

# Natural Selection

---

COMP 571 - Fall 2010  
Luay Nakhleh, Rice University

# Outline

- (1) Natural selection**
- (2) Natural selection on a diallelic locus**
- (3) Natural selection and fitness**
- (4) Viability selection with three alleles or two loci**
- (5) Alternative models of natural selection**
- (6) Combining natural selection with other processes**
- (7) Natural selection and the coalescent**

# (1) Natural Selection

- \* Charles Darwin's statement of the process of **natural selection** can be summarized as three basic observations about populations:
- \* all species have more offspring than can possibly survive and reproduce
- \* individual organisms vary in phenotypes that influence their ability to survive and reproduce, and
- \* within each generation, the individuals possessing phenotypes that confer greater survival and reproduction will contribute more offspring to the next generation

# (1) Natural Selection

- \* Once particulate inheritance was understood, the unification of genetics with the principle of natural selection took place in what is now called the **modern synthesis** or **neo-Darwinian synthesis** of evolutionary biology
- \* We'll now describe basic population genetic expectations for natural selection with (1) clonal reproduction, and (2) sexual reproduction

# (1) Natural Selection

## Natural selection with clonal reproduction

- \* At its core, natural selection is a process of population growth

$$N_{t+1} = \lambda N_t$$

↑                    ↑  
number        finite  
of             rate of  
ind.         increase

- \* The population growth rate can be thought of as the chance that an individual contributes one offspring to the next generation

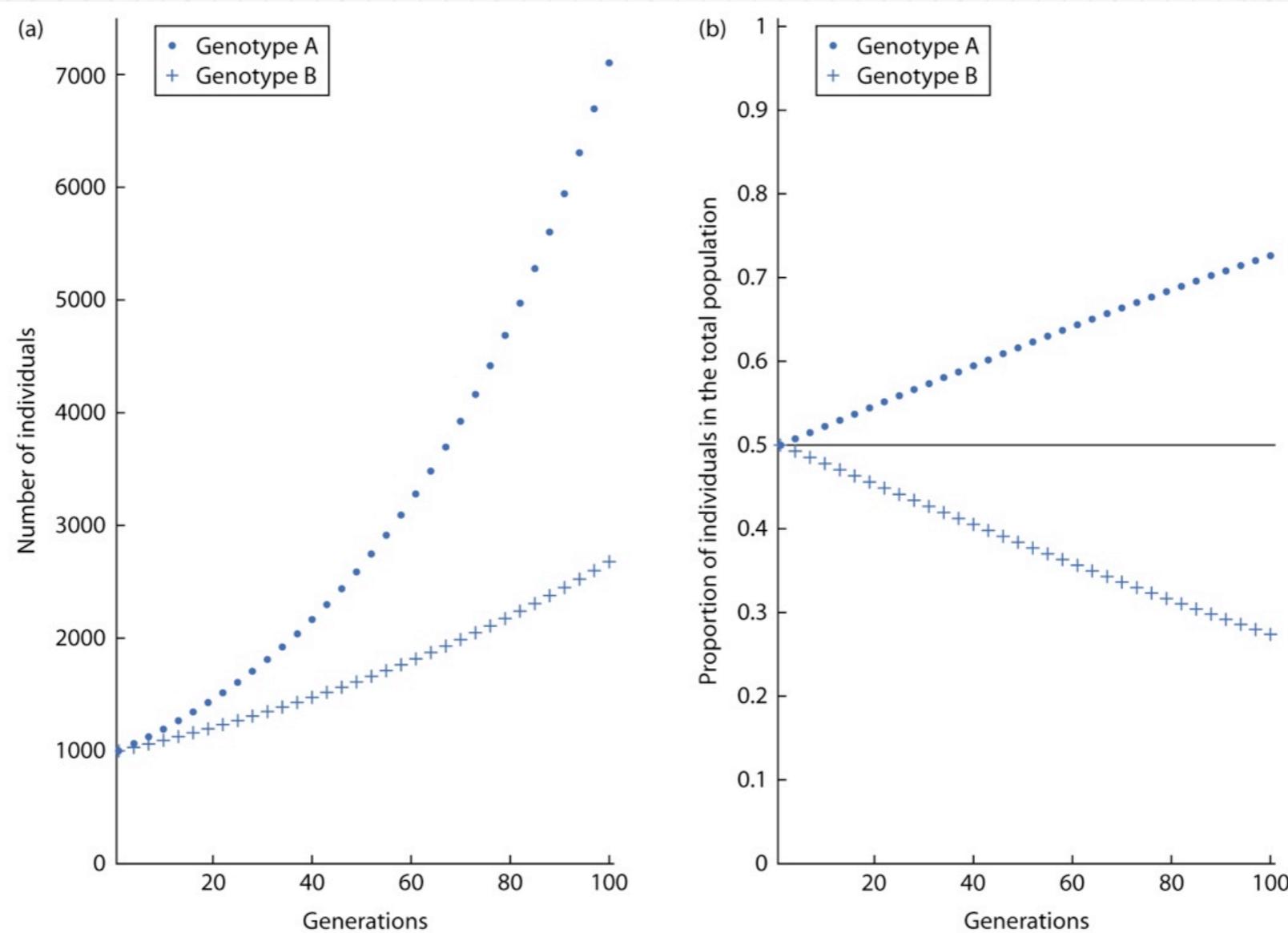
# (1) Natural Selection

- \* Let's consider a population composed of two genotypes A and B of an asexual organism, with genotype-specific growth rates, or absolute fitnesses, of  $\lambda_A$  and  $\lambda_B$
- \* The proportions of each genotype in the population in any generation are

$$p = \frac{N_A}{N_A + N_B}$$

$$q = \frac{N_B}{N_A + N_B}$$

# (1) Natural Selection



**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.

# (1) Natural Selection

\* An alternative way to represent the changing proportions of the two genotypes in the population is to follow the ratio  $N_A/N_B$  over time

\* This ratio after one generation is given by

$$\frac{N_B(t = 1)}{N_A(t = 1)} = \left( \frac{\lambda_B}{\lambda_A} \right) \frac{N_B(0)}{N_A(0)}$$

\* And, this ratio at time  $t$  is given by

$$\frac{N_B(t)}{N_A(t)} = \left( \frac{\lambda_B}{\lambda_A} \right)^t \frac{N_B(0)}{N_A(0)}$$

# (1) Natural Selection

- \* The ratio  $\lambda_B/\lambda_A$  is called the **relative fitness**, denoted by  $w$ , and using it gives us the following:

$$\frac{N_B(t)}{N_A(t)} = w^t \frac{N_B(0)}{N_A(0)}$$

- \* Dividing the nominator and denominator on both sides by the population size ( $N=N_A+N_B$ ), we obtain

$$\frac{q_t}{p_t} = w^t \frac{q_0}{p_0}$$

# (1) Natural Selection

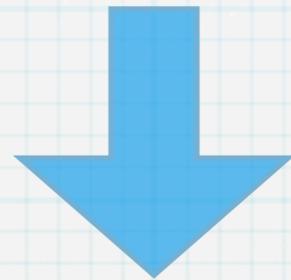
The relative fitness can be used to determine that change in frequency of a genotype over time:

**Table 6.1** The expected frequencies of two genotypes after natural selection, for the case of clonal reproduction. The top section of the table gives expressions for the general case. The bottom part of the table uses absolute and relative fitness values identical to Fig. 6.1 to show the change in genotype proportions for the first generation of natural selection. The absolute fitness of the A genotype is highest and is therefore used as the standard of comparison when determining relative fitness.

	Genotype	
	A	B
<b>Generation t</b>		
Initial frequency	$p_t$	$q_t$
Genotype-specific growth rate (absolute fitness)	$\lambda_A$	$\lambda_B$
Relative fitness	$w_A = \frac{\lambda_A}{\lambda_A}$	$w_B = \frac{\lambda_B}{\lambda_A}$
Frequency after natural selection	$p_t w_A$	$q_t w_B$
<b>Generation t + 1</b>		
Initial frequency $p_{t+1}$	$\frac{p_t w_A}{p_t w_A + q_t w_B}$	$\frac{q_t w_B}{p_t w_A + q_t w_B}$
Change in genotype frequency	$\Delta p = p_{t+1} - p_t$	$\Delta q = q_{t+1} - q_t$
<b>Generation t</b>		
Initial frequency	$p_t = 0.5$	$q_t = 0.5$
Genotype-specific growth rate (absolute fitness)	$\lambda_A = 1.03$	$\lambda_B = 1.01$
Relative fitness	$w_A = \frac{\lambda_A}{\lambda_A} = \frac{1.03}{1.03} = 1.0$	$w_B = \frac{\lambda_B}{\lambda_A} = \frac{1.01}{1.03} = 0.981$
Frequency after natural selection	$p_t w_A = (0.5)(1.0) = 0.5$	$q_t w_B = (0.5)(0.981) = 0.4905$
<b>Generation t + 1</b>		
Initial frequency $p_{t+1}$	$\frac{0.5}{0.5 + 0.4905} = 0.5048$	$\frac{0.4905}{0.5 + 0.4905} = 0.4952$
Change in genotype frequency	$0.5048 - 0.5 = 0.0048$	$0.4952 - 0.5 = -0.0048$

# (1) Natural Selection

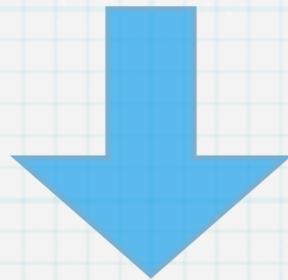
$$\Delta p = p_{t+1} - p_t \quad + \quad p_{t+1} = \frac{p_t w_A}{p_t w_A + q_t w_B}$$



$$\Delta p = \frac{p_t w_A}{p_t w_A + q_t w_B} - p_t$$

# (1) Natural Selection

$$\Delta p = p_{t+1} - p_t \quad + \quad p_{t+1} = \frac{p_t w_A}{p_t w_A + q_t w_B}$$



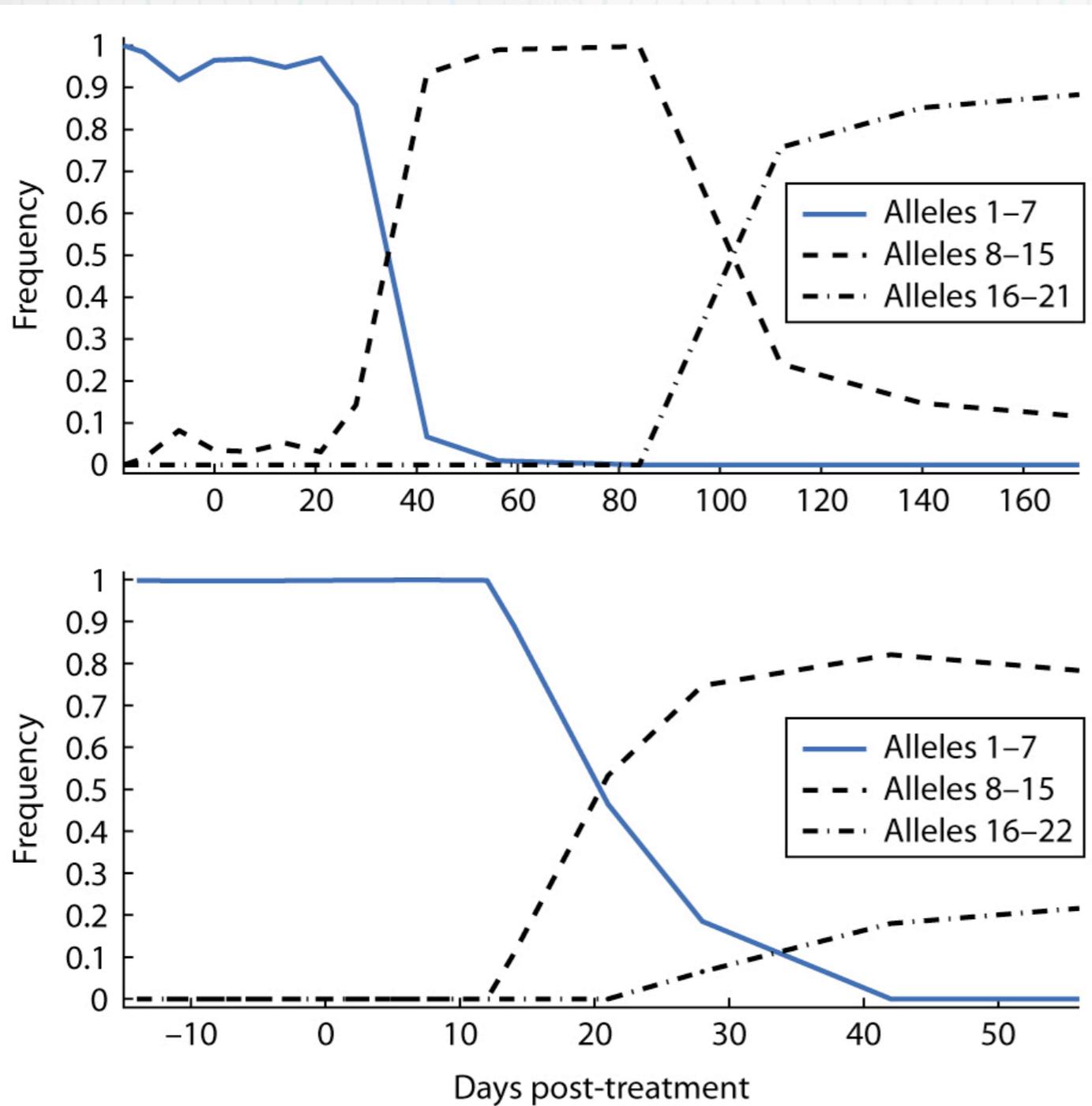
$$\Delta p = \frac{p_t w_A}{p_t w_A + q_t w_B} - p_t$$

average relative  
fitness

# (1) Natural Selection

- \* One example of natural selection “at work” is the evolution of drug-resistant genotypes in HIV

# (1) Natural Selection



**Figure 6.2** Allele frequencies at the protease locus over time in the HIV populations in two patients undergoing protease inhibitor (ritonavir) treatment (Doukhan & Delwart 2001). Alleles found at very low frequencies before drug treatment come to predominate in the HIV population after drug treatment, due to natural selection among HIV genotypes for drug resistance. Alleles are bands observed in denaturing-gradient gel electrophoresis (DGGE), a technique that is capable of discriminating single-base-pair differences among different DNA fragments. DGGE was used to identify the number of different protease locus DNA sequences present in a sample of HIV particles. Protease inhibitor treatment began on day 0. Dr. E. Delwart kindly provided the original data used to draw this figure.

# (1) Natural Selection

Natural selection with sexual reproduction

- \* To build a model of natural selection for sexual reproduction, we can combine the Hardy-Weinberg model of genotype frequencies with genotype-specific growth rates to get a general model of natural selection operating on the three genotypes produced by a single locus with two alleles

# (1) Natural Selection

**Table 6.2** Assumptions of the basic natural selection model with a diallelic 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) that lead to variable genotype-specific growth rates, termed viability selection
- \* 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
-

# (1) Natural Selection

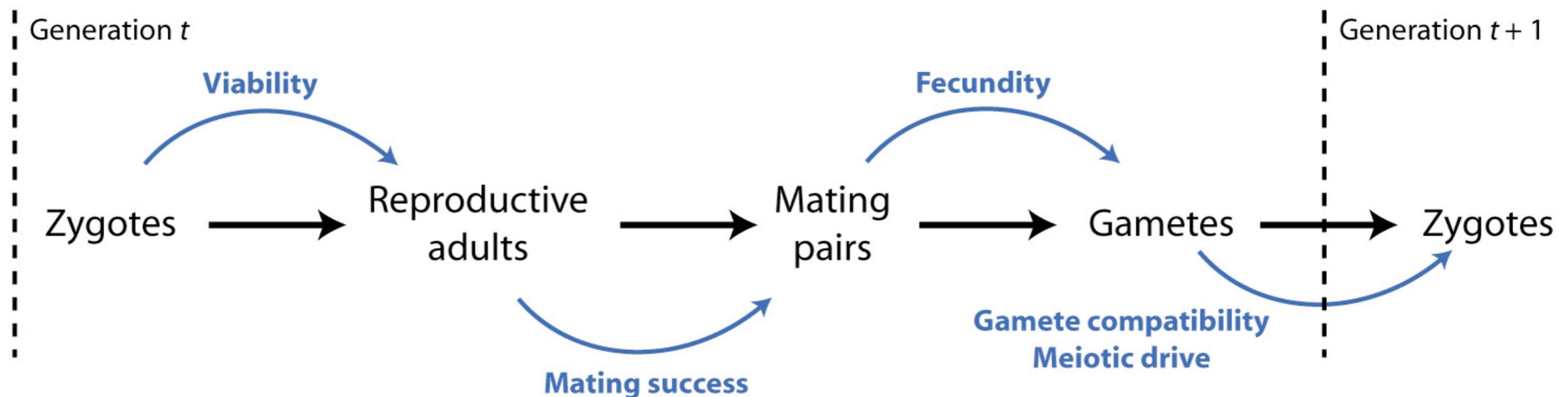
- \* Assume a population of  $N$  diploid individuals with random mating and random fusion of gametes to produce zygotes
- \* Before natural selection, the genotypes are in HW equilibrium

$$AA : p^2 N_t \quad Aa : 2pqN_t \quad aa : q^2 N_t$$

# (1) Natural Selection

- \* After the initial population of zygotes is produced, natural selection will then operate on the three genotypes
- \* Each genotype is assumed to experience genotype-specific survival and reproduction during the course of a single generation as follows

# (1) Natural Selection



**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.

# (1) Natural Selection

- \* Focusing on viability selection, let  $\ell$  represent the genotype-specific probability of survival to reproductive age
- \* The number of individuals of each genotype after viability selection at the point of reproduction is

$$AA : \ell_{AA} p^2 N_t \quad Aa : \ell_{Aa} 2pq N_t \quad aa : \ell_{aa} q^2 N_t$$

# (1) Natural Selection

- \* The total number of individuals in the population after selection is then

$$l_{AA}p^2N_t + l_{Aa}2pqN_t + l_{aa}q^2N_t$$

- \* The frequency of the AA genotype is

$$\frac{l_{AA}p^2N_t}{l_{AA}p^2N_t + l_{Aa}2pqN_t + l_{aa}q^2N_t}$$

# (1) Natural Selection

- \* The frequency of A allele in gametes is

$$\frac{\ell_{AA}p^2N_t + \frac{1}{2}(\ell_{Aa}2pqN_t)}{\ell_{AA}p^2N_t + \ell_{Aa}2pqN_t + \ell_{aa}q^2N_t}$$

- \* The frequency of a allele in gametes is

$$\frac{\ell_{aa}q^2N_t + \frac{1}{2}(\ell_{Aa}2pqN_t)}{\ell_{AA}p^2N_t + \ell_{Aa}2pqN_t + \ell_{aa}q^2N_t}$$

# (1) Natural Selection

After canceling the  $N_t$  terms, we obtain

- \* The frequency of A allele in gametes is

$$\frac{\ell_{AA}p^2 + \ell_{Aa}pq}{\ell_{AA}p^2 + \ell_{Aa}2pq + \ell_{aa}q^2}$$

- \* The frequency of a allele in gametes is

$$\frac{\ell_{aa}q^2 + \ell_{Aa}pq}{\ell_{AA}p^2 + \ell_{Aa}2pq + \ell_{aa}q^2}$$

# (1) Natural Selection

Using the relative fitness values, we obtain

$$p_{t+1} = \frac{w_{AA}p^2 + w_{Aa}pq}{w_{AA}p^2 + w_{Aa}2pq + w_{aa}q^2}$$

$$q_{t+1} = \frac{w_{aa}q^2 + w_{Aa}pq}{w_{AA}p^2 + w_{Aa}2pq + w_{aa}q^2}$$

# (1) Natural Selection

Using the mean relative fitness, we obtain

$$p_{t+1} = \frac{w_{AA}p^2 + w_{Aa}pq}{\bar{w}}$$

$$q_{t+1} = \frac{w_{aa}q^2 + w_{Aa}pq}{\bar{w}}$$

# (1) Natural Selection

Using the mean relative fitness, we obtain

$$p_{t+1} = \frac{w_{AA}p^2 + w_{Aa}pq}{\bar{w}}$$
$$q_{t+1} = \frac{w_{aa}q^2 + w_{Aa}pq}{\bar{w}}$$

marginal fitness

# (1) Natural Selection

**Table 6.3** The expected frequencies of three genotypes after natural selection for a diallelic locus with sexual reproduction and random mating. The absolute fitness of the AA genotype is used as the standard of comparison when determining relative fitness.

