New methods for construction of genetic linkage maps in hexaploid sweetpotato

Marcelo Mollinari

mmollin@ncsu.edu

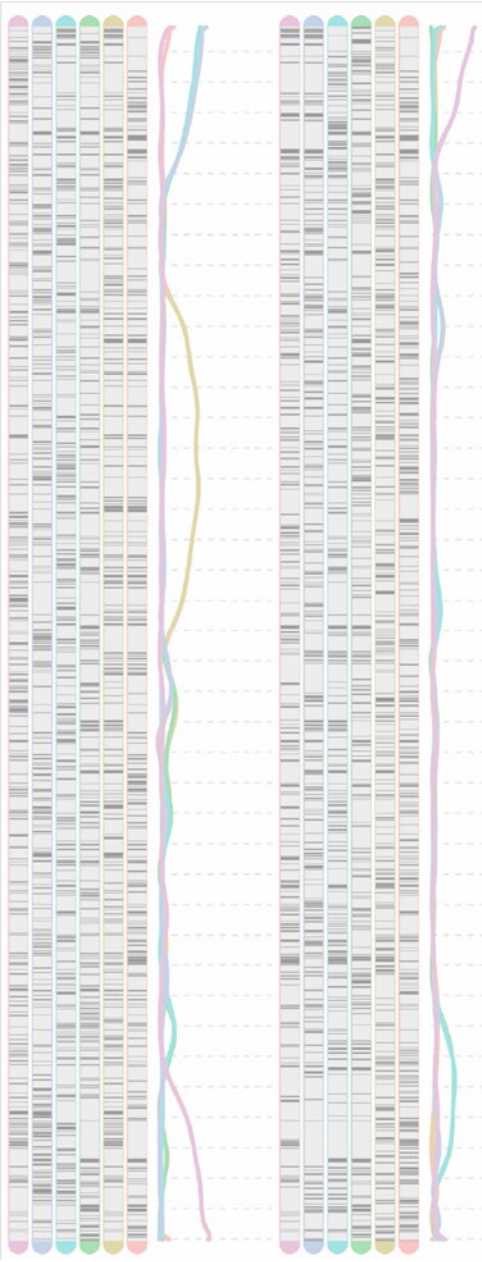
GT4SP

GENOMIC TOOLS FOR SWEETPOTATO IMPROVEMENT

BILL & MELINDA GATES foundation

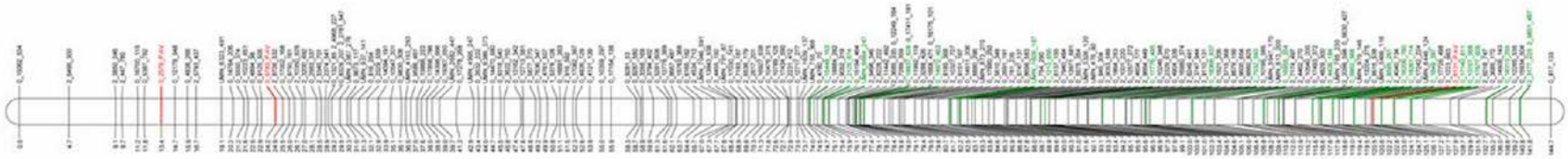
Overview

- Why do we need genetic maps?
- The general concept of genetic mapping in diploids.
  - Accessing genetic variation.
  - Genetic distance.
- Polyploid species: why are they different?
- Genotyping polyploid species.
- Genetic mapping in polyploid species.
  - The phasing problem: How to find the origin of haplotypes?
  - Recombination fraction estimation.
- Building a sweetpotato map.

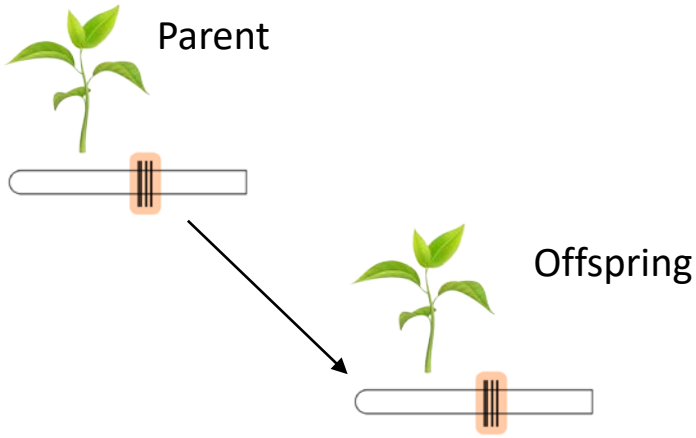


Why do we need a genetic map?

- Genetic maps are linear arrangements of markers representing chromosomes of a species

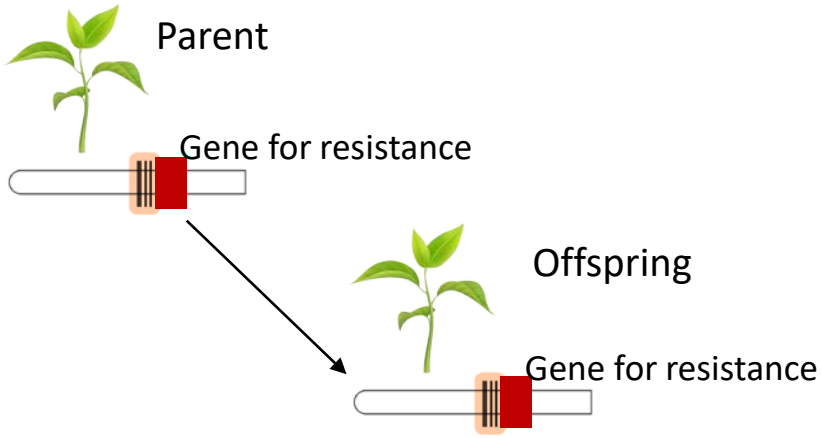


- Inheritance history of the mapping population: close markers are inherited together



Why do we need a genetic map?

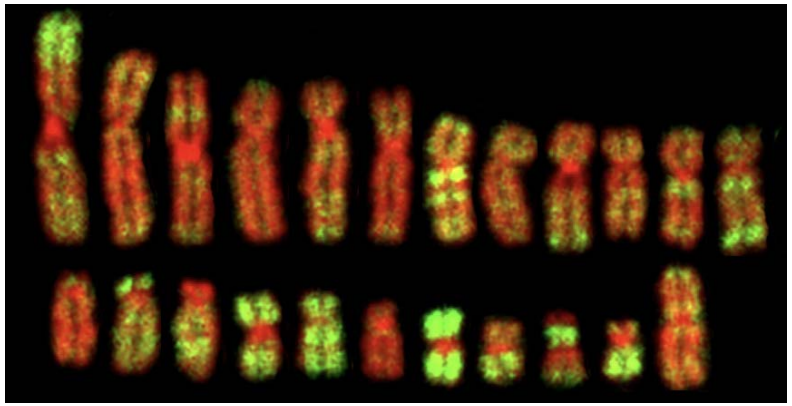
- We use markers as proxy to access important traits such asvirus and weevil resistance, beta-carotene and yield.



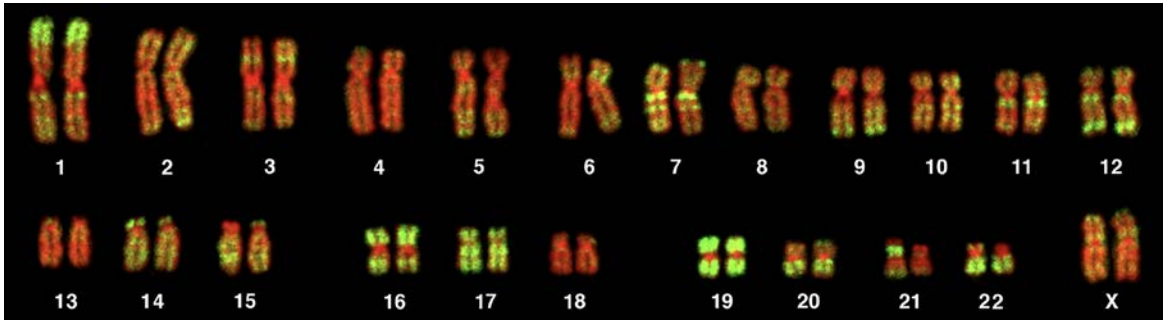
- Since we have thousands of markers spread across the genome, we try to find associations between them and traits measured in the field. The final goal, is to select better genotypes based on the markers. This can speed up the breeding process.

