Theoretical and Applied Genetics 46, 295-305 (1975)

Genetic Structures in Finite, Open-pollinated Plant Populations: A Model and its Application to Seed Orchards

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

Summary. A model has been constructed to investigate the consequences of the rate of self-fertilization, pollen-dispersal, population-size, and number of clones on the genetic structure of finite seed plant populations. Derivations have been performed for two different cases:
A) Parental genetic structure explicitly given: inferences for the expected genetic structure of the resulting seed population;
B) Extension of case A) to several non-overlapping generations.

If random cross-fertilization is assumed for case A) the genetic composition does not change and the genetic distance between the corresponding Hardy-Weinberg-structure and the expected offspring-structure is 0 if the rate of self-fertilization is equal to $\frac{1}{N}$ (N = population-size); any deviation from $\frac{1}{N}$ causes an increase in genetic distance.

In case B) the expected genetic structures have been derived for all generations and it was possible to establish a comparatively simple dependence on the coefficient of inbreeding. In addition the variance of the allele-frequency has been presented. All the above influential components can be summarized by a single quantity, called M. After proving that $\frac{1}{M}$ can be conceived as the effective population-size, all the results obtained could be presented depending on this effective size and the average rate of self-fertilization only.

Applying the findings of the model to the situation realized approximately in a seed-orchard, the following statements can be made:
Case A) Again assuming random cross-fertilization, a deviation of the parental population from the corresponding Hardy-Weinberg-proportions can, with increasing rate of self-fertilization, be exceeded by the respective deviation of the seed population. Case B) The influence of pollen dispersal on the effective population size has been investigated, assuming no variation of the individual rates of self-fertilization, pollen and seed production within the population. Only extremely small differences between effective and actual population size were obtained, which indicates that the influence of pollen dispersal is of minor importance in this case. For different rates of self-fertilization, significant differences in the increments per generation for the coefficients of inbreeding, as well as the frequency of homozygotes, were obtained for the first generation only. Decreasing number of clones influences the rate of self-fertilization and the effective population size simultaneously by increasing the first and decreasing the latter. This is transferred to the coefficient of inbreeding, frequency of the homozygotes and the variance of the allele frequency by an increase of increments for all generations.

Introduction

Any kind of breeding system may be regarded as acting upon the genetic variability contained in a real or hypothetical base population, where the pattern of variability is completely determined by the set of corresponding genotypic frequencies - denoted as the genetic structure - which in turn fix the set of gene frequencies - denoted as the genetic composition. This variability may be reduced where the predominating components of the breeding system are, for example special kinds of selecting or drift in small populations, which at the extreme can cause a high amount of inbreeding and a loss of genes originally present in the base population. A phenomenon of this kind constitutes a risk, the importance of which depends heavily on the kind of organism considered. If, for instance, plant species are considered which occupy very heterogeneous environments or have long generation cycles, reduced genetic variability will imply low adaptability and restrains separate generations from genetic exchange. Situations like these are commonly met in forest tree breeding when establishing seed orchards.

The present paper treats especially the consequences implied for their genetic structures by characteristic components of the mating system found in seed orchards. The emphasis will be put on the role played by the rate of self-fertilization, mode and range of pollen dispersal, population size and number of clones. This set of parameters determines to a great extent the amount of inbreeding and kinship and deviations from random-mating.

Of course there are many other factors which influence the breeding system displayed in seed orchards, such as variation in flowering times, seed and pollen production. But too little is known about the genetic control of components like these, so that reasonable assumptions concerning implications for the genetic structure can not be made.

An investigation of genetic structures in seed orchards can be performed from two different points of view:

A) Representations refer exclusively to the expected genetic structure within the seed production of a given seed orchard. The aim is to state how far the parental structure is reflected by the seed population.

B) Assuming that populations derived from the seed production of a seed orchard are used to establish future generations, considerations concerning the risk caused by reduction of genetic variability obtain increasing importance. To demonstrate this expected risk, investigations are extended to an arbitrary number of generations.

1. Theoretical derivations

1.1. Preliminary remark

As already mentioned, an investigation of genotype-frequencies (genetic structure) and gene-frequencies (genetic composition) in finite populations can be based on two different prerequisites: firstly, one may be interested in finding the expected genetic structure of the progeny of a parental population, whose members are explicitly described by their genotypes, thus considering merely the realized genetic structures of the parental generation and not their probabilities of occurrence; secondly, no exact knowledge about the actual genetic structure of the parental population is available, and therefore some assumptions concerning the probability distribution of this structure - which then has to be regarded as a random variable - have to be made.

Obviously, the expected genetic structure of the progeny in the latter case is obtained from the first case by taking the expectation with respect to the probability-distribution of the parental structure.

A clear distinction between these two situations is necessary when applying theoretically obtained results to actually existing populations, especially seed-plant populations, which are able to produce a comparatively large number of offspring in the form of seeds. Such a seed population may be accepted as representing sufficiently well the expected genetic structure of the progeny of a specified set of parents. It will be very interesting to find the degree of concordance between the genetic structure and the composition of the parental and offspring generations.

A comparison of this kind can not be performed if the information about the genetic structure of the parental population is not more explicitly given, but rather refers to the expected structure. In this case all the genotype- and gene-frequencies of the two generations can not be regarded as actually existing genetic structures in the sense mentioned above; they merely denote the probability of a single genotype or gene occurring in the population. The same concept underlies the definition of the coefficient of inbreeding and kinship, both of which therefore have to be treated within the scope of this concept. One of the aims of this paper shall be to demonstrate the relationship between the coefficients of inbreeding and kinship and the genotypic probability structure for the model applied.

1.2. The model

Populations are assumed to have finite size N and consist of monoecious, diploid seed-plants. The individuals are distributed over their habitat according to a specified pattern, so that each plant can be identified by its location. The following conditions shall be realized approximately: no immigration, mutation, gametic and zygotic selection; no genetically caused variation of the amounts of pollen- and seed-production, of the types of pollen-dispersal and of the rates of self-fertilization among the plants. The overall pollen production is sufficiently large to pollinate all eggs present in the population. All plants flower at the same time.

