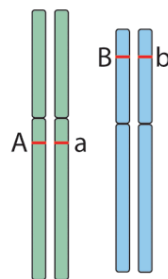


## Linkage and Recombination

- ❖ “linkage equilibrium”
  - ❖ refers to cases in which the alleles of different genes are in random association
  - ❖ expected when genes are on different chromosomes
  - ❖ gamete frequencies (genes “A” and “B”):
    - ❖  $AB: p_A \times p_B$
    - ❖  $Ab: p_A \times q_b$
    - ❖  $aB: q_a \times p_B$
    - ❖  $ab: q_a \times q_b$

Unlinked loci are generally in equilibrium

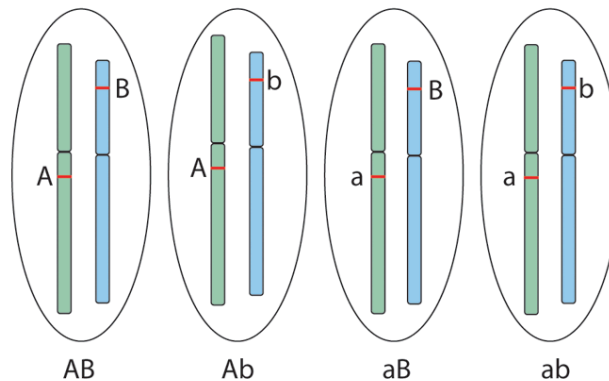


two genes on different chromosomes



four possible gametes produced in equal proportions (25% each)

- ❖ four possible gametes produced in equal proportions



## HWE example for *unlinked* genes

- ❖  $p_A = 0.7, p_a = 0.3$ 
  - ✧  $AA = 0.49, Aa = 0.42, aa = 0.09$
- ❖  $p_B = 0.2, p_b = 0.8$ 
  - ✧  $BB = 0.04, Bb = 0.32, bb = 0.64$
- ❖ 9 possible two-locus genotypes:
 

✧ $AABB = 0.0196$	$AABb = 0.1568$	$AAbb = 0.3136$
✧ $AaBB = 0.0168$	$AaBb = 0.1344$	$Aabb = 0.2688$
✧ $aaBB = 0.0036$	$aaBb = 0.0288$	$aabb = 0.0576$

## HWE example for *unlinked* genes

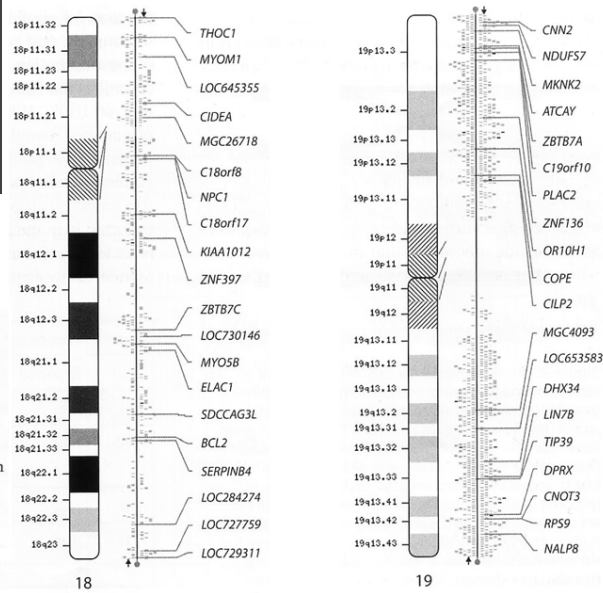
- ❖  $p_A = 0.7, p_a = 0.3$
- ❖  $p_B = 0.2, p_b = 0.8$

- randomly combining genotypes for each locus yields the same result as randomly combining two-locus gametes

