

Principles of evolution (BIO 351)



University of
Zurich^{UZH}



Instructor: Pouria Dasmeh

Group Leader in Evolutionary Medicine, Institute for Human Genetics,
Marburg University.

Long-term research fellow, Shakhnovich Lab., Harvard University.



<https://github.com/dasmeh>



@PouriaDasmeh

Dasmeh@staff.uni-marburg.de

Overview

Hardy-Weinberg equilibrium principle (review)

Selection at one locus with two alleles

Mutation-selection balance

Has the global AIDS epidemic caused an increase in the frequency of CCR5-Δ32 through natural selection?

to answer questions like this, we need to know what would happen to the allele in the absence of selection (or of any other evolutionary force).

the Hardy-Weinberg law provides the answer.

A simple, idealized life cycle

population of diploid organisms

two alleles, with frequencies p, q in the
gamete pool (gene pool):

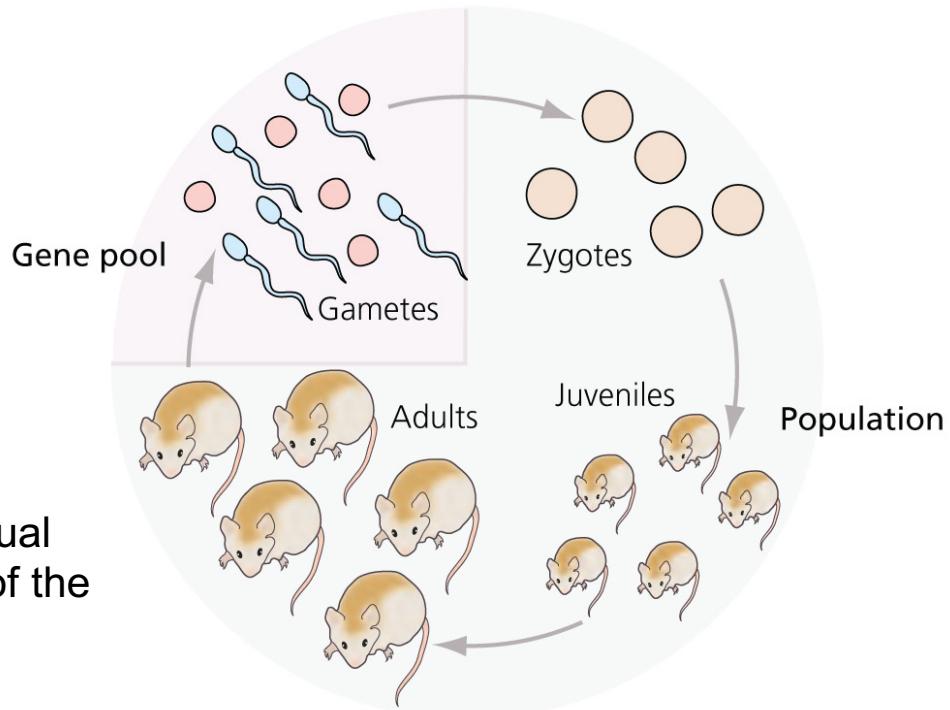
Example: A: $p=0.6$

a: $q=0.4$ (Note: $p+q=1$)

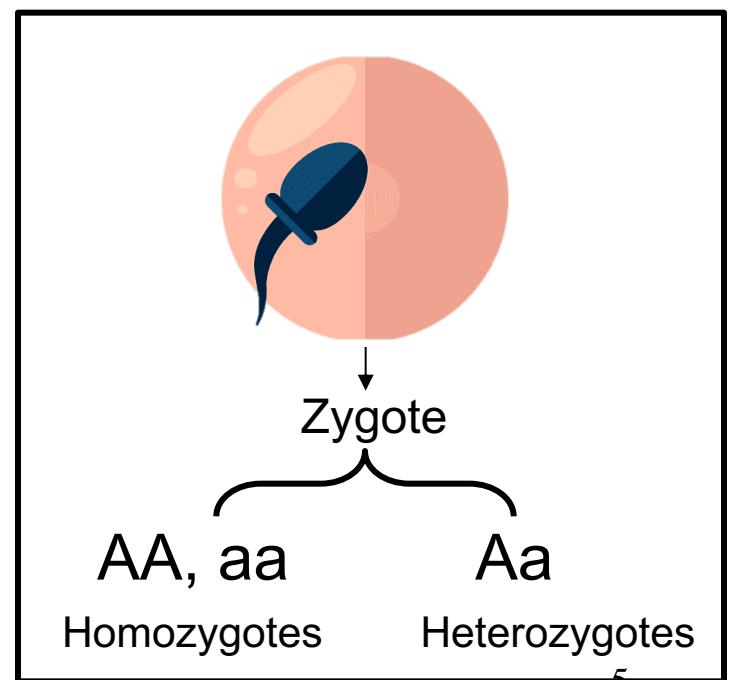
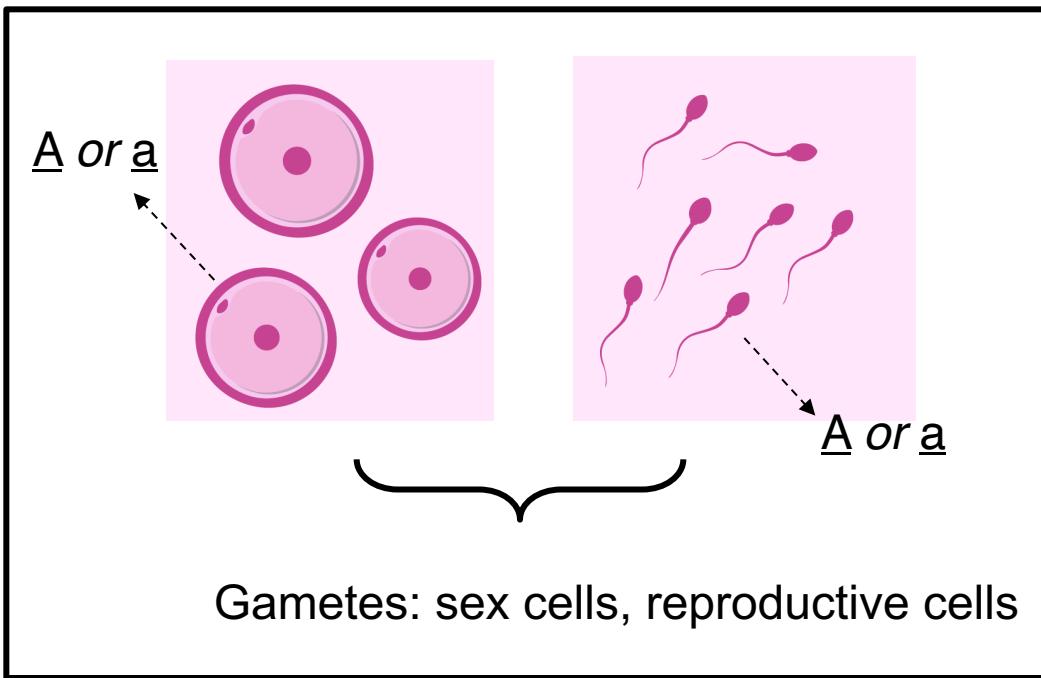
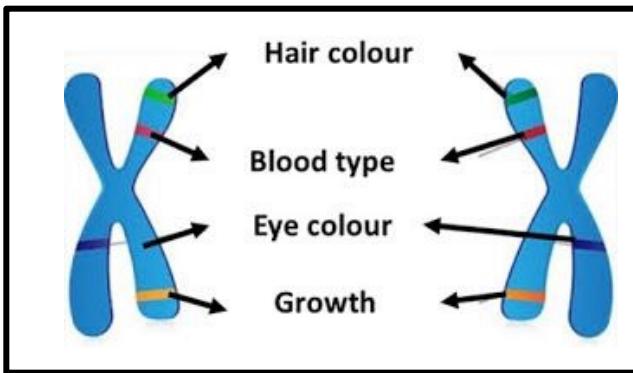
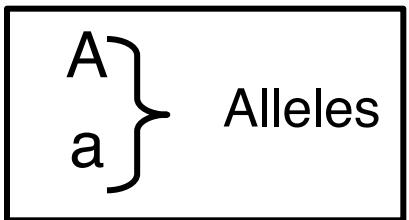
random mating produces zygotes of
three genotypes: AA, Aa, aa

all zygotes equally likely to survive

each kind of genotype contributes an equal
number of gametes to the gamete pool of the
next generation



Quick review (*biology*)



Quick review (*math*)

If two events are independent, their joint probability is equal to the product of their probability

Probability of a sunny day in Zurich ~ 0.37

Probability of a public holiday ~ (12/365) ~ 0.03

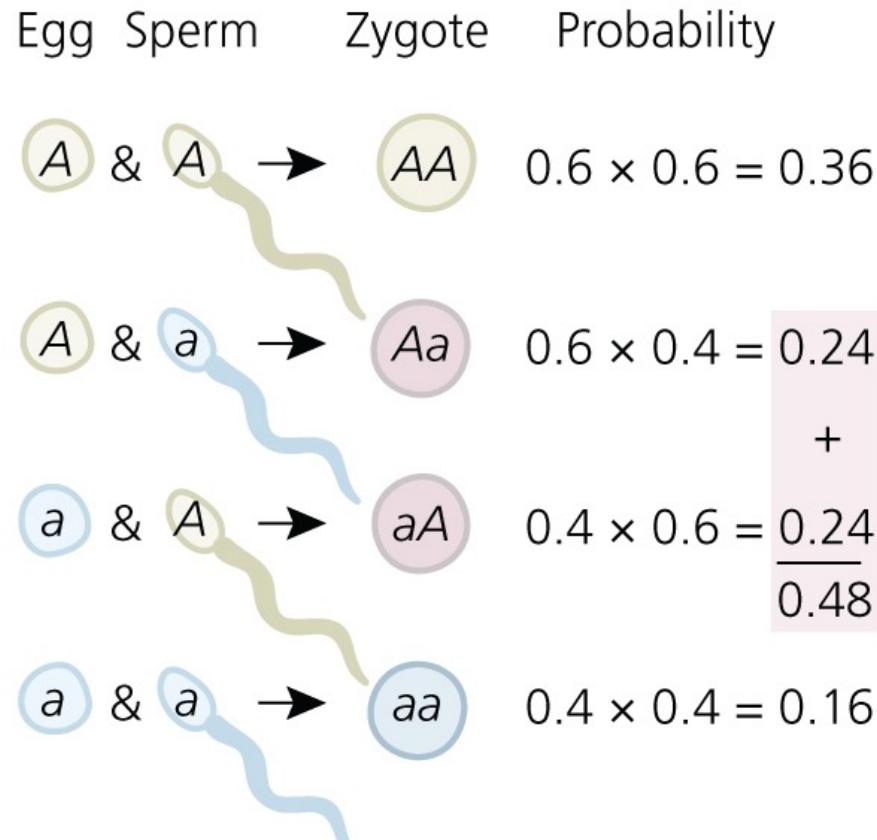
Probability of enjoying a sunny day outside ~ $0.03 \times 0.37 \sim 0.01$

Probability of a sunny day in Zurich ~ 0.37

Probability of a work day ~ (250/365) ~ 0.68

Probability of enjoying a sunny day in office ~ $0.37 \times 0.68 \sim 0.25$

Formation of genotypes from gametes



Note: Genotype frequencies add up to one
 $0.36+0.48+0.16=1$

Formation of genotypes from gametes (general case)

