

# INTRODUCTION

WHAT IS POPULATION GENETICS? WHAT DOES IT STUDY?

# WHAT IS BIOLOGICAL EVOLUTION?

# WHAT IS BIOLOGICAL EVOLUTION?

Biological evolution is the change over time in the **genetic composition of a population**, which changes due to the birth and death of individuals or the migration of individuals in or out of the population.

**WHAT ARE THE FORCES DRIVING BIOLOGICAL EVOLUTION?**

# WHAT ARE THE POPULATION EVENTS DRIVING BIOLOGICAL EVOLUTION?

# WHAT ARE THE POPULATION EVENTS DRIVING BIOLOGICAL EVOLUTION?

- birth

# WHAT ARE THE POPULATION EVENTS DRIVING BIOLOGICAL EVOLUTION?

- birth
- death

# WHAT ARE THE POPULATION EVENTS DRIVING BIOLOGICAL EVOLUTION?

- birth
- death
- migration

# MEASURING GENETIC VARIATION

# TYPES OF GENETIC VARIATION

- SNP

# TYPES OF GENETIC VARIATION

- SNP
- microsatellites

# TYPES OF GENETIC VARIATION

- SNP
- microsatellites
  - What are microsatellites?

# TYPES OF GENETIC VARIATION

- SNP
- microsatellites
  - What are microsatellites?
  - what are they useful for?

# TYPES OF GENETIC VARIATION

- SNP
- microsatellites
  - What are microsatellites?
  - what are they useful for?
- CNV

# TYPES OF GENETIC VARIATION

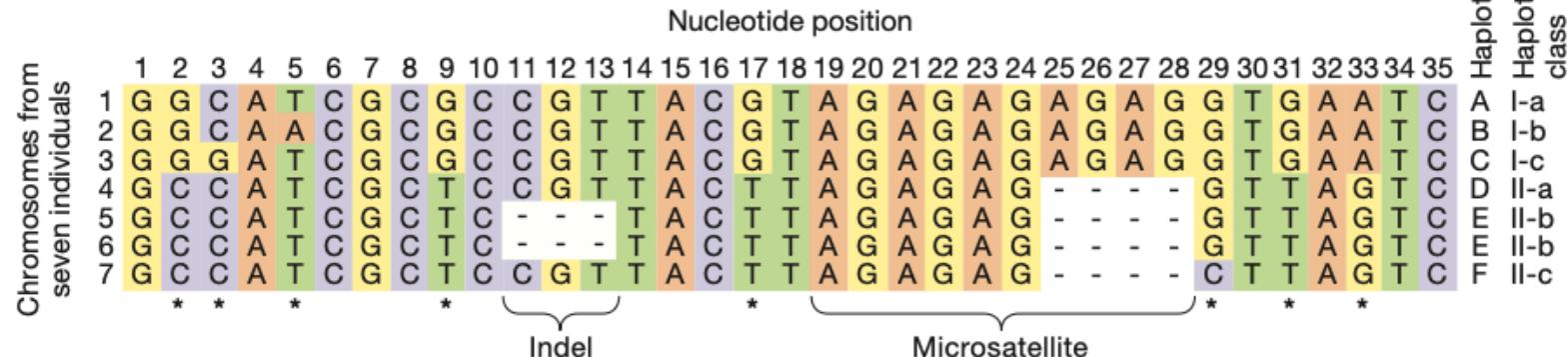
## Haplotype

refer to the combination of alleles at multiple loci on the same chromosomal homolog.

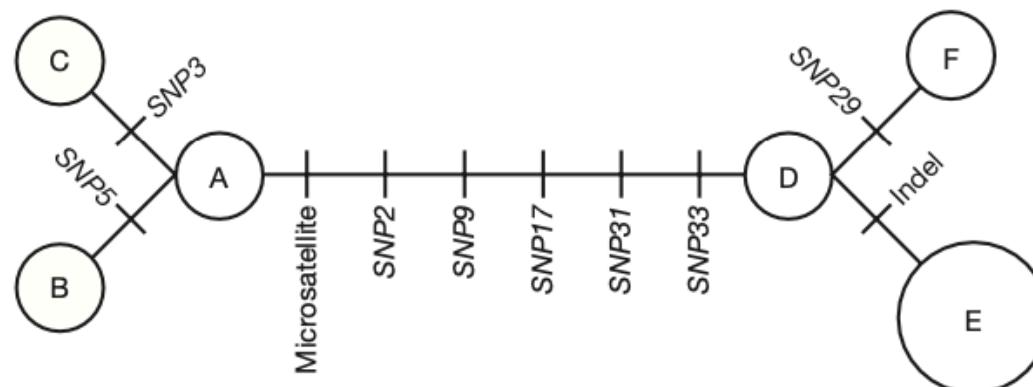
# HOW DO YOU DRAW THE NETWORK (B) BASED ON THE ALIGNMENT (A)?

A haplotype network shows the relationship among haplotypes

### (a) Haplotypes

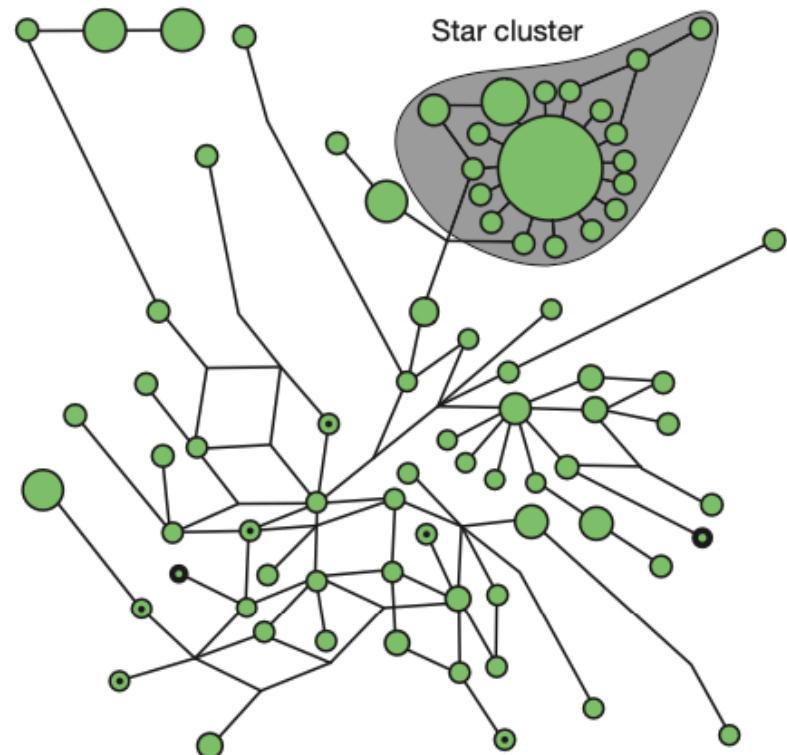


### (b) Haplotype network



A prevalent Y-chromosome haplotype among Asian men may trace back to Genghis Khan

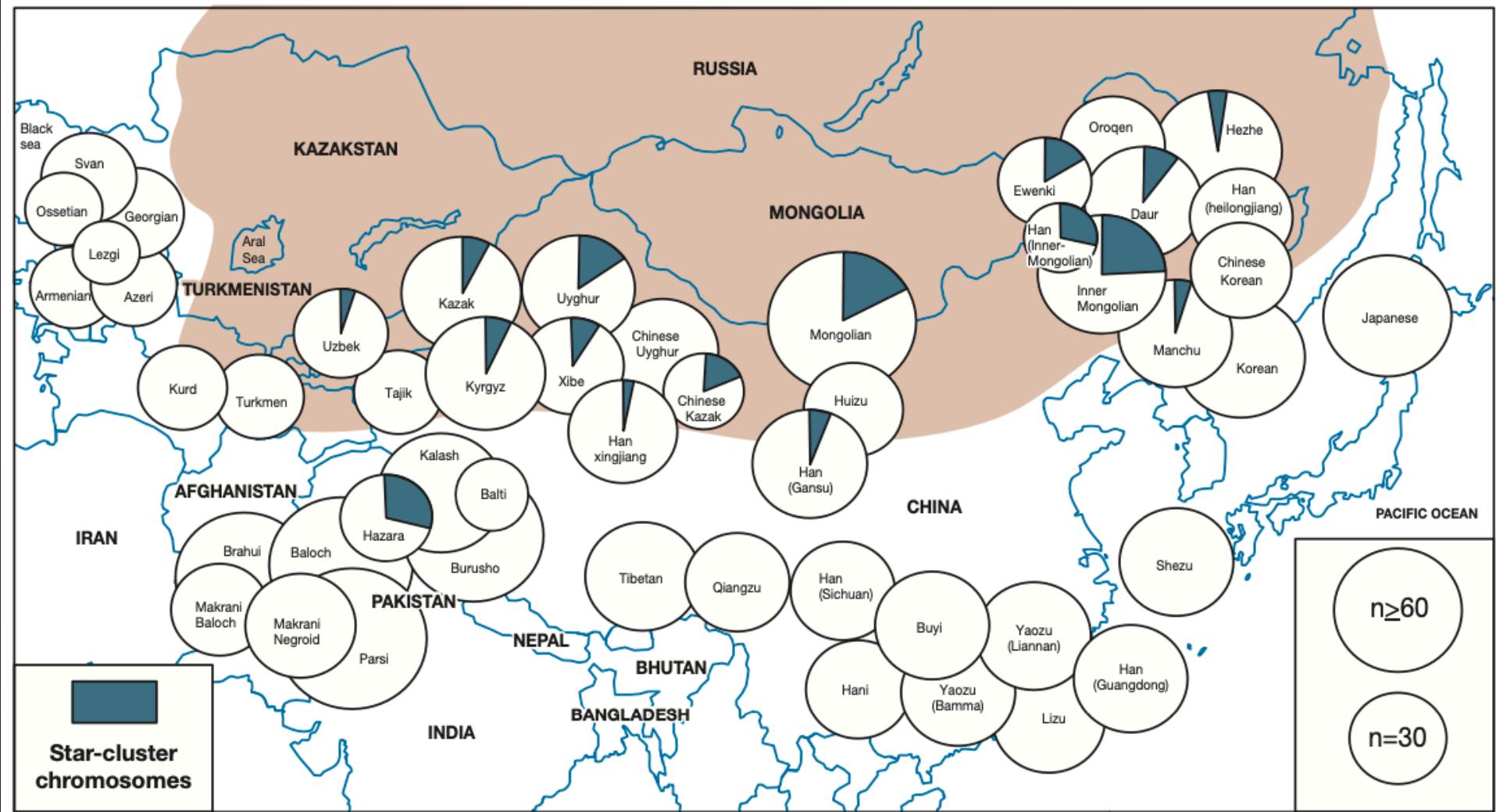
(a)



**FIGURE 18-5** (a) Haplotype network for the Y chromosomes of Asian men showing the predominance of the star-cluster haplotype thought to trace back to Genghis Khan. The area of the circle is proportional to the number of individuals with the specific haplotype that the circle represents. (b) Geographical distribution of the star-cluster haplotype. Populations are shown as circles with an area proportional to sample size; the proportion of individuals in the sample carrying star-cluster chromosomes is indicated by green sectors. No star-cluster chromosomes were found in populations having no green sector in the circle. The shaded area represents the extent of Genghis Khan's empire. [Data from T. Zerjal et al., Am. J. Hum. Genet. 72, 2003, 717–721.]

# DISTRIBUTION OF THE STAR-CLUSTER HAPLOTYPE

(b)



# HOW DO YOU QUANTIFY THE LEVEL OF GENETIC VARIATION?

A haplotype network shows the relationship among haplotypes

### (a) Haplotypes

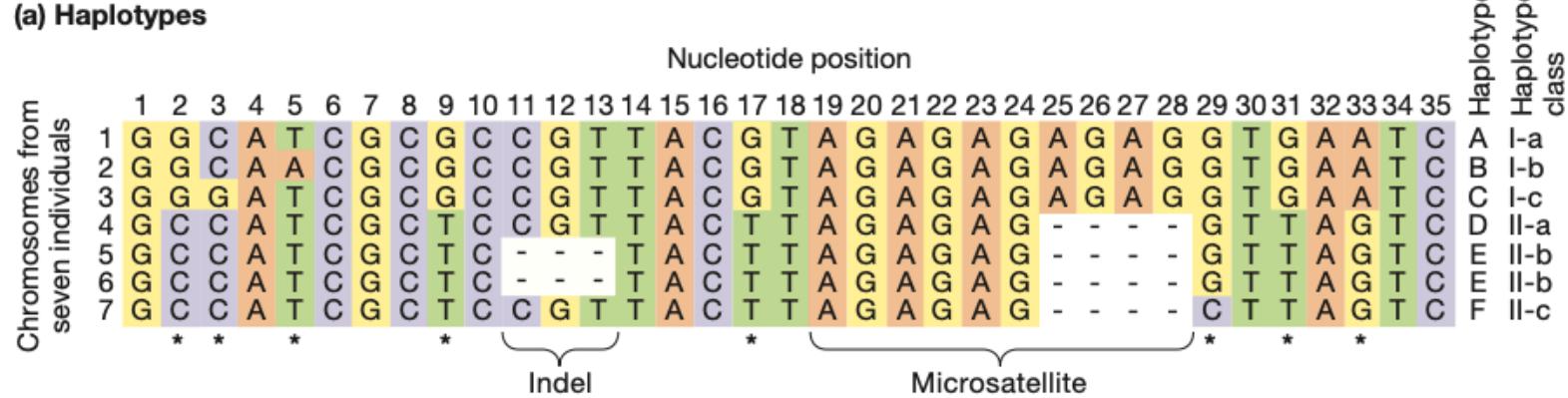
(a) Haplotypes		Nucleotide position																																	Haplotype class		
Chromosomes from seven individuals		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	Haplotype class
1	G	G	C	A	T	C	G	C	G	T	T	A	C	G	T	A	G	A	G	A	G	T	G	A	A	T	C	A	I-a								
2	G	G	C	A	A	C	G	C	G	T	T	A	C	G	T	A	G	A	G	A	G	T	G	A	A	T	C	B	I-b								
3	G	G	G	A	T	C	G	C	G	T	T	A	C	G	T	A	G	A	G	A	G	T	G	A	A	T	C	C	I-c								
4	G	C	C	A	T	C	G	C	T	C	G	T	T	A	C	T	T	A	G	A	G	-	-	-	-	G	T	T	A	G	T	C	D	II-a			
5	G	C	C	A	T	C	G	C	T	C	-	-	T	A	C	T	T	A	G	A	G	-	-	-	-	G	T	T	A	G	T	C	E	II-b			
6	G	C	C	A	T	C	G	C	T	C	-	-	T	A	C	T	T	A	G	A	G	-	-	-	-	G	T	T	A	G	T	C	F	II-b			
7	G	C	C	A	T	C	G	C	T	C	C	G	T	T	A	C	T	T	A	G	A	-	-	-	-	C	T	T	A	G	T	C	F	II-c			

- count the number of variable sites

# HOW DO YOU QUANTIFY THE LEVEL OF GENETIC VARIATION?

A haplotype network shows the relationship among haplotypes

## (a) Haplotypes

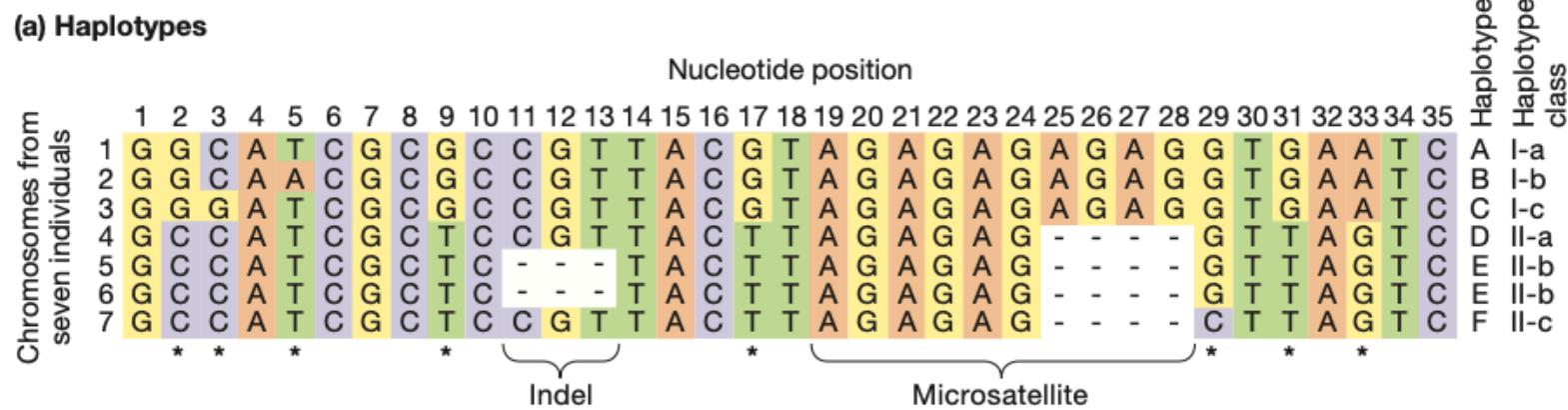


- count the number of variable sites
- account for the length of the region (sequence)

# HOW DO YOU QUANTIFY THE LEVEL OF GENETIC VARIATION?

A haplotype network shows the relationship among haplotypes

## (a) Haplotypes

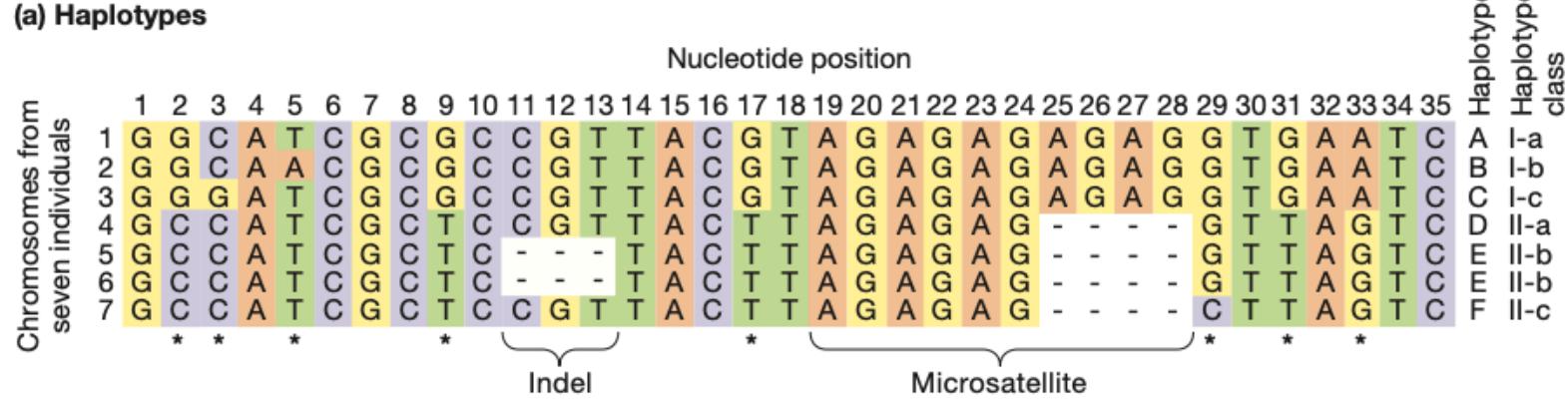


- count the number of variable sites
- account for the length of the region (sequence)
- account for the number of sequences (sample size)

# HOW DO YOU QUANTIFY THE LEVEL OF GENETIC VARIATION?

A haplotype network shows the relationship among haplotypes

## (a) Haplotypes

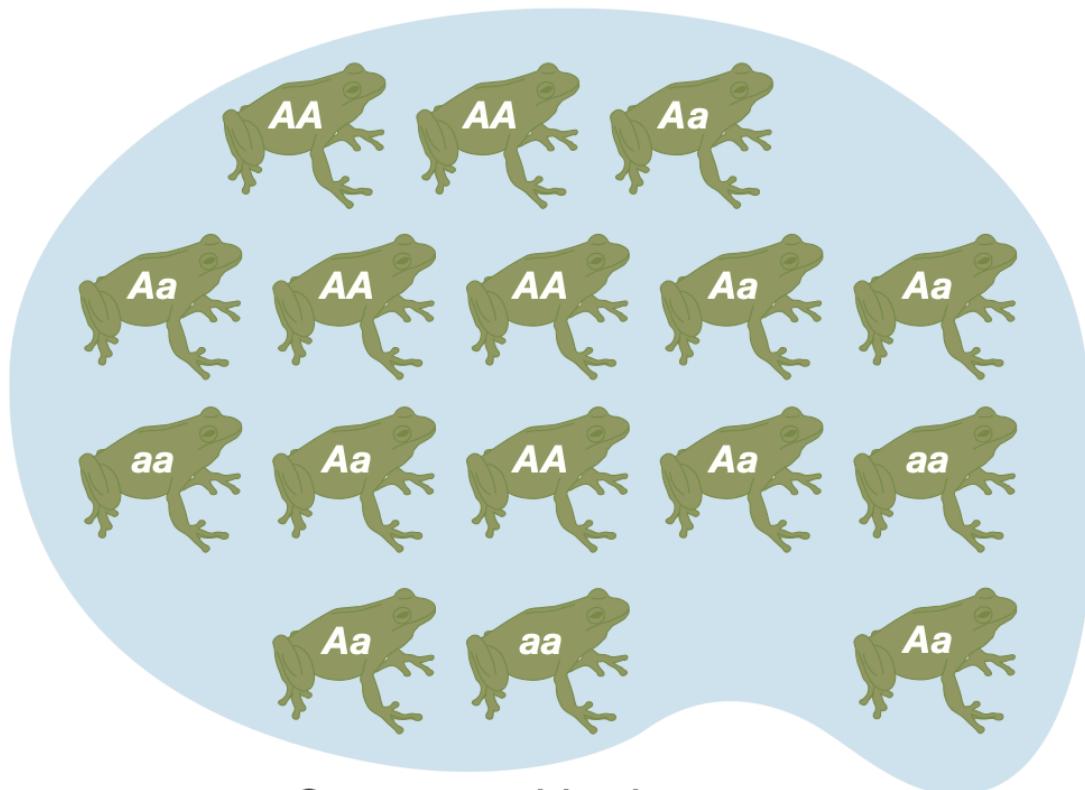


- count the number of variable sites
- account for the length of the region (sequence)
- account for the number of sequences (sample size)
- Average # of differences between two sequences

