Chapter 7 – Selection I

Selection in Haploids

Selection in Diploids

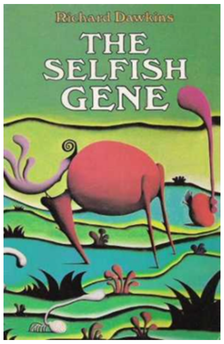
Mutation-Selection Balance

Darwinian Selection

- evolution vs. natural selection?
- evolution
  - "descent with modification"
  - change in allele frequency within a population
    - mutation, drift, selection, gene flow
- natural selection
  - "survival of the fittest" ?
  - differential survival and reproduction (**fecundity**)
- selection is an automatic consequence of differential fitness, not an "external force"

Units (or Levels) of Selection

- genes, individuals, populations ?
  - in general, selection is at the level of the individual (phenotype) but it is only the population that evolves
  - but see Dawkins, "*The Selfish Gene*"
- phenotypes versus genotypes?



Fitness

- What is fitness?
- fitness of a gene, genotype, individual?
- absolute versus relative fitness
  - the relative fitness of a genotype at a particular locus depends on complex interactions with the "genetic background" and the physical and ecological environment
- ~~selection produces adaptation~~
- adaptation is the automatic consequence of selection

Models of Natural Selection

- how do allele frequencies change over time ?
  - haploid/asexual versus diploid/sexual models
  - discrete generation versus continuous models

Selection in Haploid Organisms

- selection in asexual, haploid organisms depends only on relative population growth
  - competition between strains

Selection in Haploid Organisms

- strain A versus strain a
  - N_A, N_a - number of cells of each strain

$$N_{A(1)} = w_A N_{A(0)}$$

$$N_{a(1)} = w_a N_{a(0)}$$

  - if $w_A \neq w_a$, then the populations grow at different rates and the relative proportion of cell types changes over time

Selection in Haploid Organisms

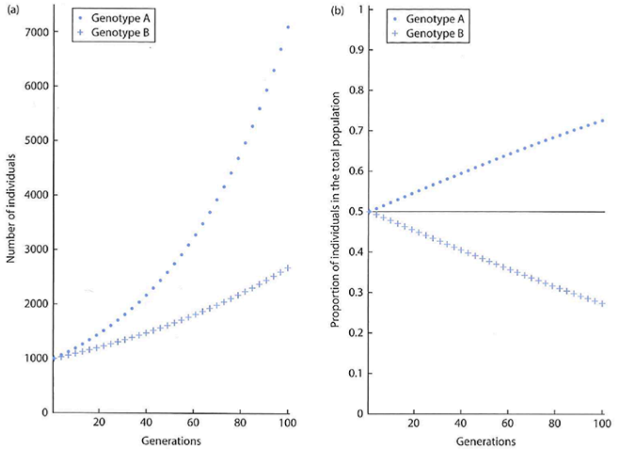


Figure 6.1 Population growth in two genotypes with clonal reproduction, starting out with equal numbers of individuals and therefore equal proportions in the total population. Genotype A grows 3% per generation ($\lambda = 1.03$) and genotype B grows 1% per generation ($\lambda = 1.01$). (a) Individuals of both genotypes increase in number over time. (b) Because the genotypes grow at different rates, their relative proportions in the total population change over time. The solid line shows the initial equal proportions. Eventually, genotype A will approach 100% and genotype B 0% of the total population. Values are plotted for every third generation.

Selection in Haploid Organisms

- frequency of type A after one generation:

$$f_A(1) = \frac{w_A N_A}{w_A N_A + w_a N_a}$$

$$= \frac{w_A f_A(0)}{w_A f_A(0) + w_a f_a(0)}$$

$$= \frac{f_A(0)}{f_A(0) + (w_a / w_A) f_a(0)}$$

Selection in Haploid Organisms

- coefficient of selection, s

$$\frac{w_a}{w_A} = 1 - s$$

$$f_A(1) = \frac{f_A(0)}{f_A(0) + (1-s) f_a(0)}$$

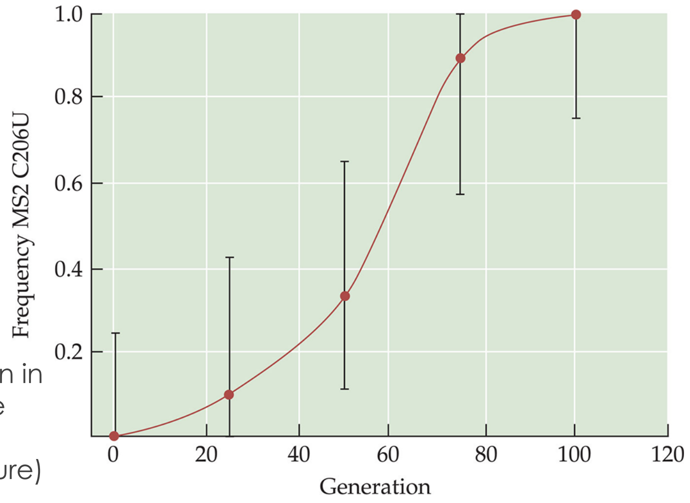
$$f_A(t) = \frac{w_A^t f_A(0)}{w_A^t f_A(0) + w_a^t f_a(0)} = \frac{f_A(0)}{f_A(0) + (1-s)^t f_a(0)}$$

