Chapter 10

Human Gene Mapping and Disease Gene Identificaton

Paul Coucke

Wouter Steyaert

Center for Medical Genetics

- The genetic landscape of the human genome
- Mapping human genes by linkage analysis
- Mapping of complex traits
- From gene mapping to gene identification

- The genetic landscape of the human genome

1. Independent assortement and homologous recombination in meiosis
2. Recombination frequency and map distance
3. Linkage equilibrium and disequilibrium
4. The hapMap

- Mapping human genes by linkage analysis

Theory

Practise

1. Interprete microsatellite results
2. Add genotypes to pedigrees
3. Create pedigree and genotype files
4. Calculate and interprete LOD-scores
5. Delineate linkage intervals

Importance of gene mapping :

- Immediate clinical application as it can be used in **prenatal diagnosis, presymptomatic diagnosis and carrier testing**.

- A first step in the identification of a disease gene (**positional cloning**).

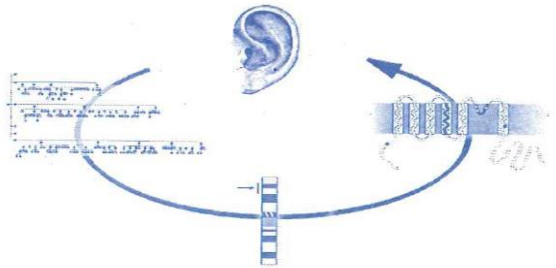
- An opportunity to characterize the disorder as to the extent for example of **locus heterogeneity**.

-Makes it possible to characterize the gene itself and the mutations involved resulting in a better understanding of **disease pathogenesis.**

UNIVERSITEIT ANTWERPEN
Faculteit wetenschappen
Departement Biochemie

LOCALISATION AND IDENTIFICATION OF GENES FOR DEAFNESS

LOKALISATIE EN IDENTIFICATIE VAN GENEN VOOR DOOFHEID



Proefschrift voorgelegd tot het behalen van de graad van
doctor in de wetenschappen aan de Universitaire Instelling Antwerpen
te verdedigen door **Paul COUCKE**

promotoren: Professor Dr. Patrick Willems
Professor Dr. Guy Van Camp

Antwerpen, 2000

Importance of gene mapping :

- Immediate clinical application as it can be used in **prenatal diagnosis, presymptomatic diagnosis and carrier testing**.

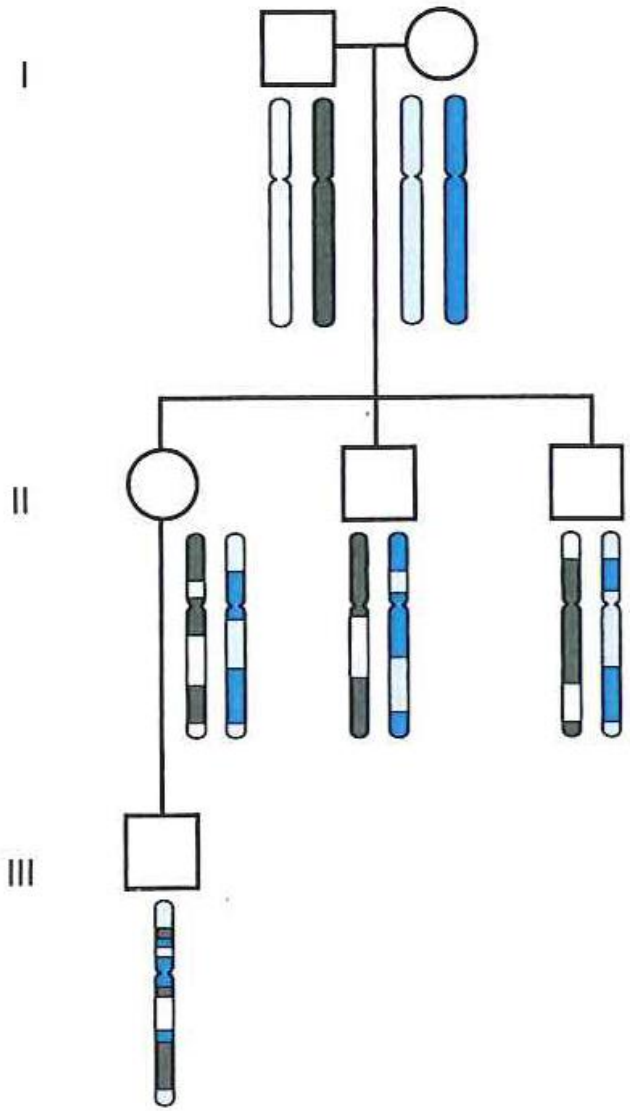
- A first step in the identification of a disease gene (**positional cloning**).

- An opportunity to characterize the disorder as to the extent for example of **locus heterogeneity**.

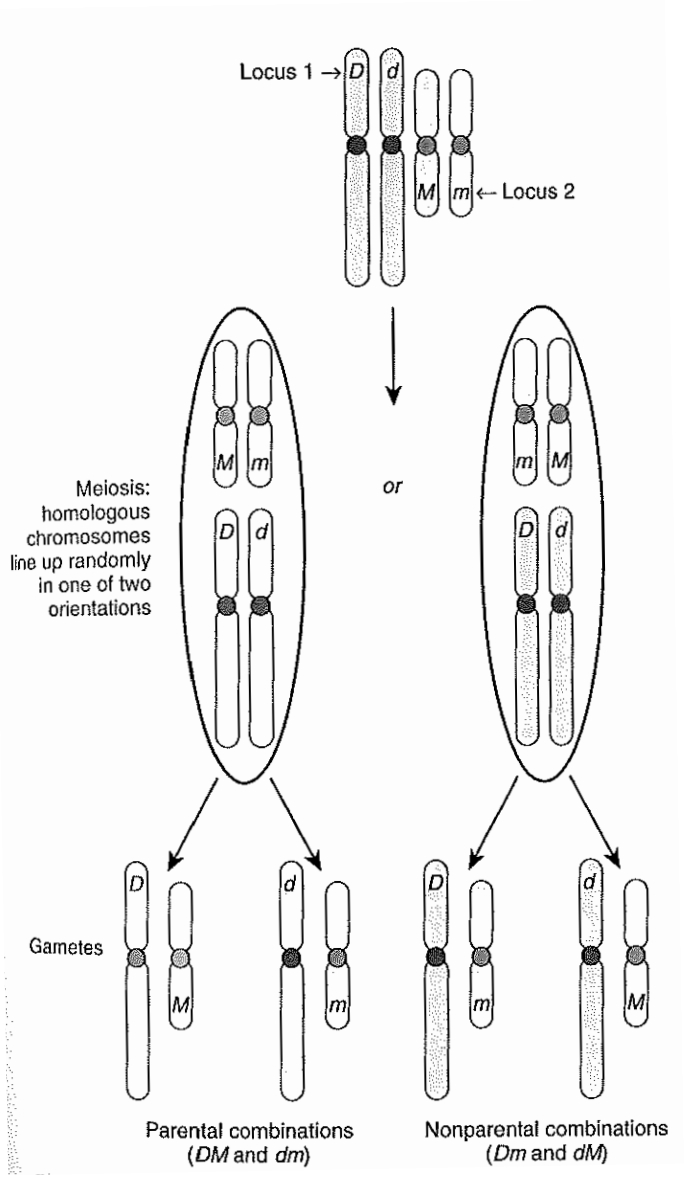
-Makes it possible to characterize the gene itself and the mutations involved resulting in a better understanding of **disease pathogenesis.**

