

Unraveling Polyploid Inheritance Using Ultra-Dense Multilocus Mapping

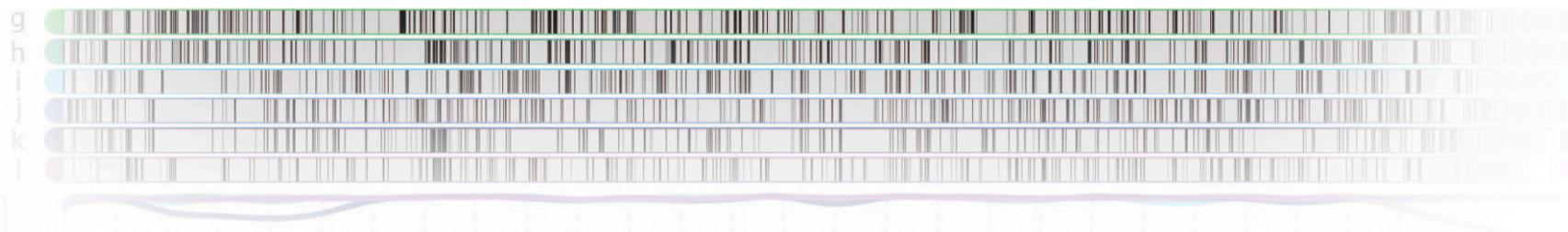
Marcelo Mollinari
mmollin@ncsu.edu



BILL & MELINDA
GATES foundation

Outline

- Introduction
- Assessing allelic variation in polyploids
- Modeling gamete formation, linkage and phasing
- Genetic mapping in hexaploid sweetpotato
- Assessing meiotic configurations, multivalent formation and preferential chromosome pairing
- Applications: QTL analysis and genomic prediction



Genomic Tools for Sweetpotato Improvement – GT4SP

- Bill & Melinda Gates Foundation has a large portfolio and 10% is dedicated to agriculture development in Sub-Saharan Africa and South Asia for food supply purposes.
- Key Crops: Cassavas, Yams, Bananas and **Sweetpotatoes**
- Phase I: GT4SP (PI: Dr. G. Craig Yencho – NCSU)
- Phase II: SweetGAINS (PI: Dr. Hugo Campos - International Potato Center)



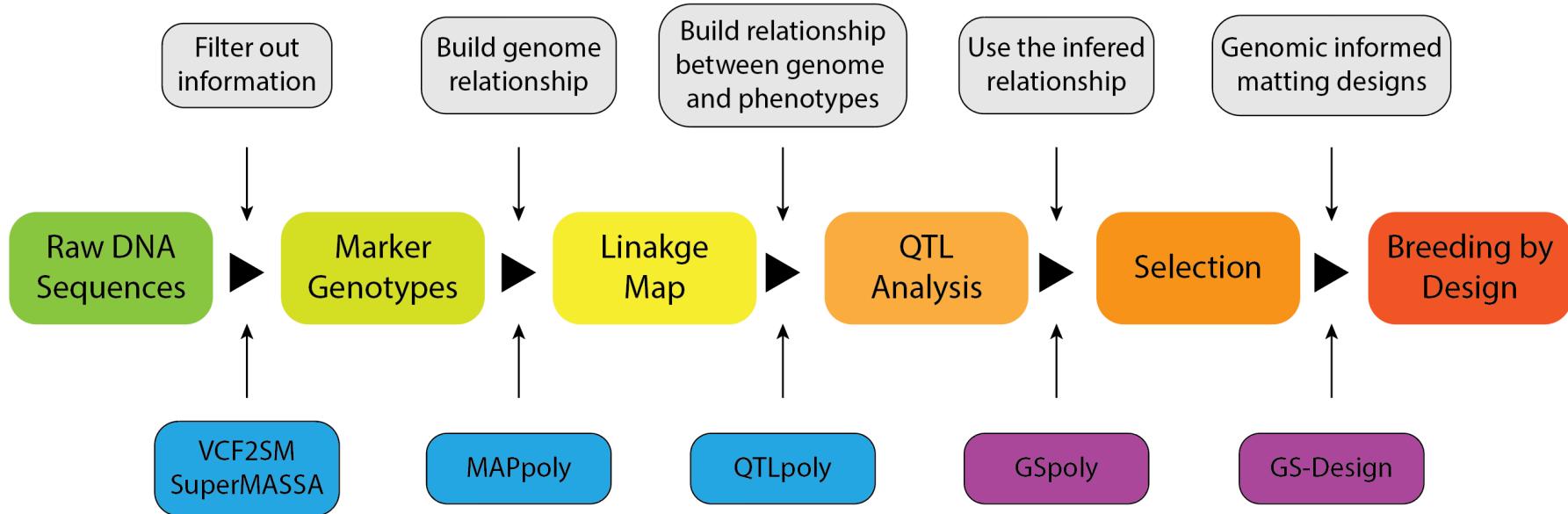
Sweetpotato for profit and health initiative

Food security for Sub-Saharan Africa



GT4SP & SweetGAINS

- Develop computational tools for the whole pipeline data analysis

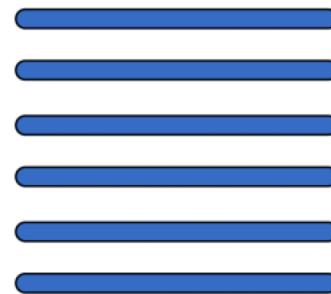


- VCF2SM:** Python scripts processing DNA calls (VCF files) for SuperMASSA
- SuperMASSA:** Call SNP dosage marker genotype probability distribution.
- MAPpoly:** An R package for constructing a complete linkage map for 2X, 4X, 6X, 8X
- QTLpoly:** An R package for QTL mapping in full-sib families for 2X, 4X, 6X, 8X
- GSpoly:** a computational tool for genomic selection (to be implemented)
- GS-Design:** a simulation tool for designing genomic selection scheme based on empirical genomic data (to be implemented)

Polyplloid Species

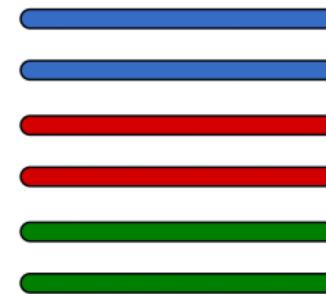
- Organisms with more than 2 complete sets of chromosomes

Autohexaploid



:

Allohexaploid

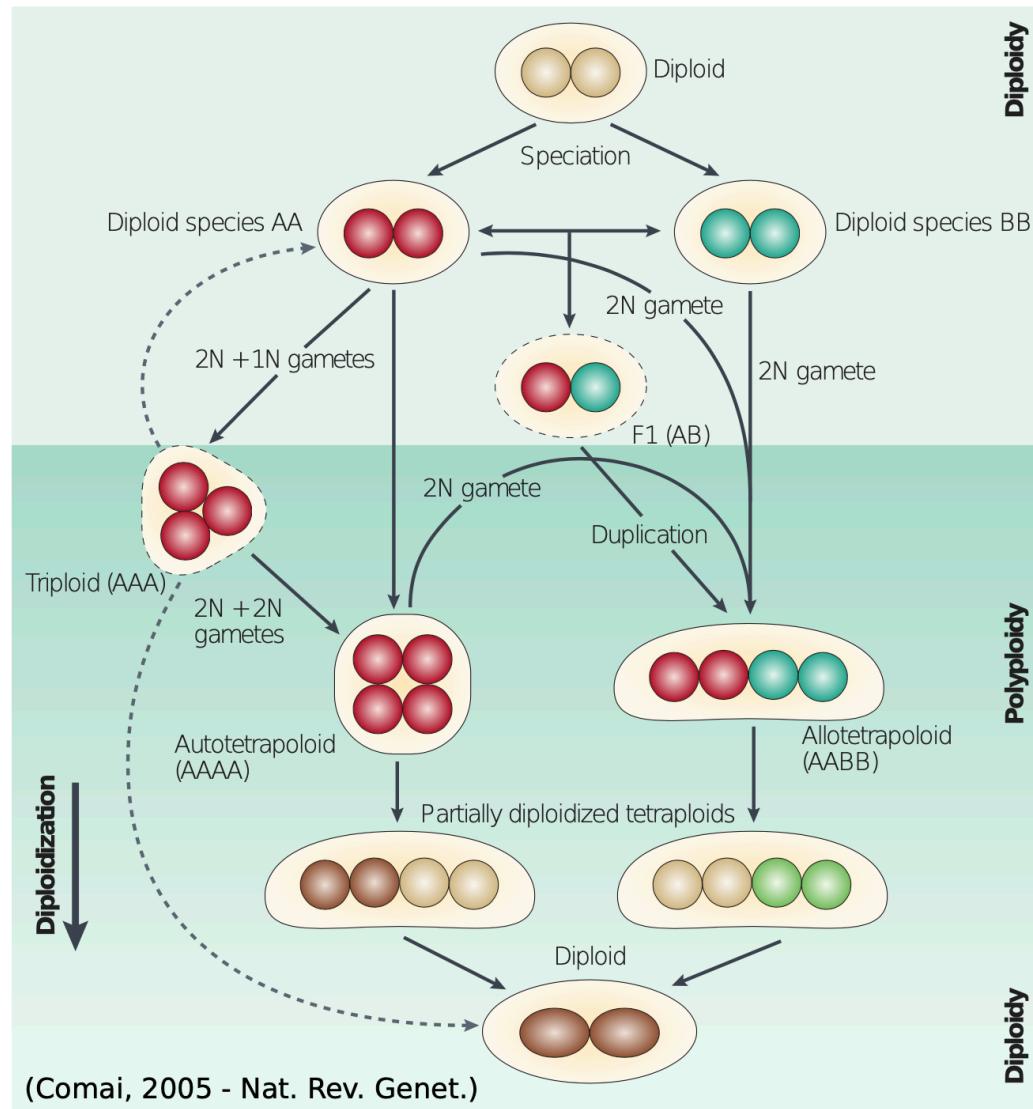


:

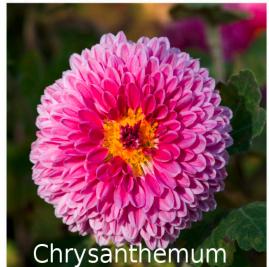
- Multiple sets have the **same** origin

- Multiple sets have **different** origins

How are polyploids formed



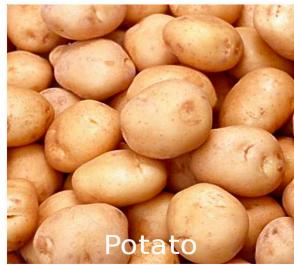
Autopolyploids and meiotic pairing



Chrysanthemum



Sweet potato



Potato



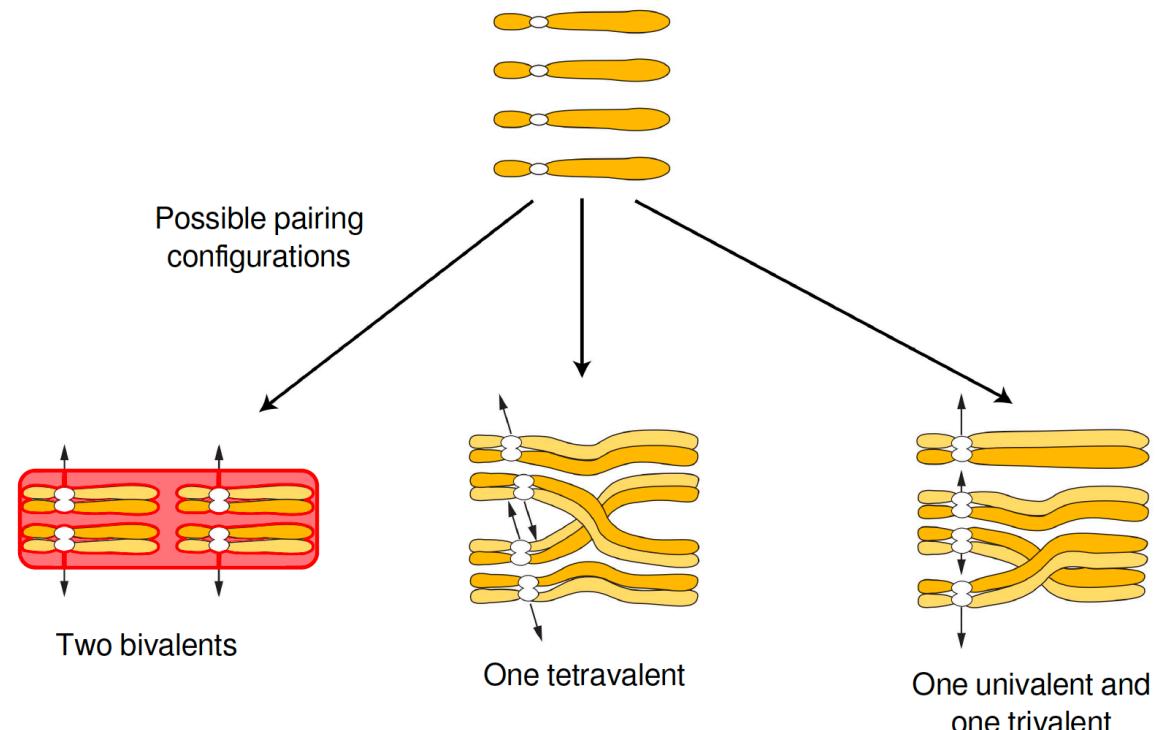
Rose



Forage crops



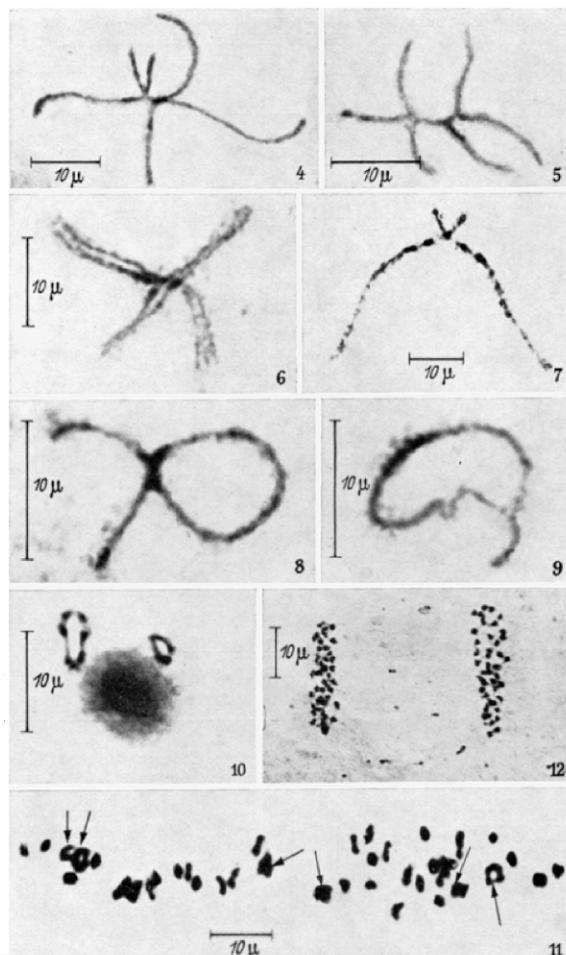
Sugarcane



Griffiths et al. (2004)

