1 The biological background

Quantitative genetics is the study of the inheritance of quantitative characters. It is based on the assumption that such characters are determined by genes which behave in the same way as the genes of major effect which control discrete characters. This introductory chapter is intended to give a brief review of the main concepts of classical, Mendelian genetics and of their application to quantitative characters. For further details the reader is referred to a standard textbook, such as Strickberger (1976) or Whitehouse (1973).

Mendel's experiments

The science of genetics began with Mendel's experiments on garden peas. As an example, we shall consider his experiments on flower colour. He had obtained from seedsmen two varieties of pea, one with purple flowers and the other with white flowers, and he verified that this character difference remained constant over several generations; one would expect different varieties to breed true since the pea is self-fertilizing. Mendel then crossed these two varieties by dusting pollen from one variety onto the stigma of the other; whichever way round this cross was done all the resulting plants in the next (F_1) generation had purple flowers. However, when these F_1 plants were allowed to self-fertilize, only three-quarters of the plants in the next (F_2) generation had purple flowers, and the remaining quarter had white flowers; the actual numbers of F_2 plants were 705 with purple and 224 with white flowers, giving a proportion of 0.759 with purple flowers.

Mendel's explanation of these results is shown diagrammatically in Fig. 1.1.

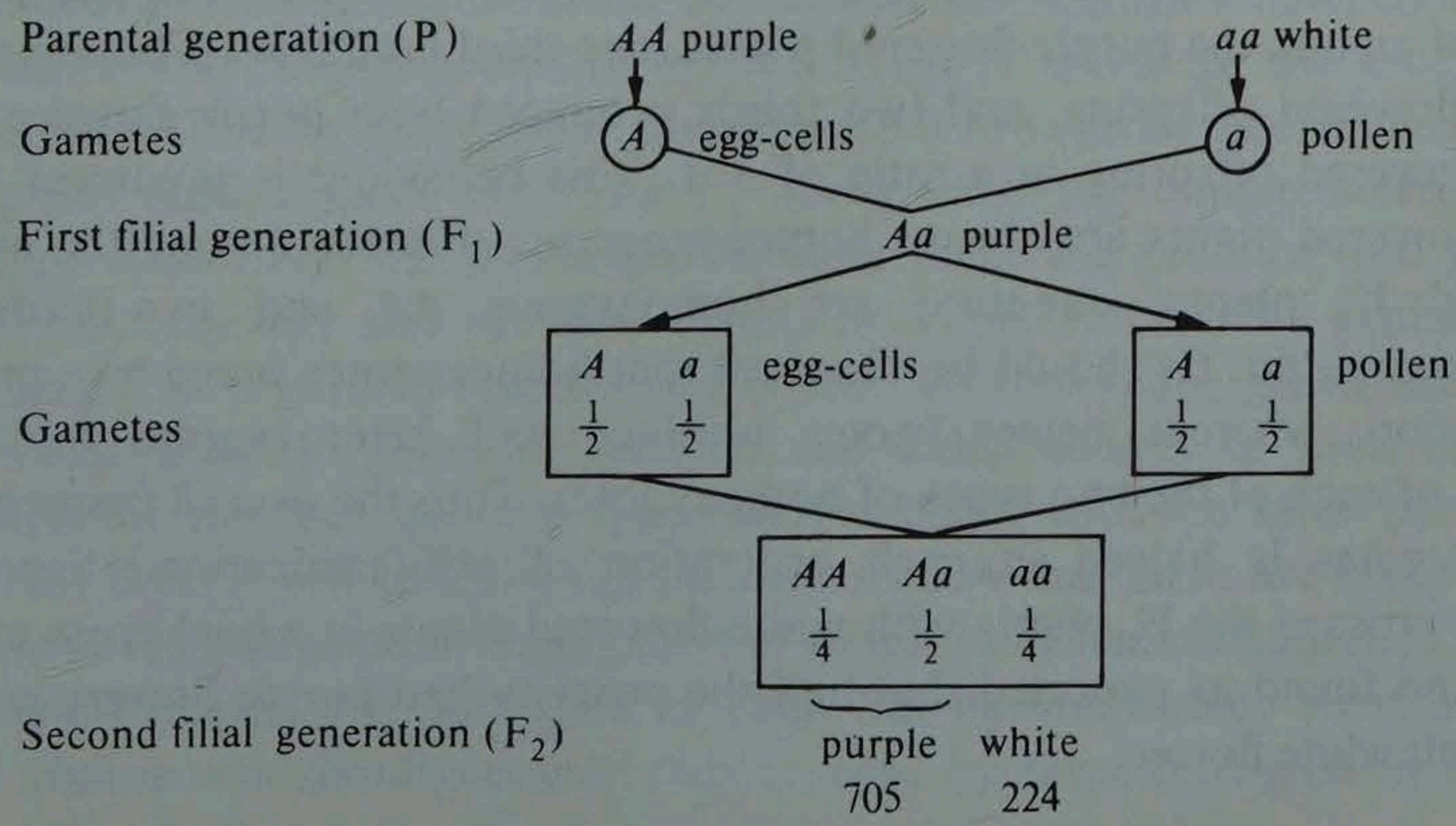

Fig. 1.1 Interpretation of Mendel's experiment on flower colour in garden peas.

