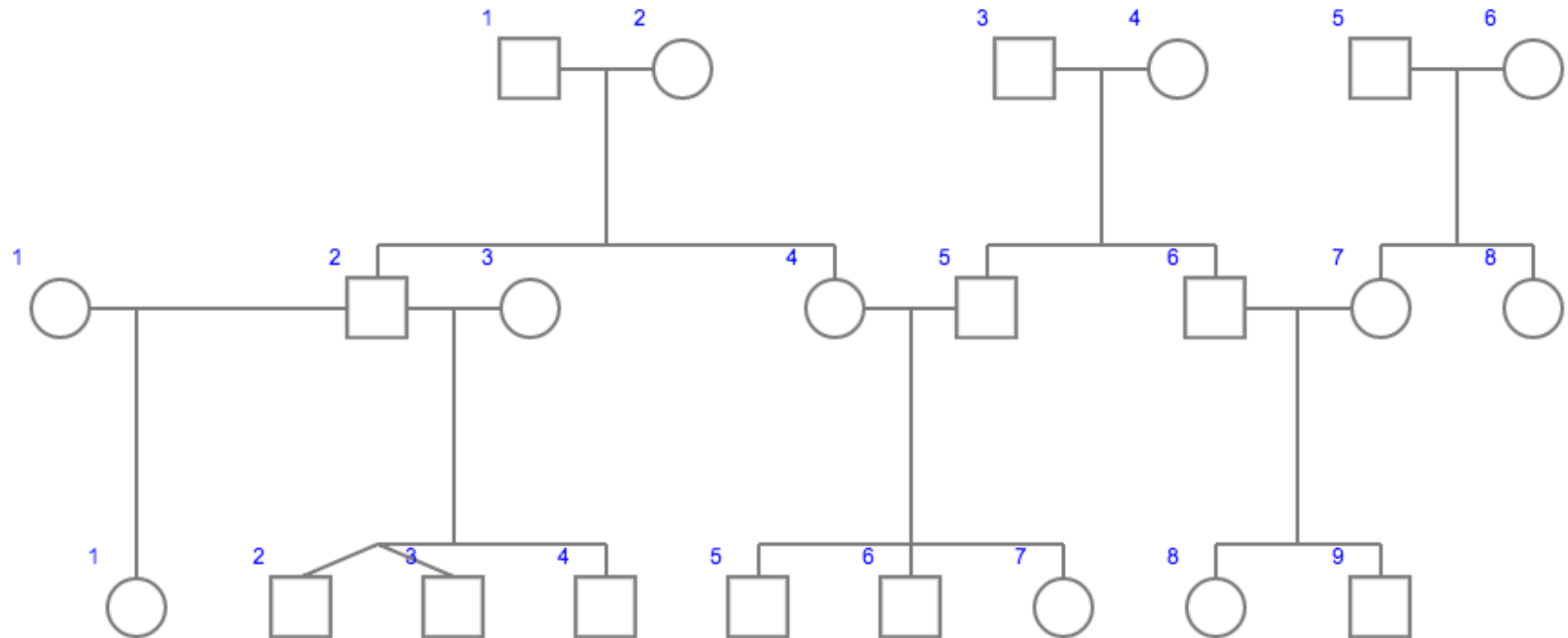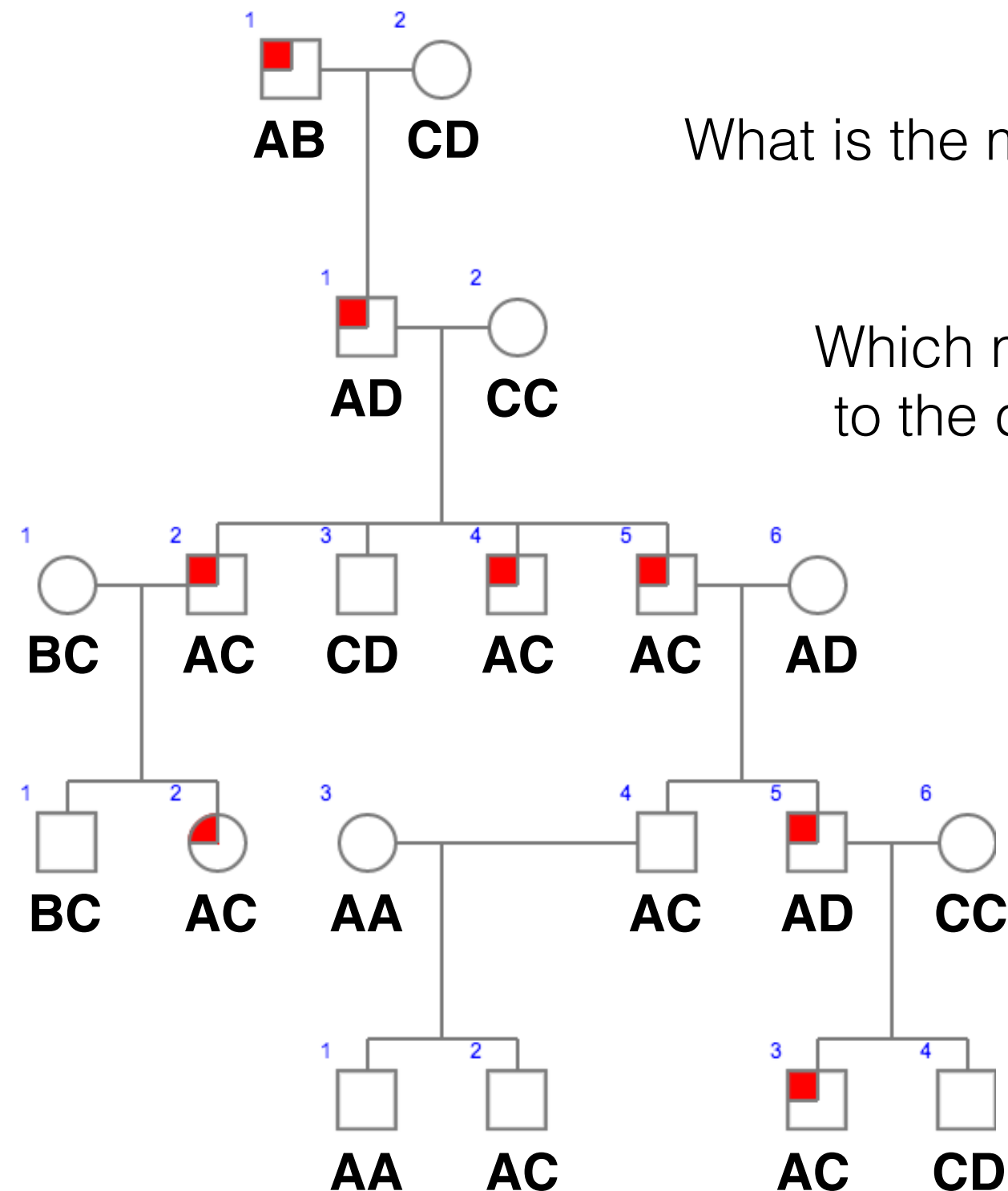# Bio393: Genetic Analysis

Linkage mapping in families

The goal of linkage analysis is to identify a marker nearby the disease-causing allele

What is the mode of inheritance?

Which marker is linked to the disease allele?

Linkage is reported as a log of the odds ratio or LOD

The odds ratio is a statistical method to measure association of some attribute with the presence or absence of another attribute.

Probability of pedigree under linkage versus no linkage

$$LOD = \log_{10} \frac{P(\text{pedigree with linkage})}{P(\text{pedigree with no linkage})}$$

LOD > 3 is good evidence of linkage
1 in 1,000 chance of data by chance

LOD < 0 means data are more likely by chance

How linked is our marker to the disease-causing allele?
Recombination frequency

Recombination frequency is written as θ

Percentage of recombinant gametes passed down

$\theta = 0$ is perfect linkage

$\theta = 0.5$ is no linkage

Linkage is reported as a log of the odds ratio or LOD

The odds ratio is a statistical method to measure association of some attribute with the presence or absence of another attribute.

