NC STATE UNIVERSITY

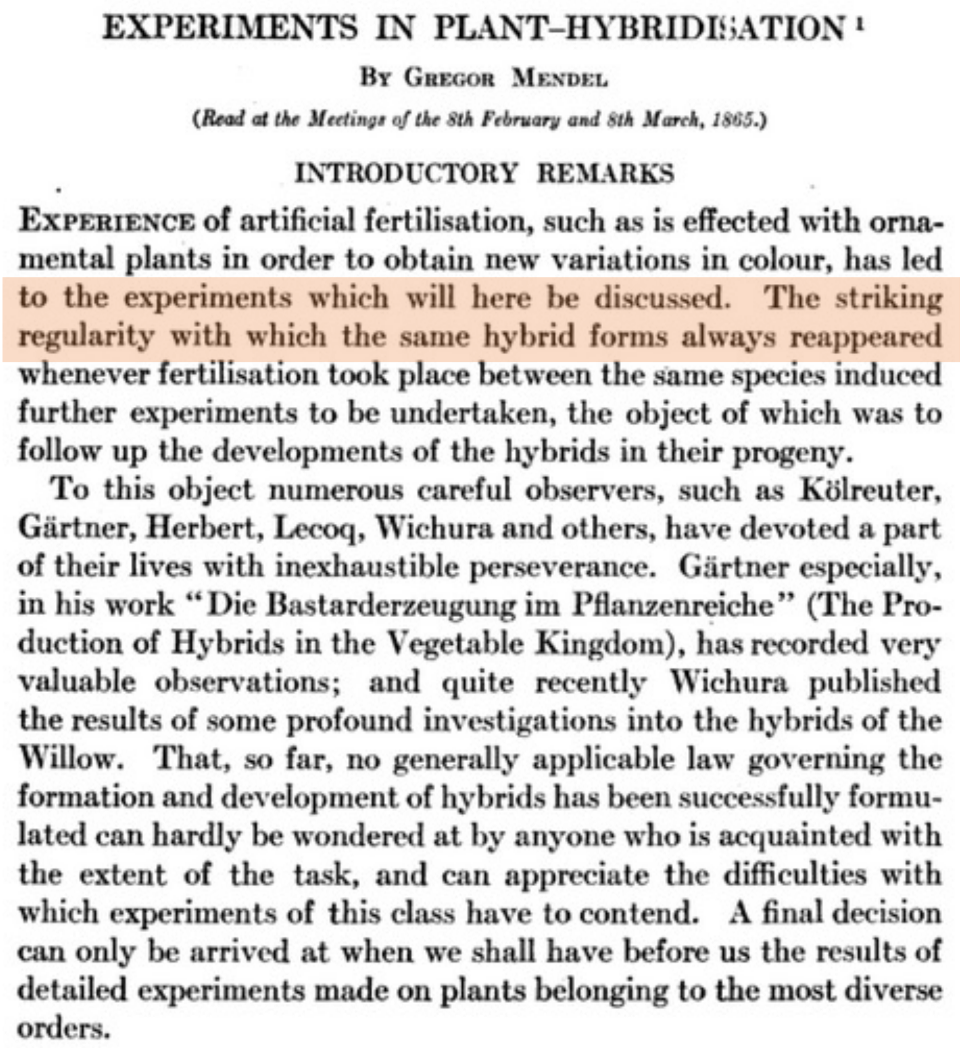
Theoretical and Practical Aspects of Linkage Analysis in Diploids and Polyploids

Marcelo Mollinari
mmollin@ncsu.edu

Breeding Insight Workshop - Ithaca – NY – August 2023

This work is licensed under CC BY 4.0

Introductory Remarks by Mendel



EXPERIMENTS IN PLANT-HYBRIDISATION[1]

BY GREGOR MENDEL

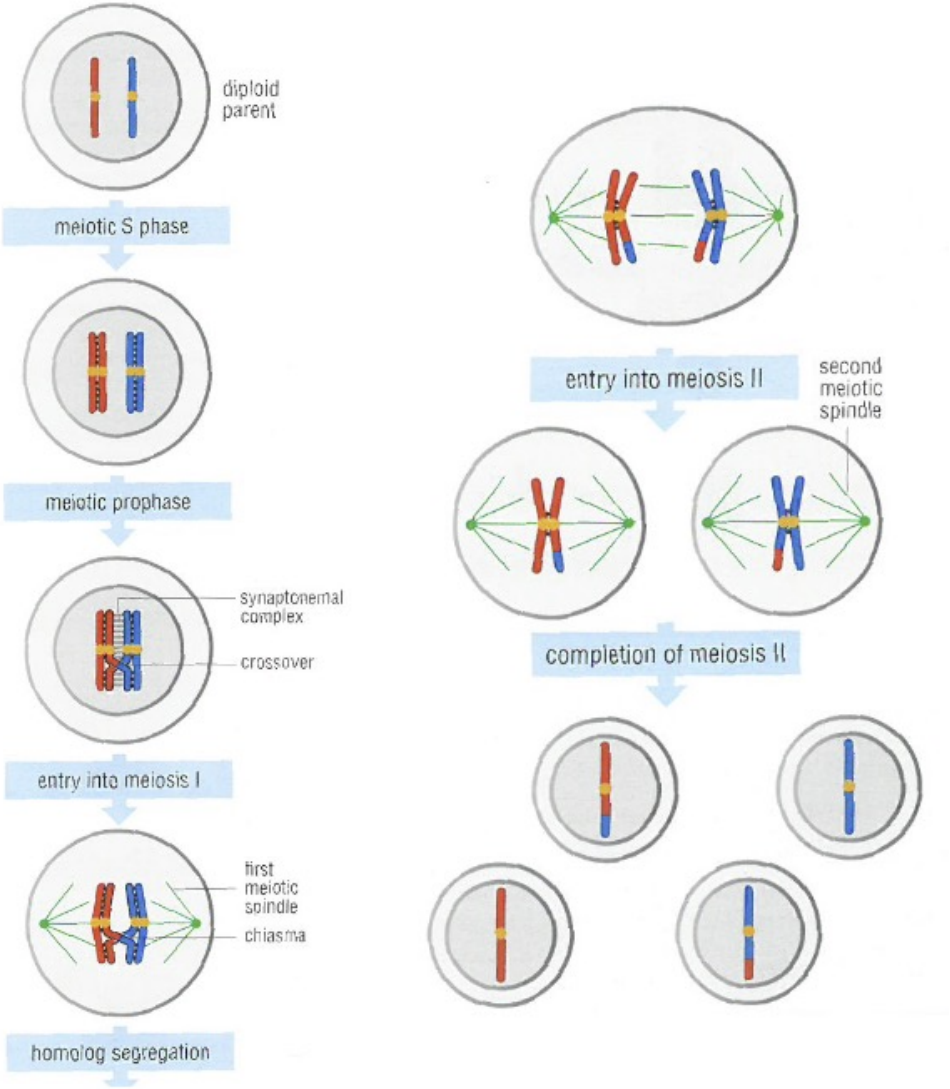
(Read at the Meetings of the 8th February and 8th March, 1865.)

INTRODUCTORY REMARKS

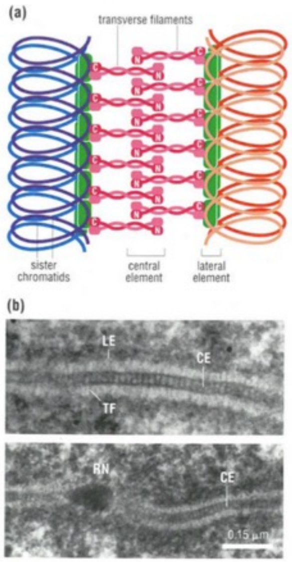
EXPERIENCE of artificial fertilisation, such as is effected with ornamental plants in order to obtain new variations in colour, has led to the experiments which will here be discussed. The striking regularity with which the same hybrid forms always reappeared whenever fertilisation took place between the same species induced further experiments to be undertaken, the object of which was to follow up the developments of the hybrids in their progeny.

To this object numerous careful observers, such as Kölreuter, Gärtner, Herbert, Lecoq, Wichura and others, have devoted a part of their lives with inexhaustible perseverance. Gärtner especially, in his work "Die Bastarderzeugung im Pflanzenreiche" (The Production of Hybrids in the Vegetable Kingdom), has recorded very valuable observations; and quite recently Wichura published the results of some profound investigations into the hybrids of the Willow. That, so far, no generally applicable law governing the formation and development of hybrids has been successfully formulated can hardly be wondered at by anyone who is acquainted with the extent of the task, and can appreciate the difficulties with which experiments of this class have to contend. A final decision can only be arrived at when we shall have before us the results of detailed experiments made on plants belonging to the most diverse orders.

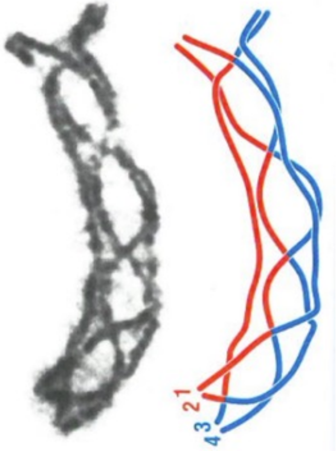
Meiosis in diploid



Meiosis is a specialized form of nuclear division that generates nuclei carrying **half** the normal complement of chromosomes.



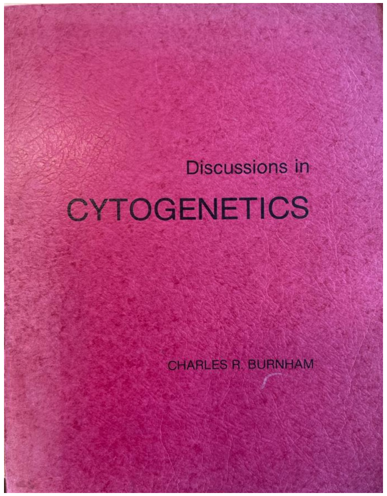
Synaptonemal complex

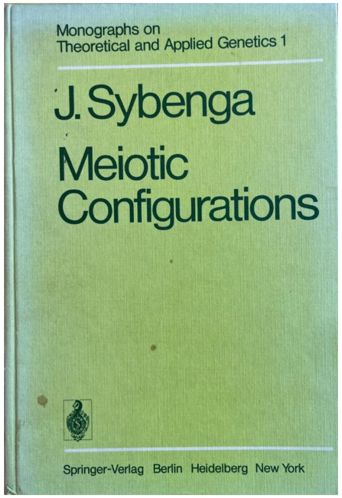


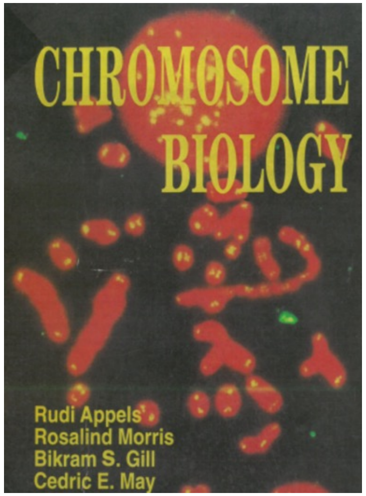
Chiasmata

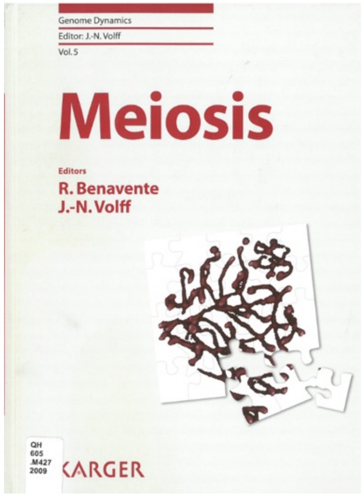
Meiosis in diploids

Cytogenetics Literature – diploids and polyploids


Discussions in
CYTOGENETICS
CHARLES R. BURNHAM

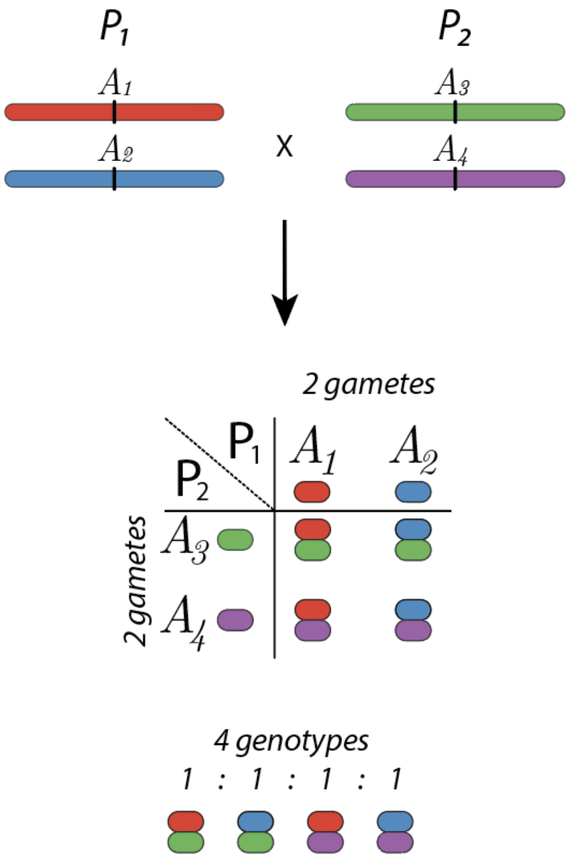

Monographs on
Theoretical and Applied Genetics 1
J. Sybenga
Meiotic
Configurations
Springer-Verlag Berlin Heidelberg New York


CHROMOSOME
BIOLOGY
Rudi Appels
Rosalind Morris
Bikram S. Gill
Cedric E. May

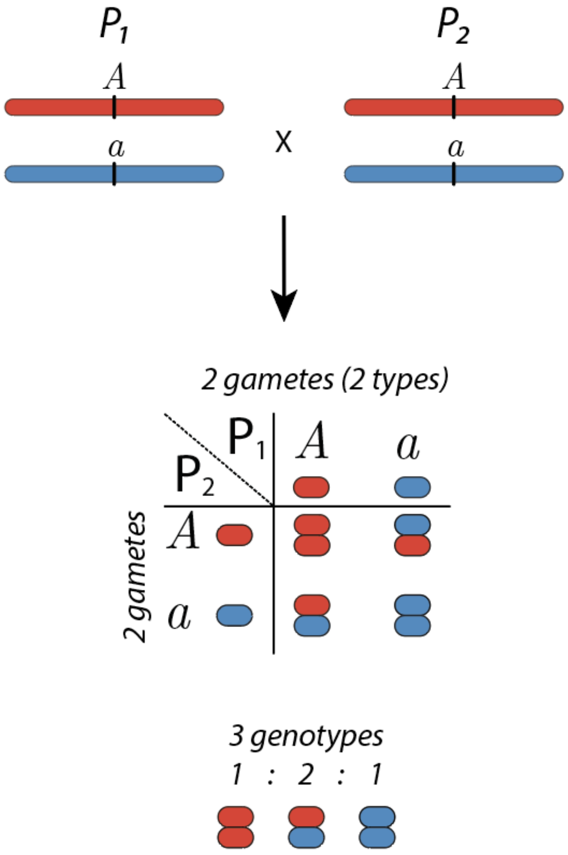

Genome Dynamics
Editor: J.-N. Volff
Vol. 5
Meiosis
Editors
R. Benavente
J.-N. Volff
QH
605
.M427
2009
KARGER

Segregation in diploids

Multiallelic



Biallelic

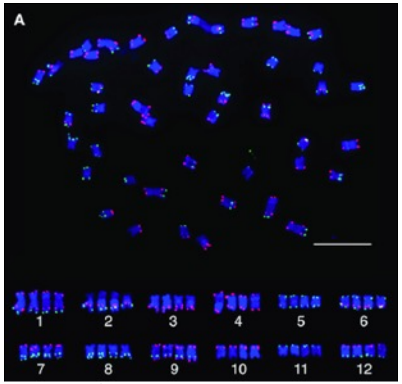


Polyploid species

Autohexaploid

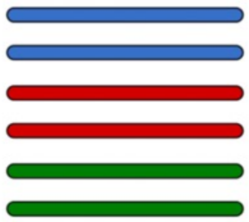
Multiple sets have the same origin

Autotetraploid potato



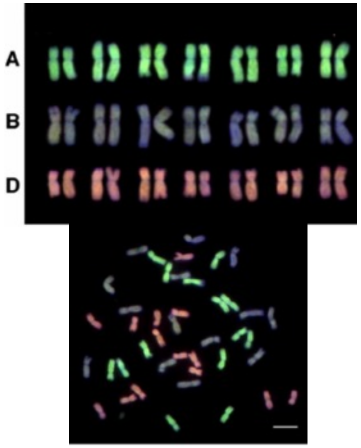
Braz et al., 2018

Allohexaploid



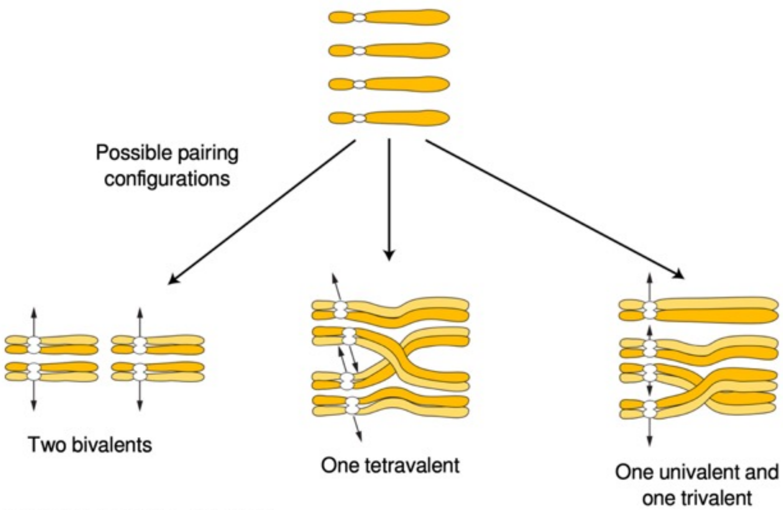
Multiple sets have different origins

Allohexaploid wheat



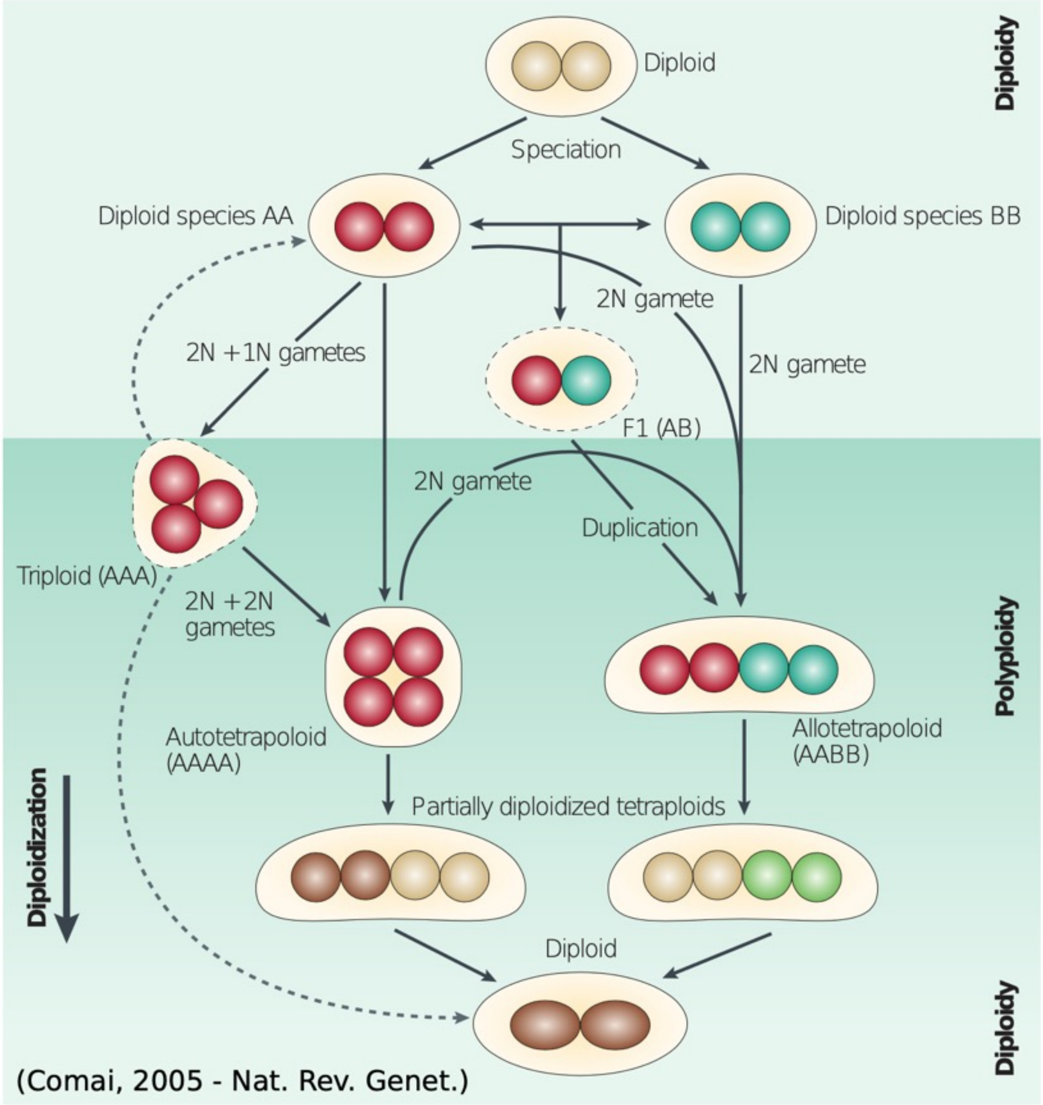
Zhang et al., 2013

Meiotic pairing in autotetraploids



Griffiths et al. (2004)

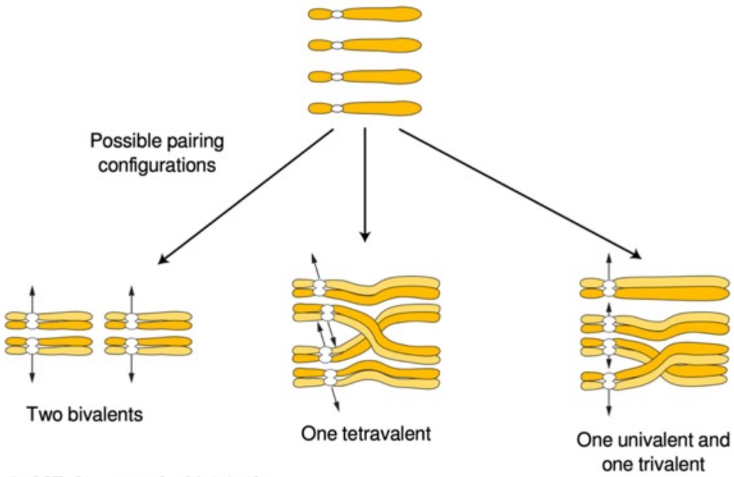
How polyploids are formed



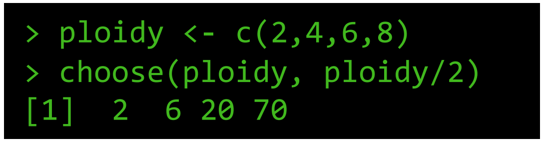
(Comai, 2005 - Nat. Rev. Genet.)

Autopolyploids

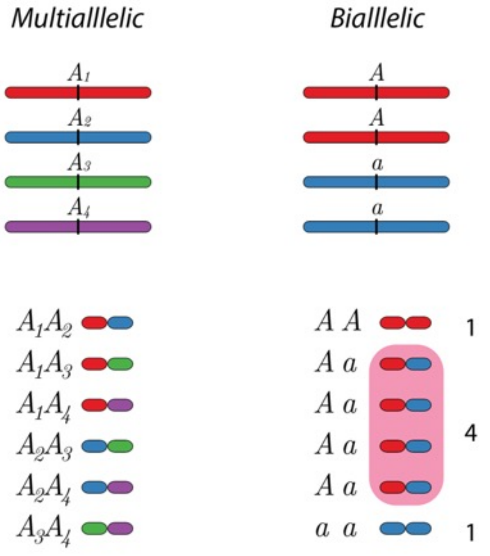
Gamete formation in autopolyploids*

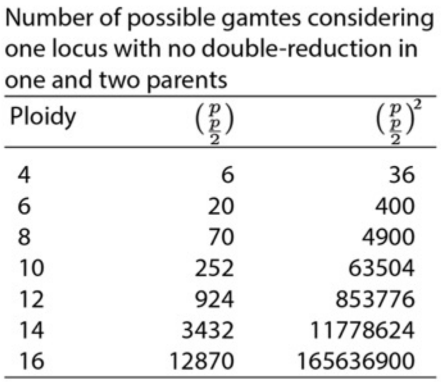


Griffiths et al. (2004)



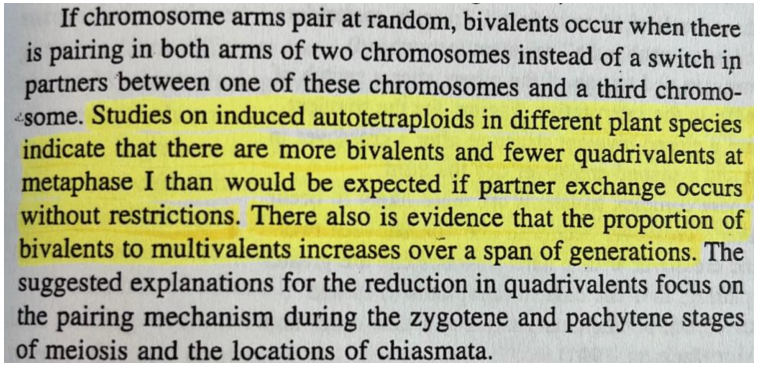
```
> ploidy <- c(2,4,6,8)
> choose(ploidy, ploidy/2)
[1]  2  6 20 70
```





Number of possible gamtes considering one locus with no double-reduction in one and two parents

Ploidy	$\binom{p}{\frac{p}{2}}$	$\binom{p}{\frac{p}{2}}^2$
4	6	36
6	20	400
8	70	4900
10	252	63504
12	924	853776
14	3432	11778624
16	12870	165636900

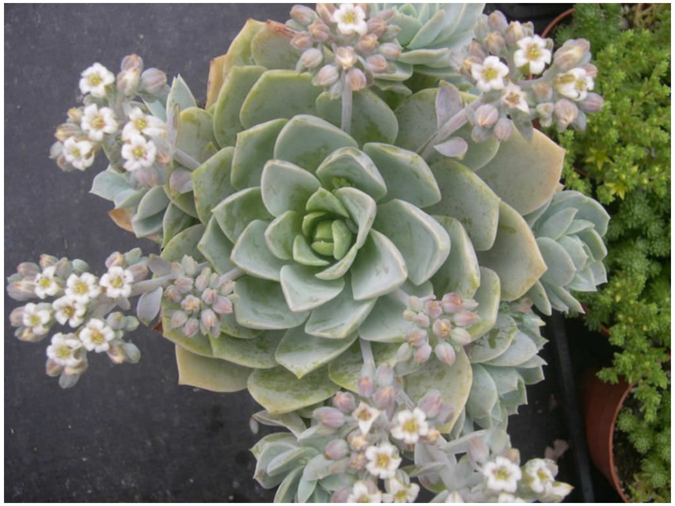


If chromosome arms pair at random, bivalents occur when there is pairing in both arms of two chromosomes instead of a switch in partners between one of these chromosomes and a third chromosome. Studies on induced autotetraploids in different plant species indicate that there are more bivalents and fewer quadrivalents at metaphase I than would be expected if partner exchange occurs without restrictions. There also is evidence that the proportion of bivalents to multivalents increases over a span of generations. The suggested explanations for the reduction in quadrivalents focus on the pairing mechanism during the zygotene and pachytene stages of meiosis and the locations of chiasmata.

Apples et al. (1998)

*random pairing and no double reduction

Extremely high poly level examples

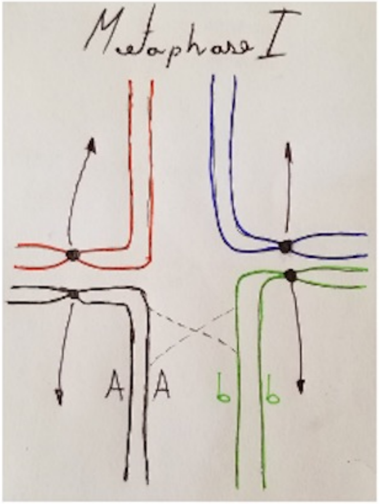


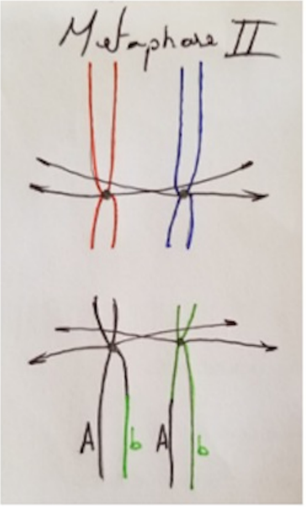
$2n = 640 \approx 80x$
Sedum suaveolens
Crassulaceae, Stonecrop

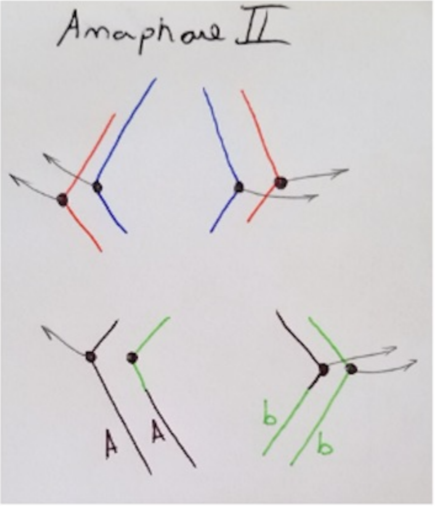
$2n = 1260 \approx 84x$
Ophioglossum pycnostichum
Ophioglossaceae , Fern

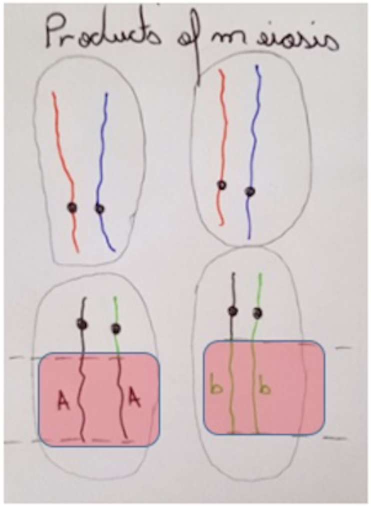
Tetravalent formation and Double Reduction

- Double reduction is a miotic phenomenon where the sister chromatids migrate to the same gamete due to a combination of multivalent pairing and crossing over occurring between the centromere and the locus
- Thus, it is a position dependent phenomenon



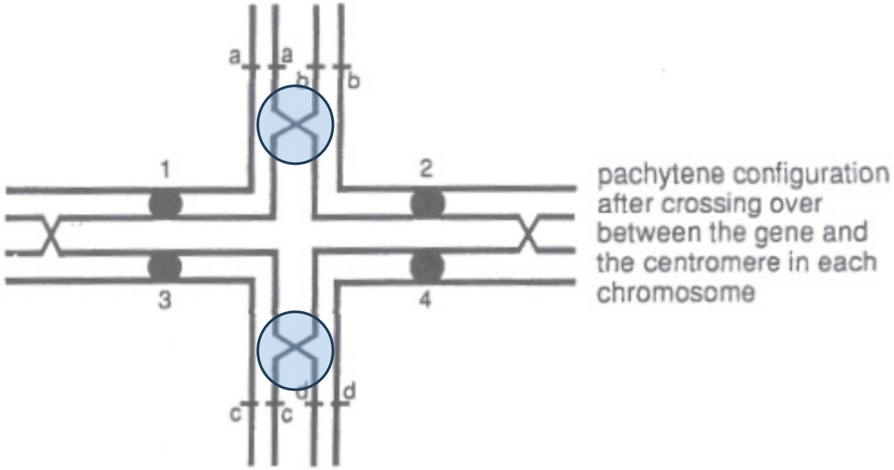






- In diploids, under stable meiotic conditions, duplicated chromatids will not appear in the same gamete.

