The Effect of Positive Assortative Mating at One Locus on a Second Linked Locus

Part 2: Limiting Characteristics of Gametic and Genotypic Structures

H. Gregorius

Lehrstuhl für Forstgenetik und Forstpflanzenzüchtung, Universität Göttingen (Federal Republic of Germany)

Summary. Considerations proceed from a model of positive assortative mating based on genotype at one locus, with an arbitrary number of alleles, assuming no selection, mutation, or migration, hypothetically infinite population size, and discrete non-overlapping generations. From these conditions, inferences are made about the genotypic structure at a linked locus, as well as about the corresponding 2-locus gametic structure.

The following main results are presented: in the course of the generations, the genotypic structure at the second locus and the 2-locus gametic structure always tend to a limit responsive to the initial conditions concerning the joint genotypic structure at the two loci and the degree of assortativity and linkage. A complete, analytical representation of the limits is given. In particular, if assortative mating is only partial and at the same time linkage is not complete, a population is not able to maintain a permanent deviation of the gametic structure from linkage equilibrium, and thus the genotypic structure at the second locus tends to Hardy-Weinberg proportions. On the other hand, if initial linkage disequilibrium is combined with partial assortative mating and complete linkage (or with complete assortative mating and unlinked loci) the population maintains this disequilibrium and thus the genotypic structure at the second locus need not tend to Hardy-Weinberg proportions. It turns out that the conditions not only of complete linkage, but also of unlinked loci together with complete assortativity, imply no change in gametic structure from the initial structure.

In order to demonstrate the influence of several parameters on the speed of convergence to and the magnitude of the respective limits, several graphs are included.

Introduction

The effect of positive assortative mating based on genotypes at a single, multiallelic locus on the genotypic structure (i.e. the set of genotypic frequencies) at a second, linked multiallelic locus has been considered earlier by the present author for one generation in order to arrive at a comprehensive understanding of the forces that transform the genetic structure of the parents into that of the progeny. To avoid unnecessary repetitions, the reader is referred to this paper (Gregorius 1980).

The extension of these considerations to an arbitrary number of generations should, in general, take into account further, evolutionary forces, such as viability selection, mutation, etc. However, introducing these additional parameters into the model makes it difficult, if not impossible, to demonstrate and evaluate the role played by the mating system in the dynamics of a population. Moreover, this role can be most effectively characterized by describing the final state, provided there is any towards which population trajectories tend. Following this generally accepted way of proceeding, the present treatise will concentrate on investigating the asymptotic behavior of the genotypic structure at a locus 'hitch-hiking' with a positive-assoratively-mating locus. Particular attention will be paid to the question as to the existence of final (equilibrium) states that (contrary to common expectation) do not exhibit Hardy-Weinberg proportions at the 'hitch-hiking' locus. Such a case would indicate the possibility that loci not participating in the genetic control of the mating system could form a separate 'apparent' system of mating.

The Model

The underlying assumptions are identical to those applied by Gregorius (1980), and thus shall be repeated only

briefly. Considerations are based on two diploid, autosomal loci, each with an arbitrary number of alleles and recombination frequency 1-c. The alleles at the A-locus are denoted A_i and at the B-locus B_j. The two-locus genotype

$A_i A_k$
$B_j B_l$

is assumed to be the result of the fusion of two gametes of genotypes A_i, B_j and A_k, B_l respectively, and the frequency of this zygotic genotype in the population is denoted P_{jl}^{ik}. It will be of no concern which of the two gametes is the male or the female, consequently $P_{jl}^{ik} = P_{lj}^{ki}$.

Further notations:

P^{ik} frequency of the genotype $A_i A_k$; the set of these frequencies is referred to as the genotypic structure at the A-locus.

P_{jl} frequency of the genotype $B_j B_l$; the set of these frequencies is referred to as the genotypic structure at the B-locus.

p^i, p_j are the allele frequencies of A_i and B_j at the A- and B-locus, respectively; they specify the allelic structure at the respective locus.

P_j^i frequency of gametes carrying the alleles A_i and B_j within the gametic production of the population; the set of these frequencies is referred to as the gametic structure.

G_j^i probability that two genes, one taken at random from the A-locus and the other taken at random from the B-locus of the same individual, are A_i and B_j; the set of these frequencies is referred to as the joint genic structure at the A- and B-locus.

$p_{j/ik}$ frequency of the allele B_j within the group of individuals posessing genotype $A_i A_k$.