Egg	Sperm	Zygote	Probability
-----	-------	--------	-------------

A_1 & $A_1 \rightarrow A_1A_1 \quad p \times p = p^2$

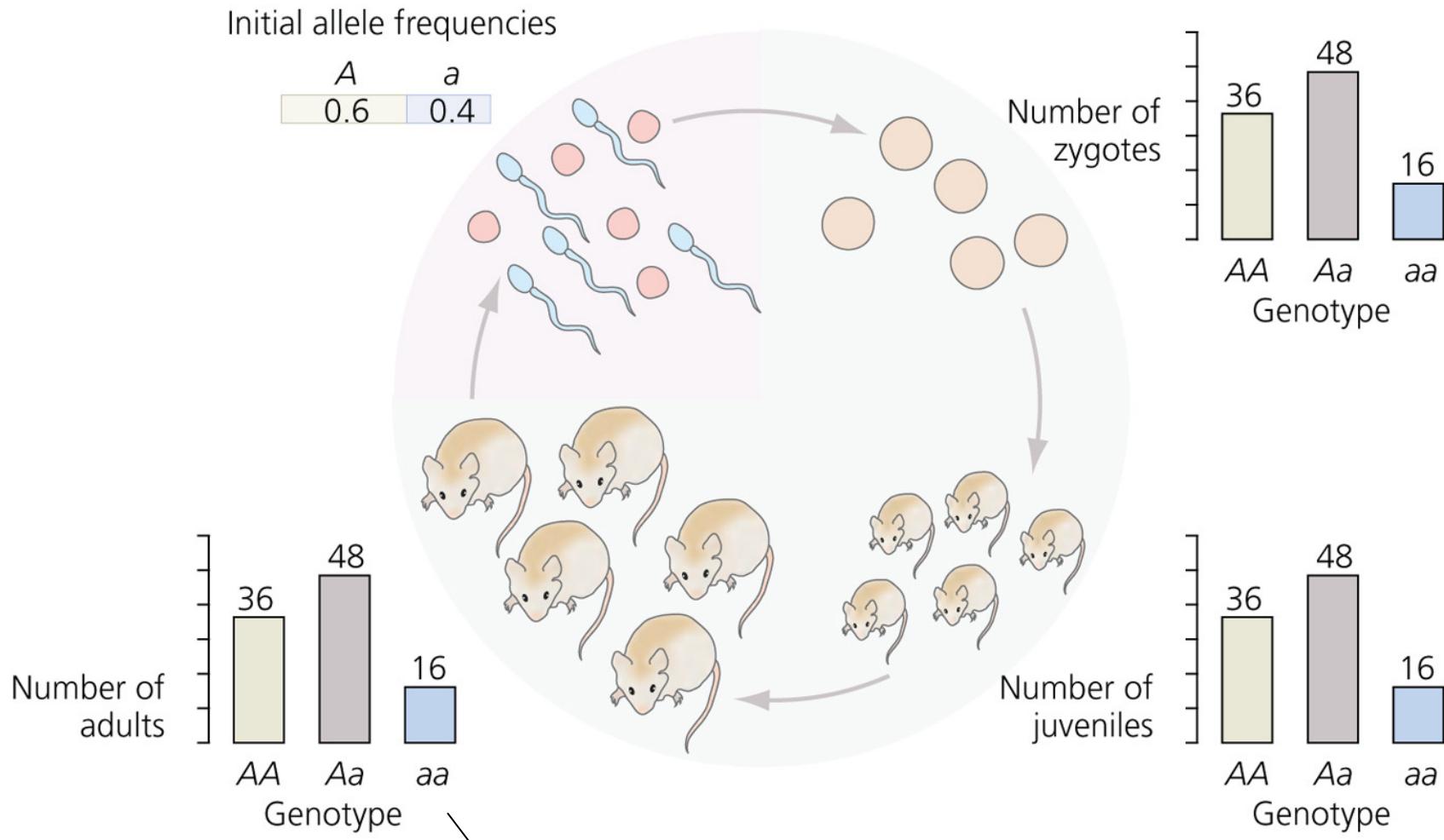
A_1 & $A_2 \rightarrow A_1A_2 \quad p \times q = pq$

A_2 & $A_1 \rightarrow A_2A_1 \quad q \times p = \frac{qp}{2pq}$

A_2 & $A_2 \rightarrow A_2A_2 \quad q \times q = q^2$

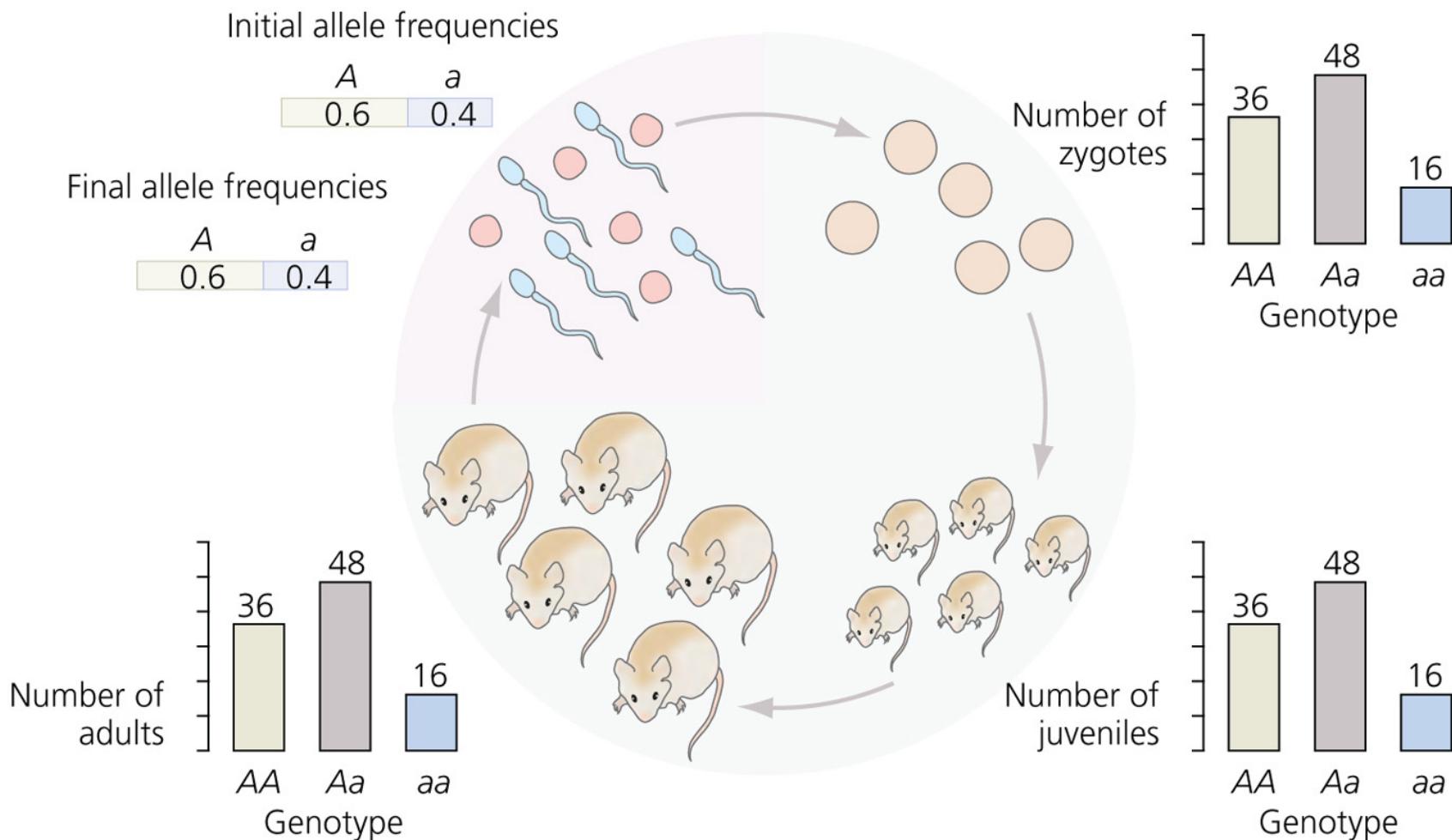
Note: Genotype frequencies add up to one
 $p^2 + 2pq + q^2 = 1$

Class activity



What are the final allele frequencies?

Class activity



The Hardy-Weinberg equilibrium principle

allele frequencies (p,q) do not change over time

for allele frequencies of p and q , genotype frequencies do not change, and they are equal to $p^2:2pq:q^2$.

analogous principles exist for

- more than two alleles at a locus (Freeman, Box 6.2)
- sex-linked loci
- more than one locus

Some assumptions behind the Hardy-Weinberg equilibrium principle

no selection

equal probability of survival and reproduction

no migration (**gene flow**)

