{\mu(z)}\,dz.\end{aligned}$$

Thus altogether

$$\begin{aligned}\Phi_1(x,y) = {} & \frac{1}{4}(1+q(x))(1+q(y))\cdot\Phi(x,y) + \frac{1}{4}(1+q(x))(1-q(y))\times \\ & \times\Big[\Big(\frac{1}{2}(1+f(x)) - \Phi(x,x)\Big)\frac{p(x|y)}{\mu(x)} + \int\Phi(z,x)p(z|y)\,dz\Big] + \\ & + \frac{1}{4}(1-q(x))(1+q(y))\Big[\Big(\frac{1}{2}(1+f(y)) - \Phi(y,y)\Big)\frac{p(y|x)}{\mu(y)} + \\ & + \int\Phi(z,y)p(z|x)\,dz\Big] + \frac{1}{4}(1-q(x))(1-q(y))\Big[\int\Big(\frac{1}{2}(1+f(z)) - \\ & - \Phi(z,z)\Big)\frac{p(z|x)p(z|y)}{\mu(z)}\,dz + \\ & + \int\Phi(u,v)p(u|x)p(v|y)\,dudv\Big]. \end{aligned} \quad (2)$$

For pure self-fertilization at places x and y, i.e. $q(x) = q(y) = 1$, equations (1) and (2), of course, specialize to the well known results $\Phi_0(x) = \frac{1}{2}(1+f(x))$ and $\Phi_1(x,y) = \Phi(x,y)$.

Computation of $f_1(x)$:

Clearly the male gamete contained in a seed can be of type ♂_s or ♂_c thus generating the following probabilities of identity between the alleles carried by the male and female gamete:

$\text{♂}_s\text{♀}$: $\frac{1}{2}(1+f(x))$; $\text{♂}_c\text{♀}$: $\int \Phi(y,x)p(y|x)dy$. Therefore

$$f_1(x) = (1-q(x)) \cdot \int \Phi(y,x)p(y|x)dy + q(x) \cdot \frac{1}{2}(1+f(x)). \tag{3}$$

Obviously there is no direct dependence of f_1 on the population density μ.

In plant breeding experiments there is usually no exact knowledge about the true values of the $\Phi(x,y)$ and $f(x)$, but rather, an idea about the magnitude of the average coefficient of kinship resp. inbreeding. This fact can be taken into account assuming the coefficients to be independent of the locations of individuals, i.e. $\Phi(x,y) = \Phi$ and $f(x) = f$. Thus Eqs. (1), (2) and (3) change into

$$\Phi_0(x) = \Phi + \left[\frac{1}{2}(1+f) - \Phi\right] \cdot \frac{1}{4}\left[(1+q(x))^2 + (1-q(x))^2 \cdot K(x)\right],$$

$$\text{with } K(x) := \int \frac{p(y|x)^2}{\mu(y)} dy; \tag{1a}$$

$$\begin{aligned} \Phi_1(x,y) = \Phi + \left[\frac{1}{2}(1+f) - \Phi\right] \cdot \frac{1}{4} \Big[& \frac{p(x|y)}{\mu(x)}(1+q(x))(1-q(y)) + \\ & + \frac{p(y|x)}{\mu(y)}(1-q(x))(1+q(y)) + \\ & + (1-q(x))(1-q(y)) \cdot K(x,y)\Big] \end{aligned} \tag{2a}$$

$$\text{with } K(x,y) := \int \frac{p(z|x)p(z|y)}{\mu(z)} dz;$$

$$f_1(x) = \Phi + \left[\frac{1}{2}(1+f) - \Phi\right] \cdot q(x). \tag{3a}$$

The meaning of the quantities $K(x)$ and $K(x,y)$, which comprise the compound influence of pollen dispersal and population density, can be stated as follows: under the condition of cross-fertilization $K(x)$ the probability that two pollen cells which come to fertilize a plant located at place x originate from the same plant, and $K(x,y)$ is the probability that two pollen cells which come to fertilization at two different plants located at place x resp. y originate from the same plant. Furthermore, $\frac{p(x|y)}{\mu(x)}$ approximately is the probability that a pollen cell which came to fertilize a plant located at place y originates from just one different plant located at place x; therefore $\frac{p(x|x)}{\mu(x)} \equiv 0$. In case the coefficients of kinship of an individual with itself, $\frac{1}{2}(1+f)$, and between two different individuals, Φ, are equal, then $\Phi_0(x) \equiv \Phi \equiv \Phi_1(x,y) \equiv f_1(x)$.

