Hardy-Weinberg genotype frequencies

- If alleles come together at random to form genotypes, then genotype frequencies will be products of allele frequencies:

  $P(AA) = p^2$

  $P(Aa) = 2pq$

  $P(aa) = q^2$

- Otherwise, there is Hardy-Weinberg disequilibrium (HWD)

Linkage Disequilibrium (LD)

- Alleles at one locus are correlated with alleles at a second locus on a population level.

A	**A**	**A**	**A**	**A**	**A**	**A**	**A**	**a**	**a**	**a**	**a**
B	**B**	**B**	**B**	**B**	**B**	**b**	**b**	**b**	**b**	**b**	**b**

If you sample a haplotype at random from the population, does knowing the allele at the first locus of this haplotype give you information about the allele at the second locus?

Linkage Disequilibrium

- One usual measure of LD is:

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

A	**A**	**A**	**A**	**A**	**A**	**A**	**A**	**a**	**a**	**a**	**a**
B	**B**	**B**	**B**	**B**	**B**	**b**	**b**	**b**	**b**	**b**	**b**

- $P_A = 8/12$
- $P_B = 6/12$
- $P_{AB} = 6/12$
- $D_{AB} = 1/6$

Linkage Disequilibrium

- One usual measure of LD is:

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

A	**A**	**A**	**A**	**A**	**A**	**A**	**A**	**a**	**a**	**a**	**a**
B	**B**	**B**	**B**	**B**	**B**	**b**	**b**	**b**	**b**	**b**	**b**

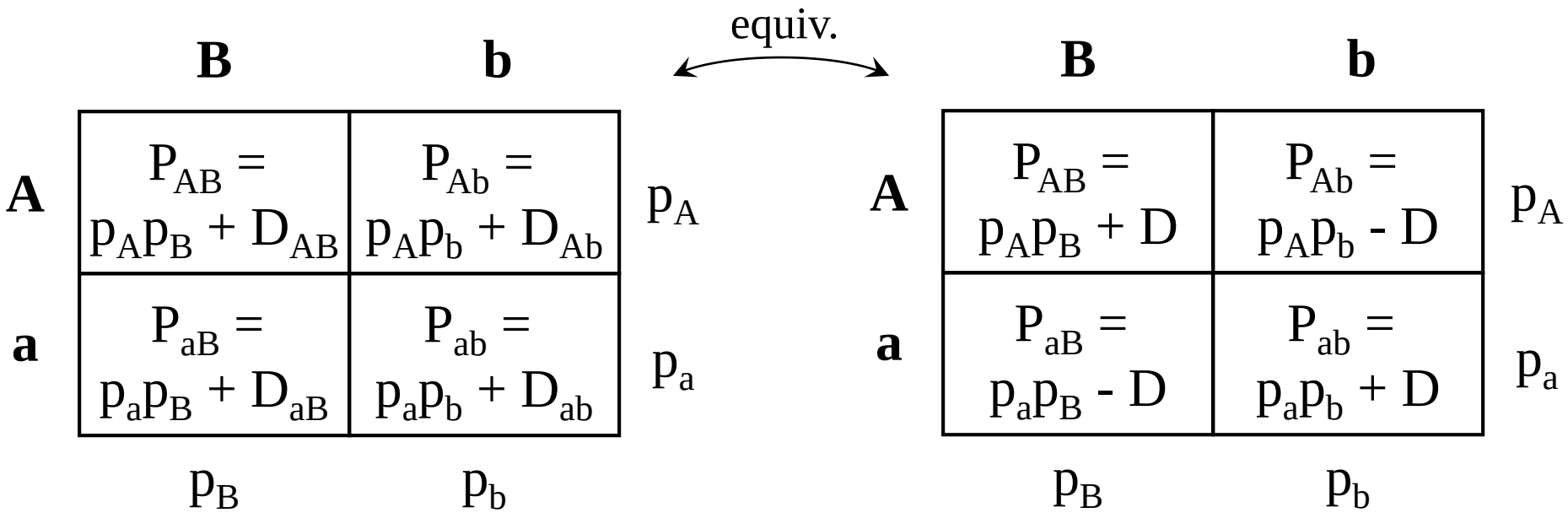
- **LD is a measure of extant haplotypes**
  - Estimates do not rely on or measure inheritance
  - LD does not measure how often alleles are *transmitted* together

