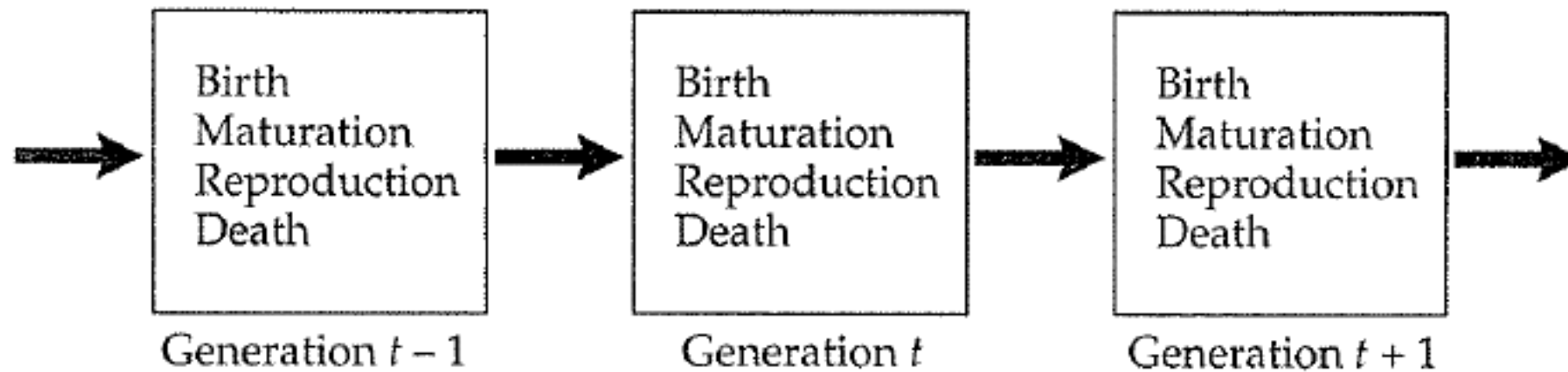# Organization of genetic variation

Non-overlapping generations

Hardy Weinberg assumptions

- **Main assumptions**

  - **Mating is random**

  - **Very large (infinite) population**

  - **No migration**

  - **No mutation**

  - **No natural selection**

- Non-necessary assumptions

  - Diploid organism

  - Reproduction is sexual

  - Non-overlapping generations

Punnett square

Male gametes

Allele	A	a
Frequency	p	q

Female gametes

Allele	Frequency		
A	p	AA p^2	Aa pq
a	q	aA qp	aa q^2

Summed frequencies in zygotes:

AA: $P' = p^2$
Aa: $Q' = pq + qp = 2pq$
aa: $R' = q^2$

Question

- In humans males have X and Y chromosomes and females two X chromosomes

- What does this imply about sex ratios in humans?

- Let's say we have a population (harem) with one male and 100 females. What is the sex ratio in the next generation?

Summed frequencies in zygotes:

$$AA:\ P' = p^2$$
$$Aa:\ Q' = pq + qp = 2pq$$
$$aa:\ R' = q^2$$

- Frequency of A in the next generation

  - $P' + Q'/2 = p^2 + pq = p(p+q) = p$

- Frequency of a in the next generation

  - $Q' + P'/2 = q$

- Thus, the frequency of alleles is the same across generations

- The frequency of genotypes will also be the same at H-W equilibrium

Question

- In the harem example above, what are the frequencies of X and Y chromosomes?

- Did they change across generations?

Dominance

- For polymorphic genes with two alleles and **dominance**, only two phenotypic classes can be observed

- E.g., humans can carry the rhesus factor protein in their blood (Rh+), or not (Rh-)

- Rh- individuals are homozygous **recessive** for the rhesus factor

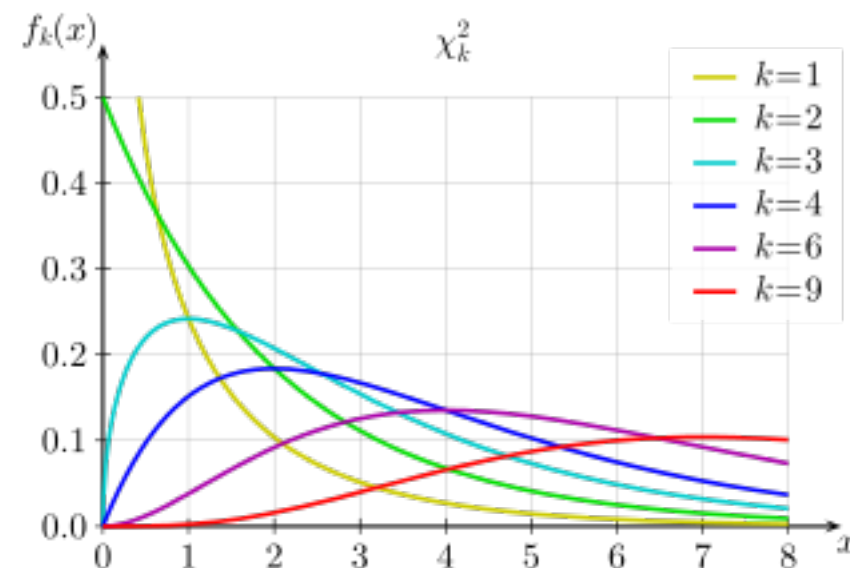- One or two copies of Rh+ give the **dominant** phenotype

Testing Hardy-Weinberg

- Chi-squared test on counts of genotypic classes is the classical approach

$$\chi^2 = \sum \frac{(observed - expected)^2}{expected}$$

df = Number of classes of data

    − Number of parameters estimated from the data   i.e., p

−1

Question

- How can we test if a locus has dominance?