# GENE POOL AND GENOTYPE FREQUENCIES

The gene pool is the sum total of alleles in a population

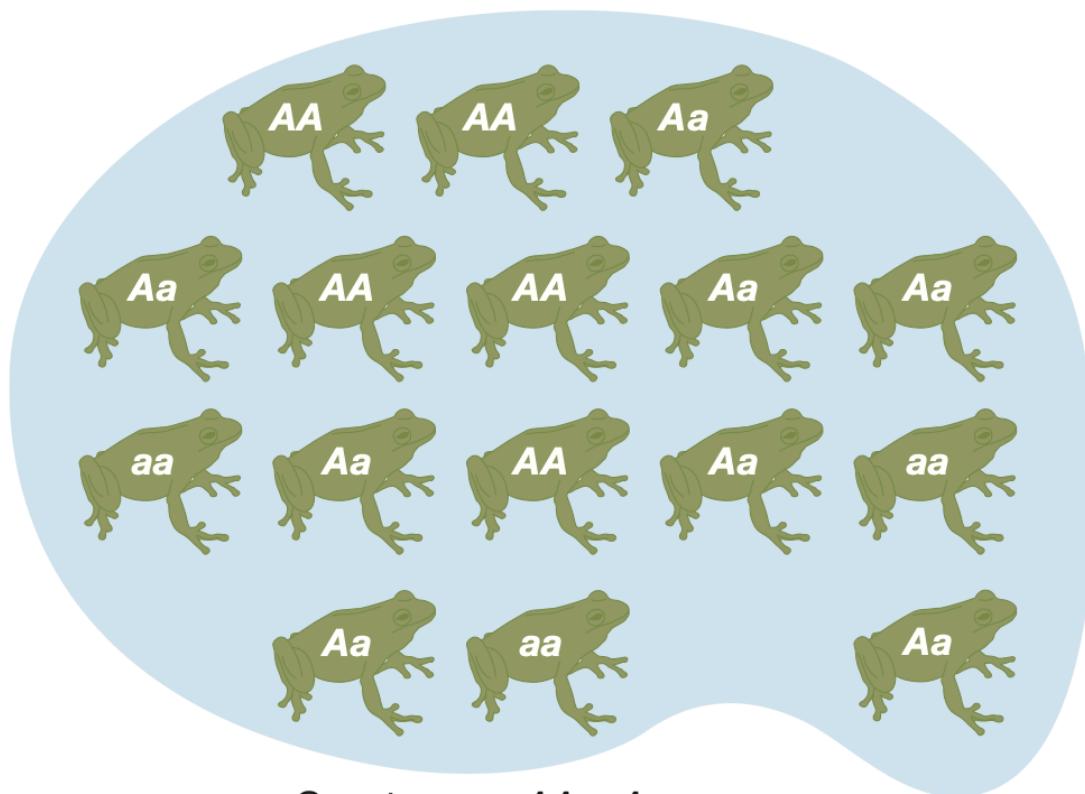
- what is a gene pool?



Genotypes	AA	Aa	aa
Number	5	8	3
Alleles	A	a	
Number	18	14	

# GENE POOL AND GENOTYPE FREQUENCIES

The gene pool is the sum total of alleles in a population

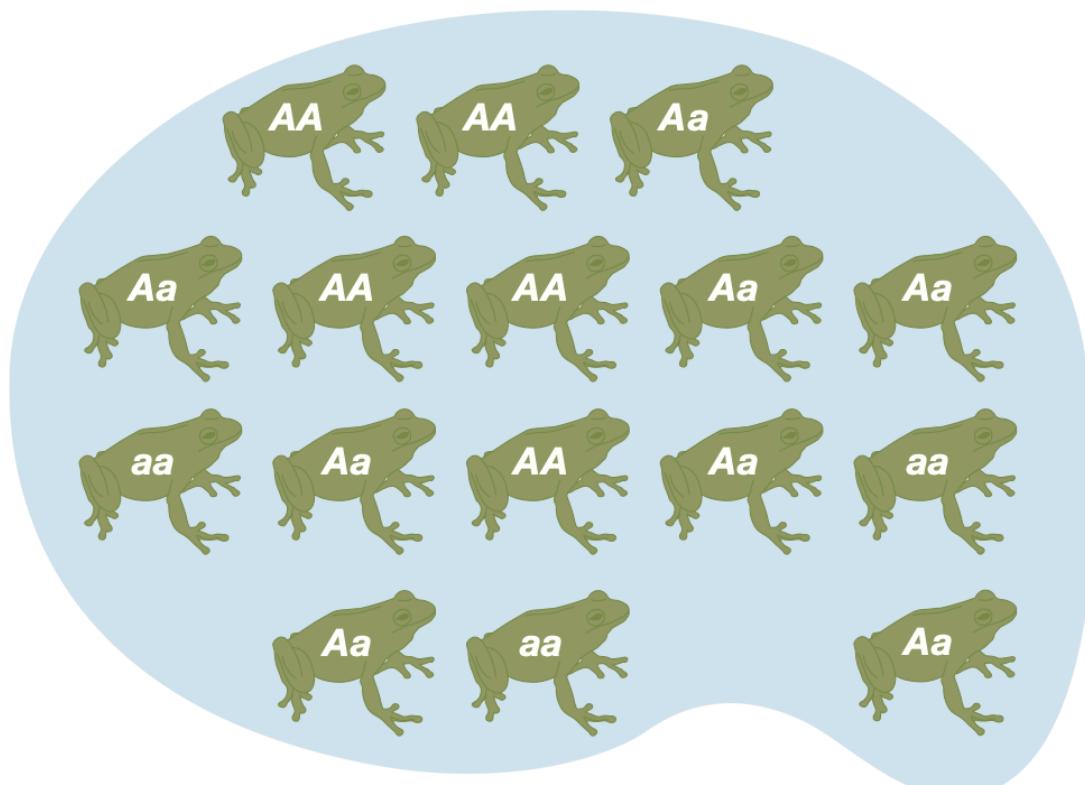


Genotypes	AA	Aa	aa
Number	5	8	3
Alleles	A	a	
Number	18	14	

- what is a gene pool?
- **Locus:** "region of interest", may be an entire gene, or a single nucleotide

# GENE POOL AND GENOTYPE FREQUENCIES

The gene pool is the sum total of alleles in a population

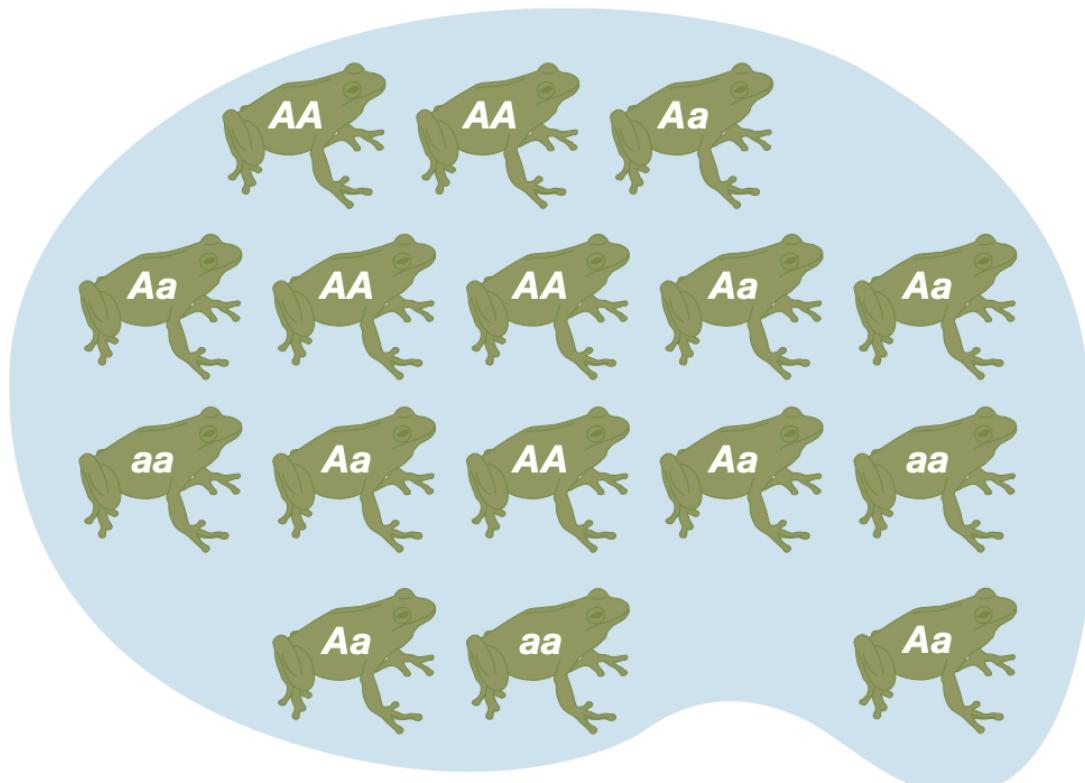


Genotypes	AA	Aa	aa
Number	5	8	3
Alleles	A	a	
Number	18	14	

- what is a gene pool?
- **Locus:** "region of interest", may be an entire gene, or a single nucleotide
- **Allele:** different "versions" of the sequence existing in a population

# GENE POOL AND GENOTYPE FREQUENCIES

The gene pool is the sum total of alleles in a population

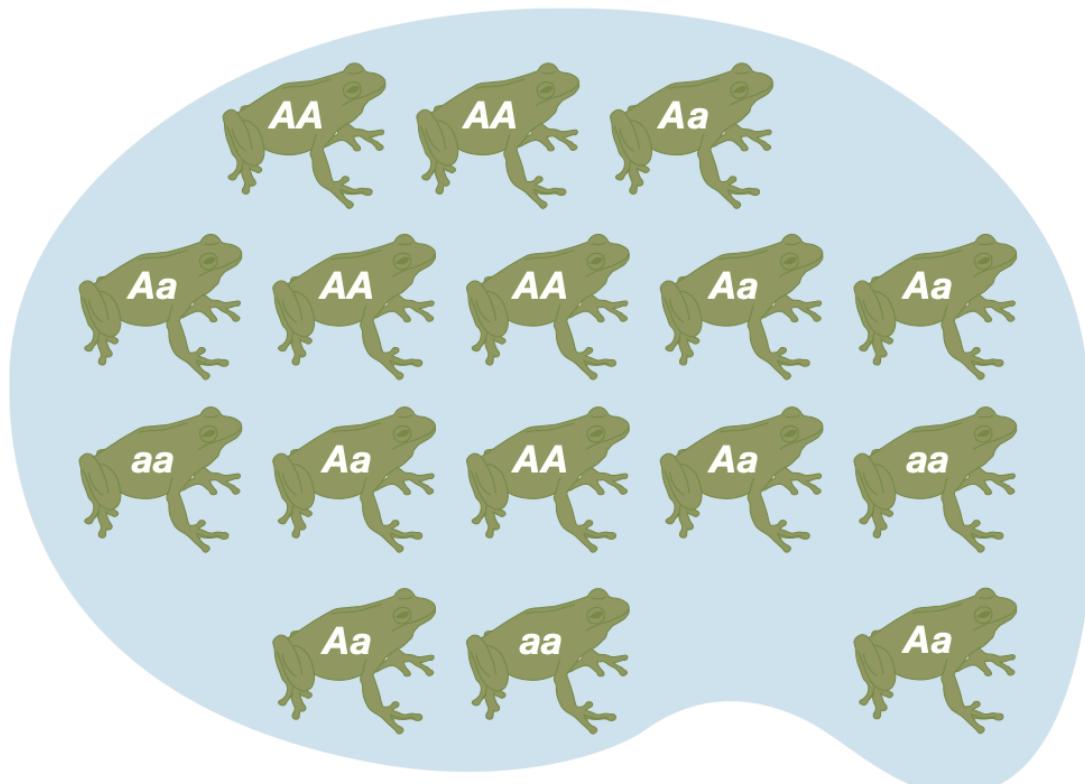


Genotypes	AA	Aa	aa
Number	5	8	3
Alleles	A	a	
Number	18	14	

- what is a gene pool?
- **Locus:** "region of interest", may be an entire gene, or a single nucleotide
- **Allele:** different "versions" of the sequence existing in a population
- **Genotype:** denotes the total allele combinations at a locus, e.g., two alleles for diploids.

# GENE POOL AND GENOTYPE FREQUENCIES

The gene pool is the sum total of alleles in a population

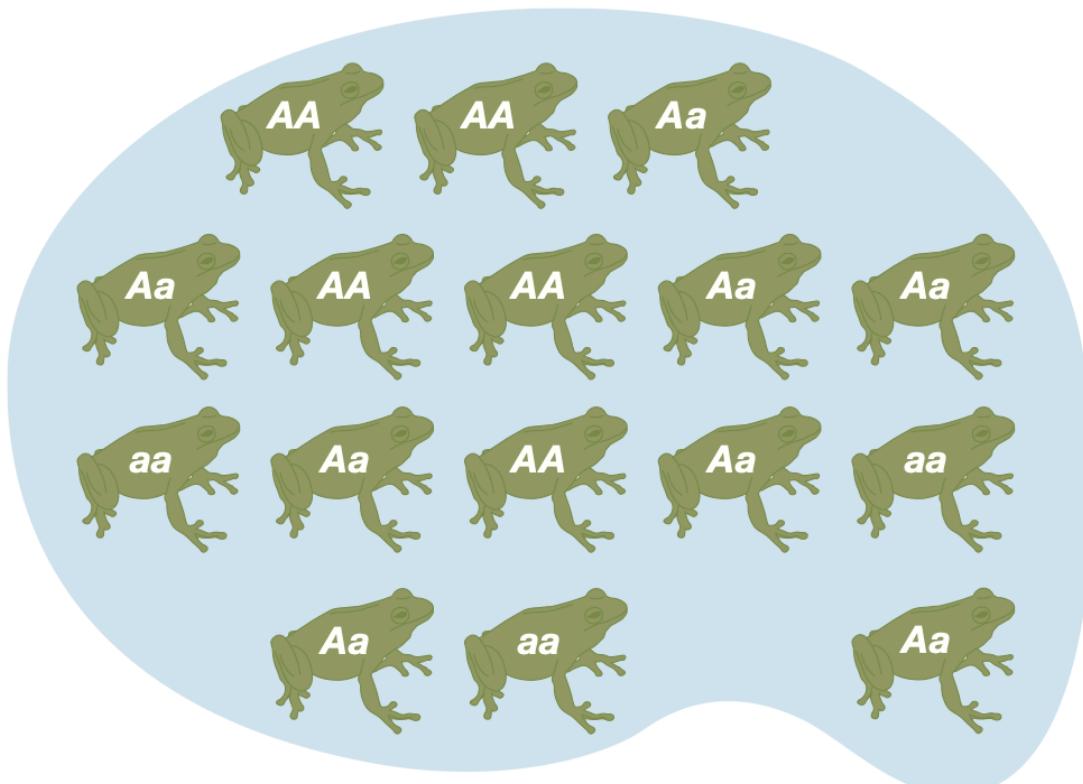


Genotypes	AA	Aa	aa
Number	5	8	3
Alleles	A	a	
Number	18	14	

- what is a gene pool?
- **Locus:** "region of interest", may be an entire gene, or a single nucleotide
- **Allele:** different "versions" of the sequence existing in a population
- **Genotype:** denotes the total allele combinations at a locus, e.g., two alleles for diploids.
- What are the genotype frequencies?

# GENE POOL AND GENOTYPE FREQUENCIES

The gene pool is the sum total of alleles in a population

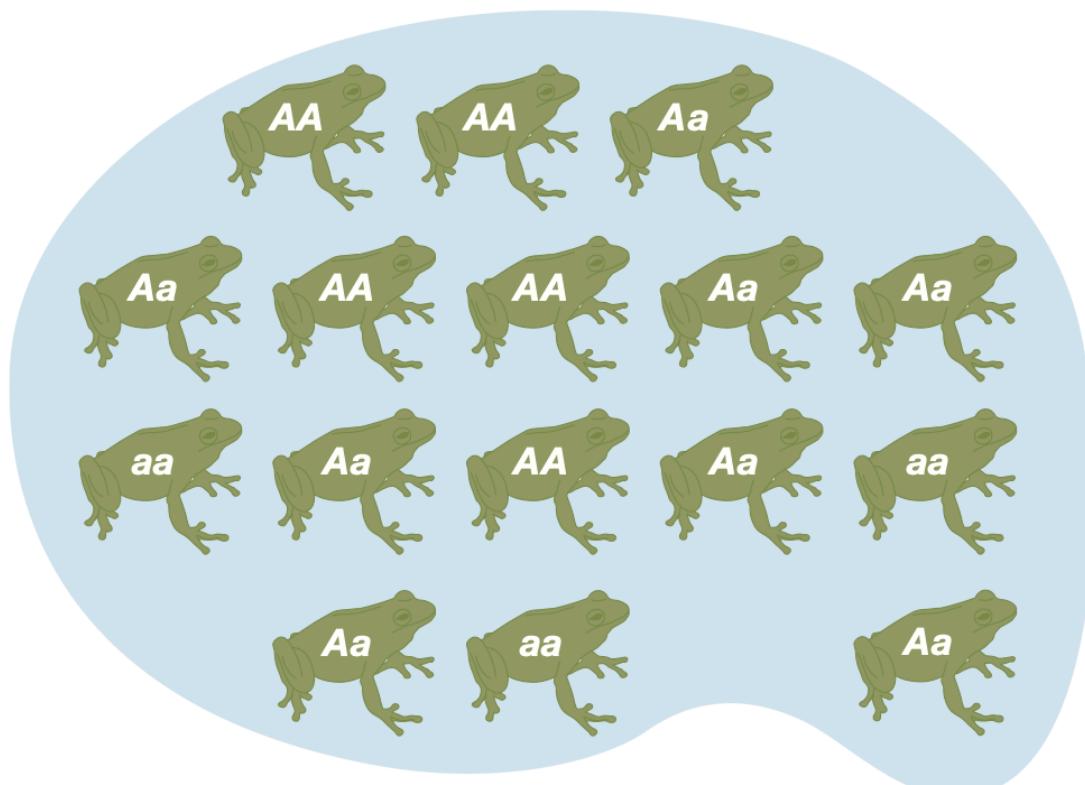


Genotypes	AA	Aa	aa
Number	5	8	3
Alleles	A	a	
Number	18	14	

- what is a gene pool?
- **Locus:** "region of interest", may be an entire gene, or a single nucleotide
- **Allele:** different "versions" of the sequence existing in a population
- **Genotype:** denotes the total allele combinations at a locus, e.g., two alleles for diploids.
- What are the genotype frequencies?
- What are the allele frequencies?

# GENE POOL AND GENOTYPE FREQUENCIES

The gene pool is the sum total of alleles in a population

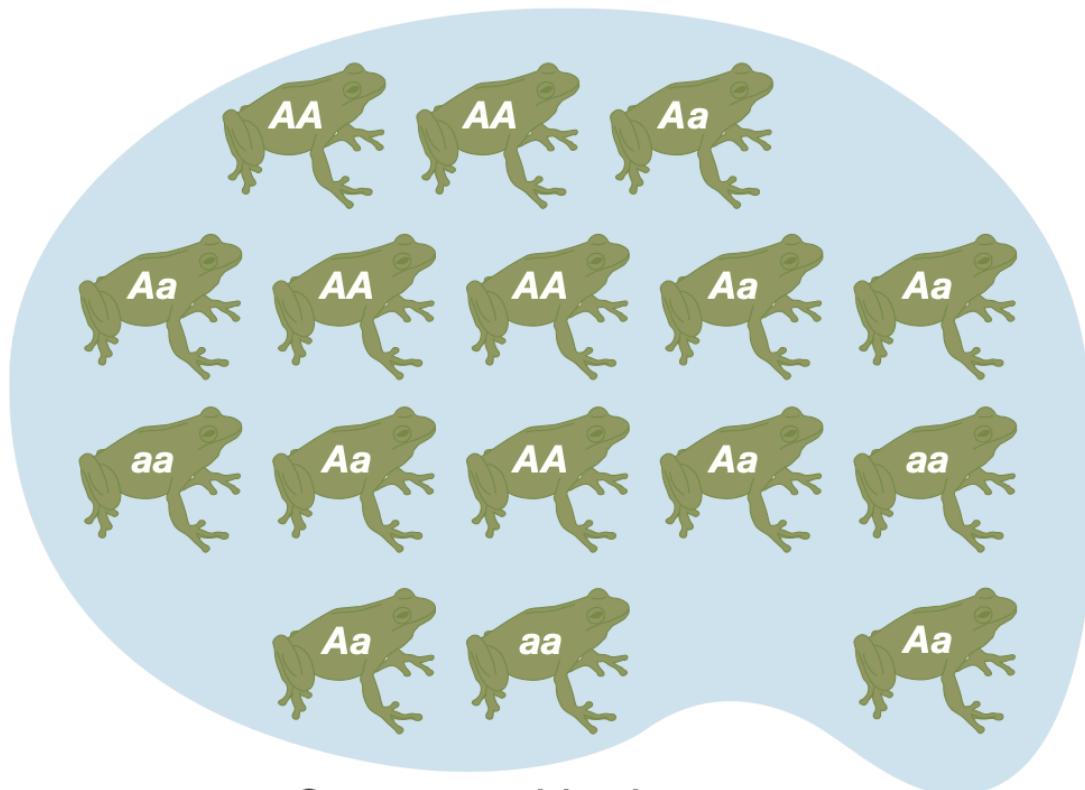


Genotypes	AA	Aa	aa
Number	5	8	3
Alleles	A	a	
Number	18	14	

- Knowing the allele frequency ( $p_A = 18/32$ ), what is a "reasonable guess" of the genotype frequency (e.g., for  $f_{A/A}$ )?