$p_{j;ik} : = p_{j/ik} \cdot P^{ik}$.

Since $P_j^i = \frac{1}{2} \cdot \sum_{k,l} (c \cdot P_{jl}^{ik} + (1 - c) \cdot P_{lj}^{ik}) + \frac{1}{2} \cdot P_{jj}^{ii}$ and

$G_j^i = \frac{1}{4} \cdot \sum_{k,l} (P_{jl}^{ik} + P_{lj}^{ik}) + \frac{1}{2} \cdot P_{jj}^{ii}$, for $c = \frac{1}{2}$ (no linkage)

we obtain $P_j^i = G_j^i$.

In order to demonstrate the pure effect of the mating system, selection, mutation, and migration are excluded, population size is assumed to be hypothetically infinite and reproduction to occur in discrete, non-overlapping generations. The mating behavior of individuals is governed exclusively by the A-locus as follows: each individual possessing genotype $A_i A_k$ has probability α of selectively (assortatively) mating with an $A_i A_k$ type, while with probability $1-\alpha$ it mates at random from the whole population; α is the same for all genotypes with regard to the

A-locus. All individuals are equally likely to take part in the process of reproduction.

As a consequence of these conditions, the allelic structures at the A- as well as at the B-locus do not change in the course of the generations. For further details, which however are not explicitly used in the succeeding derivations, the reader is referred to Gregorius (1980).

The Gametic Structure

In selection theory, it is a well known phenomenon that deviations of the gametic structure from independent assignment of genes to the different loci (frequently measured by a quantity called 'linkage disequilibrium') to a large extent determine the evolutionary course and speed of linked genes (Bodmer and Felsenstein 1967; Felsenstein 1965; Nei 1963). The usefulness of considering gametic structures to obtain the probabilities with which fixation of different pairs of genes occurs, when several systems of inbreeding are applied, has been pointed out by Kimura (1963).

The findings of Bodmer, Felsenstein, Kimura and Nei suggest that it would be worthwhile to investigate the change in gametic structures for the situation dealt with here, and in particular to raise the question as to possible implications for the genotypic structure at the B-locus. Denoting successive generations by a prime and taking note of the identity $G_j^i = p_{j;ii} + \frac{1}{2} \cdot \sum_{\substack{k \\ k \neq i}} p_{j;ik}$, the following system of difference equations can be derived from Gregorius [1980, Eqs. (5) and (1)]:

$P'^i_j = c \cdot P_j^i + \alpha(1 - c) \cdot G_j^i + (1 - \alpha)(1 - c) \cdot p^i \cdot p_j$

$G'^i_j = \frac{1}{2} \cdot [\alpha \cdot G_j^i + P_j^i + (1 - \alpha) \cdot p^i \cdot p_j]$

This system can be simplified in two steps. First we introduce $D_j^i : = P_j^i - p^i \cdot p_j$, which is the common measure linkage disequilibrium,

$R_j^i = G_j^i - p^i \cdot p_j$, and from this

$D'^i_j = c \cdot D_j^i + \alpha(1 - c) \cdot R_j^i, \quad R'^i_j = \frac{1}{2} \cdot D_j^i + \frac{1}{2} \cdot \alpha \cdot R_j^i.$ (1)

Secondly, we define a two-component vector X_j^i and a 2 × 2 matrix M by

$$X_j^i : = \begin{pmatrix} D_j^i \\ R_j^i \end{pmatrix} \quad \text{and} \quad M : = \begin{pmatrix} c, & \alpha(1 - c) \\ \frac{1}{2}, & \frac{1}{2} \cdot \alpha \end{pmatrix}.$$

From this, a matrix representation of Eq. (1) is obtained, namely

$$X'^i_j = M \cdot X_j^i. \quad (1a)$$

For random mating ($\alpha = o$) or complete linkage ($c = 1$), Eq. (1) reflect the corresponding well-known results,

while for $c = \frac{1}{2}$ and $\alpha = 1$ the gametic and joint genic structures do not change.

In the following, the indices i and j are omitted if there is no chance of confusion, and the subscript t is added to denote the generation number; upper bars denote equilibrium or limiting values.

To arrive at the possible equilibrium states for the gametic and joint genic structure, the condition $X' = X$ has to be met, which is equivalent to

$(1 - c) \cdot \bar{D} = \alpha(1 - c) \cdot \bar{R}$ and $\frac{1}{2} \cdot \bar{D} = (1 - \frac{\alpha}{2}) \cdot \bar{R}$.