- Knowing that there is dominance at a locus, how can we estimate a gene's frequency based on phenotypic data, assuming Hardy-Weinberg?

X-linked genes

		Male gametes		
		X-bearing		Y-bearing
	Allele	X^A	X^a	
	Frequency	p	q	

Female gametes	Allele	Frequency			
	X^A	p	$X^A X^A$ p^2	$X^A X^a$ pq	$X^A Y$ p
	X^a	q	$X^a X^A$ qp	$X^a X^a$ q^2	$X^a Y$ q

Summed frequencies in zygotes:

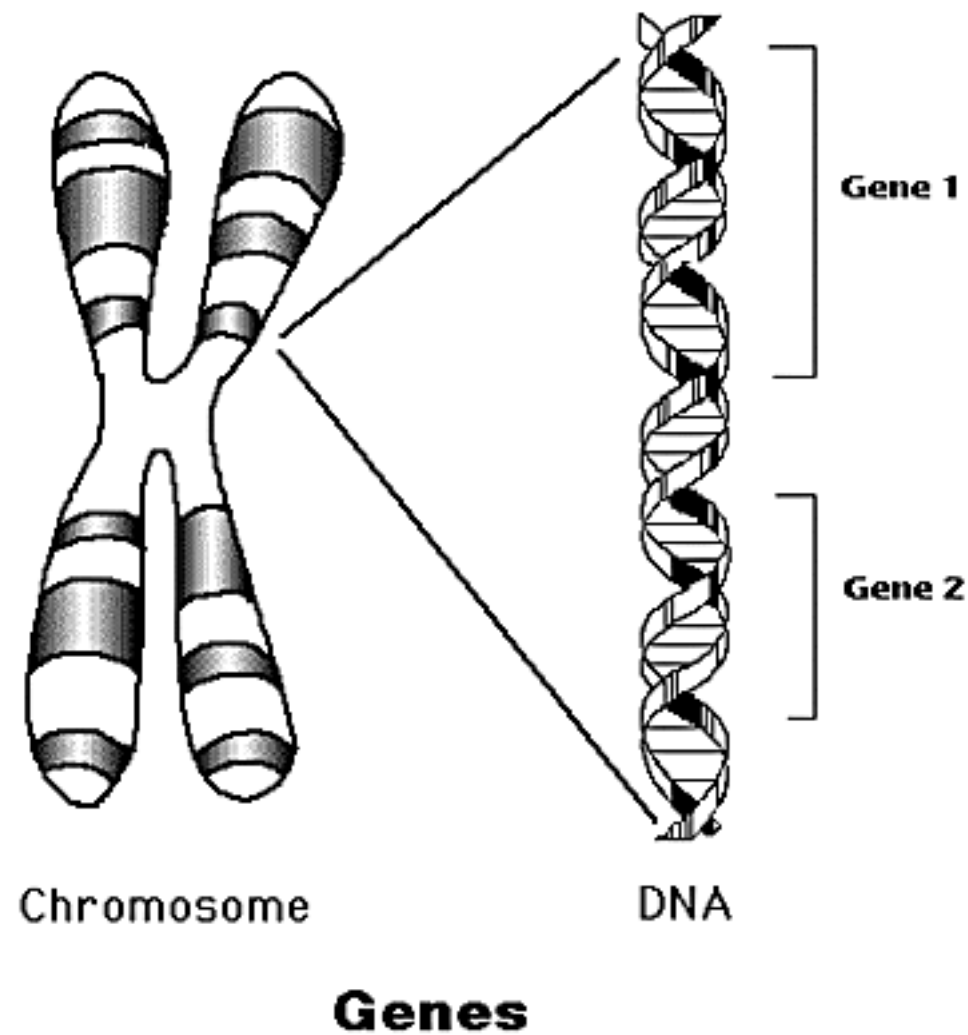| Females		Males	
$X^A X^A$:	p^2	$X^A Y$:	p
$X^A X^a$:	$2pq$	$X^a Y$:	q
$X^a X^a$:	q^2		