Gametic frequencies under Maximum Equational Segregation



Maximum Equational Segregation: Crossover between gene and centromere in each chromosome

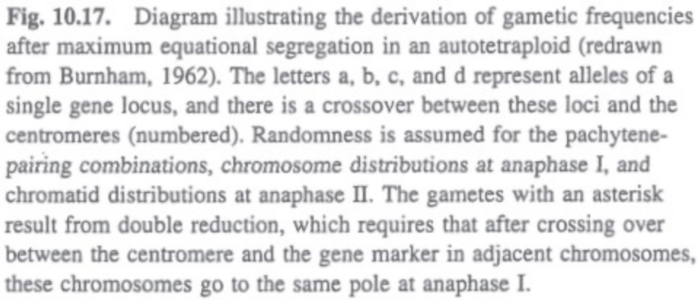
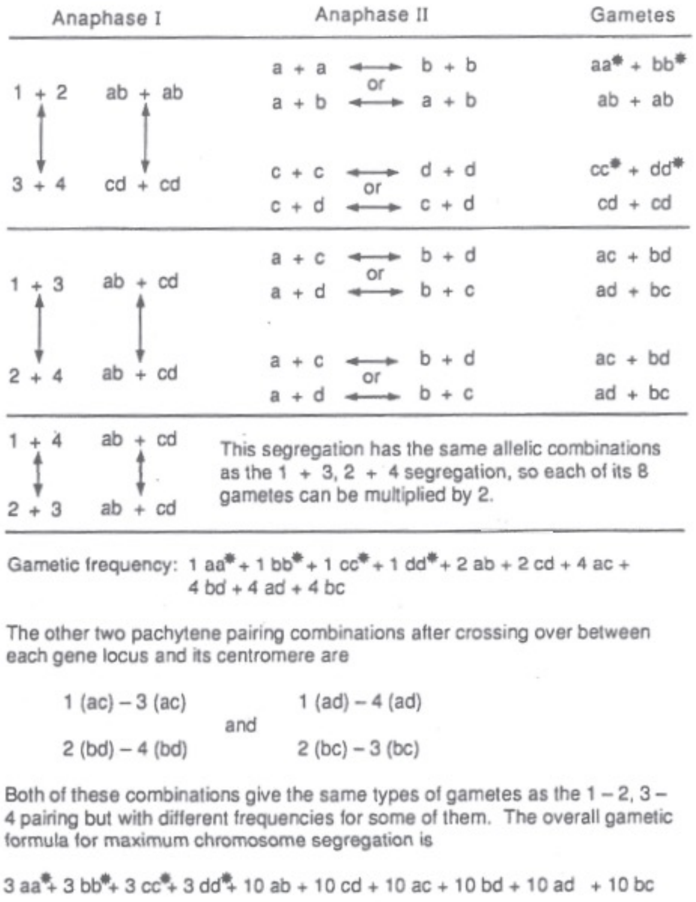


Fig. 10.17. Diagram illustrating the derivation of gametic frequencies after maximum equational segregation in an autotetraploid (redrawn from Burnham, 1962). The letters a, b, c, and d represent alleles of a single gene locus, and there is a crossover between these loci and the centromeres (numbered). Randomness is assumed for the pachytene-pairing combinations, chromosome distributions at anaphase I, and chromatid distributions at anaphase II. The gametes with an asterisk result from double reduction, which requires that after crossing over between the centromere and the gene marker in adjacent chromosomes, these chromosomes go to the same pole at anaphase I.



Anaphase I		Anaphase II	Gametes
1 + 2	ab + ab	a + a ⟷ b + b or a + b ⟷ a + b	aa* + bb* ab + ab
↕	↕		
3 + 4	cd + cd	c + c ⟷ d + d or c + d ⟷ c + d	cc* + dd* cd + cd
1 + 3	ab + cd	a + c ⟷ b + d or a + d ⟷ b + c	ac + bd ad + bc
↕	↕		
2 + 4	ab + cd	a + c ⟷ b + d or a + d ⟷ b + c	ac + bd ad + bc
1 + 4	ab + cd	This segregation has the same allelic combinations as the 1 + 3, 2 + 4 segregation, so each of its 8 gametes can be multiplied by 2.	
↕	↕		
2 + 3	ab + cd		

Gametic frequency: 1 aa* + 1 bb* + 1 cc* + 1 dd* + 2 ab + 2 cd + 4 ac + 4 bd + 4 ad + 4 bc

The other two pachytene pairing combinations after crossing over between each gene locus and its centromere are

1 (ac) – 3 (ac)
2 (bd) – 4 (bd)

and

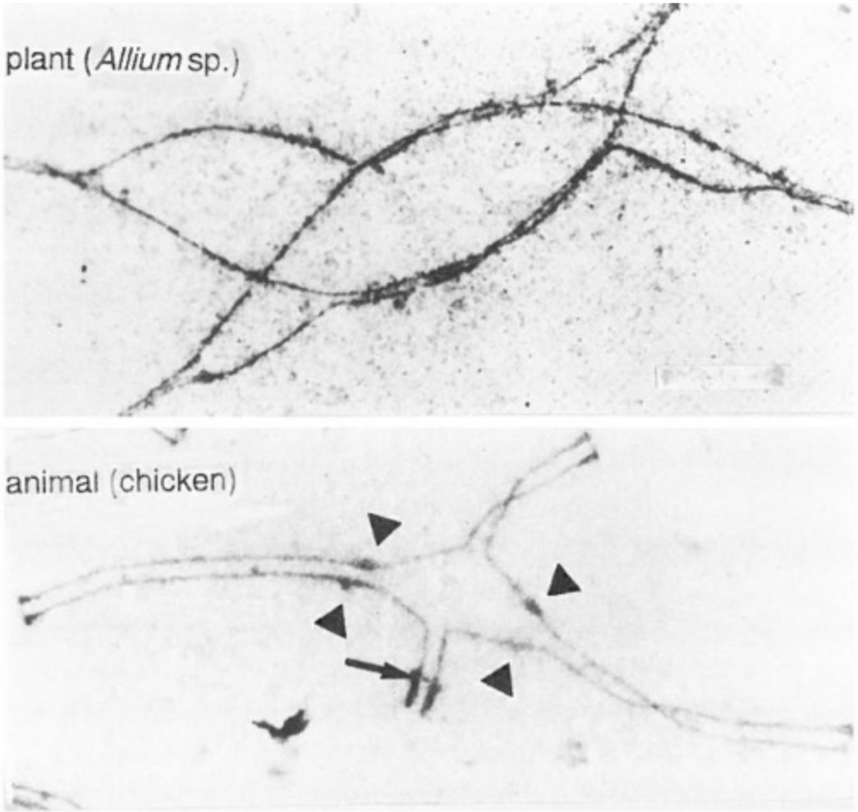
1 (ad) – 4 (ad)
2 (bc) – 3 (bc)

Both of these combinations give the same types of gametes as the 1 – 2, 3 – 4 pairing but with different frequencies for some of them. The overall gametic formula for maximum chromosome segregation is

3 aa* + 3 bb* + 3 cc* + 3 dd* + 10 ab + 10 cd + 10 ac + 10 bd + 10 ad + 10 bc

Appels et al. (1998) - Chromosome Biology

Multivalent Pairing



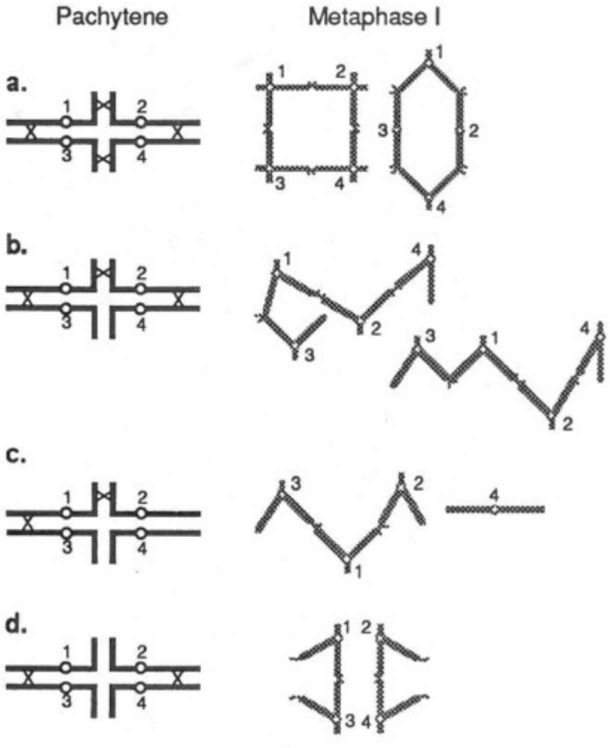
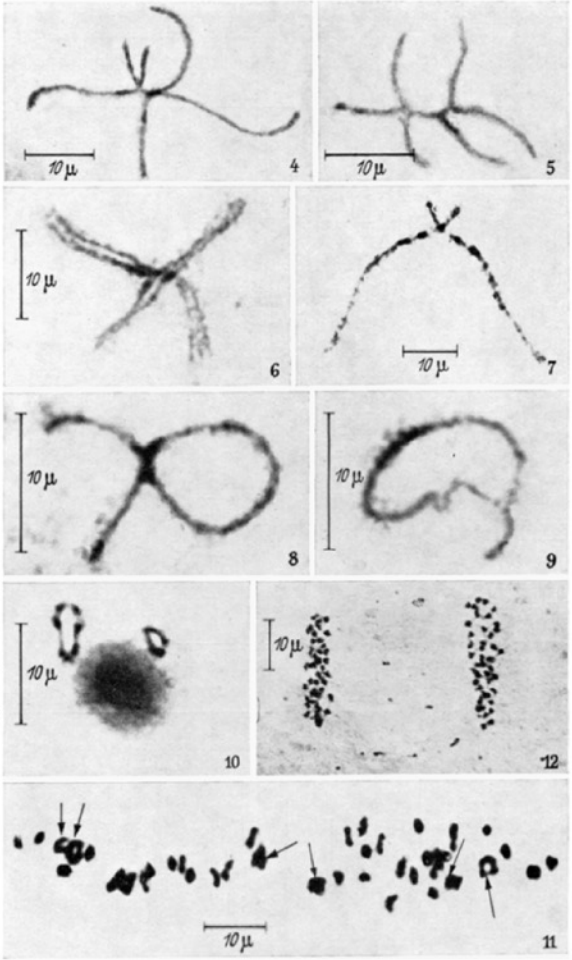


Fig. 10.13. Diagrams of quadrivalents (IVs) at pachytene in an autotetraploid (left) and some of the metaphase I configurations resulting from varying numbers of crossovers (right). Each chromosome is

Meiotic Pairing in polyploids

Sweetpotato



Hexavalents, quadrivalents and bivalents in sweetpotato (Magoon *et al.* 1970)

Sugarcane

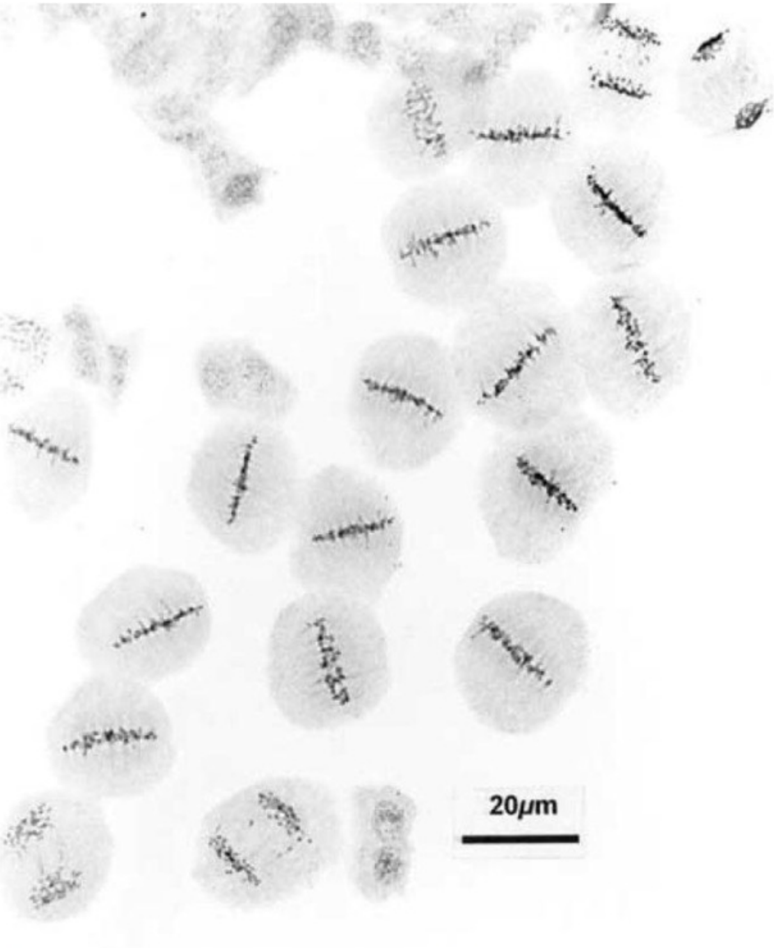
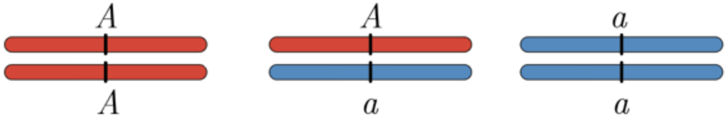
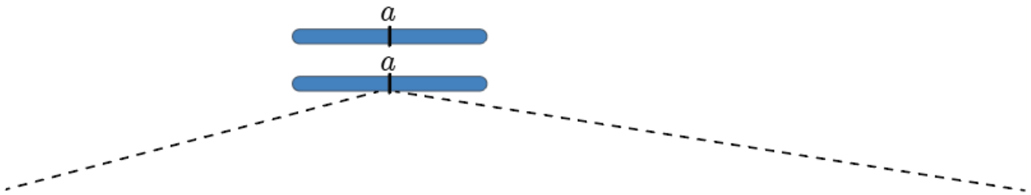


Figure 1. PMCs of *Saccharum* spp. hybrid clone 79N9059 at meiosis. As was the case in other clones, pairing was regular, bivalents generally formed. (Bielig *et al.* 2003)

Assessing allelic variation in diploids

Three genotypes

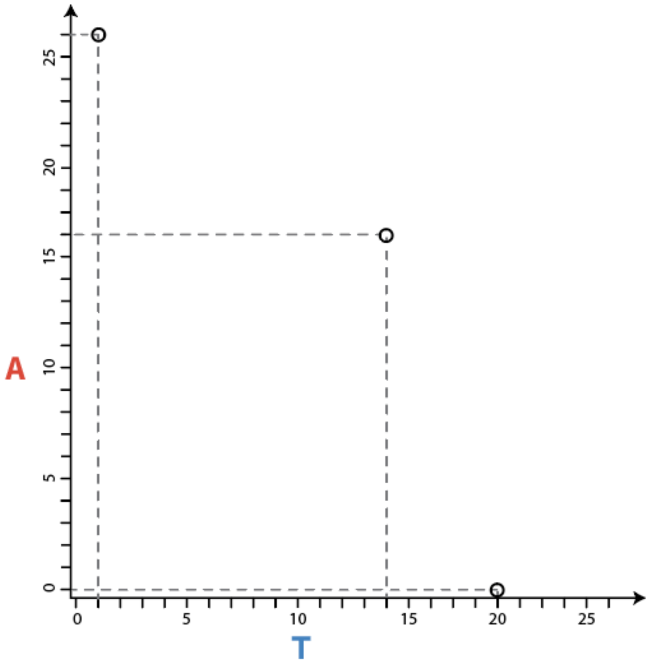




ATTGGTGTATGCGTGCGTAGACGTAGTGGCGTTGTCGAGCTGATATGTCG
ATTGGTGTATGCGTGCGTAGACGTAGTGGCGTTGTCGAGCTGATATGTCG

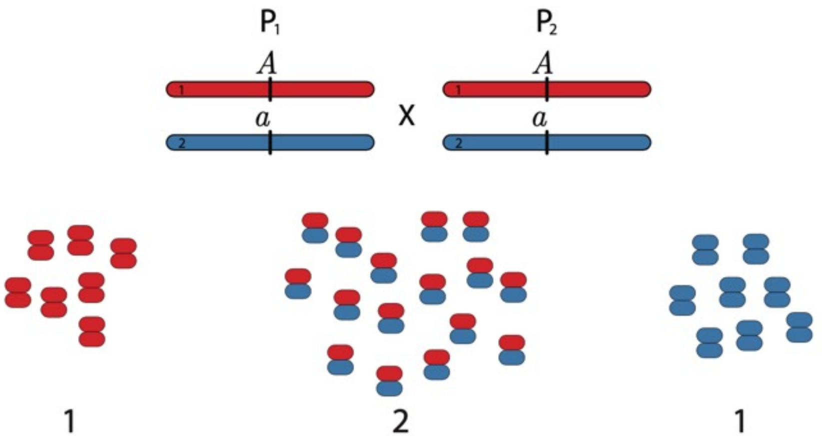
```
TATGCGTGCGTAGCGTAGTGGCGATTGTCGAGCTGATATGTCGTGATGTG
  TGCGTGCGTAGCGTAGTGGCGATTGTCGAGCTGATATGTCGTGATGTGTC
  TGCGTGCGTAGCGTAGTGGCGATTGTCGAGCTGATATGTCGTGATGTGTC
  TGCGTGCGTAGCGTAGTGGCGATTGTCGAGCTGATATGTCGTGATGTGTC
  TGCGTGCGTAGCGTAGTGGCGATTGTCGAGCTGATATGTCGTGATGTGTC
   GCGTGCGTAGCGTAGTGGCGATTGTCGAGCTGATATGTCGTGATGTGTCT
   GCGTGCGTAGCGTAGTGGCGATTGTCGAGCTGATATGTCGTGATGTGTCT
   GCGTGCGTAGCGTAGTGGCGATTGTCGAGCTGATATGTCGTGATGTGTCT
     GTGCGTAGCGTAGTGGCGATTGTCGAGCTGATATGTCGTGATGTGTCTGA
     GTGCGTAGCGTAGTGGCGATTGTCGAGCTGATATGTCGTGATGTGTCTGA
     GTGCGTAGCGTAGTGGCGATTGTCGAGCTGATATGTCGTGATGTGTCTGA
     GTGCGTAGCGTAGTGGCGATTGTCGAGCTGATATGTCGTGATGTGTCTGA
      TGCGTAGCGTAGTGGCGATTGTCGAGCTGATATGTCGTGATGTGTCTGAT
      TGCGTAGCGTAGTGGCGATTGTCGAGCTGATATGTCGTGATGTGTCTGAT
      TGCGTAGCGTAGTGGCGATTGTCGAGCTGATATGTCGTGATGTGTCTGAT
      TGCGTAGCGTAGTGGCGATTGTCGAGCTGATATGTCGTGATGTGTCTGAT
      TGCGTAGCGTAGTGGCGATTGTCGAGCTGATATGTCGTGATGTGTCTGAT
        CGTAGCGTAGTGGCGATTGTCGAGCTGATATGTCGTGATGTGTCTGATGC
        CGTAGCGTAGTGGCGATTGTCGAGCTGATATGTCGTGATGTGTCTGATGC
        CGTAGCGTAGTGGCGATTGTCGAGCTGATATGTCGTGATGTGTCTGATGC
```

