Selection Methods in Plant Breeding

Selection Methods in Plant Breeding

2nd Edition

by

Izak Bos

University of Wageningen, The Netherlands

and

Peter Caligari

University of Talca, Chile



A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 978-1-4020-6369-5 (HB) ISBN 978-1-4020-6370-1 (e-book)

Published by Springer, P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Cover photo: Bagging of the inflorescence of an oil palm

Printed on acid-free paper

© 2008 Springer Science + Business Media B.V.

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Contents

\mathbf{P}_{1}	refac	æ	ix
\mathbf{P}	refac	te to the 2nd Edition	xi
1	Int	roduction	1
2	Pop	pulation Genetic Effects of Cross-fertilization	7
	2.1	Introduction	7
	2.2	Diploid Chromosome Behaviour and Panmixis	10
		2.2.1 One Locus with Two Alleles	10
		2.2.2 One Locus with more than Two Alleles	15
		2.2.3 Two Loci, Each with Two Alleles	16
		$2.2.4$ More than Two Loci, Each with Two or more Alleles $\ .$.	26
	2.3	Autotetraploid Chromosome Behaviour and Panmixis	28
3	Poj	pulation Genetic Effects of Inbreeding	33
	3.1	Introduction	33
	3.2	Diploid Chromosome Behaviour and Inbreeding	37
		3.2.1 One locus with two alleles	37
		3.2.2 A pair of linked loci	41
		3.2.3 Two or more unlinked loci, each with two alleles	49
	3.3	Autotetraploid Chromosome Behaviour	
		and Self-Fertilization	52
	3.4	Self-Fertilization and Cross-Fertilization	56
4	Ass	sortative Mating and Disassortative Mating	59
	4.1	Introduction	59
	4.2	Repeated Backcrossing	63
5	Poj	pulation Genetic Effect of Selection with regard	
	to	Sex Expression	69
	5.1	Introduction	69
	5.2	The Frequency of Male Sterile Plants	71
		5.2.1 Complete seed-set of the male sterile plants	72
		5.2.2 Incomplete seed-set of the male sterile plants	73

6 Selection with Regard to a Trait				
	wit	h Qualitative Variation		
	6.1	Introduction		
	6.2	The Maintenance of Genetic Variation		
	6.3	Artificial Selection		
		6.3.1 Introduction		
		6.3.2 Line selection		
		6.3.3 Full sib family selection		
		6.3.4 Half sib family selection		
		6.3.5 Mass selection		
		6.3.6 Progeny testing 104		
7	Ra	ndom Variation of Allele Frequencies		
	7.1	Introduction		
	7.2	The Effect of the Mode of Reproduction on the Probability		
		of Fixation		
8	Co	mponents of the Phenotypic Value of Traits		
0	wit	h Quantitative Variation		
	8.1	Introduction		
	8.2	Components of the Phenotypic Value		
	8.3	Components of the Genotypic Value		
		8.3.1 Introduction		
		8.3.2 Partitioning of Genotypic Values According		
		to the F_{∞} -metric		
		8.3.3 Partitioning of Genotypic Values into their Additive		
		Genotypic Value and their Dominance Deviation 151		
		8.3.4 Breeding Value: A Concept Dealing		
		with Cross-fertilizing Crops		
9	Eff	ects of the Mode of Reproduction		
	on	the Expected Genotypic Value		
	9.1	Introduction		
	9.2	Random Mating		
	9.3	Self-Fertilization		
	9.4	Inbreeding Depression and Heterosis		
		9.4.1 Introduction		
		9.4.2 Hybrid Varieties		
		9.4.3 Synthetic Varieties		
10) Eff	ects of the Mode of Reproduction		
	on	the Genetic Variance		
	10.1	Introduction $\ldots \ldots 205$		

Contents

10.2 Random Mating	206
10.2.1 Partitioning of σ_a^2 in the case of open pollination	210
10.2.2 Partitioning of σ_a^2 in the case of pairwise crossing	215
10.3 Self-Fertilization	217
10.3.1 Partitioning of σ_g^2 in the case of self-fertilization	219
11 Applications of Quantitative Genetic Theory	005
	225
11.1 Prediction of the Response to Selection	225
11.2 The Estimation of Quantitative Genetic Parameters	243
11.2.1 Plant Material with Identical Reproduction	245
11.2.2 Cross-fertilizing Crops	249
11.2.3 Self-fertilizing Crops	254
11.3 Population Genetic and Quantitative Genetic Effects	
of Selection Based on Progeny Testing	257
11.4 Choice of Parents and Prediction of the Ranking of Crosses	266
11.4.1 Plant Material with Identical Reproduction	271
11.4.2 Self-fertilizing Plant Material	273
11.5 The Concept of Combining Ability as Applied to Pure Lines	277
11.5.1 Introduction \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots	277
11.5.2 General and Specific Combining Ability	279
12 Selection for Several Traits	289
12 1 Introduction	289
12.2 The Correlation Between the Phenotypic or Genotypic Values	200
of Traits with Quantitative Variation	201
12.3 Indirect Selection	201
12.5 Indirect betetion	294
12.3.2 The use of markers	200
12.3.2 The use of markers	200
Conditions Provided in Plant Production Practice	307
12.4 Estimation of the Coefficient of Phenotypic Environmental	001
Constic or Additive Constic Correlation	211
12.5 Index Selection and Independent Culling Levels Selection	218
12.5 Index Selection and Independent-Culling-Levels Selection	310
13 Genotype \times Environment Interaction	325
13.1 Introduction	325
13.2 Stability Parameters	329
13.3 Applications in Plant Breeding	333
14 Selection with Regard to a Trait	
with Quantitative Variation	330
14.1 Disalogue of Construin Values in the Case of A Trand	<i>J J J J J J J J J J </i>
14.1 Disclosure of Genotypic values in the Case of A frend	220
in the Quality of the Growing Conditions	339

14.2 Single-Plant Evaluation				341
14.2.1 Use of Plants Representing a Standard Variety				343
14.2.2 Use of Fixed Grids				343
14.2.3 Use of Moving Grids				348
14.3 Evaluation of Candidates by Means of Plots				355
14.3.1 Introduction		• •	•	355
14.3.2 Use of Plots Containing a Standard Variety	• • •	• •	•	350
14.2.2 Use of Moving Moong	• • •	• •	•	267
14.5.5 Use of moving means		• •	•	307
15 Reduction of the Detrimental Effect of Allocompet	itior	ı		
on the Efficiency of Selection				381
15.1 Introduction				381
15.2 Single-Plant Evaluation				389
15.2.1 The Optimum Plant Density				393
15.2.2 Measures to Beduce the Detrimental Effect				000
of Allocompetition				30/
15.3 Evaluation of Candidatos by Means of Plots	• • •	• •	•	308
15.5 Evaluation of Candidates by Means of 110ts		• •	•	090
16 Optimizing the Evaluation of Candidates by means				
of Plots				405
16.1 The Optimum Number of Replications				405
16.2 The Shape, Positioning and Size of the Test Plots				410
16.2.1 General considerations				410
16.2.2 Shape and Positioning of the Plots				413
16.2.3 Yardsticks to Measure Soil Heterogeneity				414
16.2.4 The Optimum Plot Size				
from an Economic Point of View				419
		• •	•	110
17 Causes of the Low Efficiency of Selection				421
17.1 Correct Selection				424
18 The Optimum Generation to Start Selection				
for Yield of a Self-Fertilizing Crop			•	429
18.1 Introduction \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots		• •	•	429
18.2 Reasons to Start Selection for Yield in an Early Gene	ratio	n.	•	430
18.3 Reasons to Start Selection				
for Yield in an Advanced Generation			•	433
10 Experimental Designs for the Evaluation				
of Candidate Varieties				437
			•	101
References				445
Index				457