Assuming that all individuals produce seed to the same extent, the average coefficient of kinship resp. inbreeding within the seed population, Φ' resp. f' say, can be defined with respect to the population density, if, for the moment, we assume the population size to be finite, i.e. $\int \mu(x)dx =: N < \infty$, and put

$$\frac{\mu(x)}{N} =: \varepsilon(x) : \Phi' = \int \Phi_1(x,y)\varepsilon(x)\varepsilon(y)dxdy + \\ + \int (\Phi_0(x) - \Phi_1(x,x)) \cdot \frac{\varepsilon(x)^2}{\mu(x)} dx,$$

which is derived with the help of the same reasoning we used in the computation of Φ_0 and Φ_1;

$$f' = \int f_1(x)\varepsilon(x)dx.$$

Some straightforward substitutions and rearrangements lead to a representation of Φ' and f' in the form

$$\Phi' = \Phi + (\frac{1}{2}(1+f) - \Phi) \cdot M, \text{ with } M := \frac{1}{4} \cdot \int \frac{1}{\mu(x)} \\ \left[\int (1-q(y))p(x|y)\varepsilon(y)dy + (1+q(x))\varepsilon(x)\right]^2 dx,$$

$$f' = \Phi + (\frac{1}{2}(1+f) - \Phi) \cdot \overline{q}, \text{ with } \overline{q} := \int q(x)\varepsilon(x)dx,$$

which are exactly the same representations I obtained earlier applying a different approach to the same situation (Gregorius (1975)). As $N \to \infty$ we arrive at the respective result for hypothetically infinite population size.

Small Size of Habitat

A special situation occurs if the range of pollen dispersal is large compared with the size of population's habitat. In this case cross-pollination, and as a consequence cross-fertilization, may be regarded as taking place approximately at random. As before, $\int \mu(x)dx =: N$ is the population size. Recalling the meanings of $K(x)$, $K(x,y)$ and $\frac{p(x|y)}{\mu(x)}$, which have been explained above, the values they adopt for random cross-fertilization can be deduced easily:

$$K(x) = \frac{1}{N-1}, \quad \frac{p(x|y)}{\mu(x)} = \frac{1}{N-1} \text{ and } K(x,y) = \frac{N-2}{(N-1)^2}.$$

With these results, Eqs. (1a) and (2a), after some simple rearrangements, attain the form

$$\Phi_0(x) = \Phi + \left[\frac{1}{2}(1+f) - \Phi\right] \cdot \frac{1}{4} \cdot \left[(1+q(x))^2 + \\ + (1-q(x))^2 \cdot \frac{1}{N-1}\right] \tag{1b}$$

$$\Phi_1(x,y) = \Phi + \frac{1}{N-1} \cdot \left[\frac{1}{2}(1+f) - \Phi \right] \times$$
$$\times \left[1 - \frac{1}{4}(1+q(x))(1+q(y)) - (1-q(x))(1-q(y)) \frac{1}{4(N-1)} \right] . \tag{2b}$$

Obviously, random cross-fertilization cancels out the influence of population density. The decrease of genetic relationship with increasing population size N is much stronger for seed originating from different plants than it is for seed produced by the same plant; to be concrete, if $\Phi_0(x)$ is considered as a function of N it never can remain under $\Phi + \frac{1}{4}(1+q(x))^2\left(\frac{1}{2}(1+f) - \Phi\right)$. Note that $N = 2$, $q(x) \equiv 0 = \Phi = f$ reflects the case of cross-fertilized full-sibs from non-related and non-inbred parents and Φ_0 as well as Φ_1 should be equal to $\frac{1}{4}$, which is confirmed by (1b) and (2b).

Large Size of Habitat

The size of the habitat is considered to be large if the reverse of the statement in the last section applies, i.e. if the range of pollen dispersal is small compared with the size of the habitat. Provided the population density is the same everywhere, i.e. $\mu(x) \equiv \mu$, we may assume the probability densities $p(x|y)$ to be independent of locations y in most parts of the population and additionally to be radially symmetric; that is, $p(x|y)$ is a function of the distance $\| x - y \|$ between locations x and y only, which can be written as $p(x|y) = p(\| x-y \|)$. With these notations

$$K(x) = \frac{1}{\mu} \cdot \int p(\|y\|)^2 dy =: K \text{ and } K(x,y) =$$
$$= \frac{1}{\mu} \cdot \int p(\|z-x\|) p(\|z-y\|)\, dz.$$

