

Directional selection.

Directional selection is a directional force which tends to increase frequencies of advantageous alleles and decrease frequencies of detrimental alleles.

Classic example: pepper moth (*Biston betularia*) in Great Britain

2 alleles, C black (*carbonaria*), c peppered

Before the industrial revolution, f(C) low and f(c) high.

After the industrial revolution, f(C) increased and f(c) decreased, although it never disappeared. f(C) remained the dominant form in industrial areas at least into the 1970's.

Explanation in terms of directional selection: bird predation.

Directional selection is the basis for most cases of Darwinian adaptive evolution, because it results in a phenotypic change that increases the fitness of the organism.

The strength of directional selection is expressed by two parameters: fitness (w) and selection coefficient (s).

Hard selection models: population size changes according to fitness.

Soft selection: population size constant.

In real life, most organisms probably have periods of hard selection and periods of soft selection.

Soft selection model:

Haploid population, $N = 200 = \text{constant}$. Initially all genes are A1 (wild type allele), and A2 is a new mutation.

Fitnesses are w_1 and $w_2 = \text{number of offspring produced}$.

$s = \text{selection coefficient of the mutant} = (w_2 - w_1)/N$; it is negative ($-1 \rightarrow 0$) for detrimental genes and positive ($0 \rightarrow +1$) for advantageous genes:

A ₁	A ₂	
-----	-----	
w_1	w_2	$s = (w_2 - w_1)/N$
200	0	-1 A ₂ is lethal detrimental mutation
150	50	-0.5 A ₂ is detrimental mutation $s < 0$ $w_2 < w_1$
100	100	0 A ₂ is neutral mutation $s = 0$ $w_2 = w_1$
50	150	0.5 A ₂ is advantageous mutation $s > 0$ $w_2 > w_1$
0	200	1 A ₂ is advantageous mutation (A ₁ is lethal)

There are many variants on the simple models we will discuss. Mostly deal with features that are very important in many, but not all, cases. e.g.

- diploidy and sexuality with different kinds of dominance
- selection on different life history stages
- measuring fitness in terms of offspring over several generations

Fitness is a quantitative trait (unless one is concerned only with lethal genotypes).

Fitness is a function of all genes that help determine fitness. However, we can often isolate a single gene for study, e.g.

- we choose to focus on effects of a single gene such as a new mutant allele
- antibiotic resistance in bacteria, where only one gene at a time has a significant effect on phenotype

Directional selection by itself leads to fixation or loss of alleles.

Haploid, deterministic model.

$p = f(\text{detrimental allele})$ $q = f(\text{advantageous allele})$

$$p = -spq / (1 - spq)$$

Change in allele frequencies is greater for

- greater $|s|$ (selection is stronger)
- p and q closer to equal frequencies

Case of special interest is a new mutation.

Neutral mutation: no selection.

Detrimental mutation: purifying or negative selection.

Advantageous mutation: positive or adaptive selection.

Combined effects of mutation and selection

Mutation-selection equilibrium (or balance)

Most new mutations are detrimental. Case of some interest, especially in human genetics, is when mutation pumps new detrimental alleles into the population and selection removes them at the same rate.