The solutions to the system are:

For $\alpha \neq 1 \neq c : \bar{D} = \bar{R} = 0$; for $\alpha = 1 : \bar{D} = \bar{R}$; and for $c = 1$:

$\bar{D} = (2 - \alpha) \cdot \bar{R}$.

Expressed in words, this result tells us that a population is not able to maintain a permanent gametic linkage disequilibrium as long as assortative mating is only partial ($\alpha \neq 1$) and linkage is not complete ($c \neq 1$).

Viewing a population whose gametic or joint genic structure still changes, some important statements about the properties and significance of equilibrium states can be derived. Primarily, it would be interesting to know which conditions must be realized so that the structures approach a limit, and furthermore to relate such a limit to corresponding equilibrium states.

For this purpose the initially stated assumption of discrete, non-overlapping generations applied to the recursively defined Eq. (1a) and

$X_t = M^t \cdot X_0$

is obtained, provided initiation is with generation 0, and the two structures in question are observed t generations later.

From the basic theory of matrices (Gantmacher 1959), it is known that the powers of M converge (in the sense of uniform convergence) for all admissable values of α and c $(0 \leq \alpha \leq 1, \frac{1}{2} \leq c \leq 1)$ to a limiting matrix $\bar{M}$. Consequently, the gametic as well as the joint genic structures converge in every case.

In particular for $\alpha \neq 1 \neq c$, $\bar{M}$ contains only zeros, and

for $c = 1$ it is easily shown that $\bar{M} = \begin{pmatrix} 1 & , 0 \\ \dfrac{1}{2 - \alpha} & , 0 \end{pmatrix}$. Further-

more, for $\alpha = 1 \neq c$, M is a stochastic matrix with positive elements; thus $\bar{M}$ has two identical rows and $\bar{M} \cdot M = \bar{M}$, which implies that the elements of the first column are

both equal to $\dfrac{1}{3-2c}$ and those of the second column both

equal to $\dfrac{2(1-c)}{3-2c}$.

Summarizing these findings, at the limit

$\bar{D} = \bar{R} = 0$ for $\alpha \neq 1 \neq c$,

$\bar{D} = D_0, \quad \bar{R} = \dfrac{1}{2 - \alpha} \cdot D_0$ for $c = 1$ (2)

$\bar{D} = \bar{R} = \dfrac{D_0 + 2(1 - c) \cdot R_0}{3 - 2c}$ for $\alpha = 1$.

If equilibrium states with limits are compared, the kind of interrelationship between them becomes apparent: for

each given equilibrium state $\bar{X} = \left(\dfrac{\bar{D}}{\bar{R}} \right)$, Eqs. (2) de-

scribe possible initial conditions for a population to converge to this $\bar{X}$. Another remarkable facet of Eqs. (2) is that in the case of initial linkage disequilibrium combined with partial assortative mating and complete linkage, a difference between the gametic and joint genic structure is always maintained, while in the remaining cases this difference gradually disappears.

The Genotypic Structure at the B-locus

Gregorius (1980) found it to be characteristic of the process of positive assortative mating at the A-locus on the genotypic structure of the B-locus that the homozygote frequencies at the B-locus always exceed the corresponding Hardy-Weinberg frequencies. On the other hand, for $\alpha \neq 1 \neq c$ there exists a unique joint equilibrium state for the A and B loci showing Hardy-Weinberg frequencies at the B-locus, a fact from which one might suspect that at least for some cases, the surplus of homozygotes vanishes in the course of the generations. Whether this is a general tendency or not shall be investigated in this section. Here, the concern is not the derivation of possible equilibrium states for the genotypic structure at the B-locus, but rather the limiting behavior of this structure. In order to facilitate examination of these considerations, some already known relations shall be recalled [Gregorius 1980, Eqs. (3), (4a), (5)]:

$P'^{ik} = \frac{1}{2} \cdot \alpha \cdot P^{ik} + 2(1 - \alpha) \cdot p^i \cdot p^k$ for $i \neq k$,

$P'^{ii} = \frac{1}{2} \cdot \alpha \cdot (p^i + P^{ii}) + (1 - \alpha) \cdot (p^i)^2$, and (3)