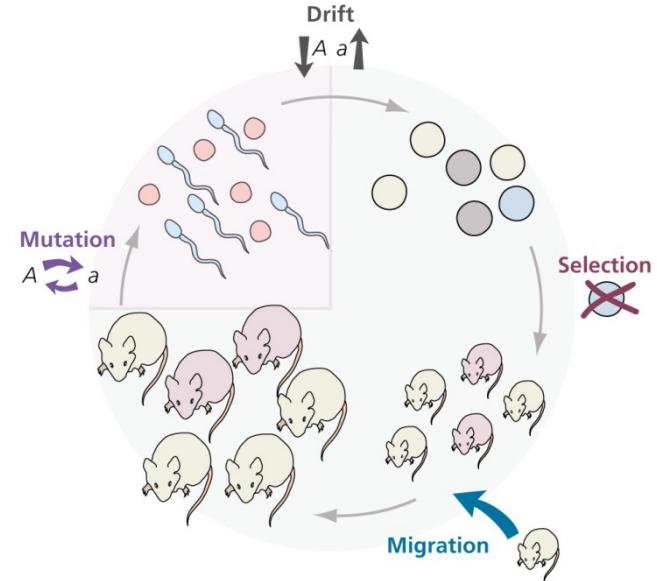
no mutation

between A and a

no **genetic drift**

chance events in gamete formation that occur in small populations

HW principle assumes that populations are effectively infinite in size

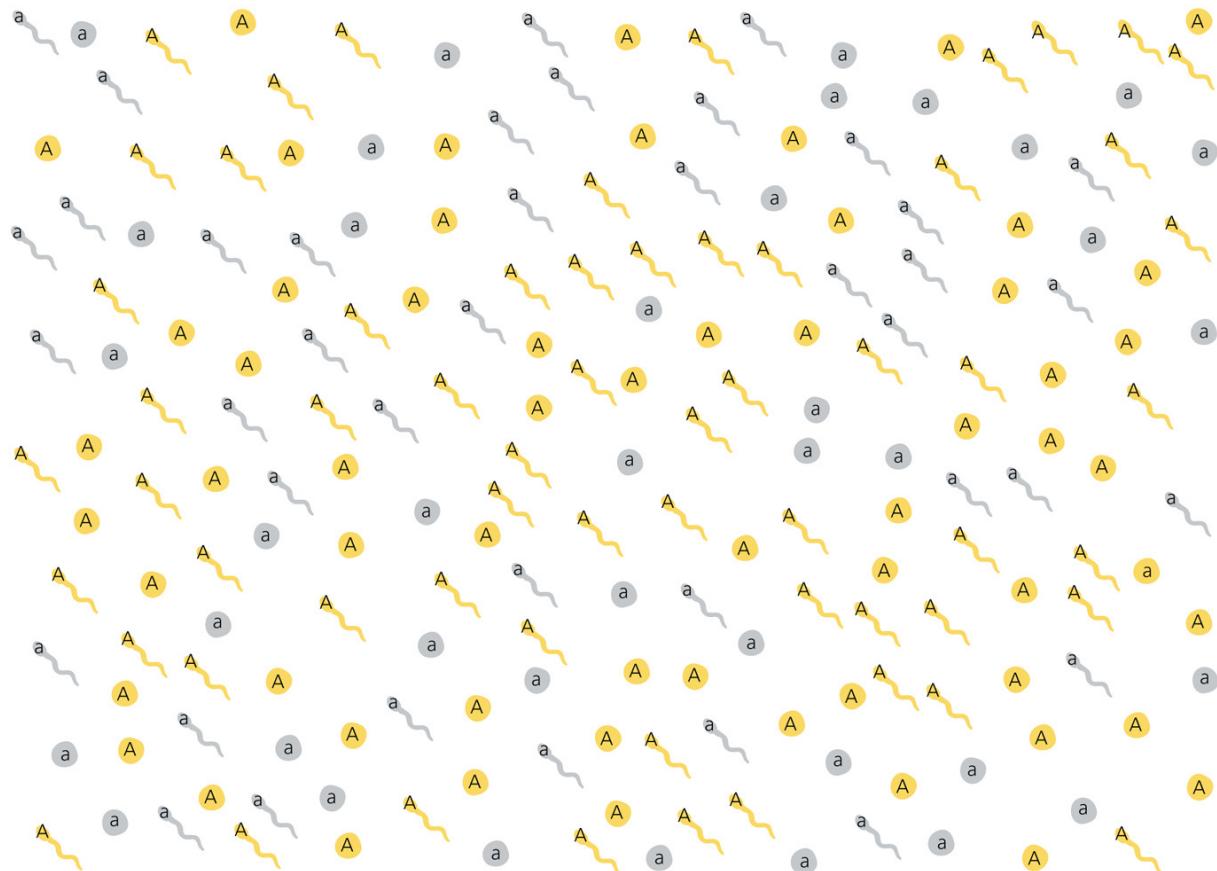


deviations from HW equilibrium can indicate that these assumptions are violated

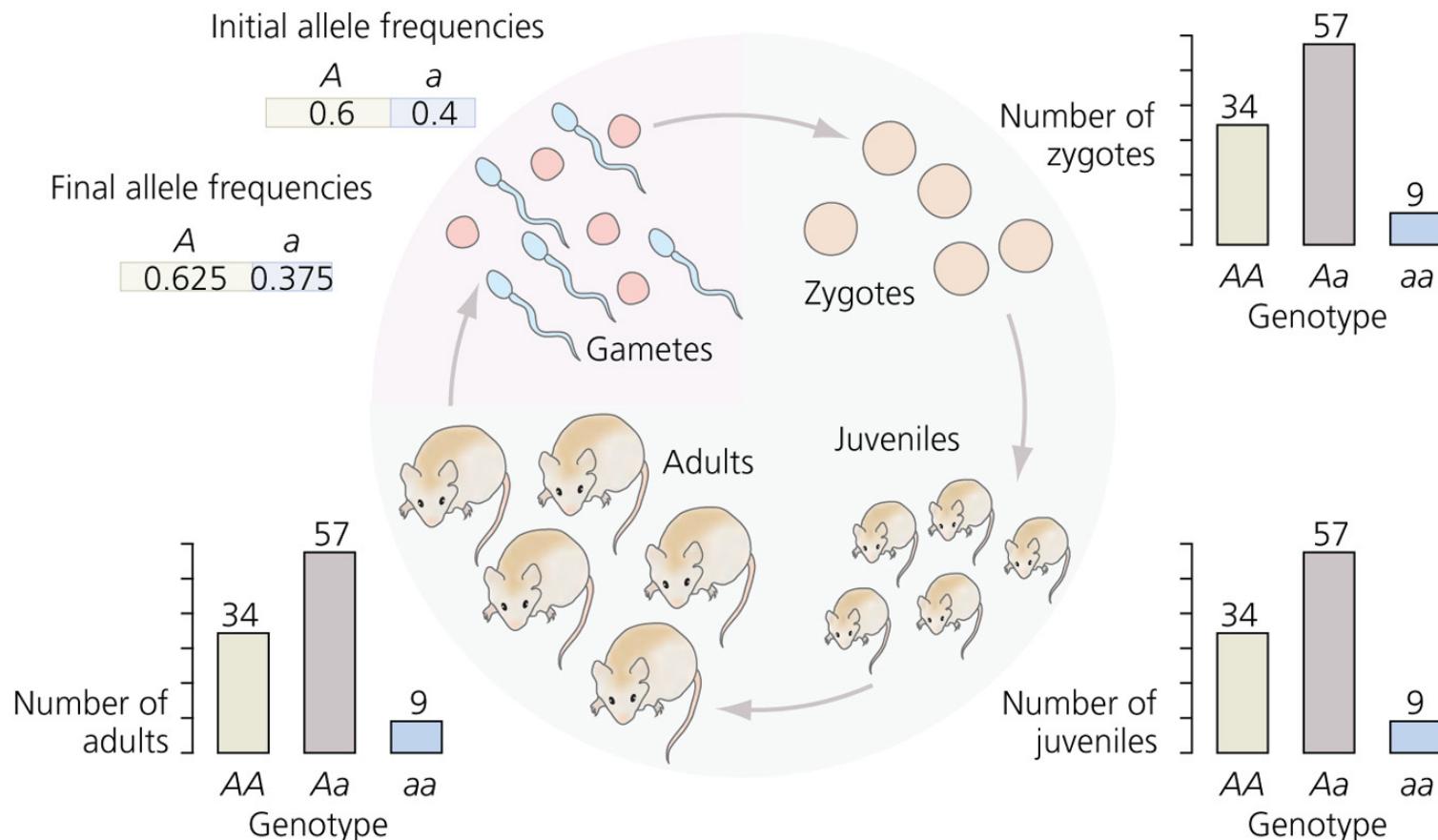
A simple example for the action of genetic drift

imagine a large gene pool
 $p=0.6$ for A
 $q=0.4$ for a

choose at random some
modest number, e.g., 100
pairs of gametes
to form genotypes

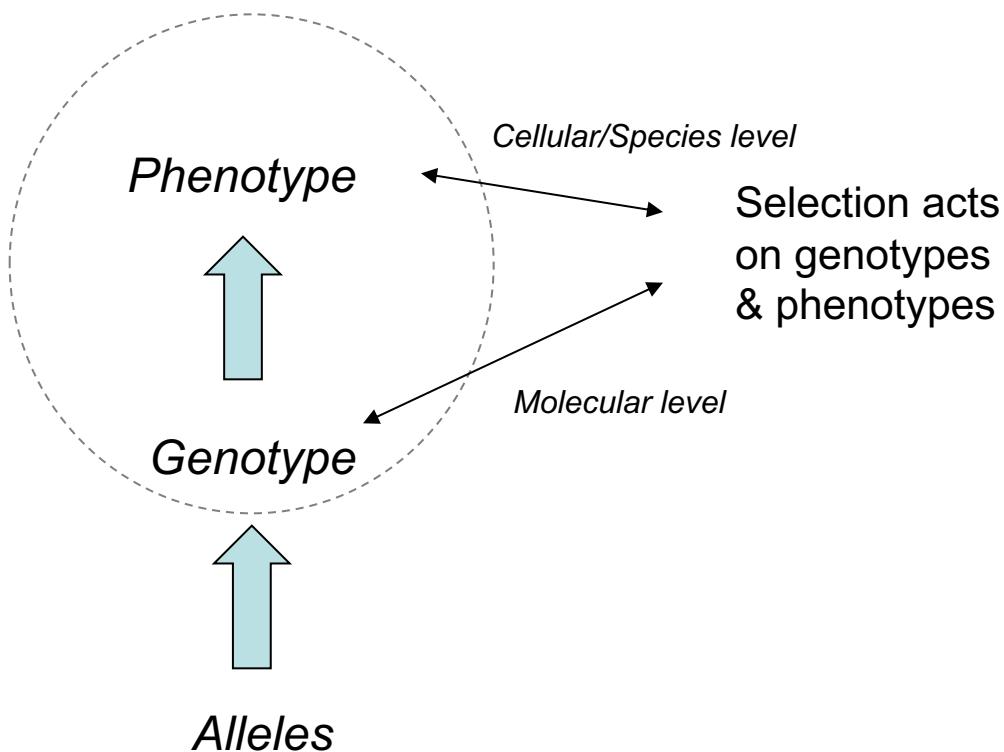


A simple example for the action of genetic drift



genetic drift can change both allele and genotype frequencies through random sampling alone in the absence of selection

A systematic way to think about all these!



A systematic way to think about all these!

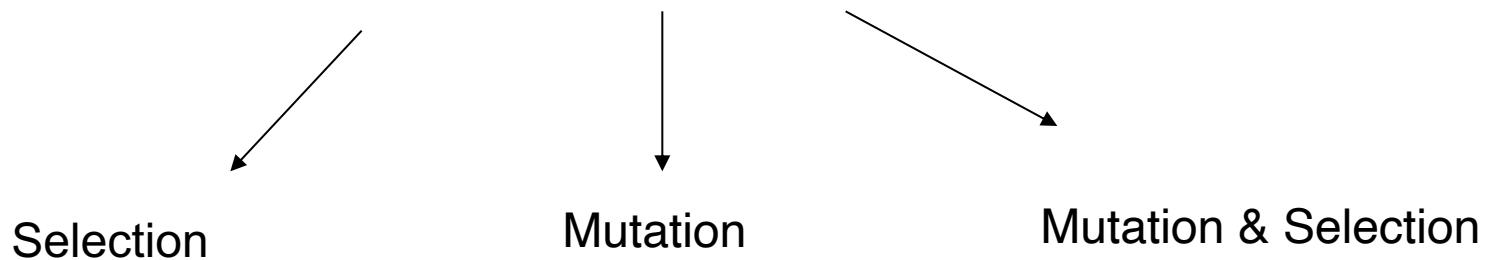
	Allele		Genotype		
	A	a	AA	Aa	aa
probability	p	q	p^2	$2pq$	q^2

