Günter Wricke · W. Eberhard Weber
Quantitative Genetics and Selection in Plant Breeding

Günter Wricke · W. Eberhard Weber

Quantitative Genetics and Selection in Plant Breeding

Walter de Gruyter
Berlin · New York 1986

Authors

Professor Dr. Günter Wricke
Professor Dr. W. Eberhard Weber
Institute of Applied Genetics
University of Hannover
Herrenhäuser Strasse 2
D-3000 Hannover 21
Federal Republic of Germany

With numerous figures and tables.

Library of Congress Cataloging-in-Publication Data

Wricke, Günter, 1928–
Quantitative genetics and selection in plant breeding.
Bibliography: p.
Includes indexes.
1. Plant-breeding. 2. Plant genetics. 3. Selection (Plant-breeding) I. Weber, Eberhard, 1941– II. Title.
SB123.W75 1986 631.5'3 85–29195
ISBN 0-89925-143-9 (U.S.)

CIP-Kurztitelaufnahme der Deutschen Bibliothek

Wricke, Günter:
Quantitative genetics and selection in plant breeding / Günter Wricke ; Eberhard Weber. –
Berlin ; New York : de Gruyter, 1986.
ISBN 3-11-007561-X

NE: Weber, Eberhard:

Preface

This book is a presentation of the principles of selection in plant breeding. It is intended for students in plant breeding and genetics, as well as for plant breeders and applied geneticists. The principles of selection described can be used by breeders of agronomic, horticultural, and forestry plants.

Most of the economically important traits are quantitatively inherited. An understanding of selection theory in plant breeding must be based upon knowledge of the inheritance of such traits. Variance and covariance components are used for the prediction of selection response. The covariance between relatives and the estimation of variance and covariance components are therefore treated in some detail. Much attention has been given to the statistical problems involved in precise estimation of such components from experiments.

Selection methods in plant breeding depend on the natural reproductive system. They can therefore be classified according to crops that reproduce by self-fertilization, cross-fertilization and by asexual means. Topics such as autotetraploidy and synthetic varieties are also taken into consideration and may seem to be overvalued. However, since autotetraploidy is very important in some crops the impact polyploidy has on selection procedures must be an integral part of a book on selection in plant breeding. The theory of synthetic varieties has been largely developed in the last few years and has therefore been treated in some detail. In hybrid breeding, cross-fertilization is used in a way not occurring in nature. Hybrid breeding is discussed in connection with cross-fertilizing crops, though it is also used in naturally self-fertilizing crops.

Success of selection in plant breeding is influenced by the types of gene action involved in the expression of the traits to be selected. The theory of selection is easier if epistatic gene action is omitted. Experiments have generally shown that epistatic variance is usually very low in comparison to additive and dominance variance. Furthermore, the precision of estimates of epistatic variance is usually low. This is mainly a consequence of the linear model used for their estimation. Nevertheless, since some investigations have shown that there is a certain amount of epistatic variance it seems justified to give the extended quantitative genetic model including epistasis in the introductory chapters. The importance of epistasis is discussed in the chapters on selection for specific types of varieties.

The theory of selection cannot be treated without a certain degree of mathematical and statistical handling, and a plant breeder who wants to compare different selection procedures must use the formulae as given in the book to determine the response to selection. To assist those less familiar with the mathematical treatment, the formulae are elucidated by simple numerical examples

throughout the book. They may be skipped over by readers more familiar with selection theory. Some knowledge of genetics and statistics is assumed; in most cases, however, this does not go beyond the well-established methods and techniques which today are part of introductory plant breeding courses.

Matrix notation is used in chapter 12, dealing with selection for several characters. The reader not familiar with matrix algebra will find a short introduction in the appendix, although this is not intended to replace a textbook on linear algebra. Other topics in the appendix are fundamental probability distributions, the general linear model, the principle of least squares, and the error of variance components.

The critical reading and helpful comments of our colleagues Prof. M. Hühn (Kiel), Dr. Sabine Franken-Bembenek (Gießen) and Dr. L. Frese (Braunschweig) are gratefully acknowledged. We also thank Dr. Q.S. Trang (Hannover) who wrote parts of the chapters on tetraploids. We are especially obliged to Prof. C.O. Qualset (Davis, California) and Dr. A. Rosielle (Ames, Iowa) for their careful reading of the manuscript, their numerous proposals and their revision of the English text. We deeply appreciate the effort and interest these colleagues devoted to our book. We are responsible for any errors or faults in treatment which remain. Special thanks go to Mrs. Ruth Peters, Miss Jutta Freye and Mr. M. Stümper (all from Hannover) for their help in the preparation of the typescript.

Finally we thank the Walter de Gruyter Publishing Company and especially Dr.-Ing. Rudolf Weber for the care he has given to the production of this work.

November 1985

Günter Wricke
W. Eberhard Weber

Contents

1. Basic population genetics

A population is a group of individuals of the same species with different genetic structures, sharing space and time. In higher animals the only way for reproduction and propagation is mating between individuals. Dobzhansky (1953) called such populations "Mendelian populations". Falconer (1981) used the term "population" in this sense. But in plants mating between individuals is not the only way that the population can survive. Other mechanisms are self-fertilization and vegetative propagation. Therefore in this book the term "population" is not restricted to Mendelian populations.

In the case of sexual reproduction the genotypes themselves are not transmitted to the next generation. Gametes are formed and the genotypes of the next generation arise from the fusion of gametes. There is a great difference between cross-fertilizing and self-fertilizing plant species. In cross-fertilizing species the genotypes of the next generation differ from the genotypes of the previous generation. But in pure self-fertilizing species the individuals are completely homozygous and all of the gametes produced by an individual are identical. Since two gametes of the same individual unite, the progenies are genotypically identical with the parents. Genetic recombination is then restricted to the rare event of cross-pollination.

In vegetatively propagated cultivated plants no sexual process is involved, and the genotypes remain unchanged through time. Genetic changes are possible only by mutations, which shall be neglected in this context. Many crops which are propagated vegetatively in culture, show a certain degree of cross-fertilization in nature and consequently genetic recombination.

Some cultivated crops are autopolyploid. The genetic behaviour of autopolyploids is quite different from diploids. This has important consequences on the plant breeding system chosen for crop improvement goals. The basic population genetics of autopolyploids is treated in a separate section.

1.1 Vegetatively propagated species

The population consists of different genotypes, which do not exchange genetic material by sexual processes. It should be noted that in most cases sexual reproduction is possible, and the breeder makes use of this fact. Single genotypes are highly heterozygous, and therefore the progeny of a cross shows large genetic

variation. This normally is the basis for selection. The progeny can be regarded as a new population.

To characterize the population, it is only necessary to describe the distribution of genotypes. We can expect, that the number of loci, which contribute to genetic variation, is high. A rough estimate of the number of loci in higher organisms is at least 50,000. Kamalay and Goldberg (1980) estimated the number of structural genes of tobacco at 60,000. Therefore the number of possible different genotypes is extremely large. As an example consider the number of different genotypes in a cross of two clones, which differ at 50 loci, and that each clone has two different alleles at each locus.

Number of different gametes of clone 1:	2^{50}
Number of different gametes of clone 2:	2^{50}
Number of different genotypes in the hybrid:	$2^{100} \approx 10^{30}$

The genetic structure of a population of vegetatively propagated plants is unknown. Even in the very simple case, where different genotypes of the population are derived from one cross, four different alleles are possible at each locus. At these loci the frequency of genotypes in the population can be calculated in the following way:

If parent 1 has the alleles B_1 and B_2 and parent 2 the alleles B_3 and B_4, the genotypes B_1B_3, B_1B_4, B_2B_3 and B_2B_4 can be produced, all with the expected frequency of $\frac{1}{4}$. No homozygous genotypes arise when the B alleles are all different from one another. Homozygous progeny at the B locus are possible only when both parents carry a common allele. For example, in the progeny of $B_1B_2 \times B_1B_3$ $\frac{1}{4}$ of homozygotes B_1B_1, and $\frac{3}{4}$ of heterozygotes ($\frac{1}{4}$ each of B_1B_2, B_1B_3 and B_2B_3) are expected. The probability that the two parents carry a common allele depends on the history of the parents and the number of different alleles which exist. Later we will discriminate between alleles identical by descent and alike in state (see page 16). Only when the breeder crosses two related parents, may a genotype be homozygous due to alleles identical by descent; otherwise it can only be homozygous for alleles alike in state or heterozygous. In general we expect that the genotypes arising from a cross between two parents will be heterozygous at a large number of loci.

1.2 Cross-fertilizing species

The theory of population genetics in standard textbooks (Li 1955, Crow and Kimura 1970, Wricke 1972, Falconer 1981, and others) can be applied to this case. Since this book is mainly concerned with selection in plant breeding, only some general remarks are made, which are necessary to understand selection theory in cross-fertilizing crops. In cross-fertilizing species the population is a

Mendelian population in the definition of Dobzhansky (1953). The genetic material is recombined each generation. Only the genes are transmitted from generation to generation, not the genotypes. In each generation they are reassorted and form new genotypes.

The population can be described in terms of genotypic or allelic frequencies. If there is no doubt, we follow other textbooks and use the term "gene frequencies" to mean allele frequencies. However this is imprecise, since it is allelic frequency rather than gene frequency which is meant.