Selection in Haploid Organisms

- What are we assuming so far?
  - ratio of growth rates remains constant over time
  - populations continue to grow indefinitely?
  - what happens when the population reaches the carrying capacity of the environment (K) or overshoots K and subsequently crashes?
- basic models of selection ignore population size and consider only changes in allele frequencies

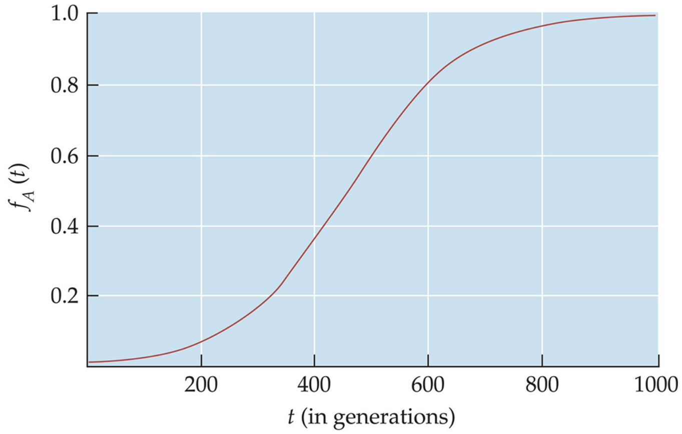
$$\Delta f_A = f_A(1) - f_A(0) = \frac{f_A(0)}{f_A(0) + (1-s) f_a(0)} - f_A(0)$$

Change in allele frequency due to selection can proceed quickly!



- C206U mutation in bacteriophage MS2 (grown at high temperature)

Change in allele frequency due to selection can proceed quickly!



- $f_A(0) = 0.01$, $s = 0.01$

Selection in Diploid Organisms

- selection (differential success) can occur at any stage in the life cycle
  - viability selection
  - sexual selection, fertility selection
  - gametic selection
    - incl. meiotic drive, segregation distortion
  - fecundity selection
- selection on (the success of) a particular genotype may vary through the life cycle
  - in fact, it's likely – tradeoffs are pervasive!

Selection in Diploid Organisms

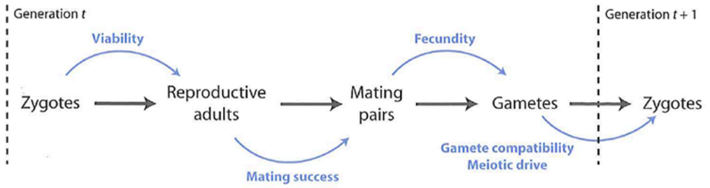


Figure 6.3 A diagram of the life cycle of organisms showing some points where differential survival and reproduction among genotypes can result in natural selection. Viability is the probability of survival from zygote to adult, mating success encompasses those traits influencing the chances of mating and the number of mates, and fecundity is the number of gametes and progeny zygotes produced by each mating pair. Gametic compatibility is the probability that gametes can successfully fuse to form a zygote whereas meiotic drive is any mechanism that causes bias in the frequency of alleles found in gametes. Most models of natural selection assume a single fitness component such as viability. In reality, all of these components of fitness can influence genotype frequencies simultaneously.

Single Locus Model of Viability Selection

- start with basic HW assumptions
- add differences in viability for different genotypes
  - assume survival probabilities (viability selection) for different genotypes remain constant over time

Absolute v. Relative Fitness

- absolute fitness (survival probability) versus relative fitness (in comparison to reference genotype)

Ospreys

$$AA:0.75,\ Aa:0.75,\ aa:0.50$$

$$AA:1.0,\ Aa:1.0,\ aa:0.67$$

vs. Oysters

$$AA:0.00075,\ Aa:0.00075,\ aa:0.00050$$

$$AA:1.0,\ Aa:1.0,\ aa:0.67$$

Single Locus Model of Viability Selection

- absolute fitness (survival probability) versus relative fitness (in comparison to reference genotype)
- outcome of selection depends on **relative fitness**
- absolute survival probabilities likely to change with population size and environmental conditions, but it may be reasonable to assume that relative fitness remains ~constant

Assumptions of the basic natural selection model with a di-allelic locus

- Genetic
  - Diploid individuals
  - One locus with two alleles
  - Obligate sexual reproduction
- Reproduction
  - Generations do not overlap
  - Mating is random
- Natural selection
  - Mechanism of natural selection is genotype-specific differences in survivorship (fitness), termed viability selection
  - Relative fitness values are constants that do not vary with time, over space, or in the two sexes
- Population
  - Infinite population size so there is no genetic drift
  - No population structure
  - No gene flow
  - No mutation

Change in Allele Frequency
w/ Viability Selection (Diploid Model)

- numbers of each genotype at time 0

$$Nf_{AA} = Nf_A^2 \qquad Nf_{Aa} = 2Nf_Af_a \qquad Nf_{aa} = Nf_a^2$$

- number of each genotype after selection given survival probabilities, v

$$Nv_{AA}f_A^2 \qquad 2Nv_{Aa}f_Af_a \qquad Nv_{aa}f_a^2$$

- total number of surviving adults:

$$N\left(v_{AA}f_A^2 + 2v_{Aa}f_Af_a + v_{aa}f_a^2\right) = N\bar{v}$$

Change in Allele Frequency w/ Viability Selection (Diploid Model)