Probability of pedigree under linkage versus no linkage

$$LOD = \log_{10} \frac{P(\text{pedigree with linkage})}{P(\text{pedigree with no linkage})}$$

$$LOD = \log_{10} \frac{P(\text{data} \mid \theta)}{P(\text{data} \mid \theta = 0.5)}$$

P(data) = probability that a particular gamete was inherited

Linkage is reported as a log of the odds ratio or LOD

$$LOD = \log_{10} \frac{P(data \mid \theta)}{P(data \mid \theta = 0.5)}$$

P(data) = probability that a particular gamete was inherited

$$P(data \mid \theta) =$$

Probability that recombination did not occur for parental $1 - \theta$

Probability that recombination did occur for recombinant θ

Probability of phase of parent

Each individual is independent.
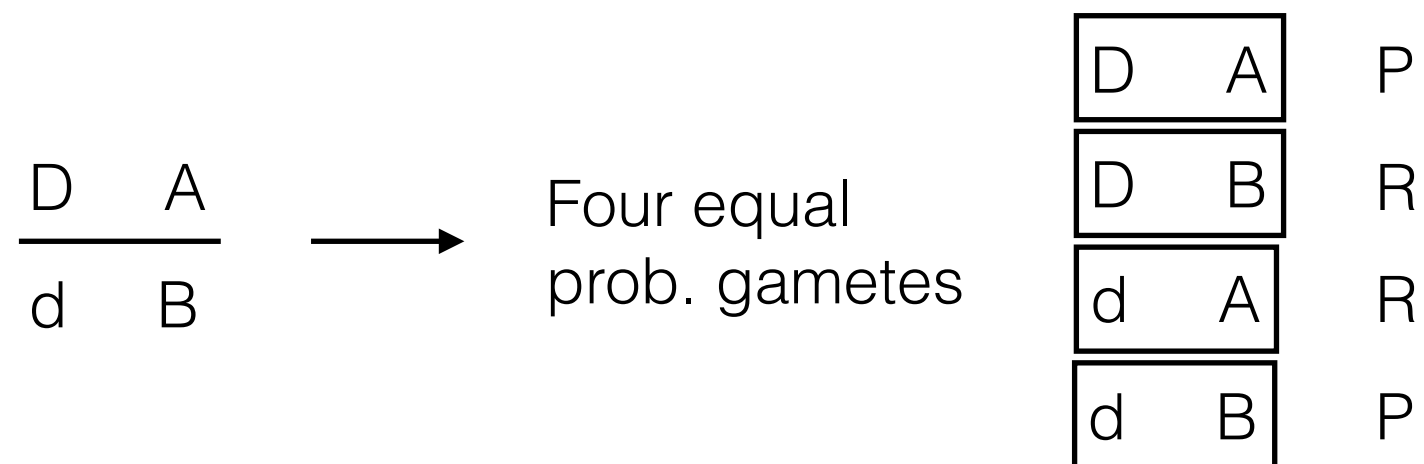Use product rule to multiply probabilities.

Linkage is reported as a log of the odds ratio or LOD

$$LOD = \log_{10} \frac{P(data \mid \theta)}{P(data \mid \theta = 0.5)}$$

P(data) = probability that a particular gamete was inherited

$$P(data \mid \theta = 0.5)$$

Equal probability of two loci independently assorting and one gamete being passed down

$$\frac{D \quad A}{d \quad B} \longrightarrow$$ Four equal prob. gametes

D	A	P
D	B	R
d	A	R
d	B	P

Each individual is independent.
Use product rule to multiply probabilities.

Linkage is reported as a log of the odds ratio or LOD

Probability of pedigree under linkage versus no linkage

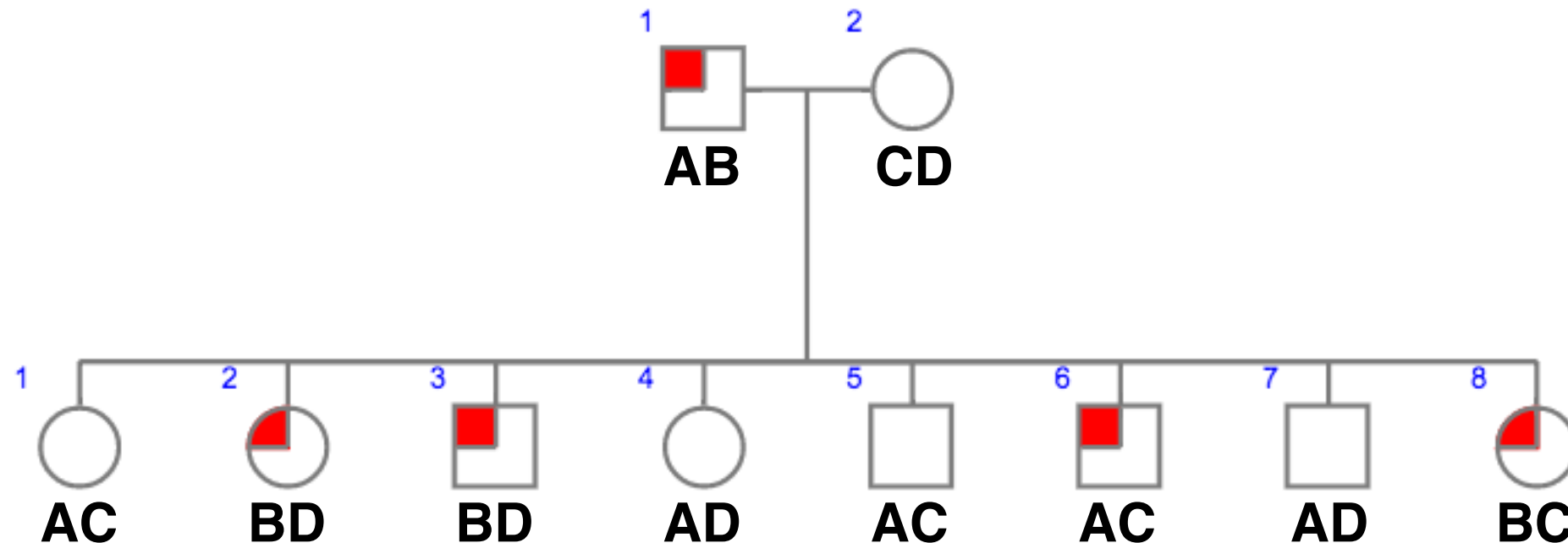$$LOD = \log_{10} \frac{P(\text{pedigree with linkage})}{P(\text{pedigree with no linkage})}$$

$$LOD = \log_{10} \frac{P(\text{data} \mid \theta)}{P(\text{data} \mid \theta = 0.5)}$$

P(data) = probability that a particular gamete was inherited

$$LOD = \log_{10} \frac{(1 - \theta)^{NR} \times \theta^{R}}{0.5^{(NR + R)}}$$

Use this equation when we know phase.

Linkage is reported as a log of the odds ratio or LOD
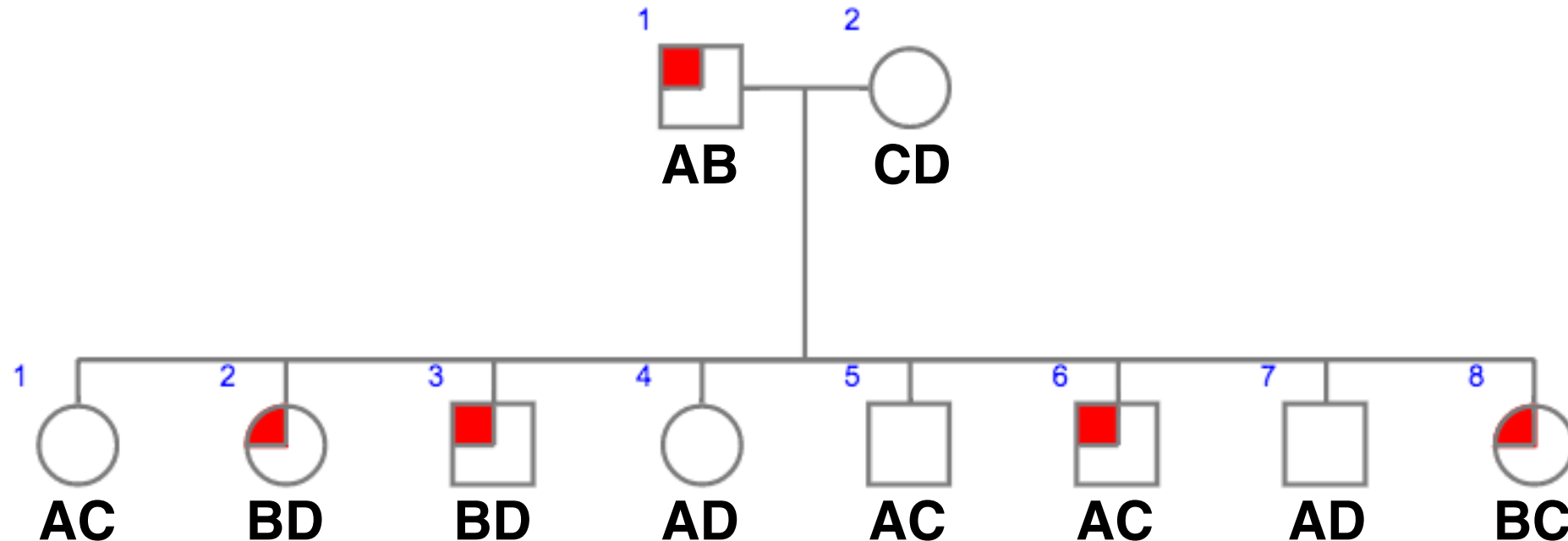

Do we know the phase of I-1?