Where several generations are considered, these are assumed to be discrete and non-overlapping, and each offspring generation is obtained by taking a random sample from the overall seed-production of the parental generation; the assignment of the individuals of this sample to their locations is at random.

Because we intend to describe the influence of the type of pollen-dispersal and the rate of self-fertilization on the genetic structure, respectively on the average coefficient of inbreeding and kinship of a population, all the following derivations assume that the probabilities of mating for all pairs of individuals are given as functions of their locations, and that the seed-production of each (mother-) plant can be split into one part resulting from self-fertilization and the remaining part resulting from cross-fertilization.

One autosomal gene locus with an arbitrary number of alleles A_i is considered.

Case A) Genetic structure of the parental population explicitly given. This situation may be described by identifying each individual by the place, say k, it is located at and the genotype A_iA_j it has. Naturally k = 1,2 ...,N. To allow exact derivations of the results we shall apply Kronecker's symbol:

$$\delta_{ij} := \begin{cases} 0 \text{ if } i \neq j \\ 1 \text{ if } i = j \end{cases} \quad \text{and}$$

$$G_{i|k} := \begin{cases} 1 \text{ if the genotype of an individual located at place } k \text{ is } A_iA_i \\ \frac{1}{2} \text{ if the genotype of an individual located at place } k \text{ is } A_iA_j\,, \\ 0 \text{ if the genotype of an individual located at place } k \text{ is } A_jA_l\,,\ j \neq i \neq l \end{cases}$$

Obviously $\sum_i G_{i|k} = 1$ for all k = 1,...,N.

Further notations:

$p(i|k)$ =: probability that a seed randomly drawn from the cross-fertilized part of the seed-production of a plant located at place k contains a pollen which has been produced by a plant located at place i.

Obviously $p(k|k) = 0$ and $\sum_{i=1}^{N} p(i|k) = 1$ for all k = 1,...,N.

q_k := rate of self-fertilization within the seed-production of a plant located at place k.

s_k := probability that a seed randomly drawn from the overall seed-production of the population originates from a plant located at place k. $\sum_{k=1}^{N} s_k = 1$.

P''_i := frequency of the allel A_i in the parental population.

P''_{ij} := frequency of the genotype A_iA_j in the parental population.

P'_{ij} := probability that a seed randomly drawn from the overall seed-production of the population has genotype A_iA_j.

p'_i := $P'_{ii} + \frac{1}{2} \cdot \sum_{\substack{j \\ j \neq i}} P'_{ij}$.

With these notations the realized genetic structure in relation to composition of the parental population is given by the set of the P''_{ij}'s and the p''_i's, respectively, while the expected genetic structure and composition of the seed-population is given by the set of probabilities, P'_{ij} and p'_i, respectively.

Clearly

$$p''_i = \frac{1}{N} \cdot \sum_{k=1}^{N} G_{i|k}$$

$$P''_{ij} = \frac{4}{N} \cdot \sum_{k=1}^{N} G_{i|k} \cdot G_{j|k} \quad \text{for } i \neq j,$$

and

$$P''_{ii} = \frac{2}{N} \cdot \sum_{k=1}^{N} G_{i|k}\left(G_{i|k} - \frac{1}{2}\right)$$

Considering that $G_{i|k}$ may be regarded as the probability that an allele randomly drawn from the genotype of an individual located at place k is A_i, the probabilities P'_{ij} can be computed in a straightforward way:

$$P'_{ij} = (2-\delta_{ij}) \cdot \left[\sum_{k=1}^{N} s_k q_k \cdot G_{i|k} \cdot G_{j|k} + \right.$$
$$+ \frac{1}{2} \cdot \sum_{k=1}^{N} s_k(1-q_k) \cdot G_{i|k} \cdot \sum_{l=1}^{N} p(l|k) \cdot G_{j|l} +$$
$$\left. + \frac{1}{2} \cdot \sum_{k=1}^{N} s_k(1-q_k) \cdot G_{j|k} \cdot \sum_{l=1}^{N} p(l|k) \cdot G_{i|l} \right], \tag{1a}$$

and from this

$$p'_i = \frac{1}{2} \cdot \sum_{k=1}^{N} s_k(1+q_k) \cdot G_{i|k} +$$
$$+ \frac{1}{2} \cdot \sum_{k=1}^{N} s_k(1-q_k) \cdot \sum_{l=1}^{N} (l|k) \cdot G_{i|l} \tag{1b}$$

This general representation is suitable for computing a great variety of different actual situations, but it is too difficult to survey if one intends to make apparent specific properties of the genetic structure relate to composition, which are due to certain influences. This is because, in the general case, the representation depends heavily on the locations of the genotypes, a complication which can be cancelled by assuming $s_k = \frac{1}{N}$,

$q_k = q$ for all k and crossfertilization to occur at random, i.e. $p(l|k) = \frac{1}{N-1}$ for all $l \neq k$. From that

$$P'_{ii} = {p''_i}^2 (1-q) \cdot \frac{N}{N-1} + \frac{1}{2}(P''_{ii} + p''_i)\left(1 - (1-q) \cdot \frac{N}{N-1}\right) \tag{2a}$$

$$P'_{ij} = 2p''_i p''_j (1-q) \cdot \frac{N}{N-1} + \frac{1}{2}P''_{ij}\left(1 - (1-q) \cdot \frac{N}{N-1}\right) \text{ for } i \neq j \tag{2b}$$

and $p_i' = p''_i$ as can be proved easily, i.e. the genetic composition does not change.

