

# Genetic models applied to polyploid species

Marcelo Mollinari  
[mmolin@ncsu.edu](mailto:mmolin@ncsu.edu)

BILL &  
MELINDA  
GATES  
foundation



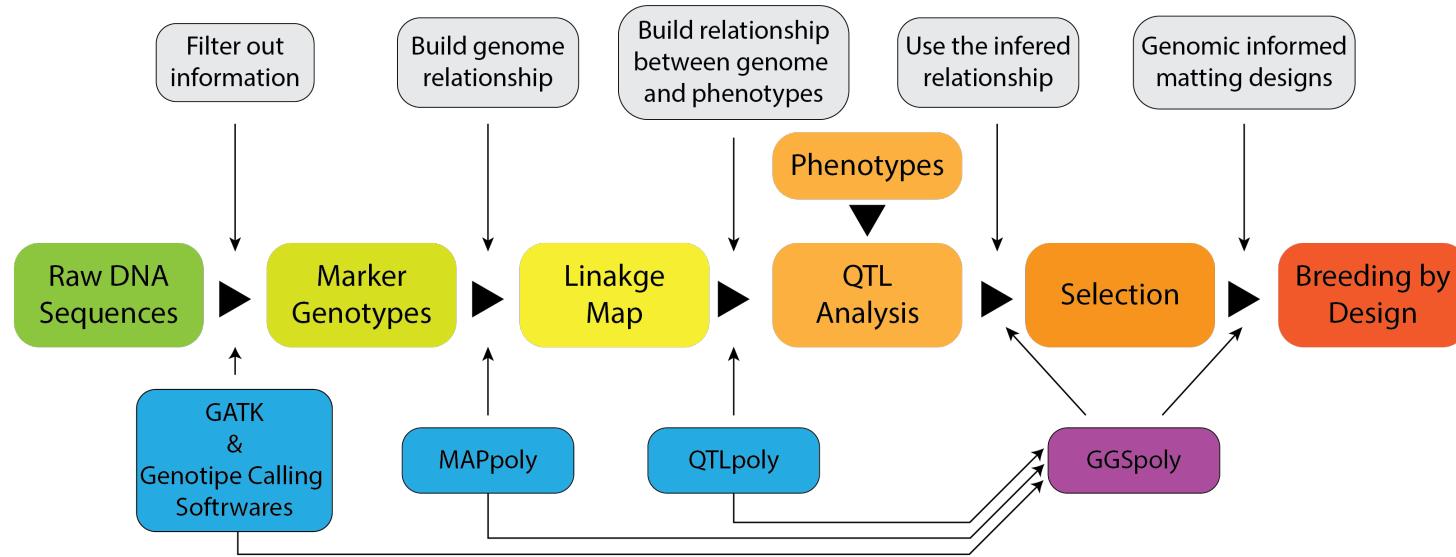
# Polyploids



Polyploids are organisms that have **multiple copies** of the complete set of chromosomes. This multiplicity imposes numerous **challenges to genomic analysis** and subsequent applications in practical breeding.

# What is our objective?

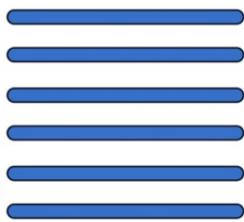
- Develop computational tools for the whole pipeline data analysis



- **MAPpoly**: R package for constructing a complete linkage map for 2X, 4X, 6X, 8X.
- **QTLpoly**: R package for QTL mapping in full-sib families for 2X, 4X, 6X, 8X
- **GGSpoly**: Working closely with polyploid breeders, we are implementing **GGSpoly**: a user-friendly **computational tool** to assist breeders in making long and short-term **breeding decisions** based on collected and learned **information** about their breeding populations.

# Polyplloid species

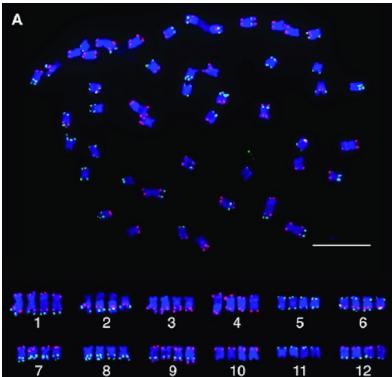
Autohexaploid



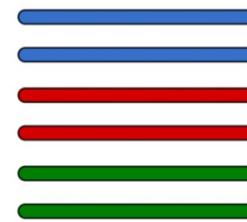
Multiple sets have the **same** origin



Autotetraploid potato



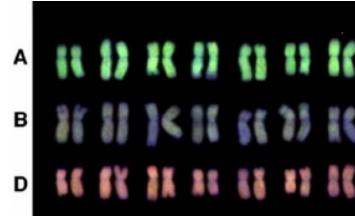
Allohexaploid



Multiple sets have **different** origins

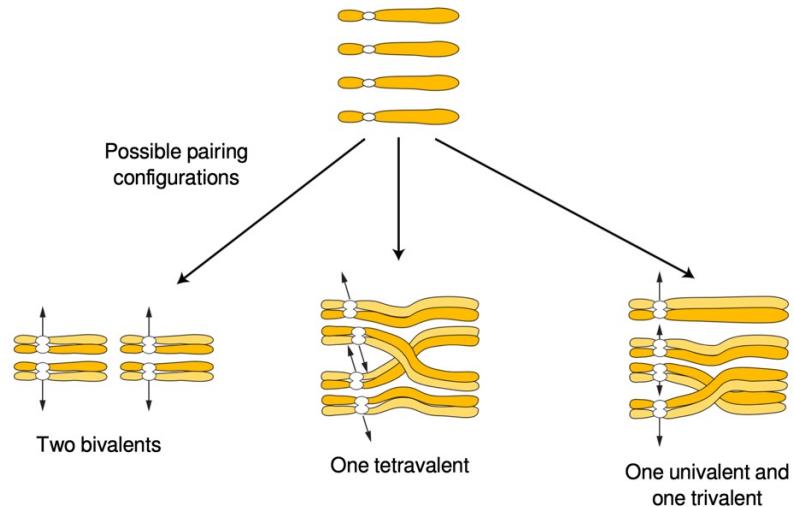


Allohexaploid wheat



Zhang et al., 2013

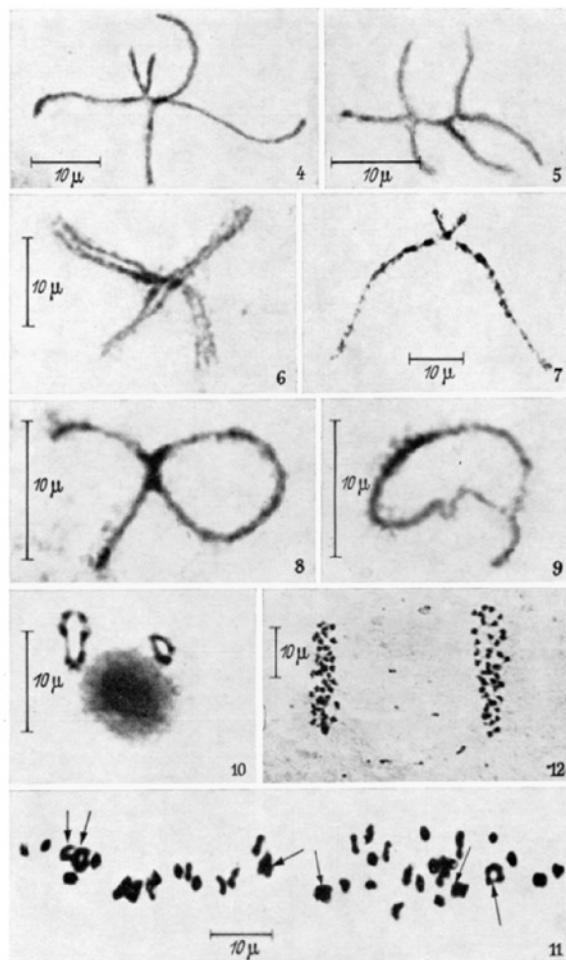
Meiotic pairing in autotetraploids



Griffiths et al. (2004)

# Meiotic Pairing in autopolyploids

Sweetpotato



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

Sugarcane

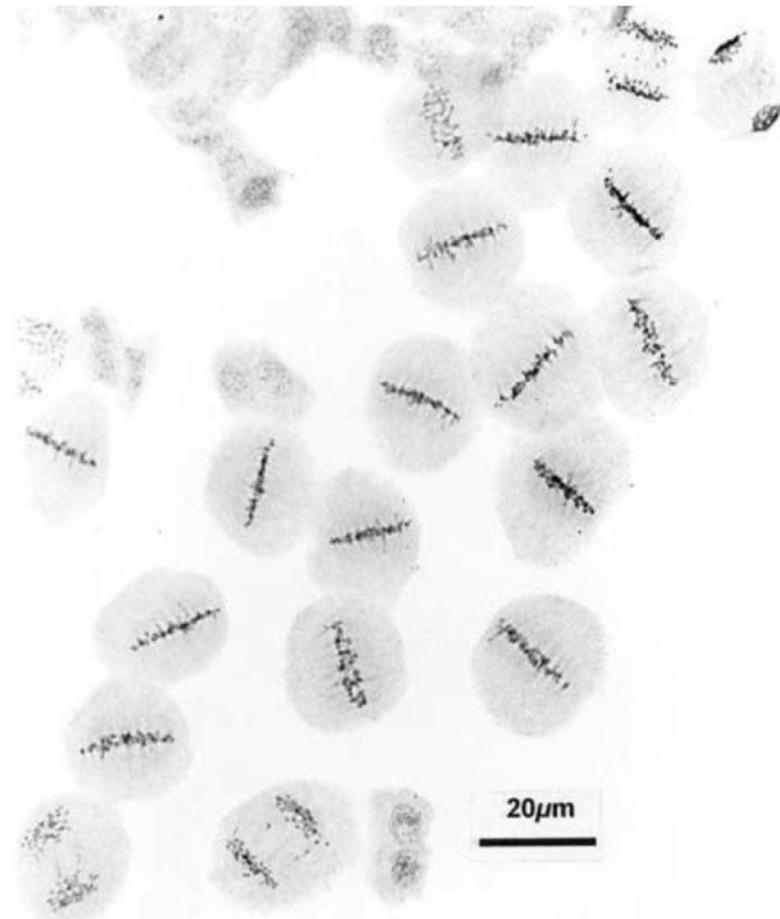


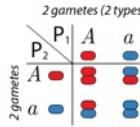
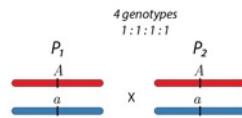
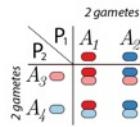
Figure 1. PMCs of *Saccharum* spp. hybrid clone 79N9059 at meiosis. As was the case in other clones, pairing was regular, bivalents generally formed. (Bielig *et al.* 2003)