# GENE POOL AND GENOTYPE FREQUENCIES

The gene pool is the sum total of alleles in a population

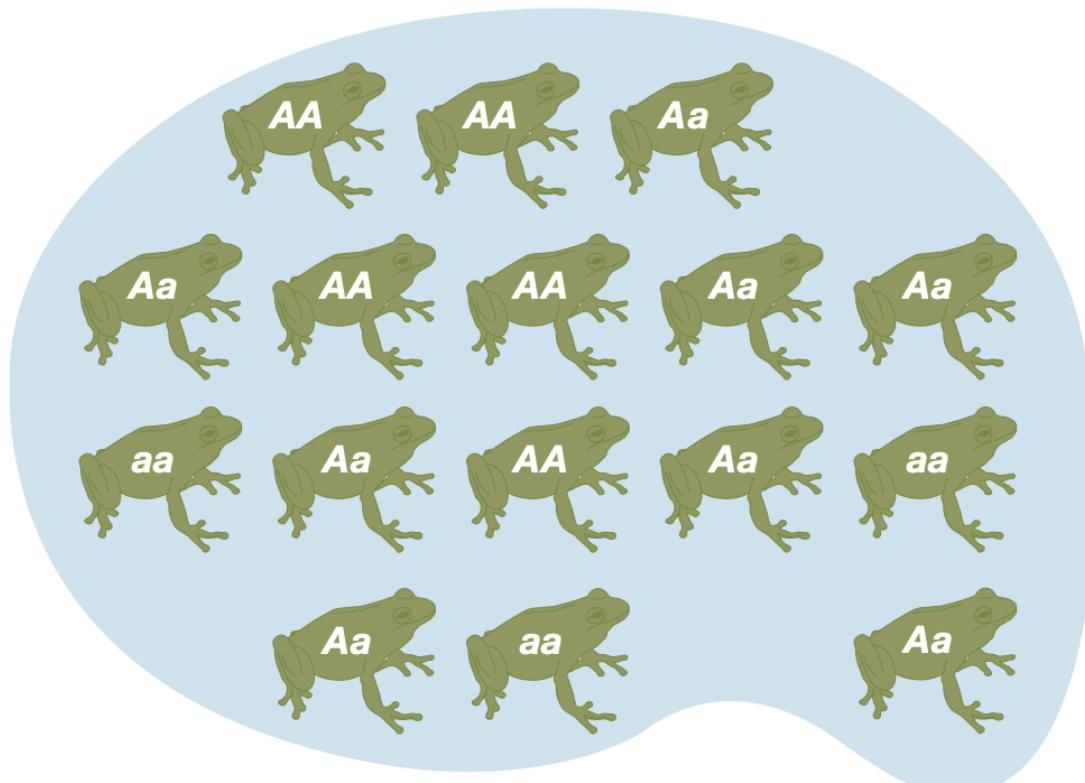


Genotypes	AA	Aa	aa
Number	5	8	3
Alleles	A	a	
Number	18	14	

- Knowing the allele frequency ( $p_A = 18/32$ ), what is a "reasonable guess" of the genotype frequency (e.g., for  $f_{A/A}$ )?
- How close is it to the observed genotype frequency?

# GENE POOL AND GENOTYPE FREQUENCIES

The gene pool is the sum total of alleles in a population



Genotypes	AA	Aa	aa
Number	5	8	3
Alleles	A	a	
Number	18	14	

- Knowing the allele frequency ( $p_A = 18/32$ ), what is a "reasonable guess" of the genotype frequency (e.g., for  $f_{A/A}$ )?
- How close is it to the observed genotype frequency?
- Now reflect on your intuitive guess, what assumptions need to be made for it to work?

# HARDY-WEINBERG EQUILIBRIUM

Suppose allele frequency of A = p, and frequency of a = q

$$\frac{f_{A/A} \quad f_{A/a} \quad f_{a/a}}{p^2 \quad 2pq \quad q^2}$$

# HARDY-WEINBERG EQUILIBRIUM

Suppose allele frequency of A = p, and frequency of a = q

$$\frac{f_{A/A} \quad f_{A/a} \quad f_{a/a}}{p^2 \quad 2pq \quad q^2}$$

- Assume the allele frequency of A is the same in females and males.

# HARDY-WEINBERG EQUILIBRIUM

Suppose allele frequency of A = p, and frequency of a = q

$$\frac{f_{A/A} \quad f_{A/a} \quad f_{a/a}}{p^2 \quad 2pq \quad q^2}$$

- Assume the allele frequency of A is the same in females and males.
- The probability of a baby frog to have A/A genotype is  $p^2$

# HARDY-WEINBERG EQUILIBRIUM

Suppose allele frequency of A = p, and frequency of a = q

$$\frac{f_{A/A} \quad f_{A/a} \quad f_{a/a}}{p^2 \quad 2pq \quad q^2}$$

- Assume the allele frequency of A is the same in females and males.
- The probability of a baby frog to have A/A genotype is  $p^2$
- What other assumptions are needed?

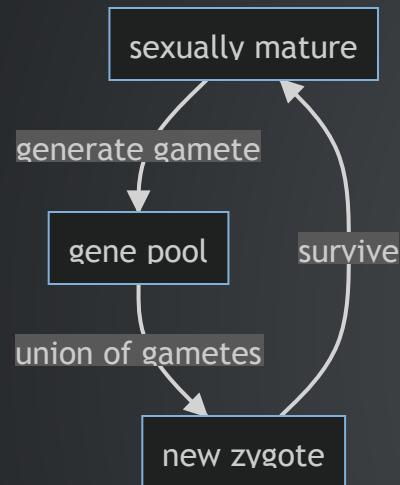
# HARDY-WEINBERG EQUILIBRIUM

Suppose allele frequency of A = p, and frequency of a = q

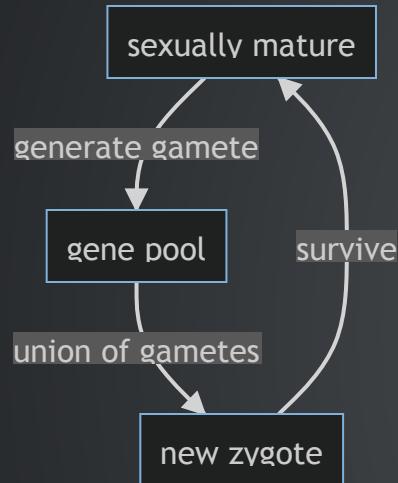
$$\frac{f_{A/A} \quad f_{A/a} \quad f_{a/a}}{p^2 \quad \quad \quad 2pq \quad \quad q^2}$$

- Assume the allele frequency of A is the same in females and males.
- The probability of a baby frog to have A/A genotype is  $p^2$
- What other assumptions are needed?
- Will any evolutionary force, such as migration and natural selection, affect the Hardy-Weinberg Equilibrium?

# HARDY-WEINBERG EQUILIBRIUM

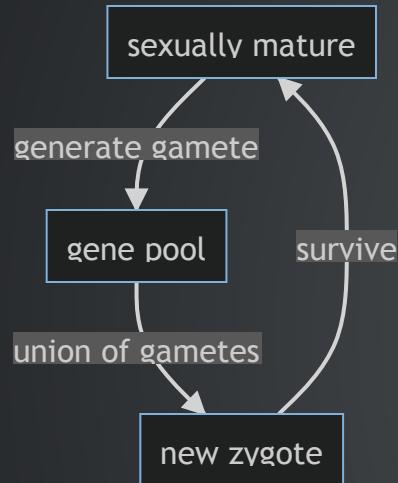


# HARDY-WEINBERG EQUILIBRIUM



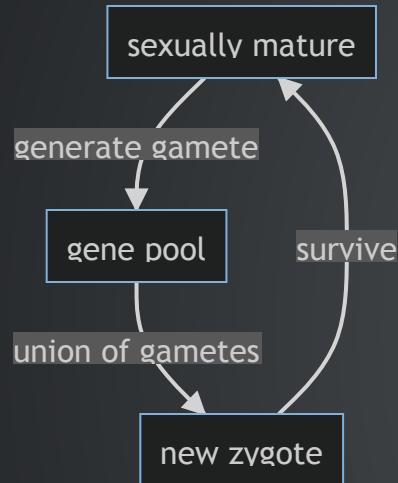
- HWE only refers to the step of forming new zygotes.

# HARDY-WEINBERG EQUILIBRIUM



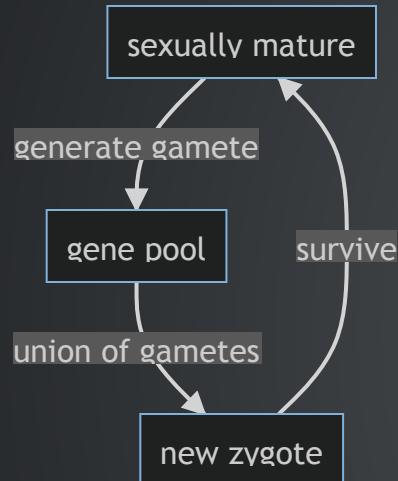
- HWE only refers to the step of forming new zygotes.
- Only assumptions needed are

# HARDY-WEINBERG EQUILIBRIUM



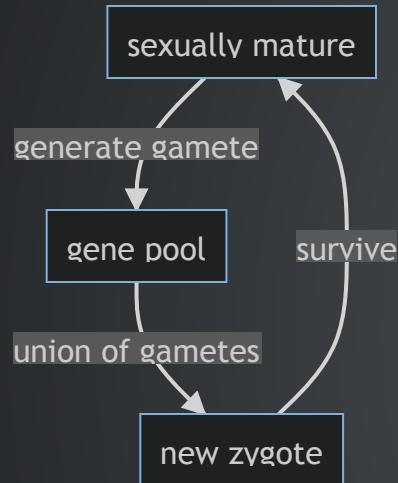
- HWE only refers to the step of forming new zygotes.
- Only assumptions needed are
  - random mating **with respect to the locus of interest**

# HARDY-WEINBERG EQUILIBRIUM



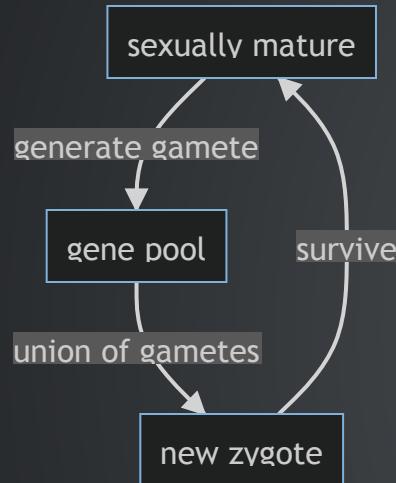
- HWE only refers to the step of forming new zygotes.
- Only assumptions needed are
  - random mating **with respect to the locus of interest**
  - same allele frequencies in both sexes

# HARDY-WEINBERG EQUILIBRIUM



- HWE only refers to the step of forming new zygotes.
- Only assumptions needed are
  - random mating **with respect to the locus of interest**
  - same allele frequencies in both sexes
- Under these assumptions, HWE is restored *instantaneously* upon mating

# HARDY-WEINBERG EQUILIBRIUM



- HWE only refers to the step of forming new zygotes.
- Only assumptions needed are
  - random mating **with respect to the locus of interest**
  - same allele frequencies in both sexes
- Under these assumptions, HWE is restored *instantaneously* upon mating
- Often hear about "no selection, no migration" only needed for HWE frequencies to represent a **long term equilibrium**.

# HWE EXERCISE

## Question

On the coastal islands of British Columbia there is a subspecies of black bear (*Ursus americanus kermodei*, Kermode's bear). Many members of this black bear subspecies are white; they're sometimes called spirit bears. These bears aren't hybrids with polar bears, nor are they albinos. They are homozygotes for a recessive change at the MC1R gene. Individuals who are GG at this SNP are white, while AA and AG individuals are black.

The genotype counts for the MC1R polymorphism in a sample of bears from British Columbia's island populations from Ritland et al. (2001) are:  $AA = 42$ ,  $AG = 24$ ,  $GG = 21$

What are the expected frequencies of the three genotypes under HWE?

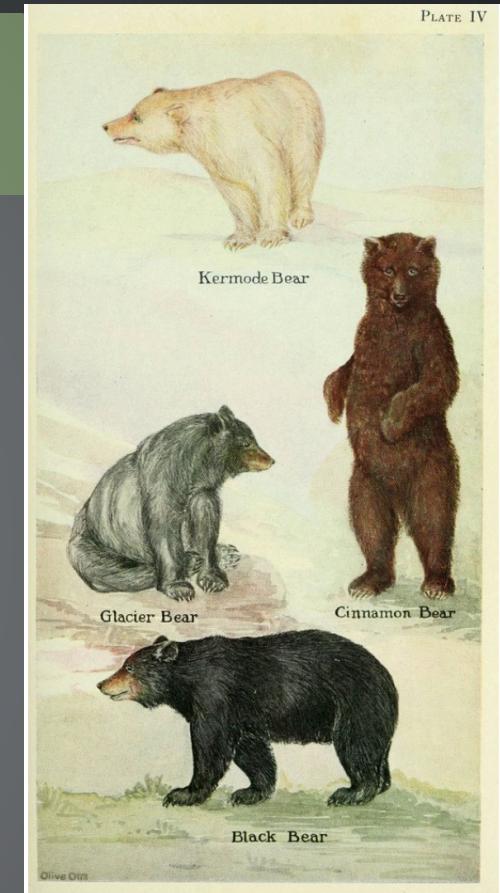
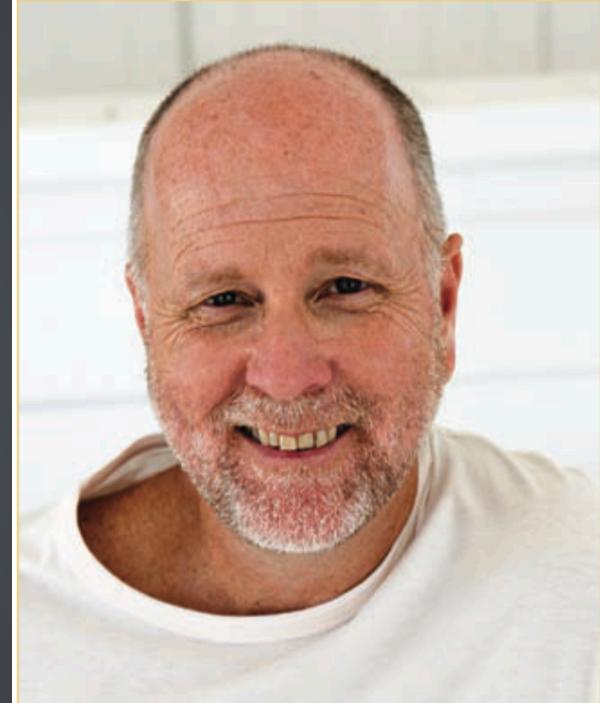


Figure 2.5: Kermode's bear (*Ursus americanus kermodei*). It's possible that this morph is favoured as the salmon these bears eat have a harder time seeing the light morph (KLINKA and REIMCHEN, 2009). The adaptive value of tasting like cinnamon is unknown.

# HARDY-WEINBERG EQUILIBRIUM

- Challenge: how does HWE apply to X-linked loci?
- 

Male pattern baldness

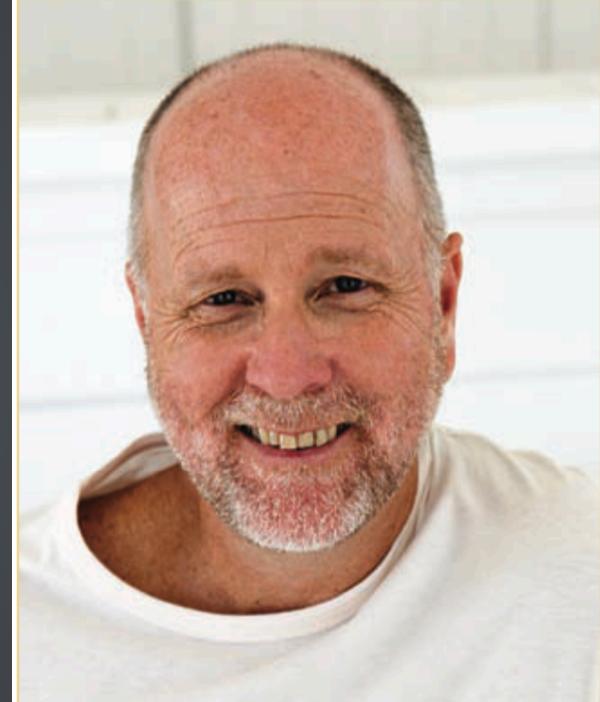


**FIGURE 18-9** Individual showing male pattern baldness, an X-chromosome-linked condition. [B2M Productions/Getty Images.]

# HARDY-WEINBERG EQUILIBRIUM

- Challenge: how does HWE apply to X-linked loci?
- 
- Male-baldness is an X-linked trait.

Male pattern baldness

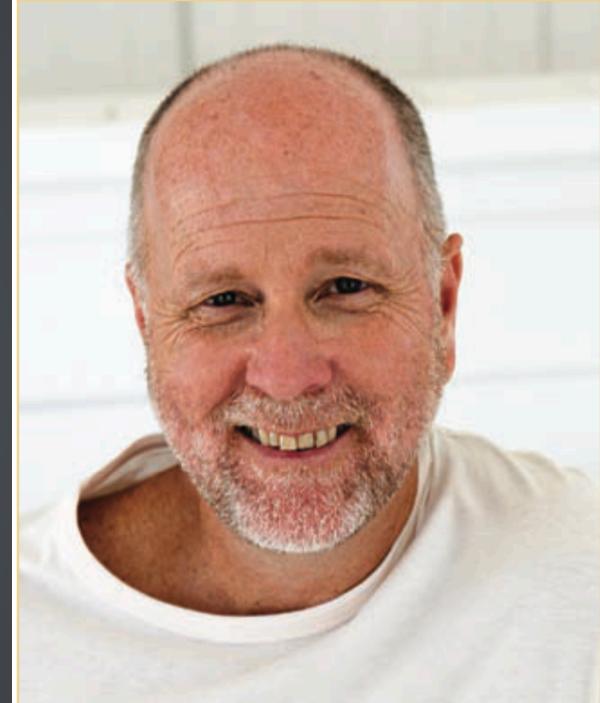


**FIGURE 18-9** Individual showing male pattern baldness, an X-chromosome-linked condition. [B2M Productions/Getty Images.]

# HARDY-WEINBERG EQUILIBRIUM

- Challenge: how does HWE apply to X-linked loci?
- 
- Male-baldness is an X-linked trait.
- Androgen Receptor (AR) is an X-linked gene involved in male development.

Male pattern baldness

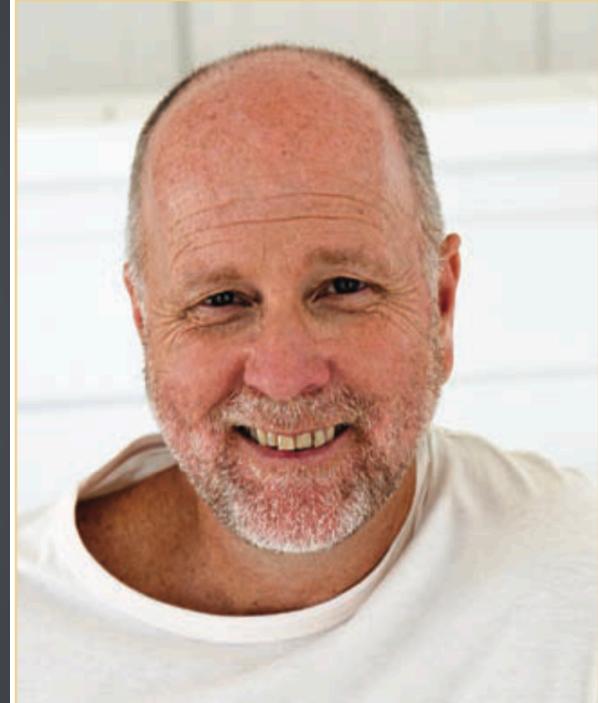


**FIGURE 18-9** Individual showing male pattern baldness, an X-chromosome-linked condition. [B2M Productions/Getty Images.]

# HARDY-WEINBERG EQUILIBRIUM

- Challenge: how does HWE apply to X-linked loci?
- 
- Male-baldness is an X-linked trait.
- Androgen Receptor (AR) is an X-linked gene involved in male development.
- an AR haplotype called *Eur-H1* is **strongly associated with pattern baldness**.

