PRENTICE-HALL FOUNDATIONS OF MODERN *Genetics* SERIES

Sigmund R. Suskind and Philip E. Hartman, Editors

AGRICULTURAL GENETICS *James L. Brewbaker*
GENE ACTION *Philip E. Hartman and Sigmund R. Suskind*
EXTRACHROMOSOMAL INHERITANCE *John L. Jinks*
DEVELOPMENTAL GENETICS* *Clement Markert*
HUMAN GENETICS *Victor A. McKusick*
POPULATION GENETICS AND EVOLUTION *Lawrence E. Mettler*
THE MECHANICS OF INHERITANCE *Franklin W. Stahl*
CYTOGENETICS *Carl P. Swanson, Timothy Merz, and William J. Young*

* Published jointly in Prentice-Hall's *Developmental Biology Series*

AGRICULTURAL GENETICS

James L. Brewbaker
University of Hawaii

PRENTICE-HALL, INC. Englewood Cliffs, New Jersey

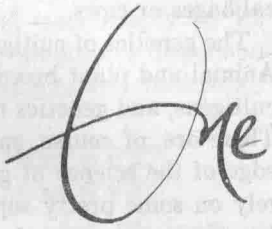
FOUNDATIONS OF MODERN GENETICS SERIES

Printed in the United States of America
Library of Congress Catalog Card Number: 64-16038

Agricultural Genetics
C-01877-P C-01878-C

PRENTICE-HALL INTERNATIONAL, INC., *London*
PRENTICE-HALL OF AUSTRALIA, PTY., LTD., *Sydney*
PRENTICE-HALL OF CANADA, LTD., *Toronto*
PRENTICE-HALL OF INDIA (PRIVATE) LTD., *New Delhi*
PRENTICE-HALL OF JAPAN, INC., *Tokyo*
PRENTICE-HALL DE MEXICO, S.A., *Mexico City*

Contents



Biological Variation

Hard cash paid down, over and over again, is an excellent test of inherited superiority.[1]

Putting inherited superiority to work is a major objective of agricultural research. Well over half the people on our procreative earth have too little to eat, and even the most profound knowledge of the gene provides little comfort to empty stomachs until it is translated into calories. Genetics has paid its way in calories in the past by demonstrating that inherited superiority can be located, transferred, and used in many cultivated plants and domesticated animals. Among perhaps thirty thousand animals and plants that enter the world's commerce, however, comparatively few species have received the benefit of controlled genetic advances. Challenging frontiers for agricultural genetic research are provided by new species, as well as by new environments for crops and animals, new pests and diseases, new uses in biochemistry, and a spiraling world population.

Cultivated plants and domesticated animals are known as cultigens. With increasing precision, man is controlling both the genetics and environments of these cultigens. Agricultural genetics is defined by its relationship to them, just as medical genetics is defined by a relationship

[1] Charles Darwin, *Animals and Plants under Domestication* (Appleton, 1897), vol. I, p. 447.

to man. The most economically important cultigens are vertebrate animals and flowering plants, to which most of the illustrations and discussions that follow are related. Of course, the basic principles of genetics are the same for a cultigen as they are for a weed. As a matter of fact agricultural genetics often attains its objectives more rapidly through studies of viruses and weeds than it does through studies of cabbages or cows.

The genetics of cultigens is not the same as the breeding of cultigens. Animal and plant breeding involve many arts and sciences relating to cultigens, and genetics may be considered the most important of these. There are, of course, successful breeders who have little or no knowledge of the science of genetics. However, today's breeder has come to rely on some pretty sophisticated genetic methods, a fact that should become apparent in the pages that follow.

Describing biological variation

The description of biological variation is a starting point for any biological inquiry. The geneticist must distinguish the genetic from the nongenetic components of the variation if he is to discern the mode of inheritance of a trait being studied. It is to this partitioning of variability that much of our agricultural genetic inquiry must be directed. To this end, the science of genetics has come to use more and more complicated tools of the statistician. Nowhere is this more evident than in studies of the economically important characters of cultigens.