20 reads



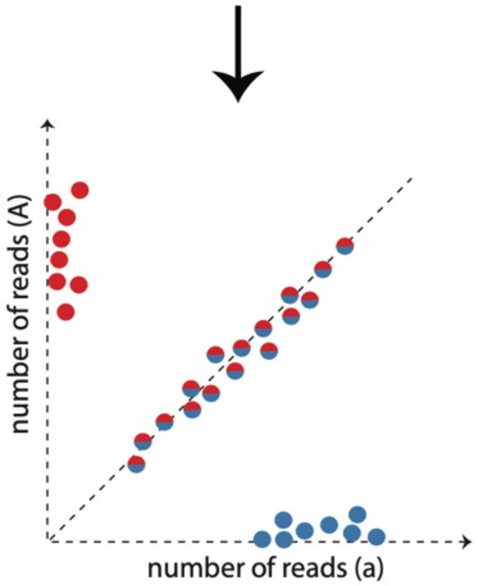
cc

Assessing allelic variation in diploids



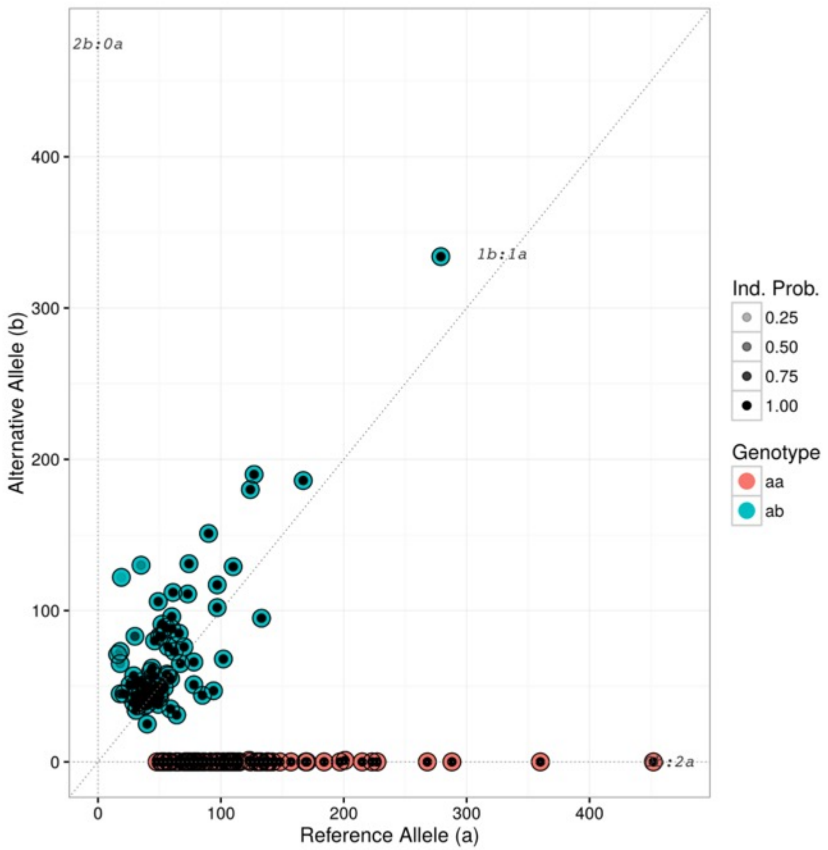
GBS

Read counts for each one of the allelic variants



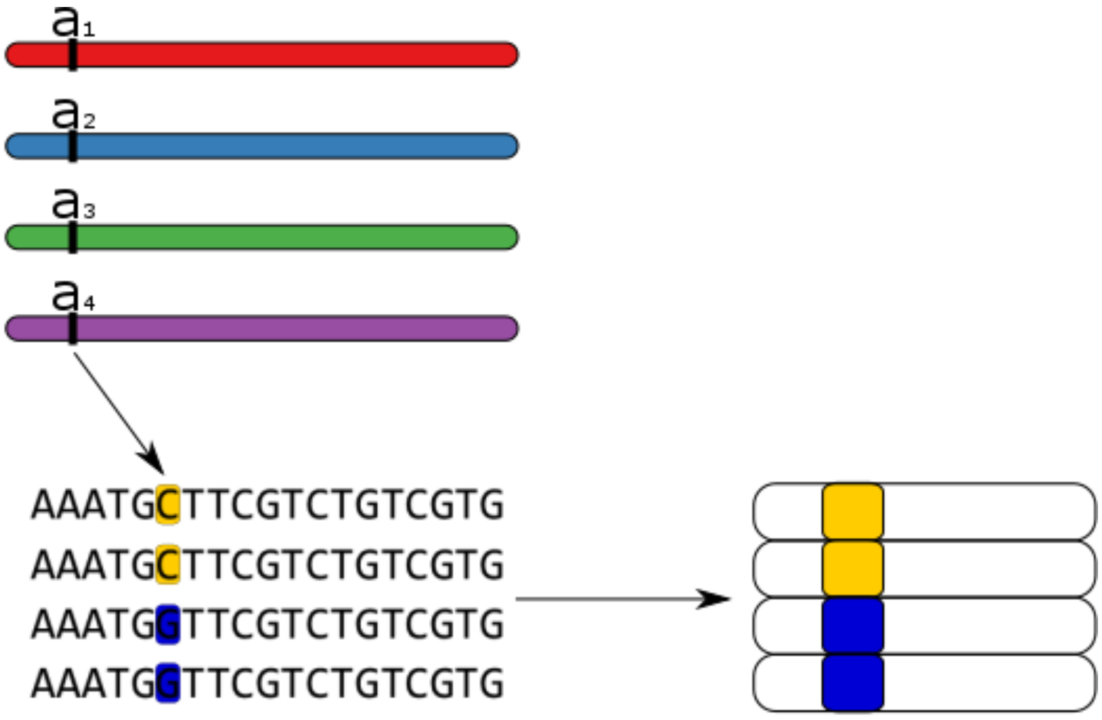
Diploid sweetpotato

Ipomea trifida



https://doi.org/10.1371/journal.pone.0030906

Assessing allelic variation in polyploids



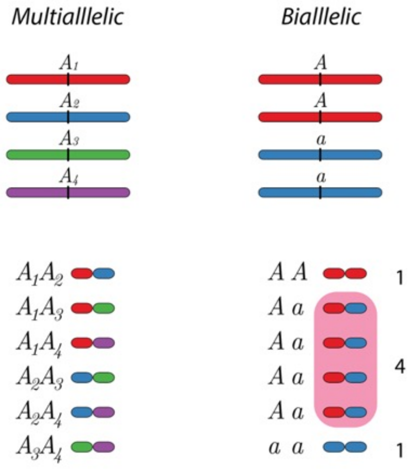
AAATGCTTCGTCTGTCGTG
AAATGCTTCGTCTGTCGTG
AAATGGTTCGTCTGTCGTG
AAATGGTTCGTCTGTCGTG

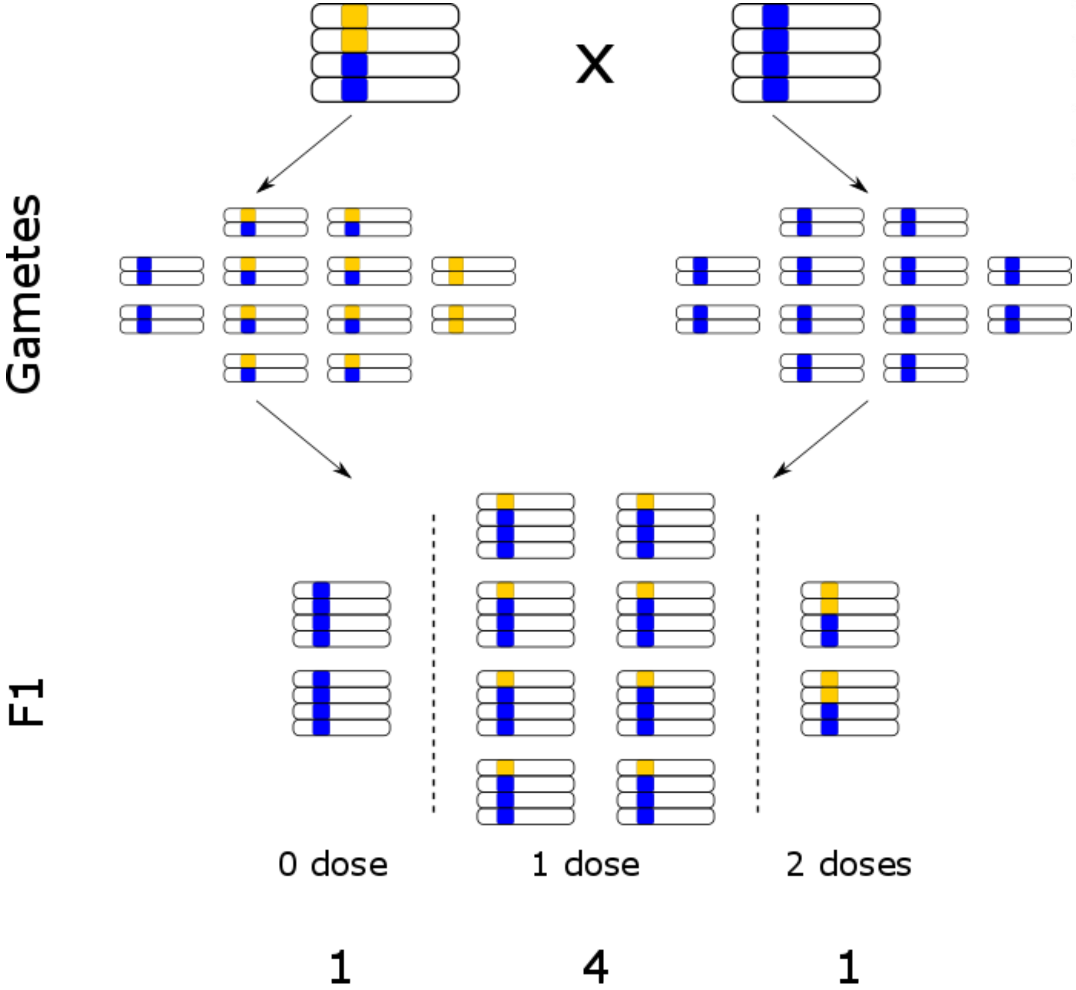
$a_1 = a_2 \neq a_3 = a_4$

2 dose marker

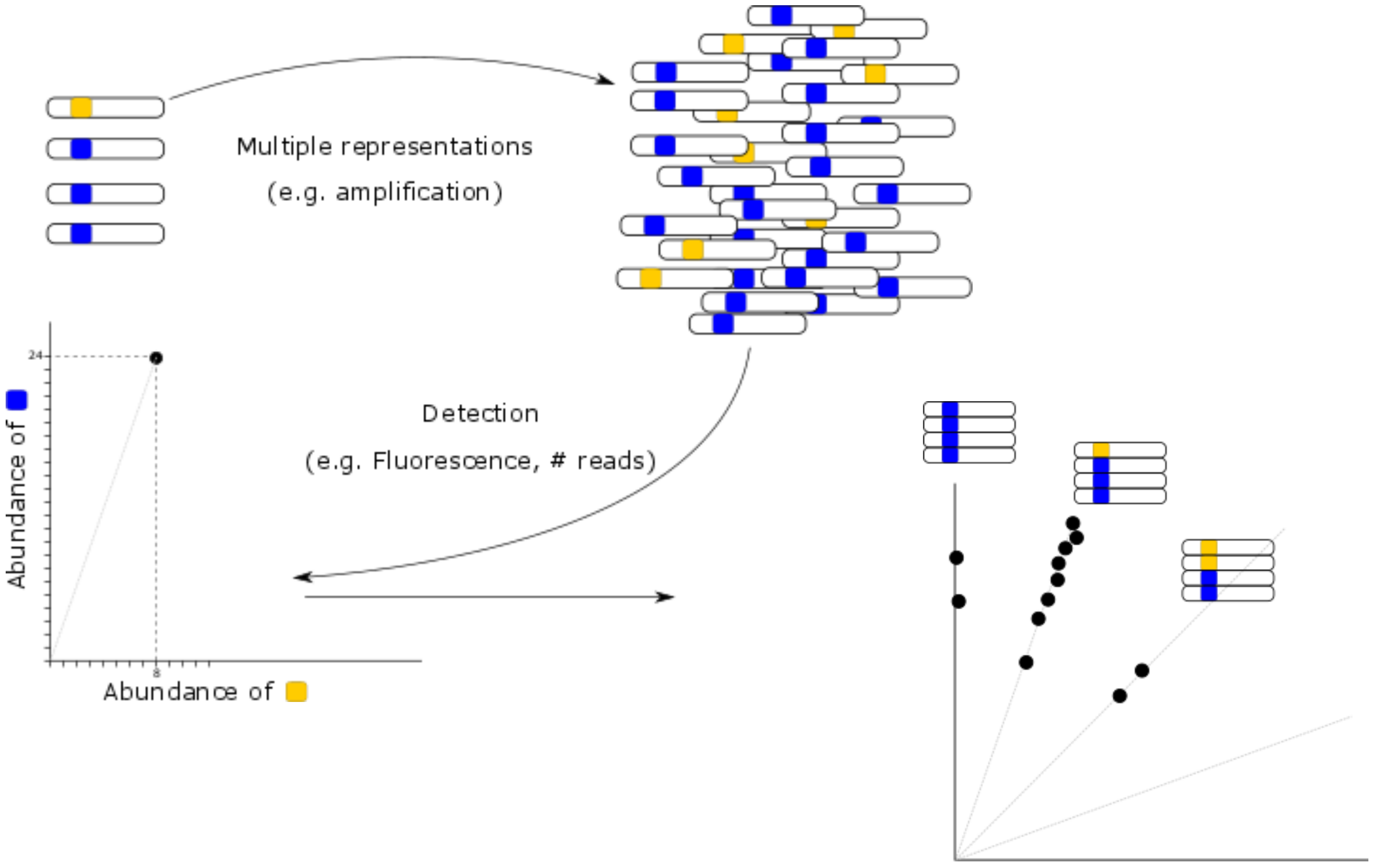
Genotype calling at SNP level

- Under polysomic inheritance



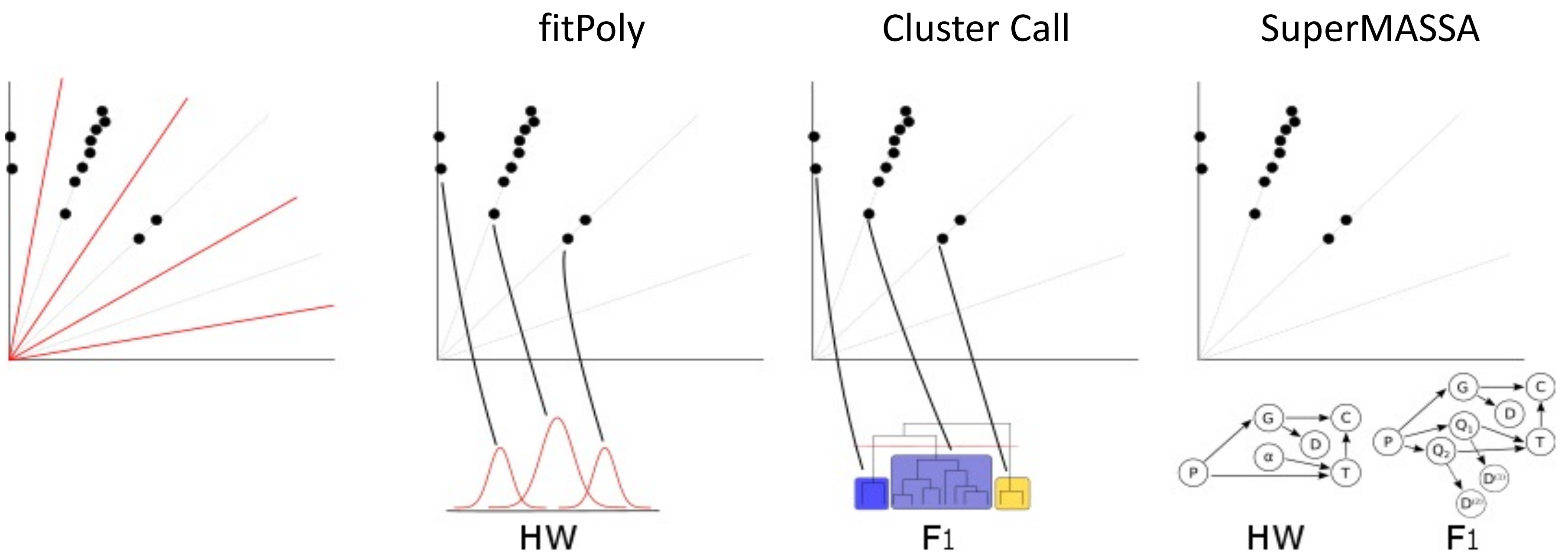


Genotype calling at SNP level



All SNP genotype calling methods are based on the relative position of ratios

Genotype calling at SNP level



Ad hoc
Thresholds

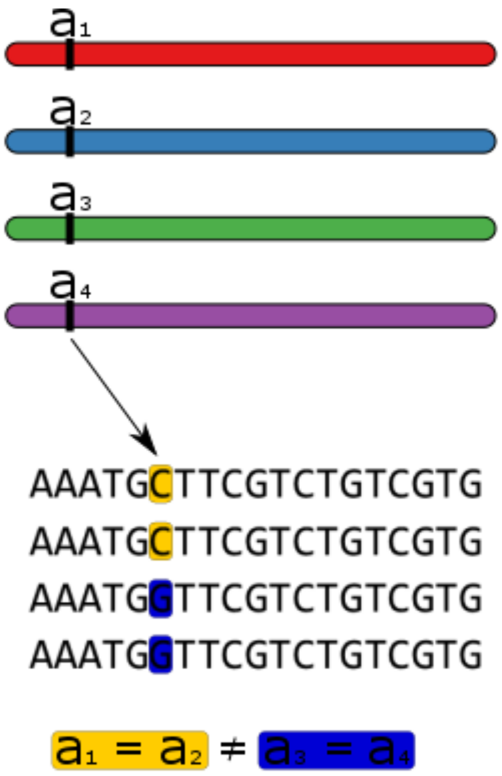
Mixture models
Population structure:
HW
Autotetraploids

Hierarchical clustering
Population structure:
F1
Autotetraploids

Maximum *a posteriori*
Population structure:
F1 and HW
Autopolyploids
Drawback: very sensitive regarding skewed data

Updog is aligned with SuperMASSA, with more features

Genotype calling at SNP level

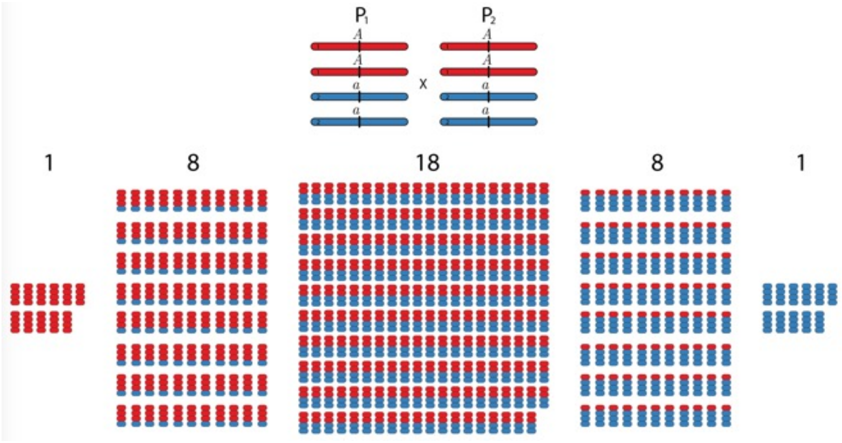


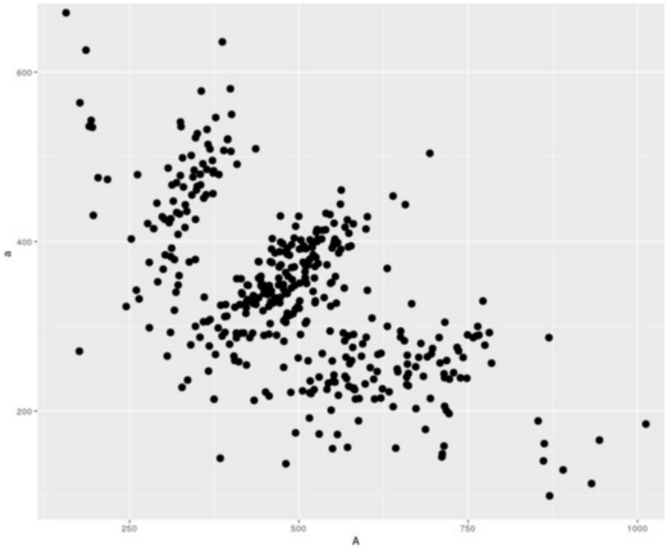
2 dose marker

- The information is assessed at SNP level (essentially biallelic).
- It is not possible to distinguish all four homologous chromosomes.
- Is this information enough for genetic mapping and QTL analysis?

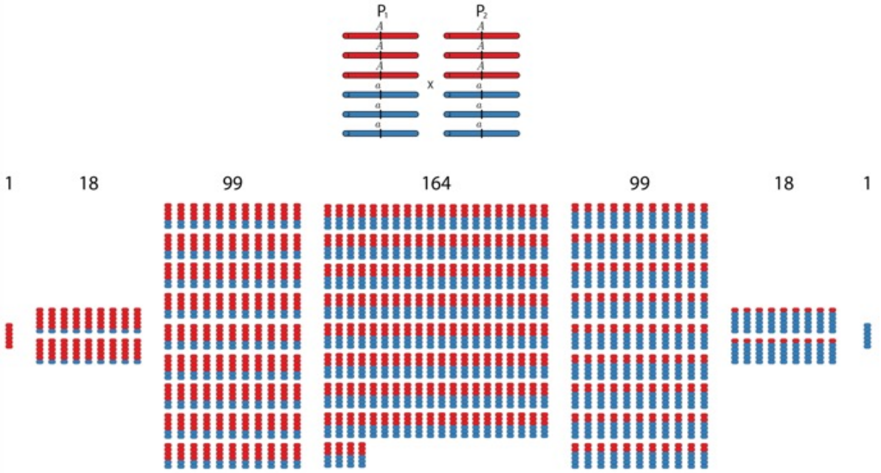
Genotype calling in polyploids

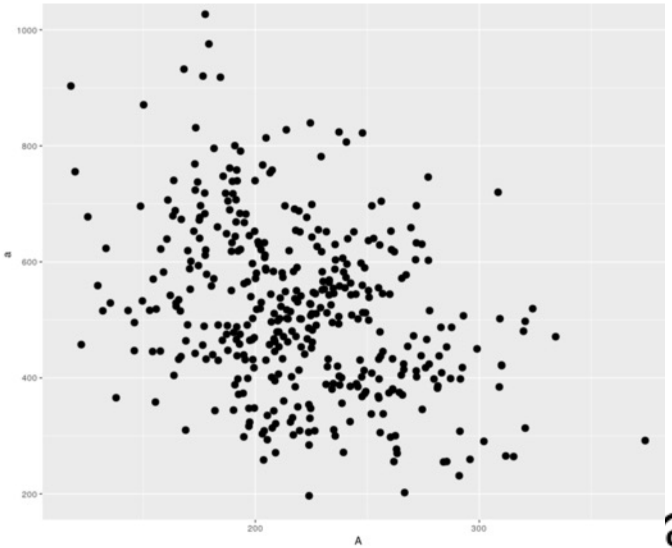
Tetraploid





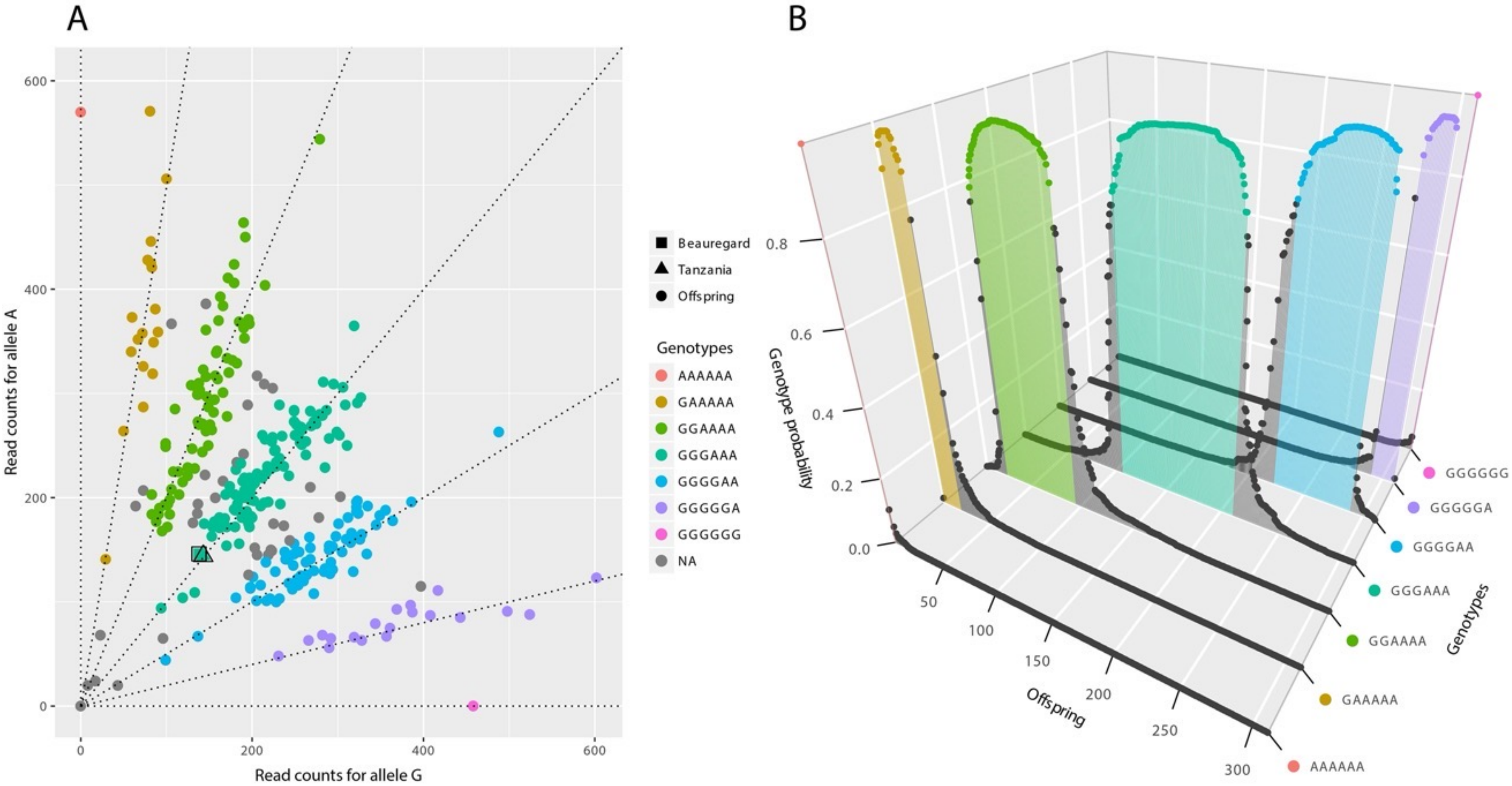
Hexaploid





Genotyping Calling using SuperMASSA

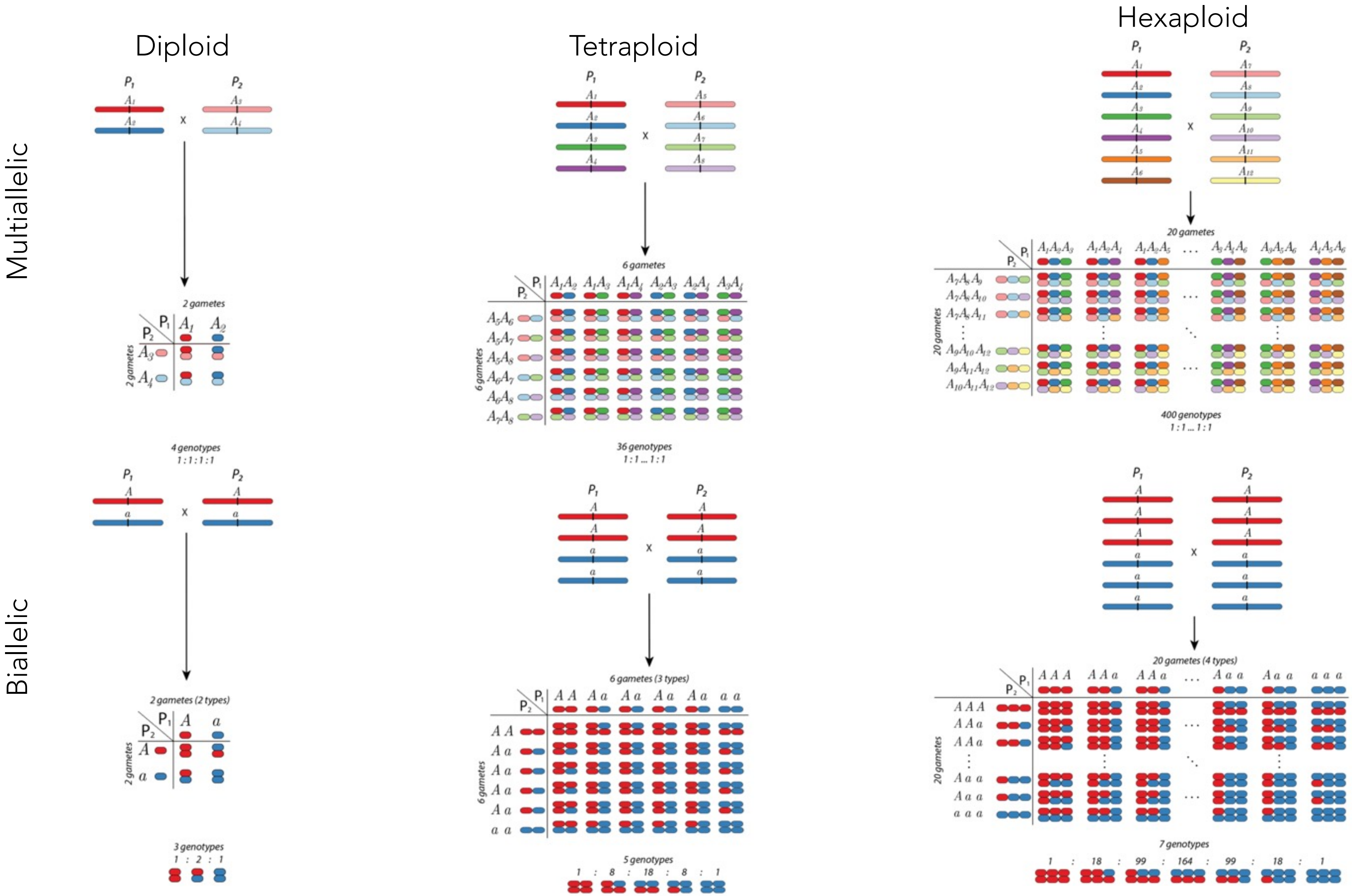
Dosage calling including the probability distribution of the genotypes



Beauregard: 3 doses
Tanzania: 3 doses

Segregation in polyploids*

```
> segreg_poly(6,3,3)
> MASS::fractions(segreg_poly(4,2,1))
```



*random pairing and no double reduction

Chi-Square Test in Genetic Segregation

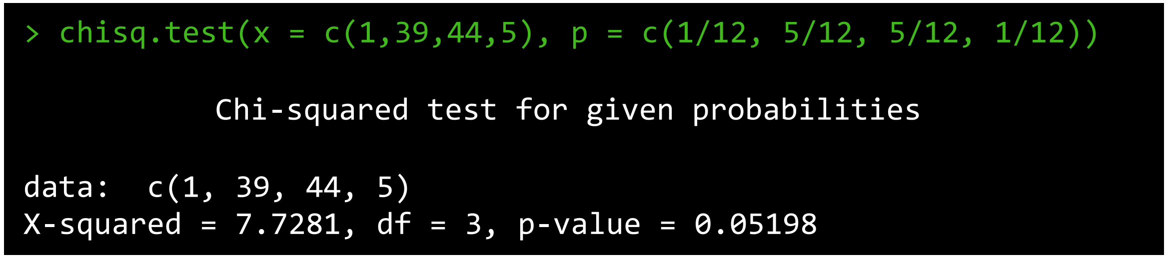
- Determines if observed genetic ratios match expected ratios, helping confirm inheritance patterns.

$$\chi^2 = \Sigma \frac{(O - E)^2}{E}$$

Where:

- O = Observed frequency
- E = Expected frequency

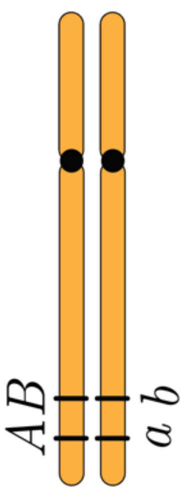
- **H0 (Null Hypothesis):** The observed genetic frequencies are consistent with the expected frequencies
- **Ha (Alternative Hypothesis):** The observed genetic frequencies are not consistent with the expected



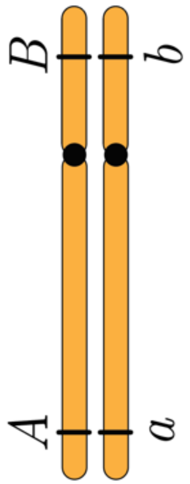
```
> chisq.test(x = c(1,39,44,5), p = c(1/12, 5/12, 5/12, 1/12))

	Chi-squared test for given probabilities

data:  c(1, 39, 44, 5)
X-squared = 7.7281, df = 3, p-value = 0.05198
```

Linkage

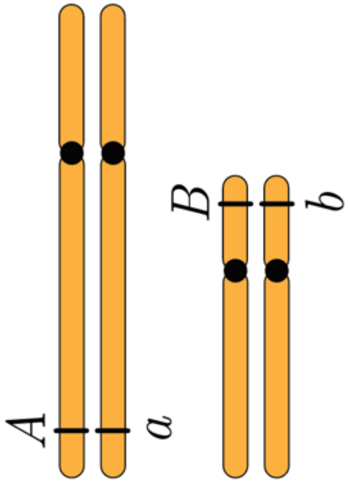
- Genetic linkage is the phenomenon where markers are likely to be inherited together.
- The closer the markers are, the lower the probability of crossing over events occur between them; consequently, the more likely they will be co-inherited.



Linked



Not linked



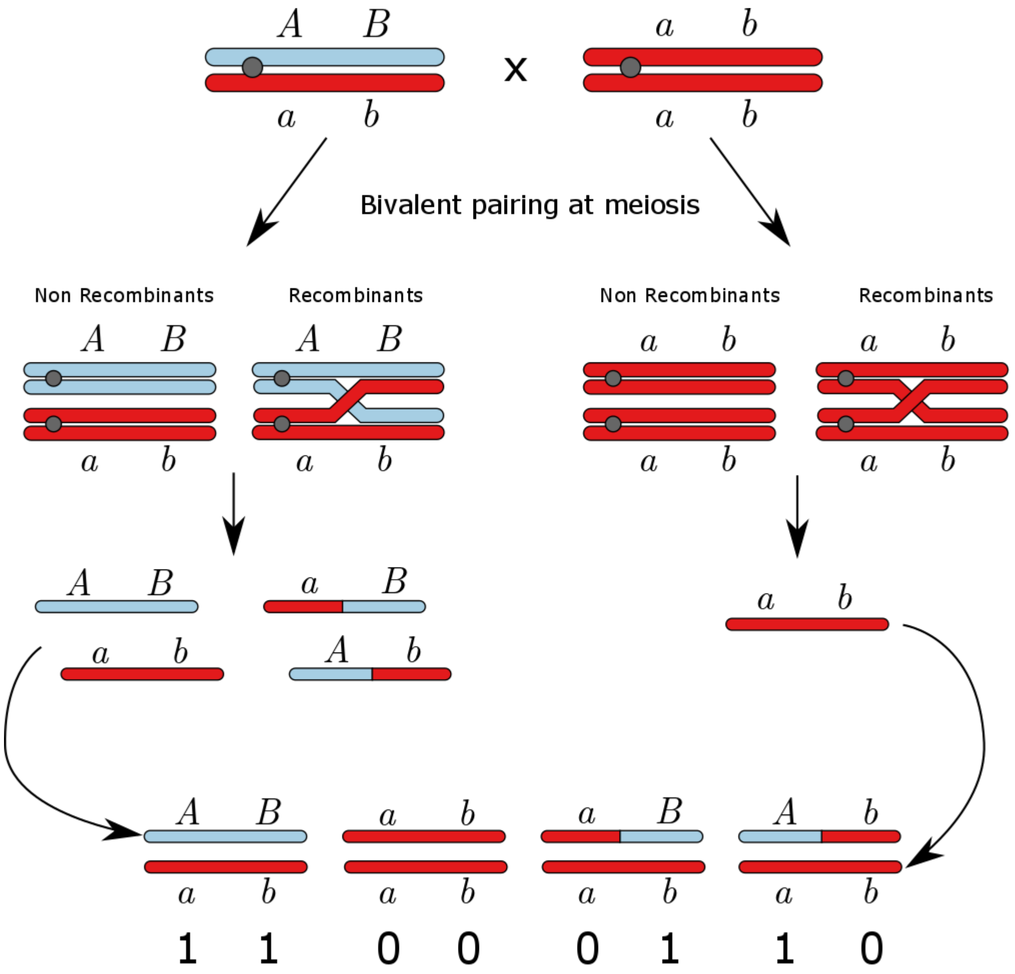
Not linked

- How can we measure how likely A and B are co-inherited?

Linkage analysis

- Genetic linkage is a concept applied to *at least two loci*.
- We measure linkage using the *recombination frequency* (or fraction) in a segregation population.
- Recombination frequency is the *probability* that an odd number of crossovers occurs between the markers. Ranges from 0.0 to 0.5 (considering double reduction this number can be higher)
- We can transform these probabilistic values into distances using various *mapping functions*, such as those developed by Morgan, Haldane, Kosambi, etc.
- By computing the recombination frequencies between pairs of markers and using mapping functions, we can construct *linkage maps* which show the linear order and relative distance between adjacent markers.

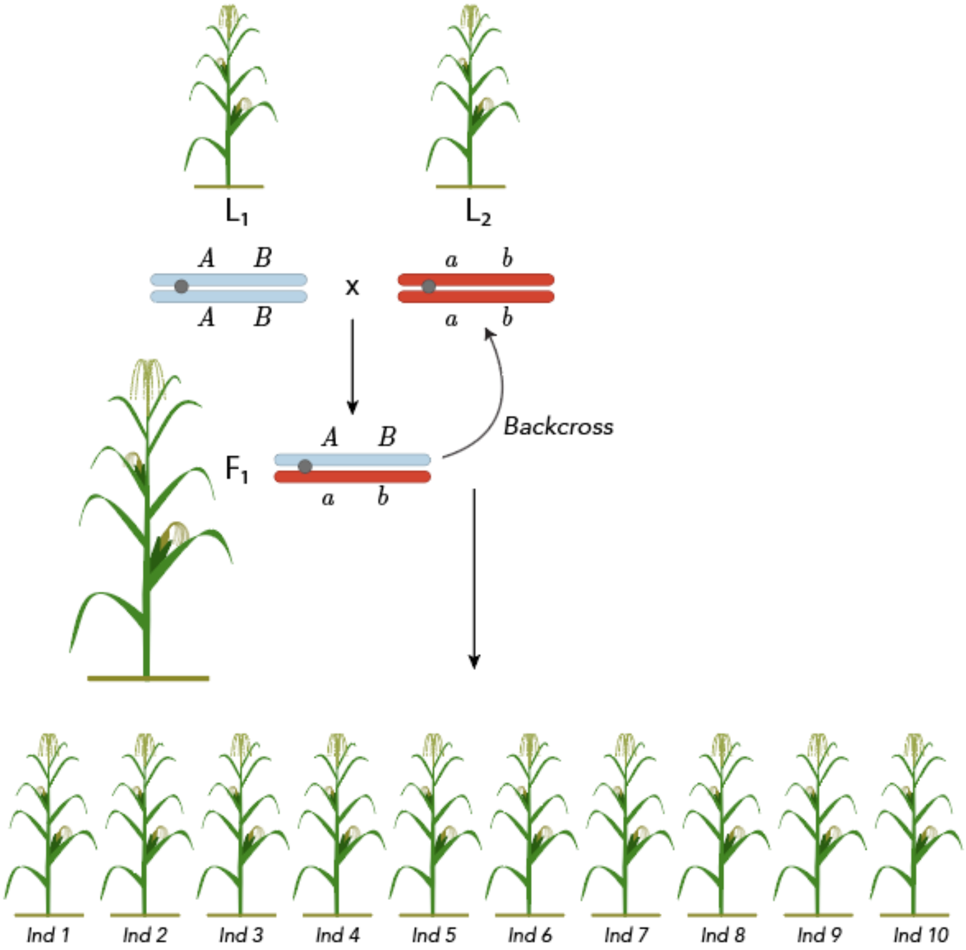
Recombination Fraction

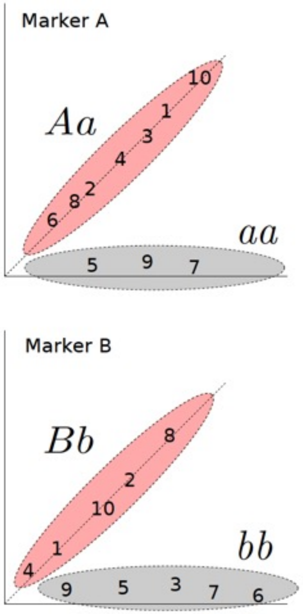


$$\hat{r} = \frac{\#\text{recombinants}}{\#\text{total}} = \frac{\#10 + \#01}{\#11{+}\#00{+}\#10{+}\#01}$$

Recombination fraction in diploids

Ten backcross individuals genotyped with two markers: A and B.

