

Linkage

Two genes are said to be linked if the transmission of alleles at one locus is stochastically dependent on the transmission of alleles at the other locus

Classical Test Cross

Parent 1 AABB x aabb **Parent 2**

F1 AaBb (100%) x aabb

F1-gametes

AB Ab aB ab

% frequencies

A and B are unlinked:

25 25 25 25

A and B linked:

35 15 15 35

A and B tightly linked

48 2 2 48

Classical Test Cross

Parent 1 AABB x aabb **Parent 2**

F1 AaBb (100%) x aabb

F1-gametes

AB Ab aB ab

% frequencies

A and B are unlinked:

25 25 25 25

A and B linked:

35 15 15 35

A and B tightly linked

48 2 2 48

R

R

R = Recombinants

Parent 1

AABB

x

aabb

Parent 2

F1

AaBb (100%)

F1-gametes

AB

Ab

aB

ab

frequencies

A and B are unlinked:

25

25

25

25

50

A and B linked:

35

15

15

35

30

A and B tightly linked

48

2

2

48

4

**% Recomb.
Frequency**

Corn example

plant 1 and plant 2 are coloured green plants
both are heterozygous

After test crossing

Progeny of plant 1 Progeny of plant 2

Coloured kernels, green plants	12	45
Coloured kernels, yellow plant	155	5
White kernels, green plants	115	3
<u>White kernels, yellow plant</u>	<u>18</u>	<u>27</u>

Recombination Fractions?

Corn example

plant 1 and plant 2 are coloured green plants
both are heterozygous

After test crossing

Progeny of plant 1 Progeny of plant 2

Coloured kernels, green plants	AaBb	12	45
Coloured kernels, yellow plant	Aabb	155	5
White kernels, green plants	aaBb	115	3
<u>White kernels, yellow plant</u>	<u>aabb</u>	<u>18</u>	<u>27</u>

Recombination Fractions?

Linkage equilibrium

$$\text{freq}(AB) = \text{freq}(Ab) = \text{freq}(aB) = \text{freq}(ab)$$

Linkage disequilibrium

$$D = \text{freq}(AB) \cdot \text{freq}(ab) - \text{freq}(Ab) \cdot \text{freq}(aB)$$

$D > 0$ through

- physical linkage
- selection (Bulmer effect)
- migration or crossing
- recent mutation

Linkage Disequilibrium is needed for mapping

- LD in a cross of inbred lines
- LD within a family
- LD population wide

after recent crossing
at very small distances

LD exists over a smaller distance only after many meiosis
(unless selection)

Mapping of genes

A.....B.....C

r_{AB}

r_{BC}

-
- Determine recombination fraction
- Distance is proportional to % recombination

Mapping functions

A.....B.....C

- recombination fractions are not additive

r_{AC} is not equal to $r_{AB} + r_{BC}$.

- With *no interference*.

$$r_{AC} = r_{AB} + r_{BC} - 2 \cdot r_{AB} \cdot r_{BC}.$$

Interference is the effect in which the occurrence of a crossover in a certain region reduces the probability of a crossover in the adjacent region.

- With *no interference*.

$$r_{AC} = r_{AB} + r_{BC} - 2 \cdot r_{AB} \cdot r_{BC}.$$

- With *complete interference*

$$r_{AC} = r_{AB} + r_{BC} \quad \text{additive}$$

- In *any case*:

With small distance between AC (<20 cM)