Thus (1a) and (2a) change into

$$\Phi_0(x) = \Phi + \left[\frac{1}{2}(1+f) - \Phi \right] \cdot \frac{1}{4} \cdot \left[(1+q(x))^2 + (1-q(x))^2 \cdot K \right] \tag{1c}$$

$$\Phi_1(x,y) = \Phi + \left[\frac{1}{2}(1+f) - \Phi \right] \cdot \frac{1}{4} \cdot \left[2\, \frac{p(\|x-y\|)}{\mu} (1 - q(x)q(y)) + \right.$$
$$\left. + (1-q(x))(1-q(y)) \cdot K(x,y) \right] . \tag{2c}$$

To get a concrete idea of the effect which a limited range of pollen dispersal has on the various coefficients of kinship, we shall investigate some examples which employ simple types of probability densities p of the kind used in (1c) and (2c). In many situations rough approximations for the actual p may be obtained by assuming the pollen dispersal to be uniform within a certain distance limit R, which, for example, in the case of a two-dimensional habitat, means that all plants within a circle of radius R, except for the one which is assumed to be located in the centre of the circle, have an equal chance to contribute as fathers to the seed production of the central mother plant; on the other hand, pollen originating from outside that circle should not reach the central mother plant, i.e. have probability 0 of coming to fertilization at this plant. From this, p has the following representation

$$\frac{1}{\mu} p(\|x\|) = \begin{cases} \frac{1}{N-1} & \text{for } \|x\| \leqslant R \\ 0 & \text{for } \|x\| > R \end{cases} ,$$

where N is the number of plants contained in a circle of radius R (two-dimensional habitat) resp. a straight line (one-dimensional habitat) of length 2R.

a) We will start with a one-dimensional habitat, i.e. individuals are distributed along a straight line according to a population density given by μ. In this case $N = 2R\mu$ and $K = \frac{1}{N-1}$.

$K(x,y)$ has to be computed for three different cases separately. First, $|x-y| > 2R$: the intervals centred around x and y respectively don't overlap and thus $K(x,y) = 0$; secondly $R < |x-y| \leqslant 2R$: the two intervals do overlap and the overlapping part does not contain the two individuals located at places x and y, so that the number of individuals which can contribute pollen at places x and y at the same time is equal to $\mu \cdot (2R - |x-y|)$; thirdly, $|x-y| \leqslant R$: the overlapping part of the two intervals contains the individuals at place x and y and thus the number of individuals which can contribute pollen at places x and y at the same time is equal to $\mu \cdot (2R - |x-y|) - 2$. Altogether we therefore have

$$K(x,y) = \begin{cases} 0 & \text{for } |x-y| > 2R \\ \dfrac{\mu \cdot (2R - |x-y|)}{(N-1)^2} & \text{for } R < |x-y| \leqslant 2R \\ \dfrac{\mu \cdot (2R - |x-y|) - 2}{(N-1)^2} & \text{for } |x-y| \leqslant R. \end{cases}$$

This means that $\Phi_1(x,y)$ can be regarded as a function of the difference $|x-y|$ and thus is composed of three straight lines with a discontinuity at the point $|x-y| = R$, which is demonstrated by Fig. 1.

b) For the case of a two-dimensional habitat, as described above, considerations are analogous. We now have $N = \mu \cdot R^2 \pi$ and $K = \frac{1}{N-1}$ as before. The number of individuals located within the intersection of two cirles which

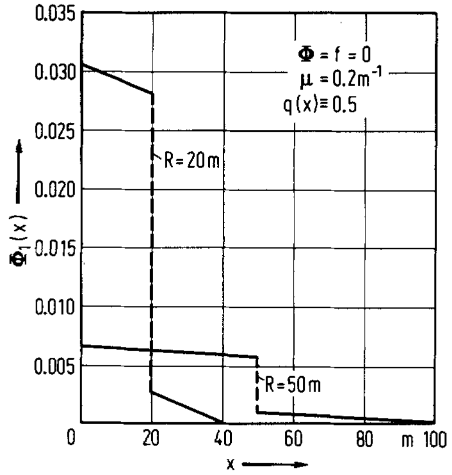

Fig. 1. Coefficient of kinship $\Phi_1(x)$ of two seed produced from plants separated by a distance x. The plants are distributed along a one-dimensional habitat with population density μ. Pollen dispersal takes place according to a uniform probability density with range R. q = rate of self-fertilization, Φ resp. f = coefficient of kinship resp. inbreeding in the parental generation. The respective coefficient of kinship Φ_0 of two seed produced by the same plant equals in case R = 20: $\Phi_0 = 0.2857$ and in case R = 50: $\Phi_0 = 0.2829$

