Coop, Chapter 10: Intro.-10.0.1

One-Locus Models of Selection

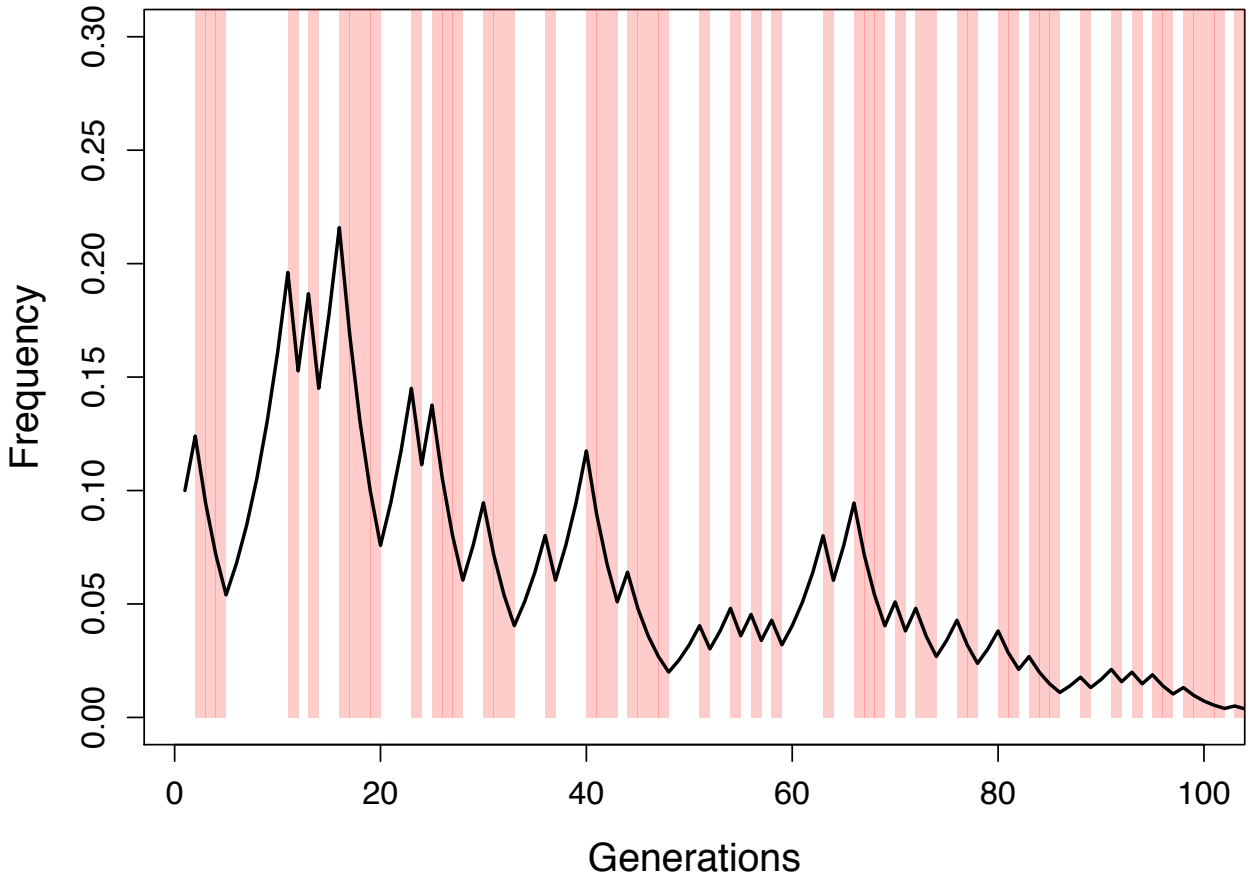
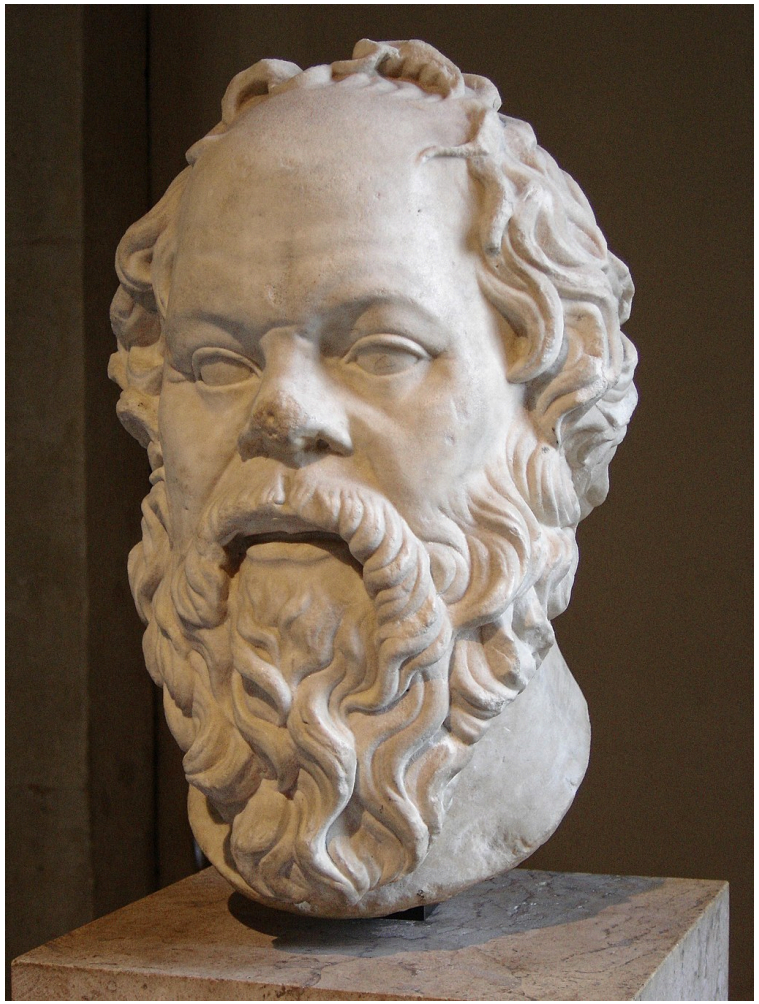


Figure 10.2: An example frequency trajectory of the A_1 allele under variable environments (using the fitnesses from Table 10.1). Wet years (generations) are shown in red, dry years in white. The environment flips at random each year. Note how the A_1 allele increases in frequency in the dry years as it has higher fitness, and yet the A_2 allele still wins out. Code here.

Introduction

“Socrates consisted of the genes his parents gave him, the experiences they and his environment later provided, and a growth and development mediated by numerous meals. For all I know, he may have been very successful in the evolutionary sense of leaving numerous offspring. His phenotype, nevertheless, was utterly destroyed by the hemlock and has never since been duplicated. The same argument holds also for genotypes. With Socrates’ death, not only did his phenotype disappear, but also his genotype.[...] The loss of Socrates’ genotype is not assuaged by any consideration of how prolifically he may have reproduced. Socrates’ genes may be with us yet, but not his genotype, because meiosis and recombination destroy genotypes as surely as death.”

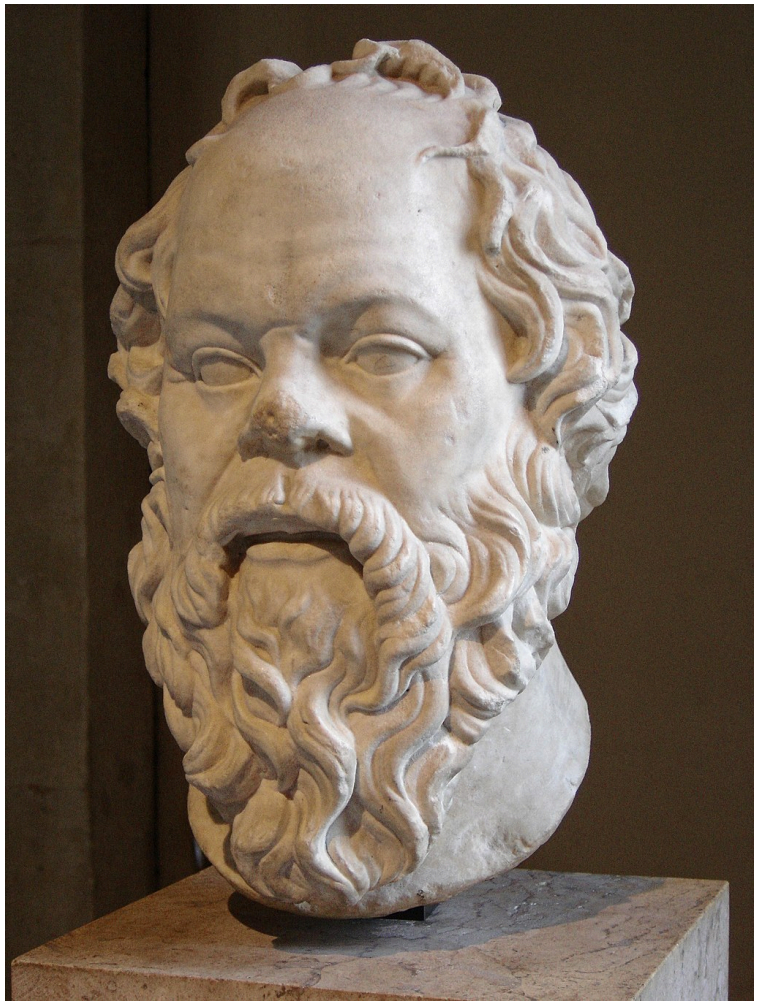
–Williams (1966)



Marble Head of Socrates, Louvre

Introduction

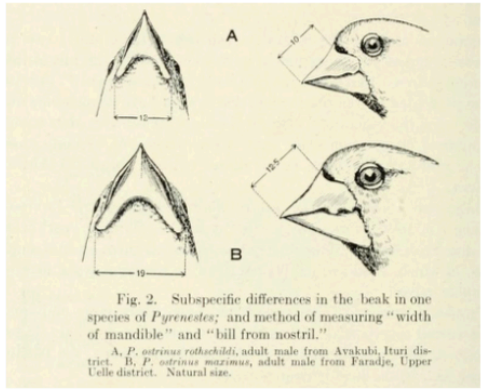
- Organisms along with their genotypes and phenotypes are temporary, yet individual alleles have more permanence
- Phenotypic changes driven by natural selection are due to changes in the allelic composition of populations
- To understand selection, we must understand changes in allele frequency

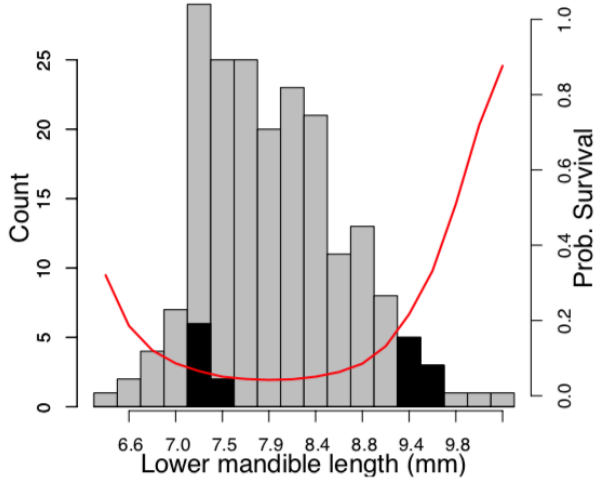


Marble Head of Socrates, Louvre

Introduction

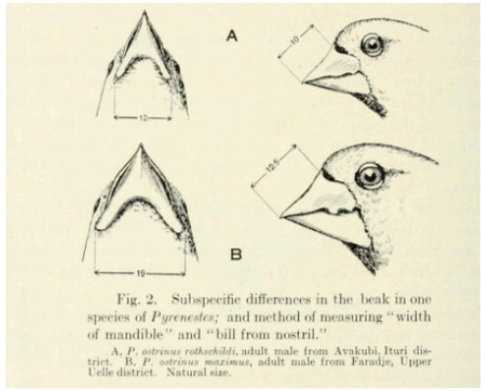
- We've seen in multiple examples over the last few chapters that natural selection occurs when there are differences across individuals in terms of fitness
- Fitness may be defined as the probability to survive long enough to reproduce (viability) or based on the number of offspring (fertility) or both
- We'll define absolute fitness of a genotype as the expected number of offspring of an individual with that genotype

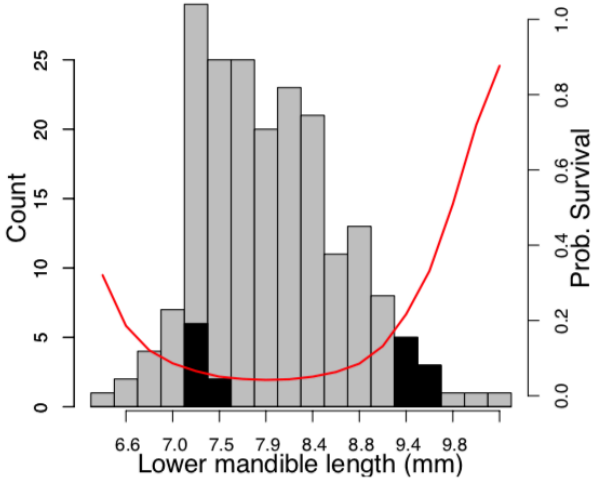




Introduction

- Differences in number of offspring will result in allele frequency changes across generations
- In this chapter, we'll study these dynamics based on single loci
- We will also be ignoring the effects of genetic drift, but will return to the interaction of selection and drift in future chapters





10.0.1 Haploid Selection Model

- We will model selection in a haploid due its relative mathematical simplicity
- The number of individuals carrying the A_1 and A_2 alleles in generation t are P_t and Q_t
- The relative frequencies of these alleles are: $p_t = P_t(P_t + Q_t)$ and $q_t = Q_t(P_t + Q_t)$
- Individuals with A_1 and A_2 alleles produce, on average, W_1 and W_2 offspring, with W_i being the absolute fitness of each genotype
- In the next generation, the absolute number of carriers of A_1 and A_2 alleles will be $P_{t+1} = W_1 P_t$ and $Q_{t+1} = W_2 Q_t$

10.0.1 Haploid Selection Model

- The mean absolute fitness of the population will then be:

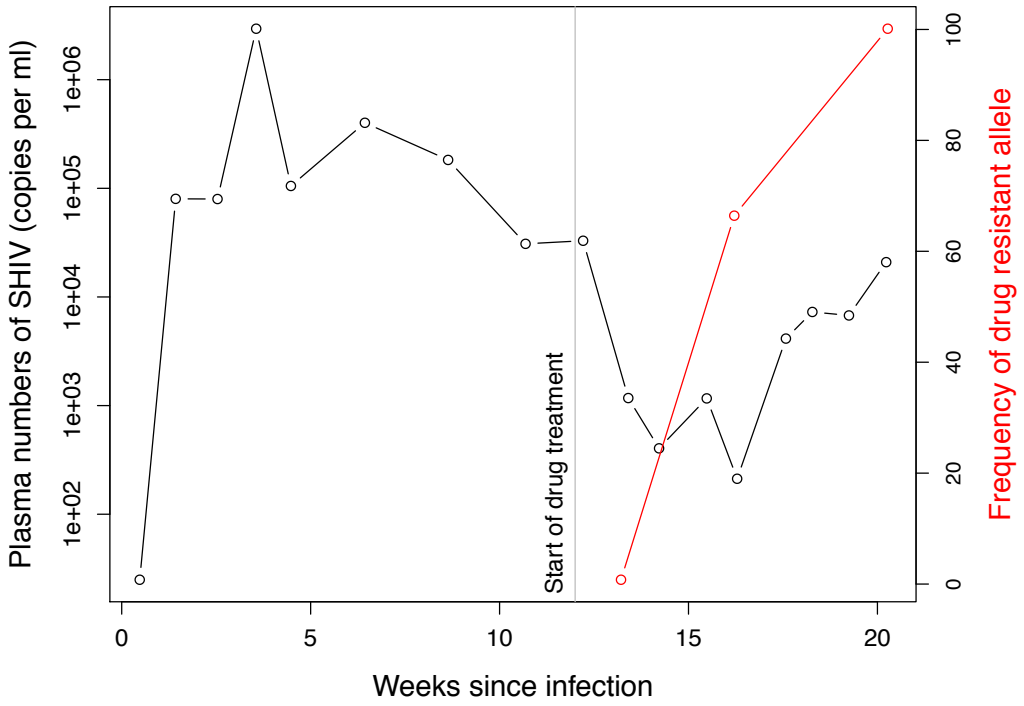
$$\overline{W}_t = W_1 \frac{P_t}{P_t + Q_t} + W_2 \frac{Q_t}{P_t + Q_t} = W_1 p_t + W_2 q_t, \qquad (10.1)$$

- This is the sum of the absolute fitness of the two allele types weighted by their relative frequencies

- Note that mean fitness is tied to time, because it reflects the frequency of genotypes at any given time