One locus with two alleles

At first we consider a single locus with two different alleles B and b. Then three genotypes BB, Bb, and bb are possible. Let the absolute frequencies of these genotypes be x, $2y$ and z. From these frequencies we can deduce the gene frequencies. BB genotypes carry two B alleles and Bb genotypes one B and one b allele. Therefore the frequency of allele B is $2x + 2y = 2(x + y)$. In a similar way the frequency of b is $2(y + z)$.

It is convenient to use relative frequencies. If the number of individuals in the population is N, the relative frequency for B is $p = \frac{2(x+y)}{2N} = \frac{(x+y)}{N}$ and for b is $q = \frac{2(y+z)}{2N} = \frac{(y+z)}{N}$ with $p + q = 1$. In the following, p and q are called gene frequencies.

We now assume, that there is *random mating*, that is that mating occurs without regard to genotype. This is equivalent to a random combination of gametes. The genotypic frequencies of the next generation can be obtained as follows:

female gametes \ male gametes	p B	q b
p B	p^2 BB	pq Bb
q b	pq Bb	q^2 bb

The genotypic frequencies therefore are given by p^2 for BB, $2pq$ for Bb and q^2 for bb. If we consider the gametic output of the genotypes of this generation we again get $p^2 + pq = p$ for B and $pq + q^2 = q$ for b, since $p + q = 1$. Therefore the genotypic frequencies after one generation of random mating are not changed in the following generations.

The genotypic frequencies following random mating can be calculated from the gene frequencies in the following way:

$$(pB + qb)^2 = p^2 BB + 2pq\,Bb + q^2 bb.$$

This principle is called the Hardy-Weinberg principle or Hardy-Weinberg law and was first shown independently by Hardy (1908) and Weinberg (1908) (see also Keeler (1968) for early contributions of W. E. Castle (1903)). The population is in equilibrium after one generation of random mating. Therefore the Hardy-Weinberg principle is sometimes called Hardy-Weinberg equilibrium. This principle holds only in the absence of mutation, migration, selection, and random drift, which lead to change in gene frequencies. The effect of mutation and migration on single loci will not be considered in this book. Random drift is negligible if the population size is large. This is not true for many breeding programs, and random drift can be important. The effect is lowered for quantitative traits which are inherited by several genes. Though the loss of some favourable alleles cannot be avoided if the population size is small, the breeder still can improve the trait if selection occurs for the remaining favourable alleles.

Selection also may be the reason that the Hardy-Weinberg relation is not reached. Since this book deals with selection in plant breeding, some remarks are necessary on selection as a force to change frequencies of alleles in a population. This type of selection plays an important role in the theory of evolution. Normally we assume that all genotypes have the same probability of contributing to the gametic pool of the next generation. If selection occurs this probability is not equal for all alleles. It is said that the genotypes have different fitness values. Genotypes with high fitness values contribute more alleles to the next generation than genotypes with low fitness values. This type of selection is called *natural selection* and should not be confounded with deliberate selection by the breeder, which is called *artificial selection*. Selection in plant breeding means artificial selection. But natural selection cannot be avoided in breeding programs. The breeder also uses selection methods related to natural selection. For example, a breeding population may be inoculated with diseases or parasites so that all nonresistant plants die. This may be regarded as natural selection in an artificial environment. This selection technique has become very important for in vitro cultures. The breeder sometimes also screens material under natural environmental pressures, as for example in a test on winter hardiness under extreme climatic conditions. Natural selection is discussed in more details in chapter 5.

For a population in Hardy-Weinberg equilibrium the frequency of heterozygotes is $2pq$ in the case of two alleles. This value reaches a maximum at $p = q = 0.5$. Then 50 per cent of the genotypes are heterozygous. Figure 1.1 illustrates the relationship between gene frequencies and the distribution of genotypes. The equilateral triangular of de Finetti (1926) is used. The genotypic frequencies are found as follows. Given the allelic frequencies p for B and q for b, the cutting point of the vertical axis at p and q ($p = 0.4$ and $q = 0.6$ in figure 1.1) with the parabola $y^2 = xz$ is M. x, $2y$, and z are the genotypic frequencies for BB, Bb, and bb, respectively (see page 3). The length of the three perpendiculars from M on the three sides are the desired genotypic frequencies x, $2y$, and z. The corresponding genotypes are given in the opposite corners.

If the frequency of one allele is rare, the genotypic frequency of the corresponding homozygous genotype is very rare, but there may be a reasonably high frequency of heterozygotes. This is illustrated in table 1.1 for low frequencies of *b*.

This example shows that the majority of rare alleles is carried by heterozygotes. If a rare allele is completely recessive, its presence is predominantly concealed in phenotypically normal heterozygotes. If all three genotypes are distinguishable it

Table 1.1 Frequencies of heterozygous (*Bb*) and homozygous (*bb*) genotypes if the allele *b* is rare.

frequency of *b*	genotypic frequencies *Bb*	*bb*	ratio of *b* (in heterozygotes *Bb*) to *b* (in homozygotes *bb*)	proportion of the *b* alleles carried by heterozygotes *Bb*
0.1	0.18	0.01	9	0.90
0.05	0.095	0.0025	19	0.95
0.01	0.0198	0.0001	99	0.99
q	$2pq$	q^2	$\frac{p}{q}$	p

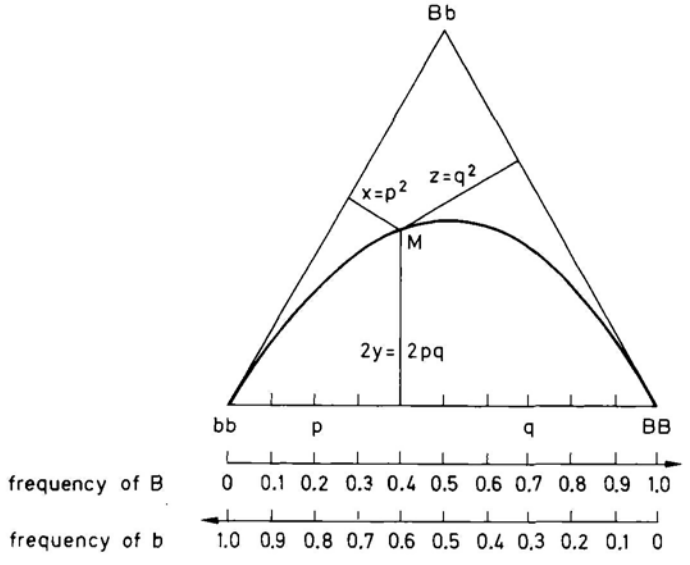

Figure 1.1 Genotypic frequencies as functions of the gene frequencies.

is possible to calculate $\frac{2pq}{\sqrt{p^2q^2}}$. This is a rough way of determing whether a population is in Hardy-Weinberg equilibrium. For completely dominant traits only two phenotypic classes exist. In this case it is possible to estimate the frequency of the recessive allele q by the square root of the proportion of recessive genotypes. This proportion is q^2 if the population is in Hardy-Weinberg equilibrium. With this information it is possible to calculate the frequencies of all genotypes for a Hardy-Weinberg equilibrium population. However, it is not possible to conclude if the population is in Hardy-Weinberg equilibrium, since dominant homozygous genotypes cannot be distinguished from heterozygous genotypes.

Multiple alleles at one locus

The number of different alleles at one locus in a population may be larger than two. With k alleles, there are k homozygous and $\frac{k\,(k-1)}{2}$ heterozygous genotypes. As in the case of two alleles, equilibrium is reached after one generation of random mating. This is easily seen as follows (see also Falconer 1981): We regard one allele and make no distinction between the other $(k-1)$ alleles. We then effectively have a two-allele situation and equilibrium is reached in one generation for the two alleles, as shown previously. Since we can regard each of the alleles in a similar fashion, the result is true for all k alleles.

With multiple alleles the proportion of heterozygous genotypes is increased. This proportion is at maximum if the frequencies of all alleles are equal, i.e. $\frac{1}{k}$. In this case the proportion of the heterozygous genotypes is $\frac{(k-1)}{k}$.

If we denote the frequencies of the alleles $B_i\,(i=1,\ldots,k)$ as $p_i\,(i=1,\ldots,k)$, we can derive genotypic frequencies from gene frequencies in a similar way to that for two alleles. We get

$$(\sum_{i=1}^{k} p_i B_i)^2 = p_1^2 B_1 B_1 + p_2^2 B_2 B_2 + \ldots + p_k^2 B_k B_k + 2p_1 p_2 B_1 B_2 + \ldots + 2p_{k-1} p_k B_{k-1} B_k .$$

Table 1.2 gives an example for three alleles with different frequencies.

Multiple alleles are widespread in natural populations, as was shown in many species for genes which can be made "visible" by electrophoresis of isoenzymes. In this book we are mainly concerned with selection on characters for which no specific genes can be identified. So we do not know how many alleles are present in the population for the loci responsible for the inheritance of quantitative characters. But we know that we can reach equilibrium genotypic frequencies with one generation of random mating even in the case of multiple alleles. This is

Table 1.2 Gene frequencies and genotypic frequencies for 3 alleles at one locus.

gene frequency			genotypic frequencies B_1B_1	B_2B_2	B_3B_3	B_1B_2	B_1B_3	B_2B_3	per cent heterozygotes
p_1	p_2	p_3	p_1^2	p_2^2	p_3^2	$2p_1p_2$	$2p_1p_3$	$2p_2p_3$	
$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	66.7
$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	62.5
0.5	0.4	0.1	0.25	0.16	0.01	0.40	0.10	0.08	58.0
0.8	0.1	0.1	0.64	0.01	0.01	0.16	0.16	0.02	34.0

important, because for many situations discussed in this book, it is assumed that the population is in equilibrium.

