Chapter 1
The Theoretical Principles of Population Genetics

Before we turn to an analysis of genetic processes in salmon populations, it will be useful to introduce some general theoretical principles of population genetics. Population genetics, the most theoretically advanced field of genetic research, occupies a special place among the fields of modern genetics and biology as a whole.

In population genetics, quantitative theory has existed for a long time and continues to improve; it is a theory of how natural or anthropogenic factors influence the stability of population changes in important biological traits over successive generations. In other words population genetics is expressed in mathematical models. The models vary in their relevance to biological systems, but they are important, since they permit research to be planned in a particular way. If a model conforms with the natural situation, the changes in populations can be quantified and their possible consequences predicted. Because a large number of diverse papers and books have been published on this theme in recent years, we shall examine only the main population genetic terms, models and approaches. We base our further discussion on works by Sewall Wright (1931, 1951, 1969, 1977, 1978), Neel & Schull (1958), Ehrlich & Holm (1963), Dobzhansky (1970), Kimura (1983), Kimura & Ohta (1971), Cavalli-Sforza & Bodmer (1971), Nei (1975, 1987), Li (1976) and several others mentioned in the text.

1.1 Estimation of gene frequencies

As a first approximation, a population may be defined as *an aggregate of freely interbreeding individuals that share a common gene pool* (Dobzhansky). Because the number of segregating loci in the genome is large, one can understand the difficulties that confront the researcher in attempting to give an adequate description of this pool of inherited information. At the same time though, it is evident that, however great these difficulties may be, there is only one way of obtaining such a description, which is by defining the frequencies of allelic genes at each single locus. Knowledge of the spatial and temporal distribution of gene frequencies enables a

quantitative assessment of how genetic processes in populations are influenced by given external and internal factors.

There are several carefully formulated methods of evaluating gene frequencies in a population. We shall examine two of them, the first applied to a situation without dominance and the second to inheritance with dominance. The situation relates to a pair of alleles at a single autosomal locus.

1.1.1 Absence of dominance

Let us assume that of N diploid individuals N_1 carry only the allele A, that is, are AA, N_2 the heterozygotes AB, and N_3 the homozygotes BB, so that $N_1 + N_2 + N_3 = N$, and the total number of genes is $2N$. Each AA homozygote has two A genes, and each AB heterozygote has only one gene of this kind. Consequently, the total number of A genes in the group under study equals $2N_1 + N_2$ and the fraction (frequency) of this gene is:

$$pA = \frac{2N_1 + N_2}{2N} = \frac{N_1 + \left(\frac{1}{2}\right)N_2}{N}$$

The frequency of gene B is defined in exactly the same way:

$$qB = \frac{2N_3 + N_2}{2N} = \frac{N_3 + \left(\frac{1}{2}\right)N_2}{N}$$

so that $p + q = 1$. We can also apply the same method to loci with more than two alleles.

Of course, a complete examination of all individuals in a natural population is generally impossible, which in practical terms means that sampling is necessary. Hence the reliability of estimates of gene frequency depends very much on the numbers sampled. Such estimates should be characterized by the least possible error or dispersion factor, that is, they should satisfy the so-called *criterion of effectiveness*. Thus the requisite sample size depends on the genetic population structure, which can be established by preliminary research. The most reliable results come from the method of 'directly calculating' genes formulated by Fisher and used in the above example.

1.1.2 Dominance

If there is *dominance* of one allele (A) over the other (a), then only two distinguishable phenotypes are present in a population, and of them only one phenotype – the homozygote (aa) for the recessive allele – corresponds to only one genotype.

The method of directly determining the allelic frequency is inapplicable to a genetic situation of this kind, and one must allow the hypothesis that the Hardy–Weinberg equation (see the next section) holds in the population; namely, the distribution of genotypes in random mating conforms to the coefficients of the binomial expansion $p^2 + 2pq + q^2 = 1$.

It follows that in order to obtain an effective estimate of the frequency (q) of a recessive gene (a) we should extract the square root of the fraction of *aa* individuals in our sampling:

$$qa = \sqrt{N_3/N}$$

Accordingly, $pA = 1 - q$. In principle this method could also be used to estimate allelic frequency in the absence of dominance, but it only gives less-biased estimates at high-frequency values (Fig. 1.1).

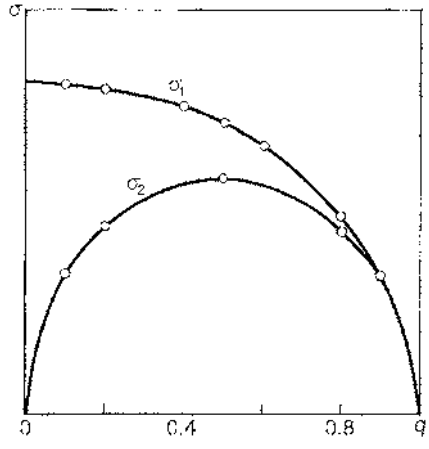


Fig. 1.1 Distributions of the standard errors of the two values of a gene frequency for a single pair of alleles without dominance (based on Neel & Schull, 1958). $\sigma_1 = \sqrt{(1-q^2)/4N}$ (the gene frequency found by extracting the square root of the fraction of homozygous genotype); $\sigma_2 = \sqrt{q(1-q)/2N}$ (the same allele frequency determined by the direct calculation method).

The reader can find more general examples of estimating gene frequencies (e.g. for multiple alleles or linked loci) in C. Li's book (1978) and other publications on theoretical population genetics.

1.2 The Hardy–Weinberg rule

The field of population genetics examines the principles regulating the maintenance and dynamics of population genotypic structure in time and space. The Hardy–Weinberg principle provides a theoretical basis for this view. It reflects the

invariability of the genetic composition of a randomly mating (panmictic) population, unlimited in number and existing extra-environmentally, i.e. in the absence of environmental pressure. In this structureless community the genotype frequencies, and hence also the gene frequencies, at an autosomal locus having a pair of alleles A and a, reach equilibrium in the generation that follows random mating.

Inasmuch as random matings signify merely a random association of gametes, it can easily be verified that this combination of $p(A)$ and $q(a)$ male gametes and $p(A)$ and $q(A)$ female gametes, when $p+q=1$, produces an invariable distribution of $p^2\mathrm{AA}+2pqAa+q^2\mathrm{aa}=1$.

		male gametes	
		$p(A)$	$q(a)$
female gametes	$p(A)$	p^2	pq
	$q(a)$	pq	q^2

The algebraic calculations supporting the Hardy–Weinberg equilibrium are given in Table 1.1. As we are concerned here with autosomal genes, the reciprocal crosses (that is of the type *male* $Aa \times$ *female* AA or *male* $AA \times$ *female* Aa etc.) may be combined, and consequently the nine possible crosses reduce to six, given that $p+q=1$.

Table 1.1 Types of matings and proportions of genotypes of population progeny at genetic equilibrium.

Type of mating	Frequency of mating	Proportions of genotypes among progeny		
		AA	Aa	aa
$AA \times AA$ $(\mathrm{p}^2 \times p^2)$	p^4	p^4	0	0
$AA \times Aa$ $2(p^2 \times 2\,pq)$	$4p^3q$	$2p^3q$	$2p^3q$	0
$Aa \times Aa$ $2pq \times 2pq)$	$4p^2q^2$	p^2q^2	$2p^2q^2$	p^2p^2
$AA \times aa$ $2(p^2 \times q^2)$	$2p^2q^2$	0	$2p^2q^2$	0
$Aa \times aa$ $2(2pq \times q^2)$	$4pq^3$	0	$2pq^3$	$2pq^3$
$aa \times aa$ $(q^2 \times q^2)$	q^4	0	0	q^4
Totals for population	1,00	p^2	$2pq$	q^2

It is clear that this equilibrium ratio of genotypes is provided by the symmetry of distribution of allelic genes into male and female gametes and by the free combination of these into the zygotes formed during reproduction. From this it follows that when there are no disturbances affecting a population of unlimited numbers, the frequencies of the genotypes and genes that characterize them remain unchanged in an infinitely long series of generations – the so-called 'absolute zero' of genetic dynamics. Ideal populations of this kind are, however, virtually never encountered in nature, which is why there are always natural factors that shift them

from the point of equilibrium, disturbing their stability – random genetic drift, mutations, migration and natural selection. These are the 'factors of evolution' or microevolutionary forces that we shall now examine.

1.3 Random genetic drift

Random genetic drift is a mathematical fact that emerges from the phenomenon of finiteness of numbers of any real population. Of special importance is the difference that exists between the total population number and that part of it which transmits the gene pool to the next generation. A reproductive population size (N_r), and even more so its genetically effective population size (N_e) (and of the species as a whole), are virtually always less, and most often considerably less, than its total or census population size (N_t). Concepts of the substantial differences between genetically effective numbers and total population number were developed both theoretically and experimentally during studies of the stochastic processes of gene frequency dynamics (Serebrovsky, 1927, 1930; Fisher, 1930; Dubinin, 1931; Romashov, 1931; Wright, 1931; Dubinin & Romashov, 1932; Kolmogorov, 1935).