constraint 1: $\sum p(\text{Alleles}) = 1$

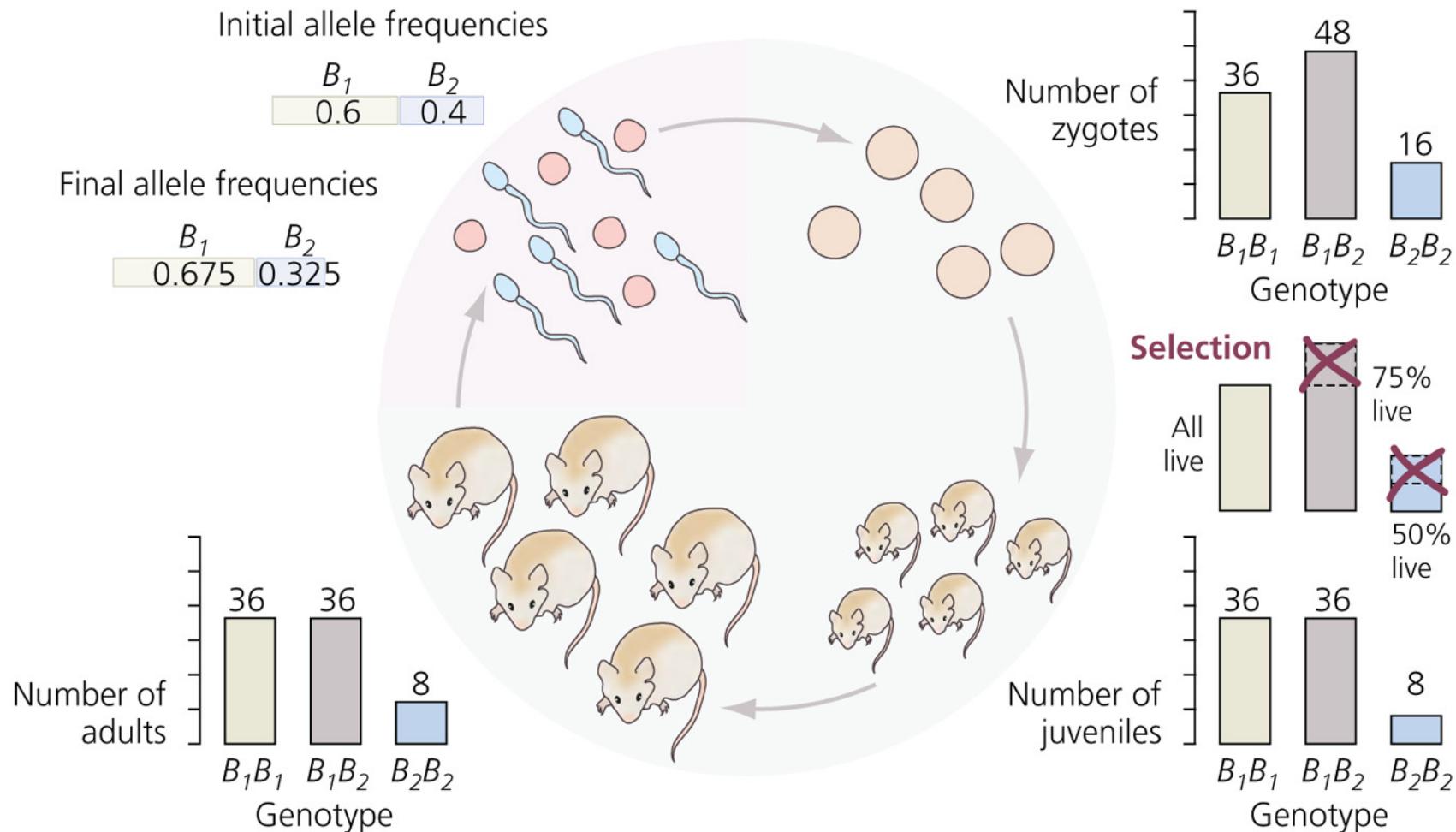
constraint 2: $\sum p(\text{Genotypes}) = 1$

Weight factors
or multipliers

Three corrections to HW equilibrium



Selection can change both genotype and allele frequencies



Class activity 1

Selection can change allele frequencies unidirectionally

fruit flies encode an alcoholdehydrogenase (Adh) enzyme

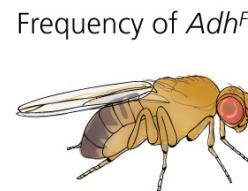
it has two allozymes:

Adh^F (migrates fast in electrophoresis)

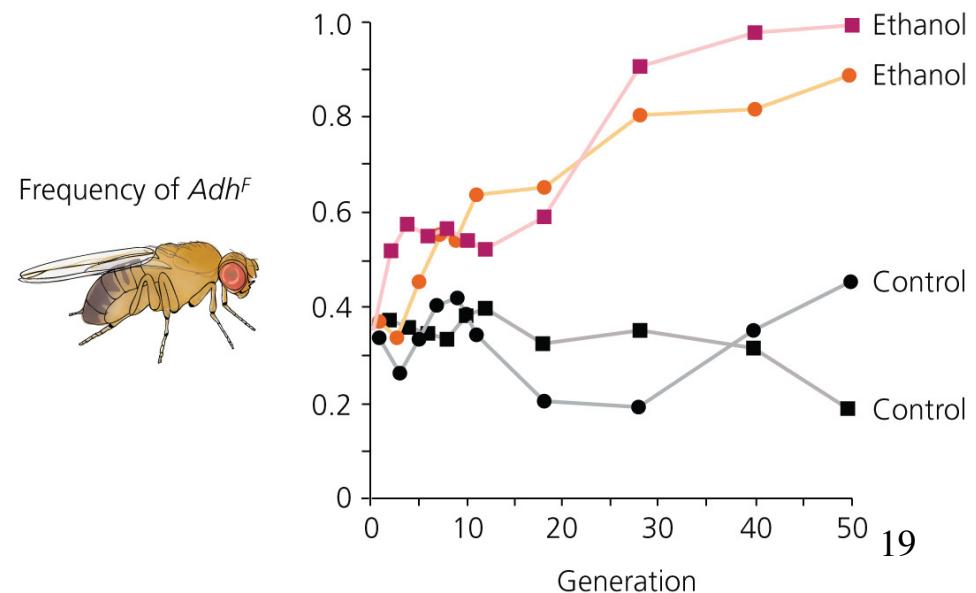
Adh^S (migrates slowly)

Adh^F is superior at detoxifying ethanol

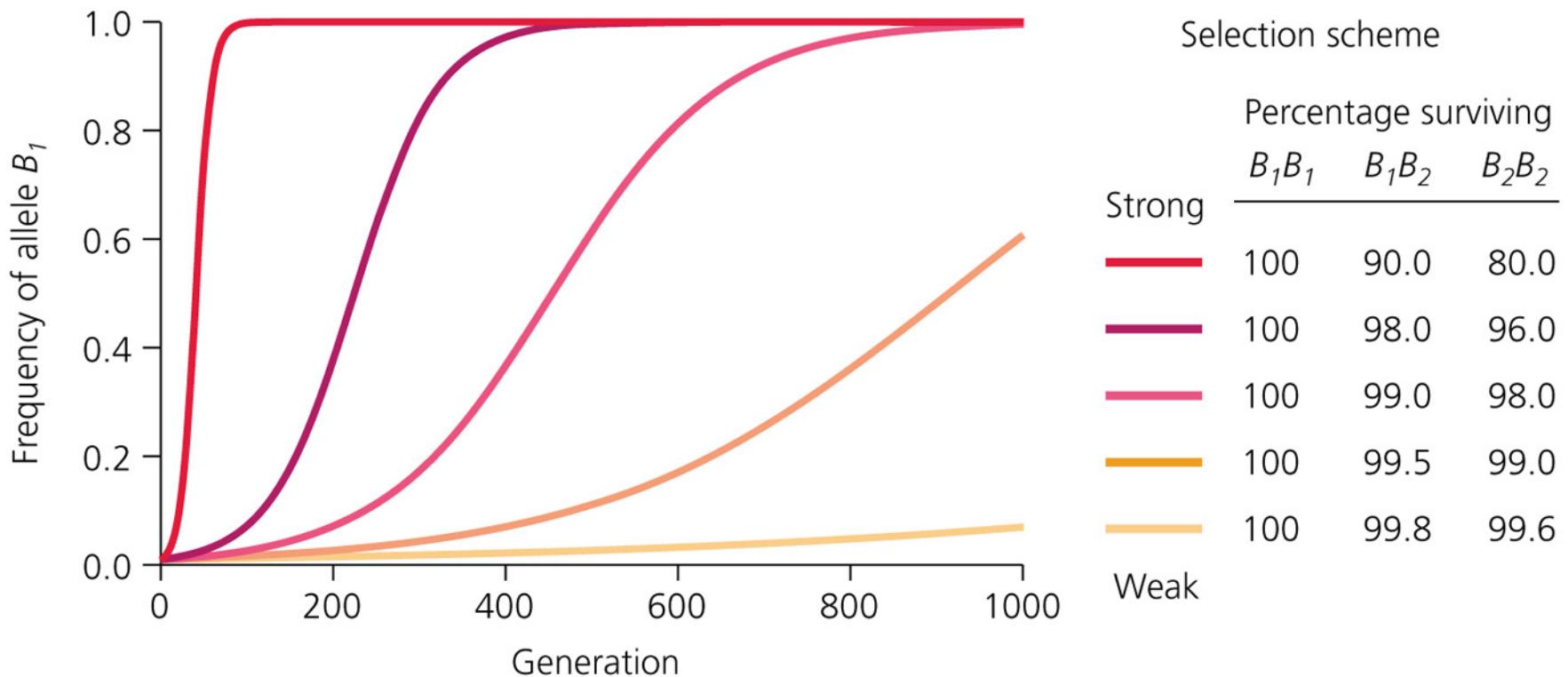
flies raised on food spiked with ethanol show a consistent increase in Adh^F frequency



Frequency of Adh^F



The changes caused by selection accumulate and depend on the strength of selection



Recall: most mutations in the wild are weakly deleterious

A quantitative treatment of selection

How do allele frequencies p, q change as a result of selection?

Viability selection

likelihood that zygotes A_1A_1 survive to adulthood: w_{11}

likelihood that zygotes A_1A_2 survive to adulthood: w_{12}

likelihood that zygotes A_2A_2 survive to adulthood: w_{22}

w_{ij} is the fitness of genotype A_iA_j

ratio of genotype frequencies after selection

$$p^2w_{11} : 2pqw_{12} : q^2w_{22}$$

What are in the boxes?

	Allele		Genotype		
	A	a	AA	Aa	aa
	p	q	p^2 	$2pq$ 	q^2 
			$p^2 w_{11}$	$2pq w_{12}$	$q^2 w_{22}$

A quantitative treatment of selection

to go from the ratio to the actual genotype frequencies, we need to ensure that genotype frequencies add up to one

we need to multiply with a normalization factor

this factor is the **mean population fitness**

$$\bar{w} = p^2 w_{11} + 2pq w_{12} + q^2 w_{22}$$

genotype frequencies after selection

$$\begin{array}{ccc} A_1A_1 & A_1A_2 & A_2A_2 \\ \frac{p^2 w_{11}}{\bar{w}} & \frac{2pq w_{12}}{\bar{w}} & \frac{q^2 w_{22}}{\bar{w}} \end{array}$$

A quantitative treatment of selection

next calculate allele frequencies p' , q' after selection

A_1A_1 contributes this fraction of A_1 gametes: $\frac{p^2 w_{11}}{\bar{w}}$

A_1A_2 contributes this fraction of A_1 gametes: $\frac{pq w_{12}}{\bar{w}}$

overall, we get $p' = \frac{p^2 w_{11} + pq w_{12}}{\bar{w}}$

and analogously $q' = \frac{q^2 w_{22} + pq w_{12}}{\bar{w}}$

note that $p'+q'=1$

A quantitative treatment of selection

it can be useful to know the per-generation change $\Delta p = p' - p$

note that $\Delta q = q' - q = (1 - p') - (1 - p) = p - p' = -(p' - p) = -\Delta p$

$$\Delta p = p' - p = \frac{p^2 w_{11} + pq w_{12}}{\bar{w}} - p = \frac{p^2 w_{11} + pq w_{12} - p \bar{w}}{\bar{w}} = \frac{p}{\bar{w}} (p w_{11} + q w_{12} - \bar{w})$$

equivalently (check for yourself)

$$\Delta p = \frac{pq[p(w_{11} - w_{12}) + q(w_{12} - w_{22})]}{\bar{w}}$$

Class activity 2

Fetal survival in malaria-infected mothers as an example of selection affecting genotype frequencies

Plasmodium falciparum that infects pregnant women invades the placenta

Causes placental inflammation and affects placental development

Can cause spontaneous abortion, premature birth, higher risk of infant death

Fetal cells in the placenta release vascular endothelial growth factor receptor 1 (VEGFR1)

VEGFR1 can influence placental inflammation

Class activity 2

Fetal survival in malaria-infected mothers as an example of selection affecting genotype frequencies

Genotype at VEGFR1 locus can affect outcome of placental malaria

VEGFR1 gene carries two kinds of alleles distinguished by an indel in a stretch of repetitive DNA (L...long, S...short)

SS and SL genotypes produce more VEGFR1 than LL genotypes

Are these genotypes subject to natural selection?