$P'_{jl} = 2\alpha \cdot C_{jl} + 2 \cdot p_j \cdot p_l$ for $j \neq l$, $P'_{jj} = \alpha \cdot V_j + (p_j)^2$,

where $C_{jl} = \sum\limits_{\substack{i,k \\ i \leq k}} (p_{j/ik} - p_j)(p_{l/ik} - p_l) \cdot P^{ik}$ is the covari-

ance of the frequencies of the alleles B_j and B_l and $V_j = C_{jj}$ is the variance of the frequency of the allele B_j between the groups of $A_i A_k$ individuals; finally

$p'_{j;ik} = \frac{1}{2} \cdot \alpha \cdot p_{j;ik} + (1 - \alpha)(P^i_j \cdot p^k + P^k_j \cdot p^i)$ for $i \neq k$,

$p'_{j;ii} = \frac{1}{2} \cdot \alpha \cdot (P^i_j + p_{j;ii}) + (1 - \alpha) \cdot P^i_j \cdot p^i$.

Now, in the preceeding section, convergence of the quantities P^i_j for all values of α and c with an increasing

Fig. 1.

Fig. 2.

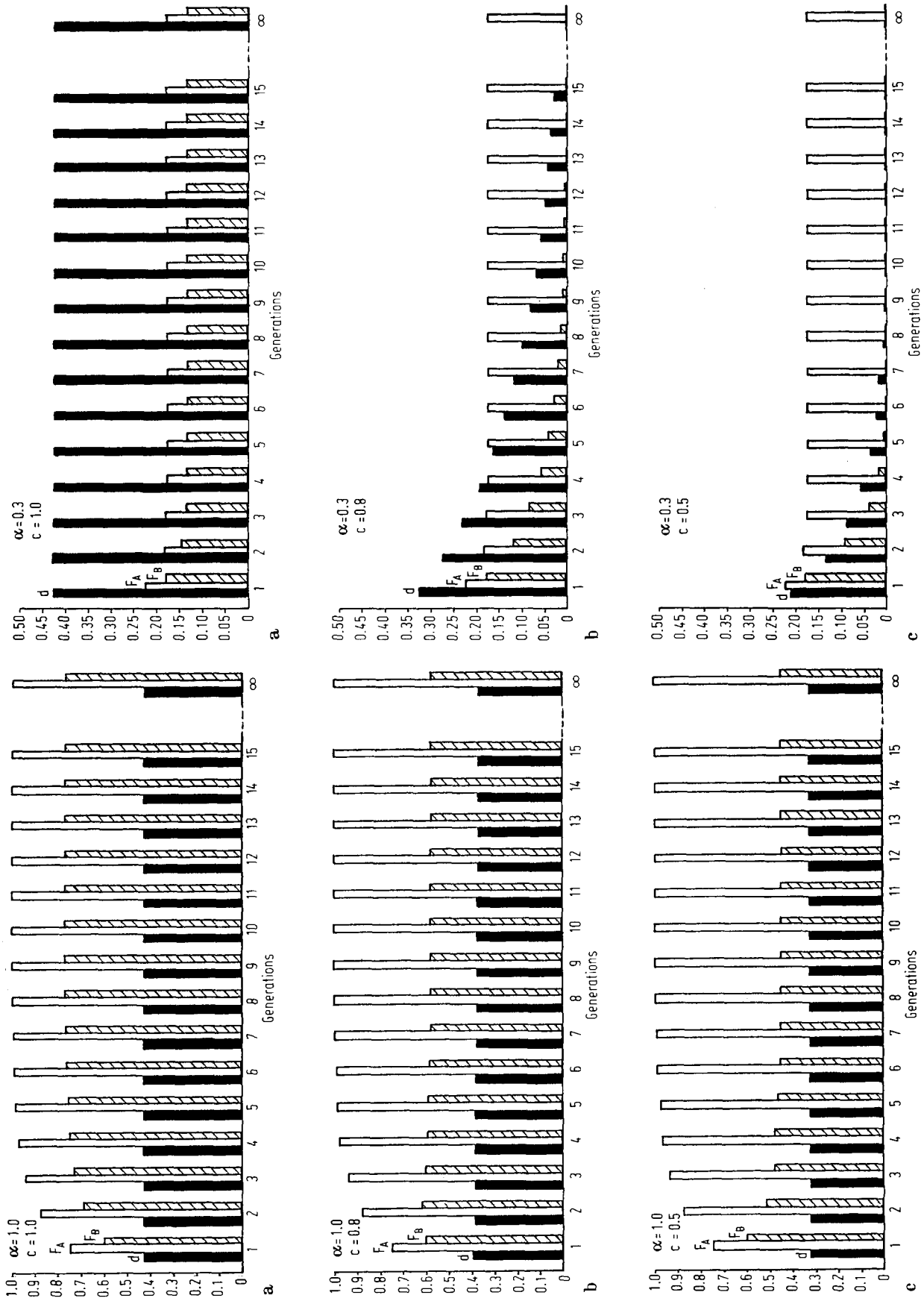**Fig. 1a-c. and Fig. 2a-c.** The excess of homozygotes as compared to the corresponding Hardy-Weinberg proportions at the A and B loci, and the deviation of the gametic structure from an independent association of genes. F_A, F_B = 'apparent coefficient of inbreeding' at the A and B loci, respectively; d = genetic distance between the allelic structure and the hypothetical structure resulting from independent association of genes; α = proportion of positive assortatively mating individuals with respect to the A-locus; 1-c = recombination frequency between the A and B loci

