Appendix A
Evolutionary and Genetics Principles

A.1 Genetics

As a start, we need to distinguish between *learning*, which we take to mean behavioral adaptation by an individual, and *evolution*, which refers to adaptation by an entire population (or system) by the accumulation of changes among the individuals that make up the population.

The word *genetics* derives from the Greek root $\gamma\epsilon\nu\nu\omega$, which means "to become" or "to grow" [1]. The classical definition of genetics introduced by William Bateson, who named the field of study in 1906 [2], is:

> Genetics is the science dealing with heredity and variation seeking to discover laws governing similarities and differences in individuals related by descent [3].

In short, genetics is the science of inheritance [4]. In higher organisms, genes are arranged on *chromosomes*. The *gene* is a chemical entity that influences an inherited trait from parents to their children. The location of a gene on a chromosome is termed its *locus*. Each gene, at a given *locus*, can take various forms called *alleles*. *Haploid* species have a single copy of each chromosome. *Diploid* species have two copies, one from each parent. Diploid species have two alleles for each locus. For example, if the color of an animal is either white, black, or gray, it is determined according to a pair of genes with the alleles "w" for white and "b" for black. If the pair of genes is the same (e.g., "ww" for a white animal or "bb" for a black one), the organism is called *homozygous* otherwise *heterozygous* (e.g., "wb" for a gray animal). If the animal is heterozygous and an allele conceals the effect of its pair member (e.g., the pair of genes is "wb" and the color is black), it is called *dominant*, "b", otherwise *recessive*. Blood groups, for instance, are determined by a single locus and fall into A, B, and O types. The alleles A and B are dominant over O, but are codominant when they occur together. Thus, the combinations AO and AA produce type A, BO and BB produce type B, AB produces type AB, and OO produces blood type O.

D. G. Green et al., *Dual Phase Evolution*, DOI: 10.1007/978-1-4419-8423-4,

A group of chromosomes constitutes the organism's *genotype* which is the genetic constitution or sometimes called the genetic makeup [5]. The observable traits of an organism constitute its *phenotype* [4].

Most importantly, the phenotype arises by a process of development from an individual's inherited genotype. Each gene contributes to the formation of phenotypic characters. In some cases, a single gene determines a phenotypic character, but in most cases, the genes function as elements of *genetic regulatory networks* (GRNs) that control the developmental processes by which phenotypic characters form. *Controller genes* determine when (and whether) a GRN becomes functional during growth and development.

The science of genetics was discovered by the Austrian monk Gregor Mendel in 1865 [6]. Mendel introduced two laws that were rediscovered in 1900 by the European botanists Correns, De-Vries, and Tschermark. The first is the law of *"segregation"*, which defines the amount of genetic materials transmitted during a mating. It states that during combination, a gene pair separates into the reproductive or sex cells formed by an individual, so that half carry one member of the gene pair and half carry the other [3]. The second is the law of *"independent assortment"*, which states that "*during gamete formation, segregating pairs of unit factors assort independently of each other*" [7]. *Unit factors* [7] are particulate for each trait which serve as the main units of heredity and are passed unchanged from generation to generation, determining various traits expressed by each individual. From these two laws, we can define what is called a *Mendelian sampling*. Since the segregation of genes occurs independently, and half of the gene pair is transmitted to the offspring during the mating, there is no guarantee which gene will be transmitted. Accordingly, although we know that the progeny has one half of their father's and their mother's genes, siblings will differ since they will take different combination of genes. The variation due to this sampling causes variation within full-sib families and is indicated as the Mendelian sampling.

In 1908, British mathematician G.H. Hardy and German physician W. Weinberg introduced the *Hardy–Weinberg* [8, 9] *equilibrium law* [6], forming the basis of population genetics. The law assumes an absence of selection and states that the process of heredity changes neither allelic frequencies nor genotypic frequencies at a given locus within a very large, closed, randomly breeding population. Additionally, the *equilibrium genotypic frequencies* at any given locus are attained in one single generation of random mating whenever the allelic frequencies are the same in the two sexes.