Here, especially, the question arises as to which conditions, in addition to random cross-fertilization, have to be fulfilled so that the expected genetic structure of the seed population is equal to the corresponding Hardy-Weinberg-proportions. A solution to this problem can be given by applying the genetic distance d between the two structures (in a version stated by Gregorius 1974):

$$d = \frac{1}{2} \cdot \left[\sum_i |P'_{ii} - {p'_i}^2| + \sum_{\substack{i,j \\ i<j}} |P'_{ij} - 2p'_i p'_j|\right].$$

If we consider that always $\frac{1}{2}(P''_{ii} + p''_i) - {p''_i}^2 \geqslant 0$, d can be written as

$$d = \frac{1}{2}\left|1 - (1-q) \cdot \frac{N}{N-1}\right| \cdot \tag{3}$$

$$\cdot \left[\sum_i \left(\frac{1}{2}(P''_{ii} + p''_i) - {p''_i}^2\right) + \sum_{\substack{i,j \\ i<j}} \left|2p''_i p''_j - \frac{1}{2}P''_{ij}\right|\right].$$

Thus $d = 0$ if the seed population attains Hardy-Weinberg-proportions, which is equivalent to $1 - (1-q)\frac{N}{N-1} = 0$ or

$\frac{1}{2}(P''_{ii} + p''_i) - {p''_i}^2 = 0$ for all i and $2p''_i p''_j - \frac{1}{2}P_{ij} = 0$

for all $i \neq j$. Thus $d = 0$ if and only if $q = \frac{1}{N}$ (complete panmixia) or all members of the parental population have the same genotype.

Therefore absence of self-fertilization ($q=0$) can not be treated as equivalent to dioecism - which would lead to Hardy-Weinberg-proportions - as is commonly done in the technical literature.

Furthermore for large population-size N, i.e. $\frac{N}{N-1} \approx 1$, the above equations reduce to a well known result for partial self-fertilization (see e.g. Crow and Kimura, 1970, p.93).

Case B) The expected genetic structures. As mentioned before, this problem can be solved by taking the expectation of equation (1a) with respect to the probability-distribution of the genetic structure of the parental population, in which case the quantities $G_{i|k}$ have to be regarded as random variables. But because this would require a lot of preliminary explanations which rather complicate the derivations, we decided to choose another approach, which of course leads to the same results. Considerations will now be extended over several generations according to the assumptions at the beginning. The following notations will be applied:

$\overline{q}_t$ =: average rate of self-fertilization within the seed-production of generation t.

M_t =: probability that two alleles randomly taken from two distinct individuals in generation t are derived from the same individual in generation t-1.

P^t_{ij} =: probability that an individual randomly taken from generation t has genotype A_iA_j.

R^t_{ij} =: probability that two alleles randomly taken from two distinct individuals in generation t are A_i and A_j.

f_t =: average coefficient of inbreeding in generation t.

φ_t =: average coefficient of kinship in generation t.

p^t_i =: $P^t_{ii} + \frac{1}{2} \cdot \sum_{\substack{j \\ j \neq i}} P^t_{ij}$; r^t_i =: $R^t_{ii} + \frac{1}{2} \cdot \sum_{\substack{j \\ j \neq i}} R^t_{ij}$.

With the help of the previously defined quantities, we obtain

$$\overline{q} = \sum_{k=1}^{N} s_k q_k \text{ and } M = \frac{1}{4} \cdot \sum_{k=1}^{N} \left[(1+q_k)s_k + \sum_{k=1}^{N} s_i \cdot (1-q_i) \cdot p(k|i)\right]^2,$$

where M is the discrete version of the continuous representation given by Gregorius (1975).

The recurrence relations for the quantities R^t_{ij} can be obtained immediately if we take into consideration that two alleles from two distinct individuals in generation t, which are derived from the same individual in generation t-1, have probability $\frac{1}{2}\left(P^{t-1}_{ii} + p^{t-1}_i\right)$ to be both A_i and the probability $\frac{1}{2}P^{t-1}_{ij}$ to be A_i and A_j:

$$R_{ii}^{t} = \frac{1}{2}M_t\left(P_{ii}^{t-1} + p_i^{t-1}\right) + (1-M_t)R_{ii}^{t-1} \tag{4a}$$

$$R_{ij}^{t} = \frac{1}{2}M_t P_{ij}^{t-1} + (1-M_t)R_{ij}^{t-1} \quad \text{for } i \neq j \tag{4b}$$

The relationship between the expected genetic structures of two successive generations t-1 and t can be derived by distinguishing between three different possibilities for the formation of an offspring's genotype: firstly, it may be produced by self-fertilization and thus has probability $\frac{1}{2}\left(P_{ii}^{t-1} + p_i^{t-1}\right)$ to be A_iA_i and probability $\frac{1}{2}P_{ij}^{t-1}$ to be A_iA_j; secondly, it may be produced by cross-fertilization and the two parental alleles are derived from the same grandparent, in which case they have probability $\frac{1}{2}\left(P_{ii}^{t-2} + p_i^{t-2}\right)$ to be both A_i and probability $\frac{1}{2}P_{ij}^{t-2}$ to be A_i and A_j; thirdly, it may be produced by cross-fertilization and the two parental alleles are derived from two distinct grandparents, in which case they have probability R_{ij}^{t-2} to be A_i and A_j ($i \neq j$ and $i=j$). In all

$$P_{ii}^{t} = \frac{1}{2}\bar{q}_{t-1}\left(P_{ii}^{t-1} + p_i^{t-1}\right) + \frac{1}{2}(1-\bar{q}_{t-1}) \times \\ \times M_{t-1}\left(P_{ii}^{t-2} + p_i^{t-2}\right) + (1-\bar{q}_{t-1})(1-M_{t-1})R_{ii}^{t-2} \tag{5a}$$

$$P_{ij}^{t} = \frac{1}{2}\bar{q}_{t-1} \cdot P_{ij}^{t-1} + \frac{1}{2}(1-\bar{q}_{t-1}) \times \\ \times M_{t-1} \cdot P_{ij}^{t-2} + (1-\bar{q}_{t-1})(1-M_{t-1})R_{ij}^{t-2} \quad \text{for } i \neq j, \tag{5b}$$

and from this

$$r_i^{t} = M_t \cdot p_i^{t-1} + (1-M_t)r_i^{t-1} \tag{6}$$

$$p_i^{t} = p_i^{t-1} \cdot \bar{q}_{t-1} + (1-\bar{q}_{t-1}) \times \\ \times M_{t-1} \cdot p_i^{t-2} + (1-\bar{q}_{t-1})(1-M_{t-1})r_i^{t-2} \tag{7}$$