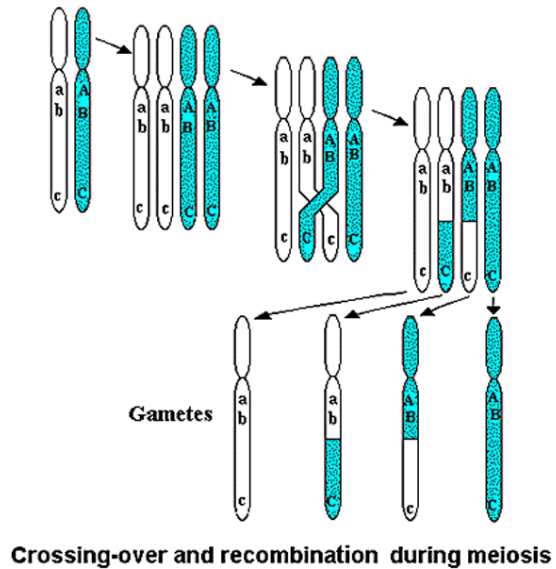
			<b>AA</b>	<b>Aa</b>	<b>aa</b>	
			0.49	0.42	0.09	
<b>BB</b>	0.04	<b>0.0196</b>	<b>0.0168</b>	<b>0.0036</b>		
<b>Bb</b>	0.32	<b>0.1568</b>	<b>0.1344</b>	<b>0.0288</b>		
<b>bb</b>	0.64	<b>0.3136</b>	<b>0.2688</b>	<b>0.0576</b>		
		<b>AB</b>	<b>Ab</b>	<b>aB</b>	<b>ab</b>	
			0.14	0.56	0.06	0.24
<b>AB</b>	0.14	<b>0.0196</b>	<b>0.0784</b>	<b>0.0084</b>	<b>0.0336</b>	
<b>Ab</b>	0.56	<b>0.0784</b>	<b>0.3136</b>	<b>0.0336</b>	<b>0.1344</b>	
<b>aB</b>	0.06	<b>0.0084</b>	<b>0.0336</b>	<b>0.0036</b>	<b>0.0144</b>	
<b>ab</b>	0.24	<b>0.0336</b>	<b>0.1344</b>	<b>0.0144</b>	<b>0.0576</b>	