The general concept of genetic mapping:
diploid organisms

- Basic chromosome number: the number of different chromosomes that make up a single complete set. In humans, this number is 23.



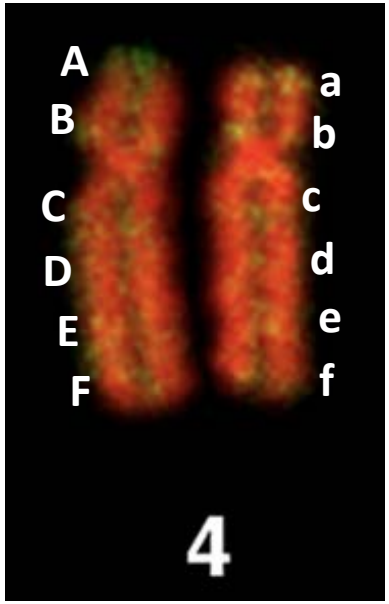
- Ploidy level: Number of basic chromosome sets. In humans, the ploidy levels is two. Thus we are diploids.

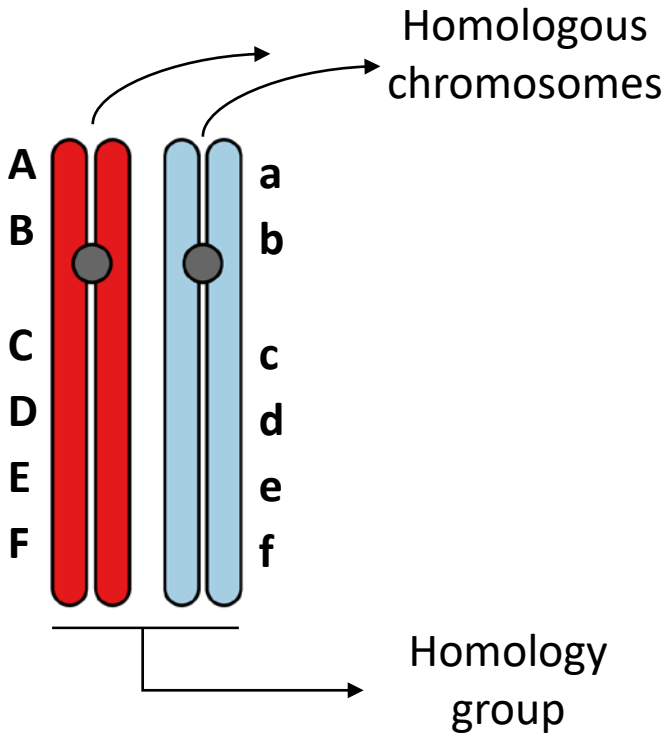


Accessing genetic variation.

- Genotyping: measurement of variations (alleles) in homologous chromosomes within a locus

- Molecular markers: access the allelic variation in each one of the homologous chromosomes

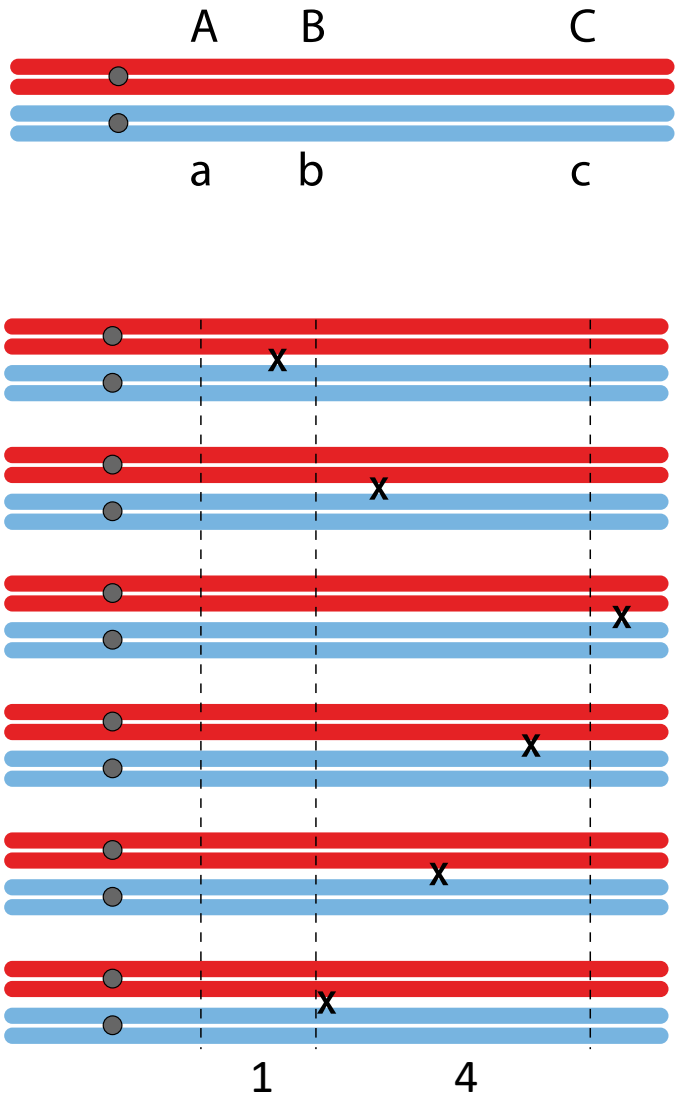




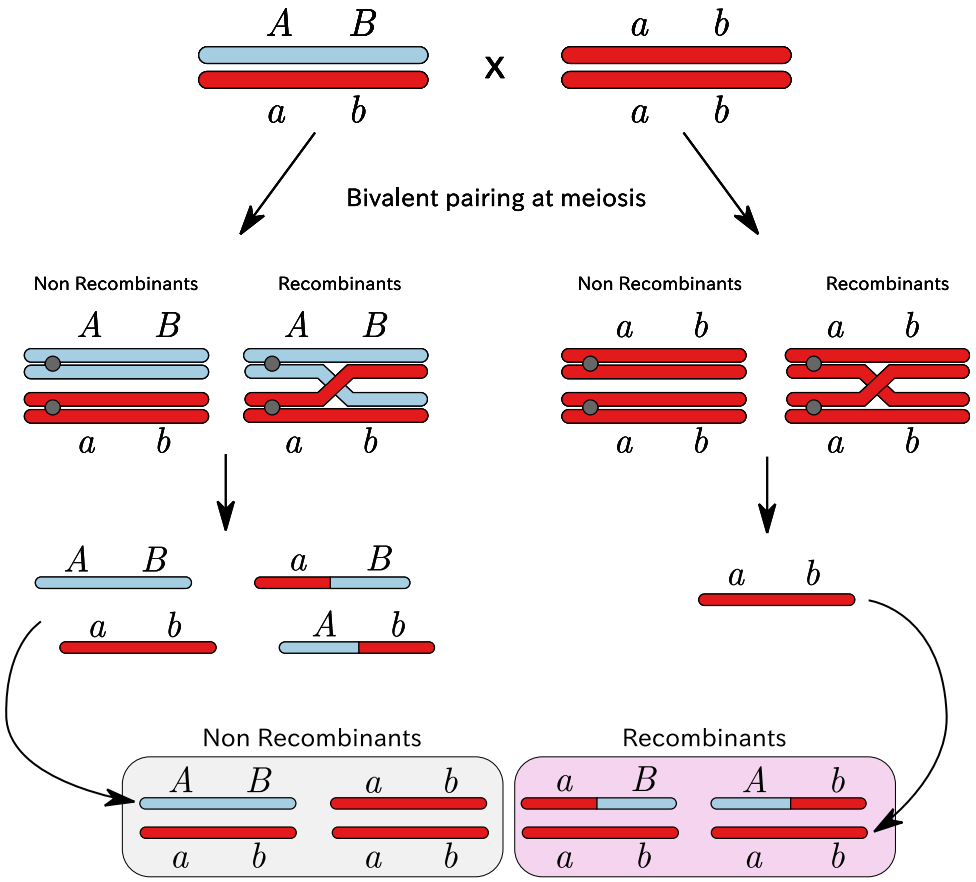
The general concept of genetic mapping:
the crossing over process

The general concept of genetic mapping

- Suppose we have a chromosome with three markers, two of them are close to each other (A and B) and the third (C) is further away.
- The process described in the video will happen several times and the gametes will be stored in pollen and ovule cells
- The number of crossing overs detected between two points will be proportional to their physical distance.