Preface

Selection procedures used in plant breeding have gradually developed over a very long time span, in fact since settled agriculture was first undertaken. Nowadays these procedures range from very simple mass selection methods, sometimes applied in an ineffective way, to indirect trait selection based on molecular markers. The procedures differ in costs as well as in genetic efficiency. In contrast to the genetic efficiency, costs depend on the local conditions encountered by the breeder. The genetic progress per unit of money invested varies consequently from site to site. This book considers consequently only the genetic efficiency, *i.e.* the rate of progress to be expected when applying a certain selection procedure.

If a breeder has a certain breeding goal in mind, a selection procedure should be chosen. A wise choice requires a wellfounded opinion about the response to be expected from any procedure that might be applied. Such an opinion should preferably be based on the most appropriate model when considering the crop and the trait (or traits) to be improved. Sometimes little knowledge is available about the genetic control of expression of the trait(s). This applies particularly in the case of quantitative variation in the traits. It is, therefore, important to be familiar with methods for the elucidation of the inheritance of the traits of interest. This means, in fact, that the breeder should be able to develop population genetic and quantitative genetic models that describe the observed mode of inheritance as satisfactorily as possible.

The genetic models are generally based, by necessity, on simplifying assumptions. Quite often one assumes:

- a diploid behaviour of the chromosomes;
- an independent segregation of the pairs of homologous chromosomes at meiosis, or, more rigorously, independent segregation of the alleles at the loci controlling the expression of the considered trait;
- independence of these alleles with regard to their effects on the expression of the trait;
- a regular mode of reproduction within plants as well as among plants belonging to the same population; and/or
- the presence of not more than two alleles per segregating locus.

Such simplifying assumptions are made as a compromise between, on the one hand, the complexity of the actual genetic control, and, on the other hand, the desire to keep the model simple. Often such assumptions can be tested and so validated or revoked, but, of course, as the assumptions deviate more from the real situation, decisions made on the basis of the model will be less appropriate. The decisions concern choices with regard to:

- selection methods, e.g. mass selection versus half sib family selection;
- selection criteria, e.g. grain yield per plant versus yield per ear;
- experimental design, e.g. testing of each of N candidates in a single plot versus testing each of only $\frac{1}{2}N$ candidates in two plots; or
- data adjustment, e.g. moving mean adjustment versus adjustment of observations on the basis of observations from plots containing a standard variety.

In fact such decisions are often made on disputable grounds, such as experience, tradition, or intuition. This explains why breeders who deal in the same region with the same crop work in divergent ways. Indeed, their breeding goals may differ, but these goals themselves are often based on a subjective judgement about the ideotype (ideal type of plant) to be pursued.

In this book, concepts from plant breeding, population genetics, quantitative genetics, probability theory and statistics are integrated. The reason for this is to help provide a basis on which to make selection more professional, in such a way that the chance of being successful is increased. Success can, of course, never be guaranteed because the best theoretical decision will always be made on the basis of incomplete and simplifying assumptions. Nevertheless, the authors believe that a breeder familiar with the contents of this book is in a better position to be successful than a breeder who is not!

Preface to the Second Edition

New and upgraded paragraphs have been added throughout this edition. They have been added because it was felt, when using the first edition as a course book, that many parts could be improved according to a didactical point of view. It was, additionally, felt that – because of the increasing importance of molecular markers – more attention had to be given the use of markers (Section 12.3.2). In connection with this, quantitative genetic theory has, compared to the first edition, been more extensively developed for loci represented by multiple alleles (Sections 8.3.3 and 8.3.4).

It was stimulating to receive suggestions from interested readers. These suggestions have given rise to many improvements. Especially the many and useful suggestions from Ir. Ed G.J. van Paassen, Ir. Joël Schwarz, Dr. Hans-Peter Piepho, Dr. Mohamed Mahdi Sohani and Dr. L.R. Verdooren are acknowledged.

Chapter 1 Introduction

This chapter provides an overview of basic concepts and statistical tools underlying the development of population and quantitative genetics theory. These branches of genetics are of crucial importance with regard to the understanding of equilibria and shifts in (i) the genotypic composition of a population and (ii) the mean and variation exhibited by the population. In order to keep the theory to be developed manageable, two assumptions are made throughout the book, i.e. absence of linkage and absence of epistasis. These assumptions concern traits with quantitative variation.

Knowledge of population genetics, quantitative genetics, probability theory and statistics is indispensable for understanding equilibria and shifts with regard to the genotypic composition of a population, its mean value and its variation.

The subject of **population genetics** is the study of equilibria and shifts of allele and genotype frequencies in populations. These equilibria and shifts are determined by five forces:

• Mode of reproduction of the considered crop

The **mode of reproduction** is of utmost importance with regard to the breeding of any particular crop and the maintenance of already available varieties. This applies both to the natural mode of reproduction of the crop and to enforced modes of reproduction, like those applied when producing a hybrid variety. In plant breeding theory, crops are therefore classified into the following categories: cross-fertilizing crops (Chapter 2), self-fertilizing crops (Chapter 3), crops with both cross- and self-fertilization (Section 3.4) and asexually reproducing crops. In Section 2.1 it is explained that even within a specific population, traits may differ with regard to their mode of reproduction. This is further elaborated in Chapter 4.

- Selection (Chapters 6 and 12)
- Mutation (Section 6.2)
- Immigration of plants or pollen, *i.e.* immigration of alleles (Section 6.2)
- Random variation of allele frequencies (Chapter 7)

A **population** is a group of (potentially) interbreeding plants occurring in a certain area, or a group of plants originating from one or more common ancestors. The former situation refers to cross-fertilizing crops (in which case the term **Mendelian population** is sometimes used), while the latter group concerns, in particular, self-fertilizing crops. In the absence of immigration the population is said to be a **closed population**. Examples of closed populations are

- A group of plants belonging to a cross-fertilizing crop, grown in an isolated field, *e.g.* maize or rye (both pollinated by wind), or turnips or Brussels sprouts (both pollinated by insects)
- A collection of lines of a self-fertilizing crop, which have a common origin, *e.g.* a single-cross, a three-way cross, a backcross

The subject of **quantitative genetics** concerns the study of the effects of alleles and genotypes and of their interaction with environmental conditions.