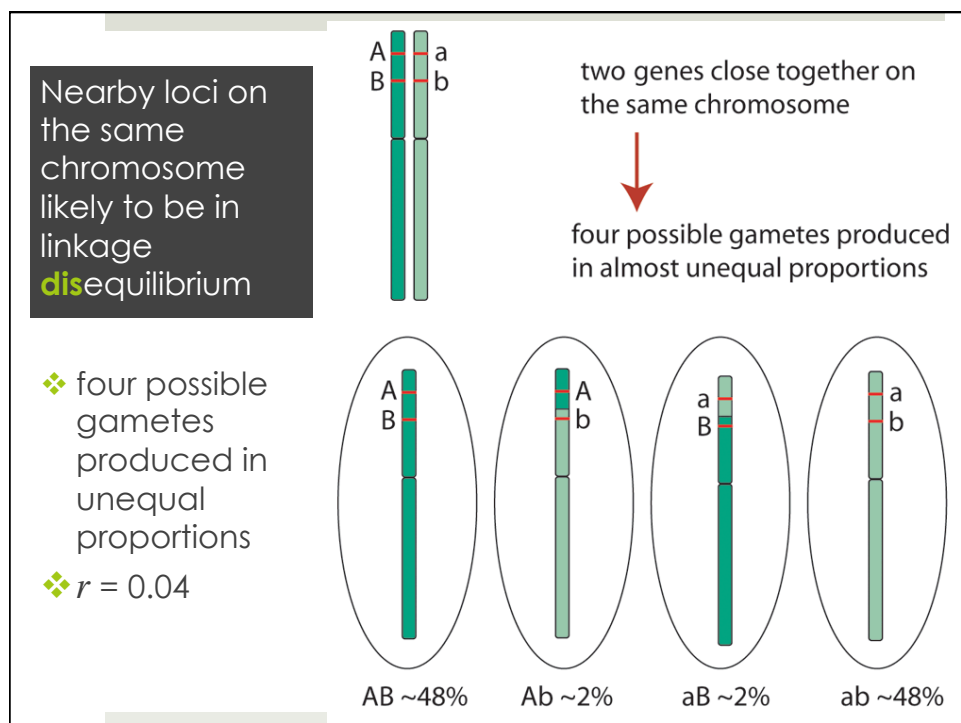
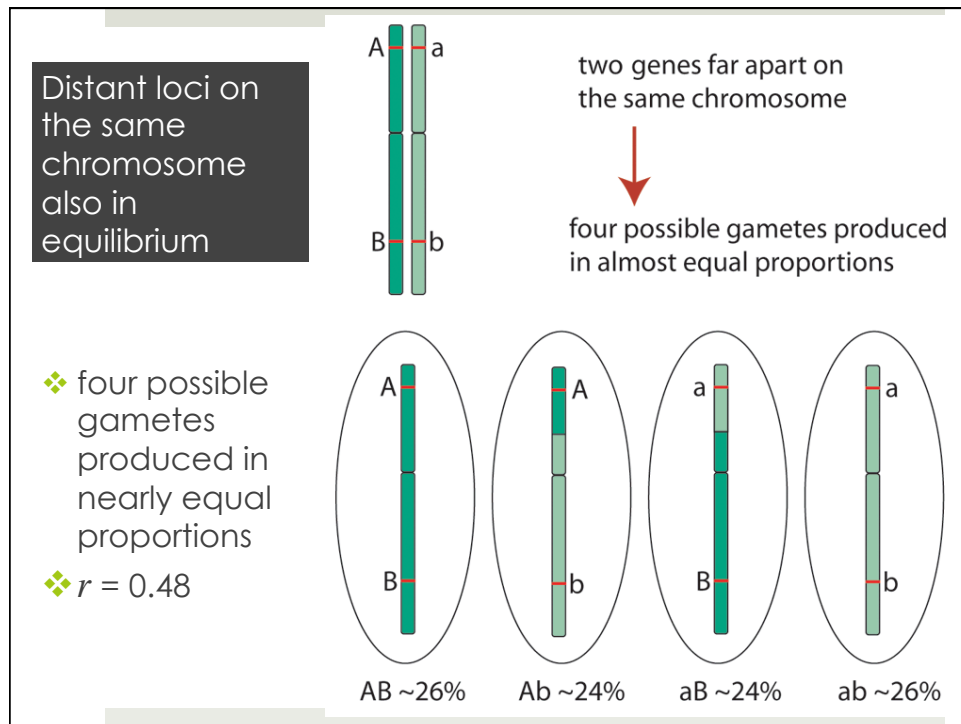
Genes on human chromosomes 18 and 19

**Figure 2.17** Maps for human chromosomes 18 (left) and 19 (right) showing chromosome regions, the physical locations of identified genes and open reading frames (labeled orf) along the chromosomes, and the names and locations of a subset of genes. Chromosome 18 is about 85 million bp and chromosome 19 is about 67 million bp. Maps from NCBI Map Viewer based on data as of January 2008.



Recombination by crossing-over during meiosis





## Frequency of recombination, $r$

- ❖  $AB/ab$  genotype produces gametes:
  - ❖  $AB$  with frequency  $(1-r)/2$
  - ❖  $ab$  with frequency  $(1-r)/2$
  - ❖  $Ab$  with frequency  $r/2$
  - ❖  $aB$  with frequency  $r/2$
- ❖  $Ab/aB$  genotype produces gametes:
  - ❖  $AB$  with frequency  $r/2$
  - ❖  $ab$  with frequency  $r/2$
  - ❖  $Ab$  with frequency  $(1-r)/2$
  - ❖  $aB$  with frequency  $(1-r)/2$

## Linkage Disequilibrium

- ❖  $D = P_{AB}P_{ab} - P_{Ab}P_{aB}$  ( $= g_{11}g_{22} - g_{12}g_{21}$ )
  - ❖ if more  $AB$  and  $ab$  chromosomes, then positive
  - ❖ if more  $Ab$  and  $aB$  chromosomes, then negative
- ❖ Note: linkage disequilibrium also possible without physical linkage (i.e., genes on different chromosomes)
  - ❖ = "gametic phase disequilibrium"
  - ❖ population admixture, selection, selfing, genetic drift

