

Lecture 9: Selection at one locus for large populations (Ch10)

Population genetic PCB4553/6685

Background

- Aim: learn how the frequency of alleles (genes) at a single locus changes over time due to natural selection, *without influence of drift, migration, and mutation*.
- Selection occurs when there are differences between individuals in fitness.
- Absolute fitness of a genotype is the expected number of offspring of an individual of that genotype. Differences in fitness among genotypes drive allele frequency change.

Part 1: Haploid selection model

- Haploid: one allele per individual
- Mathematically simpler than the diploid model

Basic notations

- P_t : no. of individuals carrying alleles A1 at time t
- Q_t : no. of individuals carrying alleles A2 at time t
- Frequencies at time t
 - $p_t = P_t / (P_t + Q_t)$
 - $q_t = Q_t / (P_t + Q_t) = 1 - p_t$
- Absolute fitness, average number of offspring for the A1 and A2 genotypes
 - W_1
 - W_2

Population size after one generation

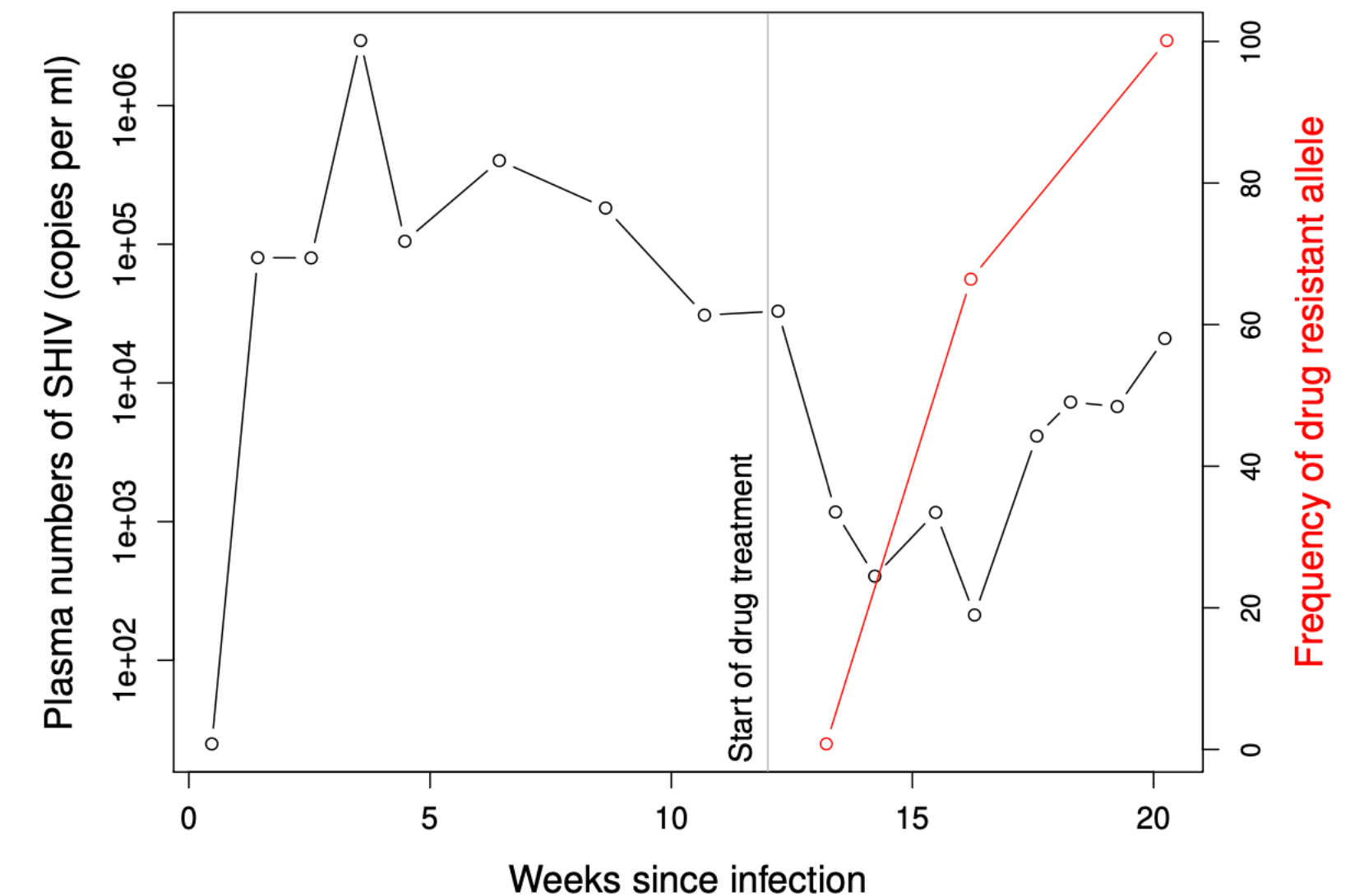
- Absolute number of A1 and A2 in the next generation
 - $P_{t+1} = W_1 P_t$
 - $Q_{t+1} = W_2 Q_t$
- No overlapping generations
- Both A1 and A2 grows exponentially
- $P_t = (W_1)^t P_0$

Mean fitness

- Mean absolute fitness

- $$\bar{W}_t = W_1 \frac{P_t}{P_t + Q_t} + W_2 \frac{Q_t}{P_t + Q_t} = W_1 p_t + W_2 q_t,$$

- Note: mean fitness depends on time, as it is a function of the allele frequencies, which are themselves time dependent.



The rapid evolution of drug-resistant SHIV. The viral load of SHIV in the blood of a macaque (black line), the frequency of a drug resistance mutation (red line). Data from Feder et al. (2017).

Mean fitness

- The frequency of allele A1 in the next generation is given by

- $$p_{t+1} = \frac{P_{t+1}}{P_{t+1} + Q_{t+1}} = \frac{W_1 P_t}{W_1 P_t + W_2 Q_t} = \frac{W_1 p_t}{W_1 p_t + W_2 q_t} = \frac{W_1}{\bar{W}_t} p_t.$$

- *change in p only depends on a ratio of fitnesses. Fitness only needs to be defined up to an arbitrary constant.