Two loci

In the preceding section it was shown, that the "Hardy-Weinberg" relation is attained with one generation of random mating irrespective of the number of alleles at one locus. For two loci, however, equilibrium is not reached in one generation.

Consider two loci, B and G, with two alleles each and denote the frequencies of the alleles B, b, G and g by p, q, r and s, where $p + q = 1$ and $r + s = 1$. p_{BG}, p_{Bg}, p_{bG} and p_{bg} are the frequencies of the four possible gametes. At equilibrium the association between the alleles of the two loci is random, which leads to the following gametic frequencies:

gamete	BG	Bg	bG	bg
frequency	$p_{BG} = pr$	$p_{Bg} = ps$	$p_{bG} = qr$	$p_{bg} = qs$

If the gametic frequencies are different from equilibrium frequencies, the population is in *gametic phase disequilibrium*, measured by the *coefficient of disequilibrium*

$$d = p_{BG}p_{bg} - p_{Bg}p_{bG}.$$

At equilibrium this coefficient is zero, since

$$(p \cdot r) \cdot (q \cdot s) - (p \cdot s) \cdot (q \cdot r) = 0.$$

A numerical example is given in table 1.3. This example indicates that the equilibrium is reached only gradually. The initial population consists of $\frac{1}{3}BBgg$ and $\frac{2}{3}BbGg$ genotypes. The gametic frequencies can be derived from the genotypic frequencies of the same generation in the usual manner from the formulae given

Table 1.3 Genotypic and gametic frequencies and coefficient of disequilibrium.

genotype	frequency	numerical example generation 0	1	2	∞	∞ (general)
BBGG	$f_1 = p_{BG}^2$	0	$\frac{1}{36}$	$\frac{49}{1296}$	$\frac{4}{81}$	p^2r^2
BBGg	$f_2 = 2p_{BG}p_{Bg}$	0	$\frac{6}{36}$	$\frac{238}{1296}$	$\frac{16}{81}$	$2p^2rs$
BBgg	$f_3 = p_{Bg}^2$	$\frac{1}{3}$	$\frac{9}{36}$	$\frac{289}{1296}$	$\frac{16}{81}$	p^2s^2
BbGG	$f_4 = 2p_{BG}p_{bG}$	0	$\frac{2}{36}$	$\frac{70}{1296}$	$\frac{4}{81}$	$2pqr^2$
BbGg	$f_5 = 2p_{BG}p_{bg} + 2p_{Bg}p_{bG}$	$\frac{2}{3}$	$\frac{8}{36}$	$\frac{68}{1296}$	$\frac{16}{81}$	$4pqrs$
Bbgg	$f_6 = 2p_{Bg}p_{bg}$	0	$\frac{6}{36}$	$\frac{238}{1296}$	$\frac{16}{81}$	$2pqs^2$
bbGG	$f_7 = p_{bG}^2$	0	$\frac{1}{36}$	$\frac{25}{1296}$	$\frac{1}{81}$	q^2r^2
bbGg	$f_8 = 2p_{bG}p_{bg}$	0	$\frac{2}{36}$	$\frac{70}{1296}$	$\frac{4}{81}$	$2q^2rs$
bbgg	$f_9 = p_{bg}^2$	0	$\frac{1}{36}$	$\frac{49}{1296}$	$\frac{4}{81}$	q^2s^2
gamete	frequency	generation 0	1	2	∞	∞ (general)
BG	$p_{BG} = (2f_1 + f_2 + f_4 + f_5/2)/2$	$\frac{1}{6}$	$\frac{7}{36}$	$\frac{135}{648}$	$\frac{2}{9}$	pr
Bg	$p_{Bg} = (f_2 + 2f_3 + f_5/2 + f_6)/2$	$\frac{3}{6}$	$\frac{17}{36}$	$\frac{297}{648}$	$\frac{4}{9}$	ps
bG	$p_{bG} = (f_4 + f_5/2 + 2f_7 + f_8)/2$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{81}{648}$	$\frac{1}{9}$	qr
bg	$p_{bg} = (f_5/2 + f_6 + f_8 + 2f_9)/2$	$\frac{1}{6}$	$\frac{7}{36}$	$\frac{135}{648}$	$\frac{2}{9}$	qs
gametic phase disequilibrium d $= p_{BG}p_{bg} - p_{Bg}p_{bG}$		$-\frac{1}{18}$	$-\frac{1}{36}$	$-\frac{1}{72}$	0	0

in the example. The genotypic frequencies of the next generation are derived with the aid of the chess-board method. The coefficient of disequilibrium is halved each generation. From this example it is seen, that from a practical point of view, equilibrium is essentially reached after a few generations.

Two special cases lead to equilibrium in one generation. In the first case all genotypes are double heterozygotes *BbGg*. This is an important case in self-fertilizing species (see chapter 1.3). In the second case the population is built up from the cross of two plants. For example, if *BBGg* is crossed with *BbGG*, we have $p = \frac{3}{4}$, $q = \frac{1}{4}$, $r = \frac{3}{4}$, and $s = \frac{1}{4}$. The genotypic frequencies of the cross are $\frac{1}{4}BBGG, \frac{1}{4}BbGG, \frac{1}{4}BBGg$ and $\frac{1}{4}BbGg$. The gametic frequencies are $p_{BG} = \frac{9}{16}, p_{Bg} = p_{bG} = \frac{3}{16}$ and $p_{bg} = \frac{1}{16}$. Therefore $p_{BG} = p \cdot r$, $p_{Bg} = p \cdot s$, $p_{bG} = q \cdot r$ and $p_{bg} = q \cdot s$ showing that the gametes derived from the cross are in gametic phase equilibrium and consequently the resulting population is in equilibrium.

Linkage

Genes are located on chromosomes (we neglect genes located elsewhere in the plasma). Since the number of different chromosomes is limited and depends on the species, many genes are located on the same chromosome. These genes are linked. The degree of linkage is a function of the distance between loci on the same chromosome and is measured by the probability that alleles of different loci are exchanged during meiosis. This probability is called *recombination frequency*, c. In case of no linkage this value is 0.5, and means an exchange between alleles occurs in 50 per cent of events. This is expected if the genes are located on different chromosomes, or, if at least one crossing-over occurs between two genes on the same chromosome. In case of linkage, c is less than 0.5 with the limiting value of $c = 0$, when no recombination occurs.

Linkage only influences the frequency of types of gametes produced by double heterozygotes. The frequency of gametes produced by $\frac{BG}{bg}$ genotypes is $p_{BG} = p_{bg} = \frac{1-c}{2}$ and $p_{Bg} = p_{bG} = \frac{c}{2}$. The gametic frequencies produced by $\frac{Bg}{bG}$ are $p_{BG} = p_{bg} = \frac{c}{2}$ and $p_{Bg} = p_{bG} = \frac{1-c}{2}$. Gametes of the parental types are more frequent if c is less than 0.5. We have therefore to distinguish between the two types of heterozygotes. Double heterozygotes $\frac{BG}{bg}$ are in *coupling phase* and $\frac{Bg}{bG}$ in *repulsion phase*.

The equilibrium condition is not changed by linkage, but the approach to the equilibrium is delayed. The disequilibrium coefficient is multiplied by $(1 - c)$ each generation. Genotypic and gametic frequencies under linkage for the same genotypic frequencies in generation 0 as in table 1.3 are given in table 1.4. It is

Table 1.4 Linkage and the coefficient of disequilibrium (example of table 1.3 with a recombination frequency of $c = 0.25$).

genotype	frequency	numerical generation	example	
		0	1	2
BBGG	f_1	0	$\frac{9}{144}$	$\frac{1225}{20736}$
BBGg	f_2	0	$\frac{30}{144}$	$\frac{4270}{20736}$
BBgg	f_3	$\frac{1}{3}$	$\frac{25}{144}$	$\frac{3721}{20736}$
BbGG	f_4	0	$\frac{6}{144}$	$\frac{910}{20736}$
BG/bg	f_{5c}	$\frac{2}{3}$	$\frac{18}{144}$	$\frac{2450}{20736}$
Bg/bg	f_{5r}	0	$\frac{10}{144}$	$\frac{1586}{20736}$
Bbgg	f_6	0	$\frac{30}{144}$	$\frac{4270}{20736}$
bbGG	f_7	0	$\frac{1}{144}$	$\frac{169}{20736}$
bbGg	f_8	0	$\frac{6}{144}$	$\frac{910}{20736}$
bbgg	f_9	0	$\frac{9}{144}$	$\frac{1225}{20736}$
gamete	frequency	generation 0	1	2
BG	p_{BG}	$\frac{3}{12}$	$\frac{35}{144}$	$\frac{137}{576}$
Bg	p_{Bg}	$\frac{5}{12}$	$\frac{61}{144}$	$\frac{247}{576}$
bG	p_{bG}	$\frac{1}{12}$	$\frac{13}{144}$	$\frac{55}{576}$
bg	p_{bg}	$\frac{3}{12}$	$\frac{35}{144}$	$\frac{137}{576}$
$d = p_{BG}p_{bg} - p_{Bg}p_{bG}$		$\frac{1}{36} = d_0$	$\frac{1}{48} = (1-c)d_0$	$\frac{1}{64} = (1-c)^2 d_0$

assumed that the two heterozygous genotypes are of the constitution $\frac{BG}{bg}$. Since we must discriminate $\frac{BG}{bg}$ from $\frac{Bg}{bG}$, the frequency f_5 of table 1.3 must be split into two frequencies f_{5c} and f_{5r} for the coupling phase $\frac{BG}{bg}$ and repulsion phase $\frac{Bg}{bG}$. From the gametic frequencies, f_{5c} and f_{5r} are calculated as follows: $f_{5c} = 2p_{BG}p_{bg}$ and $f_{5r} = 2p_{Bg}p_{bG}$.