Suppose that at a particular locus, there are two alleles P and Q, with proportions p and q, respectively. Assuming random mating and complete mixing within the population, then after a single generation, the proportions of genotypes within the population will reach equilibrium and the proportions of genotypes will be

PP	PQ	QQ
p^2	pq	q^2

Both mutation and selection can change the proportions of an allele in an otherwise stable population. Mutation introduces new alleles into a population. Selection biases the reproduction process, so that the proportion of a favoured allele can increase from generation to generation.

Mendelian genetics studies the principles of transmitting the genetic material from the parents to the offspring generation. *Population genetics* is the study of Mendelian genetics in populations. It is concerned with the frequencies of genotypes and phenotypes [5] and is limited to the inheritance of qualitative traits. *Quantitative genetics* was introduced by Fisher in 1918 and is concerned with the study of the effects of individual genes [3], as well as the study of correlation and regression between the genetic and phenotypic values. As opposed to population genetics, it concentrates on quantitative traits.

The *qualitative traits* [4] are those traits controlled by one or a few loci, in a way that allele has a marked effect on the phenotype and individuals can be phenotypically classified into one of a group. For example, the human blood groups designated A, B, O, or AB are determined by three types of alleles denoted IA, IB, and IO. The blood group of any person is determined by the particular pair of alleles present in his or her genotype. In *quantitative traits*, there are many loci, a gradation of phenotypes, and small effects of single alleles. Quantitative traits usually follow a normal distribution [5] and can be found in three categories, *continuous*, *meristic*, and *threshold* traits [6]. Continuous traits vary, with no clear-cut breaks, from one phenotypic extreme to the other, such as milk production in cattle. In meristic traits, the phenotype is determined by such counting as the number of eggs laid by a hen. Threshold traits have only two or few phenotypic classes, but their inheritance is determined by the effects of multiple genes together with the environment, such as twining in cattle.

The phenotypic value of an animal trait [2] is measured by the difference between the animal's trait value and an appropriate base group. This base group can be the average of all animals within the herd born in the same year, the herd average at the breeding program's commencement, or any other appropriate base. The phenotypic value (P) [2] of an animal for a specific trait is the sum of the animal genotypic value (G) and the environmental deviation (E); that is

$$P = G + E \tag{A.1}$$

The environmental deviation is a term used in the field to represent the effects of all non-genetic factors such as seasons, feeding systems, and management. An animal's genotypic value can be further decomposed into three measurements: the breeding or the additive genetic value (A), the dominance deviation (D), and the epistasis or interaction deviation (I).

$$G = A + D + I \tag{A.2}$$

Gene action is said to be additive if the differences between the heterozygote and the two homozygotes are equal [2]. Inheritance depends on the additive effects of genes and represents the value of an individual's genes to its progeny. Generally, the

offspring inherits only the average additive component of their parents. The dominance deviation causes a heterozygous animal to be more like one of the homozygous genotypes. For example, assume a homozygous black cow is worth \$4 and a homozygous red cow is worth \$2. If the gene for black color is dominant, a heterozygous cow will look black and will be worth \$4 although it should be worth only \$3 ($\frac{4+2}{2}$). The \$3 represents the additive genetic value, and the additional \$1 represents the dominance deviation. The epistasis is a measure of the effect of other genes on the gene in question; that is, sometimes the expression of an allele at a locus requires the presence of a particular allele at another locus. Now, we can rewrite the equation for the phenotypic value as follows:

$$P = A + D + I + E \tag{A.3}$$

A.2 Inbreeding

The relationship between animals is measured using the *inbreeding coefficient*, F, which measures the probability that both genes at a locus are identical by descent [3]. It represents the probability that two alleles will have arisen from the replication of the same piece of DNA in the common ancestor. These alleles could be at a single locus in one individual (in which case the individuals are said to be inbred) or they may be either one of two alleles present at the same locus in each of two individuals (in which case the individuals are said to be relatives). All those identical by descent alleles are alike in state, that is they occur at the same locus and are of the same type, although the reverse is not true. Therefore, two alleles may be alike in state by chance and not necessarily because they are identical by descent. Therefore, we usually need to define a reference or base population when calculating inbreeding.