Meiotic Pairing in polyploids

Sweetpotato



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

Sugarcane

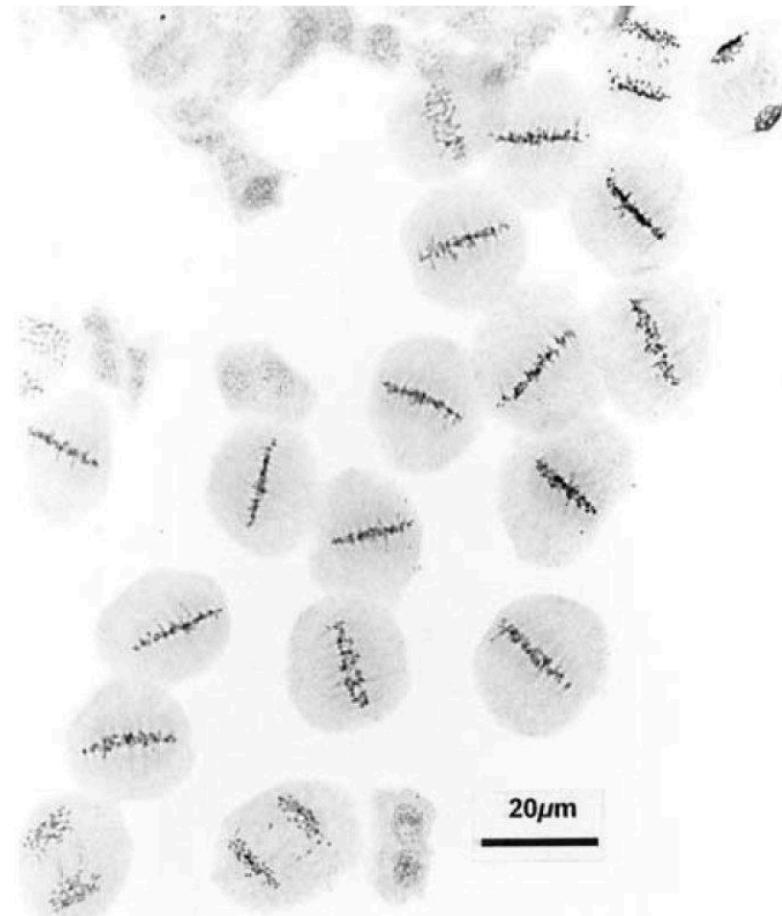
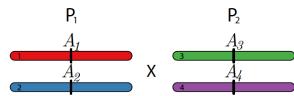


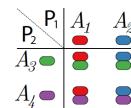
Figure 1. PMCs of *Saccharum* spp. hybrid clone 79N9059 at meiosis. As was the case in other clones, pairing was regular, bivalents generally formed. (Bielig *et al.* 2003)

Autohexaploid complexity

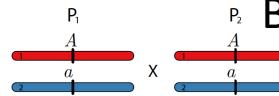
Multiallelic



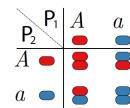
Gametes

4 possible genotypes
1:1:1:1

Diploid

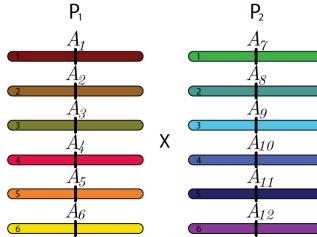


Gametes

3 possible genotypes
1:2:1

Biallelic

Multiallelic



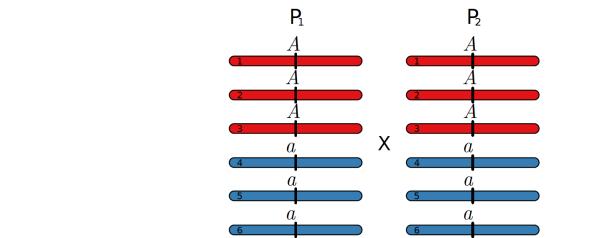
Gametes

	1	2	3	...	18	19	20
P_2	$A_1A_2A_3$	$A_1A_2A_4$	$A_1A_2A_5$...	$A_3A_4A_6$	$A_3A_5A_6$	$A_4A_5A_6$
1	$A_7A_8A_9$
2	$A_7A_8A_{10}$
3	$A_7A_8A_{11}$
...
18	$A_9A_{10}A_{12}$
19	$A_9A_{11}A_{12}$
20	$A_{10}A_{11}A_{12}$

400 possible genotypes

Hexaploid

Biallelic



Gametes

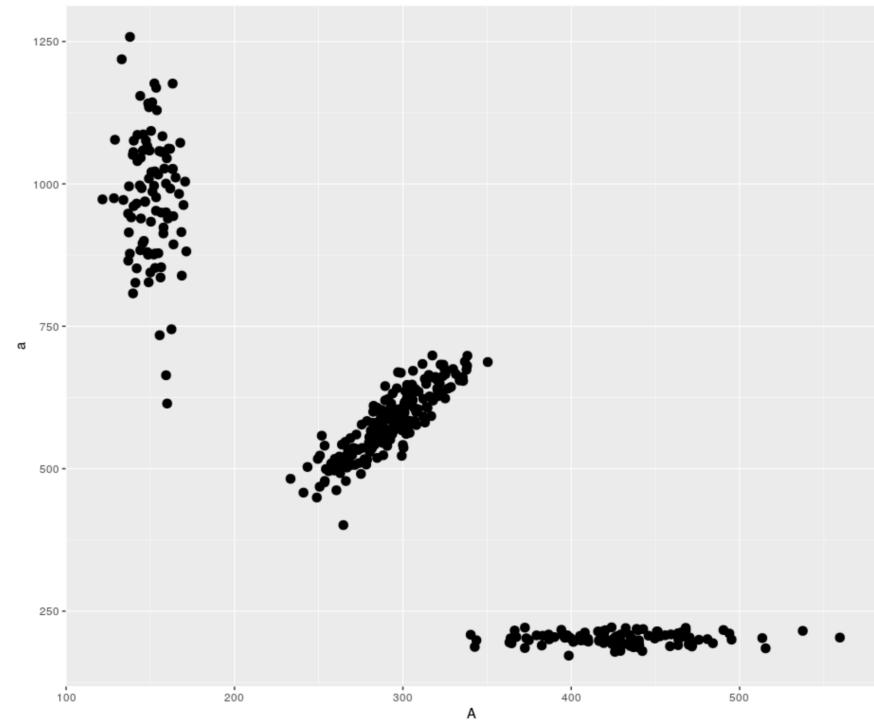
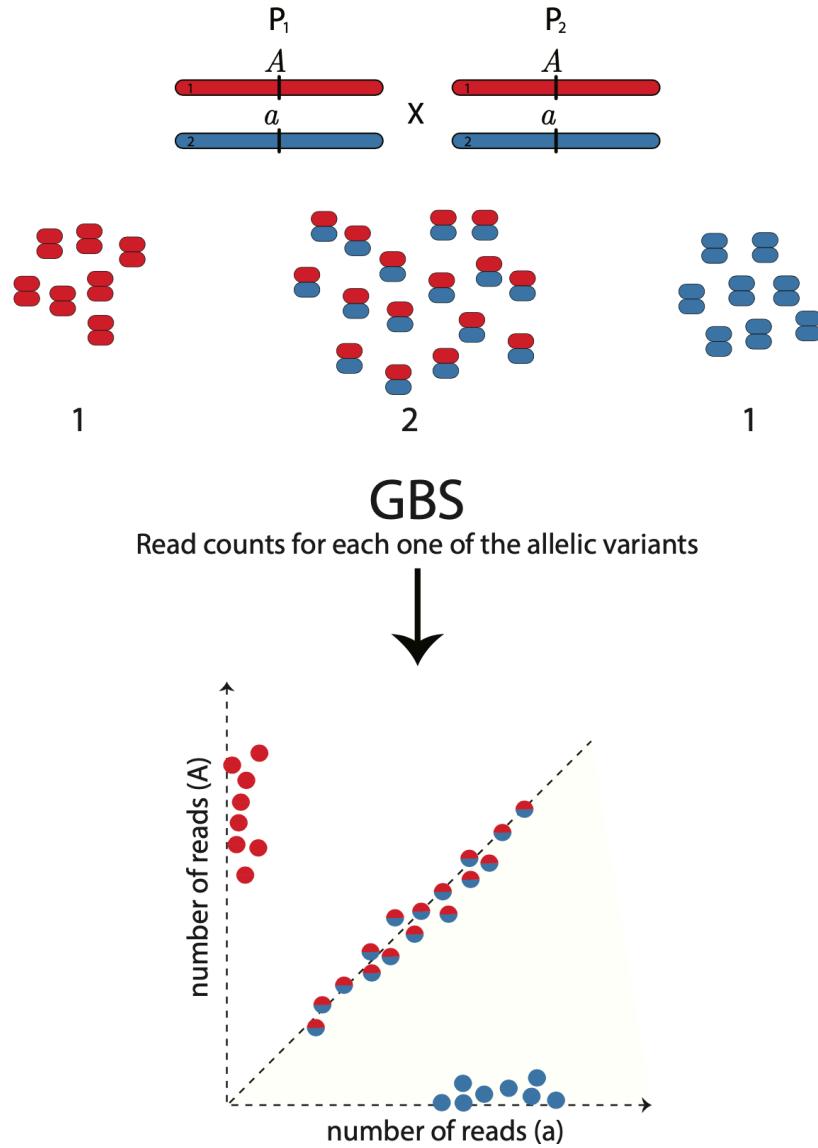
	1	2	3	...	18	19	20
P_2	$A_1A_2A_3$	$A_1A_2A_4$	$A_1A_2A_5$...	$A_3A_4A_6$	$A_3A_5A_6$	$A_4A_5A_6$
1	$A_7A_8A_9$
2	$A_7A_8A_{10}$
3	$A_7A_8A_{11}$
...
18	$A_9A_{10}A_{12}$
19	$A_9A_{11}A_{12}$
20	$A_{10}A_{11}A_{12}$

7 possible genotypes

1 : 18 : 99 : 164 : 99 : 18 : 1

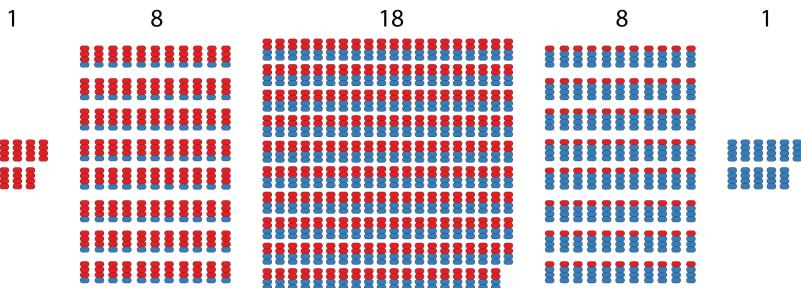
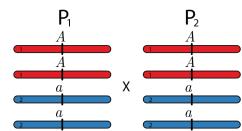


Assessing allelic variation in diploids

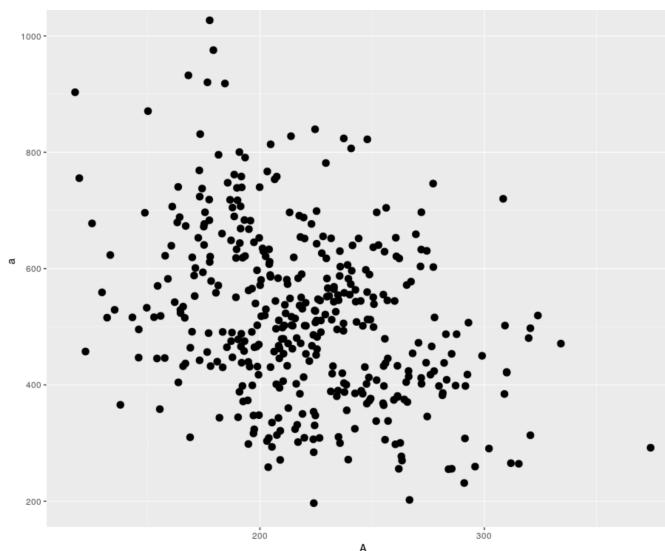
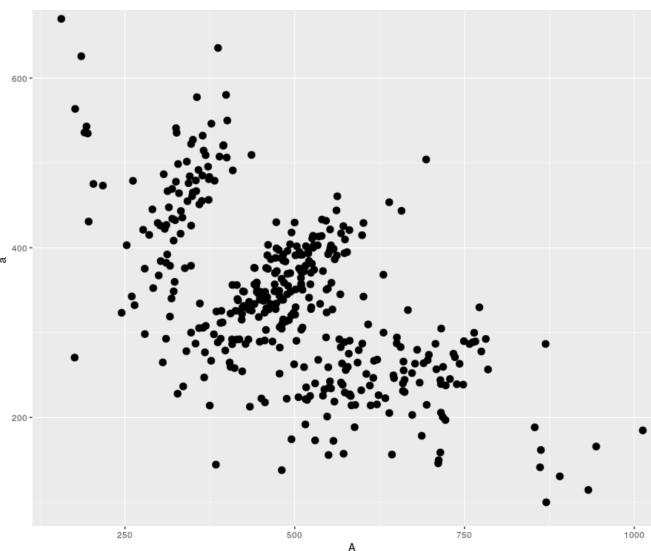
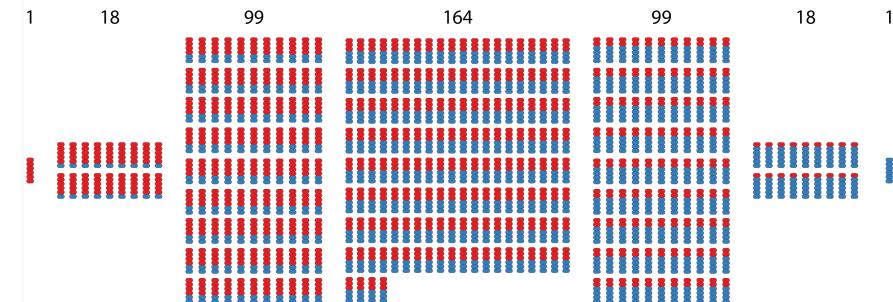
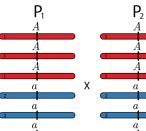