The gametic frequencies are given by the relations

$$p_{BG} = \frac{2f_1 + f_2 + f_4 + (1-c)f_{5c} + cf_{5r}}{2},$$

$$p_{Bg} = \frac{f_2 + 2f_3 + cf_{5c} + (1-c)f_{5r} + f_6}{2},$$

$$p_{bG} = \frac{f_4 + cf_{5c} + (1-c)f_{5r} + 2f_7 + f_8}{2} \quad \text{and}$$

$$p_{bg} = \frac{(1-c)f_{5c} + cf_{5r} + f_6 + f_8 + 2f_9}{2}.$$

The coefficient of disequilibrium d_t in generation t can be calculated from the coefficient in generation $(t-1)$ or 0 by the formula

$$d_t = (1-c)\,d_{t-1} = (1-c)^t d_0.$$

Figure 1.2 shows the decrease of d_t with t for several degrees of linkage.

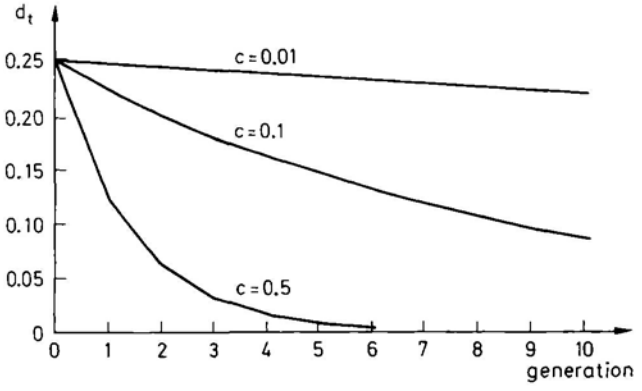

Figure 1.2 Decrease of the disequilibrium coefficient d for two linked loci in successive generations of random mating, starting with genotypes *BBGG* and *bbgg* at equal frequencies.

The number of generations necessary to halve the coefficient of disequilibrium is called *median equilibrium time* $t_{0.5}$ (Crow and Kimura 1970). This is given by

$$t_{0.5} = \frac{\ln 0.5}{\ln (1-c)}.$$

If c is small, $\ln(1-c) \approx -c$ and

$$t_{0.5} \approx \frac{0.693}{c}.$$

For example, for unlinked loci $t_{0.5} = 1$, but for $c = 0.1$ $t \approx 7$, i.e., 7 generations are necessary to halve the disequilibrium. The gametic phase disequilibrium is often called linkage disequilibrium irrespective of whether there is linkage or not.

Multiple alleles at two loci

If there are more than two alleles at one or both loci, the disequilibrium cannot be fully described with one disequilibrium coefficient. Let the number of alleles at locus B be k_B and at locus G be k_G, then $\binom{k_B}{2}\binom{k_G}{2} = \frac{k_B(k_B-1)k_G(k_G-1)}{4}$ disequilibrium coefficients between all possible pairs of alleles exists. For example, with alleles i and j at locus B and m and n at locus G the disequilibrium between the two pairs (i,j) and (m,n) is measured by

$$d_{ij\cdot mn} = p_{im}p_{jn} - p_{in}p_{jm}.$$

At equilibrium all disequilibrium coefficients are zero and the frequencies of gametes can be calculated from the marginal frequencies of the alleles, for example

$$p_{im} = p_i p_m.$$

In a population not at equilibrium the genotypic frequencies can be calculated from the gametic frequencies, and the gametic frequencies of the next generation from the genotypic frequencies. As for two alleles linkage delays the approach but does not change the equilibrium values.

With multiple alleles the change of a single disequilibrium coefficient after one generation of random mating is more complex than for two alleles.

We consider the case of three alleles at each locus in more detail. Let the alleles at the first locus be B, b and β and at the second locus be G, g and γ with a recombination value c between the two loci. Then the disequilibrium coefficient between the pairs Bb and Gg is

$$d_{BG\cdot bg} = p_{BG}p_{bg} - p_{Bg}p_{bG}.$$

After one generation of random mating this coefficient is reduced to

$$d'_{BG \cdot bg} = (1 - c) d_{BG \cdot bg} - c(1 - c)|\mathbf{A}|$$

with $|\mathbf{A}|$ as the determinant of the matrix $\mathbf{A}$ of the gametic frequencies of the previous generation

$$\mathbf{A} = \begin{pmatrix} p_{BG} & p_{Bg} & p_{B\gamma} \\ p_{bG} & p_{bg} & p_{b\gamma} \\ p_{\beta G} & p_{\beta g} & p_{\beta\gamma} \end{pmatrix}$$

(matrices and determinants are explained in the Appendix A 2).

The disequilibrium coefficient is therefore also influenced by the alleles β and γ not directly involved. If there are only two alleles at one locus, a row or column of $\mathbf{A}$ contains zeros. Then $|\mathbf{A}|$ is zero and the change of the disequilibrium coefficients is the same as in case of two alleles at each locus.

More than two loci

If more than two loci are considered, the approach to the equilibrium is very complex. As for two loci, linkage does not affect the equilibrium, but the approach is delayed. At equilibrium the frequency of a specific gamete can be calculated from the marginal frequencies of the corresponding alleles. For example, if there are three loci B, G and H, the frequency of the gamete $B_1 G_3 H_2$ is given by

$$p_{B_1 G_3 H_2} = p_{B_1} \cdot p_{G_3} \cdot p_{H_2} .$$

The genotypic frequencies can be calculated from the gametic frequencies in the same way as was described for two loci. To calculate the gametic frequencies of the next generation from the genotypic frequencies of the previous generation, the linkage relationship between the loci must be known.

Genes on the same chromosome form a linkage group. With three genes B, G, and H, the linkage is fully described by 3 recombination values c_{BG}, c_{BH} and c_{GH}. Figure 1.3 shows a linkage group for 4 genes. At first the fourth locus is not

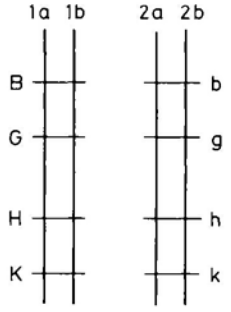

Figure 1.3 Linkage group of 4 genes.

considered. To understand why c_{BH} cannot be calculated from c_{BG} and c_{GH}, it is necessary to describe the recombination process during meiosis, where the chromosome number is halved. A recombination of alleles between loci B and G can occur if there is a crossing-over between two nonsister chromatids. There are four possibilities, 1a with 2a, 1a with 2b, 1b with 2a or 1b with 2b. If a single crossing-over has occurred the probability that a gamete has a recombination of alleles is $\frac{1}{2}$, as in the case of free recombination. For example, if a crossing-over has occured between the chromatids 1b and 2a between B and G, then the gametes BG, Bg, bG and bg are produced in equal proportions. Now the same considerations hold for the loci G and H. A recombination between B and H is possible if there was a crossing-over between B and G *or* G and H. But it is also possible if there were two crossing-overs, one between B and G *and* one between G and H, since different chromatids may be involved. For example, two crossing-overs, the first between 1b and 2a, and the second between 1a and 2a, lead to the recombination probability of $\frac{1}{2}$, while other cases, for example the first and the second crossing-over between 1b and 2a, do not lead to a recombination between B and H. Therefore it is not possible to calculate c_{BH} from c_{BG} and c_{GH}, the recombination values between B and G and between G and H. Furthermore, with four or more loci in a linkage group, all possible linkage parameters for pairs of loci cannot fully explain all recombination values. For k loci there are $\frac{k(k-1)}{2}$ parameters for pairs of loci. At all 2^k different types of gametes exist. Since the sum of all gametic frequencies is 1 and complementary gametic types have the same frequency (see table 1.5), we need $2^{k-1}-1$ parameters. For $k=4$, $\frac{k(k-1)}{2}=6$ and $2^{k-1}-1=7$. The last parameter describes the linkage between pairs of pairs. For six or more loci higher order parameters are also necessary. This concept of describing linkage between several loci in a linkage group was first introduced independently by Jones (1960) and Schnell (1961a). We only describe the case of four loci; and for a general theory the reader is referred to these papers. The parameter p of Jones is equivalent to c used in this book. Schnell used $\lambda = 1-2c$. This approach has some advantages in deriving general formulae. Let us return to four loci (figure 1.3). The six parameters c_{BG}, c_{BH}, c_{BK}, c_{GH}, c_{GK}, and c_{HK} describe the relation between pairs of loci. The last parameter c_{BGHK} is defined so that $1-c_{BGHK}$ is the probability of no recombination between loci B, G, H, and K. The use of these linkage parameters for calculation of the gametic frequencies is demonstrated in table 1.5 for the three loci B to H and the four loci B to K in figure 1.3 (for the calculation of the gametic frequencies see also Jones 1960). In both cases the gametic frequencies sum up to 1.

