Population Genetics

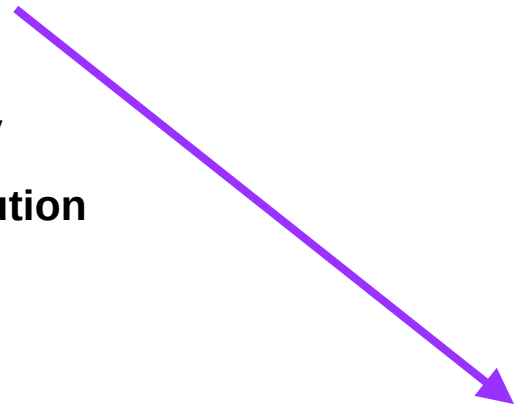
Sequence Diversity

Molecular Evolution

Physiology

Quantitative Traits

Human Diseases



Bioinformatics problems in medicine related to physiology and quantitative traits

Note: Genetics including population genetics was a serious science before DNA was clearly known to be the hereditary material and before any protein or gene sequences were known.

Books:

Molecular Evolution : A Phylogenetic Approach (1998) by Roderic D. M. Page, Edward C. Holmes.

Blackwell Science Inc; ISBN: 0865428891

Easy read, nice introduction

Molecular Evolution (1979) by Wen-Hsiung Li

Sinauer Assoc; ISBN: 0878934634

More Detailed than Page and Holmes

Human Molecular Genetics (1999) by Tom Strachan, Andrew P. Read

Wiley-Liss; ISBN: 0471330612

Wonderful human genetics text

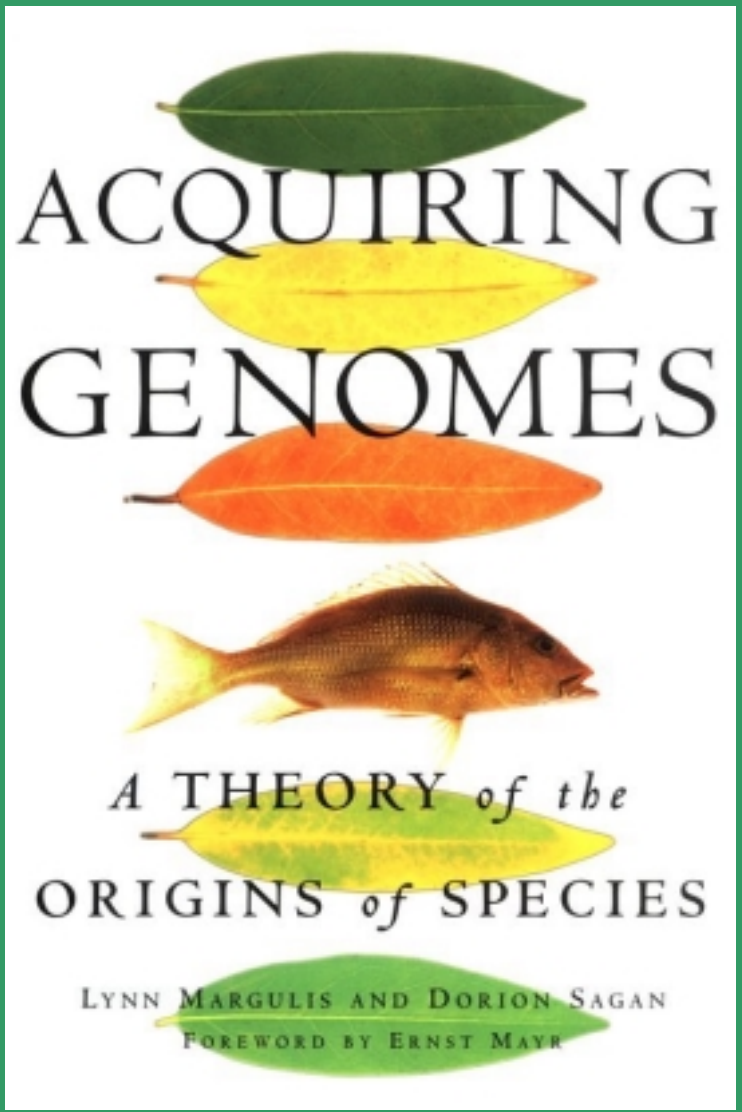
Statistical Genomics: Linkage, Mapping, and QTL Analysis (1997) by Ben-Hui Liu

CRC Press; ISBN: 0849331668

Specialized for the serious student

Important web site not previously mentioned in the course is OMIM

Online Medelian Inheritance in Man **http://www3.ncbi.nlm.nih.gov/omim/**



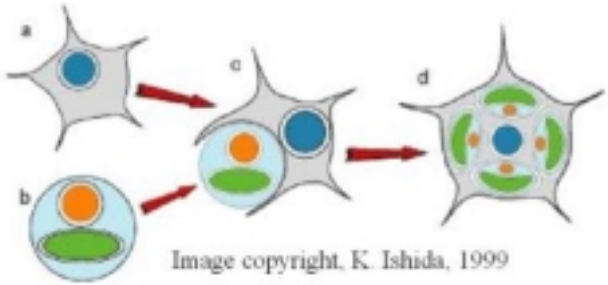
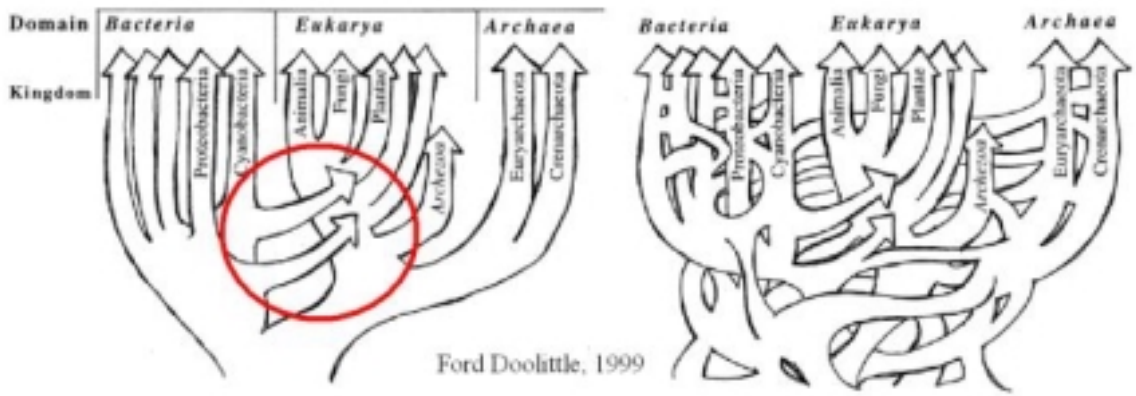
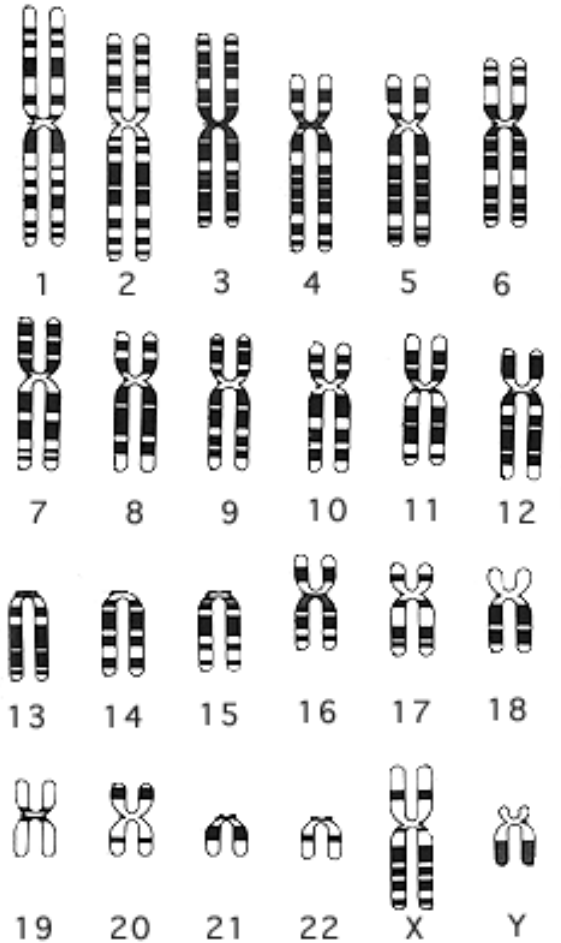


Image copyright, K. Ishida, 1999

Endosymbiosis ... *two organisms become one*



Ford Doolittle, 1999



Human genome has been sequenced

But the human genome contains many polymorphisms

There is no one sequence that is "The Human Genome"

size is 3300 Mbases. This is haploid size. 23 pieces of DNA.