Extreme age groups are excluded from reproduction and the N_e value is influenced by population parameters such as the balance in numbers of the sexes during the reproductive period, individual fertility differences and periodic fluctuations of numbers.

1.3.1 Sex ratio and fluctuations in population size

If N_m represents the male and N_f the female reproductive part of the population, then the effective population size is

$$N_e = \frac{4N_mN_f}{(N_m + N_f)} \tag{1.1}$$

When the male and female fractions deviate sharply from equilibrium, the N_e value depends more on the less numerous sex. The same effect occurs whether the mean value of N_e is valid for associated populations scattered in space and differing in numbers, or for one and the same population during fluctuations in time (generations) of the number of mating individuals (Wright, 1938; 1939). In the case of cyclical variations at an interval of n generations, $N_e = \tilde{N}$, where

$$\tilde{N} = \frac{n}{\sum_{1}^{n} 1/N_i} \tag{1.2}$$

that is, N_e is the harmonic mean. For instance, if the effective number in each of five generations of one population is 10, 10^2, 10^3, 10^4 and 10^5, the average harmonic N_e

value represents only 45 individuals. The longer the time interval of observations, or the greater the number of populations distributed in space, the more stable the N_e estimate.

1.3.2 Variability of individual fertility

When population numbers are stable (the average number of progeny reaching reproductive age per one pair of parents, $\bar{k} = 2$) and there is individual variation in the number of gametes (k) produced by the parent population (N), then:

$$N_e = \frac{(4N - 2)}{(V_k + 2)} \tag{1.3}$$

where V_k is the variance k.

When the number of progeny has a Poisson distribution ($\bar{k} = V_k = 2$), the genetically effective population size is approximately equal to its reproductive number, $N_e = N_r$. However, in the majority of natural populations $V_k > \bar{k}$; hence N_e is always less than N_r. According to the estimates of Crow and Morton (1955), the N_e/N_r ratio is approximately 0.75 for many species. There are grounds, however, for considering this estimate to be a maximum, since in those cases where this kind of definition has been applied to actual populations, the difference has been much greater (Cavalli-Sforza *et al.*, 1964; Kerster, 1964; Tinkle, 1965; Rychkov, 1968; Kerster & Levin, 1968; Frankham, 1995) and the N_e/N_r ratio could even be as low as 0.30 (Rychkov & Sheremetyeva, 1979). In salmon populations the N_e/N_t ratio may be 0.10 and still less (Bartley *et al.*, 1992a; see Chapter 2).

Crow (see Crow, 1954; Crow & Morton, 1955; Kimura & Crow, 1963) introduced the concept of *inbreeding effective number*, $N_e(f)$ and *variance effective number*, $N_e(v)$. The *inbreeding effective number* is defined as

$$N_e(f) = \frac{(N_{t-2}\ V\bar{k} - 2)}{(\bar{k} - 1 + V_k/\bar{k})} \tag{1.4}$$

in which N_{t-2} is the number of individuals two generations back, and k is the number of gametes introduced by them.

When the population number is constant, $N_{t-2} = N$, $k = 2$ and consequently

$$N_e(f) = \frac{(4N - 4)}{(V_k + 2)} \tag{1.5}$$

It is clear that when N is sufficiently large, this equation differs little from Equation 1.3

The *variance effective number* is defiend as

$$N_e(v) = \frac{p(1 - p)}{2V_{\sigma p}} \tag{1.6}$$

where p and $1 - p$ are the frequencies of any pair of alleles in a population and $V_{\delta p}$ is the change in allele frequency due to random sampling of gametes in a given generation:

$$V_{\sigma p} = \frac{p(1-p)}{2N} \tag{1.7}$$

It is obvious that if the N_e values designated by Equations 1.1, 1.2, 1.3 and 1.4 are derived from field observations of numbers, the sex ratio and other biological parameters of a real population, then Equation 1.6 represents a method of defining an N_e value by analysing the genetic parameters on the hypothesis of the selective neutrality of the allelic genes being studied.

All the preceding equations for N_e were derived for non-overlapping generations. According to Nei & Imaizumi (1966a,b), when generations overlap in time

$$N_e = \tau N_a$$

where N_a is the number of individuals who reach mean reproductive age, and τ is the generation time or mean reproductive age. For example, as these authors have noticed, the N_e value assigned in this way to cover the general Japanese human population represented approximately 40% of the total number, including non-reproductive individuals.

If a population is stable, then $N_a = Nbp$, where N is the total population size, b the annual birthrate, and p the probability of a newborn individual reaching reproductive age.

In principle, the *variance effective number* under conditions of stability is nothing other than the correlation between the random variance of gene frequencies and the effective population size; this theory was developed by S. Wright in the 1930s. The relationship of these parameters was used as a basis for estimating effective population size N_e from temporal changes in allele frequencies, i.e. from their standardized variance (Krimbas & Tsakas, 1971; Nei & Tajima, 1981; Pollak, 1983; Waples, 1990 a,b). The approaches were elaborated for applying this method to populations with discrete (Nei & Tajima, 1981) and overlapping generations (Jorde & Ryman, 1995).

Changes in gene frequency caused by sampling error during the formation of gametes in the subsequent generation has a random, stochastic character. For this reason only a probability approach is feasible when assessing genetic drift based on the genetic structure of a population, making it possible to determine only the range of variance in gene frequencies. According to Wright, these non-directional fluctuations depend exclusively on the N_e value and are described by Equation 1.7.

The distribution of probabilities of all possible q values in a series of generations of a population limited in numbers approaches normal, and an allele's ultimate fate is fixation ($q = 1$) or loss ($q = 0$); this process is essentially irreversible.

1.3.3 Modelling of random genetic drift

Dubinin and Romashov (1932) were the first to model random genetic drift by providing 'drawings' of varicoloured balls which imitated different alleles in a population of randomly crossing organisms with no selection. They used a statistical population model to demonstrate that the process of loss (or fixation) during genetic drift is dependent entirely on effective population size.

In the absence of selection, the loss of genetic diversity (that is, reduction in heterozygosity) may be described by an equation enabling one to reconstruct evolutionary time:

$$H_T = H_0 e^{-T/2N_e} \tag{1.8}$$

where H_0 and H_T are the frequency of heterozygotes at the zero point in the process and at the instant of time T, and N_e is the effective size of populations.

Although fluctuations in gene frequencies during drift are completely random, they acquire 'direction' (or vector) after a certain number of generations: the frequencies of rarer alleles in each successive generation diminish with greater probability (the loss process), whereas frequencies of other alleles increase (the fixation process).

The range of possible changes in the random variance of gene frequencies (Equation 1.7) during the life of one population may be regarded as exactly the same as that for a community of many populations roughly equal in numbers and having identical allele frequencies q_0 in the t_0 generation. In the following generation the gene frequencies in each population change randomly to become q_i. In the process, the average frequency $\bar{q}$ for the community remains equal to q_0, but the interpopulation variance will increase and in the t-th generation will be:

$$V_q = p_0 q_0 \left[1 - (1 - \frac{1}{2N})^t\right] \tag{1.9}$$

In the final stage of the process all populations become homozygous, and the intergroup variance reaches a maximum of $V_q = p_0 q_0$; that is, in some populations allele A reaches a state of fixation ($p = 1$) and in others it is lost ($p = 0$); at the same time the average allelic frequency remains unchanged as before. Thus, during random genetic drift the process of changing the probability distribution of the frequencies depends on only two factors – the population's effective size and the duration of the process, measured by the number of generations. When the effective population size is small, the distribution rapidly becomes U-shaped.

Buri (1956) modelled random genetic drift in experimental populations of *Drosophila melanogaster*. He studied the distribution frequencies of *bw* and bw^{75} alleles at the 'brown' locus in 19 consecutive generations of 107 lines with 16 individuals in each. The picture he obtained was very similar to what might have been expected (Fig. 1.2).