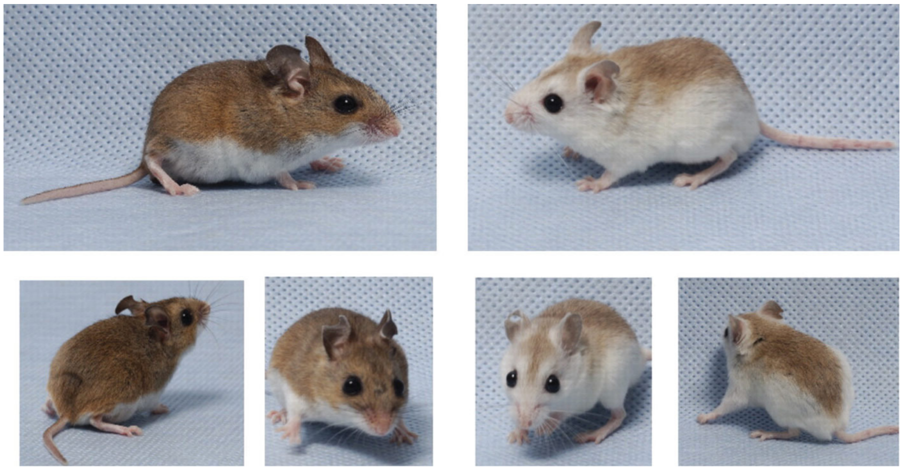
- new genotype frequencies

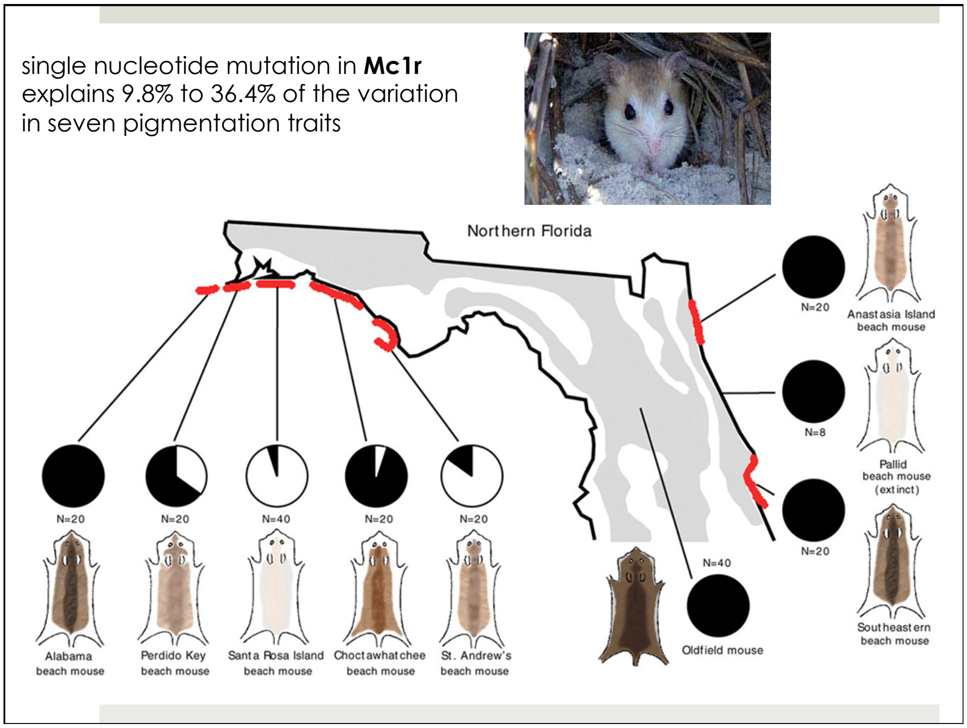
$$f'_{AA} = \frac{v_{AA} f_A^2}{\overline{v}} \qquad f'_{Aa} = \frac{2 v_{Aa} f_A f_a}{\overline{v}} \qquad f'_{aa} = \frac{v_{aa} f_a^2}{\overline{v}}$$

- …and new frequency of A allele

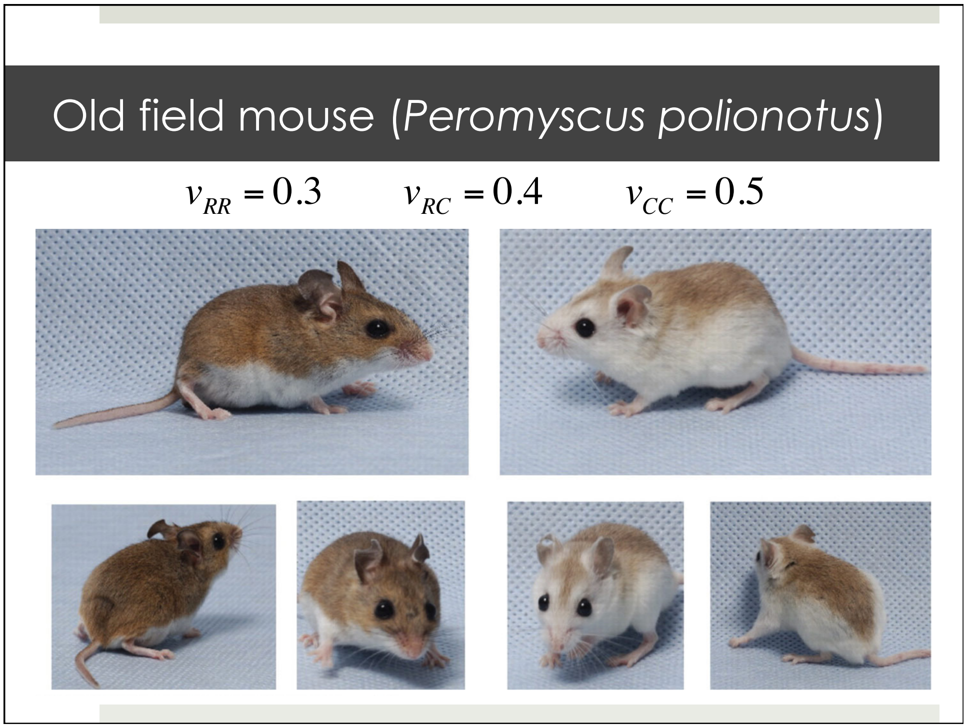
$$f'_A = \frac{v_{AA} f_A^2 + v_{Aa} f_A}{\overline{v}}$$