Assessing allelic variation in polyploids

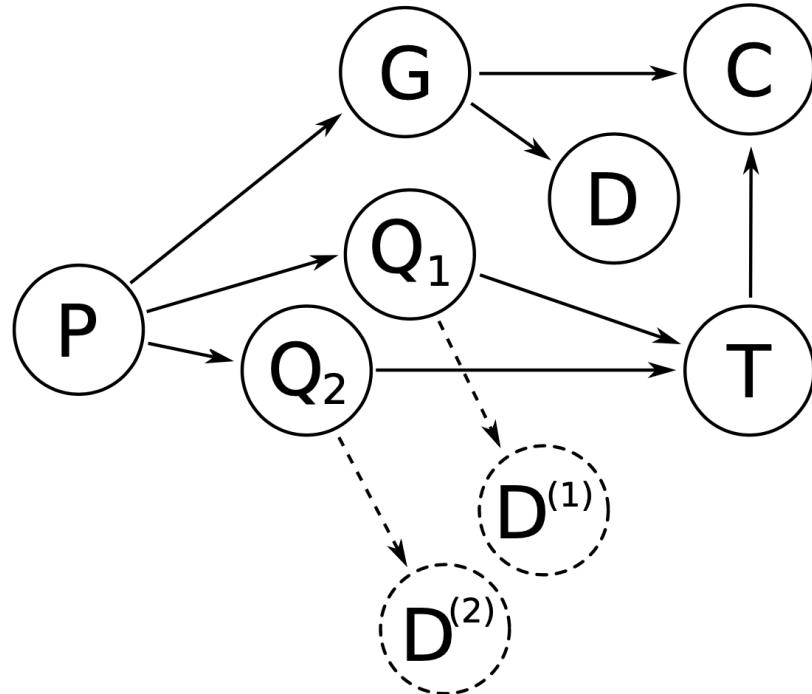
Tetraploid



Hexaploid



Assessing allelic variation in polyploids



Serang et al. (2012)

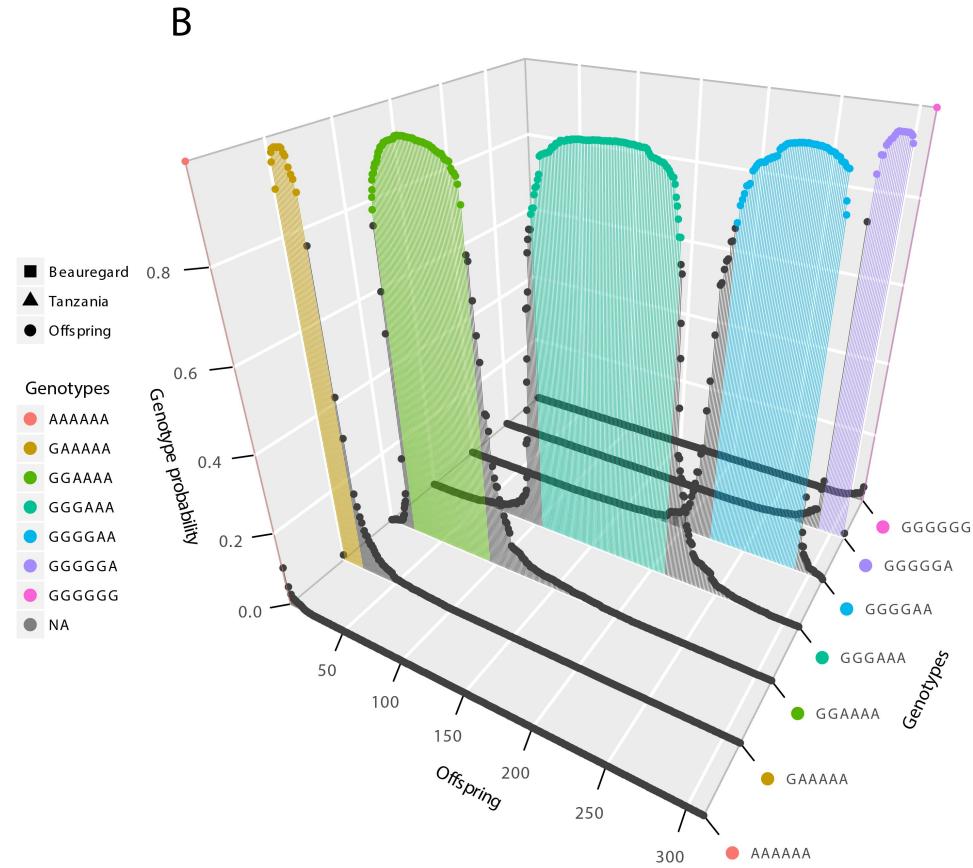
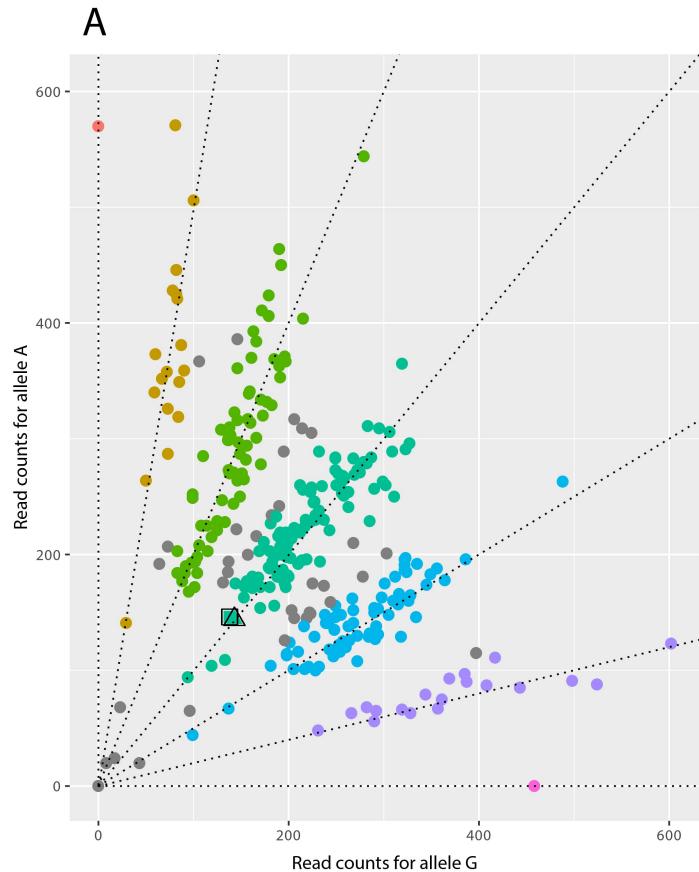
Molinari and Serang (2015)

$$\Pr(P, G, D, T, C, Q_1, Q_2, D_1, D_2)$$

- ▶ P : ploidy
- ▶ G : genotype of all individuals
- ▶ D : observed data
- ▶ T : theoretical distribution of genotypes
- ▶ C : histogram of genotypes
- ▶ Q_1 and Q_2 : parent genotypes, with data D_1 and D_2 (if available)

Genotyping Calling using SuperMASSA

Dosage calling Including the probability distribution of the genotypes



Biparental Population - BT

- Beauregard x Tanzania
- 315 individuals
- GBS – GBSpoly protocol (Bode Olukolu – U Tennessee)
- Two reference genomes *I. trifida* and *I. triloba* (Zhangjun Fei's group – BTI Cornell)



Beauregard



Tanzania



Biparental Population - BT



X



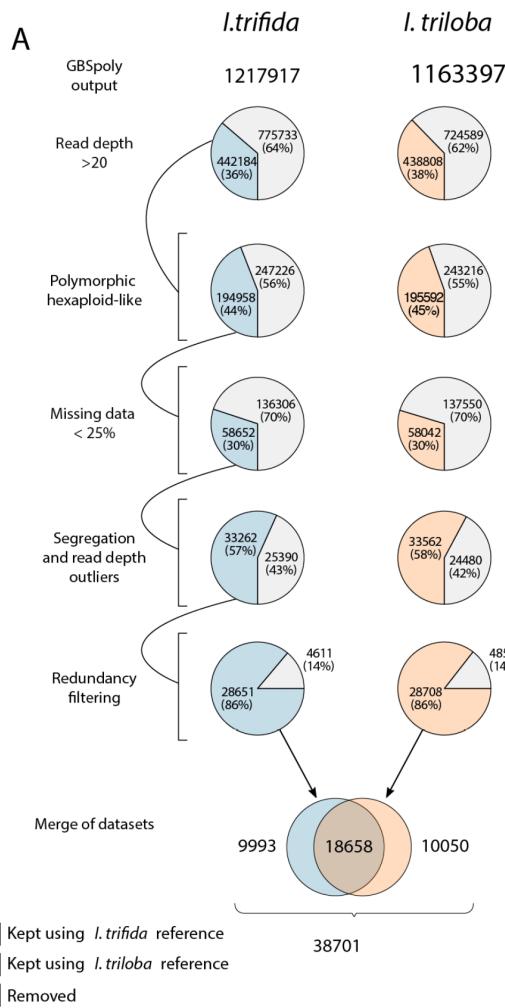
Beauregard

Tanzania

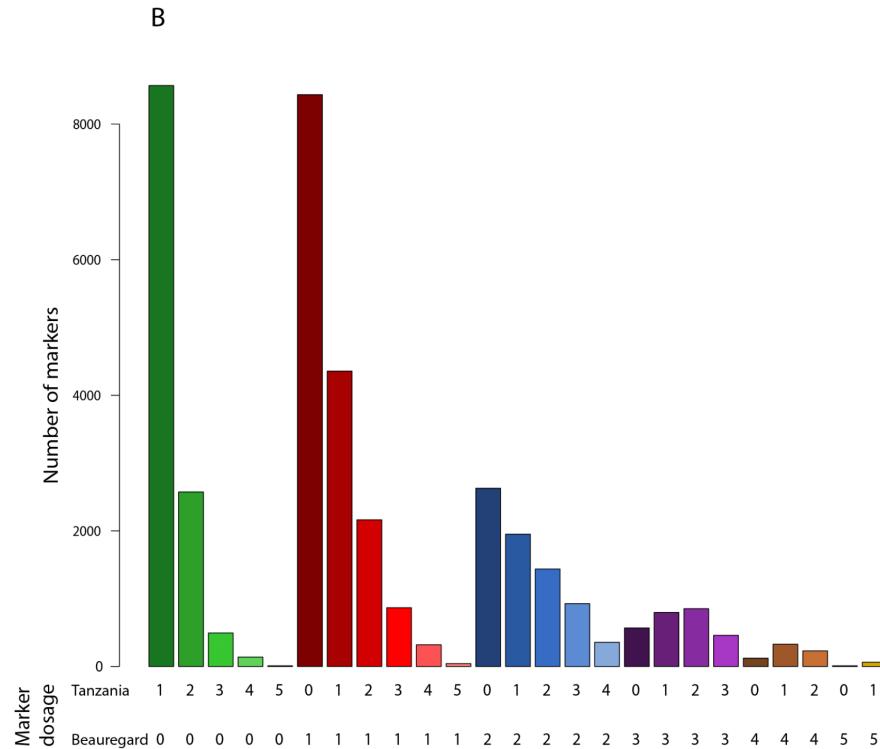


Results - Genotyping Calling - BT population

Filtering process



Distribution of the marker doses in *Beauregard* and *Tanzania* (38,701 SNPs).



Linkage and Haplotyping in polyploids

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

2019

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

[doi:10.1534/g3.119.400378](https://doi.org/10.1534/g3.119.400378)

Unraveling the Hexaploid Sweetpotato Inheritance Using Ultra-Dense Multilocus Mapping

2020

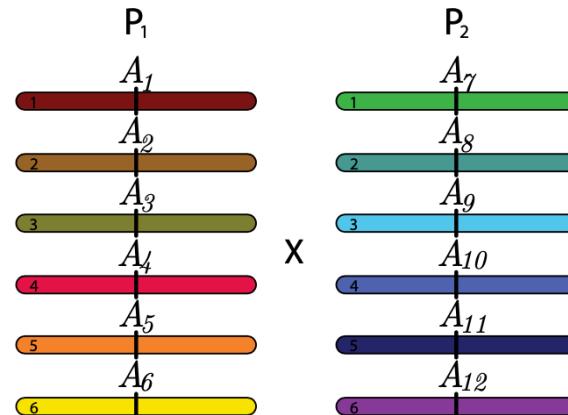
Marcelo Mollinari,^{*†,1} Bode A. Olukolu,[‡] Guilherme da S. Pereira,^{*†} Awais Khan,[§] Dorcus Gemenet,^{**}

G. Craig Yencho,[†] and Zhao-Bang Zeng^{*†}

*Bioinformatics Research Center, North Carolina State University, Raleigh, North Carolina, [†]Department of Horticultural Science, North Carolina State University, Raleigh, North Carolina, [‡]Department of Entomology and Plant Pathology, University of Tennessee, Knoxville, Tennessee, [§]Plant Pathology and Plant-Microbe Biology Section, Cornell University, Geneva, New York, and ^{**}International Potato Center, ILRI Campus, Nairobi, Kenya

[doi:10.1534/g3.119.400620](https://doi.org/10.1534/g3.119.400620)