If we assume that an animal has an inbreeding coefficient of 50 %, this indicates that 50 % of all loci in this animal are expected to be identical by descent. Other measures of inbreeding are the coefficient of relationship, f_{xy}, and the coefficient of coancestry, $F_{\text{Coancestory}}$, or the coefficient of kinship [5]. The former measures the likelihood that two individuals carry alleles that are identical by descent. The latter is the probability that the same two individuals will both produce gametes that carry identical by descent alleles. That is, the coefficient of coancestry is half the coefficient of relationship and equals exactly the expected coefficient of inbreeding for the progeny if these two animals were to be mated. Assuming that there is no selection, the inbreeding rate per generation, ΔF, is approximated classically by the following equation [10], assuming equal progeny per parent:

$$\Delta F = \frac{1}{8N_m} + \frac{1}{8N_f} \tag{A.4}$$

where N_m and N_f are the number of males and females, respectively, entering the population every year. The average coancestry among a group of offspring can be

calculated using the following equation [11]:

$$F_{Coancestory} = X^t A X \tag{A.5}$$

where X is a vector of the proportions of the contributions made by each parent (with male and female part adding to 0.5 each) in the breeding system, that is, the proportion of matings for this animal, and A is the numerator relationship matrix, which indicates the additive genetic relationship between these parents. Henderson [12] introduces a recursive function for calculating the matrix A. The algorithm depends on ordering the animals in the pedigree such that parents precede their progeny, and then, the following rules are applied for animal i where the inbreeding coefficient for animal i is simply; $F_i = a_{ii} - 1$:

- If both parents, s and d, are known, $a_{ji} = a_{ij} = 0.5(a_{js} + a_{jd})$; $j = 1, 2, \ldots, (i-1)$ and $a_{ii} = 1 + 0.5a_{sd}$.
- If one parent, s, is known, $a_{ji} = a_{ij} = 0.5a_{js}$; $j = 1, 2, \ldots, (i-1)$ and $a_{ii} = 1$.
- If both parents are unknown, $a_{ji} = a_{ij} = 0$; $j = 1, 2, \ldots, (i-1)$ and $a_{ii} = 1$.

Mating of close relatives should be avoided in breeding programs since inbreeding increases the fraction of homozygous loci (i.e., it increases expression of lethal factors). It is further estimated that each percentage of inbreeding leads approximately to a 50 kg weight decrease in milk production [2] and in more general terms, the inbreeding depression on quantitative traits. Mating of unrelated animals always results in non-inbred progeny. If the parents are inbred but are unrelated, then their progeny are not inbred. Different effects of inbreeding are summarized below from [3]:

1. It decreases the frequency of heterozygotes and increases the frequency of each homozygote by half that frequency (i.e., the population becomes more homogeneous), but it does not change the frequency of alleles. When inbreeding is at its maximum, the population is completely homozygous.
2. It increases the chance of expression of recessive lethal factor, which would allow culling of affected and carrier animals and thereby reduce the frequency of the detrimental genes. However, the cost must be balanced against the potential gain.
3. Line crosses resulting from matings between inbred lines would have mostly heterozygous loci and therefore might be superior to non-inbred animals if there is some form of dominant gene action.
4. It can be used to fix a desirable type (if the reproductive rate is sufficient to allow selection to eliminate the undesirable genes) and to achieve greater uniformity.
5. Inbreeding within a population leads to a loss of genetic variation, and therefore, a loss of future reliability to make genetic change.