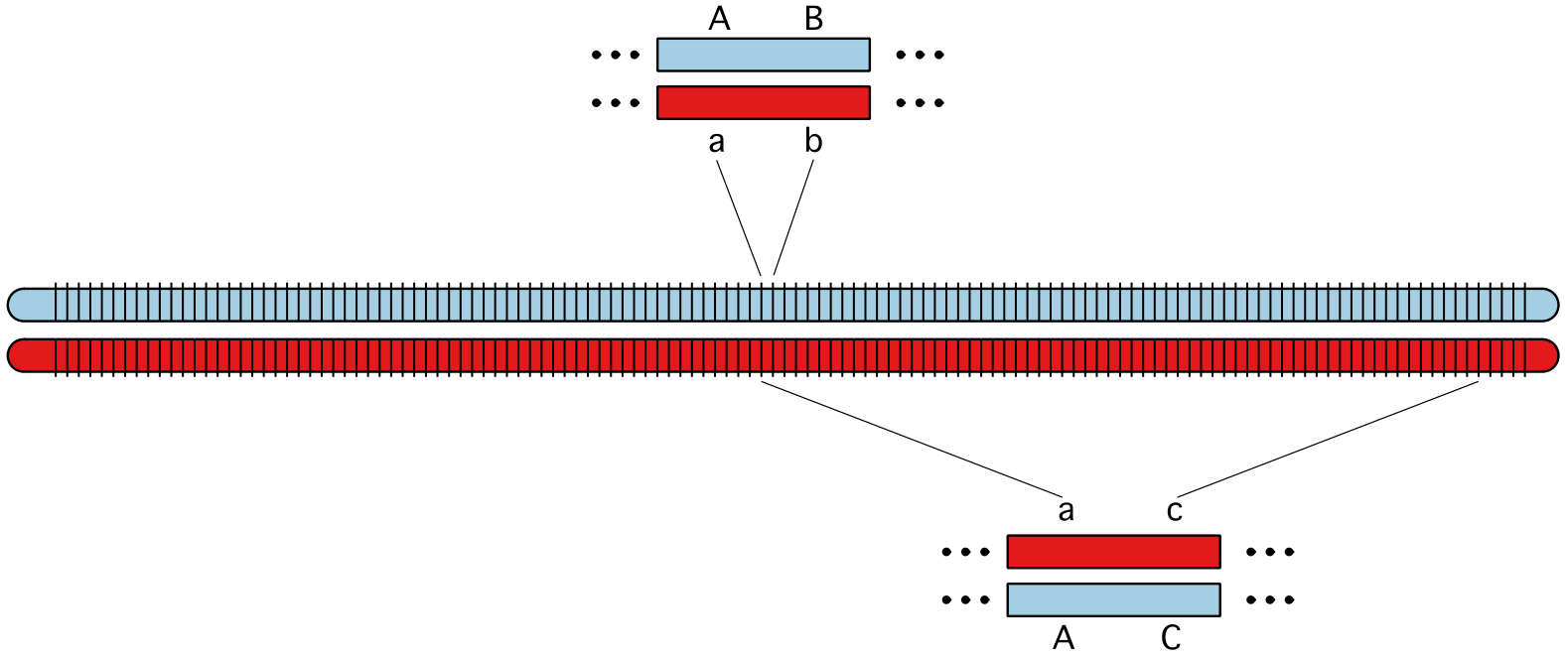
The general concept of genetic mapping



$$\widehat{r} = \frac{\#\text{recombinants}}{\#\text{total}}$$

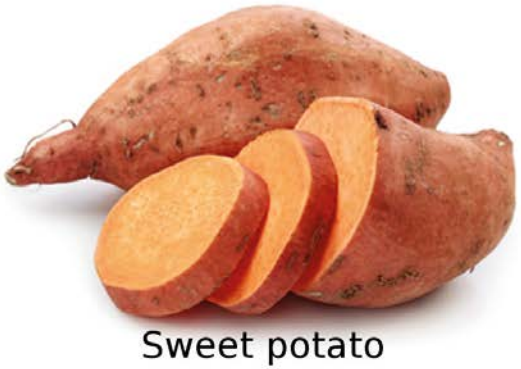
The general concept of genetic mapping

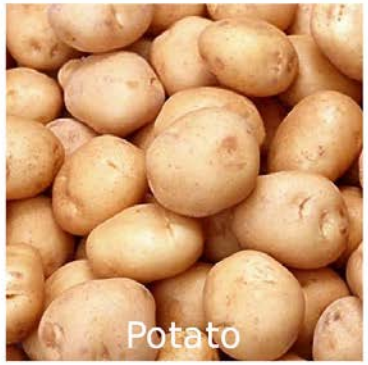
- The two markers presented are part of a bigger picture. However, the reasoning of count crossing-overs works for any pair of markers

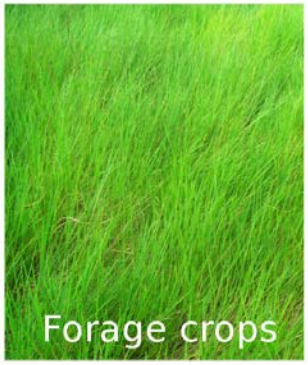


Polyploid species

Examples of polyploid crops

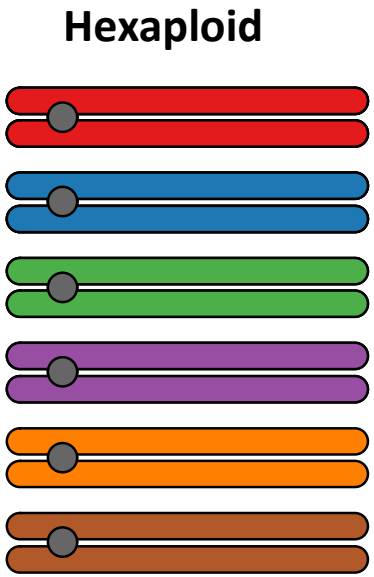






Polyploid species: why are they different?

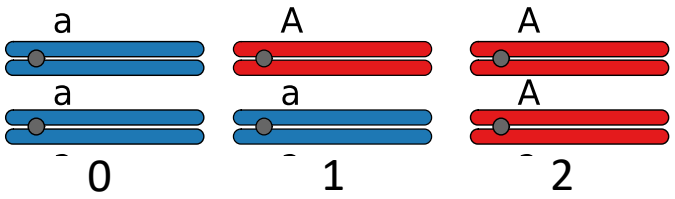
- Polyploids are organisms with multiple sets of chromosomes.