Gamete formation in polyploids

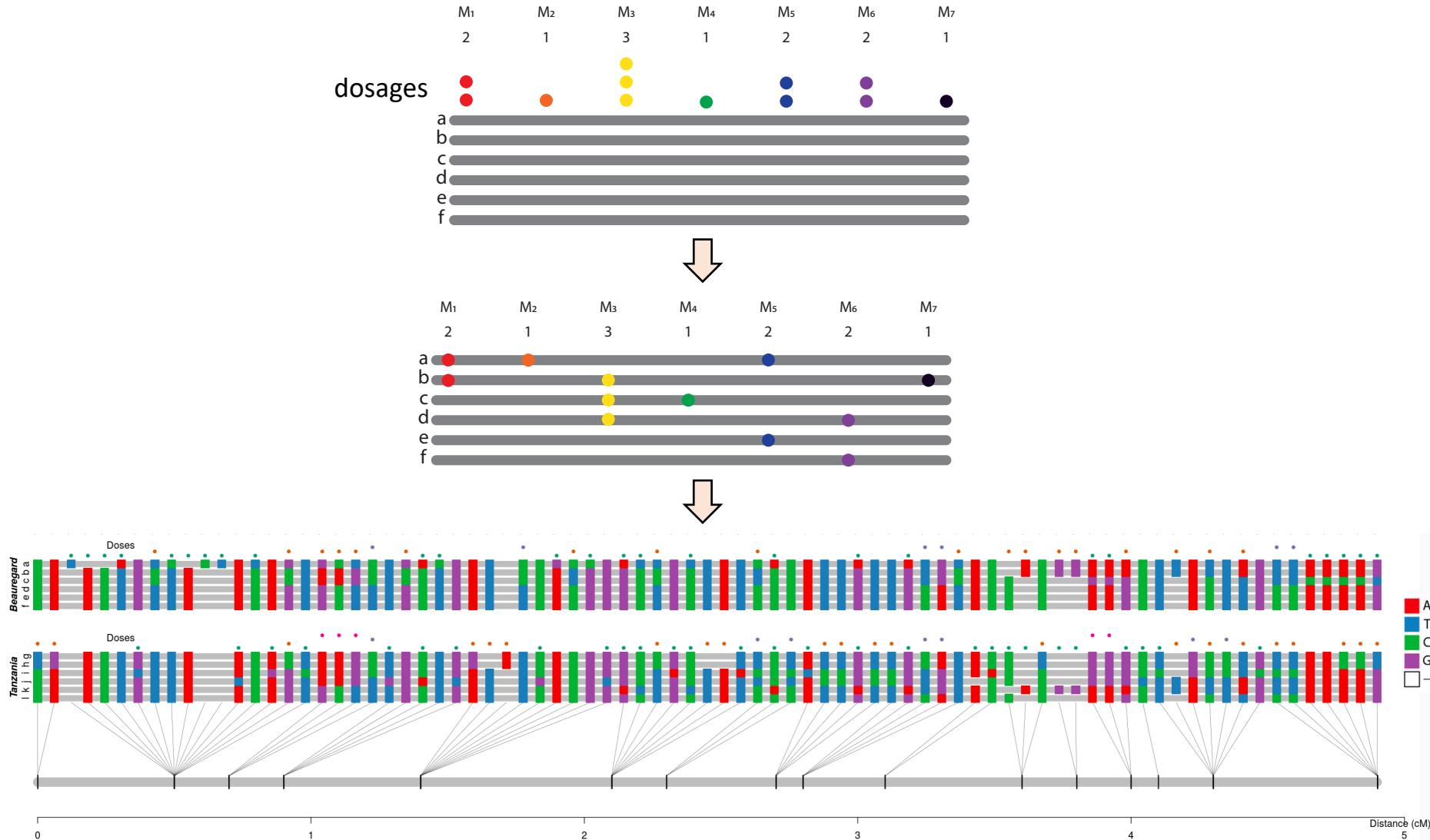


		Gametes						
		1 $A_1A_2A_3$	2 $A_1A_2A_4$	3 $A_1A_2A_5$...	18 $A_3A_4A_6$	19 $A_3A_5A_6$	20 $A_4A_5A_6$
$P_1 \backslash P_2$		$A_1A_2A_3$	$A_1A_2A_4$	$A_1A_2A_5$...	$A_3A_4A_6$	$A_3A_5A_6$	$A_4A_5A_6$
1	$A_7A_8A_9$	●●●	●●●	●●●	...	●●●	●●●	●●●
2	$A_7A_8A_{10}$	●●●	●●●	●●●	...	●●●	●●●	●●●
3	$A_7A_8A_{11}$	●●●	●●●	●●●	...	●●●	●●●	●●●
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
18	$A_9A_{10}A_{12}$	●●●	●●●	●●●	...	●●●	●●●	●●●
19	$A_9A_{11}A_{12}$	●●●	●●●	●●●	...	●●●	●●●	●●●
20	$A_{10}A_{11}A_{12}$	●●●	●●●	●●●	...	●●●	●●●	●●●

400 possible genotypes

Haplotyping in polyploids

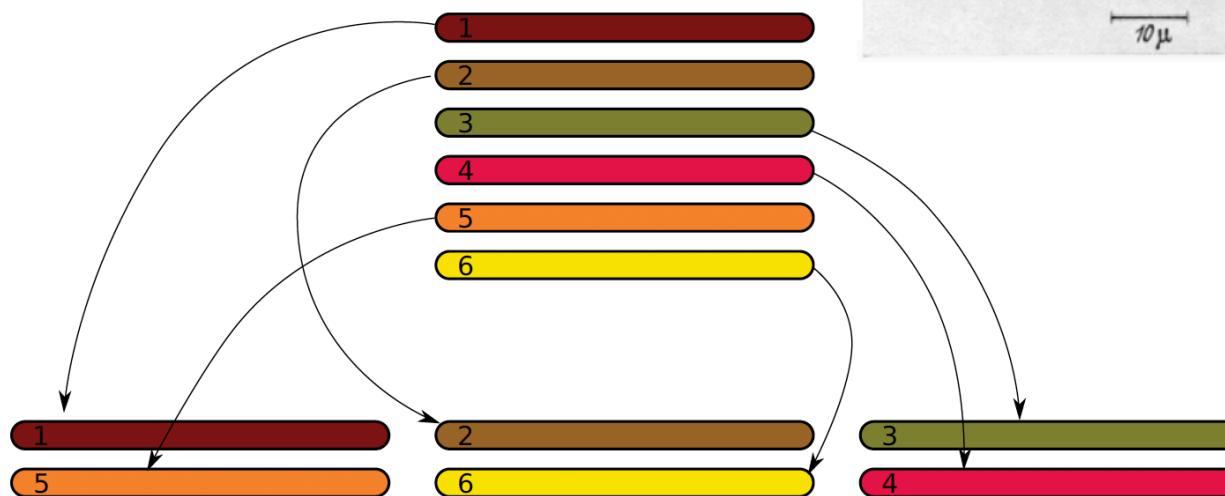
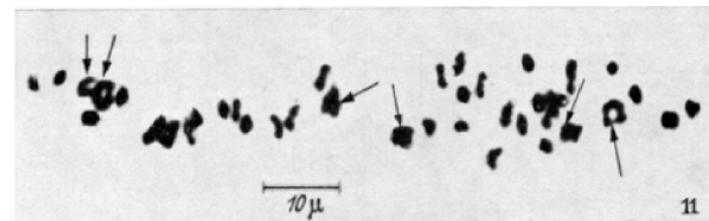
- Disposition of allelic variants in the homologs in a homology group



Gamete formation in polyploids

Random bivalent pairing

Bivalents in sweetpotato
Magoon et al. 19070



$$\psi$$

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

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

Expected gametic frequency given a bivalent configuration

$$\begin{array}{cc} A_1 & B_1 \\ \text{---} & \text{---} \\ A_6 & B_6 \end{array}
 \quad
 \begin{array}{cc} A_2 & B_2 \\ \text{---} & \text{---} \\ A_5 & B_5 \end{array}
 \quad
 \begin{array}{cc} A_3 & B_3 \\ \text{---} & \text{---} \\ A_4 & B_4 \end{array}$$

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

In general:

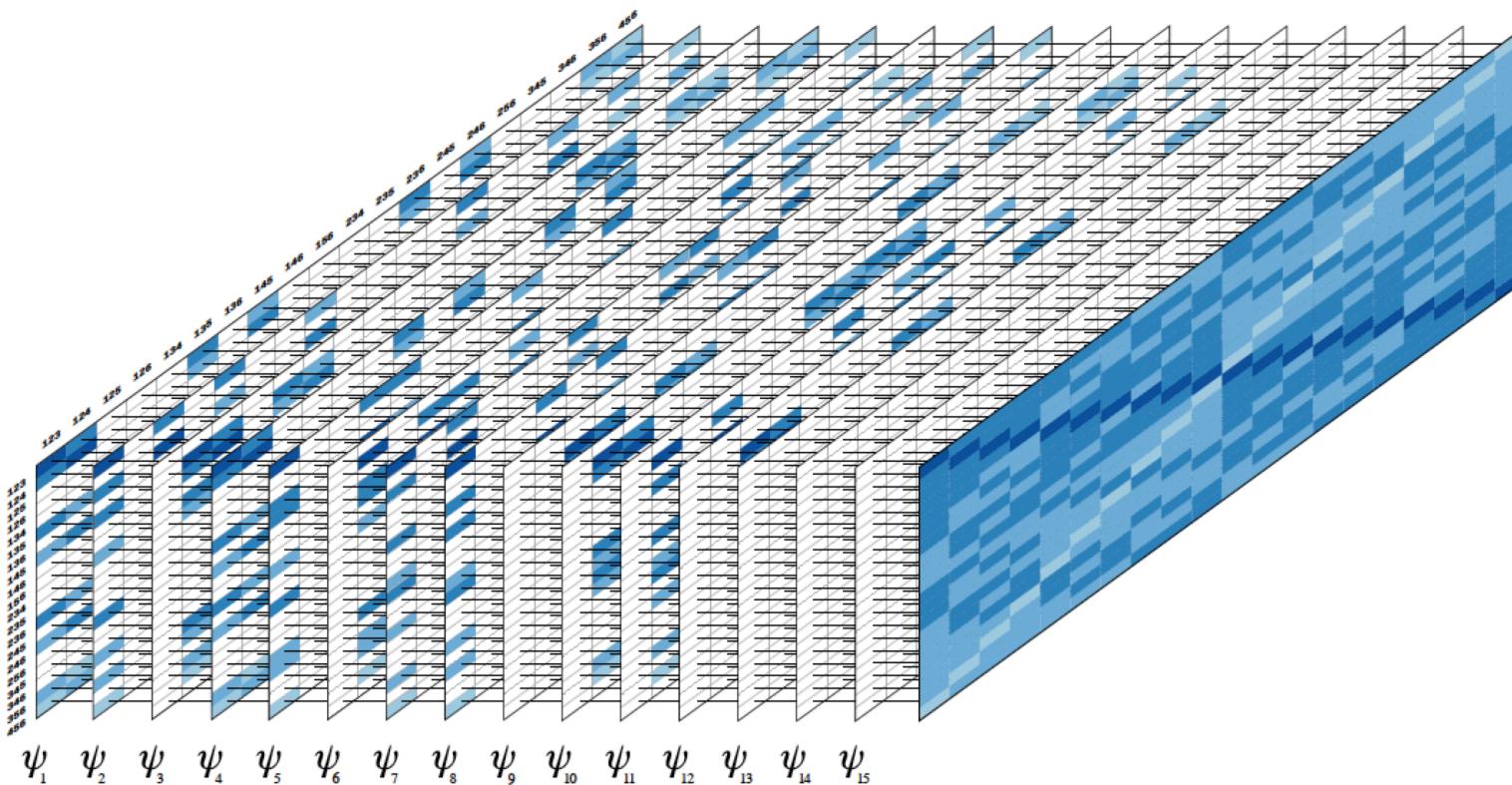
$$\mathbf{V}_1 \otimes \mathbf{V}_2 \otimes \cdots \otimes \mathbf{V}_{\frac{p}{2}}$$

All elements of this product are of the form

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

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

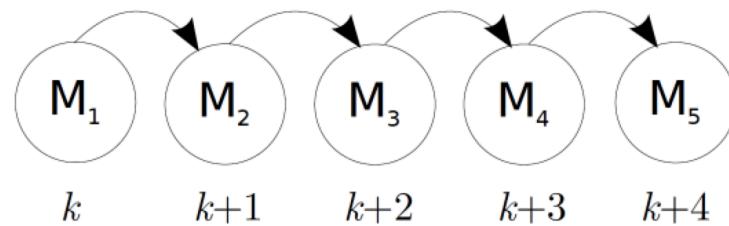
Unconditional gametic probability



$$\begin{aligned}\Pr(G_A, G_B) &= \sum_j \Pr(G_A, G_B \mid \psi_j) \Pr(\psi_j) \\ &= \frac{l! \left(\frac{m}{2} - l\right)!}{w_m} \frac{(1 - r_k)^{\frac{m}{2} - l} (r_k)^l}{2^{\frac{m}{2}}}.\end{aligned}$$

Transition probabilities and the Markov model

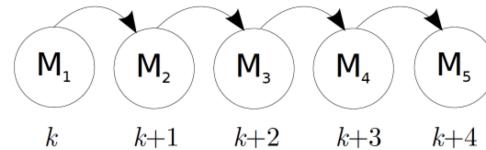
Markov Model: conditional independence



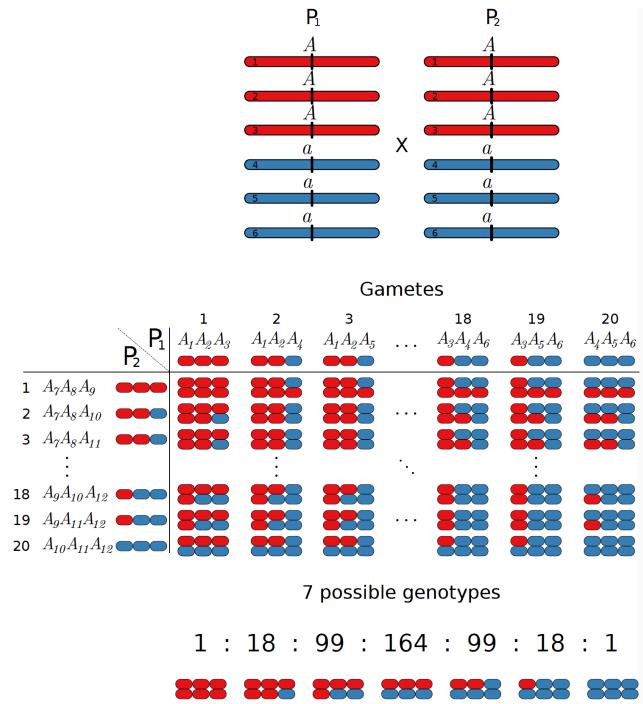
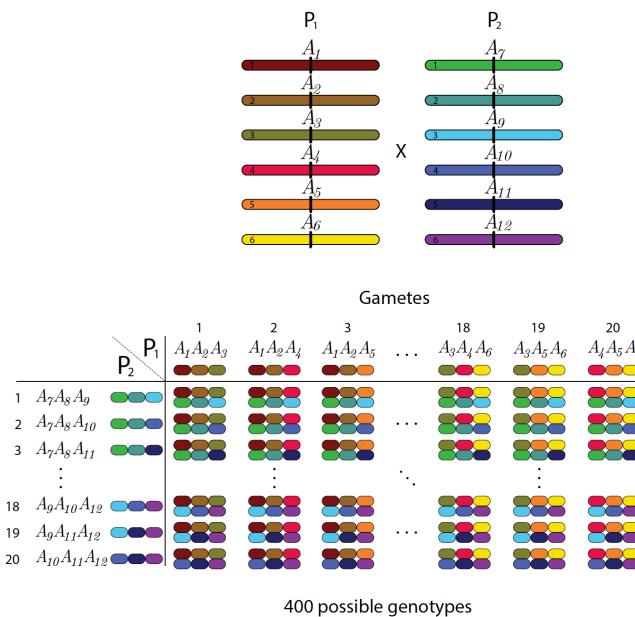
$$\Pr(G_{k+1}|G_k) = \frac{(1 - r_k)^{\frac{p}{2} - l}(r_k)^l}{\binom{\frac{p}{2}}{l}}$$