The genetic landscape of the human genome

recombination in meiosis

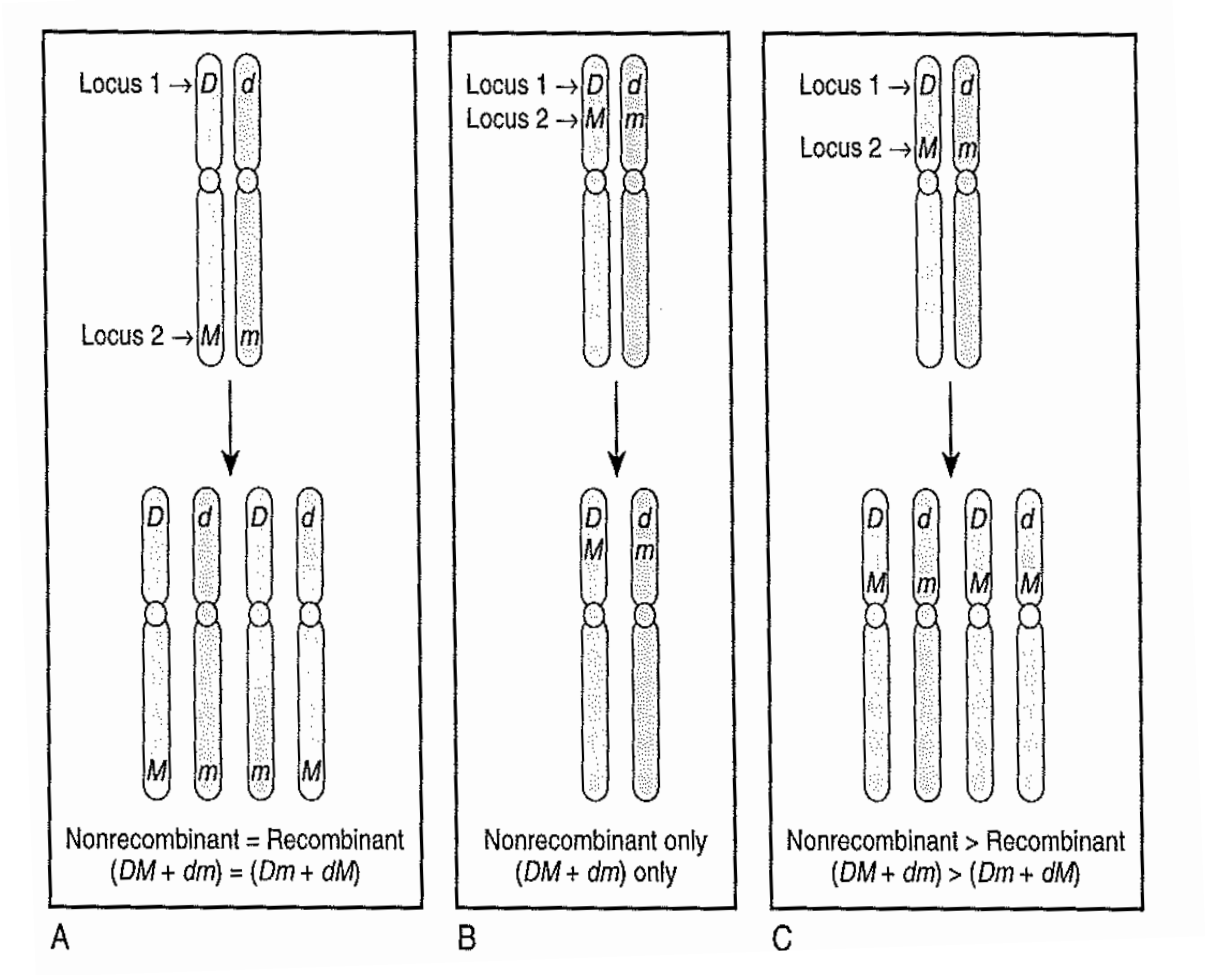


recombination in meiosis



Alleles at loci on different chromosomes assort independently

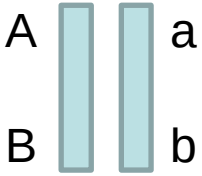
Recombination frequency (theta)

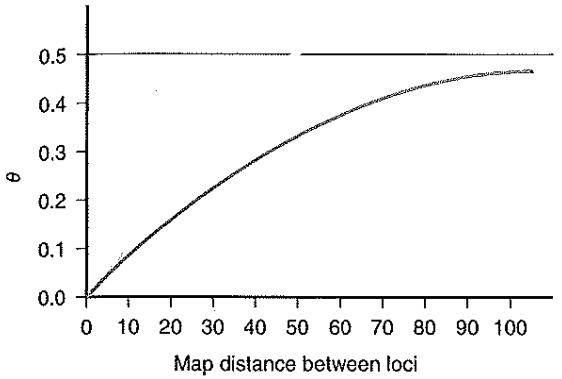


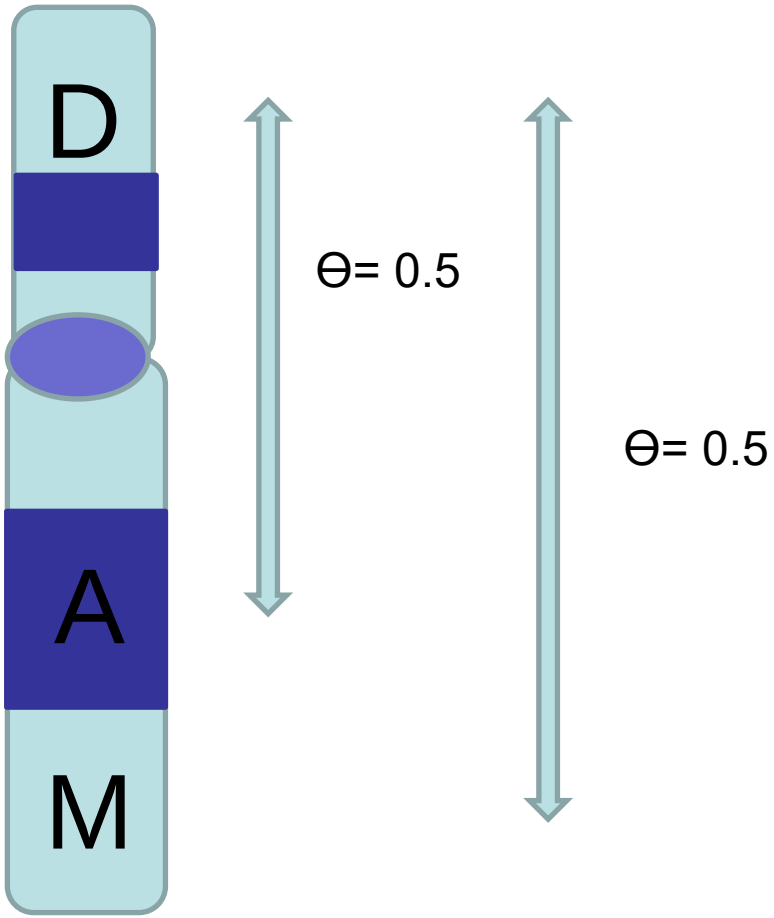
The amount of recombinations between two loci is therefore a measure for the distance between these two loci.

Recombination frequency

$$\Theta = \frac{\text{Total amount of recombinants}}{\text{Total amount of recombinants + Total amount of non-recombinants}}$$

Parent	Gametes	Theta
	50% non-rec and 50% rec	0.5
	90% non-rec and 10% rec	0.1
	99% non-rec and 1% rec	0.01
	100% non-rec	0




D
A
M
Θ= 0.5
Θ= 0.5

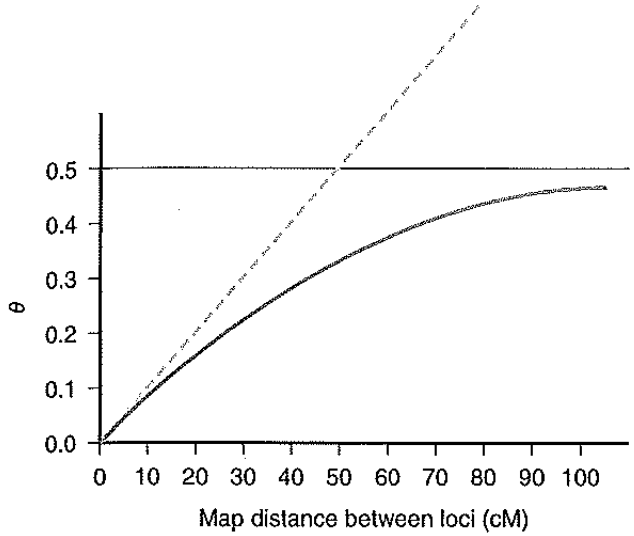
Genetic distance

Genetic distance = the genetic length over which one crossover occurs in 1% of meiosis. This distance is expressed in cMorgan.