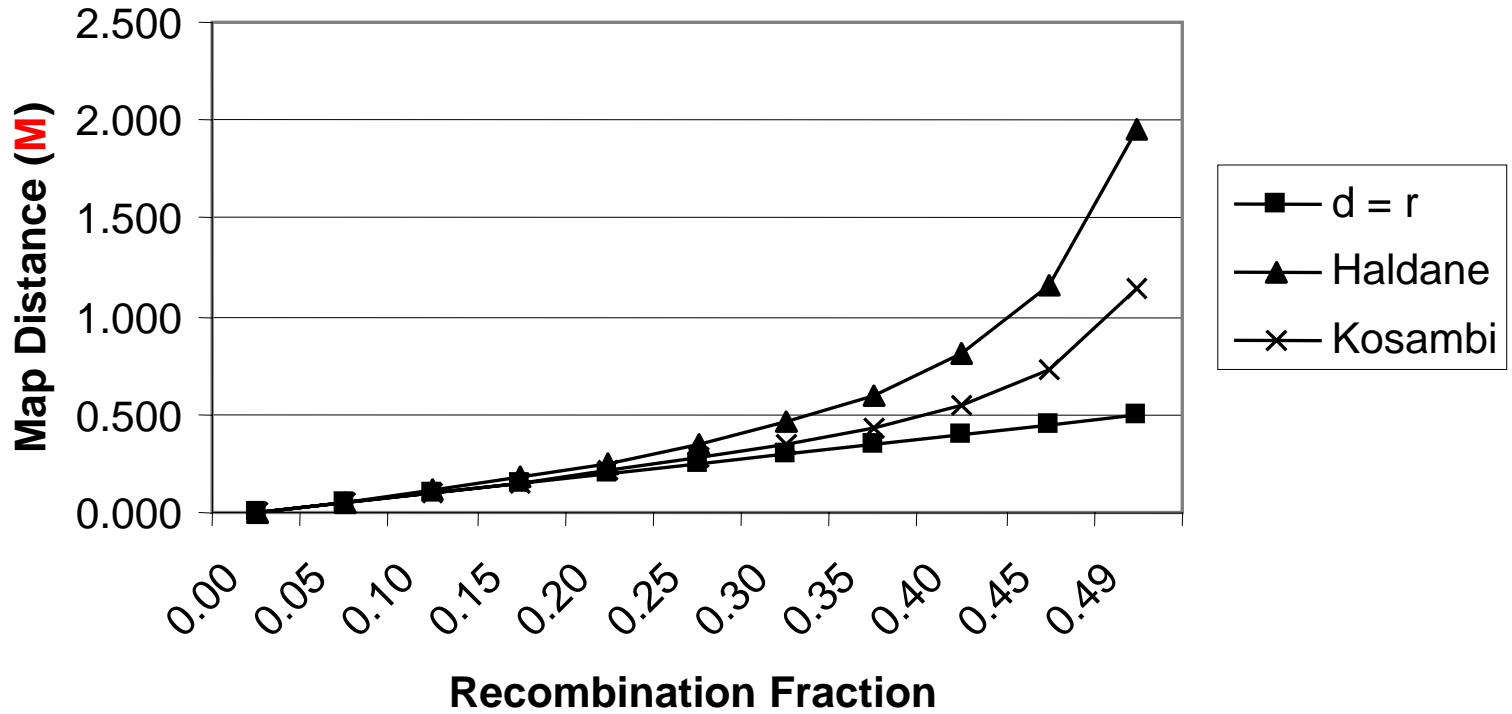
$$r_{AC} \approx r_{AB} + r_{BC} \quad \text{additive}$$

A mapping function gives relationship
between recombination fraction and distance

A mapping function should

- Assign distances that are additive
- If distance > 50 cM, this should translate into a recombination fraction of 50%.

Mapping Functions



No interference > *Haldane mapping function*

Some Interference > *Kosambi mapping function*

Complete interference > *$d = r$*

Mapping functions

- Haldane

$$d = -\frac{1}{2} \ln(1-2r).$$

$$r = \frac{1}{2} (1-e^{-2d})$$

- Kosambi

$$d = \frac{1}{4} \ln\left[\frac{1+2r}{1-2r}\right]$$

Example Haldane mapping

Distance (cM)	% Recombination
10	0.091
20	0.165
30	0.226
40	0.275

Mapping of genetic markers

- Determine recombination %
- Use a mapping function

Estimation of recombination fraction

- Linkage phase in parents AB / ab
- haplotype of the gametes $AB \ aB \ Ab \ ab$
- Linkage phase is not always known
- Haplotype is not always known
 - depends on genotypes of parents/progeny

Maximum Likelihood estimation of linkage (r)

example using 1 offspring only

$$L = \sum_{\text{phases}} \sum_{\text{gametes}} \text{Prob}(\text{phase}_i) * \text{Prob}(\text{gamete}_j | \text{phase}, \text{recomb})$$

Example: Sire: **AaBb** Dam: **AABB**

 Offspring: **AABB**

Sires genotype	Prob	Prob(AB gamete)
AB/ab	0.5	0.5*(1-r)
Ab/aB	0.5	0.5*r

$$L = 0.5 * \{0.5 * (1-r)\} + 0.5 * \{0.5 * r\} = 0.25$$

Likelihood does not depend on r >> No information about r

Maximum Likelihood estimation of linkage

example using 2 offspring

$$L = \sum_{\text{phases}} \sum_{\text{gametes}} \text{Prob}(\text{phase}_i) * \text{Prob}(\text{gamete}_j | \text{phase}, \text{recomb})$$