Class activity 2

Fetal survival in malaria-infected mothers as an example of selection affecting genotype frequencies

allele frequencies among 163 infants born from October-April when placental malaria shows an annual low

S	L
0.555	0.445

expected genotype frequencies (from Hardy-Weinberg)

$$\begin{array}{lll} SS & SL & LL \\ (0.555)^2 = 0.308 & 2(0.555)(0.445) = 0.494 & (0.445)^2 = 0.198 \end{array}$$

observed genotype frequencies

$$\begin{array}{lll} SS & SL & LL \\ 49/163 = 0.301 & 83/163 = 0.509 & 31/163 = 0.19 \end{array}$$

expected frequencies are not significantly different from observed frequencies in «off-season» infants

(Chi-square test, Freeman Box 6.4)

Class activity 2

Fetal survival in malaria-infected mothers as an example of selection affecting genotype frequencies

allele frequencies among 76 infants born from May-September when placental malaria shows an annual high

$$\begin{array}{ll} S & L \\ 0.539 & 0.461 \end{array}$$

expected genotype frequencies (from Hardy-Weinberg)

$$\begin{array}{lll} SS & SL & LL \\ (0.539)^2=0.291 & 2(0.539)(0.461)=0.497 & (0.461)^2=0.213 \end{array}$$

observed genotype frequencies

$$\begin{array}{lll} SS & SL & LL \\ 16/76=0.211 & 50/76=0.658 & 10/76=0.132 \end{array}$$

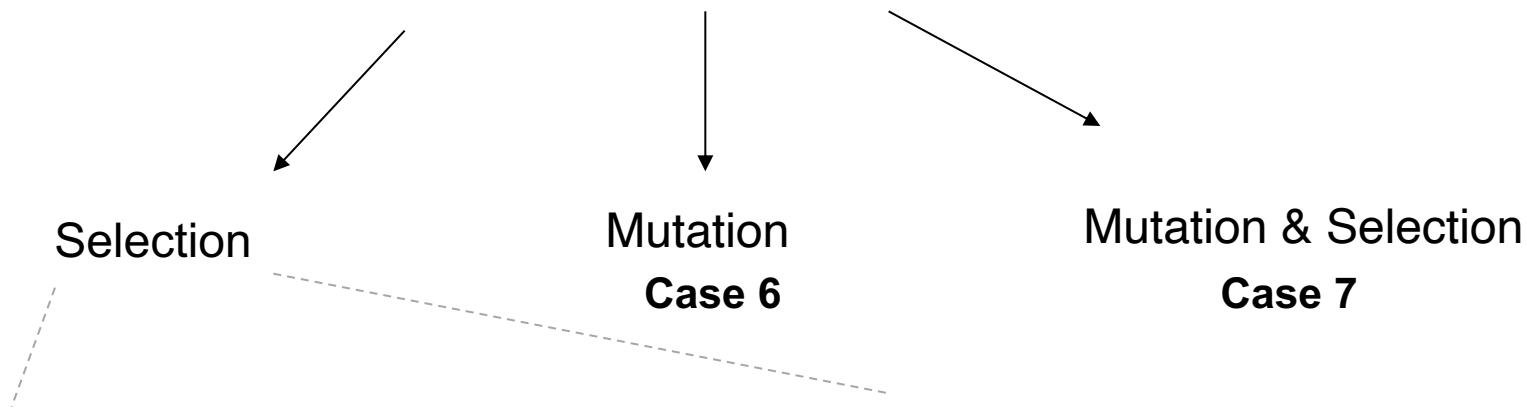
observed frequencies are significantly different from expectation in infants born during malaria season (Chi-square test, Freeman Box 6.4)

Excess of heterozygotes

Missing homozygotes may not have survived fetal development

Different forms of selection

Three “major” corrections to HW equilibrium



- Case 1** Selection against a recessive (deleterious) allele
- Case 2** Selection against a dominant (deleterious) allele
- Case 3** Selection favoring heterozygotes (overdominance)
- Case 4** Selection favoring homozygotes (underdominance)

- Case 5** Frequency-dependent selection

Selection against a recessive (deleterious) allele

recall that $q' = \frac{q^2 w_{aa} + pq w_{Aa}}{\bar{w}} = \frac{q^2 w_{aa} + pq w_{Aa}}{p^2 w_{AA} + 2pq w_{Aa} + q^2 w_{aa}}$

for a recessive allele we have

$$\begin{array}{lll} w_{AA} & w_{Aa} & w_{aa} \\ 1 & 1 & 1-s \end{array}$$

s ... **selection coefficient**, fitness difference from a reference fitness (wild-type)

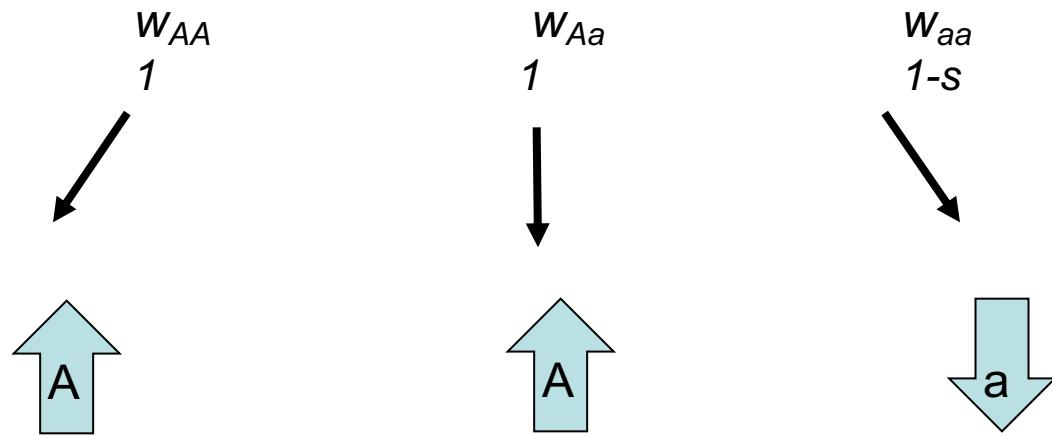
substitution into the equation yields $q' = \frac{q^2(1-s) + pq(1)}{1 - sq^2} = \frac{q(1 - sq)}{1 - sq^2}$

special case: recessive lethal ($s=1$)

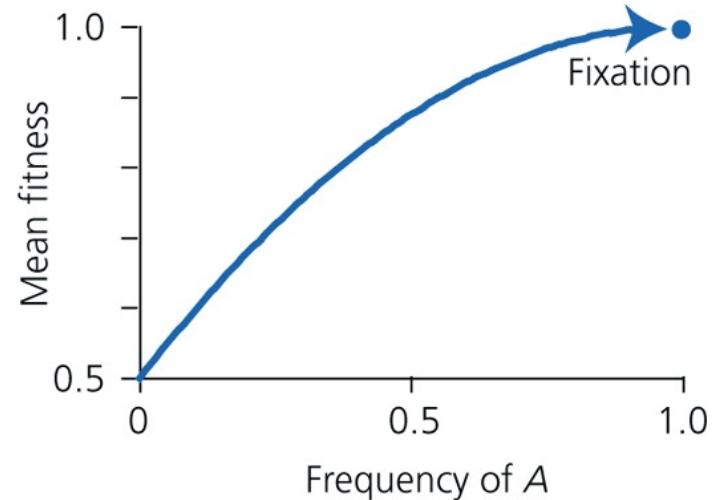
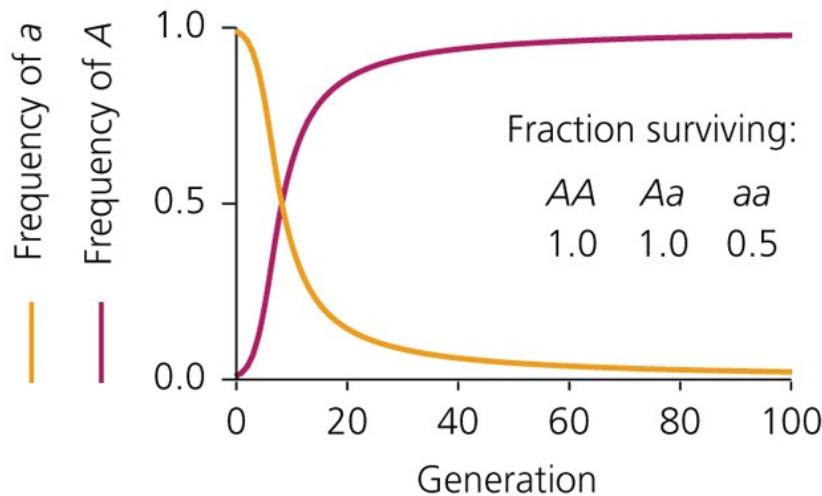
$$q' = \frac{q}{1+q}$$

Guess the behavior:

for a recessive allele we have



Selection against a recessive (deleterious) allele



allele frequency decreases very slowly when a becomes rare

allele a can «hide» from selection in heterozygotes

e.g., if $q=0.01$ and $s=1$, then $q'=0.0099$

a graph of mean fitness against allele frequency provides a complementary view

it shows that natural selection maximizes mean fitness

Class activity 3

An experiment with flour beetles to test population genetic theory

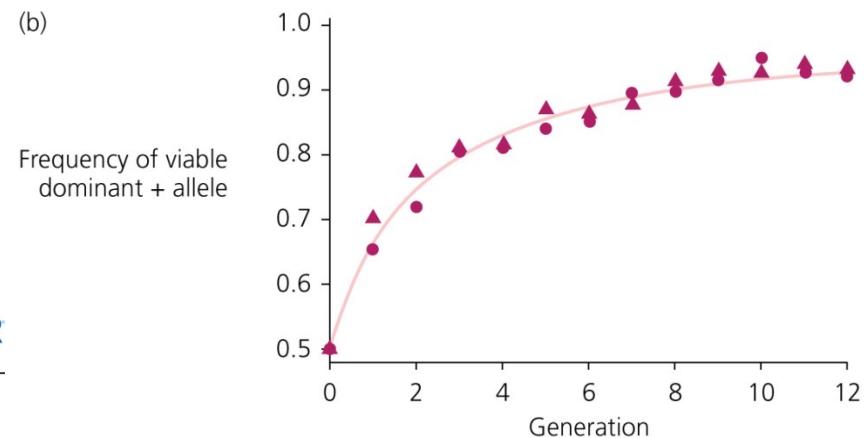
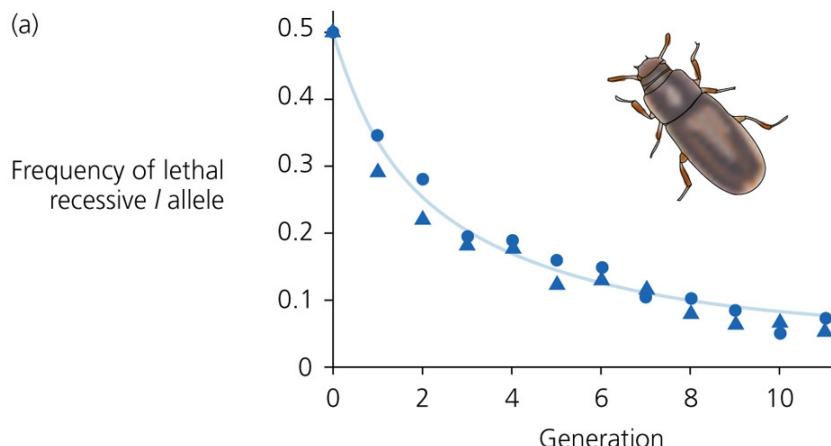
the l locus has two alleles, + and l

l is a **recessive lethal**

experiments starts with heterozygous beetles ($p=q=0.5$)

monitors allele frequency over 12 generations

data (red and blue) agrees remarkably well with theory (grey)