Population genetics is usually concerned with the probability distribution of genotypes within a population (genotypic composition), while quantitative genetics considers phenotypic values (and statistical parameters dealing with them, especially mean and variance) for the trait under investigation. In fact population genetics and quantitative genetics are applications of probability theory in genetics. An important subject is, consequently, the derivation of probability distributions of genotypes and the derivation of expected genotypic values and of variances of genotypic values. Generally, statistical analyses comprise estimation of parameters and hypothesis testing. In quantitative genetics statistics is applied in a number of ways. It begins when considering the experimental design to be used for comparing entries in the breeding programme. Section 11.2 considers the estimation of interesting quantitative genetic parameters, while Chapter 12 deals with the comparison of candidates grown under conditions which vary in a trend.

Considered across the entries constituting a population (plants, clones, lines, families) the expression of an observed trait is a random variable. If the expression is represented by a numerical value the variable is generally termed **phenotypic value**, represented by the symbol p.

Note 1.1 In this book random variables are underlined.

Two genetic causes for variation in the expression of a trait are distinguished. Variation controlled by so-called **major genes**, *i.e.* alleles that exert a readily traceable effect on the expression of the trait, is called **qualitative variation**. Variation controlled by so-called **polygenes**, *i.e.* alleles whose individual effects on a trait are small in comparison with the total variation, is called **quantitative variation**. In Note 1.2 it is elaborated that this classification does not perfectly coincide with the distinction between **qualitative traits** and **quantitative traits**.

The former paragraph suggests that the term *gene* and *allele* are synonyms. According to Rieger, Michaelis and Green (1991) a **gene** is a continuous region of DNA, corresponding to one (or more) transcription units and consisting of a particular sequence of nucleotides. Alternative forms of a particular gene are referred to as **alleles**. In this respect the two terms 'gene' and 'allele' are sometimes interchanged. Thus the term 'gene frequency' is often used instead of the term 'allele frequency'. The term **locus** refers to the site, alongside a chromosome, of the gene/allele. Since the term 'gene' is often used as a synonym of the term 'locus', we have tried to avoid confusion by preferential use of the terms 'locus' and 'allele' (as a synonym of the word gene) where possible.

In the case of qualitative variation, the phenotypic value \underline{p} of an entry (plant, line, family) belonging to a genetically heterogeneous population is a **discrete random variable**. The phenotype is then exclusively (or to a largely traceable degree) a function f of the genotype, which is also a random variable $\underline{\mathcal{G}}$. Thus

$$p = f(\underline{\mathcal{G}})$$

It is often desired to deduce the genotype from the phenotype. This is possible with greater or lesser correctness, depending for example on the degree of dominance and sometimes also on the effect of the growing conditions on the phenotype. A knowledge of population genetics suffices for an insight into the dynamics of the genotypic composition of a population with regard to a trait with qualitative variation: application of quantitative genetics is then superfluous.

Note 1.2 All traits can show both qualitative and quantitative variation. Culm length in cereals, for instance, is controlled by dwarfing genes with major effects, as well as by polygenes. The commonly used distinction between qualitative traits and quantitative traits is thus, strictly speaking, incorrect. When exclusively considering qualitative variation, *e.g.* with regard to the traits in pea (*Pisum sativum*) studied by Mendel, this book describes the involved trait as a trait showing qualitative variation. On the other hand, with regard to traits where quantitative variation dominates – and which are consequently mainly discussed in terms of this variation – one should realize that they can also show qualitative variation. In this sense the following economically important traits are often considered to be 'quantitative characters':

- Biomass
- Yield with regard to a desired plant product
- Content of a desired chemical compound (oil, starch, sugar, protein, lysine) or an undesired compound
- Resistance, including components of partial resistance, against biotic or abiotic stress factors
- Plant height

In the case of quantitative variation \underline{p} results from the interaction of a **complex genotype**, *i.e.* several to many loci are involved, and the specific growing conditions are important. In this book, by complex genotype we mean the sum of the genetic constitutions of all loci affecting the expression of the considered trait. These loci may comprise loci with **minor genes** (or polygenes), as well as loci with major genes, as well as loci with both. With regard to a trait showing quantitative variation, it is impossible to classify individual plants, belonging to a genetically heterogeneous population, according to their

genotypes. This is due to the number of loci involved and the complicating effect on \underline{p} of (some) variation in the quality of the growing conditions. It is, thus, impossible to determine the number of plants representing a specified complex genotype. (With regard to the expression of qualitative variation this may be possible!). Knowledge of both population genetics and quantitative genetics is therefore required for an insight into the inheritance of a trait with quantitative variation.

The phenotypic value for a quantitative trait is a **continuous random variable** and so one may write

$$p = f(\underline{\mathcal{G}}, \underline{e})$$

Thus the phenotypic value is a function f of both the complex genotype (represented by $\underline{\mathcal{G}}$) and the quality of the growing conditions (say **environment**, represented by \underline{e}). Even in the case of a genetically homogeneous group of plants (a clone, a pure line, a single-cross hybrid) \underline{p} is a continuous random variable. The genotype is a constant and one should then write

$$p = f(\mathcal{G}, \underline{e})$$

Regularly in this book, simplifying assumptions will be made when developing quantitative genetic theory. Especially the following assumptions will often be made:

- (i) Absence of linkage of the loci controlling the studied trait(s)
- (ii) Absence of epistatic effects of the loci involved in complex genotypes. These assumptions will now be considered.

Absence of linkage

The assumption of absence of linkage for the loci controlling the trait of interest, *i.e.* the assumption of independent segregation, may be questionable in specific cases, but as a generalisation it can be justified by the following reasoning.

Suppose that each of the *n* chromosomes in the genome contains *M* loci affecting the considered trait. This implies presence of *n* groups of $\binom{M}{2}$ pairs of loci consisting of loci which are more strongly or more weakly linked. The proportion of pairs consisting of linked loci among all pairs of loci amounts then to

$$\frac{n\binom{M}{2}}{\binom{nM}{2}} = \frac{n.M!}{2!(M-2)!} \times \frac{2!(nM-2)!}{(nM)!} = \frac{M-1}{nM-1} = \frac{1-\frac{1}{M}}{n-\frac{1}{M}}$$

For M = 1 this proportion is 0; for M = 2 it amounts to 0.077 for rye (Secale cereale, with n = 7) and to 0.024 for wheat (Triticum aestivum, with n = 21);

1 Introduction

for M = 3 it amounts to 0.100 for rye and to 0.032 for wheat. For $M \to \infty$ the proportion is $\frac{1}{n}$; *i.e.* 0.142 for rye and 0.048 for wheat.

One may suppose that loci located on the same chromosome, but on different sides of the centromere, behave as unlinked loci. If each of the *n* chromosomes contains $m(=\frac{1}{2}M)$ relevant loci on each of the two arms then there are 2n groups of $\binom{m}{2}$ pairs consisting of linked loci. Thus considered, the proportion of pairs consisting of linked loci amounts to

$$\frac{2n\binom{m}{2}}{\binom{2nm}{2}} = \frac{2n.m!}{2!(m-2)!} \times \frac{2!(2nm-2)!}{(2nm)!} = \frac{1-\frac{1}{m}}{2n-\frac{1}{m}}$$

For m = 1 this proportion is 0; for m = 2 it amounts to 0.037 for rye and to 0.012 for wheat; for m = 3 it amounts to 0.049 for rye and to 0.016 for wheat. For $m \to \infty$ the proportion is $\frac{1}{2n}$; *i.e.* 0.071 for rye and 0.024 for wheat.