Because these results hold for all t, it is possible to insert equation (4a) into (5a), (4b) into (5b) and (6) into (7). From this we obtain

$$P_{ii}^{t} = \frac{1}{2}\bar{q}_{t-1}\left(P_{ii}^{t-1} + p_i^{t-1}\right) + (1-\bar{q}_{t-1})R_{ii}^{t-1} \tag{8a}$$

$$P_{ij}^{t} = \frac{1}{2}\bar{q}_{t-1} \cdot P_{ii}^{t-1} + (1-q_{t-1})R_{ij}^{t-1} \tag{8b}$$

$$p_i^{t} = \bar{q}_{t-1} \cdot p_i^{t-1} + (1-\bar{q}_{t-1})r_i^{t-1} \tag{9}$$

In most cases for the initial generation (t=0) the condition $p_i^{o} = r_i^{o} =: p_i$ is fulfilled, which in turn implies

$$p_i^{t} = r_i^{t} = p_i \quad \text{for all } t,$$

i.e. the expected genetic composition does not change in the course of the generations. This condition is realized if, for example, the initial population is of hypothetically infinite size. As can be taken from equations (4) and (8), $M_t = \bar{q}_{t-1}$ leads to $P_{ij}^{t} = R_{ij}^{t}$ for all i, j, which, recalling the definition of R_{ij}^{t}, is equivalent to the corresponding results obtained for complete panmixia (see e.g. Li, 1963, p.209). On the other hand $\bar{q}_{t-1} = 0$ implies $P_{ij}^{t} = R_{ij}^{t-1}$ for all i, j which reflects the well-known findings for random-mating monoecious populations in the absence of self-fertilization and for random-mating dioecious populations (see e.g. Jacquard, 1974, p. 180). These statements are by no means trivial since they also cover cases where random union of gametes, because of limited pollen dispersal, can not be actually verified.

The relationship between the results stated in equations (4) and (8) and the average coefficients of inbreeding and kinship can be established as follows:

For the present situation we have the recurrence relations (Gregorius 1975):

$$f_t = \frac{1}{2}\bar{q}_{t-1}(1+f_{t-1}) + (1-\bar{q}_{t-1})\varphi_{t-1} \quad \text{and} \tag{10a}$$

$$\varphi_t = \frac{1}{2}M_t(1+f_{t-1}) + (1-M_t)\varphi_{t-1} \tag{10b}$$

If for the initial generation (t=0) we assume $f_o = \varphi_o = 0$ and $p_i^{o} = r_i^{o} = p_i$, such that $p_i^{t} = p_i$ for all t, it can be shown easily that the following statement satisfies equations (4a), (4b), (8a) and (8b):

$$P_{ii}^{t} = p_i \cdot f_t + p_i^2(1-f_t) + a_t(P_{ii} - p_i^2) \tag{11a}$$

$$P_{ij}^{t} = 2p_ip_j(1-f_t) + a_t(P_{ij} - 2p_ip_j) \quad \text{for } i \neq j \tag{11b}$$

$$R_{ii}^{t} = p_i \cdot \varphi_t + p_i^2(1-\varphi_t) + b_t(P_{ii} - p_i^2) \tag{12a}$$

$$R_{ij}^{t} = 2p_ip_j(1-\varphi_t) + b_t(P_{ij} - 2p_ip_j) \quad \text{for } i \neq j \tag{12b}$$

where the P_{ij}'s denote the genetic structure of the initial population, and the quantities a_t, b_t are determined by the recurrence relations

$$a_t = \frac{1}{2}\bar{q}_{t-1} \cdot a_{t-1} + (1-\bar{q}_{t-1})b_{t-1} \tag{13a}$$

$$b_t = \frac{1}{2}M_t \cdot a_{t-1} + (1-M_t)b_{t-1} \tag{13b}$$

with the initial conditions $a_o = 1$ and $b_o = 0$.

Where the genetic structure of the initial population agrees with the Hardy-Weinberg-proportions, equations (11) reduce to Wright's (1922) well-known formulae, while, disregarding the initial structure, for $\overline{q}_t = 0$ for all t and thus $a_t = b_{t-1}$ equations (11) and (12) coincide with the results given by Jacquard (1974, p.182). Furthermore, the above representations demonstrate clearly which aspects of the expected genetic structure (concerning the P's and R's) are directly influenced by the coefficient of inbreeding or by the coefficient of kinship.

Because deviations of the genetic structure of the initial population from Hardy-Weinberg-proportions are the rule rather than the exception, equations (11) tell us that it is necessary to specify f_t as well as a_t if we are aiming at a representation of the expected genetic structure in generation t which does not explicitly involve the initial structure.

From these findings it becomes evident once more that the interrelation between the coefficient of inbreeding and the expected genetic structure depends on the system of mating. This aspect gains particular significance when referred to what we called the initial population, because its members could already have been related and inbred to a certain degree, which thus would not agree with our initial condition $f_o = \varphi_o = 0$. But this seeming contradiction can be solved if we proceed from the concept that we exclusively intend to determine this part of the amount of genetic relationship and inbreeding which is generated by the system of mating employed, and that we furthermore represent all the preceding influences of inbreeding by the genetic structure of the initial population. This concept is indispensable if we consider - as pointed out above - that the system of inbreeding practised to obtain the initial population might determine its genetic structure in an entirely different way than does the system starting from this population.