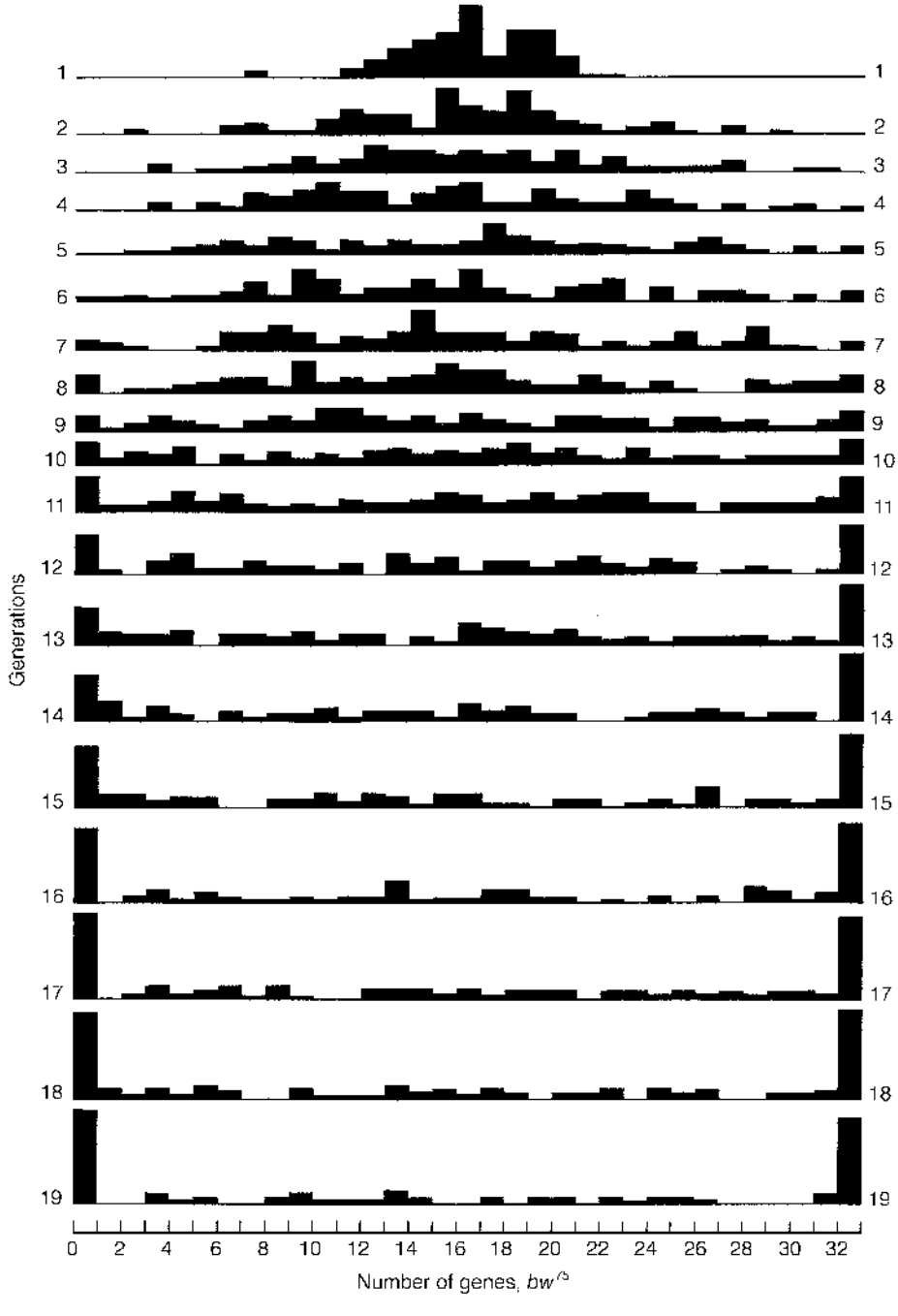


Fig. 1.2 The distribution of gene bw^{75} frequencies in 19 generations of 107 *Drosophila melanogaster* experimental populations (Buri, 1956).

The increase in the homozygosity of a population through random genetic drift indicates a direct link with *inbreeding*, since non-random association of gametes increases with time in a population with small numbers. By virtue of this fact the deviation from panmixia becomes increasingly obvious; this also implies an increase in the degree of genetic relationship between population members.

For convenience, a population with the same degree of inbreeding may be regarded as consisting of two parts, one of which is completely inbred while the other is completely panmictic. From this viewpoint the probability of combining two A gametes in random mating is p^2, whereas with inbreeding it should be more than a certain positive value of ε, namely, $p^2 + \varepsilon$. Exactly the same probability holds for combining the a gametes: $q^2 + \varepsilon$, and the probability of combining A with a is $2pq - 2\varepsilon$.

At genetic equilibrium the inbreeding coefficient (the correlation coefficient between uniting gametes) is $F = \varepsilon/pq$, that is, $\varepsilon = Fpq$. Thus the zygote fractions equal

$$\begin{array}{ccc} AA & Aa & aa \\ Q_1 = p^2 + Fpq & Q_2 = 2pq(1-F) & Q_3 = q^2 + Fpq \end{array} \tag{1.10}$$

When $Q_1 + Q_2 + Q_3 = 1$.

1.4 Mutation and migration of genes

1.4.1 Mutations

Mutations are heritable, non-directional, random changes of genetic material occurring spontaneously or under the influence of specific physical, chemical and biological factors. The process of spontaneous mutation is the ultimate source of new alleles, resulting in increased genetic diversity.

Since the overwhelming majority of *de novo* (or 'new') mutations are deleterious (Muller, 1950), one naturally asks why a considerable number of alleles at a wide range of loci are to be found in populations? Such diversity is obviously maintained at the expense of part of the selectively neutral alleles. On the other hand, alleles that have recently arisen and that display no harmful effects in the heterozygotes may gradually be incorporated, step by step, in a species' gene pool, which, as the Russian geneticist S. Chetverikov wrote, absorbs them like a sponge throughout its lengthy evolution (Chetverikov, 1926). In the latter case of mutation it is the raw material for natural selection which adjusts to coexist with ancient wild-type allelic genes that have already been ground and polished by evolution, though their effects on fitness depends critically on the environment. These questions will form the subject of discussion in Chapters 5 and 6. Here, however, it is only the dynamics of selectively neutral mutations, in generating deviations from Hardy–Weinberg equilibrium, that are discussed.

The fate of a single mutation was first analysed by Fisher (1930), who showed that when a population remains constant in size (the average number of progeny per family $m = k = 2$), the probability of total loss of a mutant gene in the first generation is approximately 0.37. If this mutant gene is not lost in the first generation, it again risks being lost in the second generation, and so on; thus the loss probability limit (l_n) is equal to 1.

Because the loss of *de novo* mutations is an essentially irreversible process, the bulk of them have no chance of becoming fixed in a population. Table 1.2 represents the overall dynamics of this process.

Table 1.2 Probability of loss of a neutral mutant gene (from Fisher, 1930).

Generation	Probability of loss, l_n	Probability of persistence, $1 - l_n$
1	0.3679	0.6321
2	0.5315	0.4685
3	0.6259	0.3741
4	0.6879	0.3121
5	0.7319	0.2691
6	0.7649	0.2351
7	0.7905	0.2095
...	...	...
31	0.9411	0.0589
63	0.9698	0.0302
127	0.9847	0.0153
...	...	...
∞	1.0000	0.0000

It should be remembered, however, that mutations constantly recur in every generation, and it is quite likely that so-called 'new' mutations have re-occurred repeatedly. When an allele has a low frequency in a population it is practically impossible to identify its mutation into another allele. It is only when a certain frequency has been reached that one can take the effects of the mutation process on this allele into account. Thus, if μ is the mutation rate from allele A to allele a, and v is the reverse mutation rate, a to A, then the value of the change in gene frequency per generation is

$$\Delta q = \mu p \text{ (increase)} - vq \text{ (decrease)} \tag{1.11}$$

Hence the increase or decrease in a gene's frequency is determined by the relative values of such an increase or decrease. For instance, if the increase is more than the decrease, then at any given moment q will grow. But as, due to an increase in q, vq also increases, it can only continue to the point of equilibrium, after which no more changes of gene frequencies will be observed in the generations, that is, at the equilibrium point $(\hat{p}, \hat{q})$ $\Delta q = 0$.

Consequently,

$$\Delta q = 0 = \mu(1 - \hat{q}) - v\hat{q}, \text{ and } \hat{q} = \frac{\mu}{\mu + v} \tag{1.12}$$

Solving this equation in exactly the same way for p, we obtain

$$\hat{p} = \frac{v}{\mu + v}$$

This equilibrium is stable; moreover the allele equilibrium frequency $\hat{q}$ is independent of its initial frequency, being determined solely by the relationship between forward and backward mutation rates: if $q > \hat{q}$ in a certain generation, the gene frequency will be reduced in the following generations; whereas if $q < \hat{q}$ it will, on the contrary, increase.

When the q value deviates from $\hat{q}$, then the Δq per generation assigned by Equation 1.11 may be expressed in terms of the deviation $(q - \hat{q})$.

If in accordance with Equation 1.12 we write $\mu = (\mu + v)\hat{q}$, then:

$$\Delta q = \mu(1 - q) - vq = -(\mu + v)(q - \hat{q}) \tag{1.13}$$

Thus, the rate at which equilibrium is reached is proportional to the deviation of the actual value of q from its value $\hat{q}$ at equilibrium.

Two important facts should be taken into account when discussing the effect of the mutation process on genetic equilibrium in a population. First, the rate of forward gene mutation is an order of magnitude higher than the rate of back mutation. Second, although the mutation rate may vary substantially for different loci, the value is extremely low, in the order of 1×10^{-5} to 10^{-6} per locus per generation. Thus when describing genetic population structure of any species, the effects of newly occurring mutations can be largely disregarded, compared with the effects of such factors as natural selection, random drift and migration.

1.4.2 Migration

Migration, or gene flow, involves the introduction of new genes into a population through the transfer of gametes and/or propagules via interbreeding with the recipient population. It is important to distinguish migration in the population genetic sense from that employed in the ecological literature; gene flow does not occur simply as a result of the movement of individuals, it is crucial that the migrants interbreed successfully with the recipient population for gene frequencies to be effected.

Migration is an important factor influencing population dynamics, since each population in nature interacts with other groupings of the same kind through exchanging genes (except in cases of extreme isolation). Thus if immigrants differ genetically from the recipient population, they cause a corresponding change in gene frequency in each generation.

$$\Delta q = -m(q - q_m) = -mq + mq_m \tag{1.14}$$

where m is the number of immigrants divided by the size of the population receiving them, q is the gene frequency of the population, and q_m is the gene frequency of the immigrants.