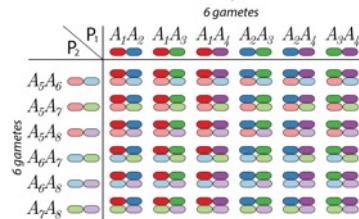
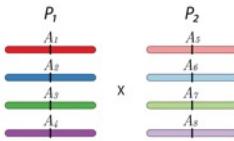
# Segregation in polyploids\*

Multiallelic

Diploid

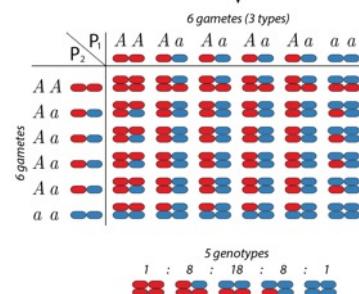
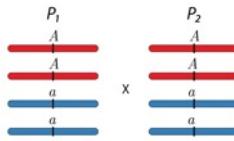
3 genotypes  
1 : 2 : 1

Tetraploid



36 genotypes

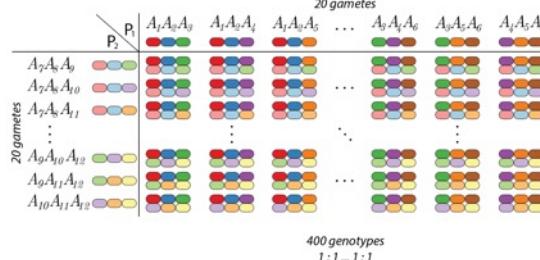
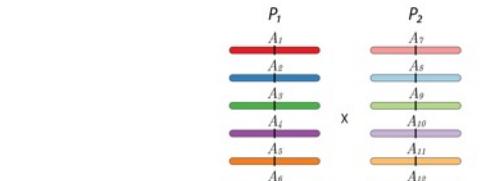
1 : 1 ... 1 : 1



5 genotypes

1 : 8 : 18 : 8 : 1

Hexaploid

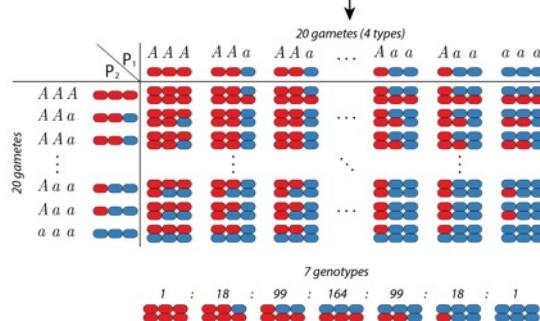
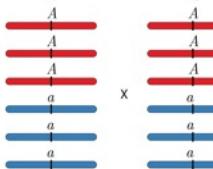


20 gametes

400 genotypes

1 : 1 ... 1 : 1

Hexaploid



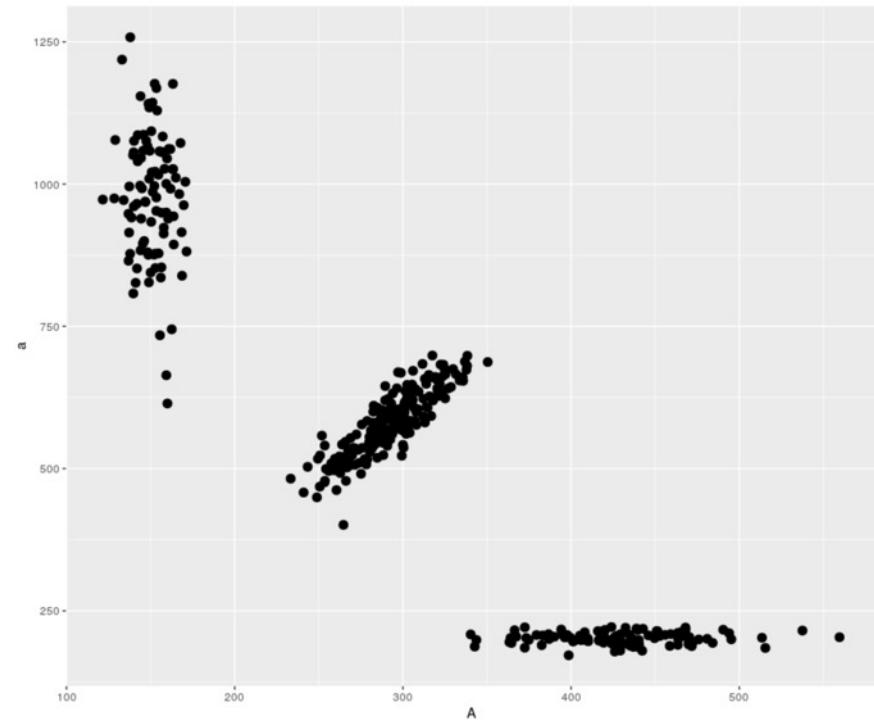
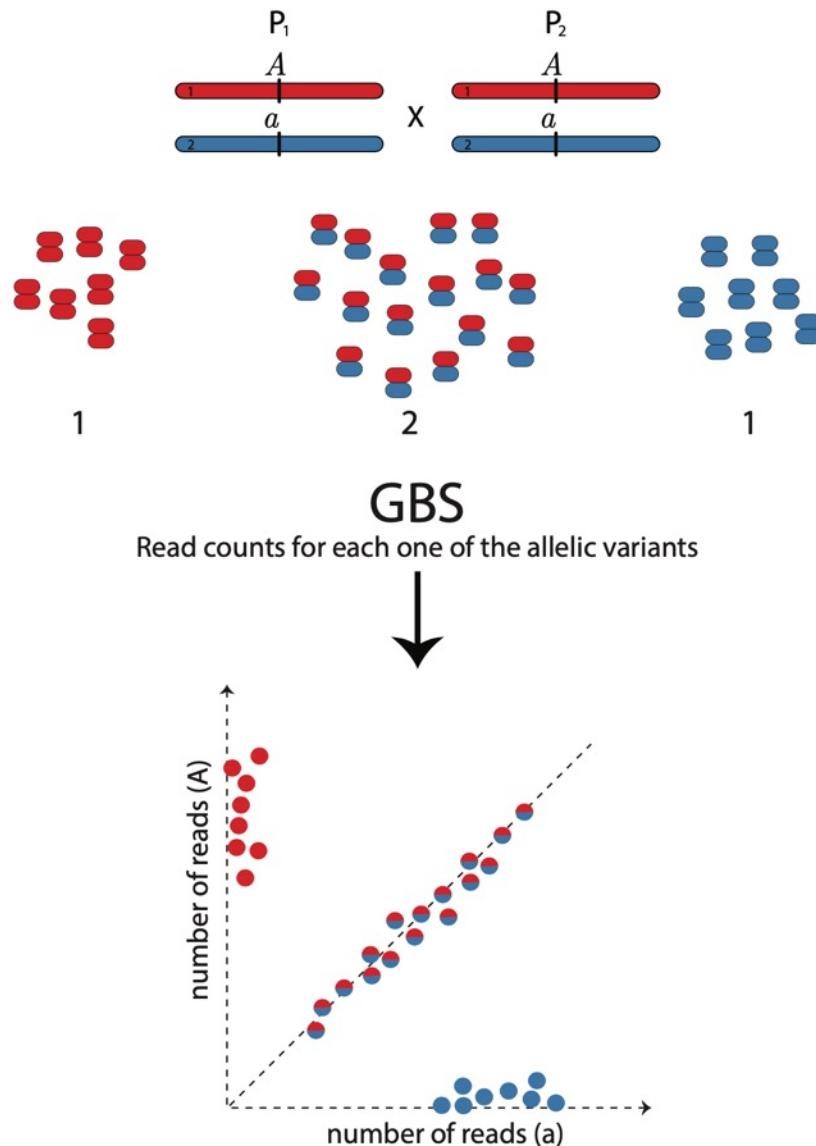
7 genotypes

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



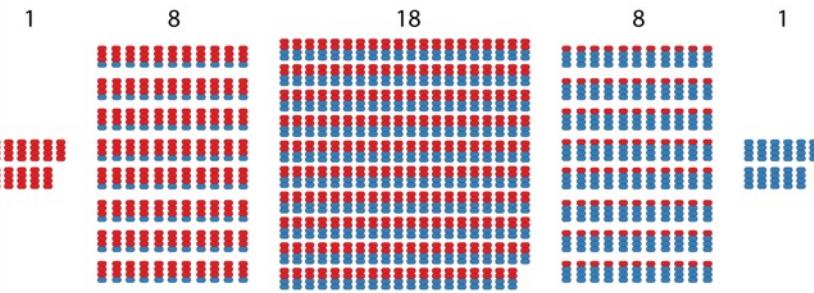
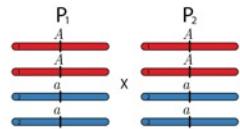
\*random pairing and no double reduction

# Genotype calling in diploids

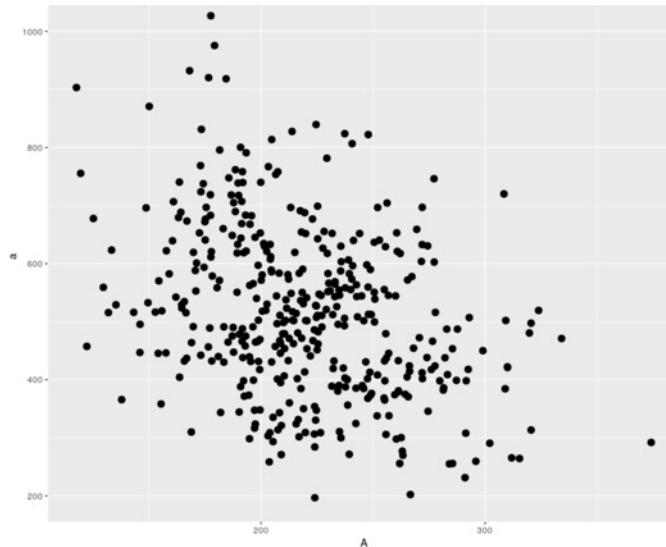
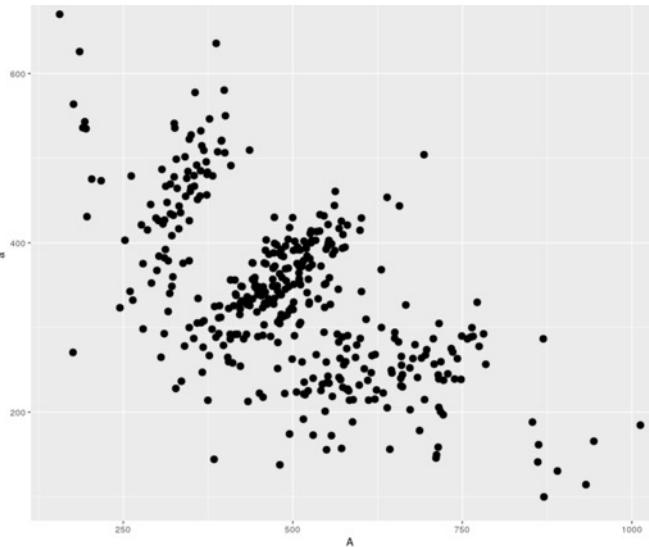
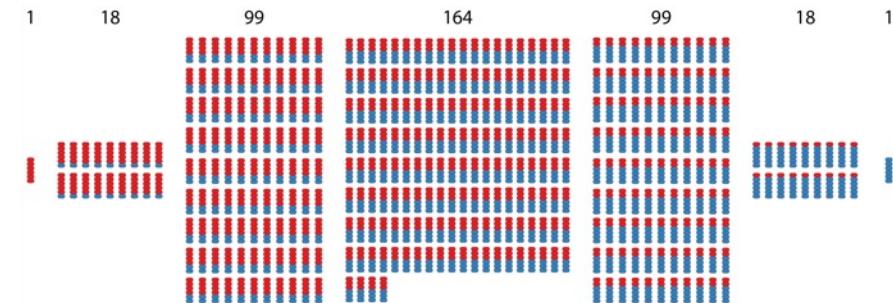
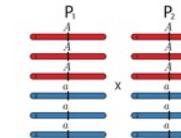


