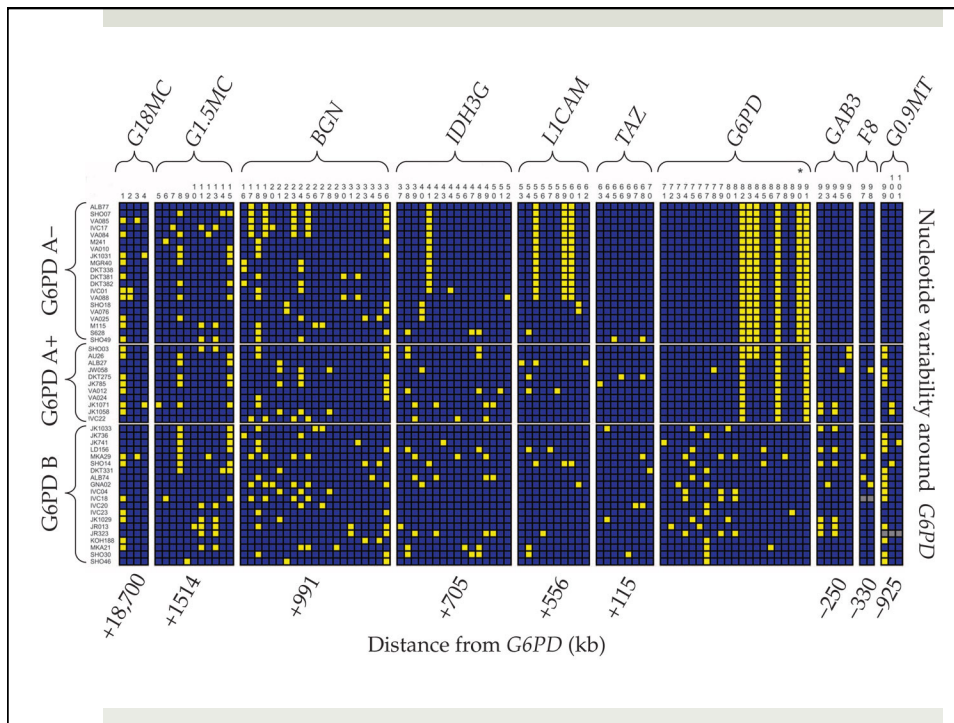


# Chapter 6 – Linkage Disequilibrium & Gene Mapping (Recombination)

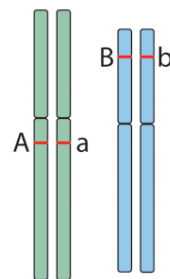
Linkage Disequilibrium  
 Genealogical Interpretation of LD  
 Association Mapping



## 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:  $f_A f_B$
    - ❖ Ab:  $f_A f_b$
    - ❖ aB:  $f_a f_B$
    - ❖ ab:  $f_a f_b$