- Relative fitness: w_i can be defined as such

- $w_1 = W_1 / W_1 = 1$

- $w_2 = W_2 / W_1$

- Mean relative fitness:

- $$\bar{w}_t = \frac{\bar{W}}{W_1} = \frac{p_t W_1 + q_t W_2}{W_1} = p_t w_1 + q_t w_2$$

- $$p_{t+1} = \frac{w_1 p_t}{\bar{w}_t}$$

- $w_1 p_t$: scaled size of A1 pop

- \bar{w}_t : scaled size of total pop

Change in frequency after one generation

- $\Delta p_t = p_{t+1} - p_t = \frac{w_1}{\bar{w}} p_t - p_t$
- $= \frac{w_1 p_t - (w_1 p_t + w_2 q_t) p_t}{\bar{w}}$
- $= \frac{w_1 - w_2}{\bar{w}} p_t q_t$
- $\Delta w_t = \Delta p_t (w_1 - w_2) = \frac{\text{var}(w)}{\bar{w}}$
- Fisher's fundamental theorem of natural selection!
- "The rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time."



RA Fisher

Allele frequencies after multiple generations

- $P_\tau = (W_1)^\tau P_0$
- $Q_\tau = (W_2)^\tau Q_0$
- $p_\tau = \frac{(W_1)^\tau P_0}{(W_1)^\tau P_0 + (W_2)^\tau Q_0} = \frac{p_0}{p_0 + (w_2/w_1)^\tau q_0}$

Allele frequencies change after multiple generations

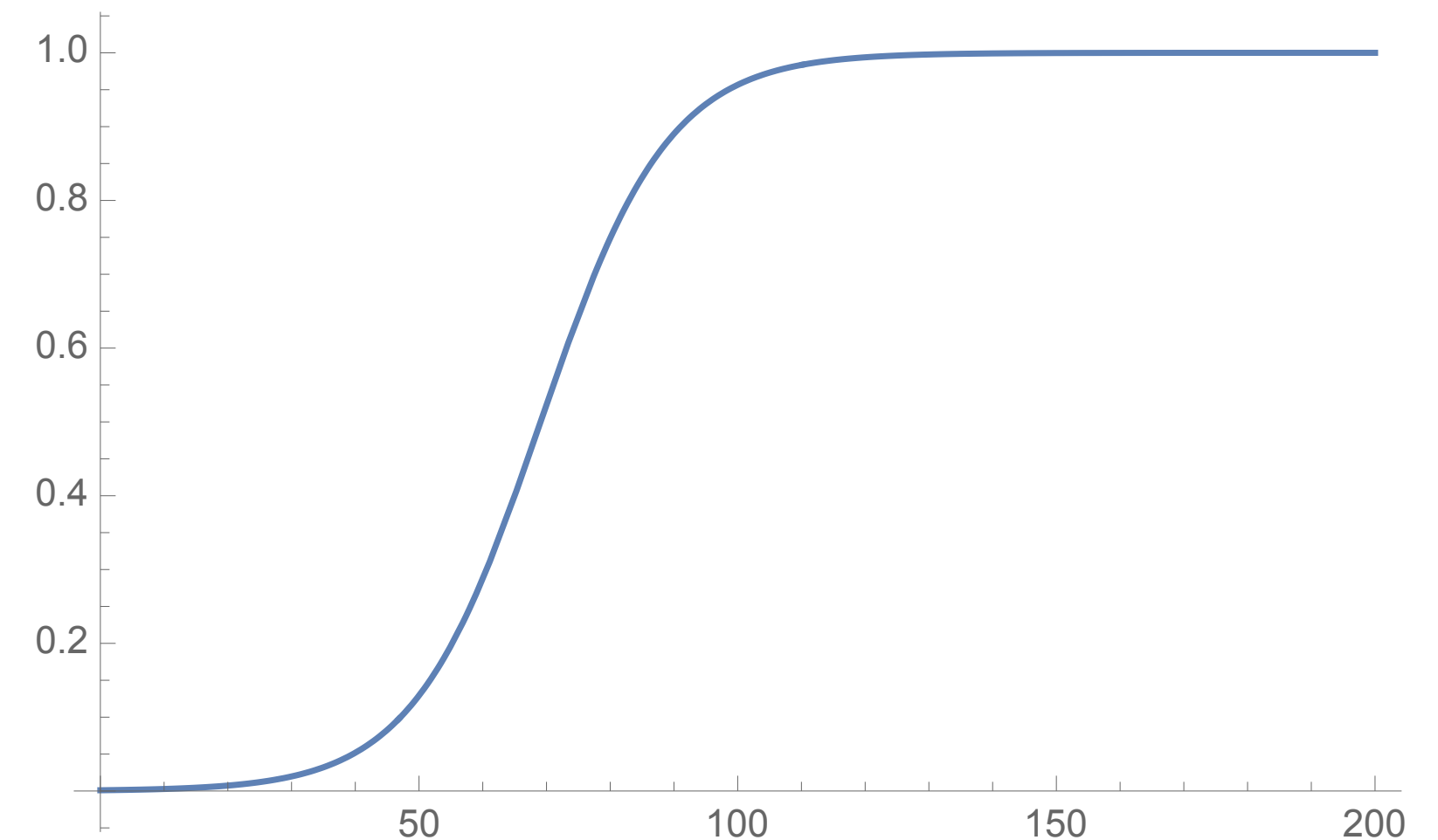
- In practice, we often parametrize the fitness in terms of selection coefficient s

- $w_1 = 1$

- $w_2 = 1 - s$

- $$p_\tau = \frac{p_0}{p_0 + (w_2/w_1)^\tau q_0} = \frac{p_0}{p_0 + (1 - s)^\tau q_0}$$

- $$p_\tau \approx \frac{p_0}{p_0 + e^{-s\tau} q_0}, \text{ if } s \ll 1$$



Calculating time to near fixation

- $p_\tau = \frac{p_0}{p_0 + (1 - s)^\tau q_0}$

- $\frac{p_\tau}{q_\tau} = \frac{p_0}{q_0} (1 - s)^\tau$

- Solve for τ :

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

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

- Use Taylor expansion of $\log(1 - s)$

- $\log(1 - s) \approx \log(1) - s = -s$

- time it takes to go from an absolute frequency of 1 to near fixation in a population of size N

- $p_0 = 1/N$

- $p_\tau = 1 - 1/N$

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

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

Question 1. I

In our example of the evolution of drug resistance, the drug-resistant SHIV virus spread from undetectable frequencies to $\sim 65\%$ frequency by 16 weeks post infection. An estimated effective population size of SHIV is 1.5×10^5 , and its generation time is ~ 1 day. Assuming that the mutation arose as a single copy allele very shortly the start of drug treatment at 12 weeks, what is the selection coefficient favouring the drug resistance allele?