The formula obtained is identical to Equation 1.13, reflecting the similarity with the results of direct and reverse mutation. To be convinced of this, one need only substitute the constant μ for mq, v for $mp = (1 - q)$, and $\mu + v$ for m. An equilibrium is established in precisely the same way during equal rates of gene flow to and from the population.

We shall return to the effects of gene migration later when we examine the effects of natural selection and the combined action of well-known evolutionary factors on the genetic population structure of salmon. Problems of the mutation process will not be considered.

1.5 Natural selection

In the theory of population genetics, natural selection is regarded as an extremely important factor of evolution that causes adaptive changes in genetic structure. These changes result from differences in the relative contributions of genotypes from reproductive individuals in a population, through differential reproduction or survival.

1.5.1 Basic equations and types of selection

If populations inhabited space of unlimited area and resources, and if at any instant of time t the birth rate a exceeded the death rate b by a constant value r, then the population number would grow continuously in an exponential manner:

$$N_t = N_0 \, e^{rt} \tag{1.15}$$

where the parameter $r = a - b$, denoting the coefficient of the population growth, is called the Malthusian parameter; Fisher (1930) introduced it in *The Genetical Theory of Natural Selection*. In nature however, space and resources are always limited and the coefficient r does not remain constant, so the exponential dependence may only be observed over limited sections of time (and space), ultimately giving way to an S-shaped logistic curve

$$\Delta N_t = \frac{k}{1 + C_0 \, e^{-rt}} \tag{1.16}$$

where k is the maximum number of individuals capable of living in a given specific

environment, and the constant $C_0 = (k - N_0)/N_0$ is a correction factor – the 'environmental resistance' to population growth (Fig. 1.3).

The curves plotted in Fig. 1.3 relate to models which are *continuous in time*, that is, across generations, and their most frequent application is in ecology. Models with *discrete time* (non-overlapping generations) are chiefly used in population genetics for quantitative descriptions of natural selection, and usually operate with a similar quantity – Wright's fitness coefficient W. Thus, if N_t is the number of adult individuals in a generation t, k and v are their fertility and viability, and $W = kv$, then the growth of numbers in a population is:

$$\Delta N_t = N_{t+1} - N_t = (W - 1)N_t, \text{ and } N_t = W^t N_0 \tag{1.17}$$

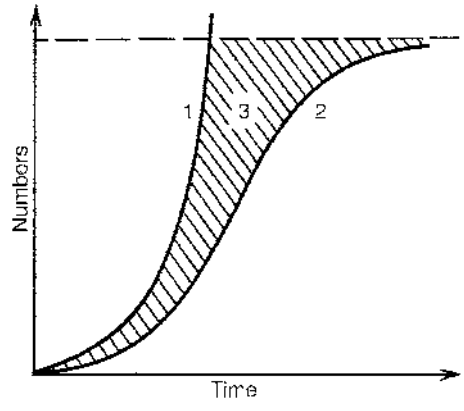


Fig. 1.3 Theoretical curves of population growth. 1 – exponential; 2 – logistic, 3 – environmental pressure.

Consequently, $W^t = N_t/N_0$, that is, the population's fitness at a certain instant of time is equal to the ratio of its numbers in the subsequent and previous generations (when the coefficient of fitness and the state of the environment are stable).

Thus, the parameter W is also important in a biological sense: when $W > 1$ the population size grows, when $W < 1$ it falls, and when $W = 1$ it remains the same.

In the theory of natural selection, developed by Fisher, Wright and Haldane, the fitness of genotypes is taken most frequently as a constant throughout the selection cycle, and its relative value is more important than its absolute value. Here we are only concerned with a change in the ratio of genotypes, which lends itself very conveniently to the study of the dynamics of gene frequencies. In this sense the symbol W_t signifies the i-th genotype's relative fitness; in other words, a quantity that reflects its reproductive contribution to the genetic pool of the following generation through differential fertility or survival, compared with other genotypes when population numbers are stable.

If the consequences of selecting out genotypes in a population with two alleles at

one locus are considered from these viewpoints, then the gene frequencies after one selection cycle are changed in the following way (Table 1.3)

Clearly, under selection only certain genotypes can have relative fitness equal to unity, and hence the average fitness of a segregating population is always less than unity (that is, less than the 'optimal' genotype's fitness):

$$\bar{W} = fW_i = p^2W_1 + 2pqW_2 + q^2W_3 \tag{1.18}$$

Table 1.3 Change in gene frequencies in a randomly mating population after one generation of selection.

Genotypes	Frequency before selection	Relative fitness	Frequency after selection	New gene frequencies	Frequency of genotypes before selection
A_1A_1	p^2	W_1	p^2W_1		p'^2
A_1A_2	$2pq$	W_2	$2pqW_2$	$\mathrm{p}' = \dfrac{p^2W_1 + pqW_2}{\bar{W}}$	$2p'q'$
A_2A_2	q^2	W_3	q^2W_3	$\mathrm{q}' = \dfrac{pqW_2 + q^2W_3}{\bar{W}}$	q'^2
Sum	1		$\bar{W}$		1

A population's fate subsequently is as follows: its evolution towards the loss ($p = 0, q = 1$) or fixation ($p = 1, q = 0$) of allele A, or else its transition to a condition in which both alleles remain, depends on the relative W_i value of the genotypes. If $W_1 > W_2 \geq W_3$ or, on the other hand, $W_3 > W_2 \geq W_1$, the population will inevitably reach a trivial equilibrium (a stationary state when either $p = 1, q = 0$ or $q = 1, p = 0$), corresponding to what is known as *directional selection*. A non-trivial equilibrium point ($0 < p < 1$) is reached if $W_1 < W_2 > W_3$ or $W_1 > W_2 < W_3$; that is, when the heterozygote fitness is more than or less than that of both homozygotes. In both cases, despite the continuing effects of selection, no genetic changes occur in the population. However, only in the first instance ('overdominance') will there be a stable ratio of genotype frequencies at equilibrium

$$\hat{p} = \frac{W_3 - W_2}{(W_1 - W_2) + (W_3 - W_2)} \tag{1.19}$$

This type of selection, one of the forms of balancing selection, is called *stabilizing*.

When, however, the heterozygote is less adaptive than both homozygotes the state of equilibrium in the population is unstable and a gene with a higher frequency moves in the direction of fixation, while a gene with lower frequency tends towards a reduction or loss. This type of selection is called *disruptive* or *diversifying*. Figure 1.4 shows how a population's average fitness changes in these cases (Li, 1976).

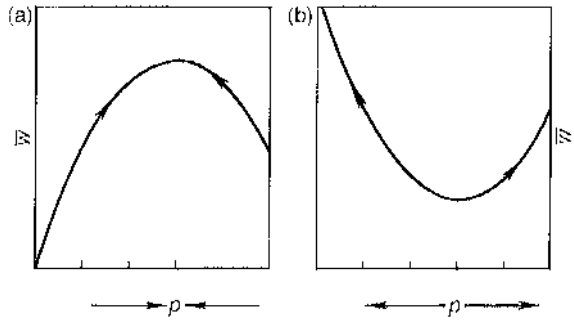


Fig. 1.4 The average fitness of a population ($\bar{W}$) as a function of gene frequencies (Li, 1967). (a) $w_1 = 2$, $w_2 = 4$, $w_3 = 1$, $\bar{w} = 2p^2 + 8pq + q^2$. Point 0.6 on the parabola corresponds to maximum fitness – stable equilibrium; the trivial equilibrium points ($P = 1.0$) are unstable. (b) $w_1 = 3$, $w_2 = 1$, $w_3 = 4$, $\bar{w} = 3p^2 + 2pq + 4q^2$. Point 0.6 corresponds to minimum fitness. As unstable equilibrium is affected by disruptive selection, a population moves towards gene frequencies of $p = 0.1$; these trivial equilibrium states are stable.

The rate at which a population approaches the equilibrium point depends on the intensity of selection ($s = 1 - W$), and is determined by the magnitude of the change in gene frequency per generation (Δp), based on the expressions for p' and q' (see Table 1.3):

$$\Delta p' = p' - p = (pq/\bar{W})[p(W_1 - W_2) + q(W_2 - W_3)] \tag{1.20}$$

Here is a numerical example: population 0.25 *AA*: 0.50 *Aa*: 0.25 *aa*, moving towards the equilibrium frequencies of $\hat{p} = 0.60$, $\hat{q} = 0.40$ and influenced by the effects of selection of the relative values of *W* for the three genotypes *AA*, *Aa* and *aa* in the four following situations: (1, 3, 0), (3, 7, 1), (8, 10, 7), (101, 103, 100). It is clear that the types of selection in these four cases differ greatly from each other. Whereas in the first case (1, 3, 0), genotype *aa* is lethal and the reproductive contribution of the heterozygote *Aa* is far in excess of that of the *AA* homozygote, the fourth type of selection (101, 103, 100) displays only very insignificant differences in the fitness of genotypes. The second (3, 7, 1) and third (8, 10, 7) models occupy an intermediate position between the two extremes. The character of the gene frequency changes in these four situations is shown in Table 1.4.