# Genotype calling in polyploids

Tetraploid

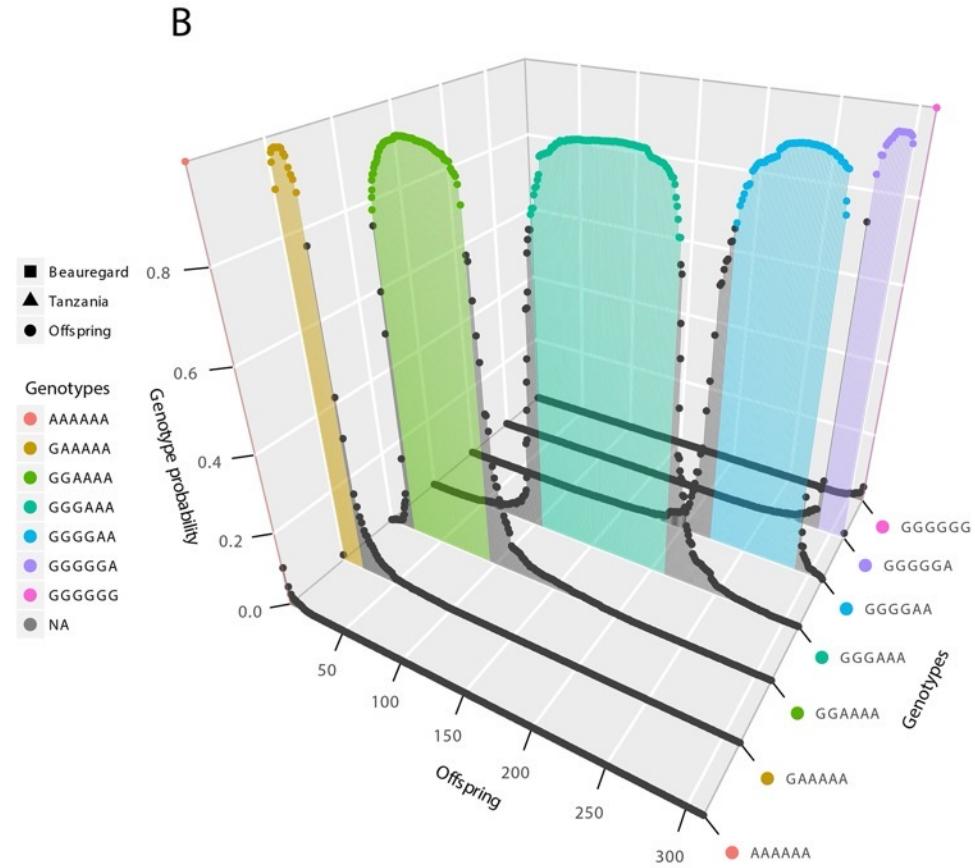
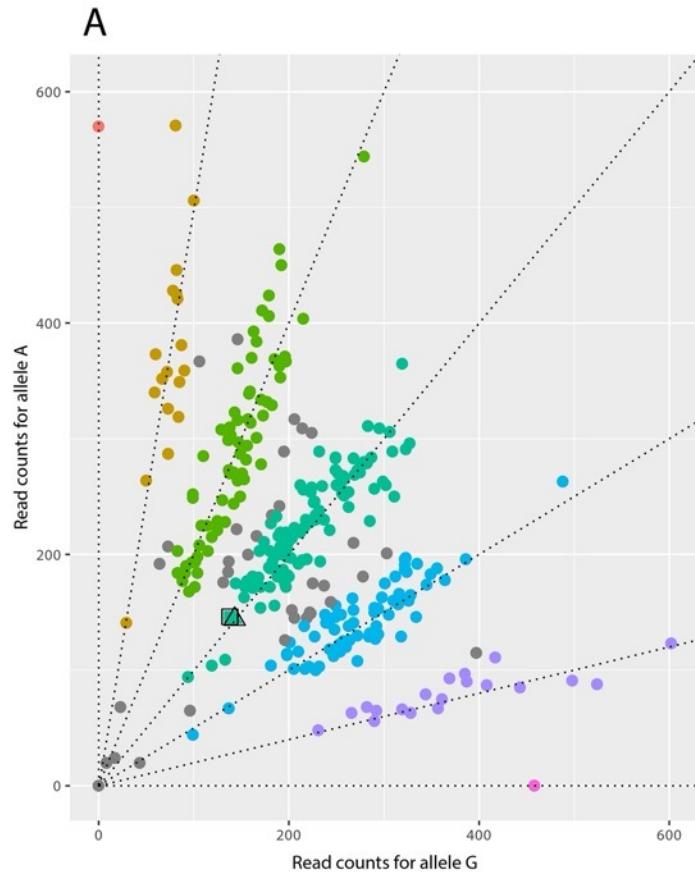


Hexaploid



# Genotyping Calling using SuperMASSA

Dosage calling including the probability distribution of the genotypes



Beauregard: 3 doses

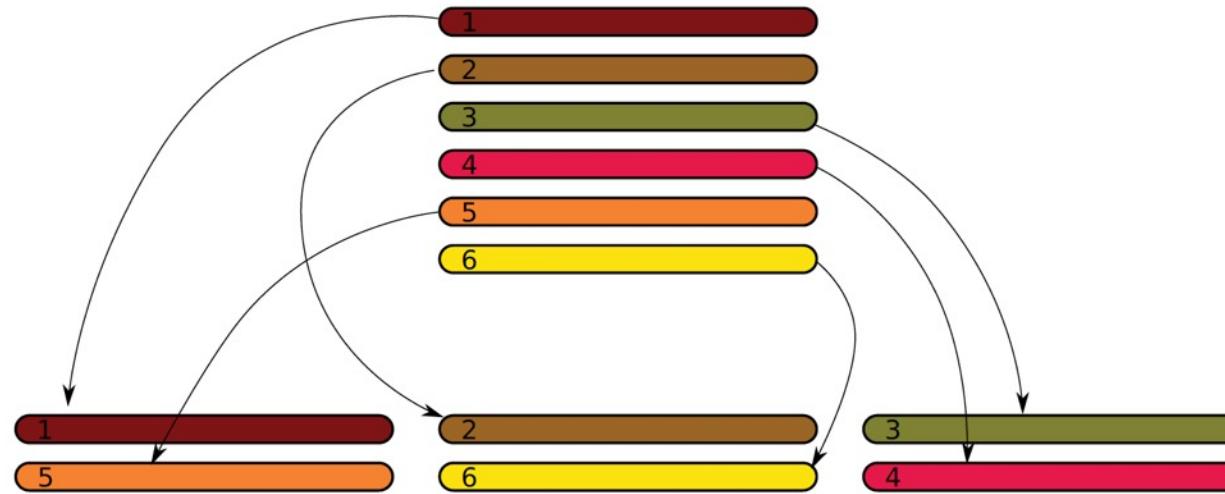
Tanzania: 3 doses

# Phasing and mapping

## The allelic transmission model



# Gamete formation in polyploids\*



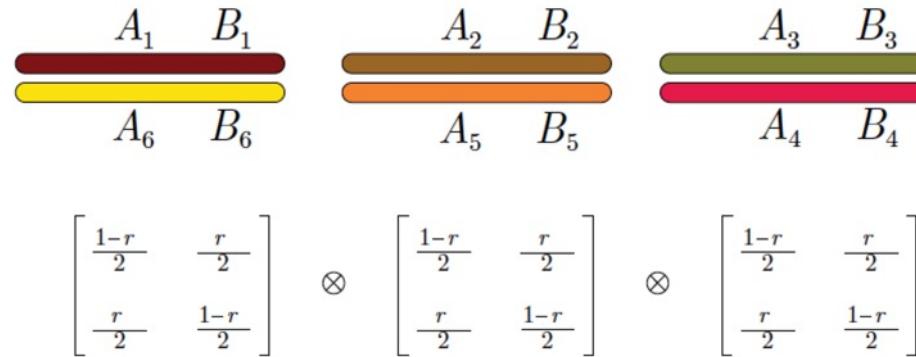
$$\psi_4$$

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

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

\*no double reduction

# Expected gametic frequency given a bivalent configuration



In general:

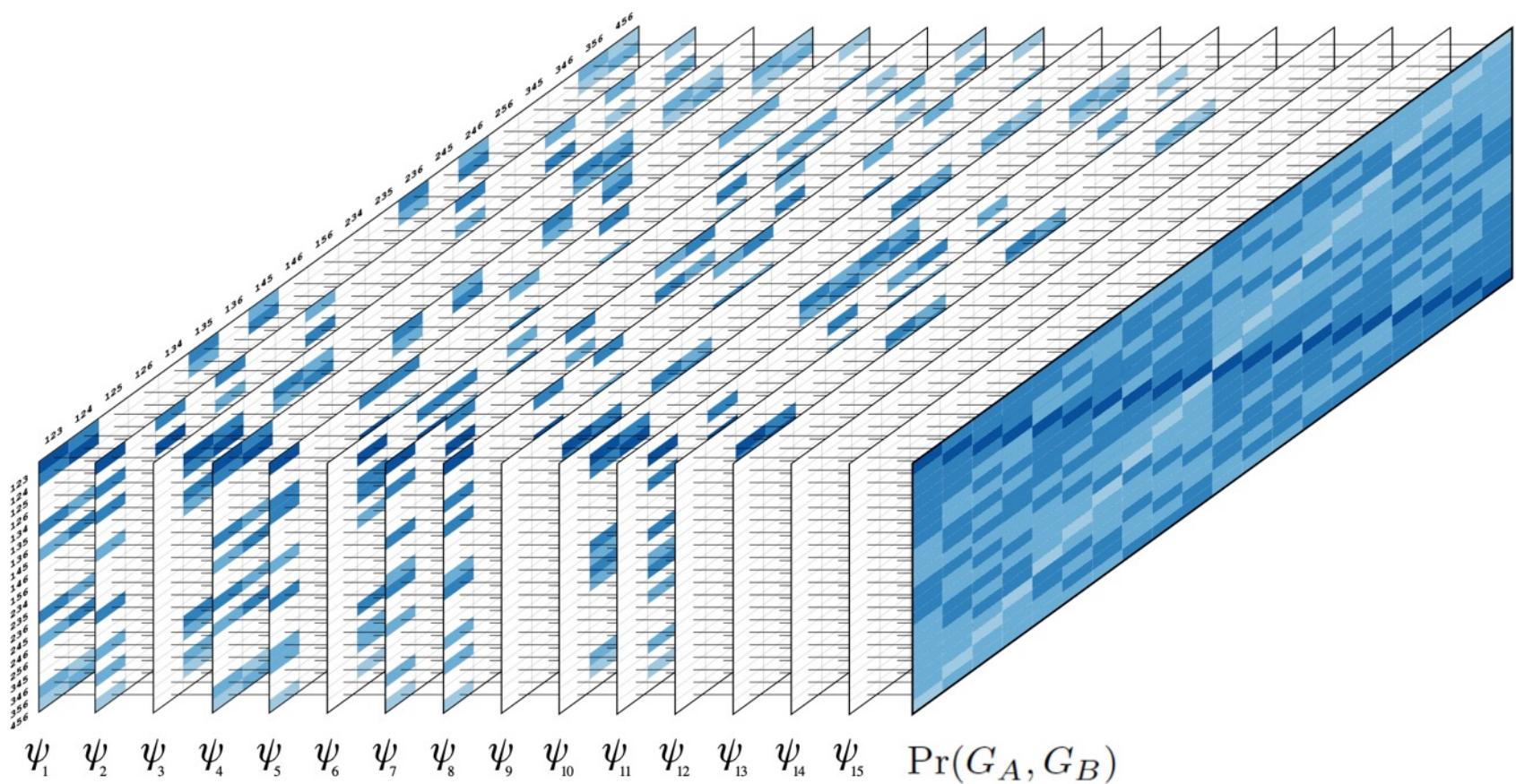
$$\mathbf{V}_1 \otimes \mathbf{V}_2 \otimes \cdots \otimes \mathbf{V}_{\frac{p}{2}}$$

All elements of this product are of the form

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

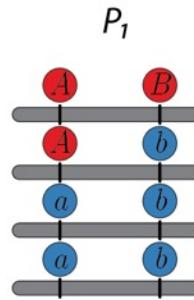
$\gamma$ : known number of recombinant bivalents between loci A and B

# Unconditional gametic probability



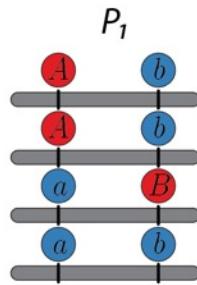
$$\begin{aligned}\Pr(G_A, G_B) &= \sum_j \Pr(G_A, G_B | \psi_j) \Pr(\psi_j) \\ &= \frac{l! \left(\frac{m}{2} - l\right)!}{w_m} \frac{(1 - r_k)^{\frac{m}{2} - l} (r_k)^l}{2^{\frac{m}{2}}}\end{aligned}$$

# Recombination Fraction – assessing linkage phases



$$\hat{r}_1 = \underset{r}{\operatorname{argmax}} L_1(r) \implies L_1(\hat{r}_1)$$

Compare likelihoods  
choosing the **most likely**  
configuration

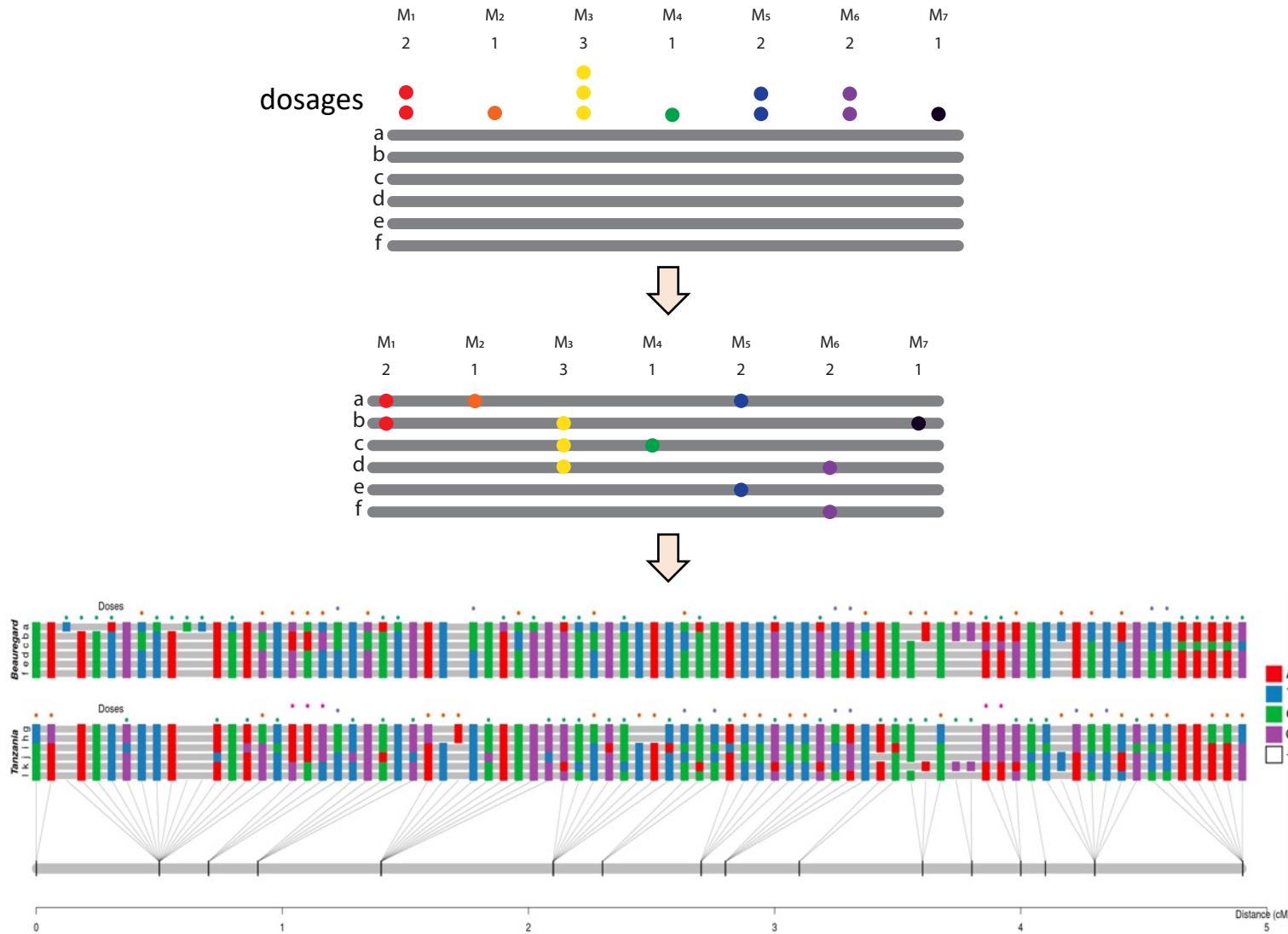


$$\hat{r}_2 = \underset{r}{\operatorname{argmax}} L_2(r) \implies L_2(\hat{r}_2)$$

- Pairwise MLE 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 likelihoods of competing linkage phases throughout multiple markers

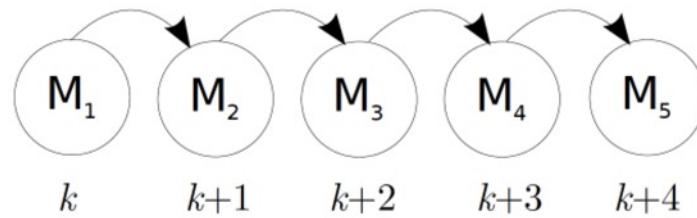
# Haplotype phasing in polyploids

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



# Multilocus linkage analysis in polyploids

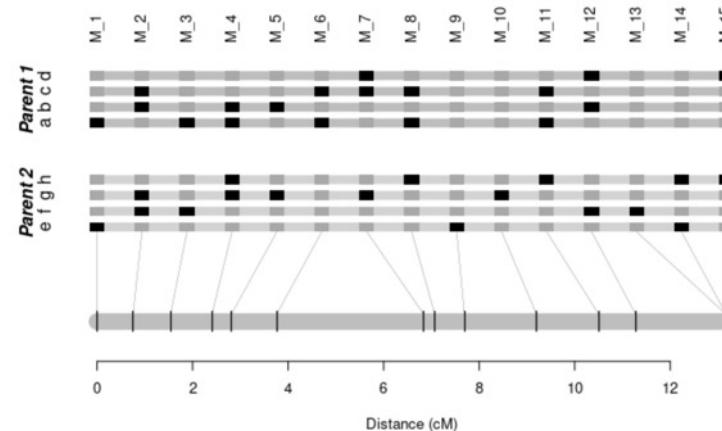
Markov Model: conditional independence



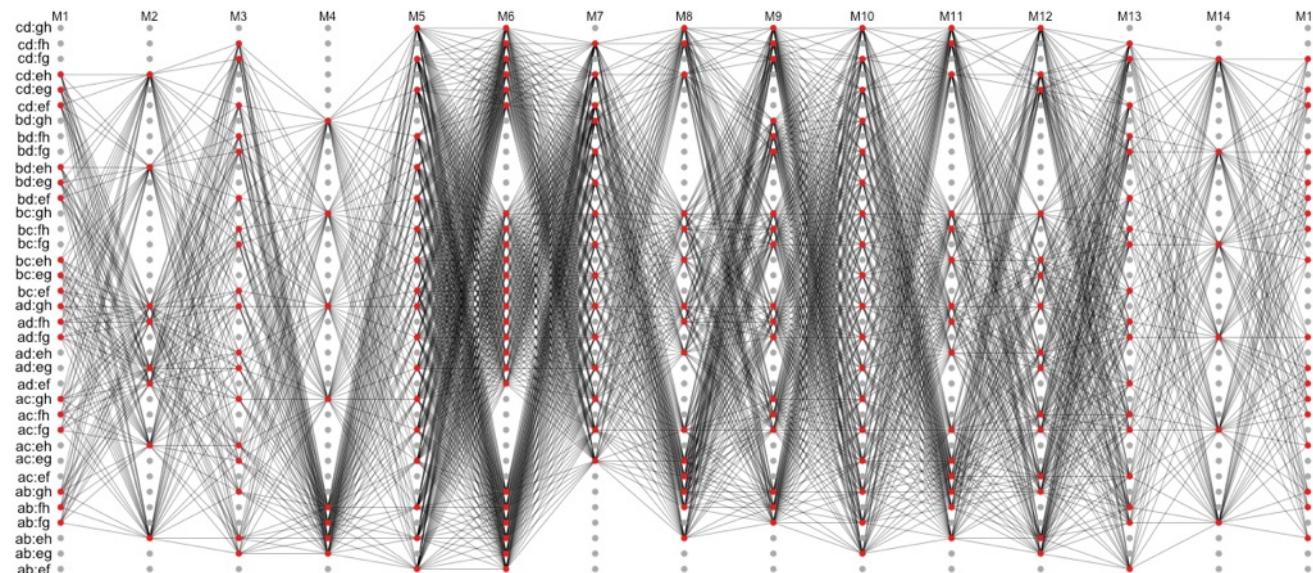
$$\Pr(G_{k+1}|G_k) = \frac{(1 - r_k)^{\frac{p}{2} - l}(r_k)^l}{\binom{\frac{p}{2}}{l}}$$

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  $k+1$ .