Old field mouse (*Peromyscus polionotus*)





single nucleotide mutation in **Mc1r** explains 9.8% to 36.4% of the variation in seven pigmentation traits



Old field mouse (*Peromyscus polionotus*)

$v_{RR} = 0.3$ $v_{RC} = 0.4$ $v_{CC} = 0.5$

Special Cases

- additive fitness (Box 7.4)

$$s_{aa} = 2s_{Aa} \qquad f_A' = \frac{f_A - sf_Af_a}{1-2sf_a}$$

- “genic selection” (Box 7.5)

$$w_{Aa} = w_{AA}\left(1-s\right), \qquad w_{aa} = w_{AA}\left(1-s\right)^2$$

  - completely equivalent to haploid selection

$$f_A' = \frac{f_A}{f_A + (1-s)f_a}$$

General scenarios for 2 alleles

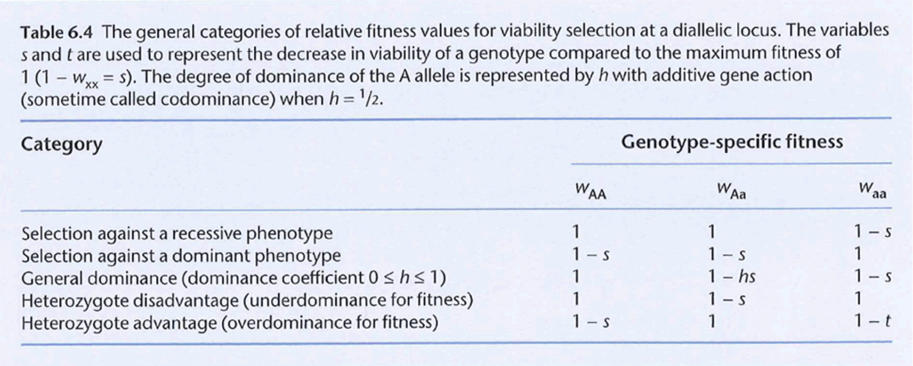


Table 6.4 The general categories of relative fitness values for viability selection at a diallelic locus. The variables *s* and *t* are used to represent the decrease in viability of a genotype compared to the maximum fitness of 1 ($1 - w_{xx} = s$). The degree of dominance of the A allele is represented by *h* with additive gene action (sometime called codominance) when $h = 1/2$.

Category	Genotype-specific fitness		
	w_{AA}	w_{Aa}	w_{aa}
Selection against a recessive phenotype	1	1	$1 - s$
Selection against a dominant phenotype	$1 - s$	$1 - s$	1
General dominance (dominance coefficient $0 \leq h \leq 1$)	1	$1 - hs$	$1 - s$
Heterozygote disadvantage (underdominance for fitness)	1	$1 - s$	1
Heterozygote advantage (overdominance for fitness)	$1 - s$	1	$1 - t$

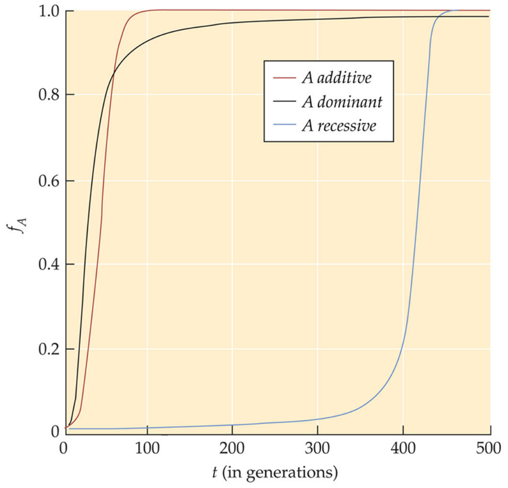
Directional Selection

$w_{AA} > w_{Aa} > w_{aa}$

- fate of new, rare allele depends on relative fitness of heterozygote
- rough approximation of time (# generations) required for a "substantial" change in allele frequency is:

$$1/s_{aa}$$

Directional Selection



$w_{AA} = 1, \quad w_{aa} = 0.8$

Heterozygote Advantage

- heterosis (hybrid advantage)
- overdominance (hybrid phenotype outside the range of the two homozygous phenotypes)
- one form of "balancing selection"
- results in stable polymorphism (assuming infinite population, etc.)
- equilibrium allele frequency:

$$\hat{f}_A = \frac{s_{aa}}{s_{AA} + s_{aa}}$$

Malaria

- **S** allele codes for a variant form of β-hemoglobin
- **SS** individuals suffer from severe anemia
  - sickle-cell disease
- **AS** individuals appear to be protected from malaria because infected cells undergo sickling and are removed from circulation
- ~fitness where malaria is prevalent
  - $w_{\mathrm{AA}} = 0.9$, $w_{\mathrm{AS}} = 1$, $w_{\mathrm{SS}} = 0.2$ ($s_{AA} = 0.075$; $s_{aa} = 0.748$)
- equilibrium allele frequency of $A = 0.89$ (~0.91)