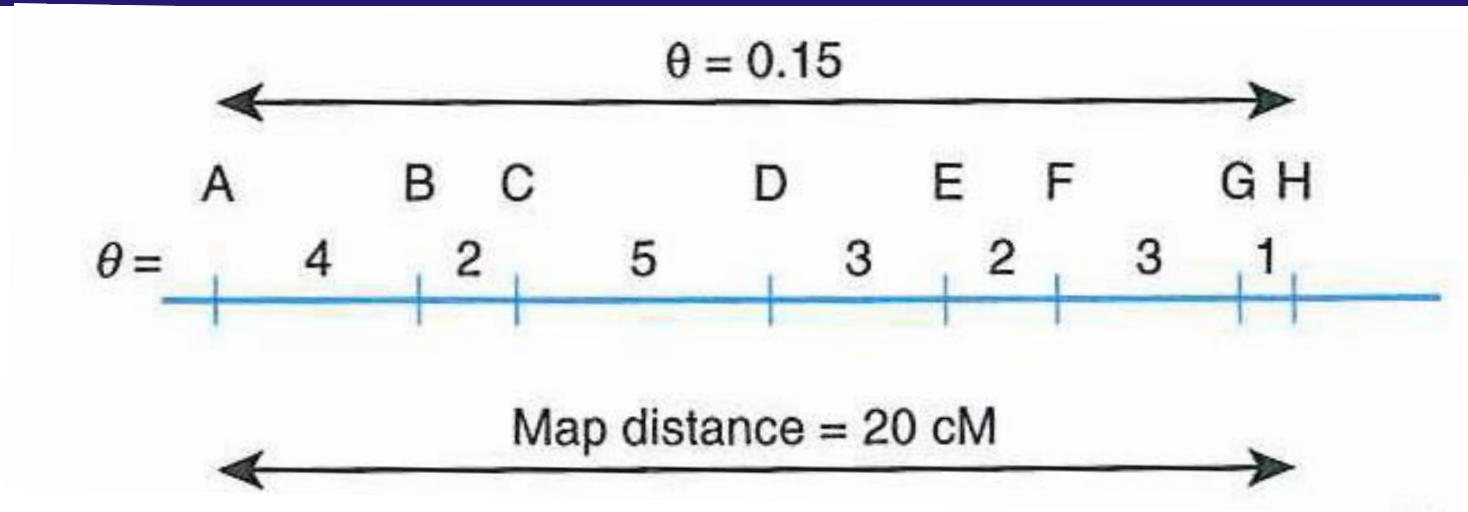
1 cMorgan = 0.01 recombinants = average of 1Mb (physical distance)

(Assuming that the recombination frequency is uniform along the chromosomes)

As double recombinants occur the further two loci are, the frequency of recombination does not increase proportionately.



recombination in meiosis



A
B

a
b

96 non-rec

A
H

a
h

80 non-rec

A
b

a
B

4 rec.

A
h

a
H

15 rec

A
H

a
h

5 double rec

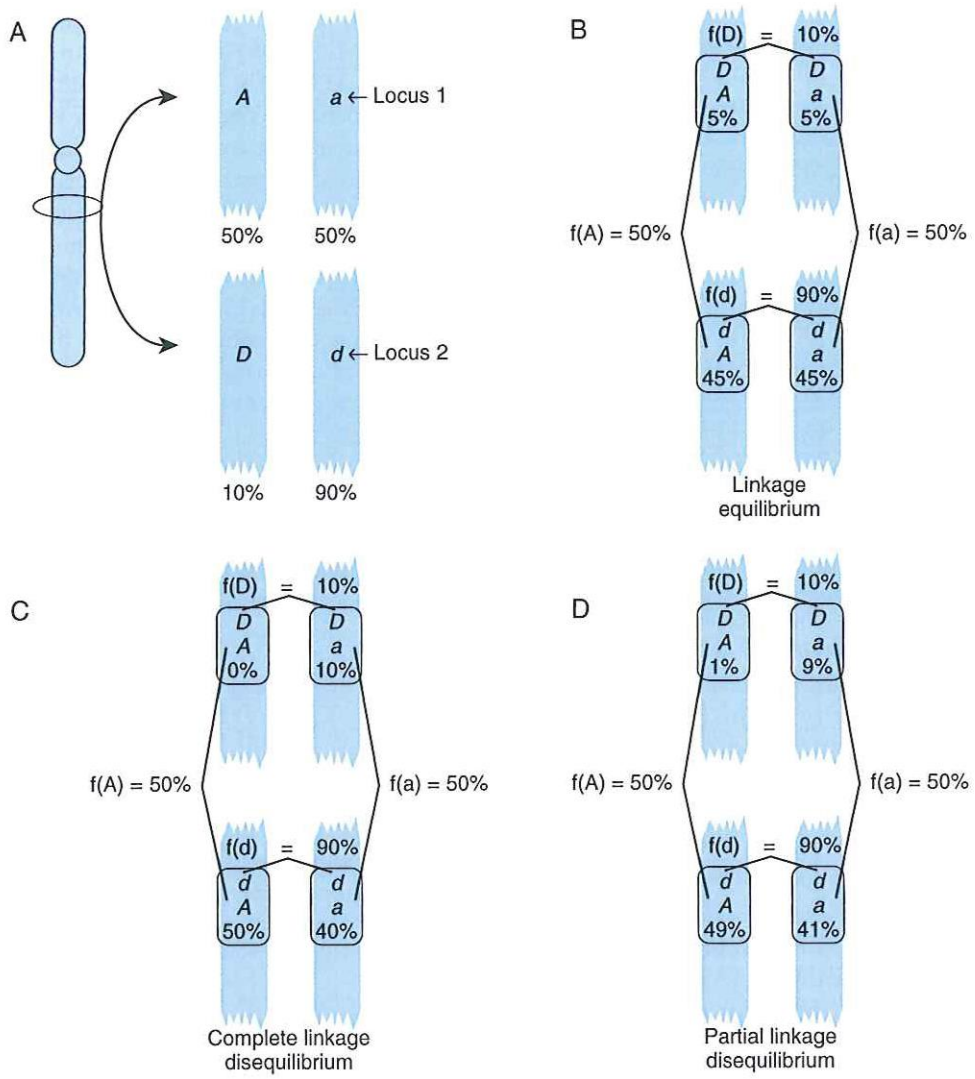
Conclusion :

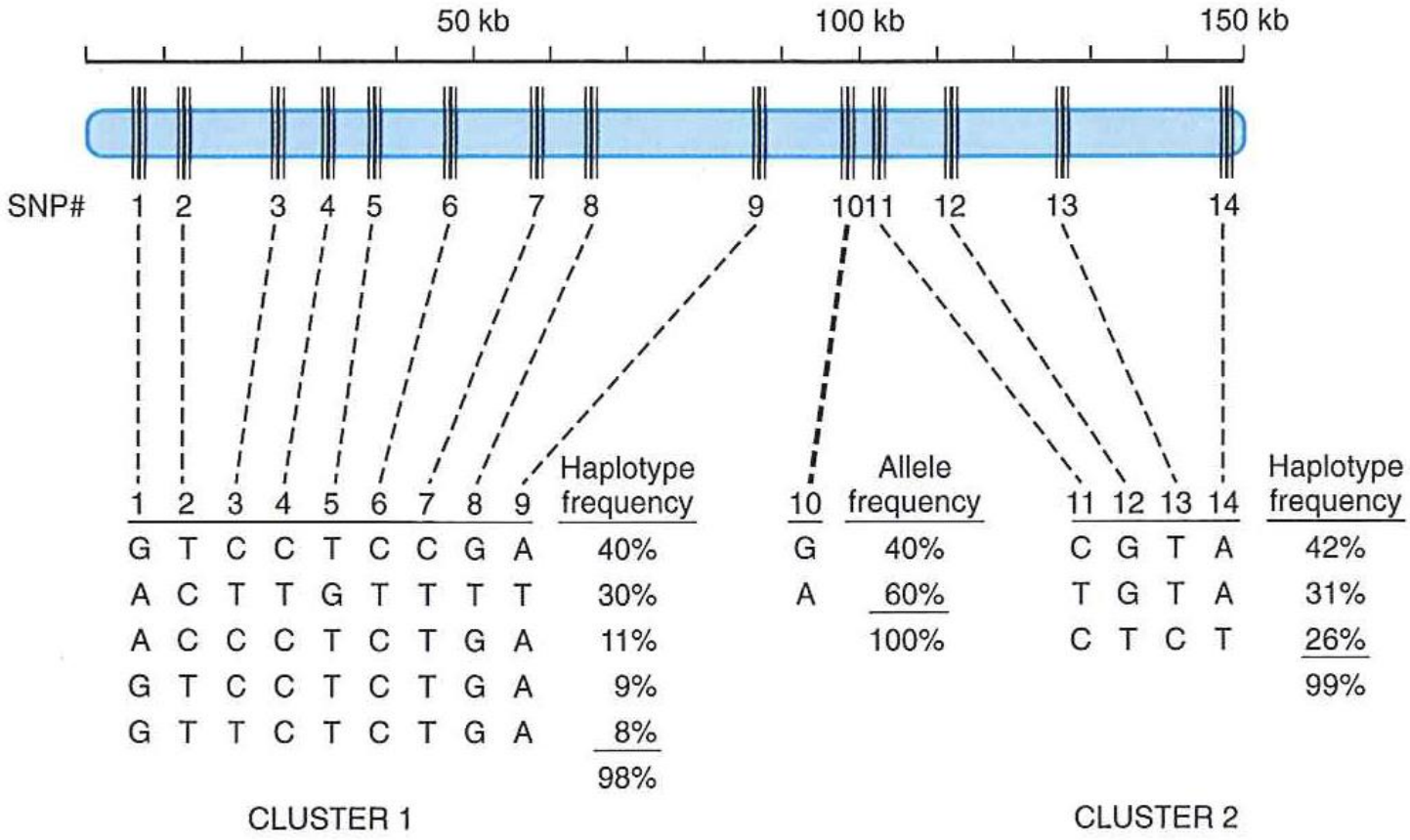
Values of theta or genetic distance are only reliable

if two loci are in the proximity of each other (max of 10 cM)