Example: Sire: **AaBb** Dam: **AABB**

Offspring 1: **AABB**

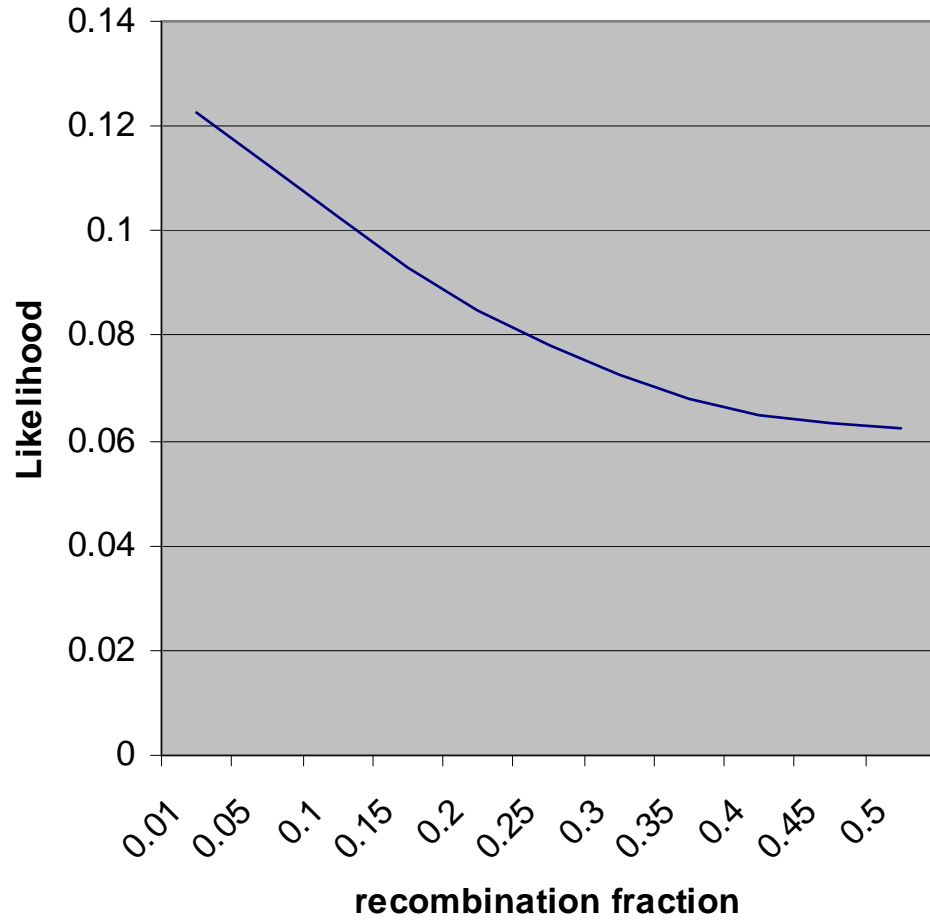
Offspring 2: **AABB**

Sires genotype	Prob	Prob(2 AB gametes)
AB/ab	0.5	$0.5 * (1-r)^2$
Ab/aB	0.5	$0.5 * r^2$

$$L = 0.5 * \{0.5 * (1-r)^2\} + 0.5 * \{0.5 * r^2\} = 0.125 * \{(1-r)^2 + r^2\}$$

