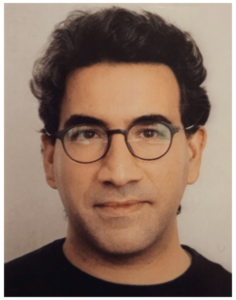
Principles of evolution (BIO 351)

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
Dasmeh@fas.harvard.edu

...which lived long before the ...rian epoch, we may feel certain that the ordinary ...ession by generation has never once been broken, ...that no cataclysm has desolated the whole world. ...ce we may look with some confidence to a secure ...e of equally inappreciable length. And as natural ...tion works solely by and for the good of each ...g, all corporeal and mental endowments will tend to progress towards perfection.

It is interesting to contemplate an entangled bank, clothed with many plants of many kinds, with birds singing on the bushes, with various insects flitting about, and with worms crawling through the damp earth, and to reflect that these elaborately constructed forms, so different from each other, and dependent on each other in so complex a manner, have all been produced by laws acting around us. These laws, taken in the largest sense, being growth with reproduction; inheritance which is almost implied by reproduction; Variability from the indirect and direct action of the external conditions of life, and from use and disuse; a ratio of increase so high as to lead to a struggle for life, and as a consequence to natural selection, entailing divergence of character and the extinction of less-improved forms. Thus, from the war of nature, from famine and death, the most exalted object which we are capable of conceiving, namely, the production of the higher animals, directly follows. There is grandeur in this view of life, with its several powers, having been originally breathed into a few forms or into one; and that, whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved.

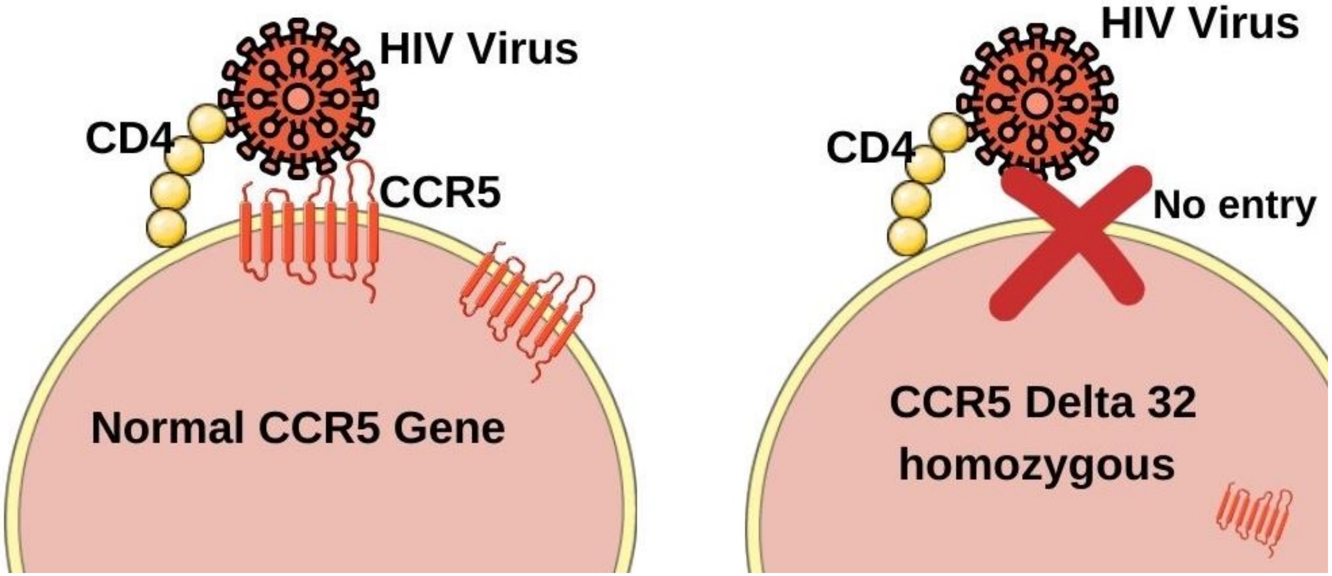
Overview

Hardy-Weinberg equilibrium principle (review)

Selection at one locus with two alleles

Mutation-selection balance

Susceptibility to HIV has a genetic basis



CCR5 (C-C chemokine receptor type 5) is a protein found on the surface of certain immune system cells, including T cells and macrophages.

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 build **Null Models.**

We need to know what would happen to the allele in the absence of selection (or of any other evolutionary force).

the Hardy-Weinberg (HW) law provides the answer.

A simple, idealized life cycle

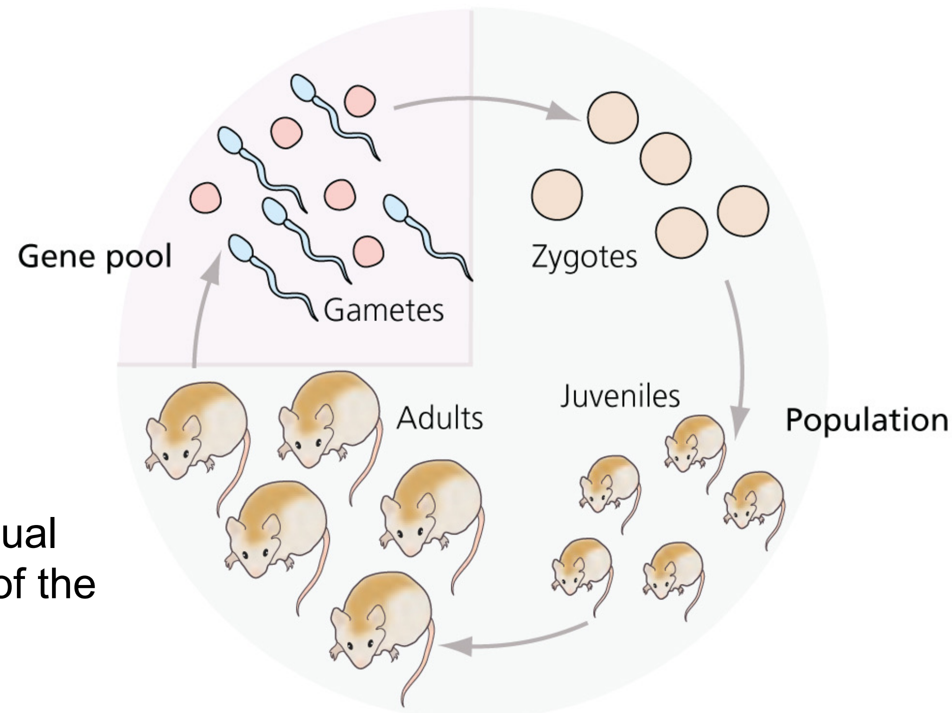
population of diploid organisms

two alleles, with frequencies p, q in the **gamete pool (gene pool)**:
Example: A: p=0.6
a: q=0.4 (Note: $p+q=1$)