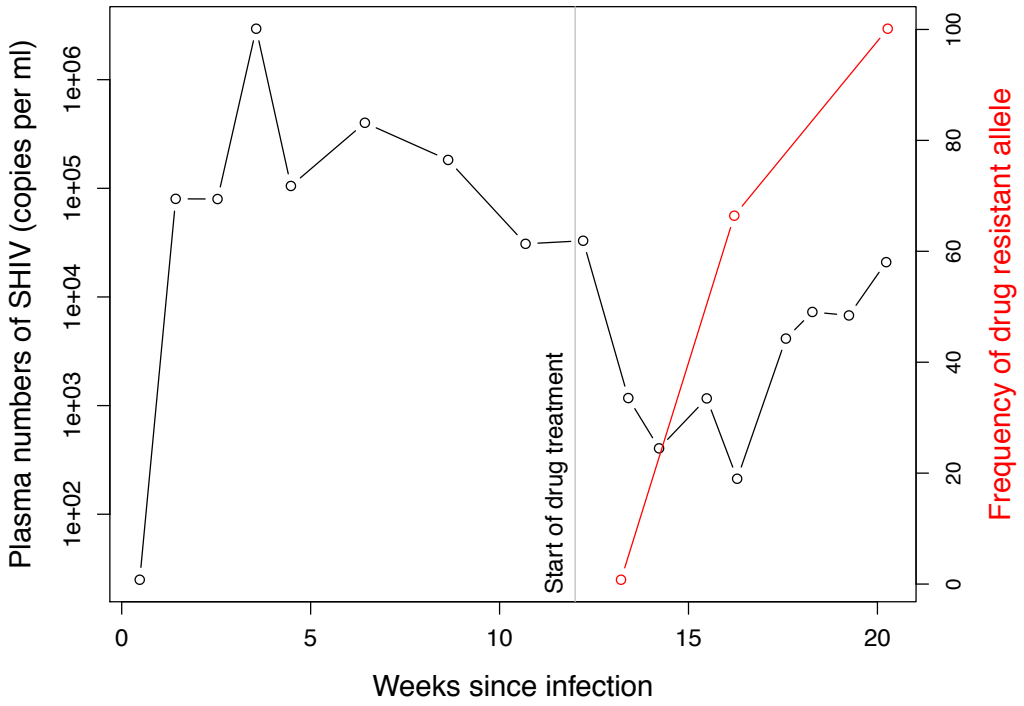
10.0.1 Haploid Selection Model

- Drug-resistant viruses are an excellent example of rapid evolution in a haploid system
- Feder and colleagues (2017) studied dynamics of the simian immunodeficiency virus (SHIV) in macaques
- 12 weeks after infection, the macaque was treated with an anti-retroviral drug
- After treatment, the viral load initially decreased as fitness of the original strain (A_2) was $W_2 < 1$ in the presence of the drug



10.0.1 Haploid Selection Model

- The virus population began to rebound upon the mutation of a new drug-resistant allele (A_1)
- Viruses with this allele have an absolute fitness of $W_1 > 1$
- Over the course of 7 weeks, the new drug-resistant allele became fixed in the virus population
- Rapid spread was driven by the much higher relative fitness of A_1 vs A_2 in the presence of the drug



10.0.1 Haploid Selection Model

- We can express the frequency of a given A_1 allele in the next generation as:

$$p_{t+1} = \frac{P_{t+1}}{P_{t+1} + Q_{t+1}} = \frac{W_1 P_t}{W_1 P_t + W_2 Q_t} = \frac{W_1 p_t}{W_1 p_t + W_2 q_t} = \frac{W_1}{\overline{W}_t} p_t. \quad (10.2)$$

- This tells us that the change in frequency of p only depends on the ratio of fitnesses
- Given this understanding, it is also common to scale absolute fitnesses by the absolute fitness of one of the genotypes (*e.g.*, the most or least fit genotype) to obtain relative fitness
- For example, if we scale by absolute fitness of A_1 then $w_1 = \frac{W_1}{W_1}$ and $w_2 = \frac{W_2}{W_1}$

10.0.1 Haploid Selection Model

- We can therefore also consider allele frequency across generations based on relative fitness:

$$p_{t+1} = \frac{w_1}{\overline{w}} p_t, \tag{10.3}$$

- And develop this as change in allele frequency:

$$\Delta p_t = p_{t+1} - p_t = \frac{w_1 p_t}{\overline{w}} - p_t = \frac{w_1 p_t - \overline{w} p_t}{\overline{w}} = \frac{w_1 p_t - (w_1 p_t + w_2 q_t) p_t}{\overline{w}} = \frac{w_1 - w_2}{\overline{w}} p_t q_t, \tag{10.4}$$

10.0.1 Haploid Selection Model

- If we make the assumption that fitness of our alleles remains constant over time, we can predict the number of our allelic types some generation τ after time 0 as
    - $P_\tau = {W_1}^\tau P_0$
    - $Q_\tau = {W_2}^\tau Q_0$
- The relative frequency of allele A_1 after τ generations then becomes:

$$p_\tau = \frac{(W_1)^\tau P_0}{(W_1)^\tau P_0 + (W_2)^\tau Q_0} = \frac{(w_1)^\tau P_0}{(w_1)^\tau P_0 + (w_2)^\tau Q_0} = \frac{p_0}{p_0 + (w_2/w_1)^\tau q_0}, \tag{10.5}$$

- where we've switched from absolute to relative frequencies and divided the whole last term by $(w_1)^\tau$

10.0.1 Haploid Selection Model

$$p_\tau = \frac{(W_1)^\tau P_0}{(W_1)^\tau P_0 + (W_2)^\tau Q_0} = \frac{(w_1)^\tau P_0}{(w_1)^\tau P_0 + (w_2)^\tau Q_0} = \frac{p_0}{p_0 + (w_2/w_1)^\tau q_0}, \tag{10.5}$$

- Equation 10.5 can be rearranged to:

$$\frac{p_\tau}{q_\tau} = \frac{p_0}{q_0}\left(\frac{w_1}{w_2}\right)^\tau . \tag{10.6}$$

- And then solved for τ:

$$\tau = \log\left(\frac{p_\tau q_0}{q_\tau p_0}\right) / \log\left(\frac{w_1}{w_2}\right) . \tag{10.7}$$

10.0.1 Haploid Selection Model

- Relative fitnesses (w_i) can also be parameterized by setting $w_1 = 1$ and $w_2 = 1 - s$ with s being known as the selection coefficient

- This allows us to simplify some of our previous equations, for example, 10.5 becomes:

$$p_{t+\tau} = \frac{p_t}{p_t + q_t(1-s)^\tau}, \qquad (10.8)$$

- Selection coefficients are often quite small and when $s \ll 1$, we can approximate $(1-s)^\tau$ in the denominator with e^{-st}:

$$p_{t+\tau} \approx \frac{p_t}{p_t + q_t e^{-s\tau}}. \qquad (10.9)$$

- Which is in the form of what is known as a logistic equation, often used to model population growth (we can think of our two alleles as populations that are growing or declining)

10.0.1 Haploid Selection Model

- We can also revisit equation 10.7 and substitute in our selection coefficient:

$$\tau = \log\left(\frac{p_\tau q_0}{q_\tau p_0}\right) / \log\left(\frac{w_1}{w_2}\right). \tag{10.7}$$

$$\tau = -\log\left(\frac{p_\tau q_0}{q_\tau p_0}\right) / \log\left(1-s\right). \tag{10.10}$$

- And then assuming again that $s \ll 1$:

$$\tau \approx \frac{1}{s}\log\left(\frac{p_\tau q_0}{q_\tau p_0}\right). \tag{10.11}$$

10.0.1 Haploid Selection Model

$$\tau \approx \frac{1}{s} \log \left(\frac{p_\tau q_0}{q_\tau p_0} \right). \qquad (10.11)$$

- One application of this equation might be wishing to know how long a new allele may take to reach fixation given a particular selective advantage

- Here, we assume the initial frequency, p_0, is $1/N$ and that fixation frequency, p_τ, is $1 - 1/N$ and after plugging these into equation 10.11 and making a few assumptions:

$$\begin{aligned} \tau &= \frac{1}{s} \log \left(\frac{1 - {}^2/_N + {}^1/_{N^2}}{{}^1/_{N^2}} \right) \\ &\approx \frac{1}{s} (\log(N) + \log(N-2)) \\ &\approx \frac{2}{s} \log(N) \qquad (10.12) \end{aligned}$$

10.0.1 Haploid Selection Model

- Up until now, we've been assuming that selection pressure and fitness of alleles stay the same over time, but this is clearly not always true

- We can make our equations more realistic by having our fitness linked to time; for example, saying $w_{1,t}$ and $w_{2,t}$ are relative fitnesses at time t

- The frequency of allele A_1 in generation $t+1$ then becomes:

$$p_{t+1} = \frac{w_{1,t}}{\overline{w}_t} p_t, \qquad (10.13)$$

10.0.1 Haploid Selection Model

- In this context, we can also consider the ratio of A_1 to A_2 at time $t+1$ as:

$$\frac{p_{t+1}}{q_{t+1}} = \frac{w_{1,t}}{w_{2,t}} \frac{p_t}{q_t}. \qquad (10.14)$$

- And, more generally, across time as:

$$\frac{p_\tau}{q_\tau} = \left(\prod_{i=1}^{\tau} \frac{w_{1,i}}{w_{2,i}} \right) \frac{p_1}{q_1}. \qquad (10.15)$$

- Based on this, the determination of which allele is increasing depends on whether $(\prod_{i=1}^{\tau} w_{1,i} / w_{2,i})$ is > 1 or < 1

10.0.1 Haploid Selection Model

- Thinking of the product of ratios over a given period of time can be a little daunting, but we can find a simple take home by taking the τ^{th} root of equation 10.15:

$$\sqrt[\tau]{\left(\prod_{i=1}^{\tau} \frac{w_{1,i}}{w_{2,i}}\right)} = \frac{\sqrt[\tau]{\prod_{i=1}^{\tau} w_{1,i}}}{\sqrt[\tau]{\prod_{i=1}^{\tau} w_{2,i}}}. \qquad (10.16)$$

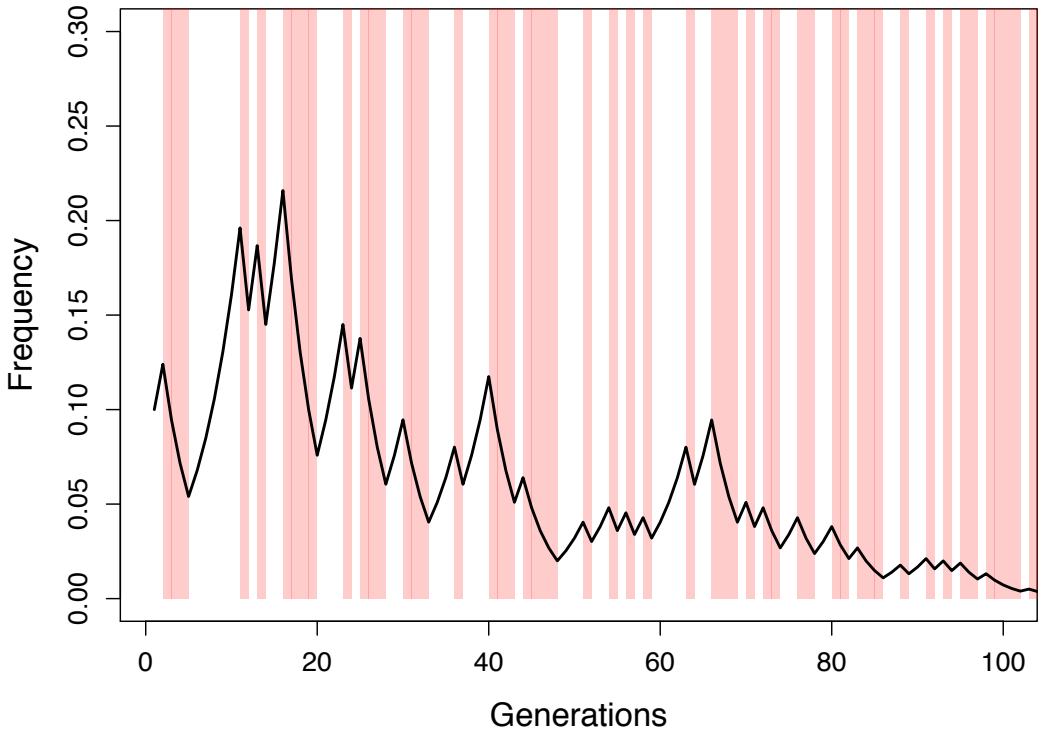
- and realizing that $\sqrt[\tau]{\prod_{i=1}^{\tau} w_{1,i}}$ is the geometric mean fitness of allele A_1 over τ generations

- This tells us that allele A_1 will increase if it has a higher geometric mean fitness than allele A_2

10.0.1 Haploid Selection Model

- In fact, an allele will increase over another allele when its geometric mean fitness is higher even when its arithmetric mean fitness is not higher

- As an example, let's consider a locus with alleles A_1 and A_2 where allele A_1 does better in dry years but suffers in wet years and A_2 is a generalist that does equally across conditions

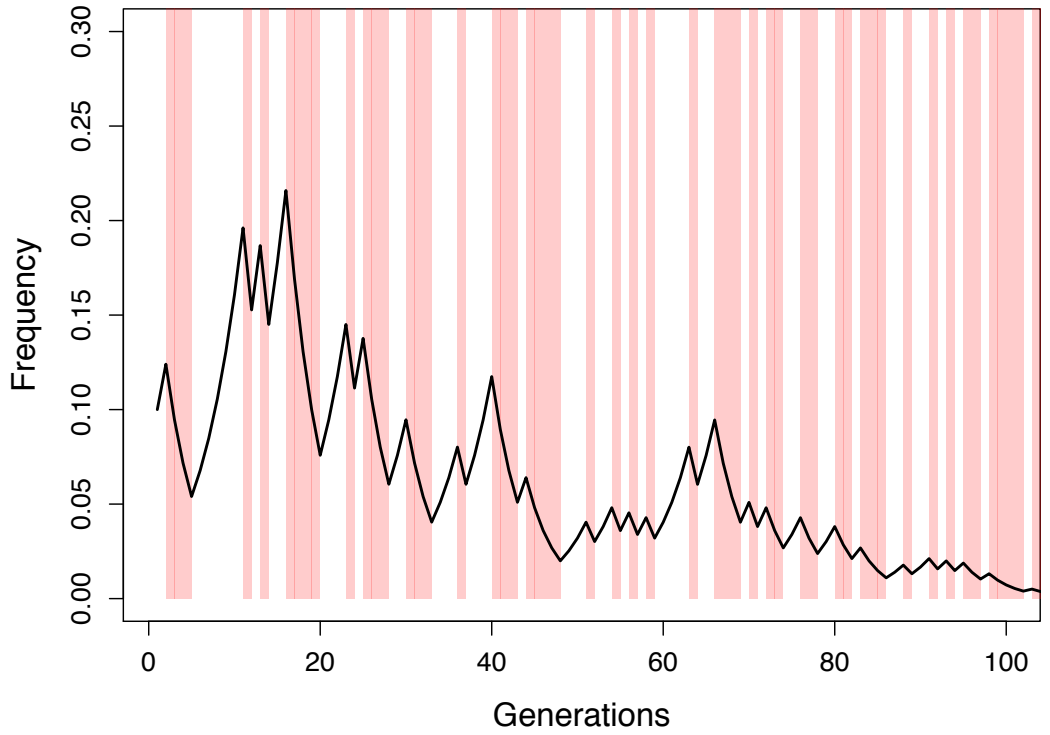


10.0.1 Haploid Selection Model

- For example, we can propose the following fitness values:

	A_1	A_2
Dry	2	1.57
Wet	1.16	1.57
Arithmetic Mean	1.58	1.57
Geometric Mean	1.52	1.57

- If wet and dry years happen with equal frequency, the A_2 allele will rise to fixation despite a lower arithmetic mean fitness because its geometric mean fitness is higher