number of generations was proved, and the limits $\overline{P}^i_j$ are given by Eqs. (2). This in turn implies convergence of the quantities $p_{j;ik}$, as can be taken from the last two recurrence relations. Since convergence of the genotypic structure at the A-locus is also warranted, a simple computation utilizing the individual limits (compare Gregorius 1980: Section about the genetic equilibrium) leads to

$$\overline{P}_{j/ik} = \tfrac{1}{2} \cdot (\overline{P}^i_j/p^i + \overline{P}^k_j/p^k) = \tfrac{1}{2} \cdot (\overline{D}^i_j/p^i + \overline{D}^k_j/p^k) + p_j.$$

Since the C_{jl} are the essential quantities in (3), convergence of the genotypic structure at the B-locus for all values of α and c is consequently proved, and furthermore,

$$\overline{C}_{jl} = \tfrac{1}{4} \cdot \sum_{\substack{i,k \\ i \leq k}} (\overline{D}^i_j/p^i + \overline{D}^k_j/p^k)(\overline{D}^i_l/p^i + \overline{D}^k_l/p^k) \cdot \overline{P}^{ik}, \quad (4)$$

where $\overline{P}^{ik} = \dfrac{2(1-\alpha)}{2-\alpha} \cdot 2 \cdot p^i \cdot p^k$ for $i \neq k$ and

$P^{ii} = \dfrac{\alpha + 2(1-\alpha)p^i}{2-\alpha} \cdot p^i$ are the well-known equilibrium

values for positive assortative mating at the A-locus, and the $\overline{D}^i_j$ are given by (2).

With the help of these statements, the initially stated question as to the conditions under which a population attains a Hardy-Weinberg structure at the limit, with respect to the B-locus can now be answered. In order for the set of $\overline{P}_{jl}$'s to form a Hardy-Weinberg structure, it is necessary and sufficient that $\overline{V}_j = 0$ for all j, which in turn is equivalent to $\overline{D}^i_j = 0$ for all i, j (linkage equilibrium); on the other hand, all situations for which linkage equilibrium is realized can be derived from (2).

From another point of view, the latter remark in particular indicates that maintenance of deviations from Hardy-Weinberg proportions at the B-locus at least requires complete positive assortative mating ($\alpha = 1$) or complete linkage combined with initial disequilibrium (c = 1, $D_o \neq 0$).

A considerable simplification of (4), and thus of the limiting representation of (3), is obtained if one assumes $\alpha = 1$:

$$\overline{C}_{jl} = \sum_i \overline{D}^i_j \cdot \overline{D}^i_l/p^i.$$

Again, considering the corresponding formula from Eqs. (2) for $\overline{D}^i_j$, this case constitutes an example of the joint influence of the initial gametic as well as joint genic structure on the limiting state of the genotypic structure at the B-locus; thus the joint genic structure cannot simply be regarded as an auxiliary variable, but rather has its own significance.

Numerical Demonstrations

So far, these presentations have been concerned exclusively with the derivation and computation of equilibrium and limiting states. This last section shall mainly be devoted to the representation of changes in gametic structure and genotypic structures in the course of generations. Since simple analytical solutions cannot be obtained in the majority of cases, a numerical treatment of the subject is employed, taking into account a certain loss of generality of the resulting conclusions.

