MUTATIONS: TRANSMISSION

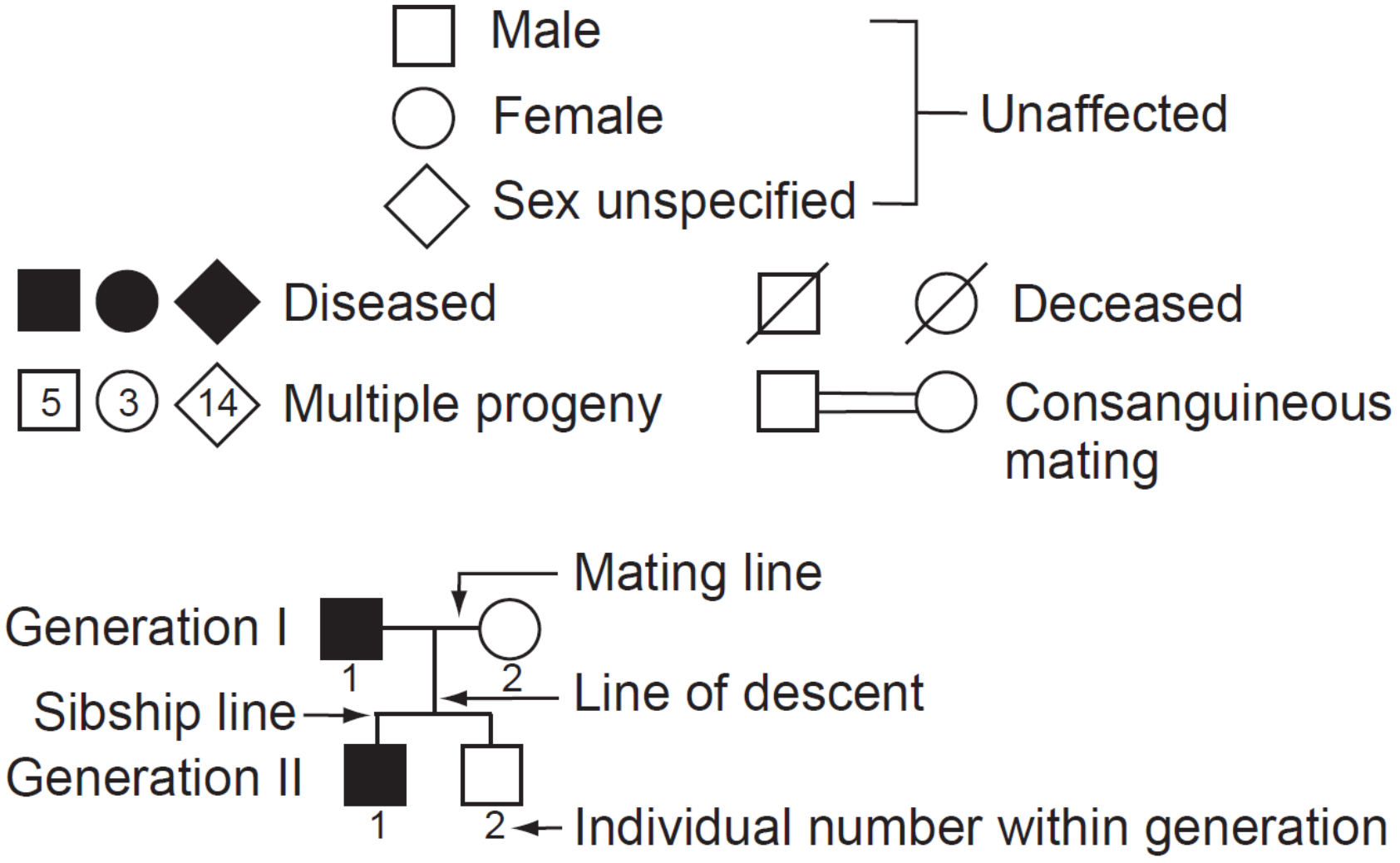
Lecture plan

- Intro: definitions, human life cycle
- Mendel's laws. Allele transmission. Genotype phase. Haplotypes and haplogroups
- Meiosis. Random distribution of chromosomes in meiosis
- Crossing over and recombination. Genetic distance and recombination probabilities
- Linkage disequilibrium and its measures

Some definitions

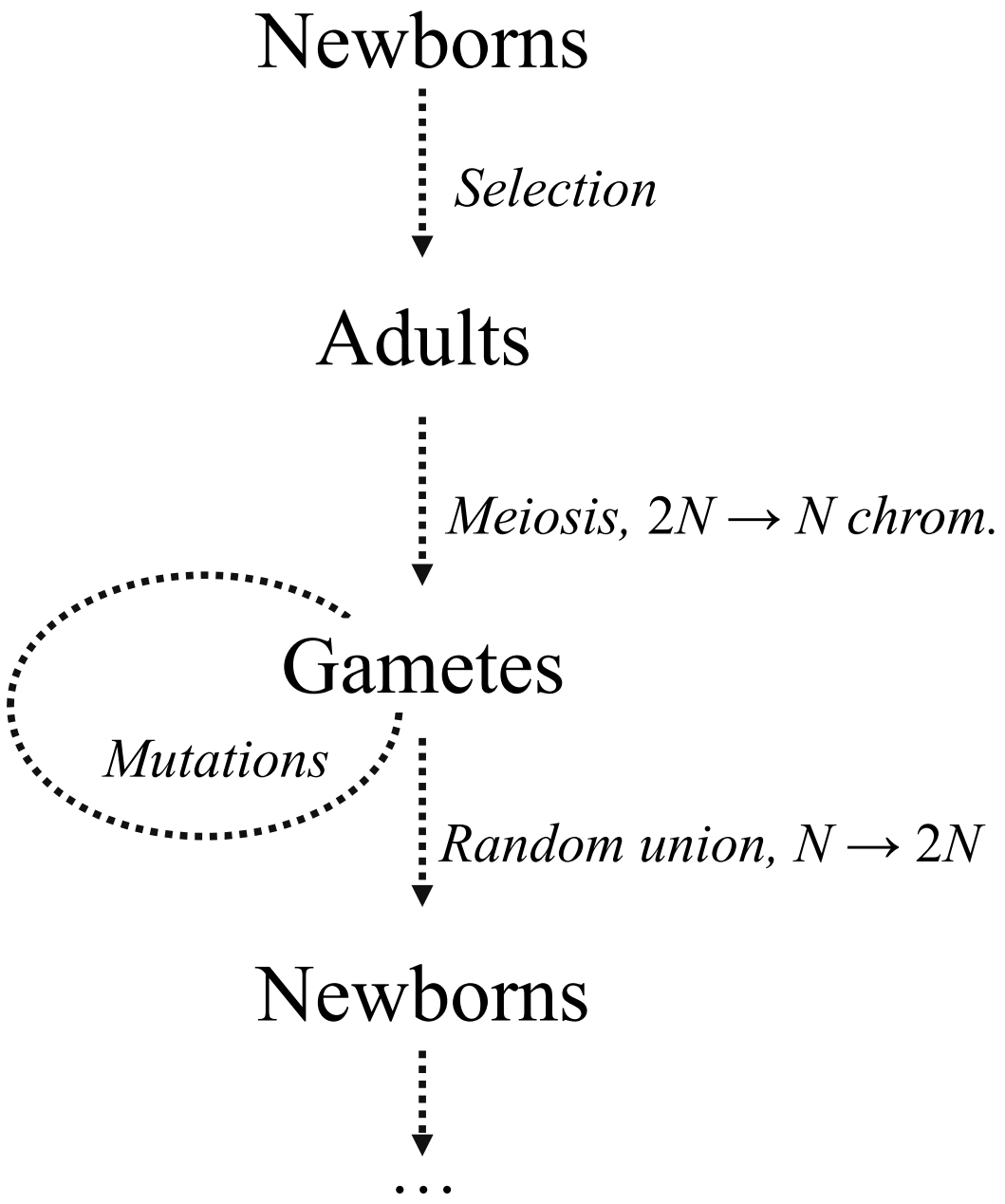
- A **locus** (plural **loci**) is a unique chromosomal location defining the position of an individual gene or DNA sequence.
- **Alleles** (A,a,B,b,…) are alternative versions of a locus (gene).
- The **genotype** is a list of the alleles present at one or a number of loci: AA, Aa, dd…
- **Phenotypes**, **characters**, or **traits** are the observable properties of an organism.
- A person is **homozygous** at a locus if both alleles at that locus are the same, and heterozygous if they are different.
- A person is **hemizygous** if they have only a single allele at a locus. This may be because the locus is on the X or Y chromosome in a male, or it may be because one copy of an autosomal locus is deleted.
- A character is **dominant** if it is manifested in a heterozygous person, **recessive** if not.

Symbols used in pedigree analysis



Hartwell – *Genetics. From genes to genomes*

Life cycle

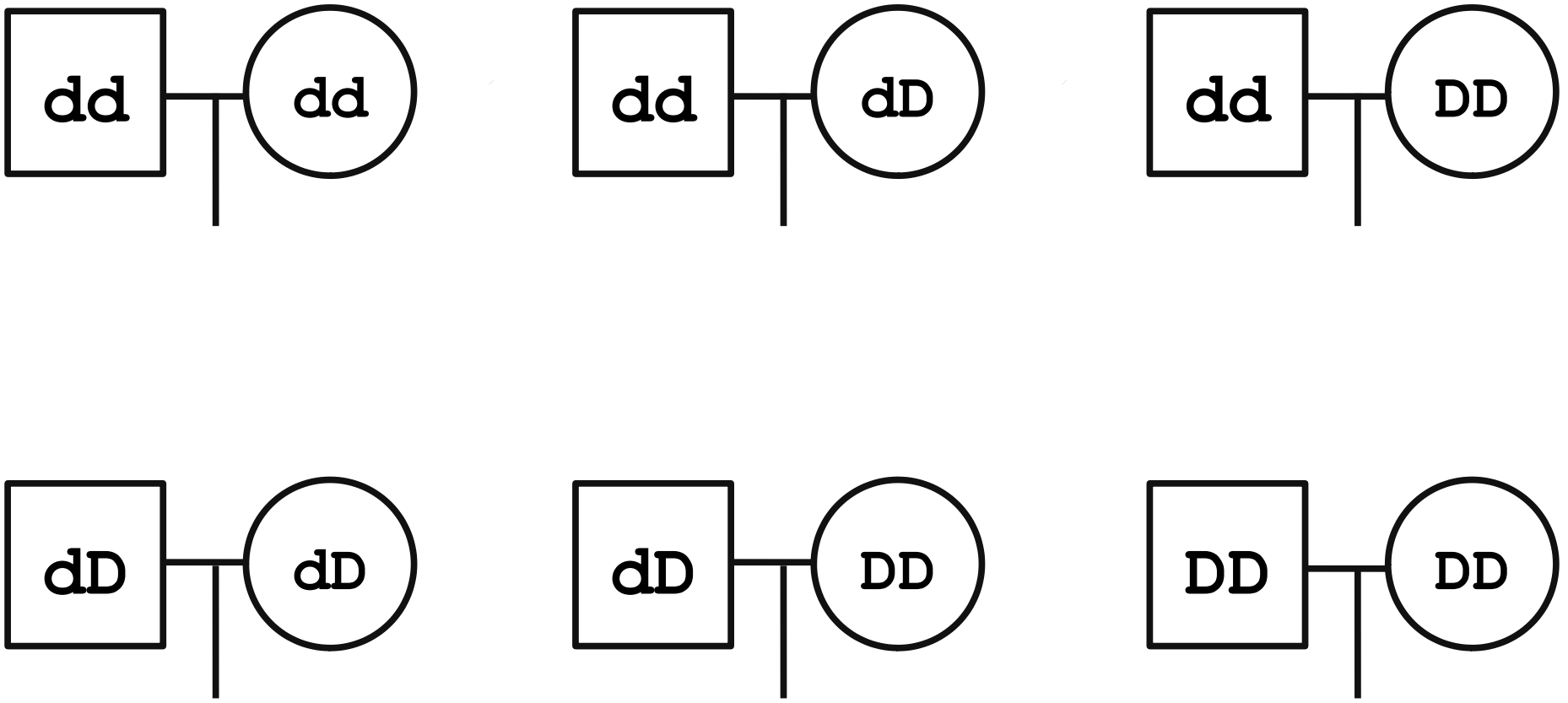

Newborns
Selection
Adults
Meiosis, 2N → N chrom.
Gametes
Mutations
Random union, N → 2N
Newborns
…

Mendel's laws of inheritance (1866)

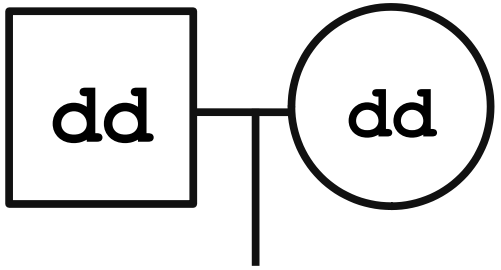
Figure 2.2 Gregor Mendel. Photographed around 1862 holding one of his experimental plants.

Law	Definition
Law of segregation	During gamete formation, the alleles for each gene segregate from each other so that each gamete carries only one allele for each gene.
Law of independent assortment	Genes of different traits can segregate independently during the formation of gametes.
Law of dominance	Some alleles are dominant while others are recessive; an organism with at least one dominant allele will display the effect of the dominant allele.

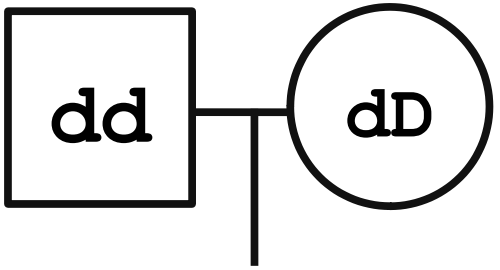
Allele transmission obeys Mendel's law



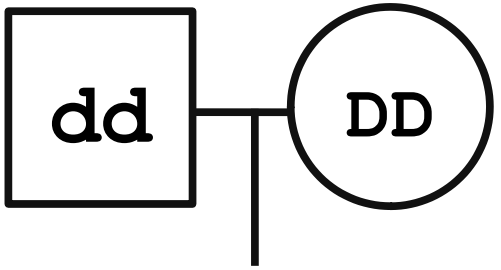
Allele transmission obeys Mendel's law



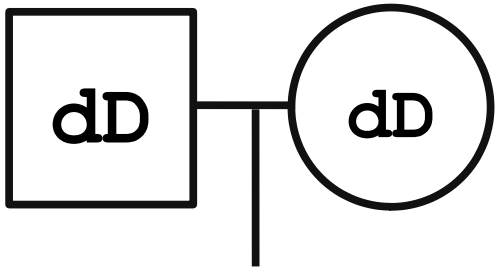
dd



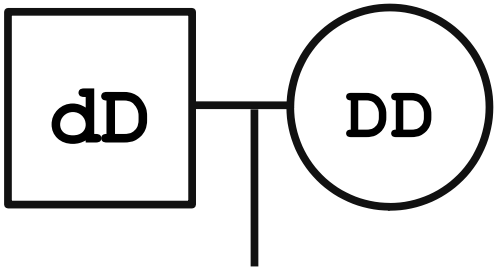
dD
dd



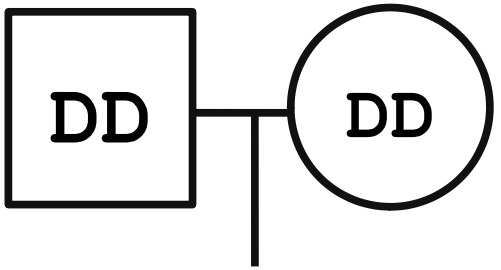
dD



dd
dD
DD

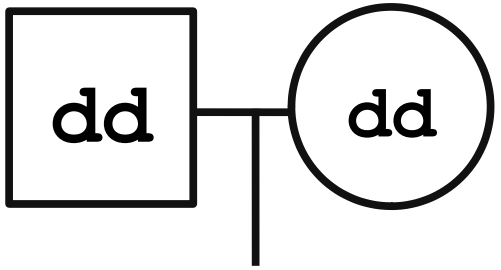


dD
DD



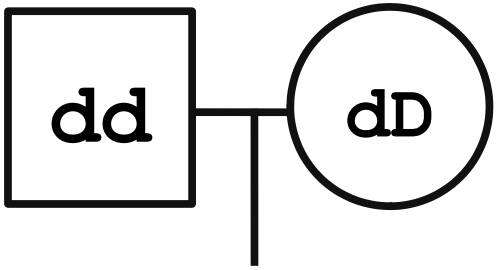
DD

Allele transmission obeys Mendel's law



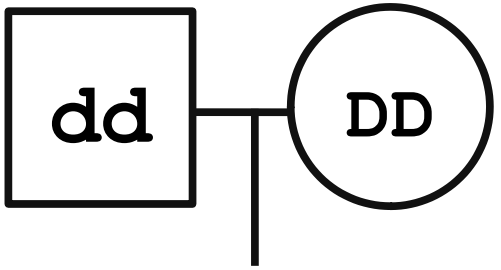
dd

dD ***DNM?***

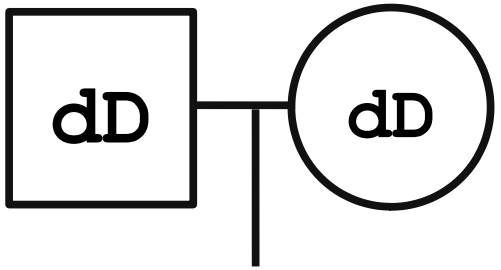


dD

dd



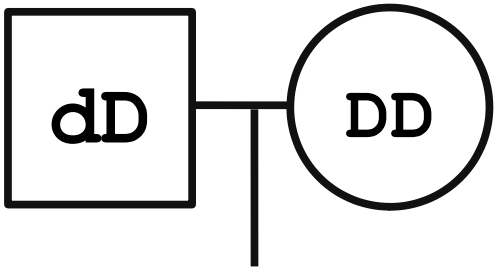
dD



dd

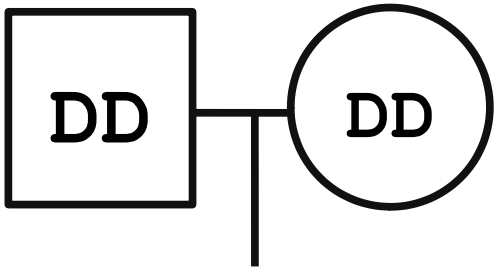
dD

DD



dD

DD



DD

Exercise: other examples of DNM?