The expected additive genetic merit a_i of the progeny i resulting from mating sire s_i and dam d_i can be calculated as [13]:

$$a_i = \frac{1}{2}a_{s_i} + \frac{1}{2}a_{d_i} + \phi_i \tag{A.6}$$

$$var(a_i) = \frac{1}{4}var(a_{s_i}) + \frac{1}{4}var(a_{d_i}) + \frac{1}{2}cov(a_{s_i}, a_{d_i}) + var(\phi_i) \tag{A.7}$$

$$\text{var}(\phi_i) = \frac{1}{2}[1 - \frac{1}{2}(F_{s_i} + F_{d_i})]\sigma^2_{a_i} \tag{A.8}$$

where ϕ_i is the Mendelian sampling, a_{s_i} and a_{d_i} are the additive genetic values of the sire and the dam, respectively, var(a_s) and var(a_d) are the additive genetic variance for both the sire and the dam, respectively, cov(a_s, a_d) is the covariance between the additive genetic values of the sire and the dam, var(ϕ) is the variance of Mendelian sampling, F_s and F_d are the inbreeding coefficients for the sire and dam, respectively, and σ^2_a is the additive genetic variance for the population.

A.3 Heritability

A basic concept in genetics is *heritability*. To illustrate this concept, the concepts of phenotypic and genotypic variances are presented. The phenotypic variance, V_P, is the sum of the population genotypic variance, V_G, and the environmental variance, V_E; that is

$$V_P = V_G + V_E \tag{A.9}$$

where

$$V_G = V_A + V_D + V_I \tag{A.10}$$

and V_A is the additive genetic variance, whereas the sum of V_D and V_I is the non-additive genetic variance (dominance and epistasis). Heritability, h^2, is the ratio between the additive genetic and phenotypic variances; that is

$$h^2 = \frac{V_A}{V_P} \tag{A.11}$$

Since V_A is always less than V_P, the heritability is always in the range [0,1]. The higher the heritability, the more important it is for the breeder to use breeding systems, which utilize the additive genetic variation. One way, although not necessary the best, to estimate the heritability of a trait is to regress the offspring on the parents, that is, the correlation coefficient between the offspring and the parents phenotypic values for the trait. However, it is better to use multitrait animal models because they use all the available information from relatives.

A.4 Variation and Random Drift

The Hardy–Weinberg Law shows that in a sufficiently large population, the genetic makeup will remain constant in the absence of mutation and selection. The effect of population genetics is therefore to inhibit random genetic variations (*random genetic drift*). However, this effect is predicated on the population being large, as well as having complete mixing and random mating. Deviations from any of these assumptions can lead to genetic drift.

According to the Wright–Fisher model of genetics for diploid populations [14], the abundance of any allele from one generation to the next follows a binomial distribution. So, if an allele Q has frequency q in a diploid population of size N (giving $2N$ alleles at the locus), then the probability p of finding m copies of the allele in the next generation is given by

$$p = \frac{(2N)!}{m!(2N-m)!} q^m (1-q)^{2N-m} \tag{A.12}$$

In a small population, repeated random sampling of an allele is effectively a random walk, in which the allele's frequency may reach either 0 or 1 as absorbing states. An allele is said to become *fixed* in a population when its frequency is 1. That is, it provides 100 % of the alleles found at its locus. This can occur when an alternative allele disappears from a population by random drift. Alternatively, an allele can disappear from the population. If an allele Q has frequency q in a diploid population of size N (giving $2N$ alleles at the locus), then most of the terms in the above binomial formula reduce to 1, so the probability of Q disappearing in the next generation is given by

$$p_{Extinction} = (1-q)^{2N} \tag{A.13}$$

From this, we can deduce that $p_{Extinction}$ is greatest when q and N are both small. Moreover, in a small population, q could readily decrease over a sequence of generations, making extinction ever more likely.