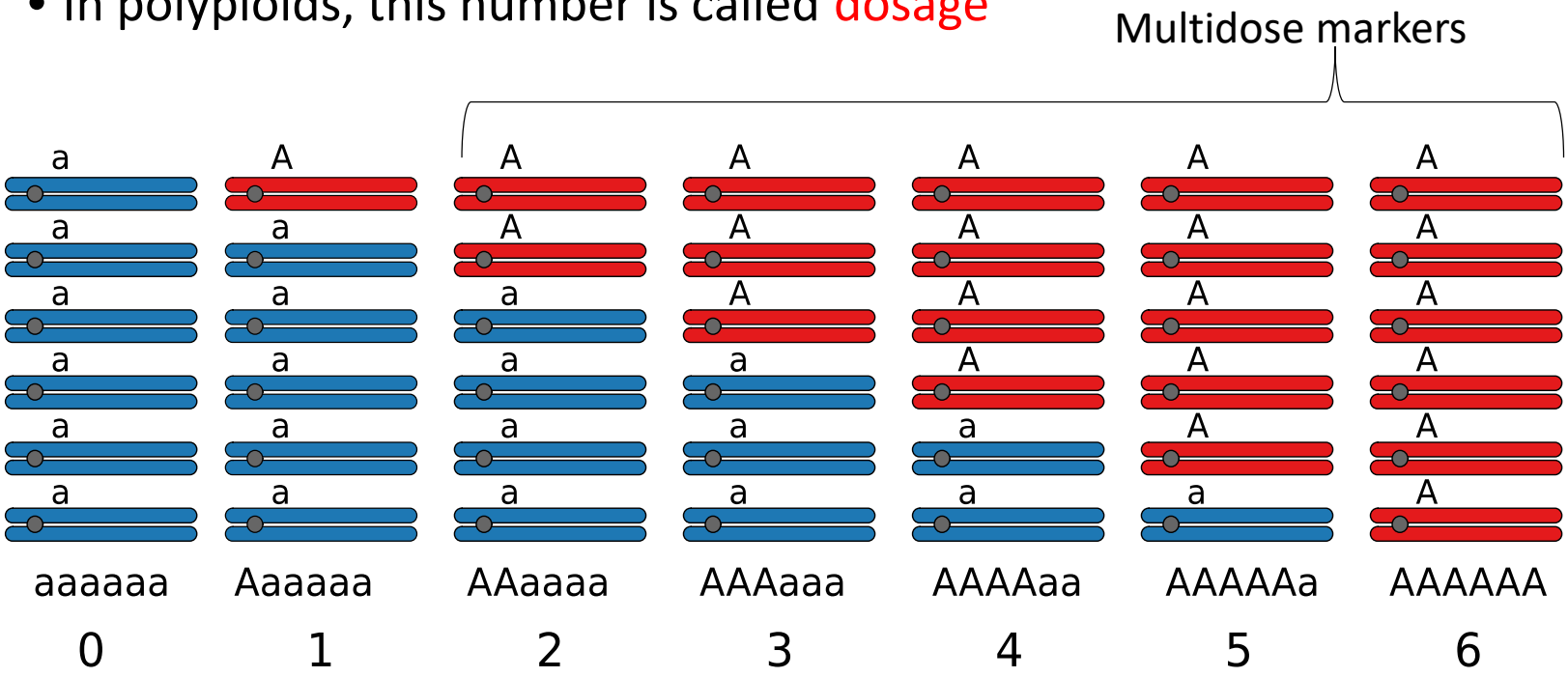
- Two problems:
  - How to access the genetic variation?
  - How to count crossing overs?

Genotyping polyploid species

- In diploids, we usually refer to the genotypes using the quantity of a reference alleles, in this case, A

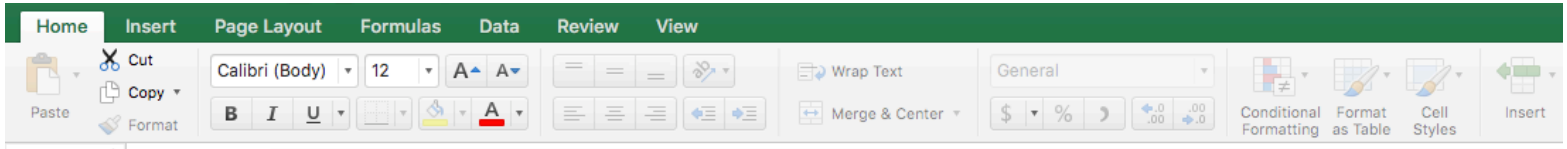


- In polyploids, this number is called dosage



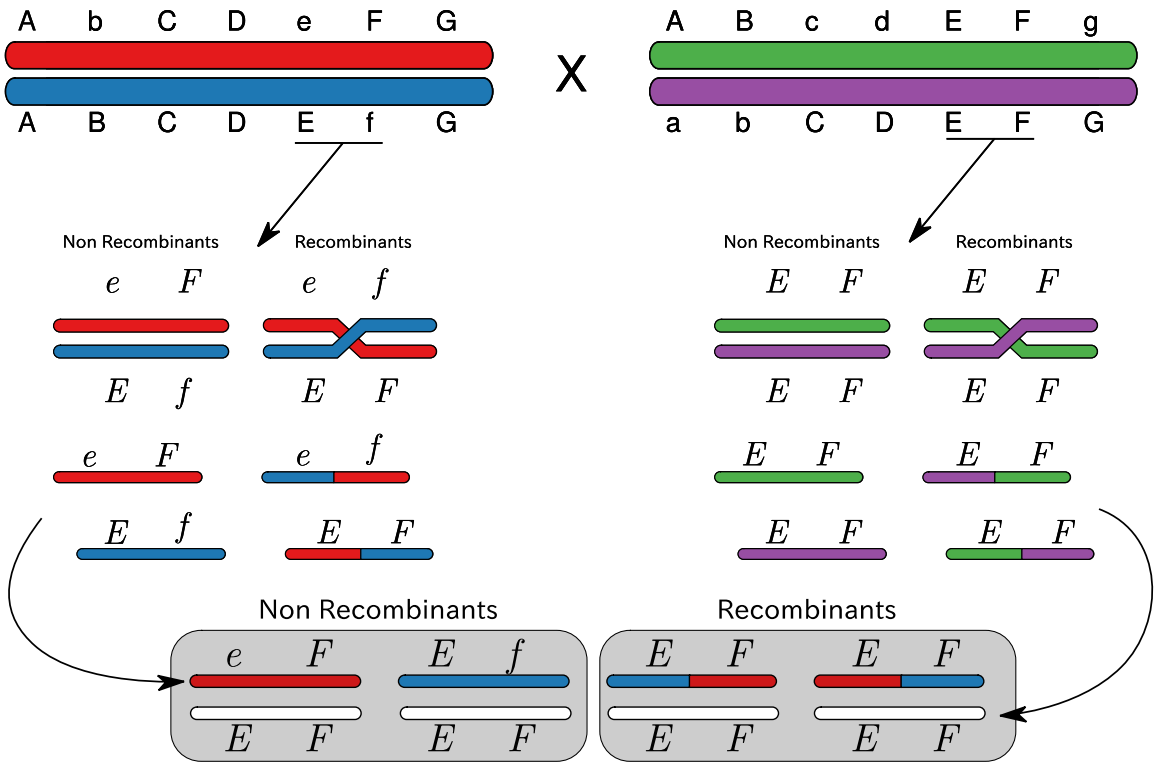
Genotyping polyploid species

- Example of input file for *MAPPoly* software



	A	B	C	D	E	F	G	H	I	J	K
1		Ind_1	Ind_2	Ind_3	Ind_4	Ind_5	Ind_6	Ind_7	Ind_8	Ind_9	Ind_10
2	M_1	1	2	1	2	2	1	1	2	2	1
3	M_2	0	1	0	1	1	1	0	0	1	1
4	M_3	2	2	1	2	2	2	1	0	1	2
5	M_4	1	0	1	0	1	1	2	1	0	1
6	M_5	0	0	0	0	0	1	1	1	0	1
7	M_6	0	0	0	0	0	1	1	1	0	1
8	M_7	2	1	1	1	1	3	0	1	2	1
9	M_8	2	4	2	4	3	2	3	3	4	3
10	M_9	1	0	1	0	0	1	0	1	1	0
11	M_10	3	2	2	3	3	3	2	1	3	1
12	M_11	1	0	1	0	0	1	0	1	1	0
13	M_12	2	1	1	1	1	3	0	1	2	1
14	M_13	2	2	1	1	3	3	2	2	1	3
15	M_14	2	1	2	1	1	1	1	1	1	1
16	M_15	0	0	1	0	0	0	0	1	0	0