Allele transmission obeys Mendel's law

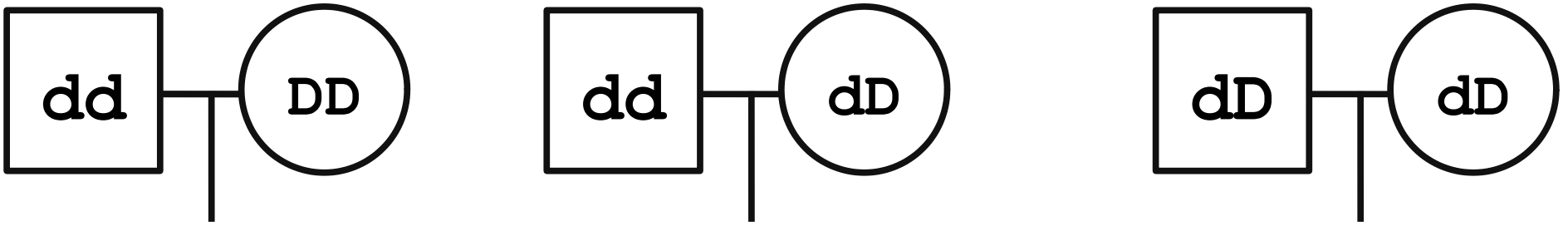
Table 3.3 Probability Distribution for Offspring's Genotype, Conditional on Parental Genotypes

Father's Genotype	Mother's Genotype	*Offspring's Genotype*		
		dd	*dD*	*DD*
dd	*dd*	1	0	0
dd	*dD*	1/2	1/2	0
dd	*DD*	0	1	0
dD	*dd*	1/2	1/2	0
dD	*dD*	1/4	1/2	1/4
dD	*DD*	0	1/2	1/2
DD	*dd*	0	1	0
DD	*dD*	0	1/2	1/2
DD	*DD*	0	0	1

Duncan Thomas – *Statistical Methods in Genetic Epidemiology*

Phased and unphased genotypes

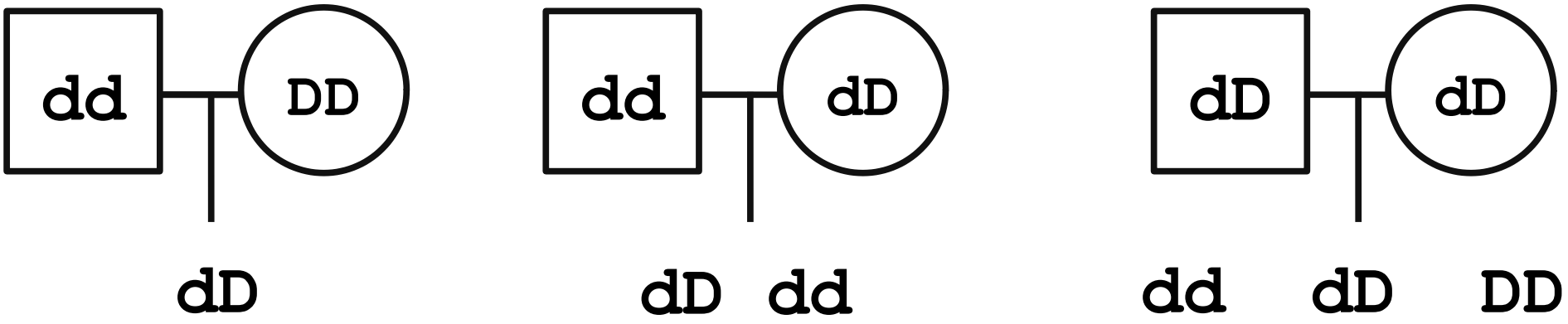
Genotype phasing: Paternal or maternal origin inference for alleles



XY: unphased genotype
X|Y: paternal | maternal

Phased and unphased genotypes

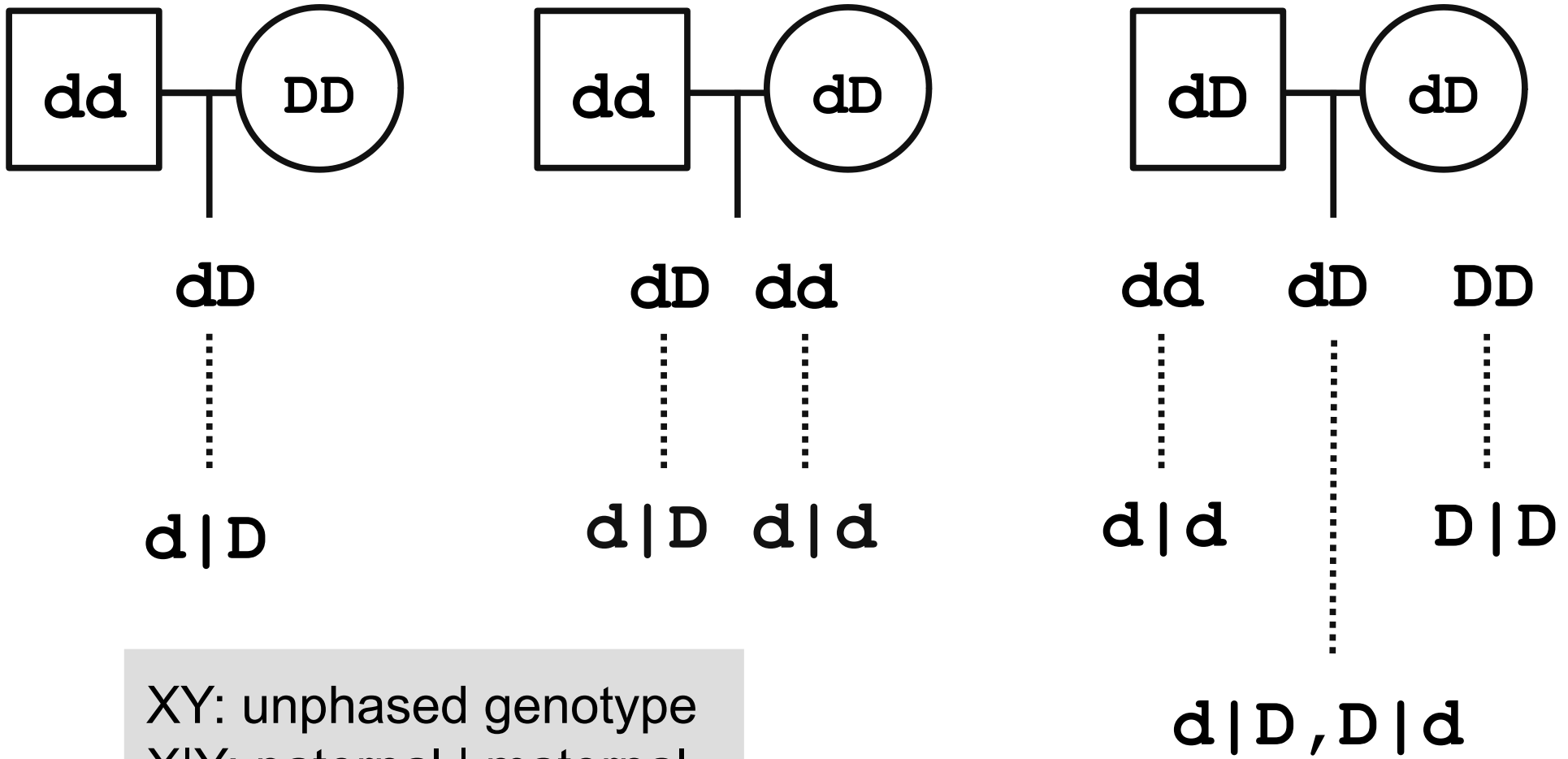
Genotype phasing: Paternal or maternal origin inference for alleles



XY: unphased genotype
X|Y: paternal | maternal

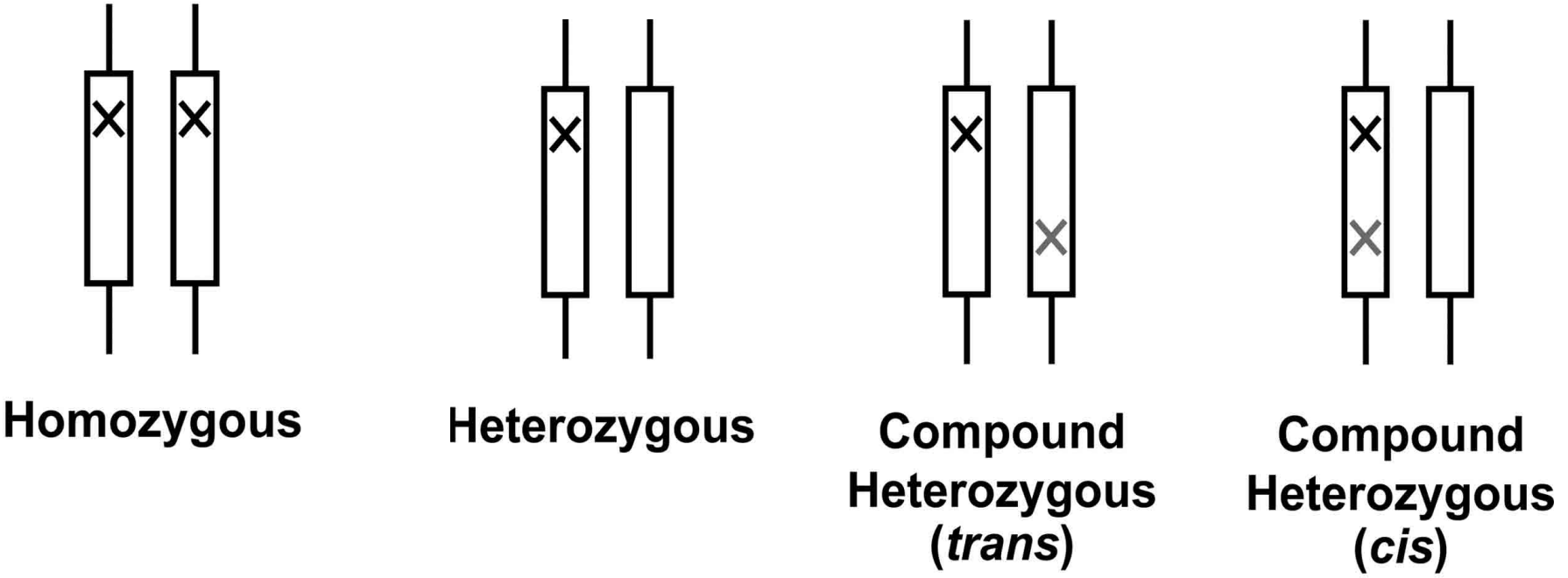
Phased and unphased genotypes

Genotype phasing: Paternal or maternal origin inference for alleles



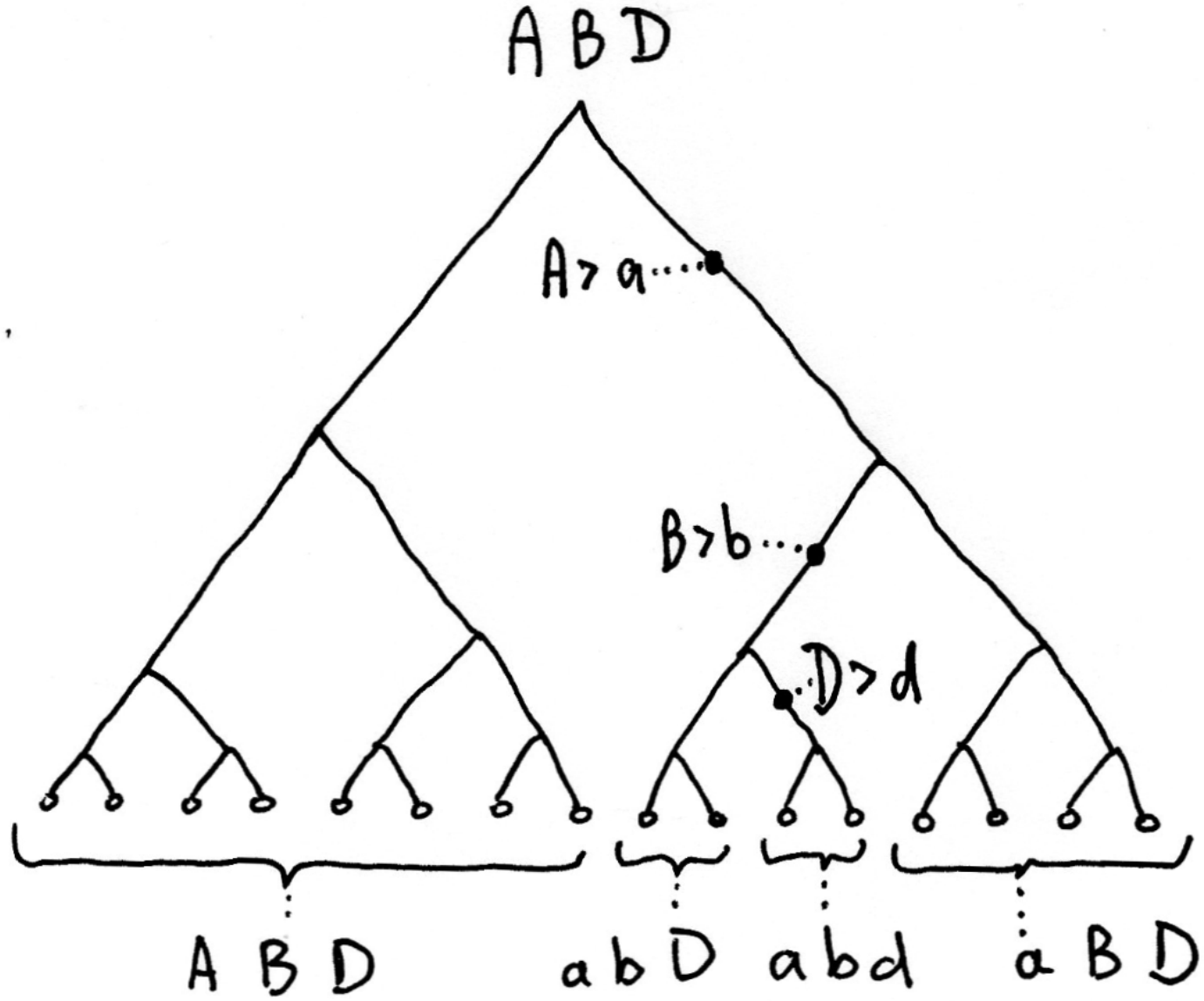
XY: unphased genotype
X|Y: paternal | maternal

Why genotype phase is important?



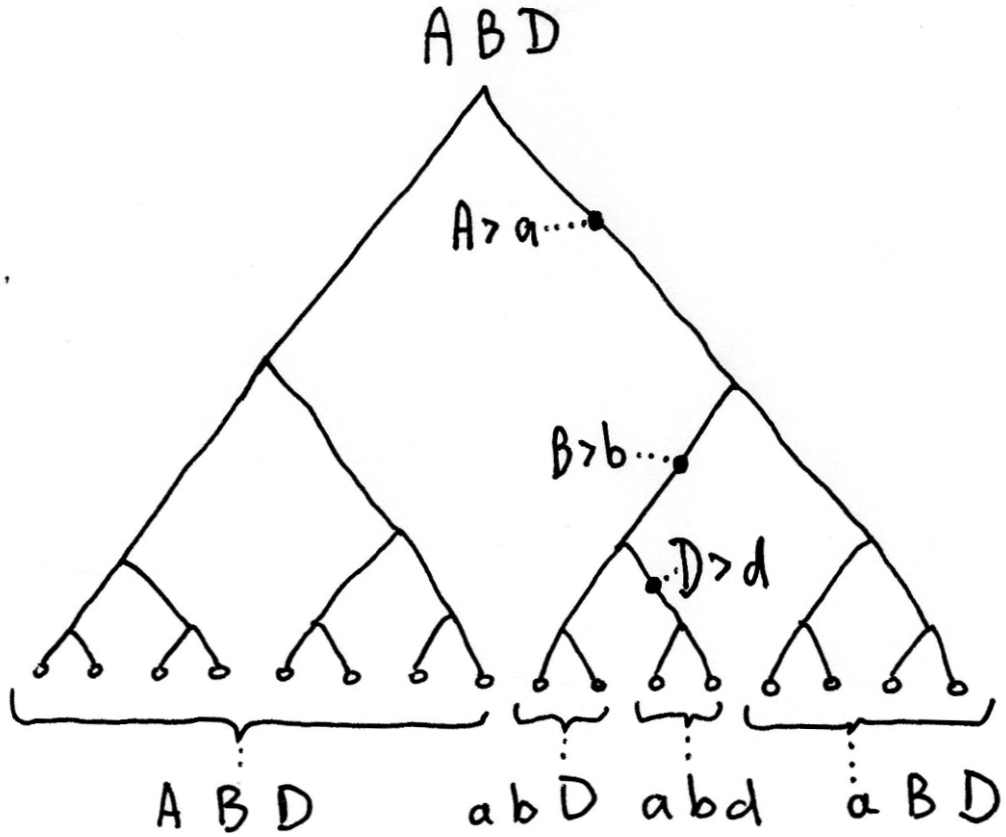
Haplotypes

Haplotype: a combination of alleles that are transmitted together



Haplotypes

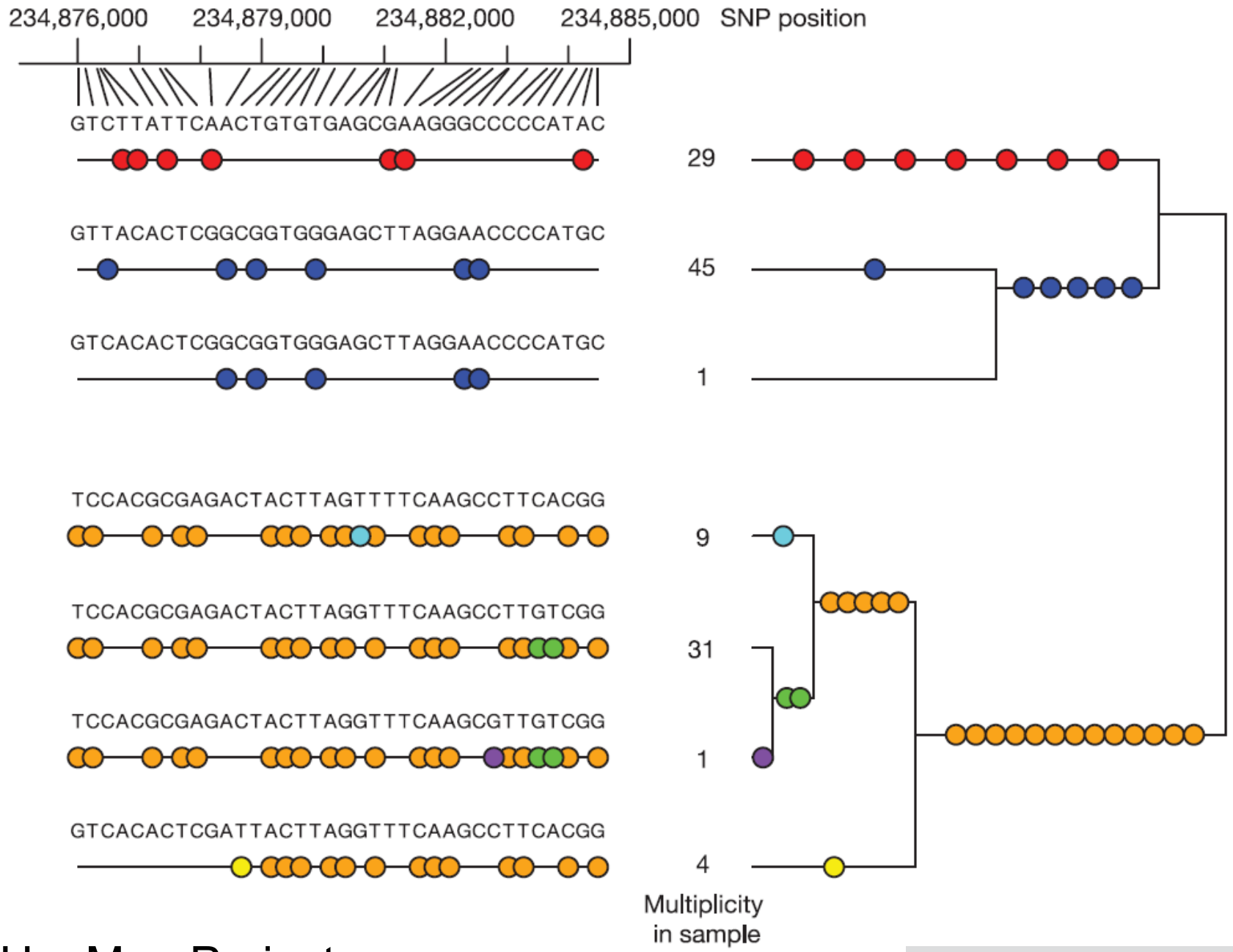
Haplotype: a combination of alleles that are transmitted together



Haplotype number	*Haplotype*	*Frequency*
0	A B D	8/16
1	a b D	2/16
2	a b d	2/16
3	a B D	4/16
-	A b D	0
-	A b d	0
-	A B d	0
-	a B d	0

– For N alleles, ~N combinations (haplotypes) instead of 2^N
– Alleles are in **linkage**: d with b etc.