New alleles in a population appear either by introduction from other populations or by mutation. At the most fundamental level, mutations appear by changes in the DNA sequence. This can occur by various means, especially errors during DNA replication, but also including genetic transposition ("jumping genes") and the effects of radiation.

If Q is a single new allele introduced into a population by mutation (or gene flow), then its initial frequency is $q = \frac{1}{2N}$. Substituting for q, the previous equation becomes

$$p_{Extinction} = \left(\frac{2N-1}{2N}\right)^{2N} \tag{A.14}$$

The value of $p_{Extinction}$ in this case is just over $\frac{1}{3}$ for a small population, and as N increases, the value converges very slowly to 0.368. The implication is that single new alleles are most likely to be eliminated within the first few generations.

For a small population, there is a chance of a random mutation becoming fixed in the population. However, extinction becomes almost certain as large population size increases because the ratio $\frac{(2N-1)}{2N}$ grows ever closer to 1.

In reality, populations are usually distributed across a landscape, so complete mixing is often impossible in any single generation. This means that the effective local breeding population is much smaller at any given location. However, gene flow will occur throughout the population over a number of generations.

If a population becomes fragmented (for instance, by the appearance of barriers to migration), then it reduces to a set of small, isolated subpopulations (a *meta-population* [15]). Under these conditions, two important changes occur. First, random mutations have a better chance of becoming fixed within individual subpopulations. Second, random drift can lead the subpopulations to diverge from another, so that genetic diversity of the overall population increases.

The above processes also lead to the *founder effect*. Suppose that the genotype of an individual contains alleles that allow it to occupy some area (or exploit some resource) that was unavailable to others. Then that individual's offspring will occupy the new resource and the frequency of the allele will increase and can become fixed within subpopulations occupying the new area.

References

1. A.M. Winchester, *Genetics: A Survey of the Principles of Heredity* (Houghton Mifflin Company, Boston, 1966)
2. G.H. Schmidt, L.D. Van-Vleck, *Principles of Dairy Science* (W.H. Freeman and Company, San Francisco, 1974)
3. L.D. Van-Vleck, E.J. Pollak, E.B. Oltenacu, *Genetics for the Animal Sciences* (W.H. Freeman and Company, New York, 1987)
4. L.H. Daniel, *Genetics*, 3rd edn. (Jones and Bartlett Publisher, London, 1994)
5. D.S. Falconer, *Introduction to Quantitative Genetics*, 3rd edn. (Longman Group Limited, London, 1989)
6. J.A. Franisco, A.K. John, *Modern Genetics*, 2nd edn. (The Benjamin/Cummings, Menlo Park, 1984)
7. W.S. Klug, M.R. Cummings, *Genetics* (Prentice-Hall, New Jersey, 1999)
8. G. Hardy, Mendelian proportions in a mixed population. Science **28**(706), 49–50 (1908)
9. W. Weinberg, Über den nachweis der vererbung beim menschen. Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg **64**, 368–382 (1908)
10. S. Wright, Evolution in mendelian populations. Genetics **16**, 97–159 (1931)
11. N.R. Wray, M.E. Goddard, Increasing long-term response to selection. Genetique Selection and Evolution **26**, 431–451 (1994)
12. C.R. Henderson, Use of all relatives in intraherd prediction of breeding values and producing abilities. Dairy Sci. **58**, 1910–1916 (1975)
13. J.H.J. Van der Werf, Models to estimate genetic parameters in crossbred dairy cattle populations under selection. Ph.D. thesis, Wageningen Agriculture University, 1990
14. R.A. Fisher, *The Genetical Theory of Natural Selection* (Clarendon Press, Oxford, 1930)
15. I. Hanski, M. Gilpin, Metapopulation dynamics: brief history and conceptual domain. Biol. J. Linn. Soc. **42**(1–2), 3–16 (1991)

Index

D. G. Green et al., *Dual Phase Evolution*, DOI: 10.1007/978-1-4419-8423-4,