Part 2: Diploid selection model

- Genotypes
 - A_1A_1 , A_1A_2 , and A_2A_2
- Fitnesses
 - W_{11} , W_{12} , and W_{22}
- W_{ij} is the probability that zygote survives
to reproduction

General theory

- Number of zygotes in generation t are

- $$Np_t^2, \quad N2p_tq_t, \quad Nq_t^2.$$

- The mean fitness of the population of zygotes is then

- $$\overline{W}_t = W_{11}p_t^2 + W_{12}2p_tq_t + W_{22}q_t^2.$$

- Expected number of individuals surviving to reproduce is

- $$NW_{11}p_t^2, \quad NW_{12}2p_tq_t, \quad NW_{22}q_t^2.$$

- Total number of individuals surviving to reproduce is

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

- Relative frequency of individuals at reproduction

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

	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	$\frac{W_{11}}{\overline{W}}p_t^2$	$\frac{W_{12}}{\overline{W}}2p_tq_t$	$\frac{W_{22}}{\overline{W}}q_t^2$

General theory

- Assuming no difference in the fecundity of the three genotypes
- Allele frequencies in the zygotes forming the next generation are simply the allele frequency among the reproducing individuals of the previous generation.
- Frequency of A_1 in generation $t + 1$ is

$$p_{t+1} = \frac{W_{11}p_t^2 + W_{12}p_tq_t}{\bar{W}}$$

-

General theory

- The change in frequency from generation t to $t + 1$ is

- $$\Delta p_t = p_{t+1} - p_t = \frac{w_{11}p_t^2 + w_{12}p_tq_t}{\bar{w}} - p_t$$

- Marginal fitnesses

- $$\begin{aligned}\bar{w}_1 &= w_{11}p_t + w_{12}q_t, \\ \bar{w}_2 &= w_{12}p_t + w_{22}q_t.\end{aligned}$$

- mean relative fitness can be expressed in terms of the marginal fitnesses as

- $$\bar{w} = \bar{w}_1p_t + \bar{w}_2q_t$$

- $$\Delta p_t = \frac{(\bar{w}_1 - \bar{w}_2)}{\bar{w}}p_tq_t.$$

General theory

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

- The frequency of A1 will
- increase if $\bar{w}_1 > \bar{w}_2$,
- decrease if $\bar{w}_1 < \bar{w}_2$

$$\Delta p_t = \frac{1}{2} \frac{p_t q_t}{\bar{w}} \frac{d\bar{w}}{dp}$$

$$\begin{aligned} \frac{d\bar{w}}{dp} &= \frac{d}{dp} (W_{11}p^2 + 2W_{12}p \\ &\quad - 2W_{12}p^2 + W_{22} - 2W_{22}p + W_{22}p^2) \\ &= 2(w_{11}p + w_{12} - 2pw_{12} - w_{22} - w_{22} + w_{22}p) \end{aligned}$$

- selection is acting to increase the mean fit- ness of the population.

Diploid directional selection

- Directional selection: one of the two alleles always has higher marginal fitness than the other one

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	$1 - sh$	$1 - s$

- Assuming that A1 is the fitter allele

- $w_{11} \geq w_{12} \geq w_{22}$

- $\bar{w}_1 > \bar{w}_2$.

- s : selection coefficient, difference in relative fitness between the two homozygotes.

- h : dominance coefficient. $0 \leq h \leq 1$

- $h = 0$

- A1 is fully dominant and A2 fully recessive.

- $h = 1$

- A1 is fully recessive and A2 fully dominant.

- $\Delta p_t = \frac{\bar{w}_1 - \bar{w}_2}{\bar{w}} p_t q_t = \frac{p_t h s + q_t s(1 - h)}{\bar{w}} p_t q_t$
- $\bar{w} = p_t \bar{w}_1 + q_t \bar{w}_2$
- $= p_t(p_t + q_t(1 - h s)) + q_t(p_t(1 - h s) + q_t(1 - s))$
- $= 1 - 2p_t q_t s h - q_t^2 s$

Directional selection on an *additive* allele

- Per generation change in A1 freq:

- $\Delta p_t = \frac{\bar{w}_1 - \bar{w}_2}{\bar{w}} p_t q_t = \frac{p_t h s + q_t s(1 - h)}{\bar{w}} p_t q_t$

- Assuming additivity, i.e. $h = 1/2$

- $\Delta p_t = \frac{\bar{w}_1 - \bar{w}_2}{\bar{w}} p_t q_t = \frac{1}{2} \frac{s}{\bar{w}} p_t q_t$

- $\approx \frac{1}{2} s p_t q_t$

- Per generation change in A1 freq for haploids

- $\Delta p_t = \frac{w_1 - w_2}{\bar{w}} p_t q_t = \frac{s}{\bar{w}} p_t q_t$

- $\approx s p_t q_t$

- Selection against each allele is equivalent to the haploid selection with $s/2$