	Genotype		
	AA	Aa	aa
Generation $t$			
Initial frequency	$p_t^2$	$2p_tq_t$	$q_t^2$
Genotype-specific survival (absolute fitness)	$\ell_{AA}$	$\ell_{Aa}$	$\ell_{aa}$
Relative fitness	$w_{AA} = \frac{\ell_{AA}}{\ell_{AA}}$	$w_{Aa} = \frac{\ell_{Aa}}{\ell_{AA}}$	$w_{aa} = \frac{\ell_{aa}}{\ell_{AA}}$
Frequency after natural selection	$p_t^2 w_{AA}$	$2p_tq_t w_{Aa}$	$q_t^2 w_{aa}$
Average fitness	$p_t^2 w_{AA} + 2p_tq_t w_{Aa} + q_t^2 w_{aa}$		
Generation $t + 1$			
Genotype frequency	$\frac{p_t^2 w_{AA}}{\bar{w}}$	$\frac{2p_tq_t w_{Aa}}{\bar{w}}$	$\frac{q_t^2 w_{aa}}{\bar{w}}$
Allele frequency	$p_{t+1} = \frac{p_t(p_t w_{AA} + q_t w_{Aa})}{\bar{w}}$	$q_{t+1} = \frac{q_t(q_t w_{aa} + p_t w_{Aa})}{\bar{w}}$	
Change in allele frequency	$\Delta p = \frac{pq[p(w_{AA} - w_{Aa}) + q(w_{Aa} - w_{aa})]}{\bar{w}}$	$\Delta q = \frac{pq[q(w_{aa} - w_{Aa}) + p(w_{Aa} - w_{AA})]}{\bar{w}}$	

# (1) Natural Selection

- \* In terms of change in allele frequency over one generation, we have

$$\Delta p = \frac{pq[p(w_{AA} - w_{Aa}) + q(w_{Aa} - w_{aa})]}{\bar{w}}$$

$$\Delta q = \frac{pq[q(w_{aa} - w_{Aa}) - p(w_{AA} - w_{Aa})]}{\bar{w}}$$

## (2) Natural Selection on a Diallelic Locus

- \* All the outcomes of the selection model can be represented by five general categories of fitness values for the three genotypes

## (2) Natural Selection on a Diallelic Locus

**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 (sometimes called codominance) when  $h = 1/2$ .

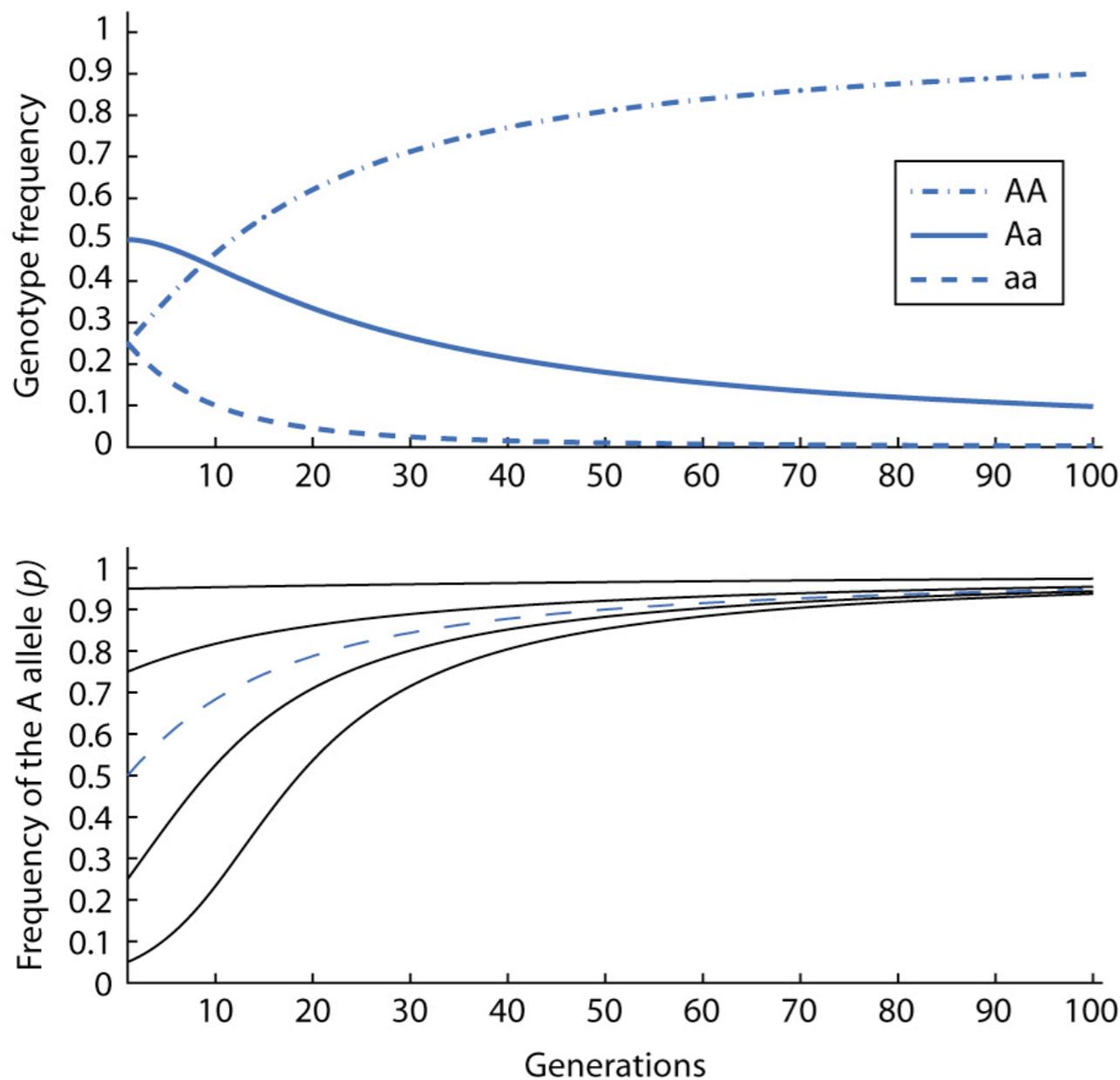
Category	Genotype-specific fitness		
	$w_{AA}$	$w_{Aa}$	$w_{aa}$
Selection against a recessive phenotype	1	1	$1 - s$
Selection against a dominant phenotype	$1 - s$	$1 - s$	1
General dominance (dominance coefficient $0 \leq h \leq 1$ )	1	$1 - hs$	$1 - s$
Heterozygote disadvantage (underdominance for fitness)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

Fitness values are given in terms of **selection coefficients**, which are the difference between a relative fitness value and one:

$$s_{xx} = 1 - w_{xx} \quad \text{OR} \quad w_{xx} = 1 - s_{xx}$$

# (2) Natural Selection on a Diallelic Locus

## Selection against a recessive phenotype



**Figure 6.4** The change in genotype and allele frequencies caused by viability selection against the aa genotype exhibiting the recessive phenotype. The top panel shows the change in genotype frequencies over time and the bottom panel shows the frequency of the dominant allele (A) over time. The colored, dashed line in the bottom panel corresponds to the allele frequencies in the top panel. Because of changes in genotype frequency caused by natural selection, the frequency of the dominant allele rapidly approaches fixation from all five initial allele frequencies. In this illustration  $w_{AA} = w_{Aa} = 1.0$  while  $w_{aa} = 0.8$ , meaning that eight individuals with the aa genotype are expected to survive to reproduce for every 10 individuals with the AA or Aa genotype that survive to reproduce each generation. Genotype frequencies assume random mating.

## (2) Natural Selection on a Diallelic Locus

Selection against a recessive phenotype

- \* The dominant allele does NOT go to fixation because the heterozygote fitness is equal to the maximum fitness and every generation heterozygotes will produce gametes that can combine to make the recessive homozygote
- \* In essence, the recessive allele is shielded from natural selection in the heterozygote due to dominance

## (2) Natural Selection on a Diallelic Locus

Selection against a recessive phenotype

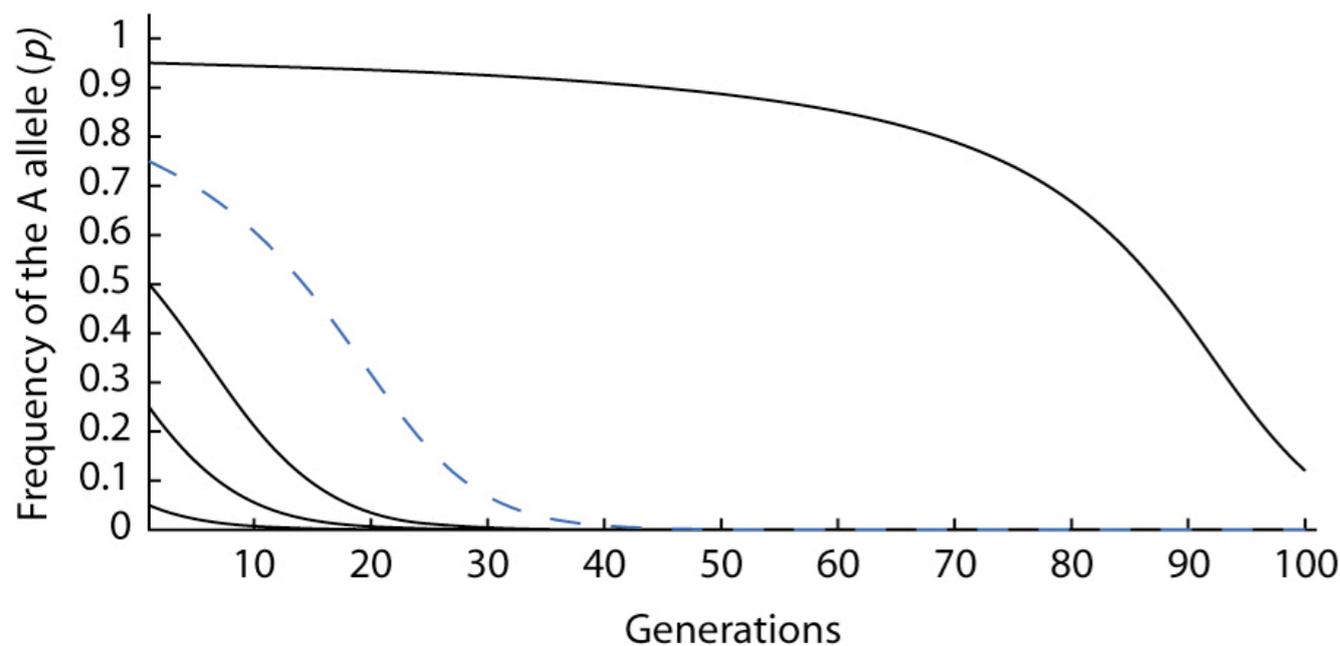
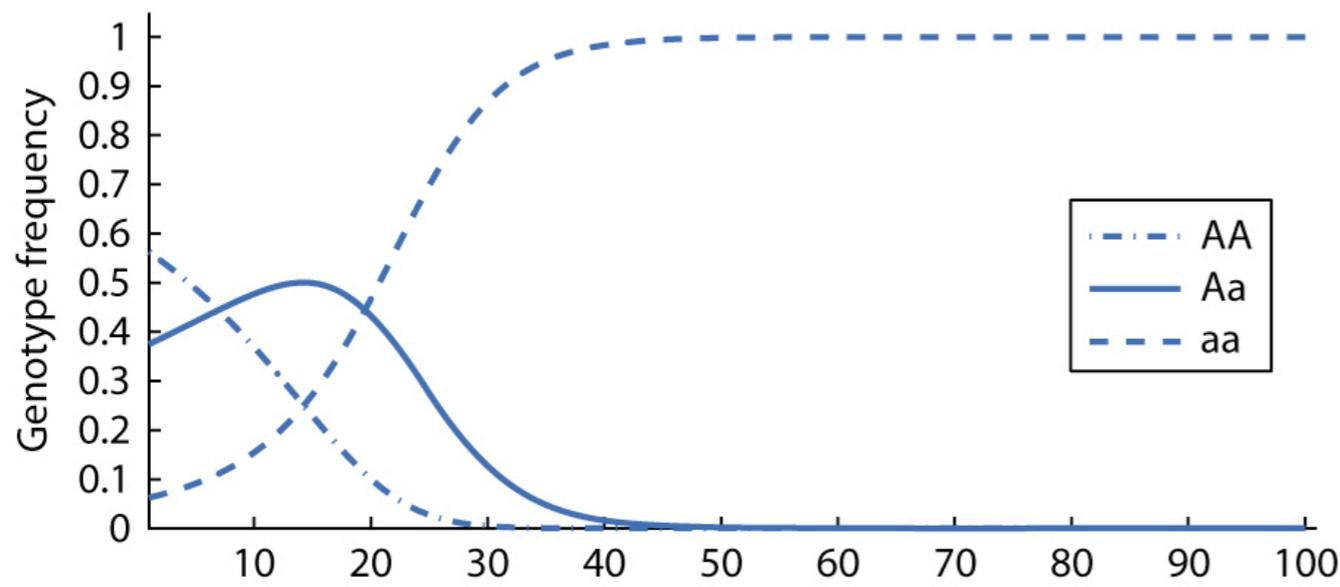
- \* One way to quantify the sheltering effect of heterozygotes is to examine the proportion of recessive alleles present in heterozygotes compared to recessive alleles present in homozygotes

$$\frac{pq}{q^2} = \frac{p}{q}$$

- \* When the frequency of the recessive allele is low, say  $q=0.05$ , the proportion of the genotype frequencies is 19. This means that there are 19 recessive alleles protected against natural selection in heterozygotes for each recessive allele impacted by natural selection in a homozygous genotype

# (2) Natural Selection on a Diallelic Locus

Selection against a dominant phenotype



**Figure 6.5** The change in the genotype and allele frequency of a completely dominant allele (A) when natural selection acts against the AA and Aa genotypes exhibiting the dominant phenotype. Notice that the frequency of the A allele decreases slowly at first when the A allele is common in the population since the aa genotype is infrequent. The colored, dashed line in the bottom panel corresponds to the allele frequencies in the top panel. In this illustration  $w_{AA} = w_{Aa} = 0.8$  while  $w_{aa} = 1.0$ . Genotype frequencies assume random mating.

# (2) Natural Selection on a Diallelic Locus

Selection against a dominant phenotype

- \* The recessive allele goes to fixation when there is natural selection against the dominant homozygote and heterozygote since both have a lower fitness than the favored homozygote and therefore the dominant allele is not shielded from natural selection in the heterozygote

# (2) Natural Selection on a Diallelic Locus

## General dominance

- \* The previous two examples cover the two extremes of dominance
- \* To consider more general cases, we can use a **dominance coefficient  $h$**
- \* Complete dominance for one allele (the homozygote and heterozygote have identical phenotypes) is represented by  $h=0$  and complete dominance for the other allele is represented by  $h=1$

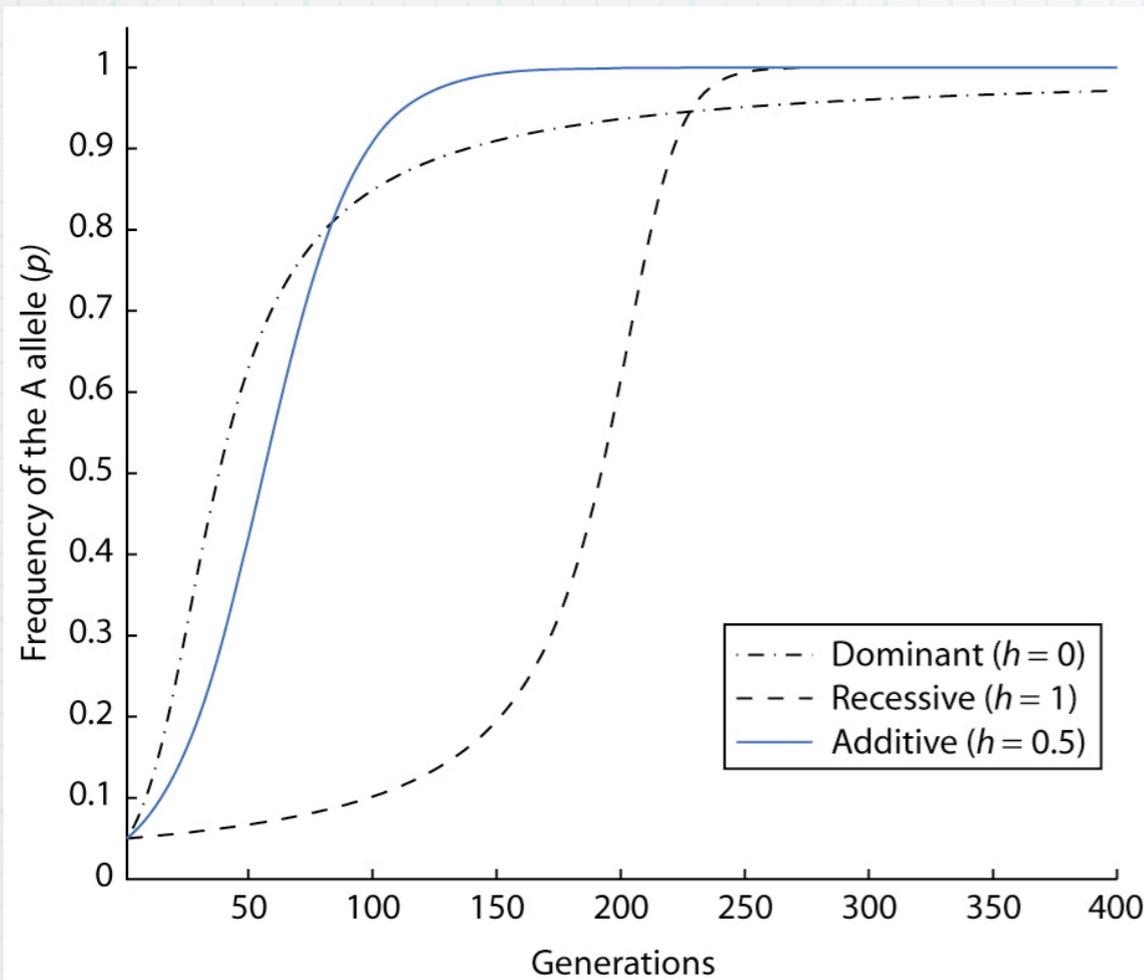
# (2) Natural Selection on a Diallelic Locus

General dominance

- \* When the heterozygote has a phenotype that is the average of the two homozygotes, then  $h=1/2$ , a situation sometimes called **codominance** or **additive gene action**

# (2) Natural Selection on a Diallelic Locus

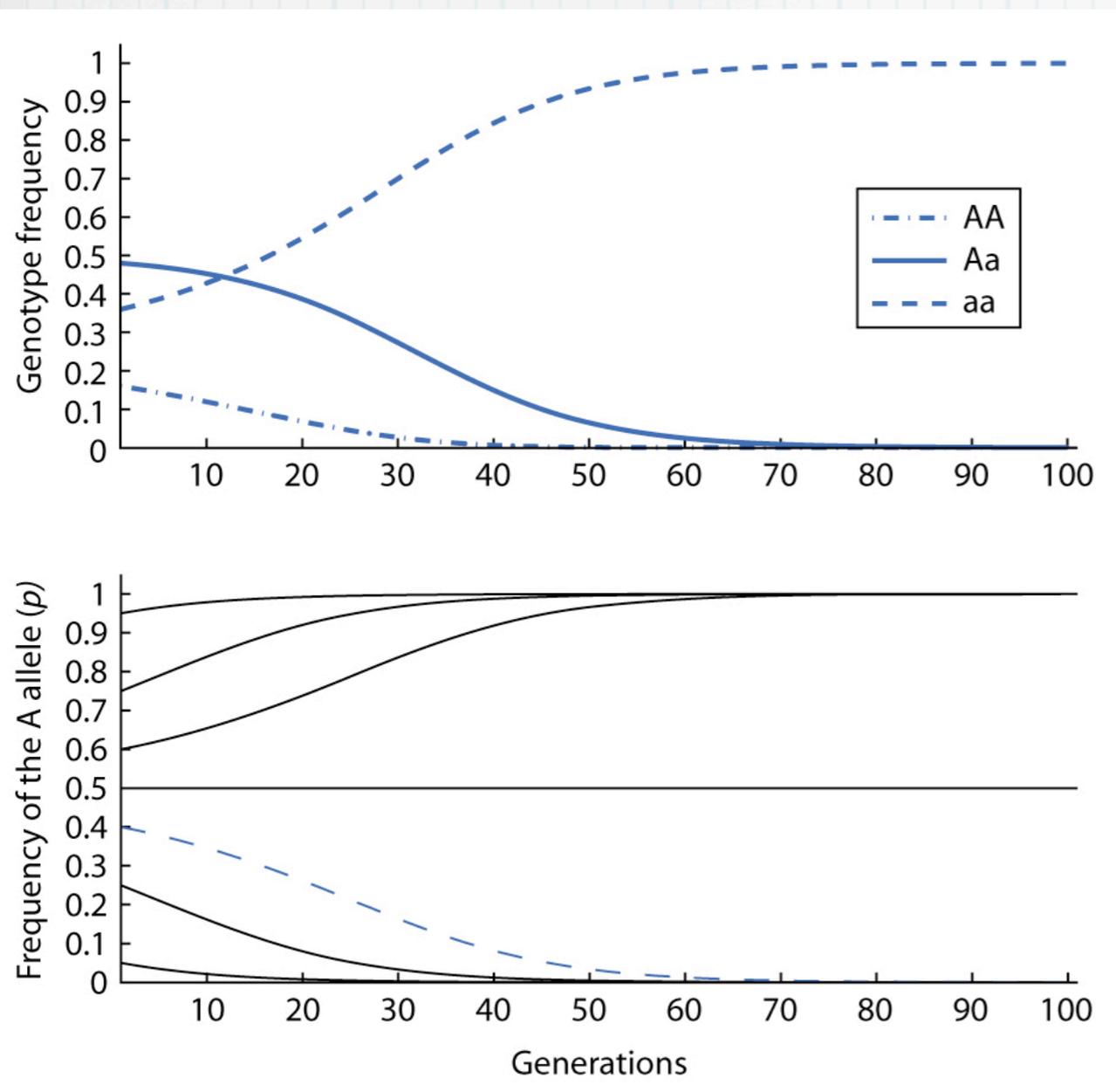
## General dominance



**Figure 6.6** Allele frequencies over time for three types of gene action with a low initial allele frequency. In all three cases the equilibrium allele frequency is fixation or near fixation for the A allele. With complete dominance, natural selection initially increases the allele frequency very rapidly. The approach to fixation for the A allele slows as aa homozygotes become rare since heterozygotes harbor a alleles that are concealed from natural selection by dominance. Natural selection initially changes the frequency of a recessive allele very slowly since homozygote recessive genotypes are very rare. As the recessive homozygotes become more common, allele frequency increases more rapidly. With additive gene action the phenotype of the heterozygote is intermediate between the two homozygotes so all genotypes differ in their viability. Additive gene action has the most rapid overall approach to equilibrium allele frequency. The degree of dominance is represented by the dominance coefficient,  $h$ . In this illustration the selection coefficient is  $s = 0.1$ .