Unlinked loci are generally in equilibrium

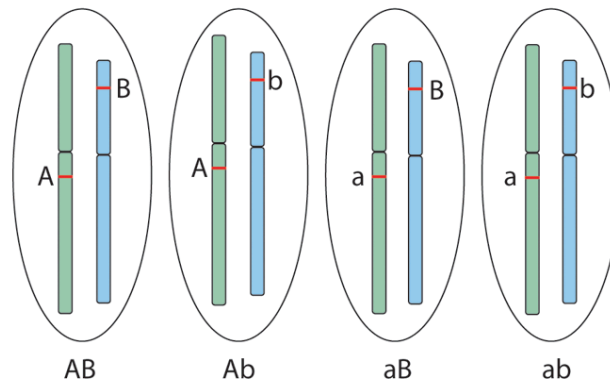


two genes on different chromosomes



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

- ❖ four possible gametes produced in equal proportions



## HWE example for *unlinked* genes

❖  $f_A = 0.7, f_a = 0.3$

❖  $AA = 0.49, Aa = 0.42, aa = 0.09$

❖  $f_B = 0.2, f_b = 0.8$

❖  $BB = 0.04, Bb = 0.32, bb = 0.64$

❖ 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

❖  $f_A = 0.7, f_a = 0.3$

❖  $f_B = 0.2, f_b = 0.8$

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

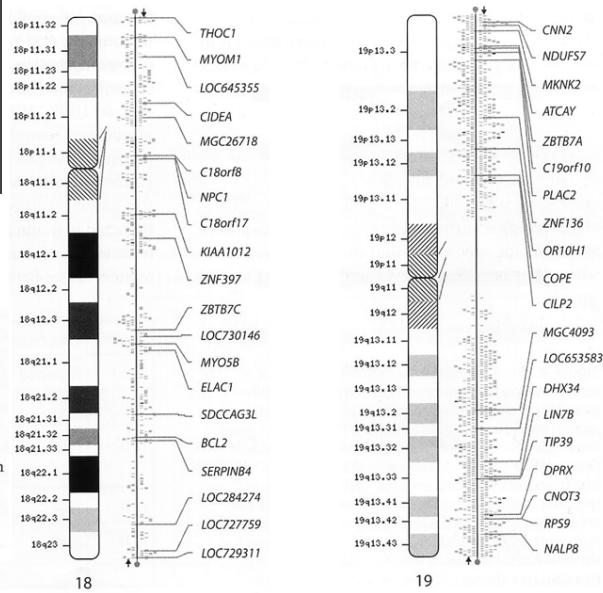
		AA 0.49	Aa 0.42	aa 0.09
BB	0.04	<b>0.0196</b>	<b>0.0168</b>	<b>0.0036</b>
Bb	0.32	<b>0.1568</b>	<b>0.1344</b>	<b>0.0288</b>
bb	0.64	<b>0.3136</b>	<b>0.2688</b>	<b>0.0576</b>

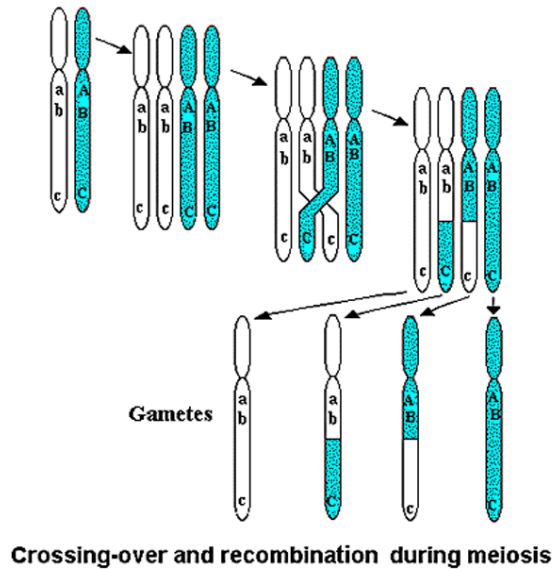
		AB 0.14	Ab 0.56	aB 0.06	ab 0.24
AB	0.14	<b>0.0196</b>	<b>0.0784</b>	<b>0.0084</b>	<b>0.0336</b>
Ab	0.56	<b>0.0784</b>	<b>0.3136</b>	<b>0.0336</b>	<b>0.1344</b>
aB	0.06	<b>0.0084</b>	<b>0.0336</b>	<b>0.0036</b>	<b>0.0144</b>
ab	0.24	<b>0.0336</b>	<b>0.1344</b>	<b>0.0144</b>	<b>0.0576</b>