Male pattern baldness

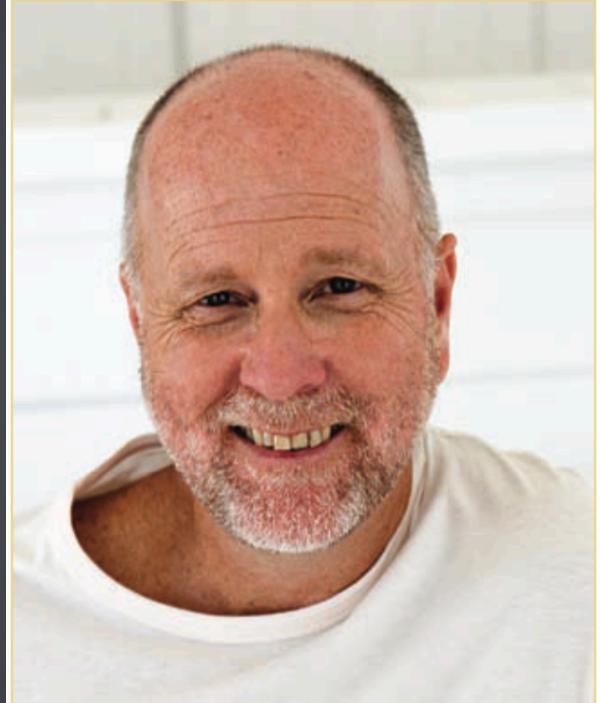


**FIGURE 18-9** Individual showing male pattern baldness, an X-chromosome-linked condition. [B2M Productions/Getty Images.]

# HARDY-WEINBERG EQUILIBRIUM

- Challenge: how does HWE apply to X-linked loci?
- 
- Male-baldness is an X-linked trait.
- Androgen Receptor (AR) is an X-linked gene involved in male development.
- an AR haplotype called *Eur-H1* is **strongly associated with pattern baldness**.
- *Eur-H1* occurs at a frequency of 0.71 in Europe. What percentage of European men are expected to have this allele?

Male pattern baldness

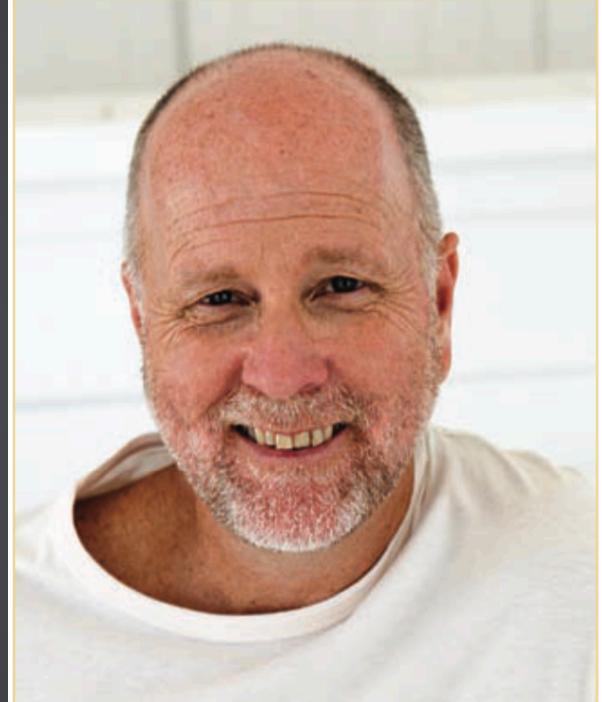


**FIGURE 18-9** Individual showing male pattern baldness, an X-chromosome-linked condition. [B2M Productions/Getty Images.]

# HARDY-WEINBERG EQUILIBRIUM

- Challenge: how does HWE apply to X-linked loci?
- 
- Male-baldness is an X-linked trait.
- Androgen Receptor (AR) is an X-linked gene involved in male development.
- an AR haplotype called *Eur-H1* is **strongly associated with pattern baldness**.
- *Eur-H1* occurs at a frequency of 0.71 in Europe. What percentage of European men are expected to have this allele?
- European women?

Male pattern baldness



**FIGURE 18-9** Individual showing male pattern baldness, an X-chromosome-linked condition. [B2M Productions/Getty Images.]

# HARDY-WEINBERG EQUILIBRIUM



- Solid line: mean genotype frequency; dashed line: HWE prediction. [Source](#)
- While HWE sounds too ideal to be a good fit to real data, researchers found that it actually holds remarkably well within populations. Since the main assumption for HWE to be true is that mating is random, how should one understand this?

# HARDY-WEINBERG EQUILIBRIUM

- Solid line: mean genotype frequency; dashed line: HWE prediction. [Source](#)
- While HWE sounds too ideal to be a good fit to real data, researchers found that it actually holds remarkably well within populations. Since the main assumption for HWE to be true is that mating is random, how should one understand this?

# **MAJOR VIOLATIONS TO HWE**

## **ASSORTATIVE MATING**

# **MAJOR VIOLATIONS TO HWE**

## **ASSORTATIVE MATING**

- Non-random mating *with respect to* the genotype at a locus

# MAJOR VIOLATIONS TO HWE

## ASSORTATIVE MATING

- Non-random mating *with respect to* the genotype at a locus
- Ex: negative or disassortative mating

# MAJOR VIOLATIONS TO HWE

## ASSORTATIVE MATING

- Non-random mating *with respect to* the genotype at a locus
- Ex: negative or disassortative mating
  - Major histocompatibility complex (MHC) locus, involved in immune response to pathogens, higher heterozygosity = more diversity ≈ higher fitness.

# MAJOR VIOLATIONS TO HWE

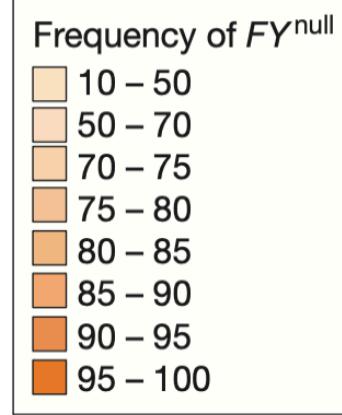
## ASSORTATIVE MATING

- Non-random mating *with respect to* the genotype at a locus
- Ex: negative or disassortative mating
  - Major histocompatibility complex (MHC) locus, involved in immune response to pathogens, higher heterozygosity = more diversity ≈ higher fitness.
  - "sweaty shirt test": MHC alleles affect mating choice, prefer individuals with different MHC alleles

# MAJOR VIOLATIONS TO HWE

## ISOLATION BY DISTANCE

(b)



Frequency variation for the  $FY^{\text{null}}$  allele of the Duffy blood group locus in Africa. [ Data from P. C. Sabeti et al., *Science* 312, 2006, 1614–1620.]

question

what's the consequence of isolation by distance?

# MAJOR VIOLATIONS TO HWE

## INBREEDING

# MAJOR VIOLATIONS TO HWE

## INBREEDING

- Mating between *close* relatives.

# MAJOR VIOLATIONS TO HWE

## INBREEDING

- Mating between *close* relatives.
- How does inbreeding violate HWE?

# MAJOR VIOLATIONS TO HWE

## INBREEDING

- Mating between *close* relatives.
- How does inbreeding violate HWE?
- Why is inbreeding potentially harmful in outcrossing populations (like humans)?

# MAJOR VIOLATIONS TO HWE

## INBREEDING

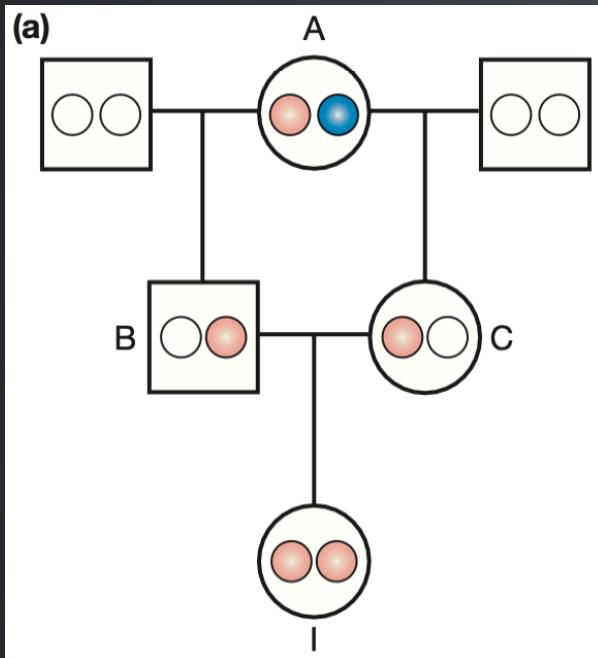
- Mating between *close* relatives.
- How does inbreeding violate HWE?
- Why is inbreeding potentially harmful in outcrossing populations (like humans)?
- Selfing (extreme form of inbreeding) is common in animals and plants. What may be some advantages?

# MAJOR VIOLATIONS TO HWE

## INBREEDING

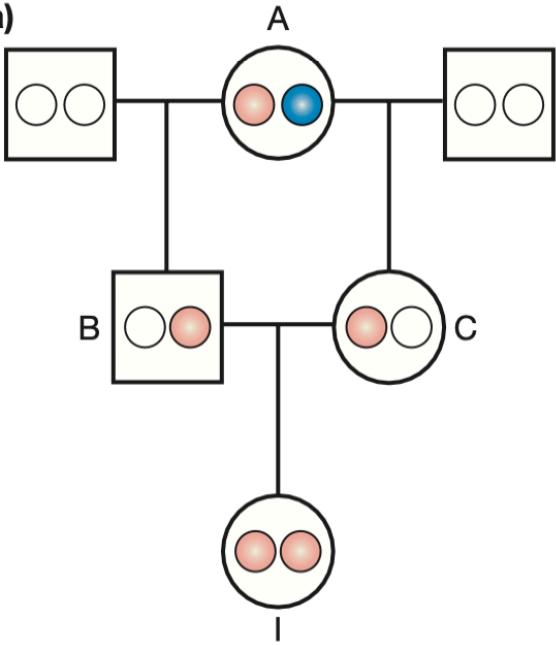
- Mating between *close* relatives.
- How does inbreeding violate HWE?
- Why is inbreeding potentially harmful in outcrossing populations (like humans)?
- Selfing (extreme form of inbreeding) is common in animals and plants. What may be some advantages?
- Identical-by-Descent (IBD)

# MEASURING INBREEDING



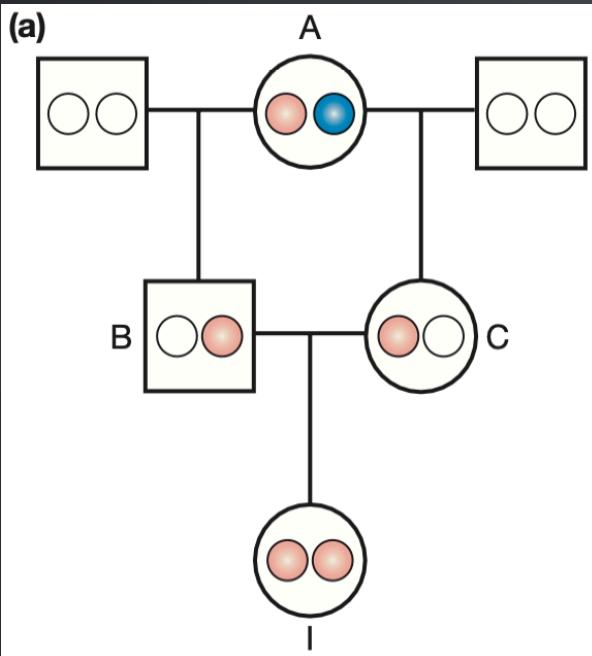
# MEASURING INBREEDING

(a)



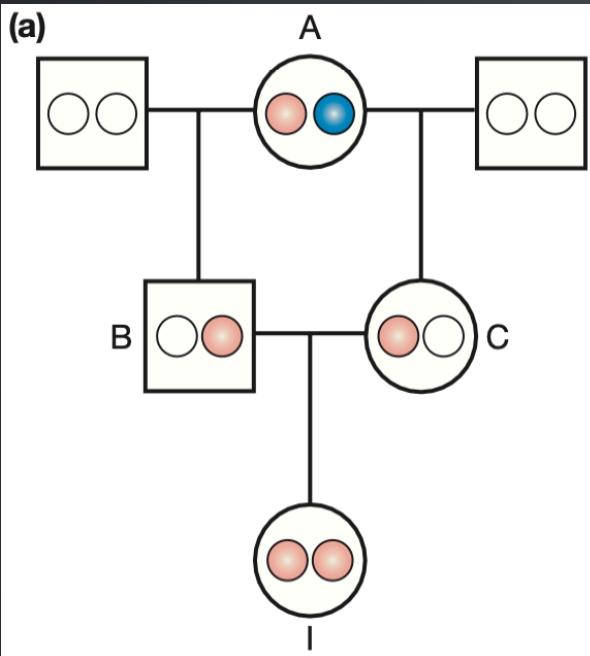
- **Inbreeding coefficient ( $F$ ):** probability that two alleles in an individual trace back to the same copy in a common ancestor.

## MEASURING INBREEDING



- **Inbreeding coefficient ( $F$ ):** probability that two alleles in an individual trace back to the same copy in a common ancestor.
- **Identical-by-Descent (IBD):** define two alleles to be IBD if they are identical due to transmission from a common ancestor in the past few generations

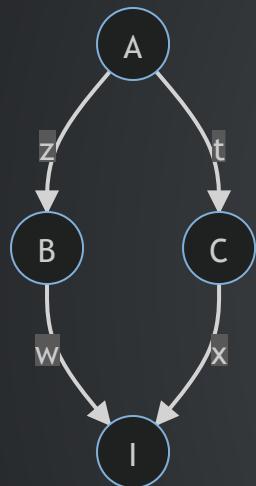
# MEASURING INBREEDING



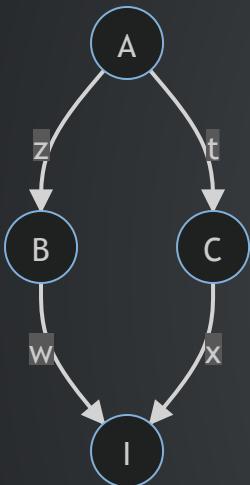
- **Inbreeding coefficient ( $F$ ):** probability that two alleles in an individual trace back to the same copy in a common ancestor.
- **Identical-by-Descent (IBD):** define two alleles to be IBD if they are identical due to transmission from a common ancestor in the past few generations
- A "closed loop" indicates inbreeding.

# MEASURING INBREEDING

- What's the probability that individual I is inbred? (calculating its inbreeding coefficient)

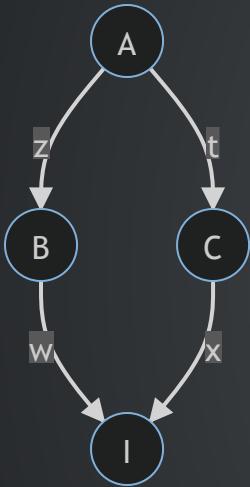


# MEASURING INBREEDING



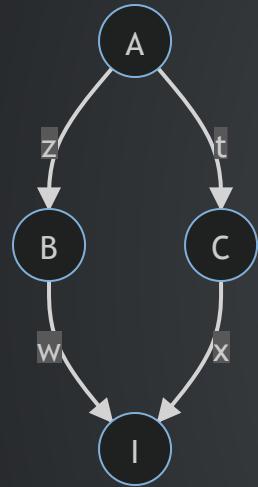
- What's the probability that individual I is inbred? (calculating its inbreeding coefficient)
- The alleles transmitted through each mating are labeled x, t, w and z. Use " $\sim$ " to indicate IBD.

# MEASURING INBREEDING



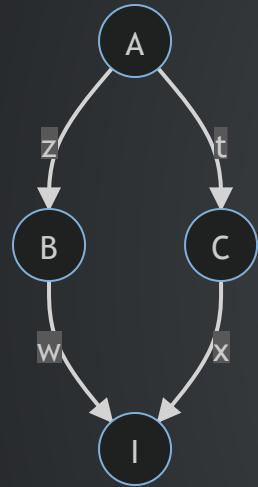
- What's the probability that individual I is inbred? (calculating its inbreeding coefficient)
- The alleles transmitted through each mating are labeled x, t, w and z. Use " $\sim$ " to indicate IBD.
- We wish to calculate  $P(w \sim x)$

# MEASURING INBREEDING



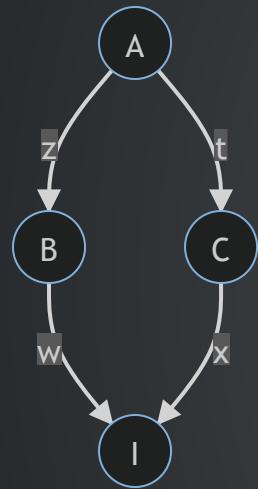
- What's the probability that individual I is inbred? (calculating its inbreeding coefficient)
- The alleles transmitted through each mating are labeled  $x$ ,  $t$ ,  $w$  and  $z$ . Use " $\sim$ " to indicate IBD.
- We wish to calculate  $P(w \sim x)$
- $P(x \sim t) = \frac{1}{2}$ ;  $P(w \sim z) = \frac{1}{2}$

# MEASURING INBREEDING



- What's the probability that individual I is inbred? (calculating its inbreeding coefficient)
- The alleles transmitted through each mating are labeled x, t, w and z. Use " $\sim$ " to indicate IBD.
- We wish to calculate  $P(w \sim x)$
- $P(x \sim t) = \frac{1}{2}$ ;  $P(w \sim z) = \frac{1}{2}$
- $P(z \sim t) = \frac{1}{2} + \frac{1}{2}F_A$

# MEASURING INBREEDING



- What's the probability that individual I is inbred? (calculating its inbreeding coefficient)
- The alleles transmitted through each mating are labeled  $x$ ,  $t$ ,  $w$  and  $z$ . Use " $\sim$ " to indicate IBD.
- We wish to calculate  $P(w \sim x)$
- $P(x \sim t) = \frac{1}{2}$ ;  $P(w \sim z) = \frac{1}{2}$
- $P(z \sim t) = \frac{1}{2} + \frac{1}{2}F_A$
- $F_I = P(z \sim t) \times P(w \sim z) \times P(x \sim t) = (\frac{1}{2})^3(1 + F_A)$

## CONSEQUENCES OF INBREEDING

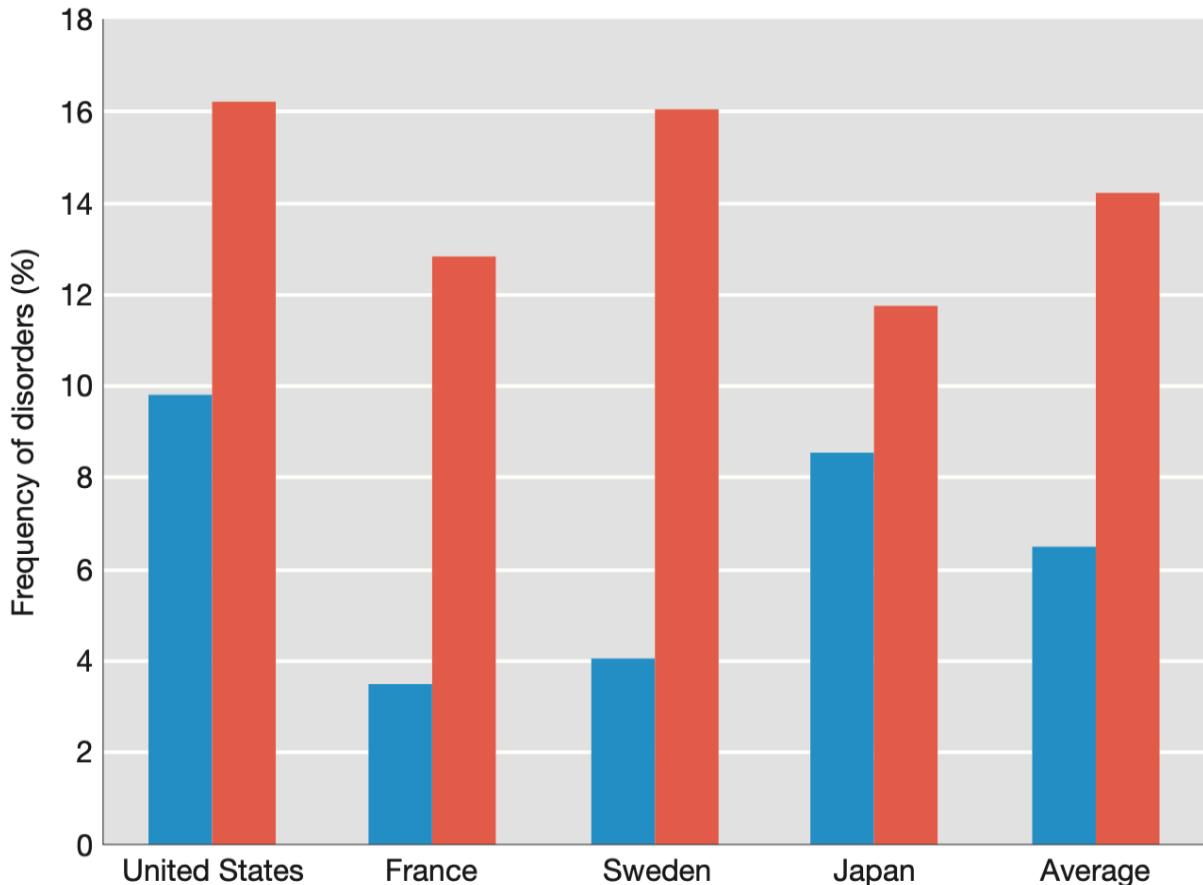
Examples of inbreeding effect on frequency of homozygous recessives

TABLE 18-3 Number of Homozygous Recessives per 10,000 Individuals for Different Allele Frequencies ( $q$ )

