HOS 6236 Molecular Marker Assisted Plant Breeding

Fall 2017

Last Class:

Linkage mapping, rationale, populations

Todays Class:

Linkage mapping construction

Developing a Linkage Map

1. Develop or identify a “mapping” population and determine sample size
2. **Genotype entire population with appropriated molecular markers (Dom or Co-dom)**
3. Perform Linkage analyses − filter markers, test segregation, calculate recombinations, establish linkage groups, estimate map distances, determine map order (software)

Linkage Map – Molecular Markers

Many types of molecular markers exist (check molecular marker review on week 2). The best markers are:

Reliable/Repeatable
Highly polymorphic
Informative
Type - dominant vs co-dominant
Cost??

Linkage Map − Molecular Marker Density

In general, marker density depends on:

Objective of the linkage map

Number of individuals in mapping pop

Relationship between linkage map and physical map

Developing a Linkage Map

1. Develop or identify a “mapping” population and determine sample size

2. Genotype entire population with appropriated molecular markers (Dom or Co-dom)

3. **Perform Linkage analyses – filter markers, test segregation, calculate recombinations, establish linkage groups, estimate map distances, determine map order (software)**

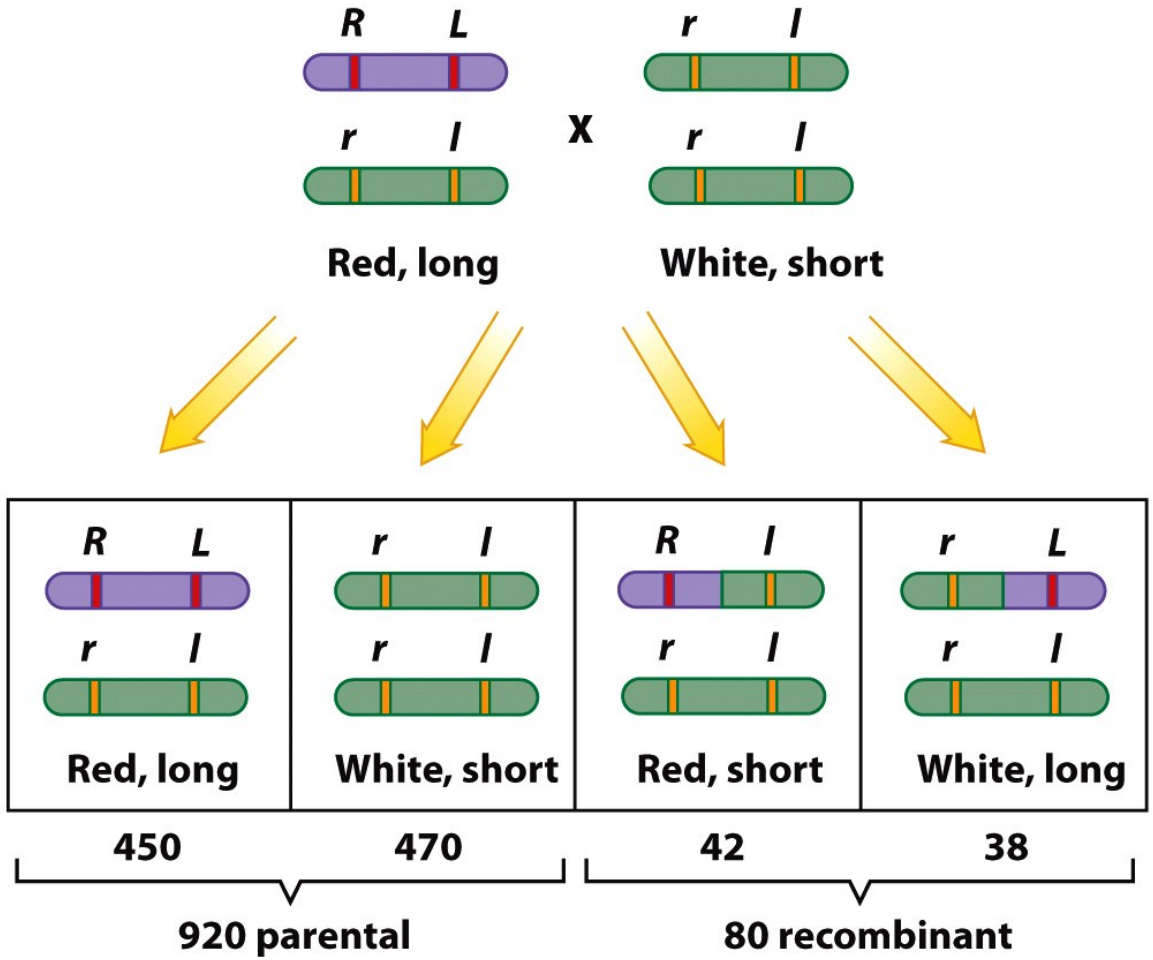
Cleaning the data

Filtering the markers

No segregation from markers homozygous on both parents
Errors in sequencing/genotyping
Segregation distortion (Chi-square test, tutorial)

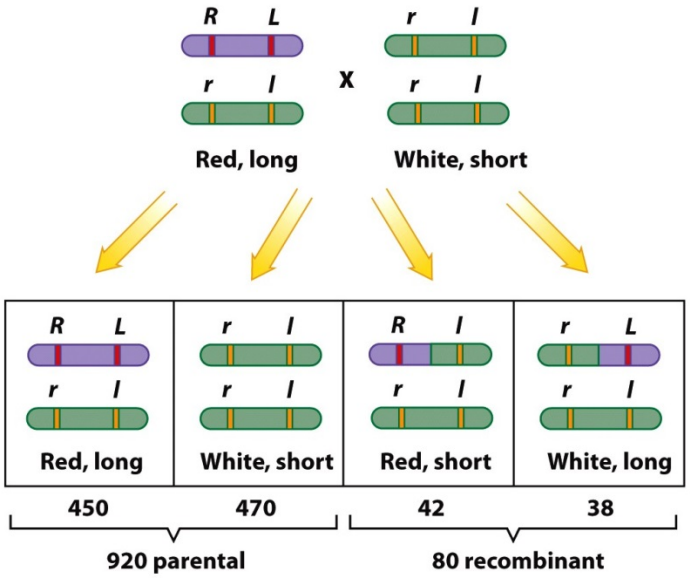
X^2 value= SUM $[(O-E)^2/E]$

degree of freedom = # categories − 1
chi-square p-value using excel
=chisq.dist.rt(X^2 value, *df*)


R L
r l
X
r l
r l
Red, long
White, short
R L
r l
r l
Red, long
r l
r l
White, short
R l
r l
Red, short
r L
r l
White, long
450
470
42
38
920 parental
80 recombinant

Determining chi-square values

1. State the hypothesis being tested and the predicted results.
2. Determine the expected numbers for each observed class (use numbers, not percentages or ratios).
3. Calculate X^2 using the formula.
4. Use the chi-square distribution table to determine the significance of the value.