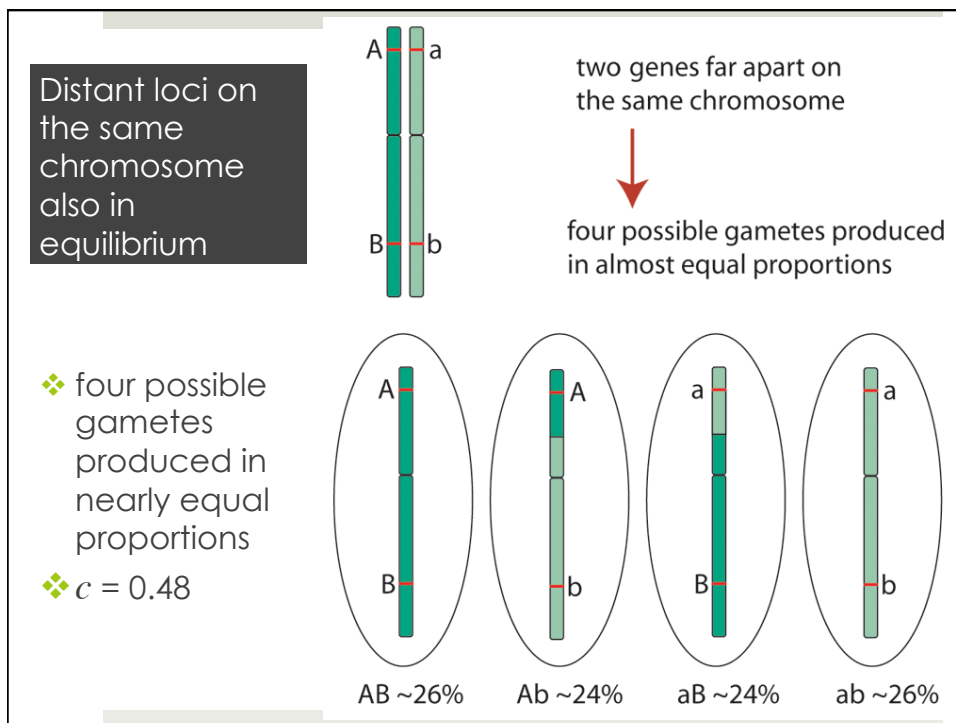
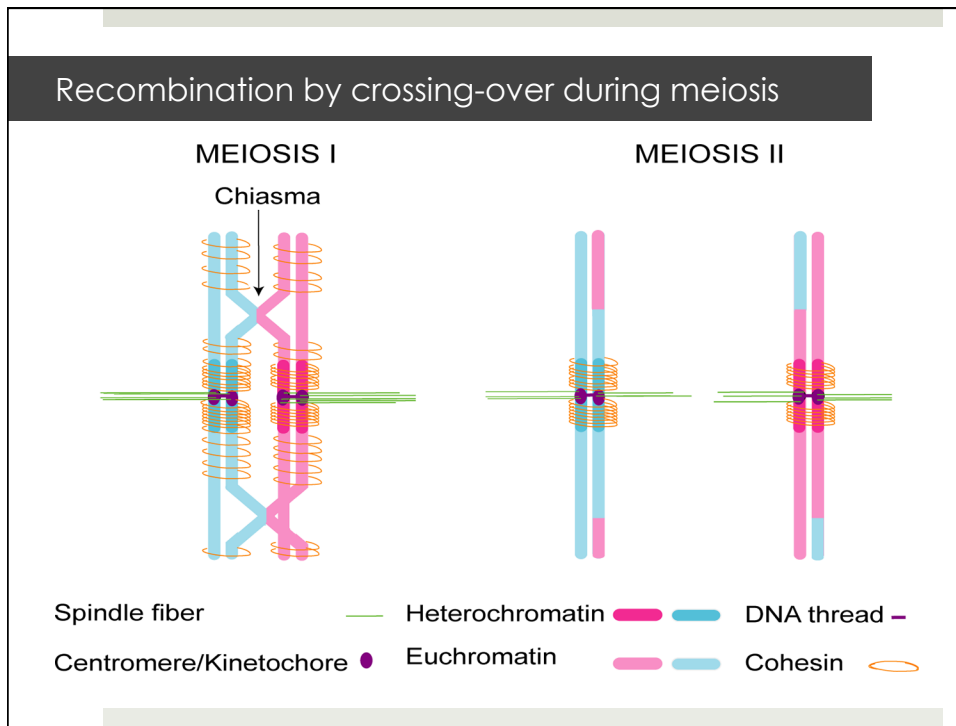
Genes on human chromosomes 18 and 19