For the case of an even distribution across all chromosomes of the polygenic loci affecting the considered trait it is concluded that the proportion of pairs of linked loci tends to be low. (In an autotetraploid crop the chromosome number amounts to 2n = 4x. The reader might like to consider what this implies for the above expressions.)

Absence of epistasis

Absence of **epistasis** is another assumption that will be made regularly in this book, notably in Sections 8.3.2 and 10.1. It implies **additivity** of the effects of the single-locus genotypes for the loci affecting the level of expression for the considered trait. The genotypic value of some complex genotype consists then of the sum of the genotypic value of the complex genotype with regard to all non-segregating loci, here represented by m, as well as the sum of the contributions due to the genotypes for each of the K segregating polygenic loci B_1-b_1, \ldots, B_K-b_K . Thus

$$\mathcal{G}_{B_1 - b_1, \dots, B_K - b_K} = m + \mathcal{G}'_{B_1 - b_1} + \dots + \mathcal{G}'_{B_K - b_K}$$
(1.1)

where \mathcal{G}' is defined as the contribution to the genotypic value, relative to the population mean genotypic value, due to the genotype for the considered locus (Section 8.3.3). The assumption implies the absence of **inter-locus interaction**, *i.e.* the absence of **epistasis** (in other words: absence of **non-allelic interaction**). It says that the effect of some genotype for some locus $B_i - b_i$ in comparison to another genotype for this same locus does not depend at all on the complex genotype determined by all other relevant loci.

In this book, in order to clarify or substantiate the main text, theoretical examples and results of actual experiments are presented. Notes provide short additional information and appendices longer, more complex supplementary information or mathematical derivations.

Chapter 2 Population Genetic Effects of Cross-fertilization

Cross-fertilization produces populations consisting of a mixture of plants with a homozygous or heterozygous (complex) genotype. In addition, the effects of a special form of cross-fertilization, i.e. panmixis, are considered. It is shown that continued panmixis leads sooner or later to a genotypic composition which is completely determined by the allele frequencies. The allele frequencies do not change in course of the generations but the haplotypic and genotypic composition may change considerably. This process is described for diploid and autotetraploid crops.

2.1 Introduction

There are several mechanisms promoting cross-pollination and, consequently, cross-fertilization. The most important ones are

• Dioecy, *i.e.* male and female gametes are produced by different plants.

Asparagus	Asparagus officinalis L.
Spinach	Spinacia oleracea L.
Papaya	Carica papaya L.
Pistachio	Pistacia vera L.
Date palm	Phoenix dactylifera L.

• **Monoecy**, *i.e.* male and female gametes are produced by separate flowers occurring on the same plant.

Banana	Musa spp.
Oil palm	Elaeis guineensis Jacq
Fig	Ficus carica L.
Coconut	Cocos nucifera L.
Maize	Zea mays L.
Cucumber	Cucumis sativus L.

In musk melon (*Cucumis melo* L.) most varieties show **andromonoecy**, *i.e.* the plants produce both staminate flowers and bisexual flowers, whereas other varieties are monoecious.

• **Protandry**, *i.e.* the pollen is released before receptiveness of the stigmata.

Leek Allium porrum L. Onion Allium cepa L. Carrot Daucus carota L. Sisal Agave sisalana Perr.

• **Protogyny**, *i.e.* the stigmata are receptive before the pollen is released.

Tea	Camellia sinensis (L.) O. Kuntze
Avocado	Persea americana Miller
Walnut	Juglans nigra L.
Pearl millet	Pennisetum typhoides L. C. Rich.

• **Self-incompatibility**, *i.e.* a physiological barrier preventing normal pollen grains fertilizing eggs produced by the same plant.

Cacao	Theobroma cacao L.
Citrus	Citrus spp.
Tea	Camellia sinensis L. O. Kuntze
Robusta coffee	Coffea canephora Pierre ex Froehner
Sugar beets	Beta vulgaris L.
Cabbage, kale	Brassica oleracea spp.
Rye	Secale cereale L.
Many grass species,	e.q. perennial ryegrass (Lolium perenne L.)

• Flower morphology

Fig	Ficus carica L.
Primrose	Primula veris L.
Common buckwheat	Fagopyrum esculentum Moench.
and probably in the Bi	rd of Paradise flower Strelitzia reginae Banks

Effects with regard to the haplotypic and genotypic composition of a population due to (continued) reproduction by means of **panmixis** will now be derived for a so-called **panmictic population**. Panmictic reproduction occurs if each of the next five conditions apply:

- (i) Random mating
- (ii) Absence of random variation of allele frequencies
- (iii) Absence of selection
- (iv) Absence of mutation
- (v) Absence of immigration of plants or pollen

In the remainder of this section the first two features of panmixis are more closely considered.

Random mating

Random mating is defined as follows: in the case of random mating the fusion of gametes, produced by the population as a whole, is at random with regard to the considered trait. It does not matter whether the mating occurs by means of crosses between pairs of plants combined at random, or by means of open pollination.

8

2.1 Introduction

Open pollination in a population of a cross-fertilizing (allogamous) crop may imply random mating. This depends on the trait being considered. One should thus be careful when considering the mating system. This is illustrated in Example 2.1.

Example 2.1 Two types of rye plants can be distinguished with regard to their epidermis: plants with and plants without a waxy layer. It seems justifiable to assume random mating with regard to this trait. With regard to time of flowering, however, the assumption of random mating may be incorrect. Early flowering plants will predominantly mate *inter se* and hardly ever with late flowering plants. Likewise late flowering plants will tend to mate with late flowering plants and hardly ever with early flowering ones. With regard to this trait, so-called **assortative mating** (see Section 4.1) occurs.

One should, however, realize that the ears of an individual rye plant are produced successively. The assortative mating with regard to flowering date may thus be far from perfect. Also, with regard to traits controlled by loci linked to the locus (or loci) controlling incompatibility, *e.g.* in rye or in meadow fescue (*Festuca pratensis*), perfect random mating will therefore probably not occur.

Selection may interfere with the mating system. Plants that are resistant to an agent (e.g. disease or chemical) will mate *inter se* (because susceptible plants are eliminated). Then assortative mating occurs due to selection.

Crossing of neighbouring plants implies random mating if the plants reached their positions at random; crossing of contiguous inflorescences belonging to the same plant (**geitonogamy**) is, of course, a form of selfing.

Random mating does not exclude a fortuitous relationship of mating plants. Such relationships will occur more often with a smaller population size. If a population consists, generation after generation, of a small number of plants, it is inevitable that related plants will mate, even when the population is maintained by random mating. Indeed, mating of related plants yields an increase in the frequency of homozygous plants, but in this situation the increase in the frequency of homozygous plants is also due to another cause: fixation occurs because of non-negligible random variation of allele frequencies. Both causes of the increase in homozygosity are due to the small population size (and not to the mode of reproduction).

This ambiguous situation, so far considered for a single population, occurs particularly when numerous small **subpopulations** form together a large **superpopulation**. In each subpopulation random mating, associated with non-negligible random variation of the allele frequencies, may occur, whereas in the superpopulation as a whole inbreeding occurs. Example 2.2 provides an illustration.