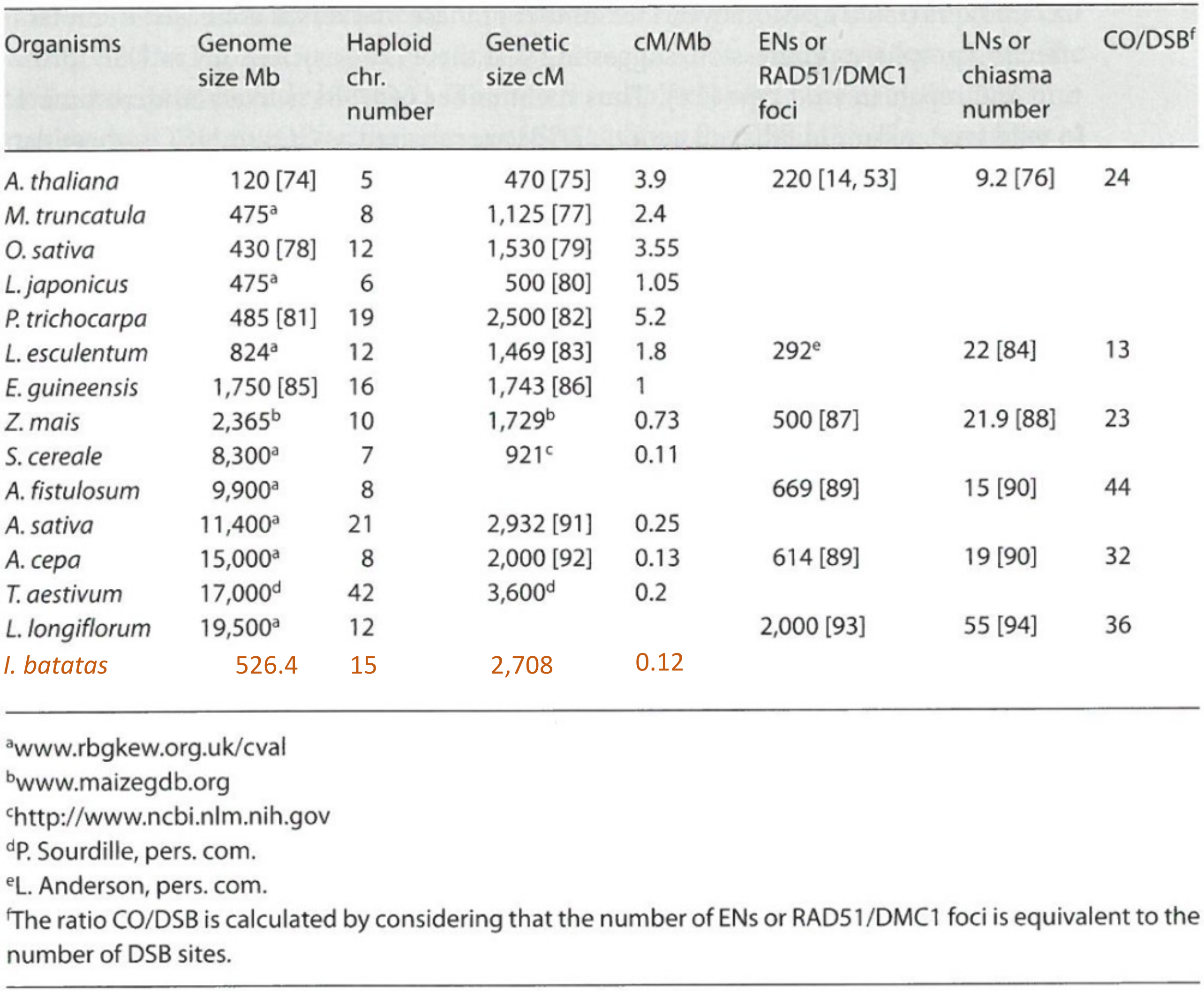




Individual	Obs. Gen.
1	(A, B)
2	(A, B)
3	(A, b)
4	(A, B)
5	(a, b)
6	(A, b)
7	(a, b)
8	(A, B)
9	(a, b)
10	(A, B)

$$\hat{r} = 2/10 = 0.2$$

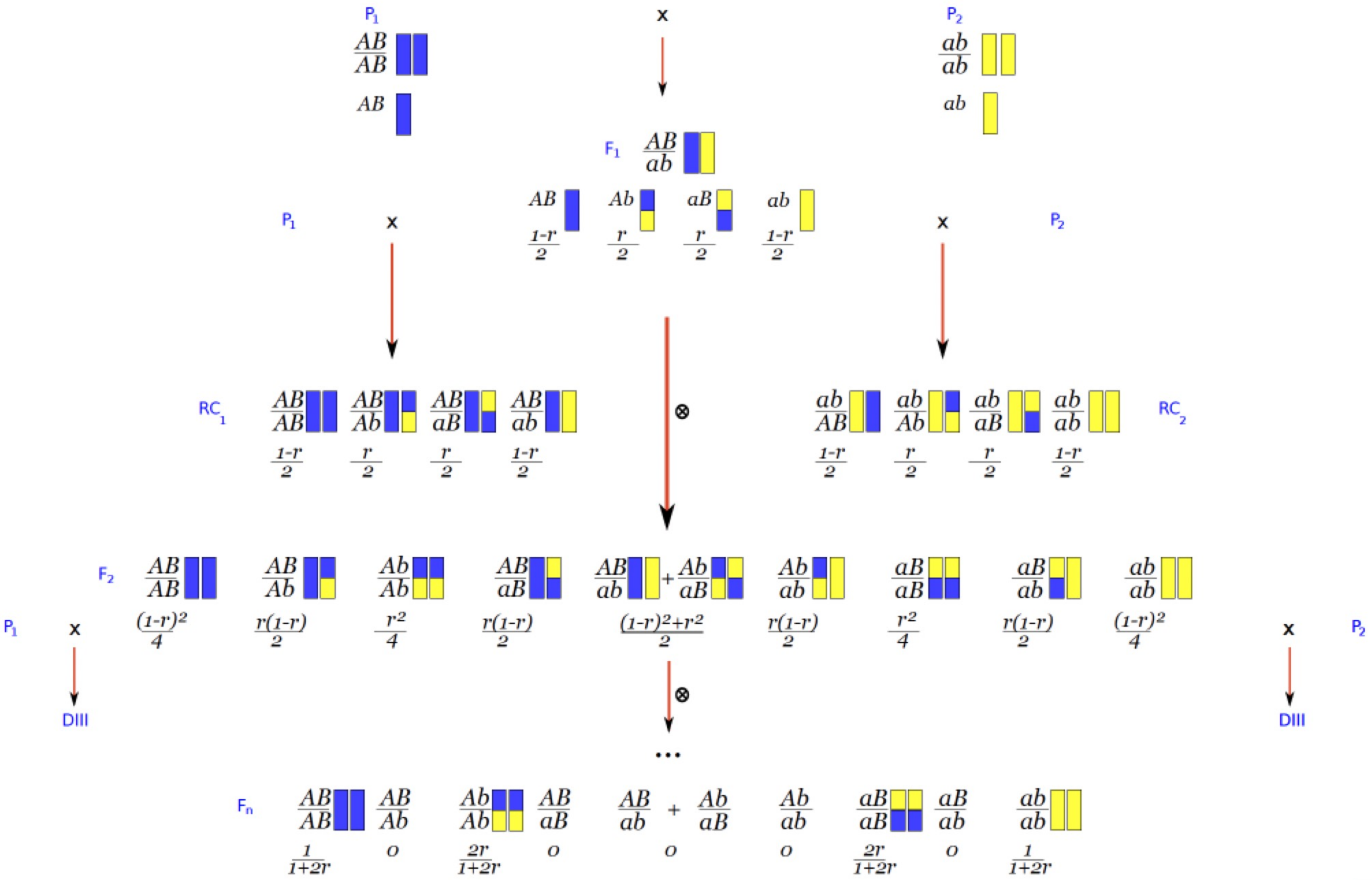
An overview of plant recombination data



	Organisms	Genome size Mb	Haploid chr. number	Genetic size cM	cM/Mb	ENs or RAD51/DMC1 foci	LNs or chiasma number	CO/DSB[f]
Arabidopsis	*A. thaliana*	120 [74]	5	470 [75]	3.9	220 [14, 53]	9.2 [76]	24
Medicago	*M. truncatula*	475[a]	8	1,125 [77]	2.4			
Lotus	*O. sativa*	430 [78]	12	1,530 [79]	3.55			
Rice	*L. japonicus*	475[a]	6	500 [80]	1.05			
Populus	*P. trichocarpa*	485 [81]	19	2,500 [82]	5.2			
Tomato	*L. esculentum*	824[a]	12	1,469 [83]	1.8	292[e]	22 [84]	13
Palm	*E. guineensis*	1,750 [85]	16	1,743 [86]	1			
Maize	*Z. mais*	2,365[b]	10	1,729[b]	0.73	500 [87]	21.9 [88]	23
Rye	*S. cereale*	8,300[a]	7	921[c]	0.11			
Green Onion	*A. fistulosum*	9,900[a]	8			669 [89]	15 [90]	44
Garlic	*A. sativa*	11,400[a]	21	2,932 [91]	0.25			
Onion	*A. cepa*	15,000[a]	8	2,000 [92]	0.13	614 [89]	19 [90]	32
Wheat	*T. aestivum*	17,000[d]	42	3,600[d]	0.2			
Lily	*L. longiflorum*	19,500[a]	12			2,000 [93]	55 [94]	36
Sweetpotato	*I. batatas*	526.4	15	2,708	0.12			

[a] www.rbgkew.org.uk/cval
[b] www.maizegdb.org
[c] http://www.ncbi.nlm.nih.gov
[d] P. Sourdille, pers. com.
[e] L. Anderson, pers. com.
[f] The ratio CO/DSB is calculated by considering that the number of ENs or RAD51/DMC1 foci is equivalent to the number of DSB sites.

Genetical design based on inbred lines



P_1 $\frac{AB}{AB}$ x P_2 $\frac{ab}{ab}$

AB ab

F_1 $\frac{AB}{ab}$

AB $\frac{1-r}{2}$ Ab $\frac{r}{2}$ aB $\frac{r}{2}$ ab $\frac{1-r}{2}$

P_1 x x P_2

RC_1 $\frac{AB}{AB}$ $\frac{1-r}{2}$ $\frac{AB}{Ab}$ $\frac{r}{2}$ $\frac{AB}{aB}$ $\frac{r}{2}$ $\frac{AB}{ab}$ $\frac{1-r}{2}$

⊗

RC_2 $\frac{ab}{AB}$ $\frac{1-r}{2}$ $\frac{ab}{Ab}$ $\frac{r}{2}$ $\frac{ab}{aB}$ $\frac{r}{2}$ $\frac{ab}{ab}$ $\frac{1-r}{2}$

F_2 $\frac{AB}{AB}$ $\frac{(1-r)^2}{4}$ $\frac{AB}{Ab}$ $\frac{r(1-r)}{2}$ $\frac{Ab}{Ab}$ $\frac{r^2}{4}$ $\frac{AB}{aB}$ $\frac{r(1-r)}{2}$ $\frac{AB}{ab}+\frac{Ab}{aB}$ $\frac{(1-r)^2+r^2}{2}$ $\frac{Ab}{ab}$ $\frac{r(1-r)}{2}$ $\frac{aB}{aB}$ $\frac{r^2}{4}$ $\frac{aB}{ab}$ $\frac{r(1-r)}{2}$ $\frac{ab}{ab}$ $\frac{(1-r)^2}{4}$

P_1 x DIII

x P_2 DIII

⊗

...

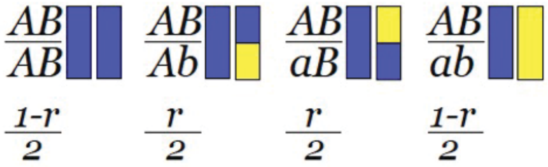
F_n $\frac{AB}{AB}$ $\frac{1}{1+2r}$ $\frac{AB}{Ab}$ 0 $\frac{Ab}{Ab}$ $\frac{2r}{1+2r}$ $\frac{AB}{aB}$ 0 $\frac{AB}{ab}+\frac{Ab}{aB}$ 0 $\frac{Ab}{ab}$ 0 $\frac{aB}{aB}$ $\frac{2r}{1+2r}$ $\frac{aB}{ab}$ 0 $\frac{ab}{ab}$ $\frac{1}{1+2r}$

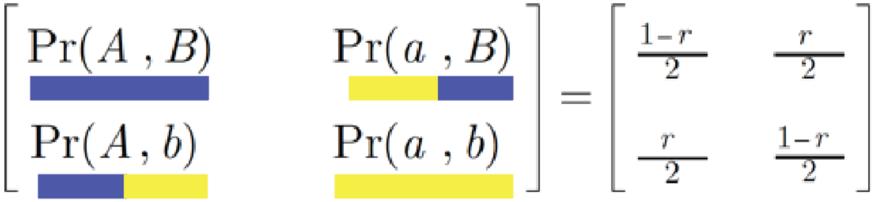
Slides from Augusto Garcia - https://github.com/augusto-garcia

BY NC SA

Recombination fraction - Likelihood

The likelihood function is a function of the parameters of a statistical model, given specific observed data. It represents how likely the observed data are, assuming a particular set of parameter values.





$$\begin{bmatrix} \Pr(A\,,B) & \Pr(a\,,B) \\ \Pr(A\,,b) & \Pr(a\,,b) \end{bmatrix} = \begin{bmatrix} \frac{1-r}{2} & \frac{r}{2} \\ \frac{r}{2} & \frac{1-r}{2} \end{bmatrix}$$

$$L(r) = \prod_{n} \Pr(G_A, G_B \mid r)$$

where n is the number of individuals. The maximum likelihood estimator of r is

$$\hat{r} = \underset{r}{\operatorname{argmax}}\, L(r)$$

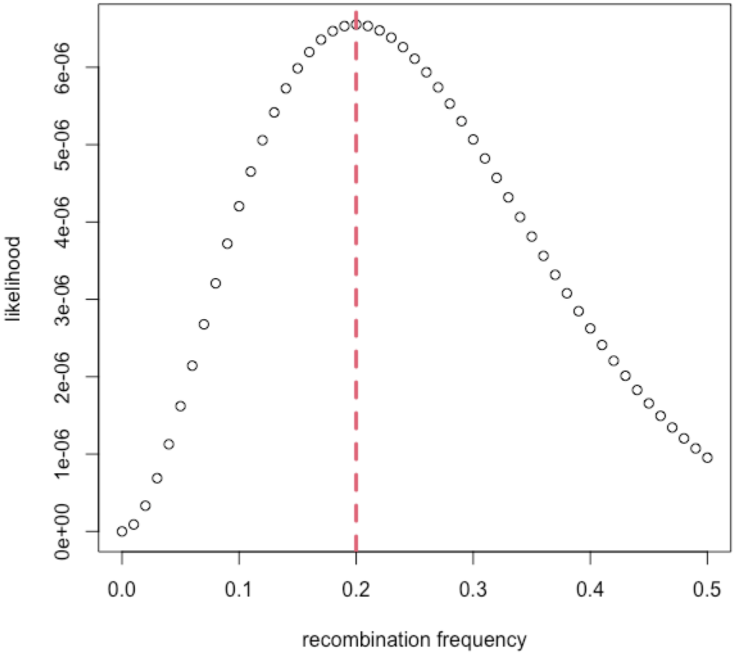
Recombination fraction in diploids

$$L = \prod_n \Pr(\text{loc}_B\ ,\ \text{loc}_A \mid Data)$$

Individual	Obs. Gen.	$\Pr(\text{loc}_B,\text{loc}_A)$
1	$(A\ ,B)$	$\frac{1}{2}(1-r)$
2	$(A\ ,B)$	$\frac{1}{2}(1-r)$
3	$(A\ ,b\ )$	$\frac{1}{2}\ (r)$
4	$(A\ ,B)$	$\frac{1}{2}(1-r)$
5	$(a\ \ ,b\ )$	$\frac{1}{2}(1-r)$
6	$(A\ ,b\ )$	$\frac{1}{2}\ (r)$
7	$(a\ \ ,b\ )$	$\frac{1}{2}(1-r)$
8	$(A\ ,B)$	$\frac{1}{2}(1-r)$
9	$(a\ \ ,b\ )$	$\frac{1}{2}(1-r)$
10	$(A\ ,B)$	$\frac{1}{2}(1-r)$

$$L = \left(\frac{r}{2}\right)^2\left(\frac{1-r}{2}\right)^8$$

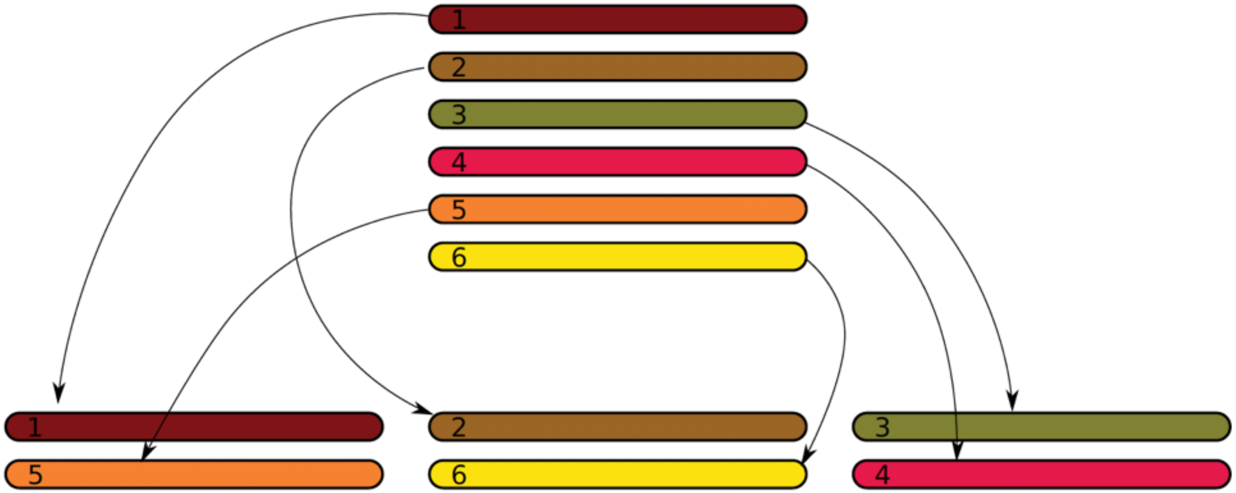
$$L = \prod_n \Pr(G_A, G_B) = \left(\frac{r}{2}\right)^2 \left(\frac{1-r}{2}\right)^8$$



The MLE (maximum likelihood estimate) of r is $\hat{r} = 0.2$

Computing recombination frequencies in diploids using R and C++
https://github.com/mmollina/Cpp_and_R

Gamete formation in polyploids*



$$\psi_4$$

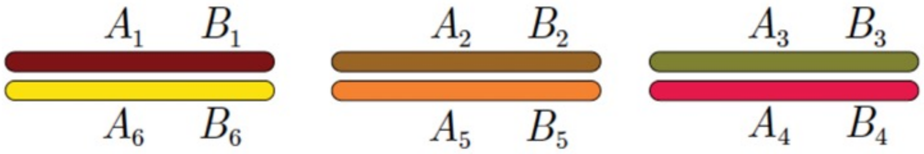
In this case: 15 possible configurations. For any ploidy level p

$$\frac{1}{\frac{p}{2}!}\prod_{i=1}^{\frac{p}{2}}\binom{2i}{2}$$

https://doi.org/10.1534/g3.119.400378

*no double reduction

Expected gametic frequency given a bivalent configuration



$$\begin{bmatrix} \frac{1-r}{2} & \frac{r}{2} \\ \frac{r}{2} & \frac{1-r}{2} \end{bmatrix} \otimes \begin{bmatrix} \frac{1-r}{2} & \frac{r}{2} \\ \frac{r}{2} & \frac{1-r}{2} \end{bmatrix} \otimes \begin{bmatrix} \frac{1-r}{2} & \frac{r}{2} \\ \frac{r}{2} & \frac{1-r}{2} \end{bmatrix}$$

In general:

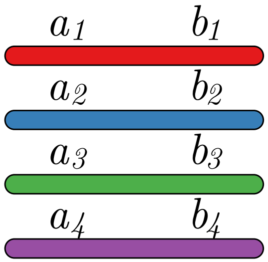
$$\mathbf{V}_1 \otimes \mathbf{V}_2 \otimes \cdots \otimes \mathbf{V}_{\frac{p}{2}}$$

All elements of this product are of the form

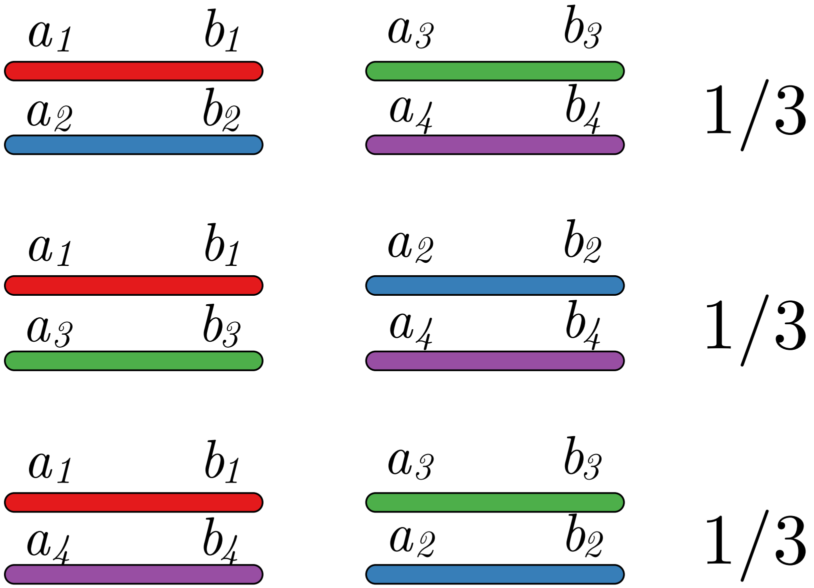
$$\Pr(G_A, G_B \mid \psi_j, r) = \frac{(1-r)^{(\frac{p}{2}-l)} p^l}{2^{\frac{p}{2}}}$$

l : known number of recombinant bivalents between loci A and B

Recombination Fraction - autotetraploid

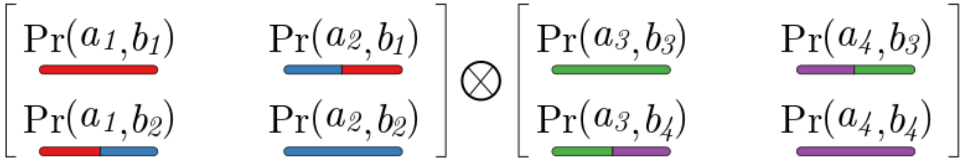


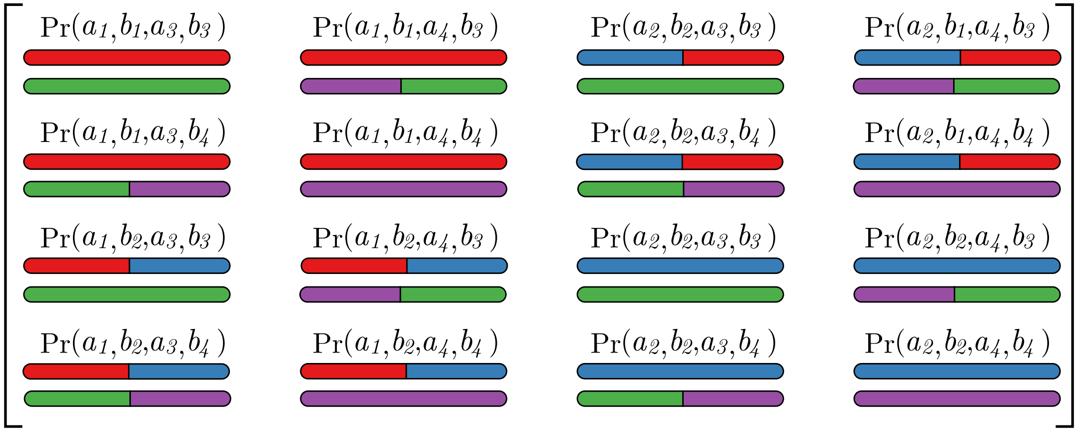
Bivalent pairing at meiosis



Recombination Fraction - autotetraploid

First bivalent pairing at meiosis



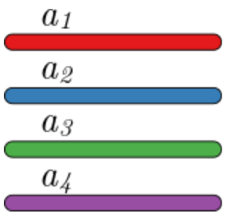


$$\begin{bmatrix}\frac{1-r}{2} & \frac{r}{2}\\ \frac{r}{2} & \frac{1-r}{2}\end{bmatrix}\otimes\begin{bmatrix}\frac{1-r}{2} & \frac{r}{2}\\ \frac{r}{2} & \frac{1-r}{2}\end{bmatrix}=\begin{bmatrix}\frac{(1-r)^2}{4} & \frac{(1-r)r}{4} & \frac{(1-r)r}{4} & \frac{r^2}{4}\\ \frac{(1-r)r}{4} & \frac{(1-r)^2}{4} & \frac{r^2}{4} & \frac{(1-r)r}{4}\\ \frac{(1-r)r}{4} & \frac{r^2}{4} & \frac{(1-r)^2}{4} & \frac{(1-r)r}{4}\\ \frac{r^2}{4} & \frac{(1-r)r}{4} & \frac{(1-r)r}{4} & \frac{(1-r)^2}{4}\end{bmatrix}$$

Recombination Fraction - autotetraploid

$1/3$

	$a_1 a_3$	$a_1 a_4$	$a_2 a_3$	$a_2 a_4$
$b_1 b_3$	$\frac{(1-r)^2}{4}$	$\frac{(1-r)r}{4}$	$\frac{(1-r)r}{4}$	$\frac{r^2}{4}$
$b_1 b_4$	$\frac{(1-r)r}{4}$	$\frac{(1-r)^2}{4}$	$\frac{r^2}{4}$	$\frac{(1-r)r}{4}$
$b_2 b_3$	$\frac{(1-r)r}{4}$	$\frac{r^2}{4}$	$\frac{(1-r)^2}{4}$	$\frac{(1-r)r}{4}$
$b_2 b_4$	$\frac{r^2}{4}$	$\frac{(1-r)r}{4}$	$\frac{(1-r)r}{4}$	$\frac{(1-r)^2}{4}$