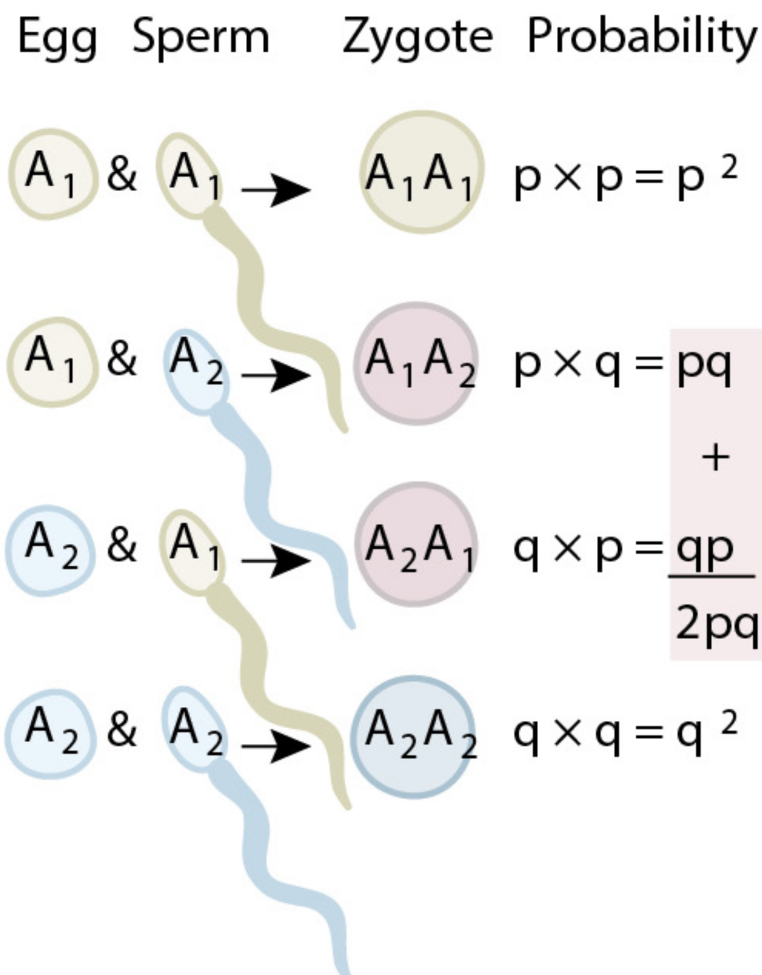
random mating produces zygotes of three genotypes: AA, Aa, aa

all zygotes equally likely to survive

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



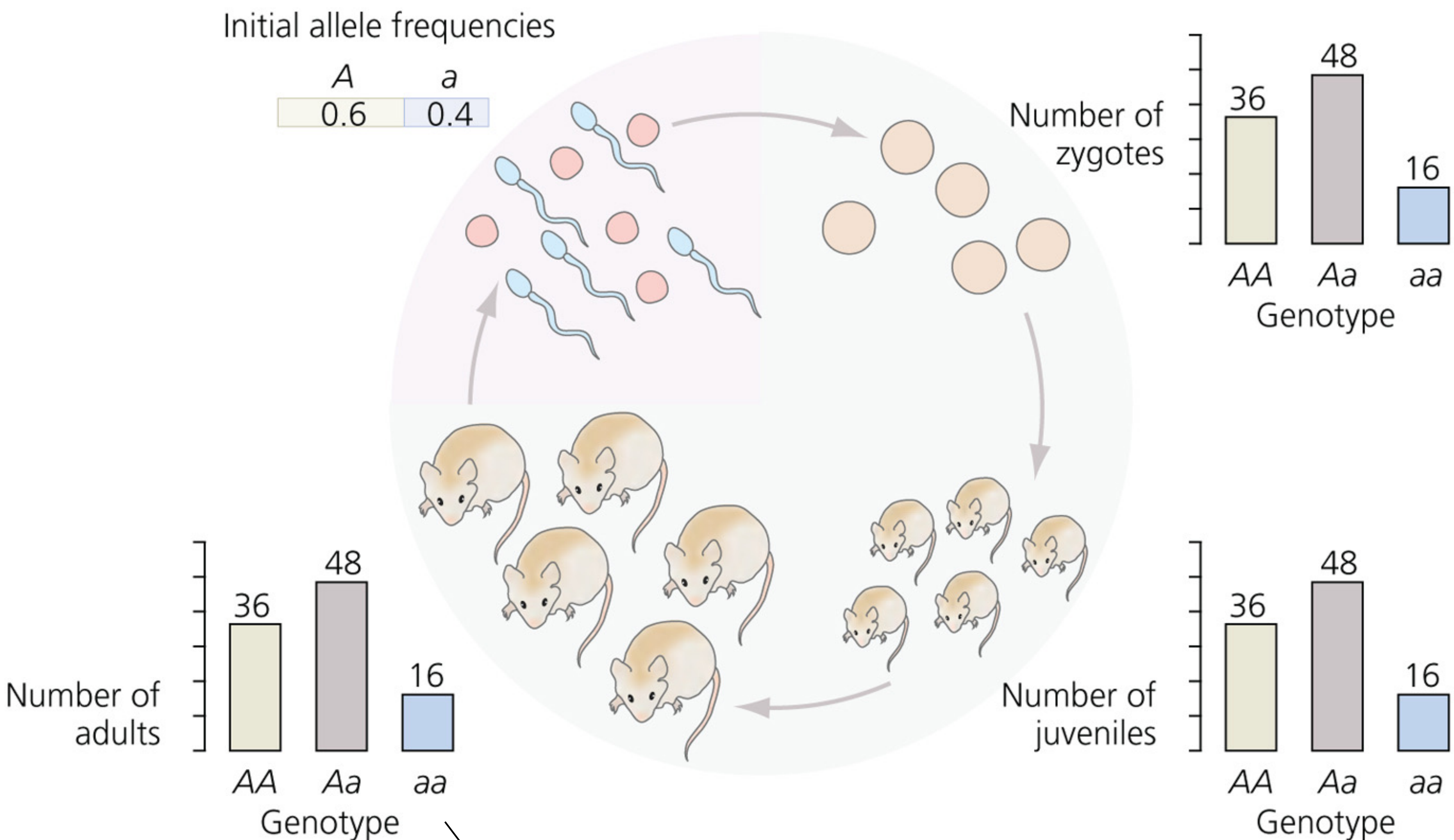
Formation of genotypes from gametes (general case)