Mating	$F$	$q = 0.01$	$q = 0.005$	$q = 0.001$
Unrelated parents	0.0	1.00	0.25	0.01
Parent-offspring or brother-sister	1/4	25.75	12.69	2.51
Half-sib	1/8	13.38	6.47	1.26
First cousin	1/16	7.19	3.36	0.63
Second cousin	1/64	2.55	1.03	0.17

# CONSEQUENCE OF INBREEDING

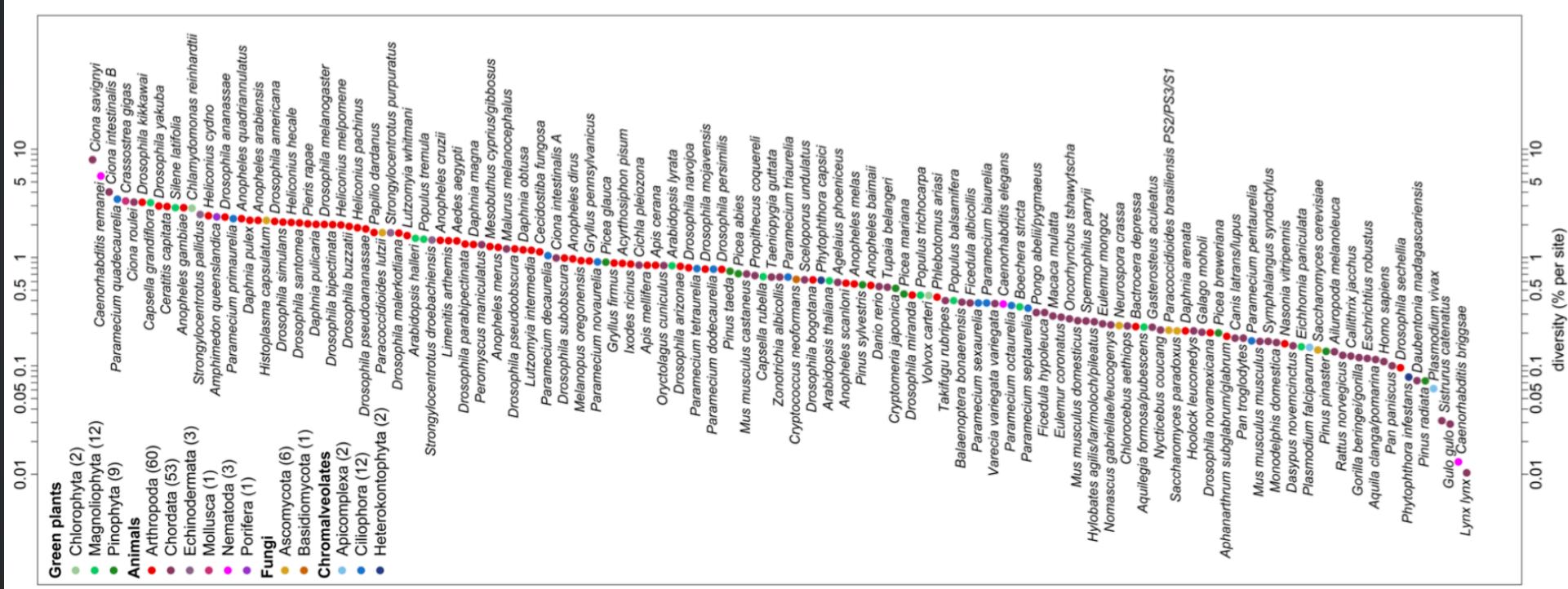
Inbreeding leads to an increase in recessive genetic disorders



*Frequency of genetic disorders among children of unrelated parents (blue columns) compared to that of children of parents who are first cousins (red columns).*

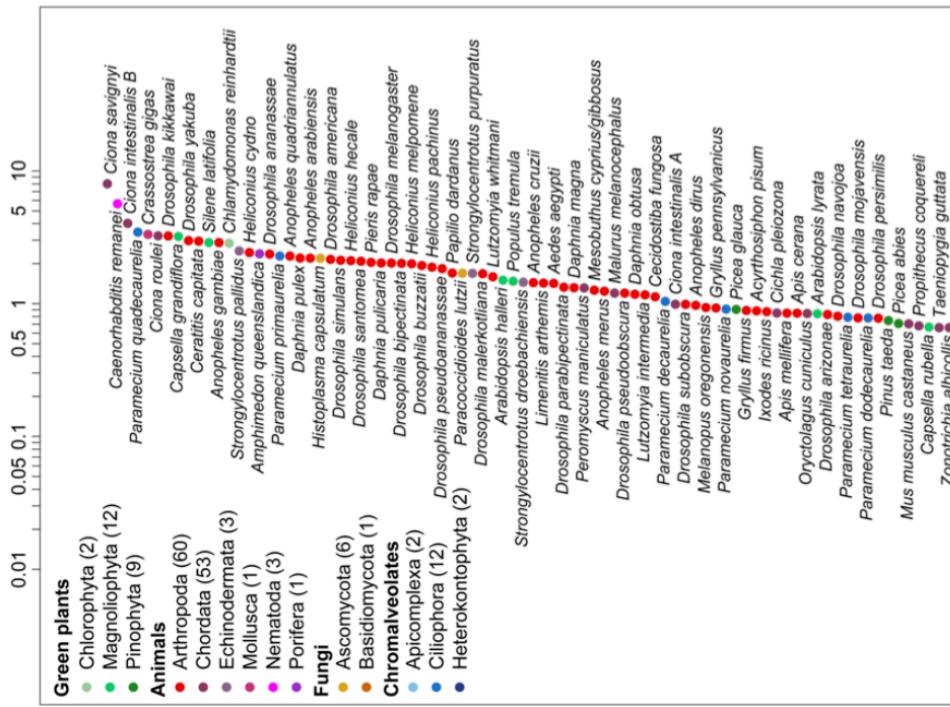
*Data from C. Stern, Principles of Human Genetics, W. H. Freeman, 1973.*

# LEVEL OF GENETIC VARIATION VARY ACROSS SPECIES

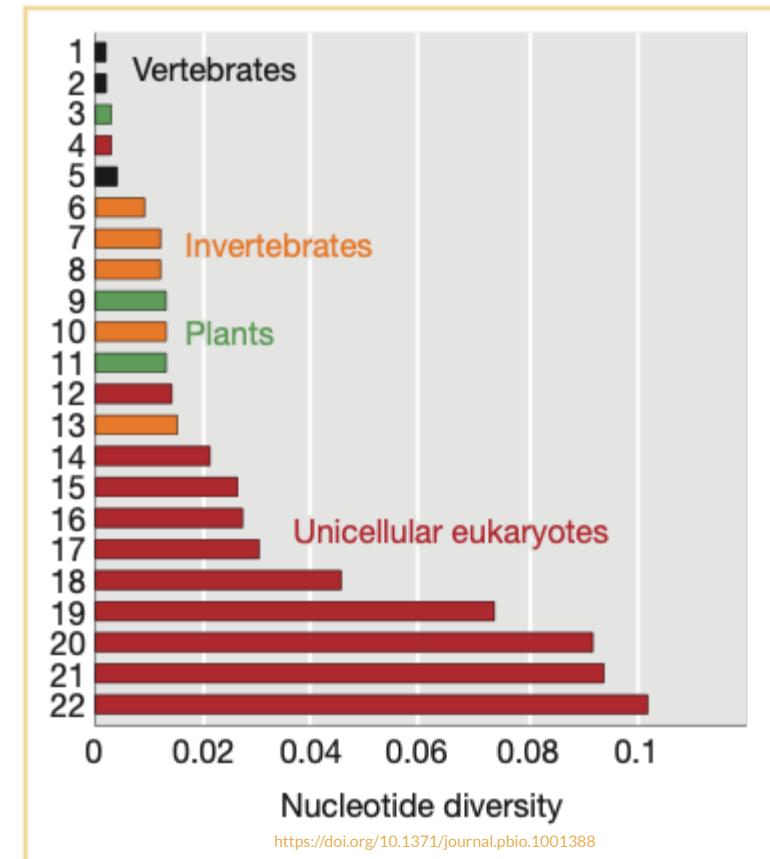


1: Leffler et al. 2012. "Revisiting an Old Riddle: What Determines Genetic Diversity Levels within Species?" *PLoS Biology* 10 (9): e1001388. <https://doi.org/10.1371/journal.pbio.1001388>.

# LEVEL OF GENETIC VARIATION VARY ACROSS SPECIES



1: Leffler et al. 2012. "Revisiting an Old Riddle: What Determines Genetic Diversity Le



M. Lynch and J. S. Conery, Science 2003

# WHAT FORCES SHAPE THE LEVEL AND PATTERN OF GENETIC VARIATION?

# WHAT FORCES SHAPE THE LEVEL AND PATTERN OF GENETIC VARIATION?

- How do new alleles enter the gene pool?

# WHAT FORCES SHAPE THE LEVEL AND PATTERN OF GENETIC VARIATION?

- How do new alleles enter the gene pool?
- What forces drive changes in the frequency of alleles?

# WHAT FORCES SHAPE THE LEVEL AND PATTERN OF GENETIC VARIATION?

- How do new alleles enter the gene pool?
- What forces drive changes in the frequency of alleles?
- How can existing genetic variation recombine to create novel combinations of alleles?

# WHAT FORCES SHAPE THE LEVEL AND PATTERN OF GENETIC VARIATION?

- How do new alleles enter the gene pool?
- What forces drive changes in the frequency of alleles?
- How can existing genetic variation recombine to create novel combinations of alleles?
- Forces examined next
  1. mutation
  2. migration
  3. drift
  4. recombination
  5. selection

# MUTATION

TABLE 18-5 Approximate Mutation Rates per Generation per Haploid Genome

Organism	SNP mutations (per bp)	Microsatellite
<i>Arabidopsis</i>	$7 \times 10^{-9}$	$9 \times 10^{-4}$
Maize	$3 \times 10^{-8}$	$8 \times 10^{-4}$
<i>E. coli</i>	$5 \times 10^{-10}$	—
Yeast	$5 \times 10^{-10}$	$4 \times 10^{-5}$
<i>C. elegans</i>	$3 \times 10^{-9}$	$4 \times 10^{-3}$
<i>Drosophila</i>	$4 \times 10^{-9}$	$9 \times 10^{-6}$
Mouse	$4 \times 10^{-9}$	$3 \times 10^{-4}$
Human	$3 \times 10^{-8}$	$6 \times 10^{-4}$

**Note:** Microsatellite rate is for di- or trinucleotide repeat microsatellites.

**Source:** Data from multiple published studies.

- **Mutation rate ( $\mu$ ):** probability that a copy of an allele changes to some other allelic form *in one generation*

# MUTATION

TABLE 18-5 Approximate Mutation Rates per Generation per Haploid Genome

Organism	SNP mutations (per bp)	Microsatellite
<i>Arabidopsis</i>	$7 \times 10^{-9}$	$9 \times 10^{-4}$
Maize	$3 \times 10^{-8}$	$8 \times 10^{-4}$
<i>E. coli</i>	$5 \times 10^{-10}$	—
Yeast	$5 \times 10^{-10}$	$4 \times 10^{-5}$
<i>C. elegans</i>	$3 \times 10^{-9}$	$4 \times 10^{-3}$
<i>Drosophila</i>	$4 \times 10^{-9}$	$9 \times 10^{-6}$
Mouse	$4 \times 10^{-9}$	$3 \times 10^{-4}$
Human	$3 \times 10^{-8}$	$6 \times 10^{-4}$

**Note:** Microsatellite rate is for di- or trinucleotide repeat microsatellites.

**Source:** Data from multiple published studies.

- **Mutation rate ( $\mu$ ):** probability that a copy of an allele changes to some other allelic form *in one generation*
- doesn't just apply to a single nucleotide position - for any locus

# MUTATION

TABLE 18-5 Approximate Mutation Rates per Generation per Haploid Genome

Organism	SNP mutations (per bp)	Microsatellite
<i>Arabidopsis</i>	$7 \times 10^{-9}$	$9 \times 10^{-4}$
Maize	$3 \times 10^{-8}$	$8 \times 10^{-4}$
<i>E. coli</i>	$5 \times 10^{-10}$	—
Yeast	$5 \times 10^{-10}$	$4 \times 10^{-5}$
<i>C. elegans</i>	$3 \times 10^{-9}$	$4 \times 10^{-3}$
<i>Drosophila</i>	$4 \times 10^{-9}$	$9 \times 10^{-6}$
Mouse	$4 \times 10^{-9}$	$3 \times 10^{-4}$
Human	$3 \times 10^{-8}$	$6 \times 10^{-4}$

**Note:** Microsatellite rate is for di- or trinucleotide repeat microsatellites.

**Source:** Data from multiple published studies.

- **Mutation rate ( $\mu$ ):** probability that a copy of an allele changes to some other allelic form *in one generation*
- doesn't just apply to a single nucleotide position - for any locus
- in a single generation is important - other forces can change allele forms, but they act after the formation of zygote stage

# MUTATION

TABLE 18-5 Approximate Mutation Rates per Generation per Haploid Genome

Organism	SNP mutations (per bp)	Microsatellite
<i>Arabidopsis</i>	$7 \times 10^{-9}$	$9 \times 10^{-4}$
Maize	$3 \times 10^{-8}$	$8 \times 10^{-4}$
<i>E. coli</i>	$5 \times 10^{-10}$	—
Yeast	$5 \times 10^{-10}$	$4 \times 10^{-5}$
<i>C. elegans</i>	$3 \times 10^{-9}$	$4 \times 10^{-3}$
<i>Drosophila</i>	$4 \times 10^{-9}$	$9 \times 10^{-6}$
Mouse	$4 \times 10^{-9}$	$3 \times 10^{-4}$
Human	$3 \times 10^{-8}$	$6 \times 10^{-4}$

**Note:** Microsatellite rate is for di- or trinucleotide repeat microsatellites.

**Source:** Data from multiple published studies.

- **Mutation rate ( $\mu$ ):** probability that a copy of an allele changes to some other allelic form *in one generation*
- doesn't just apply to a single nucleotide position - for any locus
- in a single generation is important - other forces can change allele forms, but they act after the formation of zygote stage
- data in human (circa 2009) gave an estimate of  $\sim 3.0 \times 10^{-8}$  mutations/nucleotide/generation for a part of the Y-chromosome. if we extrapolate this to the entire human genome, we get an estimate of about 100 new mutations one would inherit from each of our parents *on average*.

# MIGRATION

- In addition to mutation, migration is the other way by which new variation can be introduced into a population (not counting **new combinations**, see recombination)
- when there is isolation by distance, migration (or gene flow) is a homogenizing force, preventing allele frequencies from diverging too far

# RECOMBINATION

- Alleles are not gained or lost, but new combinations (haplotypes) are created by "mixing" existing ones.

# RECOMBINATION

- Alleles are not gained or lost, but new combinations (haplotypes) are created by "mixing" existing ones.
- Let's consider the observed and expected frequencies of the four possible haplotypes for two loci, each with two alleles (A and B, with alleles A/a and B/b)

# RECOMBINATION

- Alleles are not gained or lost, but new combinations (haplotypes) are created by "mixing" existing ones.
- Let's consider the observed and expected frequencies of the four possible haplotypes for two loci, each with two alleles (A and B, with alleles A/a and B/b)
  - the four possible haplotypes are?

# RECOMBINATION

- Alleles are not gained or lost, but new combinations (haplotypes) are created by "mixing" existing ones.
- Let's consider the observed and expected frequencies of the four possible haplotypes for two loci, each with two alleles (A and B, with alleles A/a and B/b)
  - the four possible haplotypes are?
  - if the frequencies of the alleles are  $p_A$  and  $p_B$ , what are the expected frequencies for the four haplotypes naively?

# RECOMBINATION

- Alleles are not gained or lost, but new combinations (haplotypes) are created by "mixing" existing ones.
- Let's consider the observed and expected frequencies of the four possible haplotypes for two loci, each with two alleles (A and B, with alleles A/a and B/b)
  - the four possible haplotypes are?
  - if the frequencies of the alleles are  $p_A$  and  $p_B$ , what are the expected frequencies for the four haplotypes naively?
  - $P_{AB} = p_A \times p_B$  etc.

# RECOMBINATION

- Alleles are not gained or lost, but new combinations (haplotypes) are created by "mixing" existing ones.
- Let's consider the observed and expected frequencies of the four possible haplotypes for two loci, each with two alleles (A and B, with alleles A/a and B/b)
  - the four possible haplotypes are?
  - if the frequencies of the alleles are  $p_A$  and  $p_B$ , what are the expected frequencies for the four haplotypes naively?
  - $P_{AB} = p_A \times p_B$  etc.
  - If the observed frequencies are very close to the expected, we say that the two loci are in **linkage equilibrium**

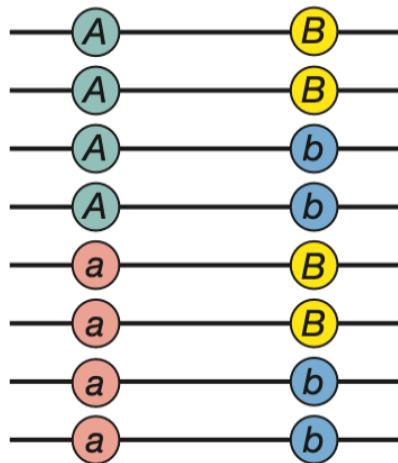
# RECOMBINATION

- Alleles are not gained or lost, but new combinations (haplotypes) are created by "mixing" existing ones.
- Let's consider the observed and expected frequencies of the four possible haplotypes for two loci, each with two alleles (A and B, with alleles A/a and B/b)
  - the four possible haplotypes are?
  - if the frequencies of the alleles are  $p_A$  and  $p_B$ , what are the expected frequencies for the four haplotypes naively?
  - $P_{AB} = p_A \times p_B$  etc.
  - If the observed frequencies are very close to the expected, we say that the two loci are in **linkage equilibrium**
  - If not, then the loci are said to be in **linkage disequilibrium (LD)**.

# LINKAGE DISEQUILIBRIUM

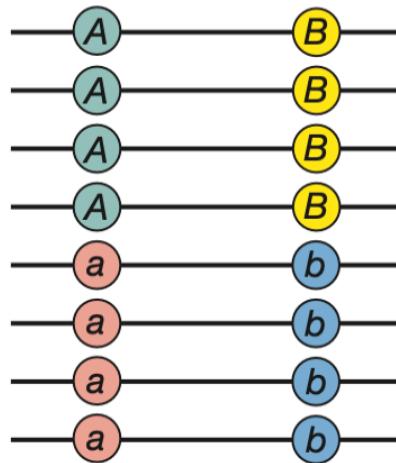
Linkage disequilibrium is the nonrandom association between two loci

(a) Linkage equilibrium



$$\begin{array}{ll} p_A = 0.5 & P_{AB} = 0.25 \\ p_a = 0.5 & P_{Ab} = 0.25 \\ p_B = 0.5 & P_{aB} = 0.25 \\ p_b = 0.5 & P_{ab} = 0.25 \end{array}$$

(b) Linkage disequilibrium



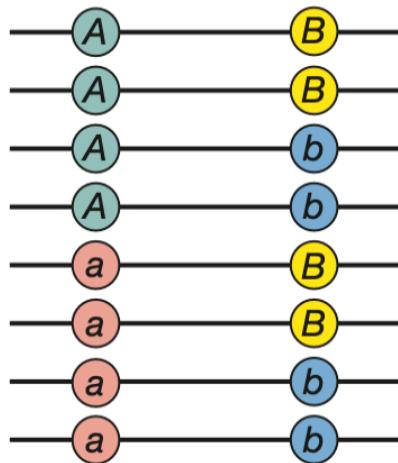
$$\begin{array}{ll} p_A = 0.5 & P_{AB} = 0.5 \\ p_a = 0.5 & P_{Ab} = 0.0 \\ p_B = 0.5 & P_{aB} = 0.0 \\ p_b = 0.5 & P_{ab} = 0.5 \end{array}$$

- Define LD as the difference (D) between the observed and expected frequencies

# LINKAGE DISEQUILIBRIUM

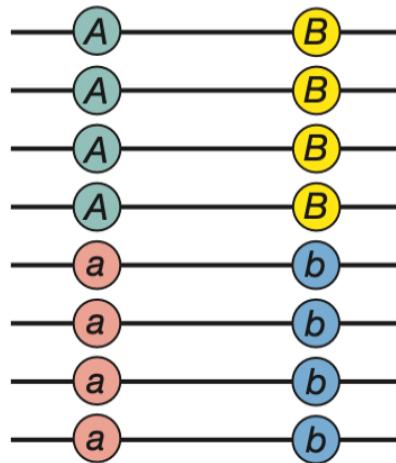
Linkage disequilibrium is the nonrandom association between two loci

(a) Linkage equilibrium



$$\begin{array}{ll} p_A = 0.5 & P_{AB} = 0.25 \\ p_a = 0.5 & P_{Ab} = 0.25 \\ p_B = 0.5 & P_{aB} = 0.25 \\ p_b = 0.5 & P_{ab} = 0.25 \end{array}$$

(b) Linkage disequilibrium



$$\begin{array}{ll} p_A = 0.5 & P_{AB} = 0.5 \\ p_a = 0.5 & P_{Ab} = 0.0 \\ p_B = 0.5 & P_{aB} = 0.0 \\ p_b = 0.5 & P_{ab} = 0.5 \end{array}$$