Selection against a dominant (deleterious) allele

For a dominant allele we have

$$\begin{array}{lll} W_{AA} & W_{Aa} & W_{aa} \\ 1-s & 1-s & 1 \end{array}$$

substitution into the equation $p' = \frac{p^2 w_{AA} + pq w_{Aa}}{\bar{w}}$

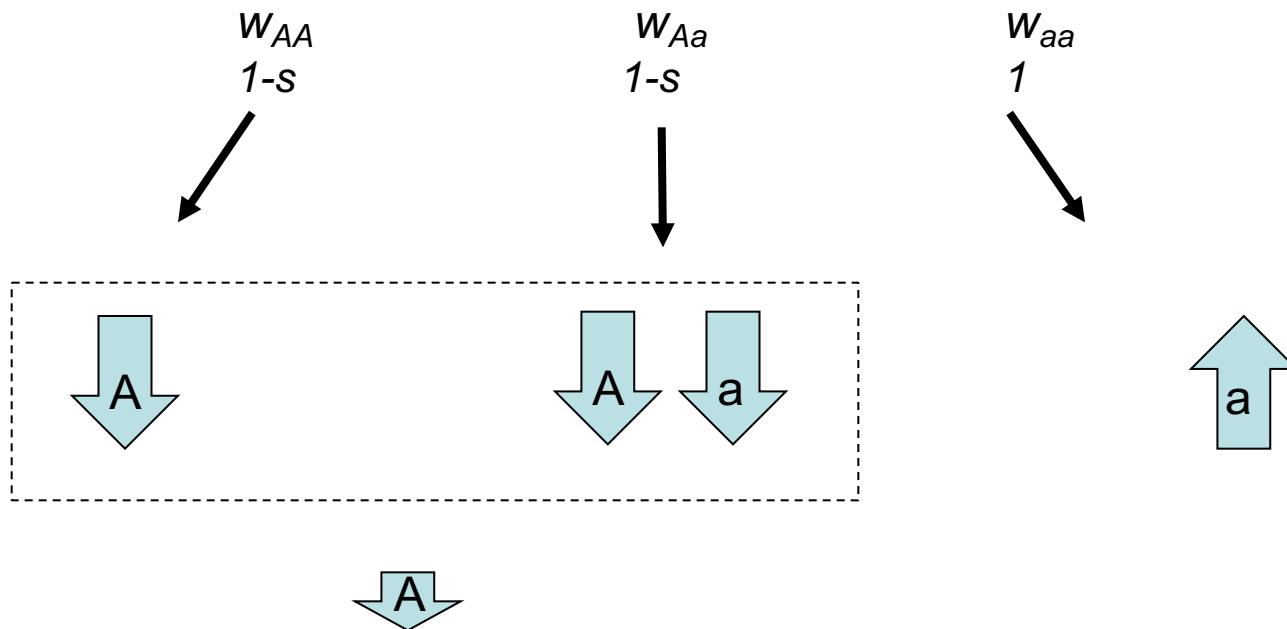
yields

$$p' = \frac{p^2(1-s) + pq(1-s)}{p^2(1-s) + 2pq(1-s) + q^2(1)} = \frac{p(1-s)}{1 - 2pqs - sp^2} = \frac{p(1-s)}{1 - p(2(1-p)s + sp)} = \frac{p(1-s)}{1 - 2sp + sp^2}$$

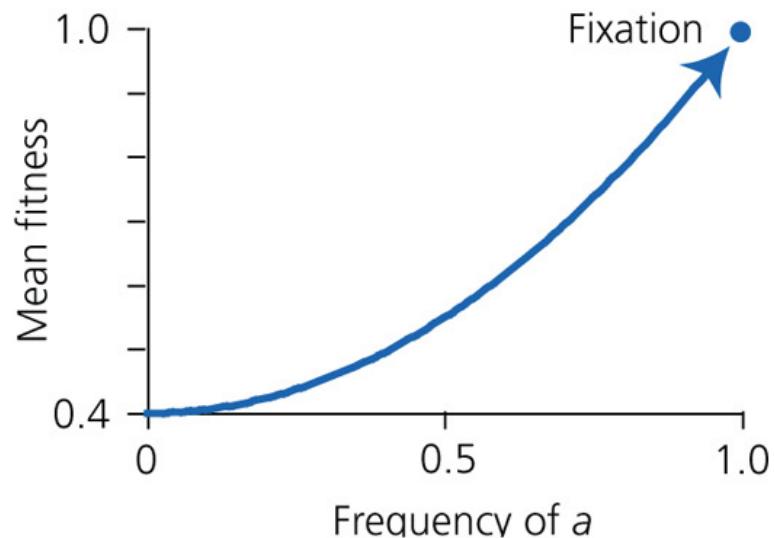
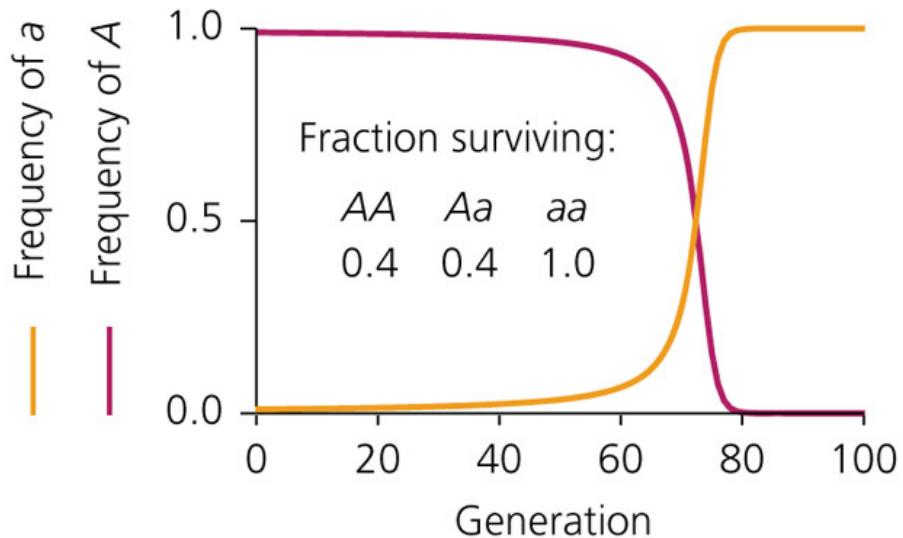
special case: dominant lethal ($s=1$): $p'=0$ (instant elimination)

Guess the behavior:

For a dominant allele we have



Selection against a dominant (deleterious) allele



allele frequency A decreases very slowly initially

thereafter, the superior allele a goes to fixation very rapidly
allele A cannot «hide» from selection in heterozygotes

Selection favoring heterozygotes (overdominance)

Here we have

$$\begin{array}{ccc} w_{11} & w_{12} & w_{22} \\ 1-s & 1 & 1-t \end{array}$$

with $s, t > 0$

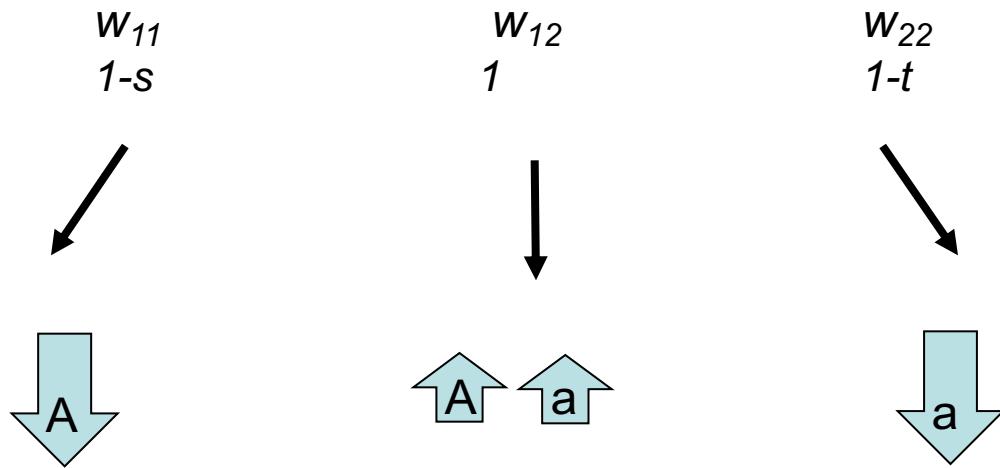
substitution into the equation $\Delta p = \frac{p}{\bar{w}}(pw_{11} + qw_{12} - \bar{w})$

yields

$$\begin{aligned}\Delta p &= \frac{p}{\bar{w}} \left((1-q)w_{11} + qw_{12} - (1-q)^2 w_{11} - 2pqw_{12} - q^2 w_{22} \right) \\ &= \frac{p}{\bar{w}} \left(w_{11} - qw_{11} + qw_{12} - (w_{11} - 2qw_{11} + q^2 w_{11}) - 2pqw_{12} - q^2 w_{22} \right) \\ &= \frac{pq}{\bar{w}} \left(w_{11} + w_{12} - qw_{11} - 2pw_{12} - qw_{22} \right)\end{aligned}$$

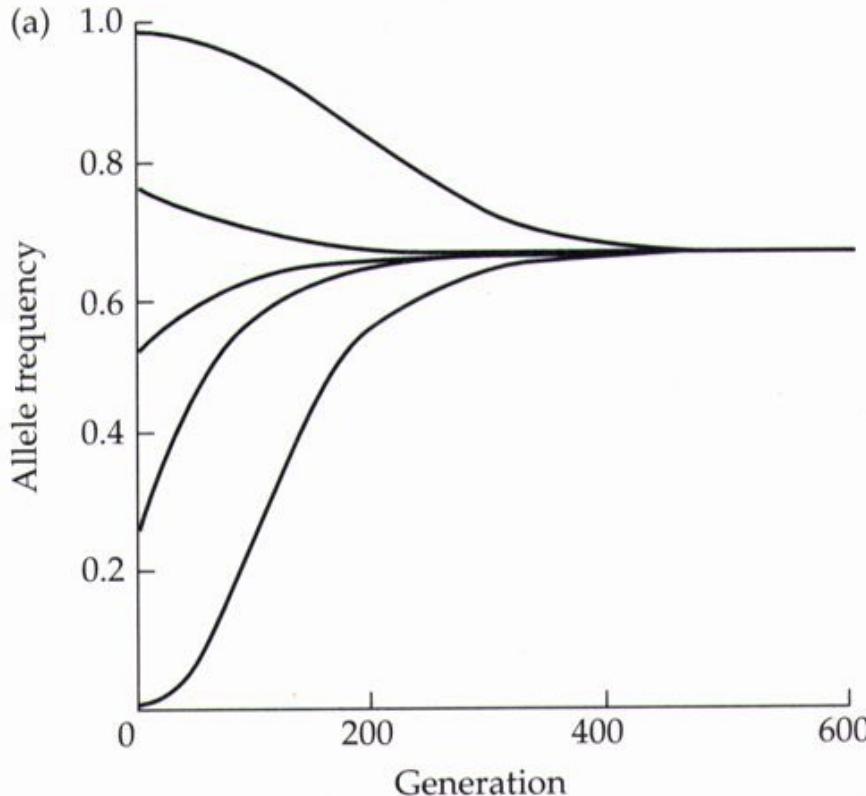
Guess the behavior:

Here we have



Overdominance/Heterozygote Superiority

Genotype	AA	Aa	aa
Fitness	$1-s$	1	$1-t$
$s = 0.02$		$t = 0.02$	



**Stable equilibrium
is reached**

**Genetic diversity
is maintained**

A

Case 3

Selection favoring heterozygotes (overdominance)

An equilibrium is attained if $\Delta p=0$

This is met under three conditions

$$p = 0$$

$$q = 0$$

$$w_{11} + w_{12} - qw_{11} - 2pw_{12} - qw_{22} = 0$$

substituting $1-p$ for q in the last equation and solving for p yields

$$\hat{p} = \frac{w_{22} - w_{12}}{w_{11} - 2w_{12} + w_{22}}$$

and in terms of selection coefficients

$$\hat{p} = \frac{t}{s+t}$$

Selection favoring heterozygotes (overdominance)

Δp as a function of p

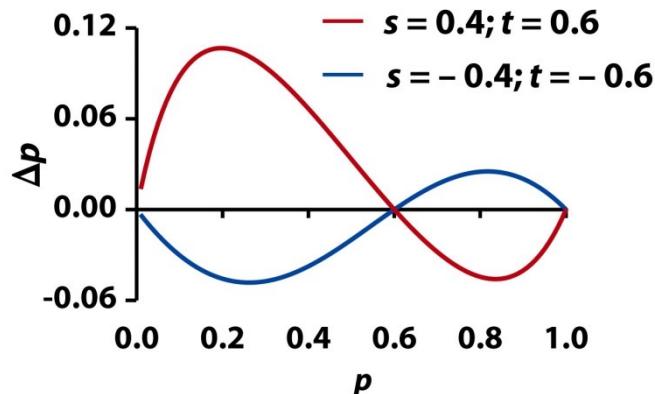


Figure 6-20a Evolutionary Analysis, 4/e
© 2007 Pearson Prentice Hall, Inc.