2 The biological background

The following assumptions are made: (1) flower colour is determined by a factor (today called a gene) which can exist in two forms (alleles), *A* and *a*, determining purple and white flowers respectively; (2) each plant contains two of these genes, one derived from each of its parents, so that it can have one of the three possible genotypes, *AA*, *Aa*, or *aa*; (3) the gametes (egg-cells and pollen, or ova and sperm in animals) contain only one gene chosen at random from the two genes of the parent plant; (4) the egg-cells and pollen unite at random, independently of the genes they carry, to produce the plants of the next generation. The essential feature of this theory is the separation or segregation of allelic genes when the gametes are formed.

In a self-fertilizing species we expect all plants to be homozygous (either *AA* or *aa*), as will become clear shortly. Thus when the two true-breeding varieties were crossed at the beginning of the experiment the mating was between two homozygotes, $AA \times aa$; all the resulting plants must have had the heterozygous genotype, *Aa*, since they received *A* from the first and *a* from the second parent. In order to account for the fact that all these plants had purple flowers like the first parent, we must suppose that both *AA* and *Aa* plants have purple flowers and only *aa* plants have white flowers. Purple flower colour is said to be dominant and white recessive. A plausible explanation might be that the *A* gene leads to the production of a purple pigment, enough of which is produced to make the flowers purple whether the gene is present in single or in double dose. It is useful to draw a distinction between the genotype of a plant (*AA*, *Aa*, or *aa*) and its phenotype (purple or white flowers).

When the heterozygous F_1 plants were allowed to self-fertilize the mating was of the type $Aa \times Aa$. In this case egg-cells containing *A* and *a* will be produced with equal probabilities, and likewise for the pollen. Hence the probabilities of the genotypes *AA*, *Aa*, and *aa* in the next generation will be $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{1}{4}$ respectively, bearing in mind that the heterozygote *Aa* can arise in two ways (*A* egg-cell and *a* pollen, or vice versa). Thus three-quarters of them will have purple flowers and one-quarter white flowers, as observed.

Mendel confirmed this theory in two ways. Firstly, he allowed the F_2 plants to self-fertilize to produce an F_3 generation. All the white-flowered plants bred true, but among the purple-flowered plants, one-third bred true, producing only purple-flowered offspring, and two thirds produced both purple-flowered and white-flowered offspring in a ratio of 3:1. This behaviour is predicted, since white-flowered plants are always homozygous *aa*, whereas among the purple-flowered F_2 plants, one-third are homozygous *AA* and two-thirds are heterozygous *Aa*. (It should be observed that homozygotes breed true on self-fertilization, whereas heterozygotes produce half heterozygotes and one-quarter of each of the two types of homozygotes. Thus the overall frequency of heterozygotes is halved in each generation of self-fertilization.) Secondly, Mendel crossed the F_1 plants with white-flowered plants in a backcross experiment, and found as expected that half the progeny had purple flowers and the other half white flowers.

Mendel examined seven contrasting pairs of characters (such as round v. wrinkled seeds and tall v. short stem) in the same way, and he obtained similar results in each case, with one character being completely dominant to the other. He also investigated what happened when plants differing in two contrasting characters were crossed. Having established that round seeds were dominant to wrinkled seeds and yellow seed colour to green seed colour, he crossed a true-breeding variety having round, yellow seeds with a variety having wrinkled, green seeds; the resulting F_1 seeds were all round and yellow as expected. (Note that both these seed characters are determined by the genotype of the seed, not by that of the mother plant.) The plants grown from these seeds were then either allowed to self-fertilize to produce F_2 seeds or crossed to the variety with wrinkled, green seeds in a backcross. The results of these two experiments are shown in Table 1.1.

TABLE 1.1

Mendel's data on the joint segregation of seed shape and colour in the garden pea

		Yellow	Green	Total
F_2	Round	315	108	423
	Wrinkled	101	32	133
	Total	416	140	556
		Yellow	Green	Total
Backcross	Round	55	51	106
	Wrinkled	49	52	101
	Total	104	103	207

The marginal totals are in the ratio of 3 : 1 in the F_2 experiment and 1 : 1 in the backcross as expected. Furthermore, it is clear that the two characters are behaving independently of each other in the sense that there is no tendency for round seeds to be yellow and wrinkled seeds green, or vice versa. It can be concluded that the genes for these two characters segregate independently of each other when the gametes are formed. Write *A* and *a* for the genes determining round and wrinkled seeds, and *B* and *b* for the genes determining yellow and green seed respectively. The F_1 seeds have the double heterozygous genotype *AaBb*, and the plants grown from them will produce four gametic types, *AB*, *Ab*, *aB*, and *ab*. The results in Table 1.1 can be explained by supposing that a gamete receives one of the two seed-shape genes and one of the two seed-colour genes at random and independently of each other; in consequence the four gametic types should occur with the same probability of $\frac{1}{4}$. However, it will be shown in the next section that there are exceptions to the rule of the independent segregation of genes for different characters.

Mendel's paper was published in 1866 but its importance went unrecognized until it was simultaneously rediscovered by three biologists in 1900. It

was quickly shown that Mendel's theory could be used to explain the inheritance of many characters in both plants and animals, but that some extension of the theory was required. Thus it soon became clear that complete dominance, though common, was not universal. For example, roan cattle are heterozygous for a pair of alleles which in homozygous form produce either red or white coat-colour. Thus red cattle mated together produce red calves, white cattle mated together produce white calves, red bulls mated to white cows (or vice versa) produce roan calves, and roan cattle mated together produce red, roan, and white calves in an average ratio of 1:2:1.