Question

- Color-blindness is an X-linked recessive trait. Which sex is it more prevalent and why?
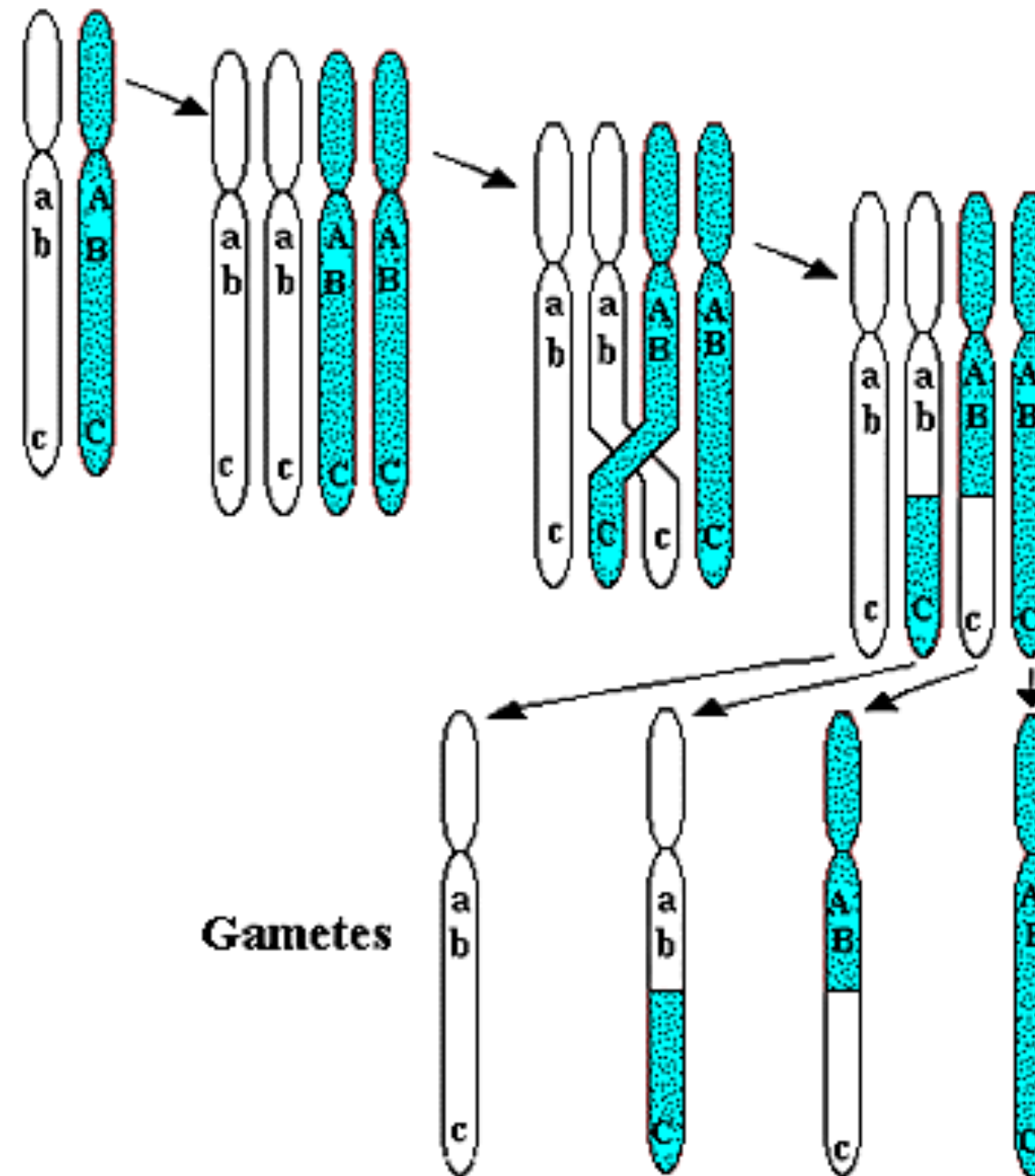
Review question

- What are the conditions of Hardy-Weinberg equilibrium?