- Generations to reach certain frequency in haploids

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

- Generations to reach certain frequency in diploids

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

- Starting frequency

- $p_0 = 1/(2N)$

- Near fixation frequency

- $p_\tau = 1 - 1/(2N)$

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

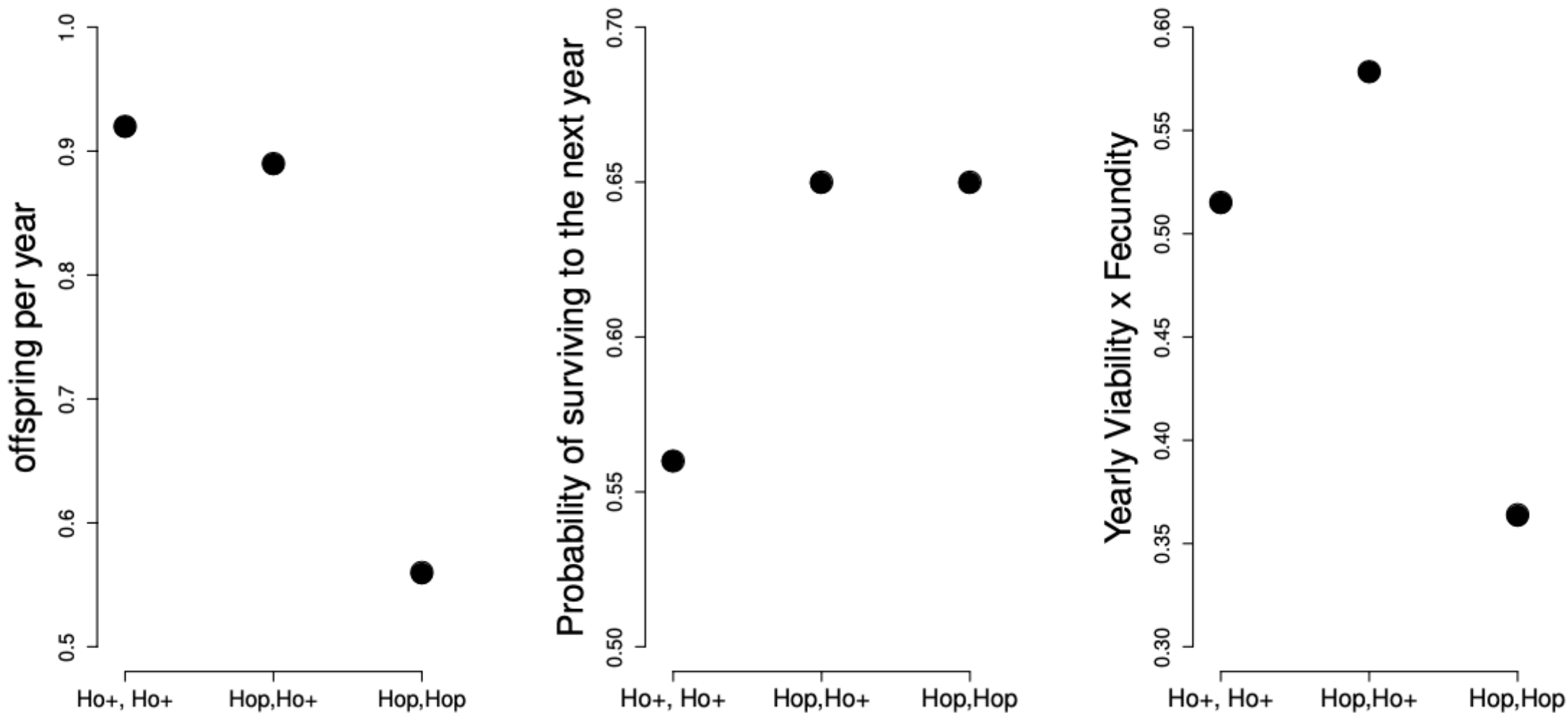
- Compared with the haploid time:

- $$\tau \approx \frac{2}{s} \log(N)$$

Balancing selection with heterozygote advantage

- Heterozygote advantage
 - Heterozygotes are fitter than either of the homozygotes.
- When allele A1 is rare
 - it is often found in A1A2, while the A2 allele is usually in the A2A2 state,
 - A1 is more fit and increases in frequency.
- Vice versa for when A2 is rare
- A form of balancing selection

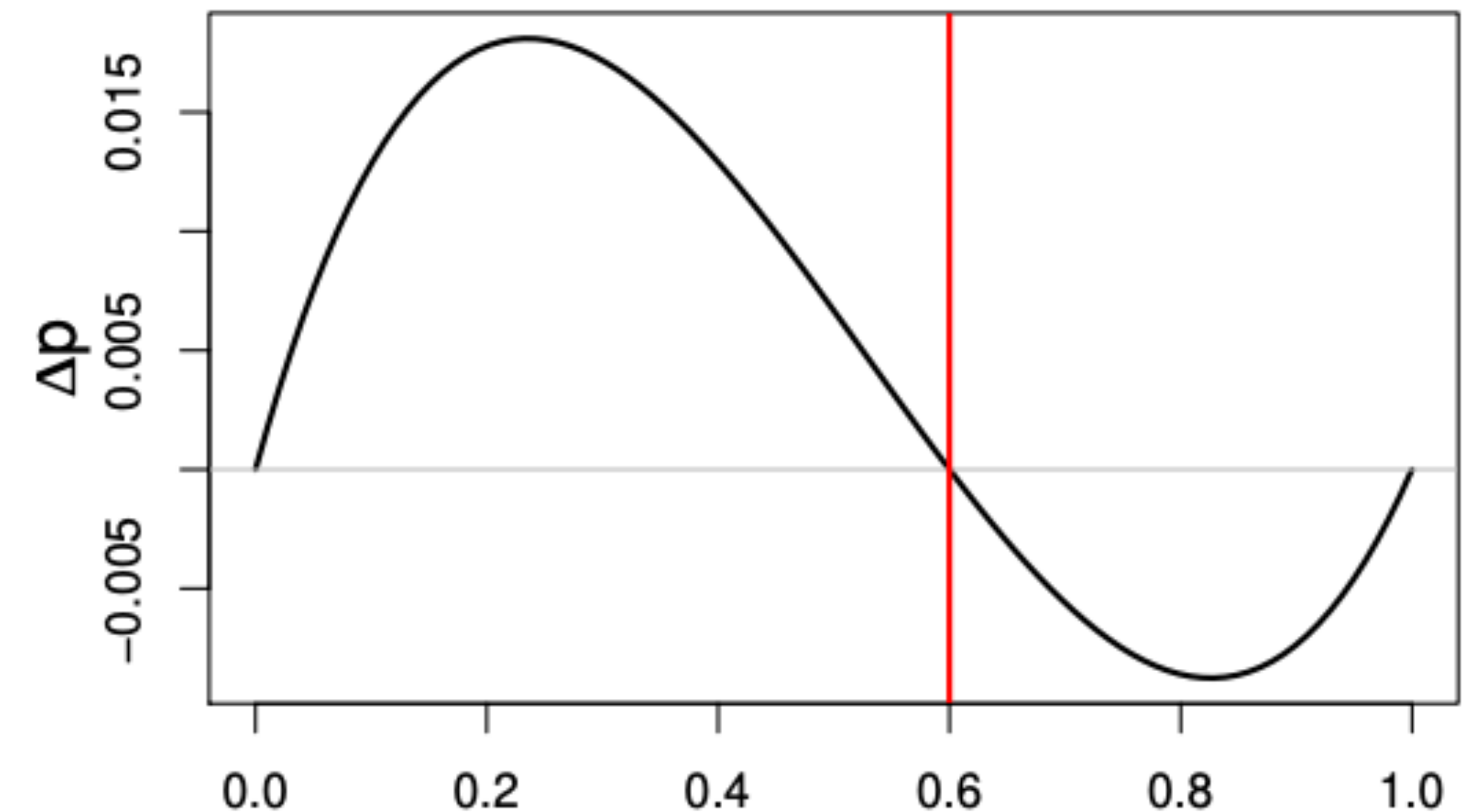
genotype	A_1A_1	A_1A_2	A_2A_2
absolute fitness	w_{11}	$\langle w_{12} \rangle$	w_{22}
relative fitness (generic)	$w_{11} = W_{11}/W_{12}$	$w_{12} = W_{12}/W_{12}$	$w_{22} = W_{22}/W_{12}$
relative fitness (specific)	$1 - s_1$	1	$1 - s_2$



For the three Soay sheep genotypes: the offspring per year (left), the probability of surviving a year (middle), and the product of the two (right). Johnston et al. (2013).