Another complicating factor was the discovery of interaction (or epistasis) between genes. Bateson, Saunders, and Punnett (1905) crossed two white-flowered varieties of the sweet pea, and found to their surprise that all the F_1 progeny had purple flowers. On self-fertilization three flower colours appeared in F_2, purple, red, and white, in the approximate ratios of 27 : 9 : 28. These ratios together with further breeding data showed that flower colour was determined by three independently segregating genes *A*, *B*, and *C* with recessive alleles *a*, *b*, and *c*, respectively. Colour is only produced if both *A* and *B* are present (in single or double dose); the colour produced in non-white flowers is purple or red depending on the presence or absence of *C*. The experimental results can be explained by supposing that the original cross was of the type *AAbbCC* × *aaBBcc*. A possible biochemical explanation of how these genes act is shown below; it is supposed that each reaction can only occur in the presence of the appropriate gene.

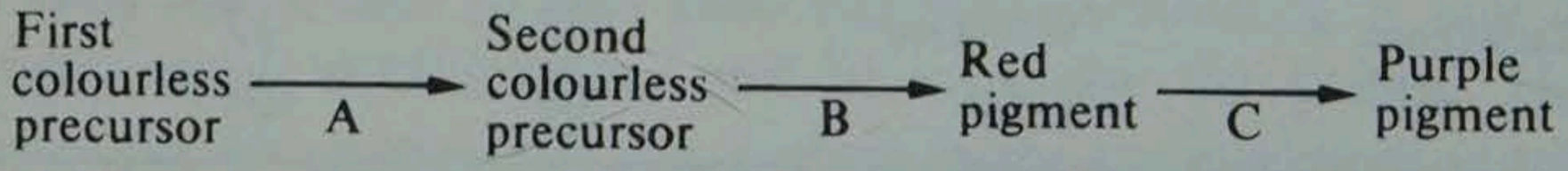

Fig. 1.2. Possible biochemical mechanism for the inheritance of flower colour in the sweet pea.

Chromosomes

The genes were originally hypothetical constructs invented to account for the results of breeding experiments. It was soon realized that there are physical objects, the chromosomes in the nuclei of cells, which behave in exactly the way in which genes are postulated to behave; it therefore became natural to suggest that the genes are carried on the chromosomes.

Briefly, every somatic cell nucleus contains a number of chromosomes which is characteristic of the species of animal or plant; there are 46 chromosomes in every human cell, 40 in the mouse, 12 in the housefly, 14 in the garden pea, 20 in maize, and so on. These chromosomes can be arranged in pairs, the members of each pair (homologous chromosomes) being physically identical, but distinguishable in shape, size, or some other characteristic from the members of any other pair. (The sex chromosomes are the exception to this rule; they will be discussed later.) In ordinary cell-division each chromosome is

replicated exactly, so that the daughter cells contain the same number of chromosomes as the parent cell. However, the gametes are produced by a special type of cell-division called meiosis in which each cell divides twice to produce four daughter cells, but the chromosomes are replicated once. In consequence, the chromosome number in the gametes is halved, only one chromosome from each homologous pair being represented in each gamete; such cells are called haploid. When two gametes unite to form the zygote the original (diploid) complement of chromosomes is restored. These cytological facts provide the physical basis for Mendelian inheritance (cf. Fig. 1.1).

The theory that the genes are carried on the chromosomes explained the phenomenon of linkage which had puzzled the early geneticists. The independent segregation of seed shape and colour in Table 1.1 can be explained by supposing that the genes controlling these factors are on different (non-homologous) chromosomes, which segregate independently at meiosis. After the rediscovery of Mendelism in 1900, it was soon found that most characters segregate independently of each other in this way, but that there are some exceptions, which can be explained by supposing that the corresponding genes are at different positions (loci) on the same chromosome; such genes are said to be linked.

As an example Table 1.2 shows the results of a backcross experiment on two seed characters in maize, coloured v. colourless and starchy v. waxy seeds. A pure-breeding variety of maize with coloured, starchy seeds was crossed to a variety with colourless, waxy seeds; all the resulting F_1 seeds were coloured and starchy, showing the dominance of these two characters. The plants grown from these seeds were then crossed back to the variety with colourless, waxy seeds with the results shown in Table 1.2. Both the marginal totals show good 1:1 segregations, but the two factors are clearly not behaving independently of each other.

TABLE 1.2

The joint segregation of two seed characters in a backcross experiment in maize (Bregger 1918; quoted by Whitehouse 1973)

	Starchy	Waxy	Total
Coloured	147	65	212
Colourless	58	133	191
Total	205	198	403

We denote the allelic genes for colour and lack of colour by *A* and *a*, and the genes for starchiness and waxiness by *B* and *b* respectively, and we suppose that these genes are at different loci on the same chromosome. The genotype of the F_1 seeds may be represented as *AB*/*ab*, indicating that *A* and *B* are on one chromosome, inherited from the first parent, while *a* and *b* are on the homologous chromosome inherited from the second parent. During meiosis

homologous chromosomes come together and lie side by side; breaks then occur, at the same randomly located places in both chromosomes, and the broken pieces rejoin with their opposite partners. This process is called crossing-over, and its effect is that the chromosomes in the gametes are mosaics of the homologous maternal and paternal chromosomes in the parent organism (Fig. 1.3). An even number of cross-overs between two loci has no net effect, but an odd number of cross-overs will change AB/ab into Ab/aB (or vice versa). This is called recombination, and the probability that it will occur is the recombination fraction, denoted by r. Thus the double heterozygote AB/ab will produce four types of gametes, the two recombinant types Ab and aB with probability $\frac{1}{2}r$ each and the two non-recombinant types AB and ab with probability $\frac{1}{2}(1 - r)$ each. In a backcross to the double recessive these four gametic types can be recognized directly; in Table 1.2 the two recombinant types are coloured waxy and colourless starchy, and the two non-recombinant types are the other two, so that the recombination fraction can be estimated as $r = (58 + 65)/403 = 0.31$.