It can be seen that in the first case the rate at which the population reaches equilibrium is very great, while it is negligibly small in the last case. When $p = 1$ or $p = 0$, $\Delta p = p' - p = 0$ (see Equation 1.20), the population reaches trivial equilibrium state. The factor pq in Equation 1.20 also shows that the rate at which equilibrium is approached is fairly high at intermediate allele frequencies and becomes smaller the smaller the value of p and q.

Let us turn again to Fig. 1.4 to convince ourselves that the mean fitness of the population $\bar{W}$ under the pressure of selection always increases to its maximum value in the equilibrium state. This rule, known as the 'fundamental theorem of natural

Table 1.4 Gene frequency in a population after one generation of selection as a function of fitness coefficients of genotypes.

Selection type	Fitness coefficient values			Initial gene frequency, p	Gene frequency after selection, p'	Change value, $\Delta p = p' - p$
	W_1	W_2	W_3			
1	1	3	0	0.5000	0.57143	0.07143
2	3	7	1	0.5000	0.55556	0.05556
3	8	10	7	0.5000	0.51429	0.01429
4	101	103	100	0.5000	0.50123	0.00123

selection', was formulated by Fisher in 1930 and was originally defined as: 'The rate increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time.' In the new edition of 1941, it was rephrased: 'The rate of increase in the average fitness of a population is equal to the genetic variance [that is, its additive component] of fitness of that population.'

Fisher based this premise on models having continuous time and logarithmic fitness. Li (1967) extended it to include models of populations with non-overlapping generations. This principle, which confirms the monotonic increase of a population's fitness at any instant of time under selection pressure, plays a major role in Wright's 'adaptive topography' – a component of his shifting balance theory of evolution.

Fisher's fundamental theorem of natural selection has several limitations (Levins, 1968; Zhivotovsky, 1981). A population's fitness may be reduced as a result of inbreeding, frequency-dependent selection, mutations and gene recombinations; hence selection is capable of making a population extinct. (Frequency-dependent selection is a form of balancing selection (see Li, 1978), in which a genotype's selective advantage changes as its frequency in a population changes. Of particular interest is the situation in which fitness and frequency are inversely correlated.)

Nevertheless, Fisher's theorem is justified for constant values of genotypic fitness (Zhivotovsky, 1984). Obviously a natural population existing today has reached maximum adaptation in previous stages of evolution that are unknown to us, and now maintains a dynamic balance with its surroundings. The totality of these adaptations to the specific environment that populations encountered previously is recorded in patterns of contemporary genotypic structure. It represents their reserve of genetic stability despite the changing conditions of the environment.

The evolution described above of the genetic structure of a single-locus, diallelic population is certainly a simplification. Each natural population of a bisexual species evolves simultaneously at a large number of loci, among which there may also be multiallelic genetic systems. The corresponding genotypes can interact in various ways; genes may be linked, or their combinations non-random, under the effects of selection (where there is very close linkage, the situation may be reduced effectively to a single locus and the complex of several or many separate genes regarded as a

supergene in inheritance). All this creates many additional difficulties, and it follows that no mathematical description is adequate to provide a complete quantitative picture of genetic processes influenced by selection. None the less, existing models are proving exceptionally useful even today: they enable one to plan research along definite lines and to quantify the results obtained. For all this information we refer the reader to several sources that discuss comprehensively these problems and the tasks ahead (see, for instance, Wright, 1969, 1970; Lewontin, 1978a; Li, 1978; Zhivotovsky, 1984; and also Chapter 5).

1.5.2 Genetic load of populations

The constant pressure of mutations and the migration of genes, as well as the segregation of biologically less adapted genotypes at balanced polymorphic loci, give rise to the problem of so-called 'genetic loads' which is of paramount scientific and practical importance.

The concept of genetic load was introduced by Muller (1950), but the first researches – revealing a saturation of natural populations with non-adaptive mutant phenotypes that are produced in each generation from apparently normal individuals – were carried out in the 1920s and 1930s (Chetverikov, 1926; Timofeev-Ressovsky & Timofeev-Ressovsky, 1927; Dubinin *et al.*, 1937; and see also Dobzhansky, 1970).

In his paper 'Our load of mutations', Muller (1950) showed that weakly deleterious mutant genes can inflict greater damage on a population than individual mutant genes having a strong negative effect. Moreover, he suggested that each of us is on average the bearer of at least eight harmful genes concealed in the heterozygous state.

Morton *et al.* (1956) determined the amount of genetic load in humans by investigating inbreeding depression. They proposed the special term 'lethal equivalent' and indicated an approach to evaluating the relative contribution to the total amount of genetic burden of the mutation processes (mutational load) and segregation (segregational load) of genes; the latter relates to loci whose polymorphism is maintained at the expense of the selective advantage of heterozygotes.

Although many unresolved points remain in the problem of genetic load (Dobzhansky, 1970; Cavalli-Sforza and Bodmer, 1971), the concept of genetic load is of fundamental interest in quantifying selection intensity and as a parameter linked with a population's fitness. From this point of view the following definition by Crow (1958) deserves attention. According to Crow (1958), the genetic load is the proportion by which the fitness of the average genotype in a population is reduced in comparison to the best (optimal) genotype.

In this instance, the genetic load for a model population with discrete time is:

$$L = \frac{W_{\max} - \bar{W}}{W_{\max}} \tag{1.21}$$

where $W_{\max}$ is the fitness of the best (optimal) genotype, and $\bar{W}$ is the average fitness of a population.

It is obvious that this approach can easily be implemented by taking into account the segregational load arising on the basis of the heterozygotes' increased fitness compared with both homozygotes ('overdominance').

Let us assume two alleles at one locus with constant and positive selection coefficients:

Genotype	A^1A^1	A^1A^2	A^2A^2
Fitness	$W_1\,(1-s_1)$	W_2	$W_3\,(1-s_2)$
Frequency	p^2	$2pq$	q^2

In random mating, equilibrium is reached when $s_1p = s_2q$ and hence the equilibrium frequencies are:

$$\hat{p} = \frac{s_2}{s_1 + s_2},\ \hat{q} = \frac{s_1}{s_1 + s_2}$$

The total amount of load equals a decrease of population fitness resulting from the selective elimination of both types of homozygote, that is, $s_1\hat{p}^2 = s_2\hat{q}^2$, and consequently:

$$L_{OD} = s_1 \frac{s_2^2}{(s_1 + s_2)^2} + s_2 \frac{s_1^2}{(s_1 + s_2)^2} = \frac{s_1(s_2)^2 + s_2(s_1)^2}{(s_1 + s_2)^2} = \frac{s_1 s_2}{s_1 + s_2} \tag{1.22}$$

The population's total load, segregating at n loci, is

$$L_T = 1 - \mathrm{e}^{-\Sigma l_i}$$

where l_i is the load at the i-locus.

The concept of genetic load in connection with the problem of inbreeding is discussed in the book of Lynch & Walsh (1998), which we recommend for interested readers. It is necessary only to emphasize in our context that in evaluating segregational (balanced) load, great importance is attached to the question of the number of loci whose polymorphism may be maintained in a population by overdominance. It is quite obvious that a population's maximum genetic load will come to bear when both homozygotes are lethal; in that event 50% of the descendants will perish in each generation.

If the homozygotes are not lethal, the load is reduced; but even in this case the number of overdominant loci is limited by selection coefficients and population numbers.

For example, if $s_1 = s_2 = 0.02$, the equilibrium frequency $\hat{p} = 0.5$ and the number of independent overdominant loci $n = 1000$, the probability of a specific individual

reaching reproductive age is 4.3×10^{-6}. In a selection type of this nature the view of a maximally heterozygous individual as the best-adapted, optimal genotype has no meaning, since such an individual is simply not encountered in a population of finite numbers. If, for instance, the number of polymorphic loci $n = 40$, then the expected frequency of a genotype, heterozygous at all loci, is less than 2^{-40}, that is, one in a trillion.

Taking these calculations into account, Kimura and Ohta (1971) estimated, on the basis of the population size N and the number n of overdominant loci with multiplicative fitness effects, by how many times ($e^{\tilde{L}}$) the fertility of the most adapted heterozygous genotype would be greater than the population mean:

$$\tilde{L} = \sqrt{\sum_{1}^{n} \left(\frac{s_1 s_2}{s_1 + s_2} \right)^2 \times 2\log_e(0.4N)} \tag{1.23}$$

If $s_1 = s_2 = 0.01$, $n = 1000$ and $N = 25\,000$, then $\tilde{L} = \sqrt{0.46} = 0.68$ and $e^{\tilde{L}} = 1.97$. This estimate is completely realistic and corresponds to many natural situations (for example, many populations of mammals). However, when $s_1 = s_2 = 0.1$, other conditions being equal, the segregational load is already exceptionally large ($\tilde{L} = \sqrt{46} = 6.8$) because a maximally adapted individual should produce $e^{6.8} \approx 898$ times more progeny compared with the population average. Clearly, this scale of fertility does not exist in populations of most species. We shall return to the problem in Chapter 6 when we examine the data on the role of selection in maintaining biochemical polymorphisms of salmon populations.