# (2) Natural Selection on a Diallelic Locus

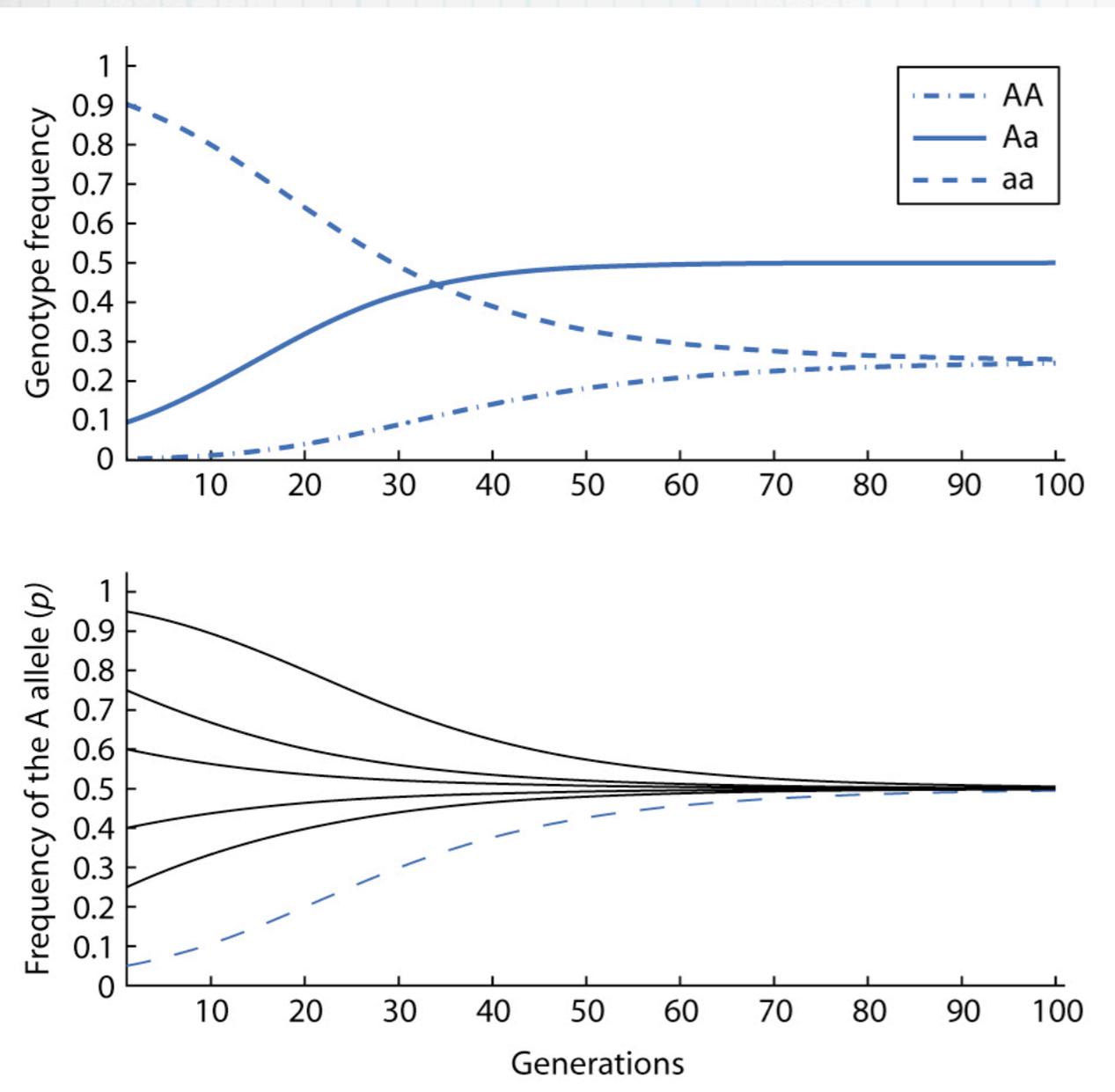
Heterozygote disadvantage  
[Disruptive selection]



**Figure 6.7** The change in the genotype and allele frequency when there is underdominance for fitness and natural selection acts against individuals with Aa genotypes. The equilibrium allele frequency depends on the initial allele frequency. Starting below 0.5 populations head toward loss while starting above 0.5 populations go to fixation. There is an unstable equilibrium at an initial allele frequency of exactly 0.5. From any initial allele frequency the population converges on a minimum frequency of heterozygotes. The colored, dashed line in the bottom panel corresponds to the allele frequencies in the top panel. In this illustration  $w_{AA} = w_{aa} = 1.0$  and  $w_{Aa} = 0.9$ . Genotype frequencies assume random mating.

# (2) Natural Selection on a Diallelic Locus

Heterozygote advantage  
[Balancing selection]



**Figure 6.8** The change in the genotype and allele frequencies when there is overdominance for fitness and natural selection favors individuals with Aa genotypes. From any initial allele frequency the population converges on a maximum frequency of heterozygotes. This corresponds to equal allele frequencies with random mating. The colored, dashed line in the bottom panel corresponds to the allele frequencies in the top panel. In this illustration  $w_{AA} = w_{aa} = 0.9$  and  $w_{Aa} = 1.0$ . Genotype frequencies assume random mating.

# (2) Natural Selection on a Diallelic Locus

## Heterozygote advantage

- \* Selection against a dominant phenotype, recessive phenotype, and a heterozygote, result in fixation or near fixation of one allele and the loss or near loss of the other allele
- \* These three forms of natural selection produce an equilibrium with little or no genetic variation, known as a **monomorphic equilibrium**

# (2) Natural Selection on a Diallelic Locus

## Heterozygote advantage

- \* In contrast, when heterozygotes have the highest fitness, natural selection maintains both alleles in the population at equilibrium, resulting in a **polymorphic equilibrium**
- \* Thus, balancing selection is one type of natural selection that is consistent with the permanent maintenance of genetic variation in populations

# (2) Natural Selection on a Diallelic Locus

## Heterozygote advantage

- \* In the case of balancing selection, and assuming  $s$  and  $t$  as the selection coefficients against the  $AA$  and  $aa$  homozygotes, respectively, the equilibrium allele frequencies are

$$p_{equilibrium} = \frac{t}{s + t}$$

$$q_{equilibrium} = \frac{s}{s + t}$$

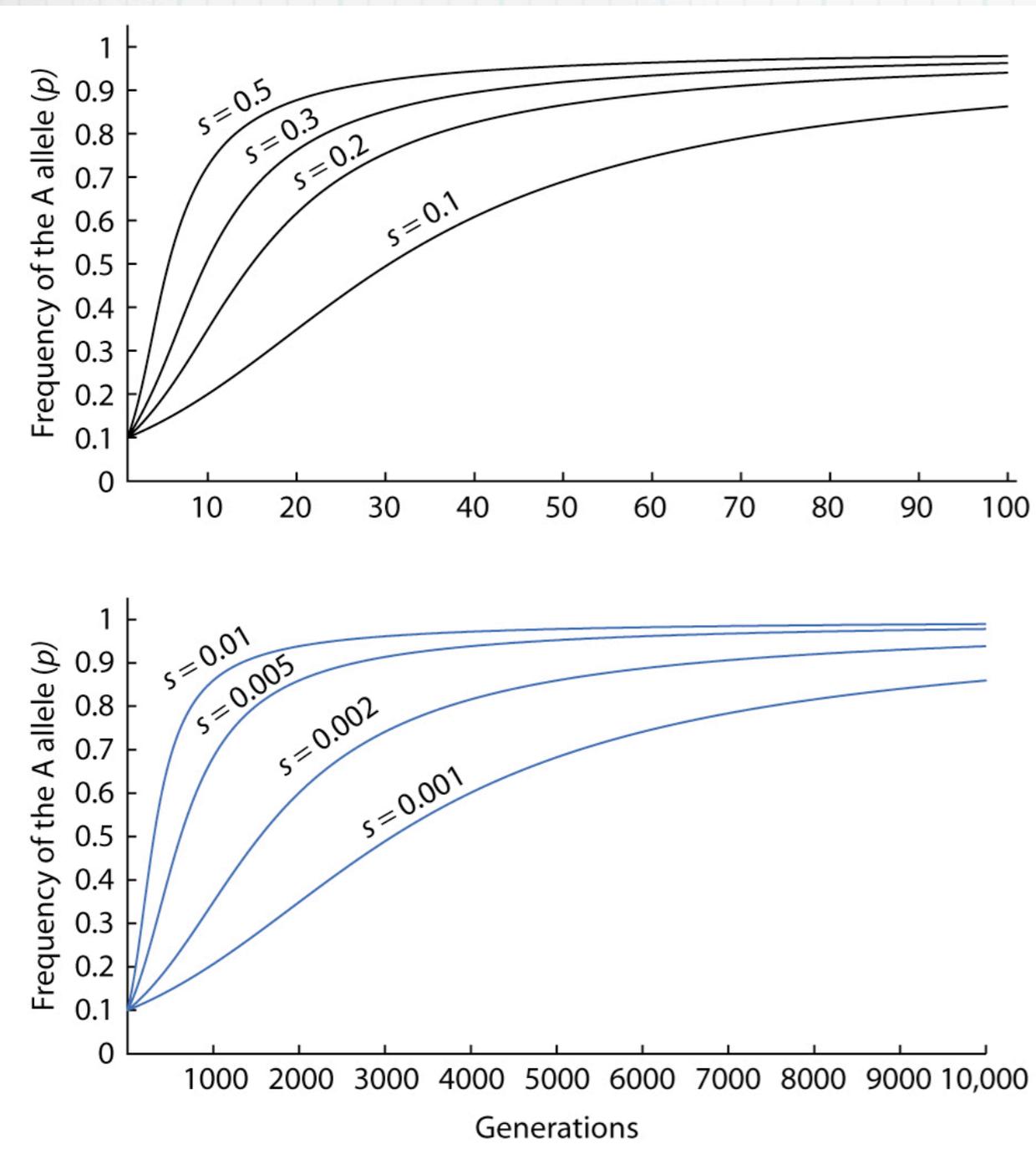
# (2) Natural Selection on a Diallelic Locus

The strength of natural selection

- \* The strength of natural selection against a genotype can vary from weak, such as a viability of 0.1% less than the most fit genotype, to very strong, such as 50% viability or even zero viability (lethality) of a genotype

# (2) Natural Selection on a Diallelic Locus

## The strength of natural selection

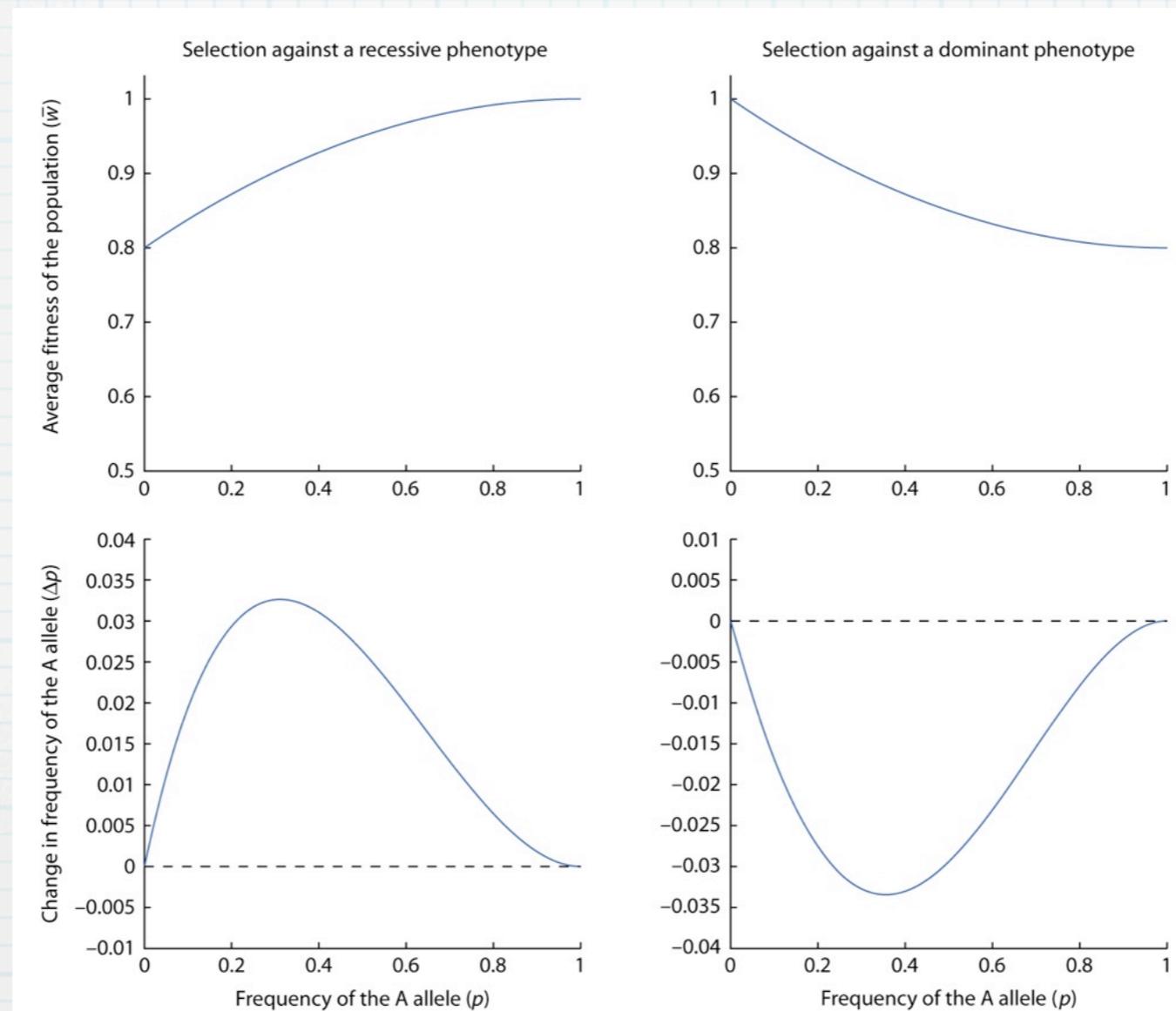


**Figure 6.9** The strength of natural selection influences the rate of change in genotype and allele frequencies. In this illustration, selection acts against the recessive homozygote (aa). The top panel shows strong natural selection where viability of the aa genotype is 10–50% less than that of the other genotypes. The bottom panel shows weak natural selection where viability of the aa genotype is 1–0.1% less than that of the other genotypes. Note the vastly different time scales in the two plots.

# (3) Natural Selection and Fitness

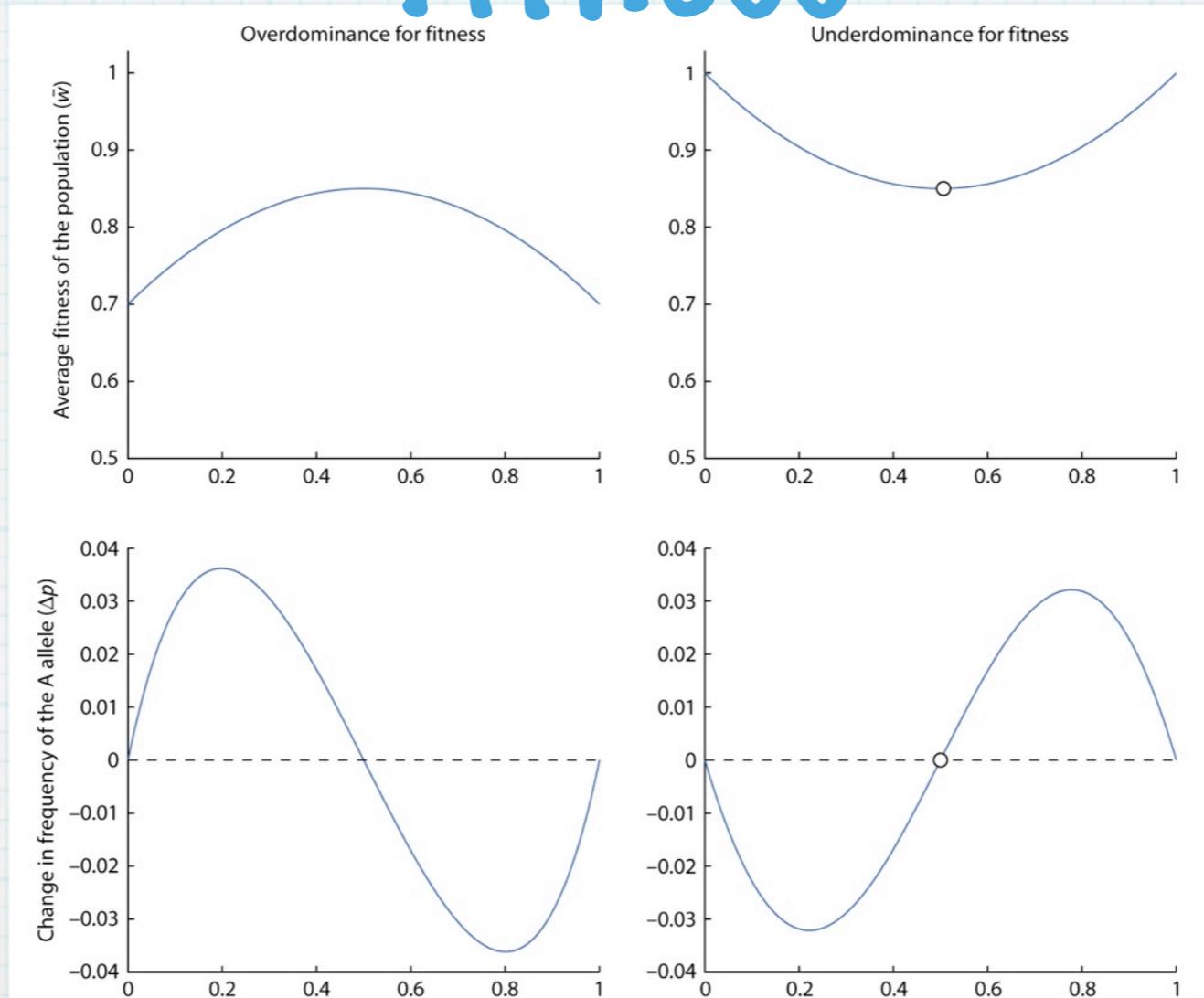
- \* In the five situations shown thus far, the two general outcomes are either fixation/loss or intermediate frequencies of both alleles
- \* This can be understood by examining the average fitness of a population as well as the rate of change in allele frequency over the entire range of allele frequencies

# (3) Natural Selection and Fitness



**Figure 6.10** Mean fitness in a population ( $\bar{w}$ ) and change in allele frequency over a single generation ( $\Delta p$ ) as a function of allele frequency for directional selection. Directional selection reaches allele frequency equilibrium at either fixation or loss, the point of highest mean fitness. Positive values of  $\Delta p$  (above the dashed line) indicate that allele frequency is increasing under selection while negative values of  $\Delta p$  (below the dashed line) indicate that allele frequency is decreasing under selection. The change in allele frequencies is faster when average fitness changes more rapidly (the slope of  $\bar{w}$  is steeper). Here  $w_{AA} = w_{Aa} = 1.0$  and  $w_{aa} = 0.8$  for selection against a recessive phenotype and  $w_{AA} = w_{Aa} = 0.8$  and  $w_{aa} = 1.0$  for selection against a dominant phenotype.

# (3) Natural Selection and Fitness



**Figure 6.11** Mean fitness in a population ( $\bar{w}$ ) and change in allele frequency over a single generation ( $\Delta p$ ) as a function of allele frequency for balancing and disruptive selection. Natural selection changes allele frequencies to increase the average fitness in each generation, eventually reaching an equilibrium when the mean fitness is highest. The change in allele frequencies is faster when average fitness changes more rapidly (the slope of  $\bar{w}$  is steeper). The dashed line in the plots of  $\Delta p$  by  $p$  shows where allele frequencies stop changing ( $\Delta p = 0$ ) and thus are allele frequency equilibrium points. With underdominance for fitness,  $\Delta p$  is zero when  $p = 0.5$  and so defines an equilibrium point marked by the circle. However, this equilibrium point is unstable since  $\Delta p$  on either side of  $p = 0.5$  changes allele frequencies *away* from the equilibrium point (below  $p = 0.5$   $\Delta p$  is negative leading toward loss and above  $p = 0.5$   $\Delta p$  is positive leading toward fixation). In contrast, with overdominance  $\Delta p$  on either side of  $p = 0.5$  changes allele frequencies *toward* the equilibrium point (below  $p = 0.5$   $\Delta p$  is positive and above  $p = 0.5$   $\Delta p$  is negative) and thus  $p = 0.5$  is a stable equilibrium point. Here  $w_{AA} = w_{aa} = 1.0$  and  $w_{Aa} = 0.7$  for underdominance and  $w_{AA} = w_{aa} = 0.7$  and  $w_{Aa} = 1.0$  for overdominance.