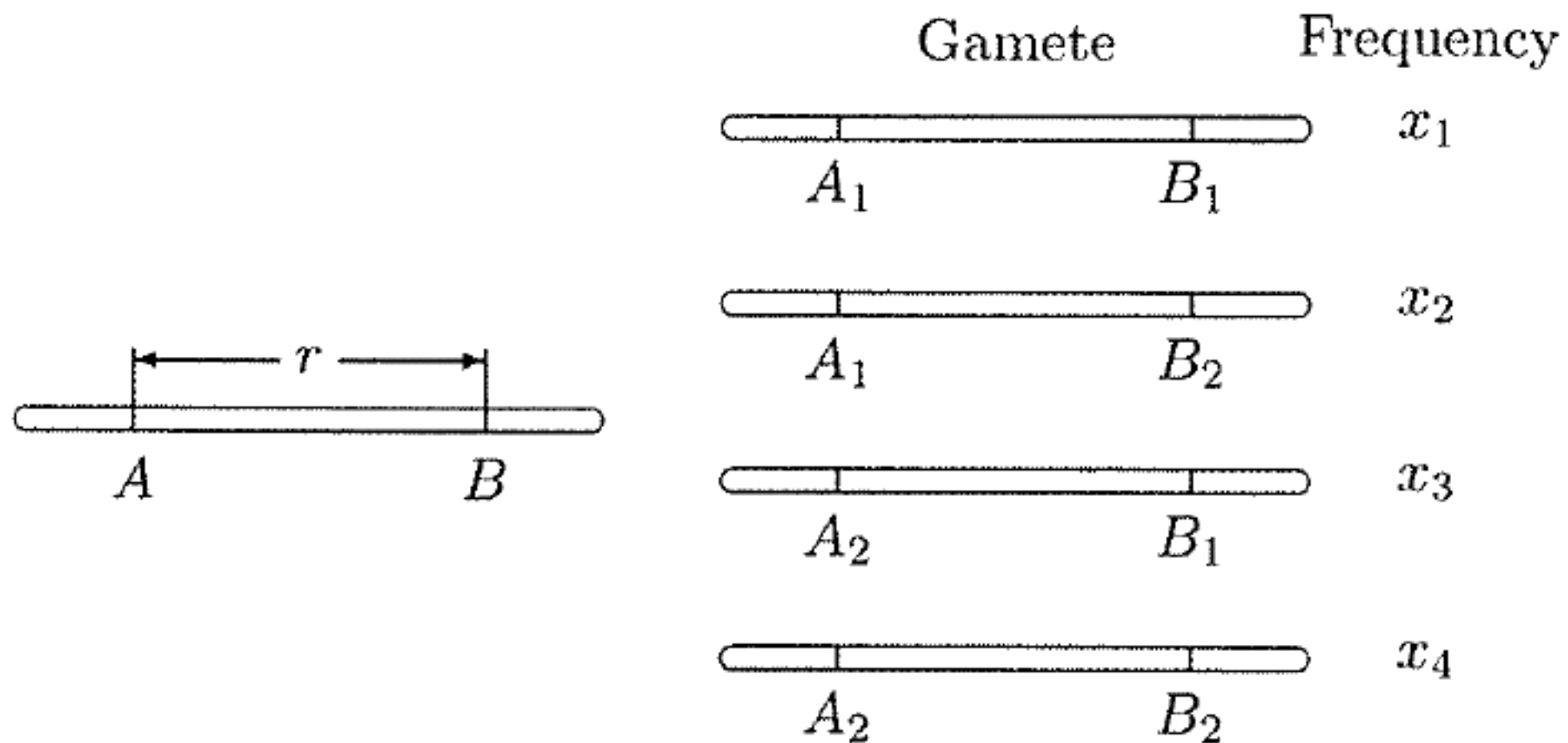- Why is the concept of Hardy-Weinberg equilibrium useful?

Accounting for chromosomes (linkage)

Chromosome   DNA

Genes

Meiosis breaks up combinations of genes

Crossing-over and recombination during meiosis

Gamete	Frequency
A_1 B_1	x_1
A_1 B_2	x_2
A_2 B_1	x_3
A_2 B_2	x_4

The frequency of x_1 in the next generation is

$$x'_1 = (1-r)x_1 \quad \textit{non-recombinant}$$
$$+ \, r \, p_1 \, p_2 \quad \textit{recombinant}$$

The change in the frequency of x_1 is

$$\Delta_r x_1 = -r(x_1 - p_1 p_2)$$

The frequency of x_1 in the next generation is

$$x_1' = (1 - r)x_1 + rp_1p_2$$

The change in the frequency of x_1 is

$$\Delta_r x_1 = -r(x_1 - p_1p_2)$$

The **linkage disequilibrium** coefficient is defined as

$$D = x_1 - p_1p_2$$

Question

- What is the maximum value for r?