Example 2.2 A large population of a self-fertilizing crop, *e.g.* an F_2 or an F_3 population, consists of numerous subpopulations each consisting of a single plant. Because the gametes fuse at random with regard to any trait, one may state that random mating occurs within each subpopulation. At the level of the superpopulation, however, selfing occurs.

Selfing is impossible in dioecious crops, *e.g.* spinach (*Spinacia oleracea*). Inbreeding by means of continued sister \times brother crossing may then be applied. This full sib mating at the level of the superpopulation may imply random mating within subpopulations consisting of full sib families (see Section 3.1).

Seen from the level of the superpopulation, inbreeding occurs if related plants mate preferentially. This may imply the presence of subpopulations, reproducing by means of random mating. If very large, the superpopulation will retain all alleles. The increasing homozygosity rests on gene fixation in the subpopulations. If, however, only a single full sib family produces offspring by means of open pollination, implying crossing of related plants, then the population as a whole (in this case just a single full sib family) is still said to be maintained by random mating.

Absence of random variation of allele frequencies

The second characteristic of panmixis is absence of random variation of allele frequencies from one generation to the next. This requires an infinite **effective size** of the population, originating from an infinitely large sample of gametes produced by the present generation. Panmixis thus implies a **deterministic model**. In populations consisting of a limited number of plants, the allele frequencies vary randomly from one generation to the next. Models describing such populations are **stochastic models** (Chapter 7).

2.2 Diploid Chromosome Behaviour and Panmixis

2.2.1 One Locus with Two Alleles

The majority of situations considered in this book involve a locus represented by not more than two alleles. This is certainly the case in diploid species in the following populations:

- Populations tracing back to a cross between two pure lines, say, a single cross
- Populations obtained by (repeated) backcrossing (if, indeed, both the donor and the recipient have a homozygous genotype)

It is possibly the case in populations tracing back to a three-way cross or a double cross. It is improbable in other populations, like populations of cross-fertilizing crops, populations tracing back to a complex cross, landraces, multiline varieties.

To keep (polygenic) models simple, it will often be assumed that each of the considered loci is represented by only two alleles. Quite often this simplification will violate reality. The situation of multiple allelic loci is explicitly considered in Sections 2.2.2 and 8.3.3.

If the expression for the trait of interest is controlled by a locus with two alleles A and a (say locus A-a) then the probability distribution of the geno-types occurring in the considered population is often described by

$$\begin{array}{c} \hline \begin{array}{c} \text{Genotype} \\ \hline aa & Aa & AA \\ \end{array} \\ \text{Probability} & f_0 & f_1 & f_2 \end{array}$$

One may represent the probability distribution (in this book mostly the term **genotypic composition** will be used) by the row vector (f_0, f_1, f_2) . The symbol f_j represents the probability that a random plant contains j A-alleles in its genotype for locus A-a, where j may be equal to 0, 1 or 2. It has become custom to use the word **genotype frequency** to indicate the probability of a certain genotype and for that reason the symbol f is used.

The plants of the described population produce gametes which have either haplotype a or haplotype A. (Throughout this book the term **haplotype** is used to indicate the genotype of a gamete.) The probability distribution of the haplotypes of the gametes produced by the population is described by

	Haplotype		
	a	A	
Probability	g_0	g_1	

The symbol g_j represents the probability that a random gamete contains j Aalleles in its haplotype for locus A-a, where j may be equal to 0 or 1. The row vector (g_0, g_1) describes, in a condensed way, the **haplotypic composition** of the gametes. The habit to use the symbol q instead g_0 and the symbol pinstead of g_1 is followed in this book whenever a single locus is considered. The term **allele frequency** will be used to indicate the probability of the considered allele.

So far it has been assumed that the allele frequencies are known and hereafter the theory is further developed without considering the question of how one arrives at such knowledge. In fact allele frequencies are often unknown. When one would like to estimate them one might do that in the following way. Assume that a random sample of N plants is comprised of the following numbers of plants of the various genotypes:

	Genotype		
	aa	Aa	AA
Number of plants	n_0	n_1	n_2

For any value for N the frequencies q and p of alleles a and A may then be estimated as

$$q = \frac{2n_0 + n_1}{2N}$$
 and $p = \frac{n_1 + 2n_2}{2N}$

Throughout the book the expressions 'the probability that a random plant has genotype Aa', or 'the probability of genotype Aa', or 'the frequency of genotype Aa' are used as equivalents. This applies likewise for the expressions 'the probability that a gamete has haplotype A', or 'the probability of A'.

Fusion of a random female gamete with a random male gamete yields a genotype specified by \underline{j} , the number of A alleles in the genotype. (The number of a alleles in the genotype amounts – of course – to 2 - j.) The probability that a plant with genotype aa results from the fusion is in fact equal to the probability of the event that \underline{j} assumes the value 0. The quantity \underline{j} assumes thus a certain value (0 or 1 or 2) with a certain probability. This means that \underline{j} is a random variable.

The probability distribution for \underline{j} , *i.e.* for the genotype frequencies, is given by the binomial probability distribution:

$$P(\underline{j}=j) = \binom{2}{j} p^j q^{2-j}$$

Fusion of two random gametes therefore yields

- With probability q^2 a plant with genotype aa
- With probability 2pq a plant with genotype Aa
- With probability p^2 a plant with genotype AA

The probabilities for the multinomial probability distribution of plants with these genotypes may be represented in a condensed form by the row vector $(q^2, 2pq, p^2)$. This notation represents also the genotypic composition to be expected for the population obtained after panmixis in a population with gene frequencies (q, p). In the case of panmixis there is a direct relationship between the gene frequencies in a certain generation and the genotypic composition of the next generation (see Fig. 2.1). Thus if the genotype frequencies f_0, f_1 and f_2 of a certain population are equal to, respectively, $q^2, 2pq$ and p^2 , the considered population has the so-called Hardy–Weinberg (genotypic) composition. The actual genotypic composition is then equal to the composition expected after panmixis. With continued panmixis, populations of later generations will continue to have the Hardy–Weinberg composition. Therefore such composition may be indicated as the **Hardy–Weinberg equilibrium**. The names of Hardy (1908) and Weinberg (1908) are associated with this genotypic composition, but it was in fact derived by Castle in 1903 (Keeler, 1968).

With two alleles per locus the maximum frequency of plants with the Aa genotype in a population originating from panmixis is $\frac{1}{2}$ for $p = q = \frac{1}{2}$ (Fig. 2.1). This occurs in F₂ populations of self-fertilizing crops. The F₂ originates from selfing of individual plants of the F₁, but because each plant of the



Fig. 2.1 The frequency of plants with genotype aa, Aa or AA in the population obtained by panmixis in a population with gene frequency P_A

 F_1 has the same genotype, panmixis within each plant coincides with panmixis of the F_1 as a whole. (The F_1 itself may be due to bulk crossing of two pure lines; the proportion of heterozygous plants amounts then to 1.)

The Hardy–Weinberg genotypic composition constitutes the basis for the development of population genetic theory for cross-fertilizing crops. It is obtained by an infinitely large number of pairwise fusions of random eggs with random pollen, as well as by an infinitely large number of crosses involving pairs of random plants. One may also say that it is expected to occur both after pairwise fusions of random eggs and pollen, and when crossing plants at random.