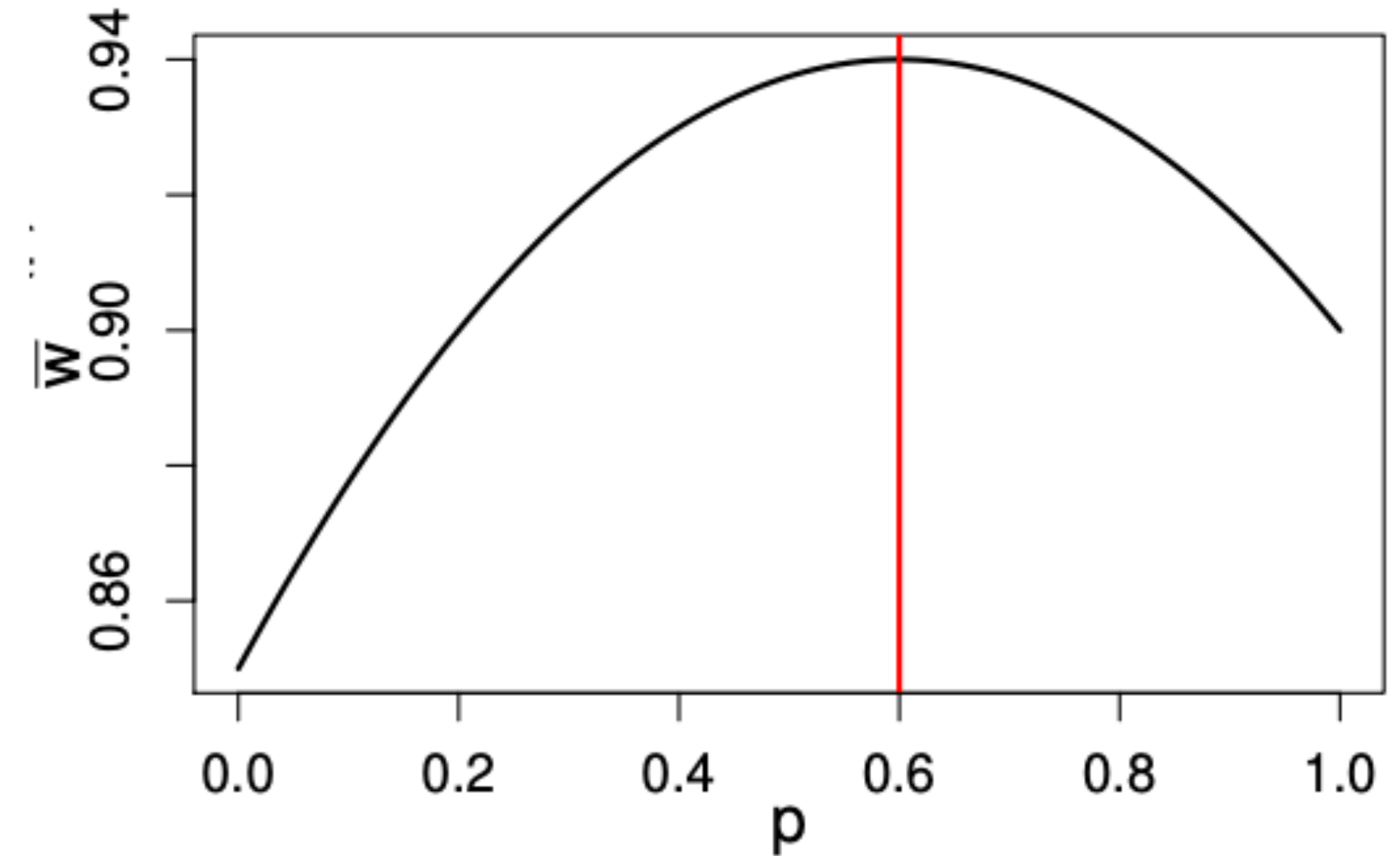
Balancing selection with heterozygote advantage

- equilibrium frequency by setting $\Delta p_t = 0$
- $p_t q_t (w_1 - w_2) = 0$
- Equilibria at
- $p = 0$ or $q = 0$
- but also
- $p_e = \frac{s_2}{s_1 + s_2}$
- Stable equilibrium!