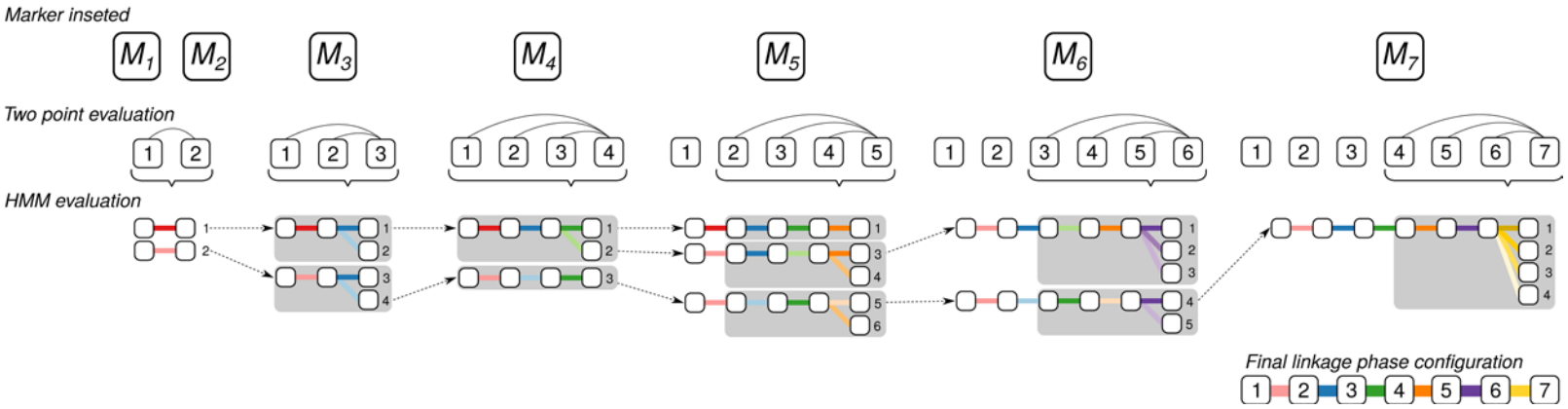
The importance of haplotype phasing in polyploids



- In this case, we know how the alleles are arranged in parental chromosomes. Thus, we can distinguish recombinant and non recombinant gametes.
- If this information is not available, we can infer that gametes e-F and E-f are non recombinants, since they will be more frequent than E-F.

Haplotype phasing and assembly of homology groups

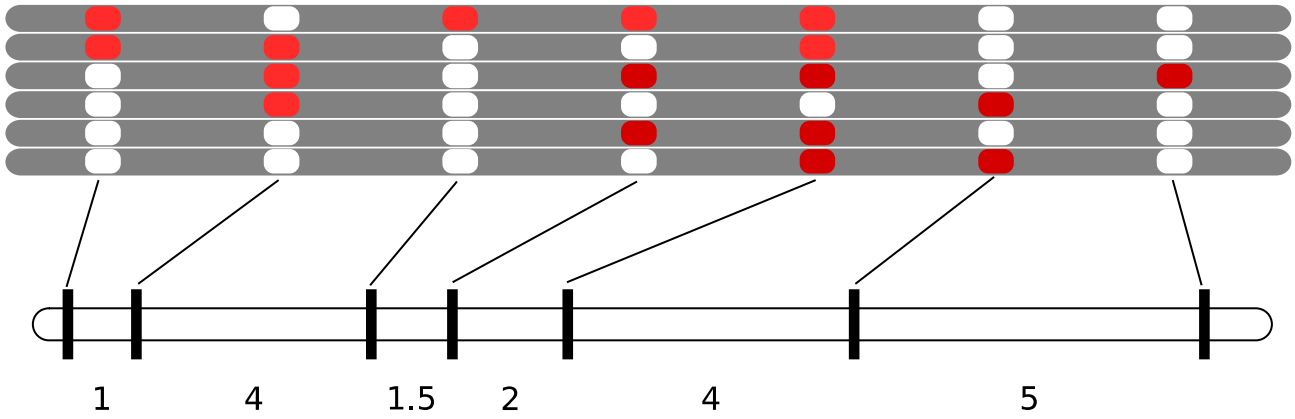
- In polyploids, we do not know how alleles are arranged in parental chromosomes
- We know the dosage, but we need to infer the linkage phase configuration of the alleles



A	b	C	D	E	f	g
A	B	c	d	E	f	g
a	B	c	D	E	f	G
a	B	c	d	e	F	g
a	b	c	D	E	f	g
a	b	c	d	E	F	g
2	3	1	3	5	2	1

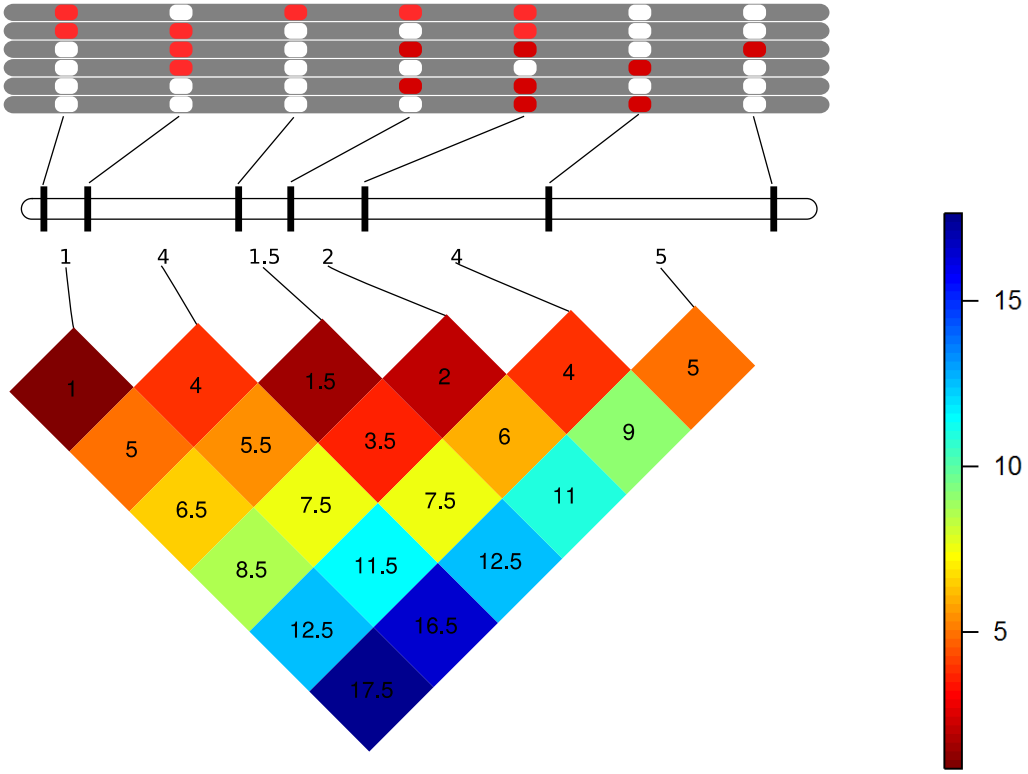
Recombination fraction estimation

- Once we have the linkage phase configuration, we can distinguish between recombinant and parental gametes, and count the number of crossing over between markers.
- The procedure is not as straightforward as the one used in diploids, but the idea is the same