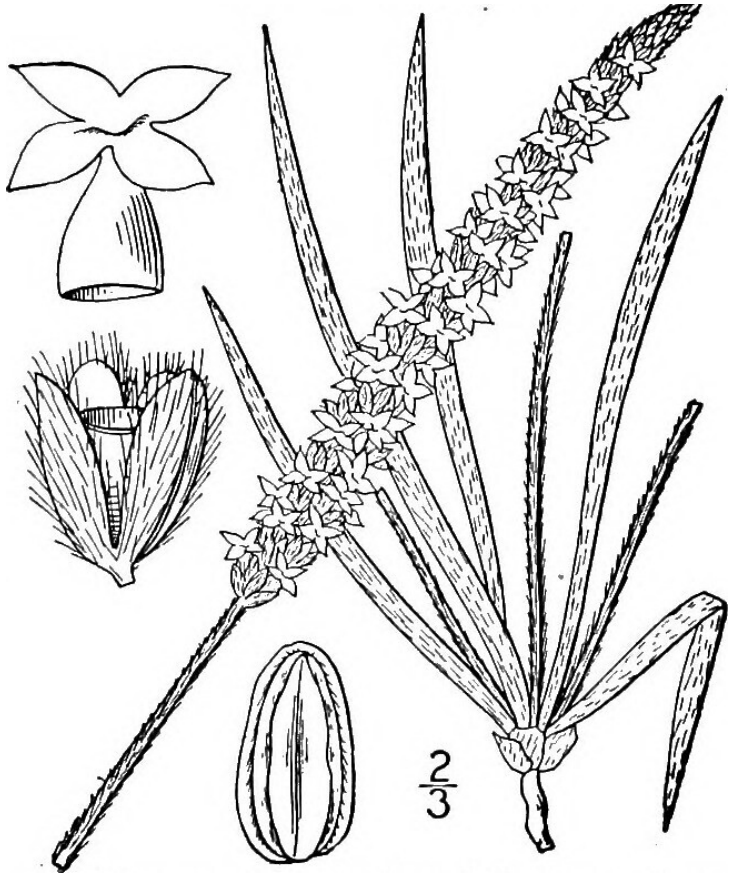


10.0.1 Haploid Selection Model

- Based on further development of expectations using the arithmetic and geometric means, in section 10.0.1 (we'll skip the mathematical details here), we can explore the concept of "bet hedging" in evolution
- Just as in managing an investment portfolio, evolution can be risk averse when there is high variance in the selection environment
- For example, birds may lay eggs in multiple nests to spread risk across multiple nesting sites
- Genotypes with a high arithmetic mean fitness will be selected against due to a low geometric mean fitness when variance is high across generations

10.0.1 Haploid Selection Model

- Some clear examples of bet hedging happens in Sonoran desert plants investigated by Gremer and Venable (2014)
- In highly variable environments like the desert it's a good strategy for plants to have only a portion of their seeds germinate the year after they are produced
- The trade-off of delayed germination is that this does come with some seed mortality
- Plants, like the woolly plantain (right), with highly variable yield from year to year also had the lowest fraction of germination each year
- This strategy maximized geometric fitness

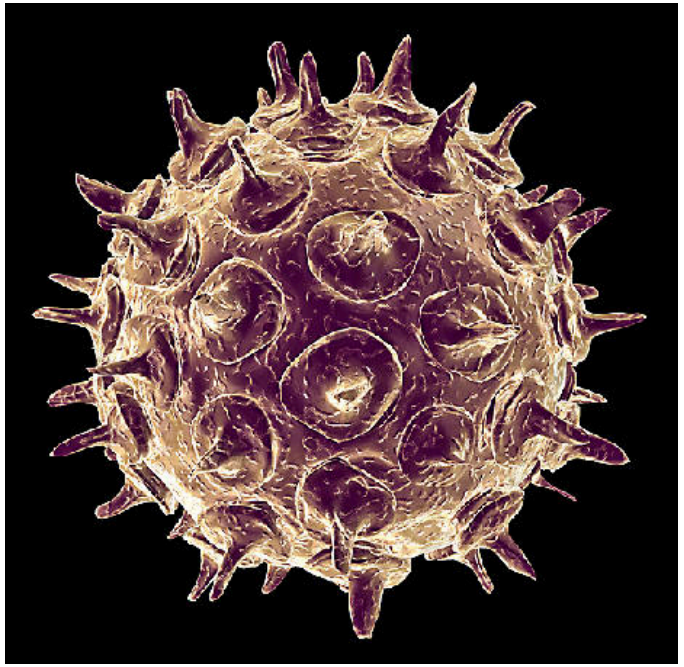


10.0.1 Haploid Selection Model

- Bet-hedging is also a frequently used strategy in micro-organisms like the varicella zoster virus that causes chicken pox.

- After causing chicken pox, the virus enters a latent phase in the neurons of the spinal cord and can emerge 5-40 years later, causing the disease known as shingles

- This strategy can help the virus encounter a greater number of susceptible hosts in the future



Source: Microbe Wiki
Contributed by Bryce S. Naberhaus

Coop, Chapter 10: 10.0.2-10.0.3

One-Locus Models of Selection

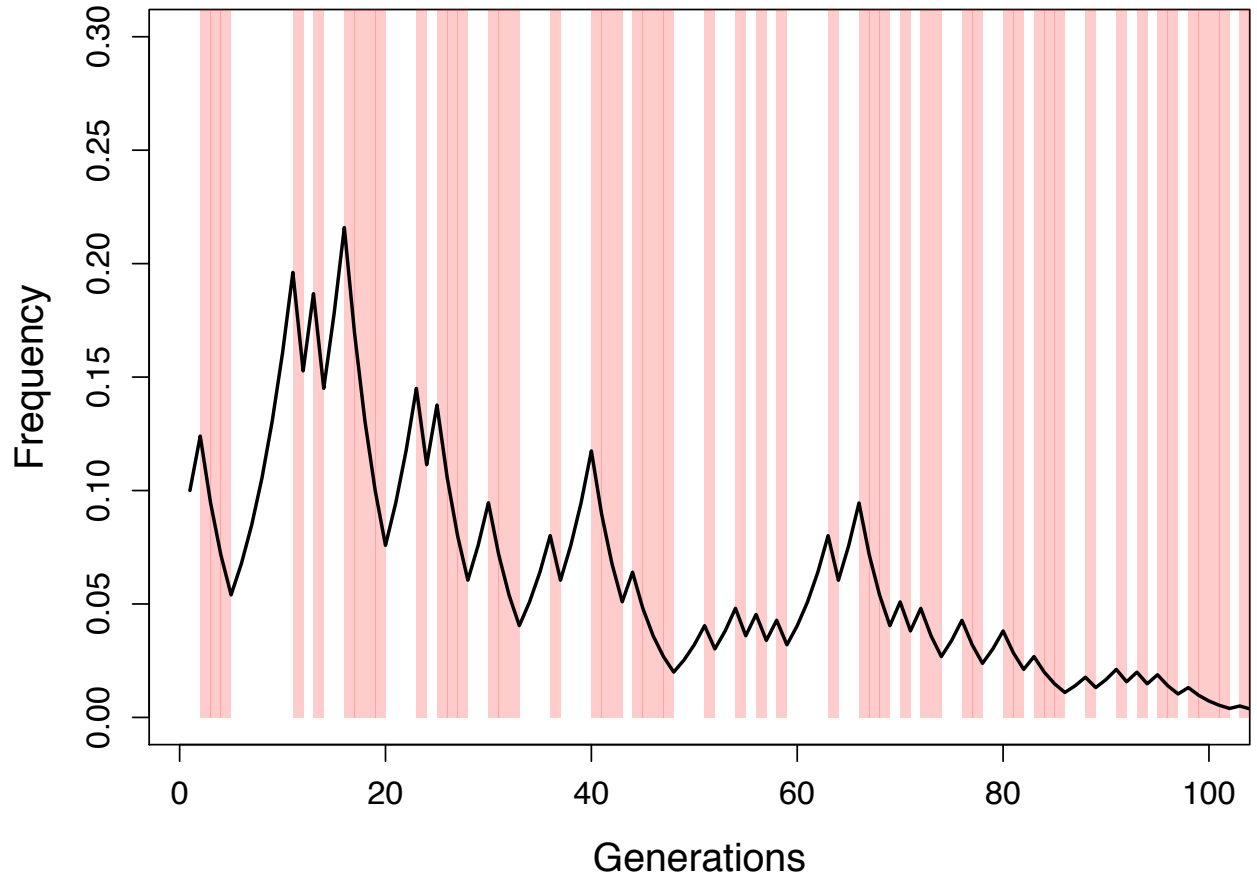
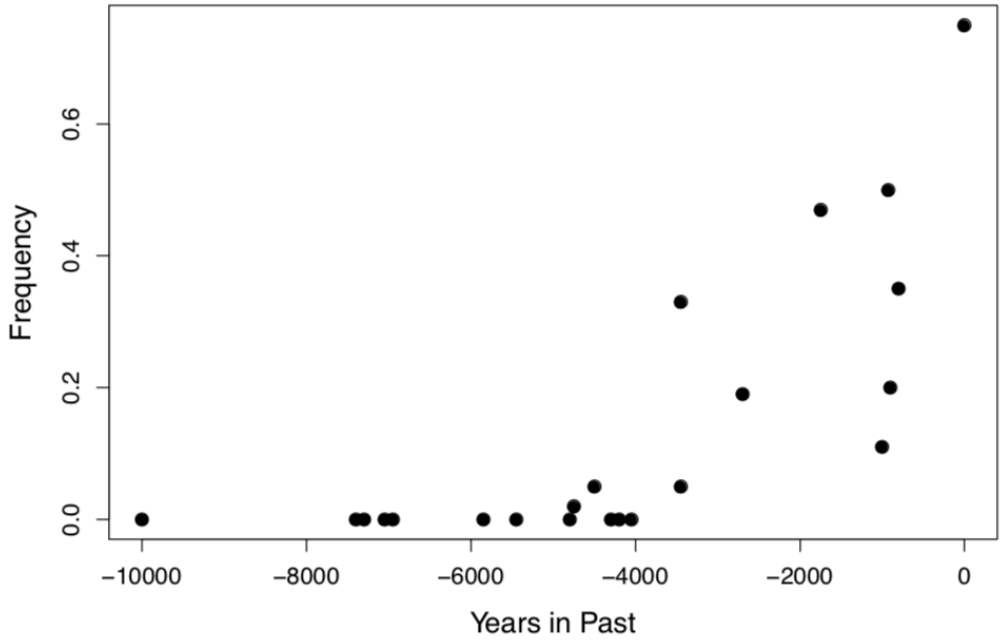
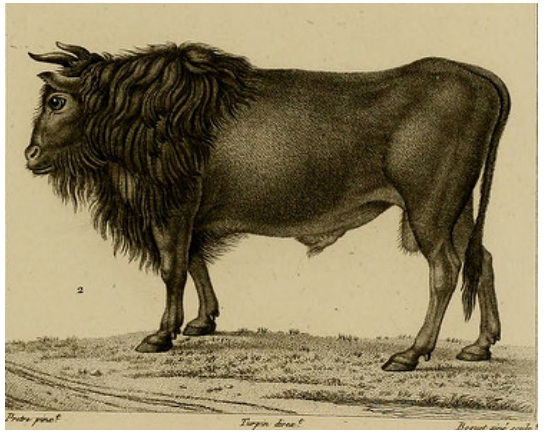


Figure 10.2: An example frequency trajectory of the A_1 allele under variable environments (using the fitnesses from Table 10.1). Wet years (generations) are shown in red, dry years in white. The environment flips at random each year. Note how the A_1 allele increases in frequency in the dry years as it has higher fitness, and yet the A_2 allele still wins out. Code here.

10.0.2 Diploid Model

- In this section we consider selection in diploid organisms at a single locus
- As an example of selection in a diploid organism, we can look at one of the best-documented instances of selection in humans: increased lactase persistence
- In most mammals, the lactase protein is no longer produced after childhood, but in multiple, independent human populations that have raised cattle and consumed milk, there is persistence of lactase production into adulthood



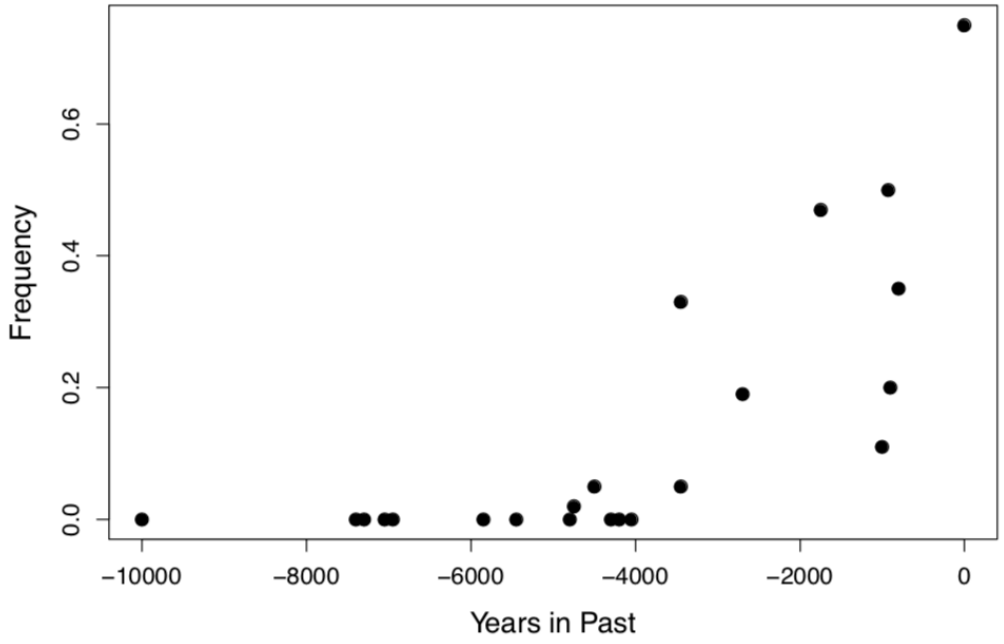


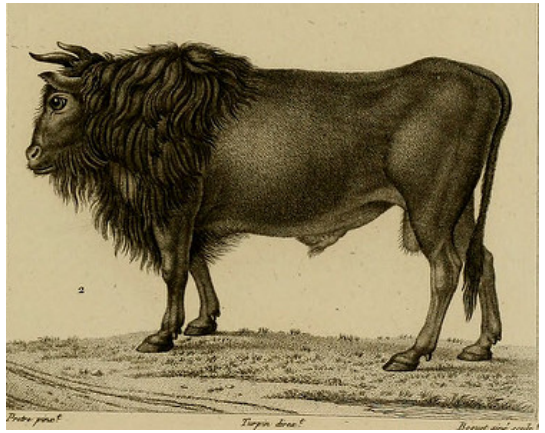
10.0.2 Diploid Model

- Production of lactase allows humans to break down lactose, the main carbohydrate in milk, and to benefit nutritionally, thereby increasing fitness

- Recent improvements in DNA sequencing technology have allowed for sequencing ancient DNA and tracking frequency of the lactase persistence allele

- The allele was absent 5,000 years ago, but is now at 70% frequency in certain European populations





10.0.2 Diploid Model

- To understand how selection occurs in diploid organisms, let's again develop models based on fitness
- In this section, we will focus on viability fitness, the differential number across genotypes of individuals that survive from zygotes until the time of reproduction
- We will denote the absolute fitnesses of genotypes A_1A_1, A_1A_2, and A_2A_2 as W_{11}, W_{12}, and W_{22}
- For example, W_{11} is the probability that an individual with the A_1A_1 survives to reproduce
- With random mating, our number of individuals with each genotype at time t will be:

$$Np_t^2, \quad N2p_tq_t, \quad Nq_t^2. \qquad (10.21)$$

10.0.2 Diploid Model

- The mean fitness of our population of zygotes will be:

$$\overline{W}_t = W_{11}p_t^2 + W_{12}2p_tq_t + W_{22}q_t^2. \qquad (10.22)$$

- And, if we want to know the probable number of individuals with each genotype that survive to reproduce:

$$NW_{11}p_t^2, \quad NW_{12}2p_tq_t, \quad NW_{22}q_t^2. \qquad (10.23)$$

- Which, in a straight-forward extension tells us the total number of individuals surviving to reproduction:

$$N\left(W_{11}p_t^2 + W_{12}2p_tq_t + W_{22}q_t^2\right). \qquad (10.24)$$

- And, even more simply: $N\overline{W}$

10.0.2 Diploid Model

- The relative frequency of individuals with each genotype that have survived to reproduce can then be summarized as:

$$\frac{NW_{11}p_t^2}{N\overline{W}}, \quad \frac{NW_{12}2p_tq_t}{N\overline{W}}, \quad \frac{NW_{22}q_t^2}{N\overline{W}} \tag{10.25}$$

- With all of this conveniently summarized in Table 10.2

	A_1A_1	A_1A_2	A_2A_2
Absolute no. at birth	Np_t^2	$N2p_tq_t$	Nq_t^2
Fitnesses	W_{11}	W_{12}	W_{22}
Absolute no. at reproduction	$NW_{11}p_t^2$	$NW_{12}2p_tq_t$	$NW_{22}q_t^2$
Relative freq. at reproduction	$\frac{W_{11}}{\overline{W}}p_t^2$	$\frac{W_{12}}{\overline{W}}2p_tq_t$	$\frac{W_{22}}{\overline{W}}q_t^2$

10.0.2 Diploid Model

- Since we are only considering viability selection and not fecundity (*i.e.*, number of offspring), then frequency of, for example, the A_1 allele in the next generation is simply:

$$p_{t+1} = \frac{W_{11}p_t^2 + W_{12}p_t q_t}{\overline{W}}. \qquad (10.26)$$

- Each genotypic class responds to selection based on its fitness relative to the mean fitness of the population (*e.g.*, $W_{11}/\overline{W}$)

- We can estimate this fitness ratio for any genotype by looking at its frequency at birth relative to adulthood