Mean fitness as a function of p for overdominance

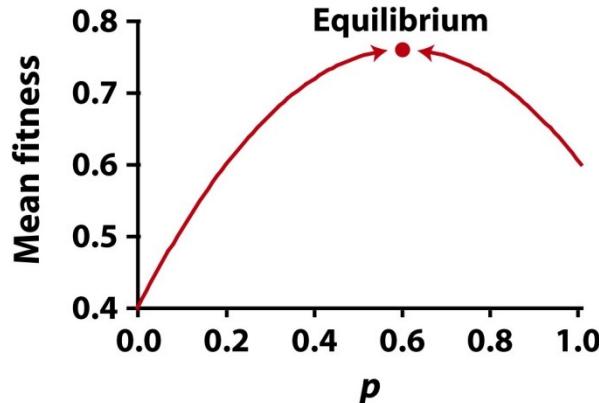


Figure 6-20b Evolutionary Analysis, 4/e
© 2007 Pearson Prentice Hall, Inc.

the sign of Δp depends in magnitude on p

p increases if $p < \hat{p}$

p decreases if $p > \hat{p}$

two **unstable boundary equilibria** ($p=0, q=0$)

one **stable interior equilibrium**

selection increases mean fitness, **but** maximum of one is unattainable

Class activity 4

An experiment demonstrating overdominance in fruit flies

an experiment by Mukai and Burdick in 1959

two alleles: V (viable), L (lethal) at one locus

expectation: V alleles attain a frequency of one

wrong: an intermediate allele frequency is attained

reason: VL heterozygotes have highest fitness

VV	VL	LL
0.735	1.0	0

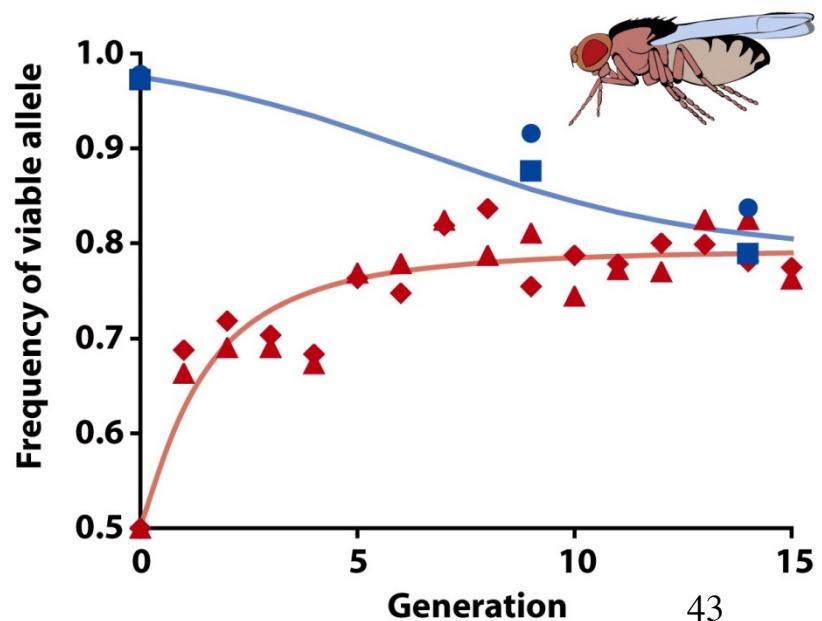


Figure 6-18 Evolutionary Analysis, 4/e
© 2007 Pearson Prentice Hall, Inc.

Selection favoring homozygotes (underdominance)

the same equations work, and yield the same equilibria

$$p = 0$$

$$q = 0$$

$$p = \frac{t}{s+t}$$

but now we have

$$w_{11}$$

$$1-s$$

$$w_{12}$$

$$1$$

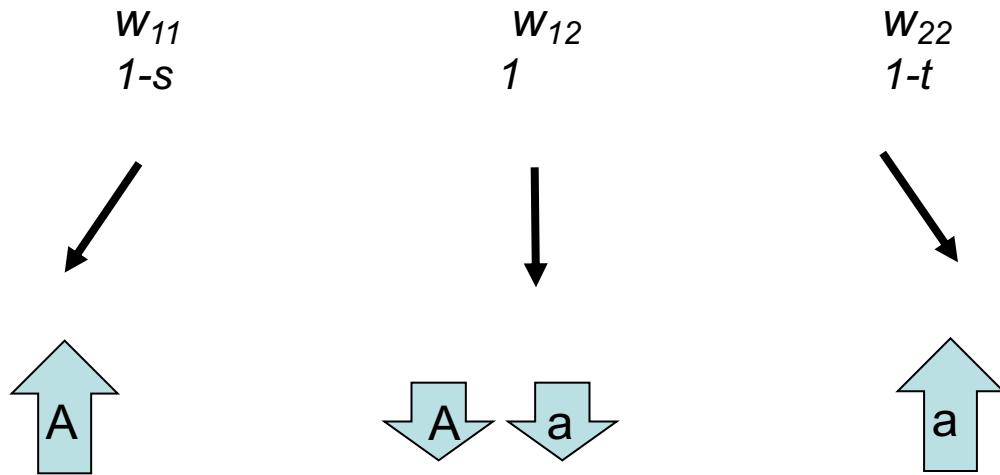
$$w_{22}$$

$$1-t$$

with $s, t < 0$, such that $w_{11}, w_{22} > w_{12}$

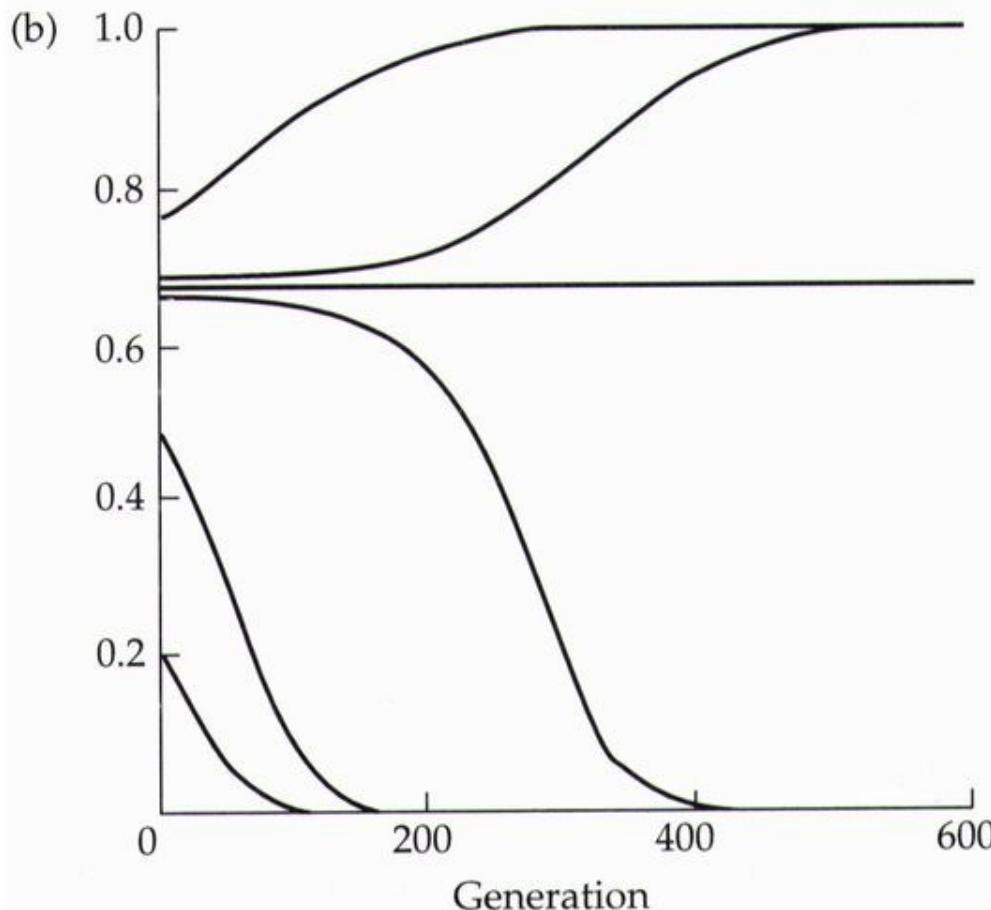
Guess the behavior:

with $s,t < 0$, such that $w_{11}, w_{22} > w_{12}$



Underdominance

Genotype	AA	Aa	aa
Fitness	$1+s$	1	$1+t$
$s=0.01 \quad t=0.01$			



Unstable equilibrium

A maybe fixed or lost from the population

Selection favoring homozygotes (underdominance)

Δp as a function of p

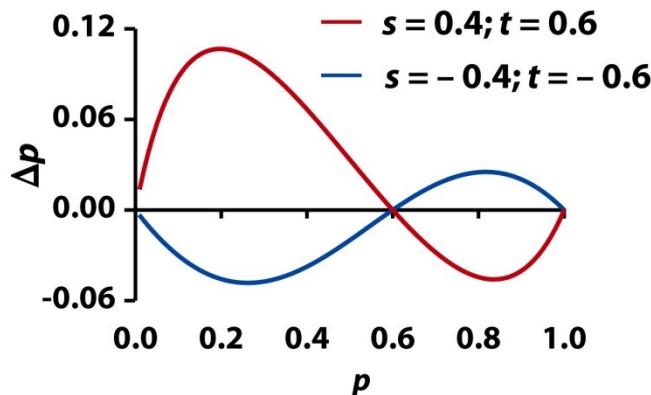


Figure 6-20a Evolutionary Analysis, 4/e
© 2007 Pearson Prentice Hall, Inc.