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

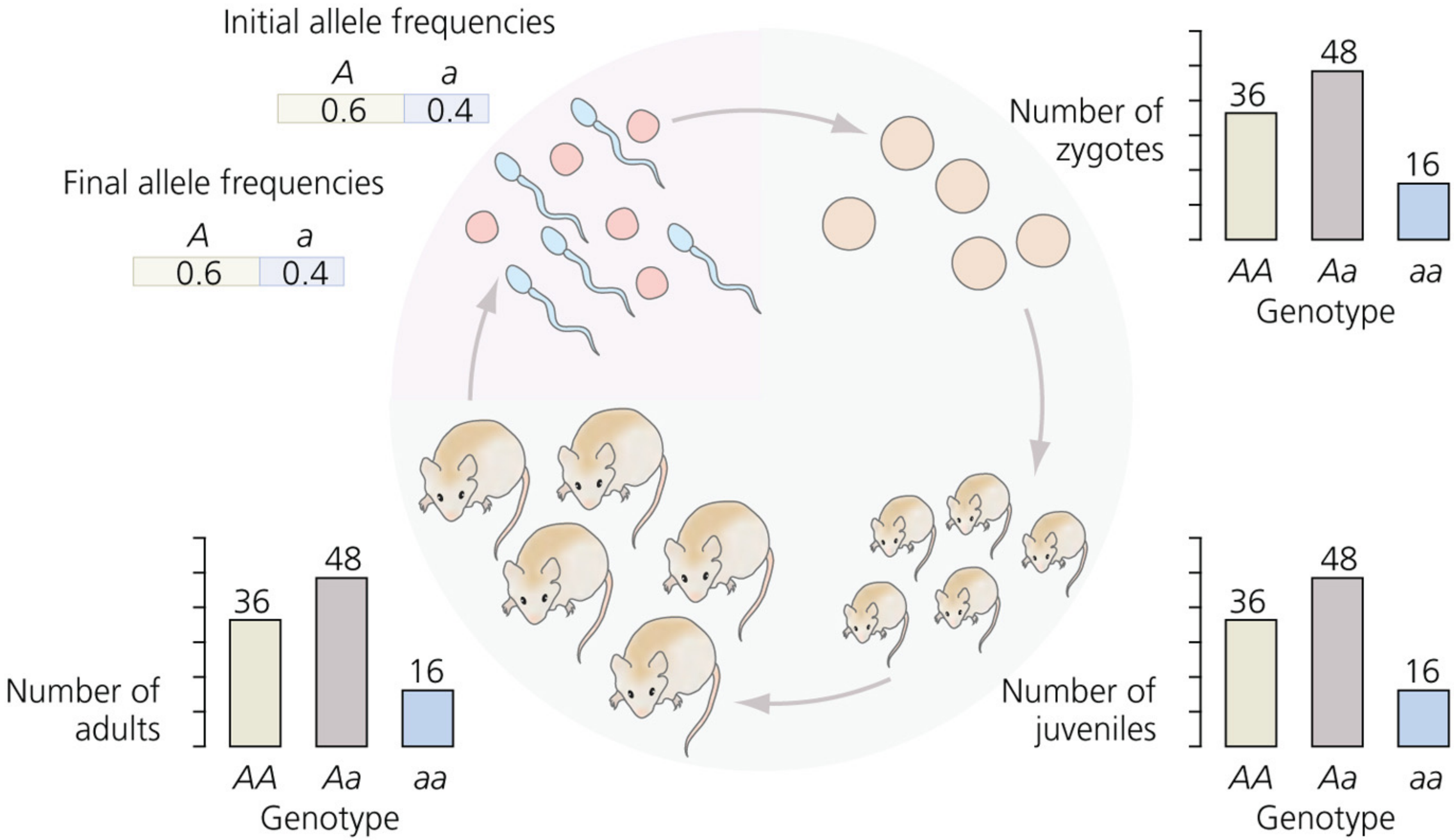
Is it obvious?
Why?

Class activity

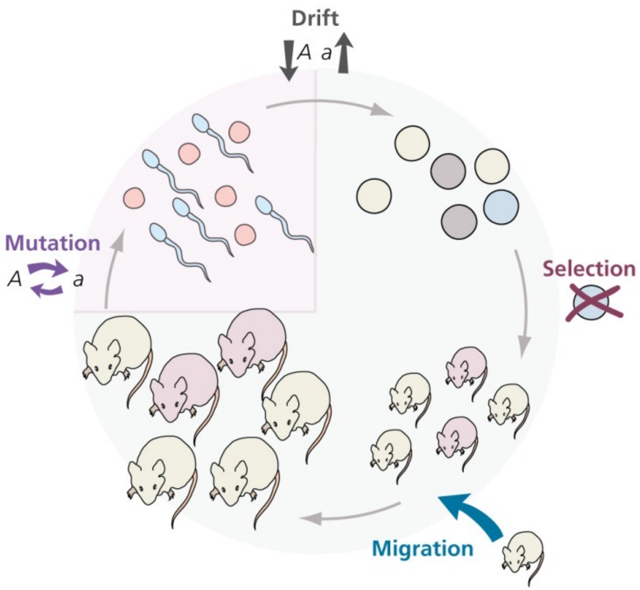


What are the final allele frequencies?

Class activity



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 systematic way to think about all these!