# Hidden Markov Model - HMM

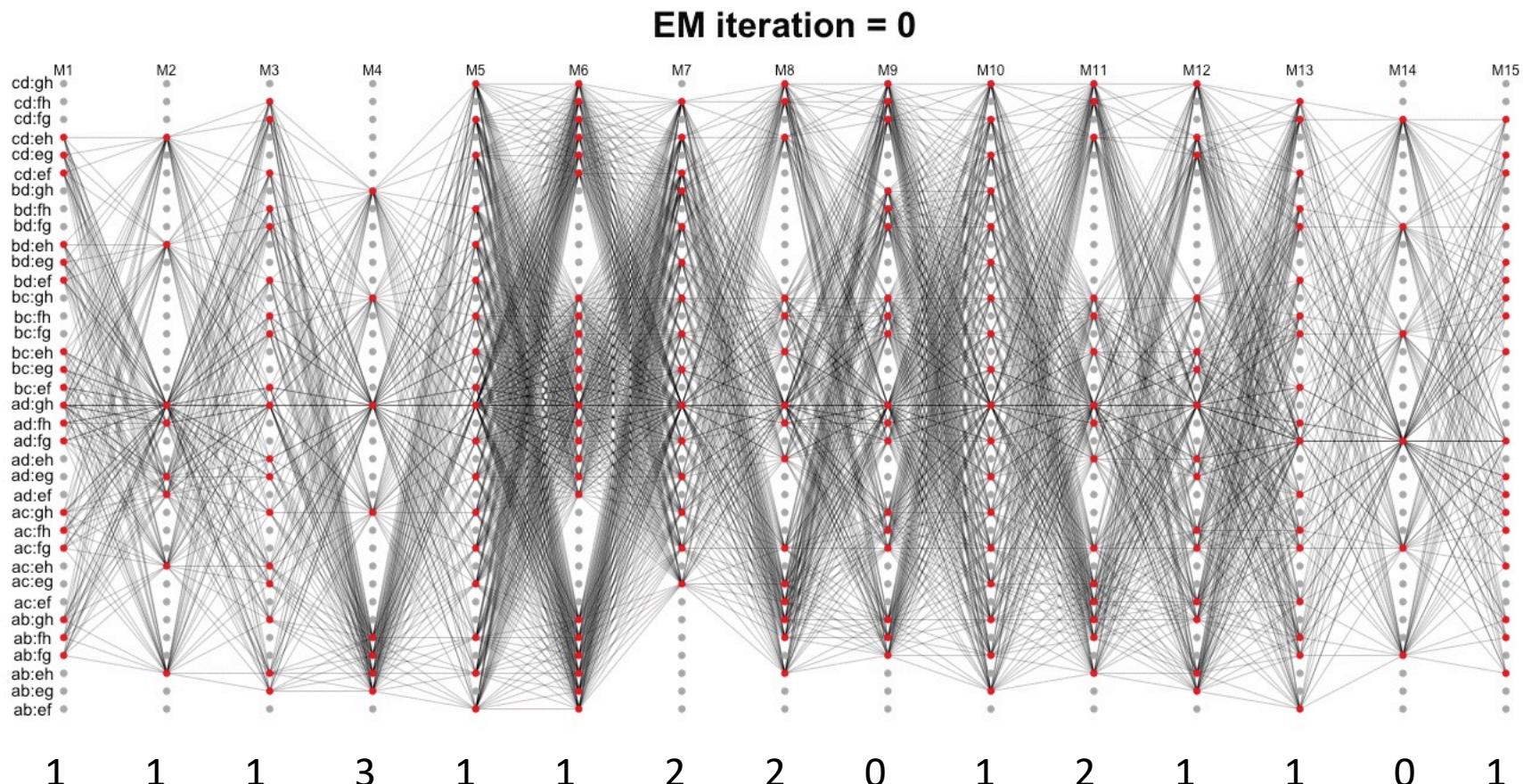


Individual 64: 1 1 1 3 1 1 2 2 0 1 2 1 1 0 1



# Hidden Markov Model - HMM

- Tetraploid example, one individual, 15 markers



# 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



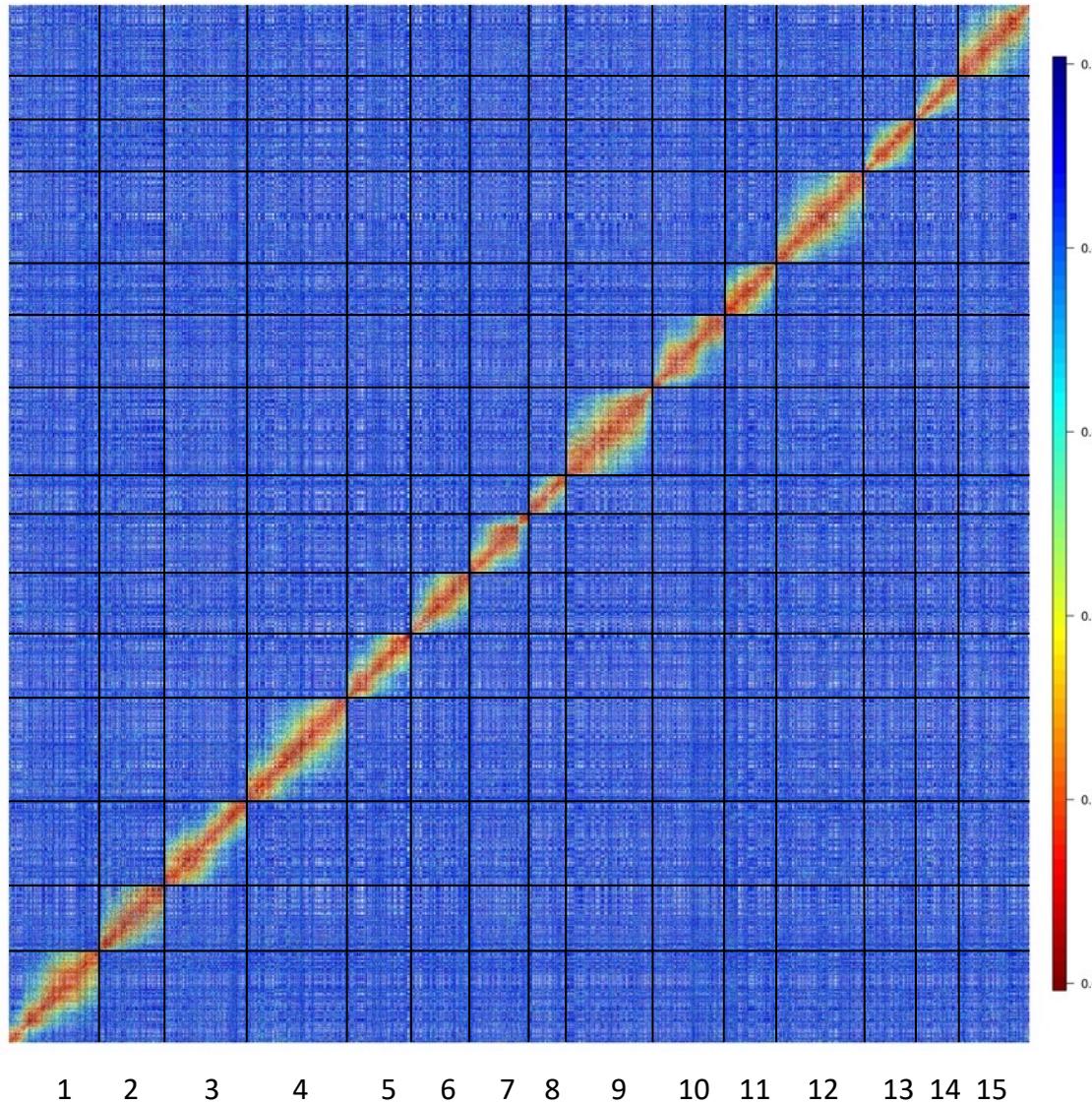
Beauregard



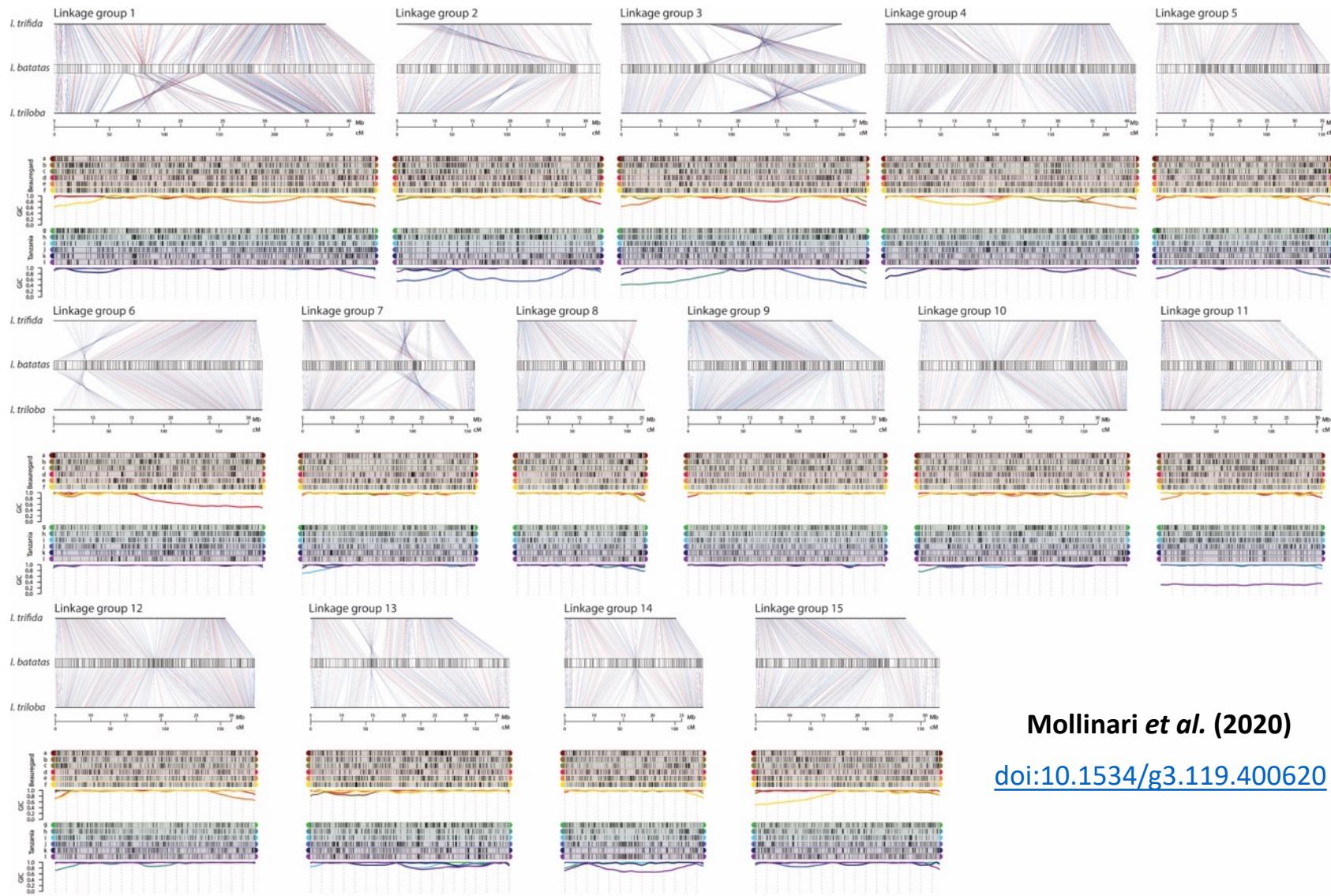
Tanzania



# Ordering with MDS – 15 linkage groups



# Sweetpotato genetic map

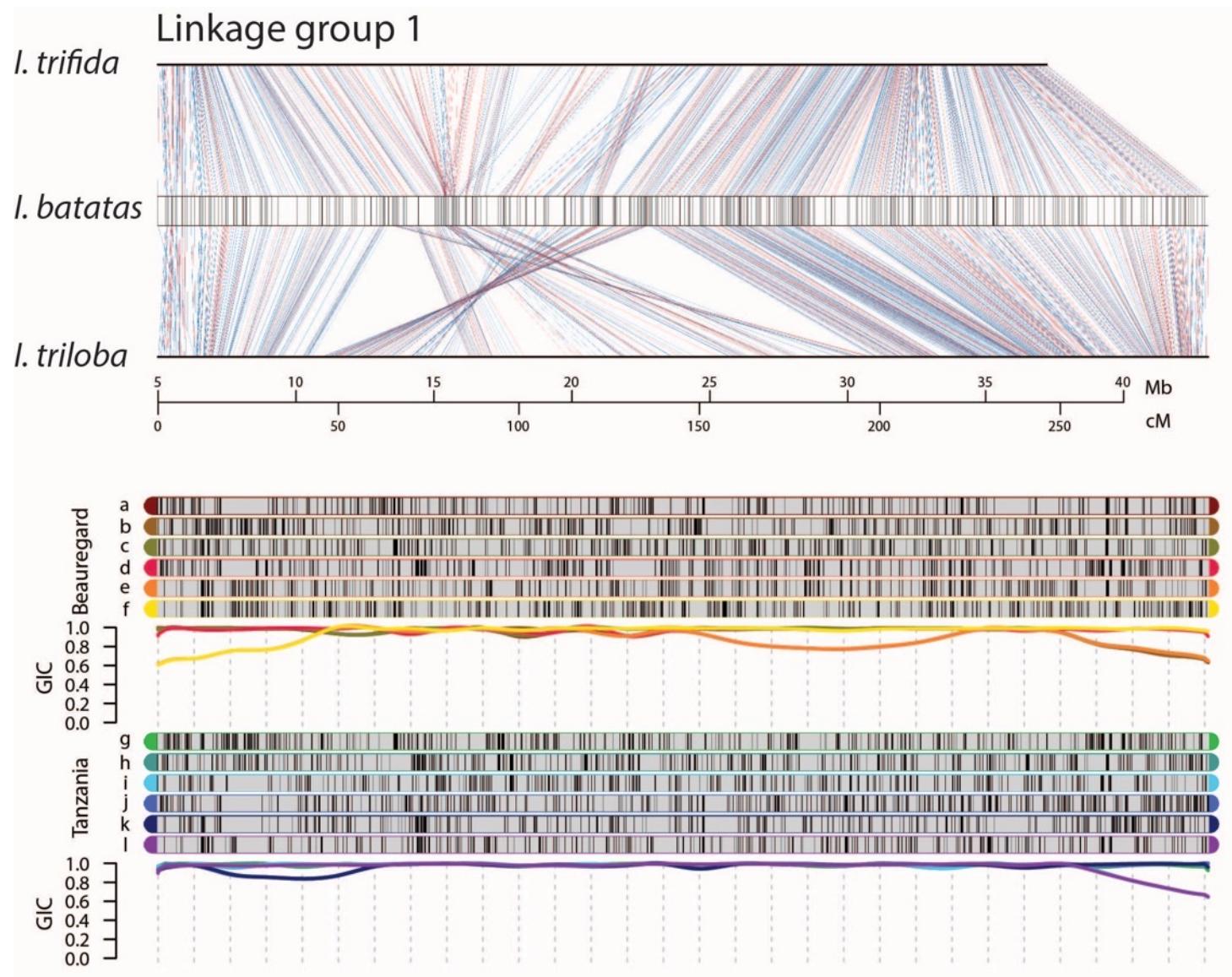