X^2 = Sum of $(O-E)^2/E$

O = Observed number for each class

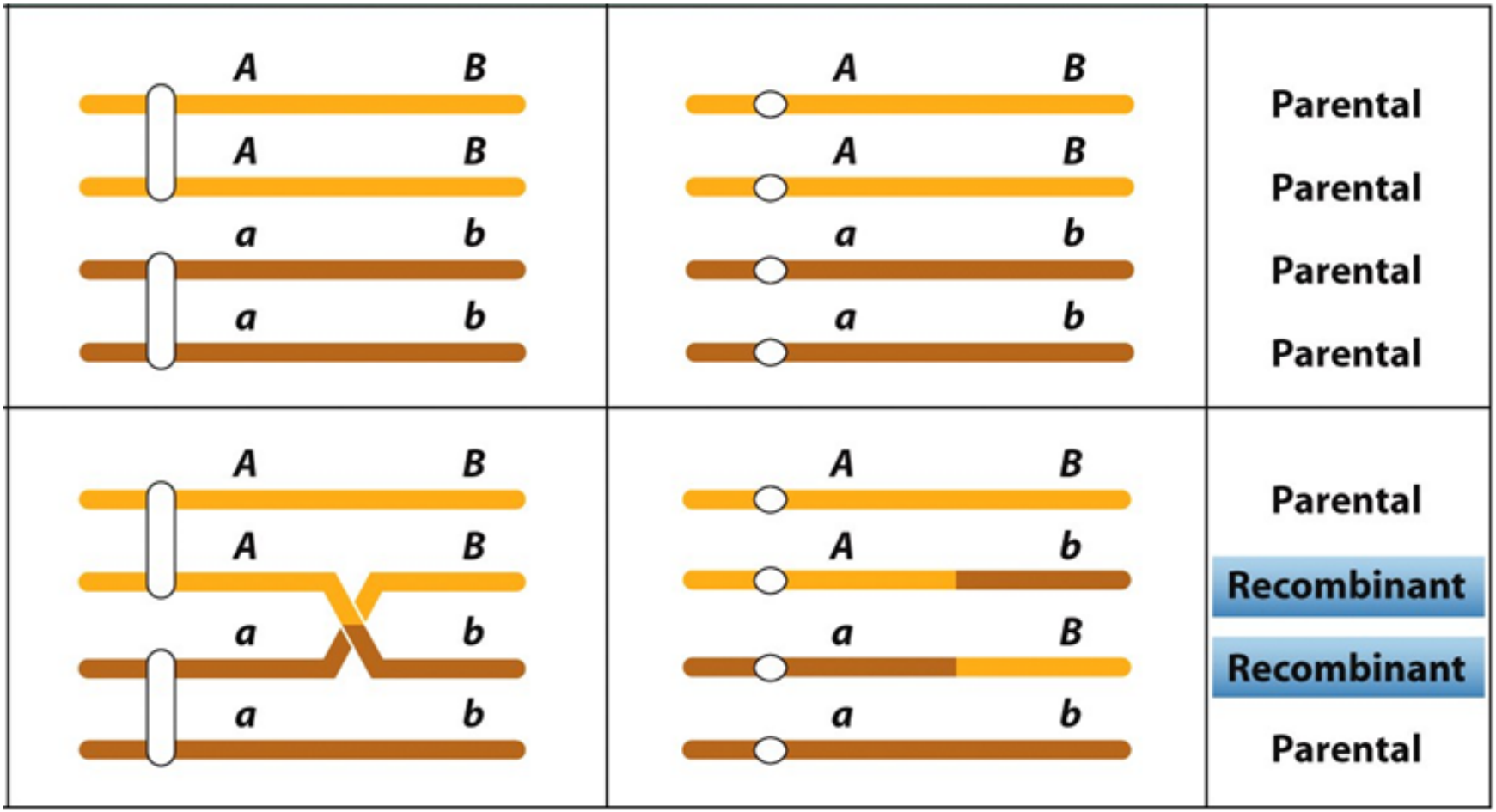
E = Expected number for each class

F_1 Gamete	Observed	Expected	$(O-E)^2/E$
R, L	450	250	160
R, l	42	250	173
r, L	38	250	180
r, l	470	250	194
Total	**1000**	**1000**	**X^2=707**

	Probability (*p*)										
df	**0.95**	**0.90**	**0.80**	**0.70**	**0.50**	**0.30**	**0.20**	**0.10**	**0.05**	**0.01**	**0.001**
1	0.004	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	6.64	10.83
2	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.60	5.99	9.21	13.82
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.82	11.34	16.27
4	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	13.28	18.47
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	15.09	20.52
6	1.63	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	16.81	22.46
7	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	18.48	24.32
8	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	20.09	26.12
9	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	21.67	27.88
10	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	18.31	23.21	29.59
	Nonsignificant								**Significant**		

Linkage Map – recombination

- Recombination is a result of physical crossing over of chromosomes during meiosis



Griffiths et al, Introduction to Genetic Analysis, 10th Ed. W. H. Freeman and Company, 2011

Linkage Map − how recombination is used to map markers/genes

A *− round fruit*
a *− elongated fruit*

C *− simple flower*
c *− composite flower*

AA x ***aa***

F1: ***Aa***

F2: ***AA Aa Aa aa***

3 round : 1 elongated

CC x ***cc***

F1: ***Cc***

F2: ***CC Cc Cc cc***

3 simple : 1 composite

AACC *(round-simple)* ***x*** ***aacc*** *(elongated-composite)*

F1: AaCc *(round-simple)*

Linkage Map – how recombination is used to map markers/genes

AACC *(round-simple)* ***x*** ***aacc*** *(elongated-composite)*

F1: AaCc *(round-simple)*

Backcross: AaCc x aacc 210 individuals

	Expected	Observed	Phenotypic type
Round-simple (AaCc)	25% = 52.5	85	Parental
Round-composite (Aacc)	25% = 52.5	23	Recombinant
Elongated-simple (aaCc)	25% = 52.5	19	Recombinant
Elongated-composite (aacc)	25% = 52.5	83	Parental