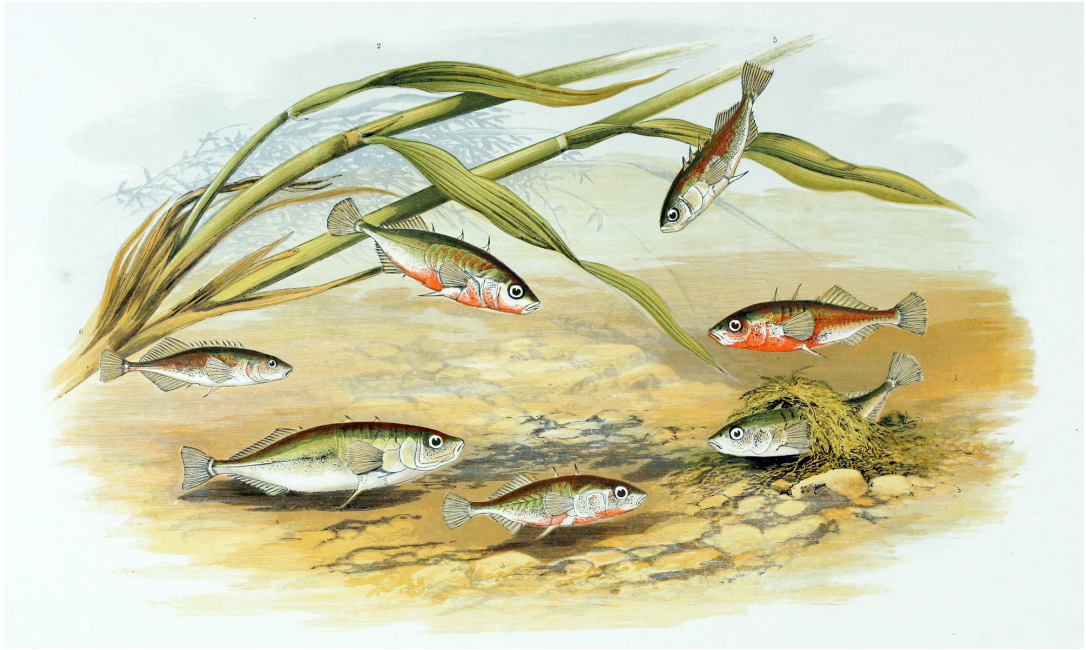
- Let's revisit our good friends the sticklebacks for an example…

10.0.2 Diploid Model

- As glaciers receded following the last ice age, stickleback colonized freshwater lakes a number of times from marine environments

- As you may remember, marine sticklebacks have substantial body armor that protects them from predation in this environment, but freshwater stickleback have repeatedly lost body armor through selection on the *Ectodysplasin* (*EDA*) gene

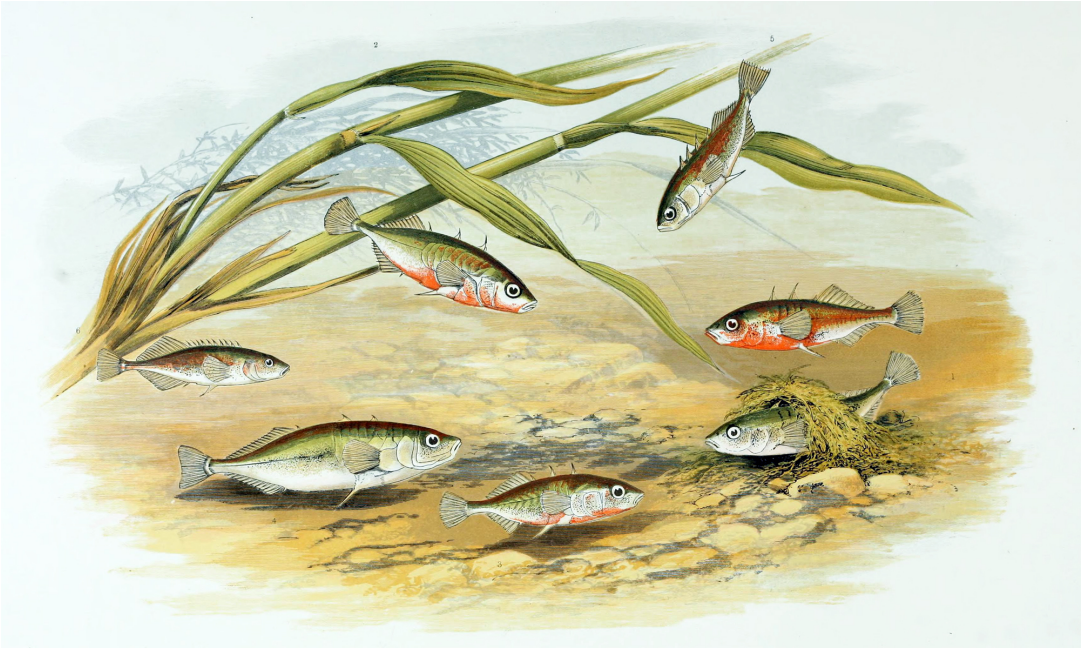
- The low armor allele (L) at this locus is found at low frequency in marine environments so selection has been on standing variation



10.0.2 Diploid Model

- Barrett and colleagues (2008) conducted an experiment in which they bred, in the lab, populations of stickleback segregating for the low armor (L) and high armor (C) alleles and introduced these into freshwater ponds

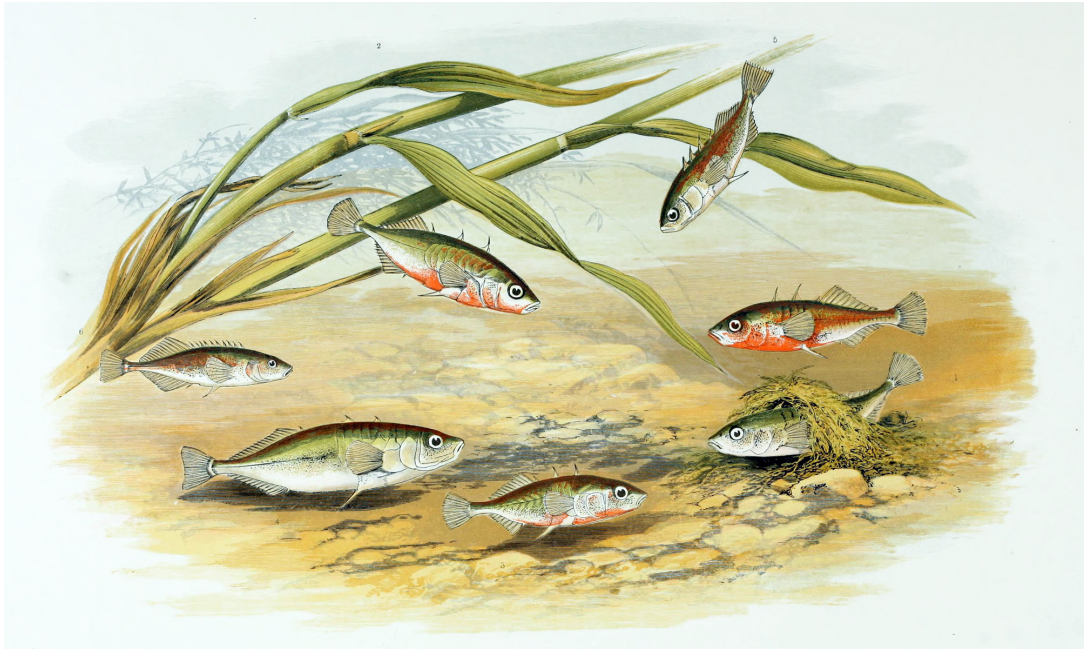
- Genotype frequencies were then measured over their lifetime:



	CC	LC	LL
Juveniles	0.55	0.23	0.22
Adults	0.21	0.53	0.26
Adults/Juv. $(W_\bullet/\overline{W})$	0.4	2.3	1.2
rel. fitness $(W_\bullet/W_{12})$	0.17	1.0	0.54

10.0.2 Diploid Model

	CC	LC	LL
Juveniles	0.55	0.23	0.22
Adults	0.21	0.53	0.26
Adults/Juv. $(W_\bullet/\overline{W})$	0.4	2.3	1.2
rel. fitness $(W_\bullet/W_{12})$	0.17	1.0	0.54



- Heterozygotes increased dramatically due to their high fitness relative to the mean of the population

- Looking at the relative fitness of the different genotypes, the CC homozygotes had a value ~ 1/5 of heterozygotes

10.0.2 Diploid Model

- The absolute value of fitness is irrelevant to the frequency of an allele and, when considering allele frequencies across time, we can replace absolute with relative fitnesses

$$\Delta p_t = p_{t+1} - p_t = \frac{w_{11}p_t^2 + w_{12}p_t q_t}{\overline{w}} - p_t. \qquad (10.27)$$

- To simplify this equation, we will define two variables:

$$\overline{w}_1 = w_{11}p_t + w_{12}q_t, \qquad (10.28)$$

$$\overline{w}_2 = w_{12}p_t + w_{22}q_t. \qquad (10.29)$$

- $\overline{w}_1$ and $\overline{w}_2$ are what are known as the “marginal fitnesses” of alleles A_1 and A_2 and represent the average fitness of these alleles

10.0.2 Diploid Model

- For example, the fitness of the A_1 allele would be its fitness as a homozygote weighted by the probability that it is in a homozygote (p_t) plus its fitness as a heterozygote weighted by the probability that it is in a heterozygote (q_t)
- Mean relative fitness can be expressed in terms of marginal fitnesses as:

$$\overline{w} = \overline{w}_1 p_t + \overline{w}_2 q_t, \qquad (10.30)$$

- And, in this context, change in allele frequencies can be expressed as:

$$\Delta p_t = \frac{(\overline{w}_1 - \overline{w}_2)}{\overline{w}} p_t q_t. \qquad (10.31)$$

10.0.2 Diploid Model

$$\Delta p_t = \frac{(\overline{w}_1 - \overline{w}_2)}{\overline{w}} p_t q_t. \qquad (10.31)$$

- From 10.31, we can observe that whether A_1 increases or decreases, depends only on $\overline{w}_1 - \overline{w}_2$
- A_1 will rise in frequency as long as its marginal fitness is higher than A_2

10.0.2 Diploid Model

Question 3. For many generations you have been studying an annual wildflower that has two color morphs, orange and white. You have discovered that a single bi-allelic locus controls flower color, with the white allele being recessive. The pollinator of these plants is an almost blind bat, so individuals are pollinated at random with respect to flower color. Your population census of 200 individuals showed that the population consisted of 168 orange-flowered individuals, and 32 white-flowered individuals.

Heavy February rainfall creates optimal growing conditions for an exotic herbivorous beetle with a preference for orange-flowered individuals. This year it arrives at your study site with a ravenous appetite. Only 50% of orange-flowered individuals survive its wrath, while 90% of white-flowered individuals survive until the end of the growing season.

A) What is the initial frequency of the white allele, and what do you have to assume to obtain this?

B) What is the frequency of the white allele in the seeds forming the next generation?

For part A, $p^2 = \frac{32}{200} = 0.16$

$$p = \sqrt{0.16} = 0.4$$

$$q = 1 - p = 0.6$$

10.0.2 Diploid Model

Question 3. For many generations you have been studying an annual wildflower that has two color morphs, orange and white. You have discovered that a single bi-allelic locus controls flower color, with the white allele being recessive. The pollinator of these plants is an almost blind bat, so individuals are pollinated at random with respect to flower color. Your population census of 200 individuals showed that the population consisted of 168 orange-flowered individuals, and 32 white-flowered individuals.

Heavy February rainfall creates optimal growing conditions for an exotic herbivorous beetle with a preference for orange-flowered individuals. This year it arrives at your study site with a ravenous appetite. Only 50% of orange-flowered individuals survive its wrath, while 90% of white-flowered individuals survive until the end of the growing season.

A) What is the initial frequency of the white allele, and what do you have to assume to obtain this?

B) What is the frequency of the white allele in the seeds forming the next generation?

For part B:

$$p_{t+1} = \frac{W_{11}p_t^2 + W_{12}p_t q_t}{\overline{W}}. \qquad (10.26)$$

$W_{11} = 0.9 \qquad W_{12} = 0.5 \qquad W_{22} = 0.5$

$w_{11} = 1 \qquad w_{12} = 0.56 \qquad w_{22} = 0.56$

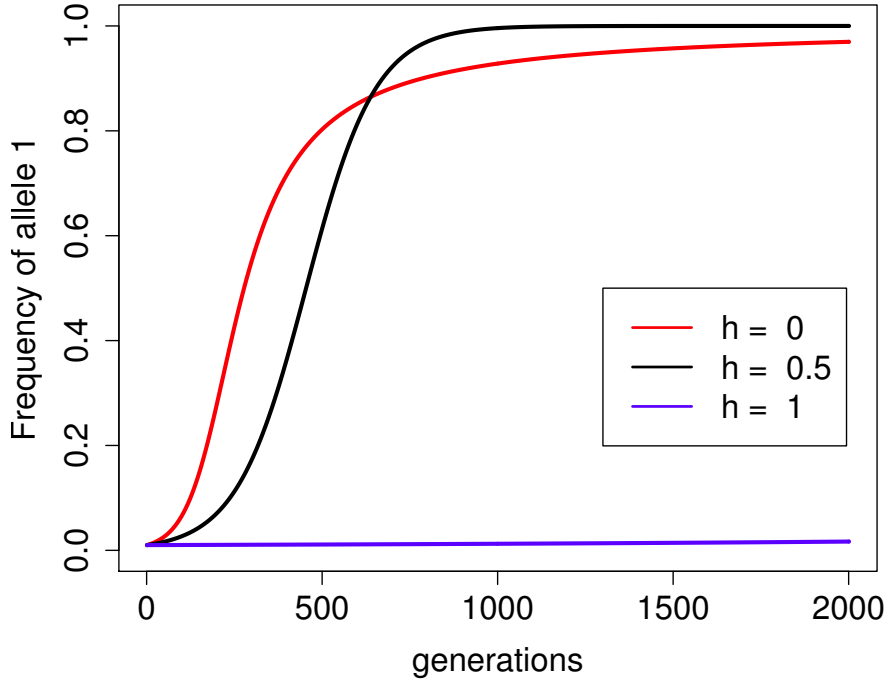
$$\overline{W}_t = W_{11}p_t^2 + W_{12}2p_t q_t + W_{22}q_t^2. \qquad (10.22)$$

$$\frac{1\times0.4^2 + 0.56\times0.4\times0.6}{1\times0.4^2 + 0.56\times2\times0.4\times0.6 + 0.56\times0.6^2}$$

0.2944/.6304=0.467

10.0.3 Diploid Model

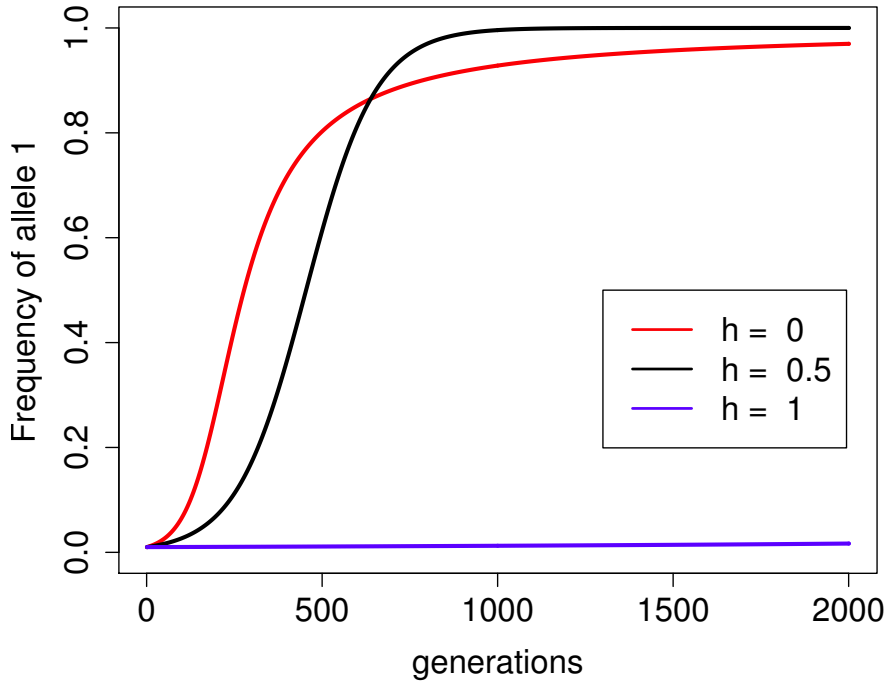
- Coming back to directional selection in the context of diploid selection, we'll see that one allele always has a larger marginal fitness
- For example, if the A_1 allele is most fit, then:
  - $\overline{w}_1 > \overline{w}_2$
- And, for mean fitnesses:
  - $w_{11} > w_{12} > w_{22}$



10.0.3 Diploid Model

- In diploid models, like haploid models, we can also parameterize reduction in fitness using a selection coefficient:

genotype	A_1A_1	A_1A_2	A_2A_2
absolute fitness	W_{11}	$\geq W_{12} \geq$	W_{22}
relative fitness (generic)	$w_{11} = W_{11}/W_{11}$	$w_{12} = W_{12}/W_{11}$	$w_{22} = W_{22}/W_{11}$
relative fitness (specific)	1	$1 - sh$	$1 - s.$