If q = frequency of detrimental mutation, at equilibrium $q = (u/s)$

Combined effects of mutation, drift, and selection

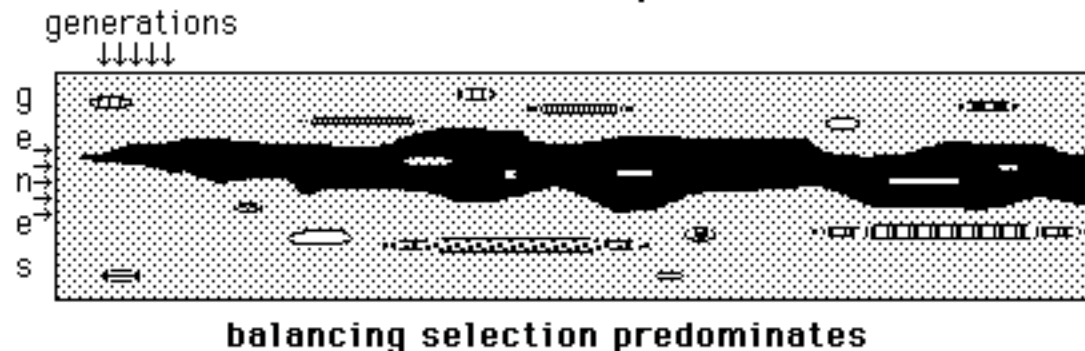
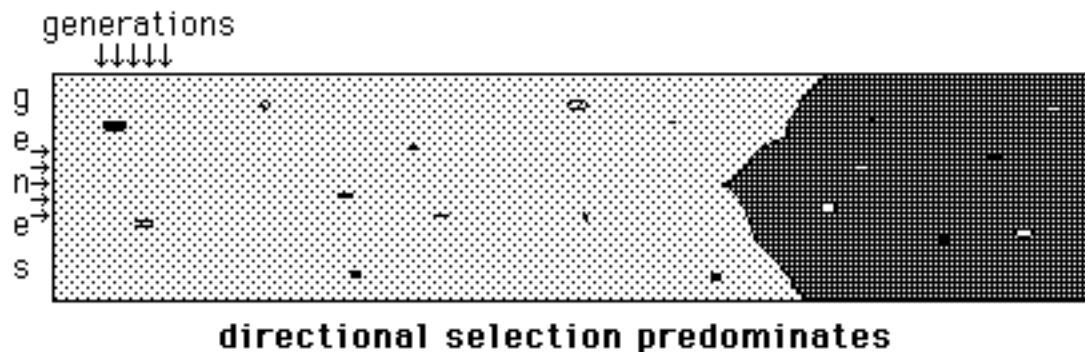
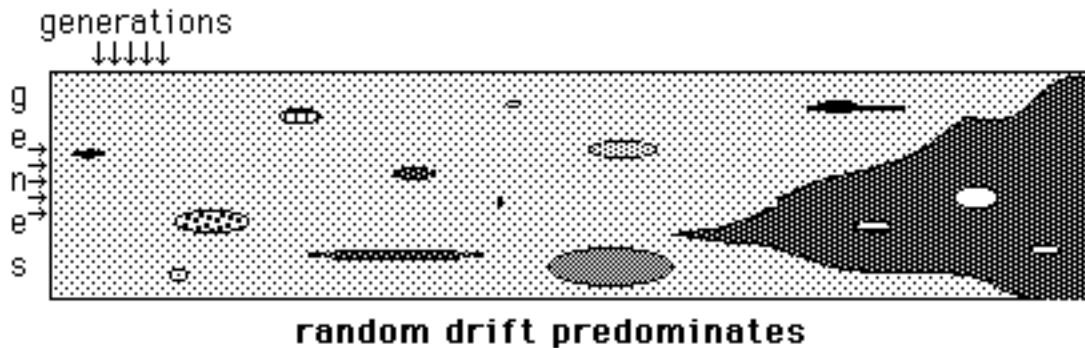
Here we consider the realistic situation in which all alleles are subject to drift, some are neutral, and of the remainder, the vast majority are detrimental.

Time to fixation or loss and expected heterozygosity or nucleotide diversity are reduced when some sites are under selection

With selection pushing a new detrimental mutation to loss, or pushing a new advantageous mutation to fixation, time to fixation or loss will be shorter than the neutral expectation.

Because mutations that are subject to selection don't stay in the population as long as neutral mutations, they contribute less to diversity than do neutral mutations.

$$H < \frac{4N_e u}{1 + 4N_e u} \quad \text{etc.}$$



For example, consider sites within a gene:

- (1) Start AUG codon: directional selection dominates, $s \approx -1$ because any mutation leads to inability to make the protein; consequently diversity ≈ 0
- (2) 3rd position in other codons: drift dominates because most substitutions don't change amino acids and hence are neutral (or nearly neutral in highly expressed genes).
- (3) 1st and 2nd positions in other codons: significant selection on amino acid substitutions causes diversity to be less than the neutral case, although in some cases (e.g. very small populations) it can be effectively neutral.

Probability of fixation or loss

Kimura's equation:

diploid $P(\text{fix mutation of freq. } x) = F = (1 - e^{-4Nesx}) / (1 - e^{-4Nes})$

new mutation $x = 1/2N$ $F = (1 - e^{-2Nes/N}) / (1 - e^{-4Nes})$

haploid $F = (1 - e^{-2Nesx}) / (1 - e^{-2Nes})$

new mutation $x = 1/N$ $F = (1 - e^{-2Nes/N}) / (1 - e^{-2Nes})$

Solving equation for various values of N_e and s shows:

- Even an advantageous mutation is more likely to be lost due to drift than it is to be fixed, unless the selective advantage is remarkably high. This illustrates the power of drift. It also shows that most new mutations will be lost regardless of the effects of selection.
- Even a detrimental mutation can be fixed. This again illustrates the power of random drift.