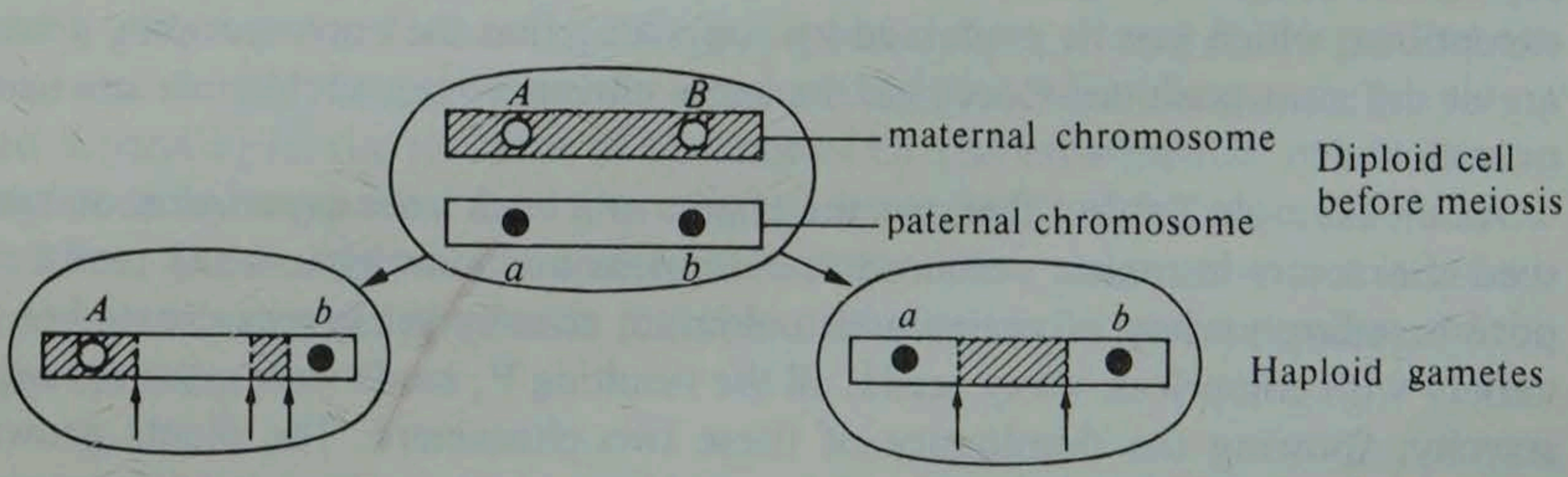

Fig. 1.3. Simplified diagram of the effects of crossing-over; arrows mark the points at which cross-overs have occurred.

In organisms which have been extensively investigated, such as the fruitfly *Drosophila melanogaster* and maize, it has been found that the genes fall into linkage groups, the genes in each group being linked to each other but not to genes in any other group. The number of linkage groups is equal to the haploid number of chromosomes, four in *Drosophila* and ten in maize; this fact provides strong support for the theory that the genes lie on the chromosomes. Furthermore, one would expect the recombination fraction between two linked loci to depend on the distance between them on the chromosome; loci close together will seldom recombine, whereas loci far apart on the same chromosome will behave almost as if they are unlinked, with a recombination fraction nearly $\frac{1}{2}$. Linkage data have been successfully used in several species to construct a consistent linear map of the chromosomes, in which the order of the genes and their approximate distances apart are determined.

We must now consider the sex chromosomes which have so far been ignored. In species which have separate sexes (most animals and a few plants) the sex of an individual is determined by a pair of sex chromosomes, denoted X and Y. In most species the female has two X chromosomes, whereas the male

has one X and one Y chromosome; the sex chromosomes in males are the exception to the rule that the chromosomes occur in homologous pairs. (The other chromosomes, which occur in homologous pairs in both sexes, are called autosomes.) All the ova produced by females contain a single X chromosome, but the male produces X-bearing and Y-bearing sperm in equal numbers; an X-bearing sperm when it fertilizes an ovum will produce an XX female zygote, whereas a Y-bearing sperm will produce an XY male zygote. (This is the typical genetic mechanism for determining sex, though there are several variants on it; for example, the situation is reversed in birds, butterflies, and moths, males being XX and females XY.)