Malaria

- glucose-6-phosphate dehydrogenase (G6PD) is an X-linked gene that helps control oxidative damage in erythrocytes
- **G6PD *A*-** has reduced activity but appears to protect carriers from severe malaria

---

distribution of sickle-cell allele

distribution of G6PD deficiency allele

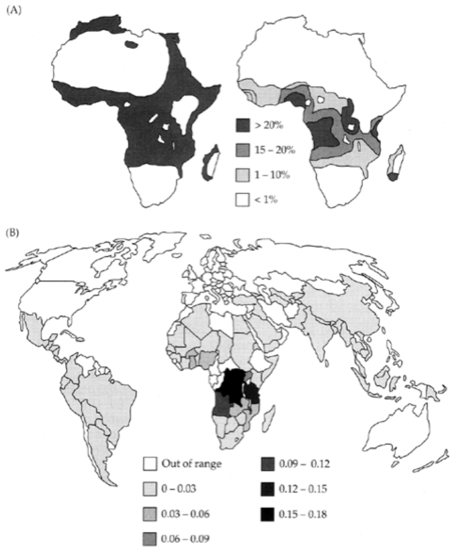
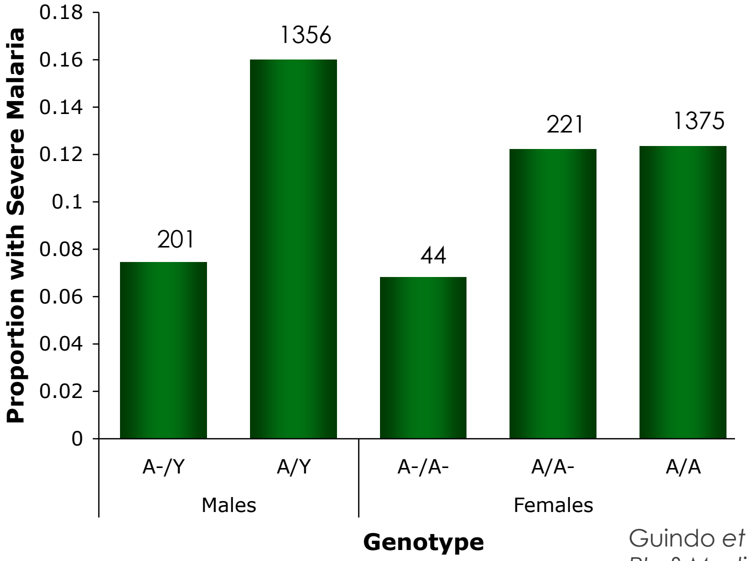


FIGURE 5.5 (A) A map of the distribution of the allele for sickle-cell anemia, and (B) map of the incidence of the *G6PD* deficiency allele. Both mutations are associated with resistance to malaria cause by *Plasmodium falciparum*. (A) The upper left map shows in gray the areas of incidence of falciparum malaria in Africa in the 1920s before mosquito control programs were implemented. The upper right figure shows the distribution of the beta-S globin allele. The global map in (B) shows the frequency of *G6PD* deficiency alleles indicated by the shading. The extensive overlap in the distributions relative to malaria was an early indication that there might be some causal connection.

X-Linked G6PD Deficiency Protects Hemizygous Males but Not Heterozygous Females against Severe Malaria



Guindo *et al*. 2007
PLoS Medicine

The C allele…

- CC homozygote has the highest fitness but is rare… why?

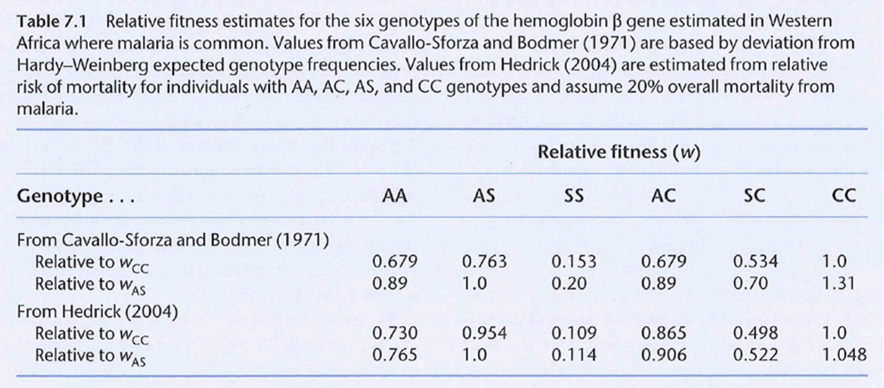


Table 7.1 Relative fitness estimates for the six genotypes of the hemoglobin β gene estimated in Western Africa where malaria is common. Values from Cavallo-Sforza and Bodmer (1971) are based by deviation from Hardy–Weinberg expected genotype frequencies. Values from Hedrick (2004) are estimated from relative risk of mortality for individuals with AA, AC, AS, and CC genotypes and assume 20% overall mortality from malaria.