The change in D in one generation is
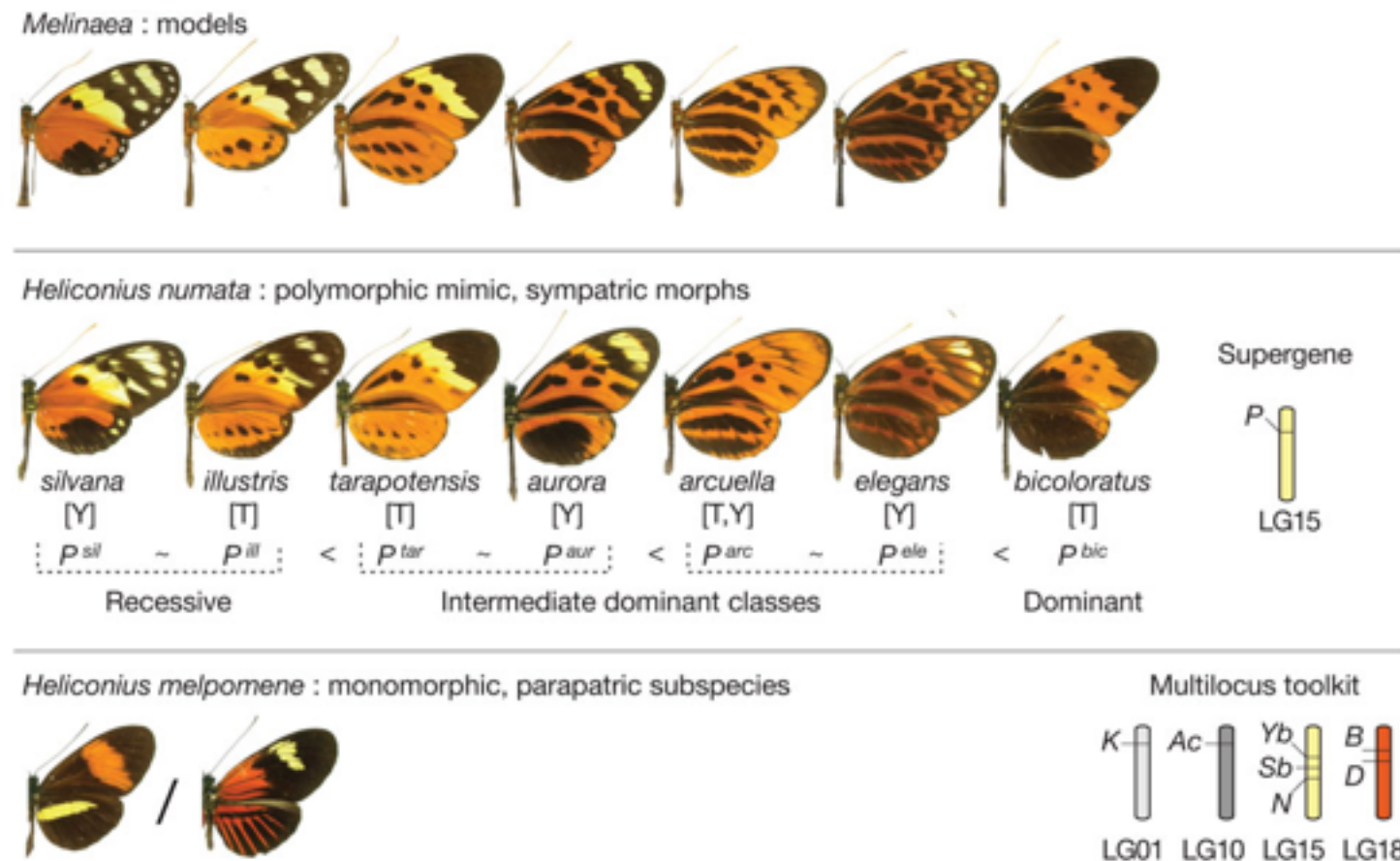
$$\Delta_r D = -rD$$

So,

$$D' = (1 - r)D$$

Over time LD approaches 0:

$$D_t = (1 - r)^t D_0$$

- LD can be under selection
- LD can be affected by drift

LD preserves combinations

Melinaea : models

Heliconius numata : polymorphic mimic, sympatric morphs

Supergene

P

LG15

silvana [Y]   illustris [T]   tarapotensis [T]   aurora [Y]   arcuella [T,Y]   elegans [Y]   bicoloratus [T]