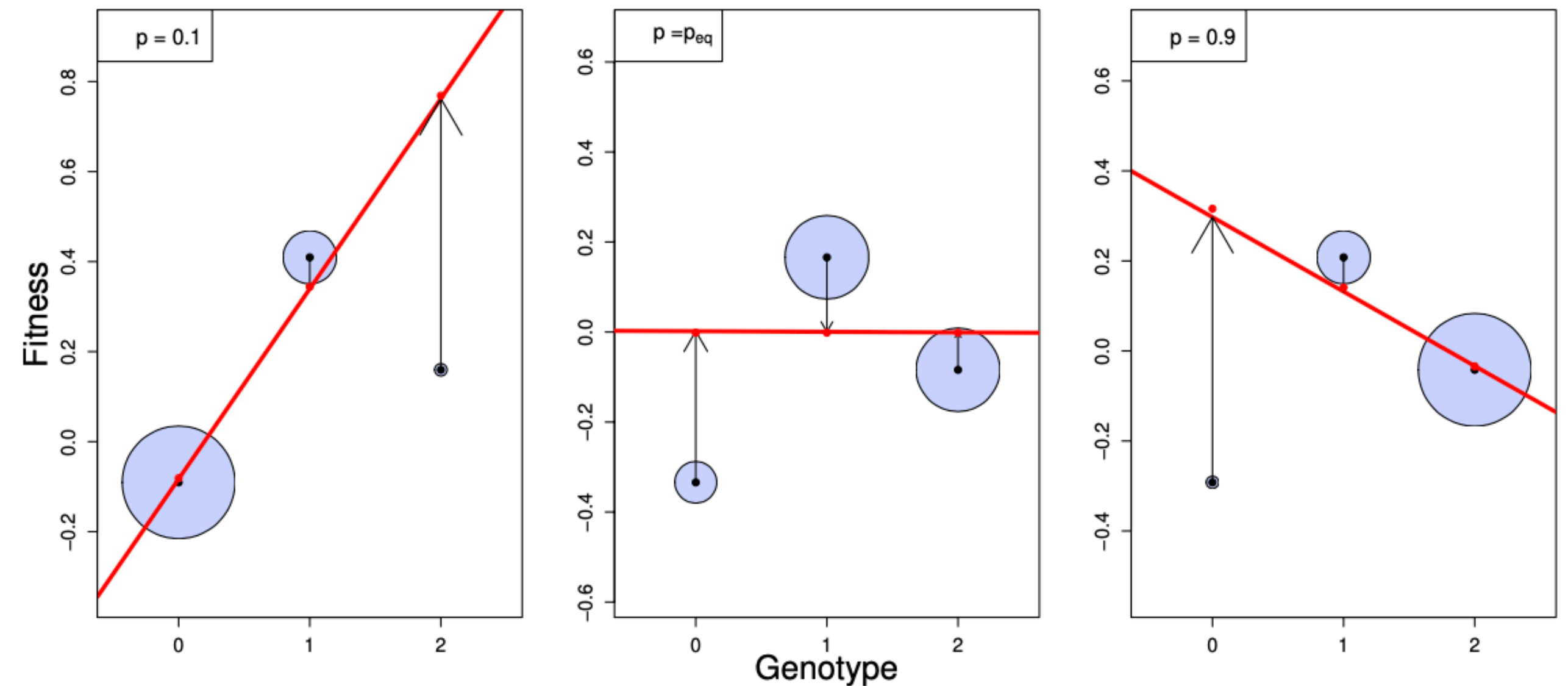
Balancing selection with heterozygote advantage

- Mean fitness \bar{w} maximized at
- $$p_e = \frac{s_2}{s_1 + s_2}$$
- The highest possible fitness of the population would be achieved if every individual was a heterozygote.
- However, Mendelian segregation of alleles in the gametes of heterozygotes means that a sexual population can never achieve a completely heterozygote population.
- This equilibrium frequency represents an evolutionary compromise between the advantages of the heterozygote and the comparative costs of the two homozygotes.



Balancing selection with heterozygote advantage

- When frequency of the A1 allele is below the equilibrium frequency.
 - Regression of fitness on the number of A1 alleles in a genotype has a positive slope.
 - $\bar{w}_1 > \bar{w}_2$
- When frequency of the A1 allele is above the equilibrium frequency.
 - Regression of fitness on the number of A1 alleles in a genotype has a negative slope.
 - $\bar{w}_1 < \bar{w}_2$



The deviations of the fitness of each genotype away from the mean population fitness (0) are shown as black dots. The area of each circle is proportion to the fraction of the population in each genotypic class (p^2 , $2pq$, and q^2). The additive genetic fitness of each genotype is shown as a red dot.

Question 7.

Assume that the frequency of the Ho^P allele is 10%, that there are 1000 males at birth, and that individual adults mate at random.

A) What is the expected number of males with each of the three genotypes in the population at birth?

B) Assume that a typical male individual of each genotypes has the following probability of surviving to adulthood:

$Ho^+ Ho^+$	$Ho^+ Ho^P$	$Ho^P Ho^P$	Making the assumptions
0.5	0.8	0.8	

from above, how many males of each genotype survive to reproduce?

C) Of the males who survive to reproduce, let's say that males with the Ho^+Ho^+ and Ho^+Ho^P genotype have on average 2.5 offspring, while $Ho^P Ho^P$ males have on average 1 offspring.

Taking into account both survival and reproduction, how many offspring do you expect each of the three genotypes to contribute to the total population in the next generation?

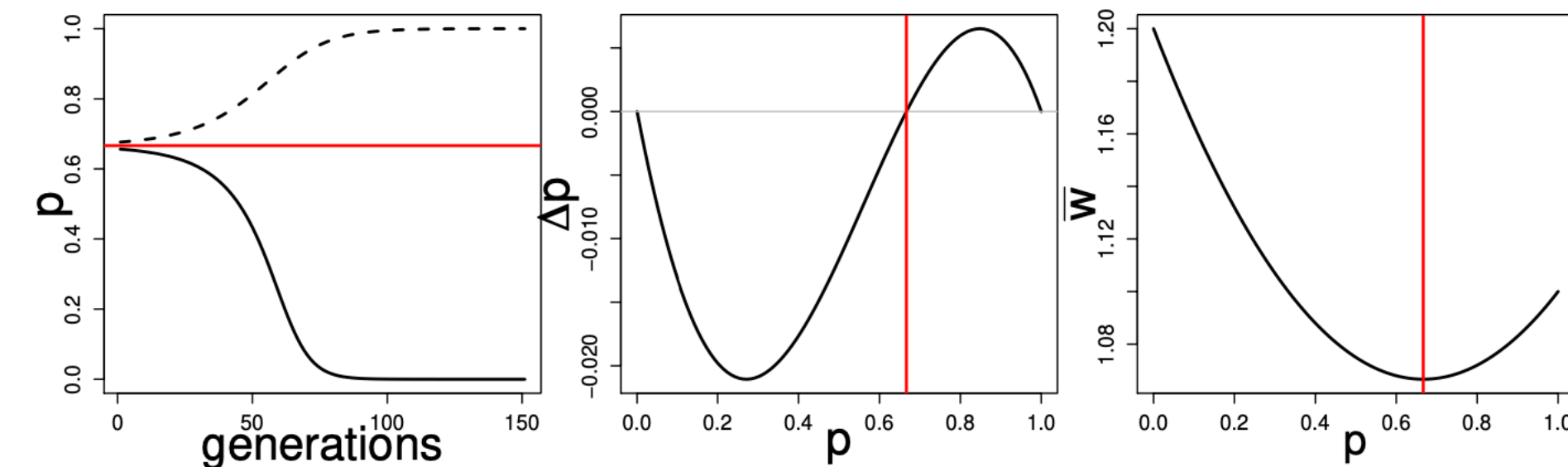
D) What is the frequency of the Ho^+ allele in the sperm that will form this next generation?

E) How would your answers to B-D change if the Ho^P allele was at 90% frequency?

Selection with underdominance

- Underdominance: the heterozygote is less fit than either of the two homozygote
- Three equilibria:
 - $p = 0$, $p = 1$, and a polymorphic equilibrium $p = p_U$
- However, p_U is unstable
 - If $p < p_U$, then Δp_t is negative and allele A1 will be lost
 - If $p > p_U$, allele A_1 will become fixed.