	Alleles		*Genotypes*		
	A	a	AA	Aa	aa
probability	p	q	$p^2 \boxed{}$	$2pq \boxed{}$	$q^2 \boxed{}$

Weight factors
or multipliers

constraint 1: $$\sum p(Alleles) = 1$$

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

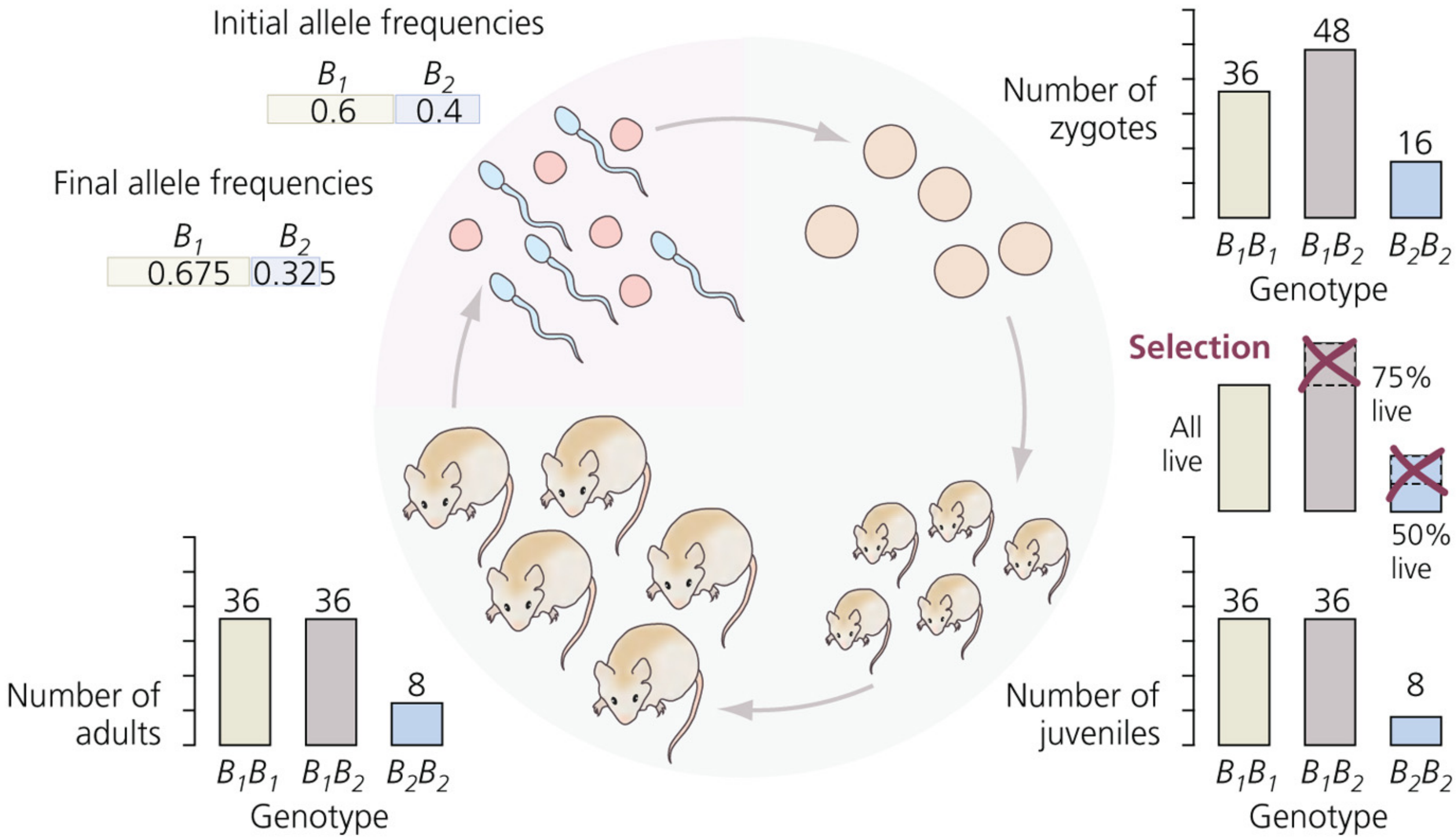
Three corrections to HW equilibrium

Selection

Mutation

Mutation & Selection

Selection can change both genotype and allele frequencies



https://bolnicklab.shinyapps.io/Geno_to_allele_frequencies/

Class Activity

Group 1:
Stefanie K.
Dominik K.
Anna T.
Gilles R.
Eric T.

Group 2:
Jonathan F.
Sarah R.
Alex P.
Krystsina S.

Class activity 1

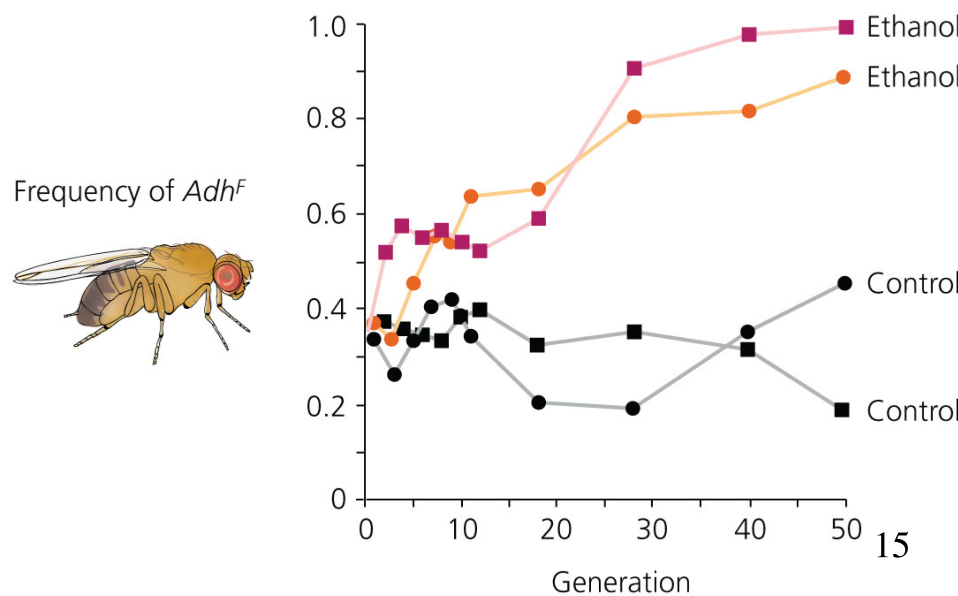
Selection can change allele frequencies unidirectionally

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



A quantitative treatment of selection

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

	Alleles		*Genotypes*		
	A	a	AA	Aa	aa
	p	q	p^2 $\square$	$2pq$ $\square$	q^2 $\square$

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

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**

$$\overline{w} = p^2 w_{11} + 2pqw_{12} + q^2 w_{22}$$

genotype frequencies after selection

A_1A_1	A_1A_2	A_2A_2
$\frac{p^2 w_{11}}{\overline{w}}$	$\frac{2pqw_{12}}{\overline{w}}$	$\frac{q^2 w_{22}}{\overline{w}}$

A quantitative treatment of selection

next calculate allele frequencies *p'*, *q'* after selection

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

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

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

note that $p'+q'=1$

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

A quantitative treatment of selection

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

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

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

equivalently (check for yourself)

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

How to go back to HW without selection?

