GENES IN POPULATIONS

Second Edition

Eliot B. Spiess

Professor of Biological Sciences
University of Illinois
at Chicago

A Wiley-Interscience Publication
JOHN WILEY & SONS
New York Chichester Brisbane Toronto Singapore

between populations, and among species. According to Lewontin (1970), the biparental level "gives the great advantage that the evolution of the entire population, rather than a single lineage, is promoted by selection." Under constant selection conditions, the multiplicity of genotypes resulting from the sexual process in biparental organisms allows for an immense diversity on which selection may act, while with clonal organisms genic diversity is more restricted to acquisition by mutation. Under changing conditions, it is obvious according to Wright "that the process of continual readaptation would be more effective if it could be based on the adaptiveness of genotypes or systems of these than merely on the momentary net effects of the separate genes." Consequently, in the adaptiveness of several gene pools to communities or several demes within a species to the entire species, interdeme selection leading to various adaptive levels including races and species is of vast importance. Of course, linear, diversifying, and stabilizing selection acts at all these levels, not as mutually exclusive forces but as overlapping ones, so that in natural populations the mode of selection may be difficult or nearly impossible to discern. Only when we apply selection experimentally in the laboratory or in breeding can we be more certain of the goal to be attained. The instances in which actual specific selective action has been ascertained in nature are very few indeed, and a considerable amount of attention should be given in future to finding out what selective differences genotypes may have in nature. At this time, more than 100 years after the Darwin–Wallace papers on the subject, we are still quite in the dark when we ask somewhat metaphorical questions: "What is selection aiming at in the population?" "What are the selective agents in our environmental niche?" Nevertheless, the few cases of which we are reasonably certain demonstrate how powerful and how subtle the action of natural selection may be. For several examples, consult Endler (1986).

While natural selection and artificial selection differ in that the first is directed by relative fitness of hereditary traits under specified conditions and the second is goal-directed by the experimenter, they share the common element that the effectiveness of selection in producing a response depends on (1) modes and intensity of selection, (2) action of genes as manifested in phenotypes being selected, and (3) the amount and architecture of genetic variability in the population. The result of these factors—the adapted population following the action of natural selection or the "improved" population following artificial selection—demonstrates the powerful yet varied property of the entire process. With numerous genetic pathways available to selection, populations with various genotypes initially may respond with parallel outcome; conversely, under various selective pressures, a single population may give diverse responses.

THE SIMPLEST CASE OF SELECTION

For selection to produce a directed change in the gene pool, only two ingredients are essential: a reproductive difference between genotypes and the duplicating ability of the genetic material. In order to establish some fundamental principles and definitions, we begin with a hypothetical example of selection in a haploid organism. We may visualize a population of haploid prokaryotic organisms stressed by a selective agent such as an antibiotic at low concentration (in order not to kill all the cells) or a low dose of radiation to separate sensitive from the more resistant cells. Further, we assume:

1. Two alleles exist in the population initially, A and a, in equal frequencies ($p = q = 0.5$).
2. Reproduction is by simple fission with no sexual phase.

3. The ability of *A* to get through its life cycle and reproduce, leaving *one* offspring per *one* parent individual, is better than the ability of *a* allele under the specified environmental conditions; for example, for every 100 new *A* individuals produced by one life cycle from *A* parents, only 90 new *a* individuals are produced in the same life cycle by *a* parents. In brief, the Darwinian "fitness" of *A* relative to *a* is 1:0.9; and we assume fitness remains constant over all generations.
4. The environment can support a constant number of individuals in a generation. If these haploid organisms double their number of their life cycle, the "struggle for existence" will cause all to be lost except that constant number. We call this number 100%. For example, if the population consisted of 1000 individuals initially (500 *A*:500 *a*), and they all divided by binary fission, making a total of 2000 progeny, the number must be rduced back to 1000 in the "struggle for existence."

Given these assumptions, the changes in frequencies of the two kinds of organisms (or genotypes) will take place as given in Table 14-1A. It is important to have clearly in mind the meaning of the calculations in the first generation. First, we might imagine that the organism only reproduces by making a single copy of itself and that the *A* individuals are 10% better at doing that than the *a* individuals. But this interpretation is not what is meant by the calculation, because such an implication would result in a *lowering* of the total progeny number. It is more meaningful to recall condition 4, above—that both *A* and *a* reproduce themselves by doubling and that the environment will support only a *constant* number. To make the relative numbers add up to a constant total, we "normalize" the relative survival numbers from the row "after selection." These relative "after-selection" numbers are merely the product of fitness × frequency of each genotype in order to calculate the mean fitness, given at the end of that line as 0.950. (The student should recall the formula for arithmetic average $\bar{Y}=\sum f Y_i/n$ for data grouped by frequency.) We now have a total progeny = 100%. There would consequently be a probability for any individual in G_1 to be *A* genotype $=0.500/0.950=0.5263$; for $a=0.450/0.950=0.4737$, or a ratio of 1:0.900.

If the fitnesses remain constant (condition 3) each generation, for each round of life cycle the mean fitness and relative proportions of progeny surviving will be occurring as in the calculated illustration. It is evident that there is a rise both in the "favored" allele (*A*) and in the mean fitness (boxed in Table 14-1). Briefly, the objective of natural selection is to maximize the mean fitness of the population by ridding the population of the less-fit allele. We may say that the population "suffers a genetic load" (using H. J. Muller's term) as long as the less fit allele is present. We merely imply by this metaphorical phrase that the population is not reproducing at its potentially maximum efficiency for the conditions under which these genotypes have the fitness values we have assumed.