# (3) Natural Selection and Fitness

- \* We now turn to Fisher's **fundamental theorem of natural selection** which states that "the rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time"
- \* The theorem has been interpreted as showing that any change in mean fitness caused by natural selection must always be positive
- \* To understand the theorem, we'll examine natural selection and the change in the average fitness of a population over time

# (3) Natural Selection and Fitness

- \* Assume a haploid population with clonal reproduction and discrete generations
- \* If  $k$  is the total number of haplotypes in the population, then the mean fitness is

$$\bar{w} = \sum_{i=1}^k (p_i w_i)$$

# (3) Natural Selection and Fitness

- \* The frequency of any single haplotype (say haplotype  $i$ ) after one generation of natural selection is

$$p'_i = \frac{p_i w_i}{\bar{w}}$$

# (3) Natural Selection and Fitness

- \* The average fitness after one generation of natural selection is then

$$\bar{w}' = \sum_{i=1}^k (p'_i w_i)$$

- \* Substituting the expression for  $p'_i$  gives

$$\bar{w}' = \frac{1}{\bar{w}} \sum_{i=1}^k (p_i w_i^2)$$

# (3) Natural Selection and Fitness

- \* The change in fitness from one generation to the next, standardized by the mean fitness in the initial generation is

$$\Delta \bar{w} = \frac{\bar{w}' - \bar{w}}{\bar{w}}$$

- \* Substituting the expression for  $\bar{w}'$  gives

$$\Delta \bar{w} = \frac{\frac{1}{\bar{w}} \sum_{i=1}^k p_i w_i^2 - \bar{w}}{\bar{w}}$$

# (3) Natural Selection and Fitness

\* After rearranging, we obtain

$$\begin{aligned}\Delta \bar{w} &= \frac{1}{\bar{w}} \left[ \left( \frac{1}{\bar{w}} \sum_{i=1}^k p_i w_i^2 \right) - \bar{w} \right] = \frac{1}{\bar{w}^2} \left[ \left( \sum_{i=1}^k p_i w_i^2 \right) - \bar{w}^2 \right] = \\ &= \frac{1}{\bar{w}^2} [\mathbb{E}(w^2) - \mathbb{E}(w)^2] = \frac{1}{\bar{w}^2} \text{var}(w)\end{aligned}$$

# (3) Natural Selection and Fitness

- \* The relative fitness values of all the haplotypes can be scaled so that  $\bar{w} = 1$
- \* This then leads to

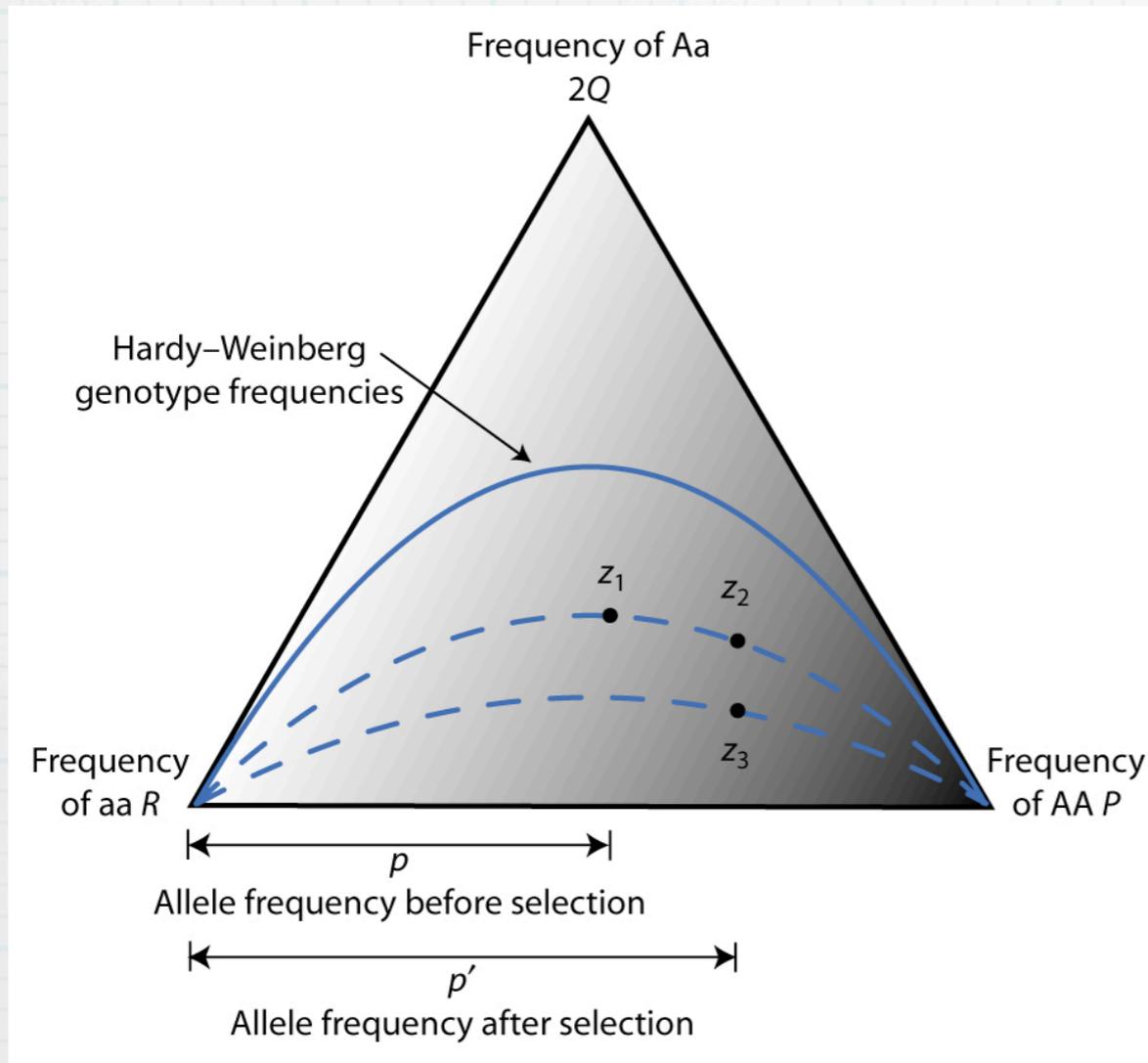
$$\Delta\bar{w} = \text{var}(w)$$

and the conclusion that the change in mean fitness of the population after one generation of natural selection is equal to the variation in fitness

# (3) Natural Selection and Fitness

- \* The variation in fitness is really genetic variation in the case of haploids, due to the frequencies of the different haplotypes in the population as well as to the different fitness values of each haplotype
- \* Therefore, the change in fitness under natural selection is equal to the genetic variation in fitness
- \* Further, since a variance can never be negative, the change in mean fitness by natural selection must then be greater than or equal to zero

# (3) Natural Selection and Fitness



**Figure 6.13** A graphical illustration of R.A. Fisher's fundamental theorem of natural selection. The curved lines represent the product of the homozygote frequencies ( $P = p^2$  and  $R = q^2$ ) as a constant proportion of the square of the product of the allele frequencies ( $Q = pq$ ) or  $\lambda = Q^2/PR$ . Hardy-Weinberg genotype frequencies produced by random mating represent the special case of  $\lambda = 1$  (solid colored line). Mean fitness is represented by the grayscale gradient with darker tones representing higher mean fitness. In this illustration, genotype frequencies start out at  $z_1$ . Suppose that natural selection over one generation changes genotype frequencies to point  $z_3$  (under the conditions that genotype AA has the highest fitness and additive gene action, for example). This change in genotype frequencies can be decomposed into two distinct parts. One part is the change from  $z_1$  to  $z_2$  moving along the curve where  $\lambda$  is constant but allele frequencies change from  $p$  to  $p'$ . The other part is the change in the genotype frequencies (changing the value of  $\lambda$ ) that occurs by moving vertically on the De Finetti diagram from  $z_2$  to  $z_3$  but keeping allele frequencies constant. The fundamental theorem says that the change in the mean fitness by natural selection is proportional to the change in allele frequency alone. Processes other than natural selection, such as mating system, dictate the change in genotype frequencies. When natural selection moves the genotype frequencies along a curve of constant  $\lambda$ , then the total change in mean fitness is completely due to changes in allele frequency and genetic variation in fitness is completely additive. Modified from Edwards (2002).

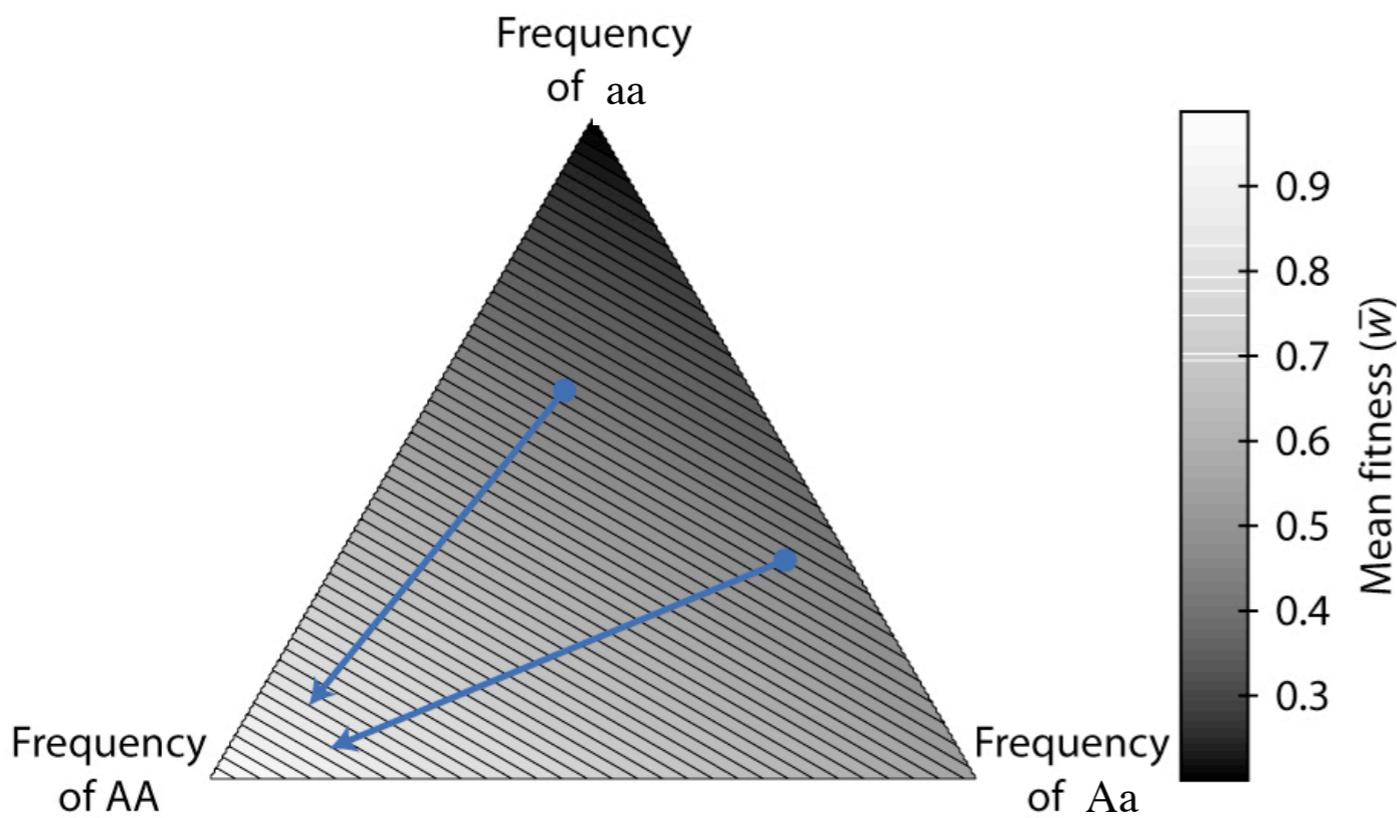
# (4) Viability Selection with Three Alleles or Two Loci

- \* So far, we've looked at natural selection when fitness is equivalent to genotype-specific viability determined by a single locus with two alleles
- \* We'll now consider the outcomes of viability selection when fitness is determined by either a single locus with three alleles or two loci with two alleles

# (4) Viability Selection with Three Alleles or Two Loci

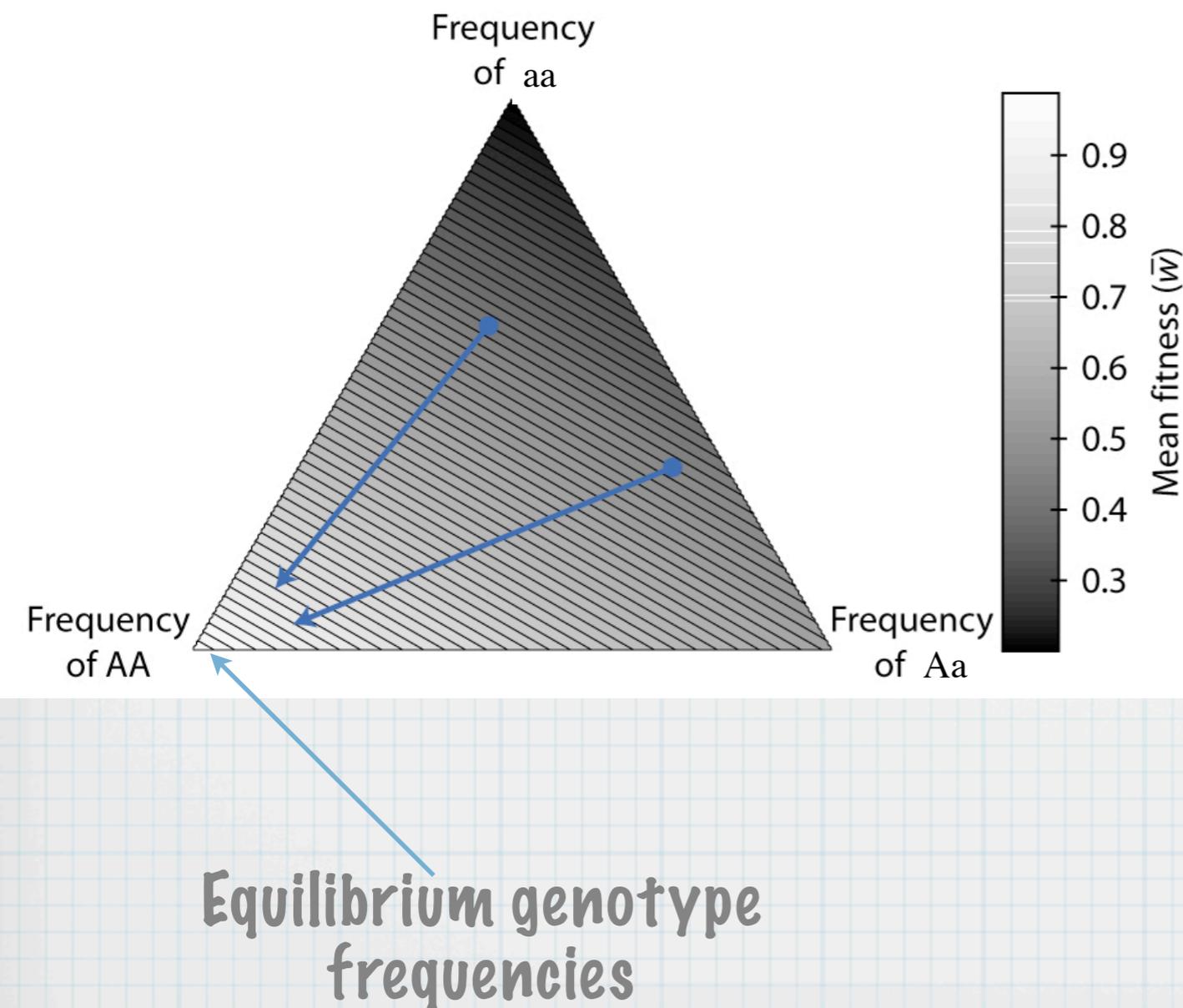
- \* To understand the dynamics of genotype frequencies, allele frequencies, and mean fitness under natural selection, we will use a tool called fitness surface
- \* A fitness surface is a graph that shows genotype frequencies of a population on some axes along with the mean fitness of the population at each possible point in the range of genotype frequencies

# (4) Viability Selection with Three Alleles or Two Loci



**Figure 7.1** A fitness surface made by including mean fitness on a De Finetti plot of the three genotype frequencies for a diallelic locus. The colored lines indicate the possible trajectories of genotype frequencies as natural selection increases the mean fitness of the population. The fitness values are  $w_{AA} = 1.0$ ,  $w_{Aa} = 0.6$ , and  $w_{aa} = 0.2$  so the highest mean fitness is found in the lower left apex when the population is fixed for the AA genotype. This highest fitness point can be reached by continually increasing mean fitness from any initial point on the surface. Gene action is additive because alleles have a constant impact fitness regardless of the allele they are paired with in a genotype. An A allele always contributes 0.5 and an a allele 0.1 toward the fitness of a genotype.

# (4) Viability Selection with Three Alleles or Two Loci



**Figure 7.1** A fitness surface made by including mean fitness on a De Finetti plot of the three genotype frequencies for a diallelic locus. The colored lines indicate the possible trajectories of genotype frequencies as natural selection increases the mean fitness of the population. The fitness values are  $w_{AA} = 1.0$ ,  $w_{Aa} = 0.6$ , and  $w_{aa} = 0.2$  so the highest mean fitness is found in the lower left apex when the population is fixed for the AA genotype. This highest fitness point can be reached by continually increasing mean fitness from any initial point on the surface. Gene action is additive because alleles have a constant impact fitness regardless of the allele they are paired with in a genotype. An A allele always contributes 0.5 and an a allele 0.1 toward the fitness of a genotype.

# (4) Viability Selection with Three Alleles or Two Loci

One locus with three alleles

- \* Let's consider the example of natural selection on three alleles at the human hemoglobin  $\beta$  gene that encodes the  $\beta$  protein, which is referred to as Hb
- \* Although several hundred Hb alleles have been identified in the human population, three alleles are of particular interest: A, S, and C

# (4) Viability Selection with Three Alleles or Two Loci

One locus with three alleles

- \* Allele A is the most common
- \* Individuals homozygous for the S allele are at higher susceptibility to chronic anemia
- \* Individuals homozygous for the C allele have mild to moderate anemia

# (4) Viability Selection with Three Alleles or Two Loci

One locus with three alleles

- \* The fitness of the Hb genotypes depends on the environment where people live
- \* In areas without the malaria parasite *P. falciparum*, genotypes that result in anemia have lower fitness
- \* In regions where malarial infection is common, certain Hb genotypes confer resistance to infection by *P. falciparum* that may partly or completely compensate for any disadvantage due to anemia

# (4) Viability Selection with Three Alleles or Two Loci

One locus with three alleles

**Table 7.1** Relative fitness estimates for the six genotypes of the hemoglobin  $\beta$  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.

Genotype . . .	Relative fitness ( $w$ )					
	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

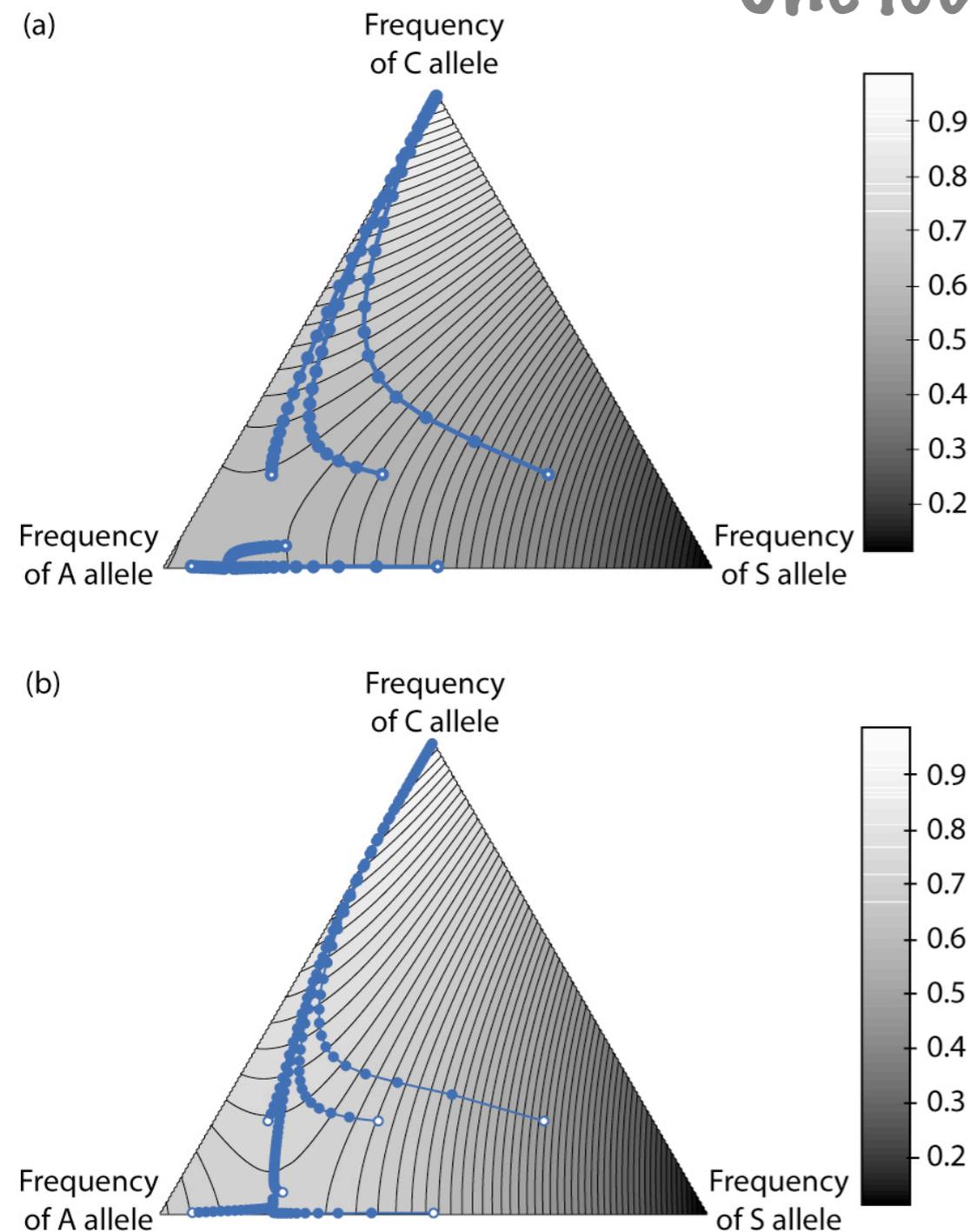
# (4) Viability Selection with Three Alleles or Two Loci

One locus with three alleles

- \* A prediction from the table is that natural selection in populations where malaria is common would increase the frequency of the CC genotype and eventually fix the C allele
- \* Is this really what will happen?

# (4) Viability Selection with Three Alleles or Two Loci

## One locus with three alleles



**Figure 7.2** Fitness surfaces for the A, S, and C alleles at the human hemoglobin  $\beta$  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.