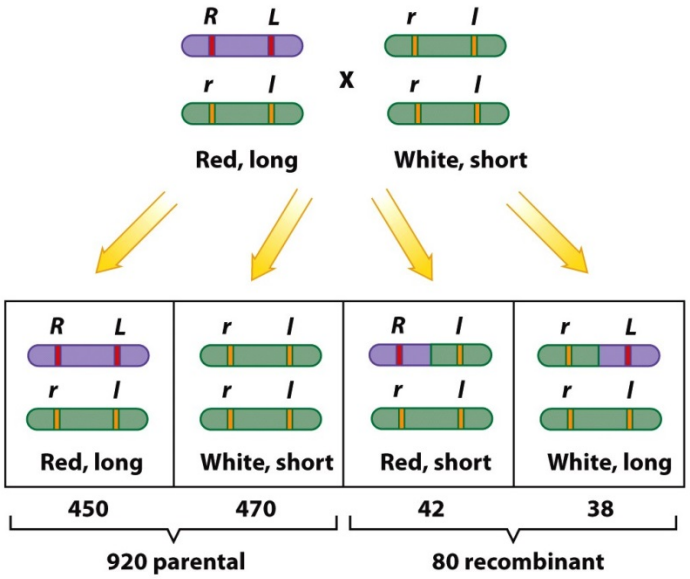
Linkage Map – how recombination is used to map markers/genes

$$\textit{Recombination Fraction} = \frac{\textit{\# of recombinants}}{\textit{total \# progeny}} \times 100$$

	Expected	Observed	Phenotypic type
Round-simple (AaCc)	25% = 52.5	85	Parental
Round-composite (Aacc)	25% = 52.5	23	Recombinant
Elongated-simple (aaCc)	25% = 52.5	19	Recombinant
Elongated-composite (aacc)	25% = 52.5	83	Parental

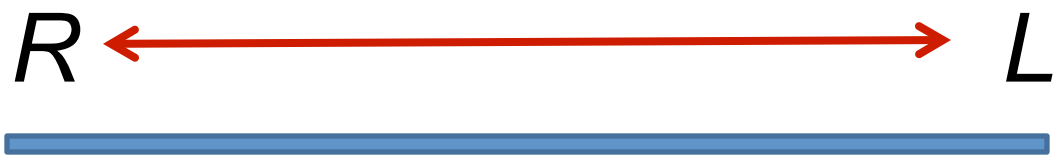
$$\textit{Recombination Fraction} = \frac{(23+19)}{210} \times 100$$

Recombination Fraction = 20 % or 20 mu or 20cM between the two loci



What would the recombination fraction (r) be?

How many cM apart would these genes be?



Genetic Distance (cM)

- Genetic distances are additive

If: $r_{RL} = 0.08 = 8$ cM

and: $r_{LC} = 0.15 = 15$ cM

then: $r_{RC} =$ 23 cM

Genetic Distance (cM)

- Genetic distances are additive

If: $r_{RL} = .08 = 8$ cM

and: $r_{LC} = .15 = 15$ cM

then: $r_{RC} = .20$

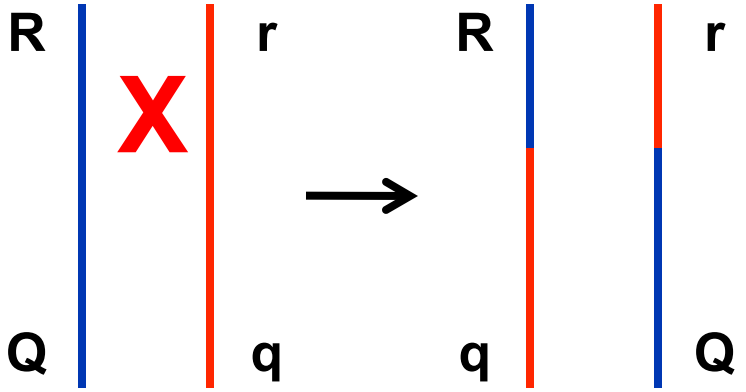
What if you calculated r for the R and C loci and got 0.20 instead of 0.23?