	Physical dist.	Genetic dist.
Chromosome 1 :	283 Mb	270 cM (0.95 cM/Mb)
q arm of chromosome 21:	30 Mb	62 cM (2.1 cM/Mb)
Human genome	3200 Mb	3615 cM (1.13 cM/Mb)
Female genome		4460 cM
Male genome		2590 cM

Linkage equilibrium and disequilibrium

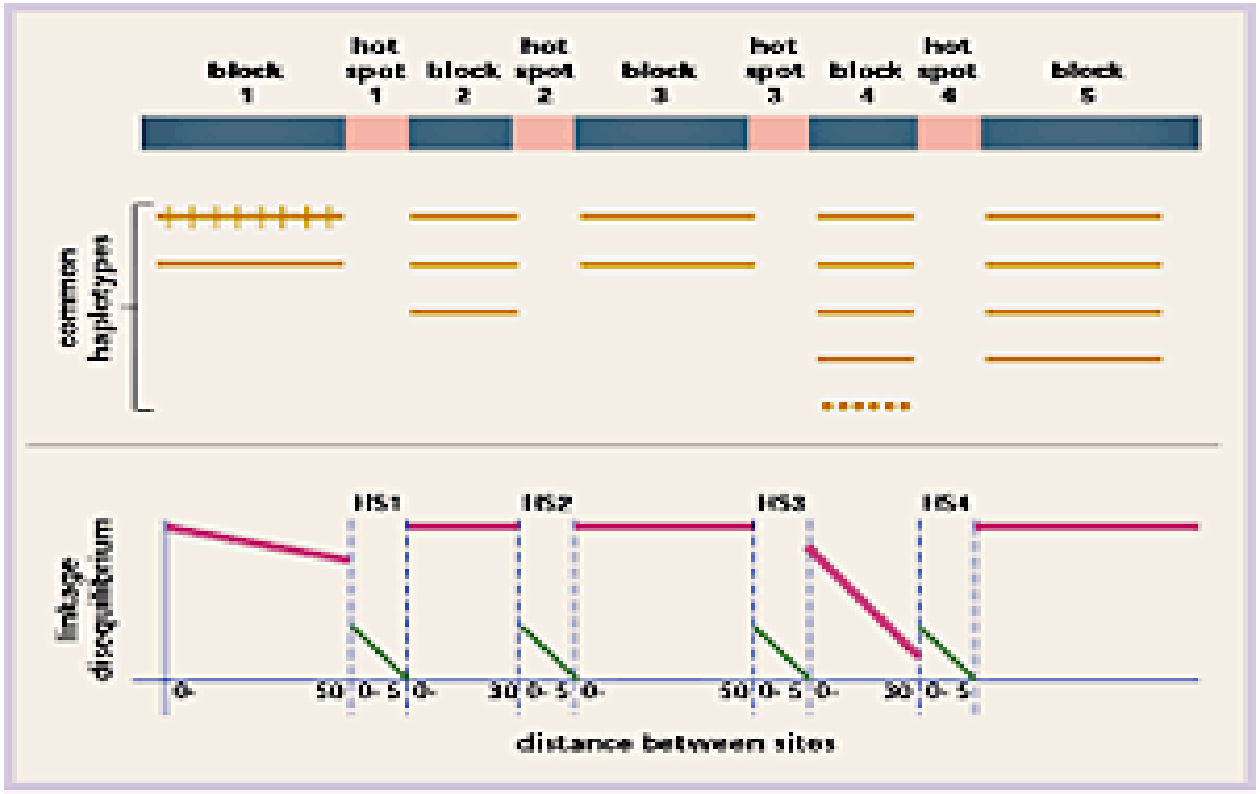




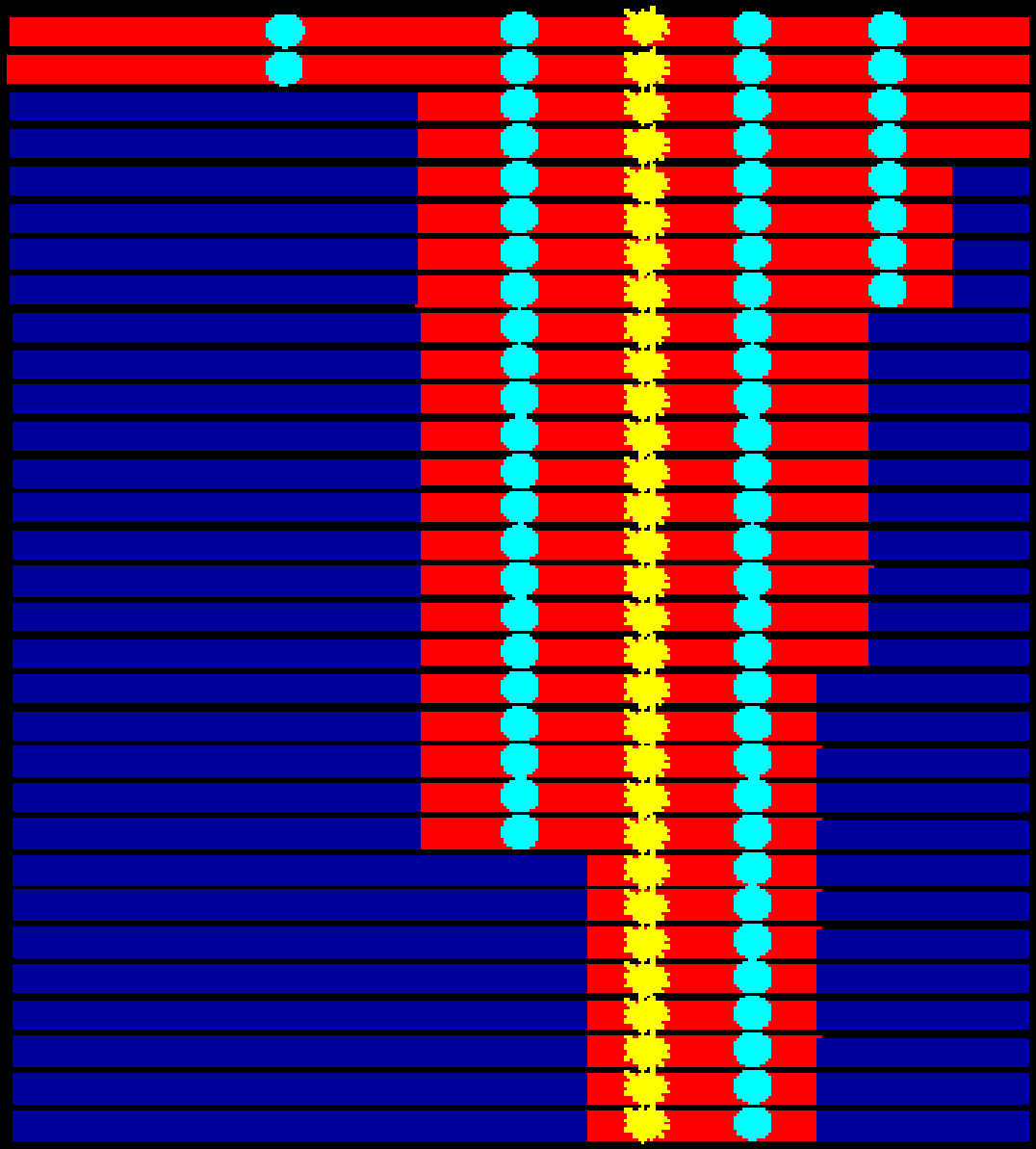
- Ratio form genetic distance to basepairs range from 0.01cM/Mb to 60 cM/Mb