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

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

Group 1: Infants born from October-April

Group 2: Infants born from May-September

<u>Presentation:</u>

1) Quick intro
2) Observation(s)
3) Important concept(s)
4) Propose follow-up research

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)

SS	*SL*	*LL*
$(0.555)^2=0.308$	$2(0.555)(0.445)=0.494$	$(0.445)^2=0.198$

observed genotype frequencies

SS	*SL*	*LL*
49/163=0.301	*83/163=0.509*	*31/163=0.19*

expected frequencies are not significantly different from observed frequencies in «off-season» infants
(Chi-square test, Freeman Box 6.4)

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

S	*L*
0.539	*0.461*

expected genotype frequencies (from Hardy-Weinberg)

SS	*SL*	*LL*
$(0.539)^2=0.291$	$2(0.539)(0.461)=0.497$	$(0.461)^2=0.213$

observed genotype frequencies

SS	*SL*	*LL*
16/76=0.211	*50/76=0.658*	*10/76=0.132*

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

Selection

Mutation
Case 6

Mutation & Selection
Case 7

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

for a recessive allele we have

w_{AA}	w_{Aa}	w_{aa}
1	1	$1-s$

a: recessive allele
A: Dominant allele

Case 1

Selection against a recessive (deleterious) allele

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

for a recessive allele we have

w_{AA}	w_{Aa}	w_{aa}
1	*1*	*1-s*

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

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

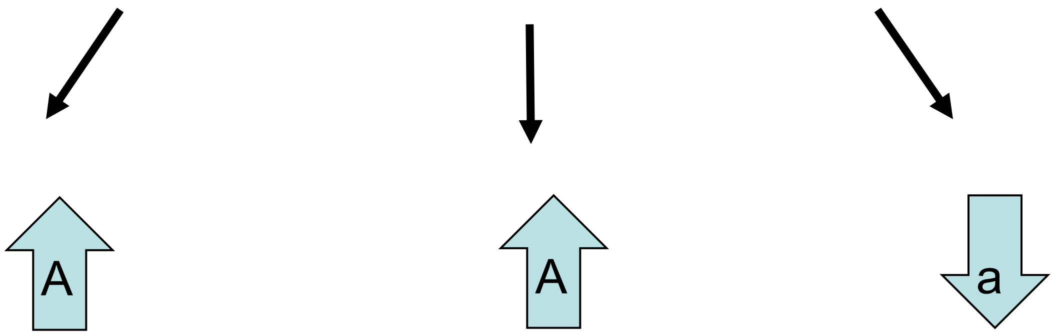
special case: recessive lethal (s=1) $\boxed{q' = \frac{q}{1+q}}$

Case 1

Guess the behavior:

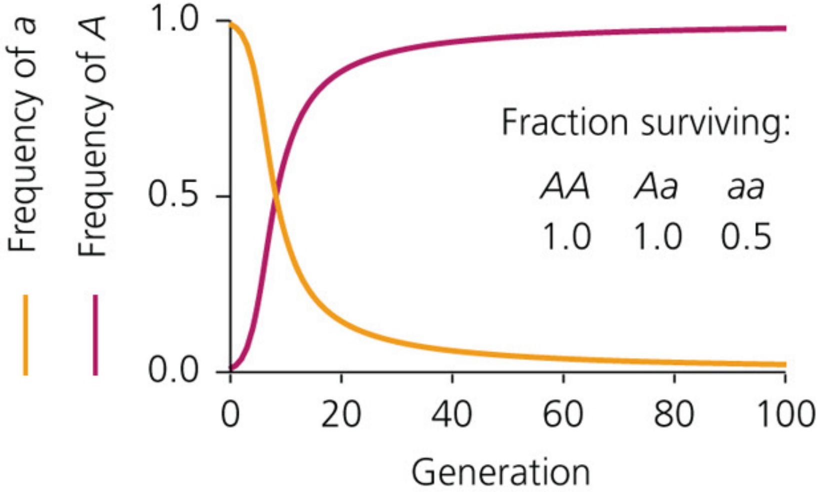
for a recessive allele we have

w_{AA}	w_{Aa}	w_{aa}
1	*1*	*1-s*



Case 1

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

Case 1

Selection against a dominant (deleterious) allele

For a dominant allele we have

w_{AA}	w_{Aa}	w_{aa}
$1-s$	$1-s$	1

a: Dominant allele
A: Recessive allele

Case 2

Selection against a dominant (deleterious) allele

For a dominant allele we have

w_{AA}	w_{Aa}	w_{aa}
1-s	*1-s*	*1*

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

yields

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

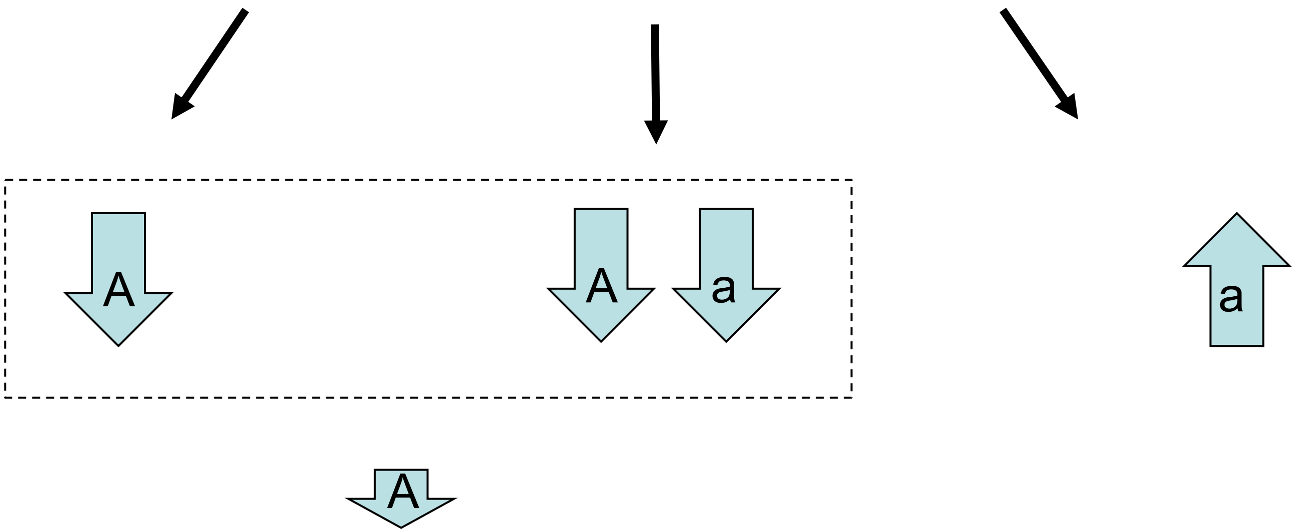
special case: dominant lethal (s=1): p'=0 (instant elimination)