Two facts underlie the importance of statistics to agricultural genetics. In the first place, many economic traits are governed by many genes, and are continuous rather than discrete in their variation. Second, it is often necessary in agricultural research to derive genetic information with maximum efficiency from small populations. In studies of milk production or fruit color, for example, considerations of economy and time dictate that we get the most information out of the fewest possible cows or apple trees. No attempt will be made here to describe in detail the methodology of statistical or quantitative genetics, but the understanding of results derived from the statistical genetic approach will require the clear understanding of several basic constants and statistical concepts.

The statistics used to describe biological variation include measures of average, dispersion, and relationships. Among those most often used are:

Measures of average: Mean
Measures of dispersion: Variance, Range, Standard deviation
Measures of relationship: Correlation, Covariance

The two most important measures are mean and variance. The mean, symbolized by $\bar{x}$, describes the average or central tendency of a population; it is calculated by summing (Σ) the observations (X) and dividing by the number (n) of observations summed.

$$\bar{x} = \frac{\Sigma X}{n}$$

The dispersion of a population can be indicated most simply by its range, or extreme values. In many instances, a population's dispersion is as important as its mean. Consider, for example, two basketball teams:

Team A: 6′1″, 6′2″, 6′3″, 6′4″, 6′5″ Mean = 6′3″
Team B: 5′5″, 5′6″, 6′3″, 7′0″, 7′1″ Mean = 6′3″

Although these two teams have identical mean heights, the range of team B is five times that of team A. Given his choice of teams, then, a newcomer would probably elect to play with team B, with its towering seven-footers.

Of greater precision than the range in describing the dispersion of a population is the variance (V, or σ^2), the average of the squared deviations (d^2) of individual observations from the mean:

$$\sigma^2 = \frac{\Sigma d^2}{n - 1}$$

Note that we divide by $n - 1$ rather than by n; this is to adjust the variance of a sample of observations for their mean. We say that a single degree of freedom has been lost from n by this adjustment. For ease in machine calculations, variance is calculated directly from original measurements by use of the formula

$$\sigma^2 = \frac{\Sigma X^2 - (\Sigma X)^2/n}{n - 1}$$

The variance is the measure most frequently used in statistical genetics, since it is the most valuable of the statistics that describe the variation of a biological trait.

The square root of variance is the standard deviation (σ), which expresses dispersion in the same units of measurement as does the mean. When it is related directly to the mean in per cent, it forms a useful coefficient of variation.

$$\text{Coefficient of variation} = \frac{\sigma}{\bar{x}}$$

Coefficients of variation can be used to compare different experiments, since they are independent of units of measure. Coefficients of variation smaller than 10 per cent are uncommon in biological data.

For the basketball teams A and B, for example, variance values are

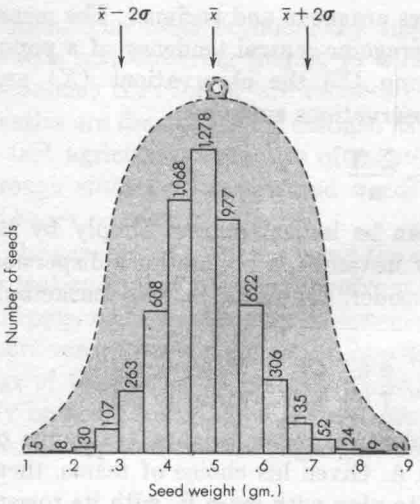


Fig. 1.1. Normal probability distribution of 5,494 kidney bean seeds, grouped according to seed weight. Based on experiments of Wilhelm Johannsen.

2.5 for team A and 90.5 for team B; standard deviations are 1.6 for team A and 9.4 for team B; and coefficients of variation are 2 per cent for team A and 13 per cent for team B. Each of these sets of values indicates the fact that the dispersion is much greater in team B than it is in team A.