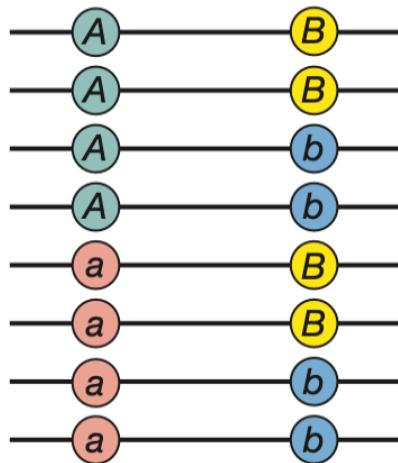
- Define LD as the difference ( $D$ ) between the observed and expected frequencies

- $$D = P_{AB} - p_A p_B$$

# LINKAGE DISEQUILIBRIUM

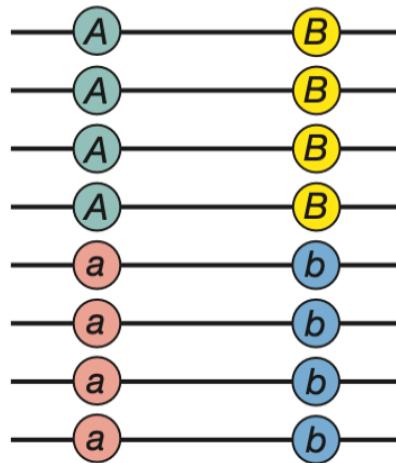
Linkage disequilibrium is the nonrandom association between two loci

(a) Linkage equilibrium



$$\begin{array}{ll} p_A = 0.5 & P_{AB} = 0.25 \\ p_a = 0.5 & P_{Ab} = 0.25 \\ p_B = 0.5 & P_{aB} = 0.25 \\ p_b = 0.5 & P_{ab} = 0.25 \end{array}$$

(b) Linkage disequilibrium



$$\begin{array}{ll} p_A = 0.5 & P_{AB} = 0.5 \\ p_a = 0.5 & P_{Ab} = 0.0 \\ p_B = 0.5 & P_{aB} = 0.0 \\ p_b = 0.5 & P_{ab} = 0.5 \end{array}$$

- Define LD as the difference (D) between the observed and expected frequencies

$$D = P_{AB} - p_A p_B$$

- Exercise:** calculate D for the two scenarios on the left

# LINKAGE DISEQUILIBRIUM

Nucleotide variation at the *G6PD* gene in humans

Individual	Origin	Allele	SNP																		Haplotype
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
1	Southern African	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
2	Central African	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
3	Central African	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
4	African American	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
5	African American	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
6	Central African	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
7	Central African	A+	G	.	G	.	.	.	.	.	.	T	.	.	.	C	.	G	.	2	
8	Central African	A+	G	.	G	.	.	.	.	.	.	T	.	.	.	C	.	G	.	2	
9	Central African	B	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	G	.	3	
10	Southern African	B	.	.	.	.	.	.	.	.	.	.	.	.	A	.	C	T	G	.	4
11	Southern African	B	.	.	.	.	.	.	.	.	.	.	.	.	A	.	C	T	G	.	4
12	Southern African	B	.	.	.	.	T	.	.	.	.	.	.	.	.	C	T	G	.	5	
13	Southern African	B	.	.	.	.	.	.	.	.	.	.	.	.	.	C	T	G	.	6	
14	Southern African	B	.	.	.	.	.	.	.	.	.	T	.	A	.	C	.	G	.	7	
15	Central African	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
16	European	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
17	European	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
18	European	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
19	Southwest Asian	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
20	East Asian	B	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	G	.	3	
21	Native American	B	.	.	.	.	A	T	.	.	.	.	.	.	.	C	.	G	.	9	
22	Southern African	B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	
23	Native American	B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	
24	Native American	B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	
25	Native American	B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	

- How LD arises

# LINKAGE DISEQUILIBRIUM

Nucleotide variation at the *G6PD* gene in humans

Individual	Origin	Allele	SNP																		Haplotype
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
1	Southern African	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
2	Central African	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
3	Central African	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
4	African American	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
5	African American	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
6	Central African	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
7	Central African	A+	G	.	G	.	.	.	.	.	T	.	.	.	C	.	G	.	2		
8	Central African	A+	G	.	G	.	.	.	.	.	T	.	.	.	C	.	G	.	2		
9	Central African	B	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	G	.	3	
10	Southern African	B	.	.	.	.	.	.	.	.	.	.	.	.	A	.	C	T	G	.	4
11	Southern African	B	.	.	.	.	.	.	.	.	.	.	.	.	A	.	C	T	G	.	4
12	Southern African	B	.	.	.	.	T	.	.	.	.	.	.	.	.	C	T	G	.	5	
13	Southern African	B	.	.	.	.	.	.	.	.	.	.	.	.	.	C	T	G	.	6	
14	Southern African	B	.	.	.	.	.	.	.	.	.	T	.	A	.	C	.	G	.	7	
15	Central African	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
16	European	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
17	European	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
18	European	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
19	Southwest Asian	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
20	East Asian	B	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	G	.	3	
21	Native American	B	.	.	.	A	T	.	.	.	.	.	.	.	.	C	.	G	.	9	
22	Southern African	B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	
23	Native American	B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	
24	Native American	B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	
25	Native American	B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	

- How LD arises

- consider a population with only AB and Ab. what happens when a new allele "a" arises at A locus, on a chromosome that already possess b. Is LD zero? or is it maximum?

# LINKAGE DISEQUILIBRIUM

Nucleotide variation at the *G6PD* gene in humans

Individual	Origin	Allele	SNP																		Haplotype
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
1	Southern African	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
2	Central African	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
3	Central African	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
4	African American	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
5	African American	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
6	Central African	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
7	Central African	A+	G	.	G	.	.	.	.	T	.	.	.	.	C	.	G	.	2		
8	Central African	A+	G	.	G	.	.	.	.	.	T	.	.	.	C	.	G	.	2		
9	Central African	B	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	G	.	3	
10	Southern African	B	.	.	.	.	.	.	.	.	.	.	.	.	A	.	C	T	G	.	4
11	Southern African	B	.	.	.	.	.	.	.	.	.	.	.	.	A	.	C	T	G	.	4
12	Southern African	B	.	.	.	.	T	.	.	.	.	.	.	.	.	C	T	G	.	5	
13	Southern African	B	.	.	.	.	.	.	.	.	.	.	.	.	.	C	T	G	.	6	
14	Southern African	B	.	.	.	.	.	.	.	.	.	T	.	A	.	C	.	G	.	7	
15	Central African	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
16	European	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
17	European	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
18	European	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
19	Southwest Asian	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
20	East Asian	B	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	G	.	3	
21	Native American	B	.	.	.	A	T	.	.	.	.	.	.	.	.	C	.	G	.	9	
22	Southern African	B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	
23	Native American	B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	
24	Native American	B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	
25	Native American	B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	

- How LD arises

- consider a population with only AB and Ab. what happens when a new allele "a" arises at A locus, on a chromosome that already possess b. Is LD zero? or is it maximum?
- other than mutation, can migration cause LD? how?

# LINKAGE DISEQUILIBRIUM

Nucleotide variation at the *G6PD* gene in humans

Individual	Origin	Allele	SNP																		Haplotype
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
1	Southern African	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
2	Central African	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
3	Central African	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
4	African American	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
5	African American	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
6	Central African	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
7	Central African	A+	G	.	G	.	.	.	.	T	.	.	.	.	C	.	G	.	2		
8	Central African	A+	G	.	G	.	.	.	.	.	T	.	.	.	C	.	G	.	2		
9	Central African	B	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	G	.	3	
10	Southern African	B	.	.	.	.	.	.	.	.	.	.	.	.	A	.	C	T	G	.	4
11	Southern African	B	.	.	.	.	.	.	.	.	.	.	.	.	A	.	C	T	G	.	4
12	Southern African	B	.	.	.	.	.	.	T	.	.	.	.	.	.	C	T	G	.	5	
13	Southern African	B	.	.	.	.	.	.	.	.	.	.	.	.	.	C	T	G	.	6	
14	Southern African	B	.	.	.	.	.	.	.	.	.	T	.	A	.	C	.	G	.	7	
15	Central African	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
16	European	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
17	European	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
18	European	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
19	Southwest Asian	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
20	East Asian	B	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	G	.	3	
21	Native American	B	.	.	.	.	A	T	.	.	.	.	.	.	.	C	.	G	.	9	
22	Southern African	B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	
23	Native American	B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	
24	Native American	B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	
25	Native American	B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	

- How LD arises

- consider a population with only AB and Ab. what happens when a new allele "a" arises at A locus, on a chromosome that already possess b. Is LD zero? or is it maximum?
- other than mutation, can migration cause LD? how?

- How LD decays

# LINKAGE DISEQUILIBRIUM

Nucleotide variation at the *G6PD* gene in humans

Individual	Origin	Allele	SNP																		Haplotype
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
1	Southern African	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
2	Central African	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
3	Central African	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
4	African American	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
5	African American	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
6	Central African	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
7	Central African	A+	G	.	G	.	.	.	.	T	.	.	.	.	C	.	G	.	2		
8	Central African	A+	G	.	G	.	.	.	.	T	.	.	.	.	C	.	G	.	2		
9	Central African	B	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	G	.	3	
10	Southern African	B	.	.	.	.	.	.	.	.	.	.	.	.	A	.	C	T	G	.	4
11	Southern African	B	.	.	.	.	.	.	.	.	.	.	.	.	A	.	C	T	G	.	4
12	Southern African	B	.	.	.	.	.	.	T	.	.	.	.	.	.	C	T	G	.	5	
13	Southern African	B	.	.	.	.	.	.	.	.	.	.	.	.	.	C	T	G	.	6	
14	Southern African	B	.	.	.	.	.	.	.	.	.	T	.	A	.	C	.	G	.	7	
15	Central African	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
16	European	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
17	European	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
18	European	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
19	Southwest Asian	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
20	East Asian	B	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	G	.	3	
21	Native American	B	.	.	.	.	A	T	.	.	.	.	.	.	.	C	.	G	.	9	
22	Southern African	B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	
23	Native American	B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	
24	Native American	B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	
25	Native American	B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	

- How LD arises

- consider a population with only AB and Ab. what happens when a new allele "a" arises at A locus, on a chromosome that already possess b. Is LD zero? or is it maximum?
- other than mutation, can migration cause LD? how?

- How LD decays

- decays proportional to time and recombination fraction

# LINKAGE DISEQUILIBRIUM

Nucleotide variation at the *G6PD* gene in humans

Individual	Origin	Allele	SNP																		Haplotype
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
1	Southern African	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
2	Central African	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
3	Central African	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
4	African American	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
5	African American	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
6	Central African	A-	G	A	G	.	G	.	.	T	.	T	.	.	.	C	.	G	.	1	
7	Central African	A+	G	.	G	.	.	.	.	T	.	.	.	.	C	.	G	.	2		
8	Central African	A+	G	.	G	.	.	.	.	T	.	.	.	.	C	.	G	.	2		
9	Central African	B	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	G	.	3	
10	Southern African	B	.	.	.	.	.	.	.	.	.	.	.	.	A	.	C	T	G	.	4
11	Southern African	B	.	.	.	.	.	.	.	.	.	.	.	.	A	.	C	T	G	.	4
12	Southern African	B	.	.	.	.	.	.	T	.	.	.	.	.	.	C	T	G	.	5	
13	Southern African	B	.	.	.	.	.	.	.	.	.	.	.	.	.	C	T	G	.	6	
14	Southern African	B	.	.	.	.	.	.	.	.	.	T	.	A	.	C	.	G	.	7	
15	Central African	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
16	European	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
17	European	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
18	European	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
19	Southwest Asian	B	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	G	.	8
20	East Asian	B	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	G	.	3	
21	Native American	B	.	.	.	.	A	T	.	.	.	.	.	.	.	C	.	G	.	9	
22	Southern African	B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	
23	Native American	B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	
24	Native American	B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	
25	Native American	B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	

- How LD arises

- consider a population with only AB and Ab. what happens when a new allele "a" arises at A locus, on a chromosome that already possess b. Is LD zero? or is it maximum?
- other than mutation, can migration cause LD? how?

- How LD decays

- decays proportional to time and recombination fraction
- can be used to estimate the age of mutation, how?

# GENETIC DRIFT AND POPULATION SIZE

- When calculating the expected genotype frequency from allele frequencies under Hardy-Weinberg Equilibrium, we made a sneaky assumption - the population size is assumed to be so large that we can "sample with replacement".

# GENETIC DRIFT AND POPULATION SIZE

- When calculating the expected genotype frequency from allele frequencies under Hardy-Weinberg Equilibrium, we made a sneaky assumption - the population size is assumed to be so large that we can "sample with replacement".
- Real populations have finite population sizes. Sampling can lead to fluctuations in the allele frequency from generation to generation

# GENETIC DRIFT AND POPULATION SIZE

- When calculating the expected genotype frequency from allele frequencies under Hardy-Weinberg Equilibrium, we made a sneaky assumption - the population size is assumed to be so large that we can "sample with replacement".
- Real populations have finite population sizes. Sampling can lead to fluctuations in the allele frequency from generation to generation
  - let's consider an extreme case: a population consisting of a single heterozygous ( $A/a$ ) individual ( $N=1$ ) at generation  $t_0$ . What's the allele frequency at  $t_0$ ?

# GENETIC DRIFT AND POPULATION SIZE

- When calculating the expected genotype frequency from allele frequencies under Hardy-Weinberg Equilibrium, we made a sneaky assumption - the population size is assumed to be so large that we can "sample with replacement".
- Real populations have finite population sizes. Sampling can lead to fluctuations in the allele frequency from generation to generation
  - let's consider an extreme case: a population consisting of a single heterozygous ( $A/a$ ) individual ( $N=1$ ) at generation  $t_0$ . What's the allele frequency at  $t_0$ ?
  - suppose this species can self-fertilize, and that the population size remains at one in the next generation ( $t_1$ ). what's the probability that the allele frequency will change ("drift") to something other than its original value?

# GENETIC DRIFT AND POPULATION SIZE

- When calculating the expected genotype frequency from allele frequencies under Hardy-Weinberg Equilibrium, we made a sneaky assumption - the population size is assumed to be so large that we can "sample with replacement".
- Real populations have finite population sizes. Sampling can lead to fluctuations in the allele frequency from generation to generation
  - let's consider an extreme case: a population consisting of a single heterozygous ( $A/a$ ) individual ( $N=1$ ) at generation  $t_0$ . What's the allele frequency at  $t_0$ ?
  - suppose this species can self-fertilize, and that the population size remains at one in the next generation ( $t_1$ ). what's the probability that the allele frequency will change ("drift") to something other than its original value?
  - what happens when we increase  $N$  to 2?

# GENETIC DRIFT AND POPULATION SIZE

## Drift

any change in allele frequencies due to sampling error, not just loss or fixation of an allele

### Question

In a population with 500 individuals ( $N=500$ ) and two alleles at a frequency of  $p = q = 0.5$ , if the next generation has 501 copies of A and 499 copies of a, **has there been genetic drift?**

# GENETIC DRIFT AND POPULATION SIZE

- Drift is caused by sampling variation In a **finite population**.

# GENETIC DRIFT AND POPULATION SIZE

- Drift is caused by sampling variation In a **finite population**.
- When drift is operating, one can **calculate the probabilities of different outcomes**, but one **cannot accurately predict the outcome that will occur**

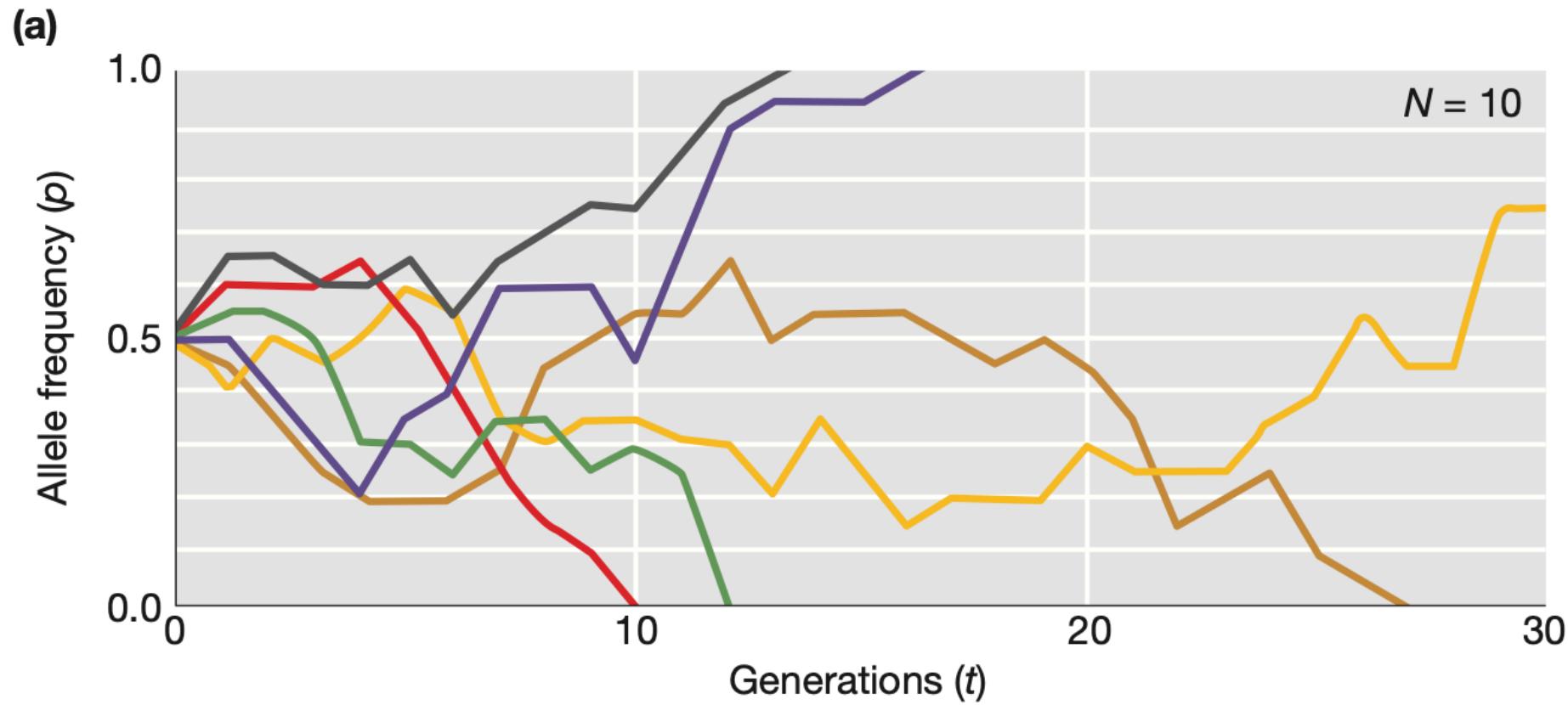
# GENETIC DRIFT AND POPULATION SIZE

- Drift is caused by sampling variation In a **finite population**.
- When drift is operating, one can **calculate the probabilities of different outcomes**, but one **cannot accurately predict the outcome that will occur**
- At a locus, drift continues to affect the allele frequency from generation to generation, until one allele is eventually fixed (and all others are lost), hence no more variation at that locus.

# GENETIC DRIFT AND POPULATION SIZE

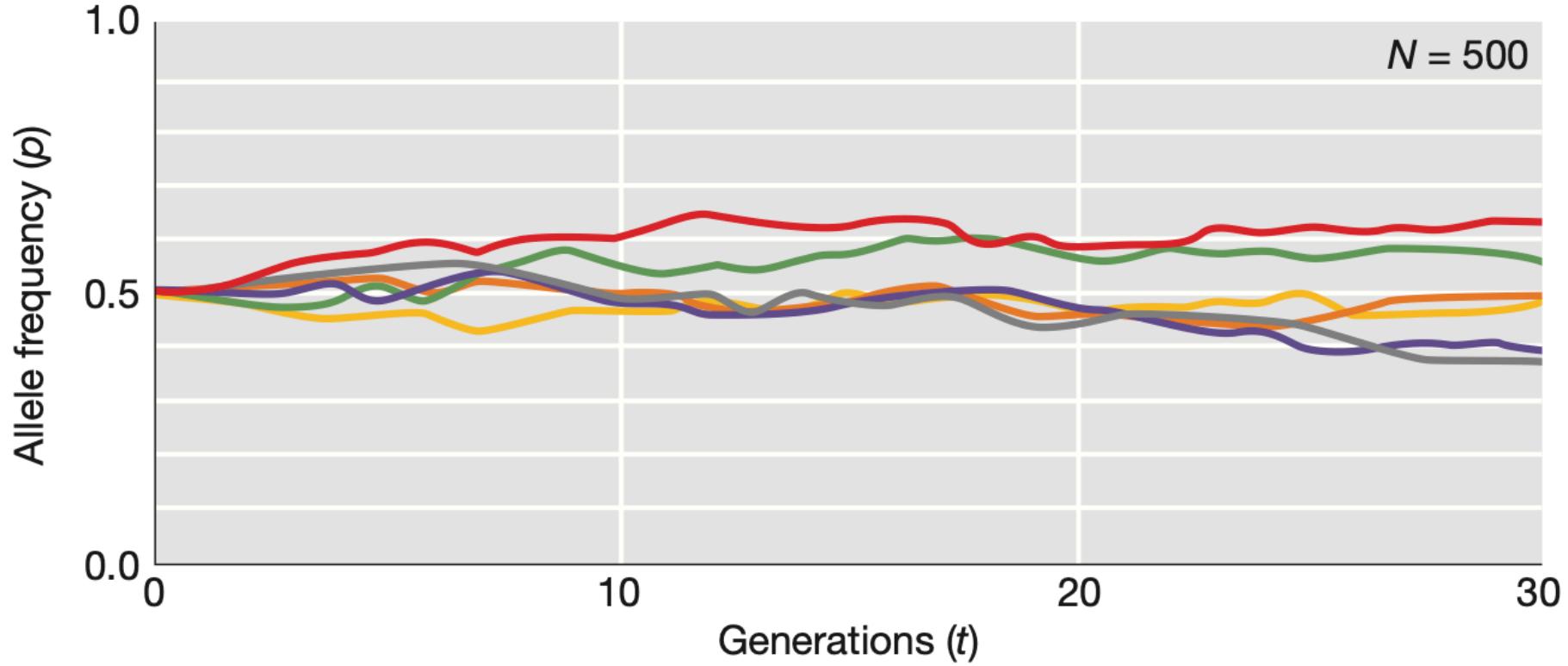
- Drift is caused by sampling variation in a **finite population**.
- When drift is operating, one can **calculate the probabilities of different outcomes**, but one **cannot accurately predict the outcome that will occur**
- At a locus, drift continues to affect the allele frequency from generation to generation, until one allele is eventually fixed (and all others are lost), hence no more variation at that locus.
- Drift **doesn't proceed in a specific direction**.