Properties of LD

- One usual measure of LD is:

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

- $\mathbf{D_{AB} = -D_{Ab} = -D_{aB} = D_{ab}}$ ← can denote as **D**



Bounds on LD

- $0 \leq p_A p_B + D_{AB} \leq \min(p_A, p_B)$
- $0 \leq p_A p_b - D_{AB} \leq \min(p_A, p_b)$
- $0 \leq p_a p_B - D_{AB} \leq \min(p_a, p_B)$
- $0 \leq p_a p_b + D_{AB} \leq \min(p_a, p_b)$

- $\max(-p_A p_B, -p_a p_b) \leq D_{AB} \leq \min(p_A p_b, p_a p_B)$

Other Measures of LD

- D' ("D-prime") normalize D by the maximum value it can obtain:

$$D'_{AB} = \begin{cases} D_{AB} / [\max(-p_A p_B, -p_a p_b)] & \text{if } D_{AB} < 0 \\ D_{AB} / [\min(p_A p_b, p_a p_B)] & \text{if } D_{AB} > 0 \end{cases}$$

- $r^2_{AB} = \dfrac{D^2_{AB}}{p_A p_B p_a p_b}$ (squared correlation coefficient, r)

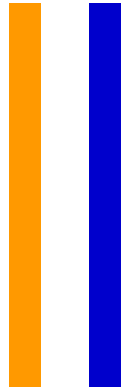
LD can be created by

- Mixing of populations.
- Population substructure (non-random mating within populations).
- Mutations creating new haplotypes.
- Selection favoring certain alleles.
- Founder effects.
- Genetic drift.

LD decays over time

- LD *decays* because of recombination
- $D_{AB}(g) = (1 - r)^g \times D_{AB}(0)$
  - r = recombination rate, g = number of generations
- Although this predicts the expected value of LD over time, there is a large variance around the mean.
- Populations with similar starting values can be quite different after time.

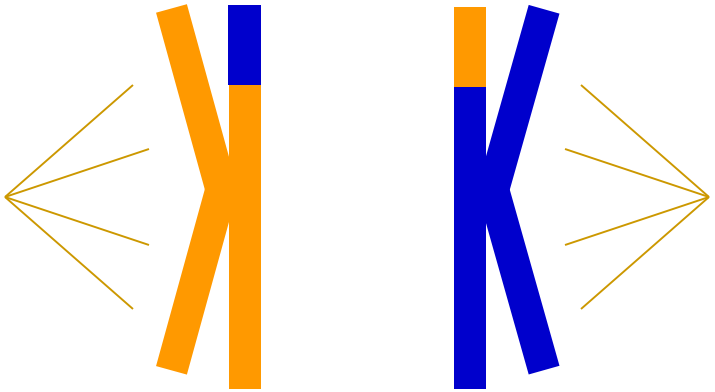
Reminder: meiosis & recombination (extremely simplified)



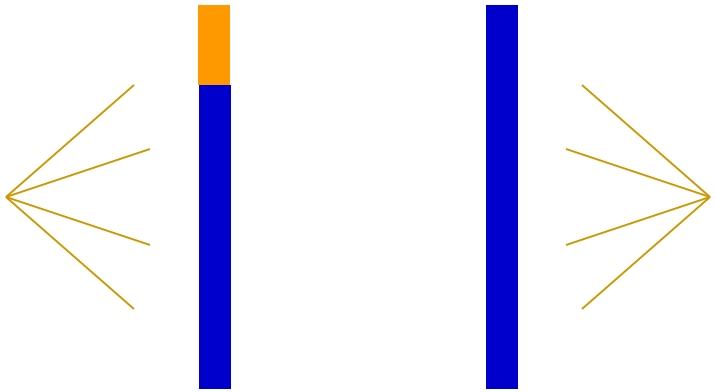
diploid progenitor

DNA replication

recombination



meiosis I



meiosis II

Consider Two Populations

- Population 1 haplotypes:

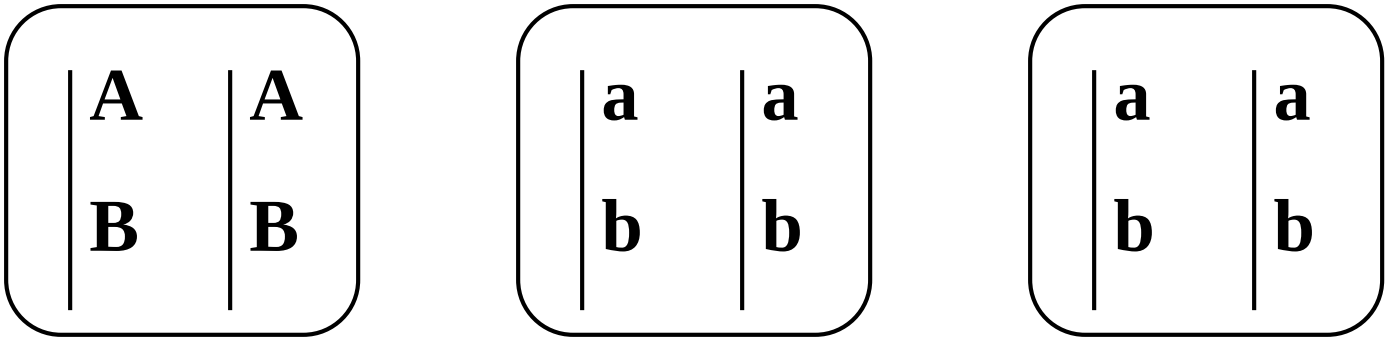
A	**A**	**A**	**A**	**A**	**A**	**A**
B	**B**	**B**	**B**	**B**	**B**	**B**

- Population 2 haplotypes:

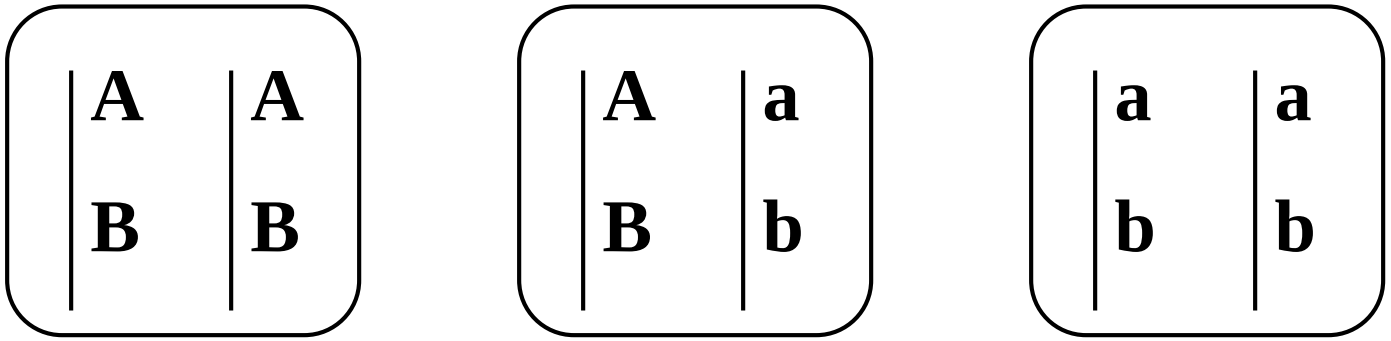
a	**a**	**a**	**a**	**a**	**a**	**a**
b	**b**	**b**	**b**	**b**	**b**	**b**