About 1400 genes/chromosome (30000/22)

We may consider two kinds of maps of the genome

physical map, Genome project

genetic map, linkage studies

- Mitochondrial genome

Alleles are different forms of a gene or DNA sequence that can exist at a single locus.
For example, there are three common alleles of Apolipoprotein E, a lipid binding protein found in the blood

```
3901 gcgcaggccc ggctgggcgc ggacatggag gacgtgtgcg gccgcctggt gcagtaccgc apo E3
     .......... .......... .......... ......t... .......... .......... apo E2
     .......... .......... .......... ......c... .......... .......... apo E4

4021 cacctgcgca agctgcgtaa gcggctcctc cgcgatgccg atgacctgca gaagcgcctg apo E3
     .......... .......... .......... .......... ....t..... .......... apo E2
     .......... .......... .......... .......... ....c..... .......... apo E4
```

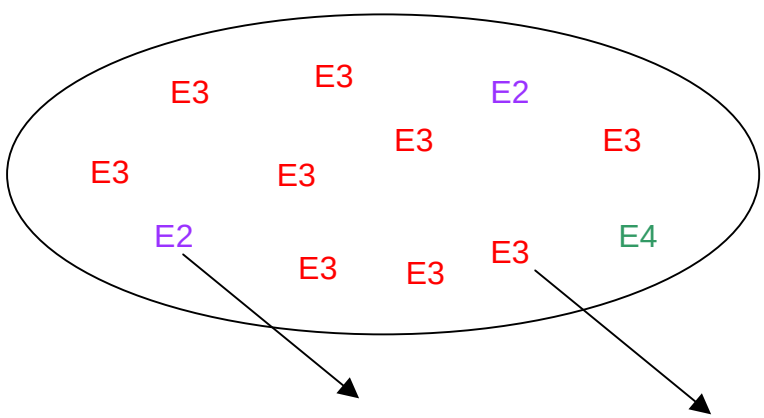
heterozygosity

$$h = 1 - \sum_{i=1}^{n} x^2$$

h = 0.364

E4 polymorphism causes arginine instead of cysteine at amino acid position 112

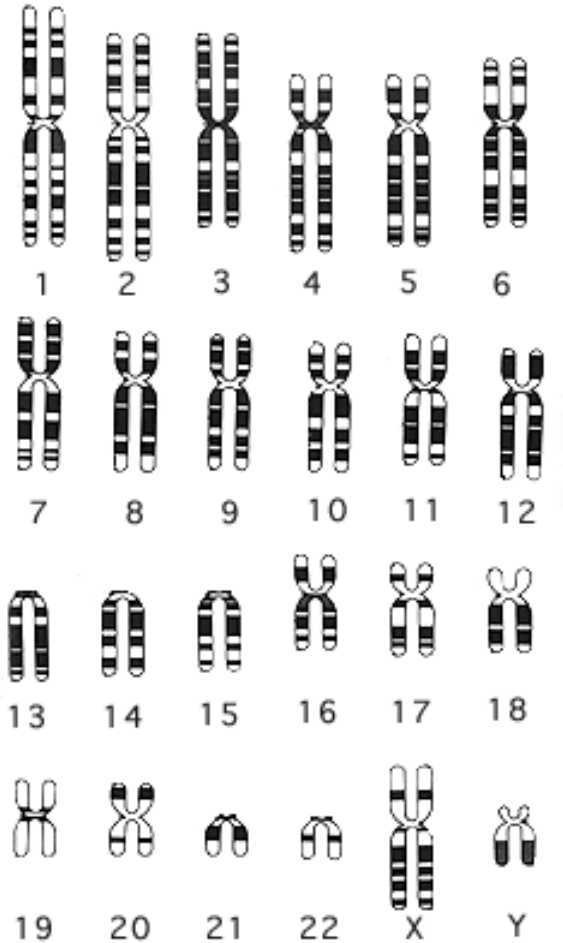
E2 polymorphism causes cysteine rather than arginine at amino acid position 158.



The most common form, E3, occurs in approximately 78% of the population, while E4 has a frequency of 15%, and E2 a frequency of 7%.

Each individual inherits two alleles from the population

Note: here alleles are identical with SNPs (single nucleotide polymorphisms)



Human genome size is 3300 Mbases. This is haploid size. 23 pieces of DNA.

About 1400 genes/chromosome (30000/22)

We may consider two kinds of maps of the genome

physical map, Genome project

genetic map, linkage studies

• Human Genome	3,000,000,000 bp
• Chromosome	150,000,000 bp
• Gene (avg.)	50,000 bp
• Coding sequence	3,000 bp
• Unit of genetic code	3 bp
• Mutation	1 bp

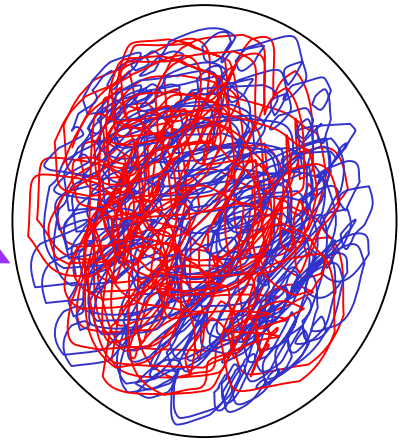
- 1 cMorgan ≅ 1 Mb males ≅ 0.7 Mb females
- A typical chromosome, about 100 Mb DNA

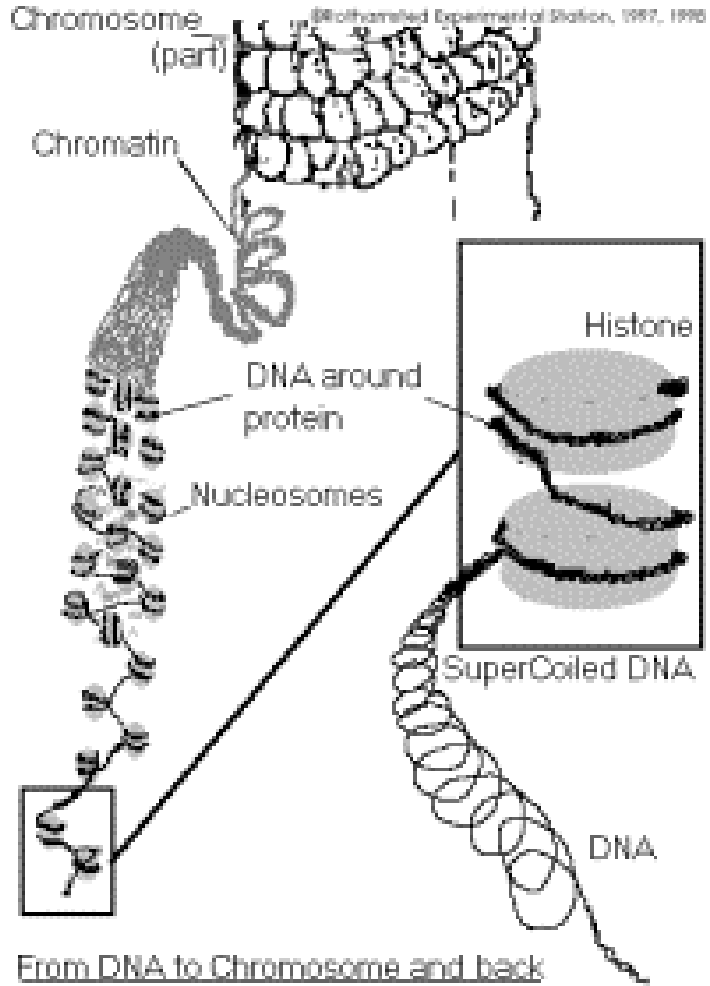
● Mitochondrial genome

Chromosomes wind up tightly during cell division at Metaphase

APOE -E2_ _APOE -E3

A human cell nucleus contains 6 feet of DNA, 46 very long molecules



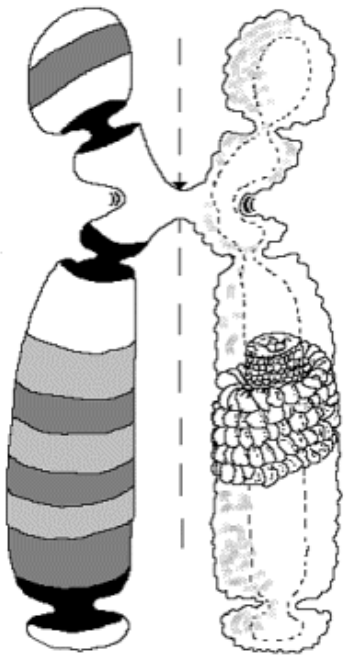


From DNA to Chromosome and back

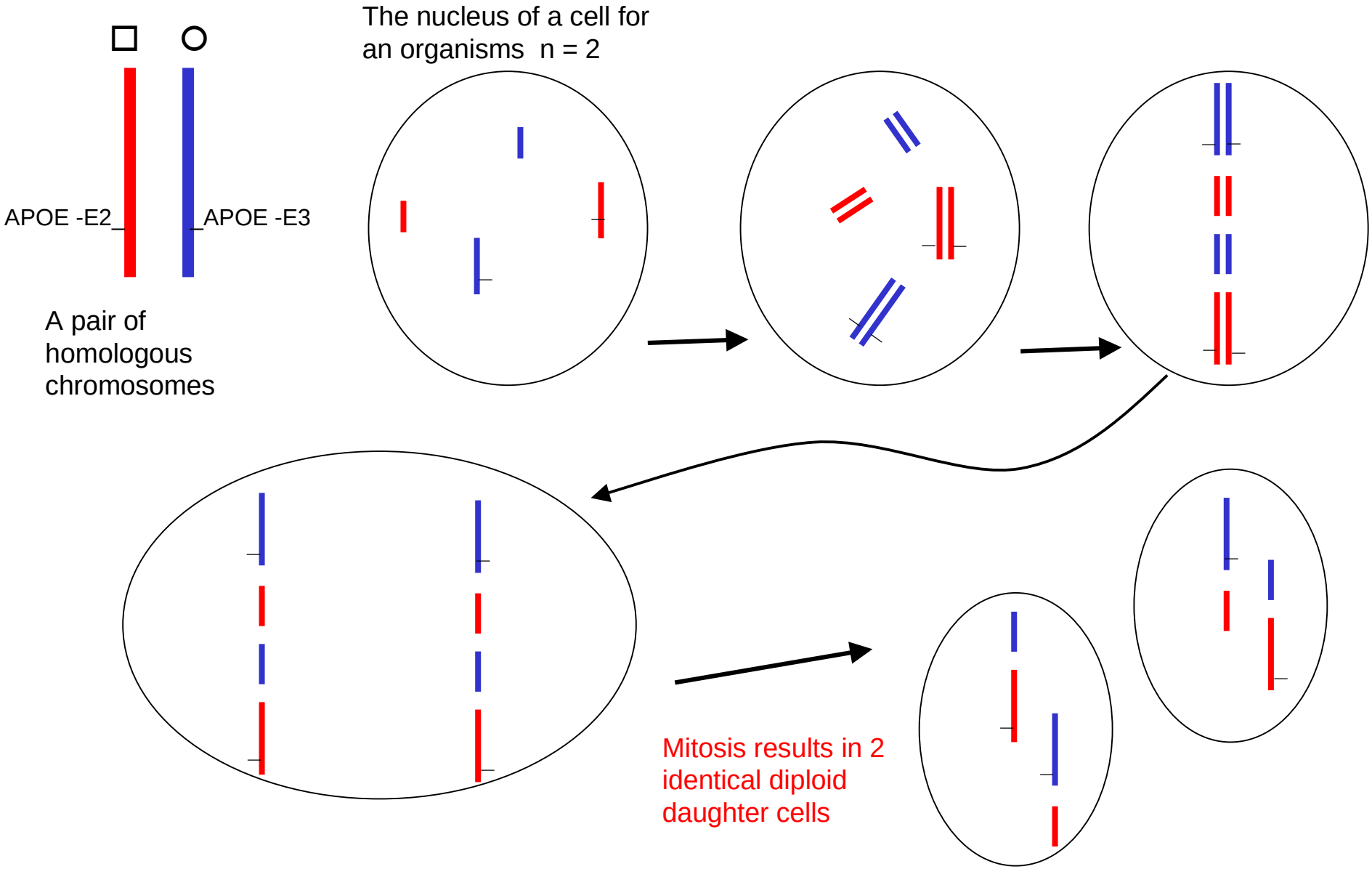
DNA coils up after doubling to ensure that each daughter cell receives one copy of each strand