# (4) Viability Selection with Three Alleles or Two Loci

One locus with three alleles

- \* Understanding how genotype frequencies will change on a fitness surface requires calculating the change in allele frequencies due to selection for a series of points on the surface
- \* Assuming  $p$ ,  $q$ , and  $r$  are the frequencies of alleles A, B, and C, respectively, in the population, the mean fitness of the population is:

$$\Delta\bar{w} = w_{AA}p^2 + w_{BB}q^2 + w_{CC}r^2 + w_{AB}2pq + w_{AC}2pr + w_{BC}2qr$$

# (4) Viability Selection with Three Alleles or Two Loci

One locus with three alleles

- \* The marginal fitness for the A allele is:

$$\bar{w}_A = \frac{w_{AA}p^2 + w_{AB}pq + w_{AC}pr}{p} = w_{AA}p + w_{AB}q + w_{AC}r$$

- \* Allele frequencies change each generation due to differences between the marginal fitness of each allele and the average fitness of the entire population
- \* The change in the frequency of the A allele is

$$\Delta p = p \frac{\bar{w}_A - \bar{w}}{\bar{w}}$$

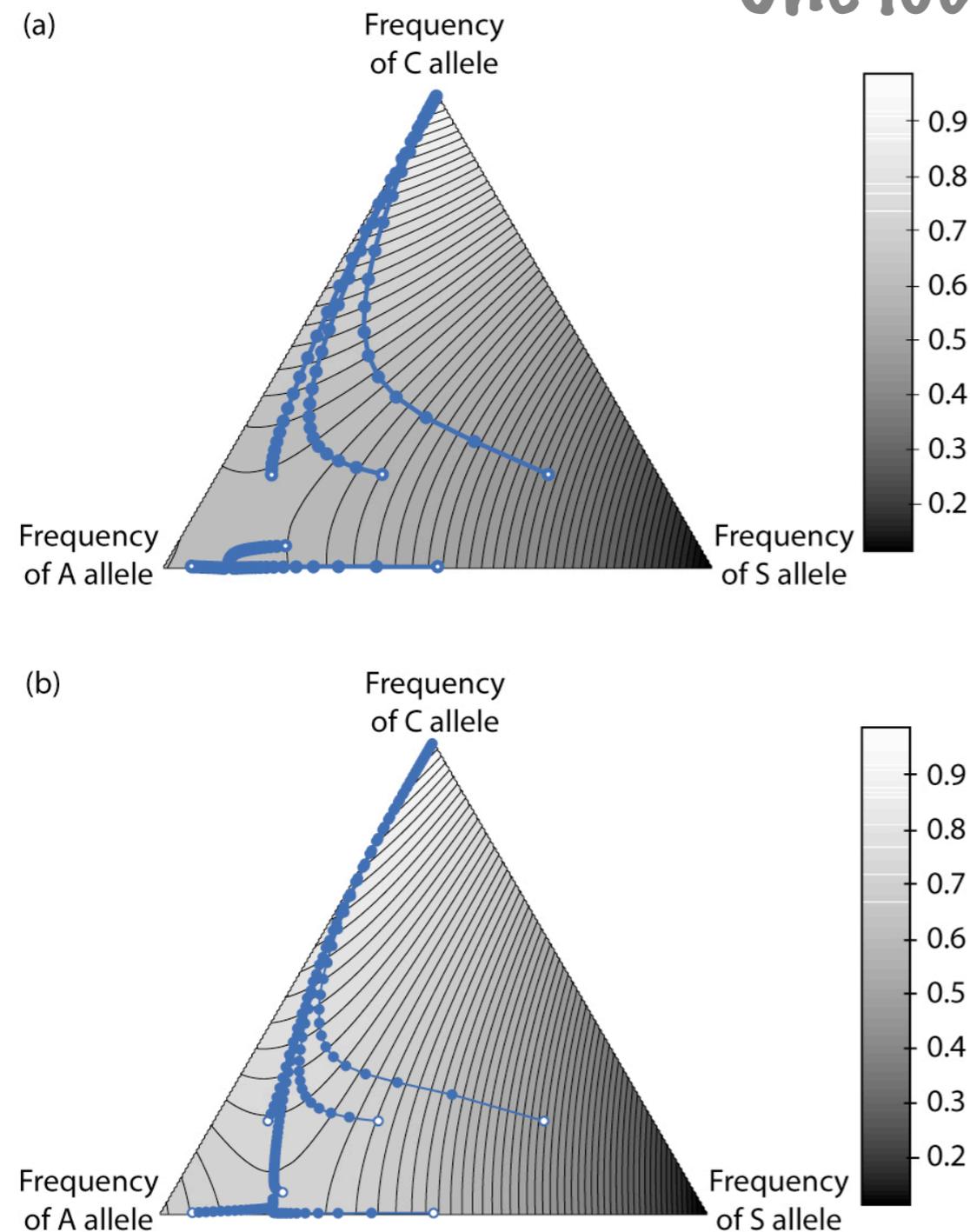
# (4) Viability Selection with Three Alleles or Two Loci

One locus with three alleles

- \* Allele frequency after one generation of selection is then simply  $p_{t+1} = p + \Delta p$
- \* Similar expressions can be obtained for the B and C alleles

# (4) Viability Selection with Three Alleles or Two Loci

## One locus with three alleles



**Figure 7.2** Fitness surfaces for the A, S, and C alleles at the human hemoglobin  $\beta$  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.

# (4) Viability Selection with Three Alleles or Two Loci

Two diallelic loci

- \* Natural selection on two loci is inherently more complicated than at a single locus because of gametic (linkage) disequilibrium
- \* The outcome of two-locus selection depends on the balance between natural selection and recombination between loci, as well as the initial genotype frequencies in the population

# (4) Viability Selection with Three Alleles or Two Loci

Two diallelic loci

- \* With two diallelic loci, there are 16 possible genotypes that result from the union of four possible gametes

# (4) Viability Selection with Three Alleles or Two Loci

Two diallelic loci

	AB	Ab	aB	ab
AB	$w_{11}$	$w_{12}$	$w_{13}$	$w_{14}$
Ab	$w_{21}$	$w_{22}$	$w_{23}$	$w_{24}$
aB	$w_{31}$	$w_{32}$	$w_{33}$	$w_{34}$
ab	$w_{41}$	$w_{42}$	$w_{43}$	$w_{44}$

	BB	Bb	bb
AA	$w_{11}$	$w_{12}$	$w_{22}$
Aa	$w_{13}$	$w_H$	$w_{24}$
aa	$w_{33}$	$w_{34}$	$w_{44}$

**Table 7.2** Matrix of fitness values for all combinations of the four gametes formed at two diallelic loci (top). If the same gamete inherited from either parent has the same fitness in a progeny genotype (e.g.  $w_{12} = w_{21}$ ), then there are 10 gamete fitness values shown outside the shaded triangle. These 10 fitness values can be summarized by a genotype fitness matrix (bottom) under the assumption that double heterozygotes have equal fitness ( $w_{14} = w_{23}$ ) and representing their fitness value by  $w_H$ . The double heterozygote genotypes are of special interest since they can produce recombinant gametes.

# (4) Viability Selection with Three Alleles or Two Loci

## Two diallelic loci

**Table 7.3** Expected frequencies of gametes under viability selection for two diallelic loci in a randomly mating population with a recombination rate of  $r$  between the loci. The expected gamete frequencies assume that the same gamete coming from either parent will have the same fitness in a progeny genotype (e.g.  $w_{12} = w_{21}$ ). Eight genotypes have non-recombinant and recombinant gametes that are identical and so do not require a term for the recombination rate. Two genotypes produce novel recombinant gametes, requiring inclusion of the recombination rate to predict gamete frequencies. Summing down each column of the table gives the total frequency of each gamete in the next generation due to mating and recombination.

Genotype	Fitness	Total frequency	Frequency of gametes in next generation			
			AB	Ab	aB	ab
AB/AB	$w_{11}$	$x_1^2$	$x_1^2$			
AB/Ab	$w_{12}$	$2x_1x_2$	$x_1x_2$	$x_1x_2$		
AB/aB	$w_{13}$	$2x_1x_3$	$x_1x_3$		$x_1x_3$	
AB/ab	$w_{14}$	$2x_1x_4$	$(1-r)x_1x_4$	$(r)x_1x_4$	$(r)x_1x_4$	$(1-r)x_1x_4$
Ab/Ab	$w_{22}$	$x_2^2$		$x_2^2$		
Ab/aB	$w_{23}$	$2x_2x_3$	$(r)x_2x_3$	$(1-r)x_2x_3$	$(1-r)x_2x_3$	$(r)x_2x_3$
Ab/ab	$w_{24}$	$2x_2x_4$		$x_2x_4$	$x_2x_4$	
aB/aB	$w_{33}$	$x_3^2$			$x_3^2$	
aB/ab	$w_{34}$	$2x_3x_4$			$x_3x_4$	$x_3x_4$
ab/ab	$w_{44}$	$x_4^2$				$x_4^2$

$x_1, x_2, x_3,$  and  $x_4$  are the frequencies of the gametes AB, Ab, aB, and ab, respectively

# (4) Viability Selection with Three Alleles or Two Loci

Two diallelic loci

- \* Using Table 7.3, we can obtain the frequency of each gamete one generation later

# (4) Viability Selection with Three Alleles or Two Loci

Two diallelic loci

$$x_{1(t+1)} = \frac{w_{11}x_1^2 + w_{12}x_1x_2 + w_{13}x_1x_3 + (1-r)w_{14}x_1x_4 + rw_{23}x_2x_3}{\bar{w}}$$



$$x_{1(t+1)} = \frac{x_1(w_{11}x_1 + w_{12}x_2 + w_{13}x_3 + w_{14}x_4) - r(w_{14}x_1x_4 - w_{23}x_2x_3)}{\bar{w}}$$



[Assuming the fitness of genotypes with the same number of A and B alleles is equal]

$$x_{1(t+1)} = \frac{x_1(w_{11}x_1 + w_{12}x_2 + w_{13}x_3 + w_{14}x_4) - rw_{14}(x_1x_4 - x_2x_3)}{\bar{w}}$$

$$[D = x_1x_4 - x_2x_3]$$



$$x_{1(t+1)} = \frac{x_1(w_{11}x_1 + w_{12}x_2 + w_{13}x_3 + w_{14}x_4) - rw_{14}D}{\bar{w}}$$

# (4) Viability Selection with Three Alleles or Two Loci

Two diallelic loci

- \* The frequency of the AB gametes after one generation of natural selection is a function of:
  - \* the viabilities of the genotypes that produce AB ( $w_{11}, w_{12}, w_{13}, w_{14}$ )
  - \* the combination of recombination, fitness values of the double heterozygotes, and initial gametic disequilibrium in the population ( $r w_{14} D$ )
  - \* the frequency of recombination and the relative fitness of the genotypes

$$x_{1(t+1)} = \frac{x_1(w_{11}x_1 + w_{12}x_2 + w_{13}x_3 + w_{14}x_4) - r w_{14} D}{\bar{w}}$$

# (4) Viability Selection with Three Alleles or Two Loci

Two diallelic loci

- \* Using the idea that the four gamete frequencies can be treated like the frequencies of four alleles at one locus, we can develop expressions for change in gamete frequencies

# (4) Viability Selection with Three Alleles or Two Loci

Two diallelic loci

- \* The marginal fitness for each of the two-locus gametes is

$$\bar{w}_i = \sum_{j=1}^4 x_j w_{ij}$$

- \* The average fitness of the population is

$$\bar{w} = \sum_{i=1}^4 \sum_{j=1}^4 x_i x_j w_{ij}$$

# (4) Viability Selection with Three Alleles or Two Loci

Two diallelic loci

$$x_{1(t+1)} = \frac{x_1(w_{11}x_1 + w_{12}x_2 + w_{13}x_3 + w_{14}x_4) - rw_{14}D}{\bar{w}}$$

# (4) Viability Selection with Three Alleles or Two Loci

Two diallelic loci

$$x_1(t+1) = \frac{x_1(w_{11}x_1 + w_{12}x_2 + w_{13}x_3 + w_{14}x_4) - rw_{14}D}{\bar{w}}$$

# (4) Viability Selection with Three Alleles or Two Loci

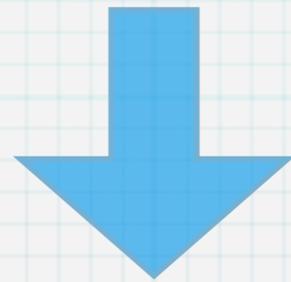
Two diallelic loci

$$x_1(t+1) = \frac{x_1(w_{11}x_1 + w_{12}x_2 + w_{13}x_3 + w_{14}x_4) - r w_{14} D}{\bar{w}}$$

# (4) Viability Selection with Three Alleles or Two Loci

Two diallelic loci

$$x_1(t+1) = \frac{x_1(\overline{w}_1 - r w_H D)}{\overline{w}}$$



$$\Delta x_1 = \frac{x_1 \overline{w}_1 - r w_H D}{\overline{w}}$$

# (4) Viability Selection with Three Alleles or Two Loci

Two diallelic loci

\* For all four gametes:

$$\Delta x_1 = \frac{x_1 \bar{w}_1 - r w_H D}{\bar{w}} \quad (7.1.1)$$

$$\Delta x_2 = \frac{x_2 \bar{w}_2 - r w_H D}{\bar{w}} \quad (7.1.2)$$

$$\Delta x_3 = \frac{x_3 \bar{w}_3 - r w_H D}{\bar{w}} \quad (7.1.3)$$

$$\Delta x_4 = \frac{x_4 \bar{w}_4 - r w_H D}{\bar{w}} \quad (7.1.4)$$

# (4) Viability Selection with Three Alleles or Two Loci

Two diallelic loci

- \* The equations show that the change in gamete frequency under natural selection is due to both fitness values and recombination
- \* If there is no recombination ( $r=0$ ), then each gamete is analogous to a single allele
- \* The outcome of selection is then like four alleles at a single locus

# (4) Viability Selection with Three Alleles or Two Loci

Two diallelic loci

- \* The process of recombination may either reinforce or oppose the changes in gamete frequencies due to natural selection
- \* For example, if gametes  $Ab$  and  $aB$  have the highest fitness values and there is no recombination, then  $\Delta x_2$  and  $\Delta x_3$  would be positive while  $\Delta x_1$  and  $\Delta x_4$  would be negative

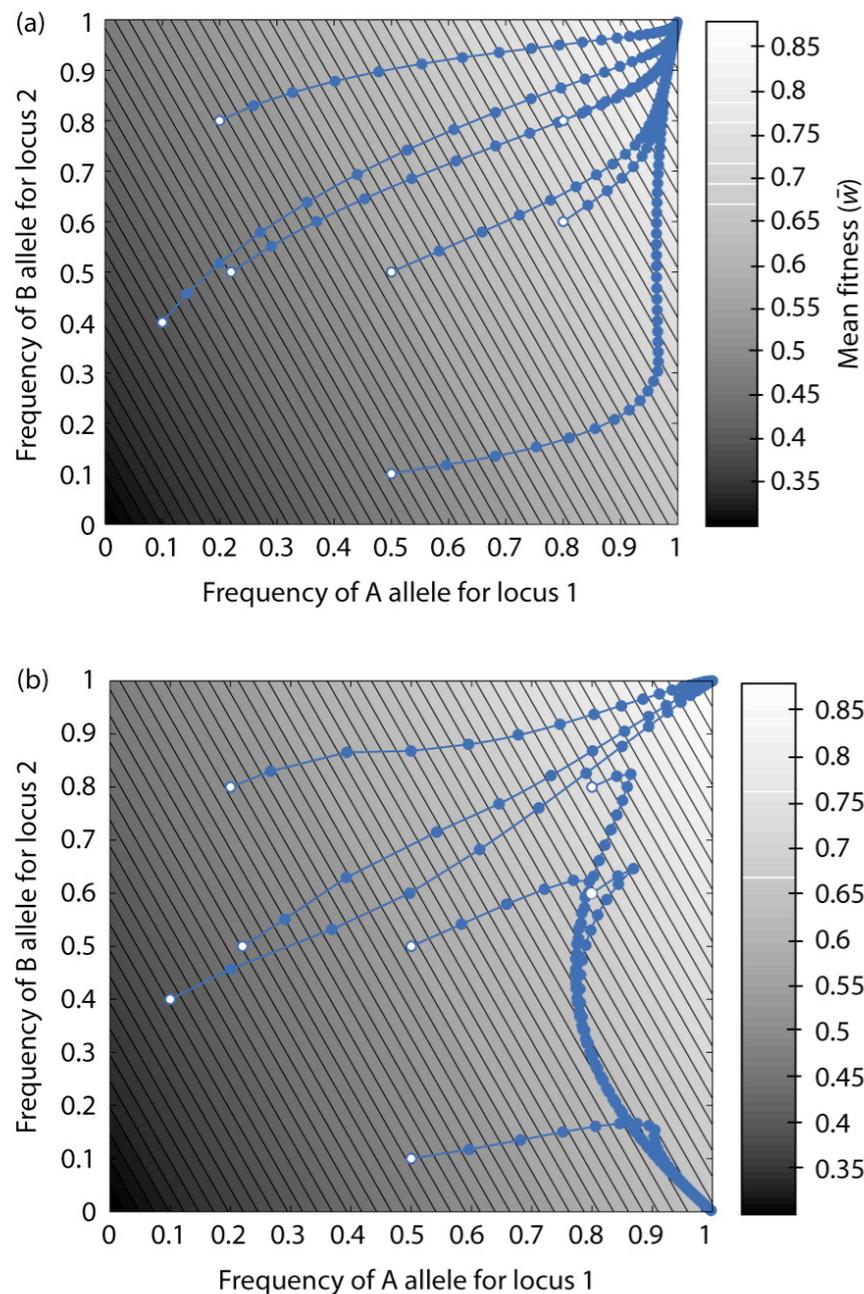
# (4) Viability Selection with Three Alleles or Two Loci

Two diallelic loci

- \* The gamete frequency changes caused by recombination would amplify the effect of natural selection on gamete frequencies since the  $rw_{HD}$  term would increase  $\Delta x_2$  and  $\Delta x_3$  but decrease  $\Delta x_1$  and  $\Delta x_4$
- \* In contrast, if gametes  $AB$  and  $ab$  have the highest fitness and there is recombination, then the  $rw_{HD}$  term would decrease  $\Delta x_1$  and  $\Delta x_4$  but increase  $\Delta x_2$  and  $\Delta x_3$  in opposition to natural selection

# (4) Viability Selection with Three Alleles or Two Loci

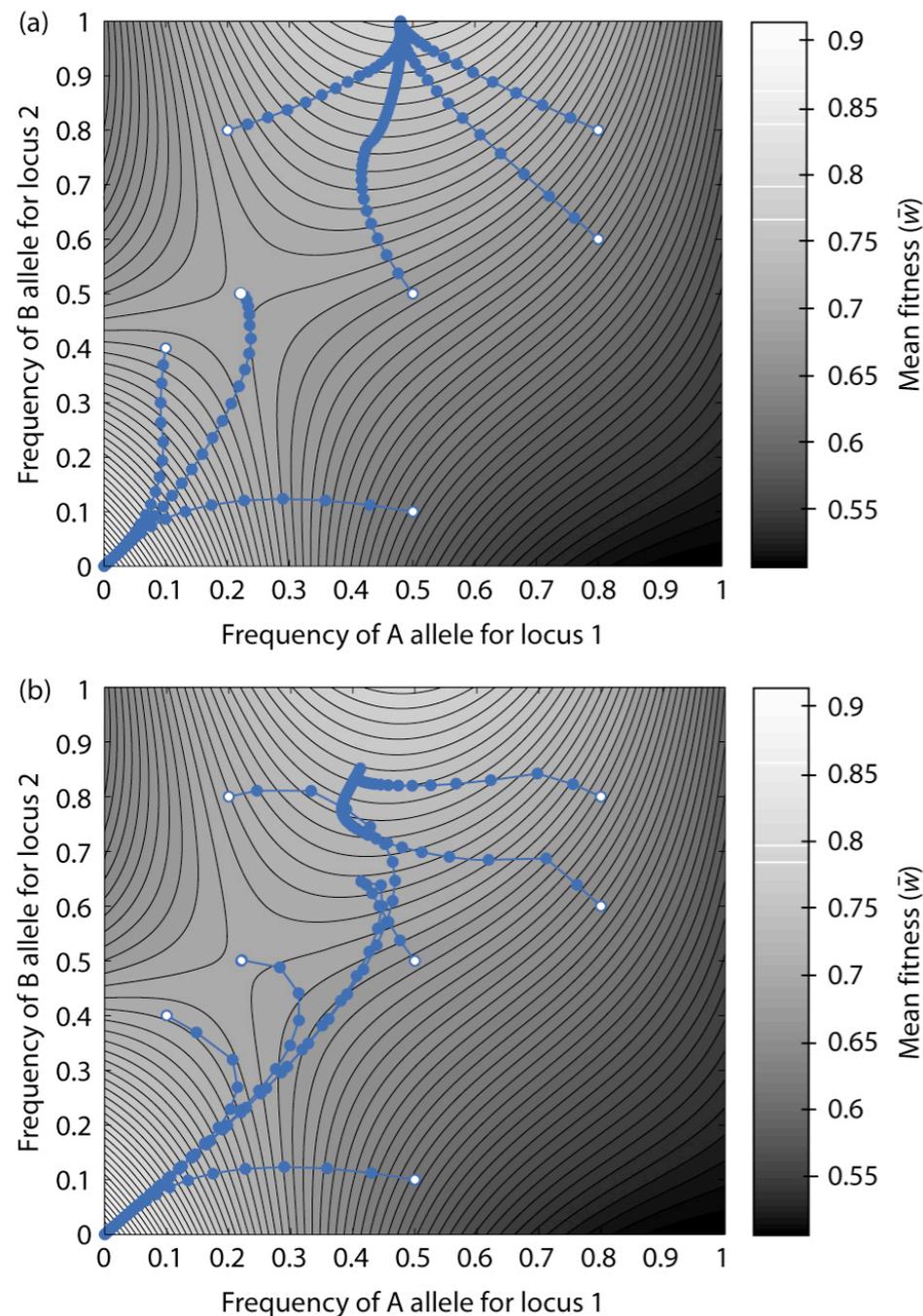
## Two diallelic loci



**Figure 7.3** A fitness surface for two loci that each have two alleles where gene action is additive. The blue dots show generation-by-generation allele frequencies based on equations 7.11–7.14 for seven different initial sets of four gamete frequencies. When recombination is a weak force ( $r = 0.05$ ), equilibrium allele frequencies are dictated by natural selection and all initial gamete frequencies eventually reach the highest mean fitness point (a). In contrast, when recombination is a strong force ( $r = 0.5$ ) then equilibrium allele frequencies depend on initial gamete frequencies (b). When recombination is strong, equilibrium allele frequencies may not correspond to the highest mean fitness. Relative fitness values are  $w_{AABB} = 0.9$ ,  $w_{AABb} = 0.8$ ,  $w_{AAbb} = 0.7$ ,  $w_{AaBB} = 0.7$ ,  $w_{AaBb} = 0.6$ ,  $w_{Aabb} = 0.5$ ,  $w_{aaBB} = 0.5$ ,  $w_{aaBb} = 0.4$ , and  $w_{aabb} = 0.3$ . The seven initial allele frequency points, shown as open circles, are identical for the two surfaces.