Table 1.5 Gametic frequencies of the genotype $BGHK/bghk$ from fig. 1.3.

3 loci (B to H)
$p_{BGH} = p_{bgh} = \frac{1}{2} - \frac{c_{BG} + c_{BH} + c_{GH}}{4}$
$p_{BGh} = p_{bgH} = \frac{c_{BH} + c_{GH} - c_{BG}}{4}$
$p_{Bgh} = p_{bGH} = \frac{c_{BG} + c_{BH} - c_{GH}}{4}$
$p_{BgH} = p_{bGh} = \frac{c_{BG} + c_{GH} - c_{BH}}{4}$
4 loci (B to K)
$p_{BGHK} = p_{bghk} = \frac{1}{2} - \frac{c_{BGHK}}{2}$
$p_{BGHk} = p_{bghK} = \frac{c_{BGHK}}{2} - \frac{c_{BG} + c_{BH} + c_{GH}}{4}$
$p_{BGhK} = p_{bgHk} = \frac{c_{BGHK}}{2} - \frac{c_{BG} + c_{BK} + c_{GK}}{4}$
$p_{BgHK} = p_{bGhk} = \frac{c_{BGHK}}{2} - \frac{c_{BH} + c_{BK} + c_{HK}}{4}$
$p_{bGHK} = p_{Bghk} = \frac{c_{BGHK}}{2} - \frac{c_{GH} + c_{GK} + c_{HK}}{4}$
$p_{BGhk} = p_{bgHK} = \frac{c_{BH} + c_{BK} + c_{GH} + c_{GK}}{4} - \frac{c_{BGHK}}{2}$
$p_{BgHk} = p_{bGhK} = \frac{c_{BG} + c_{BK} + c_{GH} + c_{HK}}{4} - \frac{c_{BGHK}}{2}$
$p_{BghK} = p_{bGHk} = \frac{c_{BG} + c_{BH} + c_{GK} + c_{HK}}{4} - \frac{c_{BGHK}}{2}$

One-locus model

In later sections gametic phase equilibrium is normally assumed. Since more than one locus is usually involved, this is not reached with one generation of random mating. Gametic phase equilibrium is also necessary for the so-called *one-locus model.* In this model, several loci may be considered, but the effects are described for each locus individually. The term "one-locus model" is derived from the assumption that no effects exist which arise from specific combinations of loci. If gametic phase equilibrium has not been reached, covariances between the effects of different loci exist which disturb the analysis of the one-locus model.

As for two loci with two alleles so in the general case the approach to equilibrium is more rapid in the first generations of random mating. Therefore, for practical purposes, populations can be regarded as close to equilibrium after a few generations of random mating even in the case of several loci.

Inbreeding

Inbreeding means intermating of individuals who have common ancestors. This has consequences on the genotypic distribution of the offspring. As long as linkage is not considered, it is sufficient to consider only one locus to examine the effects of inbreeding.

The related parents may have the same replicate of an allele of one of the common ancestors. Then there is a specified probability that both parents may transmit this allele to the offspring. In this case the offspring has two identical alleles which are replicates of the same allele of the common ancestor, and we say that the two alleles are *identical by descent*. For example, this allele may be B, and the offspring might have the genotype BB. There may be other individuals in the population with genotype BB where these alleles are not identical by descent but only *alike in state*. The following considerations may clarify what is meant by these definitions.

The number of possible ancestors depends on the number of generations traced back into the past. With every generation the number of ancestors is doubled. Ten generations back each individual has $2^{10} = 1024$ ancestors, 20 generations back about 1 million. Therefore, each individual must have related ancestors. The definitions given only make sense if we define a *base population* in which no individuals with alleles identical by descent exist, and the alleles of all homozygous genotypes are regarded only as alike in state.

Inbreeding coefficient

The effect of inbreeding is measured by the *inbreeding coefficient F*. F is the probability that two alleles at a locus of an individual are identical by descent (Malecot, 1948). The inbreeding coefficient depends on the mating system. In animals, related individuals can be mated. Most plants can also be self-pollinated and this is the strongest form of inbreeding.

There is another measurement which is closely connected with the inbreeding coefficient. This is the probability that two individuals carry alleles at a locus which are identical by descent. This probability was called the "coefficient de parenté" (Malecot, 1948). Kempthorne (1957) describes this as the *coefficient of parentage*, but Falconer (1981) uses the word *coancestry*. This is best explained by example. We regard individual X with alleles x_1 and x_2 and individual Y with alleles y_1 and y_2 and calculate the probability $P(x_i = y_j)$, where x_i may be x_1 or x_2 and y_j may be y_1 or y_2. We get

$$\varphi_{xy} = \frac{1}{4}[P(x_1 = y_1) + P(x_1 = y_2) + P(x_2 = y_1) + P(x_2 = y_2)]$$

where φ_{xy} is the coefficient of coancestry. The connection between the inbreeding coefficient F and the coefficient of coancestry is given by

$$F(\text{individual}) = \varphi(\text{parents of the individual}).$$

Two methods are proposed to find the inbreeding coefficient of an individual in the general case. The first method uses path coefficients in the pedigree, the other method uses the coefficient of coancestry and explains the coefficient between individuals in generation t by the coefficients in generation $t-1$. Both methods are explained with examples.

Path coefficients in the pedigree

We would like to compute the inbreeding coefficient of an individual Z that has parents A and B. All paths from A back to a common ancestor of A and B and forward to B must be considered. Figure 1.4 gives a hypothetical pedigree. I is a

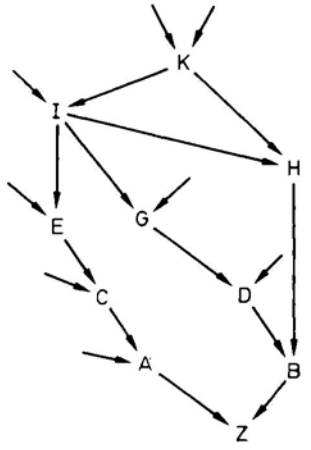

Figure 1.4 Pedigree diagram.

common ancestor of A and B, and one path is $ACEIGDB$. The contribution of this path is given by $(\frac{1}{2})^n = (\frac{1}{2})^7$, n being the number of genotypes in this path. It is assumed that I is not inbred. The probability, that I transmits the same allele to E and G is $\frac{1}{2}$, the probability that E transmits the allele received from I to C is also $\frac{1}{2}$ and so on. The inbreeding coefficient is the sum of the contributions of all paths. The three possible paths are $ACEIGDB$, $ACEIHB$ and $ACEIKHB$.

If the common ancestor itself is inbred, the contribution of the corresponding path to the inbreeding coefficient is increased. We again regard the path

$ACEIGDB$ and assume that the inbreeding coefficient of I is F_I. The probability that I transmits the same allele to E and G is $\frac{1}{2}$. If I transmits different alleles (this probability also is $\frac{1}{2}$), they are identical by descent with the probability F_I. Therefore the total probability that I transmits alleles which are identical by descent to E and G is $\frac{1}{2}(1 + F_I)$ and the contribution of the path therefore is $(\frac{1}{2})^7(1 + F_I)$. In a similar way the contribution of the path $ACEIHB$ is $(\frac{1}{2})^6(1 + F_I)$ and of the path $ACEIKHB$ is $(\frac{1}{2})^7(1 + F_K)$. The results are summarized in table 1.6.

Table 1.6 Calculation of the inbreeding coefficient of Z with paths (fig. 1.4, $F_I = 0$).

path	n	inbreeding coefficient of K 0	0.25	F_k
$ACEIGDB$	7	$\frac{1}{128}$	$\frac{1}{128}$	$\frac{1}{128}$
$ACEIHB$	6	$\frac{1}{64}$	$\frac{1}{64}$	$\frac{1}{64}$
$ACEIKHB$	7	$\frac{1}{128}$	$\frac{1+0.25}{128}$	$\frac{1+F_k}{128}$
total		$\frac{1}{32} = 0.03125$	$\frac{1}{32} + \frac{0.25}{128} = 0.03320$	$\frac{1}{32} + \frac{F_k}{128}$

Coefficient of coancestry

The coefficient of coancestry is explained using the pedigree in figure 1.5. The inbreeding coefficient of Z is equal to the coefficient of coancestry φ_{XY} of X and Y. X or Y or both can be replaced by their parents in the following way:

$$\varphi_{XY} = \frac{1}{2}(\varphi_{AY} + \varphi_{BY}) = \frac{1}{2}(\varphi_{XC} + \varphi_{XD}) = \frac{1}{4}(\varphi_{AC} + \varphi_{AD} + \varphi_{BC} + \varphi_{BD}).$$

We consider an allele of X and ask if this allele is identical by descent with an allele of Y. The probability, that the allele of X has come from parent A is $\frac{1}{2}$. The

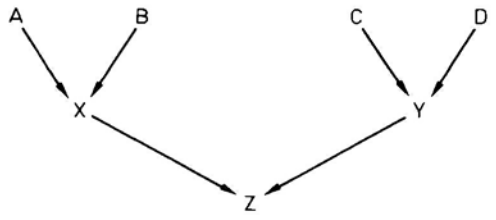

Figure 1.5 Pedigree diagram for the coefficient of coancestry.

probability that the allele of X has come from B also is $\frac{1}{2}$. Only one half of the alleles of A and B have been transmitted to X. Therefore $\varphi_{XY} = \frac{1}{2}(\varphi_{AY} + \varphi_{BY})$. The other relationships are derived in a similar way.