_ _APOE -E3

Each chromosome about 1.5 inches of DNA coiled to length about 1 micron

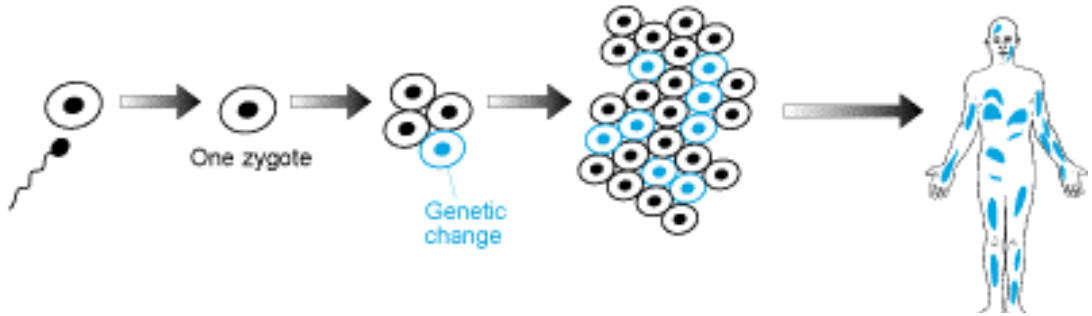

Metaphasic Chromosome

In somatic cell division the two homologous chromosomes do not associate: Mitosis



Why do gene sequences differ ? Mutations Happen

Often at replication



How frequent are mutations?

Mammalian globin pseudogene 10^{-9} nucleotide sub/(site year)

Human genome 3,000 Mb but each gene is only 0.0017 Mb coding sequence

1.7×10^{-6} - 1.7×10^{-8} per gene per cell division

A human adult has about 10^{14} cells requiring about 2×10^{14} cell divisions.

Mutations are not uncommon.

Germ line mutations are important.

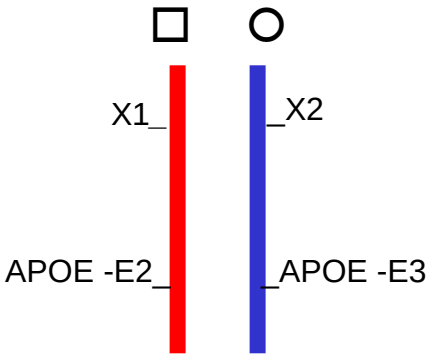
Males mutation rates higher than female.

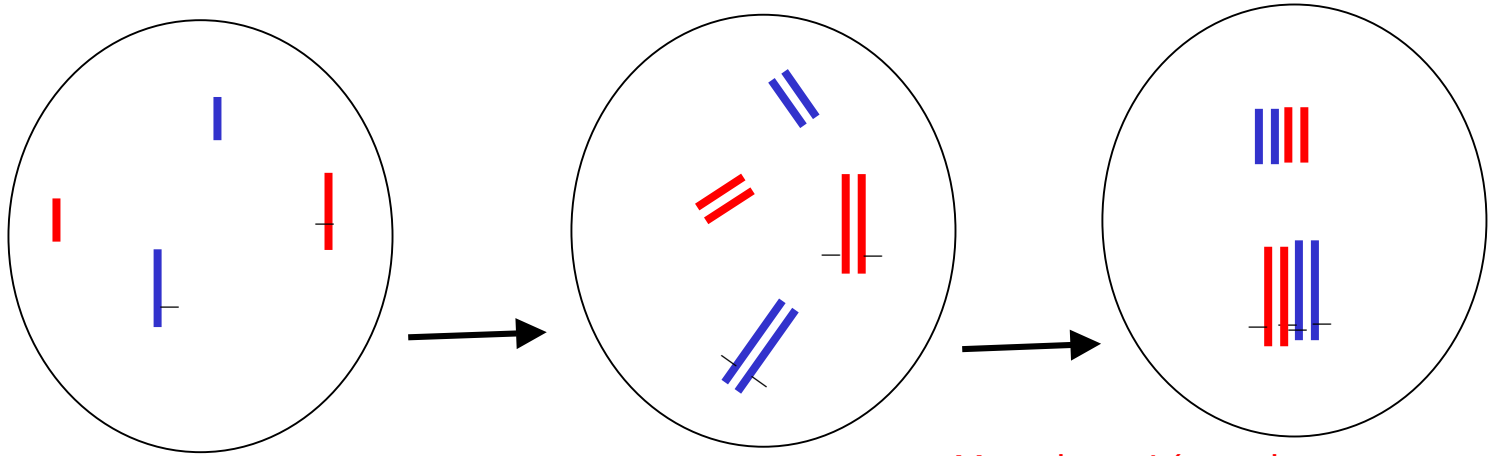
Cell divisions to gametes, Female 24-30, Male = 310 + 23/year over 25

New mutations may be important for study of disease but….

Evolution is a process of change in the genetic makeup of populations.

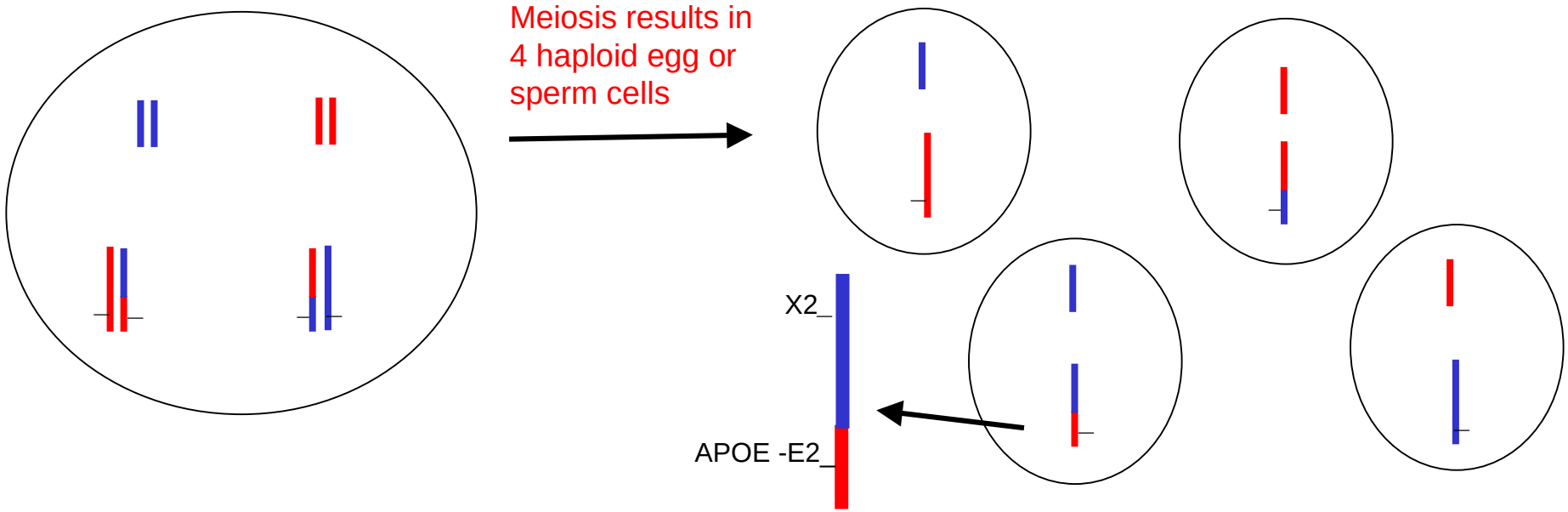
To produce gametes (eggs and sperm) the two homologous chromosomes do associate. Cells divide by Meiosis

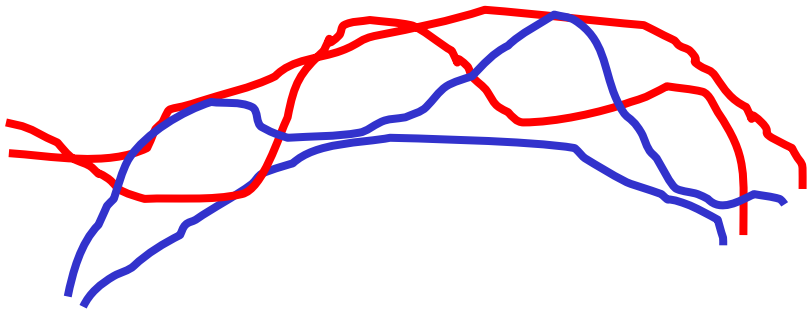




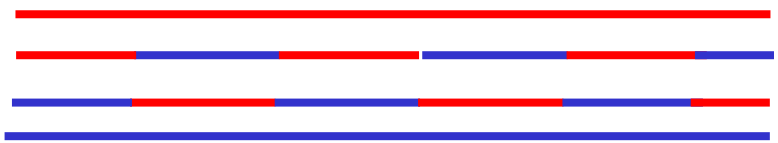
Metaphase I (crossing over between homologous chromosomes may occur here)