The Y chromosome is genetically almost inert, apart from its role in determining sex, but the X chromosome has genes for other characters on it; such genes are said to be sex-linked. A well-known human sex-linked character is colour-blindness, whose inheritance can be explained on the following assumptions: (1) there is a pair of allelic genes, *C* and *c*, on the X chromosome determining colour vision; (2) in females normal colour vision is dominant to colour blindness, so that both *CC* and *Cc* women have normal vision and only *cc* women are colour-blind; (3) there is no analogue for this gene on the Y chromosome, so that men have only one gene for colour vision, *C* men being normal and *c* men colour-blind. This model accounts for the main features in the inheritance of colour blindness: (1) it is commoner in men than women because only one *c* gene is required to produce colour blindness in men whereas two *c* genes are required in women; (2) if a normal man marries a colour-blind woman, all their sons and none of their daughters are colour-blind; (3) if a colour-blind man marries a normal woman, she will usually be homozygous *CC*; in this case all their children will be normal, but their daughters will be heterozygous carriers, *Cc*, and if they marry a normal man, then half their sons (but none of their daughters) will be colour-blind.

The nature of the gene

The classical analogy is to liken the chromosome to a string of beads, each bead representing a different gene. This model incorporates the two main features of classical genetics, the particulate nature of genes and their linear arrangement on chromosomes. Though it must be modified in detail, the string of beads still provides a valuable physical model of how genes behave. However, modern work in molecular biology has greatly increased our understanding of what genes are and how they work at the biochemical level.

The main chemical constituent of nuclei is a complex substance called nucleoprotein consisting of DNA (deoxyribonucleic acid) associated with protein. It was at first assumed that the genetic information was carried in the protein fraction, but it became clear in the 1940s that it was in fact carried in the nucleic acid fraction of nucleoprotein; two lines of evidence were the transfer of genetic information by purified DNA in bacteria (bacterial

transformation) and the fact that the DNA content of cell nuclei was constant in somatic cells from different tissues and was exactly twice the content of the gametes. The breakthrough in understanding how DNA carries genetic information came with the elucidation of its molecular structure by Watson and Crick in 1953.

DNA is a very large molecule built of components called nucleotides; a nucleotide consists of one of four possible nitrogenous bases (adenine, guanine, thymine, or cytosine, usually abbreviated as A, G, T, or C) to which is attached a sugar (deoxyribose) and a phosphate group. Watson and Crick proposed that chromosomal DNA is composed of two strands of linked nucleotides, an A in one strand always being paired with a T in the other, and a G with a C. The base composition might therefore look as follows:

$$\begin{array}{l} \text{A A G T C G G T C} \ldots \\ \text{T T C A G C C A G} \ldots \end{array} \qquad (1.1)$$

This theory was based on building chemical models to interpret X-ray studies of crystalline DNA, together with the observation that although the proportions of the four bases varied in DNA from different species, the A content and the T content were always the same, as were the G and C contents.

The Watson–Crick model of the structure of DNA has two important biological implications. First, it immediately suggests a mechanism for the self-replication of DNA; if the two strands are split open, each will attract to itself a new complementary strand because of the mutual pairwise affinities of the bases. Secondly, it suggests that the genetic information in DNA is carried by the sequence of bases. How is this information translated into the observed phenotypic effects of genes?

It has been realized for a long time that genes exert their primary effect through the production of proteins, in particular of enzymes which catalyse the many chemical reactions taking place in the body. This has been summarized in the phrase 'one gene–one enzyme', which expresses the hypothesis that each gene is responsible for the production of a specific enzyme (though it should be extended to 'one gene–one protein' to include genes which are responsible for proteins which are not enzymes). For example, Garrod (1909) showed that the rare human disease alcaptonuria ('black urine disease') was due to a recessive gene and that it was caused by an inability to break down homogentistic acid so that this substance was excreted in the urine. He suggested that the fundamental defect was the absence of a specific enzyme for metabolizing homogentistic acid; this enzyme has now been isolated and its absence in alcaptonurics confirmed (see Harris 1975). We may suppose that the normal gene produces functional enzyme, and that the abnormal or mutant allele produces an abnormal form of the enzyme which is functionless; the heterozygote apparently has enough of the normal enzyme to function normally. The mechanism of the inheritance of flower colour in the sweet pea

in Fig. 1.2 can be interpreted in the same way, the dominant gene at each locus producing an enzyme which catalyses a specific reaction.

A protein is a large molecule built up of components called amino-acids linked together in a chain; twenty different amino-acids occur naturally. To understand how genes produce their effects, we must therefore discover how the linear sequence of nucleotide bases in DNA is translated into the linear sequence of amino-acids in the corresponding protein; in brief, what is the genetic code for translating the DNA message with its four-letter alphabet into the protein message with its twenty-letter alphabet? The answer turned out to be very simple in principle. The code is a triplet code, each triplet of three nucleotides coding for a specific amino-acid. For example, AAA and AAG both code for phenyalanine, GTT and GTC both code for glutamine, and any of the six triplets, AGA, AGG, AGT, AGC, TCA, TCG code for serine. Only one of the two strands of DNA is translated, and this strand is read consecutively in non-overlapping triplets. Thus the top strand in (1.1) would be translated into a protein containing the amino-acids phenylalanine–serine–glutamine . . .. There are also signals for starting and stopping the message. A typical protein contains about 150 amino-acids, so that a typical gene may be thought of as a sequence of about 450 nucleotides, beginning with a start signal and ending with a stop signal, together with the complementary chain which is not translated. The gene leads to the production of a particular protein, having a specific, determined sequence of amino-acids.