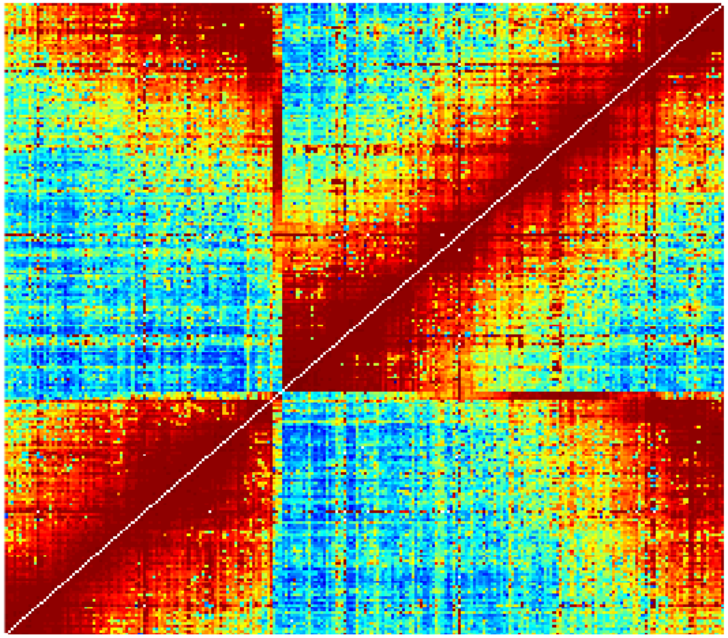
Recombination fraction matrix

- Important way to diagnose problems in genetic maps



Recombination fraction matrix

- Important diagnostic tool way to find problems in genetic maps



Map with a inversion

Map without inversion

Biparental Population - BT

- Beauregard x Tanzania
- 315 individuals, 26164 high quality SNPs

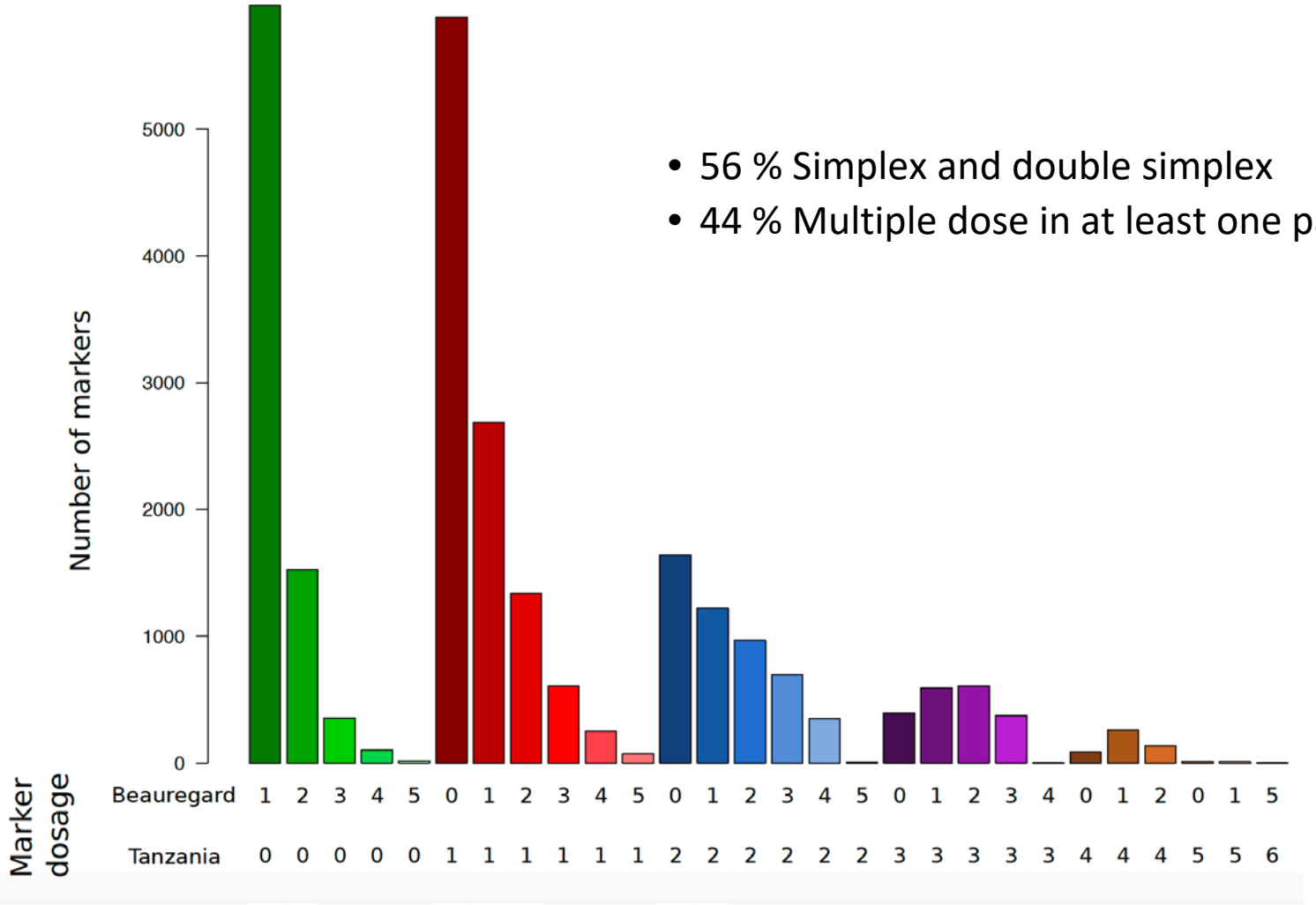
X

Beauregard

Tanzania

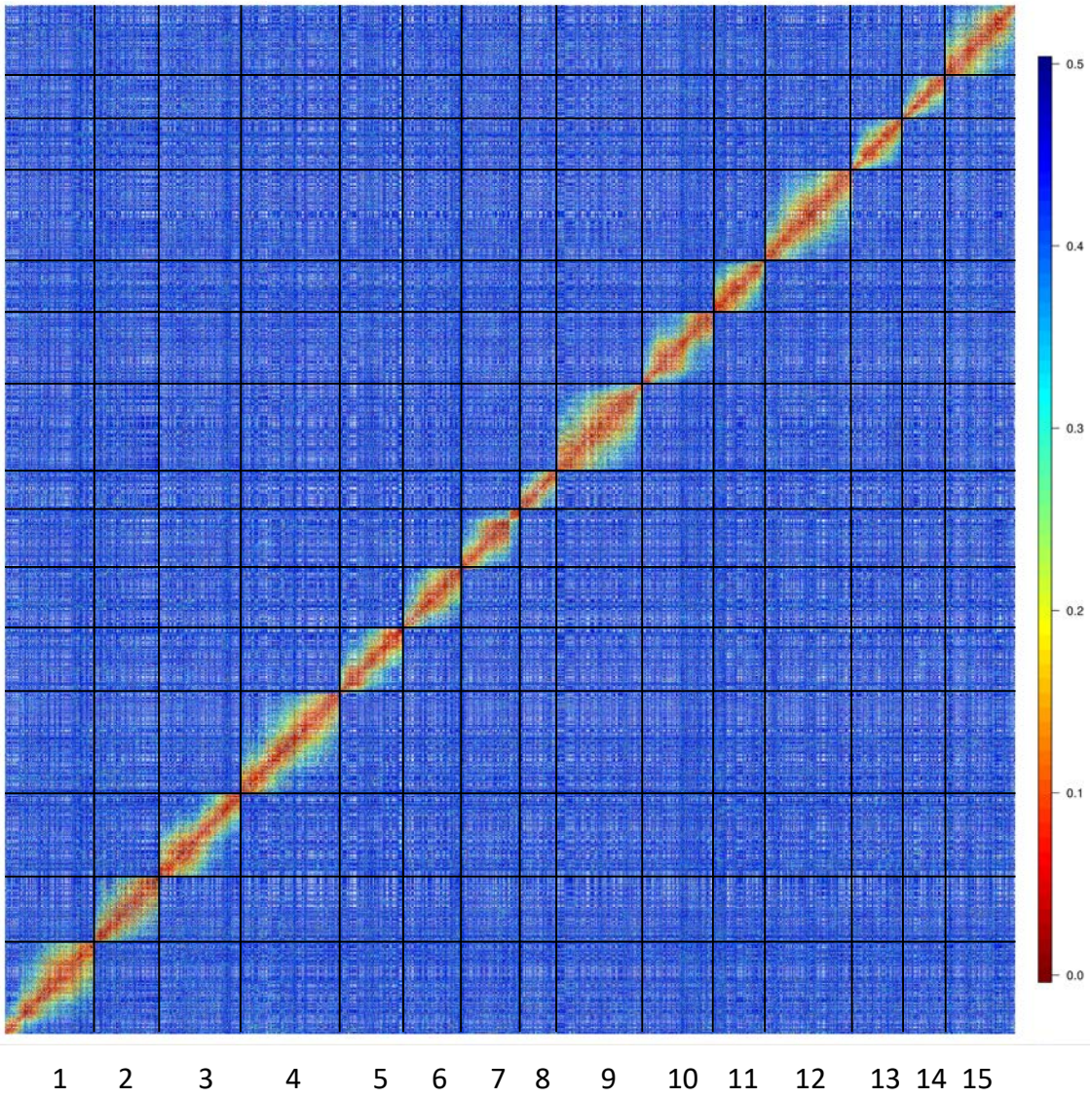
Results - Genotyping Calling - BT population