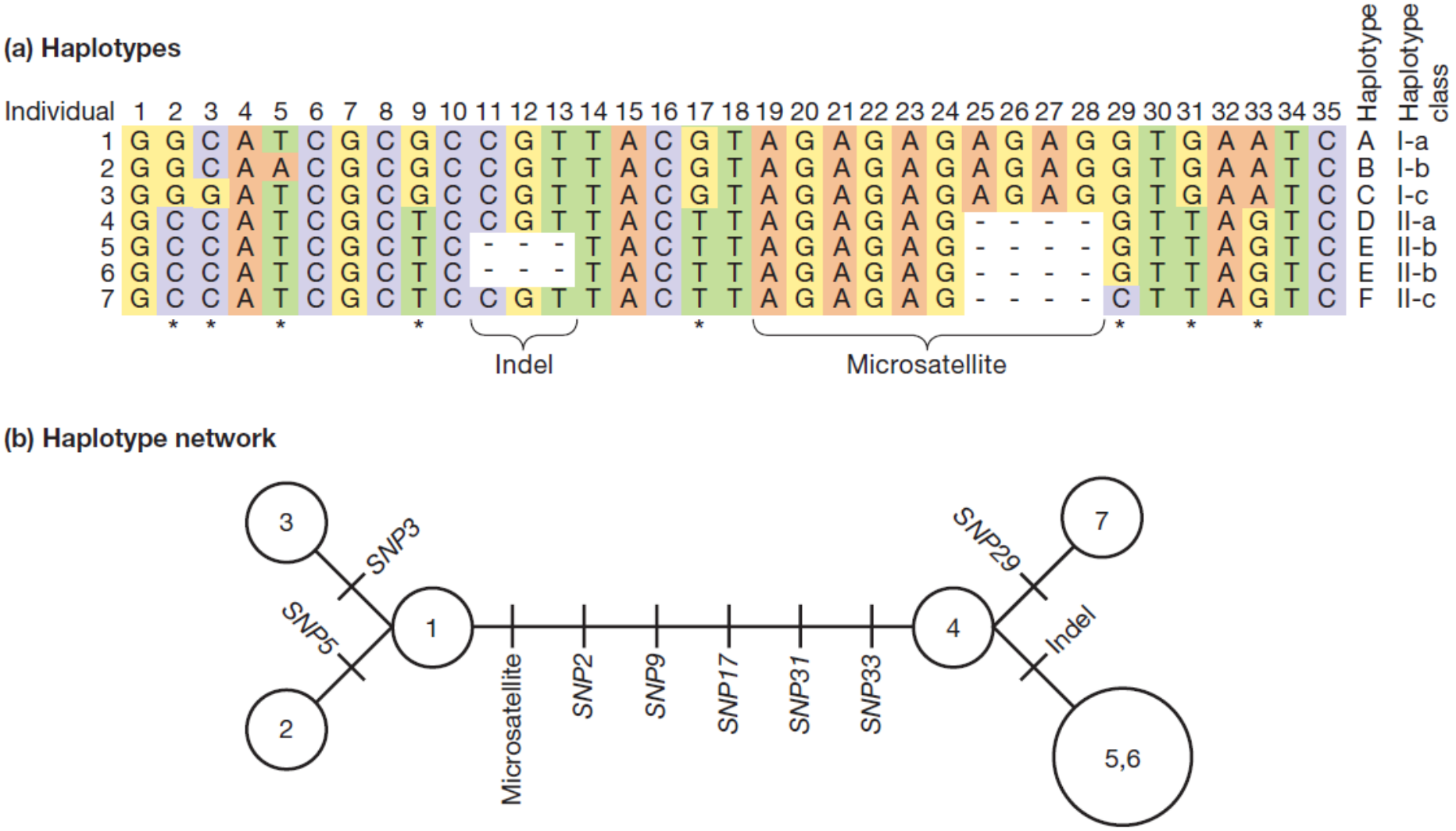
Haplotypes: more realistic example



HapMap Project

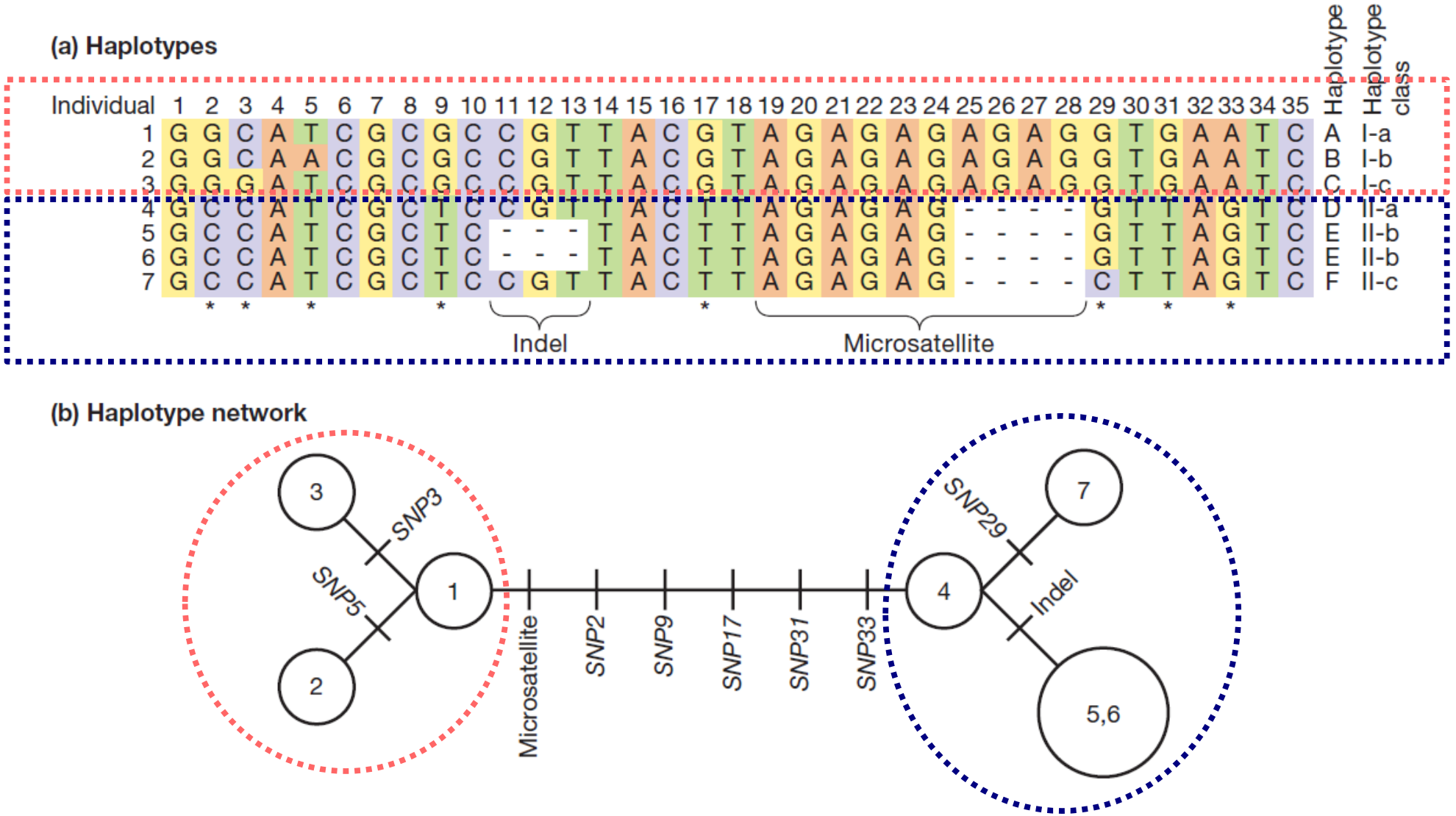
Vol 437|27 October 2005|doi:10.1038/nature04226

Haplotypes: more realistic example



Griffiths -- *Introduction to Genetic Analysis*

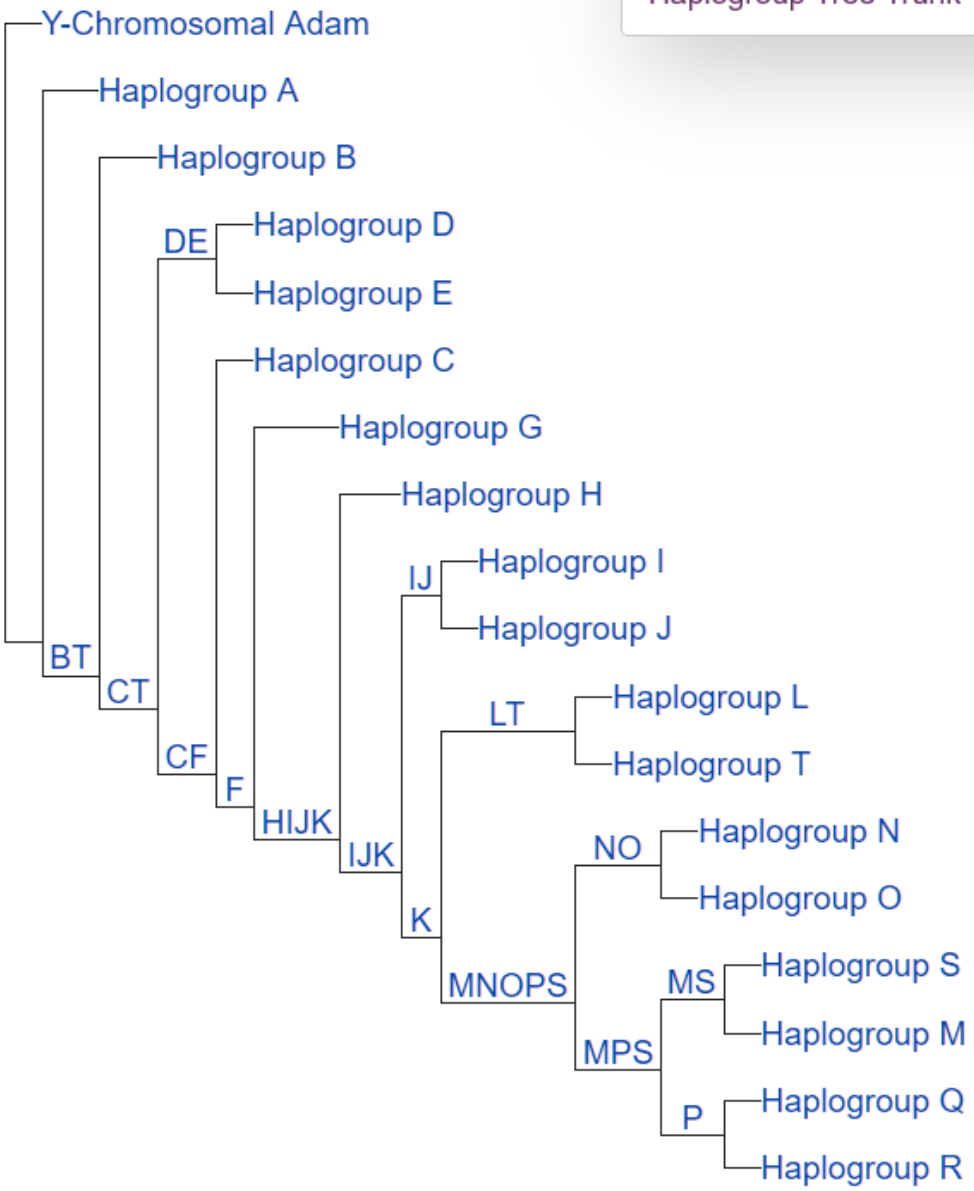
Haplotypes: more realistic example



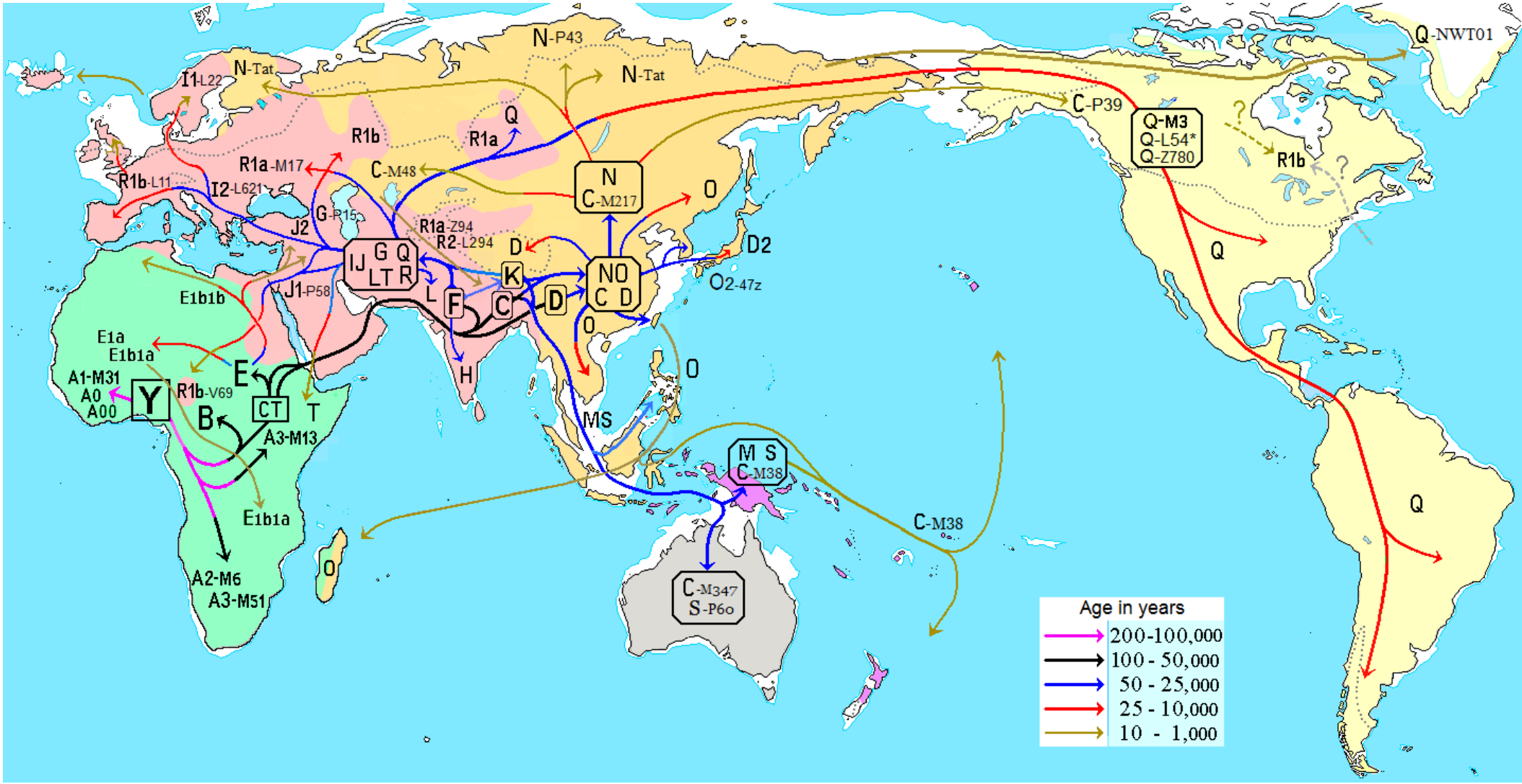
Griffiths -- *Introduction to Genetic Analysis*

Haplotypes: more realistic example

Phylogenetic tree of Y-DNA haplogroups [10]



Haplotypes: more realistic example

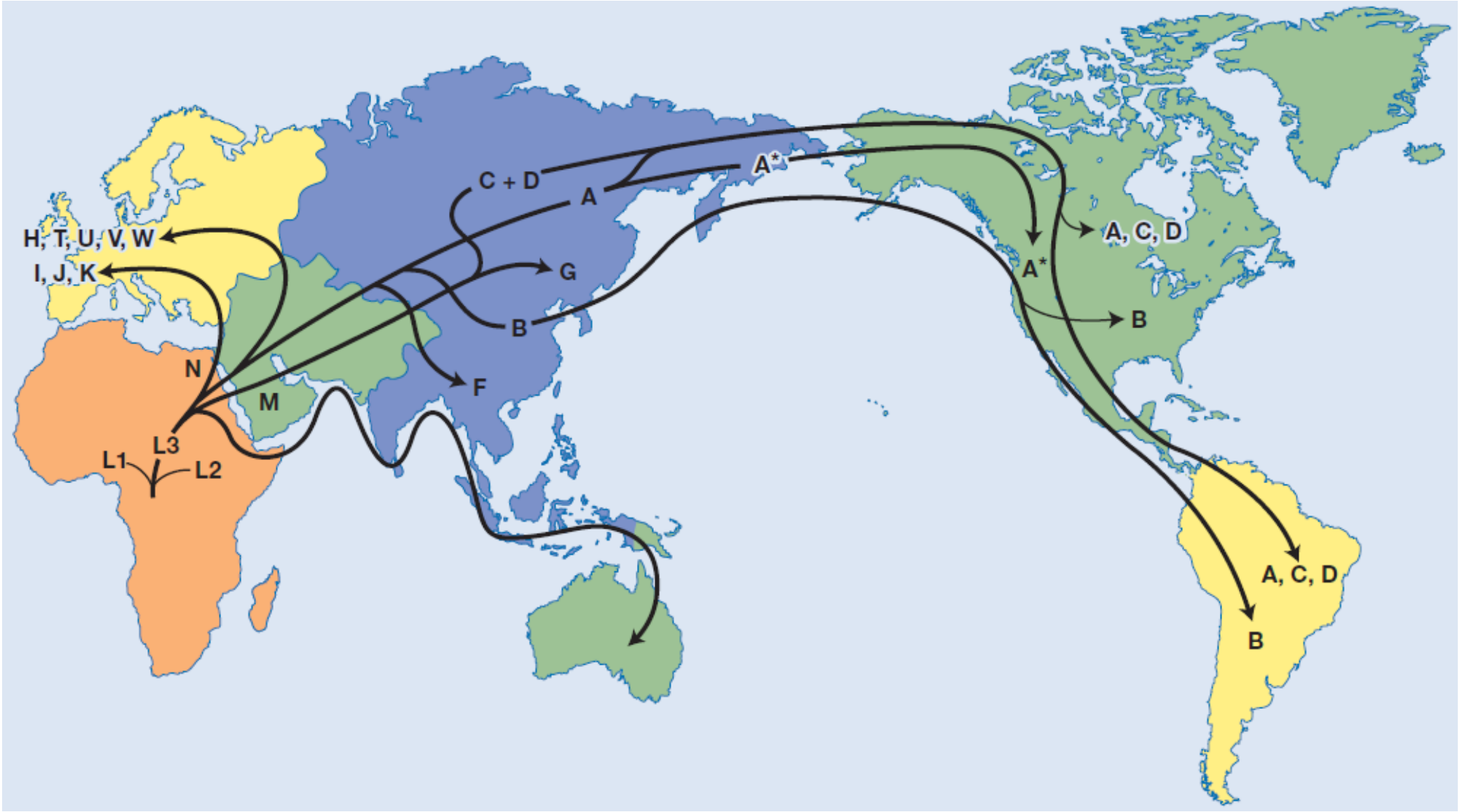


Maulucioni - Own work

World map of early migrations of modern humans based on the Y-chromosome DNA.