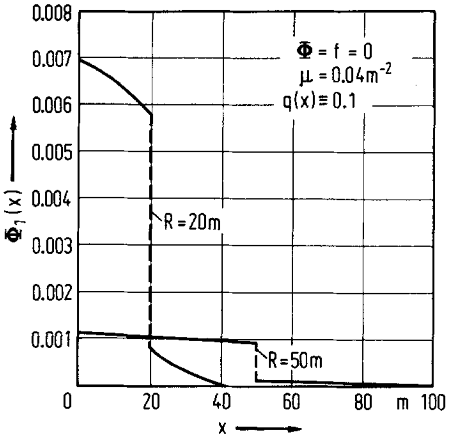

Fig. 2. Notations as given in the legend of Fig. 1. The plants are distributed over a two-dimensional habitat and pollen dispersal takes place according to a radially symmetric uniform probability density with range R. In case R = 20: $\Phi_0 = 0.1533$ and in case R = 50: $\Phi_0 = 0.1516$

have radius R and are centred around places x and y respectively ($\|x-y\| \leqslant 2R$) is equal to

$$\mu \cdot \left[2R^2 \cdot \arccos \frac{\|x-y\|}{2R} - \|x-y\| \sqrt{R^2 - \frac{\|x-y\|^2}{4}}\right] =: I(x,y)$$

as is well known from elementary geometry. With this,

$$K(x,y) = \begin{cases} 0 & \text{for } \|x-y\| > 2R \\ \dfrac{I(x,y)}{(N-1)^2} & \text{for } R < \|x-y\| \leqslant 2R \\ \dfrac{I(x,y) - 2}{(N-1)^2} & \text{for } \|x-y\| \leqslant R \ . \end{cases}$$

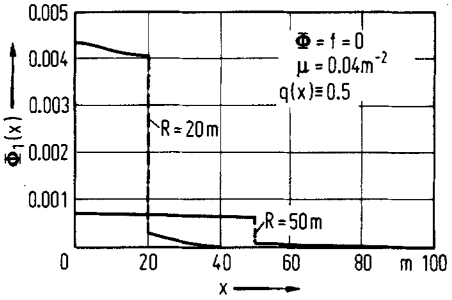

Fig. 3. Notations and preassumptions as given in the legend of Fig. 2, but now in case R = 20: $\Phi_0 = 0.2819$ and in case R = 50: $\Phi_0 = 0.2813$

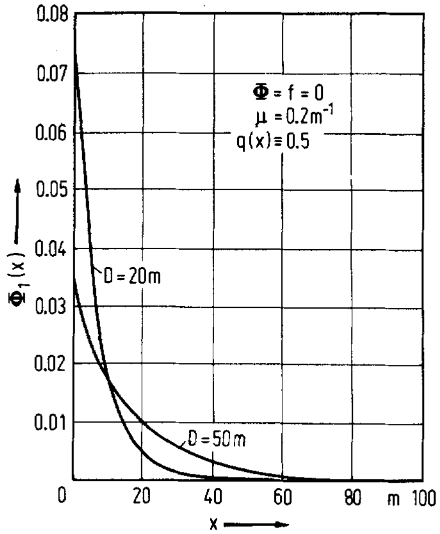

Fig. 4. Notations as given in the legend of Fig. 1. The plants are distributed along a one-dimensional habitat and pollen dispersal takes place according to an exponential probability density with 'range' D as defined in the text. In case D = 20: $\Phi_0 = 0.2871$ and in case D = 50: $\Phi_0 = 0.2836$

Again $\Phi_1(x,y)$ can be regarded as a function of $\|x-y\|$ with a point of discontinuity at $\|x-y\| = R$. The following figures additionally give an idea of the change of Φ_1 with R and q:

c) The change in shape of Φ_1 when a continuous function for p is chosen may be demonstrated by applying an exponential probability density to the case of a one-dimensional habitat:

$$p(|x|) = \frac{\alpha}{2} \cdot \exp(-\alpha|x|) \quad 0 < \alpha .$$

With the help of this function the integrals which determine K and $K(x,y)$ can be carried out easily and lead to

$$K = \frac{\alpha}{4\mu} \text{ and } K(x,y) = \frac{\alpha^2}{4\mu}\left(|x-y| + \frac{1}{\alpha}\right) \cdot \exp(-\alpha \cdot |x-y|).$$