- Distribution of SNP dosage in both parents for all 26,164 SNPs.



- 56 % Simplex and double simplex
- 44 % Multiple dose in at least one parent

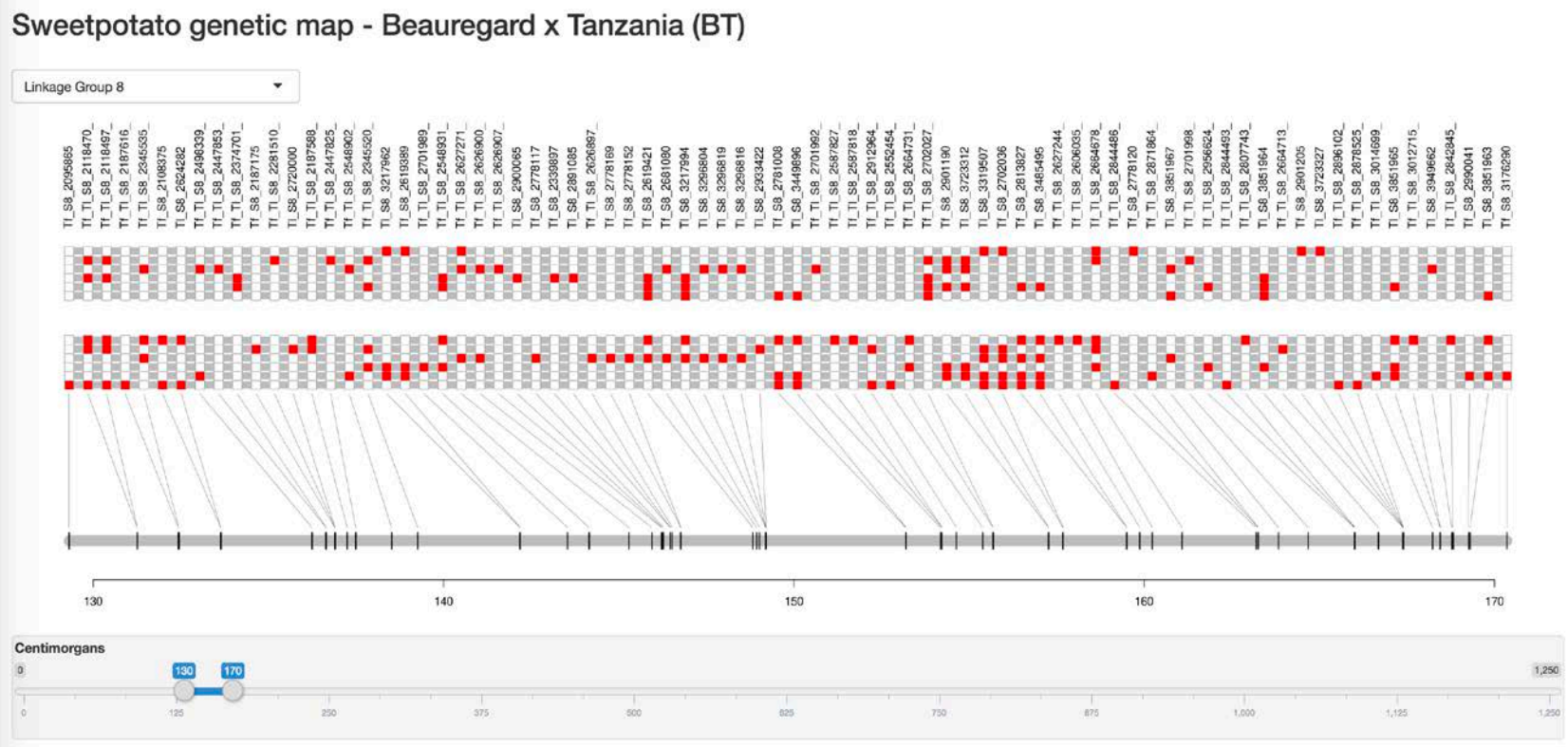
Ordering with MDS – 15 linkage groups



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

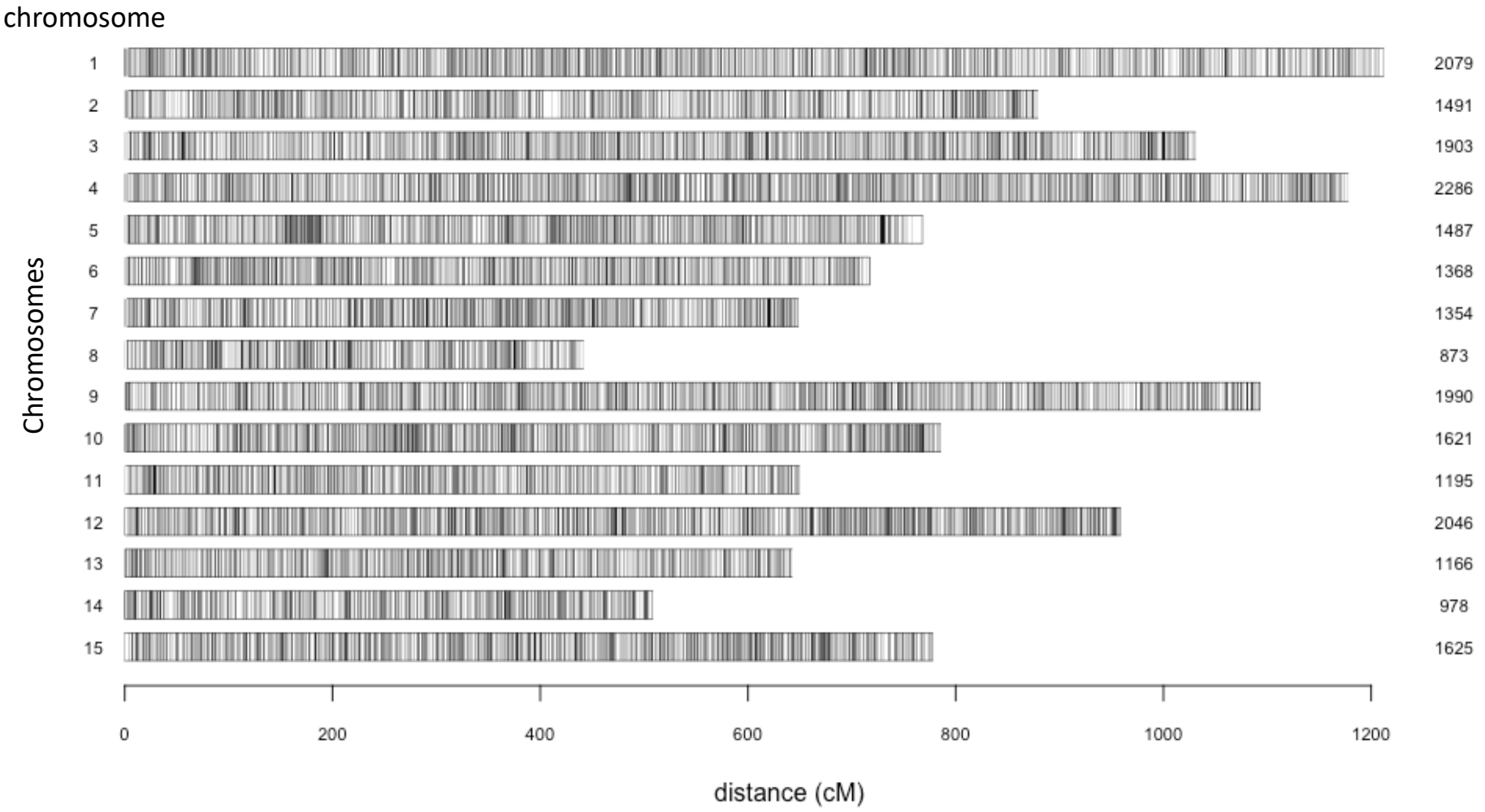
Genetic mapping – Linkage group 8 – 873 SNPs