A population's segregational load is a constant 'cost', in the form of the less adapted genotypes appearing in each generation, that a population is forced to pay for its stable existence at the maximum point of fitness.

Under directional selection, usually regarded as the most important adaptive evolutionary factor linked with the replacement of 'less adapted' alleles, the total volume of the genetic load increases still more at the expense of the so-called 'substitutional load' (Kimura's term). This problem was first posed and then mathematically investigated by Haldane (1957, 1960). His logic in solving this problem is closely allied to what forms the basis of calculating the number of overdominant loci of a population in equilibrium. By his calculations, the number of deaths resulting from a changed vector of selection are not connected with its intensity, but are determined exclusively by the initial frequency p of an unfavourable allele. Nevertheless, the number of generations to which elimination distributions apply depends on the intensity of selection.

If the effects of genes are additive, the substitutional load in one generation for a single locus is $L_i = -2 \log p_0$.

For $p = 0.5$, the approximate equation for the relationship between selection intensity and the number of generations necessary for an allele's substitution is $s = 30/n$, where 30 is the 'cost' of a single gene replacement. This quantity shows by how many times in the selection process the total number of deaths in all genera-

tions exceeds the number of individuals in a given generation. Assuming that $s = 0.1$, corresponding to 10% selective mortality per generation, then $n = 300$; in other words, about 300 generations are required to replace a single allelic gene in the adaptive evolutionary process. Thus, according to Haldane, the rate of evolution must be very slow and the numbers of simultaneously evolving genes fairly small for there not to be a sharp reduction in fitness threatening the life of a population under the conditions described above. Haldane suggested that allelic substitutions at 1000 loci are sufficient for the emergence of a new species, necessitating not less than 300 000 generations, the number of concomitantly evolving genes not exceeding 12 (see also Kimura, 1960a,b).

The speciation rate along the lines of this model generally accords with palaeontological findings about evolutionary tempos, at least insofar as they apply to mammals. However, numerous examples are known of exceedingly rapid speciation, which contradicts Haldane's calculations (see, for instance, Mayr, 1968, 1974; Dobzhansky *et al.*, 1977; Altukhov, 1990; Coyne, 1992; Fontdevila, 1992; Hurst, 1992; Orr, 1995; Gavrilets & Hastings, 1996; Slatkin, 1996).

In the preceding sections, we have concentrated on the microevolutionary factors which influence a population's genetic structure but which are essentially external to it. Such a view, however, is incomplete without taking into consideration a population's organizational structure. In many instances natural populations are not panmictic groups but represent historically evolved communities of semi-isolated subpopulations. Subpopulations are constantly exchanging genetic material with each other, being subjected as much to genetic random drift as to different forms of selection.

Several original models have been suggested to describe these more realistic situations in population genetics; Sewall Wright (1931, 1938, 1943a,b, 1951), Masatoshi Nei (1975, 1987) and other authors (Kimura, 1953; Kimura & Weiss, 1964; Malecot, 1955, 1967; Bodmer & Cavalli-Sforza, 1968; Maruyama, 1970, 1971, 1972; Nagilaki, 1985; Rannala & Hartigan, 1995) have distinguished themselves most in this field.

A population's subdivision into subpopulations of a limited size has important consequences on genetic structure. Let us examine this factor.

1.6 The influence of population subdivision on genetic structure

1.6.1 Subdivision and inbreeding – the Wahlund effect

Wahlund (1928; quoted by Li, 1955) was the first to show that if a large population is subdivided into K panmictic groups, an effect is observed in this community similar to the consequences of inbreeding in a randomly mating population: the fraction of homozygotes increases to the value of the interpopulation variance of gene frequencies through a reduction in the heterozygote fraction.

Indeed, if we denote by q_i the frequency of the gene in the i-th group ($p_i + q_i = 1$) and the frequency of this same gene in the subdivided population as a whole by q, then the average gene frequency characteristic of it and its variance will be:

$$q = \frac{\sum q_i}{K},\ V_q = \frac{\sum (q_i - q)^2}{K} = \frac{\sum q_i^2}{K} - q^2 \tag{1.24}$$

The corresponding zygote (or the genotype) frequencies equal:

$$(AA) \quad \frac{\sum p_i^2}{K} = p^2 + V_q$$

$$(Aa) \quad \frac{2\sum p_i q_i}{K} = 2pq - 2V_q$$

$$(aa) \quad \frac{\sum q_i^2}{K} = q^2 + V_q \tag{1.25}$$

By comparing the frequencies of the genotypes in Equation 1.25 with their frequencies in a population, with an inbreeding coefficient F (Equation 1.10), we obtain the equation:

$$V_q = Fq(1 - q),\ \text{or}\ F = \frac{V_q}{\bar{q}(1 - \bar{q})} \tag{1.26}$$

Since the F value characterizes the subdivided population as a whole, the corresponding frequencies of the genotypes in the population are equal to those frequencies that would characterize a *separate* inbreeding colony. In other words, 'a population's subdivision into separate crossing groups is equivalent in formal terms to inbreeding of the total population' (Li, 1978, p. 467).

The degree of this differentiation is directly connected with the scale of the interpopulation differences of the gene frequencies – the greater the genetic differences between subpopulations, the higher the variance of q.

Sewall Wright (1943a,b, 1951) played a leading part in describing the local differentiation of the gene frequencies of a subdivided population in terms of F-statistics, by establishing several F-coefficients as indicators for measuring genetic differentiation:

(1) F_{IT} – inbreeding coefficient of an individual relative to the total (T) population
(2) F_{IS} – inbreeding coefficient of an individual relative to the subpopulation (S) to which it belongs
(3) F_{ST} – inbreeding coefficient of a subpopulation (S) relative to the total (T) subdivided population.

The relationship between these values is given by the equation

$$F_{IT} = F_{ST} + (1 - F_{ST})F_{IS} \tag{1.27}$$

Nei's (1975, 1987) G_{ST} statistics are equivalent to F_{ST} statistics, relating the total and intrapopulation gene diversity (H_T and H_S) by the following formula:

$$G_{ST} = \frac{(H_T - \bar{H}_S)}{H_T} \tag{1.28}$$

Here:

$$H_T = 1 - \sum \bar{p}_i^2$$
$$\bar{H}_S = 1/n \sum H_S$$
$$H_S = 1 - \sum p_{is}^2$$

where P_{is} is the frequency of the i-th allele in subpopulation S and $\bar{p}_i$ is the mean allele frequency in the total subdivided population consisting of n subpopulations. Hence, $\bar{H}_S$ is the mean heterozygosity of the subpopulation, and H_T is the heterozygosity of the total subdivided population if it were converted to a single randomly mating unit.

The F_{ST} coefficient described in Equation 1.26 was suggested by Wright as long ago as 1943, since when it has been repeatedly used to analyse the distributions of gene frequencies in natural subdivided populations (Cavalli-Sforza *et al.*, 1964; Nei & Imaizumi, 1966a,b; Rychkov, 1969; Altukhov, 1974; Altukhov *et al.*, 1975a,b). This coefficient is of great interest as it enables one to analyse certain important effects of population subdivision on genetic structure.

Wright proposed two original population models for this purpose: an 'island model' and 'isolation by distance'.

1.6.2 The island model of population structure

There are two known versions of the island model:

(1) A species' subdivision into a large (infinite) number of randomly mating subpopulations and with the same effective size N, each of them having equal probability and identical intensity (m) of exchanging genes with the common gene pool.
(2) A large panmictic population ('mainland') surrounded by a host of isolated, genetically differentiated small colonies ('islands') each of which receives genes from the mainland with an intensity of m per generation (Fig. 1.5). The effects of back migration may be disregarded.

The intergroup variance of gene frequencies serves as a measure of the random differentiation of subpopulations in the system.

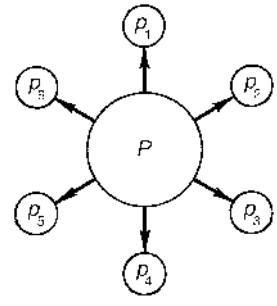


Fig. 1.5 A version of the island population model. p = the gene frequency in the system averaged out for the peripheral sub-populations ('islands', 1–6) and equal to the value of P in the population 'core' ('mainland').

$$V_q = \frac{\bar{q}(1-\bar{q})}{4N_e m + 1} \tag{1.29}$$

and hence, the state of equilibrium between genetic drift and migration in terms of F-statistics (see Equation 1.26) may be written as:

$$F_{ST} = \frac{1}{4N_e m + 1} \tag{1.30}$$

Wright's (1943a) more exact treatment gives:

$$V_q = \frac{\bar{q}(1-\bar{q})}{2N_e - (2N_e - 1)(1-m)^2} \tag{1.31}$$

At small values of m ($m << 1$) the difference between Equations 1.29 and 1.31 is negligible.

Thus the local differentiation of gene frequencies is conditioned by the parameter Nm. In other words, the deciding factor is not the migration coefficient or the size of the population itself, but their product, equalling the number of individuals entering a population in a generation.