Case 2

Guess the behavior:

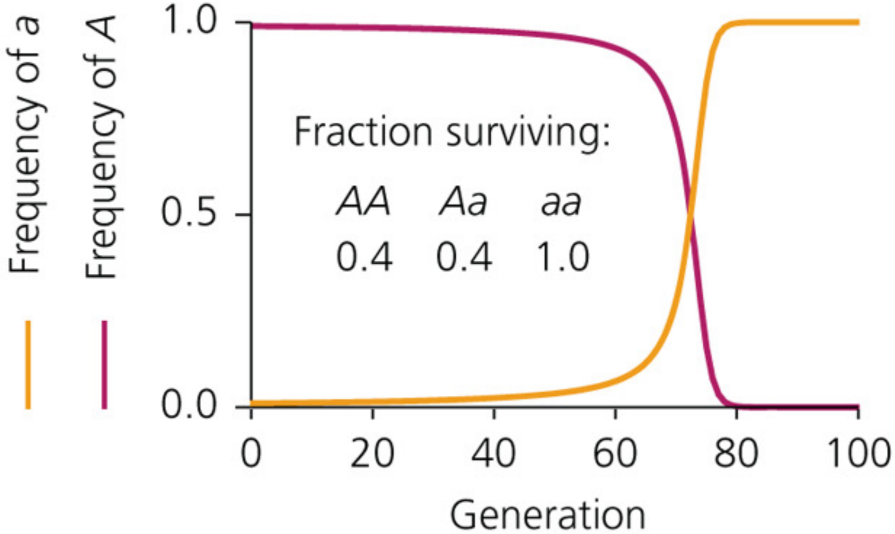
For a dominant allele we have

w_{AA}	w_{Aa}	w_{aa}
$1\text{-}s$	$1\text{-}s$	1



Case 2

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

Case 2

Class activity 3

An experiment with flour beetles to test population genetic theory

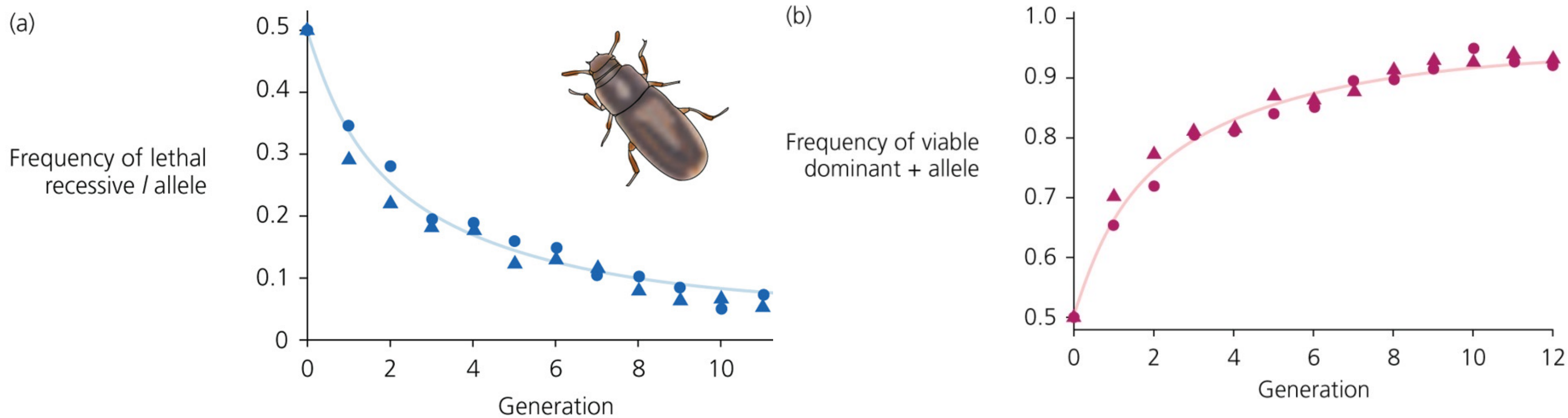
the *l* locus has two alleles, + and *l*

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)



Case 1

Selection favoring heterozygotes (overdominance)

Here we have

w_{11}	w_{12}	w_{22}
$1-s$	1	$1-t$

with $s,t>0$

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

yields

$$\Delta p = \frac{p}{\overline{w}}\left((1-q)w_{11} + qw_{12} - (1-q)^2 w_{11} - 2pqw_{12} - q^2 w_{22}\right)$$

$$= \frac{p}{\overline{w}}\left(w_{11} - qw_{11} + qw_{12} - (w_{11} - 2qw_{11} + q^2 w_{11}) - 2pqw_{12} - q^2 w_{22}\right)$$

$$= \frac{pq}{\overline{w}}\left(w_{11} + w_{12} - qw_{11} - 2pw_{12} - qw_{22}\right)$$

Case 3

Selection favoring heterozygotes (overdominance)

An equilibrium is attained if Δp=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