NO

Every child has an equal chance of being a recombinant or parental

Difference between being informative and knowing phase
Informative = Parent heterozygous at each of two loci

Linkage is reported as a log of the odds ratio or LOD

With unknown phase of parent, the LOD equation is…

$$LOD = \log_{10} \frac{\frac{1}{2}((1 - \theta)^{NR} \times \theta^{R}) + \frac{1}{2}((1 - \theta)^{NR} \times \theta^{R})}{(0.5^{(NR + R)})}$$
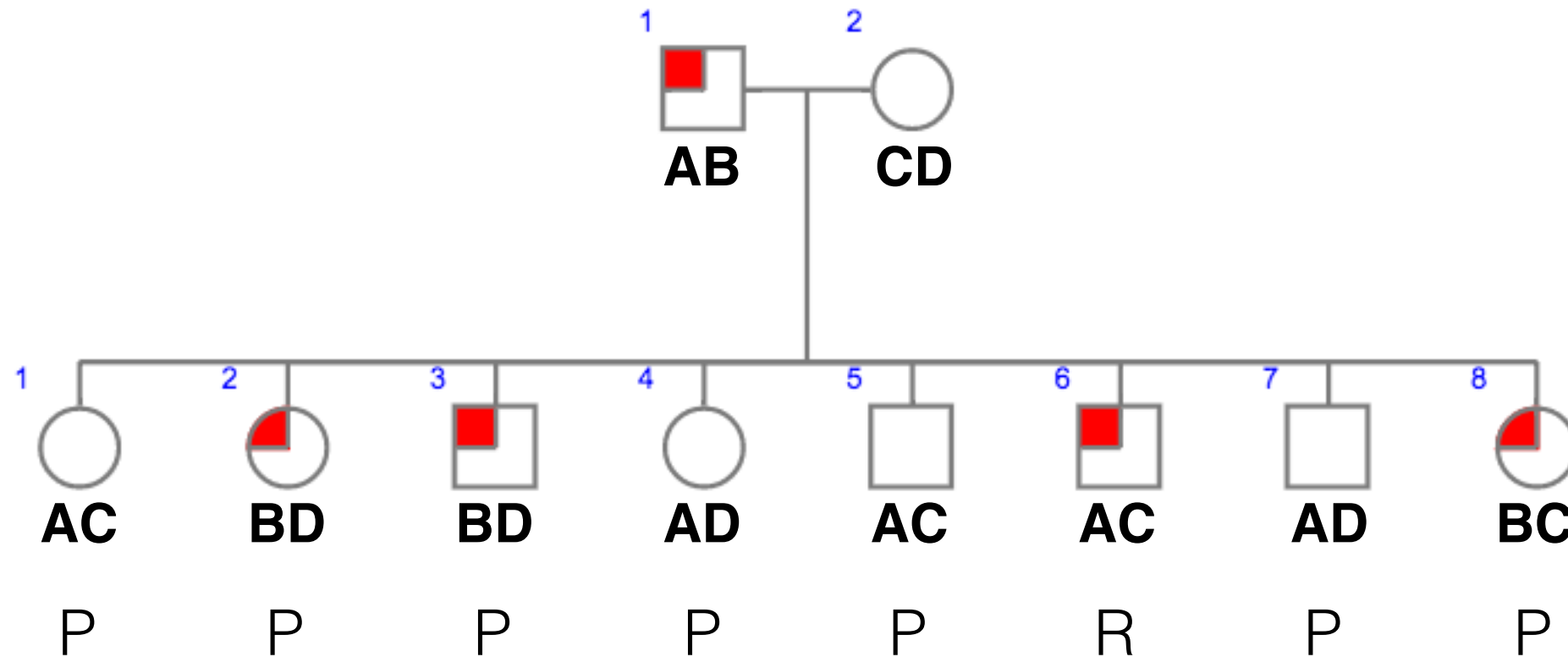
½ chance for each phase   Phase 1   Phase 2