For completeness it should be emphasized that the concept of effective population size - in this case the inbreeding effective population size (for definition see e.g. Kimura and Ohta, 1971) - generally can not be applied to our results, since for monoecious individuals it refers exclusively to the special cases $\overline{q}_t = 0$ and $M_t = \overline{q}_{t-1}$ for all t. On the other hand, if we consider that in the ideal case, in which all individuals contribute equally to the offspring generation, i.e.

$$s_k = \frac{1}{N},\ q_k = q,\ \sum_{i=1}^{N} p(k|i) = 1 \text{ for all } k = 1,\dots,N,$$

we obtain $M = \frac{1}{N}$, it is reasonable to extend the conventional concept by defining the effective population size N_e^t in generation t as $N_e^t = \frac{1}{M_{t+1}}$.

Finally, we shall briefly treat a generalization of the previous findings which, for example applies to certain forestry tree breeding programs connected with the establishment of seed orchards. It is assumed that each plant is propagated vegetatively by a certain number (called a clone) which does not depend on its genotype. Within the scope of our model, this means that each individual can be identically reproduced several times, and that, because of this, all those quantities which referred to the probability that two alleles are derived from the same individual now have to be conceived as probabilities that two alleles are derived from the same clone. If C_t denotes the probability that two distinct individuals in generation t belong to the same clone, then within all the preceding formulae

$\overline{q}_t$ has to be substituted by $\overline{q}_t' = \overline{q}_t + (1-\overline{q}_t)C_t$ and
M_t has to be substituted by $M_t' = M_t + (1-M_t)C_{t-1}$.

In particular, if in generation t all L_t clones have equal size K_t, and therefore the population size is equal to $N_t = K_t \cdot L_t$, we obtain

$$C_t = \frac{K_t - 1}{N_t - 1} .$$

The variance of the allele-frequency. Probability theory offers several possibilities to describe the implications of random sampling. Probably the most commonly used concept for recording these effects is that of the variance. To complete the preceding considerations of the expected genetic structure by taking into account the accidents of sampling, the variance of the allele-frequency for each allele A_i is computed.

With reference to the notation in case A), we now define $p_i'^{t}$, $p_i''^{t}$ and $P_{ij}'^{t}$ as p_i', p_i'' and P_{ij}' in generation t. Using these quantities, which have to be regarded as random variables, we are able to write the variance of the frequency of allele A_i in generation t as

$$V\left(p_i''^{t}\right) = E\left(\left(p_i''^{t}\right)^2\right) - E\left(p_i''^{t}\right)^2 =$$
$$E\left(E\left(\left(p_i''^{t}\right)^2 \middle| p_i'^{t-1}, P_{ii}'^{t-1}\right)\right) - p_i^2 ,$$

applying the standard notation for the expectation and the conditional expectation of random variables.

From a result obtained by one of the authors (Gregorius 1975a, equ.(2)), it can be derived, that

$$E\left(\left(p_i''^{\,t}\right)^2 \middle| p_i'^{\,t-1}, P_{ij}'^{\,t-1}\right) =$$

$$\frac{1}{2N_t}\left(p_i'^{\,t-1} + P_{ii}'^{\,t-1}\right) + \left(1 - \frac{1}{Nt}\right)\left(p_i'^{\,t-1}\right)^2 .$$

Furthermore, from the statements of case B) we conclude

$$E\left(p_i'^{\,t-1}\right) = p_i ,\ E\left(P_{ii}'^{\,t-1}\right) = P_{ii}^t \text{ and } E\left(\left(p_i'^{\,t-1}\right)^2\right) = R_{ii}^t .$$

Thus

$$V\left(p_i''^{\,t}\right) = \frac{1}{2N_t}\left(p_i + P_{ii}^t\right) + \left(1 - \frac{1}{N_t}\right) R_{ii}^t - p_i^2 \qquad (14)$$

which relates the variance within a given generation to its expected genetic structure.

Because the expected genetic structure in generation t does not depend on the population size N_t of this generation, the above equation tells us that for $N_t \to \infty$ we obtain $p_i''^{\,t} \to p_i'^{\,t}$ with probability one and thus $V\left(p_i''^{\,t}\right) \to V\left(p_i'^{\,t-1}\right) = R_{ii}^t - p_i^2$, which is the variance of the expected frequency of allele A_i within the progeny derived from generation t-1 (compare case A). It is now possible to restate equation (14) to depict this part of the total variance which is due to random sampling of individuals within one generation:

$$V(p_i''^{\,t}) = \frac{1}{Nt} \cdot \left[\frac{1}{2}\left(p_i + P_{ii}^t\right) - R_{ii}^t\right] + V\left(p_i'^{\,t-1}\right) \qquad (14a)$$

It must be emphasized that N_t refers to the number of individuals resulting from random sampling, and so does the allele-frequency $p_i''^{\,t}$. Therefore, if only L_t out of the total of N_t individuals are selected at random, while the remainder is obtained from these L_t plants by vegetative propagation performed with no regard to genotypes, then N_t in equations (14) and (14a) has to be replaced by L_t, and $p_i''^{\,t}$ describes the corresponding allele-frequency for those L_t plants. On the other hand, $p_i'^{\,t-1}$ and consequently $V\left(p_i'^{\,t-1}\right)$ are not affected by these considerations, since all the necessary information is already contained in R_{ii}^t.

2. Applications

2.1. The genetic structure of the progeny derived from an explicitly described parental population (see case A)

As already mentioned in chapter 1, the formulae were developed for the special cases where exact knowledge about the actual parental genetic structure is available and a direct determination of possible deviations between their structure and that of the progeny is needed. This may be important, for example, if among the clones in a seed orchard a few outstanding genotypes are identified, which should be obtained with the same expected frequency within the seed population. For such an explicit case the result can be obtained quite easily by using equations (2).

Accepting random cross-fertilization and the special assumptions $s_k = \frac{1}{N}$ and $q_k = q$, that is, each tree participates in seed production to the same extent and is characterized by the same average amount of self-fertilization, the genetic structure of the progeny depends only on the parental allelic and genotypic frequency (p_i'', P_{ii}'' resp. P_{ij}''), the rate of self-fertilization (q) and the number of clones (L, where L = N, if each clone is represented by one individual). Of the variety of actual situations, only a few examples within the two-allele-concept will be presented: for instance the influence of q on the P_{ii}' and P_{ij}'. As a measurement of differences between the obtained genetic structures and the corresponding Hardy-Weinberg-proportions for the parental and the seed population, the respective genetic distances d'' and d' were calculated according to equation (3).