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

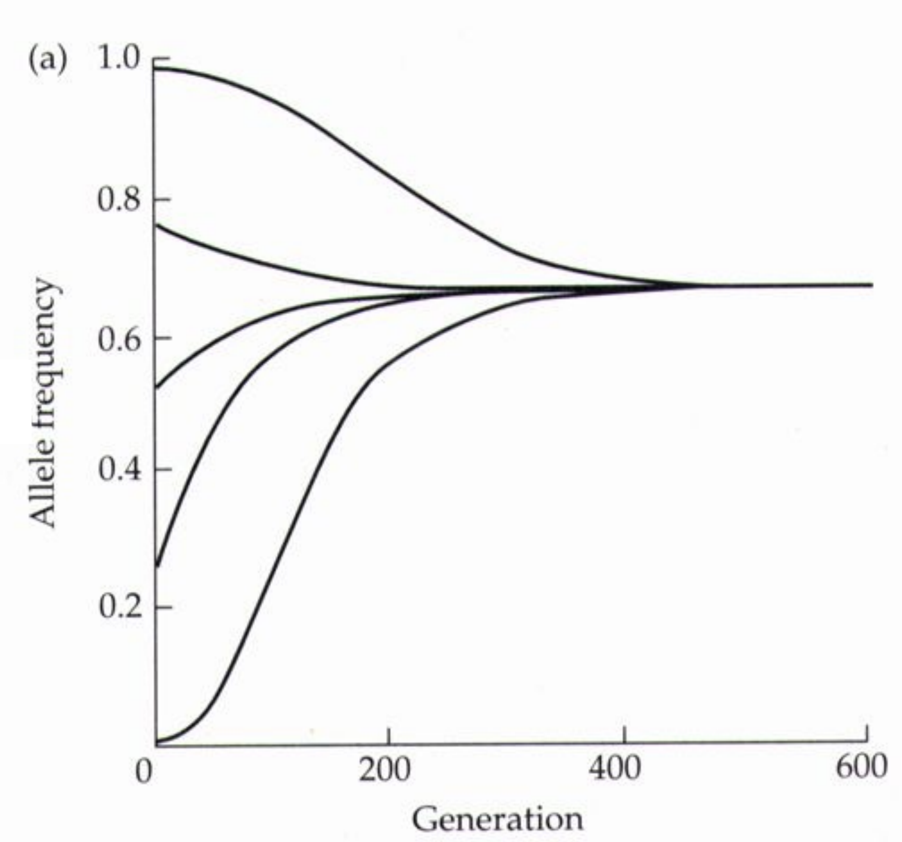
Case 3

Overdominance/Heterozygote Superiority

Genotype	AA	Aa	aa
Fitness	1-s	1	1-t

s = 0.02 t = 0.02

A



Stable equilibrium is reached

Genetic diversity is maintained

Case 3

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

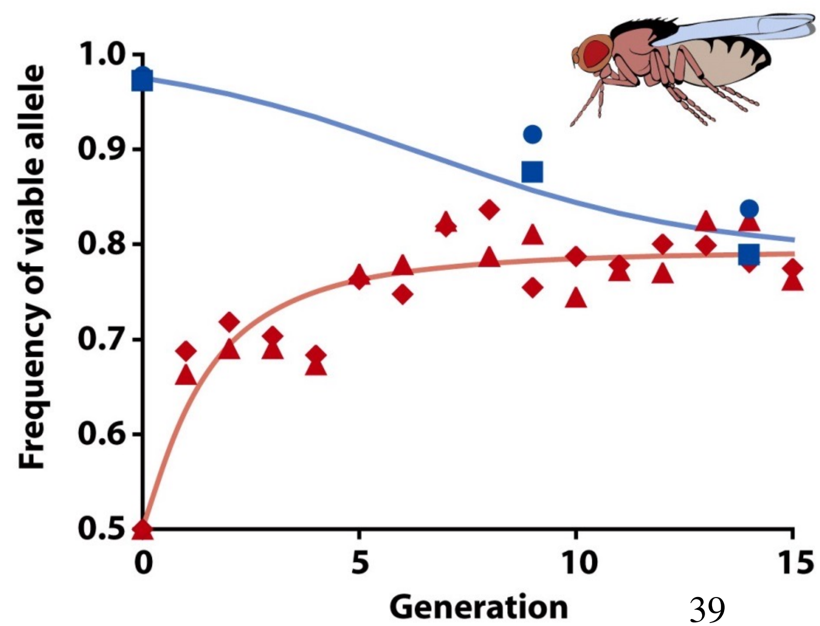


Figure 6-18 Evolutionary Analysis, 4/e

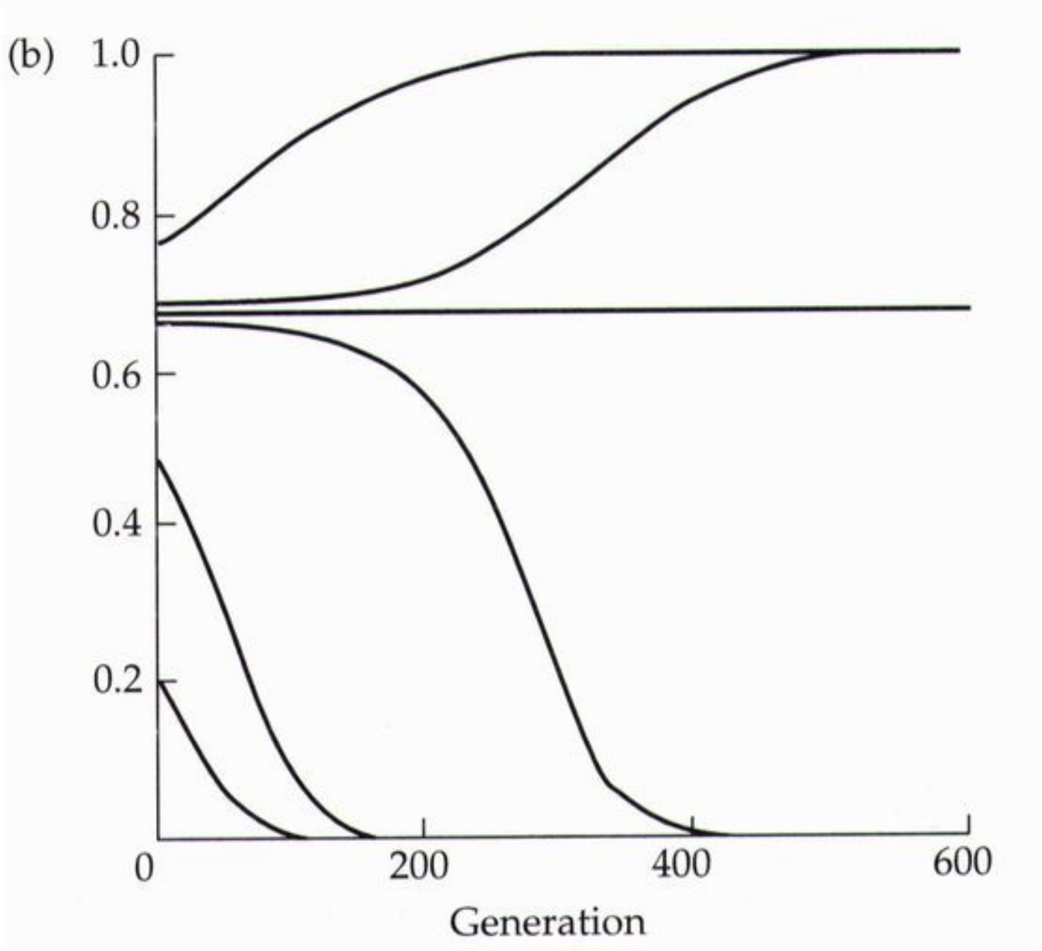
Case 3

Selection favoring homozygotes (underdominance)

Genotype	AA	Aa	aa
Fitness	1+s	1	1+t

s=0.01 t=0.01

A



Unstable equilibrium

A maybe fixed or lost from the population

Case 4

Cumpulsory sterilization

The horror of Eugenics!