$$\frac{D \quad B}{d \quad A}$$ $$\frac{D \quad A}{d \quad B}$$

Linkage is reported as a log of the odds ratio or LOD

Phase 1   P   P   P   P   P   R   P   P

$$\text{LOD} = \log_{10} \frac{\frac{1}{2}((1 - \theta)^{NR} \times \theta^{R}) + \frac{1}{2}((1 - \theta)^{NR} \times \theta^{R})}{(0.5^{(NR + R)})}$$
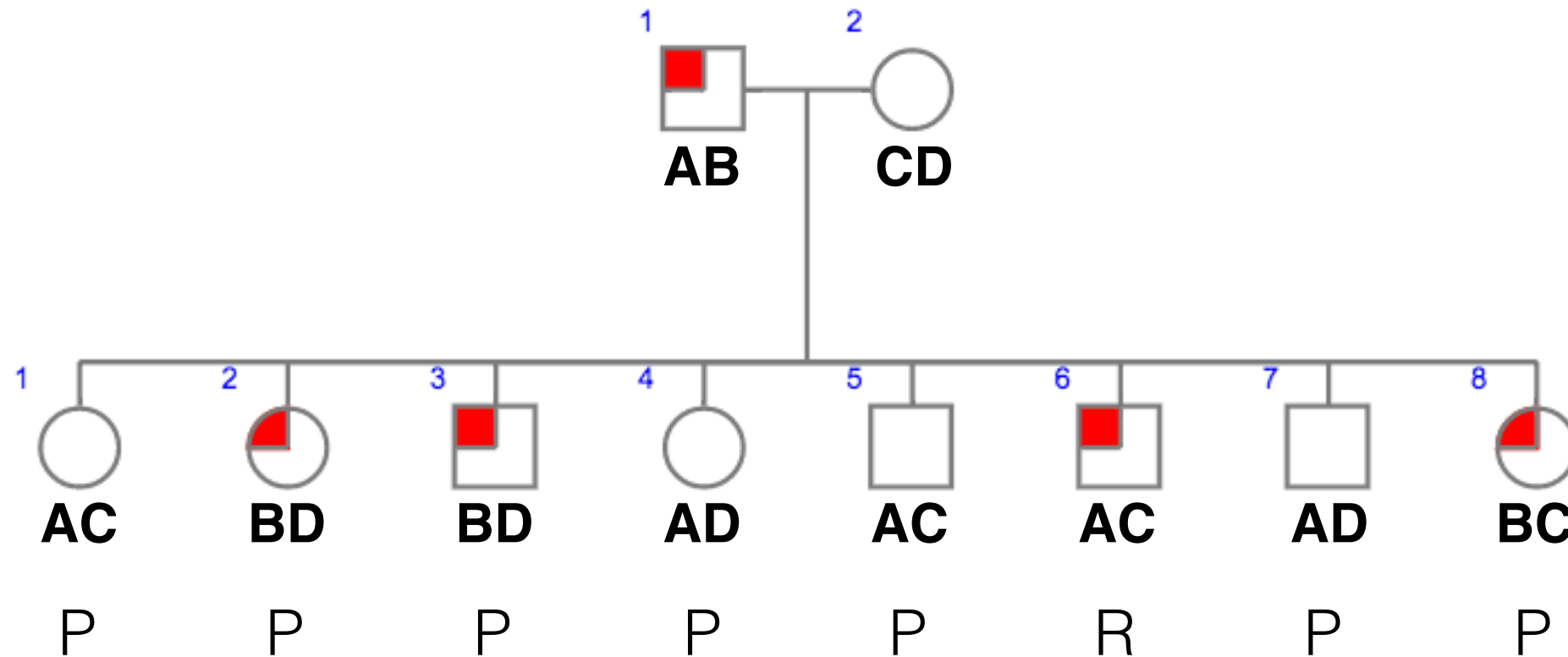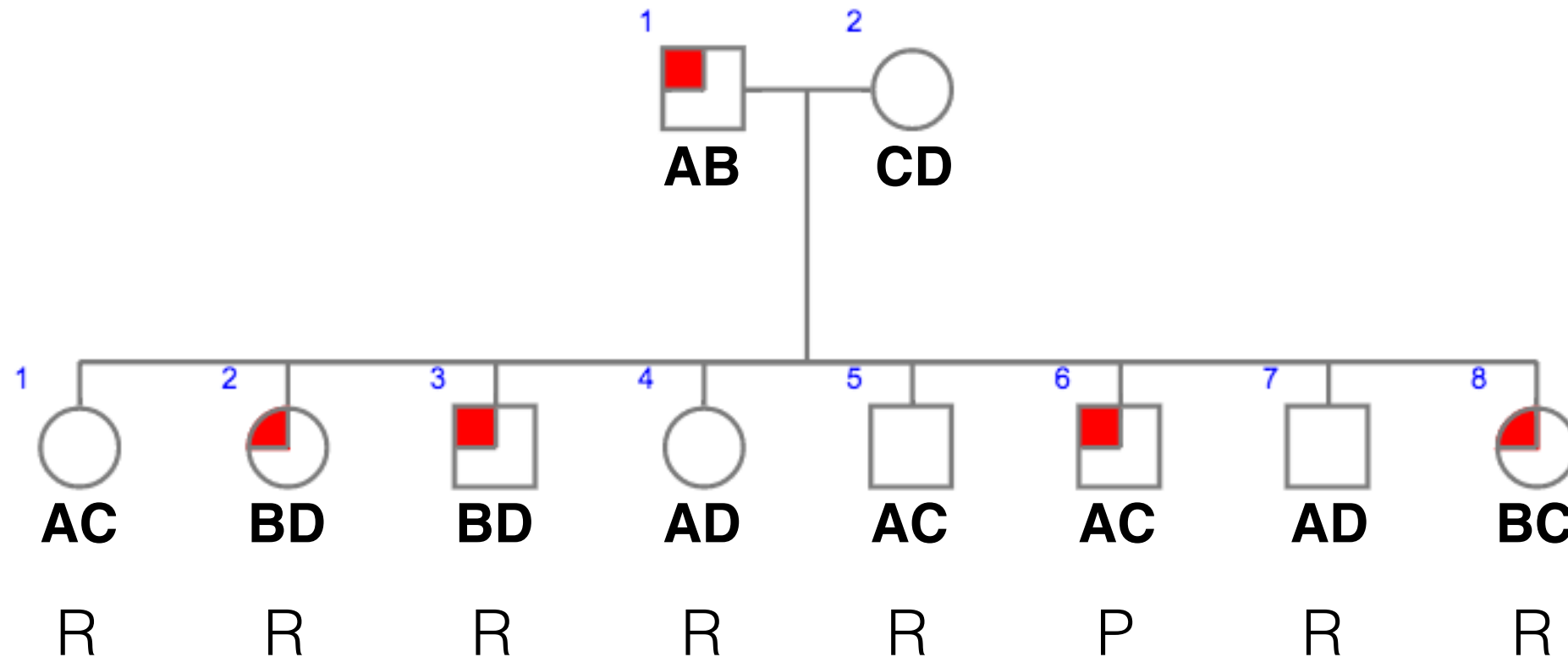
½ chance for each phase   Phase 1   Phase 2

$$\frac{D \quad B}{d \quad A} \qquad\qquad \frac{D \quad A}{d \quad B}$$

Linkage is reported as a log of the odds ratio or LOD

$$LOD = \log_{10} \frac{\frac{1}{2}((1 - \theta)^7 \times \theta^1) + \frac{1}{2}((1 - \theta)^{NR} \times \theta^R)}{(0.5^{(8)})}$$

½ chance for each phase

Phase 1

$$\frac{D \quad B}{d \quad A}$$

Phase 2

$$\frac{D \quad A}{d \quad B}$$

Linkage is reported as a log of the odds ratio or LOD

Phase 2   R   R   R   R   R   P   R   R

$$LOD = \log_{10} \frac{\frac{1}{2}((1 - \theta)^7 \times \theta^1) + \frac{1}{2}((1 - \theta)^1 \times \theta^7)}{(0.5^{(8)})}$$
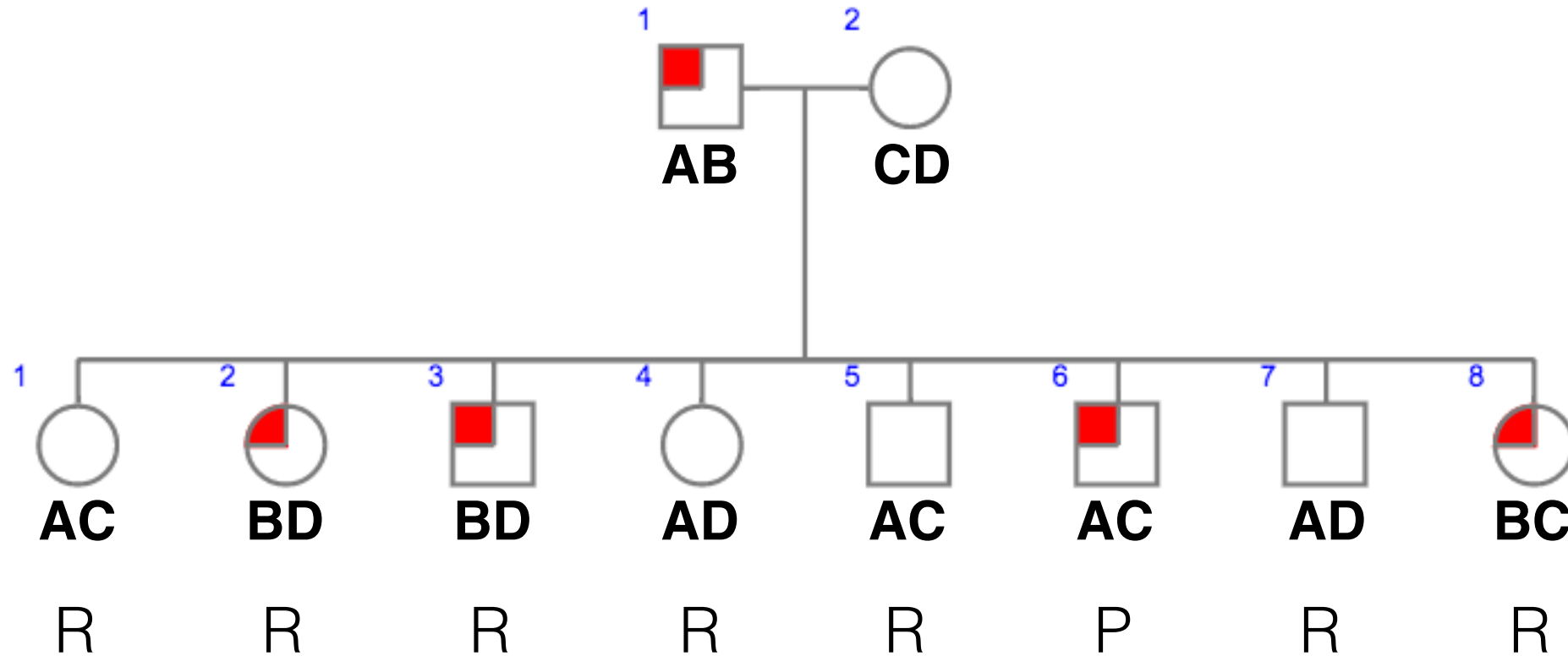
½ chance for each phase   Phase 1   Phase 2

$$\frac{D \quad B}{d \quad A} \qquad\qquad \frac{D \quad A}{d \quad B}$$

Linkage is reported as a log of the odds ratio or LOD

Phase 2 R R R R R P R R

$$LOD = \log_{10} \frac{\frac{1}{2}((1 - 0.125)^7 \times 0.125^1) + \frac{1}{2}((1 - 0.125)^1 \times 0.125^7)}{(0.5^{(8)})}$$