	Relative fitness (w)					
Genotype . . .	**AA**	**AS**	**SS**	**AC**	**SC**	**CC**
From Cavallo-Sforza and Bodmer (1971)						
Relative to w_{CC}	0.679	0.763	0.153	0.679	0.534	1.0
Relative to w_{AS}	0.89	1.0	0.20	0.89	0.70	1.31
From Hedrick (2004)						
Relative to w_{CC}	0.730	0.954	0.109	0.865	0.498	1.0
Relative to w_{AS}	0.765	1.0	0.114	0.906	0.522	1.048

The C allele…

- CC homozygote has the highest fitness but is rare… why?
  - at equilibrium:

$$f_A = 8/9, f_S = 1/9, \text{ and } \bar{w} = 0.911$$

  - rare **C** allele has mean fitness of:

$$8/9 \times 0.89 + 1/9 \times 0.7 = 0.869$$

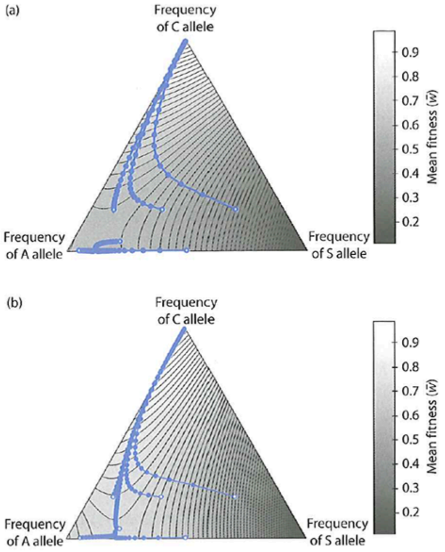


Figure 7.2 Fitness surfaces for the A, S, and C alleles at the human hemoglobin β gene when malaria is common. The surface in (a) corresponds to the top set of fitness values in Table 7.1 and (b) shows the surface for the bottom set of values. The tracks of circles represent generation-by-generation allele frequency trajectories due to natural selection over 50 generations calculated with equation 7.3. In (a), when the initial frequency of the C allele is relatively high, the equilibrium of natural selection is the fixation of the CC genotype. In contrast, when the C allele is initially rare (a frequency of less than about 7%) selection reaches an equilibrium with only the A and S alleles segregating and the C allele going to loss. In (b), selection will eventually fix the CC genotype from any initial frequency of the C allele. However, when the C allele is at low frequencies, the increase in the C allele each generation is extremely small so that selection would take hundreds of generations to fix the CC genotype. The six initial allele frequency points, shown as open circles, are identical for the two surfaces.

$$\Delta f_C = f_C \frac{(\bar{w}_C - \bar{w})}{\bar{w}} = \frac{(w_{CC} f_C + w_{AC} f_A + w_{SC} f_S) - \bar{w}}{\bar{w}}$$

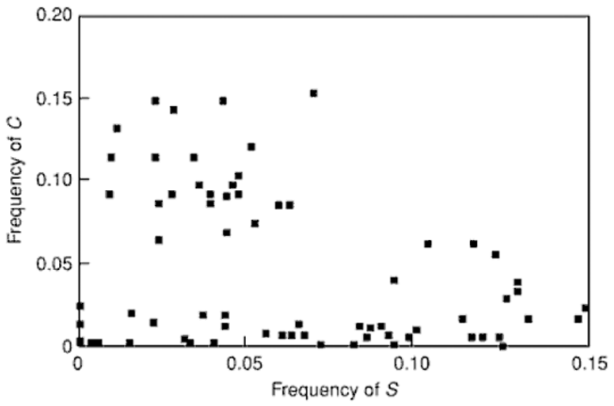


Figure 11.6. Frequencies of *S* and *C* alleles in 72 West African populations. Reprinted with permission from Fig. 4.12 in L. L. Cavalli-Sforza and W. F. Bodmer, *The Genetics of Human Populations*. Copyright ©1971 by W. H. Freeman and Co.

From: Templeton AR (2006) *Population Genetics and Microevolutionary Theory*

Heterozygote Disadvantage

$w_{AA} > w_{Aa} < w_{aa}$

- f_A = 0 and f_A = 1 are stable equilbiria
  - an unstable equilibrium occurs at:

$$f_A = \frac{w_{Aa} - w_{aa}}{2w_{Aa} - w_{AA} - w_{aa}}$$

  - outcome depends on starting point
    - population may end up at a local equilibrium with relatively low fitness