# (4) Viability Selection with Three Alleles or Two Loci

## Two diallelic loci



**Figure 7.4** A fitness surface for two loci that each have two alleles where gene action exhibits epistasis. When recombination is a weak force ( $r = 0.05$ ), equilibrium allele frequencies are dictated by natural selection. Equilibrium allele frequencies depend on initial gamete frequencies since the two highest mean fitness points are separated by a fitness valley (a). When recombination is strong ( $r = 0.5$ ), allele frequencies change such that mean fitness actually decreases for a time before increasing again to eventually reach the lower of the two mean fitness peaks (b). The two initial gamete frequencies in the upper right of the surface reach an equilibrium point where fitness is not maximized and there is gametic disequilibrium ( $D = 0.041$ ). Relative fitness values are  $w_{AABB} = 0.61$ ,  $w_{AABb} = 0.58$ ,  $w_{AAbb} = 0.50$ ,  $w_{AaBB} = 1.0$ ,  $w_{AaBb} = 0.77$ ,  $w_{Aabb} = 0.50$ ,  $w_{aaBB} = 0.64$ ,  $w_{aaBb} = 0.62$ , and  $w_{aabb} = 0.92$ . The seven initial allele frequency points, shown as open circles, are identical for the two surfaces.

# (4) Viability Selection with Three Alleles or Two Loci

Two diallelic loci

- \* We observe that:
  - \* When one process is much stronger, it will win out over the other process to determine the equilibrium
  - \* When both processes are of approximately equal strength, the result is a compromise that may produce an equilibrium that is neither gametic equilibrium nor maximum mean fitness

# (4) Viability Selection with Three Alleles or Two Loci

Two diallelic loci

- \* Although there is no general set of equilibrium gamete frequencies for two-locus selection with an arbitrary set of fitness values, many special cases have been examined that have produced some general conclusions

# (4) Viability Selection with Three Alleles or Two Loci

Two diallelic loci

- \* By itself, low frequencies of recombination (small  $r$ ) make it more likely that selection will result in LD at equilibrium gamete frequencies even with random mating
- \* The combination of non-additive gene action and infrequent recombination also make LD at equilibrium gamete frequencies more likely

# (4) Viability Selection with Three Alleles or Two Loci

Two diallelic loci

- \* High rates of self-fertilization adds an additional force on gamete frequencies that can either compliment or act in opposition to selection and recombination
- \* Since mean fitness may decrease with selection and recombination, Fisher's fundamental theorem does not hold for two-locus selection

# (4) Viability Selection with Three Alleles or Two Loci

Two diallelic loci

- \* A critical conclusion from examining two-locus natural selection is that generalizing from the results of one-locus selection models to multiple loci may be biologically misleading except in limiting cases such as when there is very little recombination and there is no epistasis

# (5) Alternative Models of Natural Selection

\* So far, we have

1. **Equated fitness with the viability of genotypes:** However, we have seen that there are numerous points in the reproductive life cycle where genotypes may have differential success of performance
2. **Assumed fitness as constant viability in an infinitely growing population:** However, it may be that fitness actually changes in response to the conditions found in different populations or in response to the changes in genotype frequency brought on by natural selection

# (5) Alternative Models of Natural Selection

- \* We now extend the basic viability model of natural selection in a variety of ways to predict how natural selection works for different components of fitness and for changing fitness values
- \* Natural selection via different levels of fecundity
- \* Natural selection with frequency-dependent fitness
- \* Natural selection with density-dependent fitness

# (5) Alternative Models of Natural Selection

Natural selection via different levels of fecundity

- \* Natural selection due to differences in genotype viability is sometimes called **hard selection** since genotype frequency changes come about from the death of individuals and their complete failure to reproduce
- \* In contrast, fecundity selection is called **soft selection** because all individuals in the parental generation reproduce, although by differing amounts

# (5) Alternative Models of Natural Selection

Natural selection via different levels of fecundity

- \* A fecundity model of natural selection on a diallelic locus requires a different approach than was taken for viability selection
- \* A major difference is that fitness depends on the pair of genotypes that mate
- \* This means that there are nine different fitness values in a fecundity selection model

# (5) Alternative Models of Natural Selection

Natural selection via different levels of fecundity

- \* Another difference is that predicting the genotype frequencies of the progeny is more complicated than for simple random mating
- \* Variation in fecundity may alter the number of progeny produced by each mating pair from the frequency expected by random mating alone
- \* This requires accounting for the expected progeny genotype frequencies that arise from each mating pair weighted by the fecundity of that mating pair

# (5) Alternative Models of Natural Selection

## Natural selection via different levels of fecundity

**Table 7.4** Fitness values based on the fecundities of mating pairs of male and female genotypes for a diallelic locus along with the expected genotype frequencies in the progeny of each possible male and female mating pair weighted by the fecundity of each mating pair. The frequencies of the AA, Aa, and aa genotypes are represented by X, Y, and Z respectively.

Male genotype	Female genotype . . .	Fitness value		
		AA	Aa	aa
AA		$f_{11}$	$f_{12}$	$f_{13}$
Aa		$f_{21}$	$f_{23}$	$f_{23}$
aa		$f_{31}$	$f_{32}$	$f_{33}$

Parental mating	Fecundity	Total frequency	Expected progeny genotype frequency		
			AA	Aa	aa
AA × AA	$f_{11}$	$X^2$	$X^2$	0	0
AA × Aa	$f_{12}$	$XY$	$1/2XY$	$1/2XY$	0
AA × aa	$f_{13}$	$XZ$	0	$XZ$	0
Aa × AA	$f_{21}$	$YX$	$1/2YX$	$1/2YX$	0
Aa × Aa	$f_{22}$	$Y^2$	$Y^2/4$	$(2Y^2)/4$	$Y^2/4$
Aa × aa	$f_{23}$	$YZ$	0	$1/2YZ$	$1/2YZ$
aa × AA	$f_{31}$	$ZX$	0	$ZX$	0
aa × Aa	$f_{32}$	$ZY$	0	$1/2ZY$	$1/2ZY$
aa × aa	$f_{33}$	$Z^2$	0	0	$Z^2$

# (5) Alternative Models of Natural Selection

Natural selection via different levels of fecundity

- \* The average fecundities of the AA, Aa, and aa genotypes, respectively, are

$$\bar{f}X_{t+1} = f_{11}X^2 + (f_{12} + f_{21})\frac{1}{2}XY + f_{22}\frac{Y^2}{4}$$

$$\bar{f}Y_{t+1} = (f_{12} + f_{21})\frac{1}{2}XY + (f_{13} + f_{31})XZ + f_{22}\frac{Y^2}{2} + (f_{23} + f_{32})\frac{1}{2}YZ$$

$$\bar{f}Z_{t+1} = f_{33}Z^2 + (f_{32} + f_{23})\frac{1}{2}YZ + f_{22}\frac{Y^2}{4}$$

- \* The total average fecundity is

$$\bar{f} = \bar{f}X_{t+1} + \bar{f}Y_{t+1} + \bar{f}Z_{t+1}$$

# (5) Alternative Models of Natural Selection

Natural selection via different levels of fecundity

- \* Since random mating does not occur by definition when there is fecundity selection, general equilibrium points cannot be found for arbitrary sets of nine fecundity values
- \* Rather, the change in genotype frequencies must be understood by considering special cases of fecundity values (see your textbook for discussion of three special cases)

# (5) Alternative Models of Natural Selection

Natural selection via different levels of fecundity

- \* It has been shown that mean fecundity does not necessarily increase with fecundity selection
- \* This means that the mean fecundity is not necessarily maximized at equilibrium genotype frequencies for fecundity selection, in contrast to the way natural selection maximizes mean fitness in the viability model for one locus with two alleles

# (5) Alternative Models of Natural Selection

Natural selection with frequency-dependent fitness

- \* In this part, we will assume that the fitness of a genotype depends on its frequency in a population
- \* The key concept in frequency-dependent selection models is creating a measure of fitness that changes

# (5) Alternative Models of Natural Selection

Natural selection with frequency-dependent fitness

- \* Suppose that the fitness of a genotype decreases as that genotype becomes more common in the population, called **negative frequency dependence**
- \* Assuming  $s_{xx}$  represents the genotype-specific selection coefficient, the relative fitness values are

$$w_{AA} = 1 - s_{AA}p^2$$

$$w_{Aa} = 1 - s_{Aa}2pq$$

$$w_{aa} = 1 - s_{aa}q^2$$

# (5) Alternative Models of Natural Selection

Natural selection with frequency-dependent fitness

- \* Genotypes have higher fitness when they are rare since relative fitness decreases as the product of the selection coefficient and the genotype's frequency increases
- \* Note that the selection coefficient itself is a constant and can be thought of as a per-capita decrease in relative fitness

# (5) Alternative Models of Natural Selection

Natural selection with frequency-dependent fitness

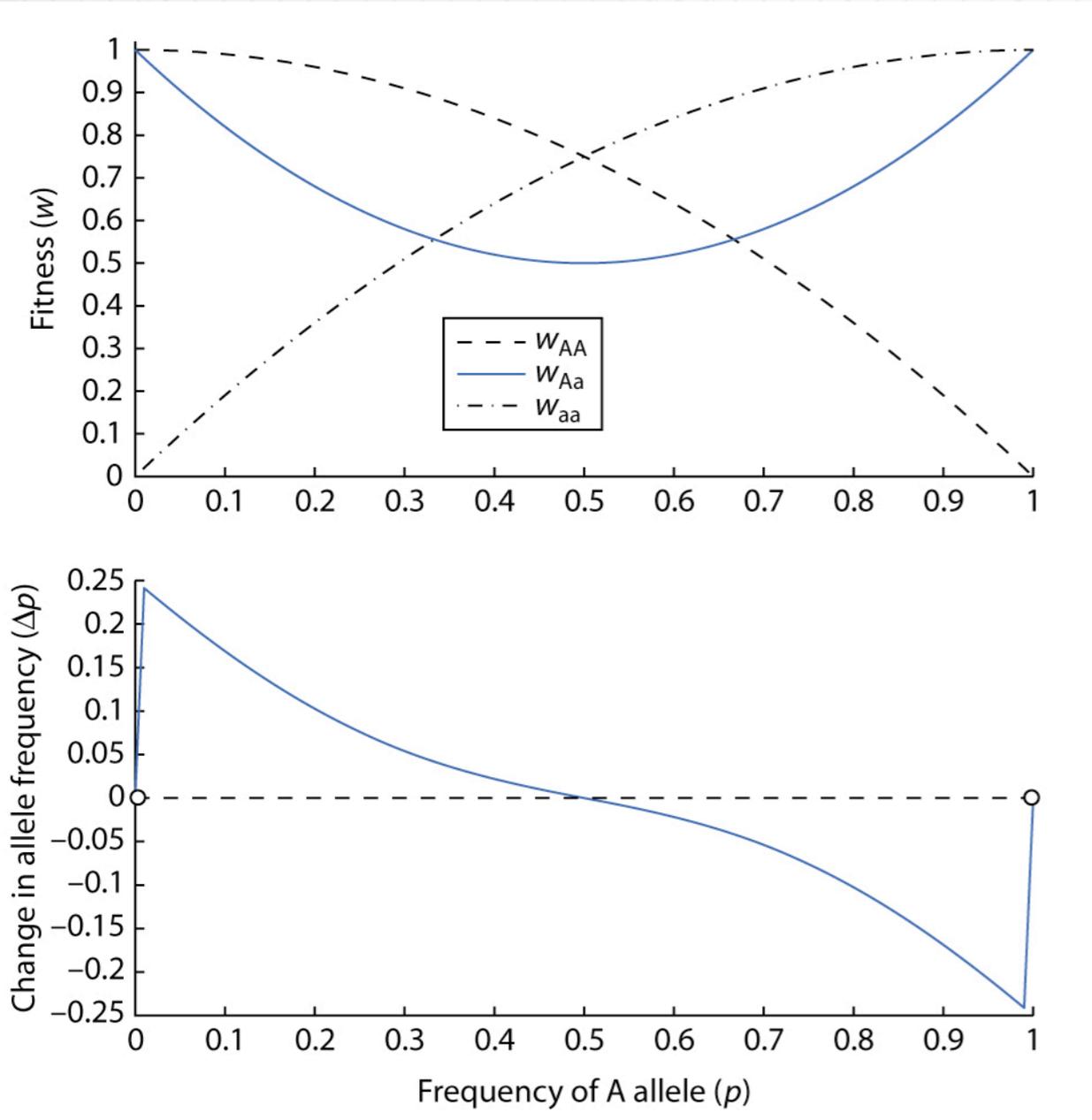
- \* For the special case of the selection coefficient being equal for all genotypes, the expression for change in allele frequency over one generation of fecundity selection is

$$\Delta p = \frac{pqs(q - p)(p^2 - pq + q^2)}{\bar{w}}$$

- \* Equilibrium points occur at fixation ( $p=1$ ), loss ( $p=0$ ), and at  $p=1/2$

# (5) Alternative Models of Natural Selection

Natural selection with frequency-dependent fitness



**Figure 7.5** The relative fitness of each genotype ( $w_{xx}$ ) and the change in allele frequency ( $\Delta p$ ) across all frequencies of the A allele under frequency-dependent natural selection. There is a stable equilibrium point at  $p = 0.5$  in this particular case, even though the heterozygote has the lowest fitness. Two unstable equilibria at fixation and loss are marked with open circles. Here the relative fitness values are  $w_{AA} = 1 - s_{AA}p^2$ ,  $w_{Aa} = 1 - s_{Aa}2pq$ , and  $w_{aa} = 1 - s_{aa}q^2$  with  $s_{AA} = s_{Aa} = s_{aa}$ .

# (5) Alternative Models of Natural Selection

Natural selection with density-dependent fitness

- \* Organisms are limited by the space and resources available to them, limitations that lead to changes in the rate of growth as the density of individuals changes over time
- \* We need to alter our basic genotype-specific population growth equations to incorporate an upper bound on the population size as well as the rate of population growth that changes with population size

# (5) Alternative Models of Natural Selection

Natural selection with density-dependent fitness

- \* A simple model where population growth has an upper bound is called **logistic growth** and the upper limit,  $K$ , is called the **carrying capacity**
- \* Denoting the population size by  $N$  and the rate of increase by  $r$ , we have

$$\lambda = 1 + r - \frac{r}{K}N$$

# (5) Alternative Models of Natural Selection

Natural selection with density-dependent fitness

- \* Logistic growth can be applied to the three genotypes at a diallelic locus by defining genotype-specific carrying capacities and rates of increase to obtain absolute fitness values for each genotype:

$$\lambda_{AA} = 1 + r_{AA} - \frac{r_{AA}}{K_{AA}} N$$

$$\lambda_{Aa} = 1 + r_{Aa} - \frac{r_{Aa}}{K_{Aa}} N$$

$$\lambda_{aa} = 1 + r_{aa} - \frac{r_{aa}}{K_{aa}} N$$

$$[ N_{AA} + N_{Aa} + N_{aa} = N ]$$

# (5) Alternative Models of Natural Selection

Natural selection with density-dependent fitness

- \* The average absolute fitness in the population is

$$\bar{\lambda} = 1 + \bar{r} - \frac{\bar{r}}{K}N$$

where

$$\bar{r} = p_t^2 r_{AA} + 2p_t q_t r_{Aa} + q_t^2 r_{aa}$$

$$\frac{\bar{r}}{K} = \frac{p_t^2 r_{AA}}{K_{AA}} + \frac{2p_t q_t r_{Aa}}{K_{Aa}} + \frac{q_t^2 r_{aa}}{K_{aa}}$$

# (5) Alternative Models of Natural Selection

Natural selection with density-dependent fitness

- \* We can now express the growth in the total size of the population as

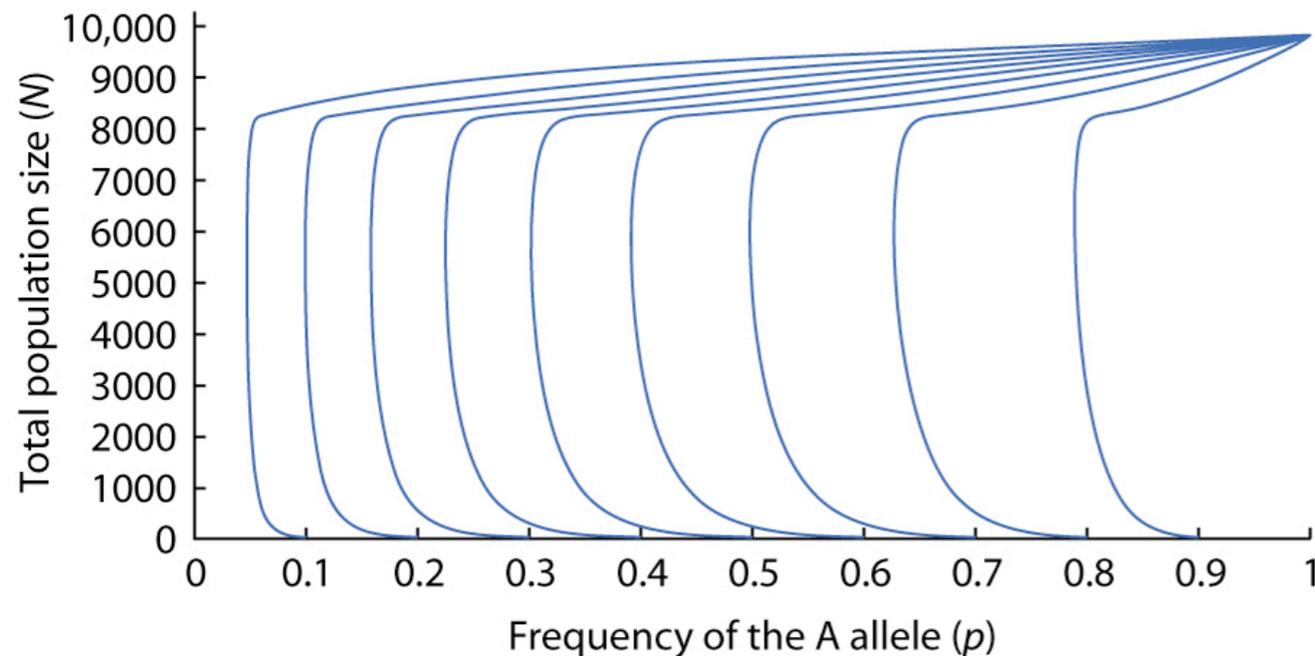
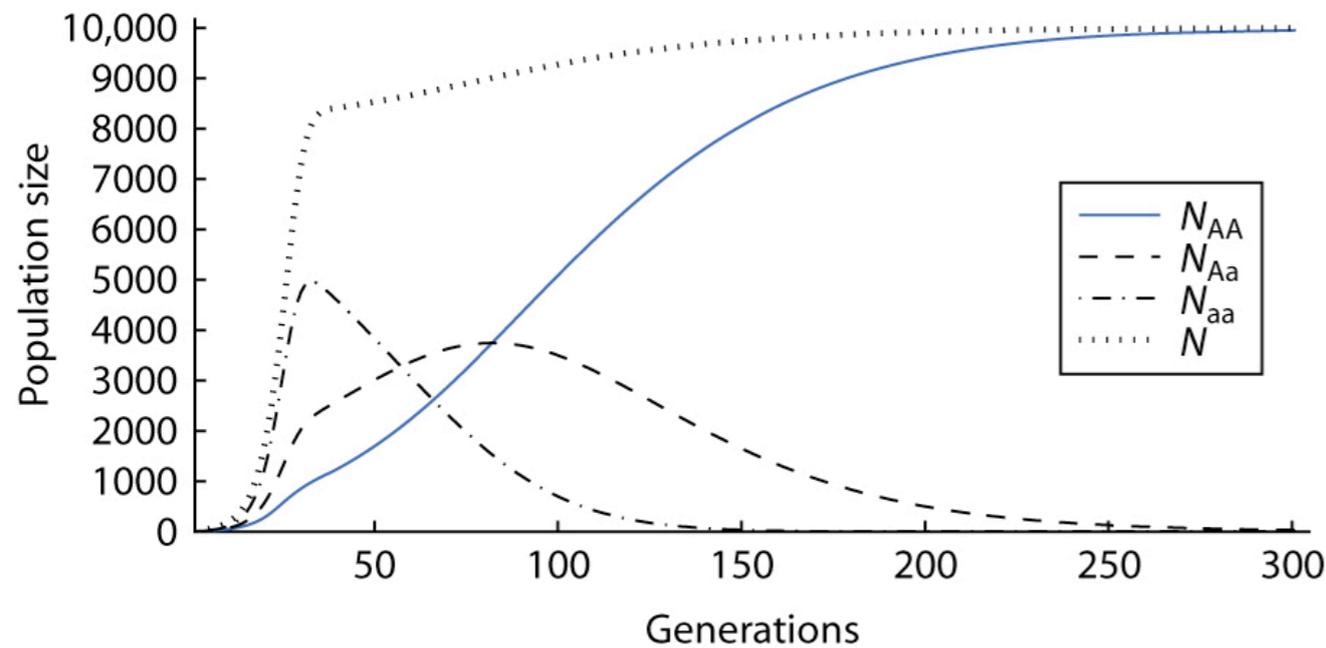
$$N_{t+1} = \bar{\lambda} N_t$$

- \* We can also follow allele frequencies over time by modifying the expression for allele frequency after one generation of selection:

$$p_{t+1} = \frac{\lambda_{AA} p_t^2 + \lambda_{Aa} p_t q_t}{\bar{\lambda}}$$

# (5) Alternative Models of Natural Selection

Natural selection with density-dependent fitness



**Figure 7.6** The results of density-dependent natural selection on the numbers of individuals of different genotypes ( $N_{AA}$ ,  $N_{Aa}$ , and  $N_{aa}$ ) and allele frequencies in a population of total size  $N$ . At the upper limit of  $N$ , the equilibrium allele and genotype frequencies are determined by the genotype with the highest carrying capacity ( $K$ ). In contrast, the genotype with the highest growth rate ( $r$ ) has the greatest impact on allele frequency when the population is small. In this example  $K_{AA} = 10,000$ ,  $K_{Aa} = 9000$ , and  $K_{aa} = 8000$  with  $r_{AA} = 0.2$ ,  $r_{Aa} = 0.25$ , and  $r_{aa} = 0.3$ .