Meiosis results in 4 haploid egg or sperm cells



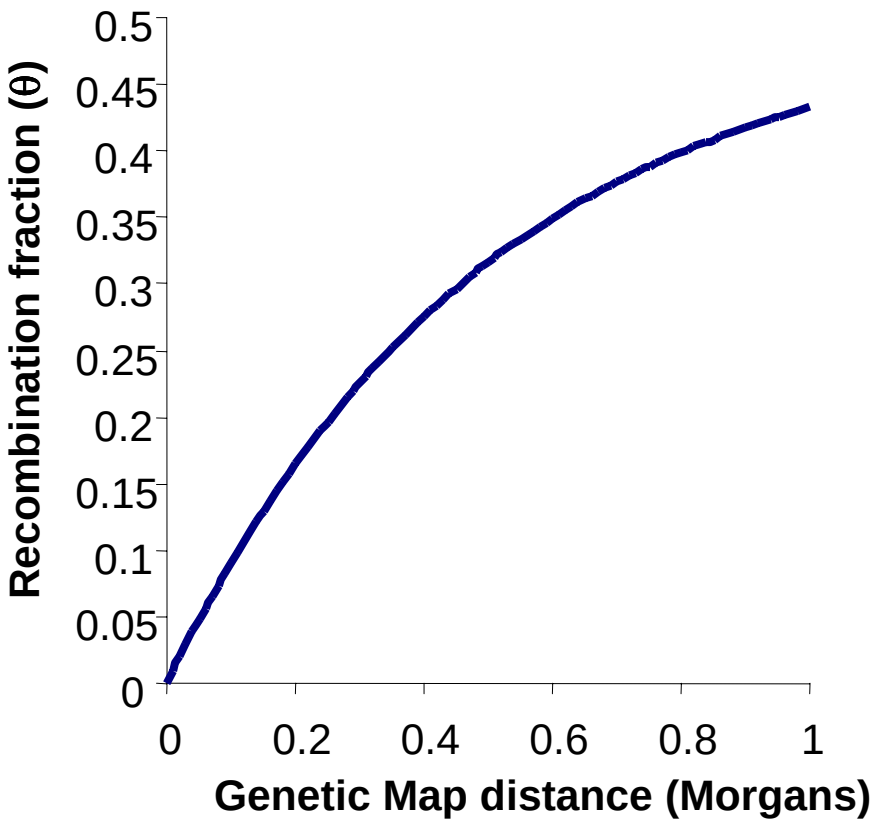


Estimates of crossing over for the 23 pairs of human chromosomes (except Y) about 50/meiosis ≅ 2/chromosome. Thus most genes far apart on chromosomes are not tightly linked.

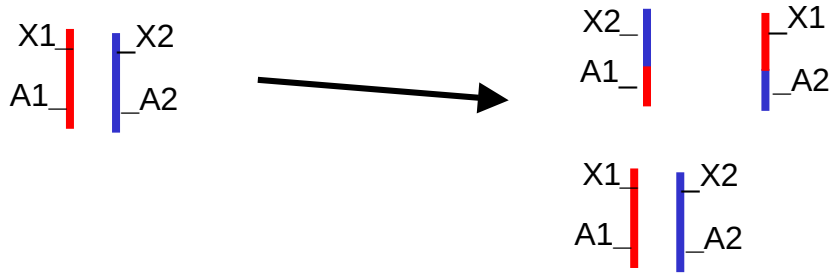


Haldane, J.B.S. 1919. "The combination of Linkage values and the calculation of distances between the loci of linked factors." J. Genet. 8:299-309.

Haldane map function



Recombination fraction = (θ)