Table 1. The genetic structure of the parental (P_{ij}'') and resulting seed (P_{ij}') population and the respective genetic distance (d'', d') to the corresponding Hardy-Weinberg-proportions for different rates of self-fertilization (q) and population-size N = 20

q	L = N	p_1''	p_2''	P_{11}''	P_{12}''	P_{22}''	d''	P_{11}'	P_{12}'	P_{22}'	d'
.0	20	.50	.50	.25	.50	.25	.0000	.243	.513	.243	.0135
1/N								.250	.500	.250	.0000
.3								.283	.434	.283	.0660
.8								.338	.324	.338	.1760
.0	20	.15	.85	.10	.10	.80	.1550	.017	.266	.717	.0110
1/N								.023	.255	.723	.0000
.3								.050	.201	.750	.0545
.8								.095	.110	.795	.1618

In the first case we assumed the parental population to be in Hardy-Weinberg-proportions, in the second case we considered a deviation from this situation. The data confirm that in both cases the corresponding Hardy-Weinberg-proportions are attained in the progeny only in the case of complete panmixa ($q = 1/N$). Any change of q from $1/N$ towards 0, as well as towards 1, causes increasing genetic distances d'. Comparing the values of d' and d'', it is obvious that a deviation of the parental population from corresponding Hardy-Weinberg-proportions d'' can, with an increasing rate of self-fertilization, be exceeded by the respective deviation of the seed population d'.

An increase in the number of trees per clone in the parental population will not change the genetic structure of the progeny.

2.2. The expected genetic structures (see case B)

If no exact determination of the parental genetic structure is available, the probability distribution of this structure has to be considered and calculations will be performed for several successive generations as described in case B.

This will be done in 4 steps:

a) description of the influence of the pollen distribution which is contained in the parameter M,

b) calculation of the coefficients of inbreeding and the parameters a_t,

c) determination of some expected genetic structures corresponding to b),

d) description of the variance of the allele-frequency.

a) Description and determination of M. The parameter M (as defined in chapter 1.2) determines the values of the coefficients of inbreeding and kinship as well as those of the genetic structure (see (10), (13), (4) and (8)). M is important because it is the only parameter which reflects the influence of the pollen dispersal, the individual rates of self-fertilization and the actual population-size on the expected genetic structures. As already demonstrated, for a clear interpretation of M, it is useful to conceive it as the reciprocal value of the effective population size N_e, i.e. $N_e = \frac{1}{M}$. Therefore, instead of the M's, the N_e's are listed in the table below.

To direct attention to the role played by pollen dispersal, it is convenient to assume no variation in the rates of self-fertilization (i.e. $q_k = q$ for all k) or in the individual seed and pollen production (i.e. $s_k = \frac{1}{N}$ for all k, where N = population size). As no exact data for pollen dispersal within seed orchards are available, we accept an exponential distribution, in accordance with published results which refer to pollen dispersal of individual trees in stands (Bateman 1947, Schmidt 1970, Müller 1974). Proceeding from an exponential pollen distribution, the mating probabilities $p(k|i)$ within M are obtained with the help of Bayes's theorem, which here attains the representation

$$p(k|i) = \frac{\exp\left(-\frac{5}{D} \cdot |k-i|\right)}{\sum\limits_{\substack{j \\ j\neq i}} \exp\left(-\frac{5}{D} \cdot |i-j|\right)}$$

where $|k-i|$ is the distance between places k and i, and D describes the range of pollen distribution for individual trees. This range is defined as the distance within which 96% of the emitted pollen sediments. The individuals of the seed orchard are supposed to be arranged according to a simple square grid, the dimensions of which will be varied. N_e is calculated for $D = 20, 50, 100$ meters, $N = 400$ (20×20, 40×10), 100 (10×10, 20×5), 25 (5×5) trees, $q = 0$ and $q = 0.3$. The grid-distance is considered to be constant for all cases, namely 7 metres.

The information contained in the following table is easier to survey if it is kept in mind that the effective population size N_e is equal to the actual population size N, if all parents contribute equally to the progeny as mentioned in chapter 1.2. This situation may be realized in different ways, one of which is that of random-cross-fertilization.

Table 2 demonstrates that the deviations between N and N_e are surprisingly small in all cases, so that the range of pollen dispersal is less effective than might have been expected.

The percentage deviations between N and N_e decrease with increasing rates of self-fertilization, a fact which confirms that a reduction in cross-fertilization suppresses the influence of pollen dispersal.

Considering that we assumed no variation in the rates of self-fertilization, pollen- and seed-production and pollen-dispersal of the individual trees, the deviations obtained can be explained exclusively as marginal effects. As seen from the table, it is not possible to interpret the effect of D, N and type of arrangement on the deviations independently of one another. An illustration of the kind of dependence between these variables would require more sophisticated investigations, which are beyond the scope of this paper.