The distribution of values about the mean for most biological characters assumes a characteristic bell-shaped pattern. This is illustrated in Fig. 1.1, which shows the weights of 5,494 kidney beans recorded by the great Danish geneticist Wilhelm Johannsen. The bell-shaped distribution can be duplicated by anyone willing to make 5,494 measurements of almost any biological trait. This type of distribution follows certain laws of probability and is known as the normal probability distribution. A normal distribution is symmetrical about the mean as midpoint, and 95 per cent of the population falls within the range from $\bar{x} - 2\sigma$ to $\bar{x} + 2\sigma$. In other words, values in a normal probability distribution exceed a deviation of 2σ, twice the standard deviation, less than 5 per cent of the time. The ratio of 95 per cent to 5 per cent, or of 19 : 1, is taken as safe betting-odds for much bioagricultural research. Thus, when values exceed a deviation of 2σ from the mean, the deviations are considered large enough to be significant, and the biologist may wisely seek a cause other than chance for the deviation. The normal probability distribution may be calculated empirically by expanding the binomial $(p + q)^n$, where n is infinitely large and $p = q$.

Measures of relationship include correlation and covariance. Each of these measures describes the change in one variable character as another one changes. Correlation is a measure of direct proportion. Thus, we say that height and weight of a growing animal are corre-

lated, since when one of these variables increases, the other also increases. The weight of a mature Holstein cow may be used accurately to estimate the size of its heart, since these traits have the high correlation of 0.94 (100 per cent correlation = 1.0). In contrast, the milk production of a fat Holstein is not necessarily more than that of a skinny one, reflecting the rather poor correlation (0.35) of milk production and body girth. Covariance is a measure of correlation among two or more individuals or populations. The degree of resemblance between any two individuals or populations is to a large degree a function of the genes they have in common. Genetic covariance statistically expresses this resemblance in terms of variances. A high degree of covariance or correlation may indicate that one variable can be used accurately to predict changes in another.

Phenotypic variation

> As has often been remarked, probably no two individuals are identically the same. All wild animals recognize each other, which shows that there is some difference between them; and when the eye is well practised, the shepherd knows each sheep, and man can distinguish a fellowman out of millions on millions of other men.[2]

Taken as a whole, the biological variation in any species is almost overwhelming. As Darwin noted, probably no two individuals (even "identical" twins) are wholly identical. Discrete genetic differences that bring hard cash to a rancher or farmer are usually quantitative and are often difficult to distinguish, measure, and evaluate. Discerning the heritable portion of this quantitative biological variation requires not only the "practised eye" (for which, however, there is no substitute) but also the best available statistical methods.

The total biological variation of a given trait is described statistically as its phenotypic variance (V_P). The components of phenotypic variance may be grouped into two major classes: genetic (V_G) and nongenetic or environmental (V_E). By definition, then,

$$V_P = V_G + V_E$$

The development of genetic laws has required the careful selection of traits in which V_E is minimized. Thus, Mendel reported in his published studies on peas that he was carefully avoiding phenotypes that showed "irregular or inconstant" variation. The thirty-four varieties of peas studied by Mendel segregated in many different characters, from which he chose only seven for his classic experiments. These seven characters, which were "constantly differentiating," were chosen because the phenotypic variation appeared to be largely genetic, with a

[2] Darwin, *Animals and Plants under Domestication*, vol. I, p. 361.