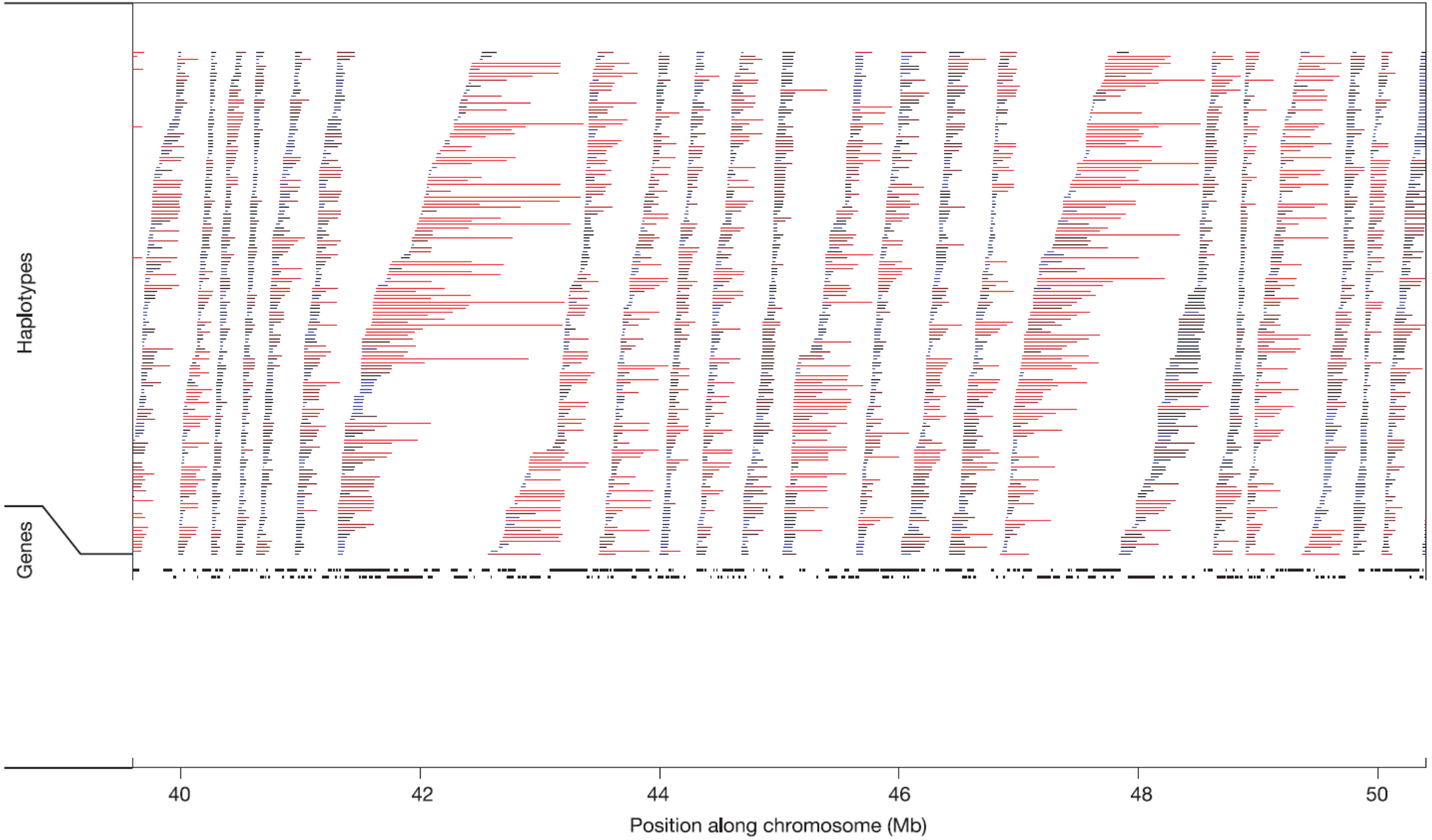
Haplotypes: more realistic example

Map of human migration based on the mitochondrial DNA



Griffiths -- *Introduction to Genetic Analysis*

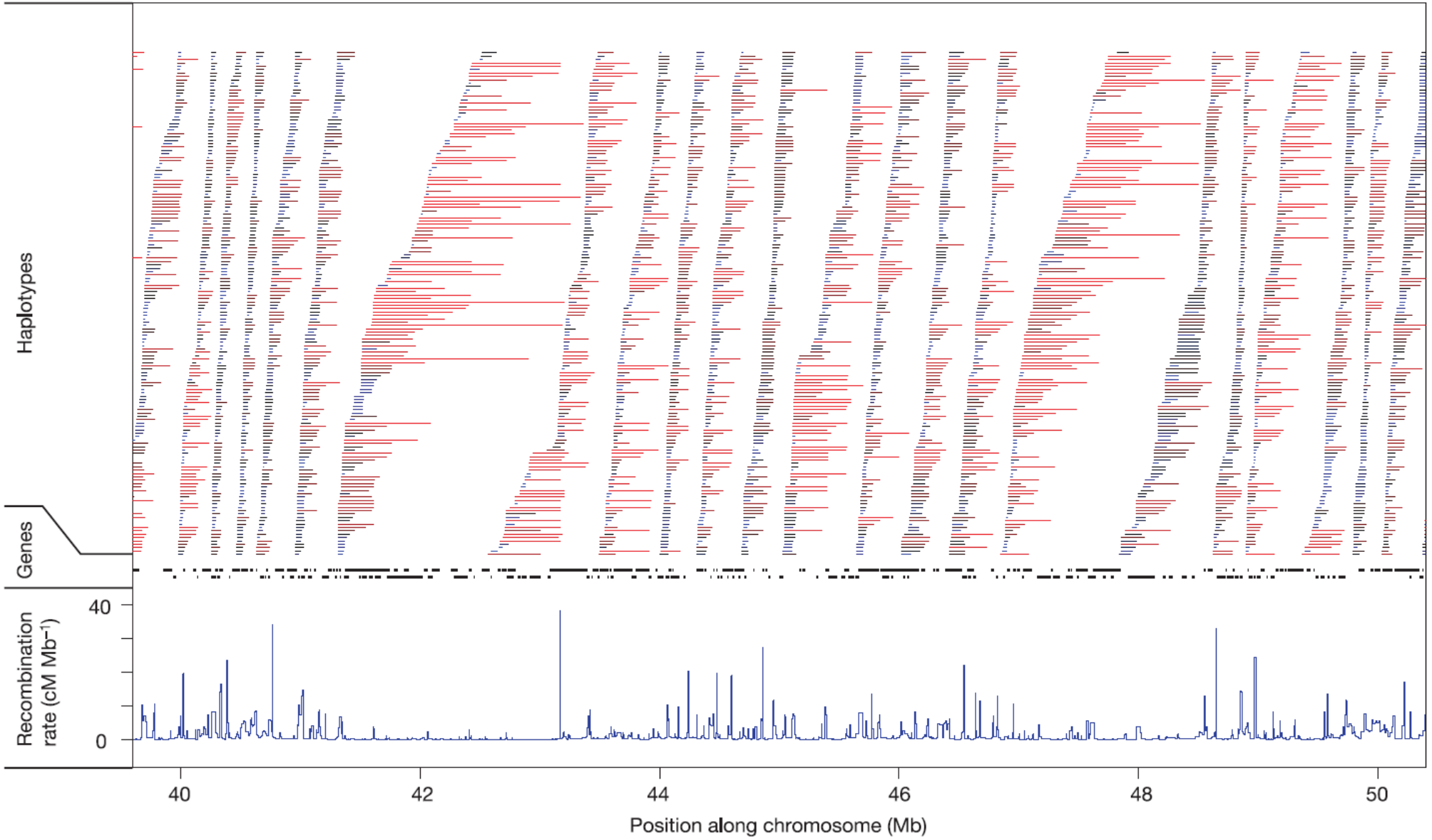
Haplotypes: even more realistic example



HapMap Project

Vol 437|27 October 2005|doi:10.1038/nature04226

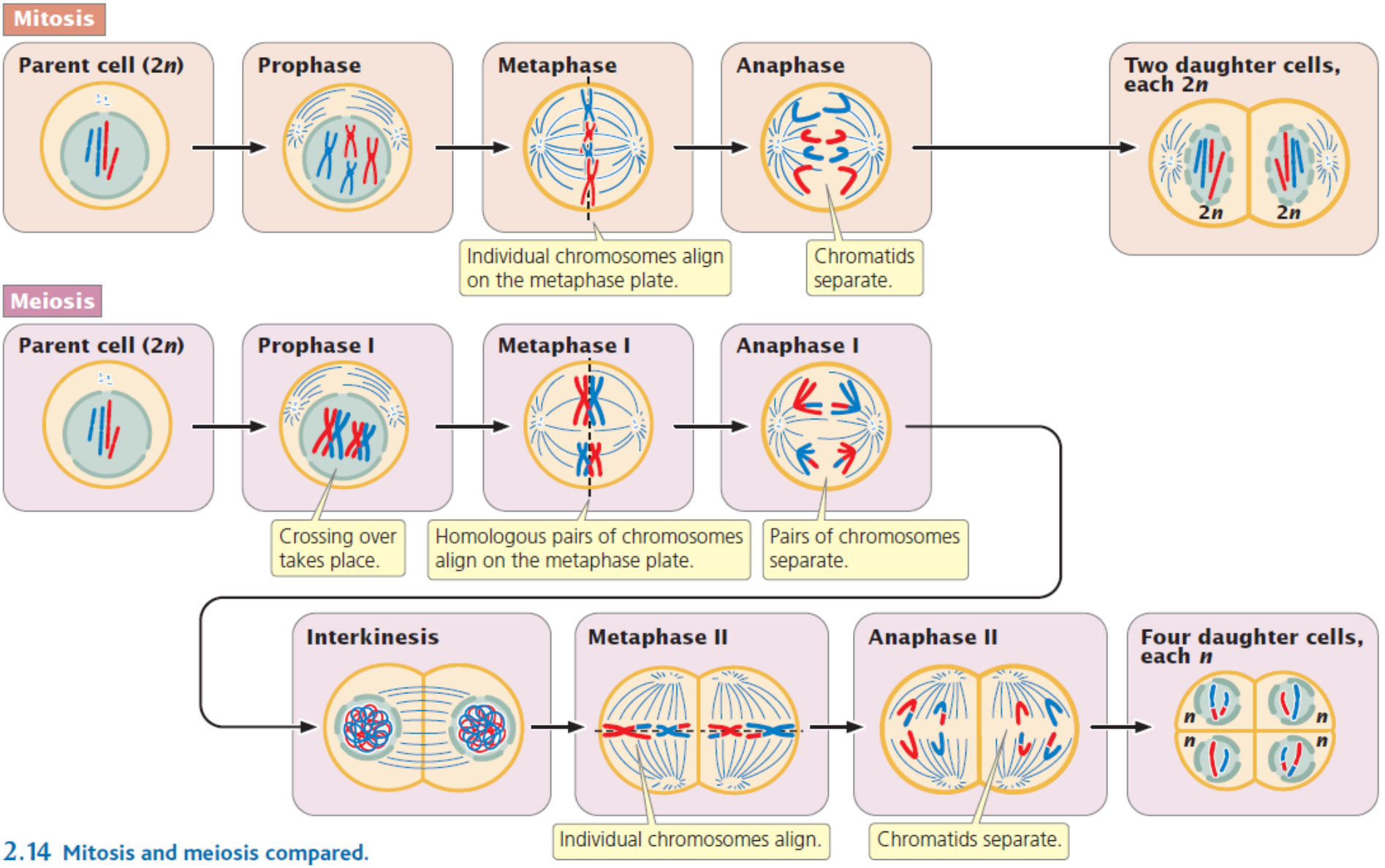
Haplotypes: even more realistic example



HapMap Project

Vol 437|27 October 2005|doi:10.1038/nature04226

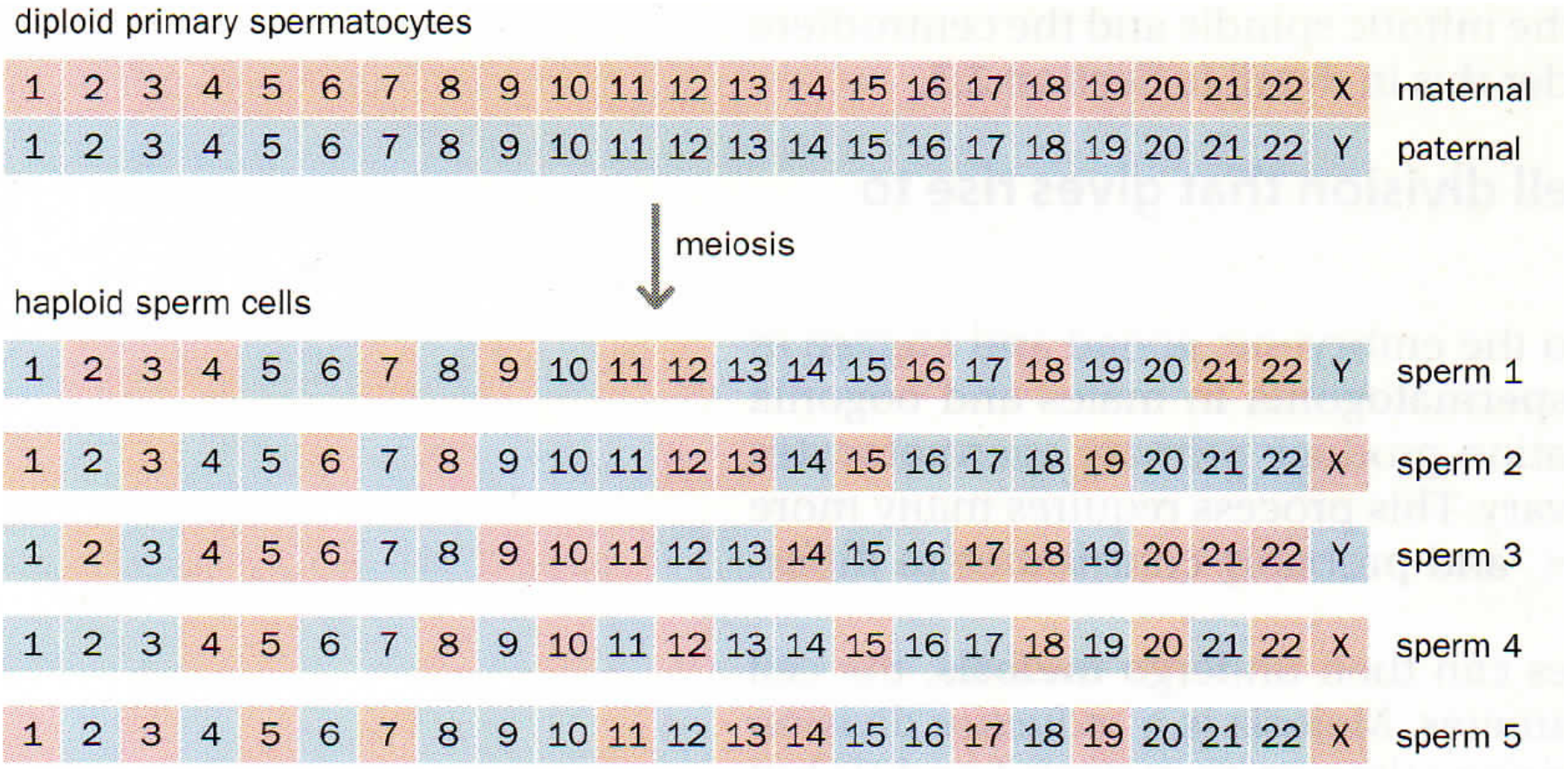
Mitosis and meiosis



2.14 Mitosis and meiosis compared.

Pierce – *Genetics Essentials. Concepts and Connections*

Random distribution of chromosomes in meiosis

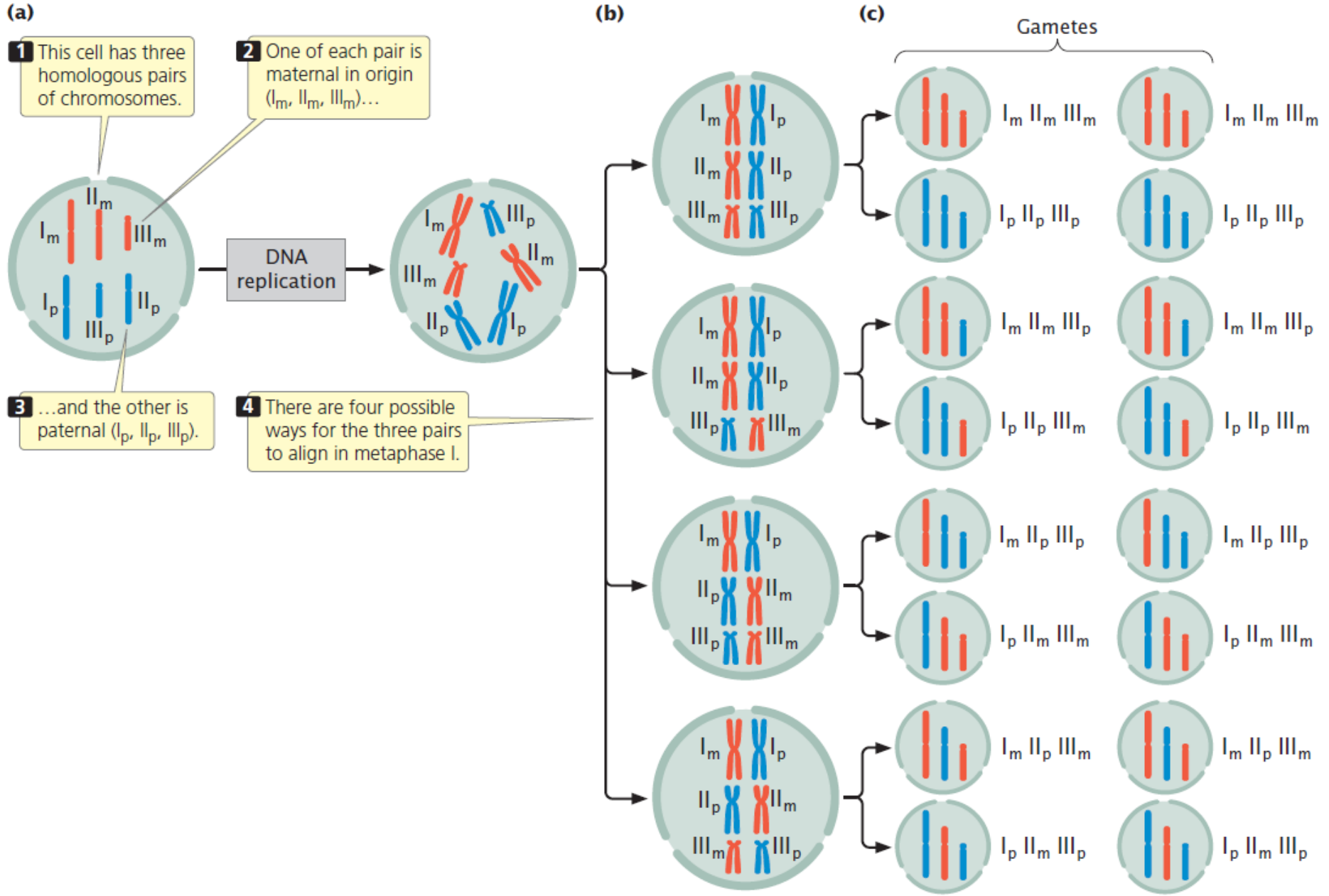


Arbitrary examples of just 5 of 2^{23} = 8,388,608 chromosome combinations in sperm cells, assuming no recombination

— *Ах, королева*, — игриво трещал Коровьев, — *вопросы крови — самые сложные вопросы в мире! <...> Я ничуть не погрешу, если, говоря об этом, упомяну о* ***причудливо тасуемой колоде карт***

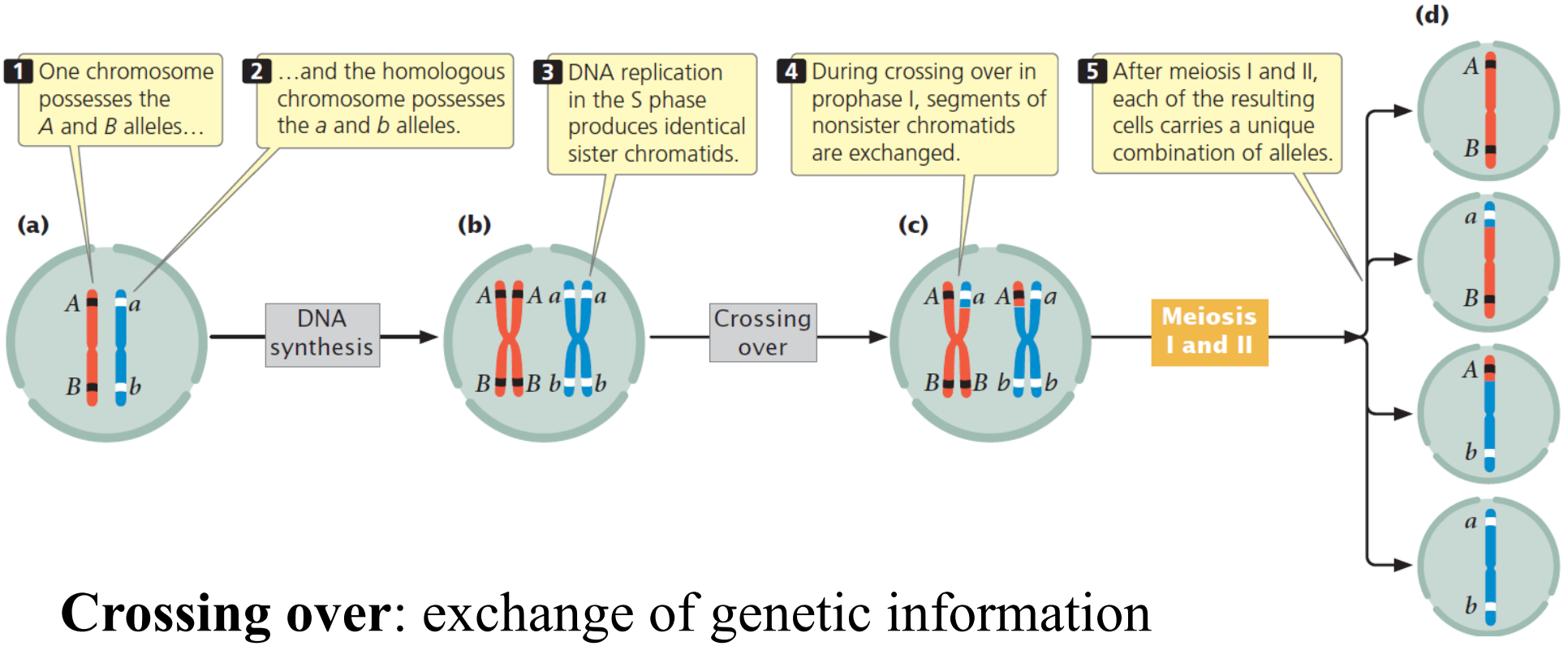
М.А.Булгаков

Random distribution of chromosomes in meiosis



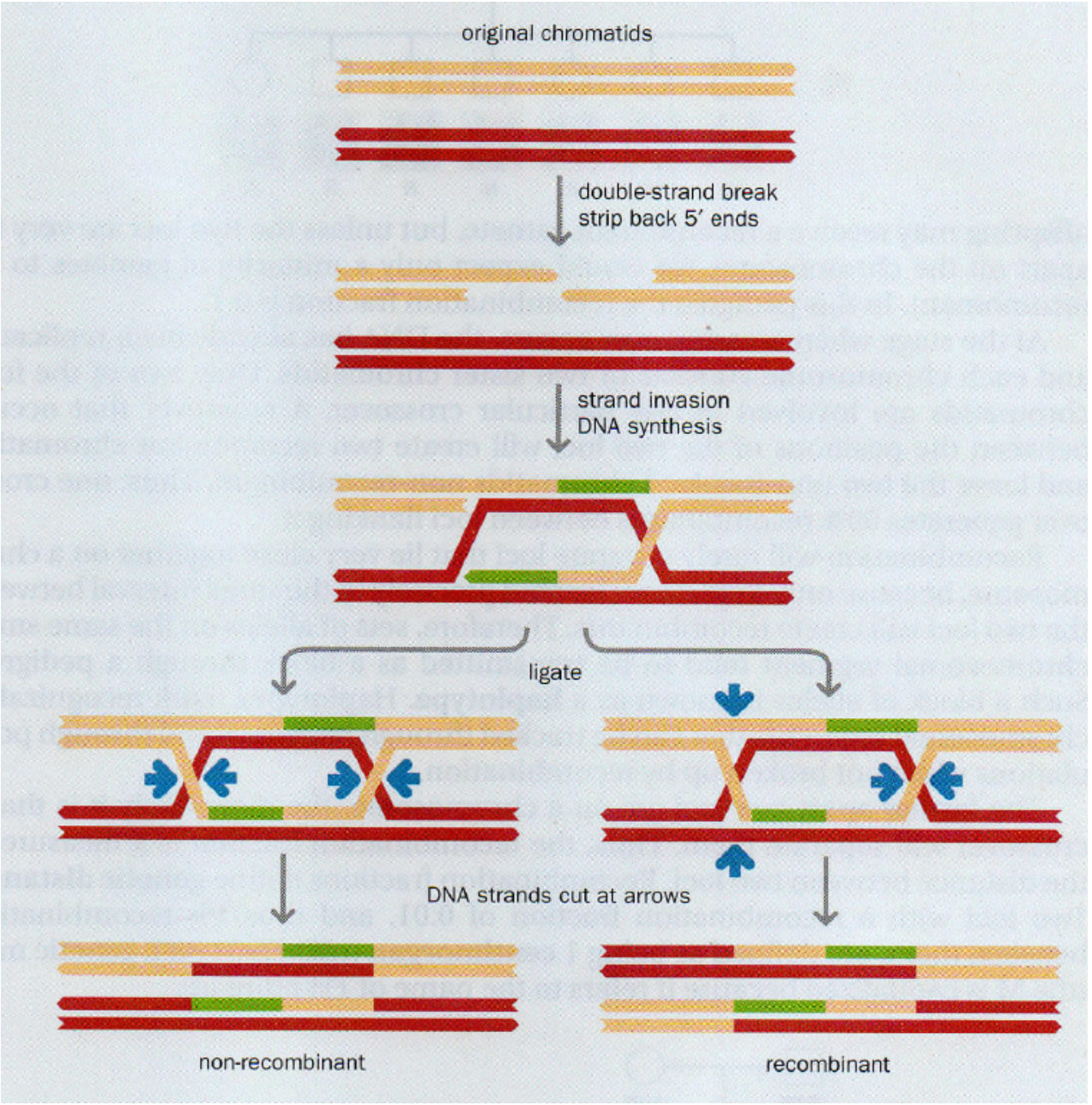
Pierce – *Genetics Essentials. Concepts and Connections*