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

Multiallelic to biallelic

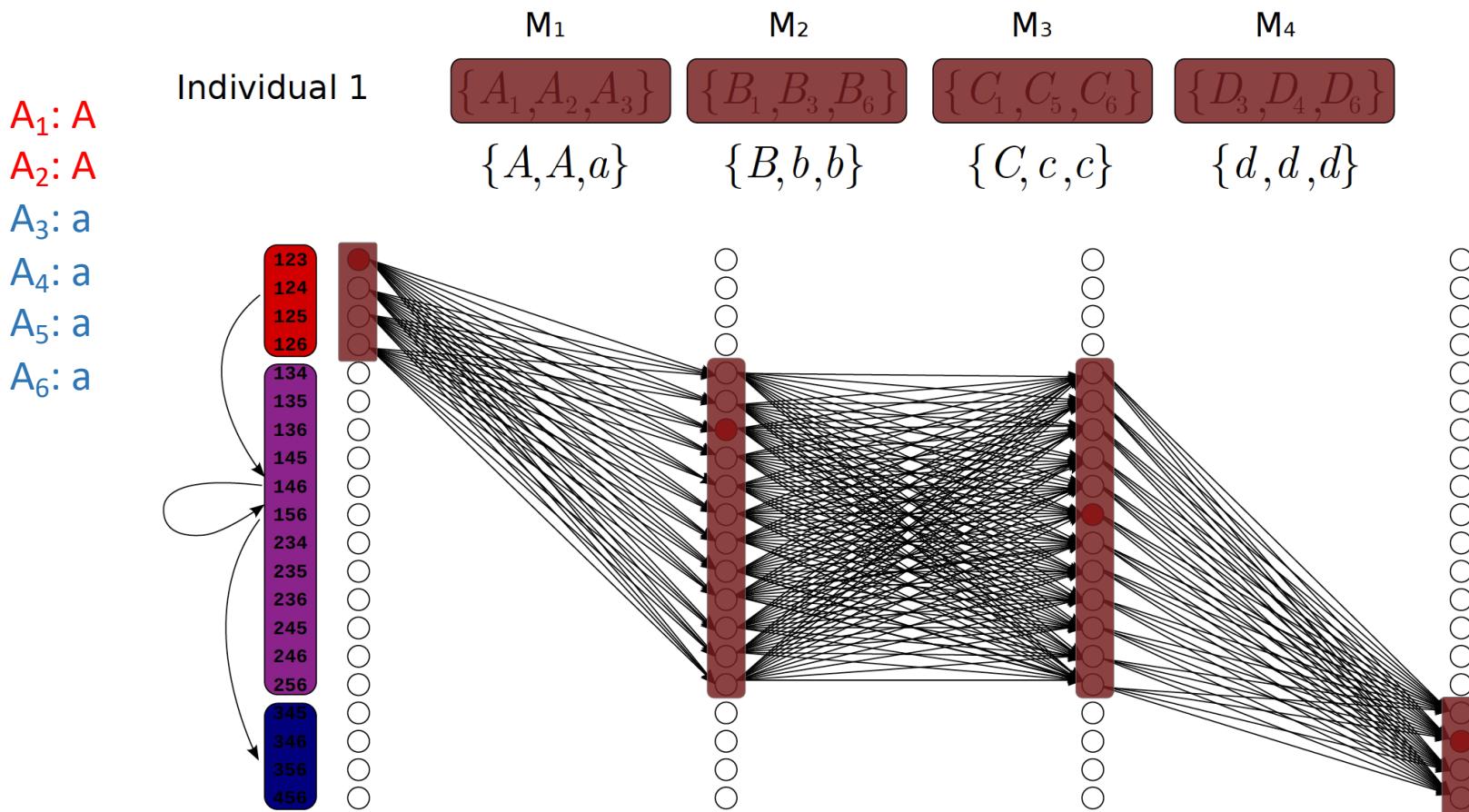


$$\Pr(G_{k+1}|G_k) = \frac{(1 - r_k)^{\frac{p}{2} - l}(r_k)^l}{\binom{\frac{p}{2}}{l}}$$



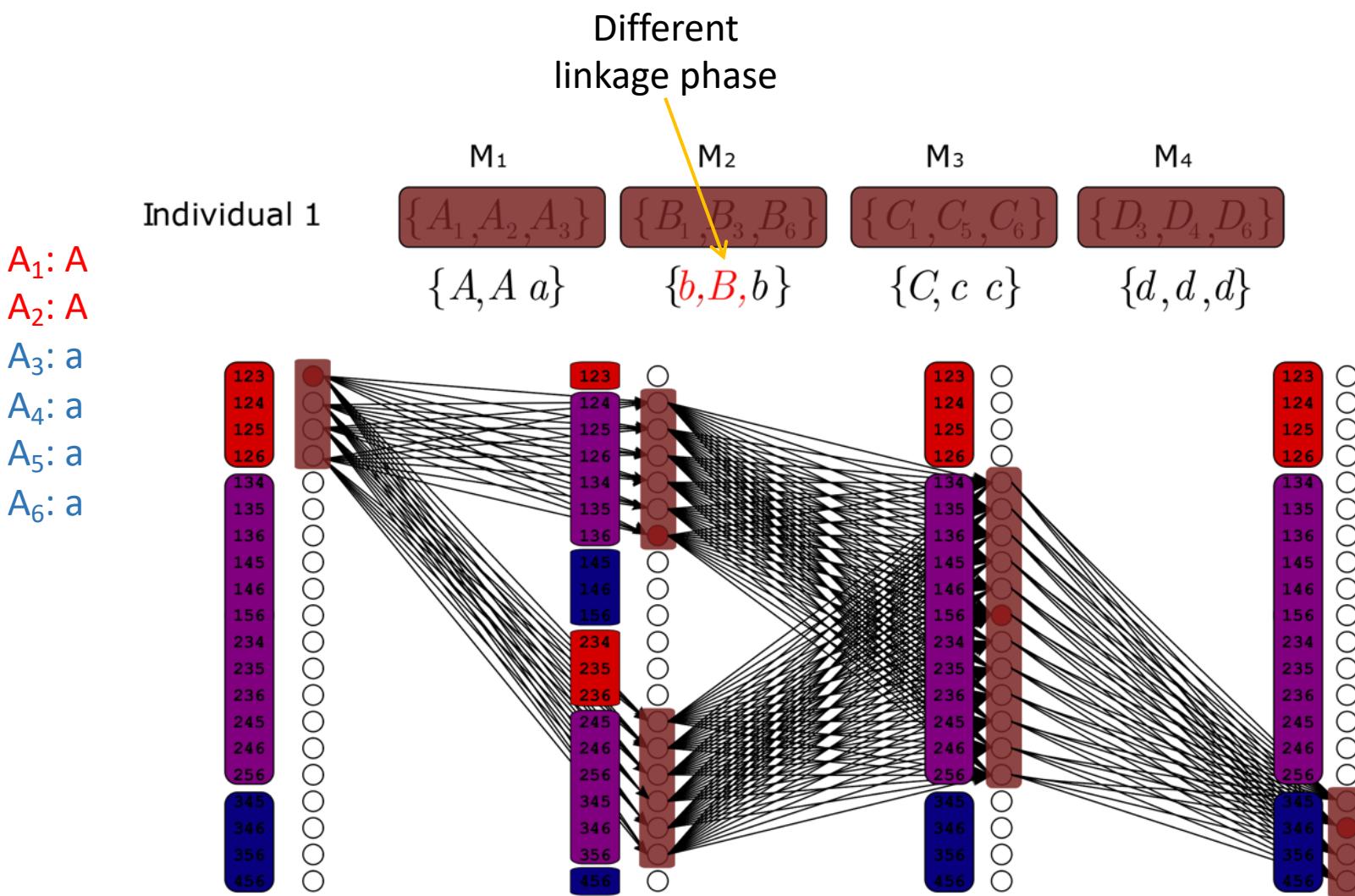
Emission function

$$b_j(O) = \Pr\left(O \mid \mathcal{G}_{k,j}^m, \varphi_P^k, \varphi_Q^k\right) = \begin{cases} 1 - \epsilon & \text{if } O = \delta(k, j) \\ \frac{\epsilon}{m} & \text{otherwise} \end{cases} \quad (8)$$



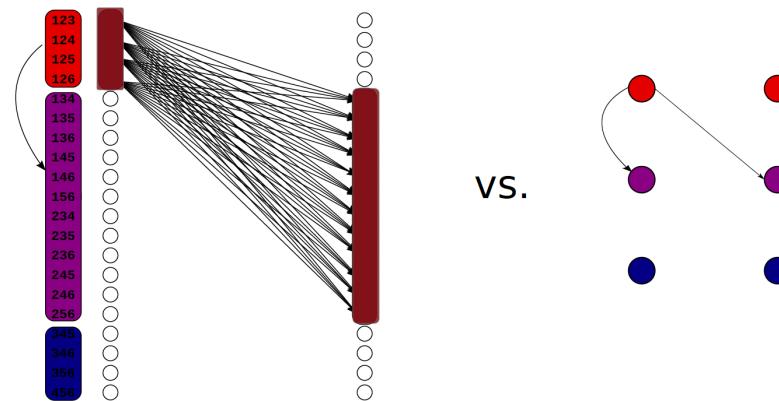
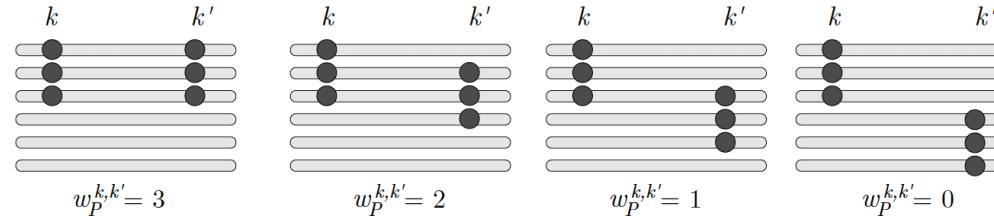
Emission function

$B_1: B$
 $B_2: b$
 $B_3: B$
 $B_4: b$
 $B_5: b$
 $B_6: b$



Linkage analysis in autopolyploids

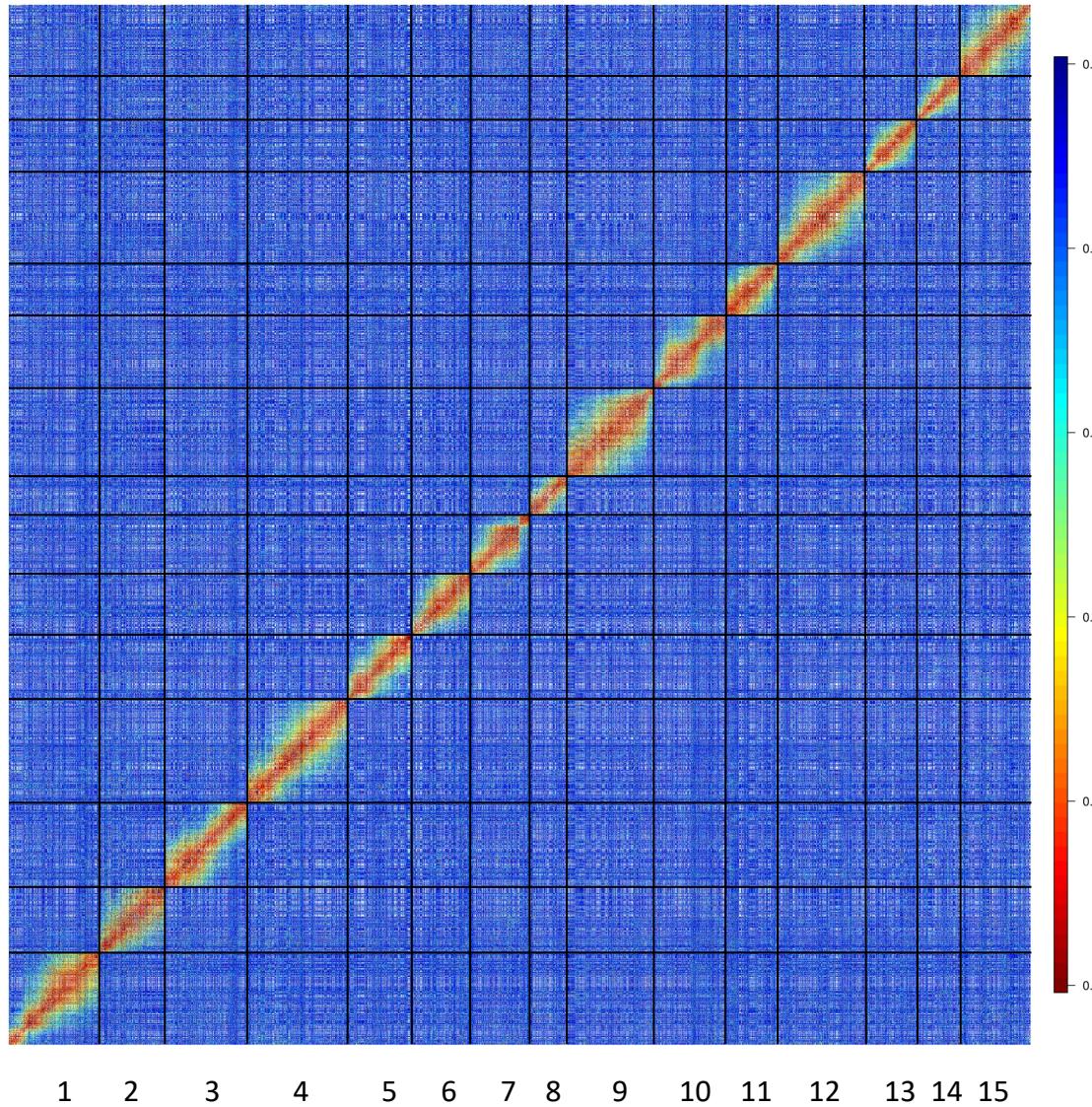
Reduction of complexity: The Two-Point Approach