Molinari *et al.* (2020)

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



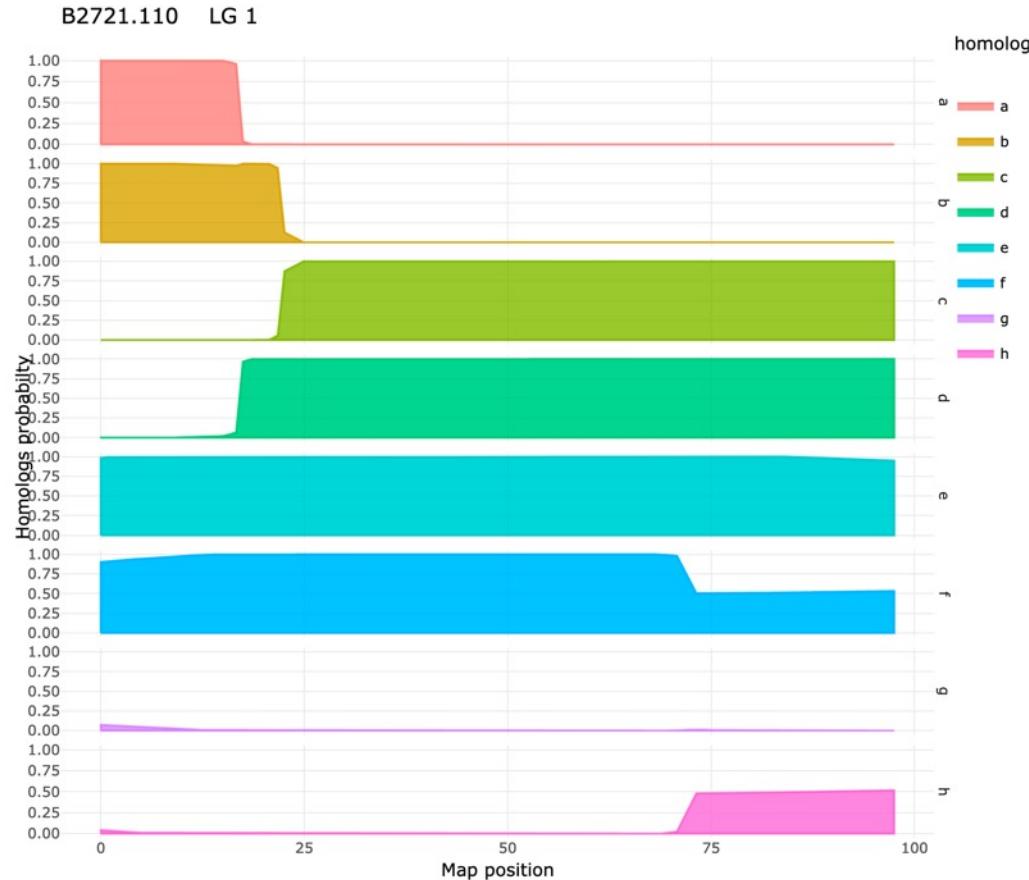
# Sweetpotato genetic map



# Probabilistic haplotype reconstruction

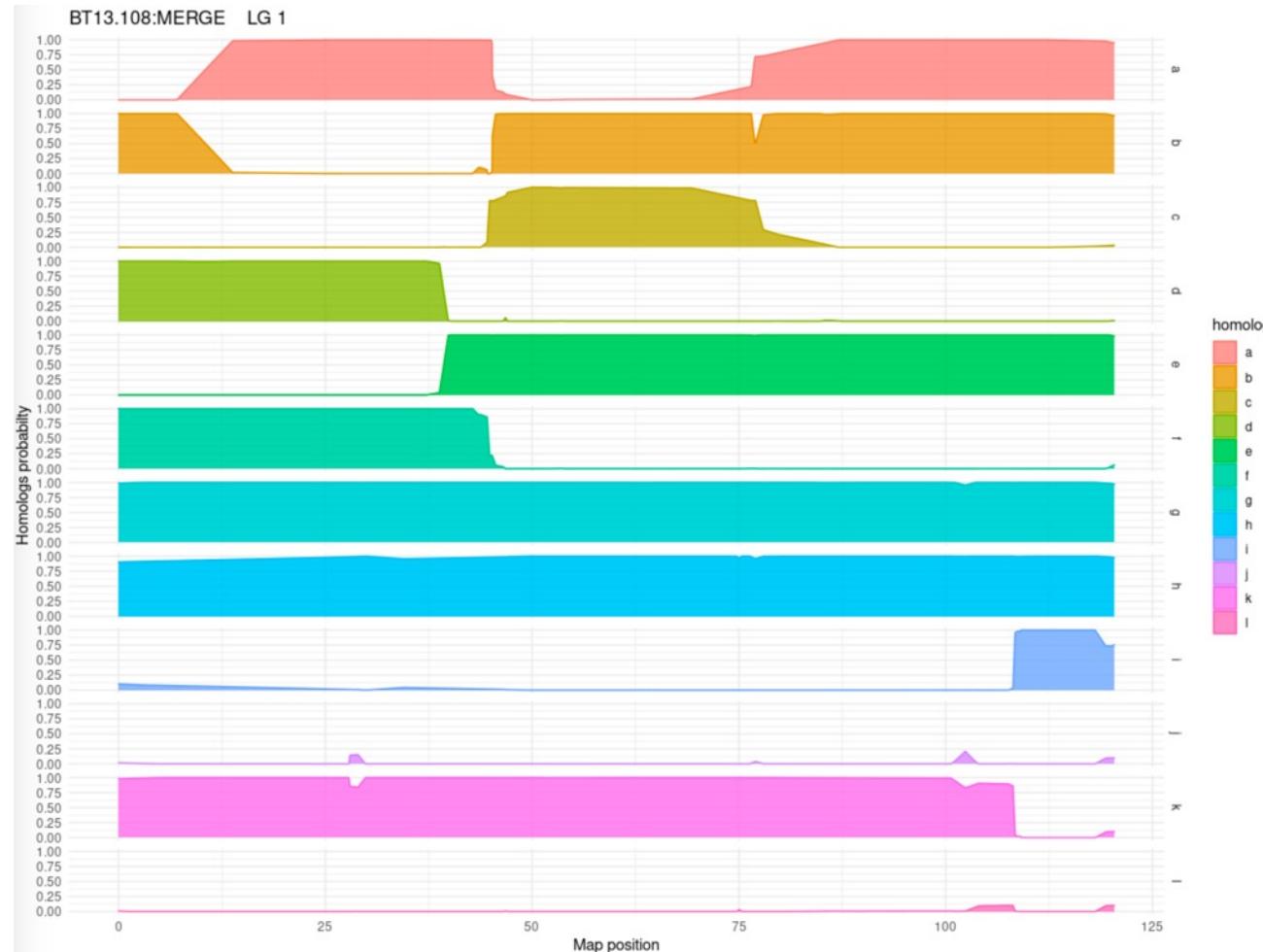
- When assuming a *prior probability* distribution of the genotypes, multilocus strategies can improve the quality of the inferred haplotypes