Reich et al. Nature Genetics May 2001

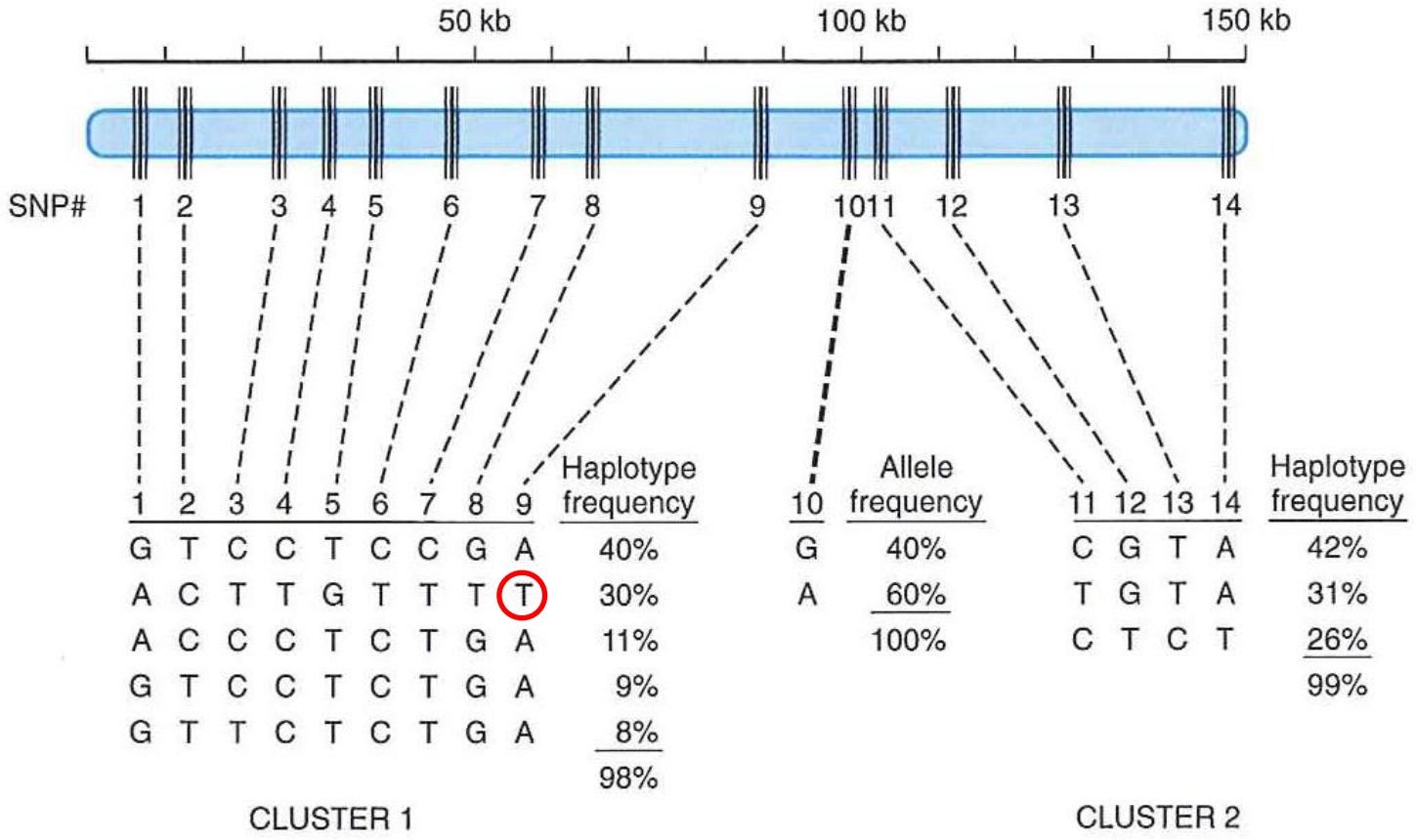
rather large blocks of LD interspersed with recombination hot spots