In order to subsume characteristic properties of the gametic structure and the genotypic structures at each of the two loci in question, it is convenient to choose parameters that reflect deviations of the gametic structure from the corresponding random association of genes and deviations of the homozygote from the corresponding Hardy-Weinberg frequencies. The difference between the actual and theoretical gametic structures shall be measured as the genetic distance between them, applying a distance measure suggested earlier (Gregorius, 1974); for this purpose it takes the very simple form

$$d := \tfrac{1}{2} \cdot \sum_{i,j} |D^i_j|;$$

d ranges between 0 and 1.

Provided that a kind of mating system or mixture of populations (as used, for example, for Wahlund's principle) is considered that implies a surplus of homozygotes as compared to the corresponding Hardy-Weinberg structure, a measure of homozygosity commonly used is the so called 'apparent coefficient of inbreeding' F. F is defined as the deviation of the homozygote frequency from the corresponding random-mating proportion, divided by the random-mating proportion of heterozygotes

$$F_A := \frac{\sum_i (P^{ii} - (p^i)^2)}{1 - \sum_i (p^i)^2}$$

for the A-locus; F_B is defined analogously. The word 'apparent' emphasizes the difference from the ordinary coefficient of inbreeding. Under the conditions mentioned above, F is a normalized quantity and ranges from 0 (Hardy-Weinberg structure) to 1 (complete homozygosity).

In the preceeding sections the decisive significance of α, c, the initial gametic structure, and the joint genic structure for the asymptotic behavior of a population were pointed out. Therefore, and in order to avoid redundancy in numerical demonstrations, these considerations will be restricted to just one initial situation for the joint genotypic structure at the A and B loci, which reflects a sufficiently representative example for a nontrivial population structure. Based upon this situation, the effect of linkage and the degree of assortativity using several values for c and α shall be displayed. The initial population structure is summarized in the following table for a model with two alleles at each locus:

Table 1. The joint genotypic structure at the A- and B-locus (P_{jl}^{ik}) and the resulting gametic as well as joint genic structure (P_j^i, G_j^i), and the genotypic structure at the A- (P^{ik}) and at the B-locus (P_{jl}) for the initial generation. 1-c = recombination frequency

P_{jl}^{ik}	A_1 B_1	A_1 B_2	A_2 B_1	A_2 B_2	P_j^i for			G_j^i
					c = 0.5	c = 0.8	c = 1.0	
A_1 B_1	0.45	0.008	0.015	0.21	0.5165	0.5465	0.5665	0.5165
A_1 B_2	0.008	0.009	0.01	0.007	0.0715	0.0415	0.0215	0.0715
A_2 B_1	0.015	0.01	0.02	0.01	0.0875	0.0575	0.0375	0.0875
A_2 B_2	0.21	0.007	0.01	0.261	0.3245	0.3545	0.3745	0.3245

	$A_1 A_1$	$A_1 A_2$	$A_2 A_2$		$B_1 B_1$	$B_1 B_2$	$B_2 B_2$
P^{ik}	0.467	0.242	0.291	P_{jl}	0.485	0.238	0.277

Before proceeding to graphic representations, the limiting values for F_A and F_B shall be investigated briefly.

$$F_A = \frac{\alpha}{2 - \alpha} \tag{5.a}$$

and from (3): $(1 - \sum_j (p_j)^2 . \overline{F}_B = \alpha \cdot \sum_j \overline{V}_j$.

If one now uses the assumption that at each of the two loci two alleles are present, it follows that

$$D : = D_1^1 = D_2^2 = -D_2^1 = -D_1^2 \quad \text{and} \quad D = P_1^1 \cdot P_2^2 - P_2^1 \cdot P_1^2,$$

which is the commonly used measure of linkage disequilibrium. Thus

$$d = 2 \cdot | D | . \tag{5b}$$

Furthermore, some straight-forward calculations show that $\overline{C}_{jl}$, as given by (4), boils down in the two-allelic case to

$$\overline{V}_1 = \overline{V}_2 = \frac{\overline{D}^2}{(2 - \alpha) \cdot p^1 \cdot p^2} = -\overline{C}_{12},$$

and therefore

$$\overline{F}_B = \frac{\alpha}{2 - \alpha} \cdot \frac{\overline{D}^2}{p^1 \cdot p^2 \cdot p_1 \cdot p_2}, \tag{5c}$$

which specifies the relationship

$$\overline{F}_B = \overline{F}_A \cdot \frac{\overline{D}^2}{p^1 \cdot p^2 \cdot p_1 \cdot p_2}.$$