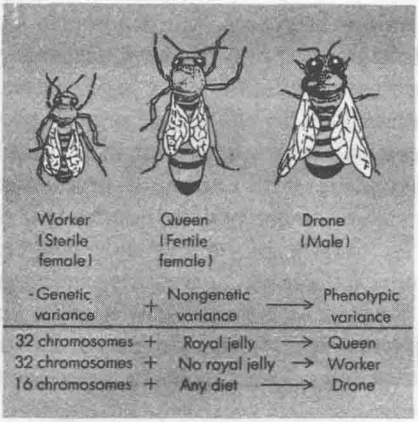


Fig. 1.2. Honeybees illustrate clearly that the genotype is only a set of potentialities upon which the environment must act to produce the phenotype we observe.

negligible nongenetic component. We may speculate that Mendel encountered greater nongenetic variation in mice and bees, which he studied in detail, but about which he never published his findings.

A detailed biochemical study of peas reveals immense variation that Mendel chose to overlook. A study of chlorophyll contents, for example, would reveal many degrees of seed color intermediate between Mendel's extremes of "pale yellow, bright yellow, and orange-colored" and "more or less intense green." It is perhaps fortunate that Mendel completed his experiments prior to this age of biochemical and statistical inquiry, else his F_2 ratio of 6,022 yellow and 2,001 green pea seeds might have been subdivided ad infinitum. At this point I encourage you to digress for thirty-seven minutes and read, if you have not done so recently, Mendel's classic paper "Experiments in Plant Hybridization." [3] These thirty-seven minutes will be among the best any student of genetics can spend, especially if he reflects while reading that the paper was written in 1865. He may also wryly meditate on the blind spots of Science, for this classic work evoked essentially no interest until its rediscovery and simultaneous substantiation in 1900 by Carl Correns, Erich von Tschermak, and Hugo de Vries.

Environmental variation

The living organism is constantly responding and adapting to its environment. In a broad sense, environment includes all intracellular and extracellular factors that influence the expression of the genotype. Any genetic description of a population of living plants or animals must, therefore, include observations (often quite detailed)

[3] Reprinted in *Classic Papers on Genetics*, J. A. Peters, ed. (Prentice-Hall, 1961), pp. 1-20.

about its environment, which are expressed in terms of environmental variance (V_E). V_E statistically incorporates all variation not directly attributable to segregating genes; hence it is often referred to as nongenetic variance.

The honeybee illustrates for us in a striking way the importance of nongenetic components in genetic analysis (see Fig. 1.2). The male bees (drones) arise from unfertilized eggs, irrespective of their diet during larval stages. The female bees (queens and workers) arise from fertilized eggs and vary greatly in their appearance, depending on their larval diet. The ability of so-called royal jelly to convert the female bee into a queen has been emphasized *ad nauseum*. It is perhaps more correct to consider the queen simply as a normal female, while the worker females are, in effect, starved queens who have been deprived of a decent diet. The workers have a short life span, are small in stature, and are usually incapable of laying eggs. Workers are thus impoverished expressions of their full genetic potentialities. The entire breeding and social structure of bees (and other insect species) therefore rests on the role of a nongenetic component, royal jelly, in the phenotypic expression of sex.

Environmental variance includes two major components, one intangible, the other controllable. Intangible variations include a statistical residue known as error and certain interactions of genetic and environmental components. Genetic experiments commonly are designed to minimize the variation from the controllable environmental component. Carried to practical extremes, this encourages the geneticist, for example, to produce germ-free chickens, to grow plants on nutrient agar, or to study microbes in nutritionally defined media.

Environmental variance can often be broken down so that the relative importance of controllable components can be assessed. Variance in the birth weights of mammals, for example, has been found to depend to a large degree on the environment, that is, on factors such as nutrition and health, of the mother. One analysis of phenotypic variance for human birth weights by L. S. Penrose gave the following results:

Source of variance	*Per cent of variance attributable to source*
Maternal environment	24
Maternal genotype	20
Age of mother	1
Numerical position of child	7
Intangible variations	30
Genotype of baby	18

The environment of the mother contributed more variance than did the baby's genotype in this study. Note that intangible variations contributed more than any other single component.