½ chance for each phase

$\theta = 0.125$

Phase 1

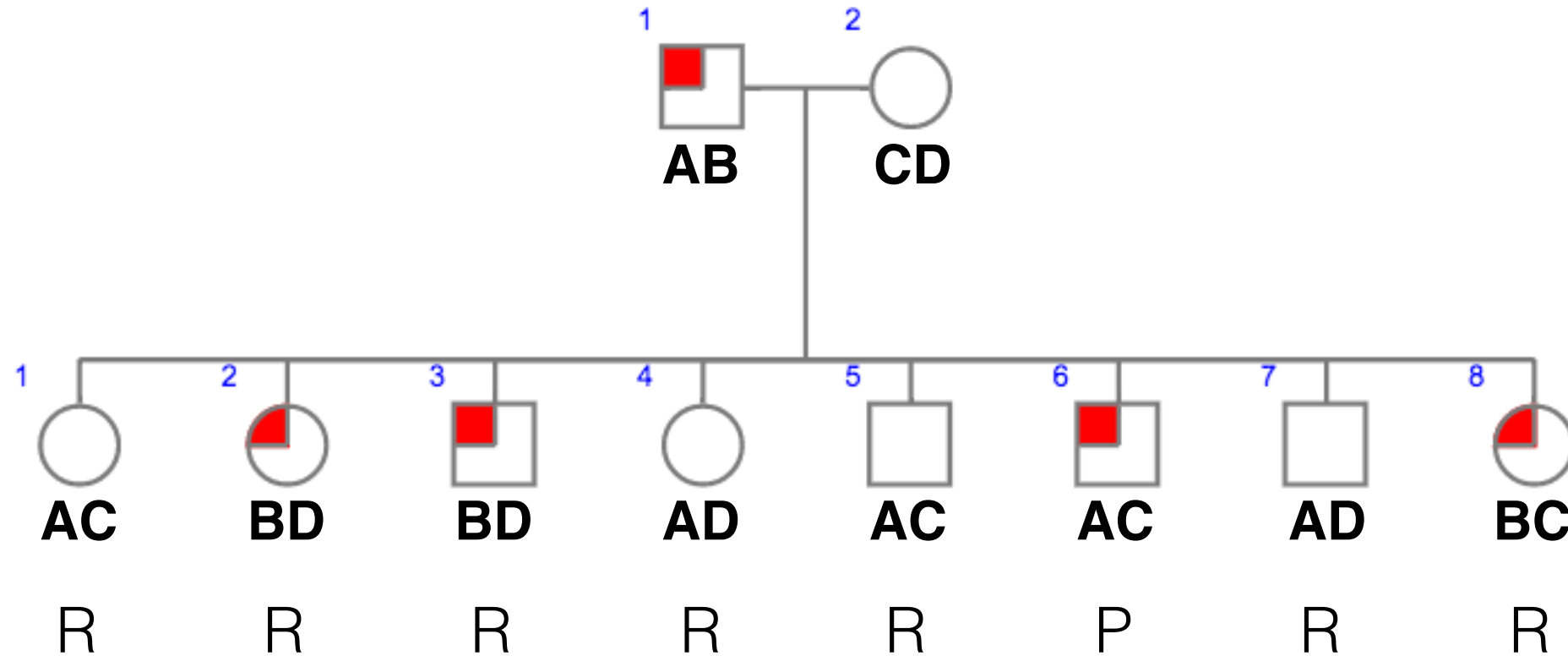$$\frac{D \quad B}{d \quad A}$$

Phase 2

$$\frac{D \quad A}{d \quad B}$$

Linkage is reported as a log of the odds ratio or LOD

Phase 2   R   R   R   R   R   P   R   R
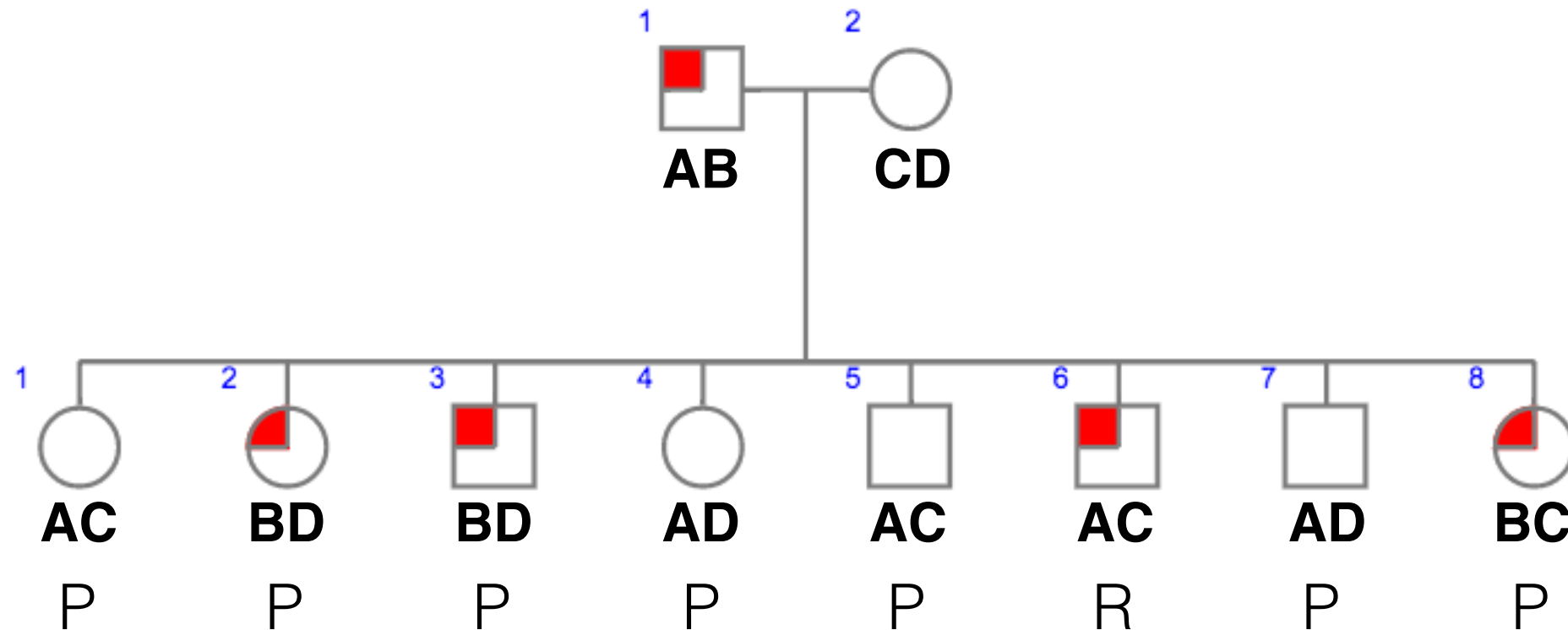
LOD = 0.79 $\theta = 0.125$

Some properties of LOD scores

LOD scores from independent families can be added (product rule with logarithms)