$a_1\ a_2$

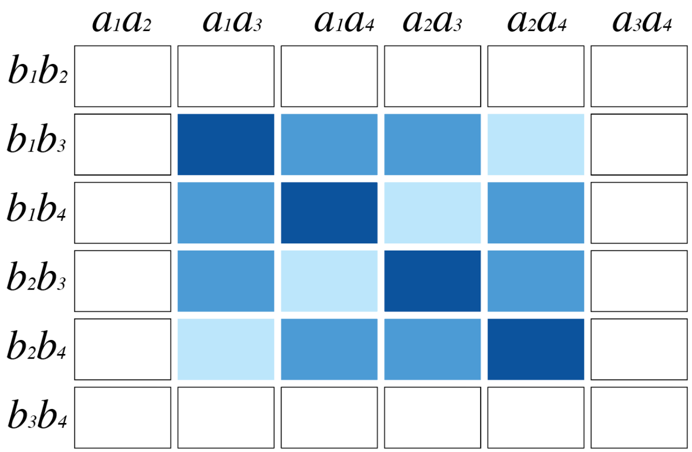
$a_1\ a_3$

$a_1\ a_4$

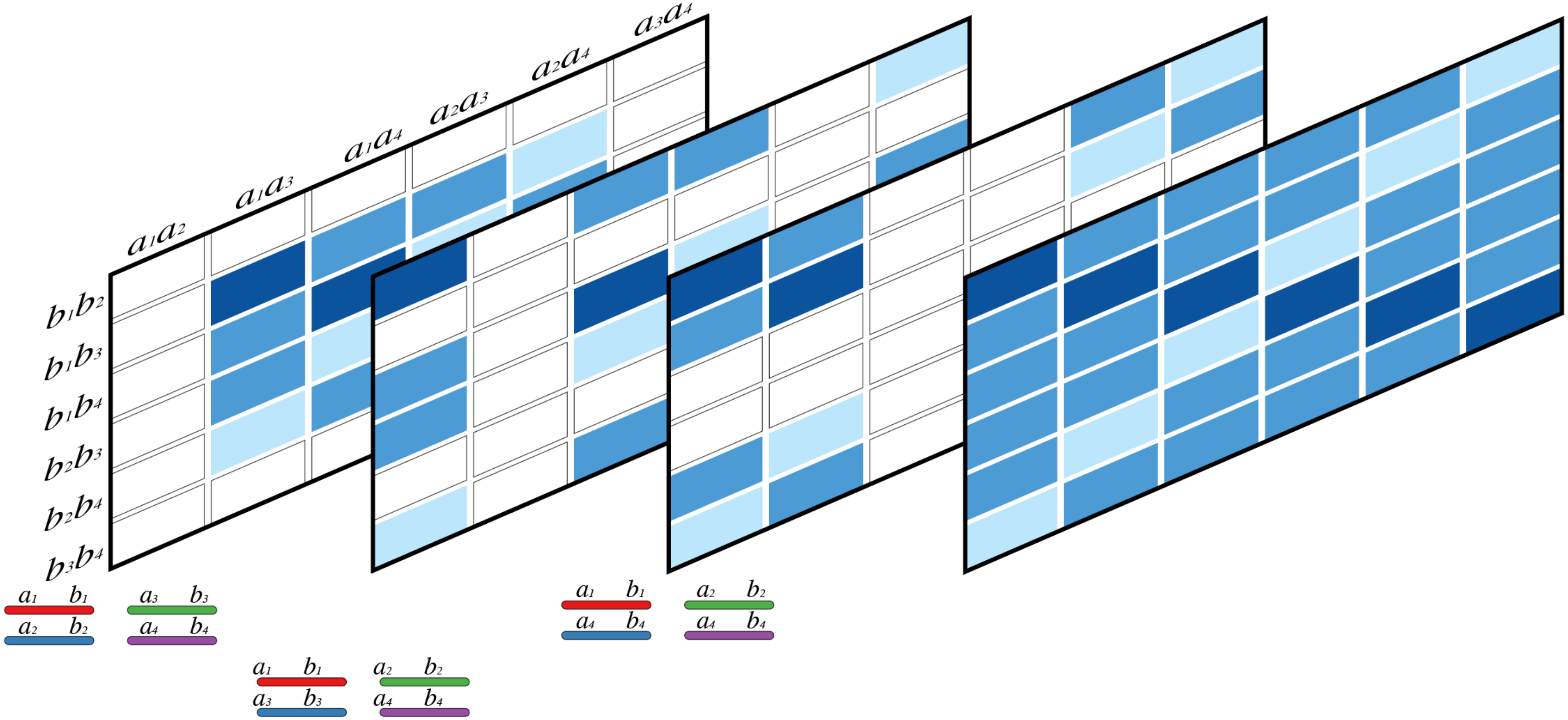
$a_2\ a_3$

$a_2\ a_4$

$a_3\ a_4$

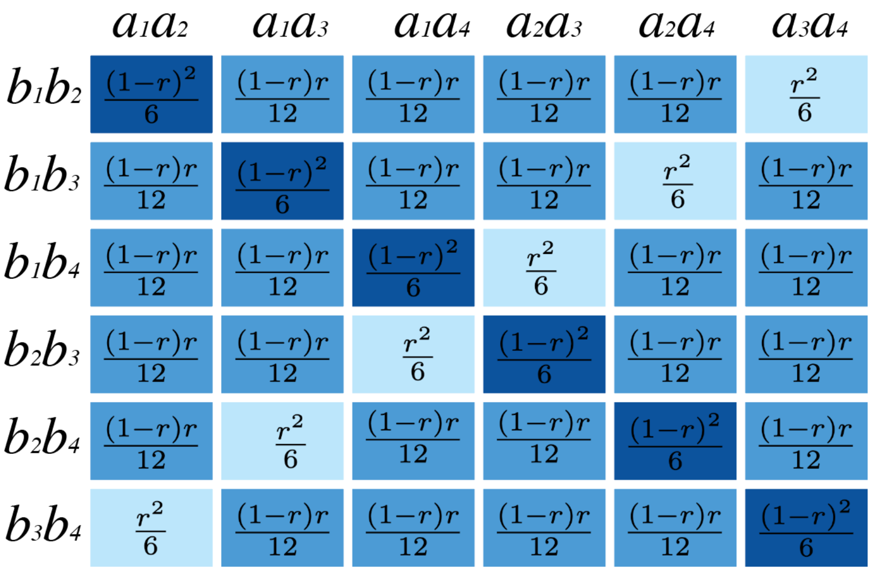


Recombination Fraction - autotetraploid



$$\Pr(G_A, G_B \mid r) = \sum_j \Pr(G_A, G_B \mid \psi_j, r) \Pr(\psi_j)$$

Recombination Fraction - autotetraploid



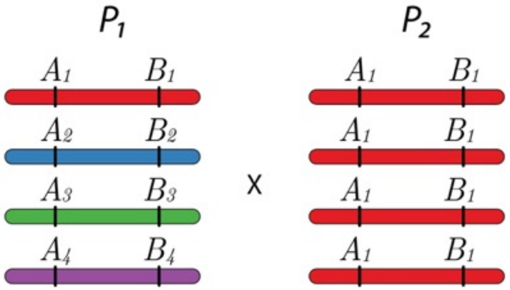
	a_1a_2	a_1a_3	a_1a_4	a_2a_3	a_2a_4	a_3a_4
b_1b_2	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$
b_1b_3	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$
b_1b_4	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$
b_2b_3	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$
b_2b_4	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$
b_3b_4	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$

$$L = \Pr(\text{model}|\text{data}) = \prod_n \Pr(\text{loc}_B\ ,\ \text{loc}_A)$$

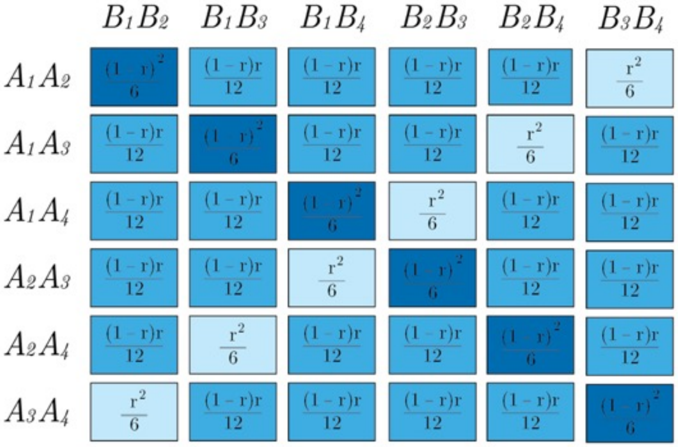
$$= \left[\frac{(1-r)^2}{6}\right]^{n_1}_{\blacksquare} \left[\frac{(1-r)r}{12}\right]^{n_2}_{\blacksquare} \left[\frac{r^2}{6}\right]^{n_3}_{\square}$$

Recombination Fraction – autotetraploid

Fully informative marker

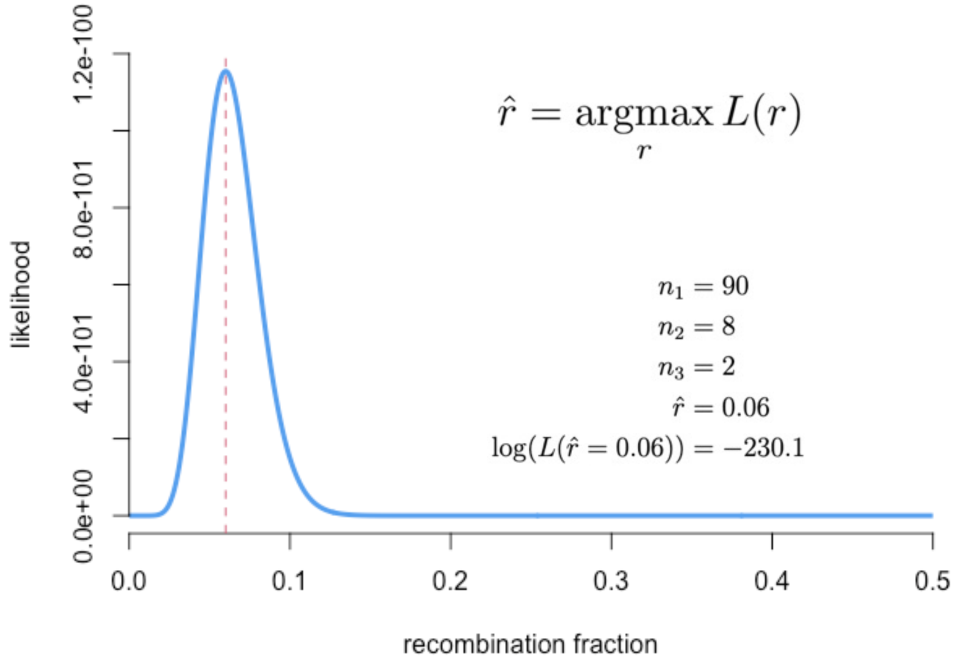


$$\Pr(G_A, G_B \mid r)$$



	B_1B_2	B_1B_3	B_1B_4	B_2B_3	B_2B_4	B_3B_4
A_1A_2	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$
A_1A_3	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$
A_1A_4	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$
A_2A_3	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$
A_2A_4	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$
A_3A_4	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$

$$L(r) = \left[\frac{(1-r)^2}{6}\right]^{n_1} \left[\frac{(1-r)r}{12}\right]^{n_2} \left[\frac{r^2}{6}\right]^{n_3}$$

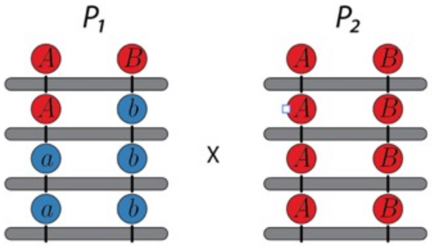


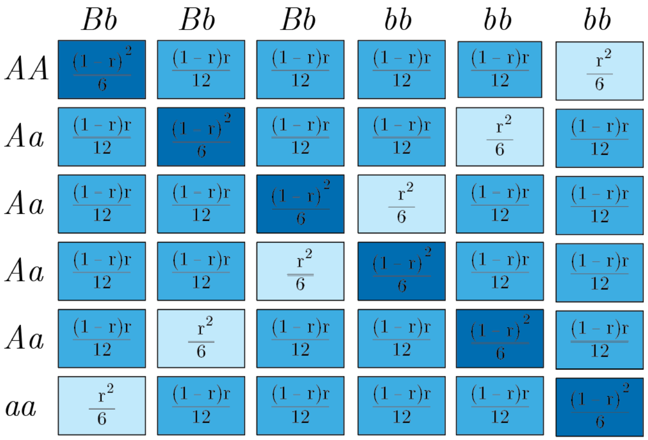
$$\hat{r} = \underset{r}{\operatorname{argmax}}\, L(r)$$

$n_1 = 90$
$n_2 = 8$
$n_3 = 2$
$\hat{r} = 0.06$
$\log(L(\hat{r} = 0.06)) = -230.1$

Recombination Fraction – autotetraploid

Partially informative marker – Duplex/simplex – Association

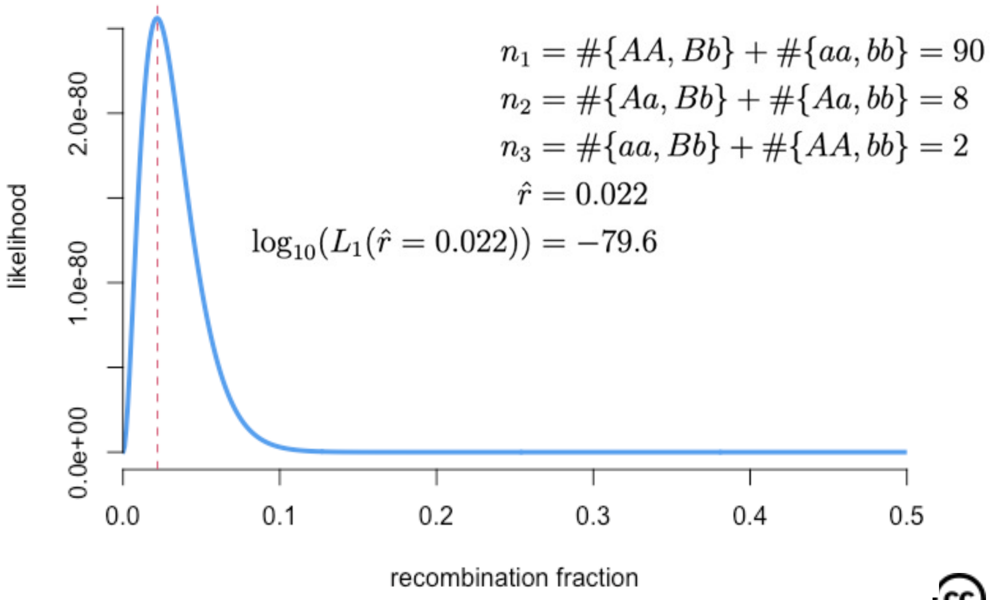




	Bb	Bb	Bb	bb	bb	bb
AA	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$
Aa	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$
Aa	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$
Aa	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$
Aa	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$
aa	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$

	Bb	bb
AA	$\frac{(1-r)}{6}$	$\frac{r}{6}$
Aa	$\frac{1}{3}$	$\frac{1}{3}$
aa	$\frac{r}{6}$	$\frac{(1-r)}{6}$

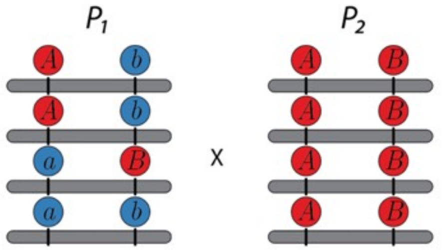
$$L_1(r) = \left[\frac{(1-r)}{6}\right]^{n_1} \left[\frac{1}{3}\right]^{n_2} \left[\frac{r}{6}\right]^{n_3}$$

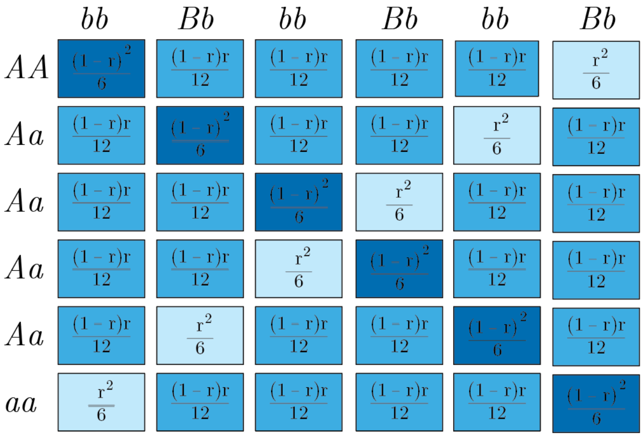


$$n_1 = \#\{AA, Bb\} + \#\{aa, bb\} = 90$$
$$n_2 = \#\{Aa, Bb\} + \#\{Aa, bb\} = 8$$
$$n_3 = \#\{aa, Bb\} + \#\{AA, bb\} = 2$$
$$\hat{r} = 0.022$$
$$\log_{10}(L_1(\hat{r} = 0.022)) = -79.6$$

Recombination Fraction – autotetraploid

Partially informative marker – Duplex/simplex – Repulsion

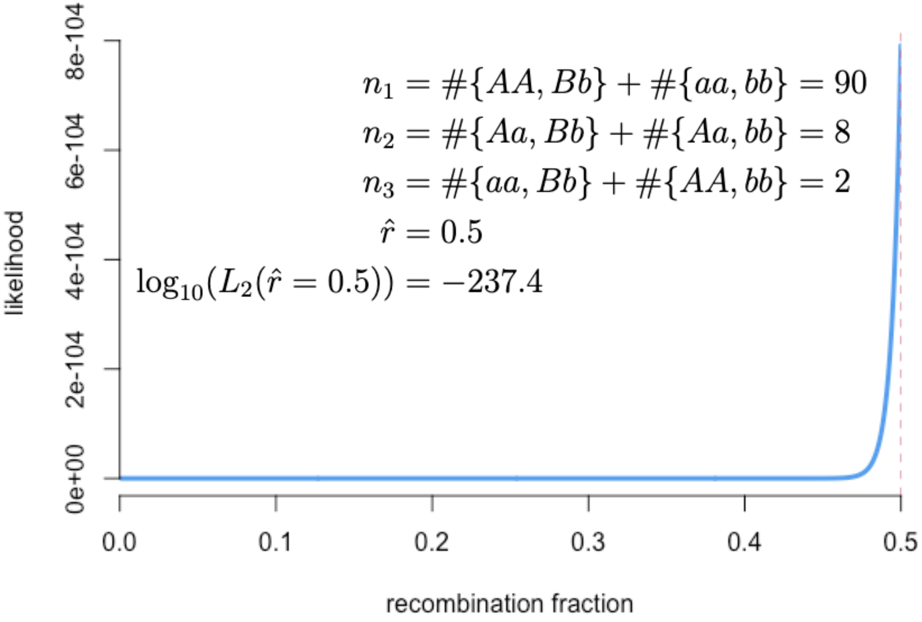




	bb	Bb	bb	Bb	bb	Bb
AA	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$
Aa	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$
Aa	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$
Aa	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$
Aa	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$
aa	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$

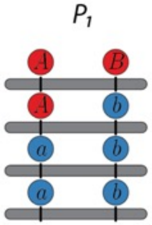
	Bb	bb
AA	$\frac{r}{6}$	$\frac{(1-r)}{6}$
Aa	$\frac{1}{3}$	$\frac{1}{3}$
aa	$\frac{(1-r)}{6}$	$\frac{r}{6}$

$$L_2(r) = \left[\frac{r}{6}\right]^{n_1} \left[\frac{1}{3}\right]^{n_2} \left[\frac{(1-r)}{6}\right]^{n_3}$$

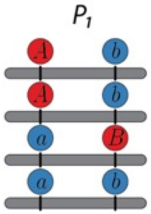


$$n_1 = \#\{AA, Bb\} + \#\{aa, bb\} = 90$$
$$n_2 = \#\{Aa, Bb\} + \#\{Aa, bb\} = 8$$
$$n_3 = \#\{aa, Bb\} + \#\{AA, bb\} = 2$$
$$\hat{r} = 0.5$$
$$\log_{10}(L_2(\hat{r} = 0.5)) = -237.4$$

Recombination Fraction – assessing linkage phases



$$\hat{r}_1 = \arg\max_{r} L_1(r) \implies L_1(0.022) = -79.6$$



$$\hat{r}_2 = \arg\max_{r} L_2(r) \implies L_2(0.5) = -237.4$$

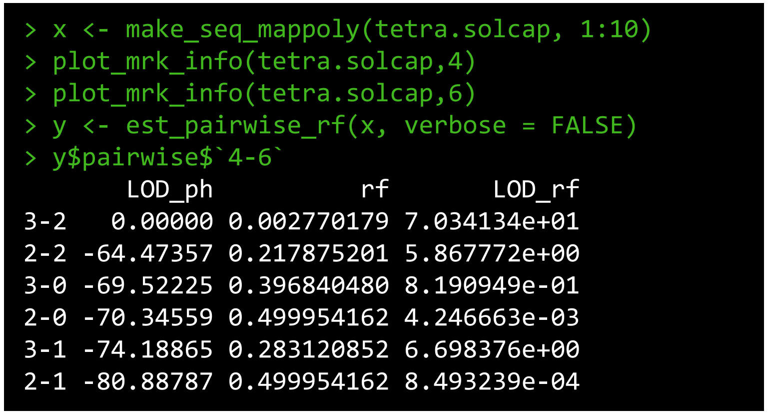
Compare likelihoods choosing the most likely configuration

logarithm-of-odds (LOD-score)

$$\begin{aligned} LOD &= \log_{10}\left(\frac{L_1}{L_2}\right) \\ &= \log_{10}(L_1) - \log_{10}(L_2) \\ &= -79.6 - (-237.4) = 157.8 \end{aligned}$$

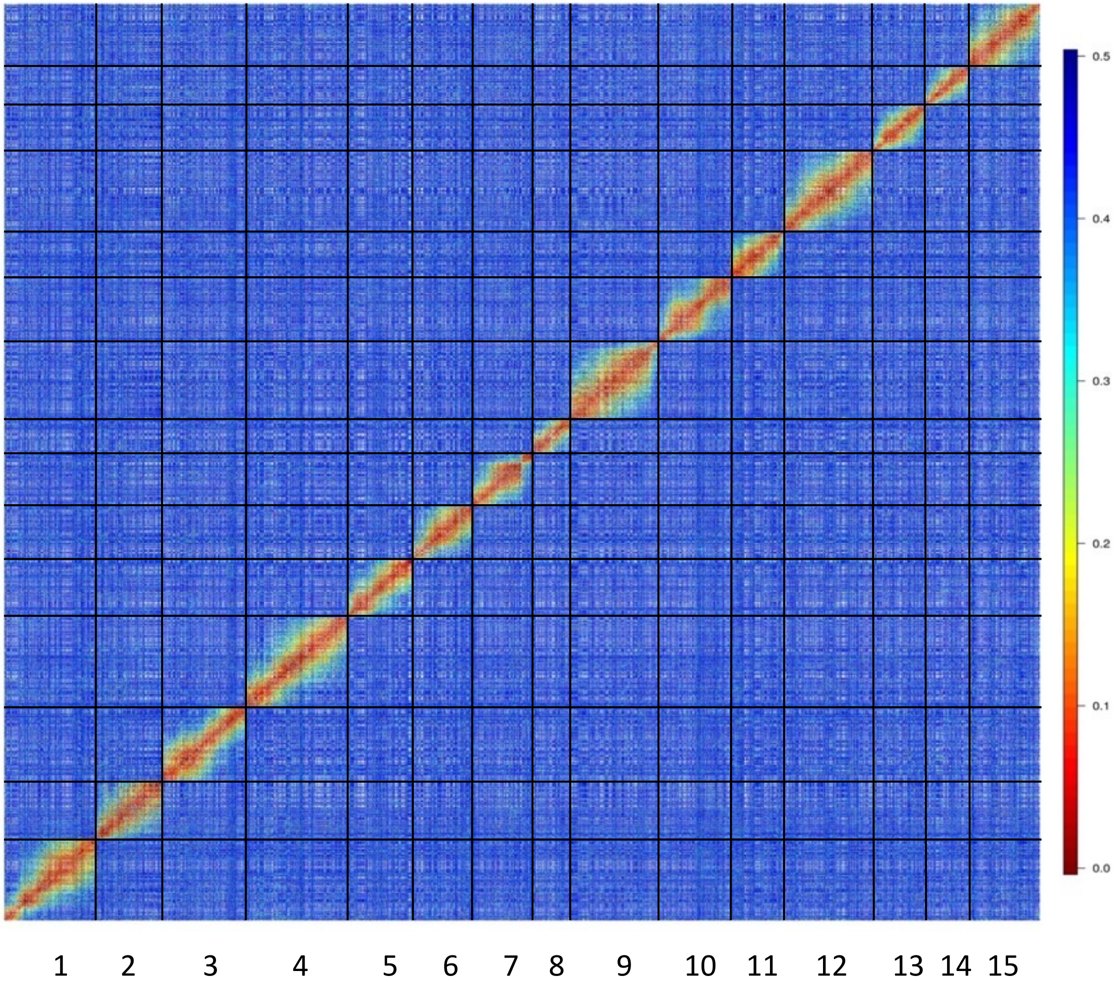
Markers A and B are most likely in association

- Pairwise MLEs of r are used to group markers into linkage groups and order markers within each linkage group using optimization algorithms such as MDS.
- Given a sequence of ordered markers, it is possible to extend the idea of comparing the likelihoods of competing linkage phases throughout multiple markers.



```
> x <- make_seq_mappoly(tetra.solcap, 1:10)
> plot_mrk_info(tetra.solcap,4)
> plot_mrk_info(tetra.solcap,6)
> y <- est_pairwise_rf(x, verbose = FALSE)
> y$pairwise$`4-6`
       LOD_ph          rf       LOD_rf
3-2   0.00000 0.002770179 7.034134e+01
2-2 -64.47357 0.217875201 5.867772e+00
3-0 -69.52225 0.396840480 8.190949e-01
2-0 -70.34559 0.499954162 4.246663e-03
3-1 -74.18865 0.283120852 6.698376e+00
2-1 -80.88787 0.499954162 8.493239e-04
```

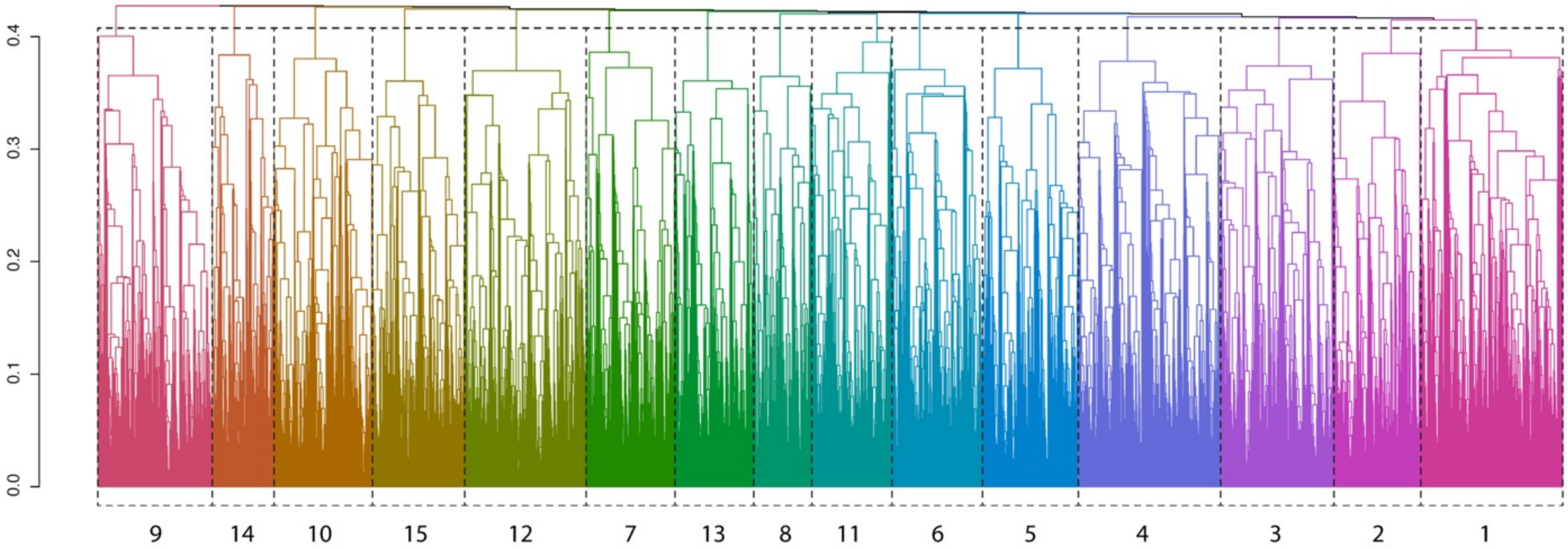
Ordering with MDS – 15 linkage groups, 30684 SNPs



Numbers indicate the associated chromosomes in *I. trifida* and *I. triloba* reference genomes

Two – point analysis and grouping

• Number of markers:	38,701
• Number of recombination fractions:	~749 million pairs

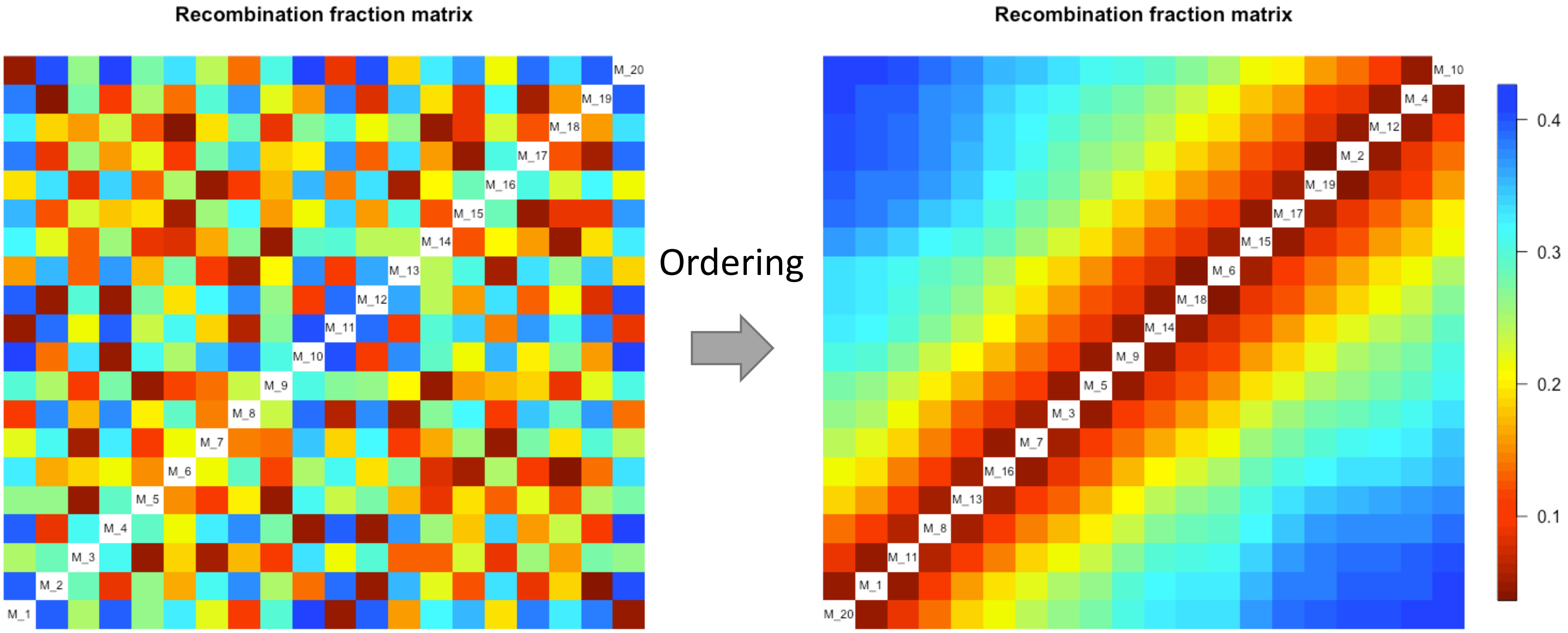


Numbers indicate the associated chromosomes in *I. trifida* and *I. triloba* reference genomes

cc

Multidimensional Scaling Algorithm (MDS)

- Reduce data from many dimensions preserving the observed distances between points by minimizing a loss function L.