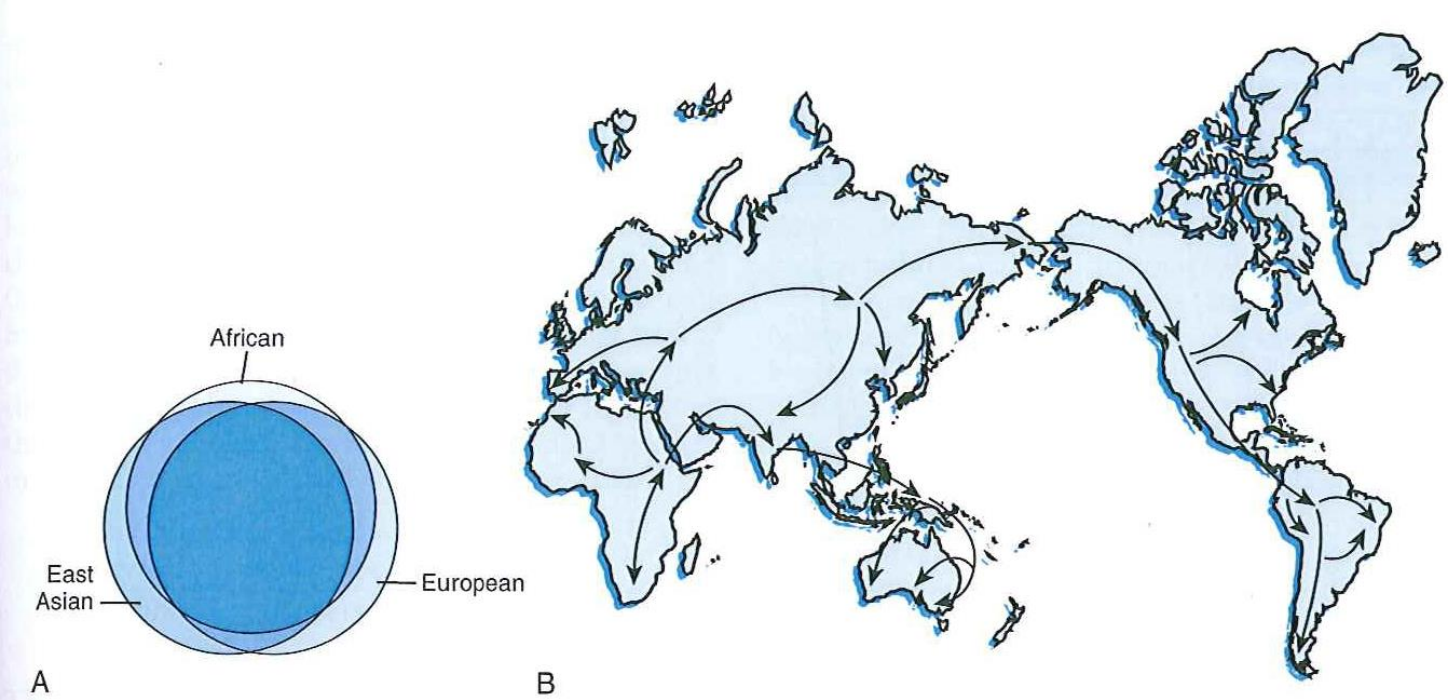
Linkage Disequilibrium



Matthew Stephens, UW

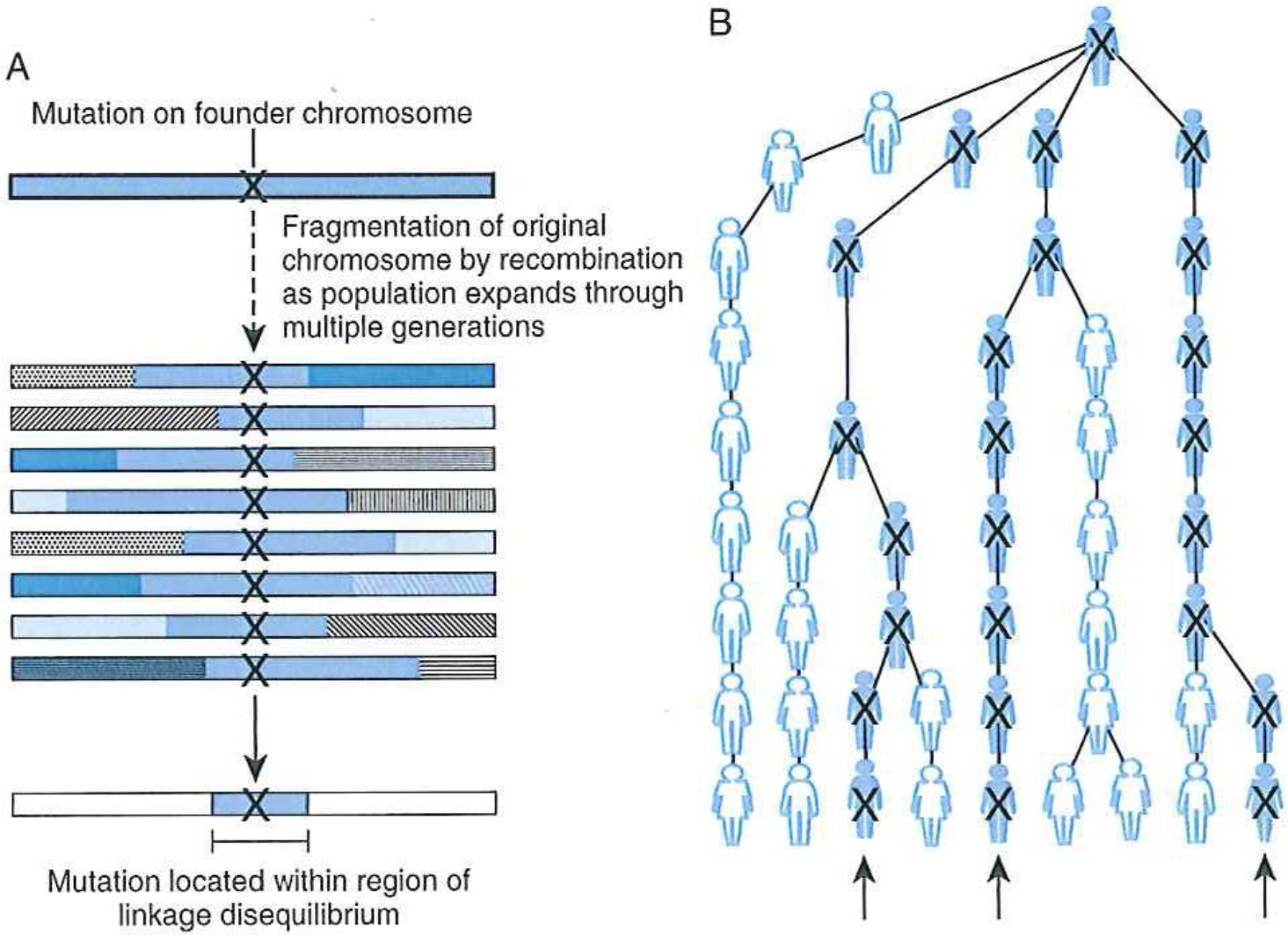


Linkage equilibrium and disequilibrium



- 90% of all SNPs are shared among disparate populations

- African populations have smallers blocks (average 7.3kb) compared with 16.3kb in Europeans whereas the Chinese and Japanese blocks have an average size of 13.2kb.



Mapping human genes by

linkage analyisis

Linkage analysis is a method that is used to decide if two loci or a loci and a disease gene are linked :

1. Ascertain whether the recombination fraction theta between two loci deviates significantly from 0.5.

2. If theta is different from 0.5, we need to make the best estimate of theta, since this parameter tells us how close the linked loci are.

Linkage is expressed as a LOD score (Z); a *" logarithm of odds"*

$$\text{LOD score } (\Theta) = \log_{10} \frac{\text{Likelihood of linkage}}{\text{Likelihood that loci are unlinked (theta = 0.5)}}$$

Positive values of Z at a given Θ suggest that two loci are linked.

Negative values of Z at a given Θ suggest that two loci are not linked.

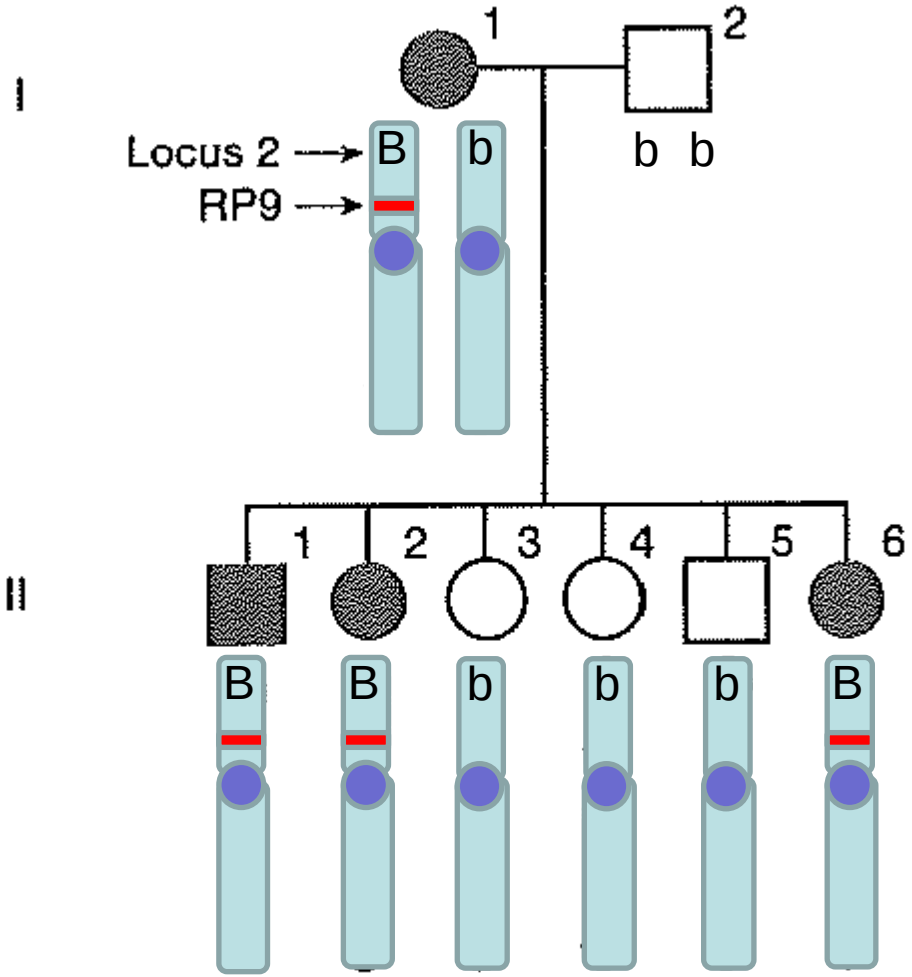
By convention, a LOD score of +3 or greater is considered definitive evidence that two loci are linked. A LOD score below -2 excludes linkage.

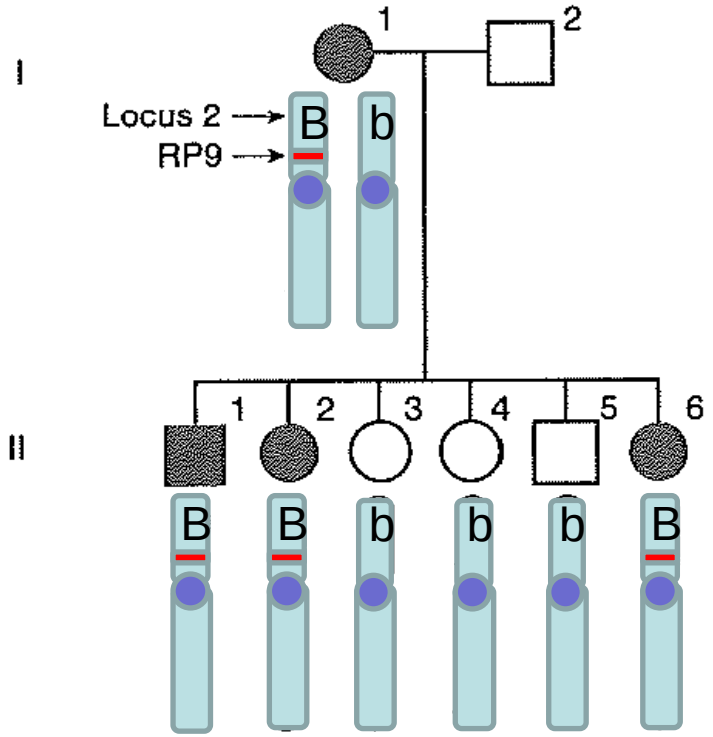
Log $\frac{1000}{1}$

Probability of a recombination is Θ
N is amount of **recombinants** in pedigree

Probability that no recombination will occur is (1-Θ)
M is amount of **non-recombinants** in pedigree

$$Z(\Theta) = \log_{10} \frac{\Theta^N}{(0.5)^N} + \log_{10} \frac{(1-\Theta)^M}{(0.5)^M} = \log_{10} \frac{\Theta^N (1-\Theta)^M}{(0.5)^N (0.5)^M}$$