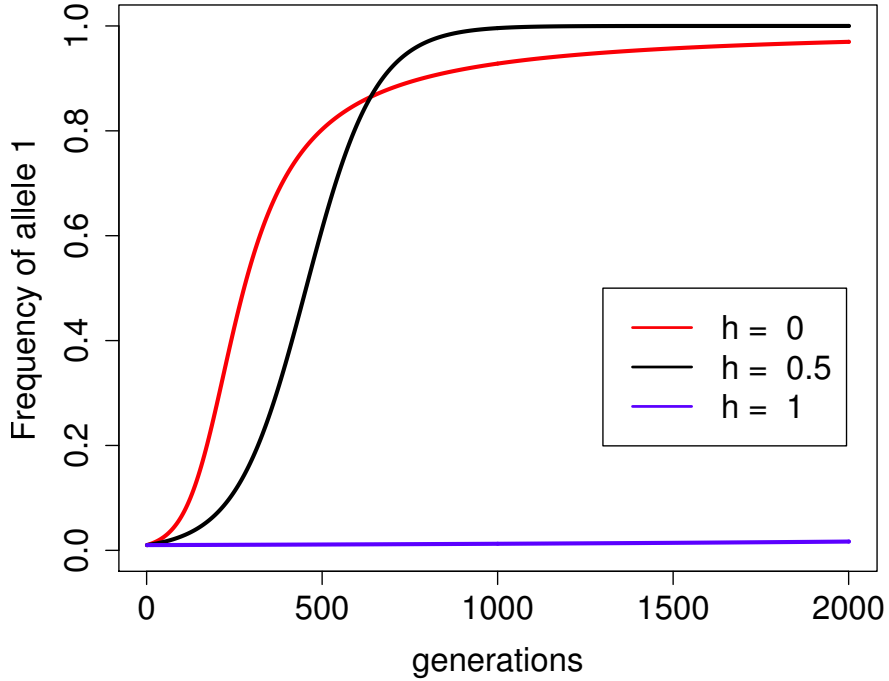


- In diploids, the selection coefficient, s, is the difference in fitness between the two homozygotes

- h is the dominance coefficient and, under directional selection, $0 \leq h \leq 1$

10.0.3 Diploid Model

- When $h = 0$, the A_1 allele is completely dominant and the A_1A_1 homozygote and A_1A_2 heterozygote have equal fitness
- When $h = 1$, the A_1 allele is completely recessive to the A_2 allele
- Our equations for change in frequency across generations and mean fitness with these selection coefficients become:



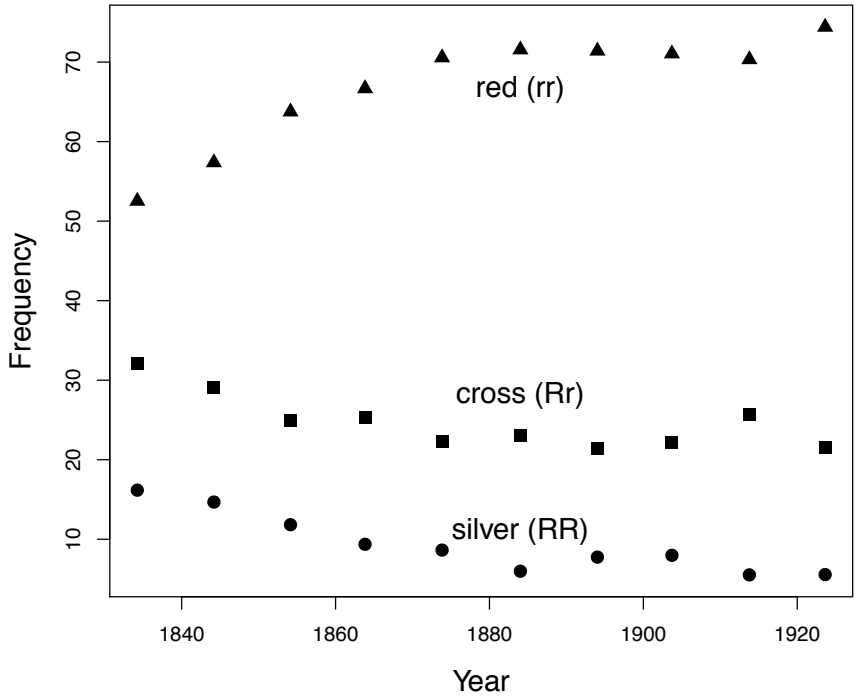
$$\Delta p_t = \frac{p_t h s + q_t s(1-h)}{\overline{w}} p_t q_t, \qquad (10.33)$$

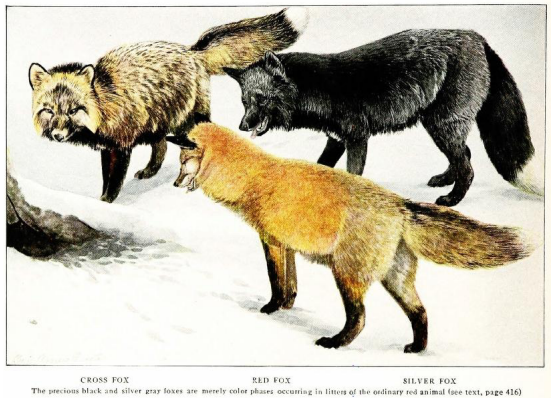
where

$$\overline{w} = 1 - 2p_t q_t s h - q_t^2 s. \qquad (10.34)$$

10.0.3 Diploid Model

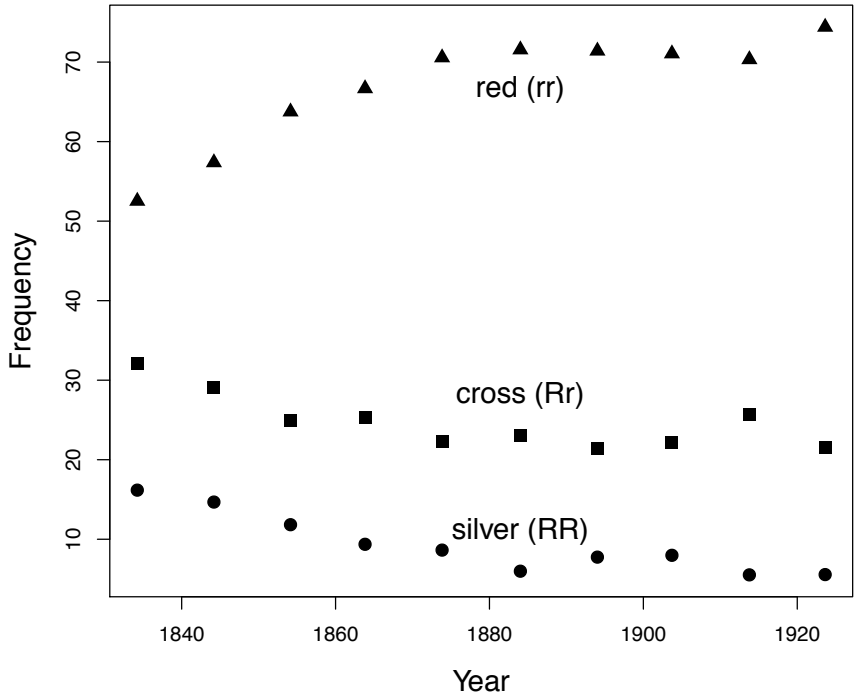
- As an example of allele trajectories under dominance, red foxes have 3 color morphs—silver, cross, and red—determined by a single polymorphism with genotypes RR, Rr, and rr
- Historically, the pelts of silver foxes were highly valued, resulting in strong selection against this particular morph
- Over the period from 1834-1937, the silver fox proportion of the population declined from 16% to 5%

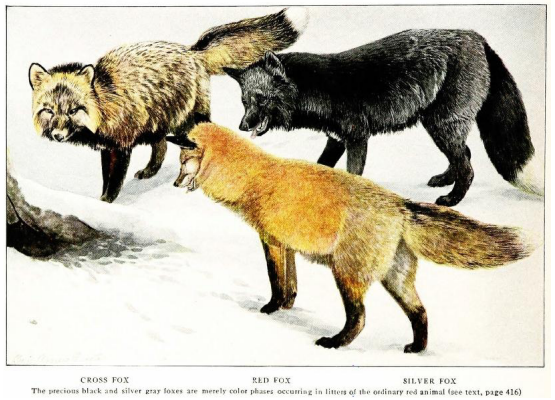




10.0.3 Diploid Model

- Analysis of population data revealed recessive selection against the silver morph
- RR homozygotes declined substantially over this period of time, but the R allele is effectively masked in heterozygotes and these decline less quickly





10.0.3 Diploid Model

- A special case of diploid, directional selection is when $h = 0.5$, in which case the interaction among alleles with respect to fitness is additive

- We can then simplify equation 10.33 by removing the dominance coefficient:

$$\Delta p_t = \frac{1}{2}\frac{s}{\overline{w}} p_t q_t. \qquad (10.35)$$

- And if we assume that $s \ll 1$ and therefore mean fitness, $\overline{w} \approx 1$:

$$\Delta p_t = \frac{1}{2} s p_t q_t. \qquad (10.36)$$

10.0.3 Diploid Model

$$\Delta p_t = \frac{1}{2} s p_t q_t. \qquad (10.36)$$

- If we return to our similar equation in haploids (10.4), set $w_1 = 1$ and $w_2 = 1 - s$ and again assume that $s \ll 1$:

$$\Delta p_t = p_{t+1} - p_t = \frac{w_1 p_t}{\overline{w}} - p_t = \frac{w_1 p_t - \overline{w} p_t}{\overline{w}} = \frac{w_1 p_t - (w_1 p_t + w_2 q_t) p_t}{\overline{w}} = \frac{w_1 - w_2}{\overline{w}} p_t q_t, \qquad (10.4)$$

- We'll see that $\Delta p_t = s p_t q_t$, and that, without dominance, the diploid model is the same as the haploid model up to a factor of $\frac{1}{2}$, a difference due to parameterization of the diploid model

10.0.3 Diploid Model

- With no dominance, we can also formulate very similar models to our haploid models of the time, τ, for our allele A_1 to go from frequency of p_0 to p_τ :

$$\tau \approx \frac{2}{s} \log\left(\frac{p_\tau q_0}{q_\tau p_0}\right) \qquad (10.37)$$

- and, similarly, in a diploid, no dominance model, to transit from new mutation to fixation in a population:

$$\tau \approx \frac{4}{s} \log(2N) \qquad (10.38)$$

Coop, Chapter 10: 10.1-10.1.1

One-Locus Models of Selection

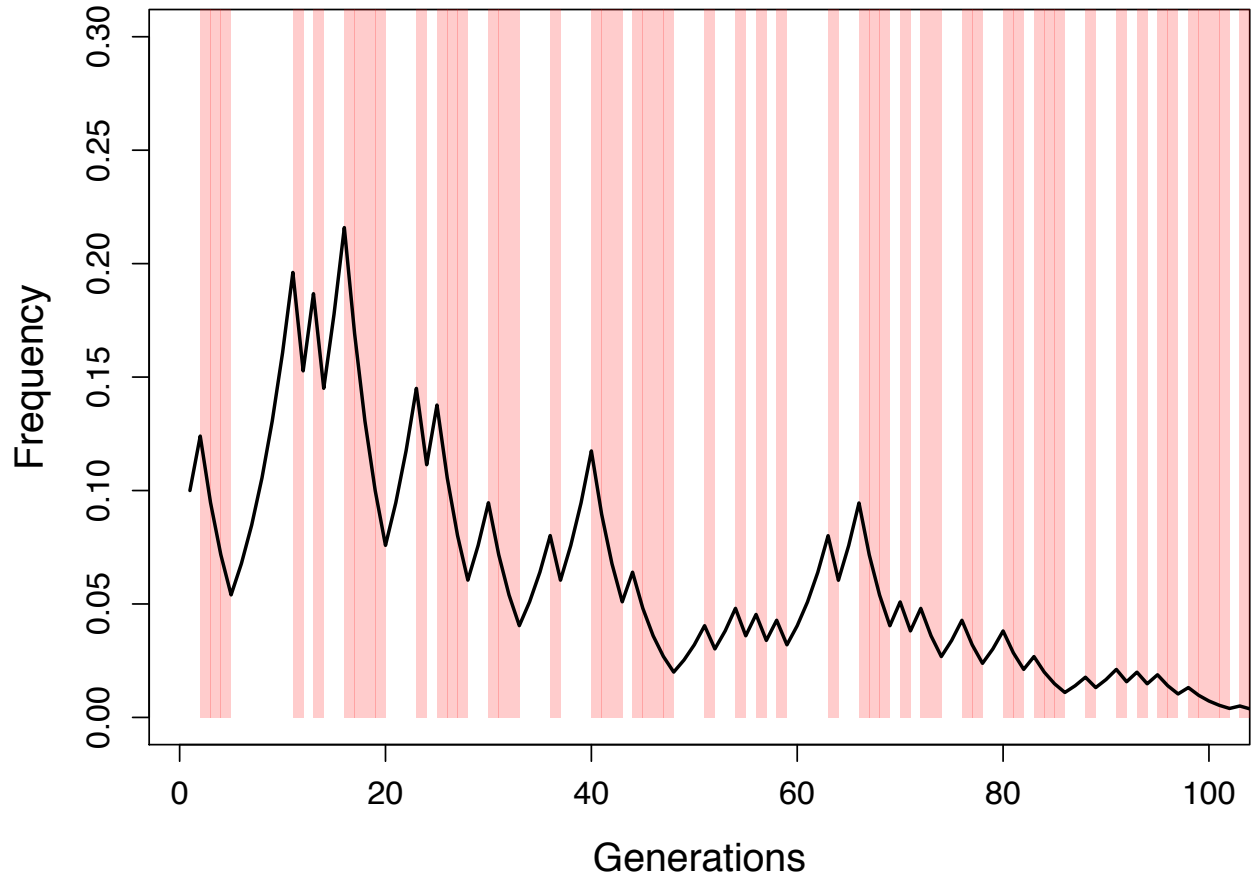


Figure 10.2: An example frequency trajectory of the A_1 allele under variable environments (using the fitnesses from Table 10.1). Wet years (generations) are shown in red, dry years in white. The environment flips at random each year. Note how the A_1 allele increases in frequency in the dry years as it has higher fitness, and yet the A_2 allele still wins out. Code here.

10.1 Balancing selection and selective maintenance of polymorphism

While directional selection should act to remove variation from a population, we see plentiful phenotypic and genotypic variation because:

1. Variation is maintained through a balance of genetic drift (removing variation) and mutation (adding variation)

2. Selection acts to maintain variation (*e.g.*, balancing selection)

3. Deleterious variation persists through a balance of natural selection (removing variation) and mutation (adding variation)

A remaining challenge for population genomics is understanding the extent to which each force shapes diversity

10.1.1 Heterozygote Advantage

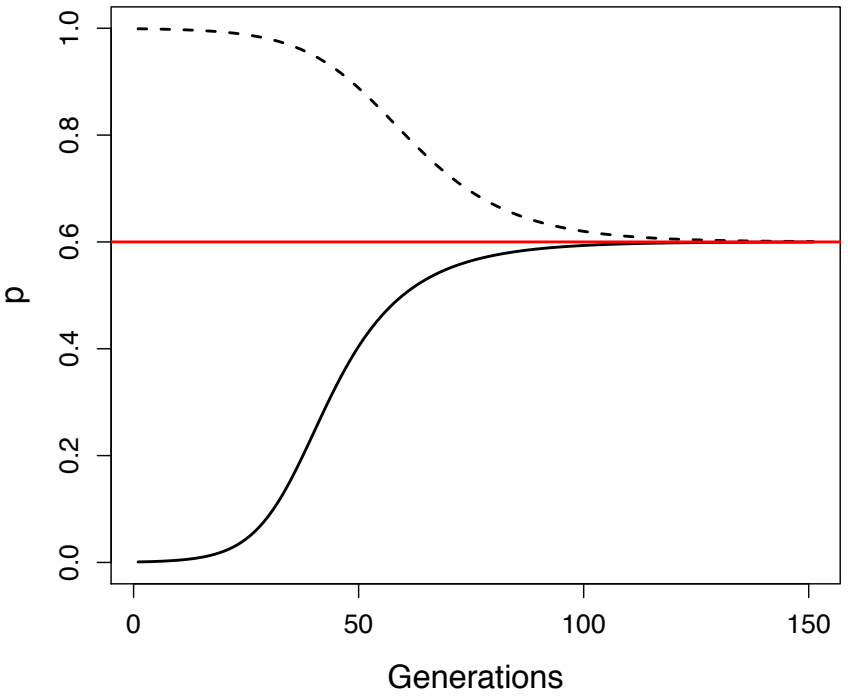
- Balancing selection can be observed when the heterozygous genotype has greater fitness than either of the homozygous genotypes:

genotype	A_1A_1	A_1A_2	A_2A_2
absolute fitness	w_{11}	$< w_{12} >$	w_{22}
relative fitness (generic)	$w_{11} = W_{11}/W_{12}$	$w_{12} = W_{12}/W_{12}$	$w_{22} = W_{22}/W_{12}$
relative fitness (specific)	$1 - s_1$	1	$1 - s_2$