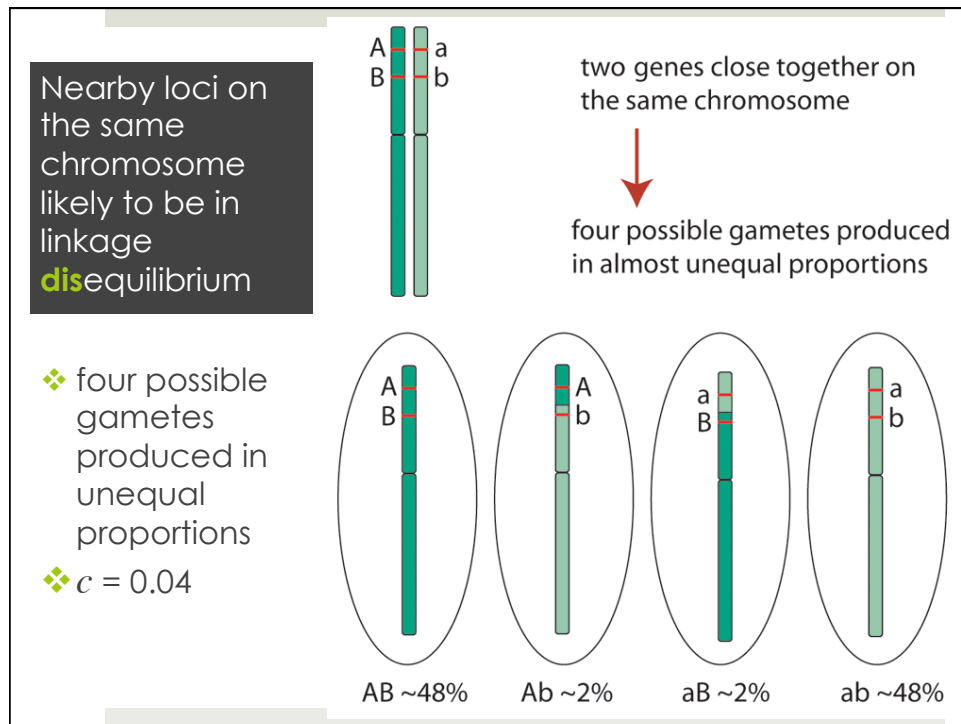
**Figure 2.17** Maps for human chromosomes 18 (left) and 19 (right) showing chromosome regions, the physical locations of identified genes and open reading frames (labeled orf) along the chromosomes, and the names and locations of a subset of genes. Chromosome 18 is about 85 million bp and chromosome 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  $(1-c)/2$
  - ❖  $ab$  with frequency  $(1-c)/2$
  - ❖  $Ab$  with frequency  $c/2$
  - ❖  $aB$  with frequency  $c/2$
- ❖  $Ab/aB$  genotype produces gametes:
  - ❖  $AB$  with frequency  $c/2$
  - ❖  $ab$  with frequency  $c/2$
  - ❖  $Ab$  with frequency  $(1-c)/2$
  - ❖  $aB$  with frequency  $(1-c)/2$

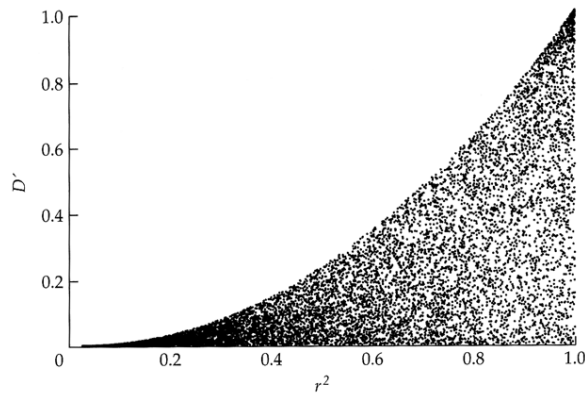
## Linkage Disequilibrium

- ❖  $D = f_{AB}f_{ab} - f_{Ab}f_{aB}$ 
  - ❖ if more  $AB$  and  $ab$  chromosomes, then positive
  - ❖ if more  $Ab$  and  $aB$  chromosomes, then negative
  - ❖ positive or negative is essentially arbitrary
- ❖ Note: linkage disequilibrium also possible without physical linkage (i.e., genes on different chromosomes)
  - ❖ = "gametic phase disequilibrium"
  - ❖ population admixture, selection, selfing

## Linkage Disequilibrium

- ❖ scaling  $D$  w.r.t. allele frequencies
- ❖  $D_{min}$  = the larger of  $-f_A f_B$  and  $-f_a f_b$
- ❖  $D_{max}$  = the smaller of  $f_A f_b$  and  $f_a f_B$
- ❖ 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" ( $r^2$  in Neilsen & Slatkin) ranges from 0 to 1
- ❖ \*\*Please note!  $\rho^2$  (also written  $r^2$ ) is a completely different measure from the frequency of recombination,  $c$ , which is sometimes written as  $r$

## 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

$$\diamond f_{AB}' = (1-c) \times f_{AB} + c \times f_A f_B$$

- ❖ Subtracting  $f_A f_B$  from both sides...