# GENETIC DRIFT AND POPULATION SIZE

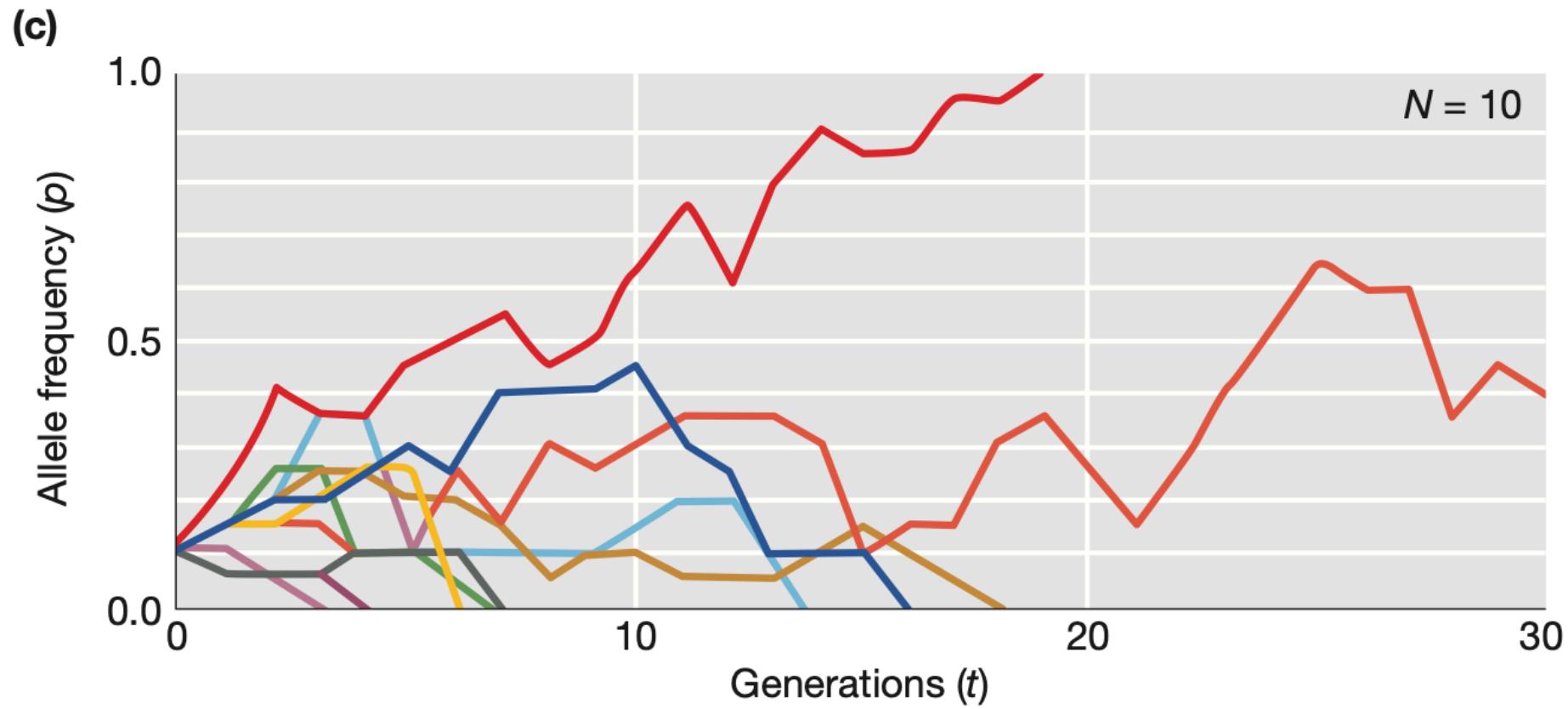


# GENETIC DRIFT AND POPULATION SIZE

(b)



# GENETIC DRIFT AND POPULATION SIZE



# GENETIC DRIFT AND POPULATION SIZE

- $\text{Prob}(\text{an allele goes to fixation}) \sim \text{frequency in the present generation}$

# GENETIC DRIFT AND POPULATION SIZE

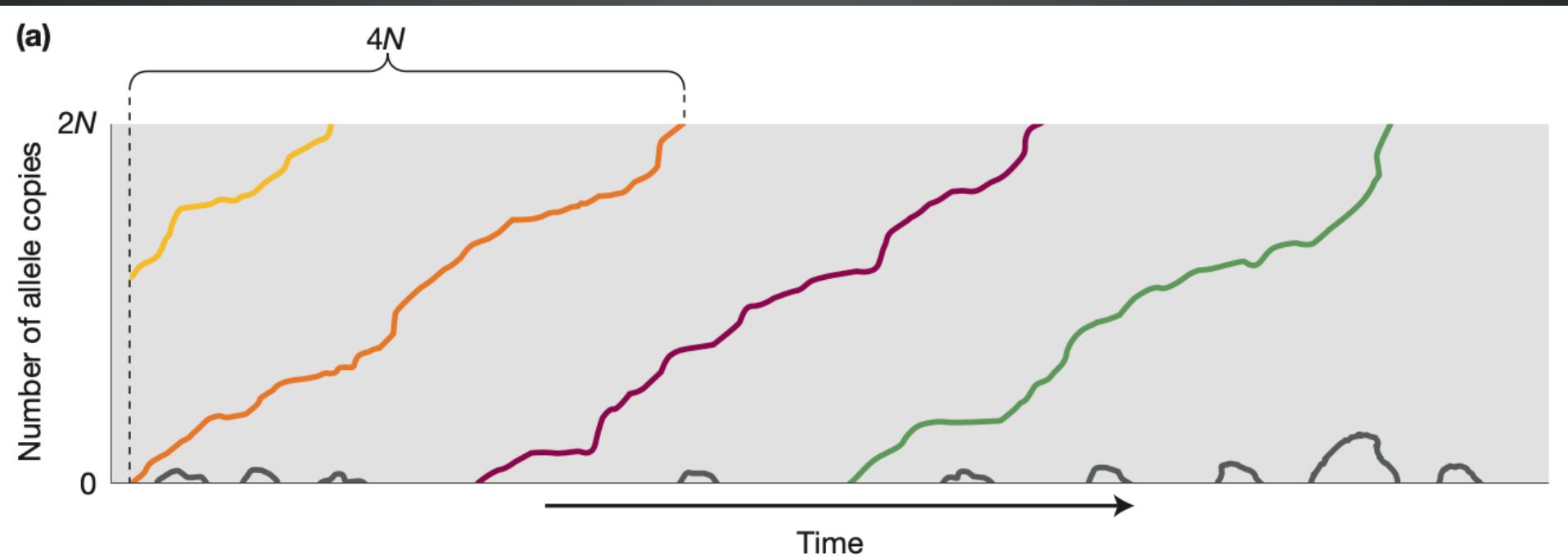
- Prob(an allele goes to fixation) ~ frequency in the present generation
- The initial frequency of a new allele in a diploid population of size N is  $\frac{1}{2N}$

# GENETIC DRIFT AND POPULATION SIZE

- Prob(an allele goes to fixation) ~ frequency in the present generation
- The initial frequency of a new allele in a diploid population of size  $N$  is  $\frac{1}{2N}$
- If  $N$  is even modestly large, e.g.,  $10^4$ , the probability of its fixation is  $5 \times 10^{-5}$ , while the probability of it eventually being lost is close to 1.0!

# GENETIC DRIFT AND POPULATION SIZE

- $\text{Prob}(\text{an allele goes to fixation}) \sim \text{frequency in the present generation}$
- The initial frequency of a new allele in a diploid population of size  $N$  is  $\frac{1}{2N}$
- If  $N$  is even modestly large, e.g.,  $10^4$ , the probability of its fixation is  $5 \times 10^{-5}$ , while the probability of it eventually being lost is close to 1.0!



# CONSEQUENCE OF GENETIC DRIFT

- 💡 Slightly deleterious alleles may be driven to fixation, while beneficial mutations are subject to loss

# CONSEQUENCE OF GENETIC DRIFT



Slightly deleterious alleles may be driven to fixation, while beneficial mutations are subject to loss

1. Strongly deleterious mutations are quickly lost. But slightly deleterious ones can drift to fixation - and when it does, it does so **quickly** (counterintuitive?)

# CONSEQUENCE OF GENETIC DRIFT



Slightly deleterious alleles may be driven to fixation, while beneficial mutations are subject to loss

1. Strongly deleterious mutations are quickly lost. But slightly deleterious ones can drift to fixation - and when it does, it does so **quickly** (counterintuitive?)
  - do you know that the fate of a mutation is not only determined by its functional impact (beneficial or deleterious), but also population size?

# CONSEQUENCE OF GENETIC DRIFT



Slightly deleterious alleles may be driven to fixation, while beneficial mutations are subject to loss

1. Strongly deleterious mutations are quickly lost. But slightly deleterious ones can drift to fixation - and when it does, it does so **quickly** (counterintuitive?)
  - do you know that the fate of a mutation is not only determined by its functional impact (beneficial or deleterious), but also population size?
2. When a new beneficial mutation arise, there is an appreciable chance that it will be lost in the first few generations!

# CONSEQUENCE OF GENETIC DRIFT



Slightly deleterious alleles may be driven to fixation, while beneficial mutations are subject to loss

1. Strongly deleterious mutations are quickly lost. But slightly deleterious ones can drift to fixation - and when it does, it does so **quickly** (counterintuitive?)
  - do you know that the fate of a mutation is not only determined by its functional impact (beneficial or deleterious), but also population size?
2. When a new beneficial mutation arise, there is an appreciable chance that it will be lost in the first few generations!
  - the individual carrying the new mutation may not have the opportunity to reproduce.

# CONSEQUENCE OF GENETIC DRIFT



Slightly deleterious alleles may be driven to fixation, while beneficial mutations are subject to loss

1. Strongly deleterious mutations are quickly lost. But slightly deleterious ones can drift to fixation - and when it does, it does so **quickly** (counterintuitive?)
  - do you know that the fate of a mutation is not only determined by its functional impact (beneficial or deleterious), but also population size?
2. When a new beneficial mutation arise, there is an appreciable chance that it will be lost in the first few generations!
  - the individual carrying the new mutation may not have the opportunity to reproduce.
  - or it may not pass on the beneficial mutation (since by definition, a new mutation must be a single copy in a diploid individual)

# MOLECULAR CLOCK: IMPLICATION OF GENETIC DRIFT

- Because of genetic drift, neutral mutations can become fixed, which we call "substitutions".

# MOLECULAR CLOCK: IMPLICATION OF GENETIC DRIFT

- Because of genetic drift, neutral mutations can become fixed, which we call "substitutions".
- note the difference between "mutation" and "substitution"

# MOLECULAR CLOCK: IMPLICATION OF GENETIC DRIFT

- Because of genetic drift, neutral mutations can become fixed, which we call "substitutions".
- note the difference between "mutation" and "substitution"
- the number of mutations that are expected to "appear" in a population in one generation is proportional to the number of copies in the gene pool, i.e.,  $2N\mu$ , where  $\mu$  denotes the mutation rate per locus per generation.

# MOLECULAR CLOCK: IMPLICATION OF GENETIC DRIFT

- Because of genetic drift, neutral mutations can become fixed, which we call "substitutions".
- note the difference between "mutation" and "substitution"
- the number of mutations that are expected to "appear" in a population in one generation is proportional to the number of copies in the gene pool, i.e.,  $2N\mu$ , where  $\mu$  denotes the mutation rate per locus per generation.
- recall that the probability of fixation for a neutral mutation equals its initial frequency. for a new, neutral mutation, this is  $\frac{1}{2N}$

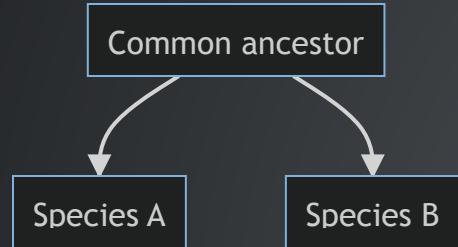
# MOLECULAR CLOCK: IMPLICATION OF GENETIC DRIFT

- Because of genetic drift, neutral mutations can become fixed, which we call "substitutions".
- note the difference between "mutation" and "substitution"
- the number of mutations that are expected to "appear" in a population in one generation is proportional to the number of copies in the gene pool, i.e.,  $2N\mu$ , where  $\mu$  denotes the mutation rate per locus per generation.
- recall that the probability of fixation for a neutral mutation equals its initial frequency. for a new, neutral mutation, this is  $\frac{1}{2N}$
- therefore, the number of **fixed neutral mutations** per generation, which we will call "substitution rate", or  $k$ , is  $k = 2N\mu \times \frac{1}{2N} = \mu$

# MOLECULAR CLOCK: IMPLICATION OF GENETIC DRIFT

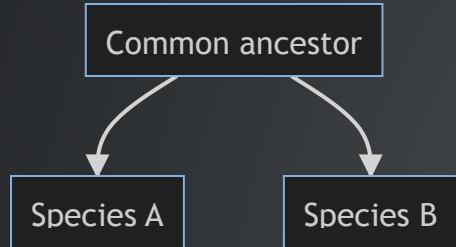
- Because of genetic drift, neutral mutations can become fixed, which we call "substitutions".
- note the difference between "mutation" and "substitution"
- the number of mutations that are expected to "appear" in a population in one generation is proportional to the number of copies in the gene pool, i.e.,  $2N\mu$ , where  $\mu$  denotes the mutation rate per locus per generation.
- recall that the probability of fixation for a neutral mutation equals its initial frequency. for a new, neutral mutation, this is  $\frac{1}{2N}$
- therefore, the number of **fixed neutral mutations** per generation, which we will call "substitution rate", or  $k$ , is  $k = 2N\mu \times \frac{1}{2N} = \mu$
- that is, the substitution rate for neutral mutations equal the rate at which they appear. this is only true for neutral mutations!

# MOLECULAR CLOCK: IMPLICATION OF GENETIC DRIFT



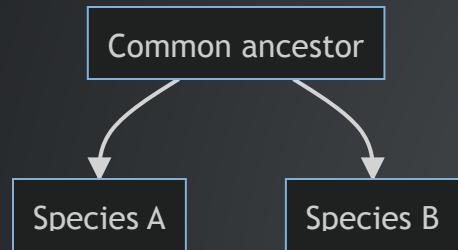
- $d$ : number of **neutral substitutions** in a gene in both species since their divergence.

# MOLECULAR CLOCK: IMPLICATION OF GENETIC DRIFT



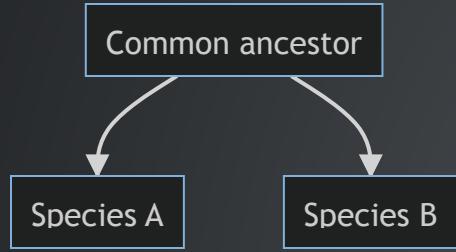
- $d$ : number of **neutral substitutions** in a gene in both species since their divergence.
- $d = 2tk$ , where  $t$  is the divergence time.

# MOLECULAR CLOCK: IMPLICATION OF GENETIC DRIFT



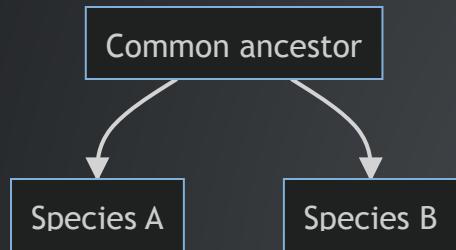
- $d$ : number of **neutral substitutions** in a gene in both species since their divergence.
- $d = 2tk$ , where  $t$  is the divergence time.
- $t = \frac{d}{2k}$ , this provides a way to estimate divergence time based on the number of substitutions and estimate of substitution rate.

# MOLECULAR CLOCK: IMPLICATION OF GENETIC DRIFT



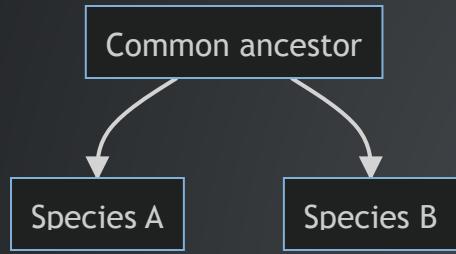
- $d$ : number of **neutral substitutions** in a gene in both species since their divergence.
- $d = 2tk$ , where  $t$  is the divergence time.
- $t = \frac{d}{2k}$ , this provides a way to estimate divergence time based on the number of substitutions and estimate of substitution rate.
- the estimated divergence time is in the unit of generations.

# MOLECULAR CLOCK: IMPLICATION OF GENETIC DRIFT



- $d$ : number of **neutral substitutions** in a gene in both species since their divergence.
- $d = 2tk$ , where  $t$  is the divergence time.
- $t = \frac{d}{2k}$ , this provides a way to estimate divergence time based on the number of substitutions and estimate of substitution rate.
- the estimated divergence time is in the unit of generations.
- how to determine which substitutions are neutral?

# MOLECULAR CLOCK: IMPLICATION OF GENETIC DRIFT



- $d$ : number of **neutral substitutions** in a gene in both species since their divergence.
- $d = 2tk$ , where  $t$  is the divergence time.
- $t = \frac{d}{2k}$ , this provides a way to estimate divergence time based on the number of substitutions and estimate of substitution rate.
- the estimated divergence time is in the unit of generations.
- how to determine which substitutions are neutral?
- can we observe all neutral substitutions?

# MOLECULAR CLOCK: IMPLICATION OF GENETIC DRIFT



- $d$ : number of **neutral substitutions** in a gene in both species since their divergence.
- $d = 2tk$ , where  $t$  is the divergence time.
- $t = \frac{d}{2k}$ , this provides a way to estimate divergence time based on the number of substitutions and estimate of substitution rate.
- the estimated divergence time is in the unit of generations.
- how to determine which substitutions are neutral?
- can we observe all neutral substitutions?
- how do we know mutation rate?

# GENETIC DRIFT AND POPULATION SIZE

## Key concept

- Population size is a key factor affecting genetic variation in populations.
- 
- 
- 
- 
-

# GENETIC DRIFT AND POPULATION SIZE

## Key concept

- Population size is a key factor affecting genetic variation in populations.
- Genetic drift is a stronger force in small populations than in large ones.
- 
- 
-

# GENETIC DRIFT AND POPULATION SIZE

## Key concept

- Population size is a key factor affecting genetic variation in populations.
- Genetic drift is a stronger force in small populations than in large ones.
- The probability that an allele will become fixed in (or lost from) a population by drift is a function of its frequency in the population and population size.
- 
-

# GENETIC DRIFT AND POPULATION SIZE

## Key concept

- Population size is a key factor affecting genetic variation in populations.
- Genetic drift is a stronger force in small populations than in large ones.
- The probability that an allele will become fixed in (or lost from) a population by drift is a function of its frequency in the population and population size.
- Most new mutations are lost from populations by drift.
-

# GENETIC DRIFT AND POPULATION SIZE

## Key concept

- Population size is a key factor affecting genetic variation in populations.
- Genetic drift is a stronger force in small populations than in large ones.
- The probability that an allele will become fixed in (or lost from) a population by drift is a function of its frequency in the population and population size.
- Most new mutations are lost from populations by drift.
- The constant rate of substitution for neutral mutations can be used to estimate divergence time.

# SELECTION

## ❶ Natural selection (*Darwinian*)

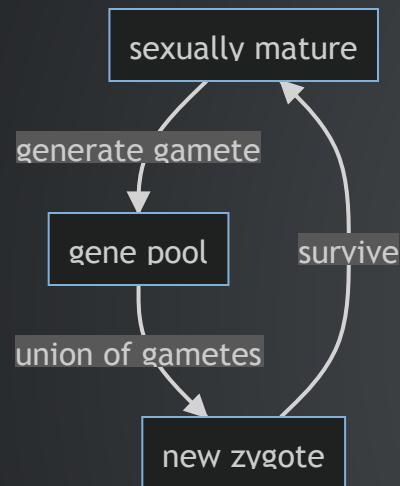
Populations change (*evolve*) over time as the environment (*nature*) favors (*selects*) features that enhance the ability to survive and reproduce.

# SELECTION

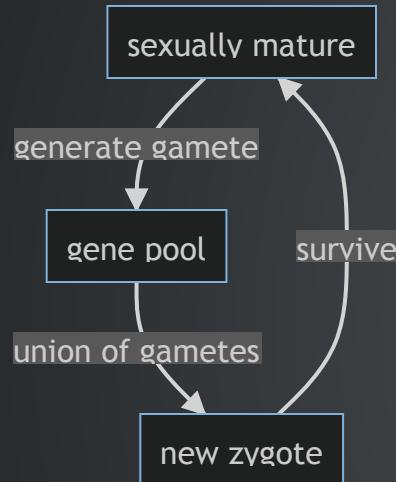
## Question

We often use the phrase "*survival of the fittest*" to describe Darwinian selection. If an individual is physically strong, resistant to disease, and lives a long life, does this individual have a higher fitness?