A graphical representation of this case for $\Phi_1(x,y)$, which again just depends on $|x-y|$, is given in Fig.4. To give the quantity α a more clear meaning, it will be related to the 'range' of pollen dispersal which now is defined as the distance D within which approximately 95% of the pollen comes to fertilization, i.e., has to be such determined such that

$$\frac{\alpha}{2} \cdot \int_{-D}^{D} \exp(-\alpha|x|)dx = 0.95.$$

It can be shown easily that $D \cong \frac{3}{\alpha}$. Thus in the preceding formula 3/D has to be substituted for α.

The figures of all three examples demonstrate a considerable difference between the maximum value of $\Phi_1(x)$ and Φ_0. The influence of the range of pollen dispersal R resp. D on $\Phi_1(x)$ is seen to be important, while there is hardly any variation of Φ_0 with changing range as well as shape of pollen dispersal (see legend of Figs. 1 and 4). Another drastic change of $\Phi_1(x)$ is due to the dimensionality of the habitat as shown in Figs. 1 and 3, where all parameters are kept constant, but the population density μ for the one-dimensional case has been chosen as just the root of the density for the two-dimensional case, to make both comparable. On the other hand, there is no strong dependence of Φ_0 on dimensionality. A remarkable change in Φ_0 obviously is caused by a change in the rate of self-fertilization, as can be seen from the legends of Figs. 2 and 3.

The Meaning of Neighbourhood

The term "neighbourhood" is due to Wright (1946), who suggested 'that the individuals are neighbours in the sense that their gametes may come together'. This coming together of gametes generally takes place with different probabilities and, at an extreme, all these probabilities may be greater than zero - even in an hypothetically infinite population - which should imply that all gametes may come together. To overcome this difficulty a finite theoretical population is constructed in which all matings in question are equally probable and for which an appropriate parameter adopts the same value as in the actual population. This parameter is widely agreed to be the probability that two successful gametes are derived from the same individual. In general these two gametes can be taken to correspond to each other in various concerns, as e.g. forming a zygote or being present in two specific individuals. The use of this parameter certainly is appropriate because it reflects the possibility of a single individual contributing to the whole process and thus makes a statement about the 'finiteness' of the number of individuals involved.

In its most elementary form a neighbourhood is defined with respect to a specified individual and describes a set of further individuals (i.e. excluding the specified individual) which are regarded as neighbours of the specified individual. Such a neighbourhood will be called effective if all its members are equally probable as mates for the specified individual, i.e. if they constitute an ideal (theoretical) population in the above sense. Thus, starting from an actual situation, it is possible to determine the size of an effective neighbourhood with the help of the probability that mating of the specified individual takes place two times with any other individual. For plant populations, this is equivalent to computing the probability that two crossfertilizing pollen at a given plant originate from the same plant. In our preceding statements this probability has been denoted by $K(x)$ for a plant located at place x. If the effective size of the neighbourhood of a plant located at place x is equal to $N_e(x)$ (not counting the plant at place x) then $K(x) = \frac{1}{N_e(x)}$ resp. $N_e(x)V = \frac{1}{K(x)}$, which can be shown easily.

A reasonable extention of the concept of effective neighbourhood to two different specified individuals may be deduced by considering the effective common neighbourhood of these two individuals which formally is defined as the intersection of the two effective neighbourhoods. The size of the effective common neighbourhood of two plants located at places x and y respectively will be denoted by $N_e(x,y)$ and allows us to determine the probability that two cross-fertilizing pollen at place x and y originate from the same plant just in terms of effective sizes, namely, $\frac{N_e(x,y)}{N_e(x) \cdot N_e(y)}$. Reflecting the actual situation, this expression has to be equated to $K(x,y)$, i.e. $K(x,y) = \frac{N_e(x,y)}{N_e(x) \cdot N_e(y)}$, and thus enables us to relate $N_e(x,y)$ to the actual situation:

$$N_e(x,y) = \frac{K(x,y)}{K(x) \cdot K(y)} \cdot$$

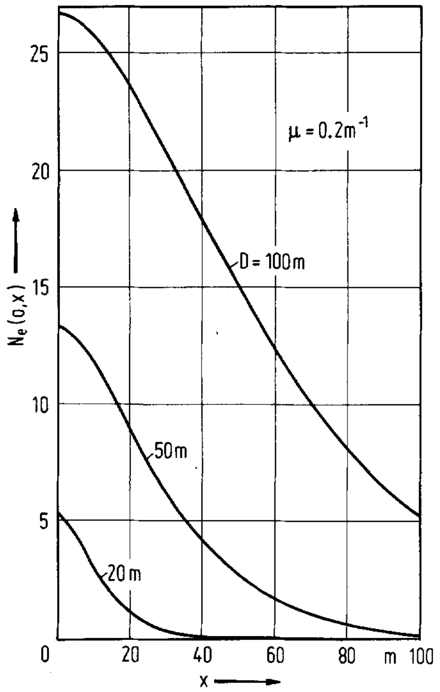

Fig.5. The size $N_e(0,x)$ of the effective common neighbourhood of two plants separated by distance x. The plants are distributed along a one-dimensional habitat with population density μ. Pollen dispersal takes place according to an exponential probability density with 'range' D as defined in the text.

The usefulness of the concept of effective neighbourhood can be judged by looking at Eqs.(1a) and (2a). It is seen that, besides the 'initial conditions' for Φ and f and the rate of self-fertilization, $\Phi_0(x)$ is governed by the effective size of neighbourhood solely. $\Phi_1(x,y)$ in addition needs knowledge about the effective size of common neighbourhood and a quantity $p(x|y)/\mu(x)$ which describes the extent of exchange of genetic material between the two plants located at places x and y respectively. Thus Malécot's claim (1969, p. 76) that Wright's concept of effective number 'does not have the weight that he attributes to it, because it does not account for the correlation with distance', is shown not to hold true in general.

If we compare the results for K and K(x,y), as given in examples a) and b), with the representations we obtain now, it can be stated that these examples directly reflect the above findings. The probability density for pollen dispersal applied in example c) now shall serve for demonstrating the connection between continuous dispersal and effective sizes. Again, as has been done in example c), 3/D is substituted for a and thus leads to

$$N_e(x) = N_e = \frac{4}{3} \cdot D\mu \quad \text{and}$$

$$N_e(x,y) = 4\mu\left(|x-y| + \frac{1}{3}D\right) \cdot \exp\left(-\frac{3}{D}|x-y|\right).$$

There is complete formal concordance with Wright's results (1946, Eq.(3)) for $N_e(x)$, provided we neglect constants and accept the standard deviation (of a normal distribution) as a measure of the range of gamete dispersal. To prevent possible misunderstanding it should be stated that in general the effective neighbourhood can not be described in terms of areas having simple geometrical structures. This becomes especially clear if we try to conceive the effective common neighbourhood as the intersection of areas, which is possible, e.g. in example a), but which proves to be impossible if we aim at a representation of example c) using the geometric constellation of example a), as can be taken from the above expression for $N_e(x,y)$. Finally, Fig.5 gives an impression of the conduct of $N_e(x,y)$ which, taking into account the above equation, can be regarded as a function of the distance $|x-y|$ only for different D's.

Literature

Fleming, W.H.; Chau-Hsing Su: Some One-Dimensional Migration Models in Population Genetics Theory. Theor. Popul. Biol. 5, 431-449 (1974)

Gregorius, H.-R.: A model for genetic relationship in areally continuous plant population. (1975, in preparation)

Malécot, G.: Les mathématiques de l'hérédité, Vol.1. Paris: Masson et Cie. 1948

Malécot, G.: Quelques schemas probabilities sur la variabilité des populations naturelles. Ann. Univ. Lyon, Sciences, Section A 13, 37-60 (1950)

Malécot, G.: Identical loci and relationship. Proc. 5th Berkeley Symp. Math. Stat. Prob. 4, 317-332 (1967)

Malécot, G.: The Mathematics of Heredity. San Francisco: Freeman & Co. 1969

Maruyama, T.: Rate of decrease of genetic variability in a two-dimensional continuous population of finite size. Genetics 70, 639-651 (1972)

Squillace, A.E.: Average Genetic Correlations Among Offspring From Open-Pollinated Forest Trees. Silvae Genetica (1975, in press)

Wright, S.: Isolation by distance. Genetics 23, 114-138 (1943)

Wright, S.: Isolation by distance under diverse systems of mating. Genetics 31, 39-59 (1946)

Received November 20, 1974
Communicated by H. Stubbe

Dr. Hans-Rolf Gregorius
Lehrstuhl für Forstgenetik und
Forstpflanzenzüchtung der Universität Göttingen
Büsgenweg 2
D-34 Göttingen-Weende (Germany/BRD)