As a result of interactions between drift and migration we have a probability distribution of gene frequencies. At any instant of time T it is a function of $\Delta q = -m(q - qT)$ as a measure of the effect of systematic migration and $V_{\delta q} = q(1-q)/2N_e$ is the sampling variance of a gene frequency in one generation through isolation, that is, random drift (Wright, 1938, 1939):

$$\Phi(q) = (C/V_{\delta q}) \exp[2 \int (\Delta q / V_{\delta q}) dq]$$

The general formula for the stationary distribution of gene frequencies in the island model represents the β-function of the probability density as follows:

$$\Phi(q) = \frac{C}{q(1-q)} \exp\,[4N \int \frac{\Delta q}{q(1-q)} dq] \tag{1.32}$$

where C is a normalizing constant chosen so that $\int_0^1 \Phi(q)dq = 1$:

$$C = \frac{\Gamma(4Nm)}{\Gamma(4Nmq)\ \Gamma[4Nm(1-q)]}$$

Equation 1.30 takes the following forms, depending on what combination of random and systematic factors is assigned it:

- For equilibrium of random drift by gene migration:

$$\Phi(q) = Cq^{4Nm\bar{q}-1}\ (1-q)^{4Nm(1-\bar{q})-1} \tag{1.33}$$

  where p and q are the allelic frequencies in the subpopulation, $\bar{p}$ and $\bar{q}$ are the mean allelic frequencies for a subdivided population as a whole, N is the effective population size, and m is the migration coefficient.
- With the combined effects of isolation, migration and selection:

$$\phi(q) = C\bar{W}^{2N}\, q^{Nm\bar{q}-1}\,(1-q)^{4Nm(1-\bar{q})-1} \tag{1.34}$$

  where the notation is the same as for the previous formula and $\bar{W}$ is the population's average intralocus fitness determined by summation of the genotypes' fitness according to their frequencies (see Equation 1.18).

The relationship between the effects of random and systematic evolutionary factors upon population genetic structure can vary widely by virtue of the fact that the corresponding stationary distributions of gene frequencies may take a different form (Fig. 1.6).

Stationary distributions can describe:

(1) The distribution of allelic frequencies of many loci in the same population in the case of neutrality or when subjected to nearly identical pressure of selection
(2) The distribution of gene frequencies at any locus in successive generations of the same stationary population
(3) The distribution of allelic frequencies of one or several loci in the community of isolated populations.

All three types of distribution are mathematically equivalent.

In the island model the gene migration coefficient is independent of the degree of population isolation. Wright (1943a) and Malecot (1955, 1967) also employed mathematical methods to investigate a population in which the intensity of exchange between the subpopulation depends on distance.

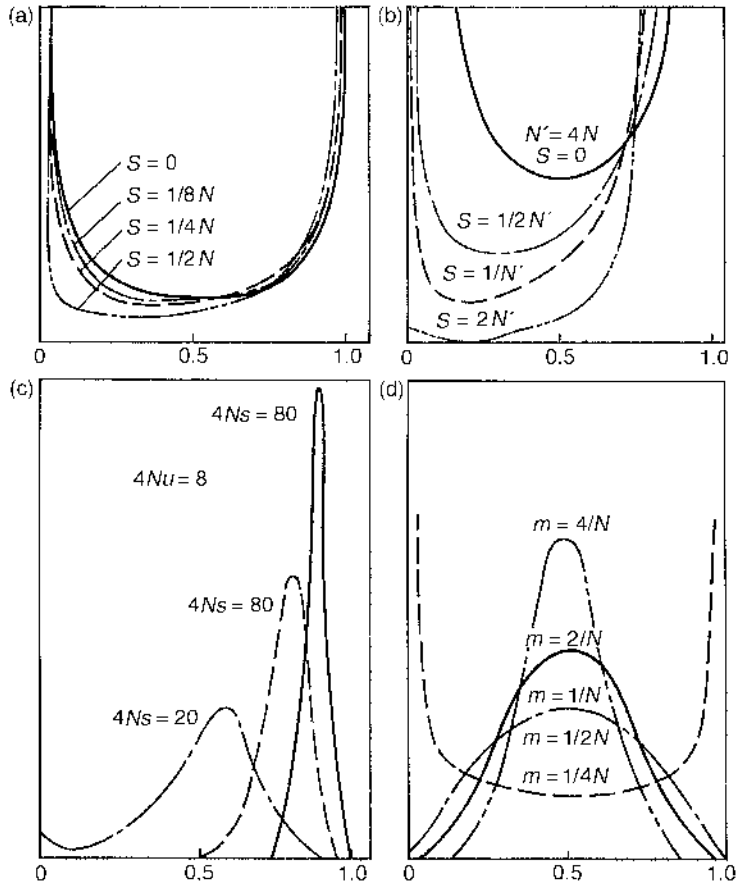


Fig. 1.6 Equilibrium (stationary) distributions of gene frequencies for different relationships between population size and the effects of migration, mutation and selection. The x-axis represents gene frequency, the y-axis represents number of subpopulations. (a) Distribution for small and completely isolated populations affected by selection. (b) Distributions for the same selection s coefficients as in (a), but population size is fourfold ($N' = 4N$). (c) Equilibrium frequency distributions of mutant alleles when $u = 2/N$; s as in (a); $N' = 40N$. (d) Interaction between genetic drift and migration at a mean frequency of 0.5; a U-shaped distribution reflecting allelic fixation in many populations is observed only when $m = 1/4N$, i.e. when a population of no more than one migrant is received for every four generations.

1.6.3 Isolation by distance

This model postulates a population spread evenly over a large territory substantially exceeding the radius of an individual's reproductive activity. The special features of local differentiation in such a system depend upon the effective size of the 'neighbourhood' N_N from which parents arise at random, as well as upon the dimensions of the area. In particular, local differentiation arises sooner or later in a one-

dimensional area, if it is sufficiently large in extent. However, on a plane (two-dimensional area) the possibility of this differentiation is much less.

According to S. Wright (1951), the size of a neighbourhood corresponds approximately to the number of genetically effective individuals within a circle whose radius equals twice the standard deviation (σ) of the extent of migration in one direction in a given generation (or, in other words, the distance between the birthplaces of parents and progeny).

Differentiation is very great when $N_N \approx 20$, somewhat less but still quite pronounced when $N_N \approx 200$, and verges on panmixia when $N_N \approx 2000$.

Kimura (1953; see also Kimura & Weiss, 1964) has proposed another model, called the 'stepping-stone model', representing a situation intermediate between Wright's island model and that devised by Wright and Malecot for evenly distributed populations.

1.6.4 The stepping-stone structure of gene migration

In this, as in the island model, colonies are considered together, but the exchange of individuals proceeds only between neighbouring colonies, as shown in Fig. 1.7; thus it is directly dependent on the colonies' distance from each other.

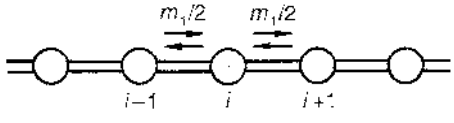


Fig. 1.7 One-dimensional stepping-stone model of gene migration.

At equilibrium the interpopulation variance of gene frequencies is:

$$V_q = \frac{\bar{p}(1-\bar{p})}{2N_e - (2N_e - 1)\{1 - 2R_1R_2/(R_1 + R_2)\}} \tag{1.35}$$

where $R_1 = [(1+\alpha)^2 - (2\beta)^2]^{\frac{1}{2}}$; $R_2 = [(1-\alpha)^2 - (2\beta)^2]^{\frac{1}{2}}$, in which $\alpha = (1-m)\,(1-\bar{m}_\infty)$ and $\beta = m_1\,(1-\bar{m}_\infty)/2$; in these equations m_1 stands for the intensity of migrations between adjacent colonies (*short-range migration*) and $\bar{m}_\infty$ for the external effects of gene migration between all the colonies together (*long-range migration* corresponding to the coefficient m in Wright's island model). If $m = 0$, then $\alpha = 1 - \bar{m}_\infty$, $\beta = 0$ and Equation 1.35 reduces to Wright's Equation 1.26. Thus the island model represents a special case of a stepping-stone model in which there is no gene exchange between neighbouring colonies.

In cases where m_1 is much higher than $\bar{m}_\infty$, Equation 1.35 reduces approximately to:

$$V_p = \frac{\bar{p}(1-\bar{p})}{1 + 4N_e\sqrt{2m_1\bar{m}_\infty}} \tag{1.36}$$

assuming that $\bar{m}_\infty << m_1 << 1$.

The standardized genetic variance in this case is:

$$F_{ST} = \frac{V_p}{\bar{p}(1-\bar{p})} = \frac{1}{1 + 4N_e\sigma\sqrt{2\bar{m}_\infty}} \tag{1.37}$$

where $\sigma = \sqrt{m_1}$, since $m_1 = V_m$ is the migration dispersion for a distance of more than one step apart.

According to the formula obtained by Kimura & Weiss (1964), under stationary conditions where there is no selection, the variance of gene frequencies may also be found from the expression:

$$V_p = \frac{\bar{p}(1-\bar{p})}{1 + 4N_e m[1 - \mathrm{r}(1)]} \tag{1.38}$$

where r(1) is the correlation coefficient of gene frequencies between adjacent subpopulations.