# SELECTION

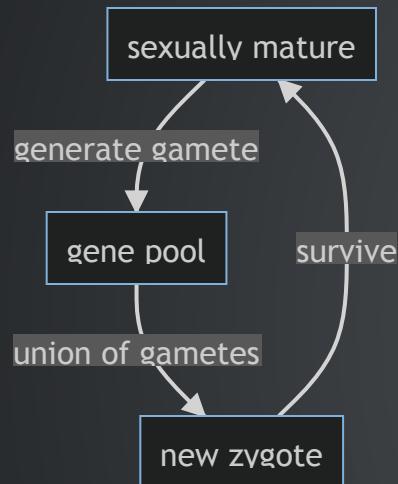


# SELECTION



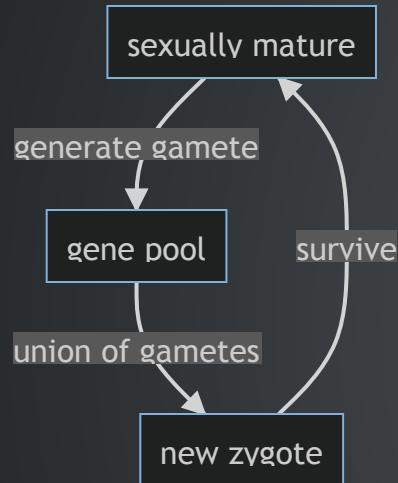
- **Natural selection** means individuals with certain features are more likely to *survive AND reproduce* than other individuals lacking these features.

# SELECTION



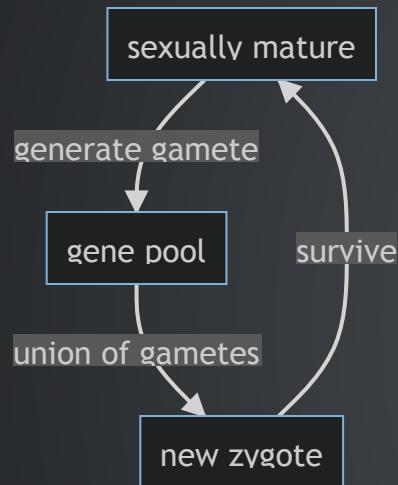
- **Natural selection** means individuals with certain features are more likely to *survive AND reproduce* than other individuals lacking these features.
- Mutation generates new variants; recombination creates new combinations (haplotypes), both of which affect the phenotype (feature).

# SELECTION



- **Natural selection** means individuals with certain features are more likely to *survive AND reproduce* than other individuals lacking these features.
- Mutation generates new variants; recombination creates new combinations (haplotypes), both of which affect the phenotype (feature).
- Individuals with higher **fitness** are more likely to survive and reproduce, i.e., their alleles will have a higher frequency in the gene pool in next round of reproduction  
-> changing the allele frequency.

# SELECTION



- **Natural selection** means individuals with certain features are more likely to *survive AND reproduce* than other individuals lacking these features.
- Mutation generates new variants; recombination creates new combinations (haplotypes), both of which affect the phenotype (feature).
- Individuals with higher **fitness** are more likely to survive and reproduce, i.e., their alleles will have a higher frequency in the gene pool in next round of reproduction -> changing the allele frequency.
- Where does selection operate in the left diagram?

# HOW NATURAL SELECTION ALTERS ALLELE FREQUENCY?

- Consider a locus A with three genotypes, A/A, A/a, and a/a in a population. The initial allele frequencies are  $p = 0.1$ ,  $q = 0.9$

# HOW NATURAL SELECTION ALTERS ALLELE FREQUENCY?

- Consider a locus A with three genotypes, A/A, A/a, and a/a in a population. The initial allele frequencies are  $p = 0.1$ ,  $q = 0.9$
- A is a favored dominant allele

Genotype	A/A	A/a	a/a
Average number of offspring (W)	10	10	5
Relative fitness (w)	1.0	1.0	0.5
Genotype frequency			

# HOW NATURAL SELECTION ALTERS ALLELE FREQUENCY?

- A is a favored dominant allele

	A/A	A/a	a/a
Average number of offspring (W)	10	10	5
Relative fitness (w)	1.0	1.0	0.5
Genotype frequency	0.01	0.18	0.81

# HOW NATURAL SELECTION ALTERS ALLELE FREQUENCY?

- A is a favored dominant allele

	A/A	A/a	a/a
Average number of offspring (W)	10	10	5
Relative fitness (w)	1.0	1.0	0.5
Genotype frequency	0.01	0.18	0.81

- The relative contribution of each genotype to the gene pool is determined by **the product of its fitness and frequency**

Genotype	A/A	A/a	a/a	Sum
Relative contribution	$1 \times 0.01$	$1 \times 0.18$	$0.5 \times 0.81$	
	= 0.01	= 0.18	= 0.405	0.595

# HOW NATURAL SELECTION ALTERS ALLELE FREQUENCY?

- The relative contribution of each genotype to the gene pool is determined by **the product of its fitness and frequency**

Genotype	A/A	A/a	a/a	Sum
Relative contribution	0.01	0.18	0.405	0.595
Genotype frequency	0.02	0.30	0.68	1.0

# HOW NATURAL SELECTION ALTERS ALLELE FREQUENCY?

- The relative contribution of each genotype to the gene pool is determined by **the product of its fitness and frequency**

Genotype	A/A	A/a	a/a	Sum
Relative contribution	0.01	0.18	0.405	0.595
Genotype frequency	0.02	0.30	0.68	1.0

- Now, let's use the genotype frequency to calculate the allele frequency, and then use that to calculate the expected genotype frequency in the next generation of zygotes under HWE

# HOW NATURAL SELECTION ALTERS ALLELE FREQUENCY?

- The relative contribution of each genotype to the gene pool is determined by **the product of its fitness and frequency**

Genotype	A/A	A/a	a/a	Sum
Relative contribution	0.01	0.18	0.405	0.595
Genotype frequency	0.02	0.30	0.68	1.0

- Now, let's use the genotype frequency to calculate the allele frequency, and then use that to calculate the expected genotype frequency in the next generation of zygotes under HWE
  - $p' = ?$  how much did  $p$  change over one round of selection?

# HOW NATURAL SELECTION ALTERS ALLELE FREQUENCY?

- The relative contribution of each genotype to the gene pool is determined by **the product of its fitness and frequency**

Genotype	A/A	A/a	a/a	Sum
Relative contribution	0.01	0.18	0.405	0.595
Genotype frequency	0.02	0.30	0.68	1.0

- Now, let's use the genotype frequency to calculate the allele frequency, and then use that to calculate the expected genotype frequency in the next generation of zygotes under HWE
  - $p' = ?$  how much did  $p$  change over one round of selection?
  - $P_{AA} = ?,$  etc.

# HOW NATURAL SELECTION ALTERS ALLELE FREQUENCY?

More generally

define  $\bar{w} = p^2w_{A/A} + 2pqw_{A/a} + q^2w_{a/a}$  is the mean relative fitness of the population

$$\text{then, } p' = \frac{p^2w_{A/A} + \frac{1}{2} \times 2pqw_{A/a}}{\bar{w}} = p \frac{pw_{A/A} + qw_{A/a}}{\bar{w}}$$

# HOW NATURAL SELECTION ALTERS ALLELE FREQUENCY?

More generally

define  $\bar{w} = p^2w_{A/A} + 2pqw_{A/a} + q^2w_{a/a}$  is the mean relative fitness of the population

$$\text{then, } p' = \frac{p^2w_{A/A} + \frac{1}{2} \times 2pqw_{A/a}}{\bar{w}} = p \frac{pw_{A/A} + qw_{A/a}}{\bar{w}}$$

we define ' $w_A = pw_{A/A} + qw_{A/a}$ ' as allelic fitness, or mean fitness of allele A

# HOW NATURAL SELECTION ALTERS ALLELE FREQUENCY?

More generally

define  $\bar{w} = p^2w_{A/A} + 2pqw_{A/a} + q^2w_{a/a}$  is the mean relative fitness of the population

$$\text{then, } p' = \frac{p^2w_{A/A} + \frac{1}{2} \times 2pqw_{A/a}}{\bar{w}} = p \frac{pw_{A/A} + qw_{A/a}}{\bar{w}}$$

we define ' $w_A = pw_{A/A} + qw_{A/a}$ ' as allelic fitness, or mean fitness of allele A

$$\Delta p = p' - p = p \frac{w_A}{\bar{w}} - p = \frac{p(w_A - \bar{w})}{\bar{w}}$$

# FORMS OF SELECTION

1. Directional selection

# FORMS OF SELECTION

## 1. Directional selection

- Positive selection

# FORMS OF SELECTION

## 1. Directional selection

- Positive selection
- Purifying (negative) selection

# FORMS OF SELECTION

## 1. Directional selection

- Positive selection
- Purifying (negative) selection

## 2. Balancing selection

# FORMS OF SELECTION

## 1. Directional selection

- Positive selection
- Purifying (negative) selection

## 2. Balancing selection

- if the heterozygous class has higher fitness than either homozygous ones, selection would favor the **maintenance of both alleles**.

# FORMS OF SELECTION

## 1. Directional selection

- Positive selection
- Purifying (negative) selection

## 2. Balancing selection

- if the heterozygous class has higher fitness than either homozygous ones, selection would favor the **maintenance of both alleles**.
- can you think of example scenarios for balancing selection?

# FORMS OF SELECTION

## 1. Directional selection

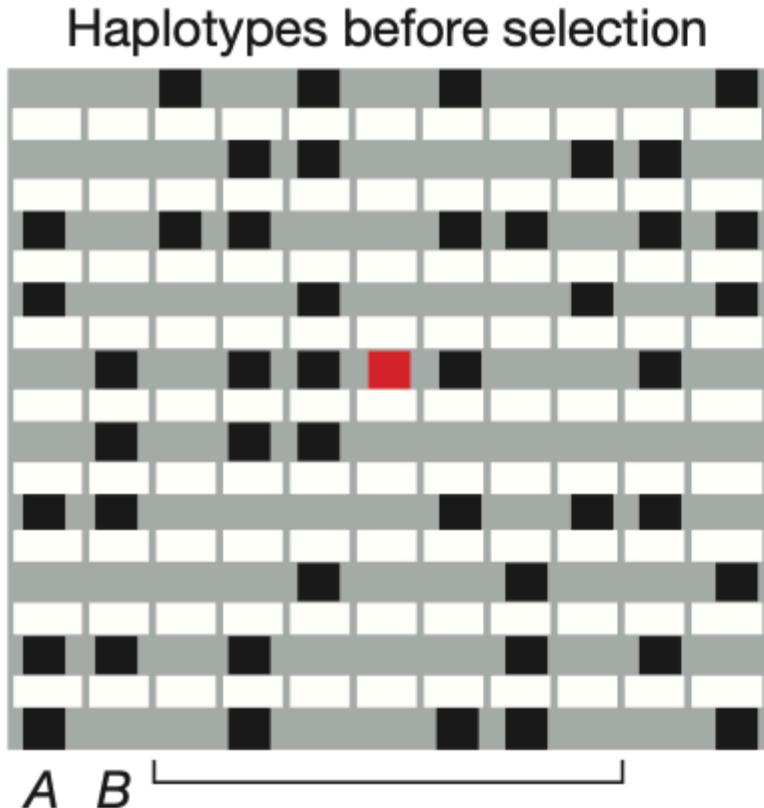
- Positive selection
- Purifying (negative) selection

## 2. Balancing selection

- if the heterozygous class has higher fitness than either homozygous ones, selection would favor the **maintenance of both alleles**.
- can you think of example scenarios for balancing selection?
- are there other types of selection that can **maintain genetic diversity** rather than **eliminating it**?

# SIGNATURES OF POSITIVE SELECTION

Positive selection leaves a distinct signature

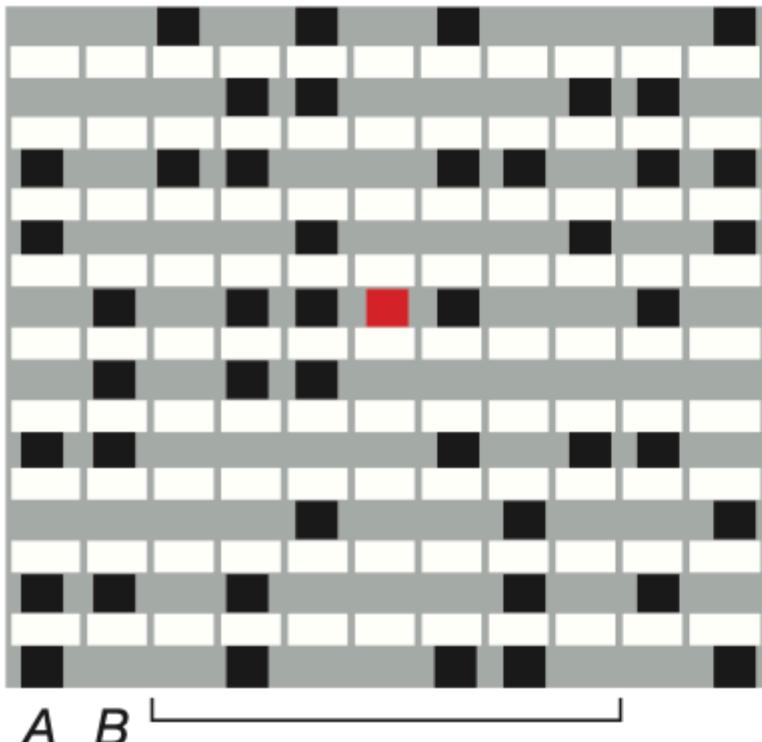


Haplotypes after selection

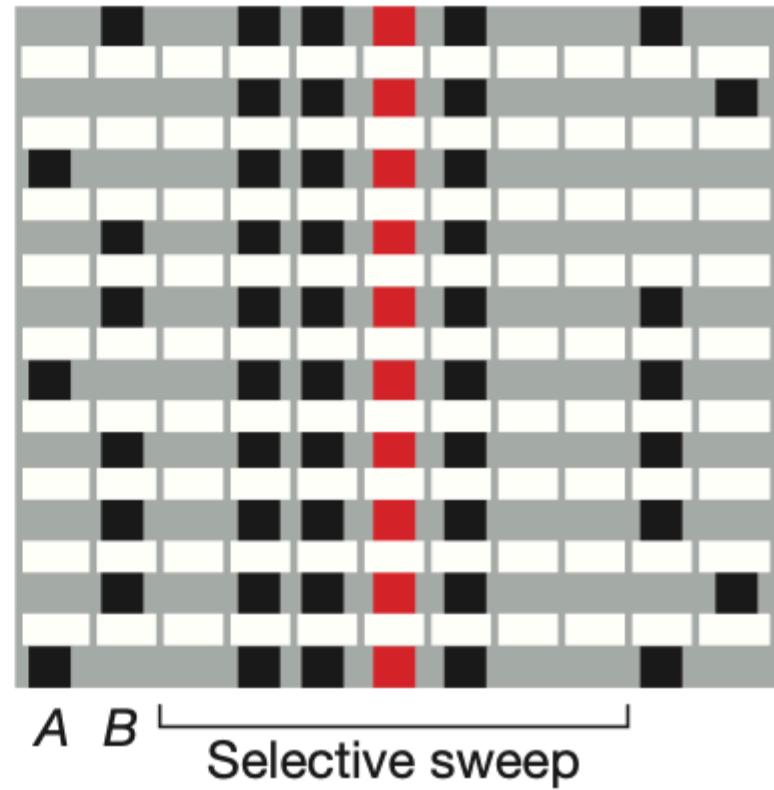
# SIGNATURES OF POSITIVE SELECTION

Positive selection leaves a distinct signature

Haplotypes before selection



Haplotypes after selection



# SIGNATURES OF POSITIVE SELECTION

- By driving the favored allele to fixation "very fast", positive selection causes a selective sweep that ends up

# SIGNATURES OF POSITIVE SELECTION

- By driving the favored allele to fixation "very fast", positive selection causes a selective sweep that ends up
  - also fix linked sites or drive the linked allele to high frequency

# SIGNATURES OF POSITIVE SELECTION

- By driving the favored allele to fixation "very fast", positive selection causes a selective sweep that ends up
  - also fix linked sites or drive the linked allele to high frequency
  - hence removing genetic variation from the population

# SIGNATURES OF POSITIVE SELECTION

- By driving the favored allele to fixation "very fast", positive selection causes a selective sweep that ends up
  - also fix linked sites or drive the linked allele to high frequency
  - hence removing genetic variation from the population
  - creates LD

# SIGNATURES OF POSITIVE SELECTION

- By driving the favored allele to fixation "very fast", positive selection causes a selective sweep that ends up
  - also fix linked sites or drive the linked allele to high frequency
  - hence removing genetic variation from the population
  - creates LD
- Note that this signature is **transient**. What does this mean, why?

# USE SELECTIVE SWEEP SIGNATURE TO INFER POSITIVE SELECTION

In Europe, a selective sweep caused a loss of all diversity at the *SLC24A5* locus

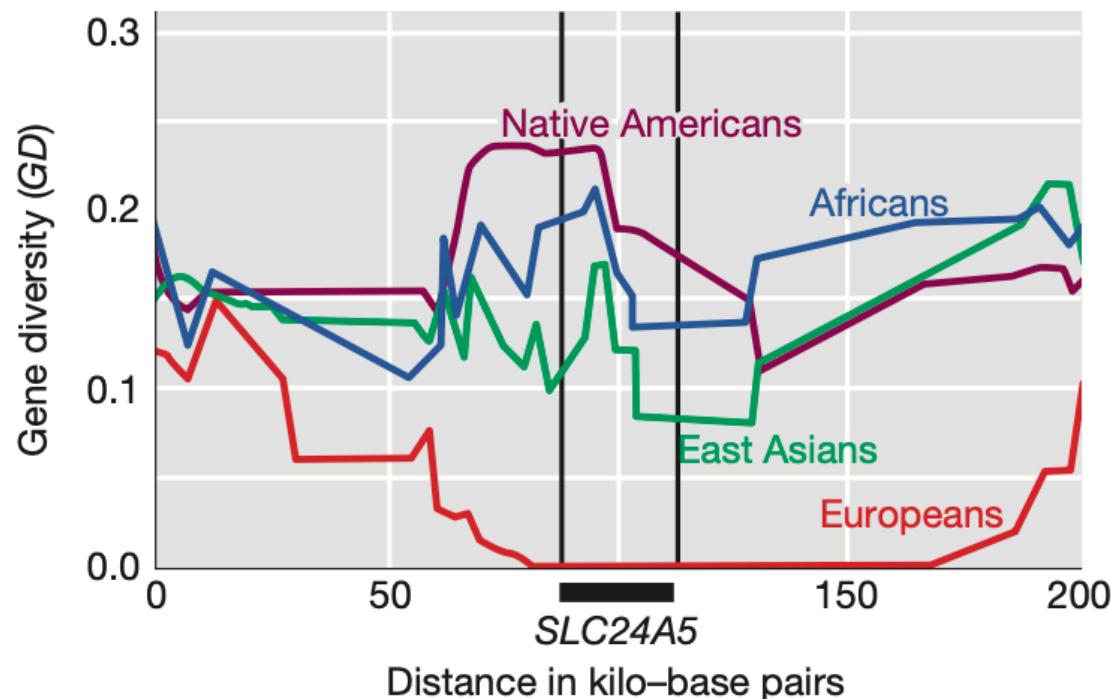


TABLE 8-6 Some Genes Showing Evidence for Natural Selection in Specific Human Populations

Gene	Presumed Trait	Population
<i>EDA2R</i> (ectodysplasin A2 receptor)	Male pattern baldness	Europeans
<i>EDAR</i> (ectodysplasin A receptor)	Hair morphology	East Asians
<i>FY<sup>null</sup></i> (Duffy antigen)	Resistance to malaria	Africans
<i>G6PD</i> (glucose-6-phosphate dehydrogenase)	Resistance to malaria	Africans
<i>Hb</i> (hemoglobin B)	Resistance to malaria	Africans
<i>KITLG</i> (KIT ligand)	Skin pigmentation	East Asians and Europeans
<i>LARGE</i> (glycosyltransferase)	Resistance to Lassa fever	Africans
<i>LCT</i> (lactase)	Lactase persistence; ability to digest milk sugar as an adult	Africans, Europeans
<i>LPR</i> (leptin receptor)	Processing of dietary fats	East Asians
<i>MC1R</i> (melanocortin receptor)	Hair and skin pigmentation	East Asians
<i>MHC</i> (major histocompatibility complex)	Infectious disease resistance	Multiple populations
<i>OCA2</i> (oculocutaneous albinism)	Skin pigmentation and eye color	Europeans
<i>PPARD</i> (peroxisome proliferator-activated receptor delta)	Processing of dietary fats	Europeans
<i>SI</i> (sucrase-isomaltase)	Sucrose metabolism	East Asians
<i>SLC24A5</i> (solute carrier family 24)	Skin pigmentation	Europeans and West Asians
<i>TYRP1</i> (tyrosinase-related protein 1)	Skin pigmentation	Europeans

**Source:** P. C. Sabeti et al., *Science* 312, 2006, 1614–1620; P. C. Sabeti et al., *Nature* 449, 2007, 913–919; B. F. Voight et al., *PLoS Biology* 4, 2006, 446–458; J. K. Pickrell et al., *Genome Research* 19, 2009, 826–837.

Balancing selection can lead to regions of unusually high genetic diversity