- Note our new selection coefficients, s_1 and s_2, which are the relative differences in the fitness of the homozygotes from the heterozygotes

10.1.1 Heterozygote Advantage

- When the A_1 allele is rare, it is typically found as a heterozygote, making it the more fit allele which will increase in frequency
- The same is true when A_2 is the more rare allele
- In this way, neither allele can reach fixation
- Both alleles will be maintained at equilibrium frequencies as a balanced polymorphism within the population



10.1.1 Heterozygote Advantage

- By going back to 10.31, our equation for Δp, we can see there are three equilibria where $\Delta p = 0$:

$$\Delta p_t = \frac{(\overline{w}_1 - \overline{w}_2)}{\overline{w}} p_t q_t. \qquad (10.31)$$

- When p_t or q_t or $\overline{w}_1 - \overline{w}_2 = 0$, an equilibrium will be obtained; the first two are trivial because they are monomorphic, but the third is a stable polymorphic equilibrium

- Using our new selection coefficients (s_1 and s_2), our marginal fitnesses will be equal, and the polymorphic equilibrium will be obtained when:

$$p_e = \frac{s_2}{s_1 + s_2} \qquad (10.39)$$

10.1.1 Heterozygote Advantage

$$p_e = \frac{s_2}{s_1 + s_2} \qquad (10.39)$$

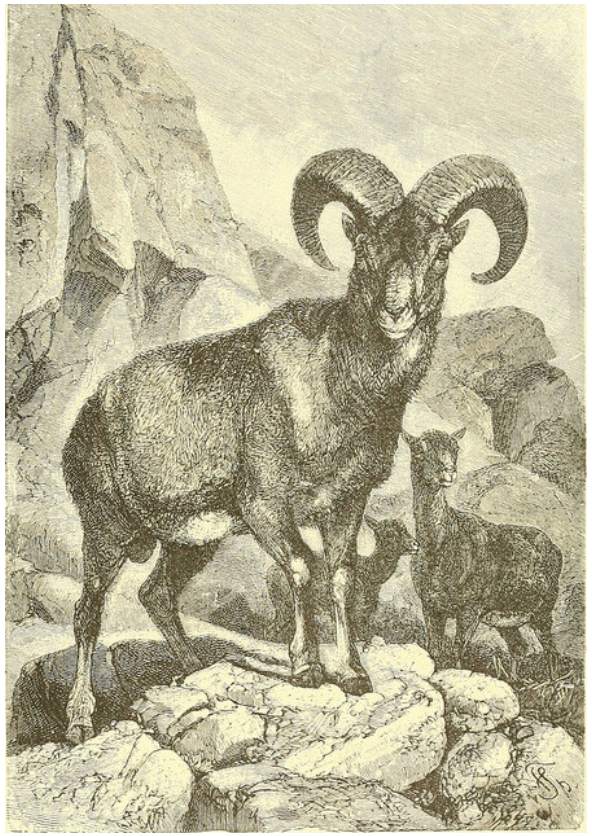
- At this equilibrium frequency of A_1 the mean fitness of the population will be maximized
- While the greatest fitness would be achieved if all individuals were heterozygotes, Mendelian segregation prevents this
- The equilibrium is an evolutionary compromise between the advantages of heterozygotes and the costs of the homozygotes

10.1.1 Heterozygote Advantage

- An example of a polymorphism maintained by heterozygote advantage is the allele for horn size in Soay Sheep on Soay island off the coast of Scotland

- Males use their horns to compete for females and Johnston and colleagues (2013) found a large-effect locus, the RXFP2 gene, that controls variation in horn size

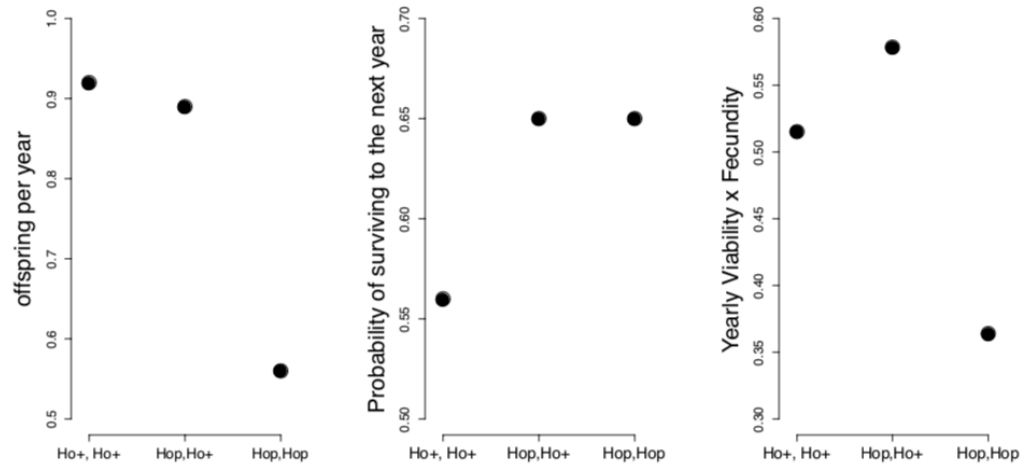
- The Ho^{+} allele at this locus leads to increased horn size and the Ho^{p} allele is associated with small horns



10.1.1 Heterozygote Advantage

- While Ho^{p} allele homozygotes show reduced offspring number, Ho^{+} allele homozygotes show reduced survival, meaning the overall highest fitness is seen in the heterozygote (right-hand panel of the figure)

- The polymorphism is therefore balanced in the population due to this instance of heterozygote advantage

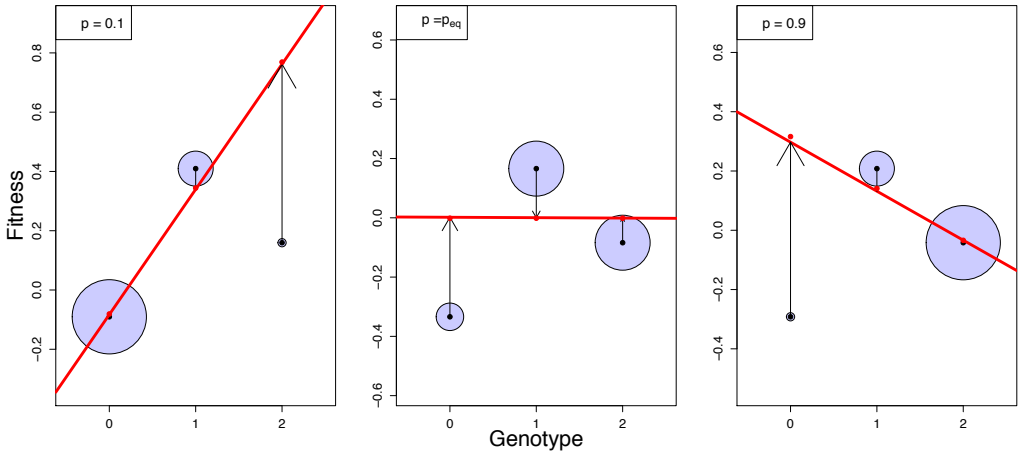


10.1.1 Heterozygote Advantage

- Let's consider that the marginal effects of alleles are equivalent to their additive effects on fitness

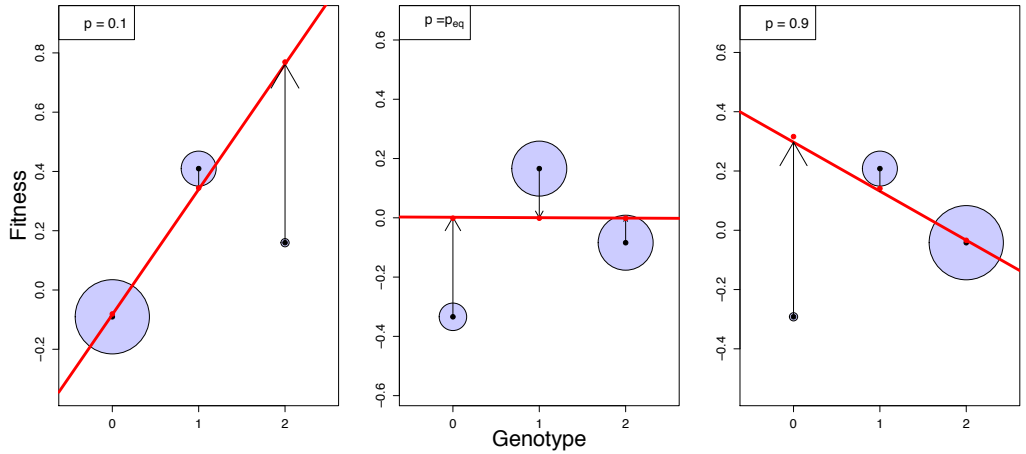
- As you may remember from Chapter 7, the difference in the additive effects of two alleles gives the slope of the regression of additive genotypes on fitness

- There is additive variance in fitness when the slope is something other than zero



10.1.1 Heterozygote Advantage

- In the heterozygote advantage model, the marginal fitness of the A_1 allele (*i.e.,* the additive effect of the A_1 allele on fitness) is greater than the marginal fitness of the A_2 allele when A_1 is at low frequency
- In this case, there is a positive slope in the regression of fitness on the number of A_1 alleles in an individual
- When the frequency of the A_1 allele rises above the equilibrium frequency, this slope becomes negative
- When the population is at equilibrium frequency, there is no additive genetic variance, and the slope is zero



10.1.1 Heterozygote Advantage (or lack of it!)

- We can also see examples in nature when the heterozygote is less fit than either of the homozygotes, known as underdominance

- For example, in the butterfly species *Pseudacraea eurytus*, the two homozygous genotypes mimic toxic orange and blue butterflies, but the heterozygous genotype mimics neither and suffers high predation

- Underdominance can be parameterized as:

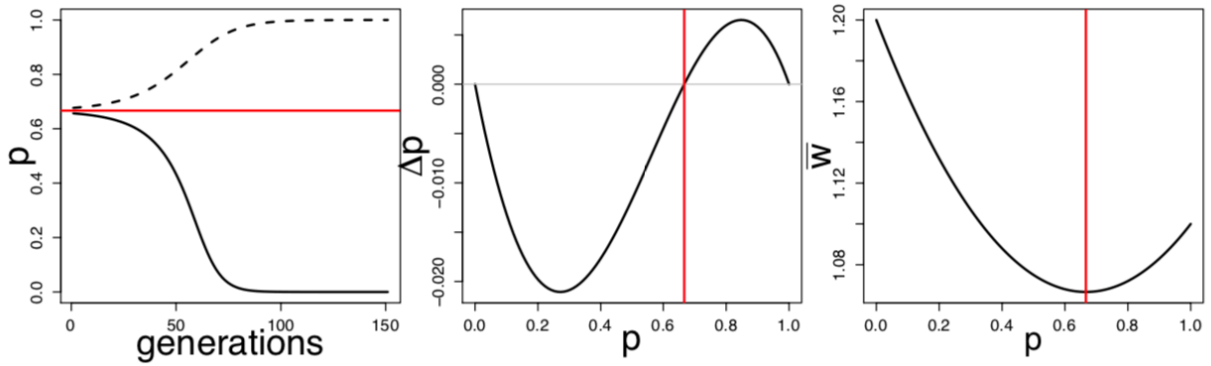
genotype	A_1A_1	A_1A_2	A_2A_2
absolute fitness	w_{11}	$> w_{12} <$	w_{22}
relative fitness (generic)	$w_{11} = W_{11}/W_{12}$	$w_{12} = W_{12}/W_{12}$	$w_{22} = W_{22}/W_{12}$
relative fitness (specific)	$1 + s_1$	1	$1 + s_2$

10.1.1 Heterozygote Advantage (or lack of it!)

- Underdominance permits three equilibria, $p = 0, p = 1$ and a polymorphic equilibrium, $p = p_U$

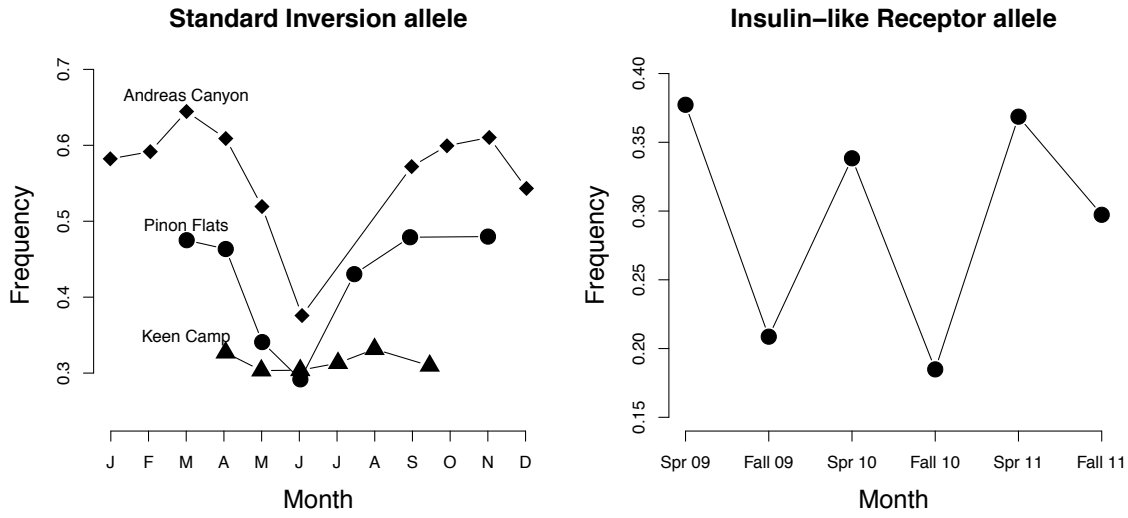
- However, the polymorphic equilibrium is unstable and if $p < p_U$ then Δp is negative and A_1 will be lost and if $p > p_U$ then Δp is positive and A_1 will be fixed

- While underdominant alleles will not typically spread within a population, they often arise in recently diverged species and may be a contributor to the maintenance of species



10.1.1 Heterozygote Advantage

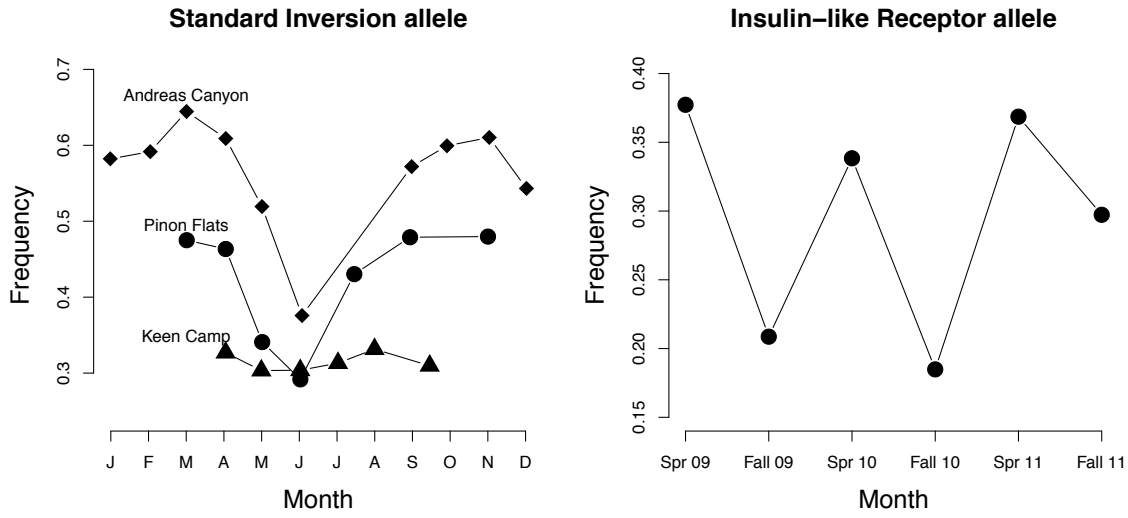
- Selection pressures can vary over time and this variability may help maintain polymorphism
- Multiple instances of this phenomenon have been observed in *Drosophila*, where short lifespan allows for observation of seasonal fluctuation in allele frequencies
- The first example is *Drosophila pseudoobscura* populations in western North America



10.1.1 Heterozygote Advantage

- Dobzhansky (1943) and Wright and Dobzhansky (1946) found an inversion polymorphism (*i.e.*, where different individuals have a different orientation of DNA within a chromosomal region) segregating in seasonal patterns

- These authors documented frequencies of inversion types in multiple populations across four year and discovered that the standard allele decreased in frequency in the summer, when the inverted allele prevailed

