

## Linkage Mapping 2

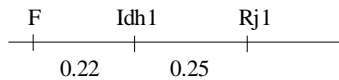
CS741  
2009  
Jim Holland

## Recombination Frequency Estimates

- They are ESTIMATES!
- Variance or standard error of the estimates can be obtained from 2<sup>nd</sup> derivative (see Mather, 1951).
- You can get better estimates (smaller std. deviations) by:
  - Increasing sample size
  - Using codominant genes/markers
  - Progeny testing F2:3s to better classify genotypes with dominant markers (e.g., distinguish AA from Aa)
  - Using F2 populations with codominant markers
  - Using backcross populations with dominant markers

## 3 – point linkage mapping

Gene Pair	Recombination Frequency
<i>F - Rj1</i>	0.41 ± 0.05
<i>F - Idh1</i>	0.22 ± 0.03
<i>Rj1 - Idh1</i>	0.25 ± 0.03

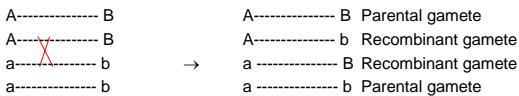


Notice: 0.22 + 0.25 > 0.41!

## Recombination frequencies are not additive!

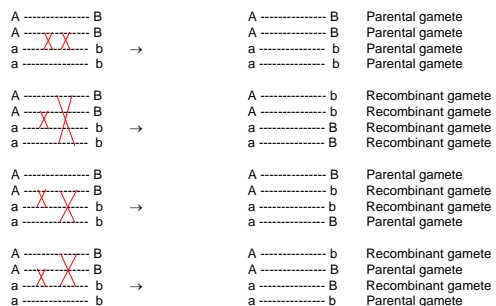
- The number of crossovers that occur along a chromosome is additive, but there is not a 1:1 relationship between crossover events and recombination.
- What is the relationship between crossovers and recombination?
- Remember that crossovers are the CAUSE and recombinations are the RESULT.

## One crossover



50% recombinant gametes produced!

## Two crossovers



Averaged over all four possibilities – 50% recombinant gametes are produced!

## Crossovers and Recombination

- Any number of crossovers greater than zero will produce 50% recombinant gametes on average.
- **This is why recombination frequency is not a linear function of the average number of crossovers between two loci.**
- If loci are widely separated on the chromosome, several crossovers may occur between them regularly at each meiosis, but they will still have only a maximum recombination frequency of 50%.

## Mapping Functions

- How to make a genetic map linear, so that map distances of separate intervals can be added up to equal distance of combined intervals?
- Mapping Functions! - genetic distances as a function of the presumed number of crossovers that occur between genes.
- **An interval in which an average of one crossover occurs every meiosis is defined to have a genetic map distance of 50 cM.**

## Mapping Functions

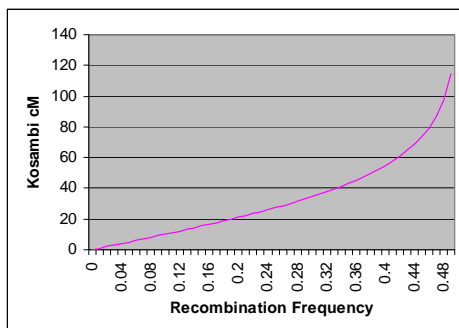
- The problem is that we do not actually observe crossovers, we observe the resulting recombinations.
- By making assumptions about the level of interference (whether crossovers are really independent along chromosome or if one crossover reduces the probability of a nearby crossover), the number of crossovers can be estimated from the observed recombination frequency.
- So, we are working backwards from the observed result (recombinations) to the unobserved cause (crossovers).
- The expected number of crossovers across different intervals then can be summed and treated as a linear map distance.

## Mapping Functions

- Haldane mapping function assumes no interference.
- Kosambi mapping function assumes a constant and specific level of interference:  

$$cM = \left(\frac{1}{4}\right) \ln\left(\frac{1+2r}{1-2r}\right) \times 100$$
- Kosambi is usually preferred, as there is evidence for interference, but in reality the level of interference is unknown, and anyway seems to vary across the genome (Sherman & Stack, 95)

## Map Distance vs. Recombination Frequency



## Map Distance: Example

<u>Gene Pair</u>	<u>Recomb. Freq.</u>	<u>cM</u>
<i>F - Idh1</i>	0.22 ± 0.03	24
<i>Rj1 - Idh1</i>	0.25 ± 0.03	27
<i>F - Rj1</i>	0.41 ± 0.05	58

Notice that the map units are supposed to be more additive than the original recombination units, but in reality they are not always so! Differences between the actual level of interference and the assumptions in the mapping function cause this.

## LOD scores

- An alternative method of measuring the evidence for linkage:
- “logarithm of odds”:

$$LOD = \log_{10} \left( \frac{L(\hat{r}_{MLE})}{L(r = 0.5)} \right)$$

## LOD scores

- From soybean example:

$$L(\hat{r} = 0.41) = \frac{250!}{144!44!39!23!} (.587025)^{144} (.162975)^{(44+39)} (.087025)^{23}$$

$$LOD = \log_{10} \left( \frac{L(\hat{r} = 0.41)}{L(r = 0.5)} \right) = \log_{10} \left( \frac{(.587025)^{144} (.162975)^{(44+39)} (.087025)^{23}}{(9/16)^{144} (3/16)^{(44+39)} (1/16)^{23}} \right)$$

$$= \log_{10} \left( \frac{8.02425 \times 10^{-124}}{9.59378 \times 10^{-125}} \right) = \log_{10}(8.364) = 0.9224$$

LOD 1 means estimated recombination frequency is 10 times more likely than 50% recombination.  
LOD 3 means estimated recombination frequency is 1000 times more likely than 50% recombination.

Typically in full genome linkage mapping, we consider genes with LOD 2.5 or more linked. Why use such a stringent level of significance?

## Multipoint Mapping in Mapmaker

- First, must select the linear order of loci in the linkage group.
- With many loci, it is not easy to know the correct order (huge numbers of possible orders).
- Tightly linked loci make it even harder!
- So, usually start with a subset of loci to make ordering simpler and a best guess of the order.
- Given the order, the maximum likelihood estimate (MLE) of the recombination frequencies is computed by extending the two locus multinomial probability function to a multiple locus probability function.

## Multipoint Mapping in Mapmaker

- Try different orders of loci.
- The LOD scores of the MLE for the different orders can be compared, the highest LOD score is selected.
- Additional loci can be added one at a time to the framework order to build up the map of the linkage group.
- At each step, compute the MLE and LOD score and also shuffle the order of loci a few times and compute LODs for each new order.
- Keep the order with best LOD score.
- You are not guaranteed the most likely order! Two different people can get two different maps from the same data set!

## Linkage Map Example

Portyanko et al., 2001