The tendency in the stepping-stone model and in isolation by distance towards local differentiation depends greatly on the dimensions of the area. Following Kimura & Weiss (1964), the correlation coefficient of gene frequencies between colonies in a one-dimensional model diminishes with distance from the exponent and is described by the formula:

$$\mathrm{r}(d) \approx \mathrm{e}^{-\left(\sqrt{\frac{\bar{m}_\infty}{m_1}}\right)d}$$

where d is the distance between subpopulations (in 'steps') and r(d) is the correlation coefficient between them.

The mathematical apparatus becomes more complicated when it involves analysis of two- or three-dimensional models, each colony exchanging genes with four or six adjacent ones (Kimura & Ohta, 1971). For our purposes it is enough to stress that local differentiation of gene frequencies, all other conditions being equal, is at a maximum in one dimension and quickly diminishes as the number of dimensions increases. This dependence manifests itself strikingly in research on the genetic correlation r(d) between colonies based on the distance between them. When the distance increases, the correlation quickly declines, a feature which is particularly characteristic of the three-dimensional model (Fig. 1.8). In accordance with Kimura & Maruyama's (1971) calculations, local differentiation is especially marked in a two-dimensional stepping-stone model when $Nm < 1$, while when $Nm > 4$ a population acts as a single panmictic unit.

For the one-dimensional case, the condition for local differentiation is $Nm < k/\pi^2$, in which k is the number of subpopulations (Maruyama, 1970).

Thus, apart from the dimensions of an area, local genetic differentiation in populations having a stepping-stone structure depends both on the intensity of

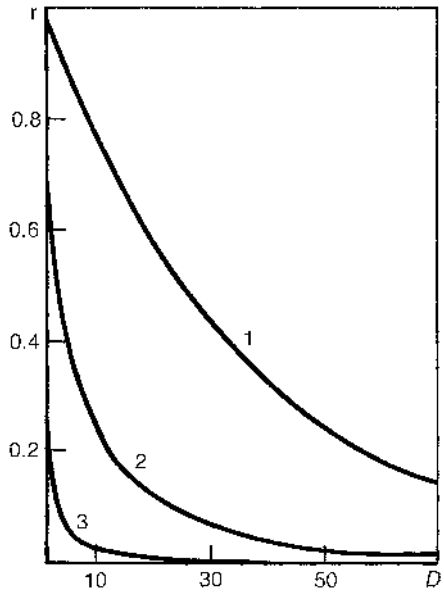


Fig. 1.8 Decrease of correlation r(d) between gene frequencies in colonies with increased distance (d) between them (Kimura & Weiss, 1964). $m = 0.1$; $m_\infty = 4 \times 10^{-5}$; 1, 2, 3 = one, two and three dimensions, respectively.

short-range (m) and long-range (m_∞) migration (the parameter m_∞ can combine all possible stabilizing factors: the mutation process, migration from a constant external gene pool and selection), and on the size of the colonies and their numbers.

Conclusion

Almost all population genetic models and methods considered above will be employed in our following analysis. Along with this, the main estimates and calculations will be connected with the subdivided population models that are mostly applicable to species structure in salmon, and at least in those sections dealing with the theory of stationary distribution of gene frequencies.

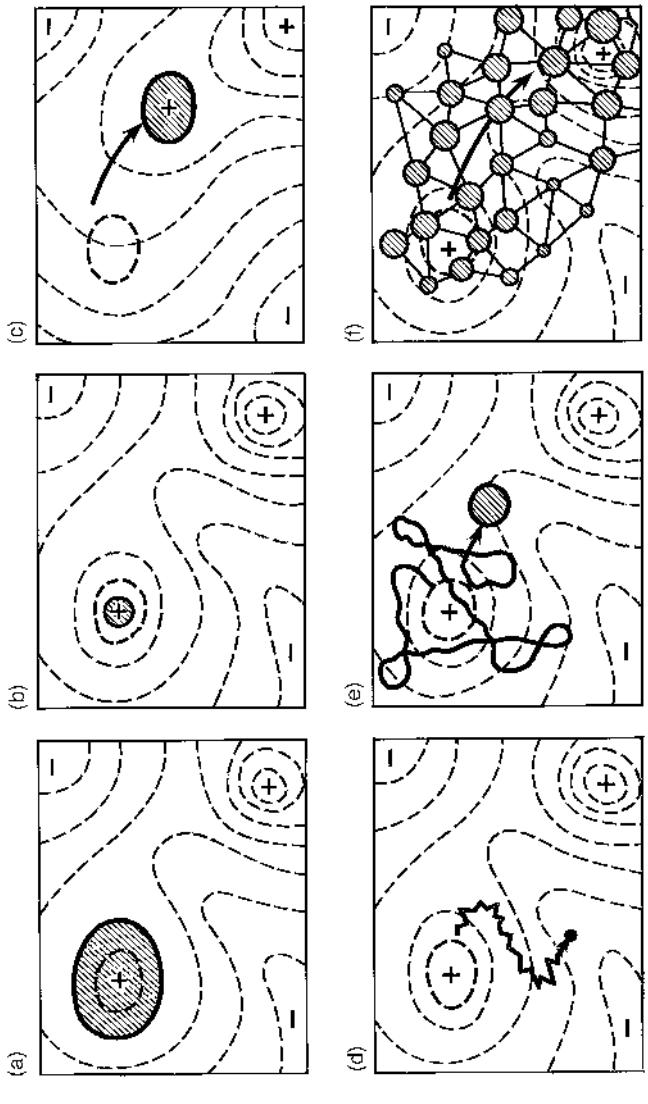
In essence the theory of stationary distributions emphasizes yet again the important fact that, although individual evolutionary factors are also capable of causing directional genetic changes, the interaction of these factors (for example, forward and reverse mutations, genetic drift and migration, etc.) leads finally to reciprocal balance, engendering a stationary type of gene frequency dynamics. This stability may be particularly great when all the known factors of evolution act simultaneously on a population.

Another important feature of subdivision, which has also been examined theoretically, is the ability of subdivided populations to maintain significantly greater genetic diversity compared with panmictic populations of similar total size (Wright, 1951; Kimura, 1968). It is believed that this diversity allows a subdivided population to react more effectively to changes in the environment, through adaptive changes

in genetic structure. This thesis, which plays a decisive role in Wright's concept of evolution, is known as the shifting balance theory in which the 'surface' W is represented on a topographical map with peaks and valleys for one landscape of gene combinations (Fig. 1.9). The most important conclusion from this model is that the 'evolutionary process depends on a continually shifting but never obliterated state of balance between factors of persistence and change, and that the most favourable condition for this occurs where there is a finely subdivided structure in which isolation and cross-communication keep in proper balance' (Wright, 1951).

Our further task is to compare this concept with what is observed in nature. The models of subdivided populations are especially expedient for studying genetic processes in salmon populations. However, before making this comparison it is necessary to examine the principal features of the biology and demography of salmonid species, that are essential for further population genetic analysis.

Fig. 1.9 Wright's two-dimensional contour maps with adaptive peaks and non-adaptive valleys in the field of gene combinations occupied by different types of populations (Wright, 1932). The isolines relate to different levels of adaptation: the thick broken line is the population's 'starting' position; arrows indicate evolutionary trends. (a) Reduced selection (or increased mutation rate) produces increased genetic variability and reduced average fitness. The population's (= species') evolutionary plasticity is fairly large, and with increased numbers it can occupy the lower slopes of another, more highly adaptive peak and then also conquer it entirely ($4Nu$ and $4Ns$ are very large). (b) The consequences of increased selection (or reduced mutation rate). The amount of the population's genetic variability is narrowed by selection, and the population's average fitness grows. Correspondingly, its evolutionary plasticity is reduced and its chances of conquering a neighbouring peak are diminished ($4Nu$ and $4Ns$ are very large). (c) Sharp changes in the adaptive landscape: peaks are transformed into valleys, and vice versa. The result of the genetic process is determined exclusively by selection intensity and environmental rates of change. A population having a small adaptive peak prior to these shifts and lacking an adequate reserve of genetic plasticity (variability) may remain in a valley and become extinct ($4Nu$ and $4Ns$ are large). (d) The consequences of drastic reduction of numbers and close inbreeding. A population falls from the adaptive peak, and as a result of drift its genetic pool undergoes random fluctuations leading, however, eventually to inbreeding and, as a rule, to degradation ($4Nu$ and $4Ns$ are very small). (e) An average-sized population subject to moderate mutations, leading to increased genetic diversity. The population shifts from its adaptive summit but cannot go far away from it. This means that the occupation of any new adaptive peak would be an exceptionally slow process ($4Nu$ and $4Ns$ have values intermediate between extreme situations). (f) The population of an extensive area subdivided into a multiplicity of interacting subpopulations, evolving rapidly by migration and mostly inadaptive. When one population or another comes into the environs of any adaptive peak it may colonize it. After increasing in numbers, it begins to 'pour' corresponding genes into other subpopulations, and entire species may become 'stretched out' in the zone of this new peak. Such a picture is considered to be evolutionarily optimal in the sense that genetic variability and the capacity for further change are preserved ($4Nm$ has an intermediate value).


(a)
(b)
(c)
(d)
(e)
(f)