Nevertheless, even though these calculations emphasize the importance of drift, directional selection often determines the ultimate fate of mutations. It is important in two general ways: it promotes the fixation of advantageous mutations that lead to greater fitness or adaptation to a new environment (positive selection), and it eliminates detrimental mutations that would otherwise decrease fitness (purifying selection).

The distinction between neutral and selected mutations is not entirely clearcut. If the population size is small, even mutations with very large values of s (or very small negative values) may have nearly the same fixation probability as a neutral mutation. Conversely, if the population is very large, even very small selection coefficients can substantially increase or decrease the fixation probability.

Note that N_e and s always appear together as $N_e s$. This expresses their antagonistic relationship. Think of drift as noise that interferes with selection, which is the signal.

- Natural selection is effective, i.e. determines the fate of a new mutation, when $2N_e |s| \geq 1$. Signal is louder than noise.
- Random drift determines the fate of a new allele when $2N_e |s| \ll 1$. Noise overwhelms the signal.

($2N_e |s| \approx 0.1$ may be small enough. Some authors use $4N_e |s| \ll 1$ or $N_e |s| \ll 1$.)

Therefore alleles with very small s , or alleles with large s in very small population, behave like neutral alleles. These are said to be effectively neutral alleles or nearly neutral alleles.

Balancing Selection

Balancing selection = any kind of selection that maintains 2 or more alleles in a population.

The classic example is sickle-cell anemia, in which the sickle-cell gene has been maintained in fairly high frequency even though it is detrimental. In countries where malaria is endemic, the genotypes and phenotypes are:

	<u>anemia</u>	<u>effects of malaria</u>
Hb ^A Hb ^A	none	severe
Hb ^A Hb ^S	mild	less severe
Hb ^S Hb ^S	severe	?

Overdominance = heterozygotes are more fit than either homozygote.

There are a number of kinds of balancing selection, notably:

1. overdominance
2. selection for different alleles in different environments: selection varies over time, or in different habitats
3. frequency-dependent selection: alleles are advantageous when rare, detrimental when common

Only a few cases of balancing selection have been clearly demonstrated. Many or most evolutionary biologists believe that it is less important in determining overall levels of diversity than directional selection, but this is still being debated. It is clearly of functional important in some specific cases, e.g. major histocompatibility locus in hominids.

Balancing selection increases heterozygosity and delays the fixation of alleles. It favors high polymorphism.

In the *Adh* locus of *Drosophila*, the fast and slow alleles are probably maintained by some unidentified kind of balancing selection. Genetic diversity for synonymous alleles is high in the vicinity of the fast-slow site; i.e. balancing selection acting on a single base pair difference delays fixation and increases heterozygosity at neighboring base pairs because they are closely linked to it.

Migration (Gene Flow)

Many species and populations are subdivided by various factors, e.g.:

- intrinsically low dispersal rates
- physical barriers
- habitat barriers

Subdivided populations tend to become genetically different; counteracted by migration.

Drift is partially independent in each subpopulation, so they may maintain or fix different alleles and increase variation in the population as a whole.

Handled theoretically by setting up models, e.g.

- continent-island
- stepping-stone

Whole population is a metapopulation.

Parameter is migration rate m .

Effect measured by Wright's F statistics or similar G statistics.

$m = 1$ migrant per generation can make metapopulation nearly identical to random-mating population with respect to some statistics.

Population subdivision important in many areas, e.g. speciation.

Polymorphism as a Double-Edged Sword: Adapted vs. Adaptable

Fisher's Fundamental Theorem

Fisher: the rate of change in mean fitness due to selection equals the additive genetic variance in fitness.

Monomorphic population can't evolve, except by selection on new mutations. The more different alleles there are in the population, the faster the population can adapt by selection for good alleles or genotypes.