# (6) Combining Natural Selection with Other Processes

\* We now consider

1. natural selection and genetic drift acting simultaneously, and
2. the balance between natural selection and mutation

# (6) Combining Natural Selection with Other Processes

Natural selection and genetic drift

- \* Wright showed the probability that a population has a given allele frequency when exposed to the simultaneous processes of natural selection, genetic drift, and mutation as given by

$$\phi(p) = C p^{(4N_e\mu-1)} q^{(4N_e\nu-1)} e^{(4N_e spq)}$$

# (6) Combining Natural Selection with Other Processes

## Natural selection and genetic drift

- \* Wright showed the probability that a population has a given allele frequency when exposed to the simultaneous processes of natural selection, genetic drift, and mutation as given by

$$\phi(p) = C p^{(4N_e\mu-1)} q^{(4N_e\nu-1)} e^{(4N_e spq)}$$

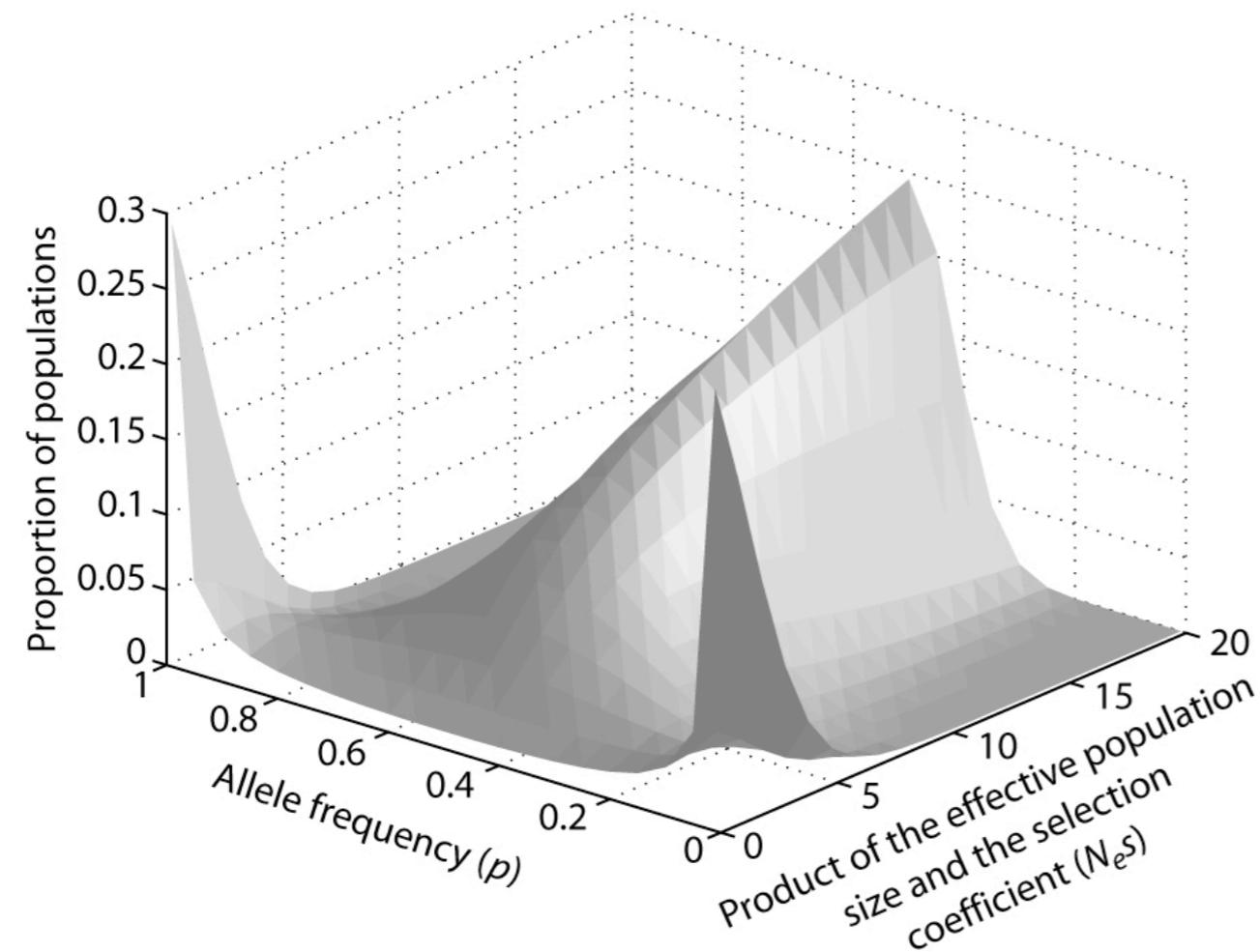
effective population size      forward/backward mutation rates      selection coefficient

probability density      allele frequencies

constant to adjust the total probability across all allele frequencies to sum to 1.0 for each value of  $N_e$

# (6) Combining Natural Selection with Other Processes

## Natural selection and genetic drift



**Figure 7.7** The expected distribution of allele frequencies for a very large number of replicate finite populations under natural selection where there is overdominance for fitness ( $w_{AA} = w_{aa} = 1 - s$  and  $w_{Aa} = 1$ ). In an infinite population the expected allele frequency at equilibrium is 0.5. However, in finite populations the equilibrium allele frequency will depend on the balance of natural selection and genetic drift. This balance is determined by the product of the effective population size and the selection coefficient ( $N_e s$ ). Low values of  $N_e s$  mean that selection is very weak compared to drift and each population reaches fixation or loss. High values of  $N_e s$  mean that selection is strong compared to drift and most populations reach an equilibrium allele frequency near 0.5. Here forward and backward mutation rates are equal ( $\mu = \nu = 0.00001$ ).

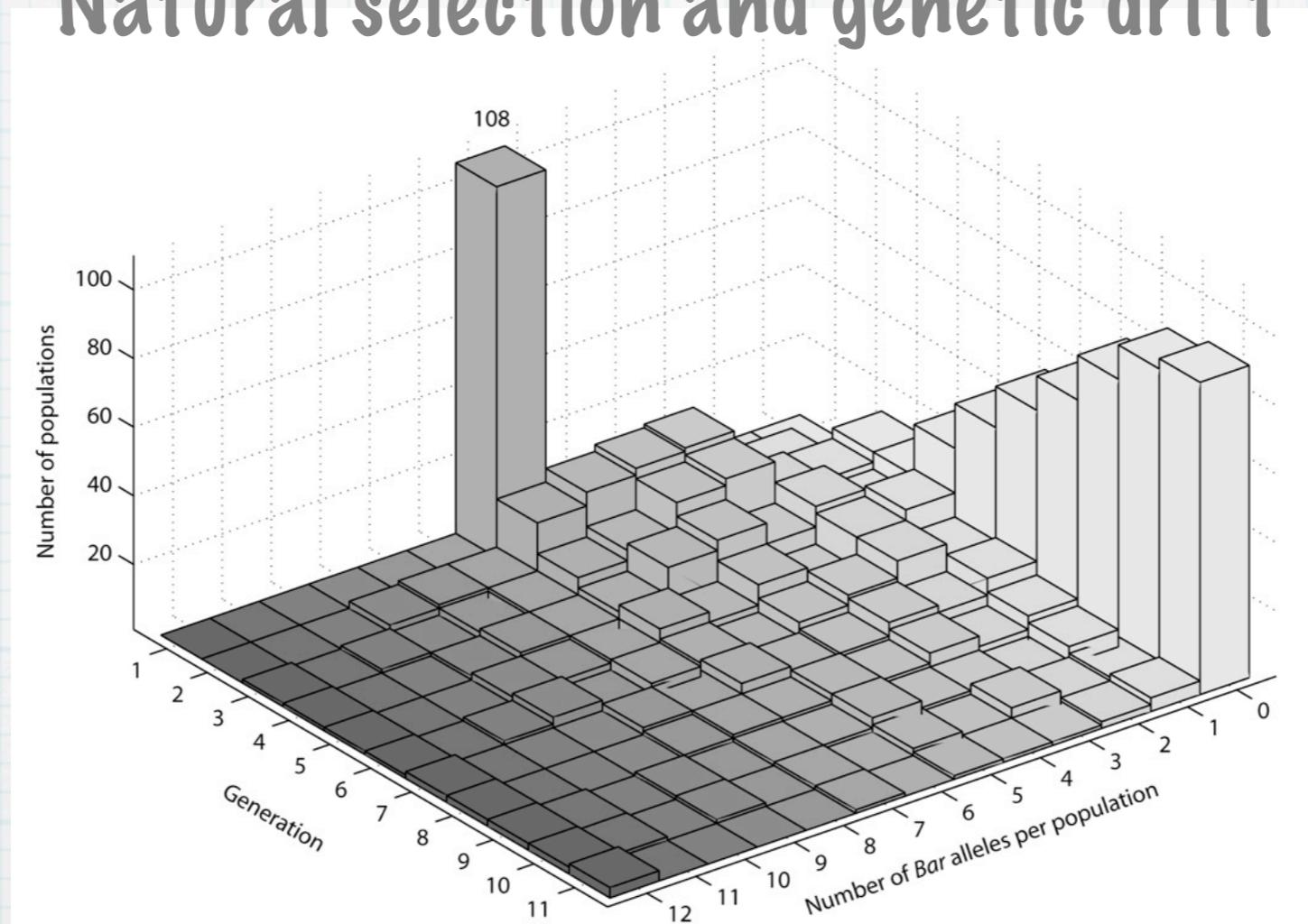
# (6) Combining Natural Selection with Other Processes

Natural selection and genetic drift

- \* Kimura suggested the following rule of thumb:
  - \* If  $4N_e s \ll 1$ , then selection is weak relative to sampling, and genetic drift will dictate allele frequencies
  - \* If  $4N_e s \gg 1$ , then selection is strong relative to sampling, and natural selection will dictate allele frequencies
  - \* If  $4N_e s \approx 1$ , then allele frequencies are unpredictable

# (6) Combining Natural Selection with Other Processes

## Natural selection and genetic drift



**Figure 7.8** Frequency of the *Bar* allele in 108 replicate *D. melanogaster* populations over 10 generations (Wright & Kerr 1954). Each population was founded from four males and four females. The *Bar* locus is found on the X chromosome and so is hemizygous in males, making the effective population size equivalent to about six individuals. The eyes of *D. melanogaster* individuals homozygous for the wild-type allele are oval, but heterozygotes and homozygotes for the partially dominant *Bar* allele have bar-shaped eyes with a reduced number of facets. Females homozygous for the *Bar* allele produced 37% of the progeny compared to females homozygous or heterozygous for the wild-type allele. Despite this strong natural selection against *Bar*, three populations fixed for *Bar* by the end of the experiment. Compare with the similar example in Figure 3.11 where the locus is selectively neutral.

# (6) Combining Natural Selection with Other Processes

Natural selection and genetic drift

- \* It is also possible to gain biological insight into  $N_e s$  by recognizing that it is analogous to the quantity  $N_e m$  that dictates the balance between genetic drift and gene flow
- \* Both quantities represent the net balance of the pressure on allele frequencies toward eventual fixation or loss due to genetic drift and the countervailing force driving allele frequencies toward a specific allele frequency caused by either natural selection or by gene flow

# (6) Combining Natural Selection with Other Processes

Natural selection and genetic drift

- \* In the case of natural selection, the specific allele frequency is dictated by the relative fitness values of genotypes, while in the case of gene flow the specific allele frequency is the average allele frequency for all demes

# (6) Combining Natural Selection with Other Processes

Natural selection and mutation

- \* Natural selection takes place at the same time that mutation is working to alter allele frequencies and reintroduce alleles that selection may be driving to loss
- \* Therefore, the process of natural selection may be counteracted to some degree by mutation
- \* What equilibrium allele frequency is expected when the opposing processes of mutation and natural selection balance out?

# (6) Combining Natural Selection with Other Processes

Natural selection and mutation

- \* Assume a diallelic locus, with allele  $a$  being completely recessive and has frequency  $q$
- \* Assume the case of selection against a recessive allele:  $w_{AA}=1$ ,  $w_{Aa}=1$ , and  $w_{aa}=1-s$
- \* We have

$$\Delta q_{selection} = \frac{pq[q((1-s) - 1) + p(1 - 1)]}{(1)p^2 + (1)2pq + (1-s)q^2} = \frac{-spq^2}{1 - sq^2}$$

# (6) Combining Natural Selection with Other Processes

Natural selection and mutation

- \* If we assume only  $A \rightarrow a$  mutations, with rate  $\mu$ , then

$$\Delta q_{\text{mutation}} = \mu p$$

# (6) Combining Natural Selection with Other Processes

Natural selection and mutation

\* At equilibrium

$$\Delta q_{\text{mutation}} + \Delta q_{\text{selection}} = 0$$

\* This yields

$$\mu p = \frac{spq^2}{1 - sq^2}$$

# (6) Combining Natural Selection with Other Processes

Natural selection and mutation

- \* If  $q$  is low,  $q^2$  is very small and  $1-sq^2$  is approximately 1, then

$$\mu p = spq^2$$

- \* This yields

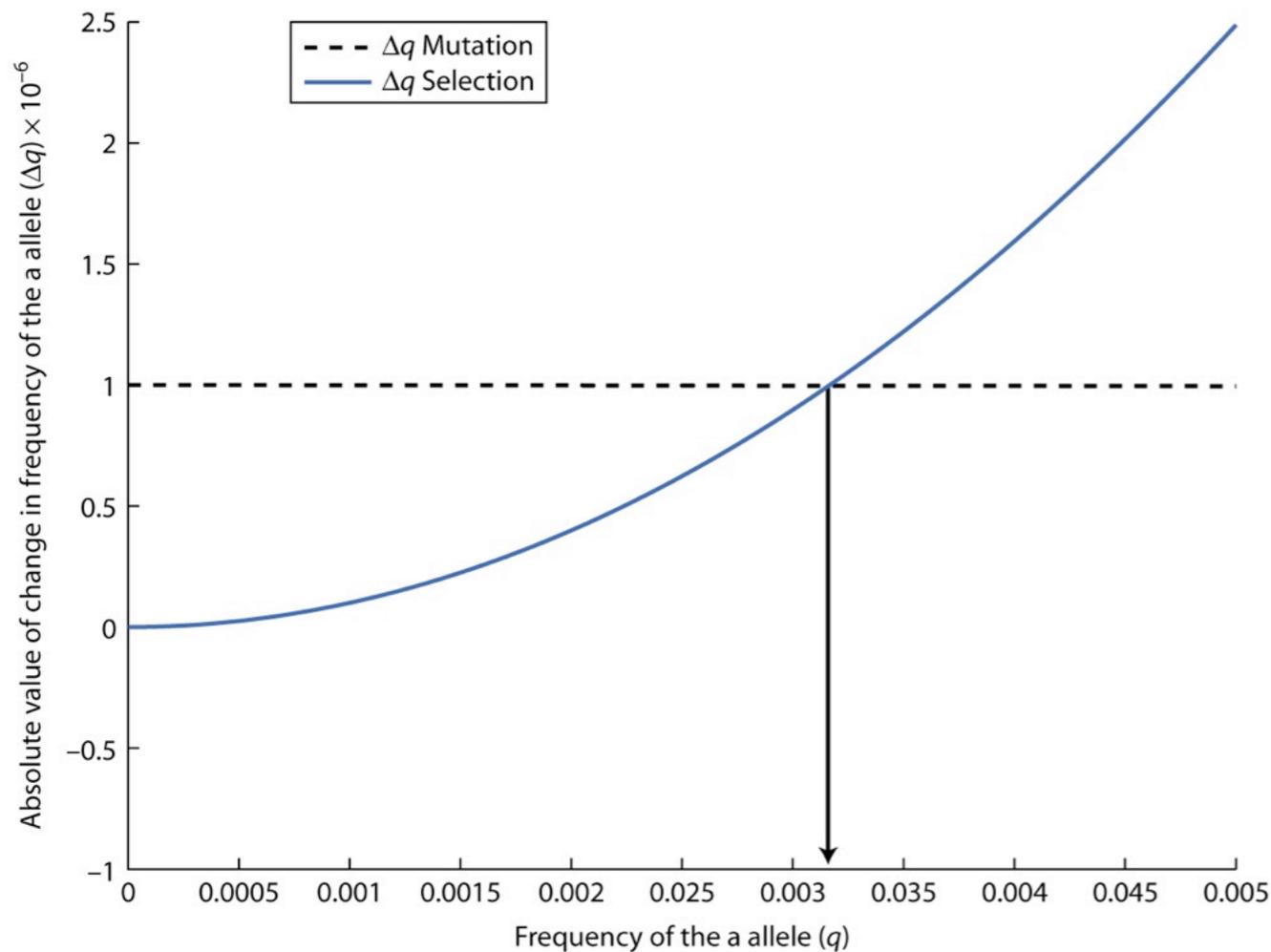
$$q^2 = \frac{\mu}{s}$$

- \* Hence

$$q_{equilibrium} = \sqrt{\frac{\mu}{s}}$$

# (6) Combining Natural Selection with Other Processes

## Natural selection and mutation



**Figure 7.9** The absolute value of the change in allele frequency due to mutation ( $\Delta q_{\text{mutation}}$ ) and due to natural selection ( $\Delta q_{\text{selection}}$ ) when there is selection against a recessive homozygote. Mutation continually makes new copies of the recessive allele while selection continually works toward loss of the recessive allele. The equilibrium allele frequency occurs when the processes of mutation and selection exactly counteract each other. Here  $s = 0.1$  and  $\mu = 1 \times 10^{-6}$  so the expected equilibrium is  $q_{\text{equilibrium}} = 0.0032$  as shown by the vertical arrow.