In a number of situations two populations are crossed as bulks. One may call this **bulk crossing**. One population contributes the female gametes (containing the eggs) and the other population the male gametes (the pollen, containing generative nuclei in the pollen tubes). In such a case, crosses within each of the involved populations do not occur. A possibly unexpected case of bulk crossing is described in Note 2.1.

Note 2.1 Selection among plants after pollen distribution, *e.g.* selection with regard to the colour of the fruits (if fruit colour is maternally determined), implies a special form of bulk crossing: the rejected plants are then excluded as effective producers of eggs (these plants will not be harvested), whereas all plants (could) have been effective as producers of pollen. The results, to be derived hereafter, in the main text, for a bulk cross of two populations with different allele frequencies, are applied in Section 6.3.5.

A bulk cross is of particular interest if the haplotypic composition of the eggs differs from the haplotypic composition of the pollen. Thus if population I, with allele frequencies (q_1, p_1) , contributes the eggs and population II, with allele frequencies (q_2, p_2) , the pollen, then the expected genotypic composition of the obtained hybrid population, in row vector notation, is

$$(q_1q_2, p_1q_2 + p_2q_1, p_1p_2) (2.1)$$

This hybrid population does not result from panmix is. The frequency of allele ${\cal A}$ is

$$p = \frac{1}{2}(p_1q_2 + p_2q_1) + p_1p_2 = \frac{1}{2}p_1q_2 + \frac{1}{2}p_1p_2 + \frac{1}{2}p_2q_1 + \frac{1}{2}p_1p_2$$

= $\frac{1}{2}p_1(q_2 + p_2) + \frac{1}{2}p_2(q_1 + p_1) = \frac{1}{2}(p_1 + p_2)$ (2.2)

as

$$q_2 + p_2 = q_1 + p_1 = 1$$

N.B. Further equations based on p + q = 1 are elaborated in Note 2.2.

Note 2.2 When deriving Equation (2.2) the equation p + q = 1 was used. On the basis of the latter equation several other equations, applied throughout this book, can be derived:

$$q^2 + 2pq + p^2 = 1 \tag{2.3}$$

$$p - q = 2p - 1 = 1 - 2q \tag{2.4}$$

$$(p-q)^{2} = (p^{2} - 2pq + q^{2}) = 1 - 4pq$$
(2.5)

$$p^{2} - q^{2} = (p+q)(p-q) = p - q = f_{2} - f_{0}$$
(2.6)

$$p - q + 2pq = p^{2} - q^{2} + 2pq = p^{2} + 2pq - q^{2} = 1 - 2q^{2}$$
(2.7)

and

$$p^{4} + p^{3}q + pq^{3} + q^{4} - (p - q)^{2} = p^{3} + q^{3} - p^{2} + 2pq - q^{2}$$

$$= p^{2}(p - 1) + q^{2}(q - 1) + 2pq$$

$$= -p^{2}q - pq^{2} + 2pq$$

$$= -pq(p + q - 2) = 2pq \qquad (2.8)$$

Panmictic reproduction of this hybrid population produces offspring with the Hardy–Weinberg genotypic composition. The hybrid population contains, compared to the offspring population, an excess of heterozygous plants. The excess is calculated as the difference in the frequencies of heterozygous plants:

$$(p_1q_2 + p_2q_1) - 2pq = (p_1q_2 + p_2q_1) - 2[\frac{1}{2}(p_1 + p_2)\frac{1}{2}(q_1 + q_2)$$

= $\frac{1}{2}(p_1q_2 + p_2q_1 - p_1q_1 - p_2q_2)$
= $\frac{1}{2}(p_1 - p_2)(q_2 - q_1) = \frac{1}{2}(p_1 - p_2)^2$ (2.9)

This square is positive, unless $p_1 = p_2$. Thus the hybrid does indeed contain an excess of heterozygous plants. Example 2.3 illustrates that the superiority of hybrid varieties might (partly) be due to this excess. This is further elaborated in Section 9.4.1. Example 2.4 pays attention to the case of both inter- and intra-mating of two populations.

Example 2.3 It is attractive to maximize the frequency of hybrid plants whenever they have a superior genotypic value. This is applied when producing single-cross hybrid varieties by means of a bulk cross between two well-combining pure lines. If $p_1 = 1$ (thus $q_1 = 0$) in one parental line and $p_2 = 0$ (thus $q_2 = 1$) in the other, the excess of the frequency of heterozygous plants will be at its maximum, because $\frac{1}{2}(p_1-p_2)^2$ attains then its maximum value, *i.e.* $\frac{1}{2}$. The genotypic composition of the single-cross hybrid is (0, 1, 0). Equation (2.2) implies that panmictic reproduction of this hybrid yields a population with the Hardy-Weinberg genotypic composition $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$. The excess of heterozygous plants in the hybrid population is thus indeed $\frac{1}{2}$. (Panmictic reproduction of a hybrid population tends to yield a population with a reduced expected genotypic value; see Section 9.4.1).

The excess of heterozygous plants is low when one applies bulk crossing of similar populations. At $p_1 = 0.6$ and $p_2 = 0.7$, for example, the hybrid population has the genotypic composition (0.12; 0.46; 0.42), with p = 0.65. The corresponding Hardy–Weinberg genotypic composition is then (0.1225; 0.4550; 0.4225) and the excess of heterozygous plants is only 0.005.

As early as 1908 open-pollinating maize populations were crossed in the USA with the aim of producing superior hybrid populations. This had already been suggested in 1880 by Beal. Shull (1909) was the first to suggest the production of single-cross hybrid varieties by crossing pure lines.

Example 2.4 Two populations of a cross-fertilizing crop, *e.g.* perennial rye grass, are mixed. The mixture consists of a portion, P, of population I material and a portion, 1-P, of population II material. In the mixture both mating between and within the populations occur. When assuming

- simultaneous flowering,
- simultaneous ripening,
- equal fertility of the plants of both populations and
- random mating

the proportion of hybrid seed is 2P(1-P); see Foster (1971). For $P = \frac{1}{2}$ this proportion is maximal, *i.e.* $\frac{1}{2}$.

2.2.2 One Locus with more than Two Alleles

Multiple allelism does not occur in the populations considered so far. However, multiple allelism is known to occur in self- and cross-fertilizing crops (see Example 2.5). It may further be expected in three-way-cross hybrids, and their offspring, as well as in mixtures of pure lines (landraces or multiline varieties). **Example 2.5** The intensity of the anthocyanin colouration in lettuce (*Lactuca sativa*), a self-fertilizing crop, is controlled by at least three alleles. The colour and location of the white leaf spots of white clover (*Trifolium repens*), a cross-fertilizing crop, are controlled by a multiple allelic locus. The expression for these traits appears to be controlled by a locus with at least 11 alleles. Another locus, with at least four alleles, controls the red leaf spots (Julén, 1959). (White clover is an autotetraploid crop with a gametophytic incompatibility system and a diploid chromosome behaviour; 2n = 4x = 32).