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	$1 + s_2$



Fluctuating selection pressures

- Selection pressures are rarely constant through time due to environmental change.
- Selection pressures on a polymorphism change the frequency of the allele can fluctuate along with them.
- This can have important implications for which alleles can survive and spread.

Haploid model with fluctuating selection

- Fitnesses depend on time
- $w_{1,t}$ and $w_{2,t}$: fitnesses of the two types in generation t
- Frequency of allele A1 in generation $t + 1$

- $$p_{t+1} = \frac{w_{1,t}}{\bar{w}_t} p_t$$

- The ratio of the frequency of allele A1 to that of allele A2 in generation $t + 1$ is

- $$\frac{p_{t+1}}{q_{t+1}} = \frac{w_{1,t} p_t}{w_{2,t} q_t}$$

Haploid model with fluctuating selection

- The ratio of the frequency of allele A1 to that of allele A2 after τ generations is

- $$\frac{p_\tau}{q_\tau} = \prod_{i=1}^{\tau} \frac{w_{1,i}}{w_{2,i}} \frac{p_1}{q_1}$$

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

- $\sqrt[\tau]{\prod_{i=1}^{\tau} w_{1,i}}$: geometric mean fitness of allele A1

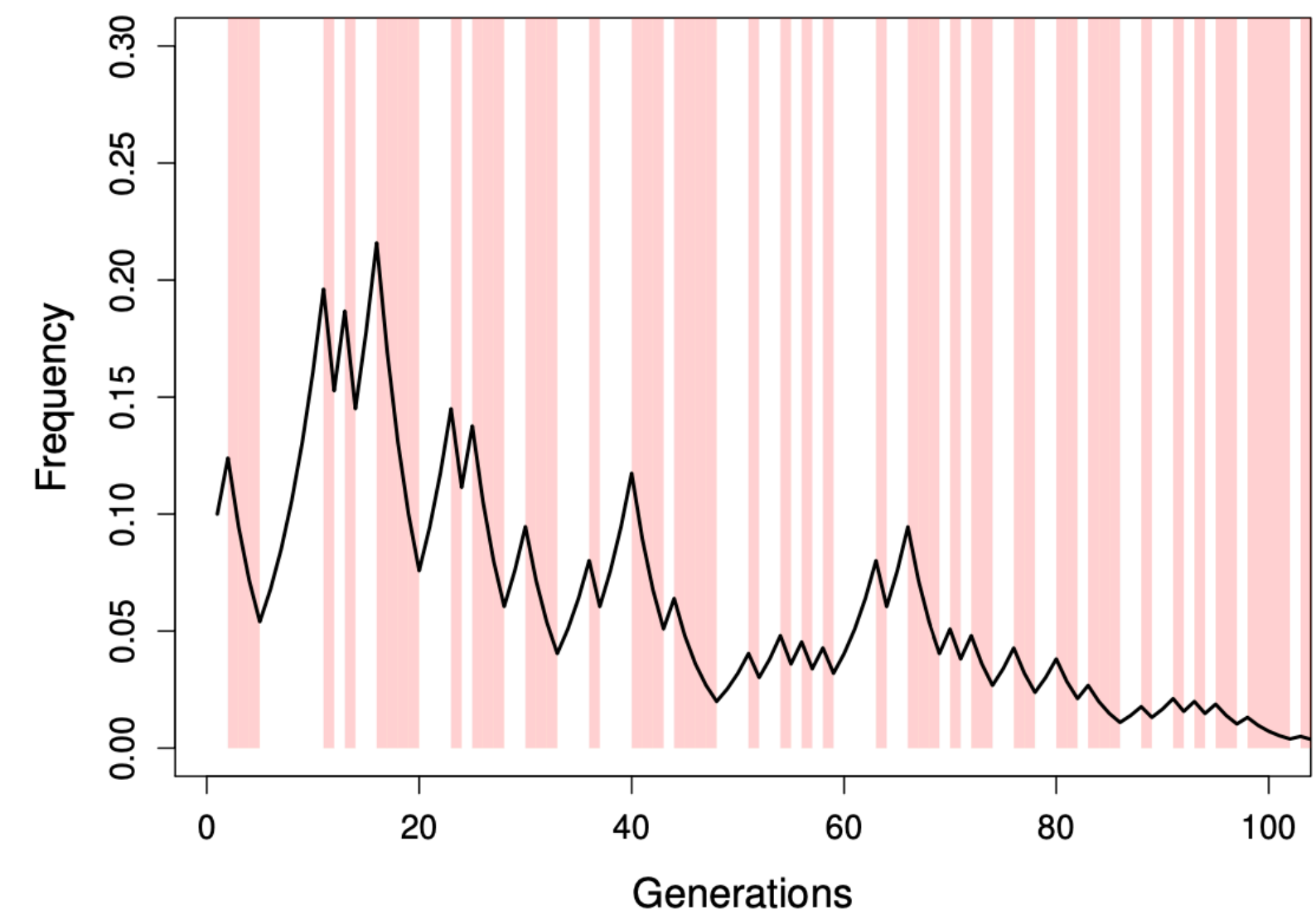
Haploid model with fluctuating selection

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

-
- A1 will only increase in frequency if it has a higher geometric mean fitness than allele A2 (at least in our simple deterministic model).
- An allele with higher geometric mean fitness can even invade and spread to fixation if its (arithmetic) mean fitness is lower than the dominant type

	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

Table 10.2: Fitnesses of two alleles in wet and dry years. Means calculated assuming equal chances of wet and dry years. The geometric mean is calculated as $\sqrt{w_{\text{wet}} w_{\text{dry}}}$. Example numbers taken from [SEGER and BROCKMANN \(1987\)](#).



Haploid model with fluctuating selection

- Write the relative fitness of an allele in a given generation: $w_i = 1 + s_i$

- $$\bar{g} = \sqrt[\tau]{\prod_{i=1}^{\tau-1} 1 + s_i}$$

- $$\begin{aligned} \log(\bar{g}) &= \frac{1}{\tau} \sum_{i=1}^{\tau-1} \log(1 + s_i) \\ &= \mathbb{E} \left[\log(1 + s_i) \right] \end{aligned}$$

- $$\begin{aligned} \log(\bar{g}) &\approx \mathbb{E} \left[s_i - s_i^2/2 \right] \\ &= \mathbb{E} \left[s_i \right] - \text{var}(s_i)/2 \end{aligned}$$

- Genotypes with high arithmetic mean fitness can be selected against, i.e. have low geometric mean fitness against, if their fitness has too high a variance across generations

Diploid fluctuating fitness

- Selection pressures fluctuate over time and can potentially maintain polymorphisms in the population.