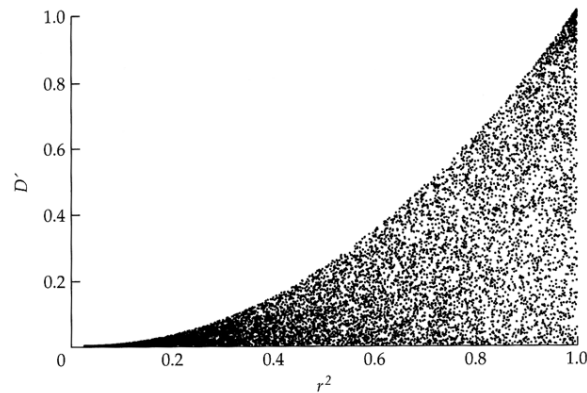
## Linkage Disequilibrium

- ❖ Scaling  $D$  w.r.t. allele frequencies
- ❖  $D_{min}$  = the larger of  $-p_A p_B$  and  $-q_a q_b$
- ❖  $D_{max}$  = the smaller of  $p_A q_b$  and  $q_a p_B$
- ❖ linkage often expressed as:
  - ❖  $D' = D/D_{max}$  [if positive] or  $D/D_{min}$  [if negative]
  - ❖  $\rho^2 = D^2 / (p_A q_a p_B q_b)$  - [for two loci with two alleles each]
    - ❖ “rho-squared” ranges from 0 to 1
- ❖ \*\*Please note!  $\rho^2$  (sometimes written  $r^2$ ) is a completely different measure from the frequency of recombination,  $r$

## Error: pg. 43

Gametic disequilibrium can be measured using several estimators, including the squared correlation coefficient ( $\rho^2$ , where  $\rho$  is pronounced “rho”), where  $\rho^2 = D^2 / (p_1 p_2 q_1 q_2)$ , which has a range of  $-1$  to  $+1$ .

## Possible Values of $D'$ and $\rho^2$



2.11 Relationship between  $D'$  and  $r^2$  for 10,000 random, uniformly distributed values of the gametic frequencies.

## Approach to Equilibrium

- ❖ gamete (= chromosome) frequencies approach random expectations with each successive generation

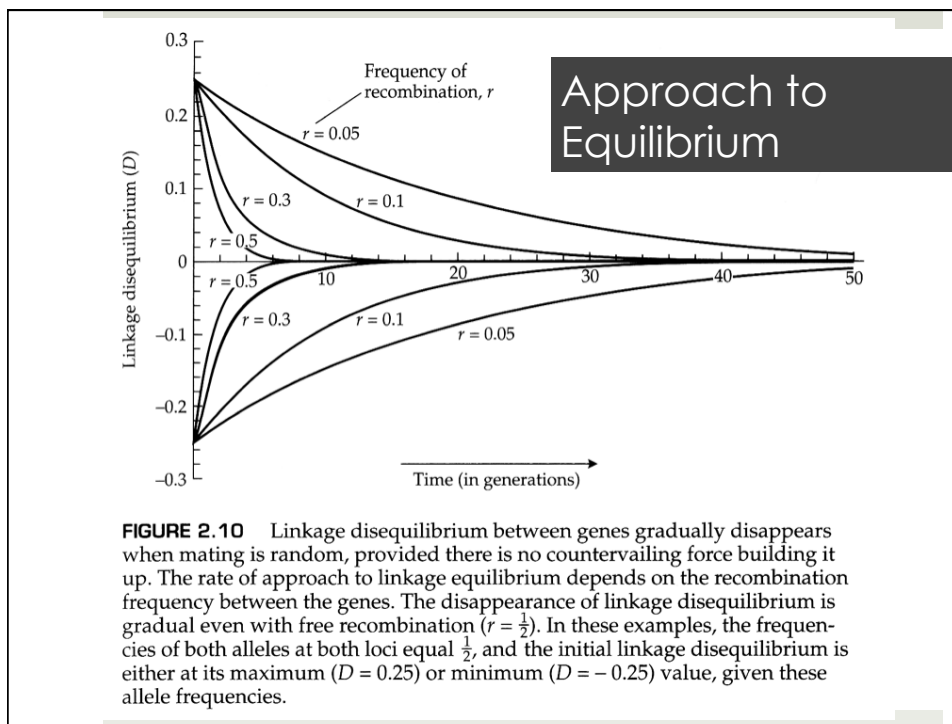
$$\diamond P_{AB}' = (1-r) \times P_{AB} + r \times p_A p_B$$