Table 2. Values of the effective population-size N_e and percentage deviations of N_e from the population-size N for different N's, ranges of pollen-dispersal (D) and rates of self-fertilization (q)

		N = 400				N = 100				N = 25	
		20 × 20	%	40 × 10	%	10 × 10	%	20 × 5	%	5 × 5	%
D = 20	q = 0	399.20	.20	398.88	.28	99.54	.46	99.39	.61	24.72	1.12
	q = 0.3	399.52	.12	399.52	.12	99.77	.23	99.70	.30	24.86	.56
D = 50	q = 0	398.88	.28	398.41	.40	99.35	.65	99.48	.52	24.79	.84
	q = 0.3	399.36	.16	399.20	.20	99.68	.32	99.75	.25	24.89	.44
D = 100	q = 0	397.93	.52	398.25	.44	99.34	.66	99.63	.37	24.91	.36
	q = 0.3	399.04	.24	399.20	.20	99.68	.32	99.82	.18	24.96	.16

As a consequence of the extremely small deviations obtained between N_e and N, the subsequent calculations of f_t, a_t and P^t_{ii} will be based on the assumption of random cross-fertilization, that is $N_e \approx N$ or equivalently $M = \frac{1}{N}$.

b) Caculation of the coefficients of inbreeding f_t and the parameters a_t. To calculate the coefficients of inbreeding f_t for generation t with the help of the derived recurrence relation (10), it is necessary to specify, besides the parameters M and q, the coefficients of kinship φ_t in generation t. On the other hand, having in mind a description of the expected genetic structure, which can be represented as a function of the coefficient of inbreeding and the parameter a_t only (see (11)), we confine our calculations to these quantities.

Calculations of the f_t's will be performed for the same cases as before and in addition the number of clones L will be reduced by a factor $\frac{N}{L} = 5$. The latter refers to the fact that in seed orchard lay-outs each clone is represented by more than one individual. The reduction of the number of clones induces an apparent q and M, denoted as q' and M', respectively, which both exceed the corresponding actual values, according to the formulae given at the end of chapter 1.2, case B). Each figure is complemented by a table of corresponding parameters a_t, which measure the reduction of the contribution from the deviation within the initial population between the actual and the corresponding Hardy-Weinberg-structure for each generation. Both parameters f_t and a_t are independent of any assumption concerning the genetic structure of the initial population, so that they reflect purely the influence of the mating system. Although a discontinuous representation would be correct, for clarity we decided in favour of a continuous one.

The fact that the effective population size N_e comprises the influences of N as well as L allows simplification of the interpretation of the above figures as follows. For any given rate of self-fertilization (actual or apparent), decreasing values of N_e cause a considerable increase in the increments per generation. To prevent possible misunderstandings it should be pointed out that the values of q given in the figures in all cases are the actual rates of self-fertilization.

For any given effective population size (actual or apparent), changing values for q cause marked differences in increments only for the first generation, while for the following generations the increments per generation become very similar.

The special influence of the number of clones on the coefficients of inbreeding can be taken from the figures, when for any given N and actual q the two different values of L are considered. In all cases the f_t for this pair of L's are diverging for the first ten generations, a tendency which is pronounced with decreasing N.

Because of the initial conditions $f_0 = \varphi_0 = 0$ and $a_0 = 1$, $b_0 = 0$, the values for a_1 always coincide with those for f_1, namely $a_1 = f_1 = \frac{1}{2}q$ and $a_1 = f_1 = \frac{1}{2}q'$, as taken from (10) and (13). The respective tables show that considerable influences of the values of a_t on the expected genetic structure are obtained for the first generation only. For all successive generations these influences are negligible.

c) Calculations of the expected genetic structures. Because of the simple functional relationship between the

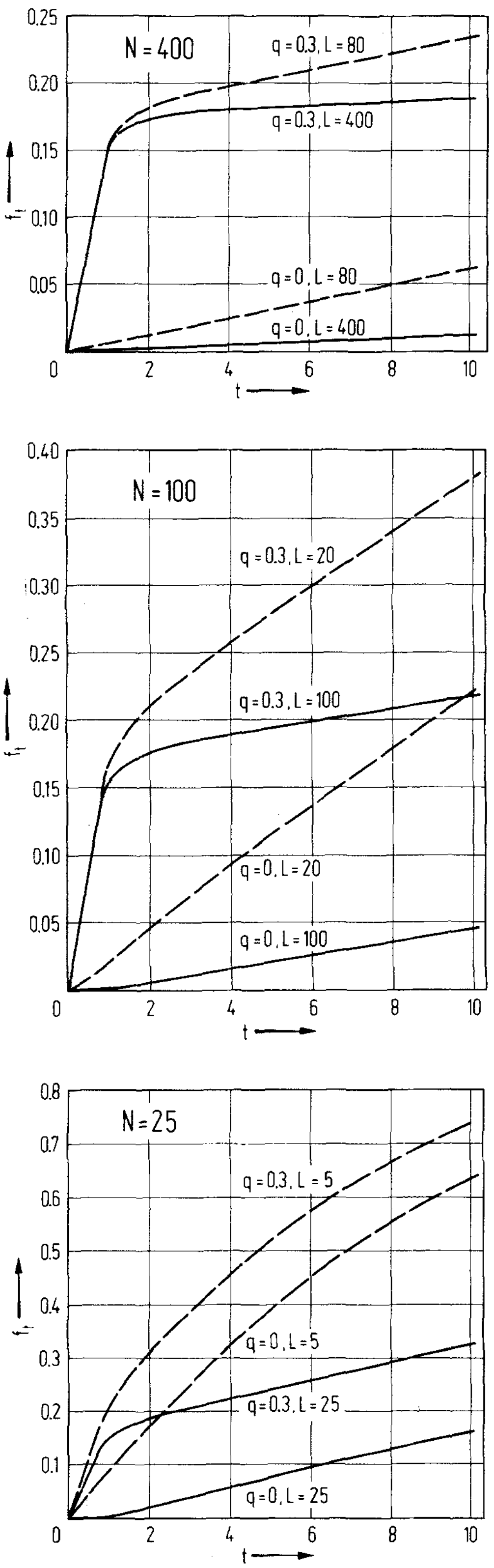

Fig. 1-3. The coefficients of inbreeding f_t in generation t as a function of t. q = rate of self-fertilization; N = population size, L = number of clones; N as well as L are assumed to be the same for all generations. All clones are considered to have the same size K, such that N = K • L

Table 3.

	N = 400			
	L = 400		L = 80	
t	q = .0	p = .3	q = .0	q = .3
1	.00000	.15000	.00501	.15351
2	.00125	.02337	.00621	.02790
3	.00125	.00451	.00617	.00922
10	.00123	.00120	.00591	.00568

Table 4.