Crossing over produces extra genetic variation



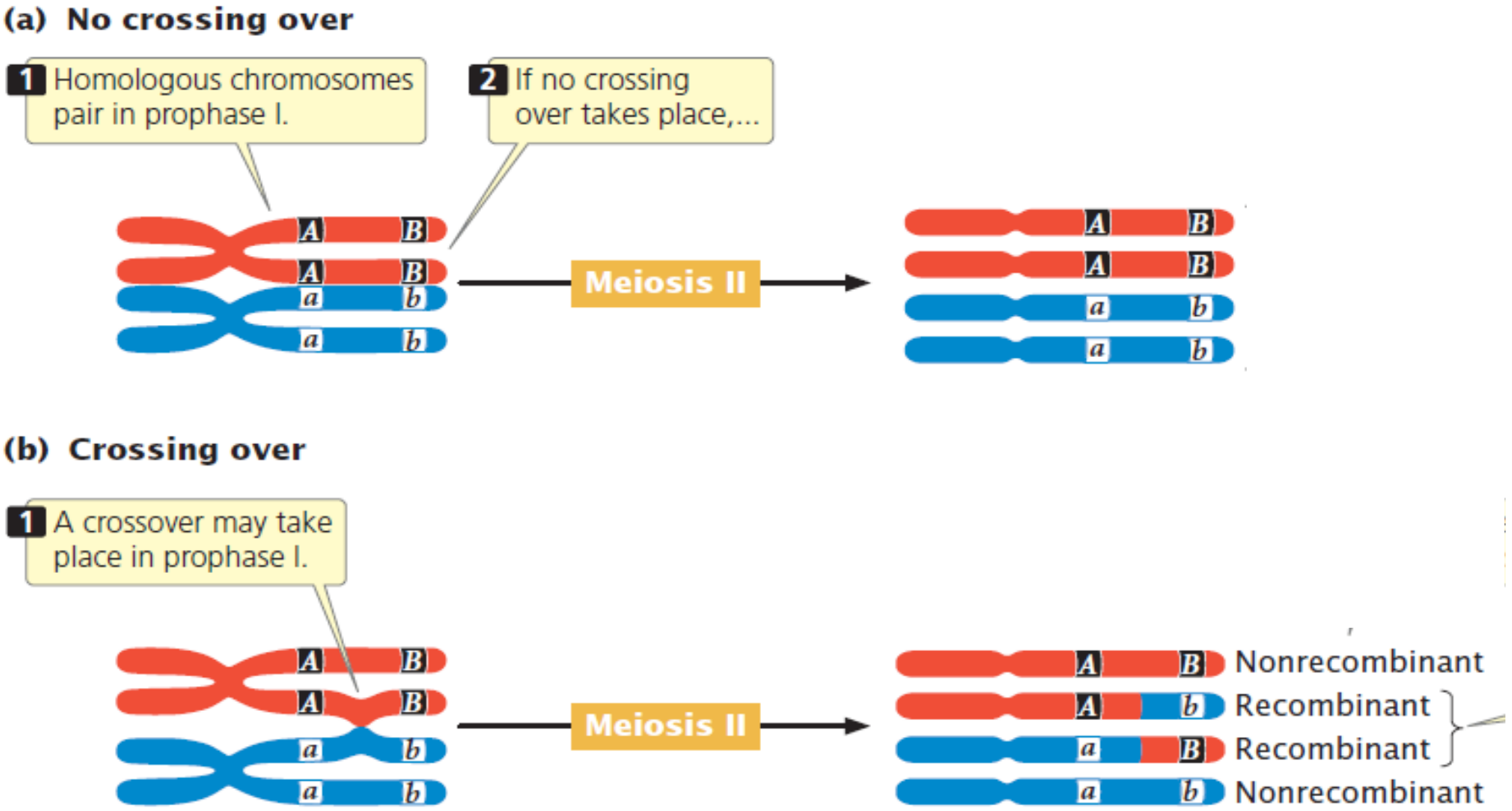
Crossing over: exchange of genetic information between homologous chromosomes. Crossing over is the basis for intrachromosomal **recombination**, creating new combinations of alleles on a chromatid.

Crossing over produces extra genetic variation



Strachan, Read – *Human Molecular Genetics*

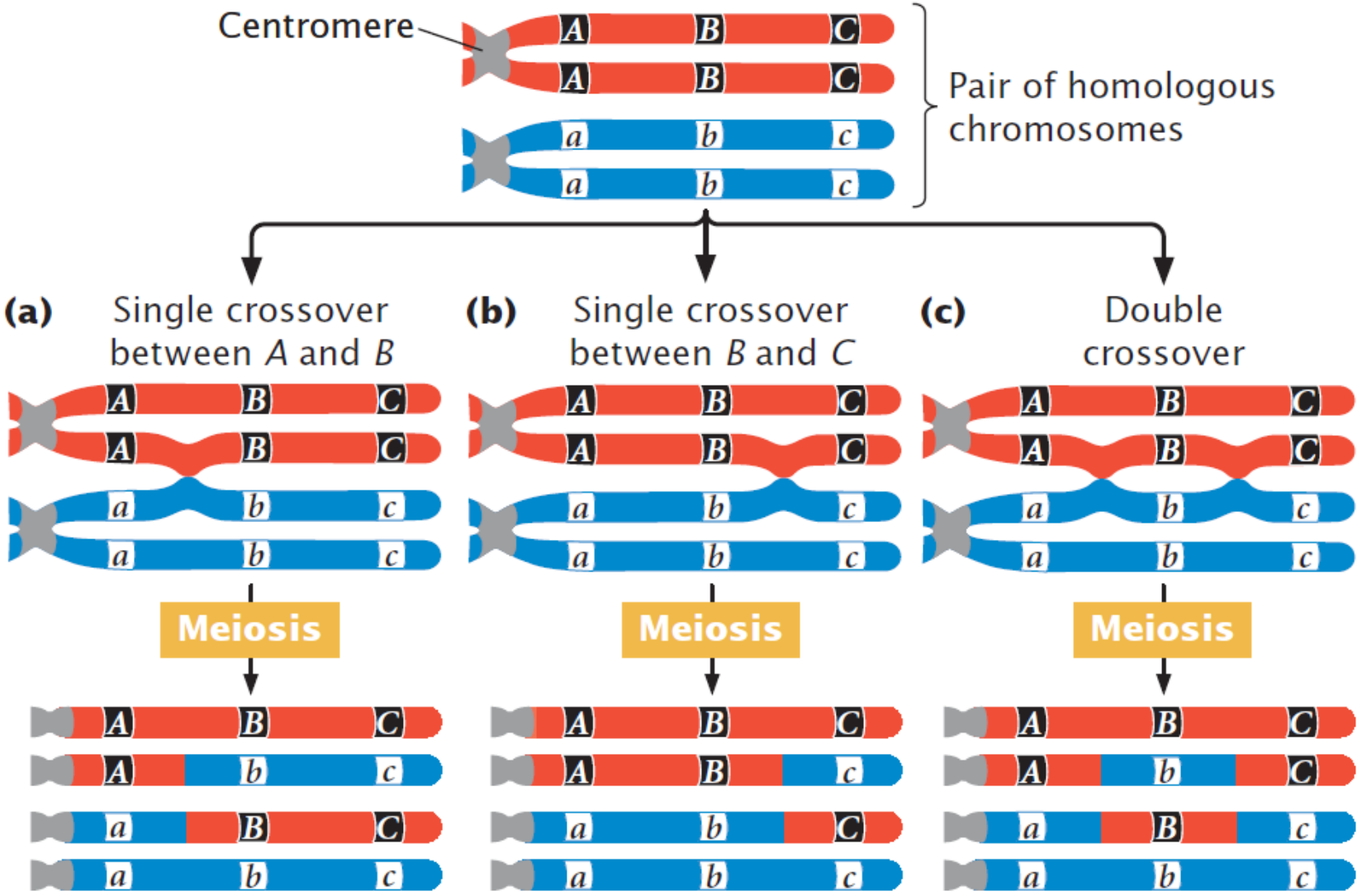
Crossing over produces extra genetic variation



5.6 A single crossover produces half nonrecombinant gametes and half recombinant gametes.

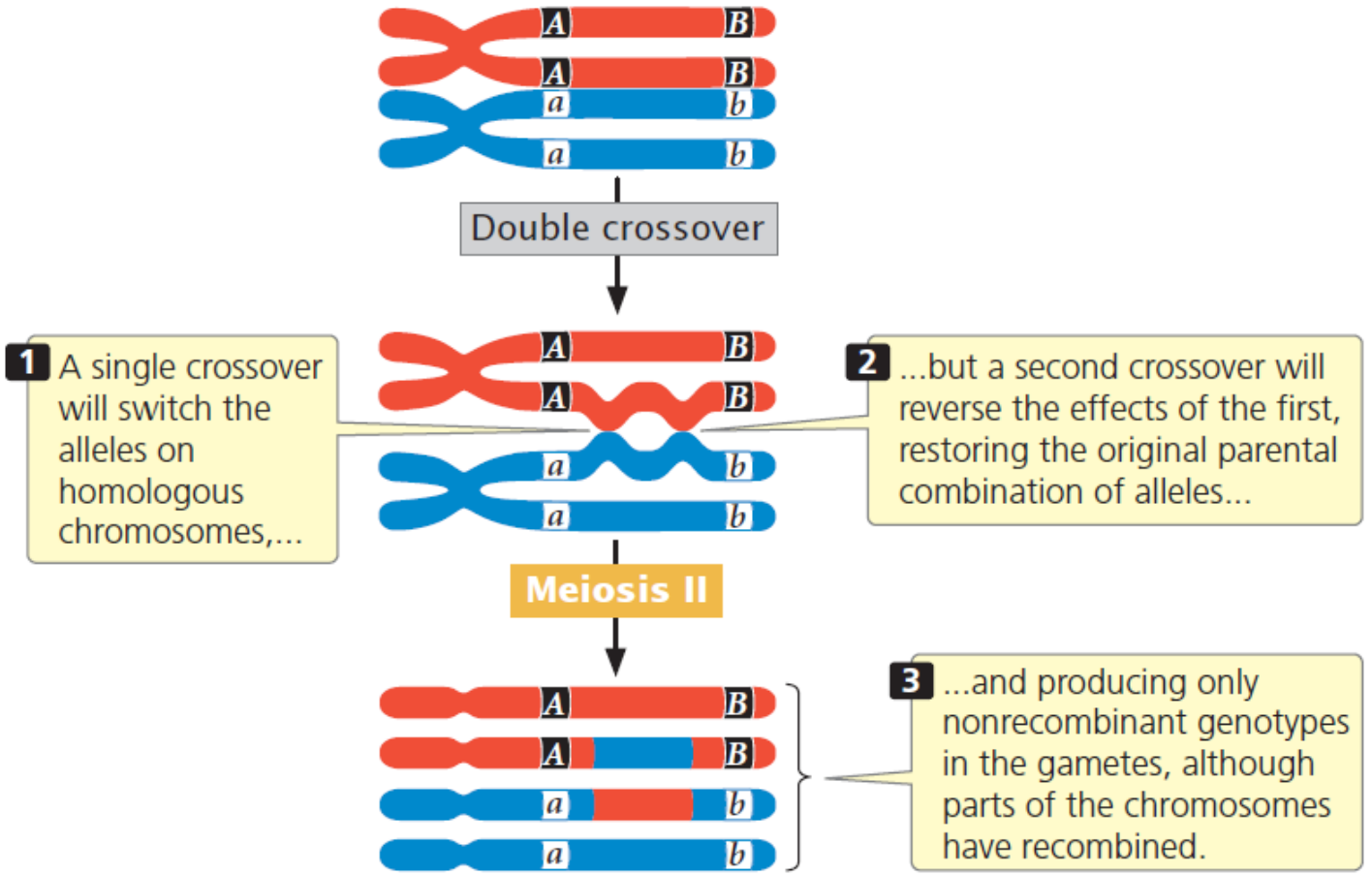
Pierce – *Genetics Essentials. Concepts and Connections*

Crossing over produces extra genetic variation



Pierce – *Genetics Essentials. Concepts and Connections*

Crossing over produces extra genetic variation



5.11 A two-strand double crossover between two linked genes produces only nonrecombinant gametes.

Pierce – *Genetics Essentials. Concepts and Connections*

Crossing over produces extra genetic variation

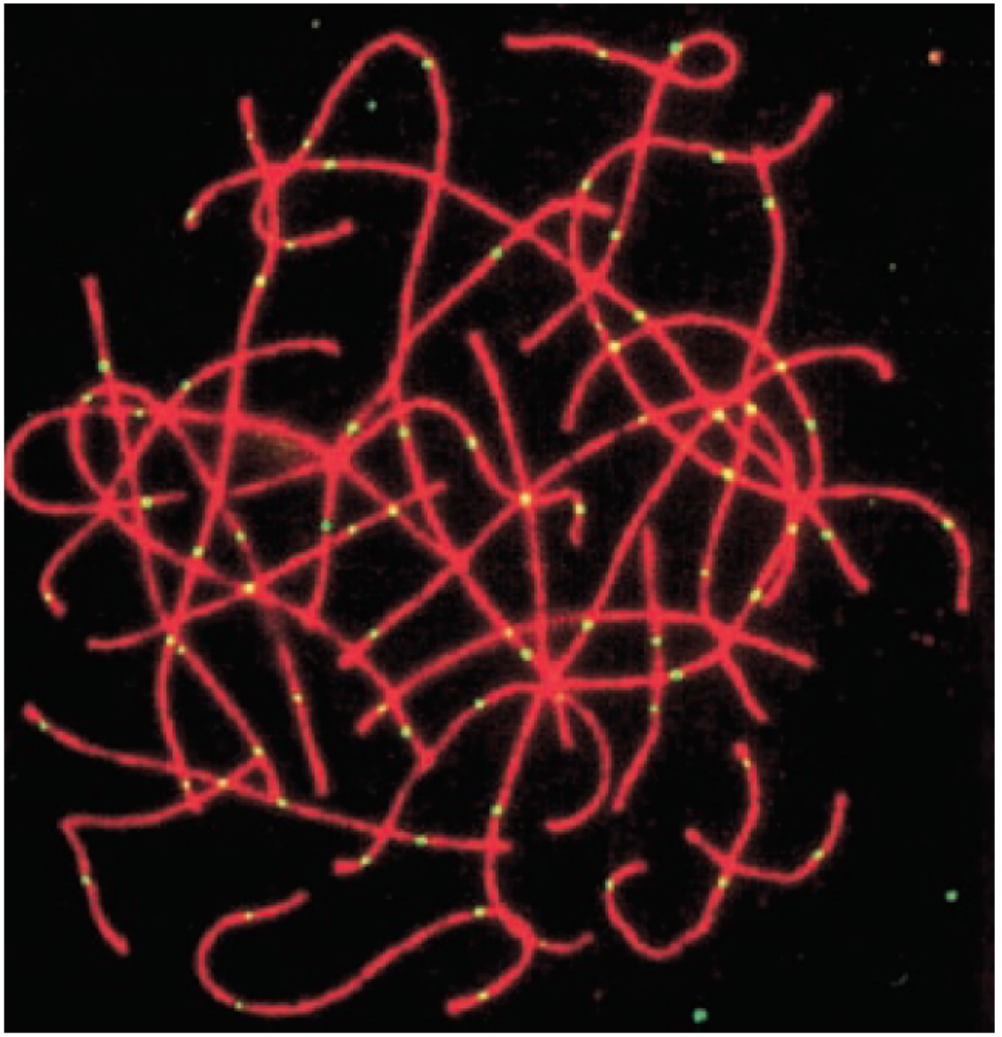
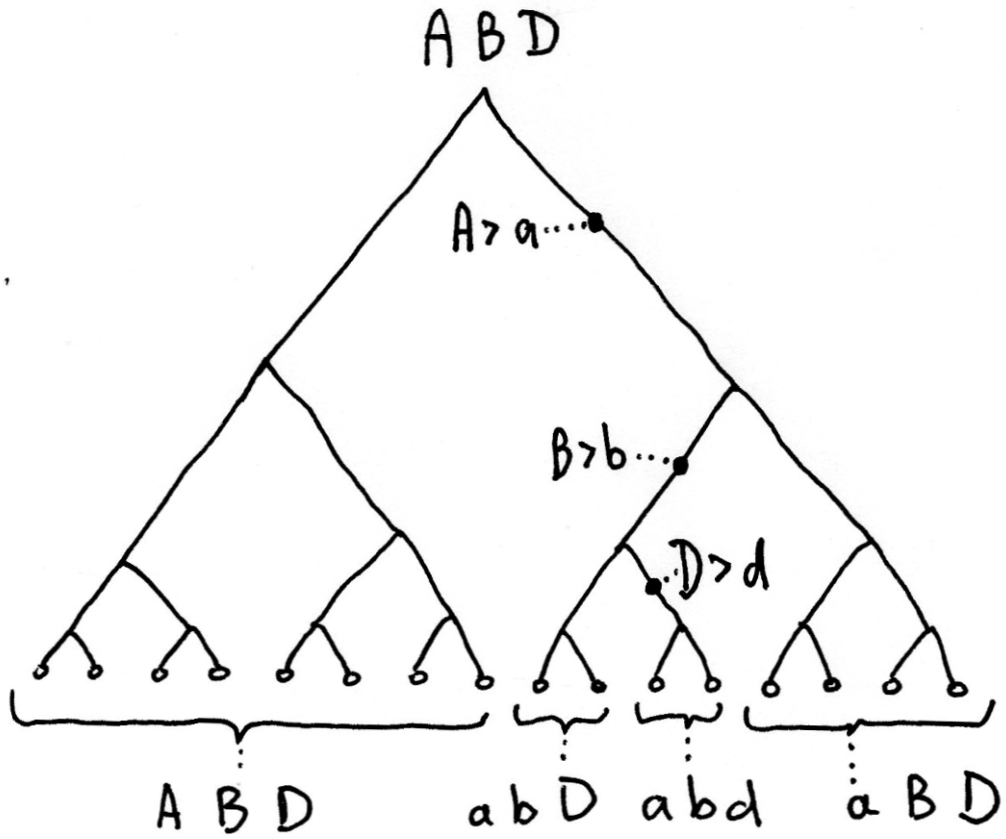


Figure 19–9 **Multiple crossovers can occur between homologous chromosomes.** Shown is a light micrograph of a spread of the chromosomes of a human oocyte (egg-cell precursor) at the stage where all four chromatids—maternal and paternal—are still tightly associated: each single long thread (stained *red*) is a bivalent containing four DNA double helices. Sites of recombination are marked by the presence of a protein (stained *green*) that is a key component of the recombination machinery. (From C. Tease et al., *Am. J. Hum. Genet.* 70:1469–1479, 2002. With permission from Elsevier.)