With the relation

$$\varphi_{XY} = \frac{1}{4}(\varphi_{AC} + \varphi_{AD} + \varphi_{BC} + \varphi_{BD})$$

we can express the coefficient of coancestry of the actual generation (X and Y) in terms of the preceding generation (A, B, C and D). This can be used to construct similar formulae for the inbreeding coefficient. Two coefficients of coancestry are of a special interest: φ_{XX} and φ_{AX}. φ_{XX} is the coefficient of coancestry of a genotype with itself. We regard one allele and ask for the probability that a randomly chosen allele is identical by descent with it. The probability of choosing the same allele is $\frac{1}{2}$. If the other allele is chosen this probability also is $\frac{1}{2}$, the probability that it is identical by descent is F_X. Therefore

$$\varphi_{XX} = \frac{1}{2}(1 + F_X).$$

φ_{AX} is the coefficient of coancestry of a genotype with its offspring. From figure 1.5 it can be shown that this coefficient is

$$\varphi_{AX} = \frac{1}{2}(\varphi_{AA} + \varphi_{AB}).$$

The coefficient of coancestry can be used to derive recurrence equations for the inbreeding coefficients of regular mating systems to produce inbred genotypes. The two most important systems are self-fertilization and full-sib mating. The increase of the inbreeding coefficient is more rapid in case of self-fertilization, but this procedure only can be used in plants and if self-fertilization is not prohibited by a natural mechanism like self-incompatibility. For both systems the inbreeding coefficient is given in terms of the inbreeding coefficients of previous generations.

a) Self-fertilization

If A is the parent in generation $(t-1)$ and Z the offspring in generation t, we get

$$F_Z = \varphi_{AA} = \frac{1}{2}(1 + F_A)$$

and the recurrence equation

$$F_t = \frac{1}{2}(1 + F_{t-1}).$$

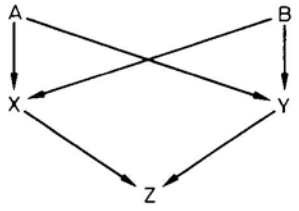

Figure 1.6 Pedigree diagram for full-sib mating.

b) Full-sib mating

To calculate the inbreeding coefficient after full-sib mating we need the inbreeding coefficients of the two preceding generations. The pedigree for full sibs is given in figure 1.6. From this figure we find

$$F_Z = \varphi_{XY} = \frac{1}{4}(\varphi_{AA} + 2\varphi_{AB} + \varphi_{BB})$$

$$= \frac{1}{4}(1 + \tfrac{1}{2}F_A + \tfrac{1}{2}F_B + 2F_X),$$

since $\varphi_{AB} = F_X$. Therefore we get the recurrence equation

$$F_t = \frac{1}{4}(1 + 2F_{t-1} + F_{t-2}).$$

Table 1.7 gives the inbreeding coefficient for the first generations under selfing and full-sib mating, for an initial inbreeding coefficient, F_0, of zero.

Table 1.7 Inbreeding coefficient after continuous selfing or full-sib mating.

generation	selfing	full-sib mating
0	0	0
1	0.5	0.25
2	0.75	0.375
3	0.875	0.500
4	0.938	0.594
5	0.969	0.672
10	0.999	0.886
20	1.000	0.986

The complement of the inbreeding coefficient is called the *panmictic index P* (Wright, 1951) with $P = 1 - F$. Sometimes the formulae are simplified if the panmictic index is used. So, for selfing we get

$$P_t = \frac{1}{2}P_{t-1} = \left(\frac{1}{2}\right)^t \tag{1.1}$$

showing, that the percentage of heterozygosity is halved every generation.

So far we have considered only one locus. The inbreeding coefficient is the same for all loci. Regarding all loci together, the inbreeding coefficient gives the expected percentage of loci of a genotype which are identical by descent. This also holds in the case of linkage.

Finally, we consider the genotypic frequencies for a population with an inbreeding coefficient F. In the case of two alleles homozygous genotypes BB and bb may carry alleles identical by descent with probabilities pF and qF. The probability is $1 - F$ that the genotypes do not carry alleles identical by descent. Therefore the total frequencies of the three genotypes are

$$pF + p^2(1 - F) = p^2 + pqF \quad \text{for} \quad BB,$$

$$2pq(1 - F) \text{ for } Bb,$$

$$qF + q^2(1 - F) = q^2 + pqF \quad \text{for} \quad bb.$$

In the general case of k alleles we find $p_iF + p_i^2(1 - F) = p_i^2 + p_i(1 - p_i)F$ for homozygous genotypes B_iB_i and $2p_ip_j(1 - F)$ for heterozygous genotypes B_iB_j.

Linked loci

We now may ask, what is the probability that two loci are simultaneously identical by descent, if the inbreeding coefficient for each locus is F? In case of unlinked loci this probability is F^2. But for linked loci no general form can be given. For two linked loci Haldane (1949) introduced a generalized inbreeding coefficient. Schnell (1961a) extended this principle to an arbitrary number of linked loci and called the probability that several loci are identical by descent simultaneously the *function of inbreeding*. For linked loci this probability is higher than for unlinked loci. The function of inbreeding plays a role in the covariance between relatives when there is interaction between loci. We come back to this problem in chapter 3.2. Here we only can say that the function of inbreeding of two linked loci is not the same if the same inbreeding coefficient is reached by different mating systems, as for example by selfing or full- sib mating. The reason is that with more generations there is more opportunity for linked loci to recombine. Gallais (1974) introduced the term *link*. The link g_t is the probability that two alleles from different loci are transmitted simultaneously over t generations. The link depends on the degree of linkage between corresponding loci, but is not zero for unlinked loci.

Hybridization

The opposite effect of inbreeding is achieved with hybridization. While with inbreeding the precentage of heterozygotes is reduced, with hybridization this percentage can be increased to a higher level than in an equilibrium population. If we consider one locus and cross only *BB* genotypes with *bb* genotypes, the whole population is completely heterozygous. Another way is to cross a single genotype to a sample of individuals from a population. If the population has a gene frequency of 0.2 for *B* and 0.8 for *b*, the population derived from a cross of this population with a single genotype *BB* contains 80 percent heterozygous genotypes *Bb*. Finally, if we cross individuals from two populations with different gene frequencies, the hybrid population also can contain more heterozygotes than expected under equilibrium conditions. If the frequency of the *B* allele is p_1 in population 1 and p_2 in population 2, and if q_1 and q_2 are the corresponding frequencies of *b*, the hybrid population contains the genotypes *BB*, *Bb* and *bb* in the frequencies p_1p_2, $p_1q_2 + p_2q_1$ and q_1q_2. The cross of two homozygous genotypes and of a homozygous genotype with a population can be regarded as limiting cases with the fixation of one allele in both or one population. Table 1.8

Table 1.8 Frequency of heterozygotes in a hybrid population (p_1 and p_2 are the frequencies of the *B* allele in the two initial populations, in brackets after one generation of random mating).

p_1 \ p_2	0	0.5	0.8	1.0	p_2
0	0.0(0.0)	0.5(0.375)	0.8(0.48)	1.0(0.5)	$p_2(p_2 - 0.5p_2^2)$
0.2	0.2(0.18)	0.5(0.455)	0.68(0.5)	0.8(0.48)	$0.2 + 0.6p_2(0.18 + 0.8p_2 - 0.5p_2^2)$
0.5	0.5(0.375)	0.5(0.5)	0.5(0.455)	0.5(0.375)	$0.5(0.375 + 0.5p_2 - 0.5p_2^2)$
1	1.0(0.5)	0.5(0.375)	0.2(0.18)	0.0(0.0)	$1 - p_2(0.5 - 0.5p_2^2)$
p_1	p_1 $(p_1 - 0.5p_1^2)$	0.5 $(0.375) + 0.5p_1 - 0.5p_1^2)$	$0.8 - 0.6p_1$ $(0.48 + 0.2p_1 - 0.5p_1^2)$	$1 - p_1$ $(0.5 - 0.5p_1^2)$	$p_1 + p_2 - 2p_1p_2$ $(p_1 + p_2 - 0.5(p_1 + p_2)^2)$

gives the frequency of heterozygotes for several values of p_1 and p_2. The frequency of heterozygotes after one generation of random mating is added in parenthesis. The hybrid population has then reached the Hardy-Weinberg equilibrium. From this table it is seen that the difference between the frequency in a hybrid population and at the Hardy-Weinberg equilibrium is always positive, since

$$(p_1 + p_2 - 2p_1p_2) - \left(p_1 + p_2 - \frac{(p_1 + p_2)^2}{2}\right) = \frac{(p_1 - p_2)^2}{2}.$$

Small populations

Two properties of small populations have considerable importance in population and quantitative genetics: (1) random drift and (2) inbreeding. Random drift is the fluctuation in gene frequencies caused by chance or sampling variation. The direction of drift is unpredictable. Random drift may lead to a loss of alleles, either desirable or undersirable. The breeder can diminish the danger of loosing favourable alleles by using large population sizes. The reader who is interested in random drift in connection with evolution, is referred to other textbooks (for example Li 1955, Crow and Kimura).