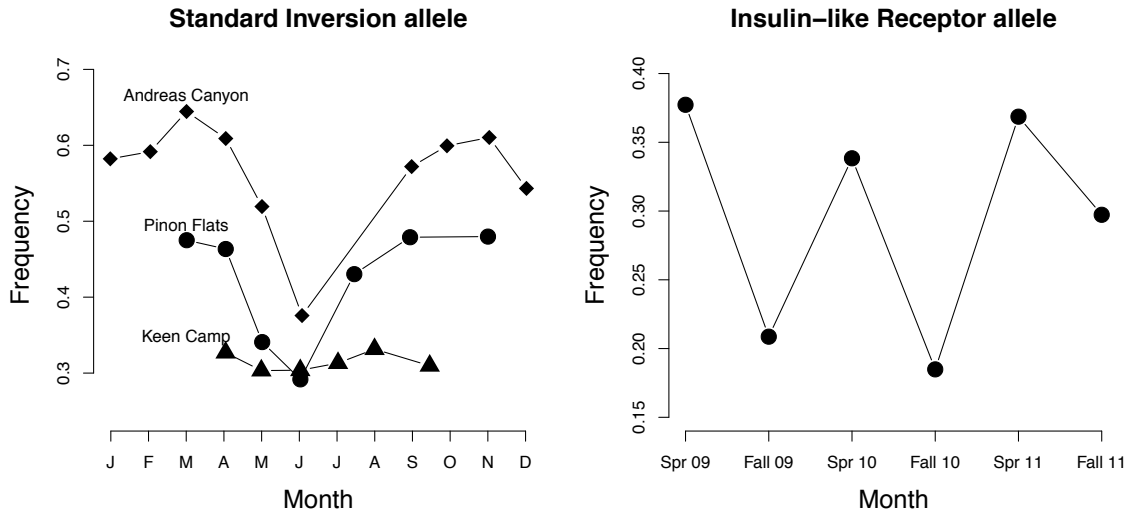


10.1.1 Heterozygote Advantage

- A second example was found by Paaby and colleagues (2014) and involved an insertion-deletion polymorphism segregating in an insulin-like receptor gene in *Drosophila melanogaster*

- Alleles at this locus also oscillated with the seasons

- These alleles had large effects on developmental time and fecundity, which may help maintain both alleles through life-history fitness trade-offs



10.1.1 Heterozygote Advantage

- In order to develop equations that capture temporal fluctuations in fitness, we'll need to make our diploid fitnesses time-dependent: $w_{11,t}, w_{12,t}, w_{22,t}$

- The diploid case is more difficult to model across allele frequencies due to segregation, but we can investigate dynamics when, for example, A_1 is rare

- When A_1 is rare ($p \ll 1$), the frequency in the next generation can be approximated in a similar fashion to equation 10.26:

$$p_{t+1} \approx \frac{w_{12}}{\overline{w}} p_t. \qquad (10.40)$$

- Where we have omitted ${p_t}^2$ terms because they are so small, and q_t in the numerator because $q_t \approx 1$

10.1.1 Heterozygote Advantage

- With similar assumptions, we can also approximate q_{t+1}:

$$\frac{p_{t+1}}{q_{t+1}} = \frac{w_{12,t}}{w_{22,t}} \frac{p_t}{q_t}. \qquad (10.41)$$

- And generalizing across any number of generations:

$$\frac{p_{t+1}}{q_{t+1}} = \left(\prod_{i=0}^{t} \frac{w_{12,i}}{w_{22,i}} \right) \frac{p_0}{q_0}. \qquad (10.42)$$

- From this we can see, similar to our haploid argument, that A_1 will increase in frequency when rare only when:

$$\frac{\sqrt[t]{\prod_{i=0}^{t} w_{12,i}}}{\sqrt[t]{\prod_{i=0}^{t} w_{22,i}}} > 1, \qquad (10.43)$$

10.1.1 Heterozygote Advantage

$$\frac{\sqrt[t]{\prod_{i=0}^{t} w_{12,i}}}{\sqrt[t]{\prod_{i=0}^{t} w_{22,i}}} > 1, \qquad (10.43)$$

- In other words, the A_1 allele will increase when rare if the heterozygote has a higher geometric mean fitness than the A_2A_2 homozygote
- But will the A_1 allele approach fixation or can a balanced polymorphism exist?
- To determine this, we can repeat our analysis for $q \ll 1$ and see:

$$\frac{p_{t+1}}{q_{t+1}} = \left(\prod_{i=0}^{t} \frac{w_{11,i}}{w_{12,i}} \right) \frac{p_0}{q_0}. \qquad (10.44)$$

10.1.1 Heterozygote Advantage

$$\frac{p_{t+1}}{q_{t+1}} = \left(\prod_{i=0}^{t} \frac{w_{11,i}}{w_{12,i}} \right) \frac{p_0}{q_0}. \qquad (10.44)$$

- Here, the A_1A_1 homozygote must be able to outcompete the heterozygote (*i.e.*, have a higher geometric mean fitness); otherwise the A_2 allele will increase when rare
- This means that we can potentially have a balanced polymorphism when the heterozygote is never the most fit genotype, as long as it is always more fit than one of the homozygotes
- The heterozygote can win out in variable environments, even when it is never the most fit genotype in any particular environment

10.1.1 Heterozygote Advantage

- As a toy example of this type of balanced polymorphism, let's imagine a plant population that half of the time experiences wet environments and the other half dry environments

- Across long periods of time, we can assume the time in wet and dry environments is equal and calculate the arithmetic means presented in the top table

- Note that the heterozygote does not have the highest fitness in either environment and it's arithmetic mean fitness is less than the AA homozygote

Environment	AA	Aa	aa
Wet	6.25	5.0	3.75
Dry	3.85	5.0	6.15
arithmetic mean	5.05	5.0	4.95

	AA	Aa	aa
Geometric mean	4.91	5.0	4.80

10.1.1 Heterozygote Advantage

- If we take the t^{th} root across generations to obtain the geometric mean fitnesses (bottom table), we will see that the heterozygote has the highest value
- Both the A_1 and A_2 alleles will increase in frequency when rare because of the higher fitness of the heterozygote
- In this way, a balanced polymorphism will be maintained

Environment	AA	Aa	aa
Wet	6.25	5.0	3.75
Dry	3.85	5.0	6.15
arithmetic mean	5.05	5.0	4.95

	AA	Aa	aa
Geometric mean	4.91	5.0	4.80

10.1.1 Heterozygote Advantage

- The heterozygote advantage examples we have considered thus far involve maintenance of polymorphism because the common allele has a disadvantage relative to the more rare allele

- However, the relative fitness of genotypes has not been dependent on the other genotypes found in the population

- This is not the case for the family of models known as “frequency-dependent selection”

- For example, negative frequency dependent selection occurs when the fitness of an allele declines as it becomes more common

10.1.1 Heterozygote Advantage

- Long-term heterozygote advantage may be rare in nature, but balancing selection through negative frequency-dependent selection is likely common
- These dynamics often arise due to interactions within or among species
- For example, in predator-prey or host-pathogen interactions, common phenotypes/genotypes may be at a disadvantage because predators and pathogens become familiar with these and learn how to counteract or target them

10.1.1 Heterozygote Advantage

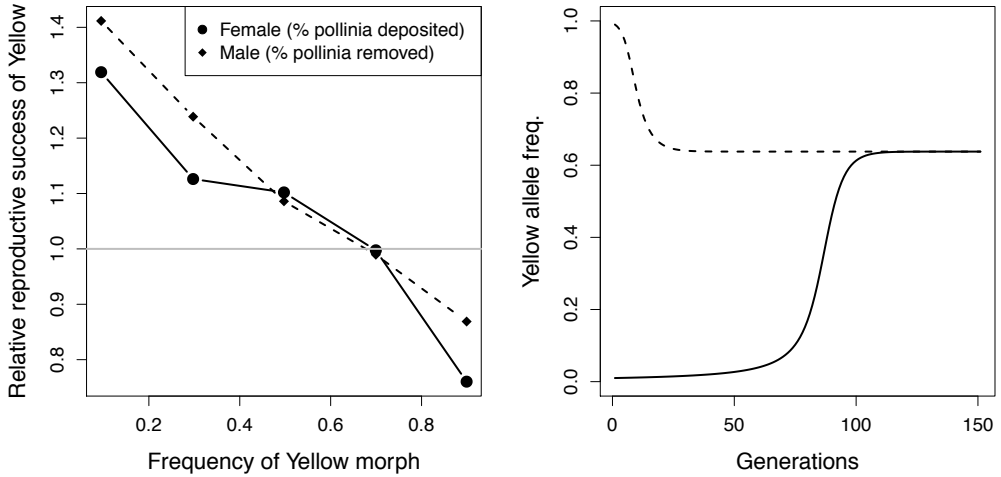
- The deceptive Elderflower orchid is a good example of negative frequency-dependent selection
- The species is polymorphic for yellow and purple flowers, with neither offering nectar or pollen rewards to their bumblebee pollinators
- The flower relies on inexperienced bumblebees for pollination and multiple flower colors help prevent pollinators from learning too quickly that the flower offers no rewards
- If one flower color were more common, the signal to pollinators would be stronger to avoid the orchid

10.1.1 Heterozygote Advantage

- Gigord and colleagues (1997) documented negative frequency-dependent selection by setting up plots with varying frequencies of the color morphs

- As a particular color morph became more common, bumblebees were less likely to visit and pollinate it because they learned there were no rewards

- This dynamic is likely responsible for maintaining the two flower morphs in natural populations across Europe



10.1.1 Heterozygote Advantage

- Negative frequency-dependent selection can also occur due to interactions across individuals within a population

- For example, ruffs, a species of sandpiper in Eurasia, exhibit lekking behavior in which males gather together on open ground to display and attract females

- There are three male morphs with different breeding strategies: Independent, Satellite, and Faeder

10.1.1 Heterozygote Advantage

- Independent males display and defend their own small territories and are most common
- Satellite males occur at 16% frequency, do not defend territories, but join in lek displays, mating opportunistically with females
- Faeder males occur at 1% frequency and their plumage and size resembles females, but they join leks and surreptitiously mate with females
- These males rely on the displays of Independents and therefore Satellites and Faeders cannot become overly common in populations

10.1.1 Heterozygote Advantage

- The dynamics of the previous examples are likely simple in comparison to those occurring amongst the hundreds of alleles found across the many genes of the Major Histocompatibility Complex (MHC)

- The MHC is key to the vertebrate immune system and is in a constant arms race with pathogens adapting to common MHC alleles, causing rare alleles to be favored

- Balancing selection has caused some MHC polymorphisms to be maintained for millions of years

- In fact, a human MHC allele may be more closely related to an allele in another primate species than that found in another human

Coop, Chapter 10: 10.2

One-Locus Models of Selection

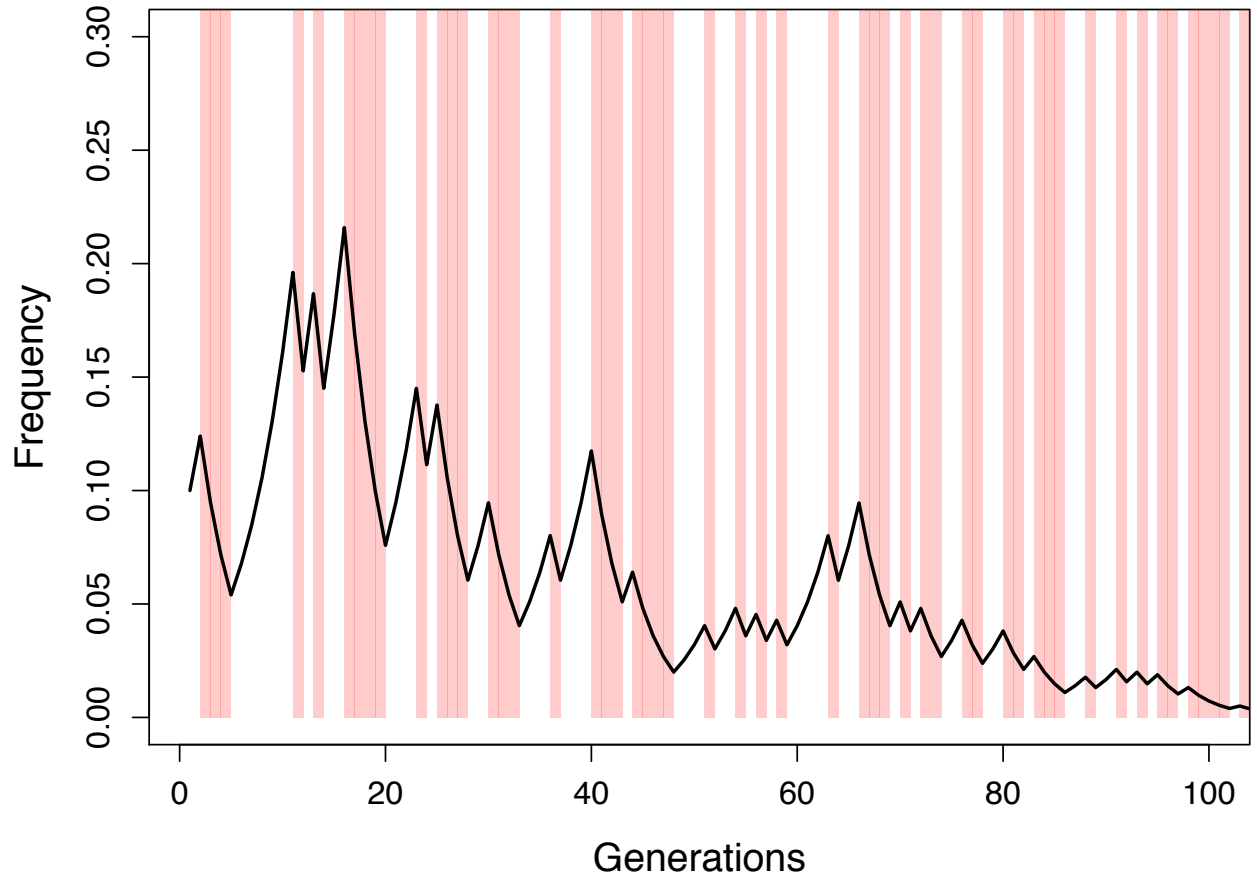


Figure 10.2: An example frequency trajectory of the A_1 allele under variable environments (using the fitnesses from Table 10.1). Wet years (generations) are shown in red, dry years in white. The environment flips at random each year. Note how the A_1 allele increases in frequency in the dry years as it has higher fitness, and yet the A_2 allele still wins out. Code here.

10.2 Sex ratios, sex ratio distorters, and other selfish elements

- Frequency dependent selection can often prevent natural selection from increasing mean fitness

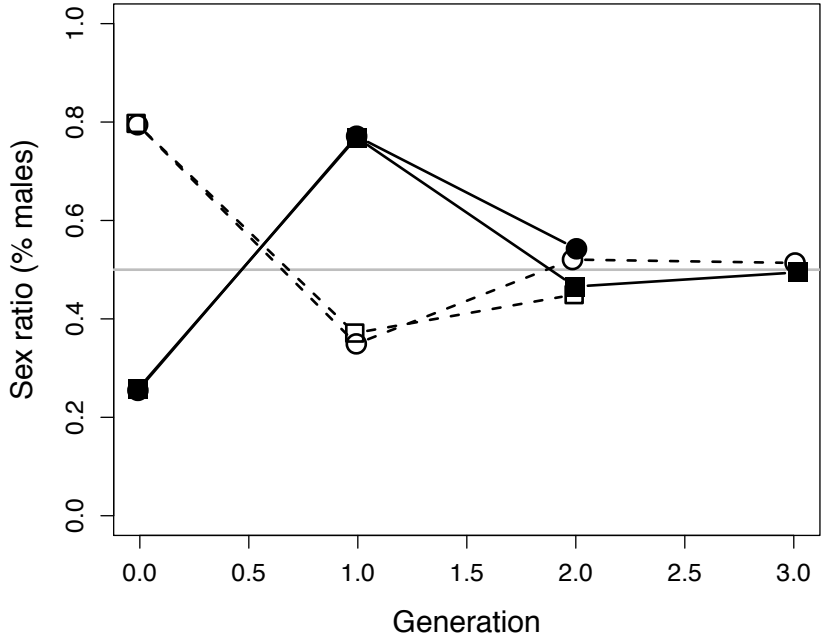
- For example, evolution of the male morphs of the Ruff species discussed in the previous section does not maximize the growth rate (*i.e.*, reproduction) of the species

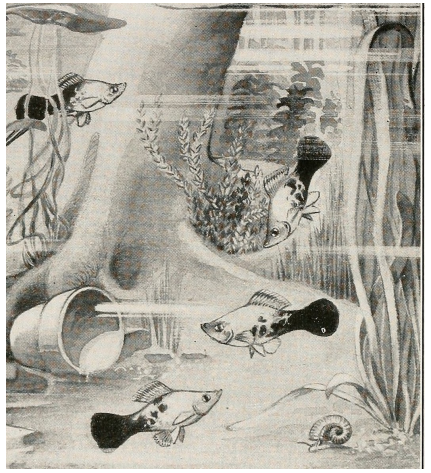
- The commonly observed 50/50 sex ratio across species is one clear example of how frequency dependent selection can drive non-optimal dynamics for a population

- In this, and other situations where selection acts below the level of the individual, we can see a compromise of fitness