Determining phase increases the LOD score
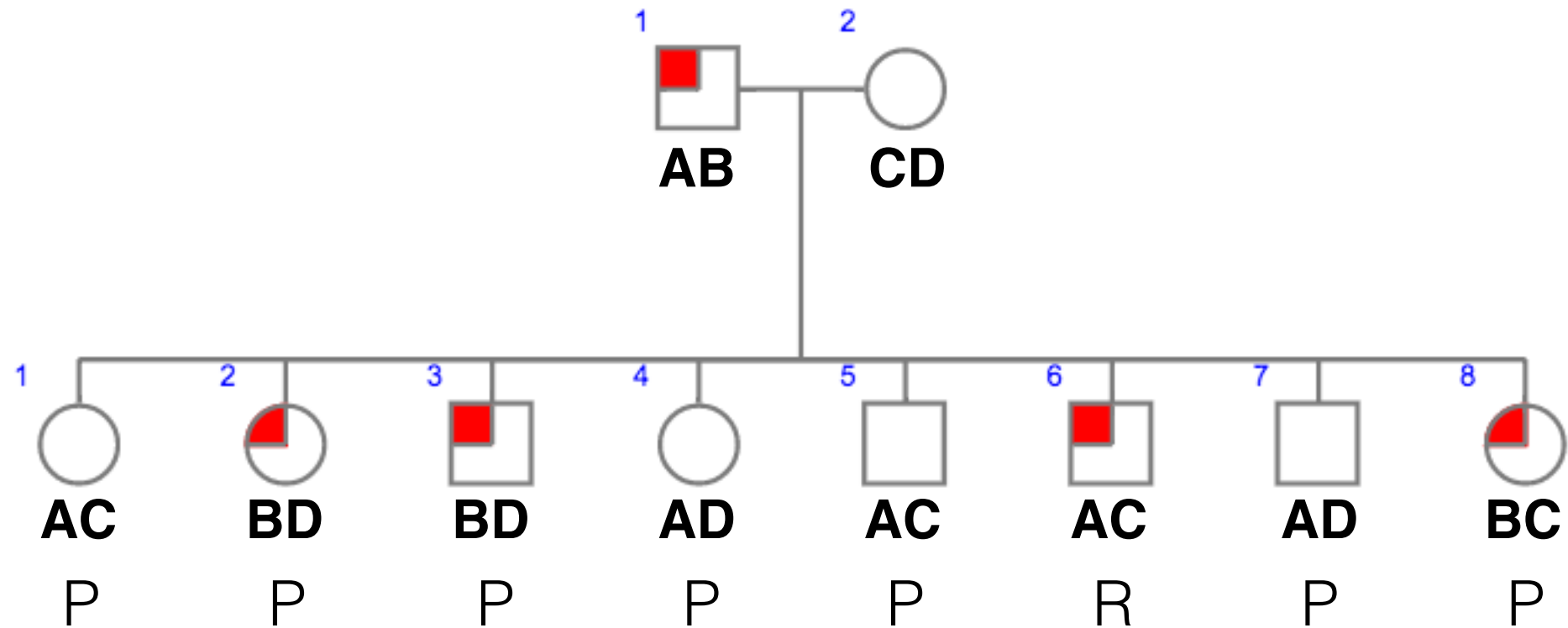
What happens when we know phase?

Phase 1   P   P   P   P   P   R   P   P

The paternal grandmother had the disease and her genotype was BD. Her husband's genotype was AA and he was not affected

$$LOD = \log_{10} \frac{(1 - \theta)^{NR} \times \theta^{R}}{0.5^{(NR + R)}}$$

½ chance for each phase
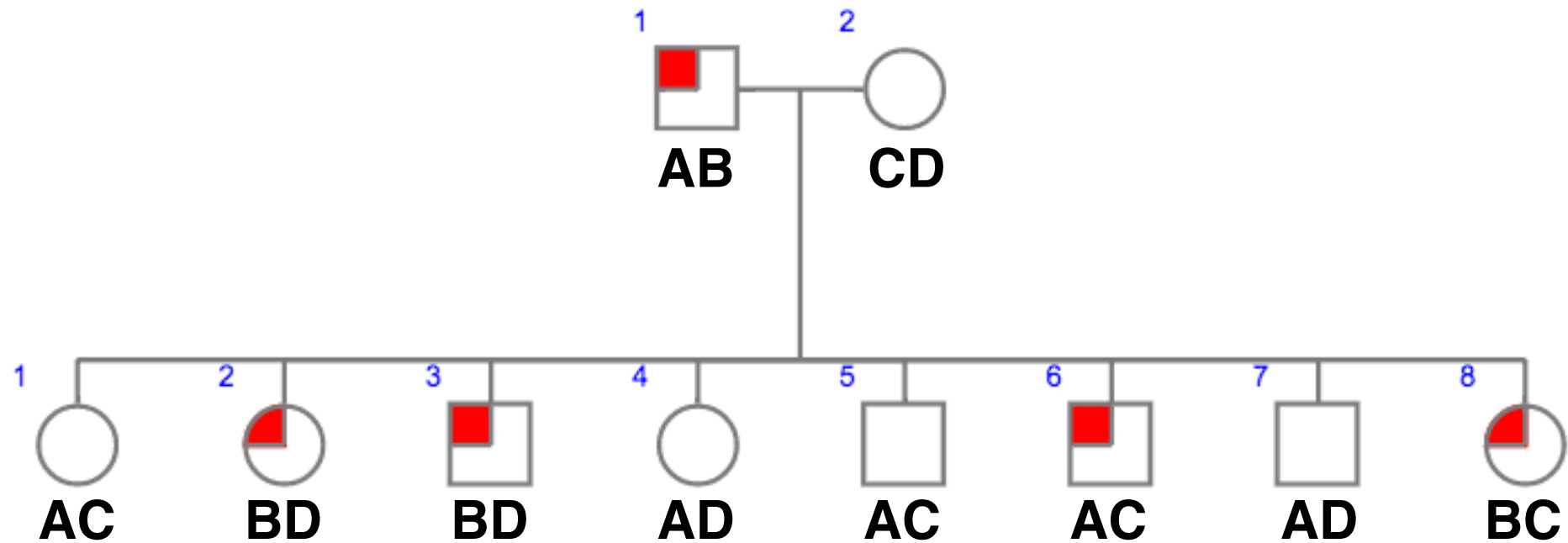and second phase probability
go away!

What happens when we know phase?

Phase 1

$$\text{LOD} = \log_{10} \frac{(1 - 0.125)^7 \times 0.125^{\,1}}{0.5^{(7 + 1)}}$$

½ chance for each phase
and second phase probability
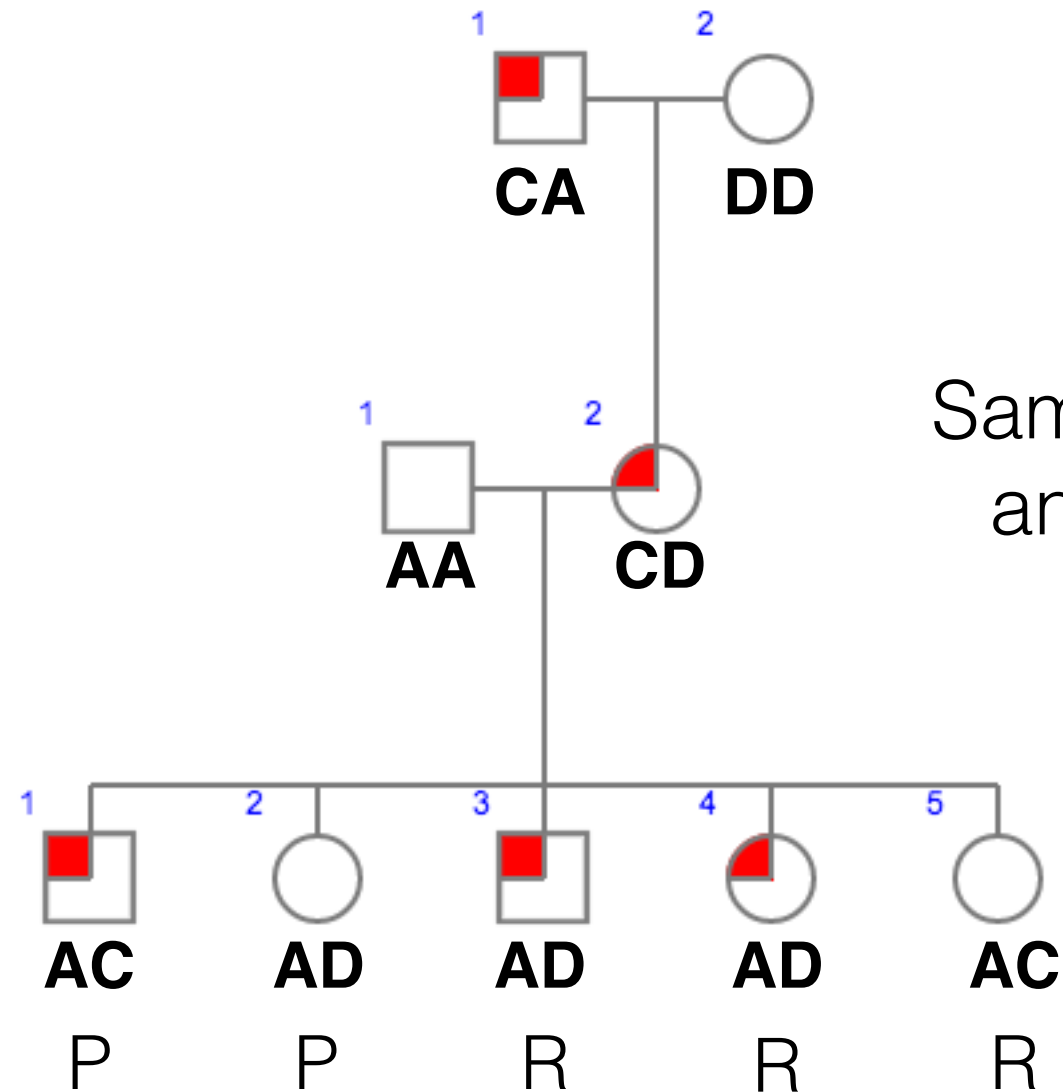go away!