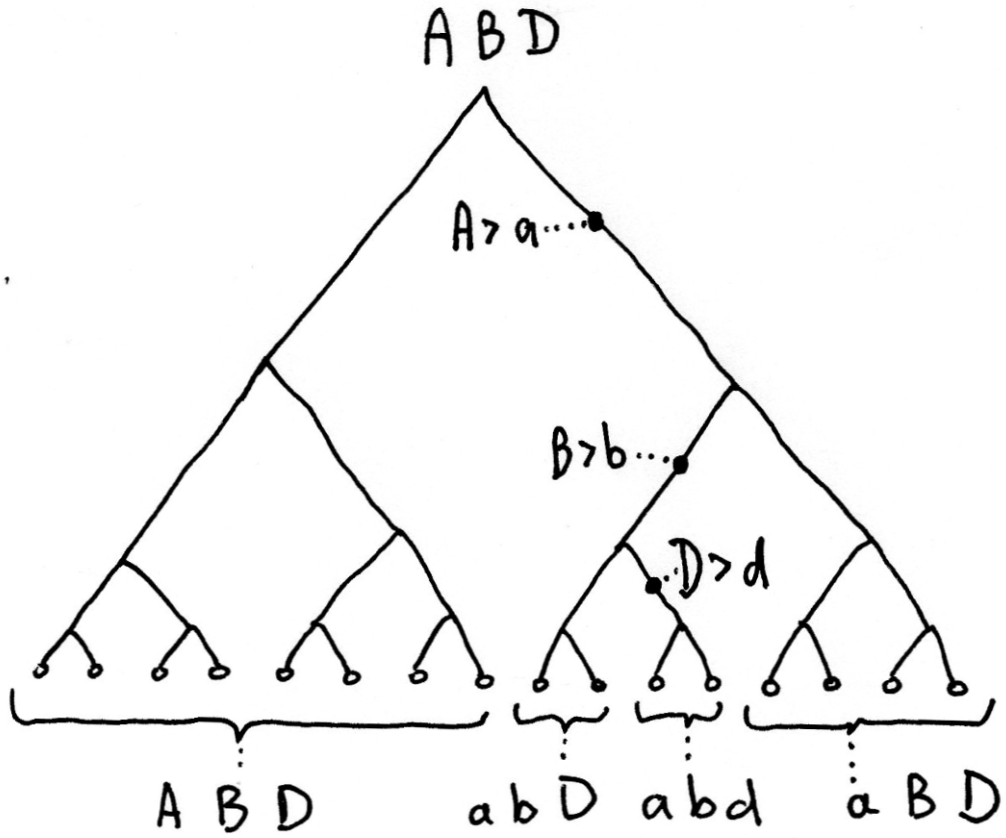
Haplotypes: now with recombination



Haplotype number	*Haplotype*	*Frequency*
0	A B D	8/16
1	a b D	2/16
2	a b d	2/16
3	a B D	4/16
-	A b D	0
-	A b d	0
-	A B d	0
-	a B d	0

Haplotypes: now with recombination

Mutation creates new alleles, recombination creates new allele combinations



A B D

a b D

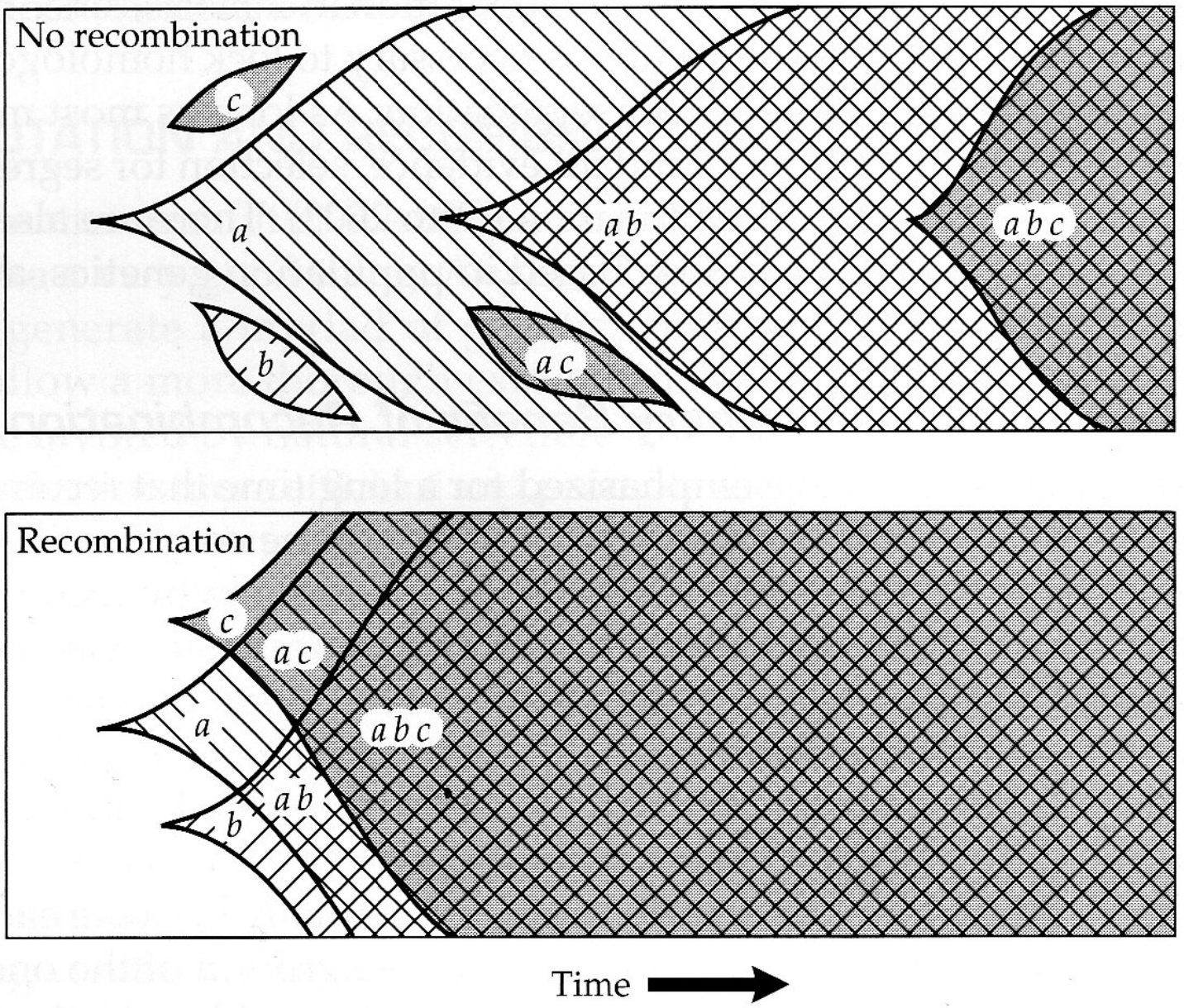
a b d

A B D

A b d

Exercise: which haplotypes recombined and where?

How to combine beneficial alleles?



Hartl, Clark – *Principles of population genetics*

Recombination: highlights

- Both double strand breaks (DSBs) and crossovers form hotspots, location: mostly intergenic. Not all DSBs) result in crossovers

- High individual variability; average: 150 male / 350 female DSBs, 50 male / 70 female crossovers per genome. Female/male ratio ~1.6

- Crossover sites are associated with: H3K4me3, nucleosome depletion, reduced DNA methylation

- 40% of crossover variation is due to *PRDM9* polymorphism. *PRDM9* is a zinc finger protein with histone methyltransferase activity that catalyzes histone H3 lysine 4 trimethylation (H3K4me3) during meiotic prophase

Zelkowski (2019) *Trends Genet*

Recombinations per meiosis

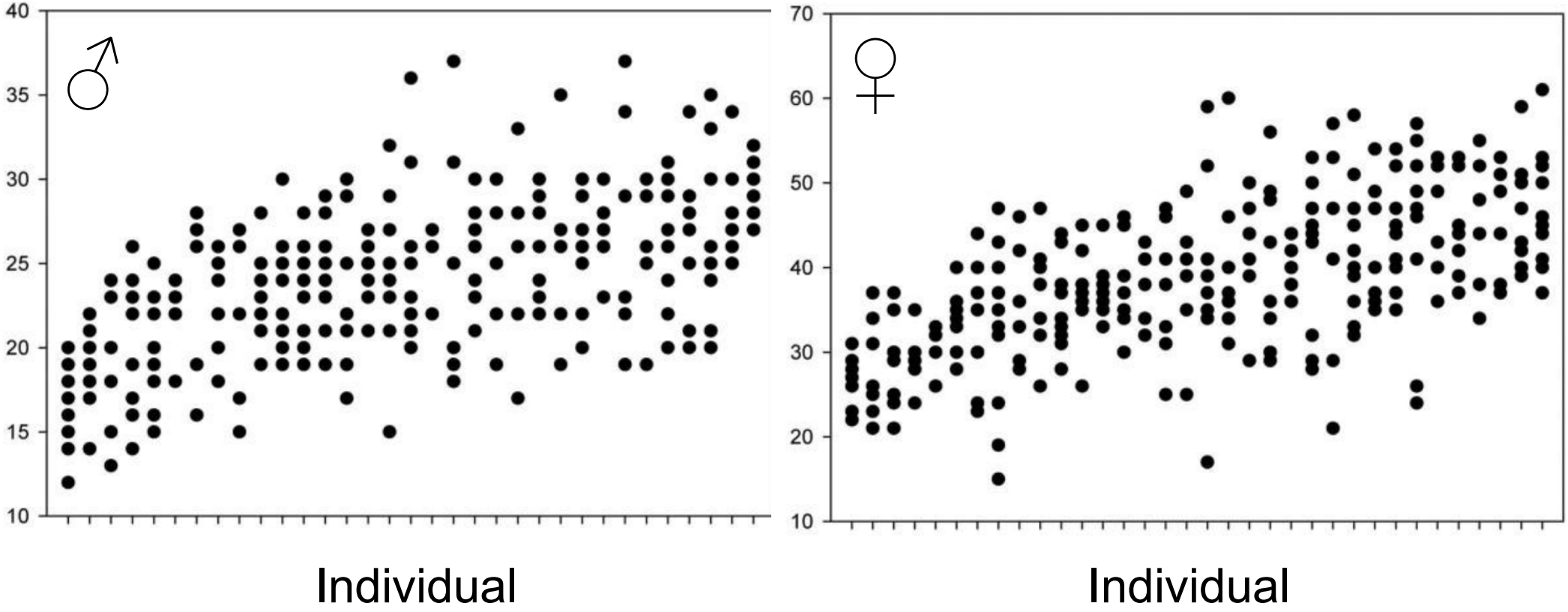
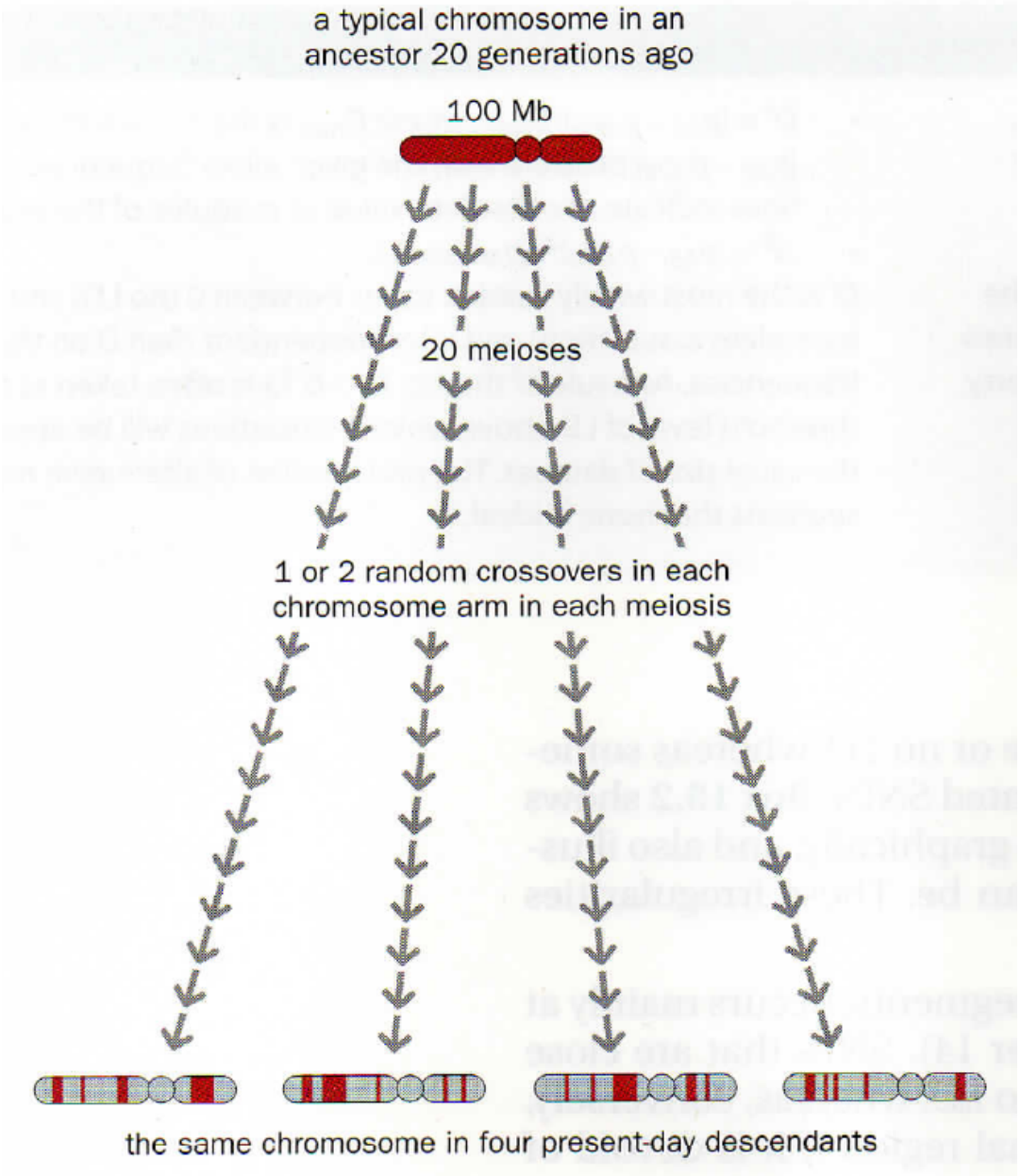


Figure 2. Individual variation in the number of recombination events per meiosis. The graphs show the number of recombination events per meiosis in each individual. The panels show data for men (*left*) and women (*right*). The number of recombination events per meiosis is shown as a dot. Individuals are arranged in ascending order of the average number of recombination events per meiosis.

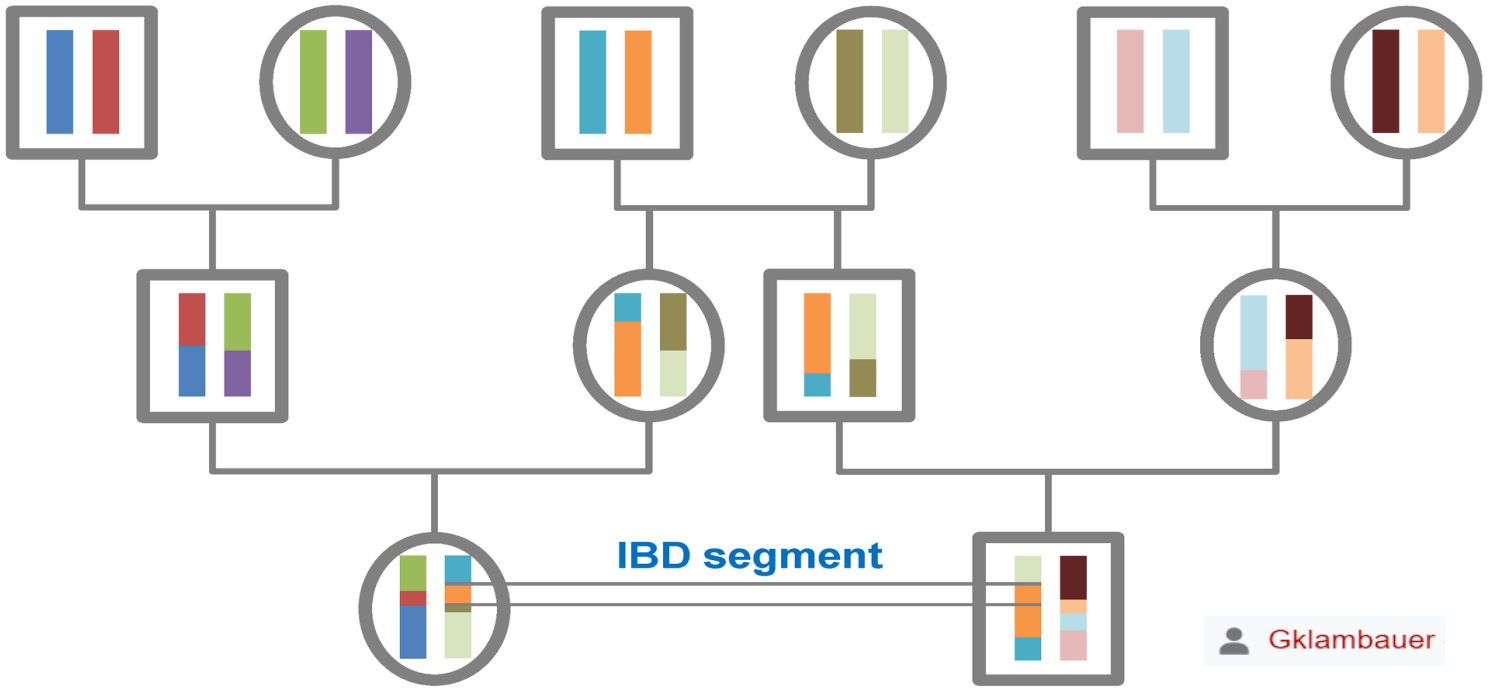
Cheung (2007) *Am J Hum Genet*

Shared ancestral chromosome segments



A typical chromosome is shown in a common ancestor, 20 generations ago, of four present-day individuals. There will be **1-2 random crossovers in each chromosome arm in each of the 20 meioses linking each present-day person to their common ancestor**. Only a small proportion of the sequence of the ancestor's chromosome will be inherited by descendants after 20 generations (red segments).

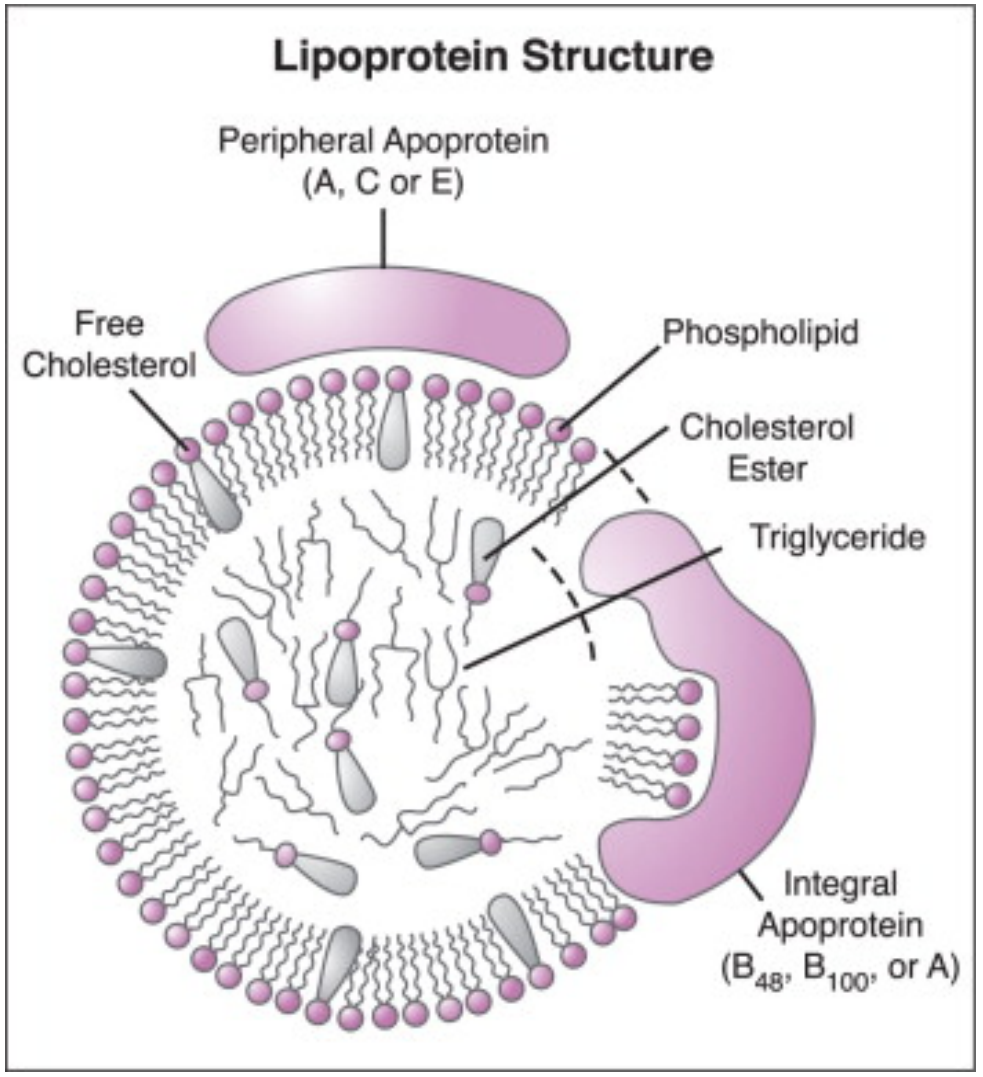
Shared ancestral chromosome segments



A DNA segment **is identical by state (IBS)** in two or more individuals if they have identical nucleotide sequences in this segment. An IBS segment is **identical by descent (IBD)** in two or more individuals if they have inherited it from a common ancestor without recombination, that is, the segment has the same ancestral origin in these individuals.

DNA segments that are IBD are IBS per definition, but segments that are *not* IBD can still be IBS due to the same mutations in different individuals or recombinations that do not alter the segment.

Lipoprotein particles



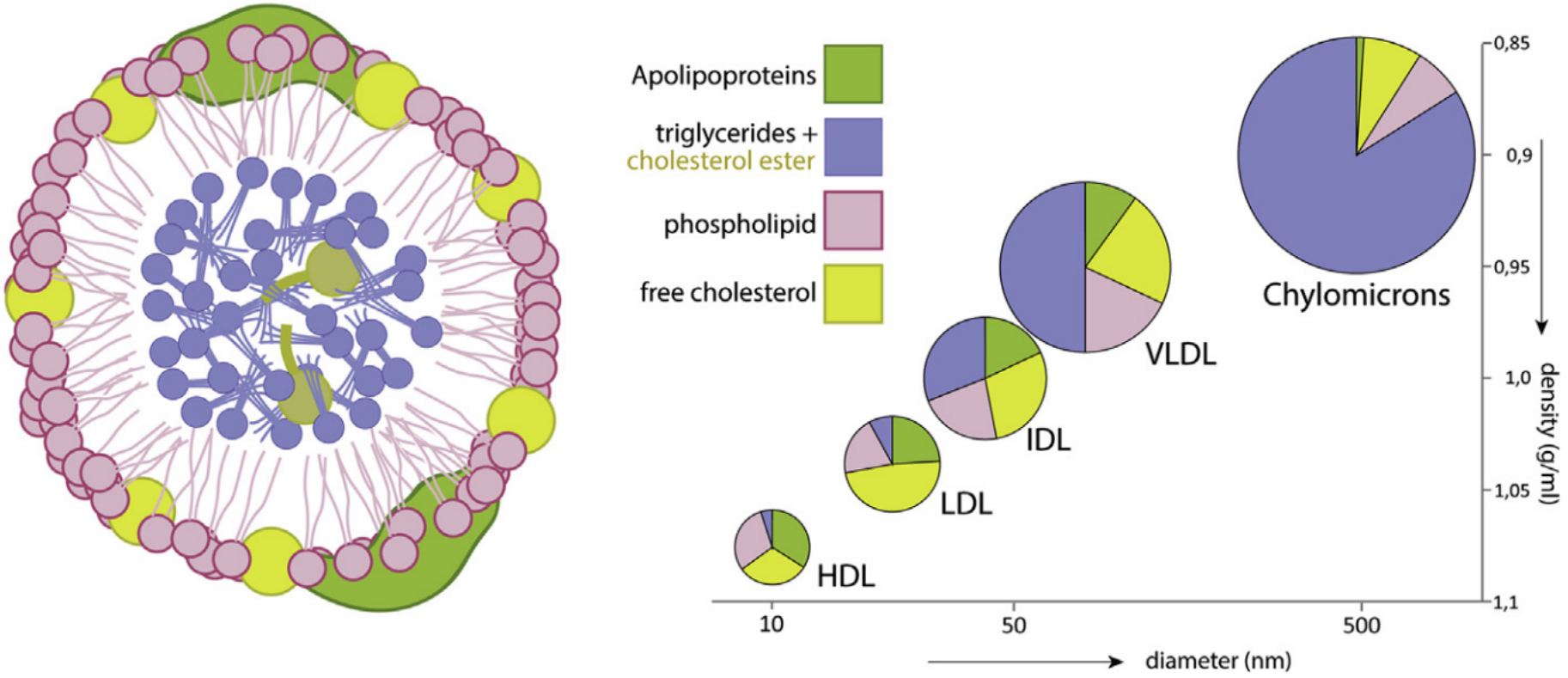
Lipoprotein: a particle that transports hydrophobic lipids in water, e.g. blood plasma.