Genetic Distance (cM)

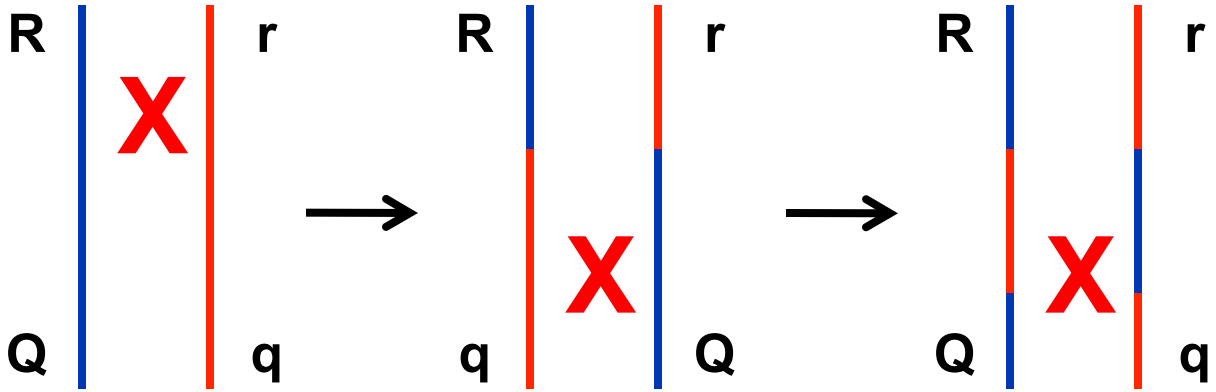
- When two genetic units are far apart, 1 cM does not correspond to 1% recombination frequency

Multiple recombination events tend to lead to an underestimation of actual linkage distance