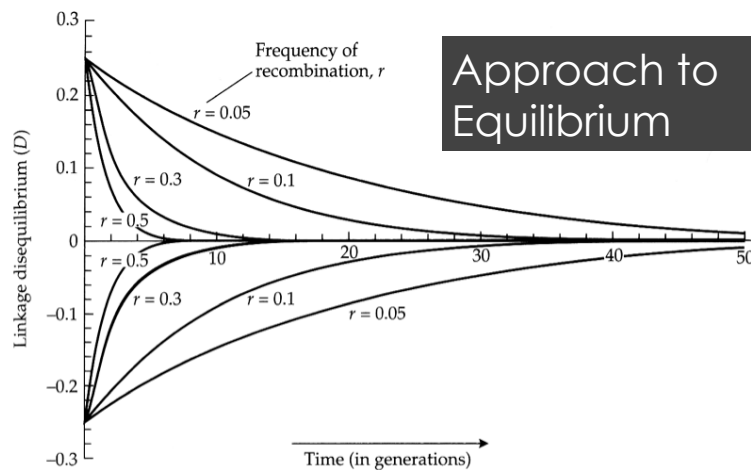
$$\diamond \underbrace{f_{AB} - f_A f_B}_{\text{recombinant AB gametes/chromosomes}} = (1-c) \underbrace{(f_{AB} - f_A f_B)}_{\text{new, recombinant AB gametes/chromosomes}} + c \underbrace{(f_A f_B - f_A f_B)}_{0}$$

- ❖ in the  $n$ th generation,  $D_{t(n)} = (1-c)^n D_{t(0)}$



## Change in D w/ Random Mating

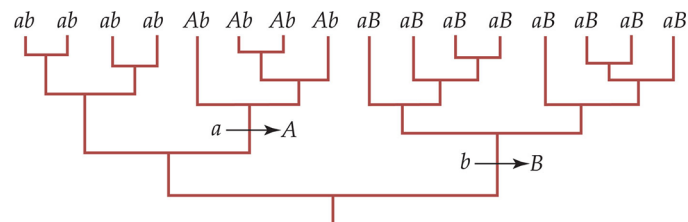
Genotype	Frequency	Gametes Produced			
		AB	Ab	aB	ab
AB/AB	$f_{AB}^2$	1	0	0	0
AB/Ab	$2f_{AB}f_{Ab}$	$\frac{1}{2}$	$\frac{1}{2}$	0	0
AB/aB	$2f_{AB}f_{aB}$	$\frac{1}{2}$	0	$\frac{1}{2}$	0
AB/ab	$2f_{AB}f_{ab}$	$(1-c)/2$	$c/2$	$c/2$	$(1-c)/2$
Ab/Ab	$f_{Ab}^2$	0	1	0	0
Ab/aB	$2f_{Ab}f_{aB}$	$c/2$	$(1-c)/2$	$(1-c)/2$	$c/2$
Ab/ab	$2f_{Ab}f_{ab}$	0	$\frac{1}{2}$	0	$\frac{1}{2}$
aB/aB	$f_{aB}^2$	0	0	1	0
aB/ab	$2f_{aB}f_{ab}$	0	0	$\frac{1}{2}$	$\frac{1}{2}$
ab/ab	$f_{ab}^2$	0	0	0	1



**FIGURE 2.10** Linkage disequilibrium between genes gradually disappears when mating is random, provided there is no countervailing force building it up. The rate of approach to linkage equilibrium depends on the recombination frequency between the genes. The disappearance of linkage disequilibrium is gradual even with free recombination ( $r = \frac{1}{2}$ ). In these examples, the frequencies of both alleles at both loci equal  $\frac{1}{2}$ , and the initial linkage disequilibrium is either at its maximum ( $D = 0.25$ ) or minimum ( $D = -0.25$ ) value, given these allele frequencies.

## History of Linkage

- ❖ within a single random breeding population, linkage disequilibrium reflects the history of mutations
- ❖ consider the case with no recombination and starting with only  $ab$  gametes...