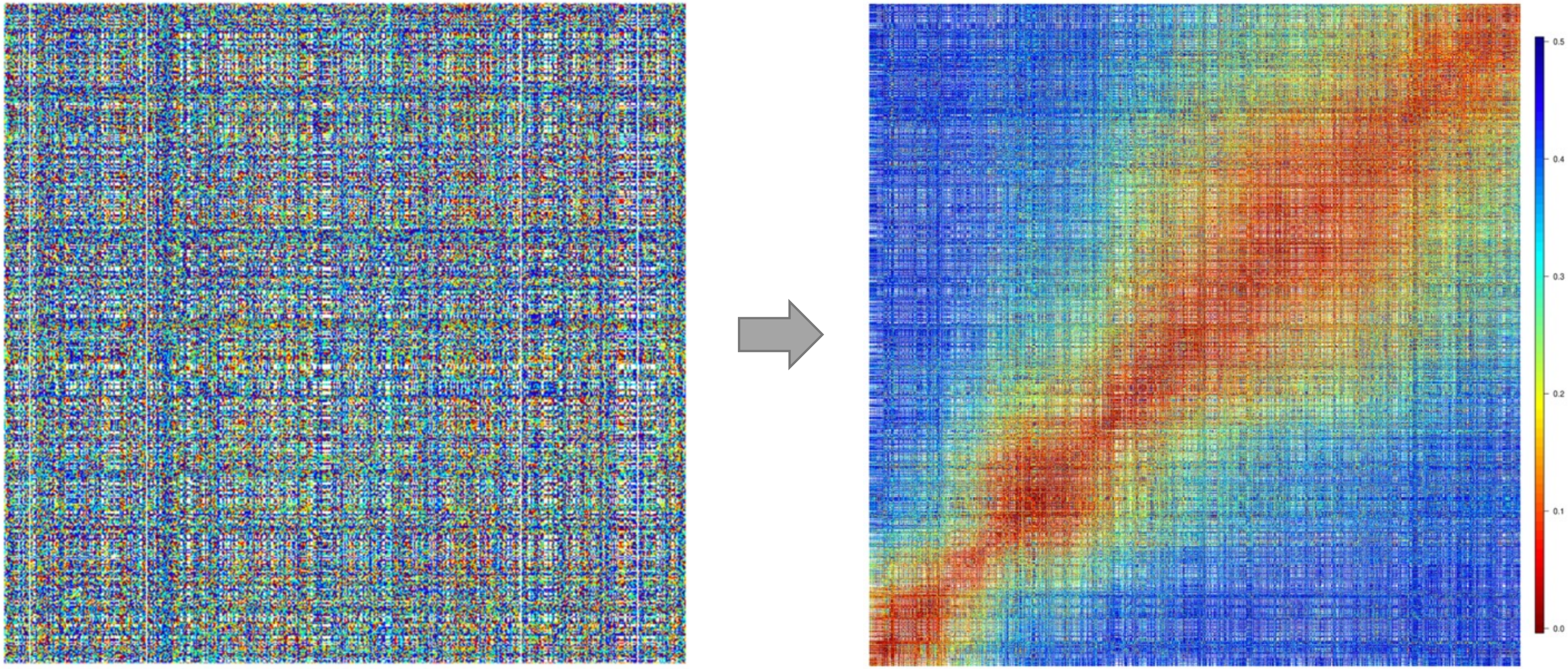
Recombination fraction matrix

Ordering

Recombination fraction matrix

Preedy and Hackett, 2016

Multidimensional Scaling Algorithm (MDS)

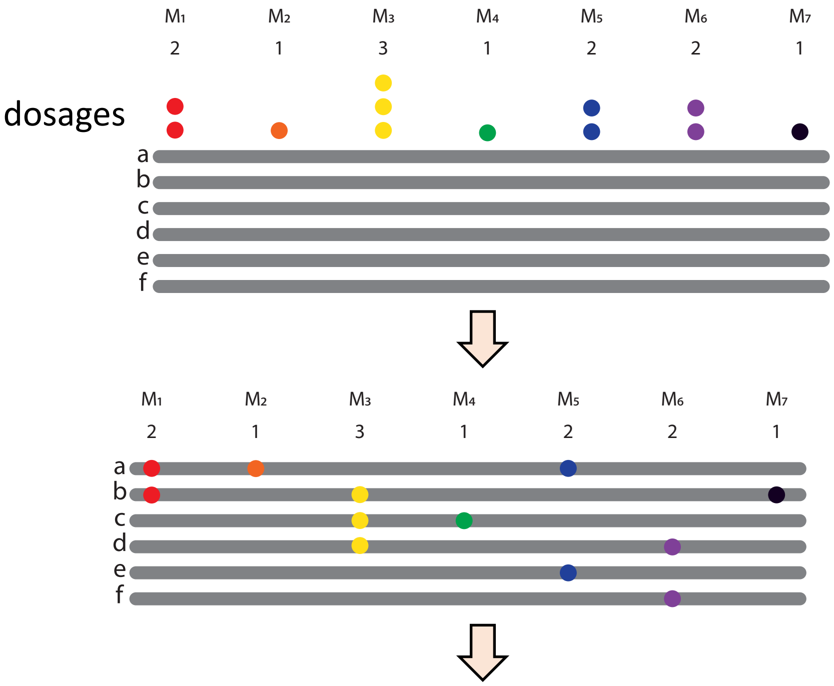


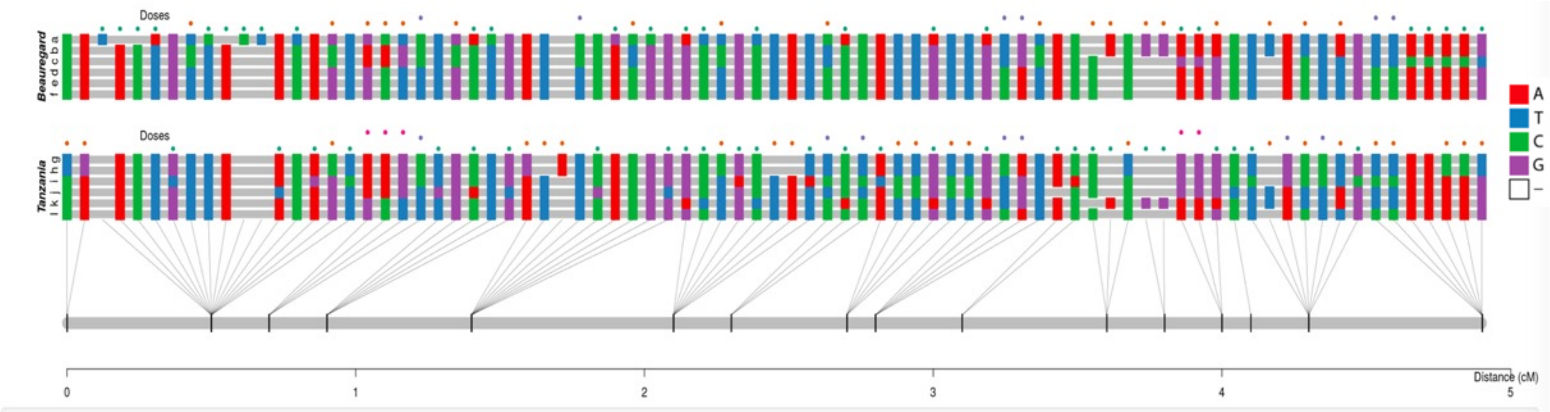
Sweetpotato linkage group 1: 2745 markers

[Preedy and Hackett, 2016]

Haplotyping in polyploids

- Placement of allelic variants in the homologs in a homology group

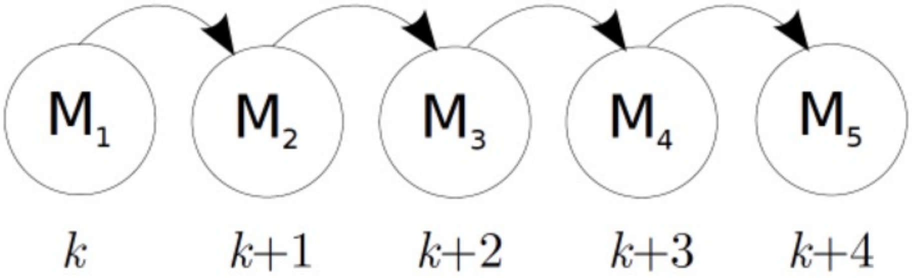




Markov Models - Intuition

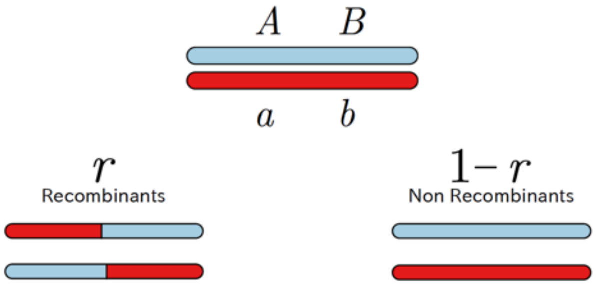
Markov Models

Markovian property: Given the present, the future does not depend on the past.



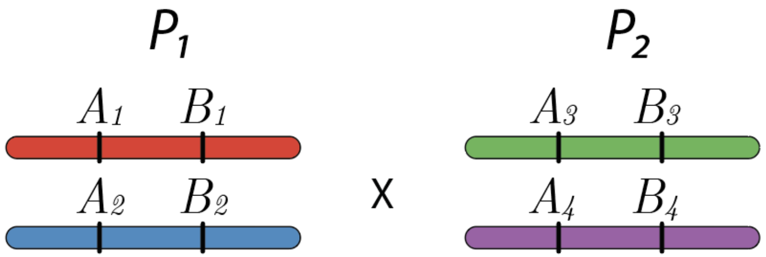
Transition probability

$$\Pr(G_{k+1}|G_k)$$

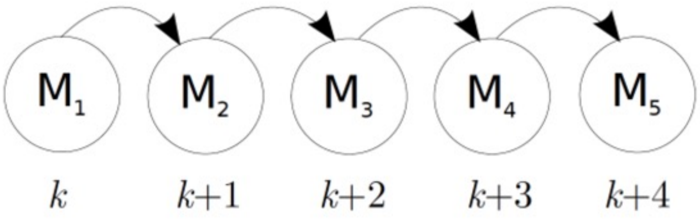


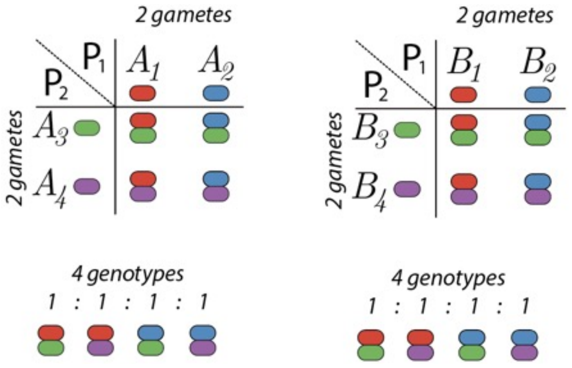
$$\mathbf{V} = \begin{bmatrix} \Pr(A\,,B) & \Pr(a\,,B) \\ \Pr(A,b) & \Pr(a\,,b) \end{bmatrix} = \begin{bmatrix} \frac{1-r}{2} & \frac{r}{2} \\ \frac{r}{2} & \frac{1-r}{2} \end{bmatrix}$$

Multilocus linkage analysis in outcrossing diploids



Markovian process

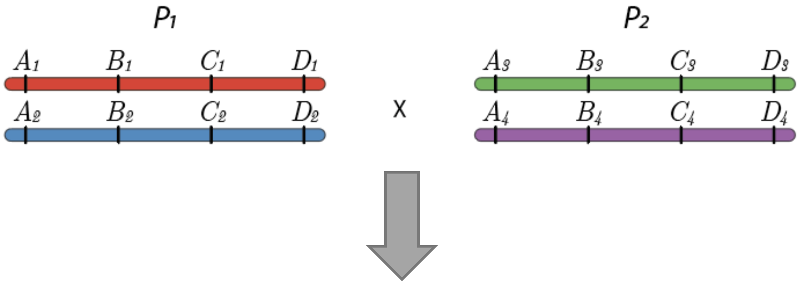




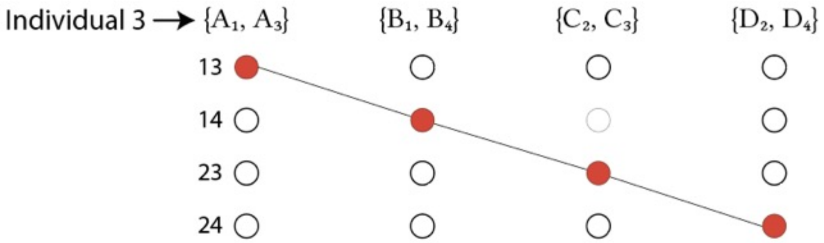
$$\Pr(G_{k+1}|G_k) = \{a_{i,j}\} =$$

$$\begin{array}{c|cccc} & 13 & 14 & 23 & 24 \\ \hline 13 & (1-r)^2 & (1-r)\,r & (1-r)\,r & r^2 \\ 14 & (1-r)\,r & (1-r)^2 & r^2 & (1-r)\,r \\ 23 & (1-r)\,r & r^2 & (1-r)^2 & (1-r)\,r \\ 24 & r^2 & (1-r)\,r & (1-r)\,r & (1-r)^2 \end{array}$$

Markov model



	M_1	M_2	M_3	M_4
Ind_1	$\{A_1, A_3\}$	$\{B_1, B_4\}$	$\{C_2, C_4\}$	$\{D_2, D_4\}$
Ind_2	$\{A_1, A_3\}$	$\{B_1, B_4\}$	$\{C_1, C_4\}$	$\{D_2, D_4\}$
Ind_3	$\{A_1, A_2\}$	$\{B_1, B_3\}$	$\{C_1, C_3\}$	$\{D_2, D_3\}$
$\vdots$		$\vdots$		
Ind_n	$\{A_2, A_3\}$	$\{B_2, B_3\}$	$\{C_2, C_3\}$	$\{D_2, D_3\}$

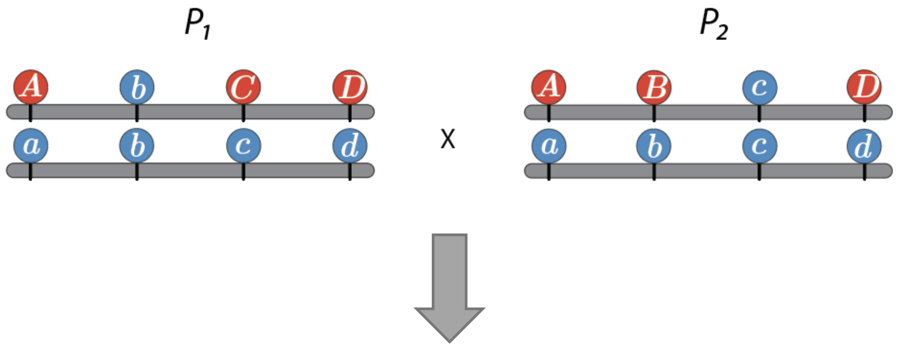


$$
\begin{aligned}
\Pr(G_A, G_B, G_C, G_D \mid \mathbf{r})_3 &= \Pr(A_1, A_3) \times \\
&= \Pr(B_1, B_4 \mid A_1, A_3) \times \\
&= \Pr(C_2, C_3 \mid B_1, B_4) \times \\
&= \Pr(D_3, D_4 \mid C_2, C_3) \times
\end{aligned}
$$

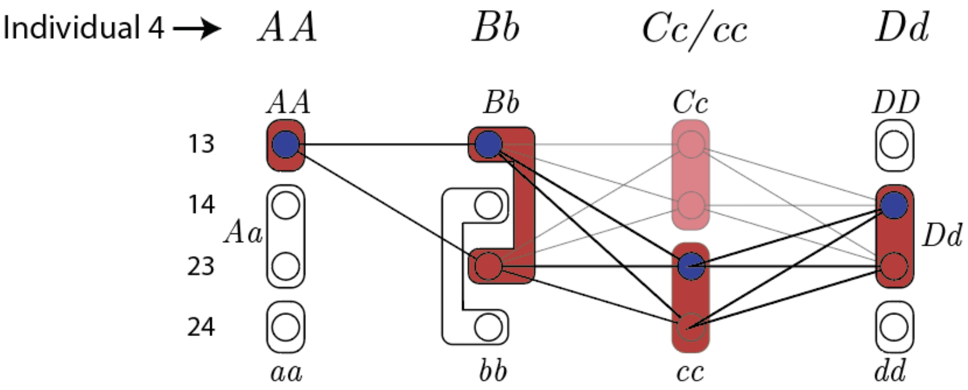
$$L(\mathbf{r}) = \prod_n \Pr(G_A, G_B, G_C, G_D \mid \mathbf{r})$$

$$\hat{\mathbf{r}} = \underset{\mathbf{r}}{\operatorname{argmax}} \, L(\mathbf{r})$$

Hidden Markov Model - HMM



	M_1	M_2	M_3	M_4
Ind_1	AA	bb	cc	dd
Ind_2	AA	bb	Cc	dd
Ind_3	AA	bb	cc	dd
Ind_4	AA	Bb	Cc/cc $0.2 \quad 0.8$	Dd
⋮		⋮		
Ind_n	Aa	Bb	Cc	Dd

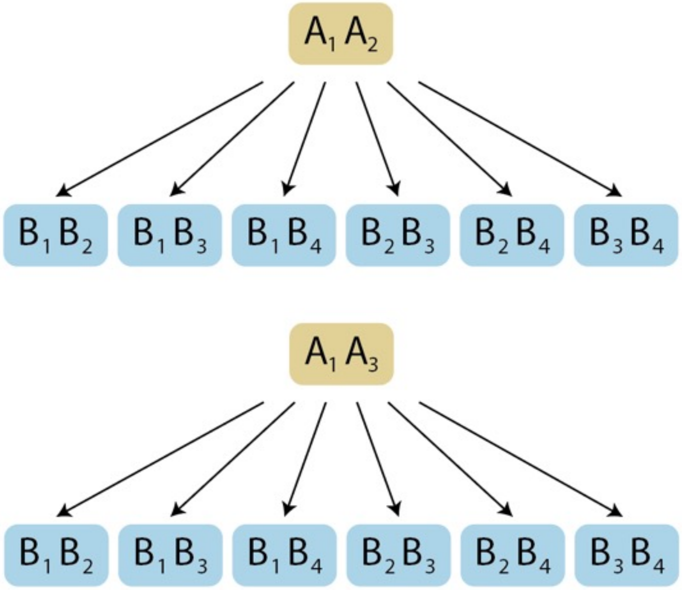


$$L(\mathbf{r}) = \prod_{n} \Pr(G_A, G_B, G_C, G_D \mid \mathbf{r})$$

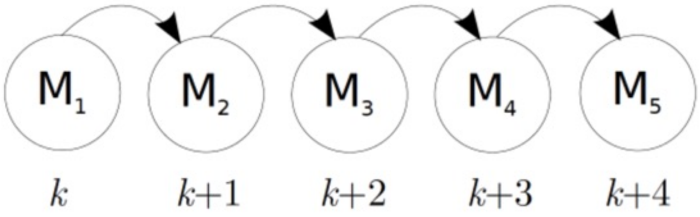
$$\hat{\mathbf{r}} = \underset{\mathbf{r}}{\operatorname{argmax}}\, L(\mathbf{r})$$

Multilocus linkage analysis in polyploids

What is the probability of observing a specific state at a moment (or position), given we observed some state in a previous moment?



Markov Model: Conditional independence

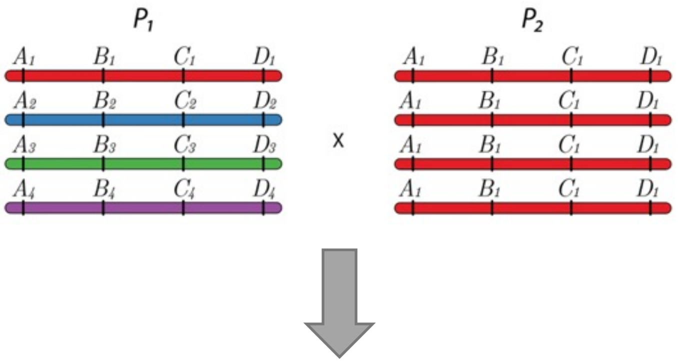


$$\Pr(G_{k+1}|G_k) = \frac{(1-r_k)^{\frac{p}{2}-l}(r_k)^l}{\binom{\frac{p}{2}}{l}}$$

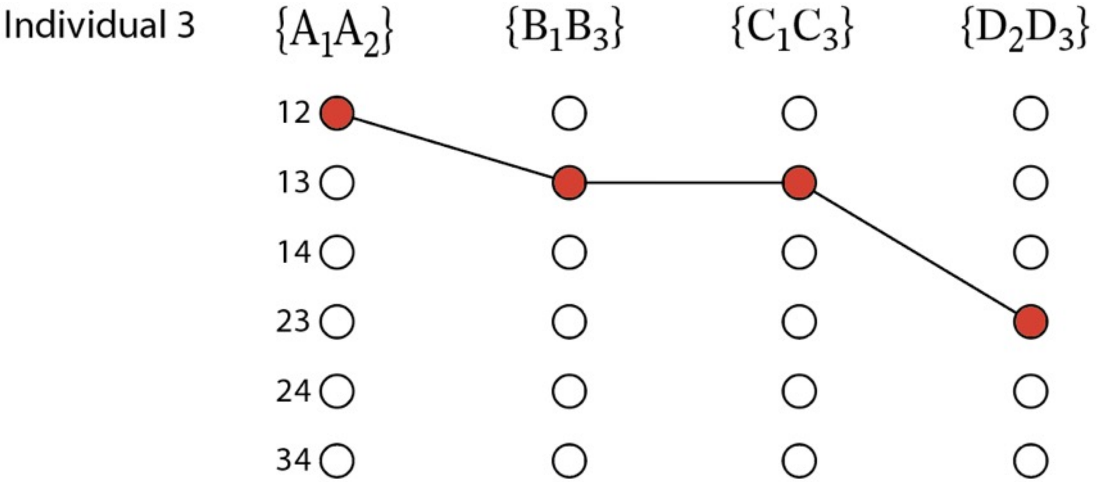
where r_k is the recombination frequency between loci k and k+1, p is the ploidy level and l is the number of recombinant events between k and k+1.

Leach et al. (2010) doi:10.1073/pnas.0908477107 **Mollinari and Garcia (2019)** doi:10.1534/g3.119.400378

Markov model



	M_1	M_2	M_3	M_4
Ind_1	{A_1, A_2}	{B_1, B_3}	{C_3, C_4}	{D_3, D_4}
Ind_2	{A_1, A_2}	{B_1, B_3}	{C_1, C_3}	{D_3, D_4}
Ind_3	{A_1, A_2}	{B_1, B_3}	{C_1, C_3}	{D_2, D_3}
⋮		⋮		
Ind_n	{A_2, A_3}	{B_2, B_3}	{C_2, C_3}	{D_2, D_3}

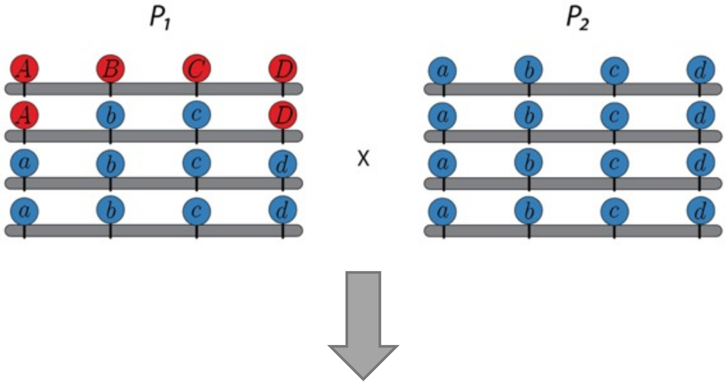


$$L(\mathbf{r}) = \prod_{n} \Pr(G_A, G_B, G_C, G_D \mid \mathbf{r})$$

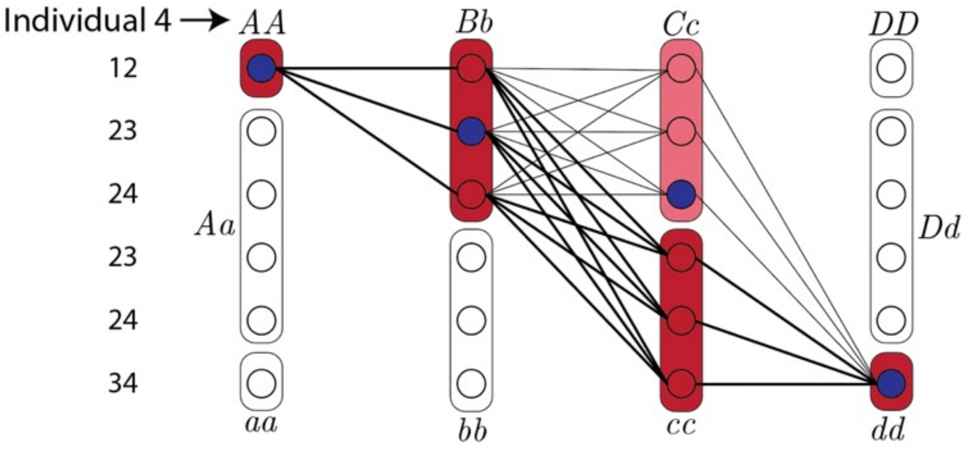
$$\hat{\mathbf{r}} = \underset{\mathbf{r}}{\operatorname{argmax}}\, L(\mathbf{r})$$

Hidden Markov Model – HMM – Emission

$$b_j(O) = \Pr\left(O \middle| \mathcal{G}^m_{k,j}, \varphi^k_P, \varphi^k_Q\right) = \begin{cases} 1-\epsilon & \text{if} \quad O = \delta(k,j) \\ \dfrac{\epsilon}{m} & \text{otherwise} \end{cases}$$



	M_1	M_2	M_3	M_4
Ind_1	AA	Bb	cc	dd
Ind_2	AA	Bb	Cc	dd
Ind_3	AA	Bb	Cc	Dd
Ind_4	AA	Bb	Cc/cc $0.2 \quad 0.8$	Dd
⋮		⋮		
Ind_n	Aa	Bb	Cc	Dd

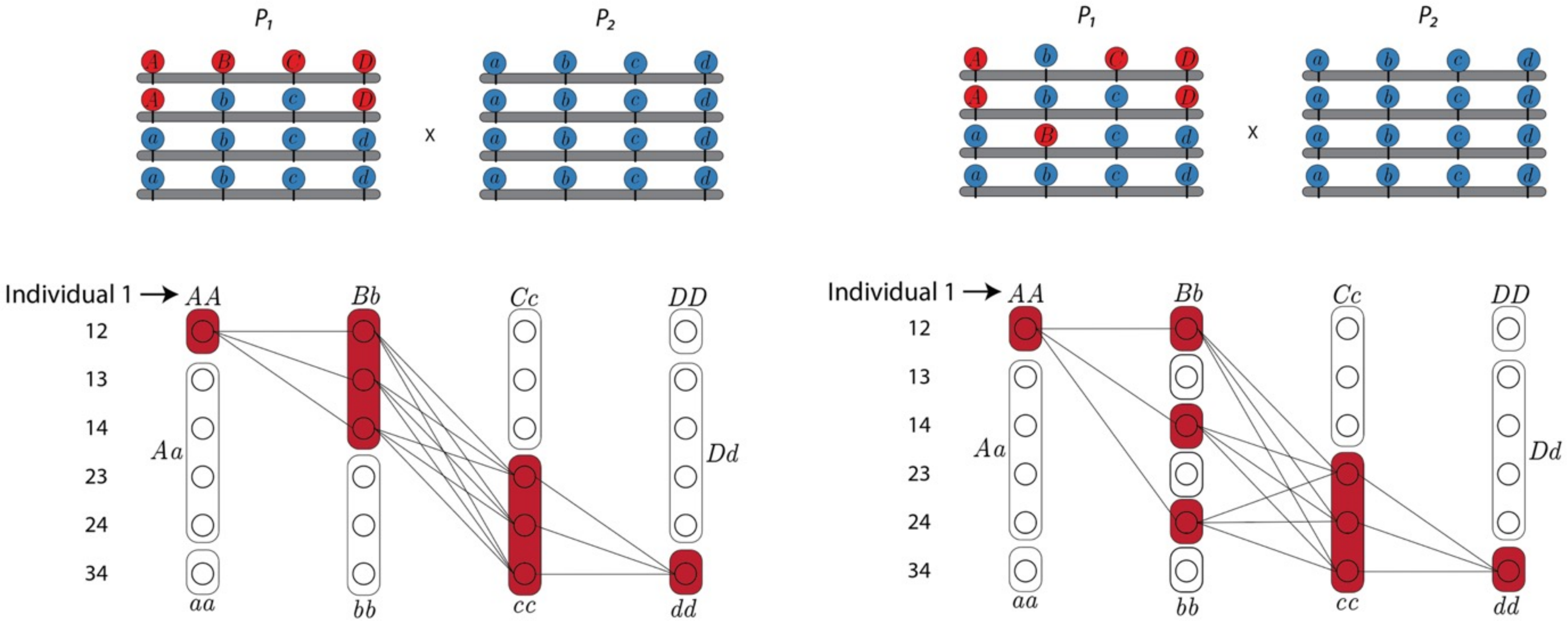


$$L(\mathbf{r}) = \prod_n \Pr(G_A, G_B, G_C, G_D \mid \mathbf{r})$$

$$\hat{\mathbf{r}} = \operatorname*{argmax}_{\mathbf{r}} L(\mathbf{r})$$

Hidden Markov Model - HMM

Assessing *different linkage phases* using multilocus analysis



HMM – Forward - Backward and EM algorithms

Forward

1. Initialization:

$$\alpha_1(j) = \gamma_j b_j(O_1),\ j = 1, \cdots, g_m$$

2. Induction:

$$\alpha_{k+1}(j') = \left[\sum_{j}^{g_m} \alpha_k(j)\, t_k(j, j')\right] b_{j'}(O_{k+1})$$

where $k = 1, \cdots, z-1$ and $j' = 1, \cdots, g_m$

3. Termination:

$$\Pr(O_1, \cdots O_z | \mathbf{r}, \Phi_P, \Phi_Q, \Pi) = \sum_{j=1}^{g_m} \alpha_z(j)$$

Then, the likelihood of the model is defined as

$$\prod_{i=1}^{n} \Pr(O_{1,i}, \cdots, O_{z,i} | \mathbf{r}, \Phi_P, \Phi_Q, \Pi_i)$$

Backward

1. Initialization:

$$\beta_z(j) = 1,\ j = 1, \cdots, g_m$$

2. Induction:

$$\beta_k(j) = \sum_{j'}^{g_m} t_k(j, j') b_{j'}(O_{k+1}) \beta_{k+1}(j')$$

where $k = z-1, z-2, \cdots, 1$ and $j = 1, \cdots, g_m$

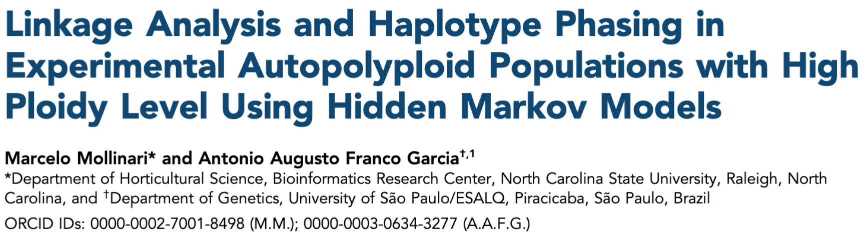
Expectation-maximization

$$\xi_k(j, j' \mid \mathbf{r}) = \Pr(\mathcal{G}^m_{k,j}, \mathcal{G}^m_{k+1,j'} \mid O_1, \cdots O_z, \Pi, \mathbf{r}, \Phi_P, \Phi_Q)$$
$$= \frac{\alpha_k(j) t_k(j, j') b_{j'}(O_{k+1}) \beta_{k+1}(j')}{\sum_{j=1}^{g_m} \sum_{j'=1}^{g_m} \alpha_k(j) t_k(j, j') b_{j'}(O_{k+1}) \beta_{k+1}(j')}$$