The second quotient in (5c) is known as the squared cor-

relation of genes at the A and B loci; because of this (or equivalently because of Eq. (iii) in the appendix):

$$\overline{F}_B \leq \overline{F}_A.$$

Since the genetic structures at each of the two loci are constant over the course of the generations and, beyond this, in most cases can be estimated from experimental data, it is of primary interest to provide information about the possible upper and lower bounds for $\overline{D}$, and thus about the possible maximum value for $\overline{F}_B$ assuming both allelic structures to be given. Some aspects of this problem have been considered earlier, for example by Sved (1971), but no sufficiently complete solutions have been obtained so far. A contribution toward such a solution is presented in the appendix. Applying these results to the situation stated in the table (in which $p^1 = 0.588$, $p^2 = 0.412$; $p_1 = 0.604$, $p_2 = 0.396$), one may obtain:

$$-0.1632 \leq \overline{D} \leq 0.2328, \text{ and thus } \overline{F}_B \leq \frac{\alpha}{2 - \alpha} \cdot 0.9357$$

and $d \leq 0.4657$.

Within the following graphic representations, Eqs. (5) shall serve as a standard, which for each generation give information about the deviation of the population from its equilibrium (limiting) state. The closeness of the limiting states to their maximum values in turn can be viewed in connection with the above numbers.

Figures 1a-2c illustrate the previously stated presumption concerning the significance of the deviations of the gametic structure from stochastically independent association of alleles (measured by d) for the genotypic structure at the B-locus. There is a remarkably close relationship between the absolute amount of linkage disequilibrium (d) and the degree to which changes at the A-locus are accompanied by changes at the B-locus. The extreme sensitivity of this relationship, at least for complete assortativity, is illustrated in Figures 1a-1c. While in Figure 1a, d is sufficiently large to induce a considerable increase of F_B in the course of the generations, this tendency in Figures 1b and 1c is reduced and even inverted, although d is only slightly diminished.

Figures 2a-2c depict the important role played by the coefficient of linkage, c, in case assortative mating is only partial. On the one hand, complete linkage (Fig. 2a) generally leads to qualitatively the same consequences as complete assortativity. However, on the other hand, incomplete linkage always causes the deviations of the genotypic structure at the B-locus from Hardy-Weinberg proportions to approach zero, and the speed of this decline is strongly associated with the magnitude of c.

Taking into account that for $\alpha = 1$ and $\alpha = 0.3$, the maximum values for $\overline{F}_B$ are 0.9357 and 0.1651, respectively, and the maximum value for d is 0.4657, it shows that the situations presented in Figures 1a and 2a are

examples of populations that are able to maintain the transmission of developments from the A- to the B-locus at a high rate.

Acknowledgements

Literature

Bodmer, W.F.; Felsenstein J., (1967): Linkage and selection: Theoretical analysis of the deterministic two locus random mating model. Genetics 57, 237-265

Felsenstein, J. (1965): The effect of linkage on directional selection. Genetics 52, 349-363

Gantmacher, F.R (1959): The theory of matrices, Part 2. Chelsea, N.Y. (1974)

Gregorius, H.-R. (1974): Genetischer Abstand zwischen Populationen. I. Zur Konzeption der genetischen Abstandsmessung. Silvae Genetica 23, 22-27

Gregorius, H.R. (1980): The effect of positive assortative mating at one locus on a second linked locus, Part 1: The genotypic structure of the offspring generation. Theor. Appl. Genet. 56, 187-192

Kimura, M. (1963): A probability method for treating inbreeding systems, especially with linked genes. Biometrics 19, 1-17

Lewontin, R.C. (1964): The interaction of selection and linkage. Part 1: General considerations; heterotic models. Genetics 49, 49-67

Nei, M. (1963): Effect of selection on the components of genetic variance. In: Statistical genetics and plant breeding. Math. Acad. Sci.-Natl. Res. Council Publ. 982, 501-515

Sved, J.A. (1971): Linkage disequilibrium and homozygosity of chromosome segments in finite populations. Theor. Popul. Biol. 2 125-141

Dr. H.-G. Gregorius
Lehrstuhl für Forstgenetik
und Forstpflanzenzüchtung der Universität
Büsgenweg 2
D-34 Göttingen-Weende (Federal Republic of Germany)

Appendix

For the case of two alleles at each of the two loci,

$$D : = P_1^1 \cdot P_2^2 - P_2^1 \cdot P_1^2$$

is the commonly used measure of linkage disequilibrium.