Tetraploid potato



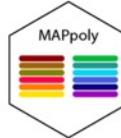
# Probabilistic haplotype reconstruction

Hexaploid sweetpotato



# 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

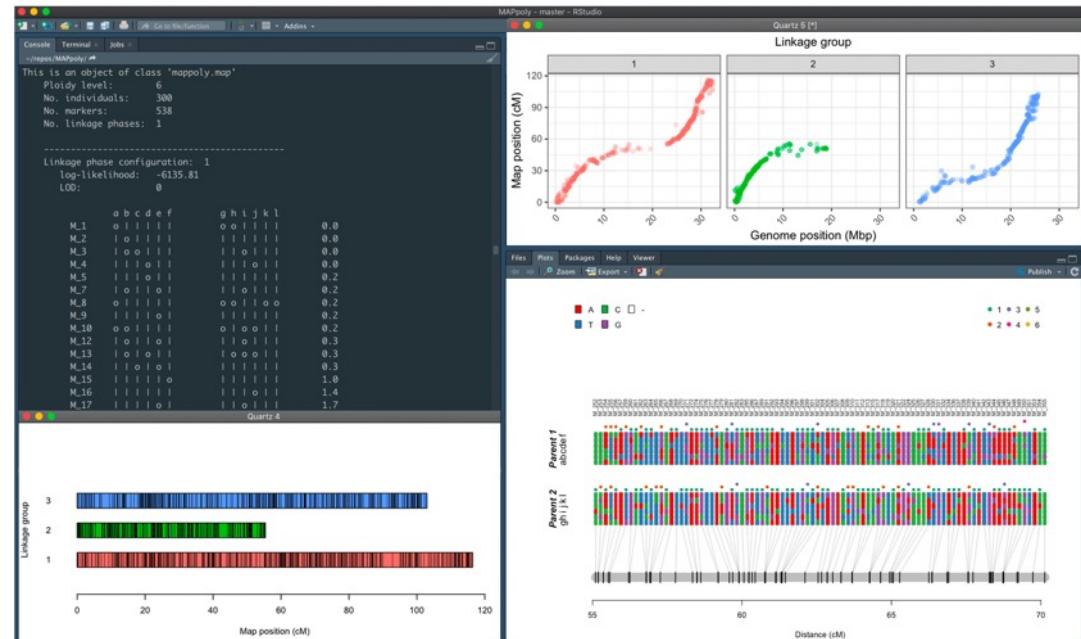
```
devtools::install_github("mmollina/mappoly")
```

## 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](#).

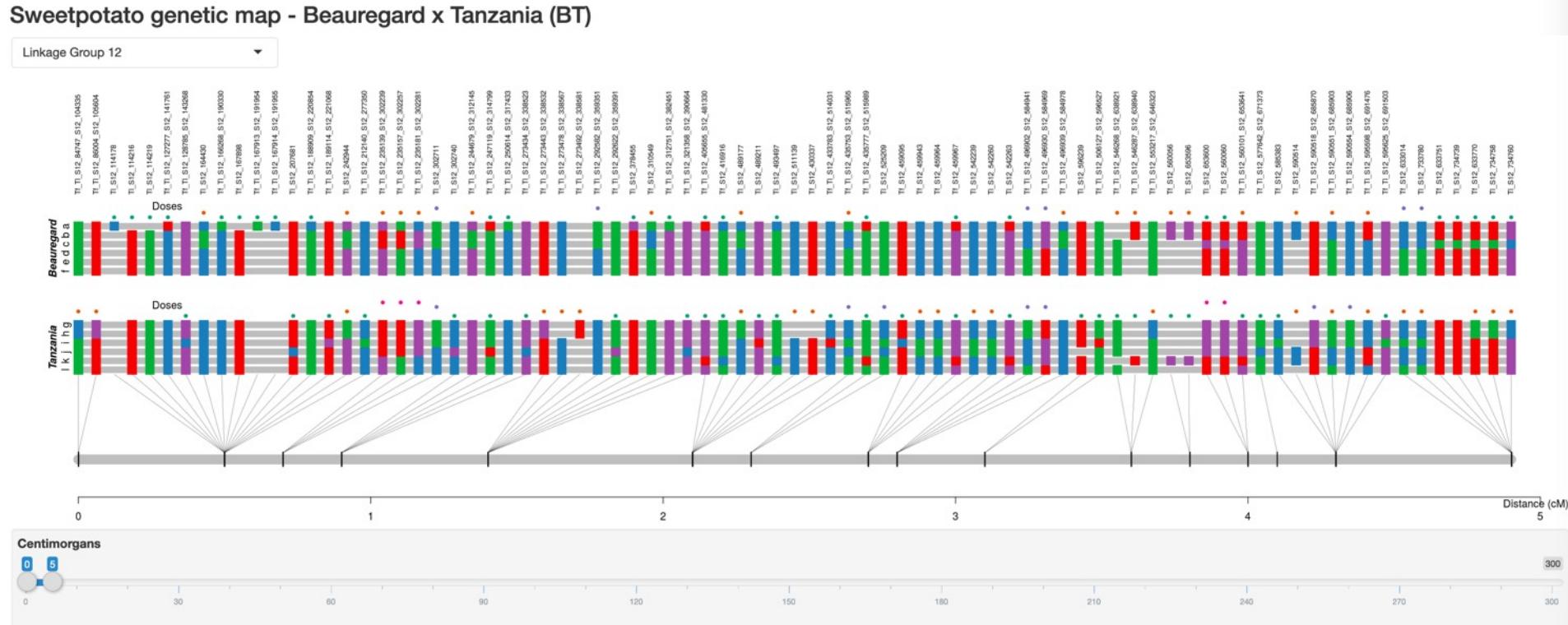


stable: <https://CRAN.R-project.org/package=mappoly>

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



# Genetic mapping – Linkage group 12 – 2661 SNPs



Show SNP names?

## Legend

Nucleotide	Doses
A	6
T	5
C	4
G	3
-	2

### 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
```

## Summary

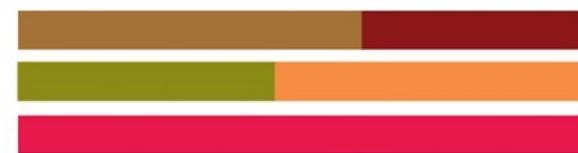
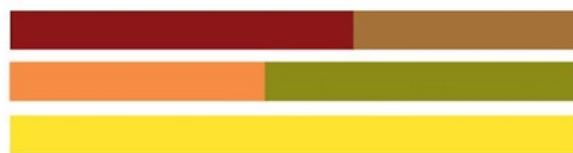
```
$number.snps  
[1] 81  
  