## 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:  $p_A = 0.95, p_B = 0.95$

$$D' = 0.81$$

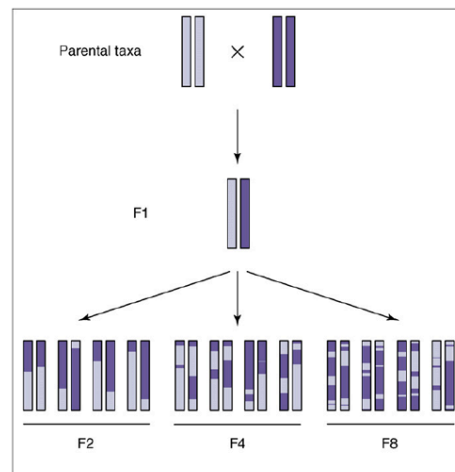
$$r^2 = 0.66$$



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

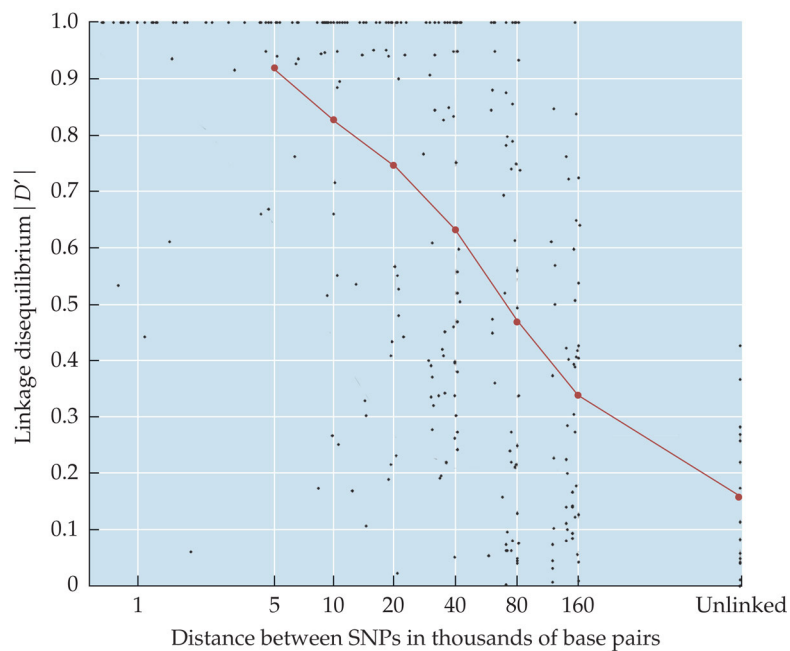
Chromosome	Frequency	Subpopulation 1	Subpopulation 2	Equal mixture
<i>AB</i>	$P_{AB}$	0.0025	0.9025	0.4525
<i>Ab</i>	$P_{Ab}$	0.0475	0.0475	0.0475
<i>aB</i>	$P_{aB}$	0.0475	0.0475	0.0475
<i>ab</i>	$P_{ab}$	0.9025	0.0025	0.4525
$D = P_{AB}P_{ab} - P_{Ab}P_{aB}$		0	0	0.2025
$D_{min}$		-0.0025	-0.0025	-0.2500
$D_{max}$		0.0475	0.0475	0.2500