Populations built up from a small number of genotypes are inbred. The average inbreeding coefficient is $\frac{1}{2N}$, where N is the number of genotypes. Consider a population of N individuals, which are not inbred and not related. The probability of self-fertilization is $\frac{1}{N}$ if there is no incompatibility system which prohibits selfing. Since the inbreeding coefficient after selfing is $\frac{1}{2}$, the average inbreeding coefficient is $\frac{1}{2N}$. For example, if a population is built up from 10 individuals, the average inbreeding coefficient is 0.05. It should be noted that the average inbreeding coefficient is not reduced if the population size is larger in later generations. This is very important, since the population often is very small in early stages of a breeding program. The stage at which the population size is the smallest is often called a bottleneck. At this stage also the probability of losing favourable alleles by random drift is maximum.

We now consider the average inbreeding coefficient after several generations of random mating in a small population. We first assume that the population size is constant N in each generation. Consider the second generation. Again the probability that genes identical by descent form a zygote is $\frac{1}{2N}$. For the remaining portion the probability of being identical by descent is given by the inbreeding coefficient of the preceding generation. For the second generation we get

$$F_2 = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) F_1$$

and generally

$$F_t = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) F_{t-1}.$$

Using the panmictic index $P = 1 - F$ we get

$$P_t = \left(1 - \frac{1}{2N}\right) P_{t-1} = \left(1 - \frac{1}{2N}\right)^t P_0.$$

Inbreeding may also occur if selfing is not possible. This is the situation in plant species with a self-incompatibility system. In the first generation no inbreeding occurs. But in the second generation for all genes not linked to the incompatibility loci the inbreeding coefficient is $\frac{1}{2N}$, where N is the population size of the first generation. Therefore self-incompatibility delays the inbreeding only by one generation. The effect of different sexes is similar. But here the situation can be more complicated if the ratio of males to females is not equal. We will not discuss this point here.

If the number of individuals is not the same in subsequent generations, the *effective population size* N_e must be considered. N_e is the harmonic mean and is defined by

$$\frac{1}{N_e} = \frac{1}{t}\left(\frac{1}{N_1} + \frac{1}{N_2} + \ldots + \frac{1}{N_t}\right).$$

N_e is more influenced by generations with lower numbers of individuals and is smaller than the arithmetic mean of the N_t. Therefore the inbreeding coefficient is higher if, with the same total number of individuals, the number of individuals is not constant over generations. This is shown in table 1.9. This also emphasizes the importance of "bottleneck" on the genetic structure of populations subsequent to a severe reduction in population size.

Table 1.9 Effective population size and inbreeding coefficient.

generation t	equal size N_t	unequal size N_t
1	10	8
2	10	3
3	10	20
4	10	9
N_e	10	6.46
inbreeding coefficient after generation 4	$1-\left(1-\frac{1}{2\cdot 10}\right)^4$ $= 0.1855$	$1-\left(1-\frac{1}{2\cdot 6.46}\right)^4$ $= .2756$

The calculation of the inbreeding coefficient with the effective population size is only an approximation. In the example above the exact inbreeding coefficient is $F_1 = \frac{1}{16} = 0.0625$ after generation 1, $F_2 = \frac{1}{6} + \frac{5}{6} F_1 = 0.2188$ after generation 2, $F_3 = \frac{1}{40} + \frac{39}{40} F_2 = 0.2383$ after generation 3 and $F_4 = \frac{1}{18} + \frac{17}{18} F_3 = 0.2806$ after generation 4. This is in good agreement with the inbreeding coefficient calculated from the effective population size.

1.3 Self-fertilizing species

In self-fertilizing species selfing is the natural reproductive mechanism. Therefore, natural populations consist of several to many nearly homozygous lines. Crosses between them occur, but generally at low frequencies. The exchange of genetic material between individuals is reduced and restricted to such cross-pollinations. The behaviour of such populations is similar to that of vegetatively propagated species because homozygous genotypes also reproduce only identical genotypes on selfing.

For breeding purpose the breeder builds up populations from crosses between homozygous lines. The cross between two homozygous lines is called the F_1-generation or more simply the F_1. (The symbol F should not be confused with the inbreeding coefficient of the preceding chapter 1.2.) The F_1-generation is homogeneous and has the highest degree of heterozygosity. For example, if we cross the homozygous lines *BBccdd* with *bbCCDD*, the genotypes of the F_1 are all of the form *BbCcDd*. To get genetic variation, it is necessary to produce the F_2 by selfing the F_1. For one locus we now have the segregation 1 *BB* : 2 *Bb* : 1 *bb*. Therefore in the F_2-generation 50 per cent of the genotypes are heterozygous for each locus at which the parental lines differed. The other 50 per cent are homozygous, and the alleles are identical by descent (see page 16). Formally an F_2-generation has an inbreeding coefficient of 0.5, but in self-fertilizing species this concept is not generally followed. In cases in which the concept of inbreeding coefficients is used in self-fertilizing species, the F_2-population is regarded as the base population with noninbred individuals, i.e. $F = 0$.

Another property of the F_2-generation is important. At each segregating locus the population is in Hardy-Weinberg equilibrium. There are only two alleles with gene frequencies of $p = q = 0.5$. As we will see in chapter 2, many formulae are simplified under these conditions.

Since reproduction is by selfing, we can use the formula (1.1) on page 21 for the panmictic index, to get the percentage of heterozygotes in generation F_t, if we bear in mind, that in the generation F_1 the panmictic index $P_1 = 1$:

$$P_t = \left(\frac{1}{2}\right)^{t-1}$$

Gene frequencies are not changed, but the number of homozygous genotypes increases with t. Therefore the population can no longer be regarded as a population in Hardy-Weinberg equilibrium. The true equilibrium is reached when all genotypes are homozygous, and this equilibrium is reached slowly even for one locus.

Linkage

We now have to consider two linked loci. For a random mating population we saw that the coefficient of disequilibrium is a function of the genotypic frequencies in the initial population, the number of generations of random mating and the recombination value. For unlinked loci the gametic output of the F_1-generation is in gametic phase equilibrium. For example, if we cross $BBGG \times bbgg$, the F_1 consists only of *BbGg* genotypes and produces gametes *BG*, *Bg*, *bG*, and *bg* in equal proportions. This is no longer true, if *B* and *G* are linked. In terms of the *recombination coefficient*, *c*, the relative frequencies are

$$\frac{1-c}{2}BG : \frac{c}{2}Bg : \frac{c}{2}bG : \frac{1-c}{2}bg\,.$$

The ratio of parental types to exchange types is $\frac{1-c}{c}$. In case of complete linkage only the parental gametes *BG* and *bg* are produced.

Table 1.10 gives the relative frequencies of genotypes in F_2 for several recombination frequencies. We now consider later generations. The formulae for F_2, F_3, F_4 and F_∞ are given in table 1.11. Often the breeder makes a cross to combine genes which are closely linked from two parents. If linkage is strong, homozygous genotypes having the desired new combination of genes are rare as can be seen in table 1.10. In such cases the progeny must be large enough to find at least one recombinant. Finally, we note that in case of dominance of *B* over *b* and *G* over *g* the gametes *BG* and *bg* are in *coupling phase* and the gametes *Bg* and *bG* in *repulsion phase*. Homozygotes *BBGG* and *bbgg* can only produce gametes in coupling phase and homozygotes *BBgg* and *bbGG* gametes in repulsion phase. Heterozygotes *BbGg* may be the result of the union of coupling phase gametes

Table 1.10 Genotypic frequencies in F_2 for various recombination frequencies (c), the parents being *BBGG* and *bbgg*.

genotype	$c = 0.5$	0.2	0.1	0	c
BBGG or *bbgg*	0.0625	0.16	0.2025	0.25	$\frac{(1-c)^2}{4}$
BBgg or *bbGG*	0.0625	0.01	0.0025	0	$\frac{c^2}{4}$
BbGG, *Bbgg*, *BBGg* or *bbGg*	0.1250	0.08	0.0450	0	$\frac{c(1-c)}{2}$
BG/*bg*	0.1250	0.32	0.4050	0.5	$\frac{(1-c)^2}{2}$
Bg/*bG*	0.1250	0.02	0.0050	0	$\frac{c^2}{2}$

Table 1.11 Genotypic frequencies in F_2, F_3, F_4 and F_∞ for two linked loci with recombination value c.
a) $BBGG \times bbgg$; b) $BBgg \times bbGG$.