What happens when we know phase?

LOD = 1.1   θ = 0.125

½ chance for each phase
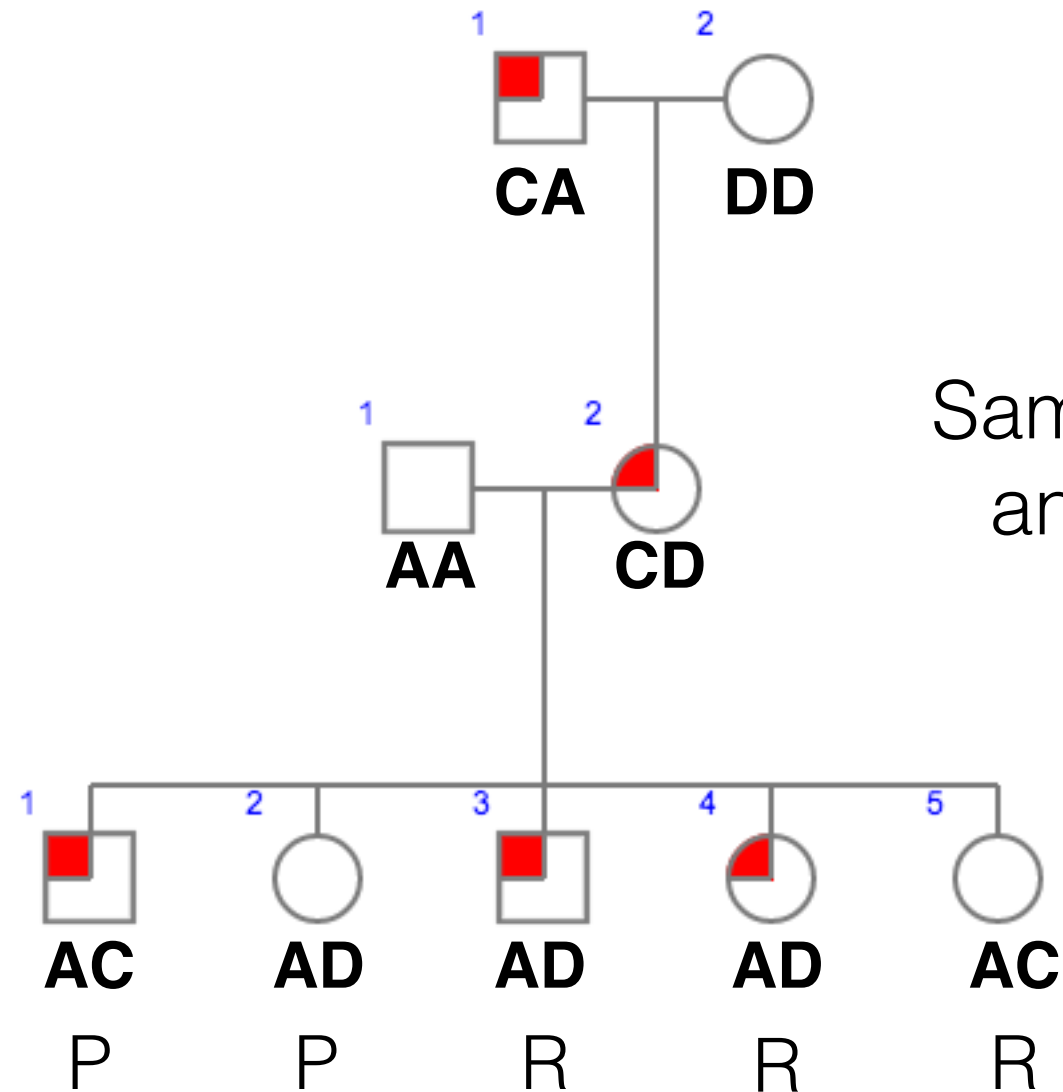and second phase probability
go away!

Let's add another family

Same marker as before and we know phase

$$LOD = \log_{10} \frac{(1 - \theta)^{NR} \times \theta^{R}}{0.5^{(NR + R)}}$$
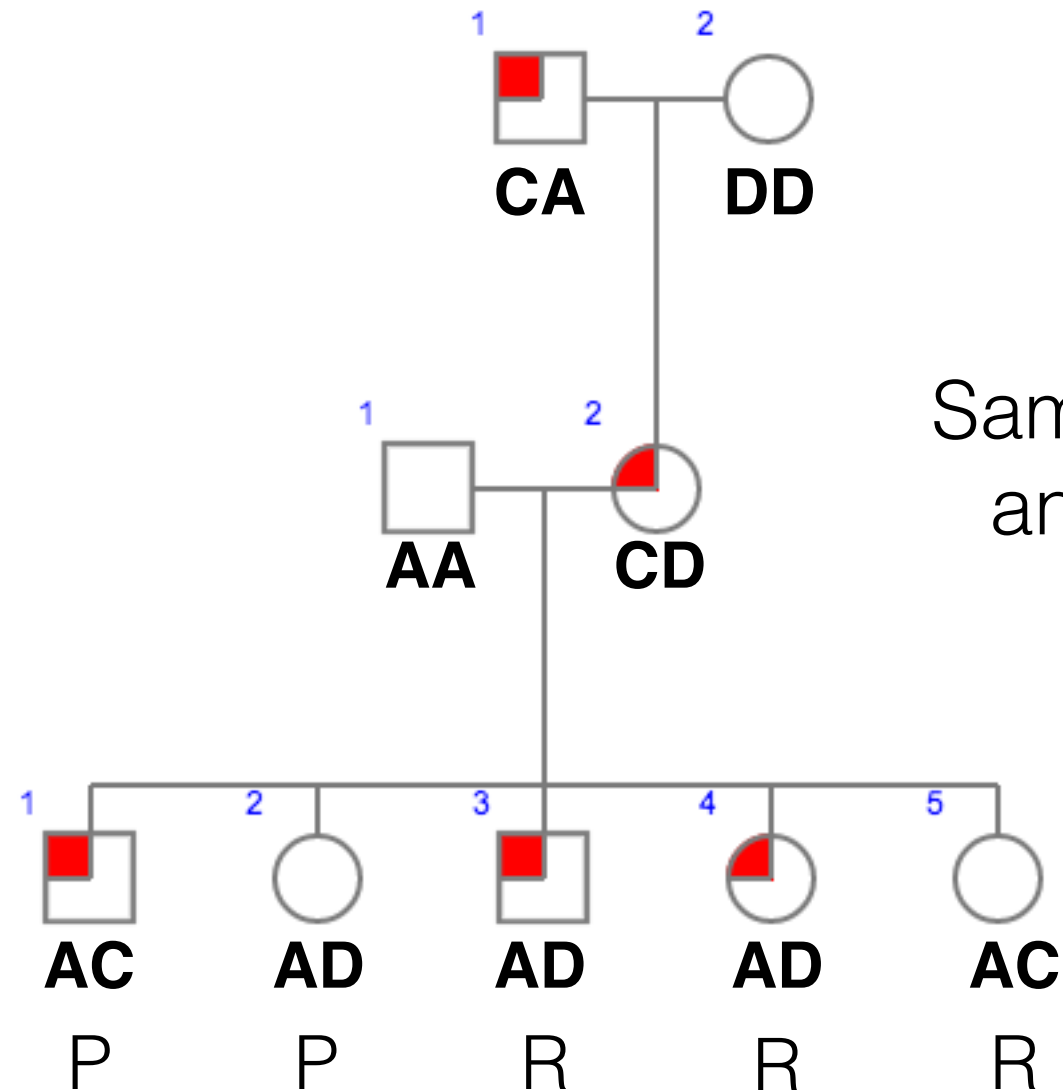
Let's add another family