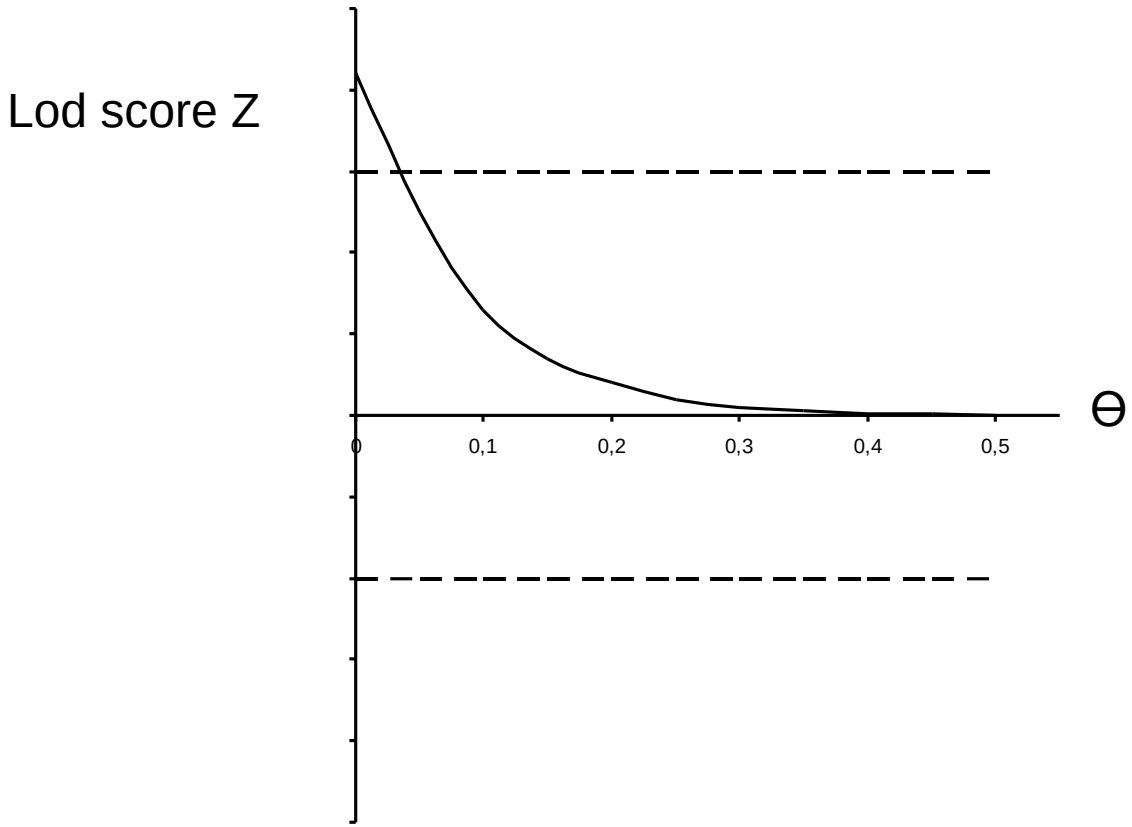


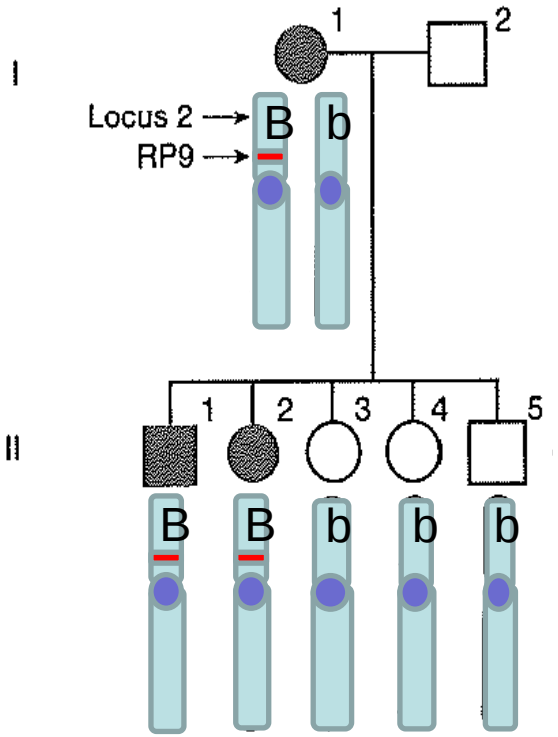
$$Z = \log_{10} \frac{\Theta^0 (1-\Theta)^6}{(0.5)^0 \ (0.5)^6} = 1{,}81 \ (\Theta=0)$$

θ =	0.00	0.01	0.05	0.10	0.20	0.30	0.40
Family 1	1.8	1.78	1.67	1.53	1.22	0.88	0.48

Z max = 1.8 at Θ max=0

Interpreting LOD plots





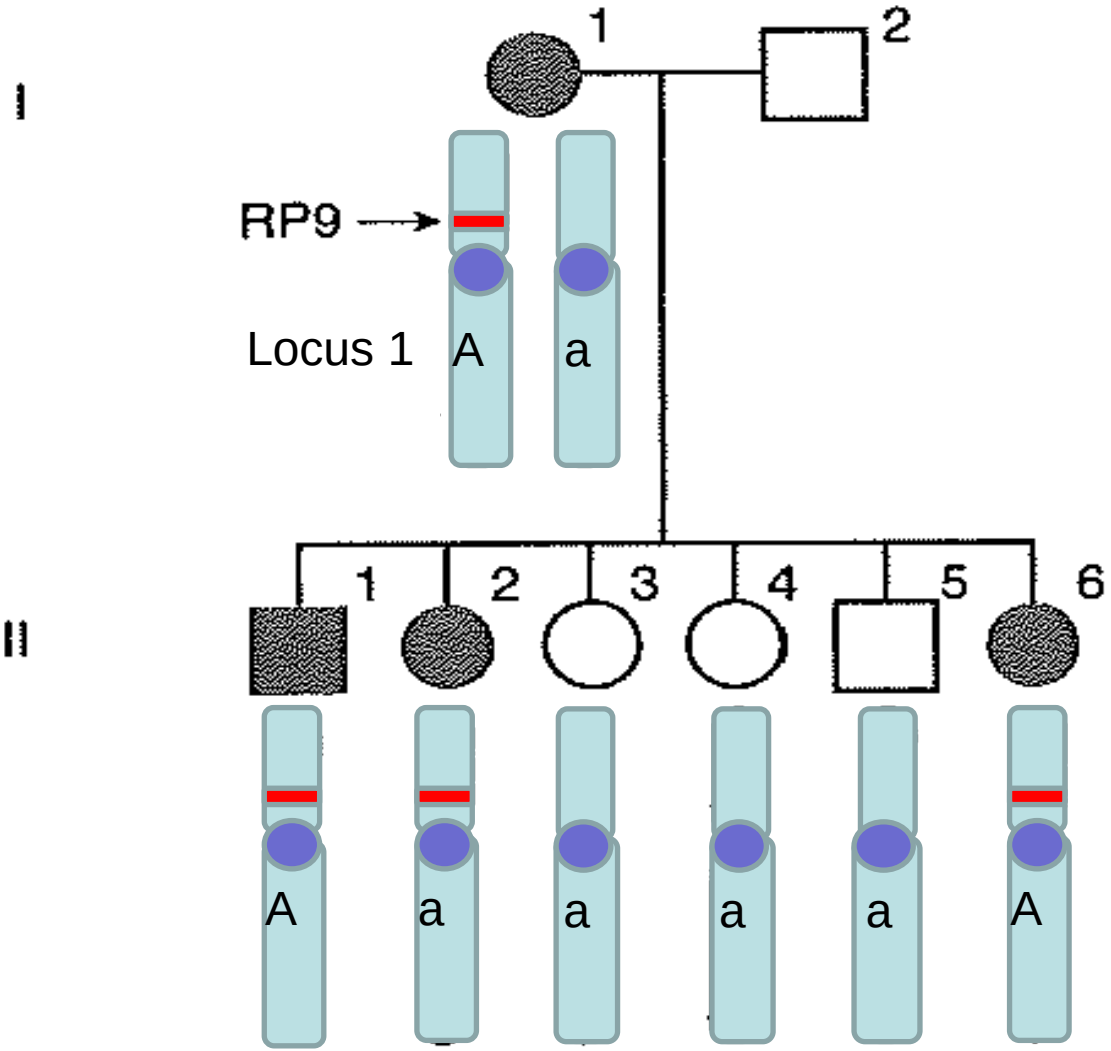
$$Z = \log_{10} \frac{\Theta^0 (1-\Theta)^5}{(0.5)^0 \ (0.5)^5} = 1{,}51 \ (\Theta=0)$$