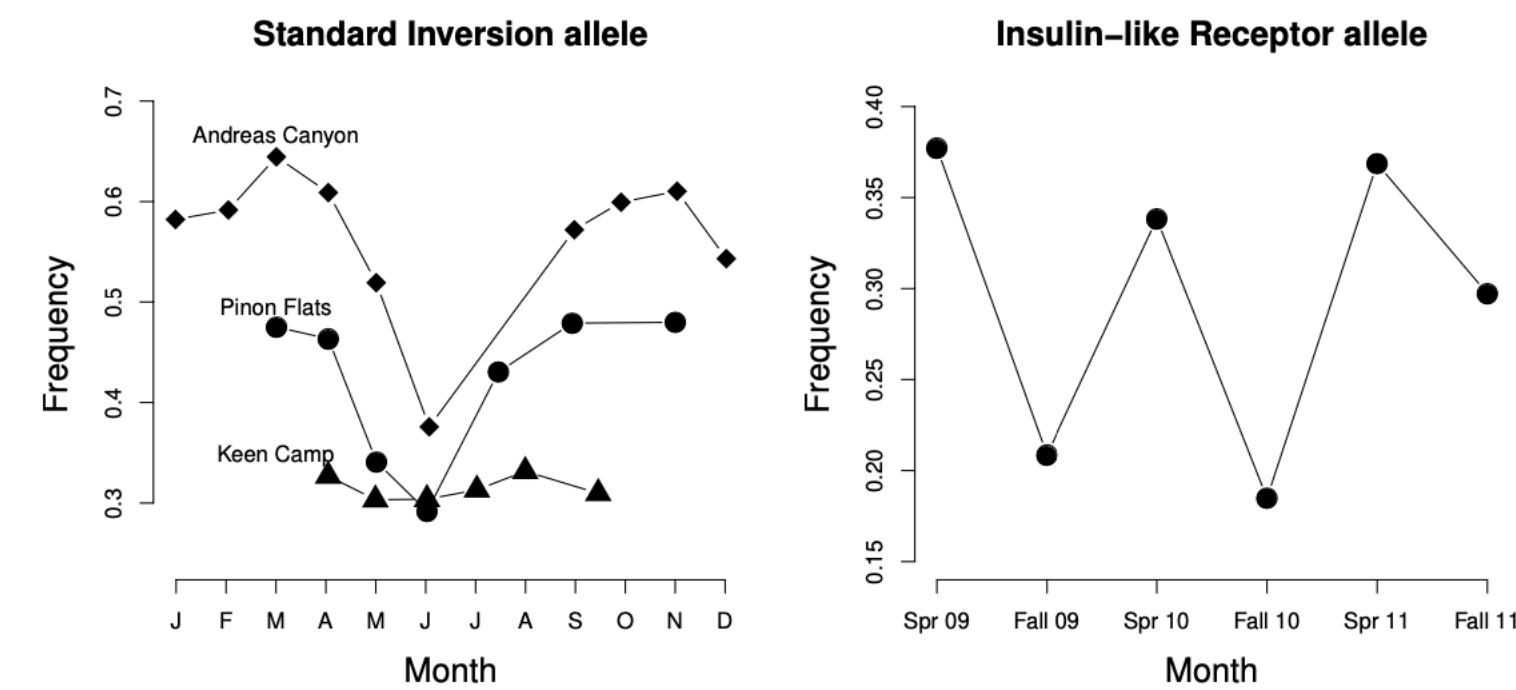


Figure 10.23: **Left)** Seasonal variation in the mean frequency of the 'Standard' inversion allele in *Drosophila pseudoobscura* for three populations from Mount San Jacinto, CA. These frequencies are an average over four years. Data from [WRIGHT and DOBZHANSKY \(1946\)](#). **Right)** The frequency of an allele at the *Insulin-like Receptor* gene over three years in *Drosophila melanogaster* samples from an orchard in Pennsylvania. Data from [PAABY et al. \(2014\)](#). Note the difference in *x* axes here. Code [here](#).

Diploid fluctuating fitness

- Modeling the diploid case with time-dependent fitness is much less tractable than the haploid case, as segregation makes it tricky to keep track of the genotype frequencies.
- $w_{11,t}$ $w_{12,t}$ $w_{22,t}$ - Relative fitness at time t
- We can think about how the frequency of allele A1 changes when it is rare following Haldane (1963)

Diploid fluctuating fitness

- **When A1 is rare ($p_t \ll 1$)**
- frequency of A1 in the next generation can be approximated as

- $$p_{t+1} \approx \frac{w_{12}}{\bar{w}} p_t.$$

- Ratio of the frequency in the next generation

- $$\frac{p_{t+1}}{q_{t+1}} = \frac{w_{12,t}}{w_{22,t}} \frac{p_t}{q_t}$$

Diploid fluctuating fitness

- $\frac{p_{t+1}}{q_{t+1}} = \left(\prod_{i=0}^t \frac{w_{12,i}}{w_{22,i}} \right) \frac{p_0}{q_0}$
- frequency of allele A1 will increase when rare only if

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

- i.e. if the heterozygote has higher geometric mean fitness than the A2A2 homozygote

Diploid fluctuating fitness

- **When A1 is common** ($1 - p_t \ll 1$)

- Can allele A1 approach fixation in the population?

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

- For allele A1 to approach fixation, we need the geometric mean of $w_{11,i}$ to be greater than the geometric mean fitness of heterozygotes $w_{12,i}$
- If instead heterozygotes have higher geometric mean fitness than the A1A1 homozygotes, then the A2 allele will increase in frequency when it is rare.
- we can thus have a balanced polymorphism even if the heterozygote is never the fittest genotype in any generation, as long as the heterozygote has a higher geometric mean fitness than either of the homozygotes.

Diploid fluctuating fitness

- Toy example
- A plant population found in one of two different environments each generation that occur randomly with equal likelihood
- $\prod_{i=0}^t w_{AA,i} \approx w_{AA,dry}^{t/2} w_{AA,wet}^{t/2}$
- Geometric mean: $w_{AA,dry}^{1/2} w_{AA,wet}^{1/2}$
- Heterozygote advantage, both alleles can invade the pop when rare

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