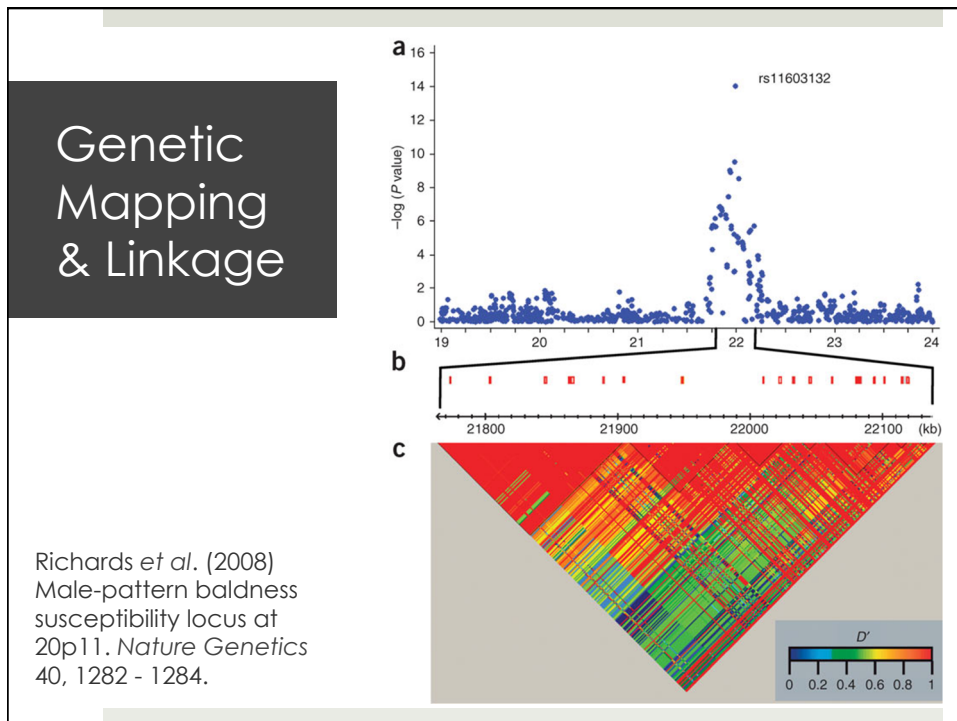
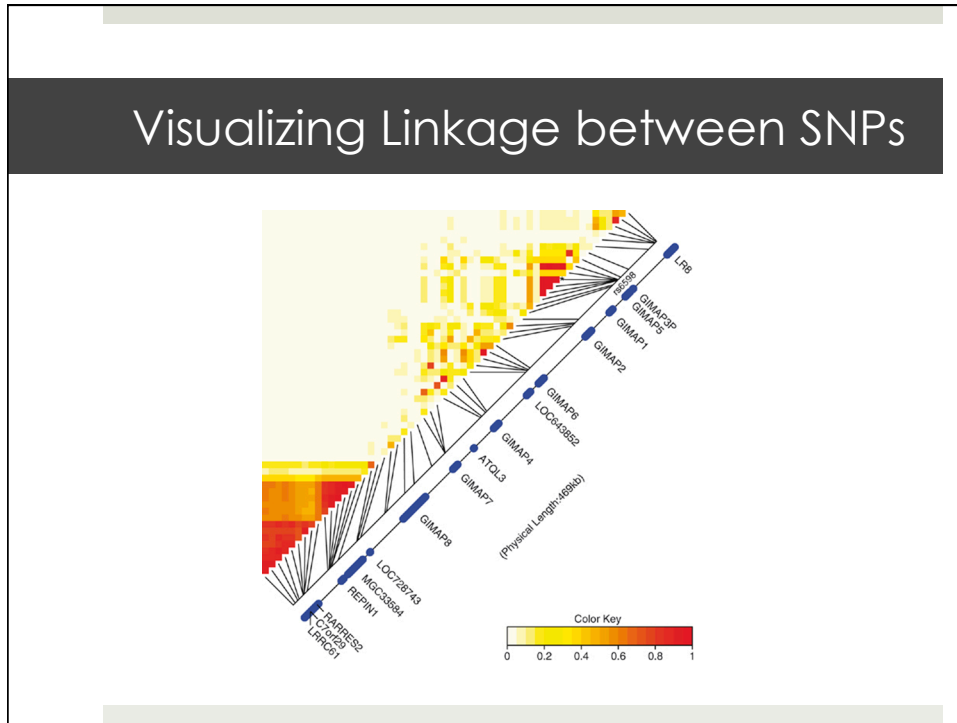
## Recombination in an admixed population



## Length of haplotypes blocks

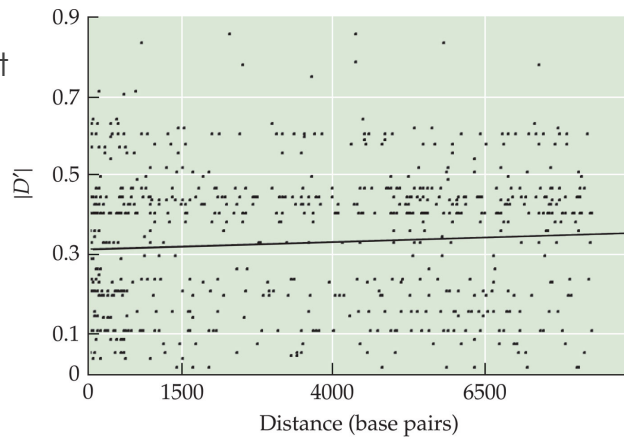
- ❖ Key point! — regions exhibiting linkage disequilibrium are generally quite short in large natural populations
  - ❖ e.g., a few kb (thousand nucleotides) in *Drosophila* and maize
  - ❖ a few tens of kb in humans and dogs
  - ❖ length varies with local recombination rates and timing and strength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





## mtDNA

- ❖ What is the line flat?
- ❖ Why is  $D'$  not equal to 1?



## 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.