	N = 100			
	L = 100		L = 20	
t	q = .0	q = .3	q = .0	q = .3
1	.00000	.15000	.02020	.16414
2	.00500	.02600	.02440	.04374
3	.00495	.00789	.02377	.02589
10	.00478	.00461	.01993	.01871

Table 5.

	N = 25			
	L = 25		L = 5	
t	q = .0	q = .3	q = .0	q = .3
1	.00000	.15000	.08333	.20833
2	.02000	.03650	.09028	.10174
3	.01920	.02102	.08113	.08001
10	.01672	.01589	.03936	.03362

Table 3-5. Values of a_t (for definition see text) for a given population size (N), different number of clones (L), rates of self-fertilization (q) and generations (t)

coefficient of inbreeding and the expected genetic structure (see (11)), we restrict representations to one of the examples treated in the preceding chapter (figure 3). As a representation of the single genotypic frequencies $\left(P^t_{ii}, P^t_{ij}\right)$ does not provide more effective information than is contained in the total amount of the homozygotes, we decided to use the frequency of the homozygotes $H_t = \sum_i P^t_{ii}$ only. Furthermore, the fact that deviations from Hardy-Weinberg-proportions within the initial population can easily be taken into account with the help of the quantities a_t suggests that the investigations can be based on a Hardy-Weinberg-structure, which in our case shall be realized by the presence of two alleles with equal frequencies.

Figure 4 proves that the tendencies of the coefficient of inbreeding for the corresponding case presented in

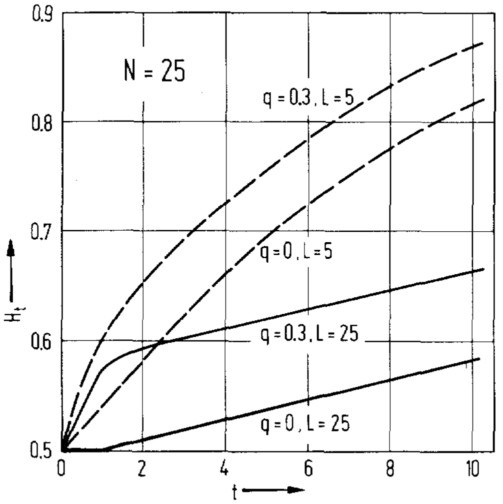

Fig.4. The proportion of homozygotes H_t in generation t as a function of t, using for the initial population a Hardy-Weinberg-structure realized for two alleles A_1, A_2, with equal frequencies. For definition of q, N, L see figures 1,2,3

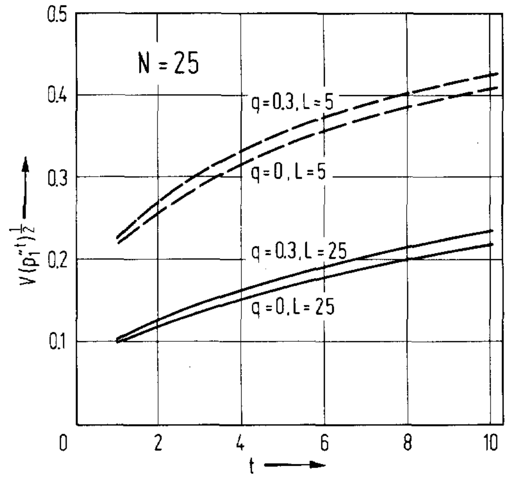

Fig.5. The standard deviation $V(p_1''^t)^{\frac{1}{2}}$ of the allele-frequency $p_1''^t$ in generation t as a function of t. Remaining notations as well as assumptions about the genetic structure of the initial population coincide with figure 4

figure 3 are directly transferred to H_t. Of course all the curves have to start at the frequency of the homozygotes assumed for the initial generation.

d) Calculations of the standard-deviation of the allele-frequency. To enable investigations to continue based on the preceding findings, the following calculations will be performed under the same assumptions chosen for obtaining the results stated in figure 4.

The interpretation of the following representation will be facilitated if we recall that the standard-deviation of the allele-frequency is a suitable measure for the risk of losing an allele because of accidents in sampling individuals in the course of generations. Thus the allele-frequency p_1'', and consequently its standard-deviation, has to be referred to the number of clones L.

Because of the evidence of the results, it shall suffice to point out that major changes in the standard-deviation are caused by the number of clones L, where decreasing values of L imply an increase in the standard-deviation.

Literature

Bateman, A.J.: Contamination in seed crops. II. Wind pollination. Heredity 1, 235-246 (1947)

Crow, J.F.; Kimura, M.: An introduction to population genetics theory. New York, London: Harper and Row 1970

Gregorius, H.-R.: Genetischer Abstand zwischen Populationen. Silvae Genetica 23 (1-3), 22-27 (1974)

Gregorius, H.-R.: Die Varianz der relativen Allelhäufigkeit in einer Population konstanter endlicher Größe. Biom. Z. 17, 105-116 (1975 a)

Gregorius, H.-R.: A model for genetic relationship among offspring from open-pollinated plant populations. Theor. Appl. Gen. 46, 109-115 (1975)

Jacquard, A.: The genetic structure of populations. Berlin, Heidelberg, New York: Springer-Verlag 1974

Kimura, M.; Ohta, T.: Theoretical aspects of population genetics. Princeton: Princeton University Press 1971

Li, C.C.: Population genetics. Chicago: The University of Chicago Press 1963

Müller, G.: Some population genetic aspects derived from new results of pollen- and seed dispersal of Spruce and Pine respectively. IUFRO-Proceedings of Working Party S 204 1-3, Stockholm (1974)

Schmidt, H.: Versuche über die Pollenverbreitung in einem Kiefernbestand. Diss. Forstl. Fak. Univ. Göttingen (1970)

Wright, S.: Coefficients of inbreeding and relationship. Amer. Natur. 56, 330-338 (1922)

Received March 27, 1975
Communicated by H. Stubbe

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