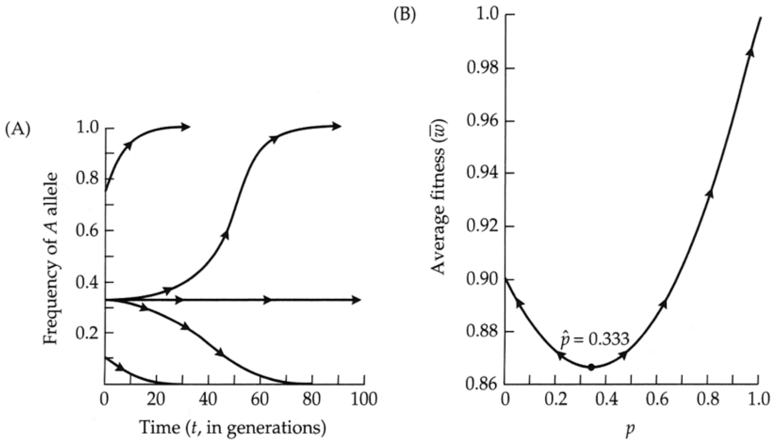


FIGURE 5.7 Selection when there is heterozygote inferiority. (A) The allele frequency goes to 0 or 1 depending on the initial frequency. In this example, $w_{11} = 1$, $w_{12} = 0.8$, and $w_{22} = 0.9$, and there is an unstable equilibrium when the frequency of the A allele is $\hat{p} = \frac{1}{3}$. An infinite population with $p = \frac{1}{3}$ maintains this frequency, but any slight upward change in the frequency of A results in eventual fixation, and any slight downward change in the frequency of A results in ultimate loss. (B) Average fitness $\bar{w}$ against p for the same example. The unstable equilibrium represents the minimum of $\bar{w}$.

Heterozygote Disadvantage

- not common
  - or at least difficult to detect and distinguish from directional selection against rare allele
- may apply to chromosomal inversions and translocations which result in mis-pairing of chromosomes during meiosis
  - potentially important in speciation, but how do they get started if disadvantageous when rare?

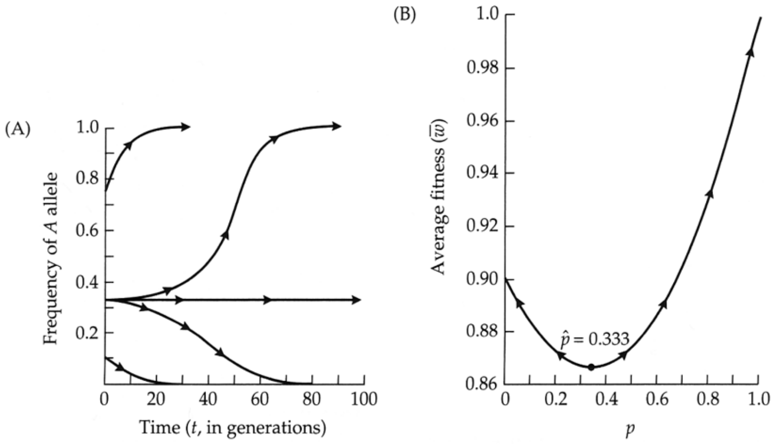


FIGURE 5.7 Selection when there is heterozygote inferiority. (A) The allele frequency goes to 0 or 1 depending on the initial frequency. In this example, $w_{11} = 1$, $w_{12} = 0.8$, and $w_{22} = 0.9$, and there is an unstable equilibrium when the frequency of the *A* allele is $\hat{p} = \frac{1}{3}$. An infinite population with $p = \frac{1}{3}$ maintains this frequency, but any slight upward change in the frequency of *A* results in eventual fixation, and any slight downward change in the frequency of *A* results in ultimate loss. (B) Average fitness $\bar{w}$ against *p* for the same example. The unstable equilibrium represents the minimum of $\bar{w}$.

Assumptions?

- infinite population size, no mutation
- all equations above are deterministic (no stochastic component)

Adaptive "Topography" & Drift

- fitness as a function of allele frequencies forms an adaptive topography or landscape