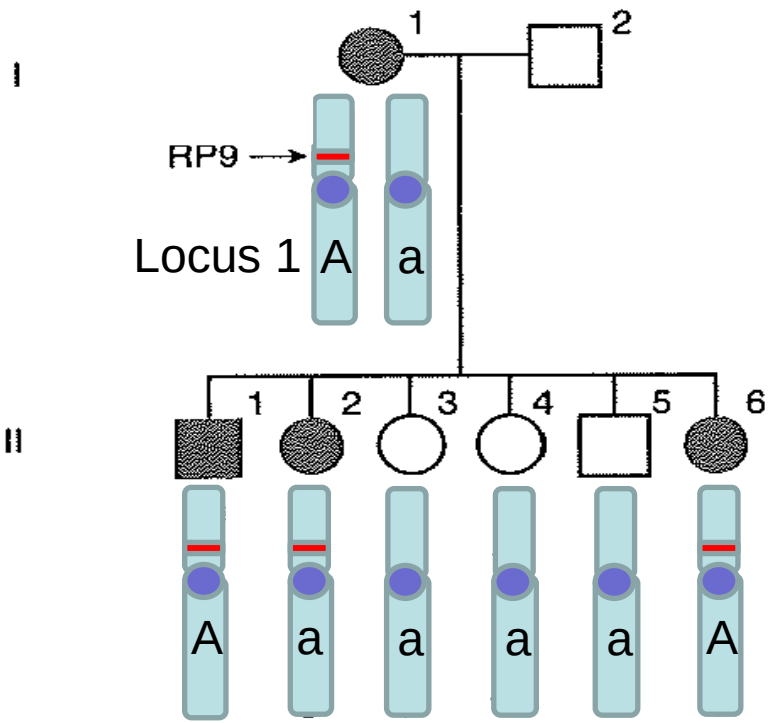
Z max = 1.5 at Θ max=0

- Ommiting one non-rec. individual lowers the LOD score with 0.3
- It does not matter if the individual is affected or not affected

Exercise : calculate LOD score at $\Theta=0$ for a similar family with 10 children without any recombinant between the disease locus and the marker.

$$Z = \log_{10} \frac{\Theta^0 (1-\Theta)^{10}}{(0.5)^0 \ (0.5)^{10}} = 3.01 \ (\Theta=0)$$

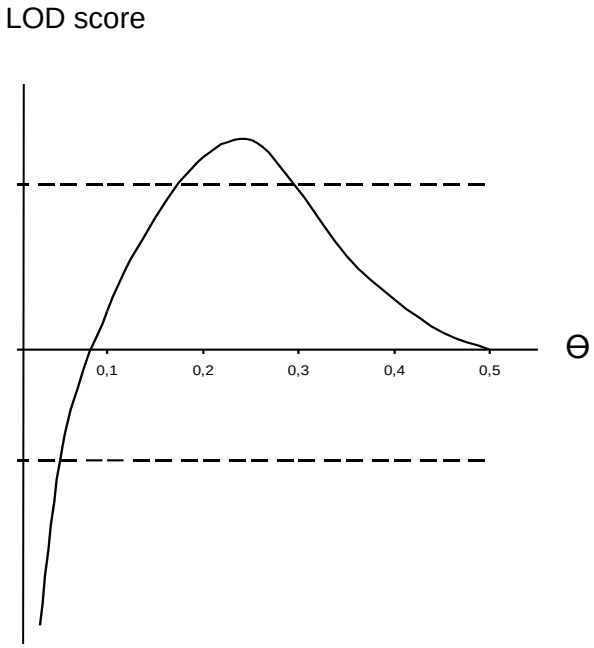


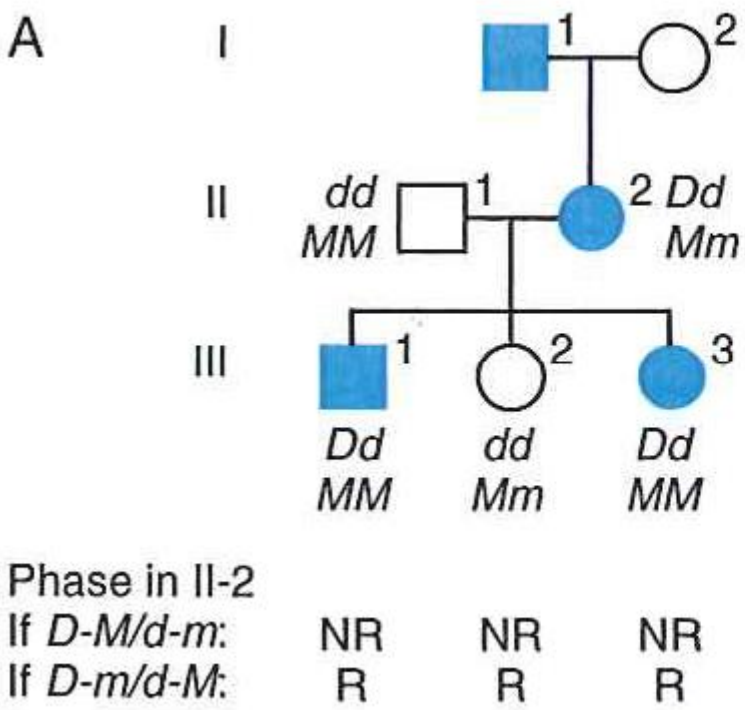


$$Z = \log_{10} \frac{\Theta^1 (1-\Theta)^5}{(0.5)^1 \ (0.5)^5} = \text{- infinity } (\Theta=0)$$

Exercise : calculate LOD scores for Θ = 0.001, 0.01, 0.1, 0.2, 0.3, 0.4 and 0.5

Θ	LOD score
0	-infinity
0.001	-1.19
0.01	-0.21
0.1	0.57
0.2	0.62
0.3	0.51
0.4	0.29
0.5	0

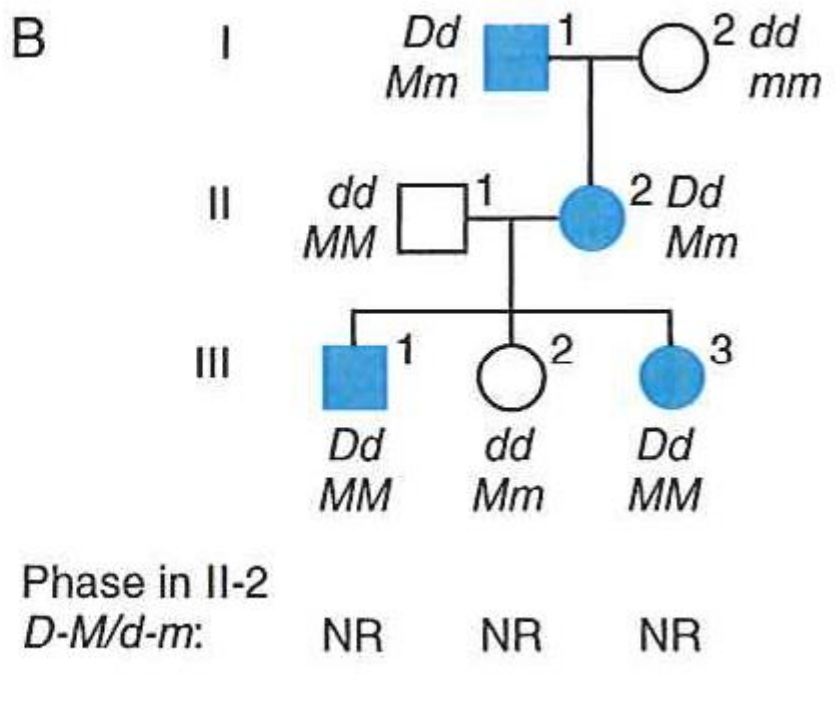




Phase in II-2

If *D-M/d-m*:	NR	NR	NR
If *D-m/d-M*:	R	R	R

$$Z = \log_{10} \frac{1/2\Theta^3 + 1/2(1-\Theta)^3}{(0.5)^3} = 0.602 \ (\Theta=0)$$



Phase in II-2

D-M/d-m:	NR	NR	NR

$$Z = \log_{10} \frac{\Theta^0 (1-\Theta)^3}{(0.5)^0 \ (0.5)^3} = 0.903 \ (\Theta=0)$$

Strength of evidence for linkage (8 to 1) is twice as great in the phase-known situation compared to the phase-unknown situation.

	LOD SCORES (Z) AT VARIOUS VALUES OF θ						
Type of Pedigree	0.00	0.01	0.05	0.10	0.20	0.30	0.40
Phase unknown $Z_{max} = 0.602$ at $\theta_{max} = 0$	.602	.589	.533	.465	.318	.170	.049
Phase known $Z_{max} = 0.903$ at $\theta_{max} = 0$	.903	.890	.837	.765	.612	.438	.237

X-linked disease