The frequencies (f) of the genotypes A_iA_j (with $i \leq j; j = 1, ..., n$) for the multiple allelic locus $A_1 \cdot A_2 \cdot ... \cdot A_n$ attain their equilibrium values following a single round of panmictic reproduction. The genotypic composition is then:

	Genotype		
	$A_1 A_1 \dots$	$A_i A_j \dots$	$A_n A_n$
f	$p_1{}^2$	$2p_ip_j$	p_n^2

The proportion of homozygous plants is minimal for $p_j = \frac{1}{n}$ (for j = 1, ..., n) and amounts then to $n\left(\frac{1}{n}\right)^2 = \frac{1}{n}$; see Falconer (1989, pp. 388–389).

2.2.3 Two Loci, Each with Two Alleles

In Section 2.2.1 it was shown that a single round of panmictic reproduction produces immediately the Hardy–Weinberg genotypic composition with regard to a single locus. It is immediately attained because the random fusion of pairs of gametes implies random fusion of separate alleles, whose frequencies are constant from one generation to the next. For complex genotypes, *i.e.* genotypes with regard to two or more loci (linked or not), however, the so-called **link-age equilibrium** is only attained after *continued panmixis*. Presence of the Hardy–Weinberg genotypic composition for separate loci does not imply presence of linkage equilibrium! (Example 2.7 illustrates an important exception to this rule.)

In panmictic reproduction the frequencies of complex genotypes follow from the frequencies of the complex haplotypes. Linkage equilibrium is thus attained if the haplotype frequencies are constant from one generation to the next. For this reason 'linkage equilibrium' is also indicated as **gametic phase equilibrium**. In this section it is derived how the haplotypic frequencies approach their equilibrium values in the case of continued panmixis. This implies that the tighter the linkage the more generations are required. However, even for unlinked loci a number of rounds of panmictic reproduction are required to attain linkage equilibrium. The genotypic composition in the equilibrium does not depend at all on the strength of the linkage of the loci involved. The designation 'linkage equilibrium' is thus not very appropriate.

2.2 Diploid Chromosome Behaviour and Panmixis

To derive how the haplotype frequencies approach their equilibrium, the notation introduced in Section 2.2.1 must be extended. We consider loci A-a and B-b, with frequencies p and q for alleles A and a and frequencies r and s for alleles B and b. The **recombination value** is represented by r_c . This parameter represents the probability that a gamete has a recombinant haplotype (see Section 2.2.4). Independent segregation of the two loci occurs at $r_c = \frac{1}{2}$, absolute linkage at $r_c = 0$. Example 2.6 illustrates the estimation of r_c in the case of a testcross with a line with a homozygous recessive (complex) genotype.

The haplotype frequencies are determined at the meiosis. The haplotypic composition of the gametes produced by generation G_{t-1} is described by

	Haplotype			
	ab	aB	Ab	AB
f	$g_{00,t}$	$g_{01,t}$	$g_{10,t}$	$g_{11,t}$

The last subscript (t) in the symbol for the haplotype frequencies indicates the rank of the generation to be formed in a series of generations generated by panmictic reproduction (t = 1, 2, ...); see Note 2.3.

Example 2.6 The spinach variety Wintra is susceptible to the fungus *Peronospora spinaciae* race 2 and tolerant to Cucumber virus 1. It was crossed with spinach variety Nores, which is resistant to *P. spinaciae* race 2 but sensitive to Cucumber virus 1. The loci controlling the host-pathogen relations are A - a and B - b. The genotype of Wintra is aaBB and the genotype of Nores *AAbb*. The offspring, with genotype AaBb, were crossed with the spinach variety Eerste Oogst (genotype aabb), which is susceptible to *P. spinaciae* race 2 and sensitive to Cucumber virus 1. On the basis of the reaction to both pathogens a genotype was assigned to each of the 499 plants resulting from this testcross (Eenink, 1974):

	Genotype				
-	aabb	aaBb	Aabb	AaBb	Total
Frequency					
• Observed	61	190	194	54	499
• Expected	124.75	124.75	124.75	124.75	499

The expected frequencies are calculated on the basis of the null hypothesis stating that the two involved loci are unlinked. The expected $\frac{1}{2}$: $\frac{1}{2}$ segregation ratio was confirmed by a goodness of fit test for each separate locus. The specified null hypothesis is, of course, rejected. The two loci are clearly linked. The value estimated for r_c is

$$\frac{61+54}{499} = 0.23$$

Note 2.3 In this book the last subscript in the symbols for the genotype and haplotype frequencies indicate the generation number. If it is t it refers to population G_t , *i.e.* the population obtained by panmictic reproduction of t successive generations.

Population G_1 , resulting from paninctic reproduction in a single-cross hybrid, has the same genotypic composition as the F_2 population resulting from selfing plants of the single-cross hybrid. To standardize the numbering of generations of cross-fertilizing crops and those of self-fertilizing crops, the population resulting from the first reproduction by means of selfing might be indicated by S_1 (rather than by the more common indication F_2). To avoid confusion this will only be done when appropriate, *e.g.* in Section 3.2.1.

The last subscript in the symbols for the haplotype frequencies of the gametes giving rise to S_1 are taken to be 1. The same applies to the frequencies of the genotypes in S_1 . This system for labelling generations of gametophytes and sporophytes was also adopted by Stam (1977).

Population G_0 is thus some initial population, obtained after a bulk cross or simply by mixing. It produces gametes with the haplotypic composition $(g_{00,1}; g_{01,1}; g_{10,1}; g_{11,1})$.

In the absence of selection, allele frequencies do not change. This implies

$$g_{10,1} + g_{11,1} = g_{10,2} + g_{11,2} = \ldots = p$$

for allele A, and similar equations for the frequencies of alleles a, B and b.

It was already noted that the haplotype frequencies in successive generations will be considered. In the appendix of this section it is shown that the following recurrent relations apply:

$$g_{00,t+1} = g_{00,t} - r_c d_t \tag{2.10a}$$

$$g_{01,t+1} = g_{01,t} + r_c d_t \tag{2.10b}$$

$$g_{10,t+1} = g_{10,t} + r_c d_t \tag{2.10c}$$

$$g_{11,t+1} = g_{11,t} - r_c d_t \tag{2.10d}$$

where the definition of d_t follows from

$$2d_t := f_{11C,t} - f_{11R,t} \tag{2.11}$$

where ':=' means: 'is defined as', and $t = 1, 2, 3, \ldots$

N.B. In Note 3.6 it is shown that Equations (2.10a–d) also apply to self-fertilizing crops. The recurrent equations show that the haplotype frequencies do not change from one generation to the next if $r_c = 0$ or if $d_t = 0$. Such constancy of the haplotypic composition implies constancy of the genotypic

composition. It implies presence of linkage equilibrium. Linkage equilibrium is thus immediately established by a single round of panmictic reproduction for loci with $r_c = 0$. This situation coincides with the case of a single locus with four alleles.

The symbol f_{11C} indicates the frequency of AB/ab-plants, *i.e.* doubly heterozygous plants in **coupling phase (C-phase)**; the symbol f_{11R} represents the frequency of Ab/aB-plants, *i.e.* doubly heterozygous plants in **repulsion phase (R-phase)**.