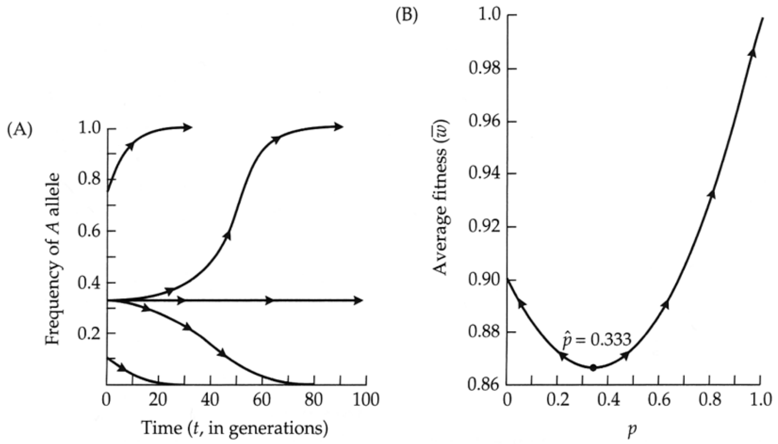
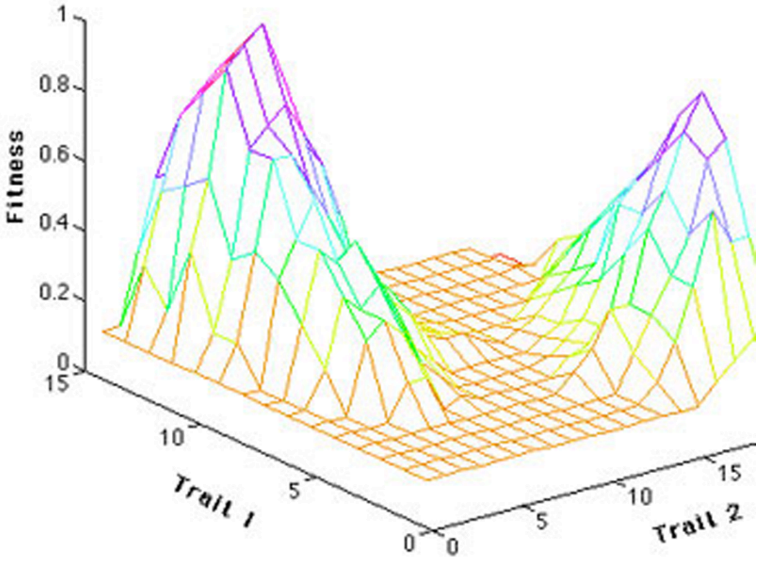
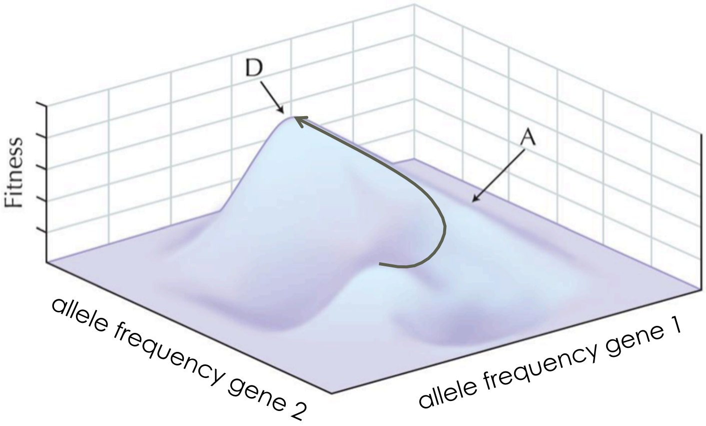


FIGURE 5.7 Selection when there is heterozygote inferiority. (A) The allele frequency goes to 0 or 1 depending on the initial frequency. In this example, w_{11} = 1, w_{12} = 0.8, and w_{22} = 0.9, and there is an unstable equilibrium when the frequency of the *A* allele is $\hat{p} = \frac{1}{3}$. An infinite population with $p = \frac{1}{3}$ maintains this frequency, but any slight upward change in the frequency of *A* results in eventual fixation, and any slight downward change in the frequency of *A* results in ultimate loss. (B) Average fitness $\bar{w}$ against *p* for the same example. The unstable equilibrium represents the minimum of $\bar{w}$.



Adaptive “Topography” & Drift

- fitness as a function of allele frequencies forms an adaptive topography or landscape
- how can populations escape non-optimal, local equilibria?
  - Wright – “shifting balance theory”
    - important role for genetic drift in moving sub-populations between adaptive peaks
  - the fitness landscape may also change with environmental conditions, population density, etc.



Mutation-Selection Balance

- approximate change in allele frequency due to the combined action of mutation and selection on a deleterious allele(s) with reduced fitness in heterozygote:

$$\Delta f_a = \mu - s_{Aa} f_a$$

- equilibrium allele frequency:

$$\hat{f}_a = \frac{\mu}{s_{Aa}}$$

Mutation-Selection Balance

- approximate change in allele frequency due to the combined action of mutation and selection on a deleterious allele(s) with reduced fitness only in homozygote:

$$\Delta f_a = \mu - s_{aa} f_a^2$$

- equilibrium allele frequency:

$$\hat{f}_a = \sqrt{\frac{\mu}{s_{aa}}}$$

Mutation-Selection Balance

- the deleterious "a" allele in above equations is actually a heterogeneous set of non-functional alleles
- μ is the rate of mutation from "normal, functional" allele to dysfunctional allele

Fertility Selection

- occurs when offspring production of a mated pair depends on the genotypes of the parents rather than the genotypes of the offspring
- e.g., Rh blood system and hemolytic disease
    - Rh negative mothers can produce antibodies against their developing Rh+ fetuses (if father is Rh+)

Fertility Selection

TABLE 7.1 **Genotypes of families of the *Rh* system**

Father	Mother	Frequency	Offspring viability
RR	*rr*	$f_{RR}f_{rr}$	$1–2s$
Rr	*rr*	$f_{Rr}f_{rr}$	$1–s$
RR, *Rr*	*RR*, *Rr*	$(1–f_{rr})^2$	1
rr	*RR*, *Rr*, *rr*	f_{rr}	1

Rh allele frequencies

Population	Rh(D) Neg	Rh(D) Pos	Rh(D) Neg alleles
Basque people	21–36%[13]	65%	approx 60%
other Europeans	16%	84%	40%
African American	approx 7%	93%	approx 26%
Native Americans	approx 1%	99%	approx 10%
African descent	less 1%	over 99%	3%
Asian	less 1%	over 99%	1%

Fertility Selection

- occurs when offspring production of a mated pair depends on the genotypes of the parents rather than the genotypes of the offspring
- e.g., Rh blood system and hemolytic disease
  - Rh negative mothers can produce antibodies against their developing Rh+ fetuses (if father is Rh+)
- e.g., self-compatibility loci in plants
  - genotypes at S-locus determine ability of pollen to grow pollen tube
  - frequency dependent selection leads to high levels of polymorphism