P^{sil} ~ P^{ill} < P^{tar} ~ P^{aur} < P^{arc} ~ P^{ele} < P^{bic}

Recessive   Intermediate dominant classes   Dominant

Heliconius melpomene : monomorphic, parapatric subspecies

Multilocus toolkit

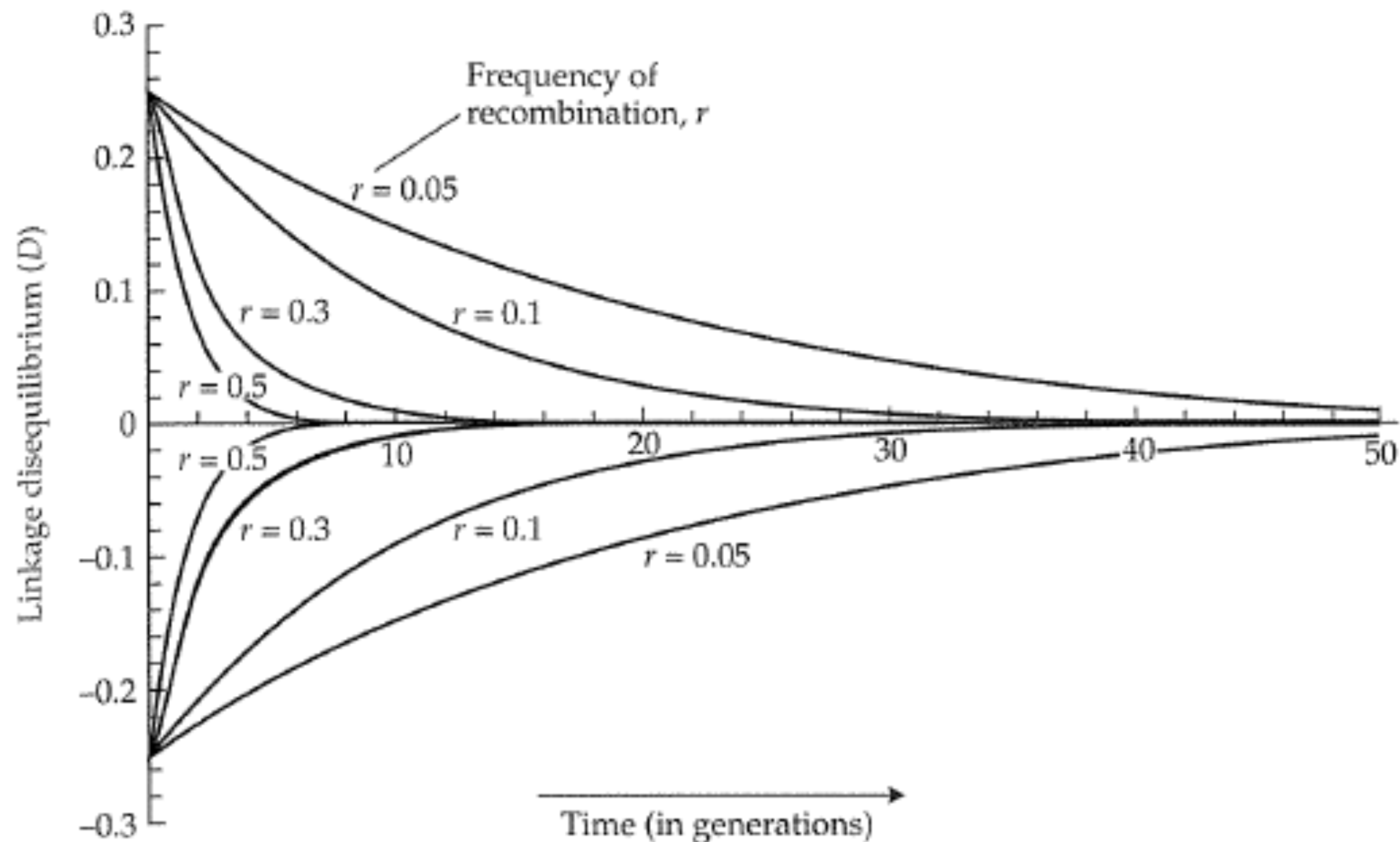K   Ac   Yb Sb N   B D