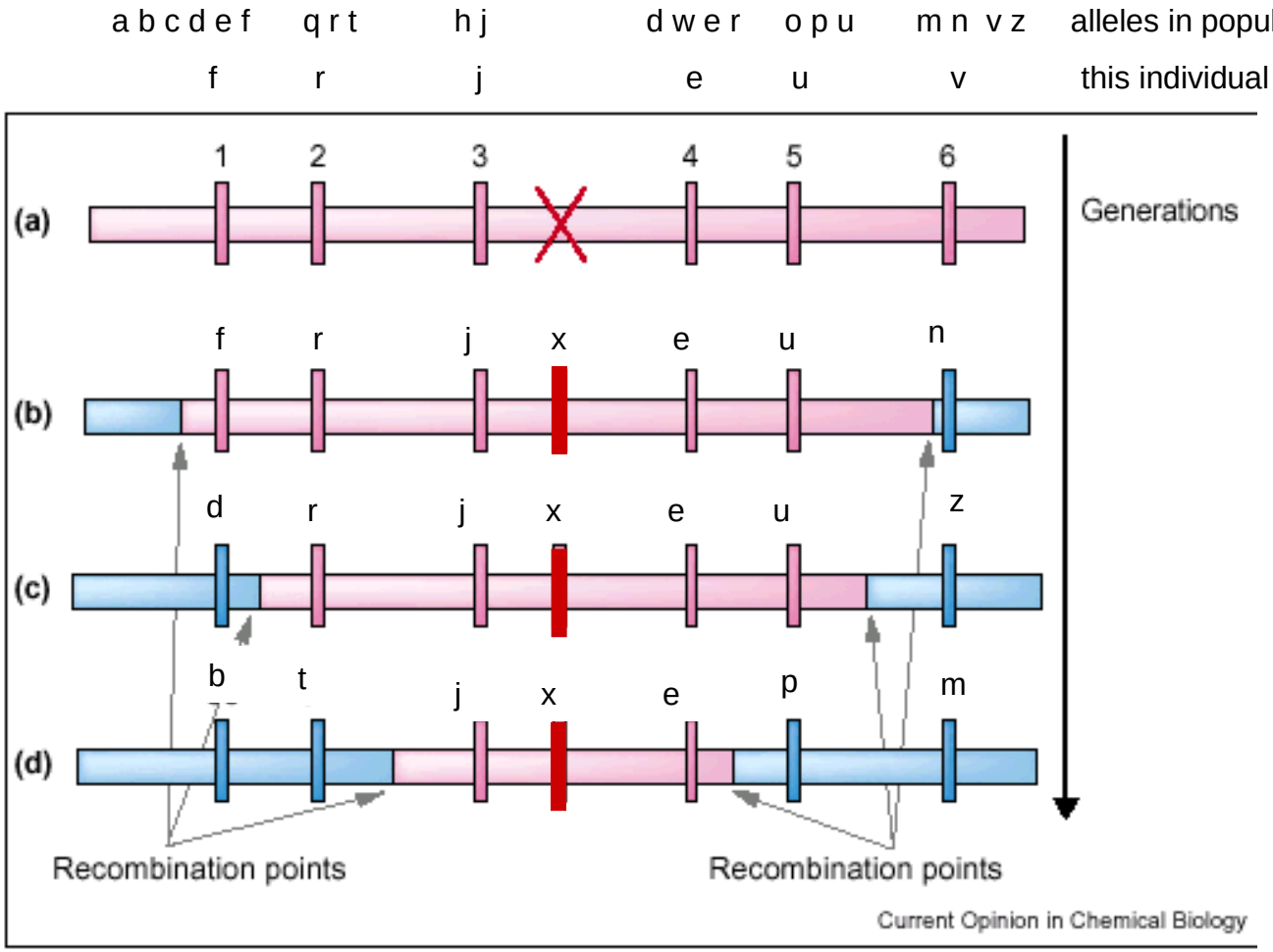


$$\theta = \frac{1 - e^{-2m}}{2}$$

$$m = -1/2 \ln(1-2\theta)$$

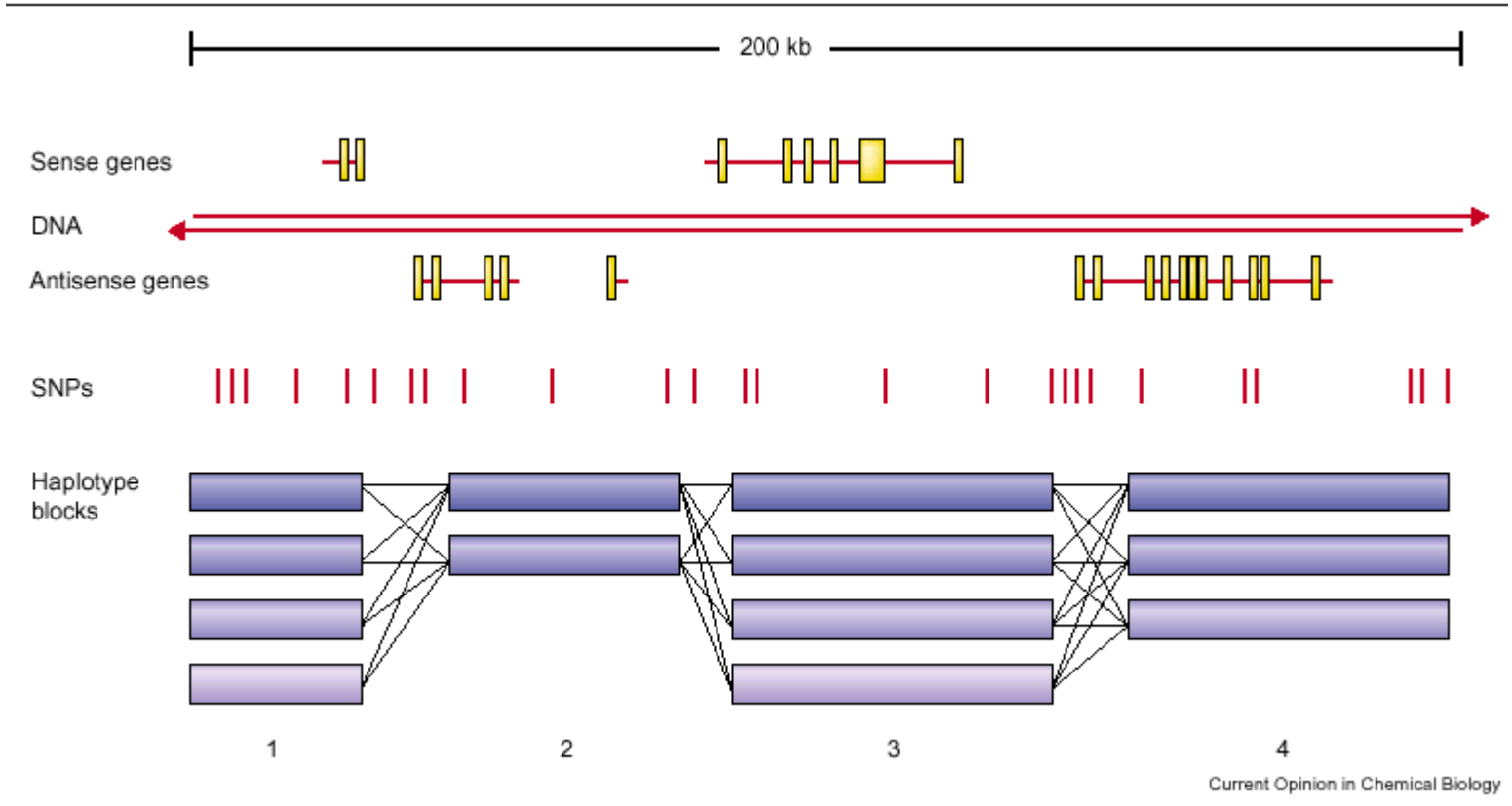
Note: Map distances are additive recombination fractions are not additive

(Morgans) = Expected number of crossovers per meiosis



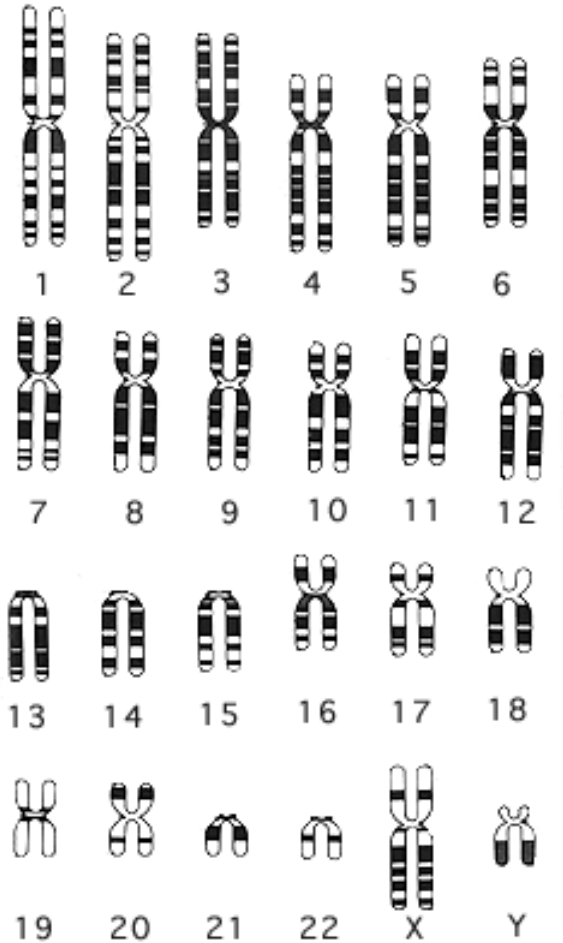
Description of LD. (a) Situation shortly after a mutation event (shown as a red cross) has generated a new SNP. At this point, the new allele is in complete LD with other marker alleles 1–6 on the same chromosome. (b,c) The situation changes as generations pass by. The extent of LD decays because of recombination between chromosomes. The recombination shuffles marker alleles and they loose their association with mutation allele (x). (d) The LD between mutation (x) and nearby markers is observed only in a short region around the mutation.

High-density genotyping and linkage disequilibrium in the human genome using chromosome 22 as a model, Current Opinion in Chemical Biology, Volume 6, February 2002, Pages 24-30



Haplotype blocks. This figure illustrates the characteristic pattern of SNPs in the human genome. Within haplotype blocks, the diversity is low; 2–4 variants are typically representing 90–95% of population. Haplotype blocks are interrupted with regions of frequent recombination, which are responsible for ‘shuffling’ of blocks between chromosomes.

High-density genotyping and linkage disequilibrium in the human genome using chromosome 22 as a model, Current Opinion in Chemical Biology, 6, 24-30 February 2002, Pages 24-30



Human genome size is 3300 Mbases. This is haploid size. 23 pieces of DNA.

About 1400 genes/chromosome (30000/22)

We may consider two kinds of maps of the genome

physical map, Genome project

genetic map, linkage studies

• Human Genome	3,000,000,000 bp
• Chromosome	150,000,000 bp
• Gene (avg.)	50,000 bp
• Coding sequence	3,000 bp
• Unit of genetic code	3 bp
• Mutation	1 bp

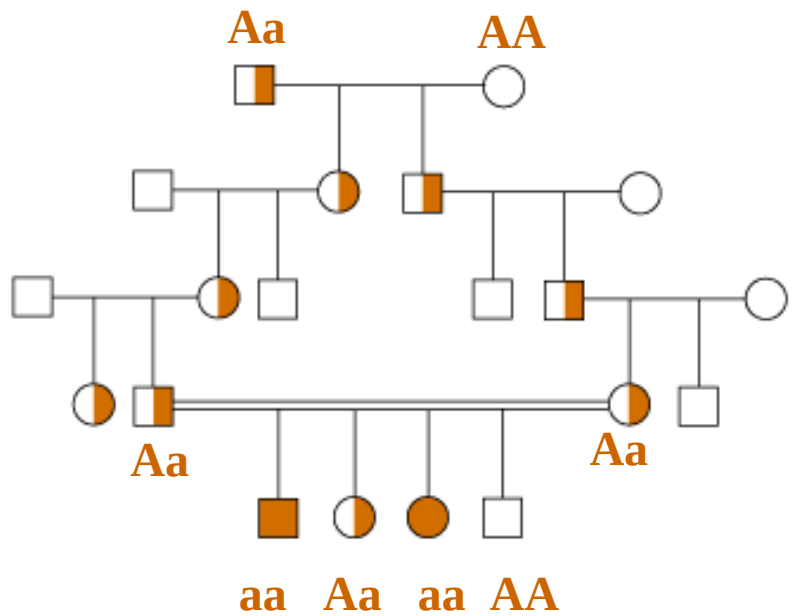
- 1 cMorgan $\cong$ 1 Mb males $\cong$ 0.7 Mb females
- A typical chromosome, about 100 Mb DNA

● Mitochondrial genome

Single gene defects and multiple gene defects cause disease.

- Cystic fibrosis
- Huntington disease
- Early onset breast cancer (BRCA1, BRCA2)
- Alzheimer disease (chr14, chr1)
- Maturity-onset diabetes of the young (MODY) (chr12)

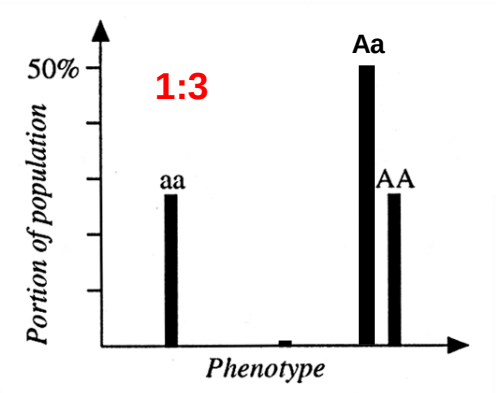
Class	Example	Comment
Chromosomal	Trisomy 21 (Down syndrome) (MIM 190685)	Prevalence increases with maternal age
	Chronic myeloid leukemia (MIM 151410)	Presence of Philadelphia chromosome
Monogenic	Familial hypercholesterolemia (MIM 143890)	Autosomal dominant; mutation in gene for LDL receptor
	Huntington disease (MIM 143100)	Autosomal dominant; diagnostic probe now available
	Cystic fibrosis (MIM 219700)	Autosomal recessive; majority of cases due to deletion of a Phe residue in a membrane protein regulating Cl^- transport (the CFTR protein)
	Sickle cell anemia (MIM 141900)	Autosomal recessive; mutation of Glu → Val at β^6 position of globin
	Phenylketonuria (MIM 261600)	Autosomal recessive; mutation of gene for Phe hydroxylase
	Duchenne muscular dystrophy (MIM 310200)	X-linked; affects synthesis of dystrophin.
	Hemophilia A (MIM 306700)	X-linked; affects synthesis of factor VIII (AHG)
Multifactorial	Ischemic heart disease	Complex genetics; study of DNA polymorphisms for lipoproteins holds promise for resolving genetic susceptibility
	Essential hypertension (MIM 145500)	A single-gene theory has its proponents