In order to simplify derivations, set

$$a : = D/(p^1 \cdot p^2) \quad \text{and} \quad a_i : = P_i^i/p^i \quad (i = 1, 2).$$

From this,

$$a = a_1 \cdot a_2 - (1 - a_1)(1 - a_2) = a_1 + a_2 - 1 \quad \text{and}$$

$$p_2 = a_2 \cdot p^2 + (1 - a_1) \cdot p^1,$$

which implies

$$a = a_1 + p_2/p^2 - (1 - a_1) \cdot p^1/p^2 - 1 = (a_1 - p_1)/p^2.$$

Now, $0 \leq a_1 \leq 1$ and $0 \leq a_2 = (p_2 - (1 - a_1)p^1)/p^2 \leq 1.$

Therefore

$$I_1 : = \max [0, 1 - p_2/p^1] \leq a_1 \leq \min [1, p_1/p^1] = : S_1 ;$$

and $I_1 \leq p_1 \leq S_1.$

Hence $(I_1 - p_1)/p^2 \leq a \leq (S_1 - p_1)/p^2,$

which is equivalent to

$$p^1 \cdot (I_1 - p_1) \leq D \leq p^1 \cdot (S_1 - p_1). \tag{i}$$

In fact, this is the same result Lewontin (1964) arrived at by using a different proof technique.

Since a is a linearily increasing function of a_1, D attains its lower and upper bound if and only if $P_1^1 = p^1 \cdot S_1$ and $P_1^1 = p^1 \cdot S_1$ respectively. Note that $I_1 = 0$ and $S_1 = 1$ simultaneously if and only if $p^1 \leq \min [p_1, p_2].$

Consequently, D attains its lower bound if and only if

$$P_1^1 = p^1 \cdot I_1, P_1^2 = p_1 - p^1 \cdot I_1,$$

$$P_2^1 = p^1 \cdot (1 - I_1), P_2^2 = p_2 - p^1 \cdot (1 - I_1),$$

and D attains its upper bound if and only if

$$P_1^1 = p^1 \cdot S_1, P_1^2 = p_1 - p^1 \cdot S_1, P_2^1 = p^1 \cdot (1 - S_1),$$

$$P_2^2 = p_2 - p^1 \cdot (1 - S_1).$$

With regard to (5c) the following derivation makes a statement about an upper bound for D^2.

Define $I_2 : = \max [0, 1 - p_1/p^2]$ and $S_2 : = \min [1, p_2/p^2].$

As before, $I_2 \leq p_2 \leq S_2$ and $p^1(I_1 - p_1) = p^2(I_2 - p_2),$

$$p^1(S_1 - p_1) = p^2(S_2 - p_2).$$

If this is applied to (i),

$$D^2 \leq p^1 \cdot p^2 \cdot (p_1 - I_1)(p_2 - I_2) \quad \text{or}$$

$$D^2 \leq p^1 \cdot p^2 \cdot (S_1 - p_1)(S_2 - p_2),$$

which implies

$$D^2 \leq p^1 \cdot p^2 \cdot \max [(p_1 - I_1)(p_2 - I_2),$$

$$(S_1 - p_1)(S_2 - p_2)]. \tag{ii}$$

Furthermore, $(p_1 - I_1)(p_2 - I_2) \leq p_1 \cdot p_2$ and

$(S_1 - p_1)(S_2 - p_2) \leq (1 - p_1) \cdot (1 - p_2) = p_1 \cdot p_2,$ and thus

$$D^2 \leq p^1 \cdot p^2 \cdot p_1 \cdot p_2 \qquad \text{(iii)}$$

It is particularly interesting to investigate the conditions under which D^2 attains a value that is equal to the right hand side of (iii). According to (ii), this is equivalent to finding the conditions for max $[(p_1-I_1)(p_2-I_2),$ $(S_1-p_1)(S_2-p_2)] = p_1 \cdot p_2$, which holds if and only if $I_1 =$ $I_2 = 0$ or $S_1 = S_2 = 0$. On the other hand, $I_1 = I_2 = 0$ is equivalent to $p^1 = p_2$, and $S_1 = S_2 = 0$ is equivalent to $p^1 = p_1$. Thus: a necessary and sufficient condition for the maximum value of D^2 (given p^1, p_1) to be equal to $p_1 \cdot p_2 \cdot p^1 \cdot p^2$ is $p^1 = p_2$ or $p^1 = p_1$.