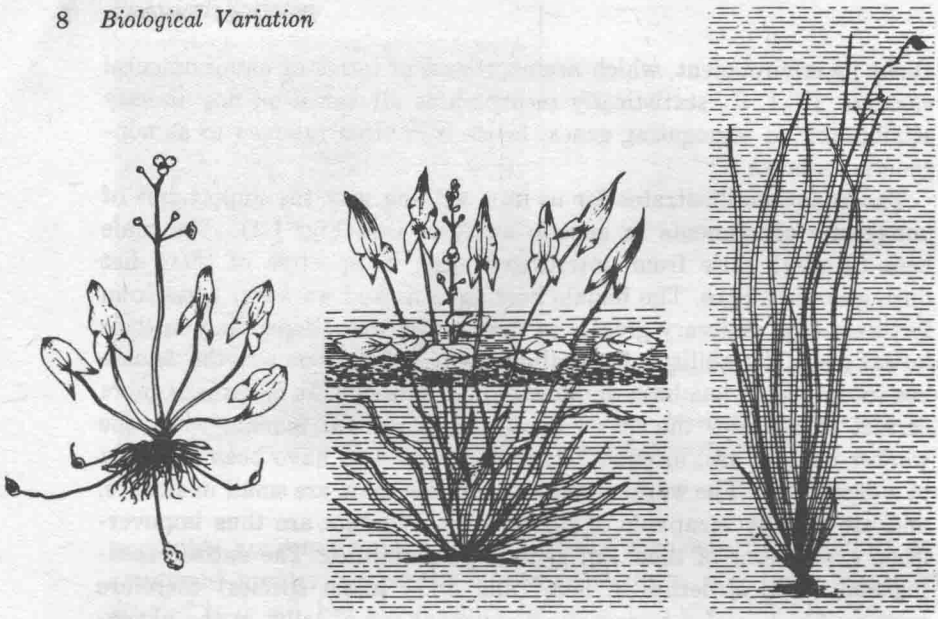


Fig. 1.3. The response of the arrowleaf to changes in its environment. Left, terrestrial growth; right, submersed growth; center, the arrowleaf as it is commonly found, partially submersed. From Bruce Wallace and Adrian M. Srb, *Adaptation*, 2nd ed. (Prentice-Hall, 1964), reproduced by permission of Prentice-Hall, Inc.

The environmental component of variance can be estimated most easily in studies of subjects or populations that have no genetic variance, such as identical twins or vegetative offspring of a single plant. In such studies, V_G equals zero, and V_P equals V_E. Populations with low genetic variance, such as highly inbred (pure) lines or the hybrids of two such lines, can also be effectively used. Caution must be exercised, however, lest the use of different, genetically uniform populations lead to different estimates of environmental variance. When tested under identical conditions, for example, corn hybrids show almost 30 per cent less environmental variance than do their inbred parents. This occurs in spite of the fact that their genetic components of variance are identical. Because of reasons explored in Chapter 5, inbreds interact more violently with environmental variations, while hybrids appear to be better "buffered" against these variations.

The adaptive responses of aquatic plants illustrate the unusual range of variations permitted by a single genotype in differing environments (see Fig. 1.3). Statistically, these variations in growth are entirely environmental; however, they reflect to a great extent the activity of certain genes in one environment and their inactivity or impaired activity in another. Environment acts as the "trigger" for the action of many genes of this type. Adaptive responses in segregating populations thus may reflect important interactions of genotype and environment. There are exceptions to every generalization, however,

and studies have shown that some genotypes possess an unusual adaptability, thriving in many different environments. Plant and animal breeders have recognized this adaptability in certain breeds, such as Golden Cross Bantam corn and Holstein-Freisian cattle, and have prized them accordingly. It is clear that certain genes or genetic systems contribute directly to such adaptability.