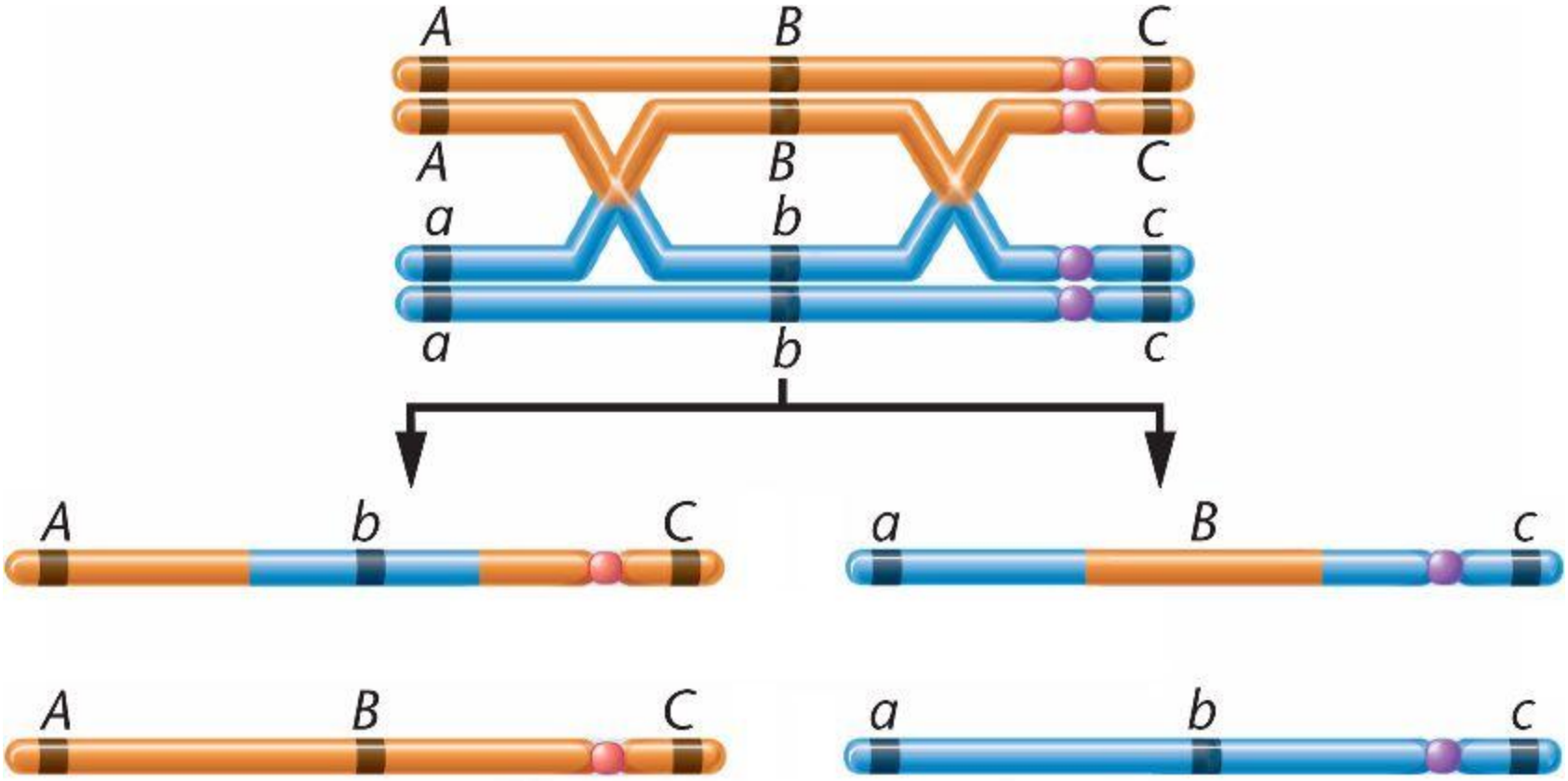
Double Crossovers



- Double crossover is hidden recombination



Double cross-over underestimate distance between distant markers



Griffiths et al, Introduction to Genetic Analysis, 10th Ed. W. H. Freeman and Company, 2011

Mapping Functions

- Mapping functions establish the relationship between recombination (r) observed and the map distance (cM). They may or not use interference estimates to take into account double crossovers
  - Interference = the phenomenon where a single crossover in a chromosome region will reduce the probability of a second crossover occurring in the same region

Mapping Functions

- No interference
  - Double or multiple crossovers occur at random
- Complete interference
  - The occurrence of one exchange (cross-over) between homologous chromosomes prevents another in its vicinity

Haldane's Function

- Assumes no interference

$$\text{cM} = -\frac{1}{2}\ln(1 - 2r)$$

  - When $r = 0$, cM = 0 (complete linkage)
  - When $r = 0.5$, cM = ∞ (unlinked)

Kosambi's Function

- Assumes interference

$$cM = \frac{1}{4} \ln\left(\frac{1 + 2r}{1 - 2r}\right)$$

  - As *r* increases, the amount of interference allowed by the Kosambi function decreases

Haldane or Kosambi?

- To most accurately choose, the degree of crossover interference must be known
  - For practical purposes – if marker density is less than 10 cM, Haldane's and Kosambi's functions are equal
  - Kosambi's function is often used for higher organisms when the degree of interference is unknown