The rate at which *a* is eliminated changes as the limit is approached. We symbolize these changes algebraically to generalize them, using the symbols employed by Wright (1931) for relative fitness (selective) values of the two genotypes: W_1 and W_2 (1 and 0.90, respectively), which may stand for "weight" given by relative selective effects of conditions on the two genotypes. Alternatively, the less-fit genotype may be thought of as having a disadvantage (0.10), which may be symbolized as the selection coefficient $s=1-W_2$ in this case. More generally, the genotype fitness is $W_i=1-s_i$. Mean fitness $\bar{W}=\sum f W_i$, with frequencies usually p, q values (or genotype frequencies in diploids). Calculations would proceed then as in Table 14-1B. The student should verify the calculations in the A portion of the table by applying the symbols from the B portion.

TABLE 14-1 Haploid Selection

A Simple haploid selection: assuming initial $p = q = 0.5$, relative fitness values of $1A:0.9a$ per generation, and a constant number of individuals (100 percent) per generation. Mean fitness values ($\sum f W_i$) are in boxes at right.

Generation	Genotypes		Totals
	A	a	
Initial frequency (parents)	$0.500 = p_0$	$0.500 = q_0$	1.000
G_0 fitness	1	0.90	
After selection	0.500	0.450	0.950
G_1 (relative)	$0.526 = p_1$	$0.474 = q_1$	1.000
Fitness	1	0.90	
After selection	0.526	0.427	0.953
G_2 (relative)	$0.552 = p_2$	$0.448 = q_2$	1.000
Fitness	1	0.90	
After selection	0.552	0.403	0.955
G_3 (relative)	$0.578 = p_3$	$0.422 = q_3$	1.000
⋮			⋮
Limit G_∞	$1.000 = p_\infty$	$0 = q_\infty$	1.000

B Symbolically, these data may be represented as follows

G_0	p	$+ q$	1.00
Fitness (W_i)	W_1	W_2	
Or	1	$(1 - s)$	
Product (after selection)	pW_1	$+ qW_2$	$\sum f W_i = \bar{W}$
Or	p	$+ q - sq$	$(1 - sq)$
G_1 (relative)	$p/(\bar{W})$	$+ (q - sq)/(\bar{W})$	1.00
Or	$p/(1 - sq)$	$+ (q - sq)/(1 - sq)$	1.00

Change in q per generation: $\Delta q = \dfrac{q - sq}{(1 - sq)} - q = \dfrac{-sq(1 - q)}{(1 - sq)}$

or more generally in terms of W_i: $\Delta q = \dfrac{qW_2}{pW_1 + qW_2} - q = \dfrac{(W_2 - W_1)pq}{\bar{W}}$

The change in allele frequency per generation (Δq) is easily calculated by subtracting the original q value from the new q value $[q(1-s)/(1-sq)]$ at the bottom of Table 14-1B.

$$\Delta q = \frac{-sq(1-q)}{1-sq} = \frac{-(W_1 - W_2)q(1-q)}{\bar{W}} \tag{14-1}$$

(rate of change in q under haploid selection). From Exercise 1 the student should note that this rate of change decreases as the less-fit allele becomes rarer. The denominator (14-1) approaches a value of 1 as $q \to 0$, and the rate approaches the allelic frequency product (pq) $\times$ the selection coefficient s. It is constantly negative (downward in q), so that the limit can only be $q=0$. If we ask whether there can be a stable equilibrium by setting $\Delta q=0$, we see that no solution except $q=0$ or 1 will satisfy the equation. We call this equilibrium "trivial" because it is more or less self-evident.

Finally, we may consider the number of generations it would take for selection to produce a given change in q. From (14-1), if we dismiss the denominator as negligible when the product sq is very small, the rate of change can be envisioned as a continuous function so that (14-1) may be expressed as a differential equation where t refers to time in generations.

$$\frac{dq}{dt} = -sq(1-q) \quad \text{or} \quad \frac{dq}{q(1-q)} = -s\,dt$$

Integrating both sides gives

$$\int_{q_0}^{q_t} \frac{dq}{q(1-q)} = -s \int_0^t dt$$

where t = generations for a given change in q value from q_0 to q_t. The solution is

$$t = \log_e \left[\frac{q_0(1-q_t)}{q_t(1-q_0)} \right] \Big/ s \tag{14-2}$$

(where $\log_e = 2.3026 \log_{10}$). For example where $s=0.10$, if we wish to estimate the number of generations to change from $q_0=0.05$ to $q_t=0.005$, we should have as follows:

$$t = \log_e \left[\frac{0.05(0.995)}{0.005(0.95)} \right] \left(\frac{1}{0.1} \right)$$

$$t = \log_e [10.4737] \left(\frac{1}{0.1} \right) = 23.5 \text{ generations}$$

SELECTION AGAINST RECESSIVE HOMOZYGOTES

With diploidy and the accompanying increase in genotypes at a locus, the relative differences in fitness between genotypes can be described in several different ways, depending on dominance, partial or incomplete dominance, or overdominance in fitness. We first examine the outcome of selection when the less-fit genotype is recessive, that is $AA = Aa > aa$ in fitness. Simplifying assumptions must be made here to include (1) standard conditions (a large population, random mating to produce genotypes from gametes, Mendelian segregation, and no other forces acting on the population), (2) constant Darwinian fitness throughout, (3) a single pair of alleles at the locus, and (4) selection acting between the zygotic stage and the adult.