Genetic Load

H. J. Muller pointed out that if there is more than one genotype with different fitnesses, all but one will be of less than optimal fitness, and the mean population fitness will be lower than if the population was monomorphic for the the optimal genotype.

e.g. segregation load: when alleles of different fitnesses are maintained by balancing selection, recombination continually produces less-fit homozygotes.

e.g. mutation load: mutation continually introduces detrimental mutations, which aren't eliminated immediately.

etc.

$$\text{load } L = (w_{\max} - w_{\text{mean}}) / w_{\text{mean}}$$

Mutation load is a function only of the total detrimental mutation rate per genome U , and is independent of the magnitude of the selection coefficient.

Large genetic load with hard selection can reduce fitness of population so that it becomes extinct (meltdown). The population can in theory be rescued by back mutations (very rare) or compensating mutations (more common).

When molecular methods showed very large proportion of loci are polymorphic, Jukes & King and Kimura argued that if polymorphism is maintained by balancing selection, the load would cause extinction. Their solution: most of the variation seen at the molecular level is selectively neutral.

Argument depends on specific kinds of selection; there are other solutions involving different kinds of selection.

Sex and Variation within Species

Sex = putting genes from two individuals together in one and recombining them. Meiotic also involves segregation of alleles, which is important in organisms with a significant diploid phase.

One locus: Hardy-Weinberg equilibrium

With random mating, frequency of heterozygotes is given by $2pq$. Although the conditions for H-W equilibrium are stringent and probably very rarely met, it turns out that this prediction is a good approximation in many cases. Used to estimate frequency of carriers of recessive human hereditary diseases.

Inbreeding reduces the proportion of heterozygotes. Measured by coefficient of inbreeding = F = probability that two copies of a gene in an individual are identical by descent. F ranges 0 (random mating) to 1.

Inbreeding can result from mechanics of fertilization (e.g. in peas the flower is enclosed by the keel which favors selfing), from behavior and mate choice (e.g. preference for individuals of similar phenotype), or from small effective population size.

Obligate inbreeding: observed heterozygosity goes to 0, fastest with autogamy (1 generation), fast with selfing, slower with sib mating, etc.

If some individuals self and others are outcrossing, heterozygosity goes to a low level but not to zero in infinite populations.

One locus: Selection on recessive detrimental mutations

Most new mutations are recessive. Consequently the mutant allele is subject to reduced or no selection in heterozygotes. This greatly retards selection, especially when the mutant allele is rare. It is easy to reduce the frequency of a detrimental mutation by selection, but very hard to completely eliminate it.

Two or more loci: Linkage (dis)equilibrium

Effectiveness of sex commonly measured by linkage disequilibrium.

Two or more loci: Hitchhiking and background selection

Consider the behavior of two loci (A, B), each with two alleles (1, 2), in a random-mating population.

Allele	Allele Frequency	Gamete Genotype	Gamete Frequency	Zygote Genotype	Zygote Frequency
A ₁	p _A	A ₁ B ₁	p _A p _B	A ₁ A ₁ B ₁ B ₁	p _A ² p _B ²
A ₂	q _A	A ₁ B ₂	p _A q _B	A ₁ A ₁ B ₁ B ₂	p _A ² p _B q _B
B ₁	p _B	A ₂ B ₁	q _A p _B	A ₁ A ₁ B ₂ B ₂	p _A ² q _B ²
B ₂	q _B	A ₂ B ₂	q _A q _B	A ₁ A ₂ B ₁ B ₁	2 p _A q _A p _B ²

etc.

Zygotic and gametic disequilibrium refers to deviations from these values.

Note that at equilibrium the genes are behaving independently. Anything that causes deviations from independence can result in disequilibrium:

- linkage
- asexual reproduction
- preferential inbreeding or outbreeding
- selection with epistasis (interactions between loci in determining fitness)

Also anything that causes gene frequencies to change can cause linkage disequilibrium:

- mutation
- drift (e.g. founder effects)

Hitchhiking

Hitchhiking means that an advantageous mutation (driver) goes to fixation and carries with it linked neutral mutations or even mutations that are detrimental (hitchhikers).

Hitchhiking reduces genetic variation. Variation slowly recovers due to new mutations.

Extreme case: periodic selection in asexual organism in which an advantageous mutation is fixed so quickly that genetic variation goes to zero at all loci.

Background selection (hitchhiking with lots of drivers going a little ways backward): detrimental mutations are lost and carry with them neutral mutations. If it happens at many loci, variation is reduced.