- Differently from several polyploid maps, we have an unified map which considers both parents at the same time.
- The map is constructed for the population, not for the parents. Also, it enables to detect QTLs considering both parents in the same model

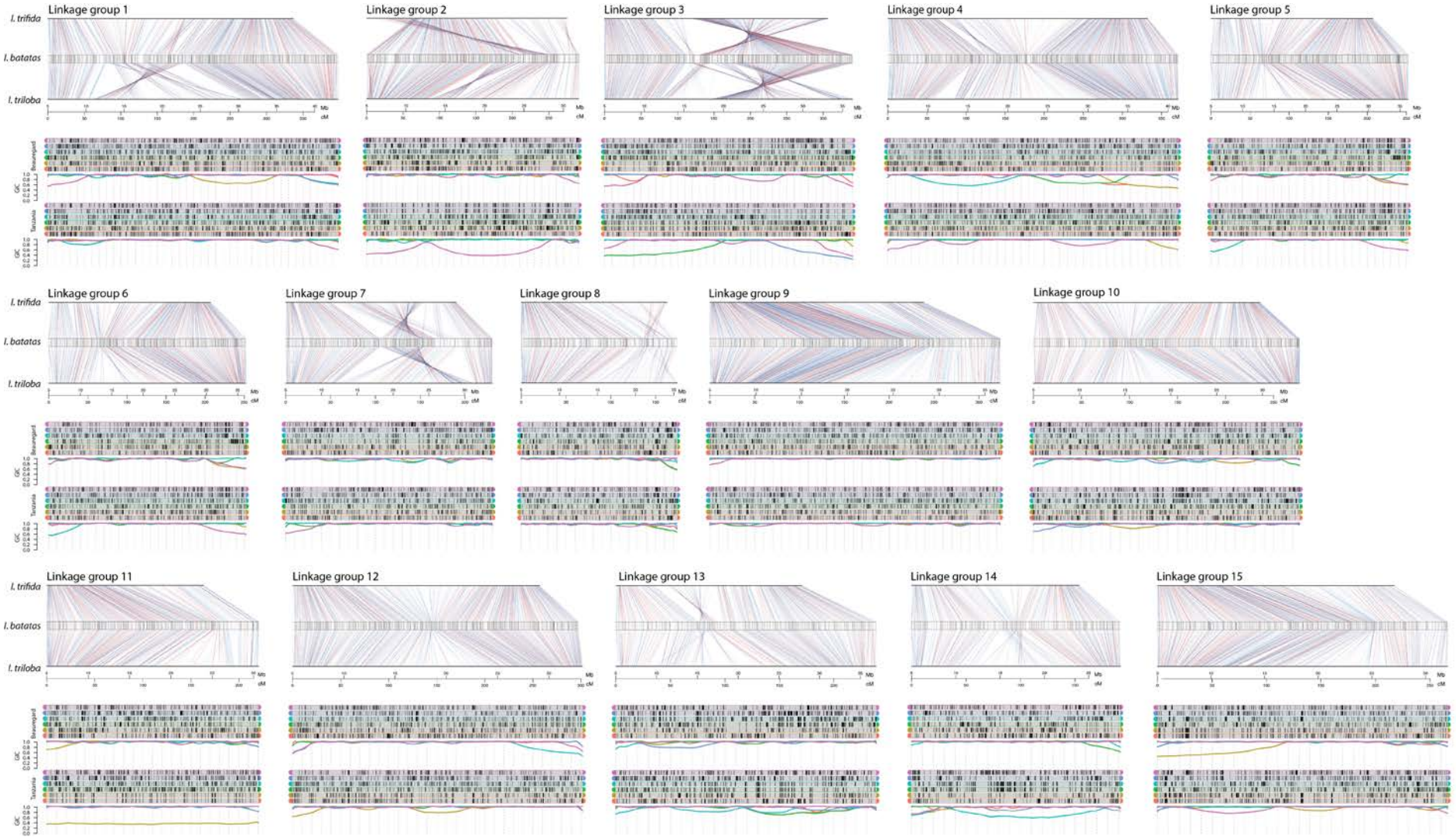


Interactive version: https://gt4sp-genetic-map.shinyapps.io/polymap_shiny/

Genetic mapping – 15 chromosomes – 23,462 SNPs

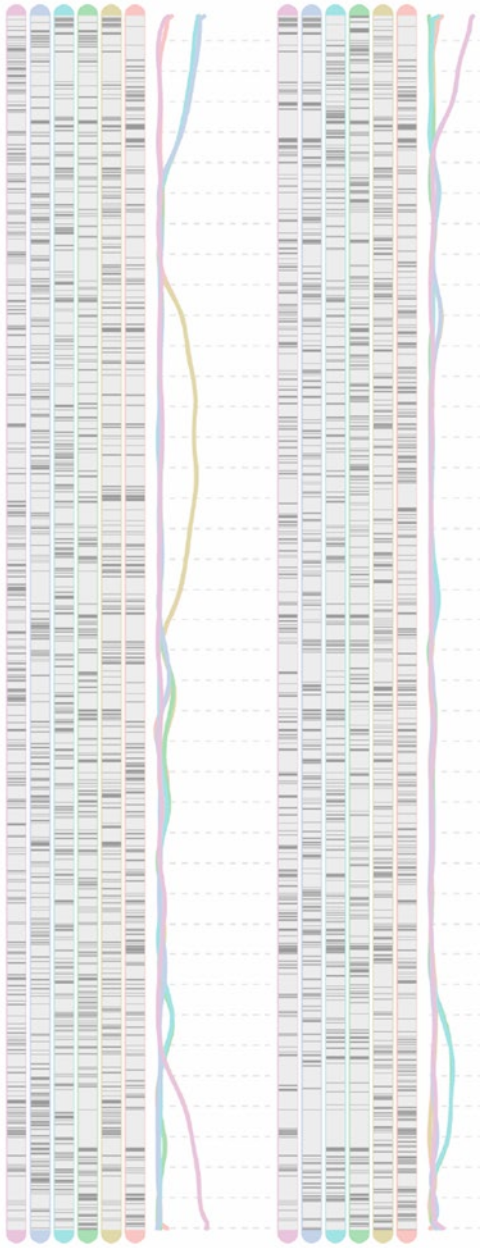


Sweetpotato map - 22,793 SNPs



Summary

- Use adequate models and software to construct maps in polyploids.
- Dosage-based markers (single dose + multiple dose) are essential in polyploid mapping
- Maps should be constructed integrating both parents
- Only maps based in homology groups will provide the adequate framework to QTL mapping



Demonstration in RStudio®