- ❖ subtracting  $p_A p_B$  from both sides ..  
 remaining, non-  
 recombinant

$$\diamond P_{AB} - p_A p_B = (1-r)(P_{AB} - p_A p_B) = (1-r)D$$

AB gametes
AB gametes  
chromosomes

- ❖ in the  $n$ th generation,  $D_{tn} = (1-r)^n D_{t0}$



## Testing for Linkage Equilibrium

- ❖ issues to consider:
  - ❖ many more categories of genotypes may result in low statistical power
  - ❖ difficulty of determining “gametic phase” in double heterozygotes
- ❖ nonetheless, strong deviations from random expectations can be detected



## History of Linkage

- ❖ within a single random breeding population, linkage disequilibrium reflects the history of mutations
- ❖ consider an example with no recombination...
  - ❖ start with only  $AB$ 
    - ❖ [denoting the alleles at two linked genes]
  - ❖ mutation results in  $aB$  chromosomes
  - ❖ a second mutation results in  $Ab$  chromosomes
  - ❖ so, there are no  $ab$  chromosomes

## Processes producing linkage disequilibrium

- ❖ new mutations
  - ❖ occur on a particular genetic background
- ❖ founding events (i.e., genetic drift)
- ❖ selection
  - ❖ selective sweeps
  - ❖ selection for coadapted genes and reduced recombination
    - ❖ e.g., chromosomal inversions
- ❖ extreme inbreeding - e.g., selfing plants
- ❖ admixture of differentiated populations

## Population Admixture Example

❖ Subpop1:  $p_A = 0.05, p_B = 0.05$

❖ Subpop2:  $p_A = 0.95, p_B = 0.95$

$$D' = 0.81$$

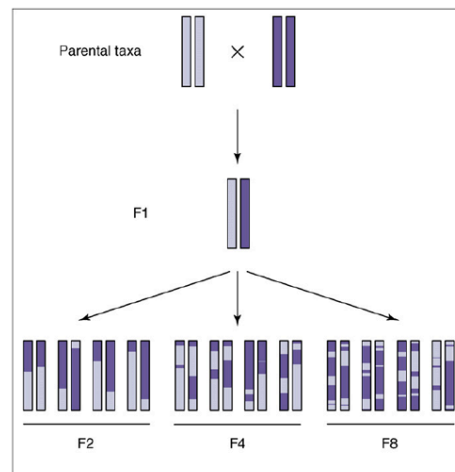
$$r^2 = 0.66$$



**TABLE 2.2** Linkage Disequilibrium from Admixture of Subpopulations

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
$D_{min}$		-0.0025	-0.0025	-0.2500
$D_{max}$		0.0475	0.0475	0.2500

## Recombination in an admixed population



## Length of haplotypes blocks

- ❖ Key point! — regions exhibiting linkage disequilibrium are generally quite short in large natural populations
  - ❖ e.g., a few kb (thousand nucleotides) in *Drosophila* and maize
  - ❖ a few tens of kb in humans and dogs
  - ❖ length varies with local recombination rates and timing and strength of natural selection
- ❖ Why is this significant for population genetics?

## Recombination and Selection

**Figure 2.20** The decay of gametic disequilibrium ( $D$ ) over time when both strong natural selection and recombination are acting. Initially, there are only coupling ( $P_{11} = P_{22} = 1/2$ ) and no repulsion gametes ( $P_{12} = P_{21} = 0$ ). The relative fitness values of the AAbb and aaBB genotypes are 1 while all other genotypes have a fitness of  $1/2$ . Unlike in Fig. 2.19, gametic disequilibrium does not decay to zero over time due to the action of natural selection.