In the case of panmixis the following equations apply:

$$f_{11C,t} = 2(g_{11,t} g_{00,t})$$
$$f_{11R,t} = 2(g_{10,t} g_{01,t})$$

In that case we get

$$d_t = (g_{11,t} g_{00,t}) - (g_{10,t} g_{01,t})$$
(2.12)

This parameter is called **coefficient of linkage disequilibrium**. It appears in the following derivation:

$$g_{11,t} = g_{11,t}(g_{10,t} + g_{01,t} + g_{11,t} + g_{00,t}) = (g_{10,t} g_{01,t} + g_{10,t} g_{11,t} + g_{11,t} g_{01,t} + g_{11,t}^2) + (g_{11,t} g_{00,t} - g_{10,t} g_{01,t})$$
$$= (g_{10,t} + g_{11,t})(g_{01,t} + g_{11,t}) + d_t = pr + d_t$$

Equation (2.10d) may thus be rewritten as

$$pr + d_{t+1} = (pr + d_t) - r_c d_t$$

which implies not only

$$d_{t+1} = (1 - r_c)d_t$$

but of course also

$$d_t = (1 - r_c)^{t-1} d_1 (2.13)$$

for t = 2, 3, ...

The derivation above (and similar derivations for the other haplotype frequencies) implies

$$d_t = g_{11,t} - pr = -(g_{10,t} - ps) = -(g_{01,t} - qr) = g_{00,t} - qs$$

Because $\frac{1}{2} \leq (1-r_c) \leq 1$, continued panmixis implies continued decrease of d_t . The decrease is faster for smaller values of $1-r_c$, *i.e.* for higher values of r_c . Independent segregation, *i.e.* $r_c = \frac{1}{2}$, yields the fastest reduction, *viz.* halving of d_t by each panmictic reproduction. The value of d_t eventually attained,

i.e. $d_t = 0$, implies that linkage equilibrium is attained, *i.e.* constancy of the haplotype frequencies. The haplotype frequencies have then a special value, *viz.*

$$g_{00} = qs$$
$$g_{01} = qr$$
$$g_{10} = ps$$
$$g_{11} = pr$$

The equilibrium frequencies of the haplotypes are equal to the products of the frequencies of the alleles involved, and the equilibrium frequencies of the complex genotypes are equal to the products of the Hardy–Weinberg frequencies of the single-locus genotypes for the loci involved. The strength of the linkage between the loci is irrelevant with regard to the genotypic composition in the equilibrium. It only affects the number of generations of panmictic reproduction required to 'attain' the equilibrium.

Table 2.1 presents the equilibrium frequencies of complex genotypes and phenotypes for the simultaneously considered loci A-a and B-b.

continue	ed panmictic reprod	luction		
(a) Gen	otypes			
	bb	Bb	BB	
aa	$q^{2}s^{2}$	$2q^2rs$	$q^{2}r^{2}$	q^2
Aa	$2pqs^2$	4 pqrs	$2pqr^2$	2pq
AA	p^2s^2	$2p^2rs$	$p^{2}r^{2}$	p^2
	s^2	2rs	r^2	1
(b) Phe	notypes			
	bb	B^{\cdot}		
aa	$q^{2}s^{2}$	$q^2(1-s^2)$	q^2	
A^{\cdot}	$(1-q^2)s^2$	$(1-q^2)(1-s^2)$	$(1-q^2)$	
	s^2	$1 - s^2$		

Table 2.1 Equilibrium frequencies of (a) complex genotypes and (b) phenotypes in the case of complete dominance. The equilibrium is attained after continued panmictic reproduction

The foregoing is illustrated in Example 2.7, which deals with the production of a single-cross hybrid variety and the population resulting from its offspring as obtained by panmictic reproduction. Example 2.8 illustrates the production of a synthetic variety and a few of its offspring generations as obtained by continued random mating.

Example 2.7 Cross $\frac{AB}{AB} \times \frac{ab}{ab}$ yields a doubly heterozygous genotype in the coupling phase, *i.e.* $\frac{AB}{ab}$, whereas cross $\frac{Ab}{Ab} \times \frac{aB}{aB}$ yields a doubly heterozygous genotype in the repulsion phase, *i.e.* $\frac{Ab}{aB}$. In both cases the single-cross hybrid variety, say population G₀, is heterozygous for the loci *A*-*a* and *B*-*b*. It produces gametes with the following haplotypic composition:

		Haplotype				
		ab	aB	Ab	AB	d_1
f	in general for G_0 in C-phase: for G_0 in R-phase:	$\begin{array}{c} g_{00,1} \\ \frac{1}{2} - \frac{1}{2} r_c \\ \frac{1}{2} r_c \end{array}$	$\begin{array}{c} g_{01,1} \\ \frac{1}{2} r_c \\ \frac{1}{2} - \frac{1}{2} r_c \end{array}$	$\begin{array}{c} g_{10,1} \\ \frac{1}{2} r_c \\ \frac{1}{2} - \frac{1}{2} r_c \end{array}$	$\begin{array}{c} g_{11,1} \\ \frac{1}{2} - \frac{1}{2} r_c \\ \frac{1}{2} r_c \end{array}$	$\frac{\frac{1}{4}(1-2r_c)}{-\frac{1}{4}(1-2r_c)}$

The quantity d_1 is calculated according to Equation (2.12). This yields for G_0 in C-phase

$$d_1 = \frac{1}{4}(1 - r_c)^2 - \frac{1}{4}r_c^2 = \frac{1}{4}(1 - 2r_c)$$

The value for d_1 is in the interval $(0, \frac{1}{4})$ or in the interval $(-\frac{1}{4}, 0)$. In G₁ the absolute value of d_1 is at a maximum. Continued pannictic reproduction gives, in G_{∞}, the linkage equilibrium pertaining to $p = q = r = s = \frac{1}{2}$. Table 2.2 presents the genotypic composition of population G₁ resulting from a single pannictic reproduction of either G₀ in C-phase or in R-phase, as well as the genotypic composition of population G_{∞} resulting from continued pannixis.

Starting with a single-cross hybrid, the quantity d_1 is equal to zero for loci with $r_c = \frac{1}{2}$. Then a single generation of panmictic reproduction produces a population in linkage equilibrium. This remarkable result applies even in the case of selfing of the hybrid variety. (In Section 2.2.1 it has already been indicated that the result of selfing of F_1 plants coincides with the result of panmixis among F_1 plants). Thus for unlinked loci panmictic reproduction (or selfing) of a single-cross hybrid immediately yields a population in linkage equilibrium. Continued panmictic reproduction does not yield further shifts in haplotype and genotype frequencies. This means that it is useless to apply random mating in the F_2 of a self-fertilizing crop with the goal of increasing the frequency of plants with a recombinant genotype.

On the basis of the frequencies of the phenotypes for two traits (each with two levels of expression) showing qualitative variation, one can easily determine whether or not a certain population is in linkage equilibrium. It is, however, impossible to conclude whether or not the loci involved are linked. Only test crosses between individual plants with the phenotype $A \cdot B \cdot$ and plants with genotype aabb will give evidence about this.

N.B. By 'phenotype $A \cdot B \cdot$ ' is meant the phenotype due to genotype AABB, AaBB, AABb or AaBb.