	LOD_ph	rf	LOD_rf
1	0.000000	0.07239439	6.7773070858
2	2.302351	0.09233424	4.8678664544
3	2.855213	0.36392079	4.0093145291
0	6.864653	0.49994804	0.0001254893

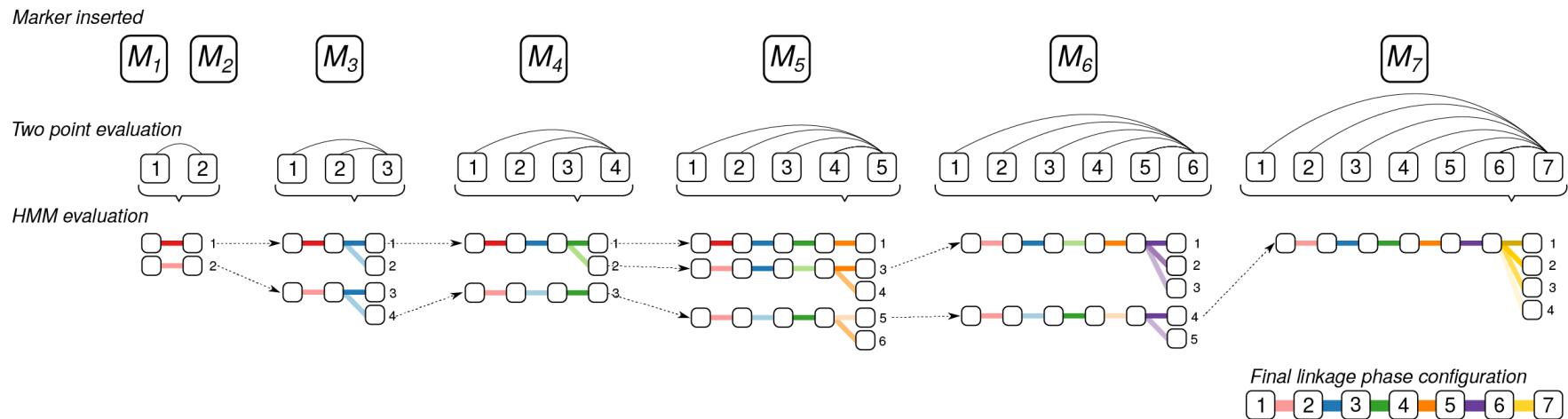
- Since the information do not propagate through the chain, it is possible to combine the multiple paths in a single edge using a computational approach.

Ordering with MDS – 15 linkage groups

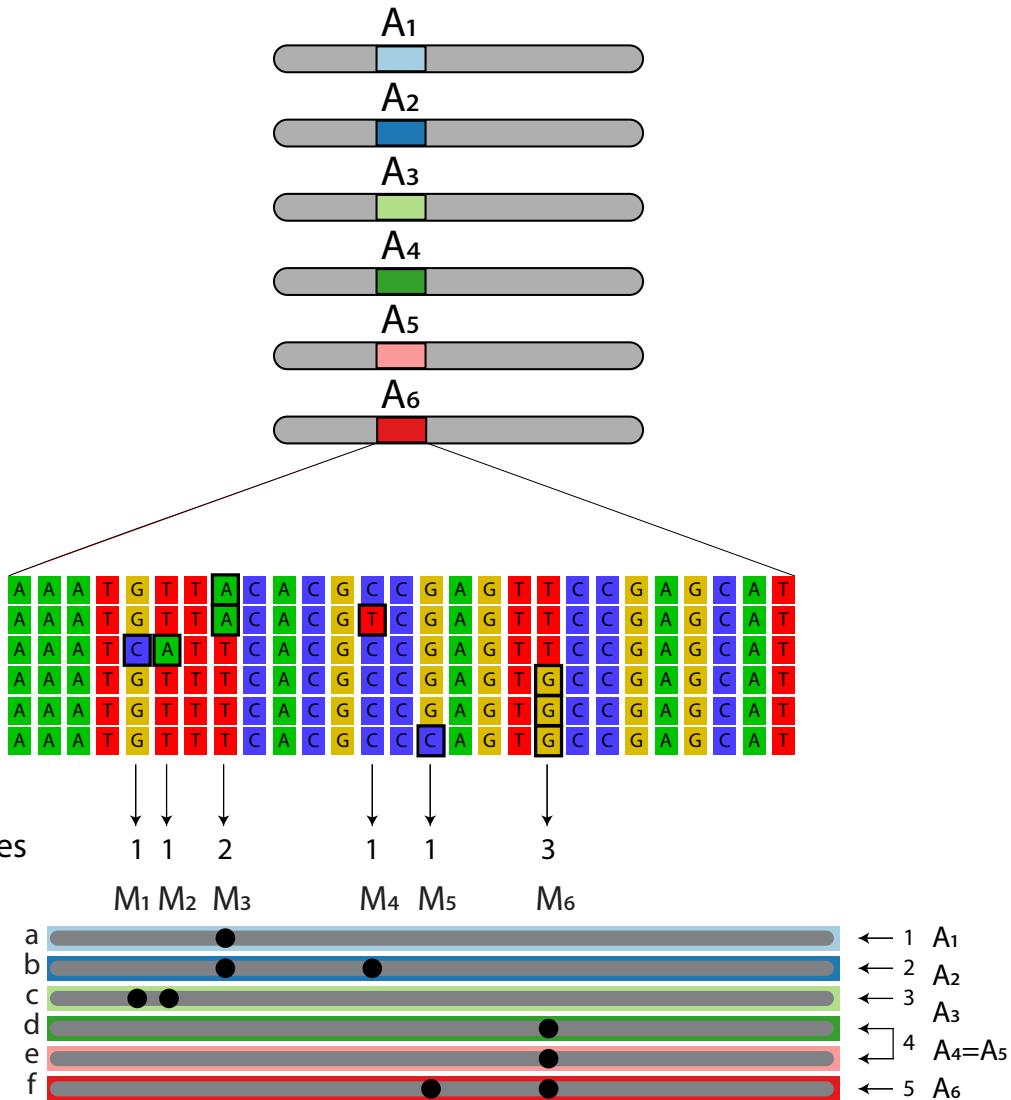


Haplotype phasing in polyploids

- Step 1: Use of two-point information to reduce the search space
- Step 2: Evaluate the remaining configurations, using HMM likelihood

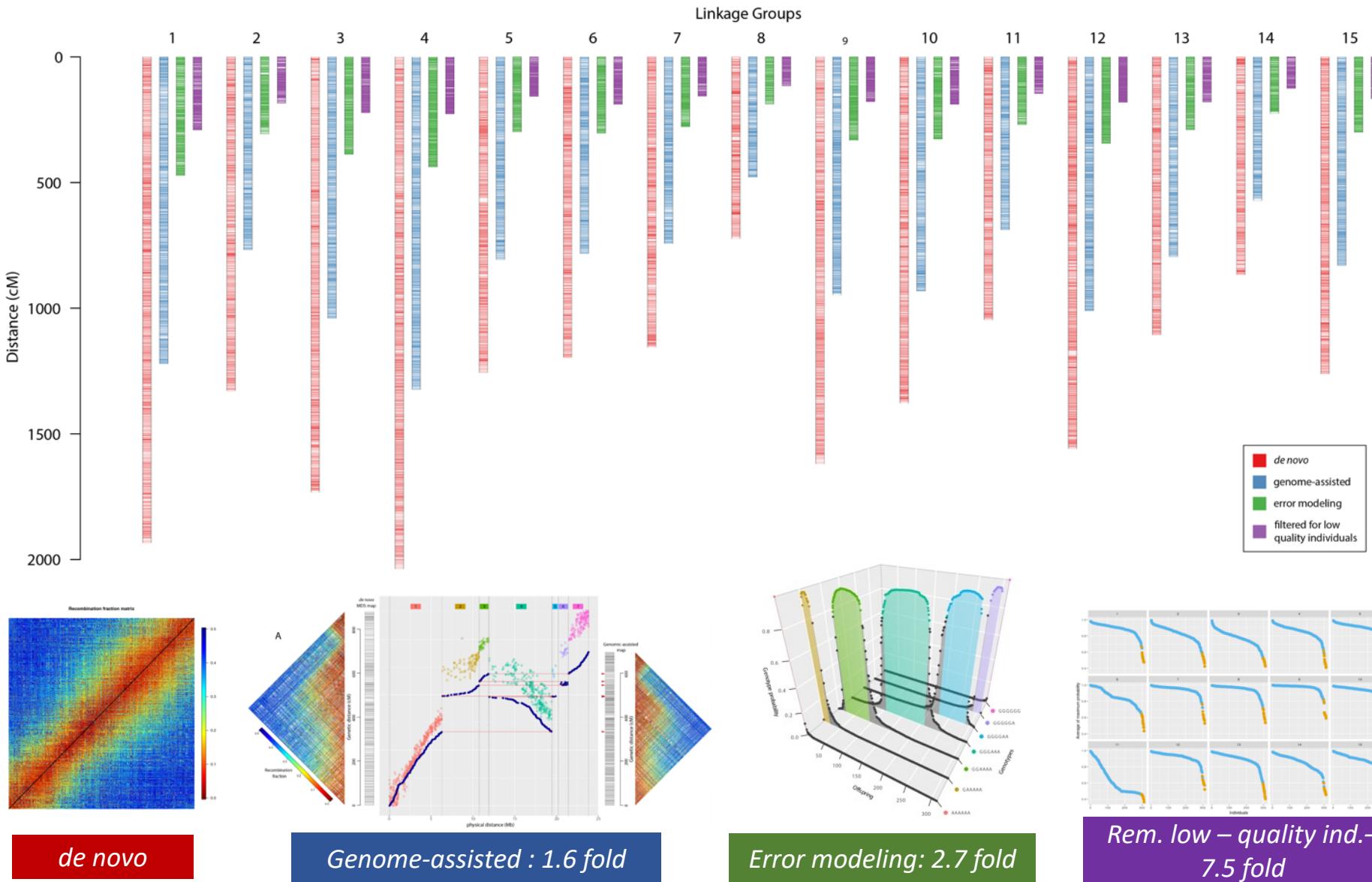


SNP calling and local haplotyping

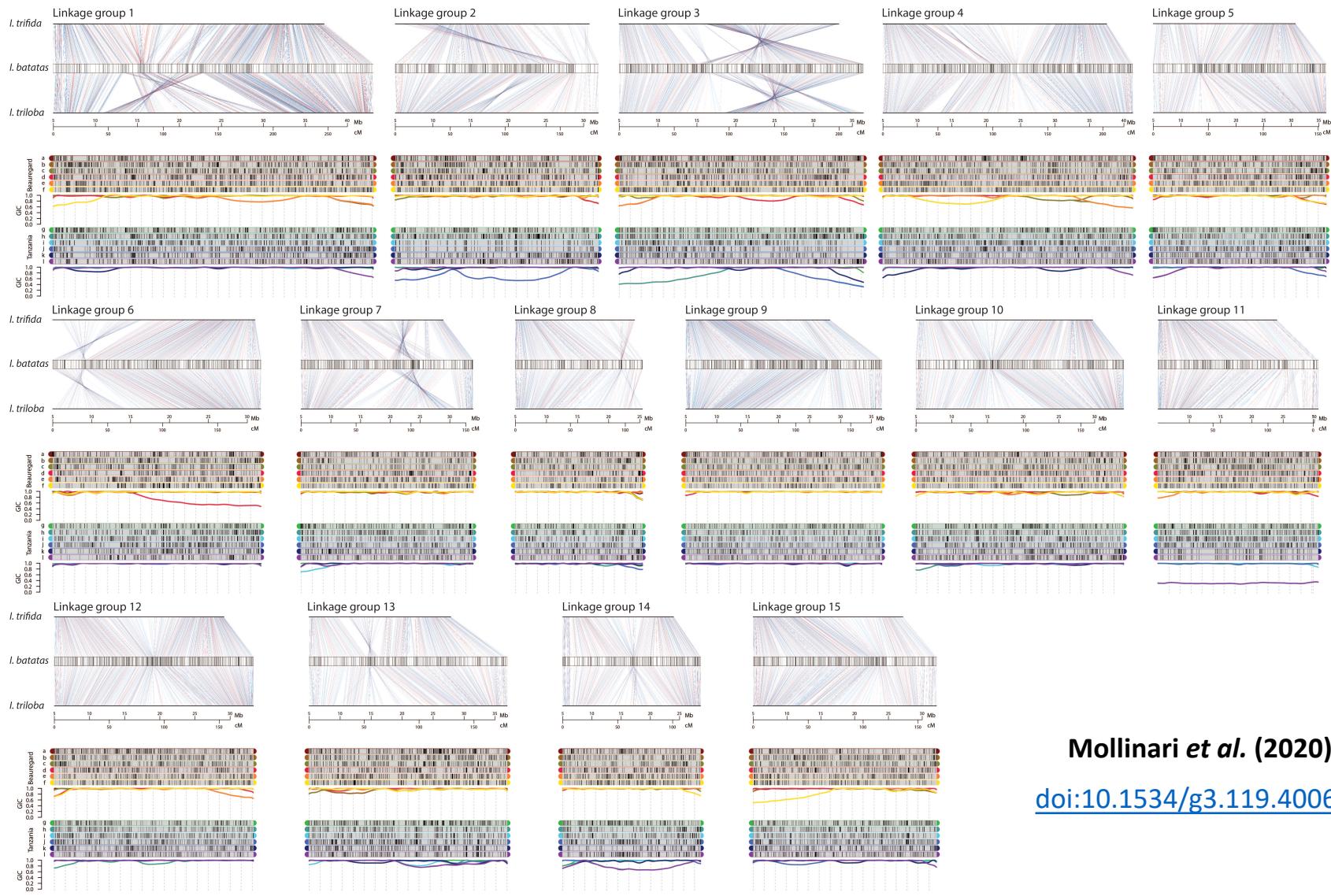


- The information is assessed at SNP level (essentially biallelic).
- It is not possible to distinguish more than two subsets of homologs.
- Combining information of several SNPs result in **haplotype-based markers**
- Mitigate **genotyping errors** (posterior state probabilities - tomorrow)