The recombination frequency r_k can be estimated through an iterative process using

$$r_k^{s+1} = \sum_{i=1}^{n} \sum_{j=1}^{g_m} \sum_{j'=1}^{g_m} \frac{\xi_k(j, j' \mid \mathbf{r}^s) \phi(j, j')}{n}$$

where $\xi_k(j, j' \mid \mathbf{r}^s)$ is calculated for individual i, $\phi(j, j') = \frac{(l_P + l_Q)}{m}$



Linkage Analysis and Haplotype Phasing in Experimental Autopolyploid Populations with High Ploidy Level Using Hidden Markov Models

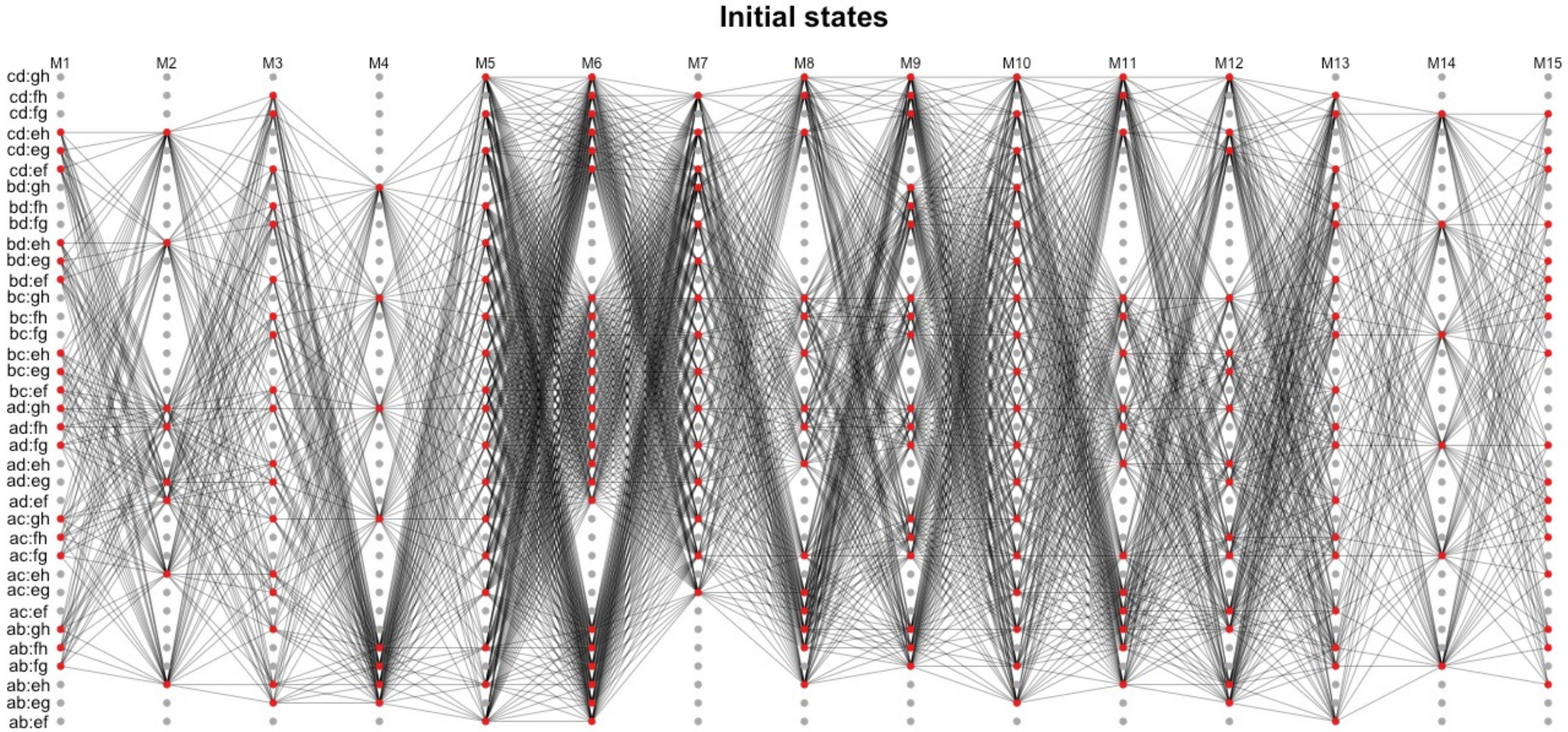
Marcelo Mollinari* and Antonio Augusto Franco Garcia[†,1]

*Department of Horticultural Science, Bioinformatics Research Center, North Carolina State University, Raleigh, North Carolina, and †Department of Genetics, University of São Paulo/ESALQ, Piracicaba, São Paulo, Brazil

ORCID IDs: 0000-0002-7001-8498 (M.M.); 0000-0003-0634-3277 (A.A.F.G.)

HMM – Backward-forward

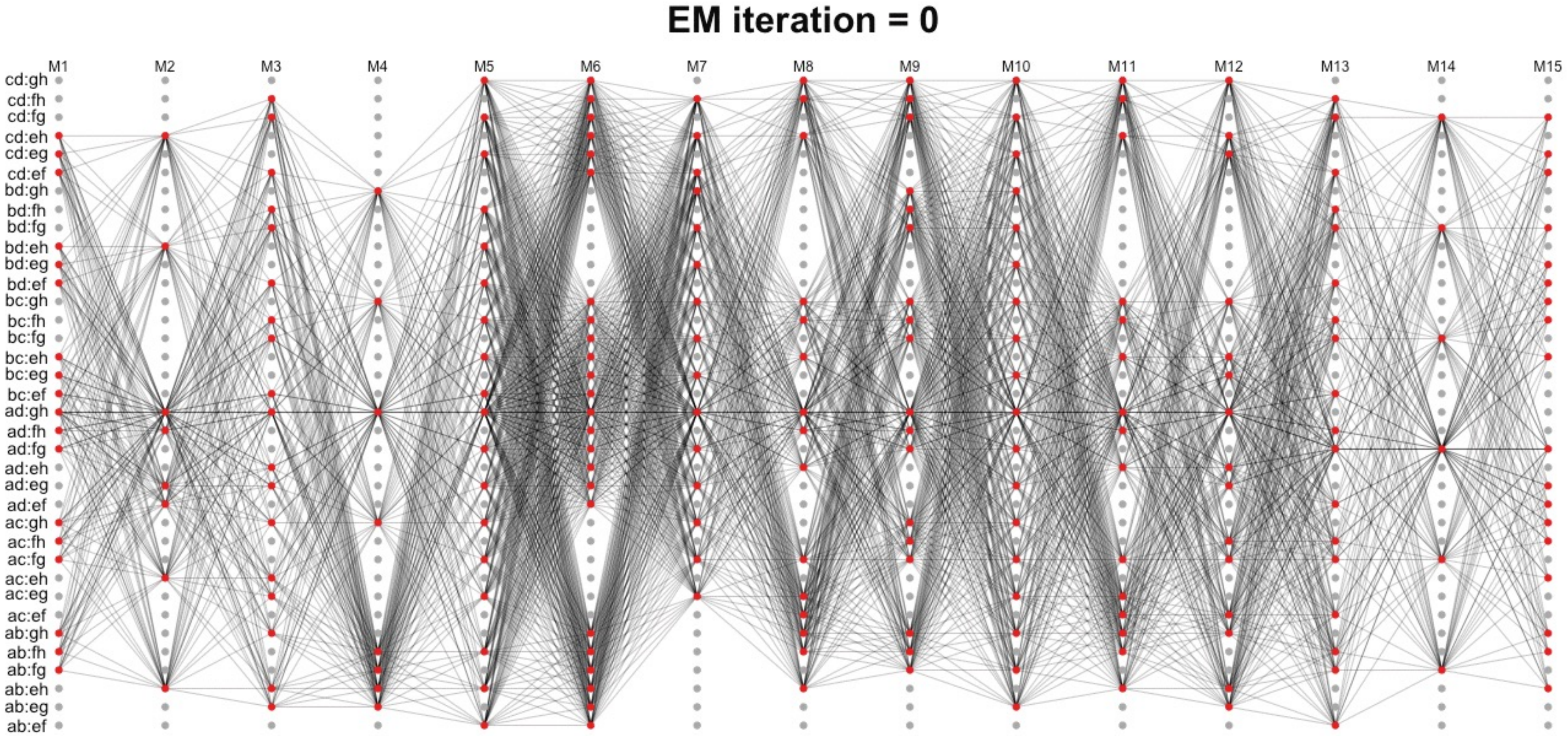
- Tetraploid example, one individual, 15 markers



1 1 1 3 1 1 2 2 0 1 2 1 1 0 1

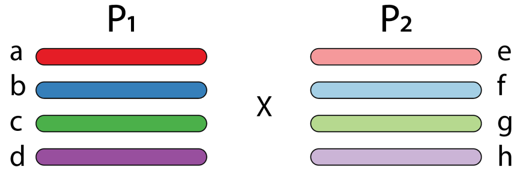
HMM – EM algorithm

- Tetraploid example, one individual, 15 markers

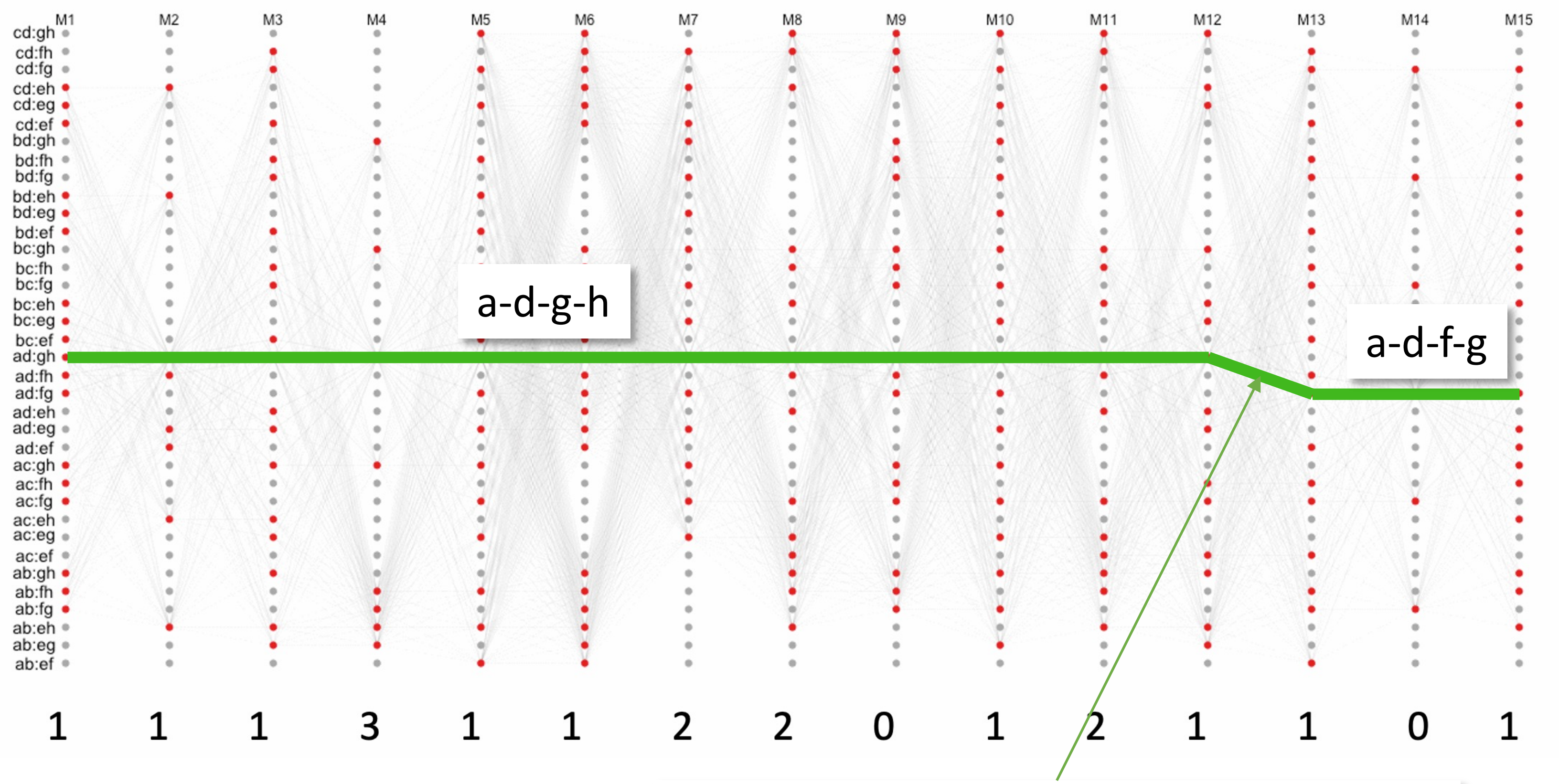


1 1 1 3 1 1 2 2 0 1 2 1 1 0 1

Hidden Markov Model - HMM



- Tetraploid example, one individual, 15 markers



a-d-g-h

a-d-f-g

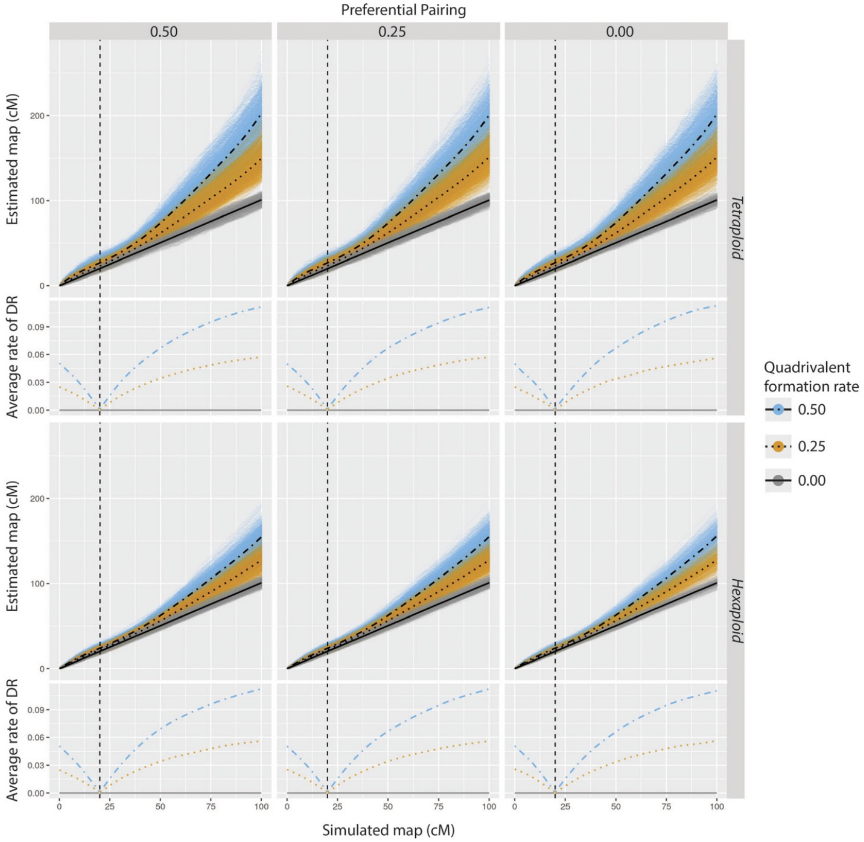
Crossing over between homologs **f** and **h**

Robustness of the HMM

(under preferential pairing and double reduction)

Table 2 Percentage of data sets where linkage phase configuration was correctly estimated for parents *P* and *Q* in simulation 2

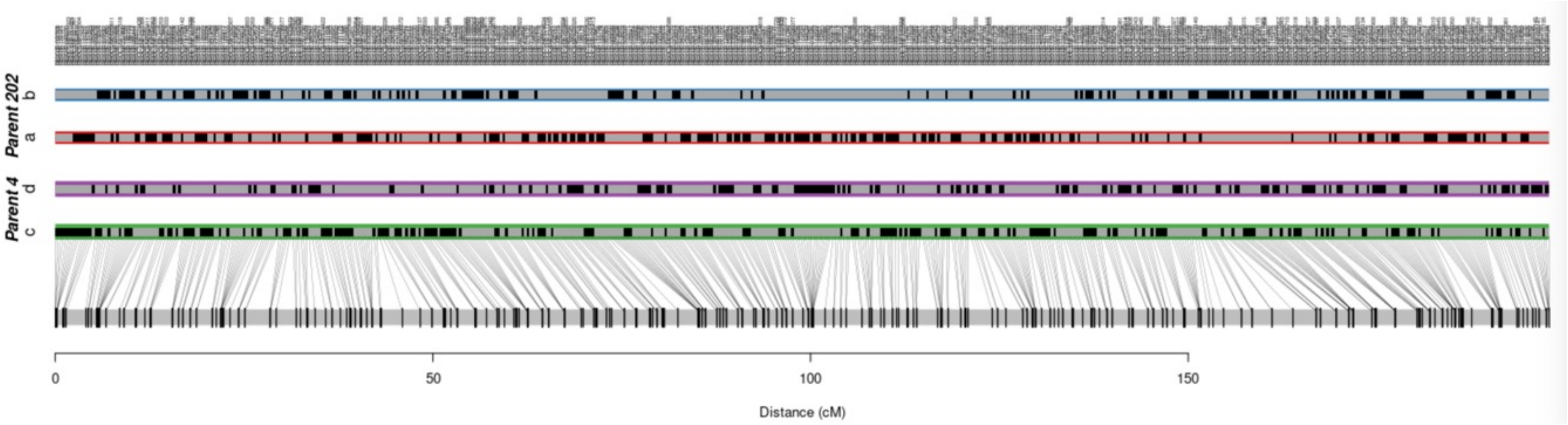
Preferential pairing		0.00			0.25			0.50		
Quadrivalent formation		0.00	0.25	0.50	0.00	0.25	0.50	0.00	0.25	0.50
Autotetraploid										
$\eta = 3$	*P*	100.0	99.0	91.5	98.5	98.5	90.0	80.5	93.0	87.5
		100.0	99.5	99.5	98.5	99.5	97.5	57.5	88.5	97.0
		99.5	97.5	98.5	100.0	98.5	94.0	55.0	85.5	94.5
		100.0	100.0	99.5	99.0	98.0	98.0	60.5	86.5	93.0
		99.5	99.5	97.0	98.5	97.0	95.5	67.5	84.5	97.5
	Q	100.0	98.5	90.0	100.0	97.0	90.0	60.0	91.5	86.0
		100.0	100.0	98.0	99.5	100.0	99.0	65.0	89.0	93.5
		100.0	98.5	98.0	97.0	98.5	94.5	41.0	82.0	93.5
		100.0	100.0	99.0	99.5	98.0	98.0	56.5	84.5	90.0
		99.5	99.5	98.0	99.0	98.5	94.5	58.0	82.0	94.0
$\eta = 5$	*P*	100.0	99.5	93.0	100.0	99.5	95.0	98.0	99.0	95.0
		100.0	100.0	100.0	100.0	100.0	100.0	90.0	99.5	99.0
		100.0	99.5	100.0	100.0	100.0	99.5	86.0	98.5	100.0
		100.0	100.0	100.0	99.5	100.0	99.5	86.5	98.5	100.0
		100.0	100.0	100.0	100.0	100.0	100.0	90.5	96.0	100.0
	Q	100.0	99.5	93.0	100.0	99.0	94.0	88.0	98.5	95.5
		100.0	100.0	100.0	100.0	100.0	100.0	91.5	99.5	99.5
		100.0	99.5	100.0	99.5	100.0	99.0	85.0	98.0	100.0
		100.0	100.0	100.0	99.5	100.0	99.5	86.0	97.5	98.5
		100.0	100.0	99.5	100.0	100.0	100.0	92.0	96.0	99.0
Autohexaploid										
$\eta = 3$	*P*	84.0	78.5	70.5	69.0	63.5	61.0	2.5	10.5	19.0
		99.0	94.0	91.0	93.0	84.5	80.0	6.5	16.0	22.0
		89.0	94.0	88.0	80.0	84.0	80.5	10.5	16.0	32.5
		93.0	90.5	86.0	88.5	84.0	80.0	9.0	16.5	28.5
		96.0	92.5	91.5	89.5	94.0	87.5	19.0	30.5	44.5
	Q	85.0	81.0	71.0	68.0	52.5	57.5	1.5	3.5	8.5
		99.0	95.0	91.0	86.5	90.0	88.5	9.0	28.0	37.5
		90.0	90.0	86.0	79.0	82.0	77.0	9.5	18.0	28.0
		96.5	92.5	89.5	90.0	89.0	89.0	25.5	35.5	41.0
		95.0	92.0	92.5	89.5	91.0	88.0	16.0	23.0	39.0
$\eta = 5$	*P*	86.0	84.5	75.5	77.5	69.5	72.5	27.0	36.5	52.5
		100.0	97.5	96.5	98.5	98.0	91.0	55.5	70.5	74.5
		91.5	95.5	93.0	90.5	94.5	89.5	68.0	68.5	77.5
		96.5	94.0	91.0	99.5	99.0	96.5	65.0	78.5	85.0
		98.0	98.5	100.0	97.5	99.0	99.0	73.0	87.5	91.0
	Q	86.5	83.5	75.0	69.5	68.5	72.0	17.5	20.0	39.5
		100.0	99.5	99.0	100.0	99.5	100.0	74.0	81.0	92.5
		91.5	95.5	93.0	91.0	95.0	89.5	67.5	71.5	77.0
		99.0	97.5	93.5	100.0	100.0	99.5	80.0	89.0	92.0
		98.0	98.5	100.0	97.5	99.0	99.0	83.0	83.0	90.5

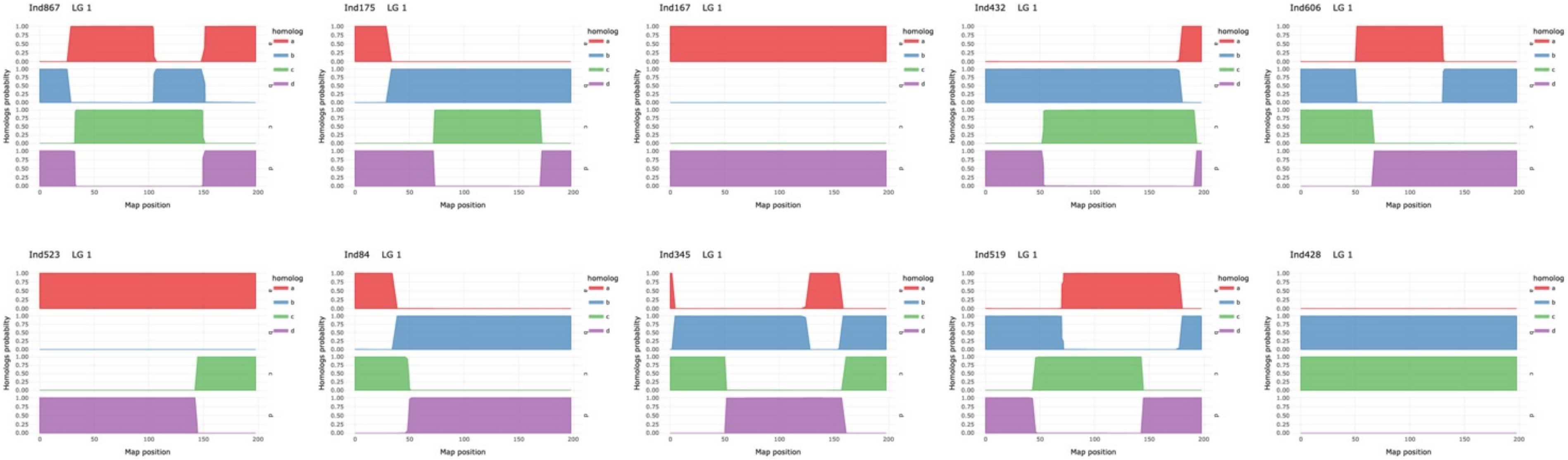


- Phase estimation was affected by preferential pairing
- RF were overestimated in the presence of quadrivalent formation.

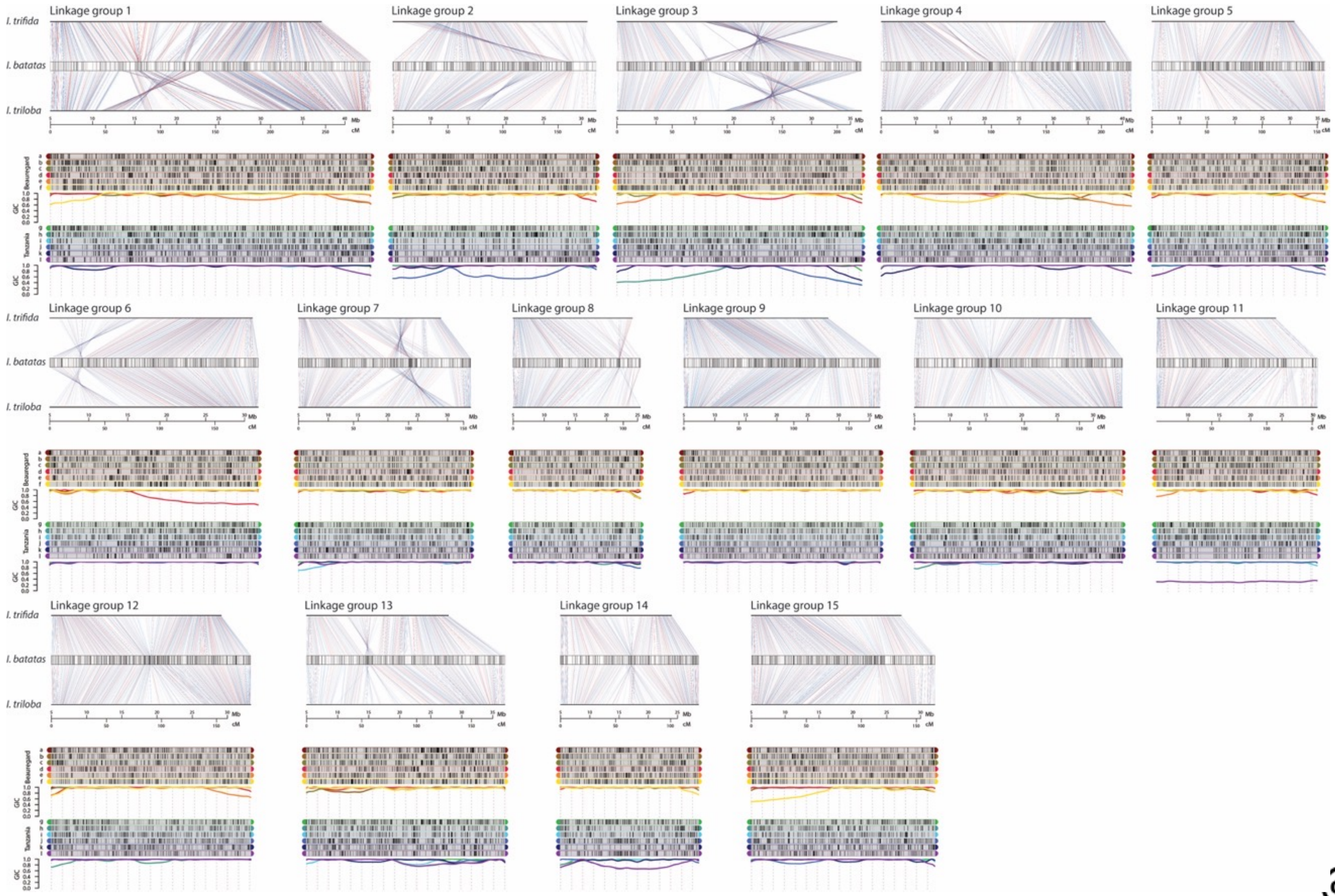
Pinus - Haplotypic composition

chromosome 1, 10 individuals

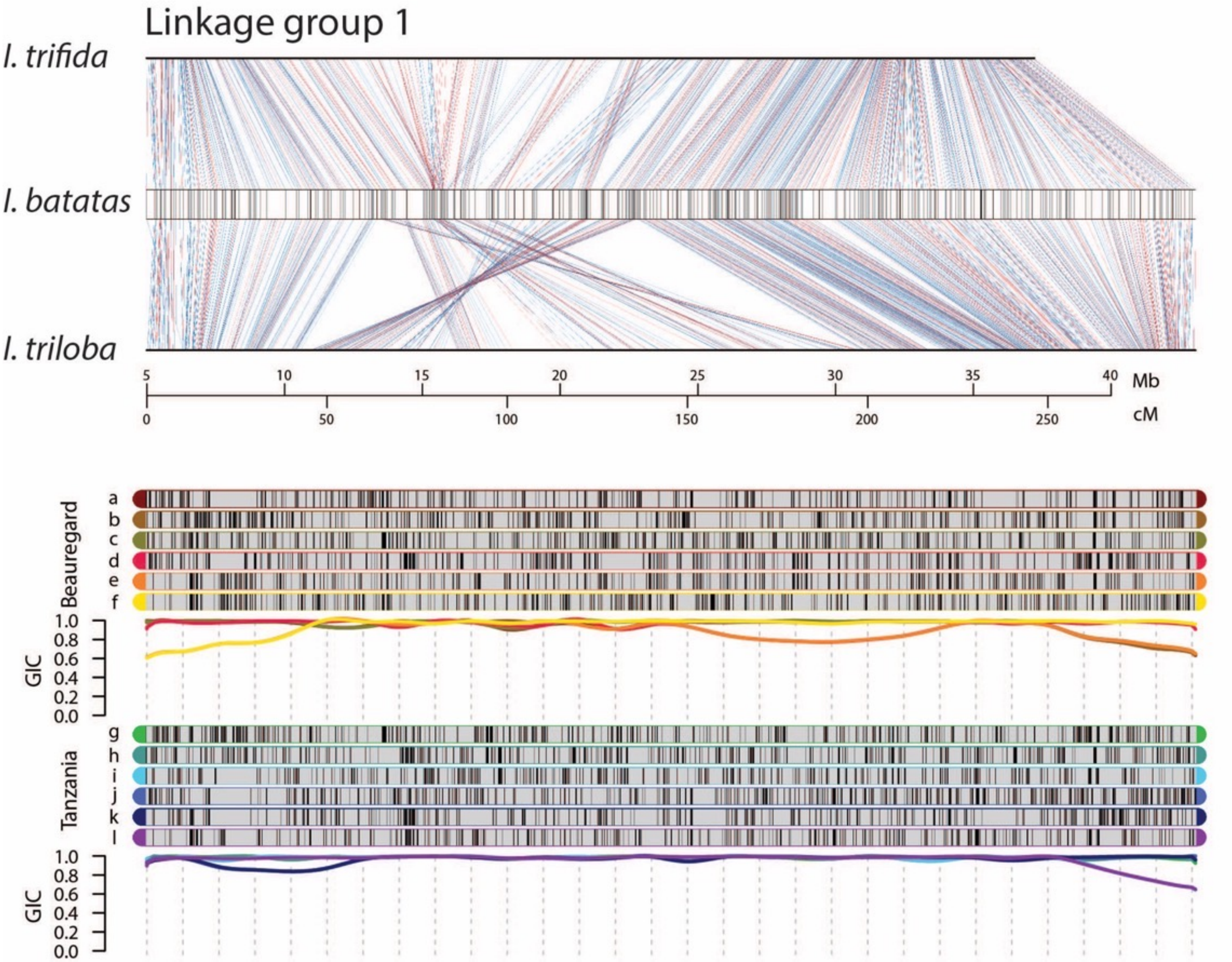




Sweetpotato genetic map



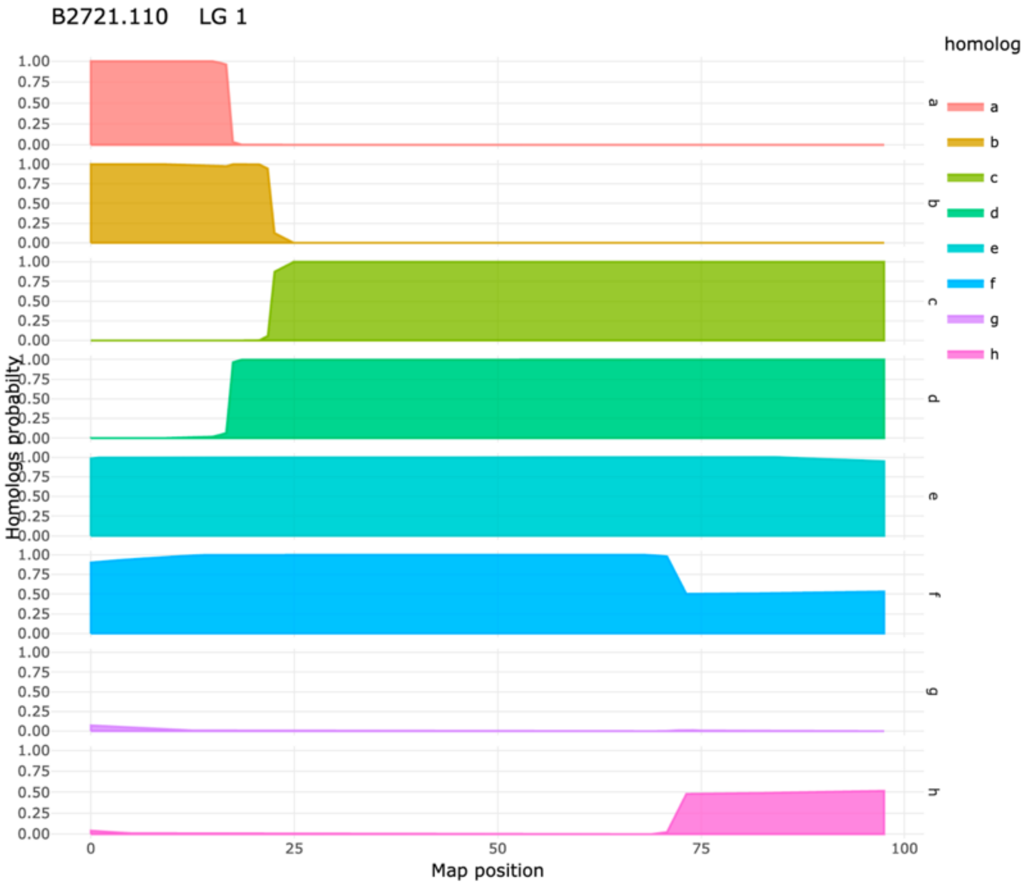
Sweetpotato genetic map



Probabilistic haplotype reconstruction

- When assuming a *prior probability* distribution of the genotypes, multilocus strategies can improve the quality of the inferred haplotypes