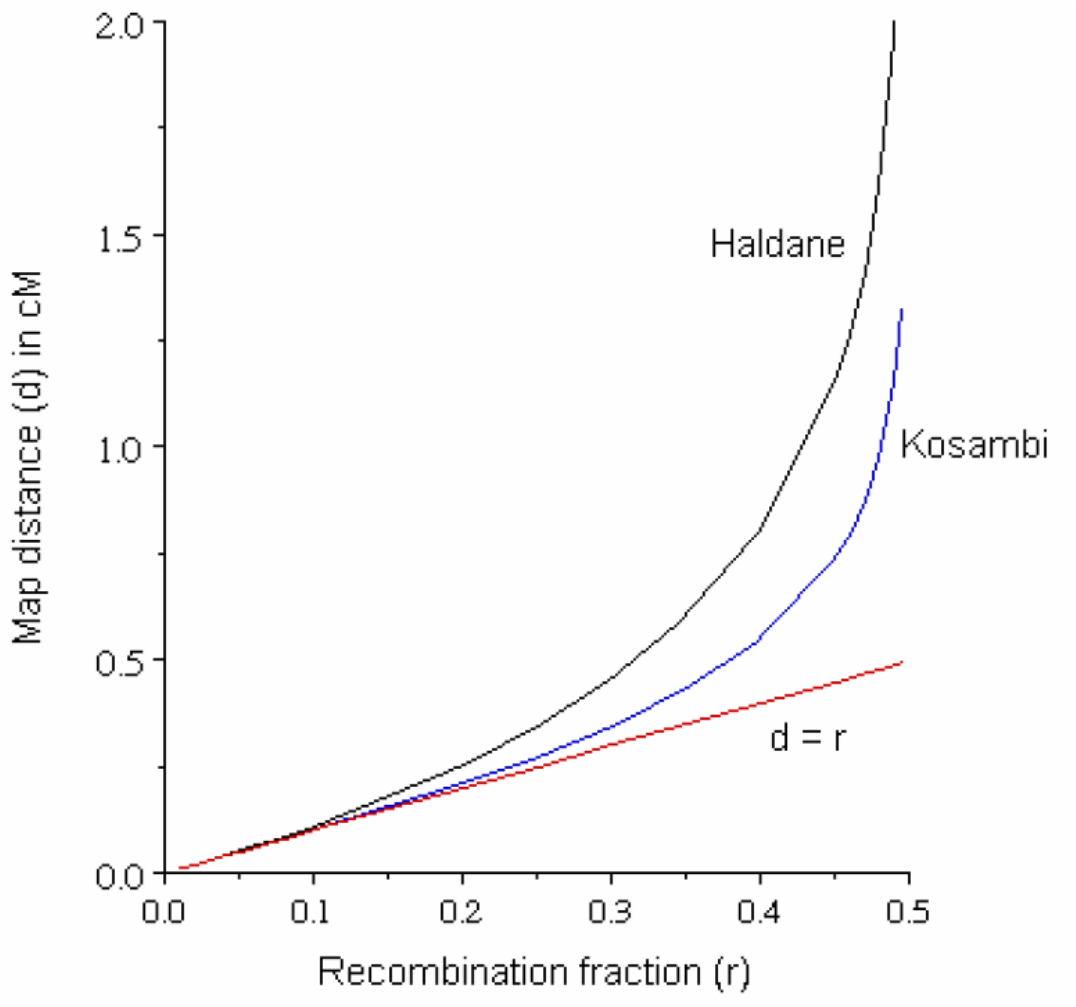


Figure 7. Comparison of Haldane's and Kosambi's mapping functions. Below a recombination frequency of 0.1 (10%), there is almost no difference between the two mapping functions (r = d). For recombination frequencies above 0.1 or 10%, Haldane mapping function gives higher map distance than Kosambi mapping function.