Strategies to improve map quality



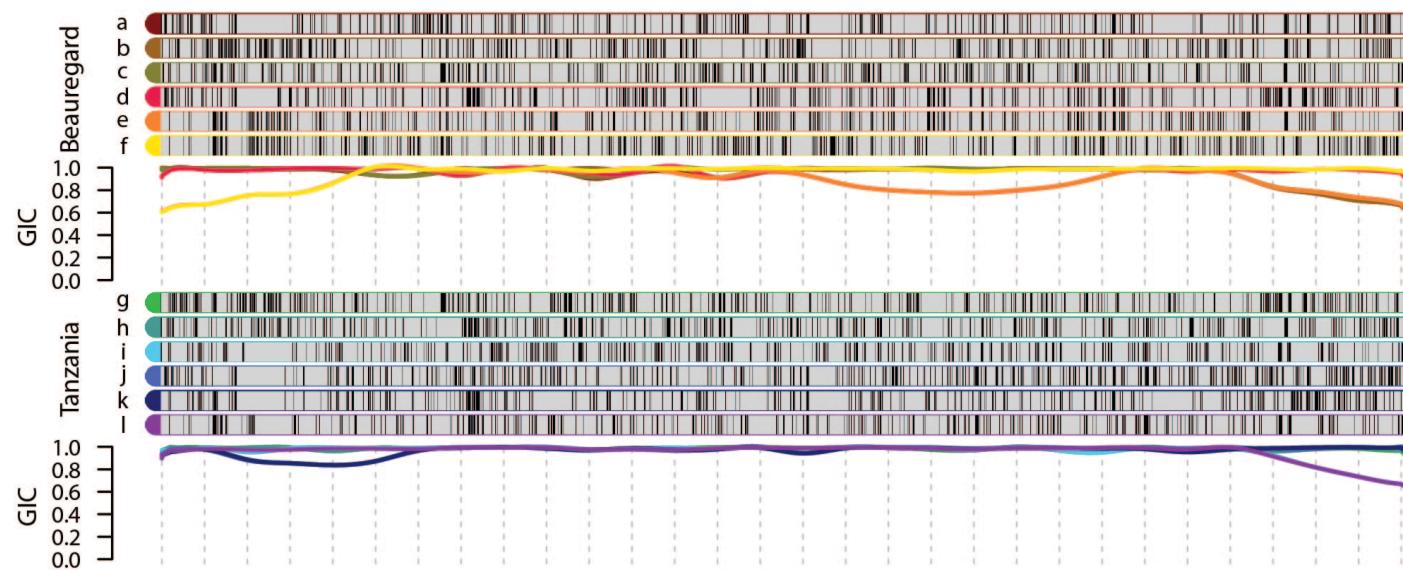
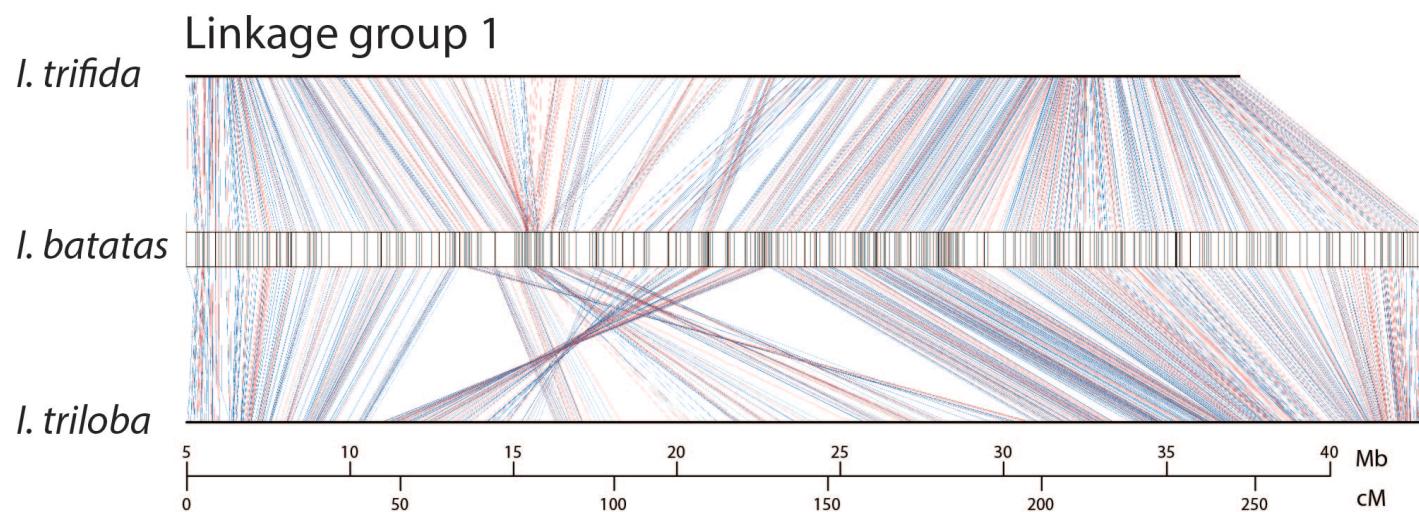
Sweetpotato genetic map



Molinari *et al.* (2020)

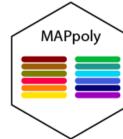
[doi:10.1534/g3.119.400620](https://doi.org/10.1534/g3.119.400620)

Sweetpotato genetic map



MAPpoly – R package to construct multipoint genetic maps in polyploids

[build passing](#) [build passing](#) [development active](#) [License GPL v3](#) [codecov 77%](#) [CRAN 0.2.0](#)



Introduction

MAPpoly (v. 0.2.0) is an R package to construct genetic maps in 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 computation.

In its current version, MAPpoly can handle three different types of datasets:

1. CSV files
2. MAPpoly files
 - Dosage based
 - Probability based
3. VCF files (beta)

The derivation of the HMM used in MAPpoly can be found in [Mollina and Garcia, 2019](#). Recently, we used MAPpoly to built an ultra-dense multilocus integrated genetic map containing ~30k SNPs and characterized the inheritance system in a sweetpotato full-sib family ([Mollina et al., 2019](#)). See the resulting map [here](#) and the haplotype composition of all individuals in the full-sib population [here](#).

MAPpoly is not available from CRAN, but you can install it from GitHub. Within R, you need to install and load the package `devtools`:

```
install.packages("devtools")
```

To install MAPpoly from GitHub use

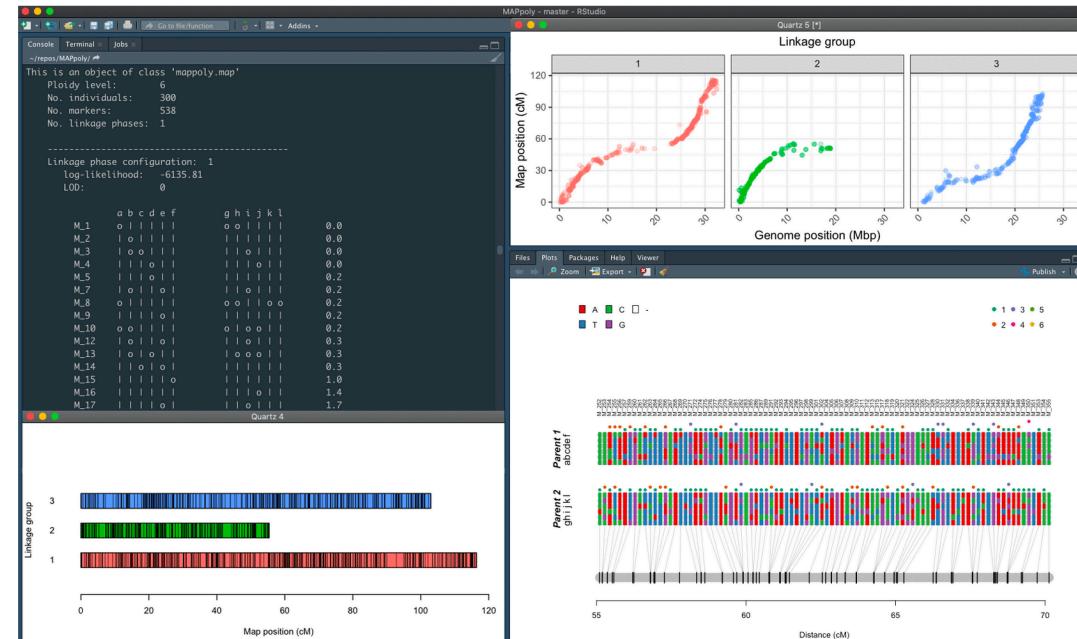
```
devtools::install_github("mmollina/mappoly")
```

Vignettes

- Building a genetic map in an hexaploid full-sib population using MAPpoly
- Building a genetic map using potato genotype data from SolCAP
- Dataset examples
 - Hexaploid simulation with dosage call in MAPpoly format
 - Hexaploid simulation with dosage probabilities in MAPpoly format
 - Tetraploid potato with dosage call in MAPpoly format
 - Tetraploid potato with dosage call in CSV format
 - Tetraploid potato with dosage probabilities in MAPpoly format

Acknowledgment

This package has been developed as part of the [Genomic Tools for Sweetpotato Improvement project \(GT4SP\)](#), funded by [Bill & Melinda Gates Foundation](#).

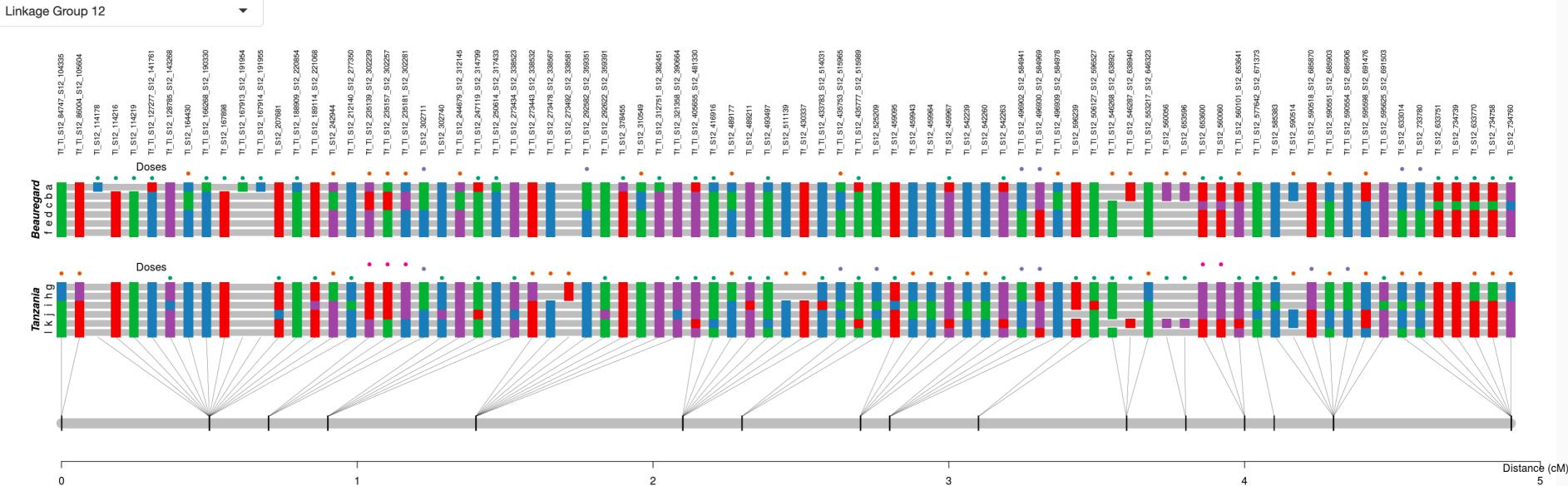


stable: <https://CRAN.R-project.org/package=mappoly>

development: <https://github.com/mmollina/MAPpoly>

Genetic mapping – Linkage group 12 – 2661 SNPs

Sweetpotato genetic map - Beauregard x Tanzania (BT)



Show SNP names?

Legend

Nucleotide	Doses
■ A	6
■ T	5
■ C	4
■ G	3
□ -	2
● 1	1

Number of SNPs per dosage

\$doses	0	1	2	3	4	5	6
0	16	12	3	0	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