Center: triglyceride, cholesterol

Outer shell: phospholipids, apolipoproteins ApoA, ApoB, ...

Engelking (2015) *Textbook of Veterinary Physiological Chemistry*

Lipoprotein particles

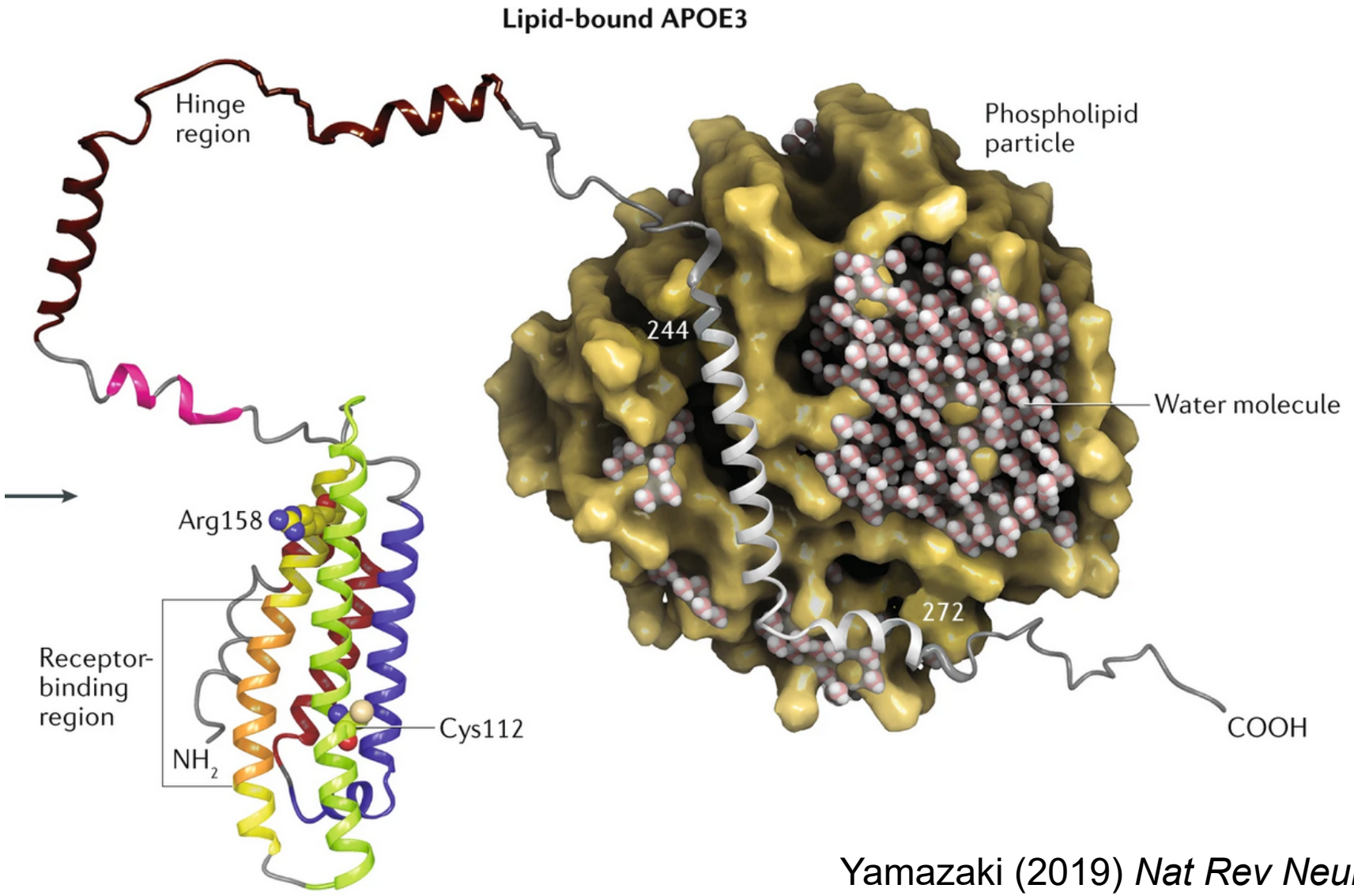


Composition and main physical-chemical properties of major lipoprotein classes Left: The outer shell of lipoproteins consists of a phospholipid and cholesterol, combined with apolipoproteins, which defines that type, function and/or destination of the lipoprotein. Hydrophobic lipids (triglycerides, cholesterol esters) are in the core of the lipoprotein. Right: Lipoproteins are classified according to their size, density and composition. HDL, high-density lipoprotein; LDL, low-density lipoprotein; IDL, intermediate-density lipoprotein; VLDL, very low-density lipoprotein.

van Leeuwen (2018) Prog Retin Eye Res

Apolipoprotein E

ApoE: a key regulator of plasma lipid levels; promotes clearance of TG-rich lipoproteins (chylomicrons and VLDL) from circulation



Yamazaki (2019) *Nat Rev Neurology*

Apolipoprotein E

	ApoE4 →	ApoE3 →	ApoE2
Haplotype	Arg112, Arg158	Cys112, Arg158	Cys112, Cys158
NFE frequency	14.9%	77.5%	7.6%
Functional	Normal binding to LDLR, stronger biding to VLDL, weaker binding to HDL	Normal binding to LDLR and lipids	Reduced LDLR binding → impaired clearance of chylomicron and VLDL remnants
Biochemical	Pro-atherogenic lipoprotein distribution	Normal plasma lipid levels and TG clearance	Increased plasma TG and cholesterol
Clinical	Premature atherosclerosis, ischemic heart disease, Alzheimer's disease	Anti-atherogenic	Familial type III hyperlipoproteinemia, premature atherosclerosis, ischemic heart disease. Protective against Alzheimer's disease

Atherogenesis: plaque development in arteries. **Hyperlipoproteinemia type III**, aka dysbetalipoproteinemia: hyperlipidemia due to accumulation of remnants of the TG-rich lipoproteins: very low density lipoproteins (VLDL) and chylomicrons.

Apolipoprotein E

	ApoE4 →	ApoE3 →	ApoE2
Haplotype	Arg112, Arg158	Cys112, Arg158	Cys112, Cys158
NFE frequency	14.9%	77.5%	7.6%
Frequency in the Ivanovo region	11.8%	79.8%	8.4%

Atherogenesis: plaque development in arteries. **Hyperlipoproteinemia type III**, aka dysbetalipoproteinemia: hyperlipidemia due to accumulation of remnants of the TG-rich lipoproteins: very low density lipoproteins (VLDL) and chylomicrons.

Apolipoprotein E

Isoform \ Position	112	158
ApoE4	Arg (C)	Arg (C)
ApoE3	Cys (T)	Arg (C)
ApoE2	Cys (T)	Cys (T)

Apolipoprotein E

Isoform \ Position	112	158
ApoE4	Arg (C)	Arg (C)
ApoE3	Cys (T)	Arg (C)
ApoE2	Cys (T)	Cys (T)
ApoE1	Arg (C)	Cys (T)

Exercise: determine haplotype combination by actual genotypes

Apolipoprotein E

Genotype	Carriers	LDL, mmol/l	HDL, mmol/l	TG, mmol/l
E3/E3	1013	3.30	1.41	1.21
E2/E3	215	2.64	1.34	1.21
E2/E2	13	2.15	1.23	**2.25**
E3/E4	295	3.47	1.36	1.17
E2/E4	33	2.82	1.30	1.38
E4/E4	20	**4.12**	1.38	1.45

Recombination probability and genetic distance

Genetic distance between two loci measures how frequently these loci recombine. The genetic distance d between two chromosomal loci equals **one centimorgan (cM)** if probability of crossover in a single meiosis is 0.01

- cM is a unit of genetic distance, 1cM ≈ 1Mbp (*physical distance*)

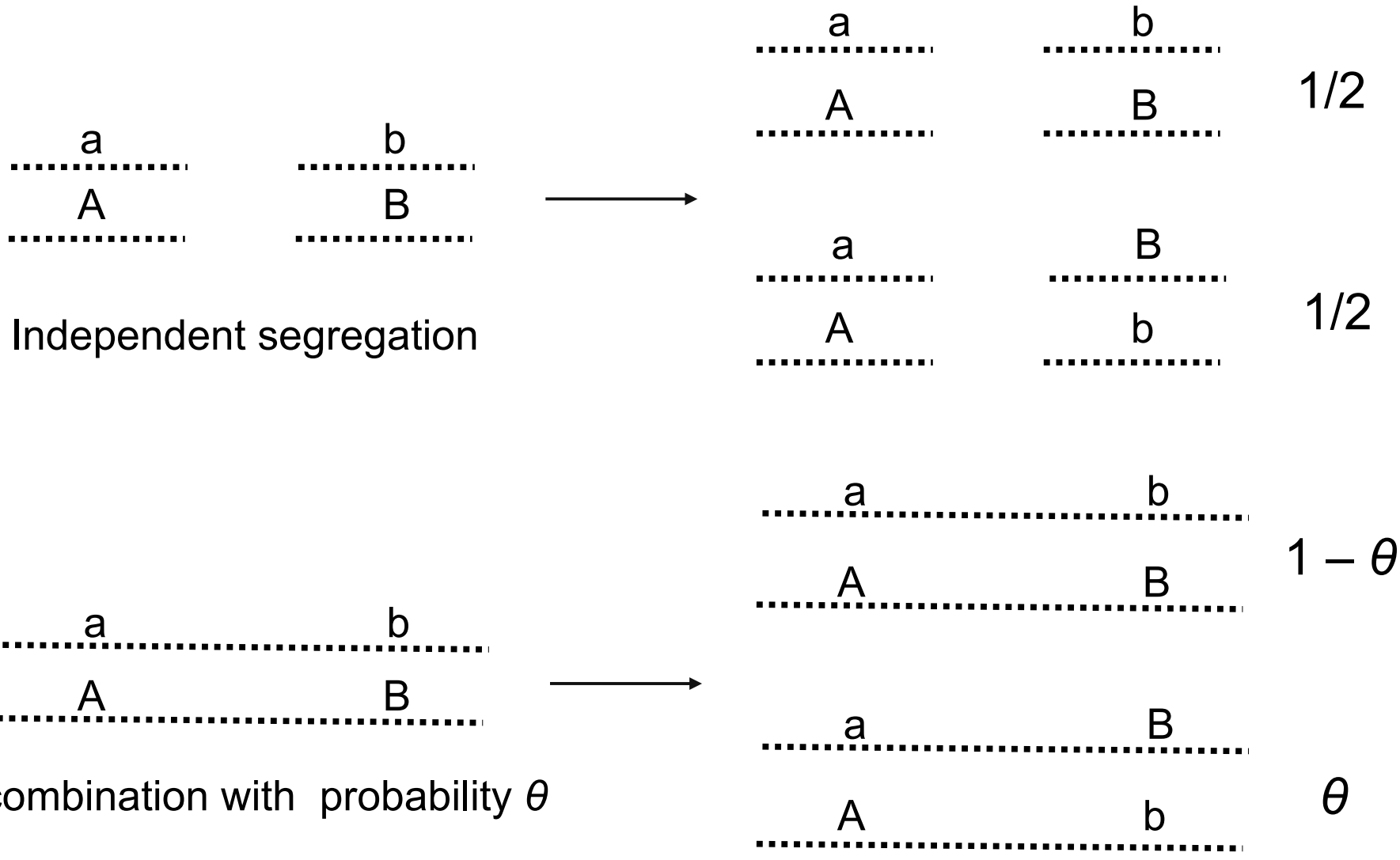
Haldane function: genetic distance $x \rightarrow$ recombination probability θ

- (a) Probability of k crossovers within a region of unit genetic distance x is $e^{-x}x^k/k!$
- (b) Two crossovers $\Longrightarrow$ no recombination, so **recombination rate**, or probability $\theta(x) = e^{-x}x + e^{-x}x^3/3! + e^{-x}x^5/5! \ldots = e^{-x}(x + x^3/3! + x^5/5! + \ldots) = e^{-x}\mathrm{sh}(x) = e^{-x}(e^x - e^{-x})/2 = (1 - e^{-2x})/2 = (1 - e^{-2d/100})/2$, d is measured in cM

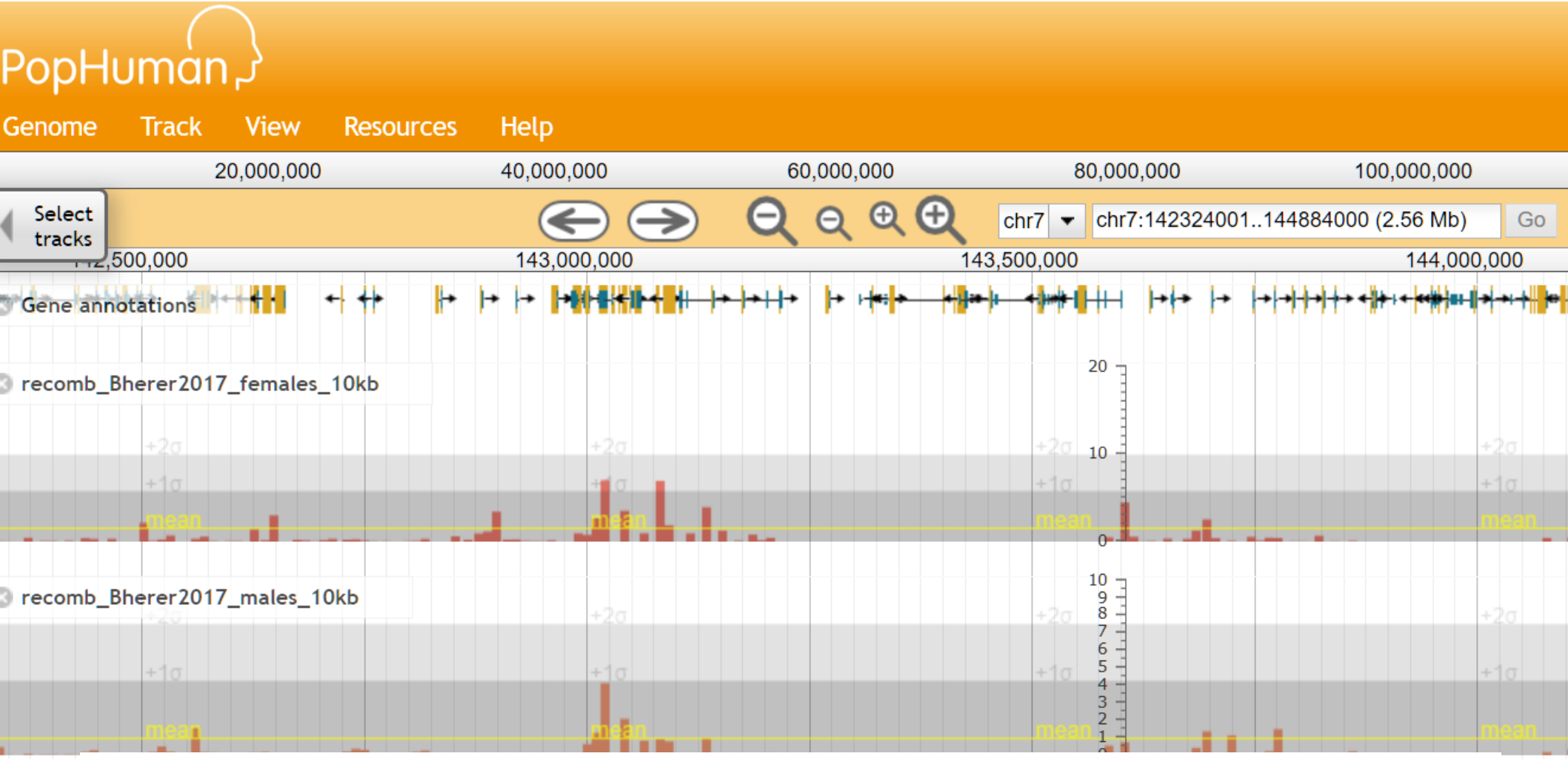
- Features of recombination rate: $0 \le \theta \le 1/2$, $\theta \approx x$ for $x \approx 0$, $\theta = 0.22$ for $x = 0.3$, $\theta \approx 1/2$ for $x \rightarrow \infty$

Exercise: draw $\theta(x)$

Gamete frequencies



Recombination frequencies (cM/Mb)



Received 8 Apr 2016 | Accepted 20 Feb 2017 | Published 25 Apr 2017

DOI: 10.1038/ncomms14994 OPEN

Refined genetic maps reveal sexual dimorphism in human meiotic recombination at multiple scales

Claude Bhérer[1,†], Christopher L. Campbell[1] & Adam Auton[1]

https://pophuman.uab.cat

Linkage disequilibrium

LD: non-random association of alleles at two loci

Alleles: A, a; B, b. Frequencies: $P_A + P_a = 1$; $P_B + P_b = 1$

Haplotypes: AB, Ab, aB, ab. Frequencies: $P_{AB} + P_{Ab} + P_{aB} + P_{ab} = 1$

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

Complete equilibrium:

$P_A = P_a = P_B = P_b = ½$

$P_{AB} = P_{Ab} = P_{aB} = P_{ab} = ¼$

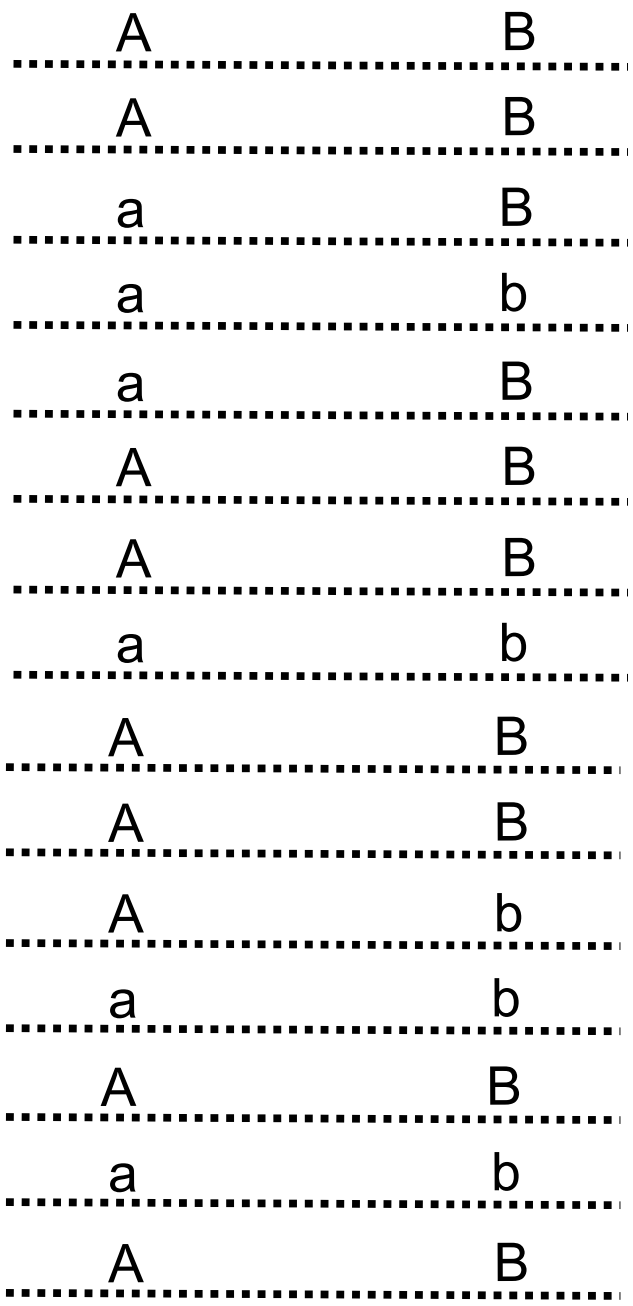
A	B
A	B
A	B
A	B
a	b
a	b
a	b
a	b

Complete disequilibrium:

$P_A = P_a = P_B = P_b = ½$

$P_{AB} = P_{ab} = ½, \quad P_{aB} = P_{Ab} = 0$

Linkage disequilibrium measures



Raw LD coefficient:

$$D_{AB} = P_{AB} - P_A P_B$$

Lewontin's LD coefficient:

$$D' = D_{AB}/D_{max}, \quad \text{where}$$

$$D_{max} = \begin{cases} \max\{-p_A p_B,\ -(1-p_A)(1-p_B)\} & \text{when } D < 0 \\ \min\{p_A(1-p_B),\ (1-p_A)p_B\} & \text{when } D > 0 \end{cases}$$

Correlation coefficient

$$r_{AB} = \frac{D_{AB}}{\sqrt{p_A p_a p_B p_b}},$$

Exercise: for the example to the left, calculate

- actual haplotype frequencies
- expected haplotype frequencies with no LD
- LD coefficients

Linkage disequilibrium decay in time

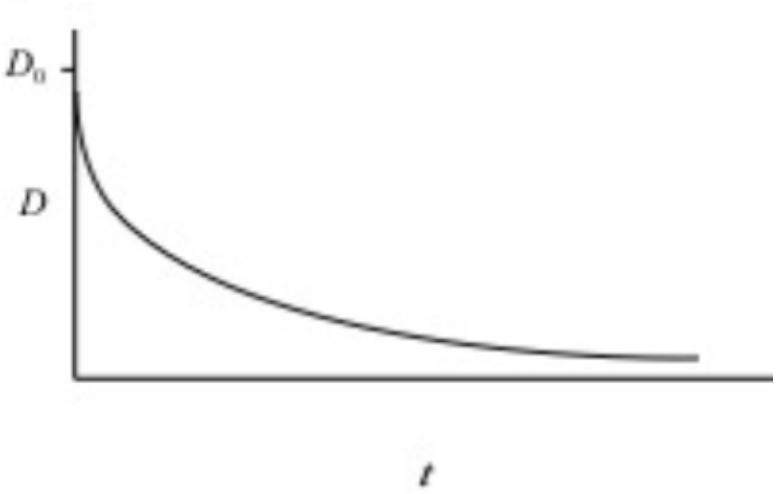
Recombination rate θ vs linkage D?