The above theory gives a new insight into the nature of mutations which produce new alleles and hence generate genetic variability. Occasional mistakes may occur in the replication of DNA. Such mistakes will usually involve single bases, so that for example AAG might become AGG in the twenty-third triplet of a particular gene; the twenty-third amino-acid of the corresponding protein would be changed from phenylalanine to serine. It seems likely that most mutations are of this kind. A well-known example is sickle-cell haemoglobin which differs from normal haemoglobin only in having valine substituted for glutamic acid in the sixth position (out of 146 amino-acids) in the beta chain. (Sickle-cell haemoglobin is less soluble than normal haemoglobin, which leads to severe anaemia in individuals who are homozygous for the sickle-cell gene; heterozygotes have a mixture of the two types of haemoglobin, which seems to give them some protection against malaria, and hence maintains the gene in the population by heterozygous advantage in malarial parts of the world.) It is clear that a very large number of different mutations is possible at each locus since a mistake may occur at any one of the nucleotide sites. It is an open question how many of them are likely to be severely deleterious through causing loss of biological activity and how many may be selectively neutral or even mildly beneficial.

Finally, we will consider dosage compensation at sex-linked loci. At autosomal loci the amount of gene product is proportional to the number of active genes; for example, heterozygotes for the alcaptonuria gene have about

half the enzyme activity for metabolizing homogentistic acid as normal homozygotes. If this rule held for sex-linked loci then females would produce twice as much gene product at all loci on the X chromosome as males, which would lead to metabolic imbalance between the sexes. In fact, males and females have the same level of activity for enzymes controlled by sex-linked loci. This phenomenon is called dosage compensation; it extends to all X-linked genes not involved in sex determination.

Different ways of achieving dosage compensation have evolved in different groups of animals. In placental mammals it is brought about by the inactivation of one of the two X chromosomes in each female cell at any early stage in the development of the embryo. It is a matter of chance whether the paternal or the maternal X chromosome is inactivated, so that female mammals are a mosaic, some cells having an active paternal and others an active maternal X chromosome. This is illustrated by the tortoise-shell cat which is phenotypically a mosaic of black and ginger fur (or tabby and ginger fur), and which is genetically a heterozygote for a sex-linked gene producing ginger fur. In kangaroos and other marsupials there is a similar mechanism but it is always the paternal X chromosome which is inactivated. In *Drosophila* it seems that dosage compensation has been achieved in a different way by the evolution of separate modifiers for different genes. These modifiers depress the activity of sex-linked genes in females so that each gene in the female produces half as much gene product as the single gene in the male.

Quantitative characters

The multiple factor hypothesis

Mendel's success in obtaining simple, clear-cut results derived in part from his careful choice of simple, discrete characters to investigate. When his experimental methods were applied to quantitative characters, results were often obtained which required a more complicated hypothesis to explain and more sophisticated methods to analyse. A good example is provided by an experiment of East (1916) on two varieties of an ornamental species of tobacco with different flower lengths. Table 1.3 shows the frequency distribution of flower length in the two parental variations, in the F_1 cross between then, and in the F_2 generation obtained by allowing F_1 plants to self-fertilize. Tobacco is normally self-fertilizing so that the experiment is comparable with Mendel's experiments. (Results for three different years have been pooled since there is no evidence of any difference between them.)

Since tobacco is self-fertilizing the two parents may be assumed to be homozygous at all loci; the F_1 plants will all be heterozygous at any loci at which the two parents differ. Thus the variability in the parents and in F_1 must be entirely of environmental origin, but the great increase in variability in F_2, which is typical of such experiments, can be attributed to the segregation of Mendelian genes. If only a single gene were involved, then the F_2 distribution

TABLE 1.3

Frequency distributions of flower length in a cross between two varieties of tobacco (East 1916)

Flower length (mm) (class centre)	P_1	P_2	F_1	F_2
34		1		
37		21		
40		140		
43		49		
46				
49				
52				3
55			4	9
58			10	18
61			41	47
64			75	55
67			40	93
70			3	75
73				60
76				43
79				25
82				7
85				8
88	13			1
91	45			
94	91			
97	19			
100	1			
Total number	169	211	173	444
Mean	93.1	40.4	63.5	68.8
Variance	4.9	2.3	7.9	41.6

should be a mixture of the distributions of P_1, F_1, and P_2 in proportions 1:2:1; thus the F_2 distribution should consist of three non-overlapping distributions which is clearly not the case. East and others suggested that the facts could be explained by supposing that the parents differed in several genes controlling the character; this became known as the multiple factor hypothesis.

Suppose that the two parents differ at n loci which affect flower length. Assume for simplicity that these loci are equivalent in their effect, that they act additively without dominance or epistasis, and that at each locus P_1 has two + alleles, each of which on average adds an amount a to flower length, while P_2 has two − alleles, which have no effect. Thus P_1 has $2n$ + alleles and P_2 has $2n$ − alleles, so that $2na$ can be equated to the difference between the two parental means:

$$2na = \bar{P}_2 - \bar{P}_1 = 52.7.$$

Every F_1 plant has n + alleles and n − alleles, so that the mean value in F_1 should be the average of the two parental means, which is approximately true. In F_2 the number of + alleles in different plants will follow a binomial

distribution with probability $\frac{1}{2}$ and index $2n$, provided that the loci are unlinked and so segregate independently of each other. Thus the mean value in F_2 should be the same as in F_1 (nearly but not quite true), whereas the variance should increase by an amount $\frac{1}{2}na^2$ because of the binomial variability in the number of + alleles between plants. Estimating the environmental variance in F_2 as 5.8 (calculated as a weighted average of the variances in P_1, P_2, and F_1 with F_1 receiving double weight), we obtain the equation