Summary

```
$number.snp
[1] 81

$length
[1] 4.9

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

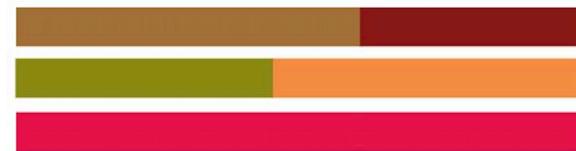
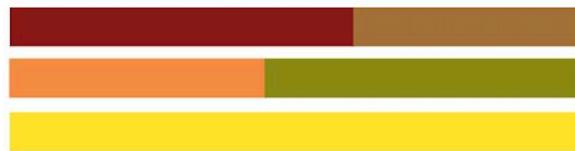
Notes

- The detailed mapping procedure is described in [Molinari et al. \(2019\)](#)
- Use the slide bar to resize or move through the map.
- The estimation of the offspring haplotype is available [here](#)

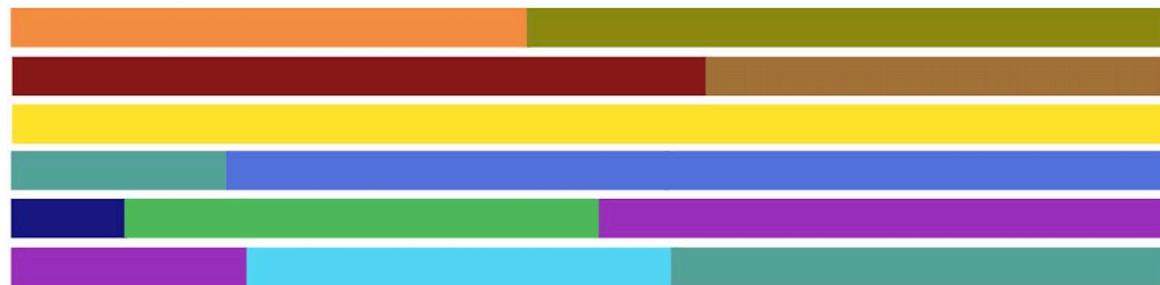
rows: Beauregard

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

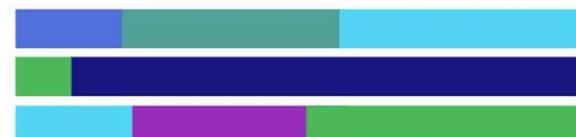
Haplotype reconstruction in the offspring



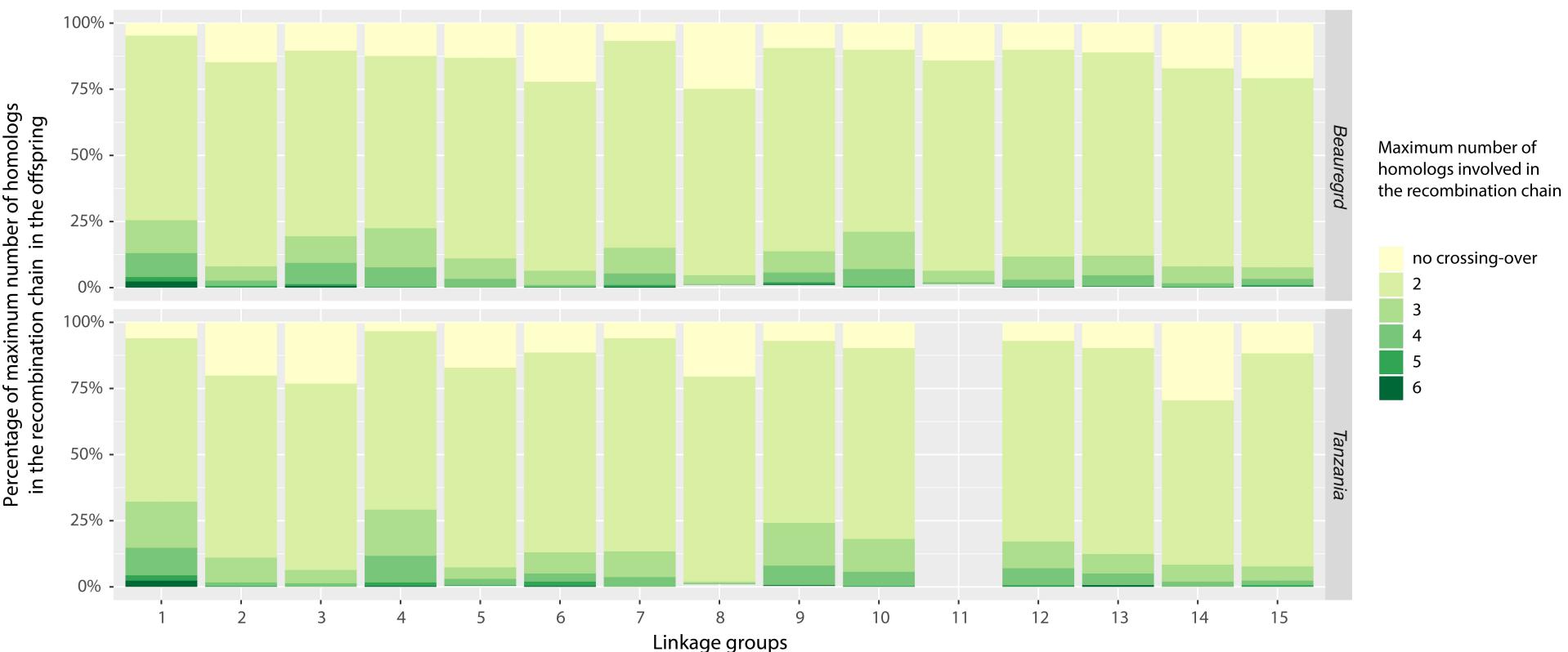
Haplotype of individual BT05:320



Resulting gametes

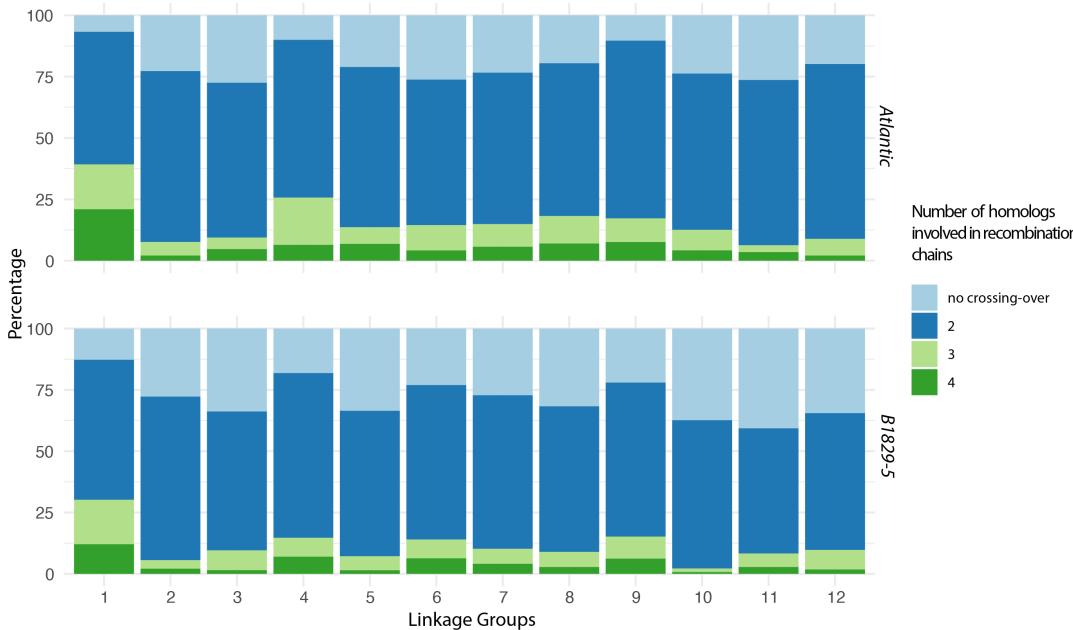


Number of homologs involved in recombination chanins



Homologs in recombination chains – potato

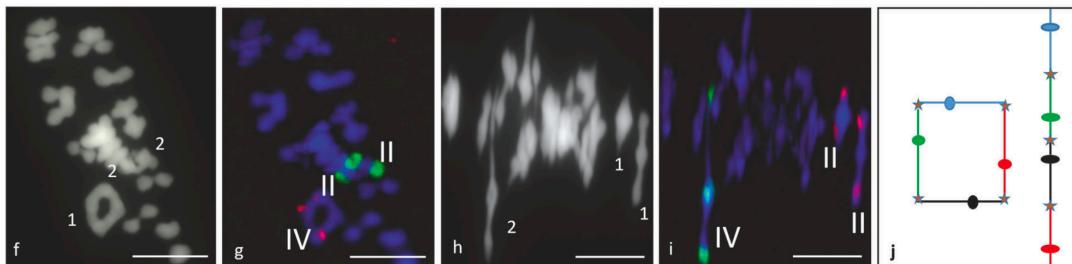
Pereira et al. (2020) - Recombination landscape in a *Solanum tuberosum* cv.



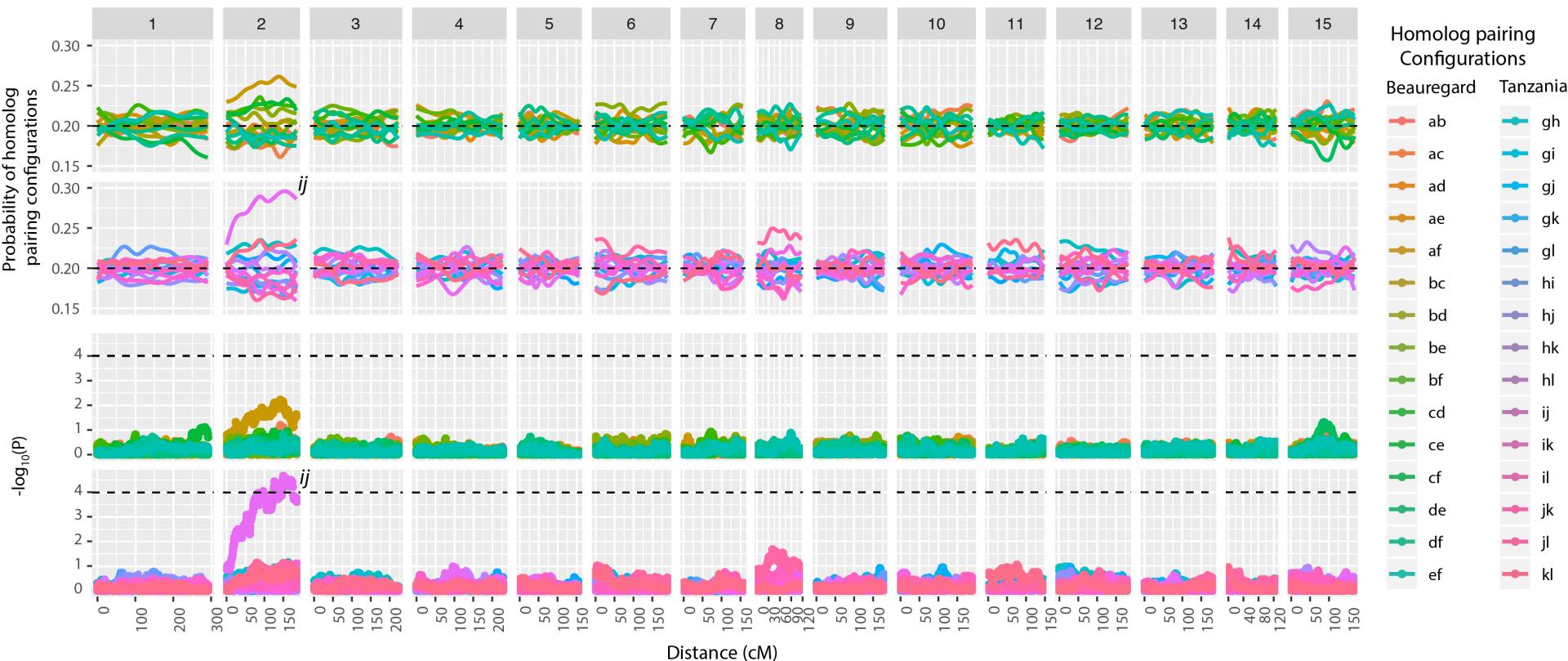
	Choudhary et al. (2020)	Pereira et al. (2020)
bivalents	predominant	62.3 %
multivalents	7~48%	2.2~39.2%

Choudhary et al. (2020) – Using fluorescence *in situ* hybridization (FISH)

5S rDNA probe (red) and 45S rDNA probe (green)



Preferential pairing profiles: Sweetpotato is vastly **auto**hexaploid



Haplotype based QTL mapping

HIGHLIGHTED ARTICLE
GENETICS | INVESTIGATION

Multiple QTL Mapping in Autopolyploids: A Random-Effect Model Approach with Application in a Hexaploid Sweetpotato Full-Sib Population

Guilherme da Silva Pereira,^{*,†,1,2} Dorcus C. Gemenet,^{‡,1} Marcelo Mollinari,^{*,†} Bode A. Olukolu,[§]
Joshua C. Wood,^{**} Federico Diaz,^{††} Veronica Mosquera,^{††} Wolfgang J. Gruneberg,^{††} Awais Khan,^{‡‡}
C. Robin Buell,^{**} G. Craig Yencho,[†] and Zhao-Bang Zeng^{*‡}

2020

*Bioinformatics Research Center and [†]Department of Horticultural Science, North Carolina State University, Raleigh, North Carolina 27695, [‡]International Potato Center, ILRI Campus, Nairobi, Kenya 25171-00603, [§]Department of Entomology and Plant Pathology, University of Tennessee, Knoxville, Tennessee 37996, ^{**}Department of Plant Biology, Michigan State University, East Lansing, Michigan 48824, ^{††}International Potato Center, Peru, Lima 1558, and ^{‡‡}Plant Pathology and Plant-Microbe Biology Section, Cornell University, Geneva Campus, New York 14456

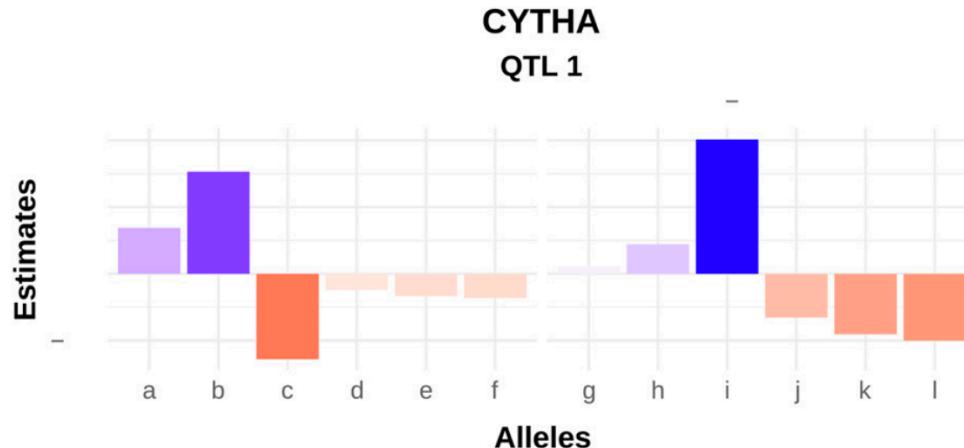


Figure 4 Additive allele effects from the decomposed best linear unbiased predictions (BLUPs) for the QTL 1 (on linkage group 15 at 5.27 cM) of commercial root yield in $t\ ha^{-1}$ (CYTHA) in a hexaploid sweetpotato full-sib family ('Beauregard' × 'Tanzania'). Letters represent each of the six haplotypes from each parent.

Conclusions

- Although compute-intensive, the construction of **multipoint genetic maps** in polyploids has several advantages, including error modeling, computation of preferential pairing profiles, and comprehensive construction of haplotypes in the offspring.
- The correct assessment of haplotypes is fundamental to further genetic studies (QTL, genomic prediction, etc)
- All procedures are implemented in the R package **MAPpoly** available at <https://CRAN.R-project.org/package=mappoly>

Acknowledgements

- Zhao-Bang Zeng (Bioinf. Res. Center - NCSU)
- Guilherme Pereira (CIP)
- Gabriel Gesteira (ESALQ/USP)
- Antonio Augusto Franco Garcia (ESALQ/USP)
- Bode Olukolu (University of Tennessee)
- Craig Yencho (Horticultural Science - NCSU)
- Dorcus C. Gemenet (CGRI - Excellence in Breeding)



BILL & MELINDA
GATES foundation