Mean fitness as a function of p for underdominance

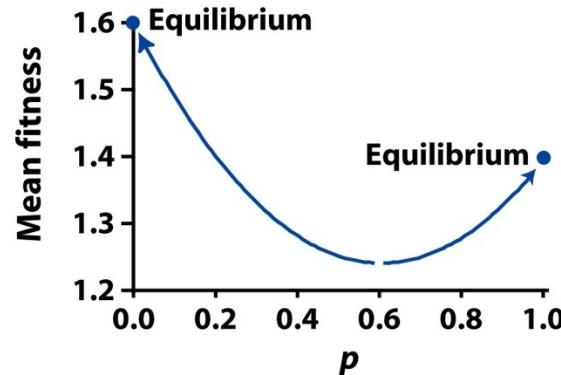


Figure 6-20c Evolutionary Analysis, 4/e
© 2007 Pearson Prentice Hall, Inc.

the sign of Δp depends in magnitude on p

p decreases if $p < \hat{p}$

p increases if $p > \hat{p}$

two **stable boundary equilibria** ($p=0, q=0$)

one **unstable interior equilibrium**

selection increases mean fitness, **but** depending on initial allele frequency, maximum fitness may not be attained

Class activity 5

Frequency dependent selection

fitness of a genotype depends on its frequency

the Elderflower orchid (*Dactylorhiza sambucina*)
has two color-morphs (yellow and purple)

both kinds of flowers carry very little nectar-reward
for pollinators

many pollinators «naive», recently emerged bumblebees

bumblebees tend to visit flowers of different
color alternately

before flying off to more rewarding plants

**this can give rare color-morphs a
reproductive advantage**

on a per-plant basis, they are visited more often



Frequency dependent selection

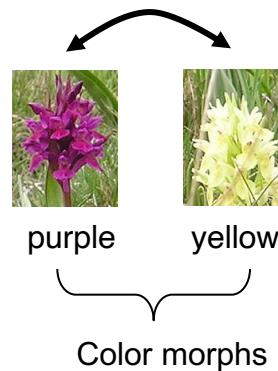
fitness of a genotype depends on its frequency



Elderflower orchid (*Dactylorhiza sambucina*)



bumble bees



Frequency dependent selection

fitness of a genotype depends on its frequency

more pollinia (pollen masses) are removed from rare color-morphs

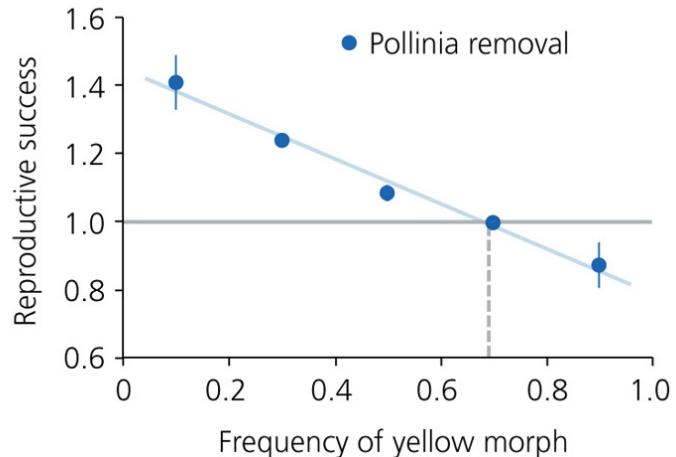
more pollinia are deposited onto rare color-morphs

rare color-morphs show greater fruit set (production of fruit)

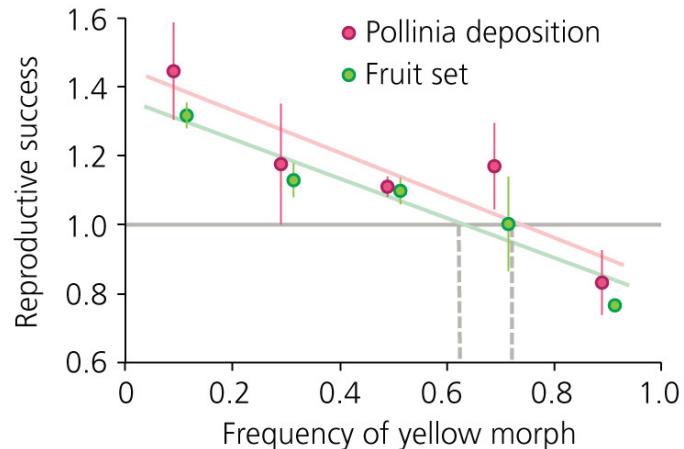
rare flowers tend to become more frequent, and frequent flowers tend to become rarer

frequency-dependent selection can explain stable polymorphisms in populations

(b) Relative male reproductive success



(c) Relative female reproductive success



vertical lines: equilibrium frequency 50 predicted by theory

Case 5

Mutation alone slowly alters allele frequencies

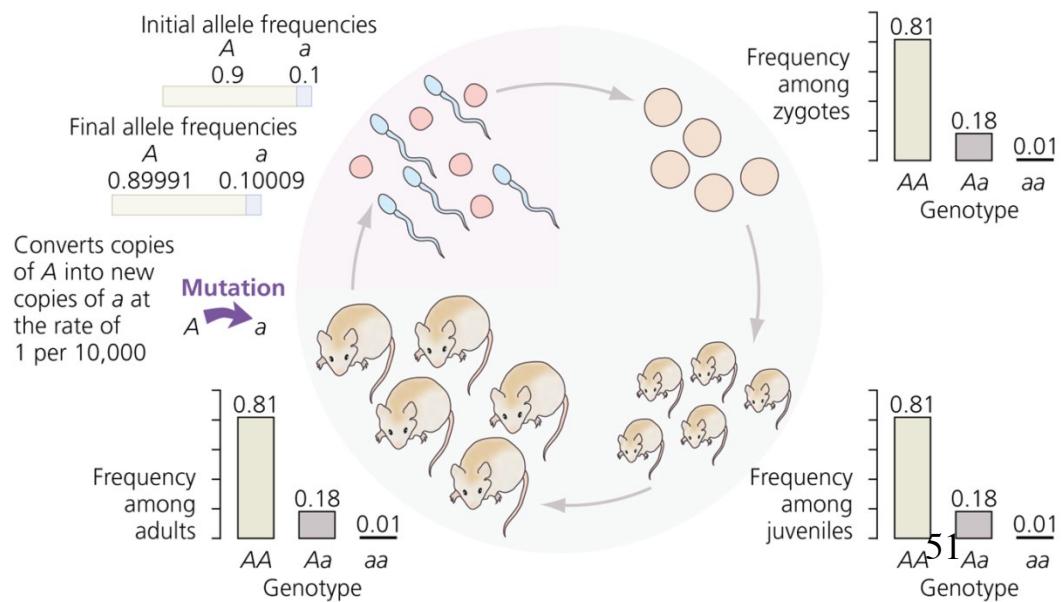
consider a locus with two alleles A and a

mutation (e.g., loss-of-function) converts A into a at a rate μ
the rate of back-mutation is negligible

frequency of A and a after mutation

$$\begin{aligned} p' &= p - \mu p \\ q' &= q + \mu p \end{aligned} \quad \longrightarrow \quad p_n = p_0 e^{-\mu n}$$

allele frequency after n generations
of mutation (Freeman Box 6.8)

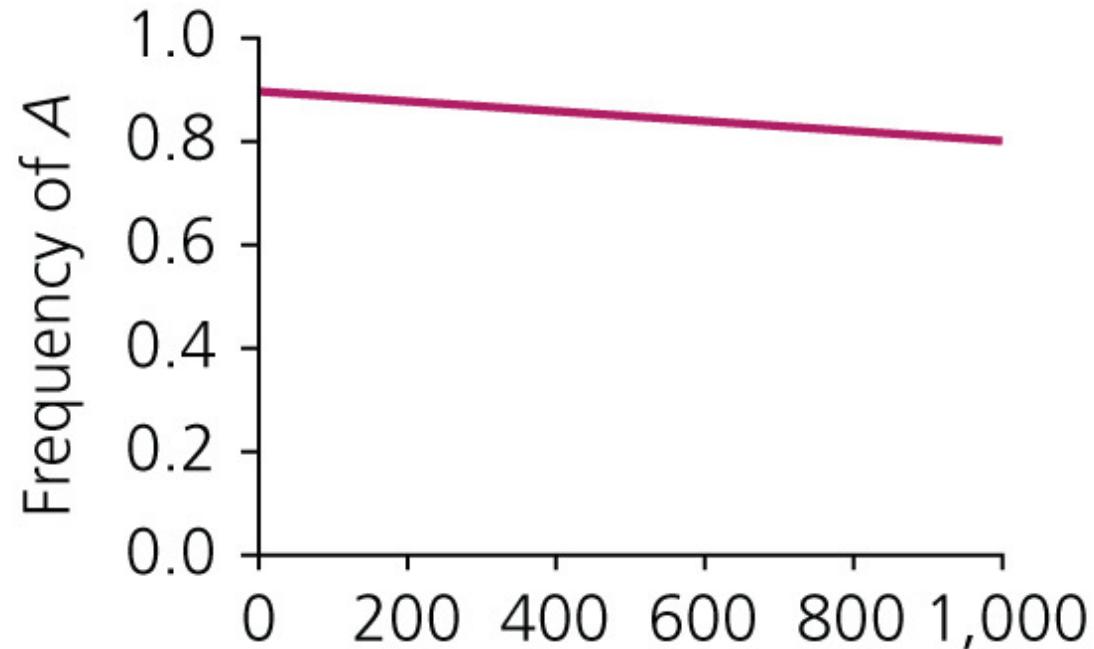


Case 6

Mutation alone slowly alters allele frequencies

$$p_0=0.9, \mu=10^{-4}$$

It takes $g>1000$ generations for the allele frequency to decrease to $p_g=0.8$



Allele frequencies under mutation-selection balance

mutation and selection will eventually balance each other

consider a deleterious recessive allele

$$\begin{array}{ccc} w_{11} & w_{12} & w_{22} \\ 1 & 1 & 1-s \end{array}$$

where mutation converts allele A_1 into A_2 at a rate μ

$$\hat{q} = \sqrt{\frac{\mu}{s}}$$

Deleterious recessive alleles

$$\hat{q} = \mu$$

Deleterious dominant alleles

Equilibrium
frequency
under
mutation-selection balance

Class activity 6

Mutation-selection balance can explain a high frequency of some disease alleles

Example: spinal muscular atrophy

- neurodegenerative disease (weakness and wasting of muscles)
- caused by deletions in *telSMN* (telomeric survival motor neuron) gene
- Recessive
- strong selection against deletion (estimated $s=0.9$)

allele frequency is ~ 0.01

under the assumption of mutation-selection balance, we have $0.01 = \sqrt{\mu / 0.9}$
or which predicts that $\mu \approx 0.9 \times 10^{-4}$.

experimental estimates of the mutation rate are close: $1.1 \times 10^{-4} \hat{q} = \sqrt{\mu / s}$
(Freeman p 214, Box 6.11)

The high frequency of this disease allele is consistent with mutation-selection balance

Equations needed to solve the programming exercises

$$p + q = 1 \text{ (or } q = 1 - p\text{)} \quad (\text{Eq. P1})$$

Allele frequencies sum up to 1

$$\bar{w} = p^2 w_{11} + 2pq w_{12} + q^2 w_{22} \quad (\text{Eq. P2})$$

Population's mean fitness

$$p_{t+1} = (p_t^2 w_{11} + p_t q_t w_{12}) / \bar{w} \quad (\text{Eq. P3})$$

Changes in allele frequency by selection

$$p_{t+1} = (1 - \mu)(p_t^2 w_{11} + p_t q_t w_{12}) / \bar{w} \quad (\text{Eq. P4})$$

Changes in allele frequency by mutation and selection

$$\hat{q} = \sqrt{\frac{\mu}{s}} \quad (\text{Eq. P5})$$

Equilibrium frequency of deleterious recessive alleles under mutation-selection balance