# (6) Combining Natural Selection with Other Processes

Natural selection and mutation

- \* Consanguineous mating results in an excess of homozygosity and a deficit of heterozygosity
- \* If  $f$  is the degree of departure from HW expectations, the the equilibrium allele frequency is

$$q^2 + fpq = \frac{\mu}{s}$$

# (6) Combining Natural Selection with Other Processes

Natural selection and mutation

- \* Under the assumption that  $q$  is small compared to  $f$ , the approximate equilibrium allele frequency when mutation and selection reach a balance with consanguineous mating is

$$q_{\text{equilibrium}} = \frac{\mu}{fs}$$

# (6) Combining Natural Selection with Other Processes

Natural selection and mutation

- \* Since recessive deleterious mutations are only perceived by natural selection when homozygous, consanguineous mating increases the effectiveness of selection by increasing the proportion of homozygous genotypes in the population
- \* This means that selection is more effective at eliminating the recessive homozygote and the equilibrium allele frequency for mutation-selection balance occurs at a lower allele frequency

# (6) Combining Natural Selection with Other Processes

Natural selection and mutation

- \* It is counterintuitive that populations which cease consanguineous mating and engage in random mating may temporarily experience an increase in deleterious allele frequencies and a decrease in average fitness due to less effective natural selection

# (7) Natural Selection and the Coalescent

- \* In the basic coalescent model where alleles are selectively neutral, each lineage within a generation has an equal and constant probability of becoming an ancestral lineage when working back in time to find the MRCA
- \* **Natural selection violates the basic assumption that all lineages have equal and constant probabilities of coalescence**

# (7) Natural Selection and the Coalescent

- \* When natural selection operates, some lineages tend to increase in frequency over time, whereas other lineages tend to decrease in frequency over time due to fitness differences among the lineages caused by differences in haplotype relative fitnesses
- \* These changes in the frequencies of lineage copies translate into probabilities of coalescence that change over time as well

# (7) Natural Selection and the Coalescent

- \* A lineage bearing a **haplotype favored by selection** will **increase in frequency over time** and therefore will have a **decreasing probability of coalescence** working back in time
- \* Similarly, a lineage bearing a **haplotype of lower fitness** will **decrease in frequency over time** and therefore have an **increasing probability of coalescence** moving back in time

# (7) Natural Selection and the Coalescent

- \* In other words, **natural selection presents a fundamental contradiction to the sampling process built into the genealogical branching model**
- \* We'll now discuss extensions of the coalescent model under directional selection and under balancing selection

# (7) Natural Selection and the Coalescent

Directional selection and the ancestral selection graph

- \* There's a relatively simple way to modify the genealogical branching model to accommodate directional natural selection
- \* Coalescence and natural selection are treated as distinct processes that can both possibly occur working back in time from the present toward the MRCA in the past

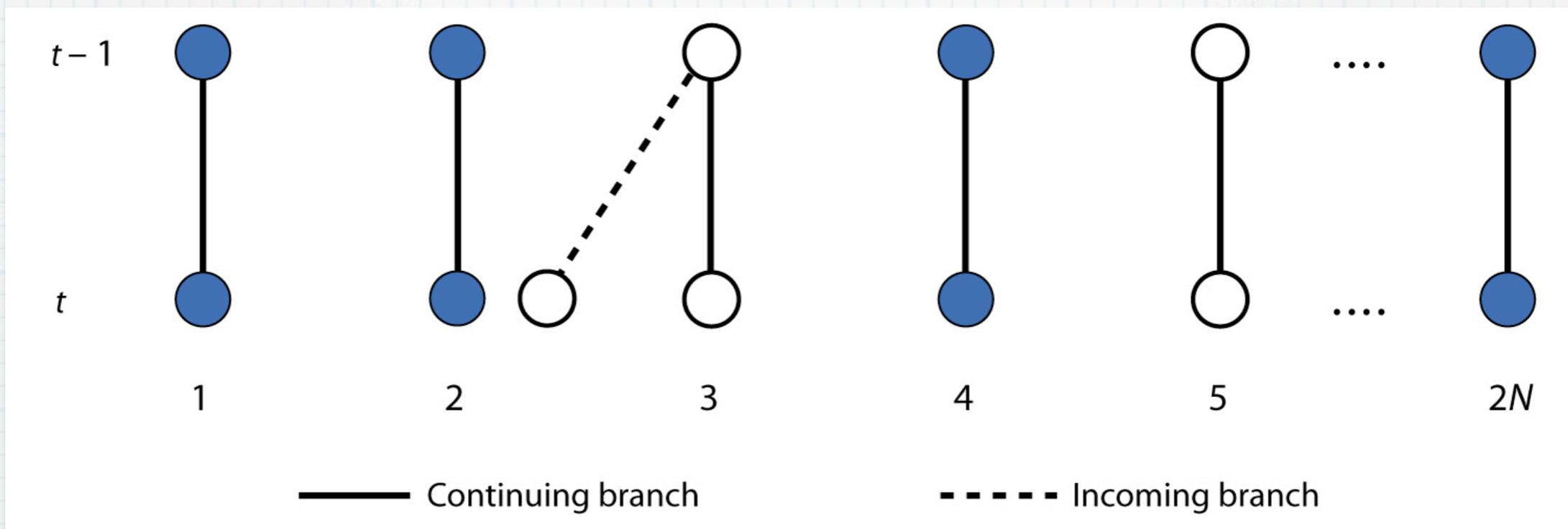
# (7) Natural Selection and the Coalescent

Directional selection and the ancestral selection graph

- \* We'll alter slightly our view of the sampling process
- \* We'll distinguish between a **"continuing branch"**, which corresponds to selectively neutral lineages, and an **"incoming branch"** that displaces a lineage of the lower-fitness haplotype to model natural selection

# (7) Natural Selection and the Coalescent

## Directional selection and the ancestral selection graph



**Figure 7.10** Haploid reproduction with the possibility of coalescence and natural selection events. Each haploid lineage replicates itself and is included in the next generation if there are no coalescence events (solid lines). A lineage makes an extra copy of itself (dashed line) and has the potential to displace one copy of another lineage. If the lineage making the extra copy of itself (open circle) has a higher-fitness haplotype than a randomly chosen lineage (blue circle), then it will displace the lineage of the lower fitness haplotype. Therefore, the outcome of a lineage-duplication event that may result in natural selection depends on the haplotype states of the specific lineages involved. The solid lines are continuing branches and the dashed line is an incoming branch.

# (7) Natural Selection and the Coalescent

Directional selection and the ancestral selection graph

- \* We can now treat the dual continuing/incoming branching process as two independent parts of the overall coalescent process
- \* As before, we assume that coalescent and natural selection events are rare, or that  $N_e$  is large and the selection coefficient is small
- \* This assumption allows us to treat the two events as mutually exclusive and to assume that whenever an event occurs that it is either coalescence or natural selection

# (7) Natural Selection and the Coalescent

Directional selection and the ancestral selection graph

- \* The exponential approximation for the chance that a natural selection event occurs at generation  $t$  for a single lineage and for  $k$  lineages, respectively, are

$$P(T_{\text{incoming branch}} = t) = e^{-t \frac{\sigma}{2}}$$

and

$$P(T_{\text{incoming branch}} = t) = e^{-t \frac{\sigma}{2} k}$$

where  $\sigma = 4N_e s$

# (7) Natural Selection and the Coalescent

Directional selection and the ancestral selection graph

- \* The chance that an incoming branch due to natural selection displaces one of  $k$  lineages at or before a certain time can then be approximated with the exponential cdf

$$P(T_{\text{incoming branch}} \leq t) = 1 - e^{-t \frac{\sigma}{2} k}$$

- \* Combining natural selection with coalescence and mutation, we get

$$P(T_{\text{event}} \leq t) = 1 - e^{-t \left( \frac{k(k-1)}{2} + \frac{\sigma}{2} k + \frac{\theta}{2} k \right)}$$

# (7) Natural Selection and the Coalescent

Directional selection and the ancestral selection graph

- \* The total chance that an event occurs is

$$\frac{k(k-1)}{2} + \frac{\sigma}{2}k + \frac{\theta}{2}k$$

- \* From this we can get the chance that an event is either a coalescence, due to natural selection, or a mutation

# (7) Natural Selection and the Coalescent

Directional selection and the ancestral selection graph

- \* The chance that an event is a coalescence:

$$\frac{\frac{k(k-1)}{2}}{\frac{k(k-1)}{2} + \frac{\sigma}{2}k + \frac{\theta}{2}k} = \frac{k-1}{k-1 + \sigma + \theta}$$

- \* The chance that an event is due to natural selection:

$$\frac{\frac{\sigma}{2}k}{\frac{k(k-1)}{2} + \frac{\sigma}{2}k + \frac{\theta}{2}k} = \frac{\sigma}{k-1 + \sigma + \theta}$$

- \* The chance that an event is a mutation:

$$\frac{\frac{\theta}{2}k}{\frac{k(k-1)}{2} + \frac{\sigma}{2}k + \frac{\theta}{2}k} = \frac{\theta}{k-1 + \sigma + \theta}$$

# (7) Natural Selection and the Coalescent

Directional selection and the ancestral selection graph

- \* Using these equations to determine the chance that an event occurs and the type of event, it is possible to construct what is known as an **ancestral selection graph**
- \* In this model, natural selection events result in an addition of branches and thereby serve to visualize selection events that are not apparent on a genealogy alone

# (7) Natural Selection and the Coalescent

Directional selection and the ancestral selection graph

- \* When a branching occurs due to a natural selection event (going back in time), the resulting branch is called the **incoming branch** to represent a possible lineage displacement
- \* The lineage that the incoming branch splits off from is called the **continuing branch**
- \* The incoming branch coalesces with a randomly chosen lineage at a later time determined by the waiting time distribution and assumes the state of the branch where it coalesces

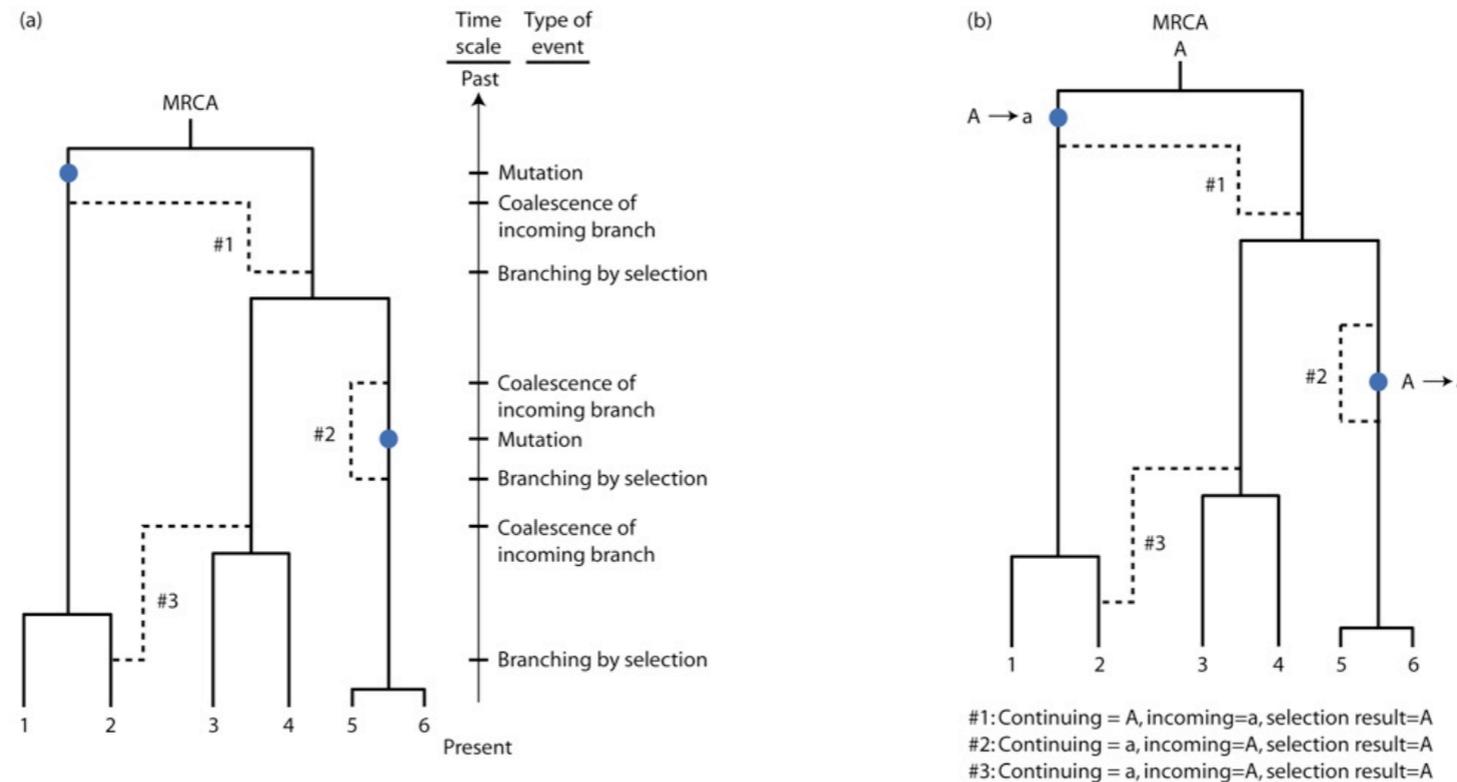
# (7) Natural Selection and the Coalescent

Directional selection and the ancestral selection graph

- \* Even though natural selection events make more branches, the coalescence process is faster and will eventually result in coalescence to the MRCA

# (7) Natural Selection and the Coalescent

## Directional selection and the ancestral selection graph



**Figure 7.11** The ancestral selection graph used to include natural selection in the genealogical branching model. In (a), the waiting times between events and the types of events are determined until the MRCA is reached by working backward in time from six lineages in the present. Branching and coalescence events due to natural selection (dashed lines) and mutation events (blue circles) are identified. Natural selection causes the addition of one “incoming” branch to the number of lineages that can coalesce and this incoming branch can then coalesce with any lineage. In (b), a haplotype state is assigned to the ultimate ancestor and allelic states are traced forward in time to determine the outcome of mutation and natural selection events. At each of the selection events the state of the continuing branch and the incoming branch are compared. In this example A is the fitter haplotype and it displaces the a haplotype when continuing and incoming branches coalesce. When the haplotypes of continuing and incoming branches are identical there is no change in haplotype state. In (c), the haplotype states of the lineages in the present are assigned once all of the selection events have been resolved. In this example selection causes a slight increase in the total branch length because the selection event at #3 displaces a shorter branch.

# (7) Natural Selection and the Coalescent

Directional selection and the ancestral selection graph

- \* Weak to moderate directional natural selection tends to have only a minor impact on average times to coalescence
- \* When the selection coefficient and mutation rate are approximately equal, the mean time to the MRCA is shortened slightly
- \* Strong natural selection for advantageous alleles or selection against deleterious mutations is expected to reduce the total height of genealogical trees

# (7) Natural Selection and the Coalescent

Genealogies and balancing selection

- \* **Balancing selection** (heterozygotes have the highest fitness) can also be incorporated in genealogical branching models
- \* Balancing selection is expected to maintain both alleles at a diallelic locus segregating in the population at equilibrium
- \* Hence, balancing selection is a special case of natural selection because it works counter to the fixation and loss due to genetic drift

# (7) Natural Selection and the Coalescent

## Genealogies and balancing selection

- \* In a genealogical branching model, genetic drift is represented by the process of coalescence
- \* To approximate the overall effect of balancing selection, we need a process that will delay coalescence to the same degree that selection favors heterozygotes in the diploid selection model
- \* This can be obtained by modeling selection along the lines of population structure with two demes

# (7) Natural Selection and the Coalescent

Genealogies and balancing selection

- \* In structured populations, two lineages cannot coalesce unless they are in the same deme
- \* Gene flow events that move lineages into different demes tend to delay coalescence events
- \* Using this logic, balancing selection can be modeled in a single panmictic population as a process where there are two lineage types
- \* A switching process (akin to gene flow) changes lineage types at random, while the coalescence process operates

# (7) Natural Selection and the Coalescent

Genealogies and balancing selection

- \* If two lineages must be of the same type to coalesce, then the switching process will prevent coalescence among the lineages that are of different types

# (7) Natural Selection and the Coalescent

Genealogies and balancing selection

- \* Let  $A$  and  $B$  be the two lineage types and their respective frequencies be  $p$  and  $q$  ( $p+q=1$ )
- \* Every generation, lineages of one type may switch to the other type with rate  $\mu$
- \* The expected number of lineages that switch type each generation is  $2N\mu$ , or  $v/2$  if we use  $v=4N\mu$

# (7) Natural Selection and the Coalescent

Genealogies and balancing selection

- \* The expected number of lineages that switch type each generation serves as a surrogate for the strength of balancing selection

# (7) Natural Selection and the Coalescent

Genealogies and balancing selection

- \* Using the switching rate, the expected waiting time until an A lineage switches to a B lineage is

$$P(T_{A \rightarrow B} \leq t) = 1 - e^{-k_A \frac{\nu}{2} \left(\frac{q}{p}\right)}$$

- \* And the expected waiting time until a B lineage switches to an A lineage is

$$P(T_{B \rightarrow A} \leq t) = 1 - e^{-k_B \frac{\nu}{2} \left(\frac{p}{q}\right)}$$

# (7) Natural Selection and the Coalescent

## Genealogies and balancing selection

- \* Next we need to express the waiting time until a coalescence event, keeping in mind that lineages can coalesce only if they are of the same type
- \* Given that there are  $2Np$  lineages of type A and  $2Nq$  lineages of type B and coalescence events are mutually exclusive, the expected waiting time until a coalescence event is

$$P(T_{\text{coalescence}} \leq t) = 1 - e^{-\frac{k_A(k_A-1)}{2} \left(\frac{1}{p}\right) + \frac{k_B(k_B-1)}{2} \left(\frac{1}{q}\right) t}$$

# (7) Natural Selection and the Coalescent

## Genealogies and balancing selection

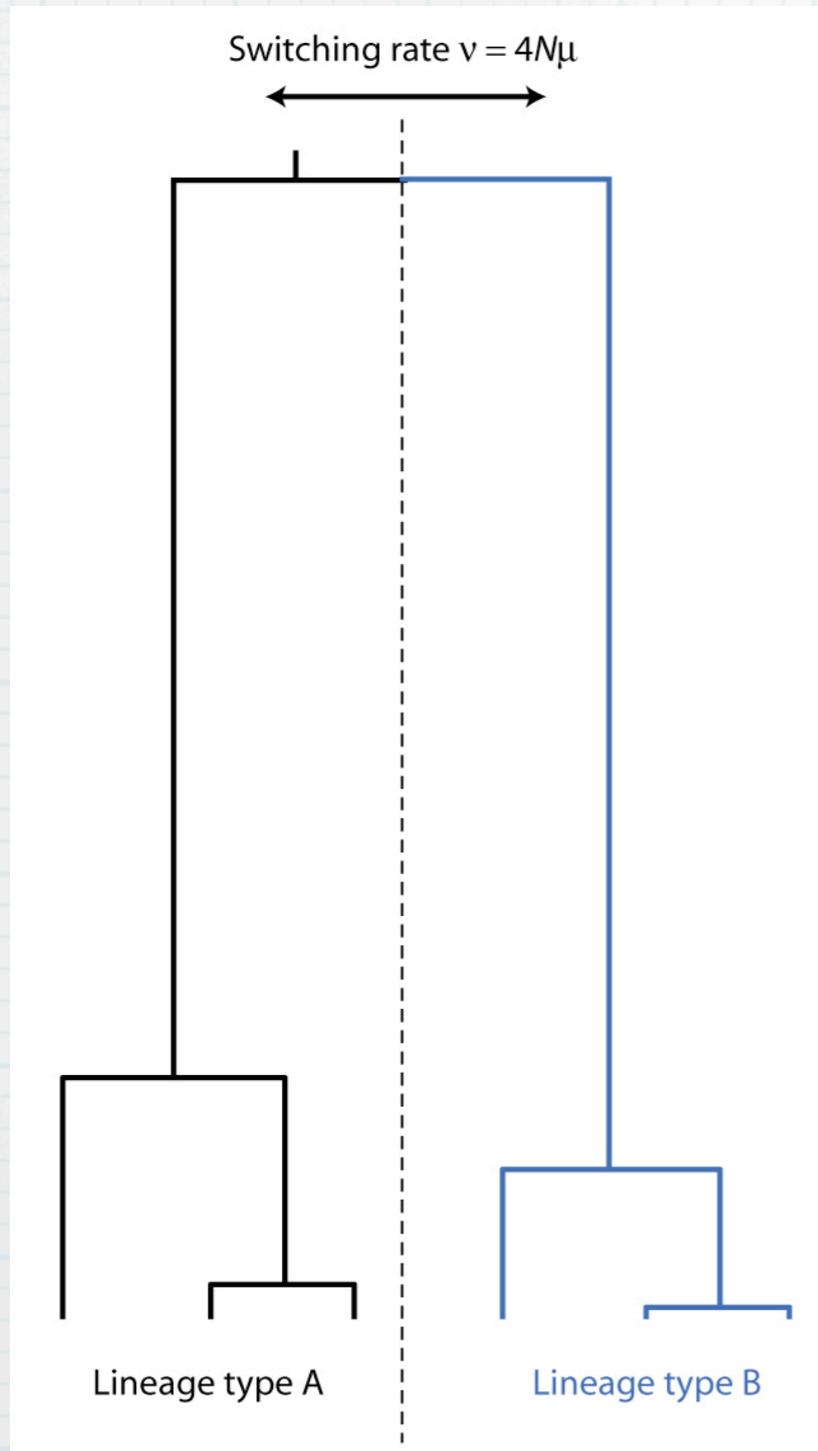
- \* If switching events and coalescence events are all mutually exclusive, the individual exponents can be added together to obtain the total waiting time to any event:

$$P(T_{\text{coalescence}} \leq t) = 1 - e^{-\frac{k_A(k_A-1)}{2} \left(\frac{1}{p}\right) + \frac{k_B(k_B-1)}{2} \left(\frac{1}{q}\right) + k_A \frac{\nu}{2} \left(\frac{q}{p}\right) + k_B \frac{\nu}{2} \left(\frac{p}{q}\right)}$$

- \* Given that an event has occurred with a known waiting time, the type of event can be determined by drawing a random number between zero and one and comparing it with the cumulative total of the chance of each event divided by the total probability of all events

# (7) Natural Selection and the Coalescent

## Genealogies and balancing selection



**Figure 7.12** A genealogy where balancing natural selection is modeled by type switching. Every generation, lineages of one type (here A and B) may switch to the other type with rate  $\mu$ . Twice the expected number of the  $2N$  total lineages in the population that switch types each generation is then  $v = 4N\mu$ . Since lineages can only coalesce when they are of the same type, type switching increases the average time to coalescence. This is analogous to natural selection favoring heterozygotes because overdominance also extends the segregation times of alleles. Genealogical trees that result from balancing selection modeled as type switching tend to have longer branches compared to genealogies that result from genetic drift or directional natural selection.

# Summary

- \* For haploid organisms, natural selection is a population growth process where different genotypes vary in genotype-specific population growth rates
- \* The ratio of genotype-specific growth rates is the relative fitness and it predicts the genotype that will approach fixation
- \* Natural selection in diploid organisms also relies on the relative fitness, with the addition of sexual reproduction such that pairs of parents can produce a predictable frequency of genotypes in their progeny under random mating
- \* The outcomes of viability selection: directional selection (a homozygote most fit) results in (nearly)fixation/loss, balancing selection maintains both alleles forever, and disruptive selection results in fixation/loss depending on initial genotype frequencies

# Summary

- \* The fundamental theorem of natural selection: the change in mean fitness by natural selection is proportional to the additive genetic variation in fitness
- \* Natural selection changes genotype frequencies fastest when gene action is additive
- \* Natural selection acts as an uphill climber on a fitness surface, moving genotype frequencies uphill based on the slope at the current genotype frequencies
- \* Fitness surfaces will have multiple peaks and valleys if there is dominance or epistasis
- \* When considering more than a single allele, recombination must be taken into account

# Summary

- \* In addition to viability, fitness can be defined in terms of fecundity or carrying capacity, and its values may vary in time and space
- \* Natural selection and genetic drift determine the fate of genotypes/alleles depending on the value of  $4N_e s$
- \* The fate of a new mutation is determined by natural selection depending on its fitness
- \* Ancestral selection graphs model directional selection with the coalescent, and balancing selection can be modeled in the same fashion as the coalescent on a two-deme structured population