LD coefficient decays each generation at a rate determined by the degree of recombination:

$$D_{t+1} = (1-\theta)D_t$$

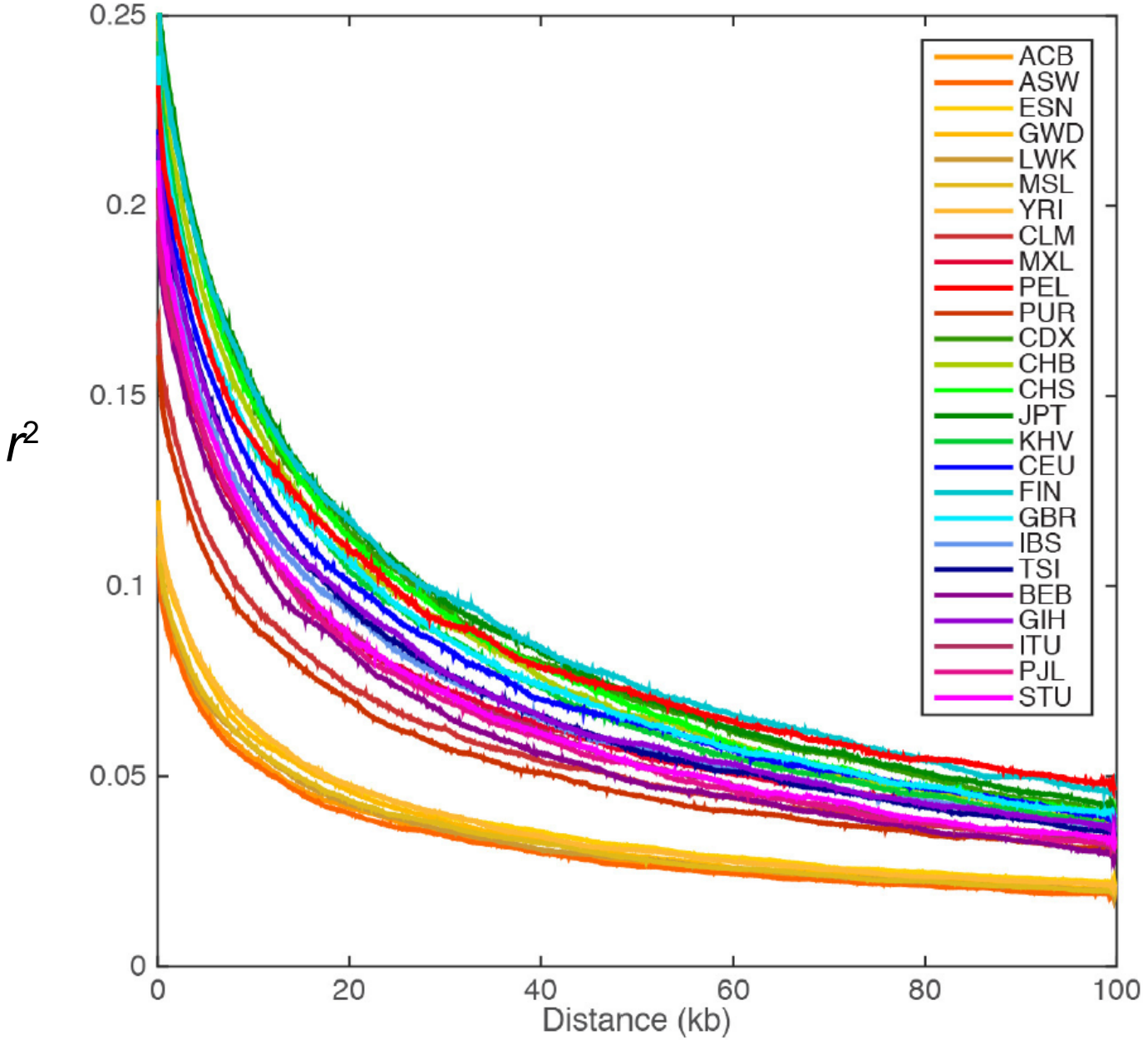
$$D_t = (1-\theta)^t D_0$$

$$D_t = e^{-\theta t} D_0 \; if \; \theta \approx 0$$



Exercise: suppose two loci separated by 500 Kbp are in partial linkage with D = 0.1. Give an estimate of how many generations it may take to reduce D to 0.05.

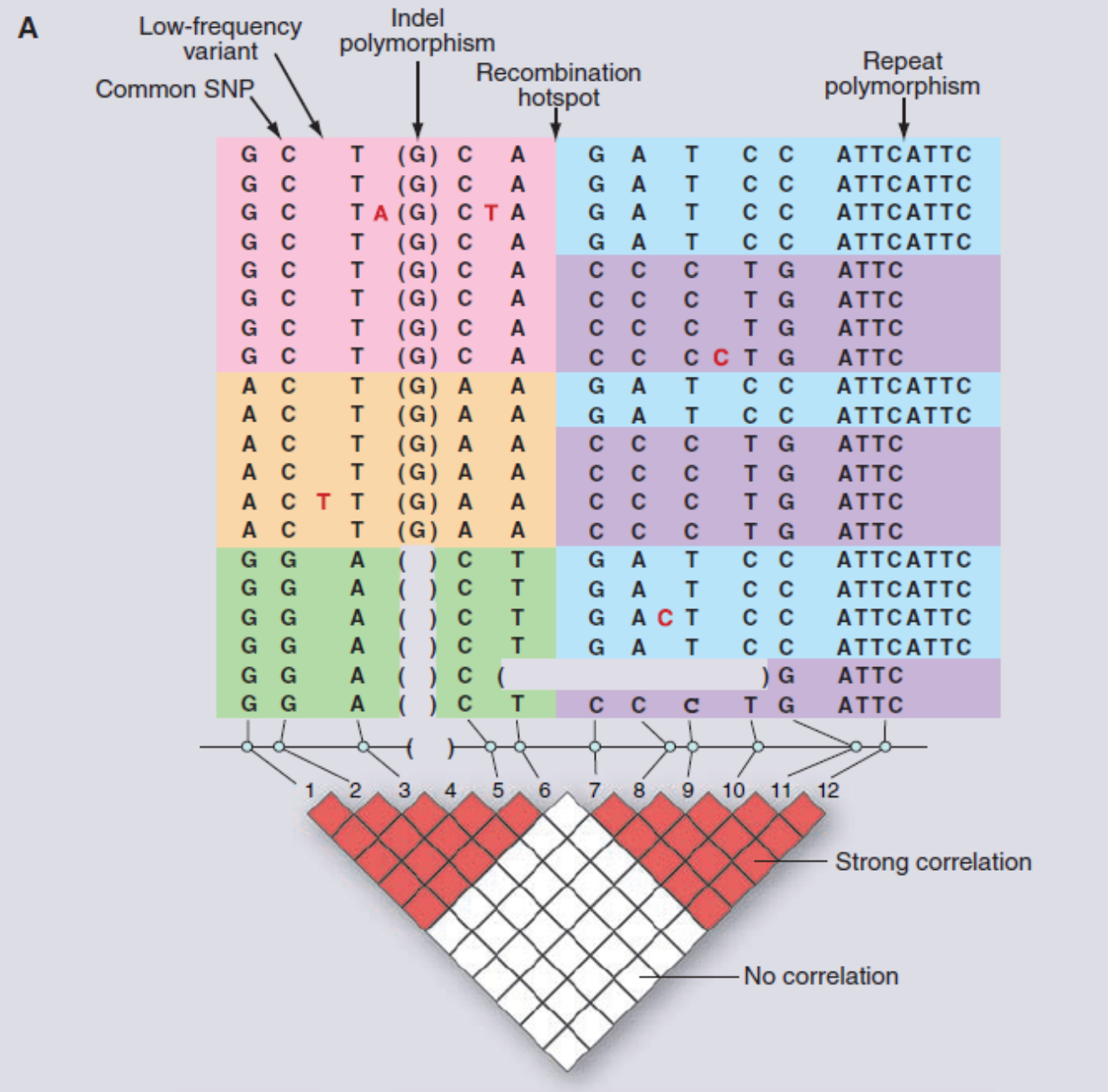
Linkage disequilibrium decay in space



Exercise: what are the populations with the lowest r^2 values?

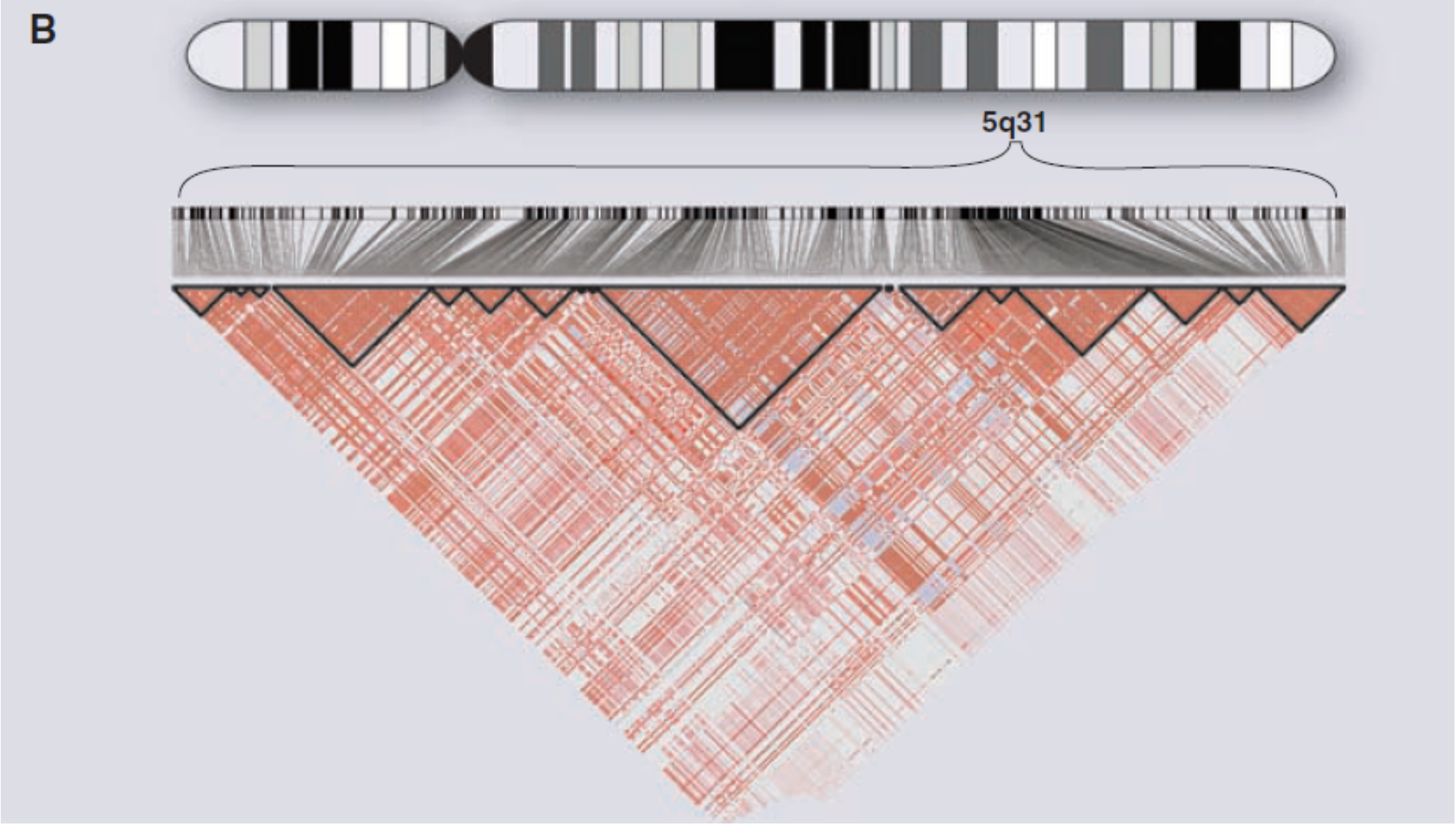
The 1000 Genomes Project Consortium (2015) *Nature* doi:10.1038/nature15393

Haplotypes: another realistic example



Altshuler (2008) *Science*

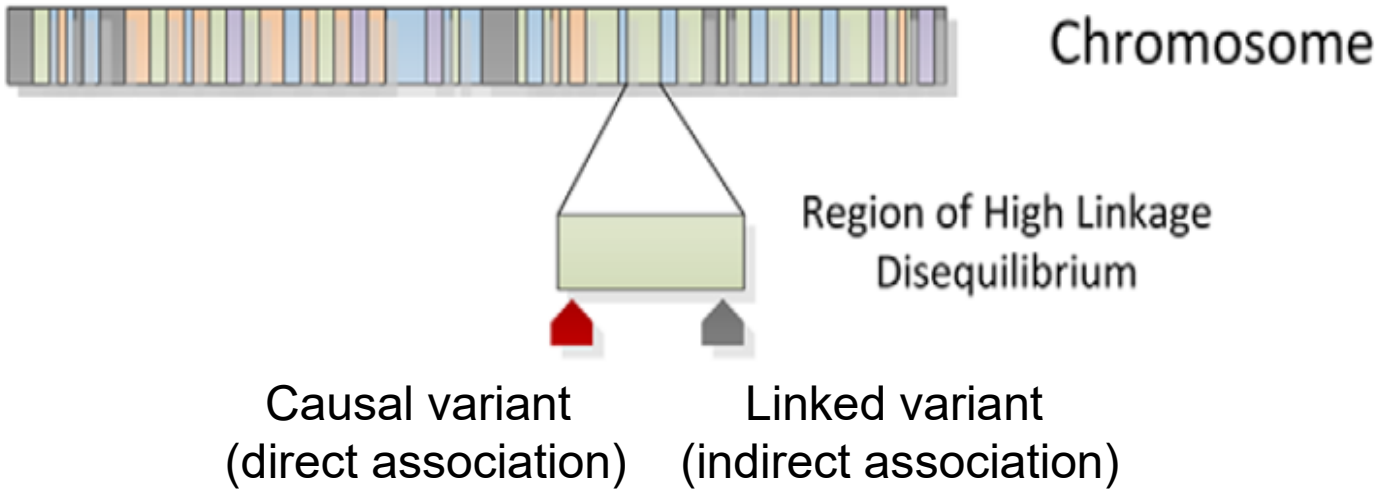
Haplotypes: another realistic example



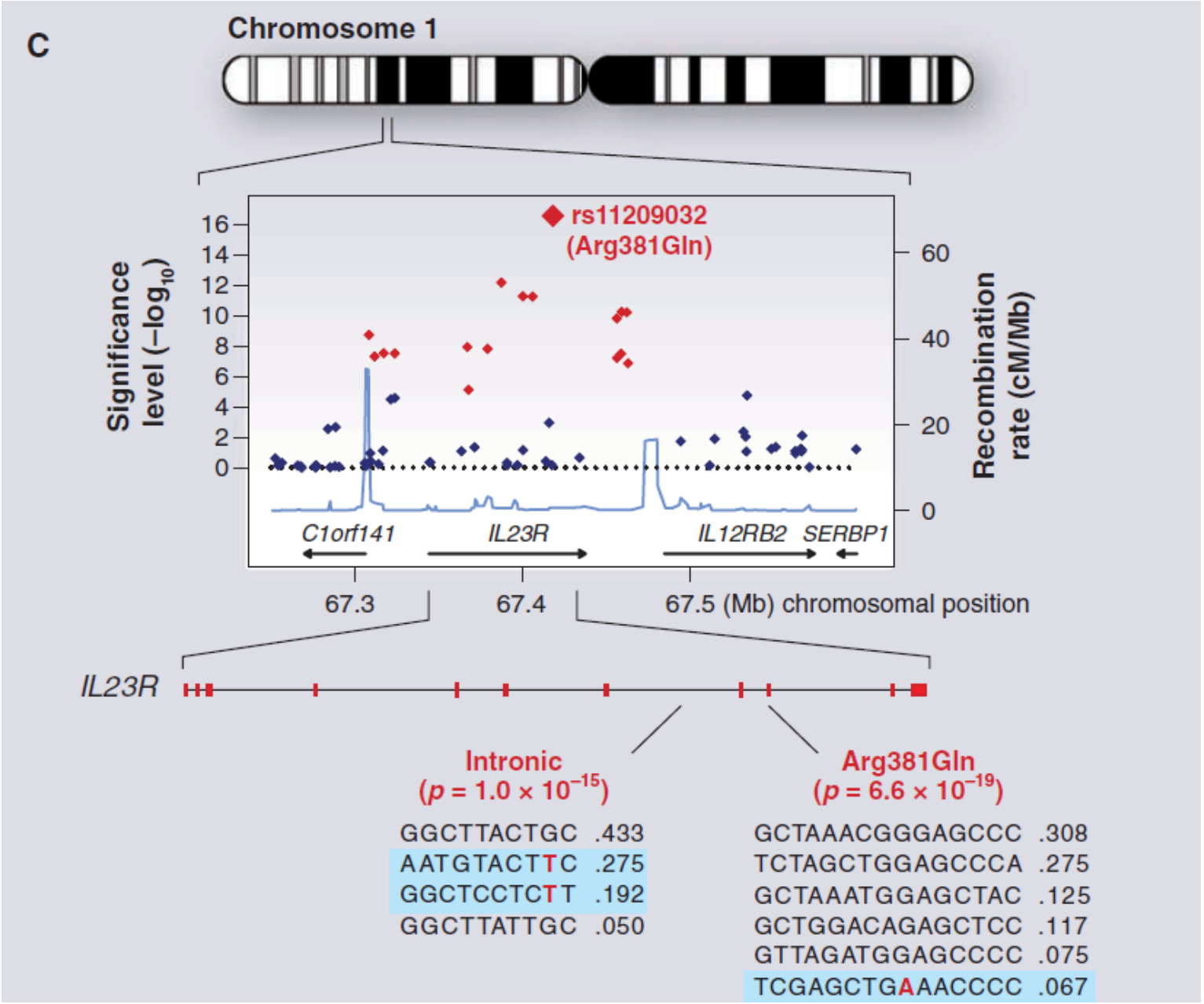
Altshuler (2008) *Science*

Why knowing linkage is important?

- Genome segments tend to remain together through many generations.
- A causal allele is transmitted along with other alleles that can be used as tags, or markers.
- However, this complicates identification of the truly causal allele and its direct association with phenotype



Why knowing linkage is important?



Altshuler (2008) *Science*

Summary

- Allele transmission obeys the Mendel's law. In some cases we can inference the paternal or maternal origin of an allele.
- Haplotypes are allele combinations
- Meiosis halves the cell's ploidy and introduces genetic diversity by independent segregation and recombination
- There is considerable variation in crossover rates. On average there are 50 male and 70 female crossovers per genome
- Genome fragments may be identical by descent or only by state IBS
- Genetic distance between two loci measures how frequently these loci recombine.
- Recombination destroys linkage of genomic loci
- Knowing linkage structure in a region is important for analysis of association

Further reading

- The International HapMap Consortium (2005). A haplotype map of the human genome. *Nature* 437, 1299–1320.
- Cheung, V.G., Burdick, J.T., Hirschmann, D., and Morley, M. (2007). Polymorphic Variation in Human Meiotic Recombination. *Am J Hum Genet* 80, 526–530.
- Altshuler, D., Daly, M.J., and Lander, E.S. (2008). Genetic Mapping in Human Disease. *Science* 322, 881–888.
- The 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature* 526, 68–74.
- Bhérer, C., Campbell, C.L., and Auton, A. (2017). Refined genetic maps reveal sexual dimorphism in human meiotic recombination at multiple scales. *Nat Commun* 8, 1–9.
- Zelkowski, M., Olson, M.A., Wang, M., and Pawlowski, W. (2019). Diversity and Determinants of Meiotic Recombination Landscapes. *Trends Genet*. 35, 359–370
- https://en.wikipedia.org/wiki/Identity_by_descent