Genetic variation and heritability

Genetic variation arises from the contribution of segregating genes and their interactions with other genes. The term *heritability* (H) expresses the proportion of the total phenotypic variance that is genetic:

$$H = \frac{V_G}{V_P} \quad \text{or} \quad \frac{V_G}{V_G + V_E}$$

This proportion is commonly expressed in percentages; heritability is equal to 100 per cent whenever there is no environmental variance. As the environmental component of variance increases, the heritability drops. Genetic advance through selection is the primary objective of animal and plant improvement. Effective selection of genetically superior individuals requires that two conditions be met: (1) phenotypic variation must be adequate in the original population and (2) heritability must be sufficiently high for effective selection. In general, as heritability and phenotypic variations increase, genetic advance through selection (see Chapter 9) also increases.

The carefully designed experiments of Wilhelm Johannsen—from which we inherit our terms *gene, genotype,* and *phenotype*—laid the foundation for the interpretation of genetic advance under selection with continuously varying characters. At the time of the rediscovery of Mendel's paper, it was considered highly improbable that such characters could be governed by genes at all. Sir Francis Galton's studies of height and other continuously varying characters in man had given great quantitative precision to the ideas of variation and inheritance proposed by his cousin Charles Darwin, but Galton was unable to distinguish clearly the genetic and nongenetic components of variation—simply because *Homo sapiens* provided unfavorable material for this study. Like Mendel, however, Johannsen made an unusually judicious selection of species for his study of the common kidney bean.

Johannsen observed that the kidney bean seeds from any one plant or variety varied greatly in weight. Starting with 5,494 seeds of a single variety, Johannsen determined the average seed weight to be 495 mg. When plotted, the individual seed weights had the bell-shaped normal distribution shown in Fig. 1.1. A few of the largest and smallest beans from this population were sown, and 19 plants were grown. The

average weights of beans differed greatly among these 19 plants, ranging from a plant averaging 350 mg. per bean to a mammoth-seeded type averaging over 640 mg. per bean. When seeds were sown from the large-seeded plants, the offspring were similarly large seeded; offspring of the small-seeded forms were small seeded, etc. The 19 lines established from these original selections were grown for 6 generations.

By a simple but classic experiment, Johannsen showed that further selections in the 19 lines were not effective. To establish each of the 6 generations following his original selection, Johannsen planted a few of the largest and a few of the smallest beans from each line. (See Fig. 1.4, which shows data from one of the large-seeded selections.)

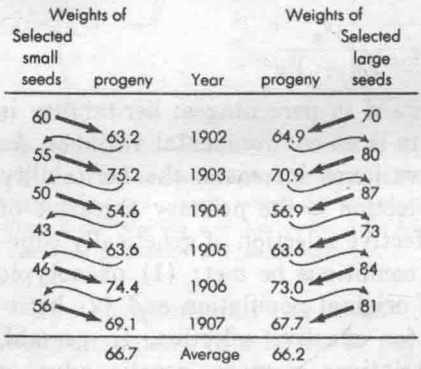


Fig. 1.4. The ineffectiveness of selection in a pure line, illustrated by bean seed weights in centigrams. Based on experiments of Wilhelm Johannsen.

With convincing regularity, the large and small sister seeds from any one line grew into plants having seeds of similar average weight. However, some variation in seed weight continued to occur each generation; Johannsen concluded that this variation was nongenetic. Note in Fig. 1.4 that the environmental variation from year to year was large, but affected large-seeded and small-seeded selections alike.

We can interpret Johannsen's results more easily knowing that kidney beans, like Mendel's peas, are regularly inbred by self-fertilization. *Continued inbreeding leads to genetic homozygosity.* The original kidney bean variety sampled by Johannsen was, in effect, a mixture of highly homozygous lines.[4] Selection of 19 seeds of different sizes had effectively separated this variety into 19 genotypically distinct lines. Self-fertilization led to no further genetic variability in the 19 lines, which Johannsen referred to as pure lines. We should not infer that

[4] It is customary to qualify the term *homozygous* when referring to the entire genotype; e.g., the "highly homozygous" line is one in which a great majority of the loci are homozygous.