LG01 LG10 LG15 LG18

- Heliconius butterflies have several genes for color linked on the same chromosome, allowing stable color patterns
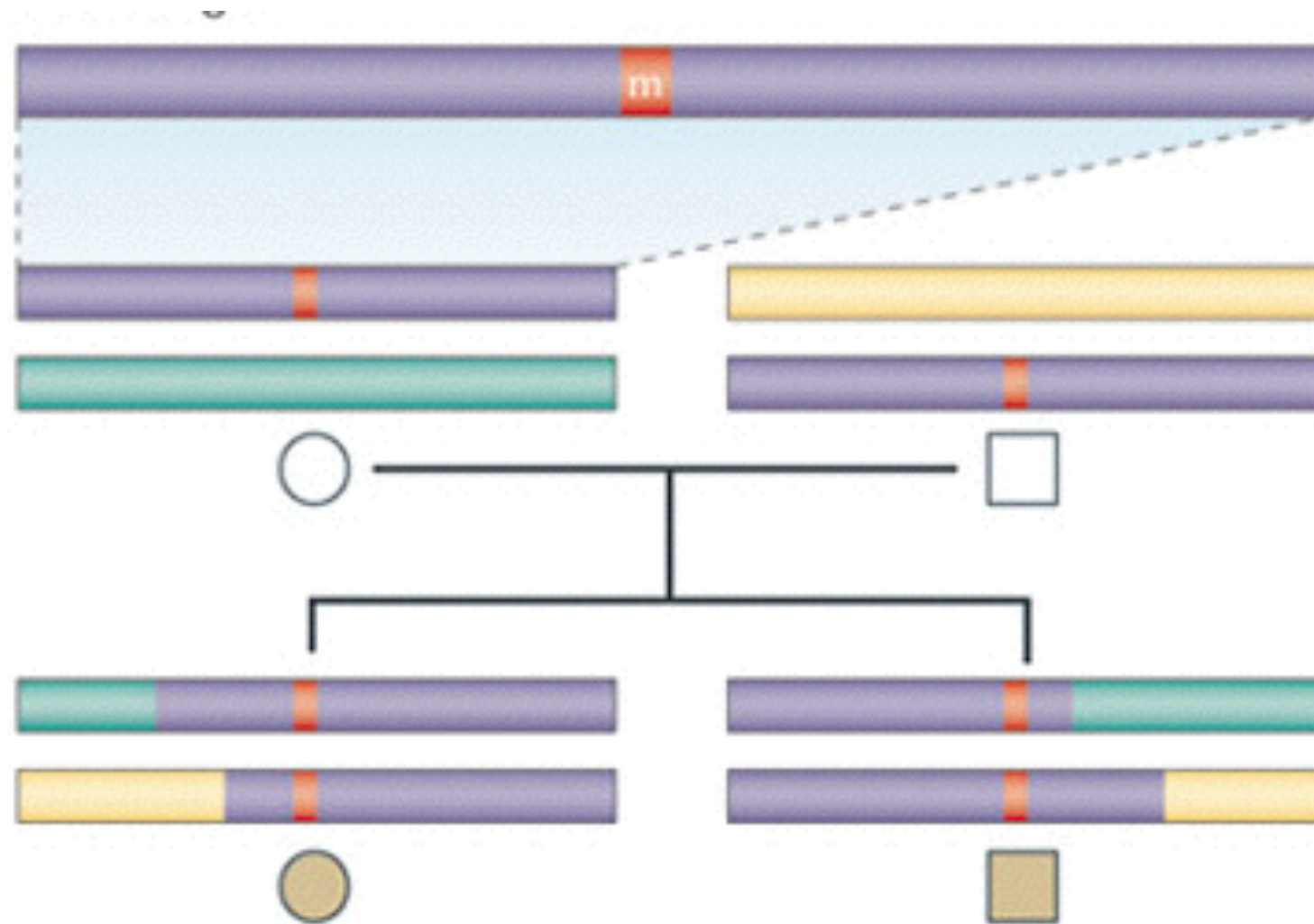
Fire ant social structure is controlled by one locus
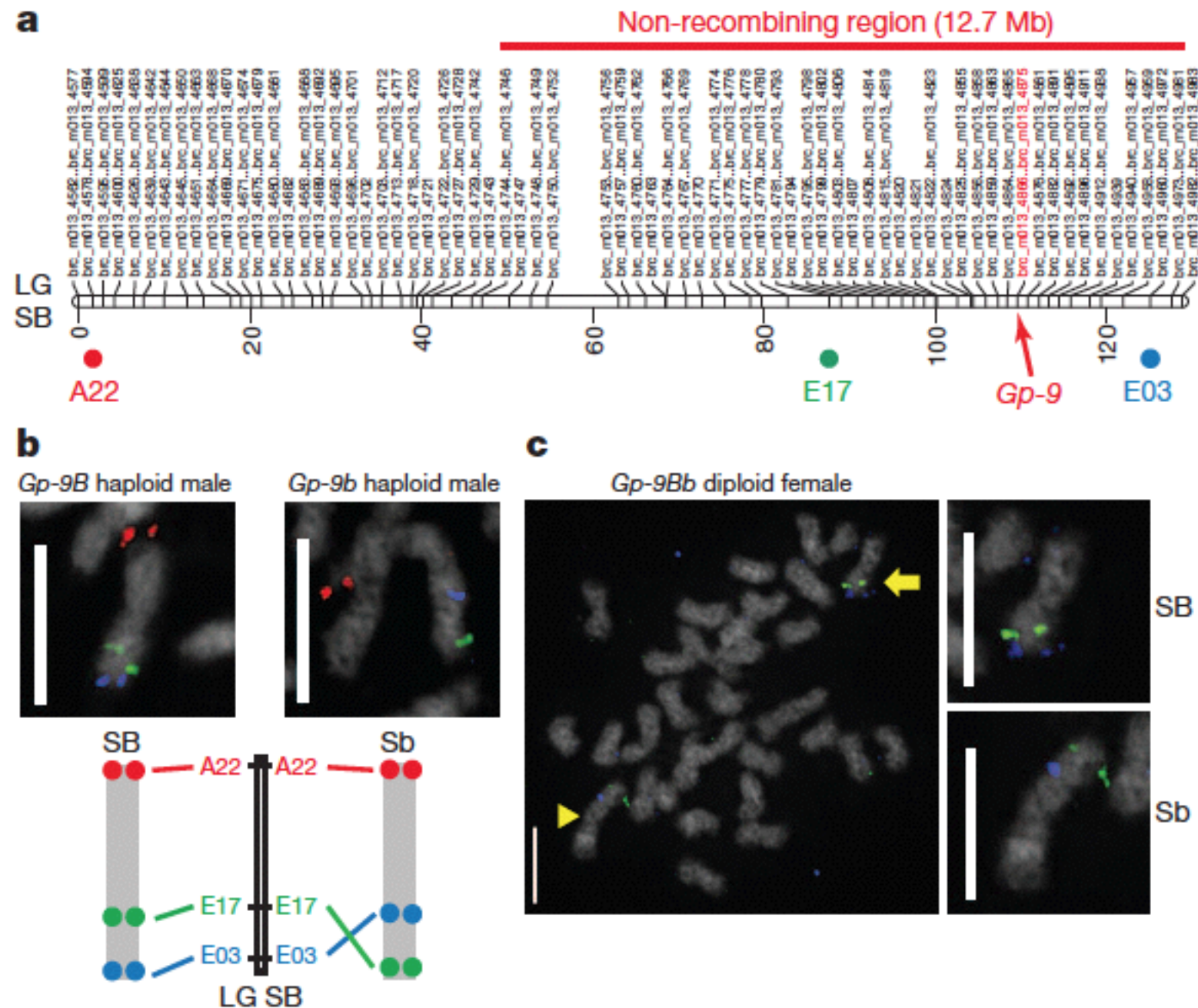
LD decay