genotypes	F_2	F_3	F_4	F_∞
BBGG or *bbgg*	a) $\frac{(1-c)^2}{4}$	$k_1 = \frac{3-6c+6c^2-4c^3+2c^4}{8}$	$k_1 + \frac{k_3}{2} + \frac{(1-c)^2}{4} k_4 + \frac{c^2}{4} k_5$	$\frac{1}{2(1+2c)}$
	b) $\frac{c^2}{4}$	$k_2 = \frac{c+c^2-2c^3+c^4}{4}$	$k_2 + \frac{k_3}{2} + \frac{c^2}{4} k_4 + \frac{(1-c)^2}{4} k_5$	$\frac{2c}{2(1+2c)}$
BBgg or *bbGG*	a) $\frac{c^2}{4}$	$k_2 = \frac{c+c^2-2c^3+c^4}{4}$	$k_2 + \frac{k_3}{2} + \frac{c^2}{4} k_4 + \frac{(1-c)^2}{4} k_5$	$\frac{2c}{2(1+2c)}$
	b) $\frac{(1-c)^2}{4}$	$k_1 = \frac{3-6c+6c^2-4c^3+2c^4}{8}$	$k_1 + \frac{k_3}{2} + \frac{(1-c)^2}{4} k_4 + \frac{c^2}{4} k_5$	$\frac{1}{2(1+2c)}$
BBGg, *bbGg*, *BbGG* or *Bbgg*	$\frac{c(1-c)}{2}$	$k_3 = \frac{c-2c^2+2c^3-c^4}{2}$	$\frac{k_3}{2} + \frac{c(1-c)}{2}(k_4+k_5)$	0
BG/bg	a) $\frac{(1-c)^2}{2}$	$k_4 = \frac{1-4c+6c^2-4c^3+2c^4}{4}$	$\frac{(1-c)^2}{2} k_4 + \frac{c^2}{2} k_5$	0
	b) $\frac{c^2}{2}$	$k_5 = \frac{c^2-2c^3+c^4}{2}$	$\frac{c^2}{2} k_4 + \frac{(1-c)^2}{2} k_5$	0
Bg/bG	a) $\frac{c^2}{2}$	$k_5 = \frac{c^2-2c^3+c^4}{2}$	$\frac{c^2}{2} k_4 + \frac{(1-c)^2}{2} k_5$	0
	b) $\frac{(1-c)^2}{2}$	$k_4 = \frac{1-4c+6c^2-4c^3+2c^4}{4}$	$\frac{(1-c)^2}{2} k_4 + \frac{c^2}{2} k_5$	0

$\left(\frac{BG}{bg}\right)$ or of repulsion phase gametes $\left(\frac{Bg}{bG}\right)$. The two types of heterozygotes $\left(\frac{BG}{bg}\right.$ and $\left.\frac{Bg}{bG}\right)$ produces gametes of the coupling type and of the repulsion type in the ratio of $(1-c):c$ and $c:(1-c)$ respectively. The selfed progeny or the backcross to the *bbgg* genotype of the double heterozygotes are used to estimate the recombination value c (see for example Bailey 1961 or Srb et al. 1965).

More than two loci

For one locus 3 genotypes are possible in F_2 and subsequent selfing generations. In F_∞ finally no heterozygotes exist and two types remain. Now, if we have k loci, the number of possible genotypes in F_2 simply is 3^k. The breeder of line varieties is searching for the best homozygous combination. With one locus, this best genotype has a probability of $\frac{1}{4}$ in F_2, $\frac{2^{t-1}-1}{2^t}$ in F_t and $\frac{1}{2}$ in F_∞. Therefore the probability of finding the best homozygous combination for k unlinked loci, is $(\frac{1}{4})^k$ in F_2, $\left(\frac{2^{t-1}-1}{2^t}\right)^k$ in F_t and $(\frac{1}{2})^k$ in F_∞. This chance is extremely low, even in F_∞, if k is not very small. If the breeder makes selections in several steps, he may be content, not to lose a favourable allele. This probability is $\frac{3}{4}$ for one locus in F_2, $\frac{2^{t-1}+1}{2^t}$ in F_t and again $\frac{1}{2}$ in F_∞, since in F_∞ there are no heterozygotes. As can be seen, this probability is somewhat better in early generations, but also low if many loci are

Table 1.12 Probability of desired genotypes in self-fertilizing species for k loci.

k	F_2 *a*	F_2 *b*	F_3 *a*	F_3 *b*	F_4 *a*	F_4 *b*	F_∞
1	0.25	0.75	0.375	0.625	0.4375	0.5625	0.5
2	0.0625	0.5625	0.141	0.391	0.1914	0.3164	0.25
3	$1.6 \cdot 10^{-2}$	0.422	0.053	0.244	0.0837	0.1780	0.125
4	$3.9 \cdot 10^{-3}$	0.316	0.020	0.153	0.0366	0.1001	0.0625
5	$9.8 \cdot 10^{-4}$	0.237	$7.4 \cdot 10^{-3}$	0.095	0.0160	0.0563	0.0313
10	$9.5 \cdot 10^{-7}$	0.056	$5.5 \cdot 10^{-5}$	$9.1 \cdot 10^{-3}$	$2.6 \cdot 10^{-4}$	$3.2 \cdot 10^{-3}$	$9.8 \cdot 10^{-4}$
20	$9.1 \cdot 10^{-13}$	$3.2 \cdot 10^{-3}$	$3.0 \cdot 10^{-9}$	$8.3 \cdot 10^{-5}$	$6.6 \cdot 10^{-8}$	$1.0 \cdot 10^{-5}$	$9.5 \cdot 10^{-7}$
50	$7.9 \cdot 10^{-31}$	$5.7 \cdot 10^{-7}$	$5.0 \cdot 10^{-22}$	$6.2 \cdot 10^{-11}$	$1.1 \cdot 10^{-18}$	$3.2 \cdot 10^{-13}$	$8.9 \cdot 10^{-16}$
100	$6.2 \cdot 10^{-61}$	$3.2 \cdot 10^{-13}$	$2.5 \cdot 10^{-43}$	$3.9 \cdot 10^{-21}$	$1.3 \cdot 10^{-36}$	$1.0 \cdot 10^{-25}$	$7.9 \cdot 10^{-31}$
k	$(\frac{1}{4})^k$	$(\frac{3}{4})^k$	$(\frac{3}{8})^k$	$(\frac{5}{8})^k$	$(\frac{7}{16})^k$	$(\frac{9}{16})^k$	$(\frac{1}{2})^k$

a: best homozygous type
b: no desired allele lost

considered. If there is linkage between the loci, the chance for some combinations is higher and reduced for other combinations.

Table 1.12 shows the probability, to find the desired types for k unlinked loci. From this table it can be seen that for more than 10 loci no real chance exists of finding the desired genotype.

Partial selfing

Some plant species like *Brassica napus* (rape) and *Vicia faba* (field beans) show self- and cross-fertilization simultaneously. For these species there is a continous exchange of genetic material within the population. We treat these species under self-fertilizing species, since they are bred like self-fertilizing crops. In this case cross-fertilization must be prevented in breeding programs. Natural populations show behaviour more similar to cross-breeding species, and an equilibrium is reached with homozygous and heterozygous plants. Compared with cross-breeding species, the percentage of heterozygotes in equilibrium populations is reduced. We consider only one locus. The equilibrium is reached only gradually. We assume that the amount of selfing is constant over generations and is denoted by s. The limiting values of s are 1 for self-fertilizing and 0 for cross-fertilizing species.

On page 19 the recurrency relation for continuous selfing was given. The inbreeding coefficient F_t can be calculated from the inbreeding coefficient F_{t-1}. These inbreeding coefficients should not be confused with the designation for generations after a cross of homozygous lines which are also denoted by F in the literature. The relation is

$$F_t = \frac{1}{2}(1 + F_{t-1}).$$

If the proportion s is self-fertilized, this relation is

$$F_t = \frac{s}{2}(1 + F_{t-1}).$$

If the inbreeding coefficient of the initial population is F_0, the inbreeding coefficient after t generations following Kempthorne (1957) is

$$F_t = \frac{s}{2-s}\left[1 - \left(\frac{s}{2}\right)^t\right] + \left(\frac{s}{2}\right)^t \cdot F_0.$$

From this formula it is seen, that equilibrium is approached slowly and that in the limit as t tends to ∞

$$F_\infty = \frac{s}{2-s}.$$

Note that the inbreeding coefficient of each generation, and therefore also at equilibrium, depends only on the rate of self-fertilization, but not on the gene frequencies.

If two alleles B and b with frequencies p and q are regarded, the genotypic frequencies at equilibrium are (Wright 1921)

$$p^2 + pqF = p^2 + pq \cdot \frac{s}{2-s} \quad \text{for } BB \text{ genotypes}$$

$$2pq(1-F) = 2pq\left(1 - \frac{s}{2-s}\right) \quad \text{for } Bb \text{ genotypes and}$$

$$q^2 + pqF = q^2 + pq \cdot \frac{s}{2-s} \quad \text{for } bb \text{ genotypes.}$$

These equations show that with partial self-fertilization, the frequencies of the two homozygous genotypes are increased by the same amount and the frequency of the heterozygous genotypes is reduced.

The limiting values of s lead to the Hardy-Weinberg equilibrium ($s = 0$) or to a ratio of $p:q$ for BB and bb homozygotes ($s = 1$).

The general equilibrium for arbitrary values of s was given by Wright (1921) and therefore sometimes is called "Wright's equilibrium law" (Li 1955). Table 1.13 gives the inbreeding coefficient and the frequency of heterozygotes at equilibrium for various values of s and p.

Table 1.13 Inbreeding coefficient (F) and frequency of heterozygotes for various frequencies of genes (p) and degrees of selfing (s).

s	F	frequency of heterozygotes		
		$p = 0.1$	$p = 0.3$	$p = 0.5$
0	0	0.18	0.42	0.50
0.2	0.11	0.16	0.37	0.44
0.5	0.33	0.12	0.28	0.33
0.8	0.67	0.06	0.14	0.17
0.9	0.82	0.03	0.08	0.09
1.0	1.00	0	0	0

1.4 Autotetraploid species

A diploid genotype contains two homologous sets of chromosomes, one from the male parent and one from the female parent. During meiosis, gametes with one homologous set of chromosomes are formed, and the parental chromosomes are