10.2 Sex ratios, sex ratio distorters, and other selfish elements

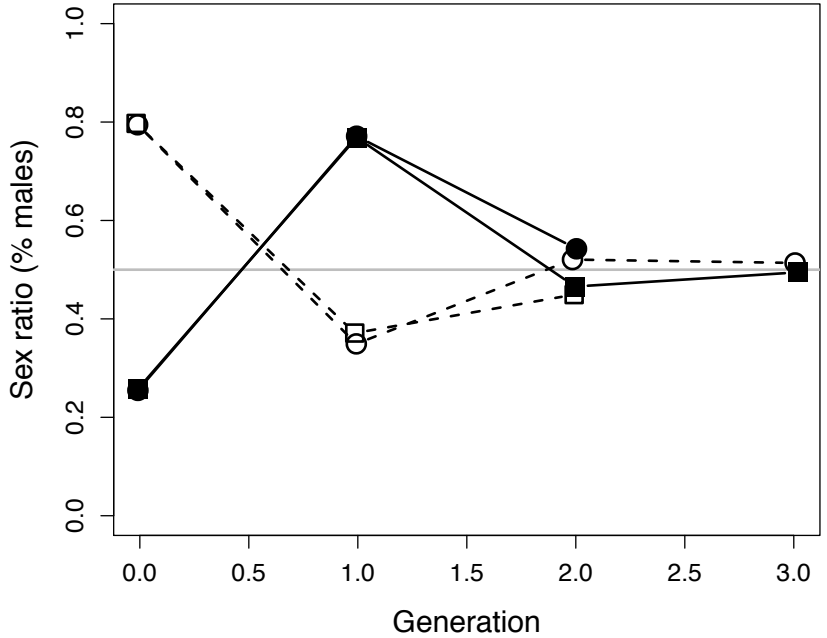
- In many species, the sex ratio is close to 50/50 even though this does not maximize fitness (reproduction) for a species
- Females are typically limiting in terms of population growth, so a sex ratio favoring females would be expected to be the optimum
- However, if populations deviate from the 50/50 ratio they will typically quickly return
- For example, the platyfish work of Basolo and colleagues (1994) in which populations with experimentally manipulated ratios quickly returned to 50/50

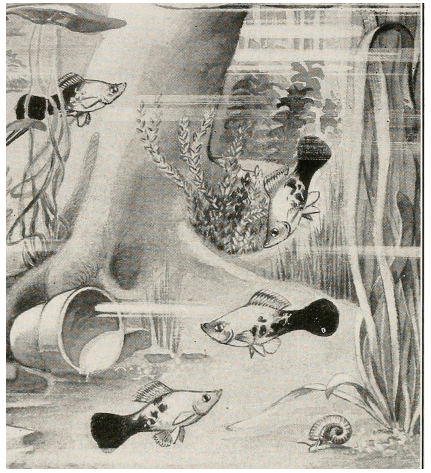




10.2 Sex ratios, sex ratio distorters, and other selfish elements

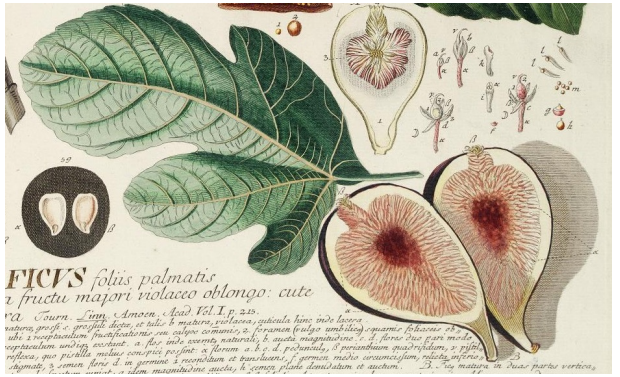
- Why do we then see such a stable ratio of 50/50?
- Imagine a population in which females predominate 80/20; an autosomal mutation favoring higher production of males would confer an advantage to that individual because their offspring would father a much greater proportion of the population
- The same would be true if the ratio was skewed in the other direction (20/80 favoring males)
- Selection on autosomal alleles will favor production of the rare sex (*i.e.*, negative frequency dependent selection)

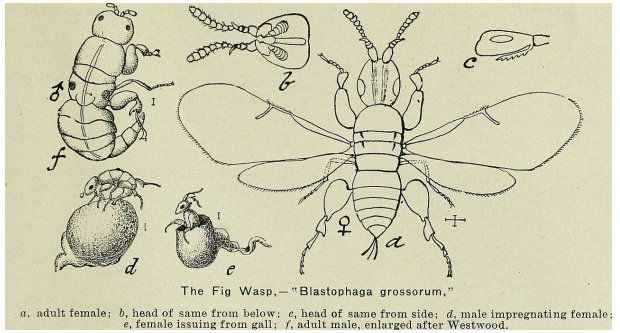




10.2 Sex ratios, sex ratio distorters, and other selfish elements

- However, there are notable deviations from the 50/50 ratio, for example, when the ratio can be controlled among offspring that reproduce together
- Fig wasps are a haplo-diploid species in which fertilized eggs lead to females and unfertilized to males
- Females can control the number of fertilized vs. unfertilized eggs that they lay in figs where they develop and therefore can control the sex ratio
- After hatching, mating occurs within the fig, with the wingless males never leaving the fig in which they were born

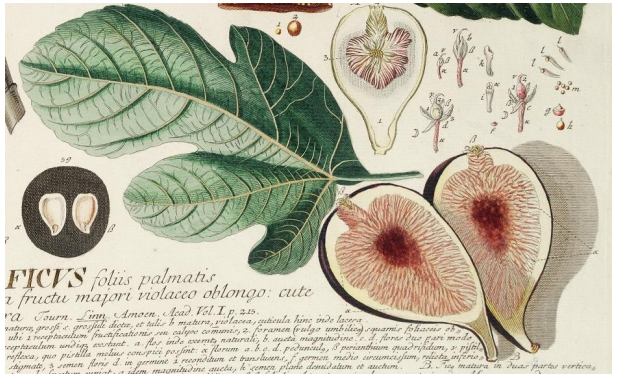


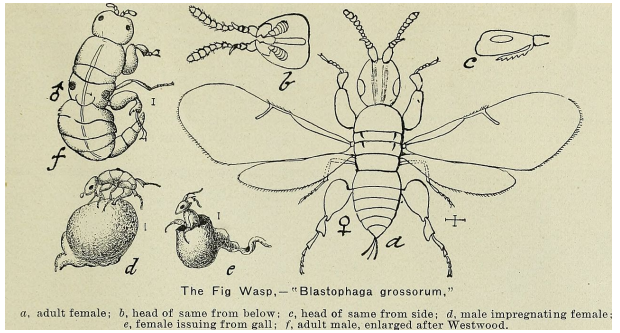


10.2 Sex ratios, sex ratio distorters, and other selfish elements

- Since females can tightly control these ratios, it would be to their advantage to produce as many females as possible and just enough males to ensure fertilization

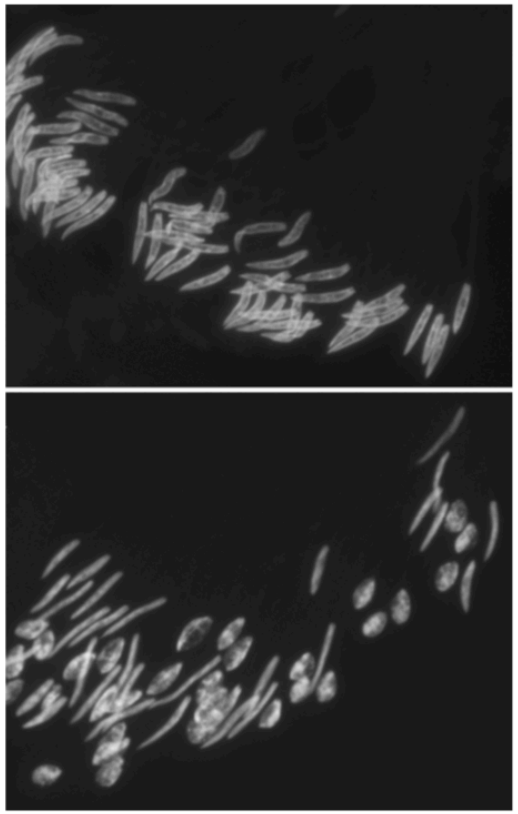
- In many species of fig wasp, 95% of individuals that are born are female





10.2 Sex ratios, sex ratio distorters, and other selfish elements

- Additional selfish strategies can occur below the level of the individual, leading to low fitness and actively harming individuals
- For example, in multiple species with XY mating systems, “selfish X” dynamics have evolved, in which an allele on the X chromosome has arisen to release a poison killing Y-chromosome-bearing sperm during gametogenesis
- This results in a strong skew toward females in the population
- This selfish mechanism can compromise the fitness of the individual
- A selfish X allele has been observed in *Drosophila simulans* populations near the town of Winters, California

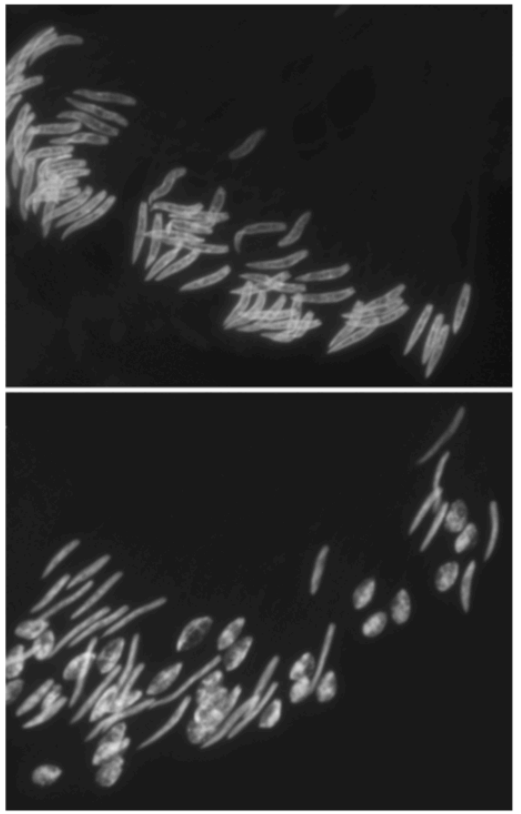


10.2 Sex ratios, sex ratio distorters, and other selfish elements

- The gene responsible for the selfish X dynamic in *D. simulans* is known as *Dox (Distorter on the X)* and results in males having >80% daughters

- Male spermatids (those that look like rice krispies in the lower figure) do not fully develop and fertilize eggs

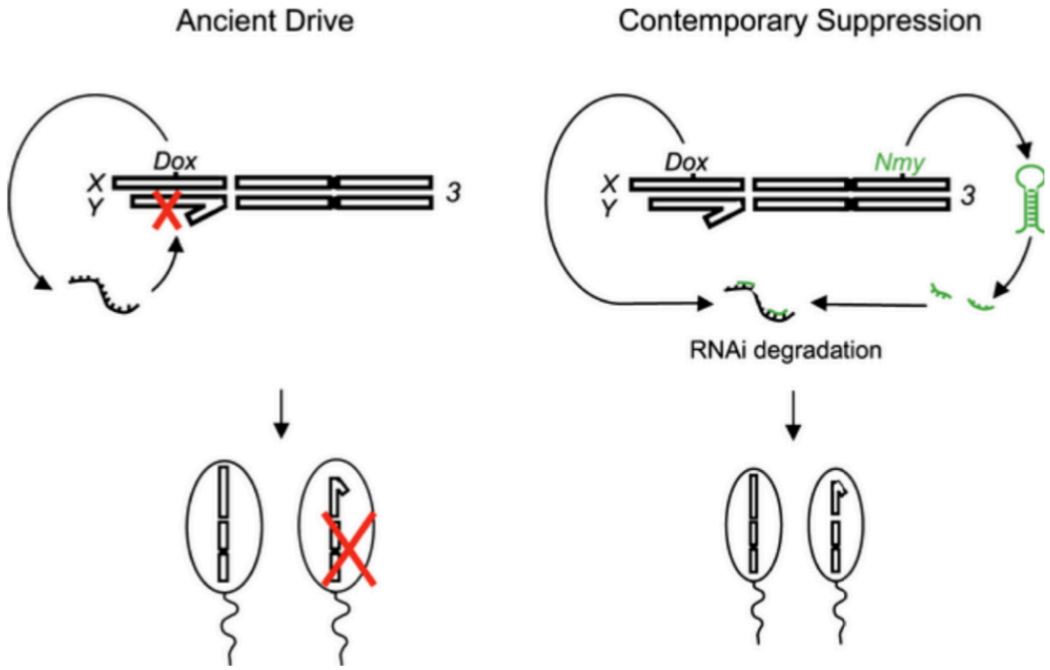
- Selfish X mechanisms are often suppressed by new mutations that occur on autosomes and are favored because individuals carrying a suppressor will have more male offspring when this sex is rare and contribute disproportionately to the next generation



10.2 Sex ratios, sex ratio distorters, and other selfish elements

- An autosomal allele has arisen that suppresses selfish X caused by the *Dox* gene in *D. simulans* and has restored a sex ratio of 50/50

- The gene, *NMY*, is actually a duplicate of the *Dox* gene that was transposed to an autosome and now blocks the action of *Dox* through RNA-interference degradation of the *Dox* transcripts



10.2 Sex ratios, sex ratio distorters, and other selfish elements

- Maternally transmitted DNA in cytoplasmic organelles can also be bad actors
- For example, from the mitochondrion's perspective, pollen and male gametes are a waste of energy because they don't transmit mitochondria
- Alleles have arisen in mitochondrial DNA that target male pollen, shunting energy to other processes beneficial to mitochondria
- This phenomenon is known as "Cytoplasmic Male Sterility" and usually creates populations of females or hermaphrodite plants as in the Bladder Campion plant to the right

10.2 Sex ratios, sex ratio distorters, and other selfish elements

- CMS is typically counteracted by autosomal suppressor alleles that arise and are favored because they produce the more rare gamete (pollen)
- The CMS mechanism has been used as a strategy in agricultural species, because it allows for much more tightly controlled crosses between different accessions of a species
- There will be no danger that CMS-induced female plants will self pollinate

10.2 Sex ratios, sex ratio distorters, and other selfish elements

- In fact, mitochondria can serve as a hotspot for mutations deleterious to males (an effect referred to as the “Mother’s Curse”)

- Another example is a mutation underlying Leber’s hereditary optic neuropathy (LHON) which causes loss of vision in teenage males

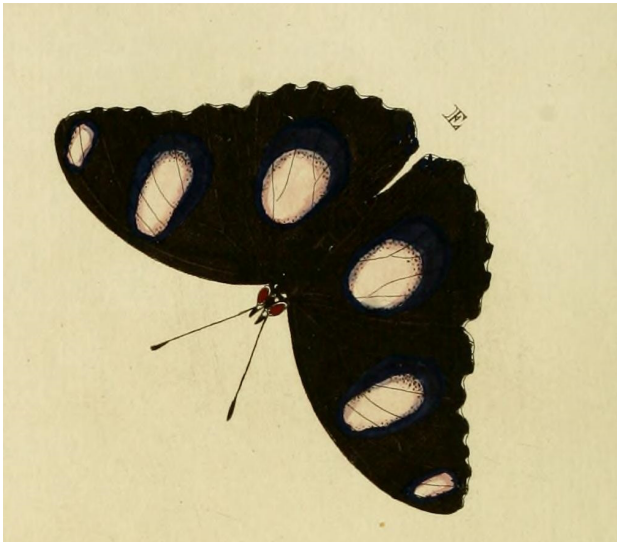
- This mutation is at low frequency in the Canadian province of Quebec, having arrived in a single woman who was one of the “*fille de roi”,* women sent to Quebec by Louis XIV of France to help balance gender ratios in the new colony.

10.2 Sex ratios, sex ratio distorters, and other selfish elements

- Milot and colleagues (2017) tracked mitochondrial descendants of this woman (those whose mothers were in her matrilineal line)
- No difference in fitness was found in women that carried or did not carry the allele
- However, males that carried this allele had only 65% of the fitness of males who did not carry the allele
- This allele has continued to segregate in the population because it has no fitness consequence for females

10.2 Sex ratios, sex ratio distorters, and other selfish elements

- It is not only mitochondria and nuclear DNA that participate in the battle of the sexes; the intracellular bacteria *Wolbachia* also plays a role, particularly in insect species



- These bacteria are passed to offspring through the cytoplasm, and thus are maternally inherited

- Similar to mitochondria, males are of no advantage to *Wolbachia* and many mutations have arisen in these bacteria that increase their transmission through feminization of males or killing of male embryos

- One dramatic example is the eggspot butterfly in which *Wolbachia* caused a 100/1 ratio of females to males until an autosomal suppressor arose that restored a 50/50 sex ratio in the years 2001-2006

10.2 Sex ratios, sex ratio distorters, and other selfish elements

- While many of the examples we have discussed have been on sex chromosomes, selfish genetic mechanisms can also arise on autosomes
- These dynamics are driven by competition among alleles for inclusion in the gamete to be passed on to the next generation
- For example, the four products of meiosis can be seen in the asci for spores of the SxS and TxT individuals of *Podospora anserina* below, and in the SxT individuals, there are only two products because the T allele releases a toxin that poisons the S spores