Same marker as before and we know phase

$$LOD = \log_{10} \left(\frac{(1 - 0.125)^2 \times 0.125^3}{0.5^{(2 + 3)}} \right)$$
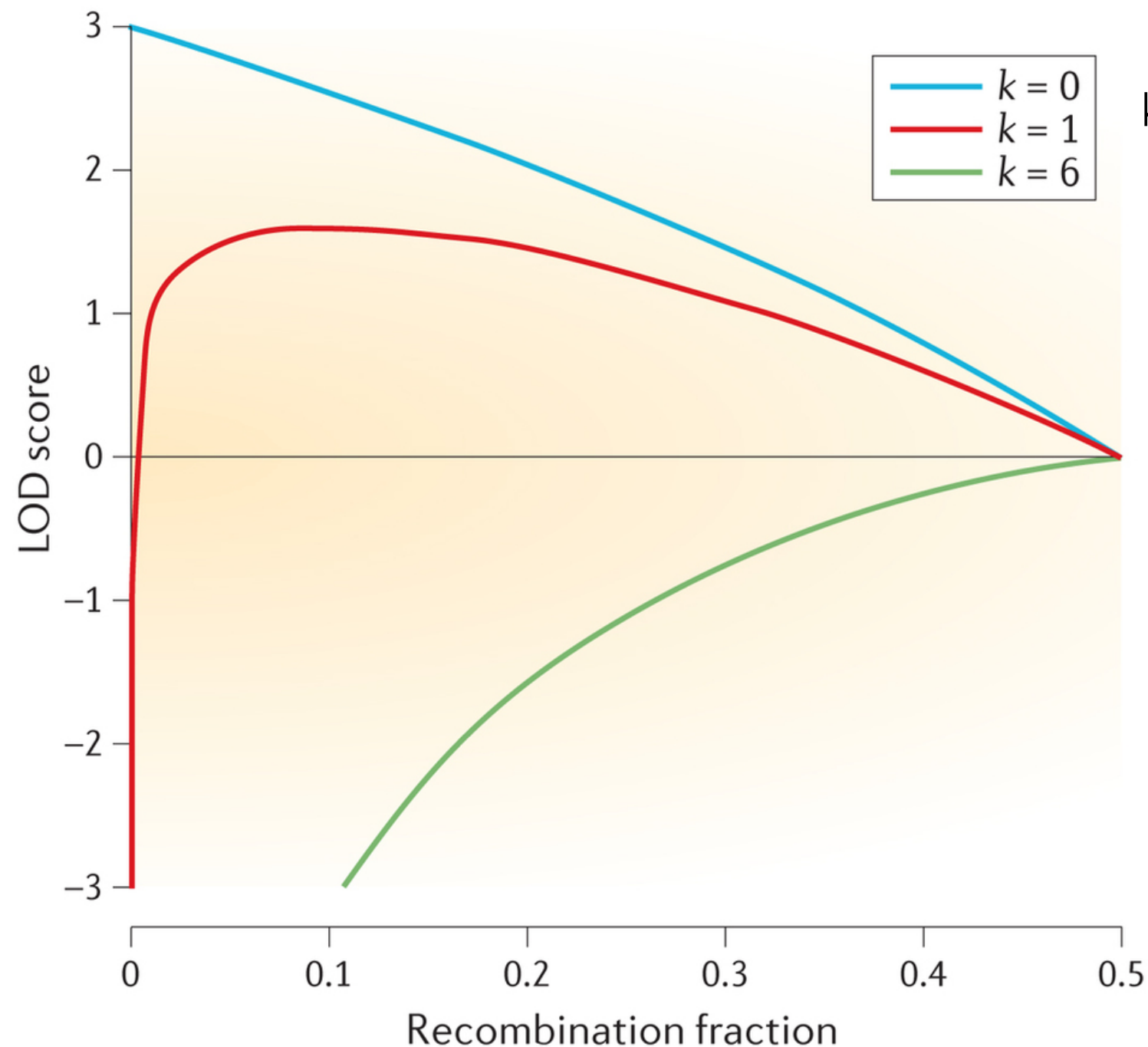
Let's add another family

Same marker as before and we know phase

LOD = -0.1159   θ = 0.125

Change theta to 0.6? No, 0.5 is unlinked

What if we try all possible thetas between 0 and 0.5?

k = # of recombinants
out of 10 individuals

Legend: k = 0, k = 1, k = 6

Nature Reviews | Genetics

# The good and the bad of family-based linkage analysis

Positives:
- Less allelic heterogeneity in families
- Clearly tell recombination events
- Powerful method to find rare variant effects

Negatives:
- Large families are rare
- Different families with the "same" disease could have different genetic causes
- Mapping resolution is 5 cM or 5 megabase pairs
- Difficult for late-onset diseases

With whole-exome and whole-genome sequencing, family linkage analysis gets even more powerful
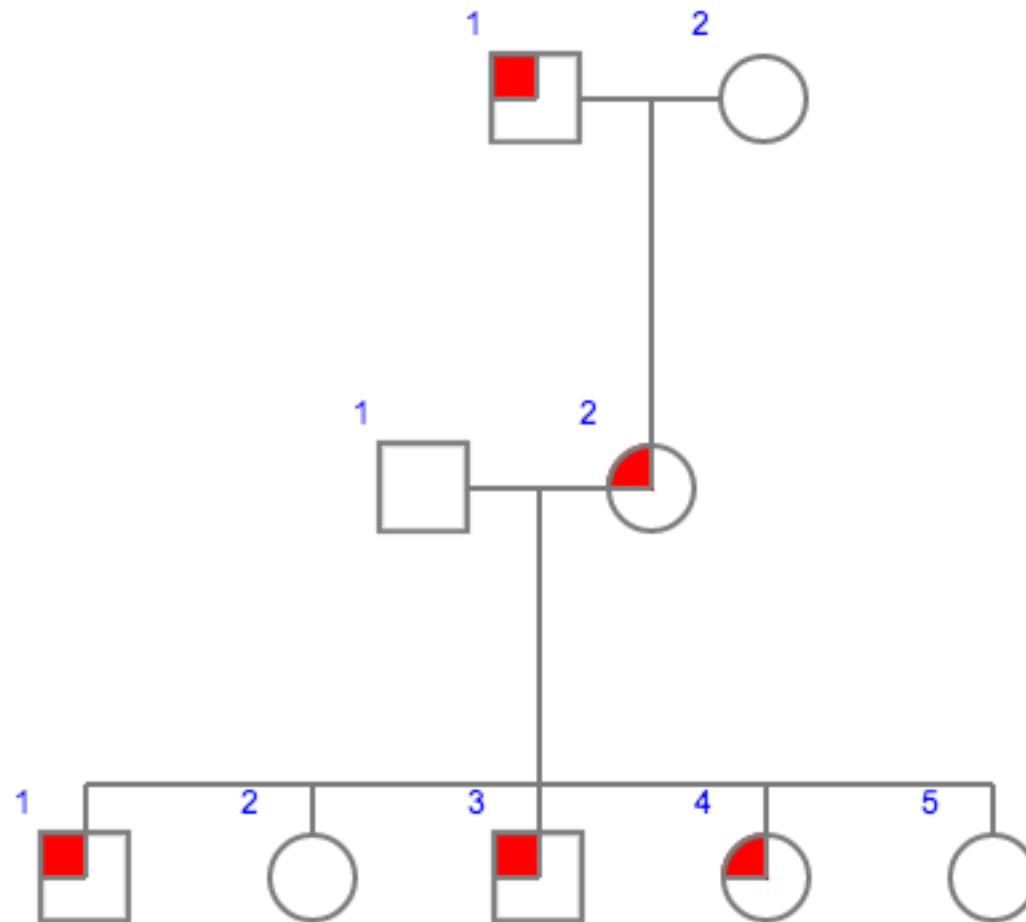
- Look at "all" markers in the genome simultaneously
- Dominant disorders mean look for heterozygous mutations linked to shared blocks of variants (haplotypes)
- X-linked lets you focus on the X chromosome
- Variants should be private to the family and deleterious

Linked markers on chromosomes allow for haplotype mapping