If These Populations Mix

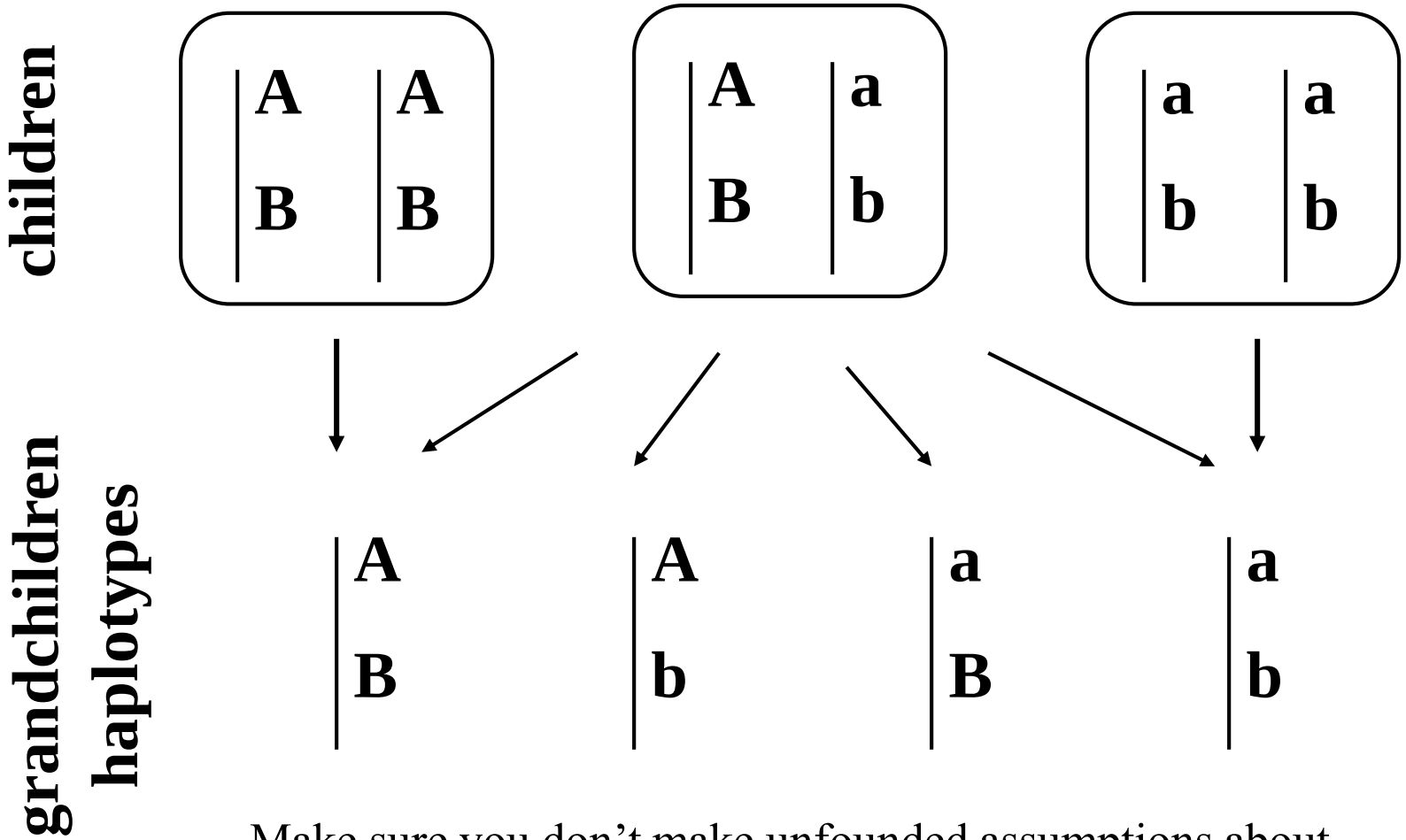
- Some individuals directly after mixing:



- Their possible children:

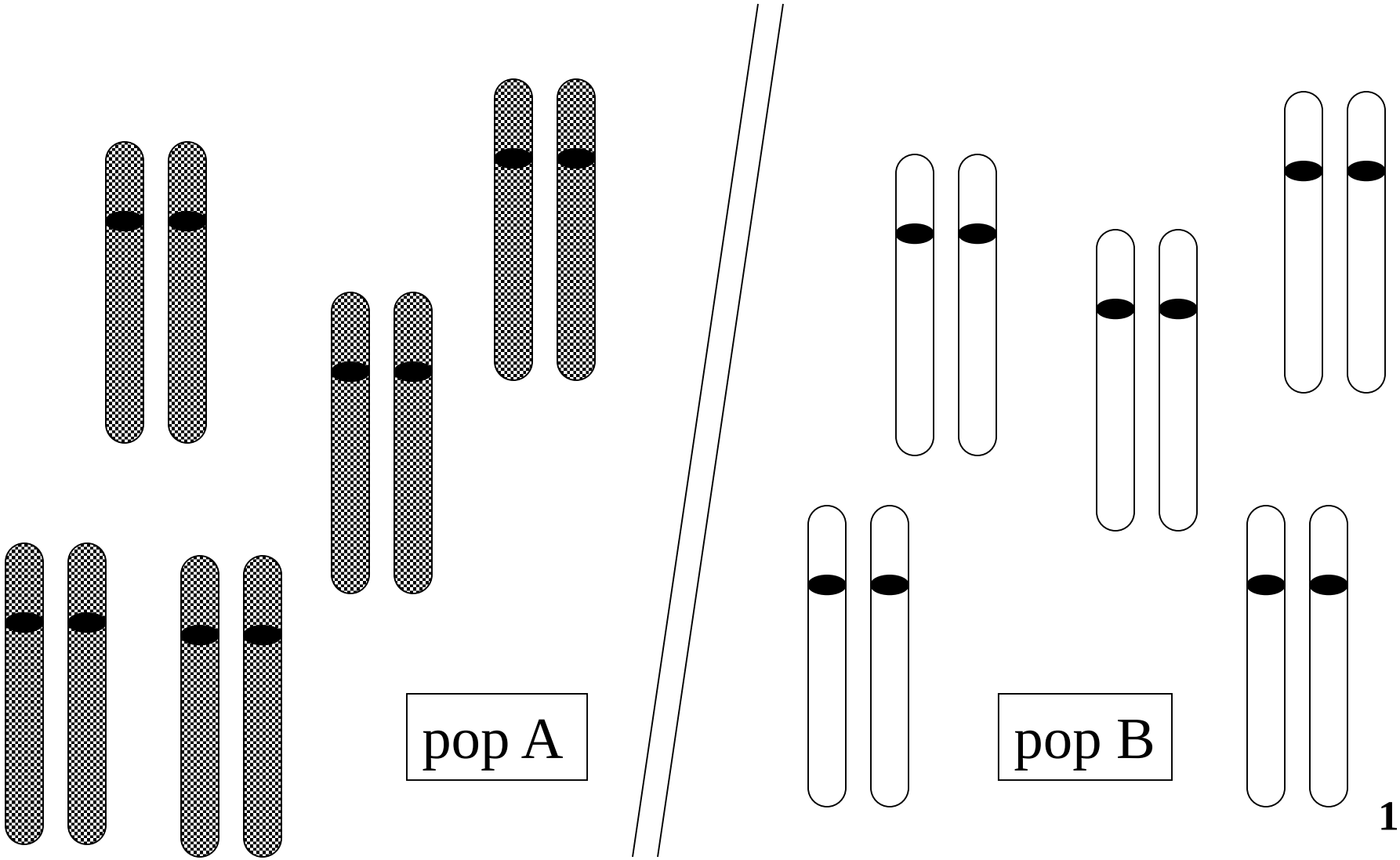


Haplotype Possibilities for the Grandchildren

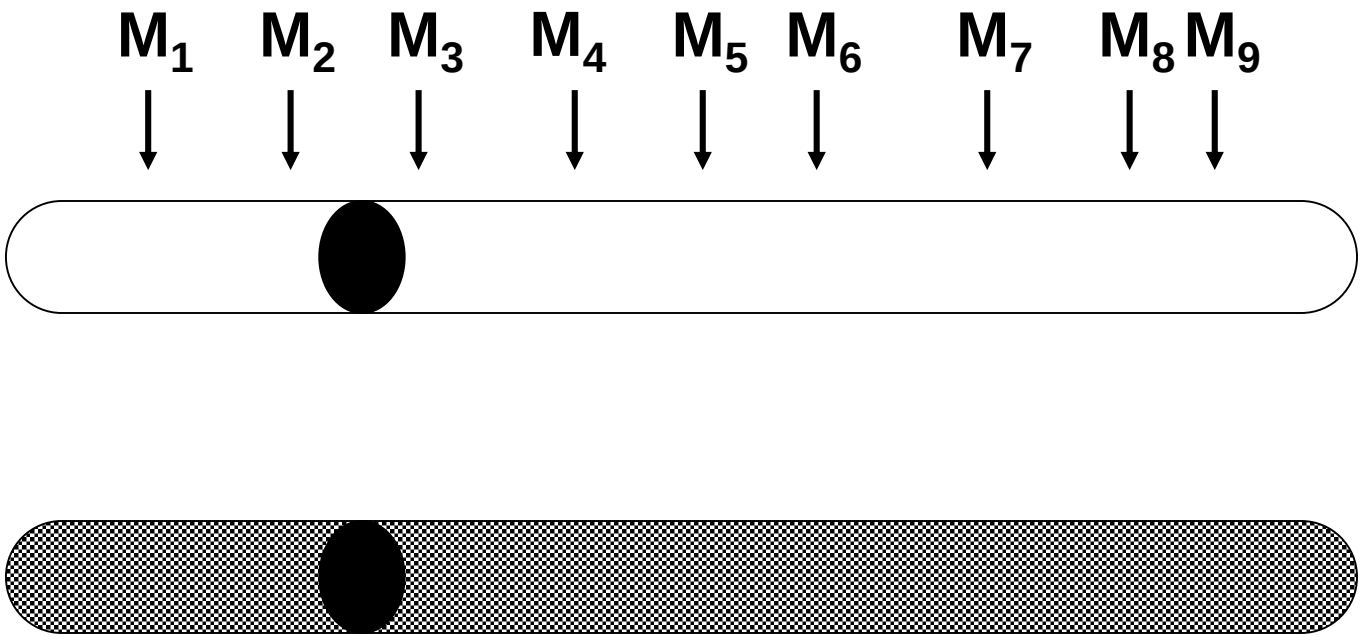


Make sure you don't make unfounded assumptions about the probability of seeing any given haplotype! (e.g. 0.5)