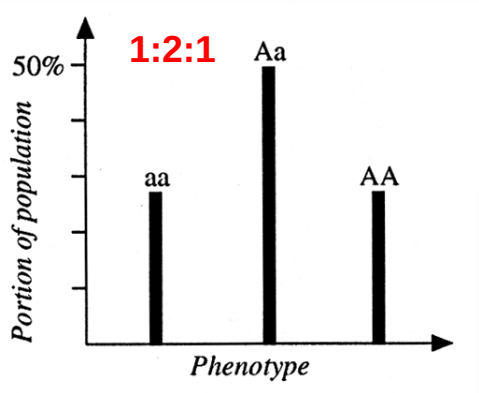


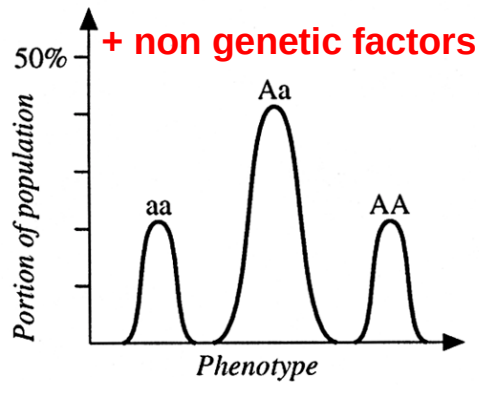
Autosomal recessive

Quantitative traits are important for many human diseases. These cannot be described by simple Mendelian ratios

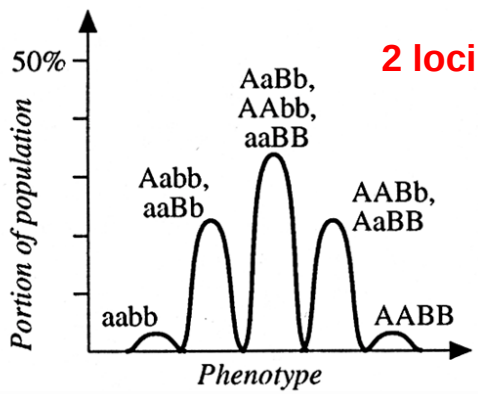
A) single genetic locus

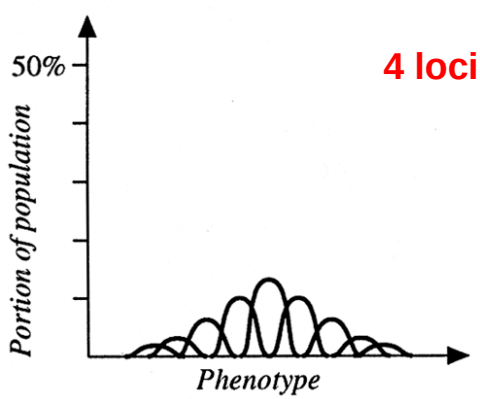


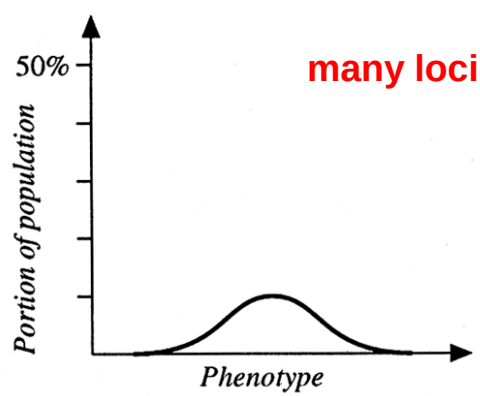




B) two and more unlinked genetic loci







Quantitative traits and disease: A grand challenge

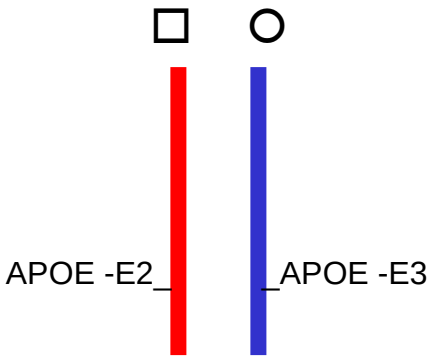
Evolution is a process of change in the genetic makeup of populations.

Natural populations of organisms are not identical in genetic makeup, they are polymorphic.

Evolution is a change in the frequency of alleles within the gene pool of a population from one generation to the next

Humans have 23 pairs of chromosomes one from each pair from each parent

3300 Mb of DNA is haploid amount



For each allele

An individual may be homozygous or heterozygous.

A Population may be polymorphic or not. (In non-polymorphic populations all individuals are homozygous).

Example: Apolipoprotein E

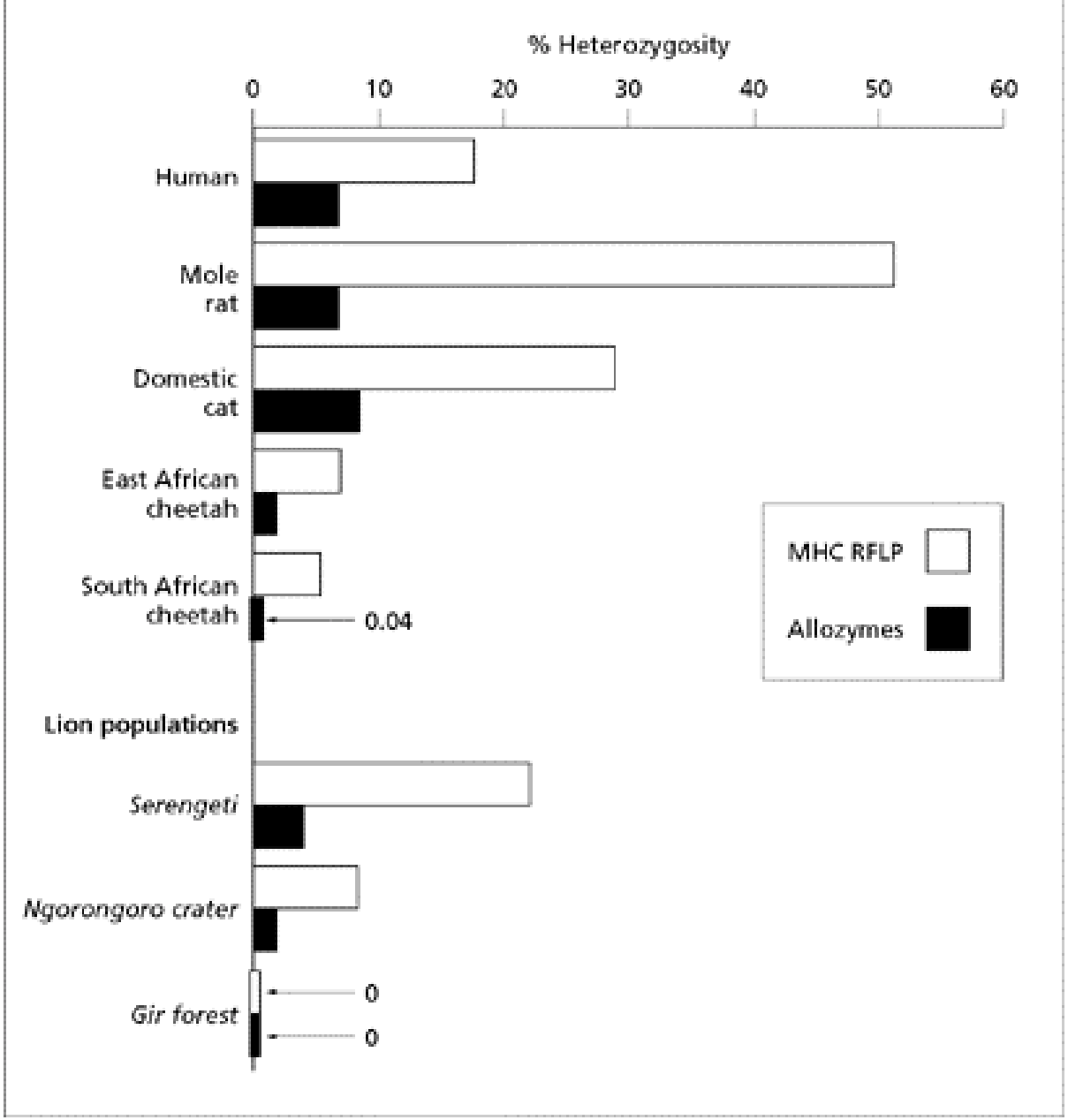
Polymorphisms in guppies.

Not all polymorphisms are known to be single nucleotide base changes.

Evolution acts on populations of organisms. Biodiversity is often advantageous within a species

However, most lab animals are highly inbred homozygous at all alleles.

At right SHR (spontaneously hypertensive) rat from Kyoto



Heterozygosity in different animal populations

How common are polymorphisms ? - a debate in the 1950s

How much genetic variation within a species

Classical school concern about genetic load - most mutations are expected to be deleterious

H.J. Muller predicted that only one locus (protein) in 1000 would be polymorphic

J.B.S. Haldane thought rate of aa substitutions in proteins low, genetic load of deleterious mutations

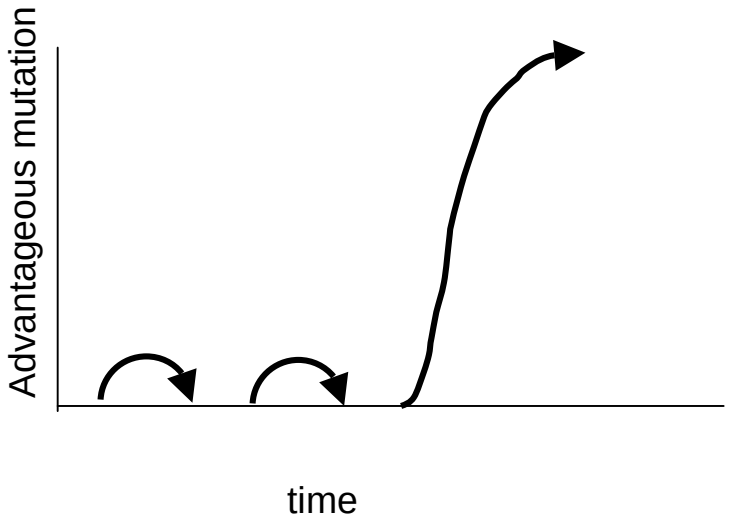
Balance School, polymorphisms high, beneficial to populations, heterozygous advantage

What the past 50 years have revealed

- No such thing as THE human genome sequence
- ~3 million differences between any two human genomes
- Common variant sequences (alleles) every 1,500 bp
- Estimated 2-3 common variants per gene (on average)

Why were Muller and Haldane wrong ?