Wall Street,
Eugenics demonstration
1915

Eugenics is based on the erroneous claim that strength of mind behaved like a simple Mendelian trait. Normal-mindedness was believed to be dominant and feeblemindedness recessive.

This is of course morally wrong and scientifically inaccurate!

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

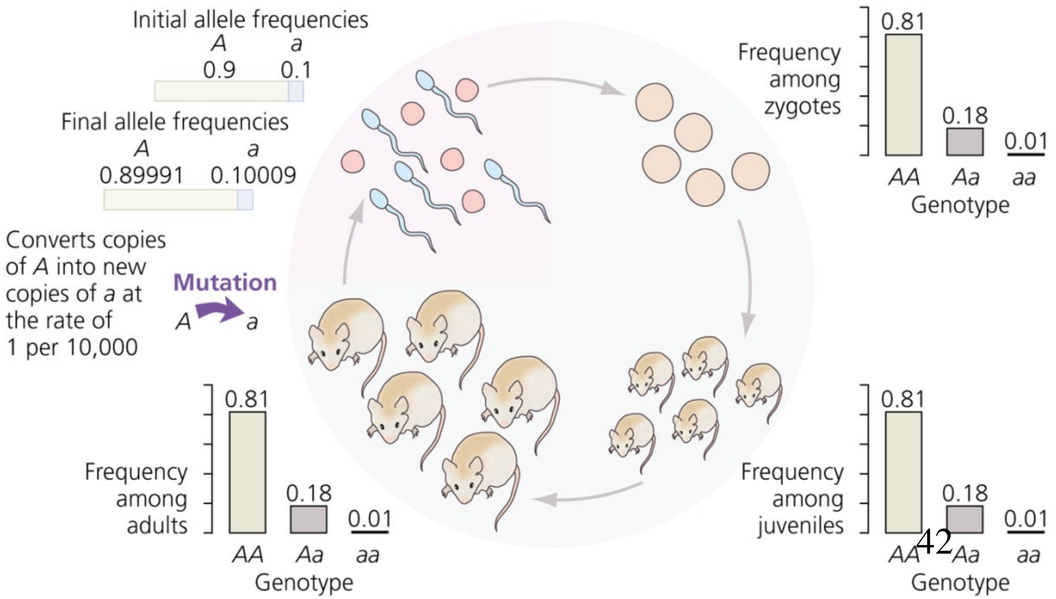
the rate of back-mutation is negligible

frequency of *A* and *a* after mutation

$$p' = p - \mu p$$
$$q' = q + \mu p$$

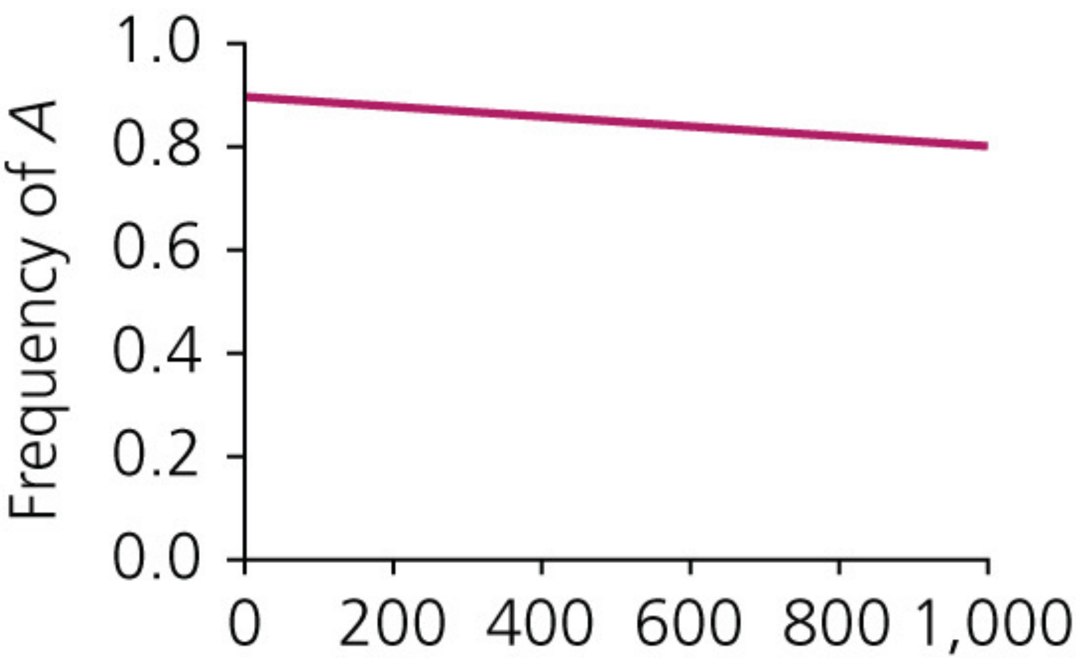
$$\longrightarrow \quad p_n = p_0 e^{-\mu n}$$

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



Mutation alone <u>slowly</u> alters allele frequencies

p_0=0.9, μ=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

w_{11}	w_{12}	w_{22}
1	*1*	*1-s*

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

Equations needed to solve the programming exercises

$$p + q = 1 \ (or \quad q = 1 - p)$$ **(Eq. P1)**

Allele frequencies sum up to 1

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

Population's mean fitness

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

Changes in allele frequency by selection

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

Changes in allele frequency by mutation and selection

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

Equilibrium frequency of deleterious recessive alleles under mutation-selection balance