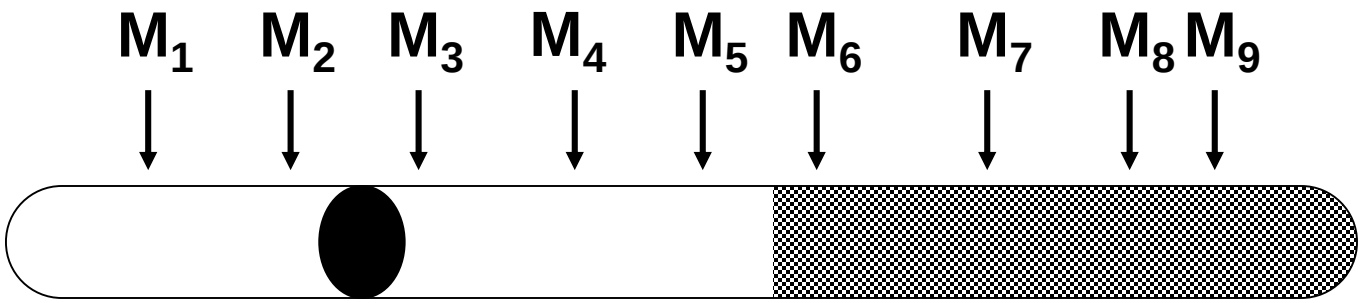
Consider this on a chromosome-wide basis



Possible full-chromosome haplotypes

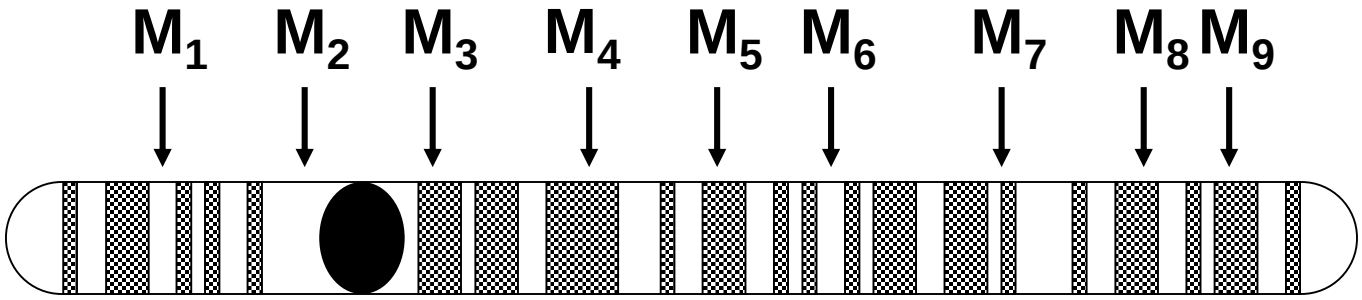


After a few generations ...