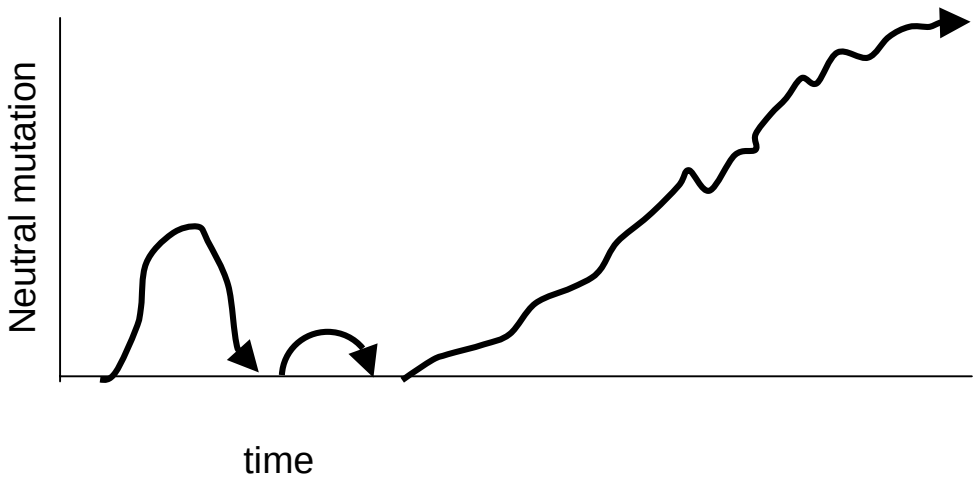
Why were Muller and Haldane wrong ?



Most mutations are not deleterious but neutral.

Neutral vs Selectionist debate 1968-69

Kiumura & King and Jukes



If a neutral mutation appears "A2" in a homozygous population where "A1" is wt

The probability, p, that A2 will replace A1 is:

$p \cong$ A2 frequency for A2 $<<1$

Mean time to fixation = 4 N (generations)

N is effective population size

HIV enters cells via a cooperation with human chemokine receptors. An important chemokine receptor for HIV-1 is CCR-5.

 a 32-bp deletion mutation in the coding region of the human *CCR5* gene has been found that results in an inactive protein. Homozygotes are highly resistant to HIV-1 infection.

The allele is found predominantly on a single haplotype, consistent with the notion that it arose once in the population.

TABLE 3 *CCR5-Δ32* frequency in worldwide populations

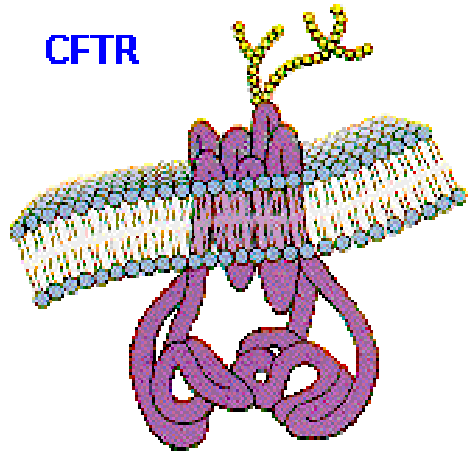
Population	+/+	+/Δ	Δ/Δ	Sum	Freq. Δ-*32*
Europe					
Mordvinian	58	28	0	86	0.16
Iceland	75	24	3	102	0.15
Sweden	251	74	10	335	0.14
Slovakia	22	8	0	30	0.13
Estonia	116	42	0	158	0.13
Russian	141	43	2	186	0.13
Ashkenazi	721	209	19	949	0.13
Denmark	387	104	7	498	0.12
.					
.					
.					
Punjab	33	1	0	34	0.01
East Siberian	214	5	0	219	0.01
India	99	1	0	100	0.01
Central Asian	106	1	0	107	0.00
Thailand	1154	7	0	1161	0.00
China	446	1	0	447	0.00
Bengal	25	0	0	25	0.00

But what about deleterious genes. They should be eliminated. Yet disease genes persist in the human population.

Why?

Heterozygous advantage, Over dominance

Example cystic fibrosis:

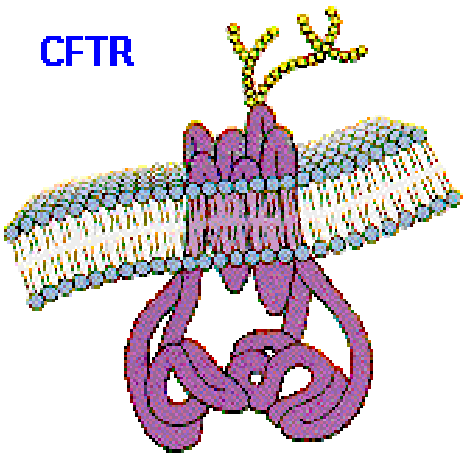


	Cystic Fibrosis in US						
U.S. population	frequency	frequency	gene frequency				
Group	affected	carriers	wt	mutant			
			A1	A2	A1A1	A1A2	A2A2
White	1/2,400	1/25	0.98	0.02	1-s	1	1-t
Hispanic	1/8,400	1/46					
Black	1/14,000	1/60			(s and t are + terms indicating		
Asians	1/89,000	1/150			fractional reduction in fitness)		

Equilibrium frequency of A1 = t/(s+t)

What is "s" for US White, Northern European, derived population:

Heterozygote may have resistance to typhoid fever.



An early example of finding a genetic disease by linkage.

Early markers, enzyme polymorphisims were not automatically associated with chromosomal location.

CF linked to paraoxonase but not helpful.

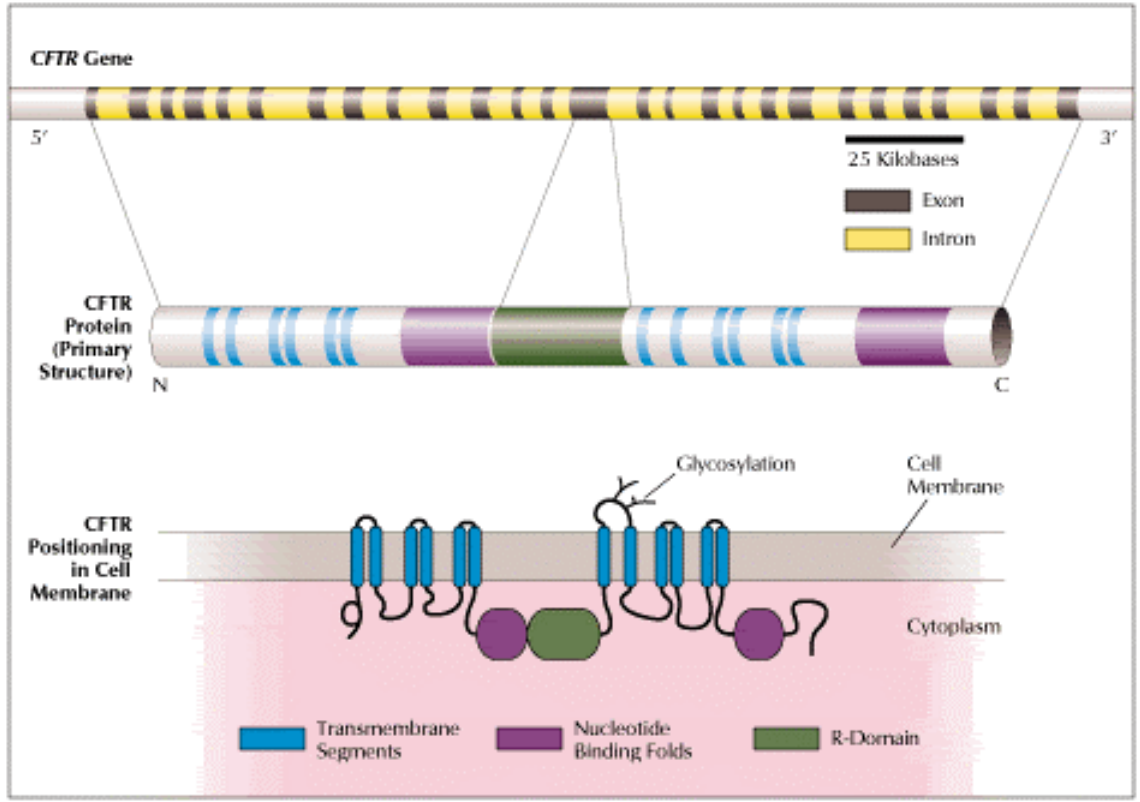


Figure 2. Human *CFTR* gene (top), identified in 1989 on the long arm of chromosome 7, uses 27 exons to specify a polypeptide consisting of 1,480 amino acids (middle). On the basis of its dual versions of a nucleotide-binding fold (NBF), the polypeptide has been classified as an ATP-binding-cassette protein (where "cassette" signifies a functional module). The polypeptide also has dual sets of six membrane-spanning segments. Unique to CFTR is a central region coded by the gene's longest exon. Suspected of having a regulatory function, it is called the R-domain. Analysis of the primary sequence of CFTR suggests that the only part of it protruding from a cell (bottom) is a short loop between transmembrane segments 7 and 8, which has attachment sites for two sidechains.

New mutations in humans are important for medical genetics but they do not become fixed in the current human population at a rapid rate (N is large).