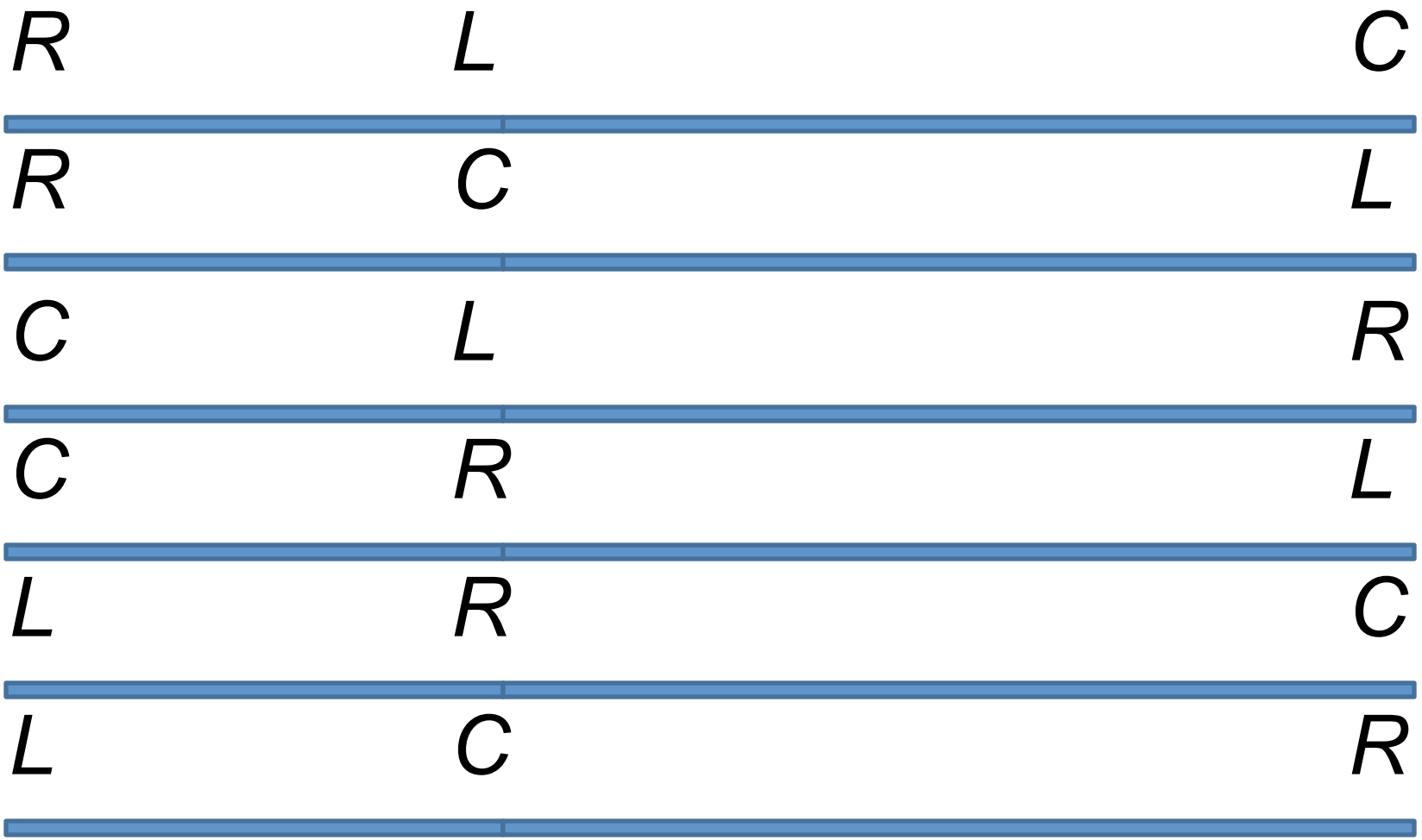
Semagn et al., 2006

Extending the Linkage Map

r_{RL} = .08
r_{LC} = .15
r_{RC} = .20

- The order for these three loci is unambiguous based on observed recombination

Extending the Linkage Map



- 3! potential orders = 6

Extending the Linkage Map

- What if a fourth locus *D* is added: 4! = 24 possible orders.

- What if you have a map of 100 marker loci?

Extending the Linkage Map

- Use a stepwise addition to add markers

LOD – Logarithm of the odds

- Used to determine statistical significance of genetic linkages

$$\text{LOD} = \log \left[\frac{\text{prob. of obtaining the observed data under linkage}}{\text{prob. of obtaining the observed data under random assortment}} \right]$$

$p < 0.01$ ~ LOD 2.0

$p < 0.001$ ~ LOD 3.0

LOD Score

LOD (Logarithm of the Odds) score is a statistic that describes the **strength of evidence** for linkage. It is based on **likelihood ratio (LR)**.

The LOD score compares the likelihood of obtaining the test data (**the two loci are indeed linked**) to the likelihood of observing the same data purely by chance (**the markers are unlinked**).

Using a certain **threshold value** of the LOD score for considering two markers **significantly linked**, will then group the markers in **linkage groups**.

LOD Score

The likelihood ratio (LR) is defined by

$$LR = \frac{\theta^{R} \times (1-\theta)^{NR}}{(0.5)^{NR+R}} = \frac{\text{Likelihood if linked}}{\text{Likelihood if unlinked}}$$

The **LR** is known as **odds ratio**

where θ represents the recombination fraction

R, the number of recombinant offspring

NR, the number of non-recombinant offspring

0.5 at the denominator refers to the 50% chance of recombination due to independent assortment (markers are completely unlinked)

The **LOD score** is the **$\log_{10}$ (LR)**

Using **log10** allows a more intuitive interpretation

LOD Score

Interpreting LOD score:

LOD values increases on a logarithmic scale, i.e., each increase of 1 unit in LOD implies a 10-fold increase in the likelihood ratio.

Example:

LOD = 2 indicates that the occurrence of linkage is **100 times more likely** than that of independent segregation

LOD = 3 indicates that the occurrence of linkage is **1,000 times more likely** than that of independent segregation.

…

LOD Score

LOD threshold for the establishment of linkage groups:
- Ranging from **4** to **6** (literature consensus)

When there is a large number of markers:
LODs with low values (less than 4) may cause **spurious linkage**,
LODs with high values (more than 6) could result in **fragmentation of real linkage groups**.