With free recombination (r = 1/2), LD does not go to 0 in one generation
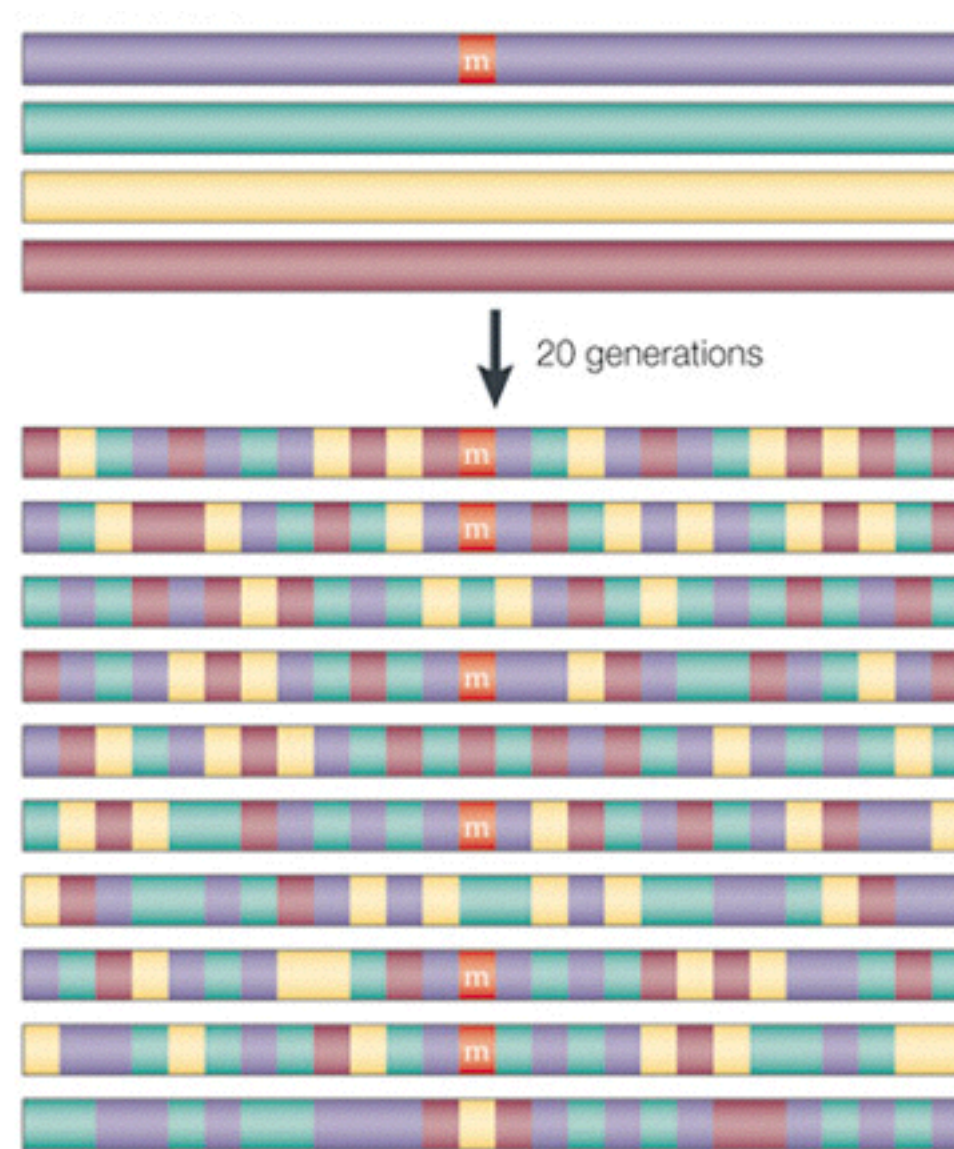
- Populations approach linkage equilibrium between two loci over time, as there is more recombination

- Linkage disequilibrium can be used to find genes associated with specific traits

Linkage

Association

20 generations

Admixture can create LD`

Chromosome	Frequency	Subpopulation 1	Subpopulation 2	Equal mixture
AB	P_{AB}	0.0025	0.9025	0.4525
Ab	P_{Ab}	0.0475	0.0475	0.0475
aB	P_{aB}	0.0475	0.0475	0.0475
ab	P_{ab}	0.9025	0.0025	0.4525
$D = P_{AB}P_{ab} - P_{Ab}P_{aB}$		0	0	0.2025

TABLE 2.2 Linkage Disequilibrium from Admixture of Subpopulations

Sub 1: 0.0025*0.9025 - 0.0475*0.0475 = 0

Sub 2: 0.0025*0.9025 - 0.0475*0.0475 = 0

Merged: 0.4525*0.4525 - 0.0475*0.0475 = 0.2025

For next week

- Rosalind problems

- Reading

- Start thinking about project