$$\tfrac{1}{2}na^2 = 41.6 - 5.8 = 35.8.$$

Solving these two equations we obtain the estimates

$$n = 9.7$$

$$a = 2.7.$$

East also gives the distribution of flower length in nine F_3 families derived from nine F_2 plants with different flower lengths. The means and variances of these distributions are shown in Table 1.4. There is a close relationship between the mean of an F_3 family and the flower length in its F_2 parent, which confirms that much of the variability in the F_2 plants is of genetic origin. (By contrast all F_2 families should have the same distribution regardless of the flower length of their F_1 parent because all the variability in F_1 is of environmental origin.) The variance within F_3 families is considerably smaller than the F_2 variance, but larger than the F_1 or the parental variances, as expected as a result of the progressive loss of genetic variability under continued selfing.

Thus the main features of East's results can be explained by supposing that the difference in flower length between the two varieties is of genetic origin and is controlled by about ten loci, each of which has a rather small effect. It can easily be seen how an effectively continuous, unimodal distribution is produced in F_2 under this model when a small amount of environmental variability is superimposed, in contrast to the non-overlapping classes typical of single gene inheritance. In consequence the statistical analysis of such polygenic characters

TABLE 1.4

Mean and variance of flower length in nine F_3 families

Flower length in F_2 parent	Mean of F_3 family	Variance of F_3 family
46	53.5	13.9
50	50.2	10.0
50	53.0	9.2
60	56.3	16.5
72	73.1	14.5
77	73.0	24.9
80	74.0	23.4
81	76.3	25.5
82	80.2	22.6

relies heavily on the calculation of means, variances and similar quantities used to characterize continuous frequency distributions. The above model is of course based on several simplifying assumptions, some of which will be relaxed in the more detailed discussion in Chapter 5.

Outbred populations

Experiments like the above presuppose the existence of homozygous true-breeding varieties (pure lines) between which crosses can be made. The occurrence of such pure lines is confined to naturally self-fertilizing plants and to artificial populations on which selfing, brother–sister mating or some other system of close inbreeding has been imposed by the breeder. In natural populations of many plants and most animals outbreeding is ensured either by the existence of separate sexes or by mechanical, physiological or genetic devices which prevent self-fertilization. An outbred population forms a single reproductive unit in which the genes are reshuffled each generation by recombination and segregation. Its structure is in marked contrast to an inbred population broken up into a large number of reproductively isolated pure lines.

The analysis of genetic variability of quantitative characters in outbred populations is based on observing correlations between relatives. The statistical theory of correlation and regression arose from Francis Galton's work on the inheritance of human stature shown in Fig. 1.4. Galton collected data on