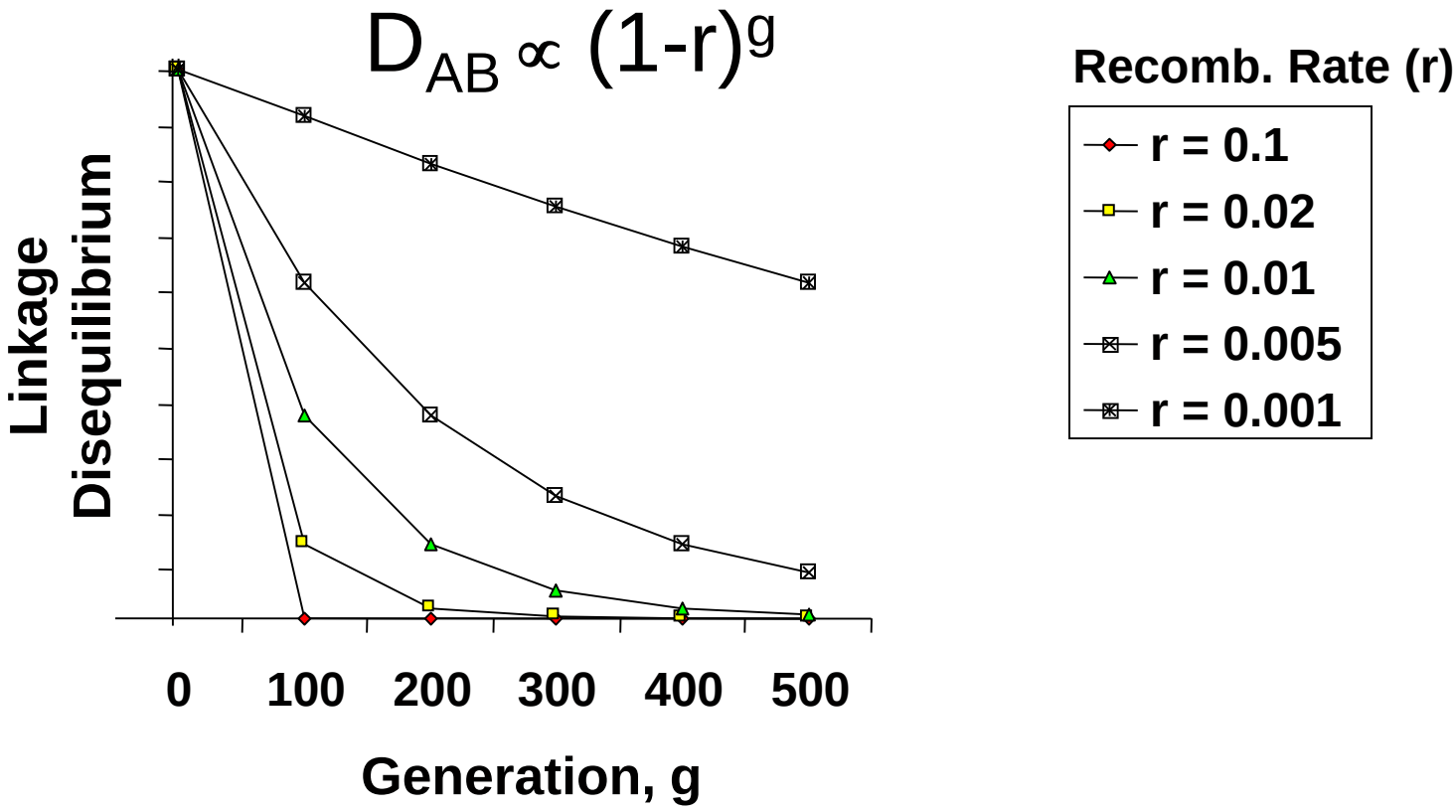


After many generations

(many recombination events)



Linkage Disequilibrium versus Generations Since its Creation



Linkage versus LD

Linkage is *defined by* recombination.

- Recombination occurs during meiosis and thus, in turn, is observed via inheritance.
- Loci whose recombination rates are < 0.5 are linked.
- Loci whose recombination rates are = 0.5 are unlinked.
- Linkage is measured via correlated *transmission* of alleles.

LD is *affected by* recombination (over time).

- LD measures correlation between alleles in a population.
- LD is based on extant haplotypes and *estimates do not rely on or measure inheritance*.
- LD breaks down over time via recombination.
- LD does ***not*** measure correlated *transmission* of alleles.