Likelihood now depends on r >> find MaxL

Likelihood map

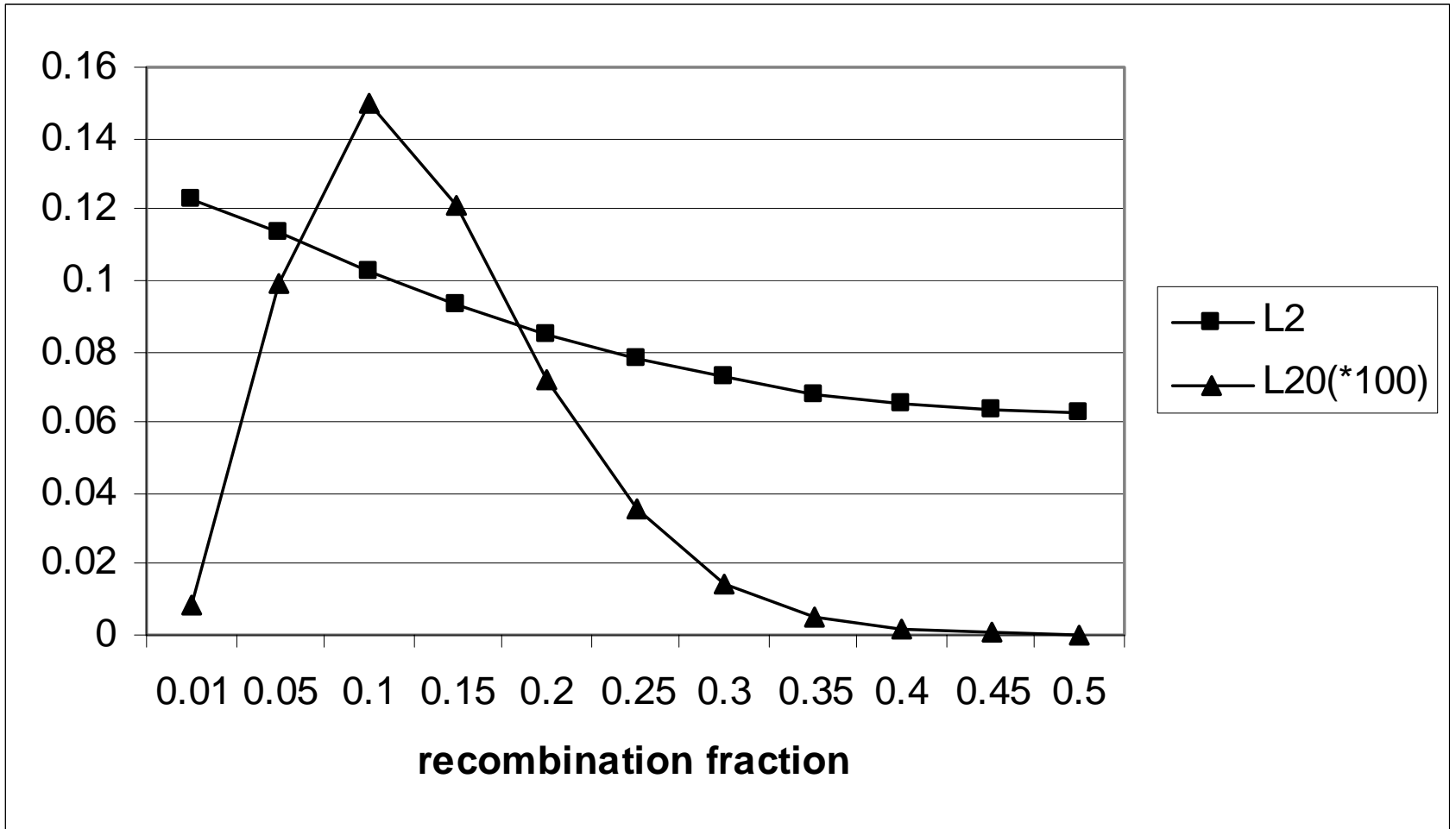


Maximum Likelihood estimation of linkage

$$L = \sum_{\text{phases}} \sum_{\text{gametes}} \text{Prob}(\text{phase}_i) * \text{Prob}(\text{gamete}_j | \text{phase}, \text{recomb})$$

Example:	Sire: AaBb	Dam: AABB	
	20 Offspring:	AABB	9
		AaBB	1
		AABb	1
		AaBb	9

$$L = \binom{20}{2} (1-r)^{18} \cdot r^2 + \binom{20}{2} (1-r)^2 \cdot r^{18}$$



Likelihood Ratio test

- To compare the likelihood of two alternative models, and test the difference

$$\text{Log} \frac{\textit{Likelihood}(r = \hat{r})}{\textit{Likelihood}(r = 0.5)}$$

The LOD score

$$^{10}\text{Log} \frac{\textit{Likelihood}(r = \hat{r})}{\textit{Likelihood}(r = 0.5)} = \frac{\textit{LR}(r)}{4.61}$$

LOD score > 3 means: < 1 : 1000

CRI-map

LOD scores

estimates of rec.fract.

Possibly many markers

many families

Uses Kosambi's function

Designs needed for mapping

- Amount of information depends on number of informative meioses
- An efficient design minimizes the number of genotypings (minimizes uninformative meioses)
- Full sibs better than half sibs
- Use more families to avoid uninformative sites