Recombination occurring at meiosis may be more important than new mutations is producing genetic diversity in human population.

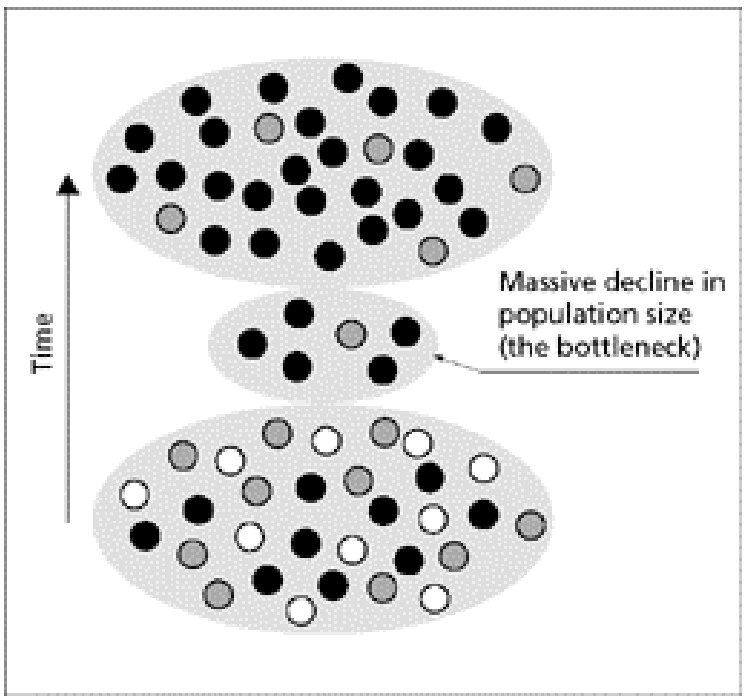
Just to review: Genetic approach to discovering mechanism of human disease

Find linkage of disease with marker on genetic map, narrow search, find the gene

Homo sapiens **A powerful idea: We are a very young species. 1 - 1.5 million years old but population bottleneck 200,000 years ago. We are closely related**

10,000 meiosis = 200,000 years

In Great Britain estimates predict that two unrelated people share common ancestor not more than 22 generations ago. Bottleneck 1500 AD, 5 x10^6 individuals. Thus only 44 meioses separate two unrelated people in this population



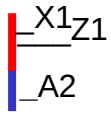
Page, Holmes
Molecular Evolution

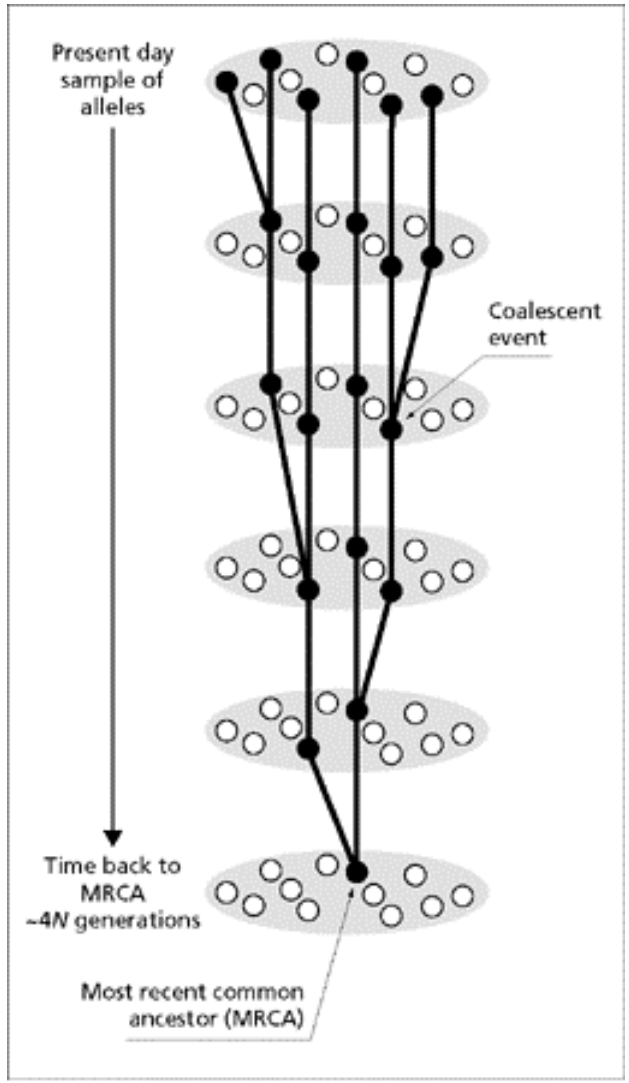
A British human geneticist may ask what is probability that two alleles with recombination frequency of 1% are preserved in linkage through 44 meioses

$(0.99)^{44} = 0.64$

so often closely spaced ancestral allele linkage is preserved in this population.

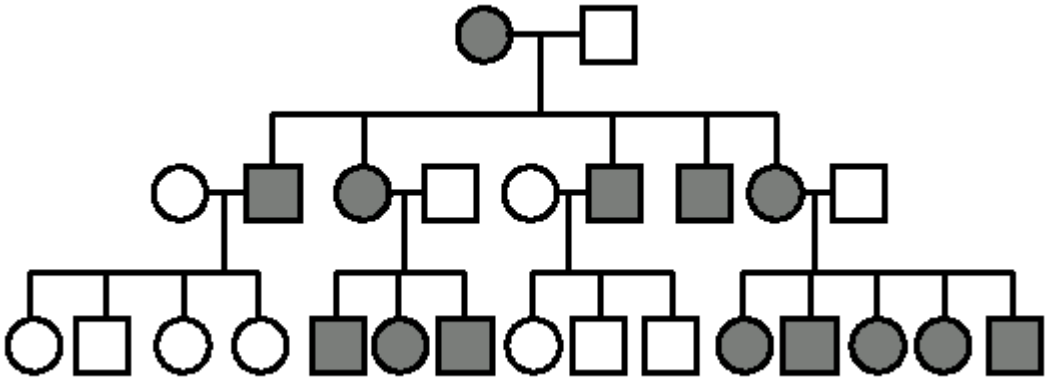
Linkage Disequilibrium

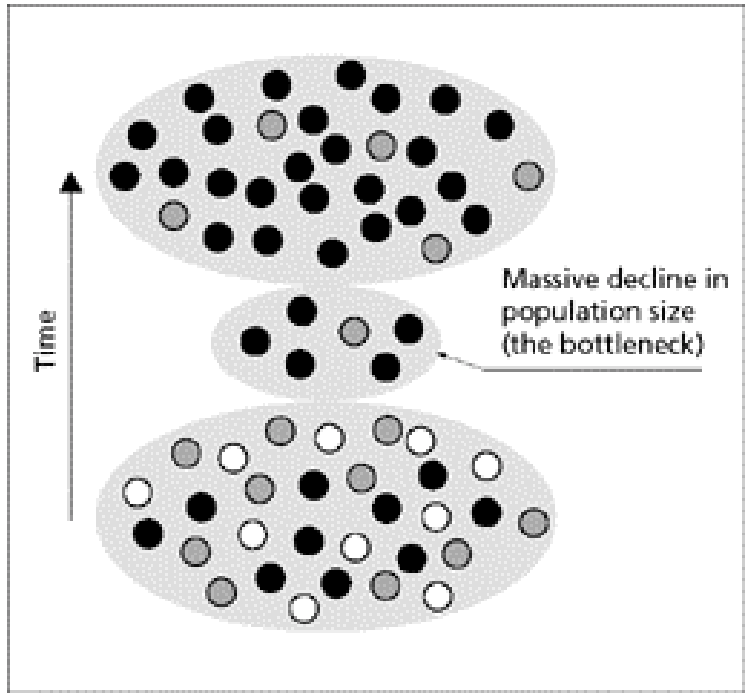




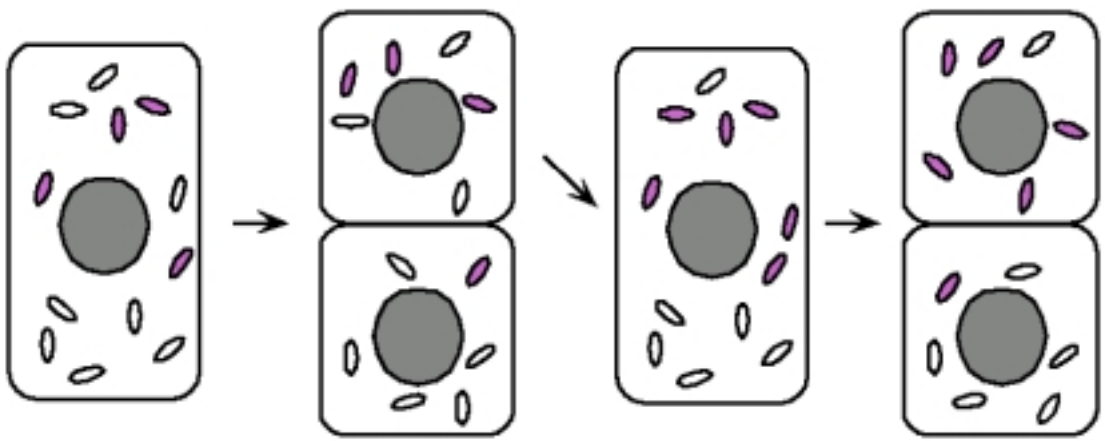
Page, Holmes
Molecular Evolution

Mitochondrial inheritance pattern





b Page, Holmes
Molecular Evolution



Mitochondrial diseases
homoplasmy from **heteroplasmy**

Common thread? Correlation between manifestation of disorder and treatment with aminoglycosides

Human Disease

What can SNP catalogs of humans tell us?

How can Pattern discovery be used here?

Animal studies -

a special case of congenic rats and blood pressure to follow.

End April 1, 2003