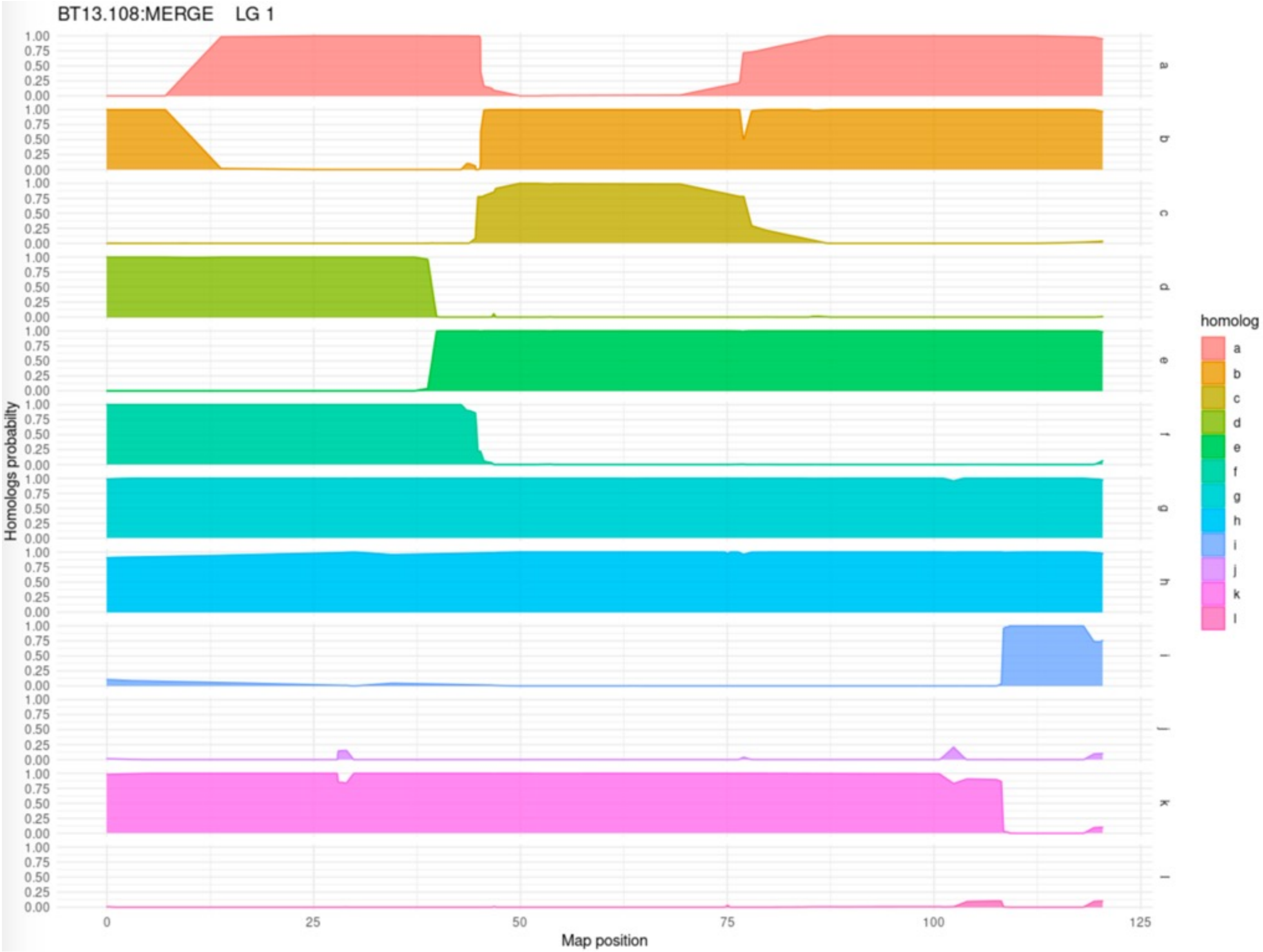
Tetraploid potato



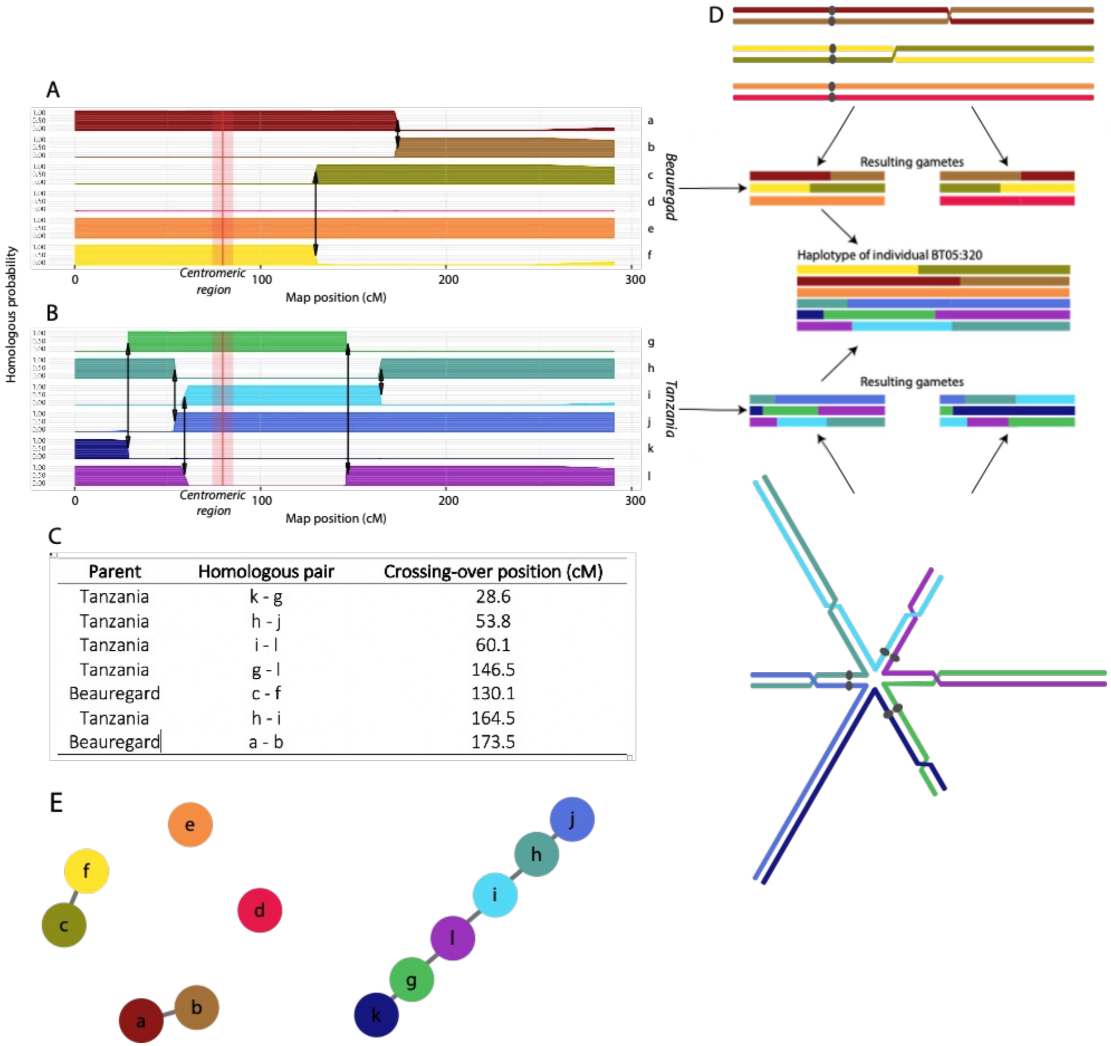
cc

Probabilistic haplotype reconstruction

Hexaploid sweetpotato

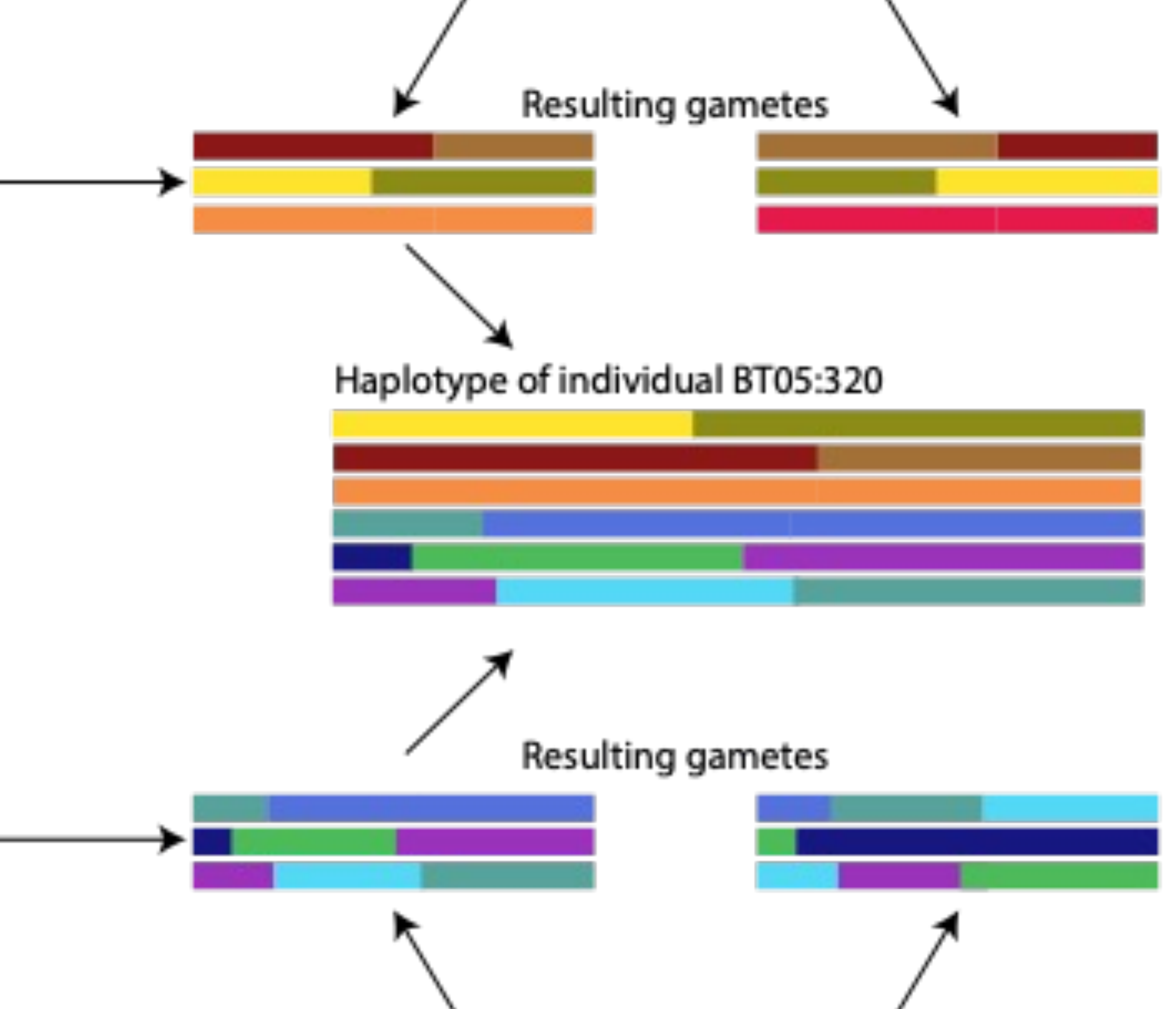


Haplotype reconstruction in the offspring

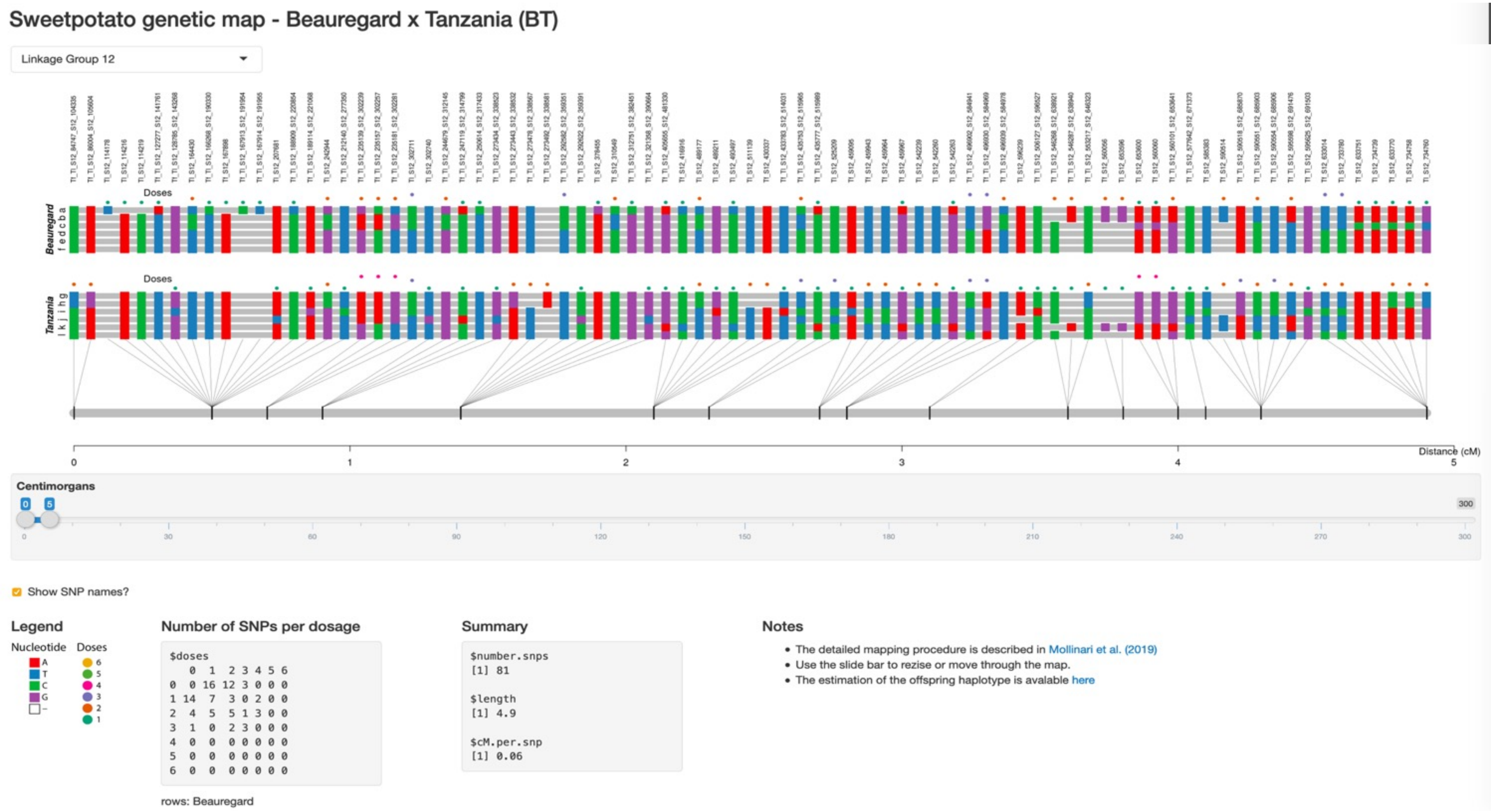


Parent	Homologous pair	Crossing-over position (cM)
Tanzania	k - g	28.6
Tanzania	h - j	53.8
Tanzania	i - l	60.1
Tanzania	g - l	146.5
Beauregard	c - f	130.1
Tanzania	h - i	164.5
Beauregard	a - b	173.5

https://gt4sp-genetic-map.shinyapps.io/offspring_haplotype_BT_population/


Resulting gametes
Haplotype of individual BT05:320
Resulting gametes

Genetic mapping – Linkage group 12 – 2661 SNPs



Sweetpotato genetic map - Beauregard x Tanzania (BT)

☑ Show SNP names?

Legend

Nucleotide
- 🟥 A
- 🟦 T
- 🟩 C
- 🟪 G
- ☐ –

Doses
- ● 6
- ● 5
- ● 4
- ● 3
- ● 2
- ● 1

Number of SNPs per dosage

```
$doses
   0  1  2 3 4 5 6
0  0 16 12 3 0 0 0
1 14  7  3 0 2 0 0
2  4  5  5 1 3 0 0
3  1  0  2 3 0 0 0
4  0  0  0 0 0 0 0
5  0  0  0 0 0 0 0
6  0  0  0 0 0 0 0
```

rows: Beauregard

Summary

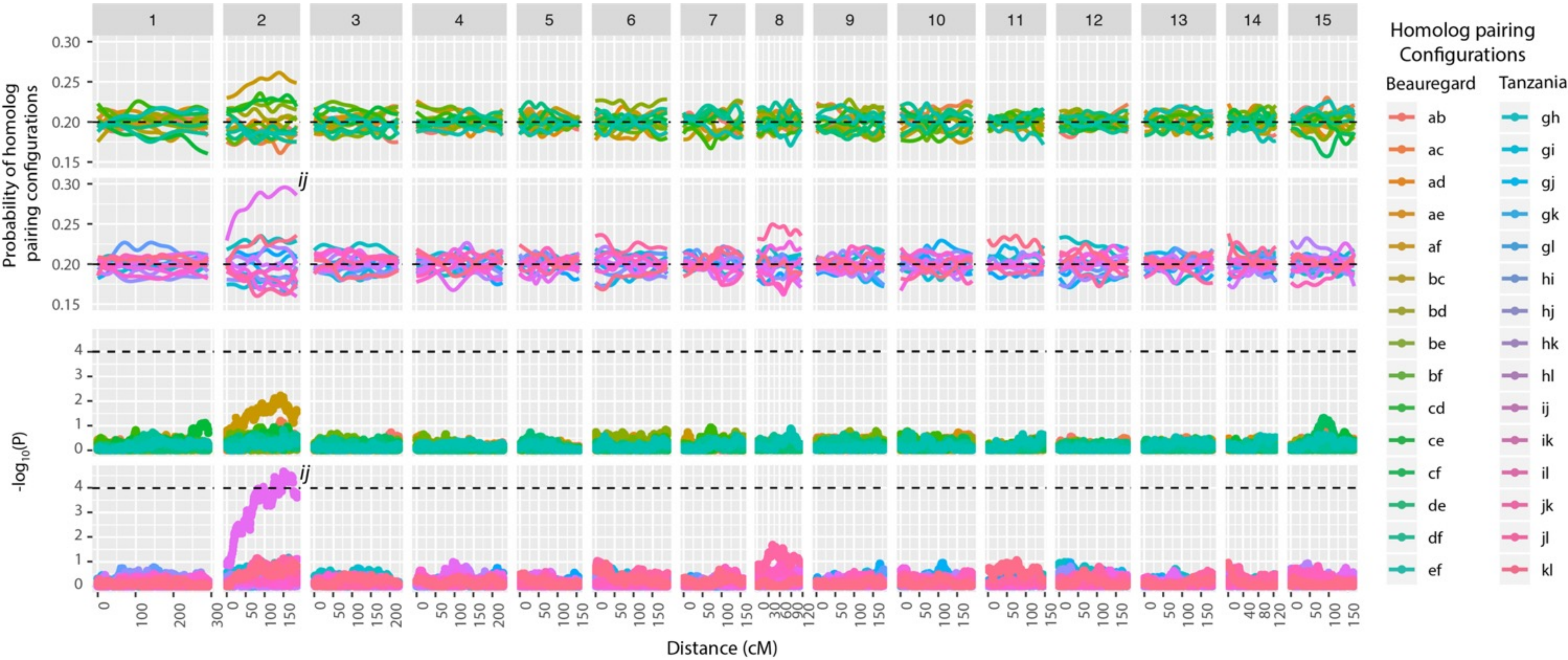
```
$number.snps
[1] 81

$length
[1] 4.9

$cM.per.snp
[1] 0.06
```

Notes

- The detailed mapping procedure is described in Mollinari et al. (2019)
- Use the slide bar to rezise or move through the map.
- The estimation of the offspring haplotype is avalable here

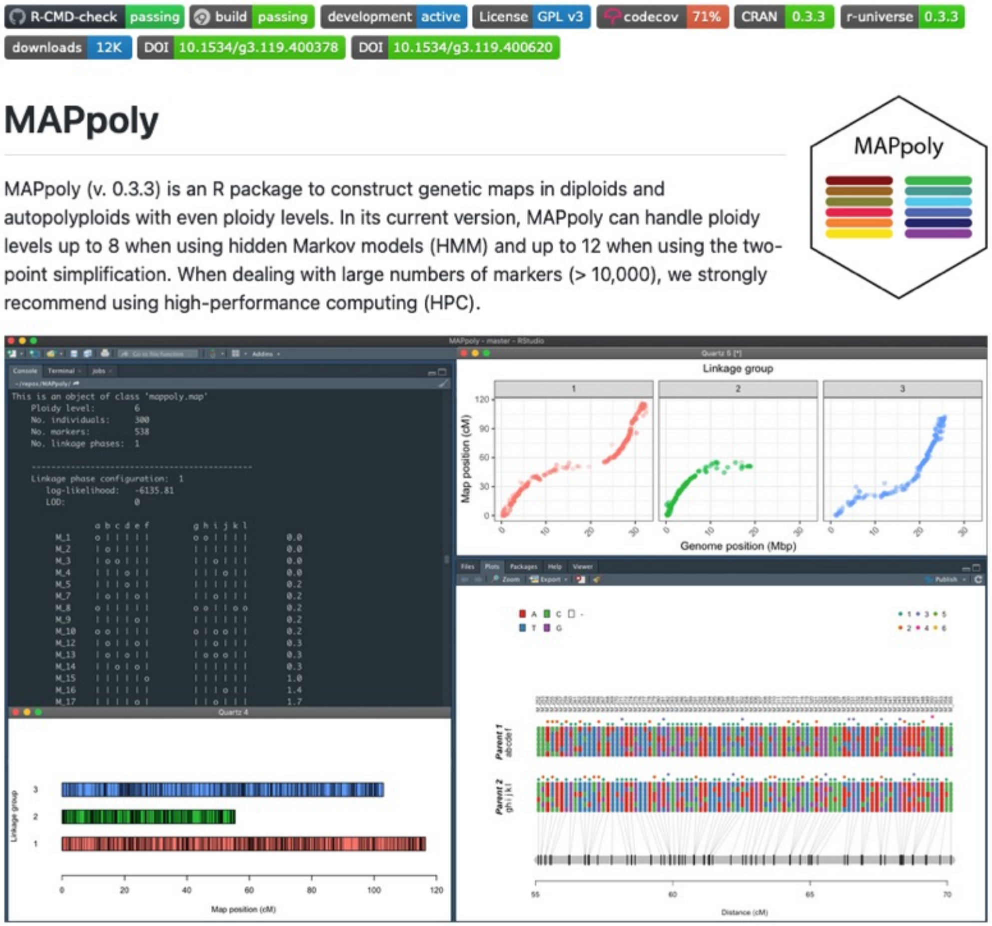
Interactive version: https://gt4sp-genetic-map.shinyapps.io/bt_map/

Preferential pairing profiles: Sweetpotato is vastly autohexaploid



© Zhao-Bang Zeng @ NC State

Software to construct genetic maps



R-CMD-check passing | build passing | development active | License GPL v3 | codecov 71% | CRAN 0.3.3 | r-universe 0.3.3
downloads 12K | DOI 10.1534/g3.119.400378 | DOI 10.1534/g3.119.400620

MAPpoly

MAPpoly (v. 0.3.3) is an R package to construct genetic maps in diploids and autopolyploids with even ploidy levels. In its current version, MAPpoly can handle ploidy levels up to 8 when using hidden Markov models (HMM) and up to 12 when using the two-point simplification. When dealing with large numbers of markers (> 10,000), we strongly recommend using high-performance computing (HPC).

```
This is an object of class 'mappoly.map'
    Ploidy level:        6
    No. individuals:     300
    No. markers:         538
    No. linkage phases:  1

    ---------------------------------------------
    Linkage phase configuration:  1
       log-likelihood:   -6135.81
       LOD:              0

             a b c d e f      g h i j k l
       M_1   o | | | | |      o o | | | |      0.0
       M_2   | o | | | |      | | | | | |      0.0
       M_3   | o o | | |      | | o | | |      0.0
       M_4   | | | o | |      | | | o | |      0.0
       M_5   | | | o | |      | | | | | |      0.2
       M_7   | o | | o |      | | o | | |      0.2
       M_8   o | | | | |      o o | | o o      0.2
       M_9   | | | | o |      | | | | | |      0.2
       M_10  o o | | | |      o | o o | |      0.2
       M_12  | o | | o |      | | o | | |      0.3
       M_13  | o | o | |      | o o o | |      0.3
       M_14  | | o | o |      | | | | | |      0.3
       M_15  | | | | | o      | | | | | |      1.0
       M_16  | | | | | |      | | | o | |      1.4
       M_17  | | | | o |      | | o | | |      1.7
```

- MAPMAKER/EXP
- Onemap/BatchMap
- GUSmap
- JoinMap
- CarthaGene
- R/QTL
- MSTmap
- polymapR
- …

Development version and other resources: https://github.com/mmollina/MAPpoly
Stable version: https://cran.r-project.org/package=mappoly

References

- Haldane, J. Theoretical Genetics of Autopolyploids. *J. Genet.* **22**, 359–372 (1930).
- Mather, K. Reductional and equational separation of the chromosomes in bivalents and multivalents. *J. Genet.* **30**, 53–78 (1935).
- Mather, K. Segregation and linkage in autotetraploids. *J. Genet.* **32**, 287–314 (1936).
- Mather, K. *The measurement of linkage in heredity*. (1938).
- Fisher, R. A. The theory of linkage in polysomic inheritance. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **233**, 55–87 (1947).
- Ripol, M. I., Churchill, G. A., Silva, J. A. G. Da & Sorrells, M. Statistical aspects of genetic mapping in autopolyploids. *Gene* **235**, 31–41 (1999).
- Luo, Z. W., Zhang, R. M. & Kearsey, M. J. Theoretical basis for genetic linkage analysis in autotetraploid species. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 7040–7045 (2004).
- Leach, L. J., Wang, L., Kearsey, M. J. & Luo, Z. Multilocus tetrasomic linkage analysis using hidden Markov chain model. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 4270–4274 (2010).
- Hackett, C. a., McLean, K. & Bryan, G. J. Linkage Analysis and QTL Mapping Using SNP Dosage Data in a Tetraploid Potato Mapping Population. *PLoS One* **8**, (2013).
- Zheng, C. *et al.* Probabilistic Multilocus Haplotype Reconstruction in Outcrossing Tetraploids. *Genetics* **203**, 119–131 (2016).
- Bourke, P. M. Genetic mapping in polyploids. (Wageningen University, 2018).
- Mollinari, M. & Garcia, A. A. F. Linkage Analysis and Haplotype Phasing in Experimental Autopolyploid Populations with High Ploidy Level Using Hidden Markov Models. *G3; Genes|Genomes|Genetics* **9**, 3297–3314 (2019).