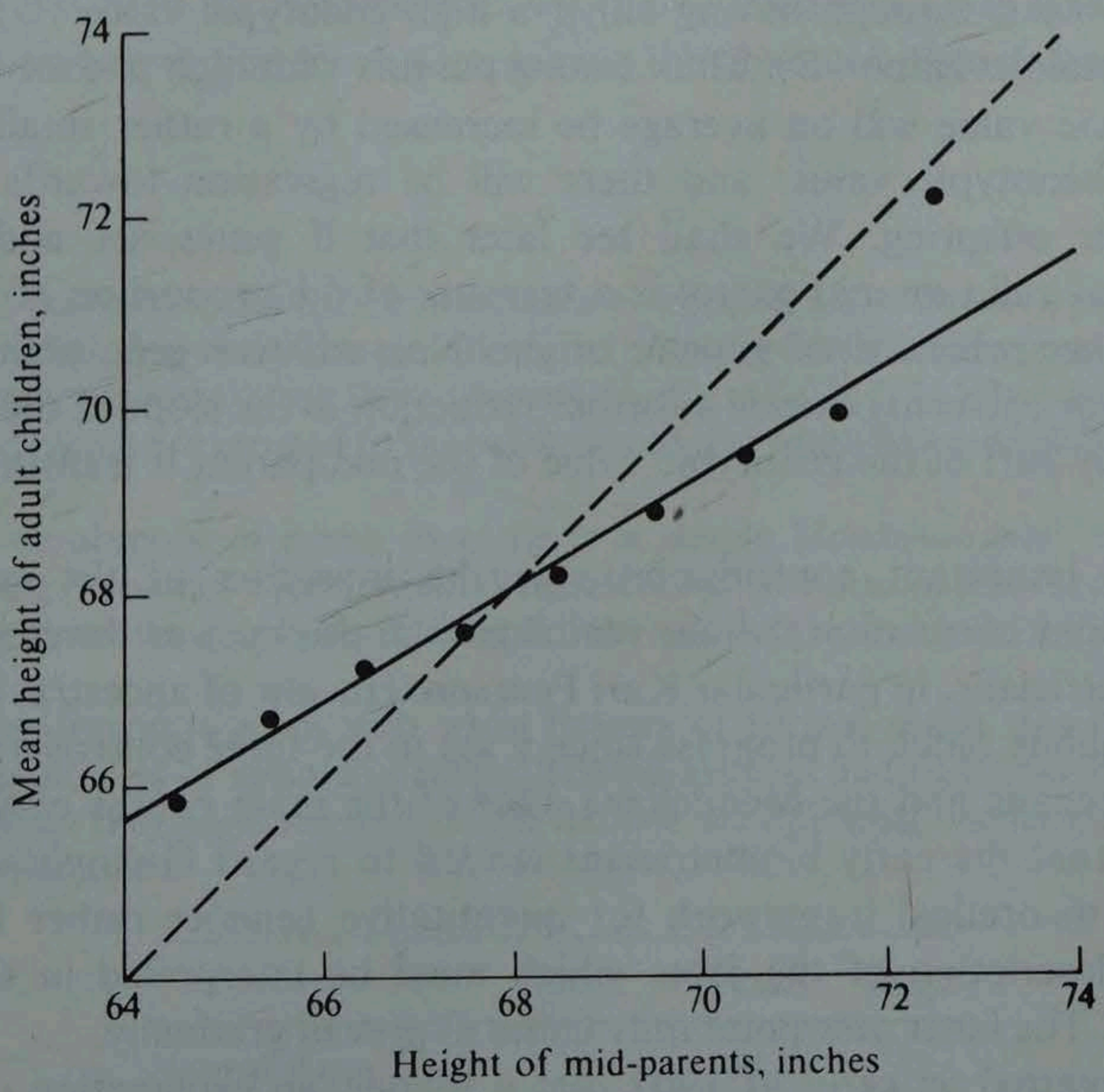

Fig. 1.4. Regression of height of adult children on their mid-parents (Galton 1887).

the heights of adult children and their parents in 205 families in 1884; after multiplying the female heights by 1.08 to make them comparable with male heights, he considered the relationship between the mean height of the children and the average height of their parents (their mid-parent). He found that there was a strong association between the heights of parents and children, but that the average deviation from the mean among the children was less than the corresponding deviation in their mid-parents; this is reflected by the differences between the observed points and the dotted line with unit slope in Fig. 1.4. Galton expressed this by saying that the children showed a 'regression towards mediocrity', and he estimated this filial regression (the slope of the solid line through the points) as 2/3–'that is to say the proportion in which the son is, on average, less exceptional than his mid-parent'. Galton was of course unaware of Mendel's laws. By a chain of rather dubious arguments he inferred from the fact of regression towards the mean his law of ancestral inheritance which he stated as follows: 'The two parents contribute between them on average one-half of the total heritage of the offspring, the four grandparents one-quarter, the eight grandparents one-eighth, and so on.'

Yule (1902) pointed out the ambiguity of this law and suggested a simpler explanation of regression towards the mean. Suppose that an individual's actual height is determined partly by his genotype and partly by the environment; we write $Y = G + E$, where Y is the observed phenotypic value, G is the genotypic value (the mean value among all individuals with the same genotype), and E is a deviation due to environmental factors. Now an individual may have a high phenotypic value through having either a high genotypic value (G) or a high environmental deviation (E). Thus among parents with high phenotypic values, the genotypic value will on average be increased by a rather smaller amount than the phenotypic value, and there will be regression towards the mean among their offspring. We shall see later that if genes act additively the regression of child on mid-parent is a measure of the proportion of the phenotypic variance which is of genetic origin. Non-additive gene action (due to dominance or epistasis) causes a further reduction in the slope of the regression because only part of the genotypic value of the mid-parent is transmitted to the offspring.

Galton's important contribution was the invention of the concepts of regression and correlation, whose mathematical theory was developed by the early biometricians, in particular Karl Pearson. His law of ancestral inheritance was a stumbling block to progress since it led to the futile controversy between the biometricians and the Mendelians. One of the main causes of this controversy was that the early biometricians tended to regard Galton's work as an alternative theoretical framework for quantitative genetics rather than as an empirical description of the facts which must be interpreted in the light of Mendelism. The latter viewpoint only came to prevail gradually.

Yule suggested as early as 1902 that a Mendelian explanation of Galton's work was possible. In 1904 Karl Pearson worked out a detailed theory for a

quantitative character determined by an arbitrary number of loci each with two equally frequent alleles of which one was completely dominant to the other. He first showed that on the assumption of random mating the frequencies of the three genotypes AA, Aa, and aa at each locus would retain the constant values of $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{1}{4}$ from one generation to the next. This is a special case of the law which bears the names of Hardy and Weinberg, who stated it independently in 1908, but which had already been foreshadowed by Yule (1902) and Castle (1903). Karl Pearson then derived the predicted regressions between relatives under the above model. He showed that the theoretical regression of offspring on one parent was a straight line with a slope of 1/3, and that the regression of offspring on mid-parent was a hyperbola, which tended to a straight line with a slope of 2/3 when the number of loci was large. He pointed out that these theoretical values were close to the actual values obtained by Galton, but that subsequent work had shown that the observed correlations were higher than this, and also varied from character to character. He concluded that the theory 'was not sufficiently elastic to cover the observed facts', but he did not attempt to generalize it since he considered complete dominance to be an essential feature of Mendelism. Yule (1906) pointed out that sufficient flexibility could be introduced into the theory by allowing dominance to be incomplete or absent and by introducing a term for the effect of the environment on the phenotype, but the precise formulation of the general theory did not come until 1918, when Fisher published his paper on 'The correlation between relatives on the supposition of Mendelian inheritance'. We shall return to this subject in Chapters 6 and 8. In the next chapter we shall consider in more detail the roles of genotype and environment in determining the observed phenotypic value.