$length  
[1] 4.9  
  
$cM.per.snp  
[1] 0.06
```

## Notes

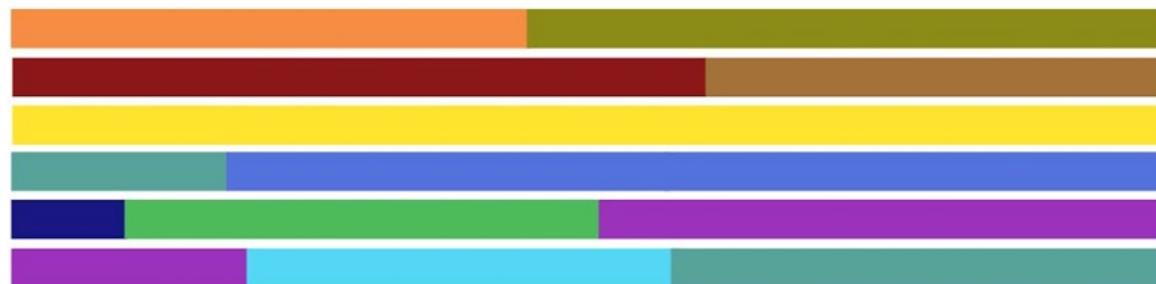
- The detailed mapping procedure is described in Mollinari et al. (2019).
  - Use the slide bar to resize or move through the map.
  - The estimation of the offspring haplotype is available [here](#).

Interactive version: [https://gt4sp-genetic-map.shinyapps.io/bt\\_map/](https://gt4sp-genetic-map.shinyapps.io/bt_map/)

# Haplotype reconstruction in the offspring



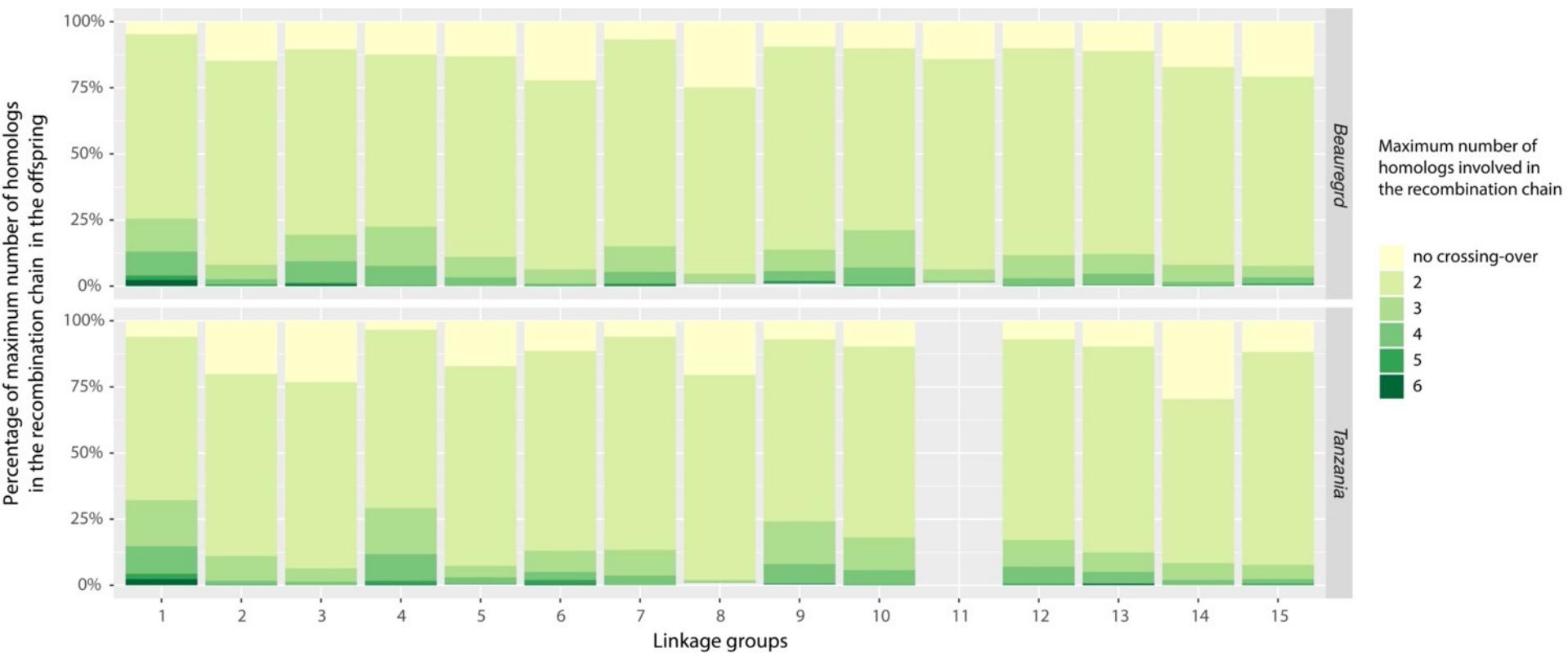
Haplotype of individual BT05:320



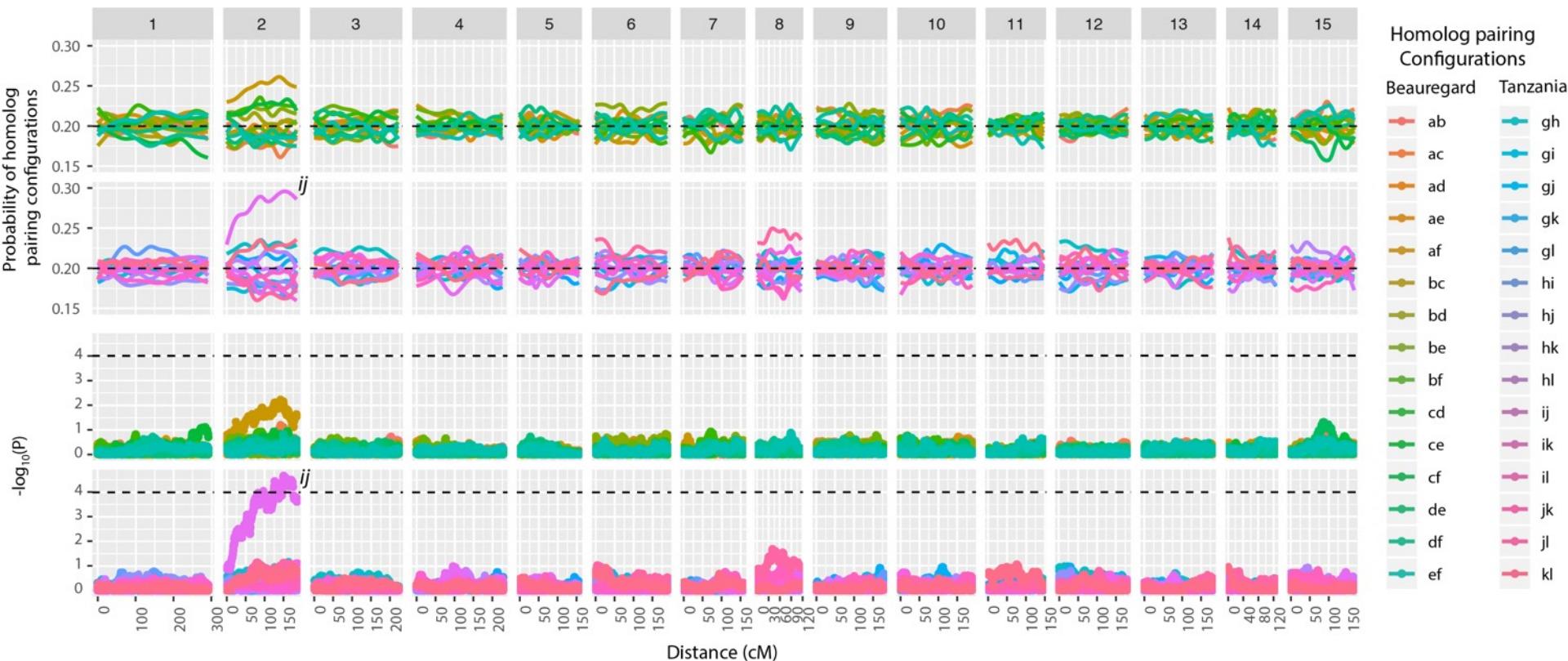
Resulting gametes



# Number of homologs involved in recombination chains



# Preferential pairing profiles: Sweetpotato is vastly autohexaploid



# QTL studies

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,<sup>\*,†,2</sup> Dorcus C. Gemenet,<sup>1,4</sup> Marcelo Mollinari,<sup>\*,†,4</sup> Bode A. Olukolu,<sup>5</sup> Joshua C. Wood,<sup>\*\*</sup> Federico Diaz,<sup>††</sup> Veronica Mosquera,<sup>††</sup> Wolfgang J. Grunberg,<sup>††</sup> Awais Khan,<sup>††</sup> C. Robin Buell,<sup>\*\*</sup> G. Craig Yencho,<sup>†</sup> and Zhao-Bang Zeng<sup>\*,†</sup>  
<sup>1</sup>Bioinformatics Research Center and <sup>2</sup>Department of Horticultural Science, North Carolina State University, Raleigh, North Carolina 27695; <sup>†</sup>International Potato Center, ILRI Campus, Nairobi, Kenya 25171-00603; <sup>§</sup>Department of Entomology and Plant Pathology, University of Tennessee, Knoxville, Tennessee 37996; <sup>\*\*</sup>Department of Plant Biology, Michigan State University, East Lansing, Michigan 48824; <sup>††</sup>International Potato Center, Peru, Lima 1558, and <sup>††</sup>Plant Pathology and Plant-Microbe Biology Section, Cornell University, Geneva Campus, New York 14456  
ORCID IDs: 0000-0002-7106-8630 (G.d.S.P.); 0000-0003-4901-1694 (D.C.G.); 0000-0002-7001-8498 (M.M.); 0000-0003-4143-8909 (B.A.O.); 0000-0001-6583-0628 (G.C.Y.); 0000-0002-3115-1149 (Z.-B.Z.)

Theoretical and Applied Genetics (2021) 134:1945–1955  
<https://doi.org/10.1007/s00122-021-03797-z>

ORIGINAL ARTICLE



## Discovery of a major QTL for root-knot nematode (*Meloidogyne incognita*) resistance in cultivated sweetpotato (*Ipomoea batatas*)

Bonny Michael Oloka<sup>1,2</sup> · Guilherme da Silva Pereira<sup>3</sup> · Victor A. Amankwah<sup>1,4</sup> · Marcelo Mollinari<sup>1</sup> · Kenneth V. Pecota<sup>1</sup> · Benard Yada<sup>2</sup> · Bode A. Olukolu<sup>5</sup> · Zhao-Bang Zeng<sup>1</sup> · G. Craig Yencho<sup>†</sup>

Theoretical and Applied Genetics (2020) 133:23–36  
<https://doi.org/10.1007/s00122-019-03437-7>

ORIGINAL ARTICLE



## Quantitative trait loci and differential gene expression analyses reveal the genetic basis for negatively associated β-carotene and starch content in hexaploid sweetpotato [*Ipomoea batatas* (L.) Lam.]

Dorcus C. Gemenet<sup>1</sup> · Guilherme da Silva Pereira<sup>2</sup> · Bert De Boeck<sup>3</sup> · Joshua C. Wood<sup>4</sup> · Marcelo Mollinari<sup>2</sup> · Bode A. Olukolu<sup>2,11</sup> · Federico Diaz<sup>3</sup> · Veronica Mosquera<sup>3</sup> · Reuben T. Ssali<sup>5</sup> · Maria David<sup>3</sup> · Mercy N. Kitavi<sup>1</sup> · Gabriela Burgos<sup>3</sup> · Thomas Zum Felde<sup>3</sup> · Marc Ghislain<sup>1</sup> · Edward Carey<sup>6</sup> · Jolien Swanckaert<sup>6</sup> · Lachlan J. M. Coin<sup>7</sup> · Zhangjun Fei<sup>8</sup> · John P. Hamilton<sup>4</sup> · Benard Yada<sup>9</sup> · G. Craig Yencho<sup>2</sup> · Zhao-Bang Zeng<sup>2</sup> · Robert O. M. Mwanga<sup>5</sup> · Awais Khan<sup>3,10</sup> · Wolfgang J. Grunberg<sup>3</sup> · C. Robin Buell<sup>4</sup>

Heredity (2021) 126:817–830  
<https://doi.org/10.1038/s41437-021-00416-x>

ARTICLE

the  
**genetics**society



## The recombination landscape and multiple QTL mapping in a *Solanum tuberosum* cv. 'Atlantic'-derived F<sub>1</sub> population

Guilherme da Silva Pereira<sup>1,2,4</sup> · Marcelo Mollinari<sup>1,2</sup> · Mitchell J. Schumann<sup>1,5</sup> · Mark E. Clough<sup>1,3</sup> · Zhao-Bang Zeng<sup>1,2</sup> · G. Craig Yencho<sup>1</sup>

Plant Disease • 2021 • 105:3048–3054 • <https://doi.org/10.1094/PDIS-10-20-2270-RE>

e-Xtra\*

## Quantitative Trait Locus Mapping for Common Scab Resistance in a Tetraploid Potato Full-Sib Population

Guilherme da Silva Pereira,<sup>1,†</sup> Marcelo Mollinari,<sup>2,3</sup> Xinshun Qu,<sup>4</sup> Christian Thill,<sup>5</sup> Zhao-Bang Zeng,<sup>2,3</sup> Kathleen Haynes,<sup>6</sup> and G. Craig Yencho<sup>2</sup>

<sup>1</sup> International Potato Center, ILRI Campus, Nairobi 25171-00603, Kenya

<sup>2</sup> Department of Horticultural Science, North Carolina State University, Raleigh, NC 27695, U.S.A.

<sup>3</sup> Bioinformatics Research Center, North Carolina State University, Raleigh, NC 27695, U.S.A.

<sup>4</sup> Department of Plant Pathology and Environmental Microbiology, The Pennsylvania State University, University Park, PA 16802, U.S.A.

<sup>5</sup> Department of Horticultural Science, University of Minnesota, St. Paul, MN 55108, U.S.A.

<sup>6</sup> Genetic Improvement of Fruits and Vegetables Laboratory, U.S. Department of Agriculture-Agricultural Research Service, Beltsville, MD 20705, U.S.A.



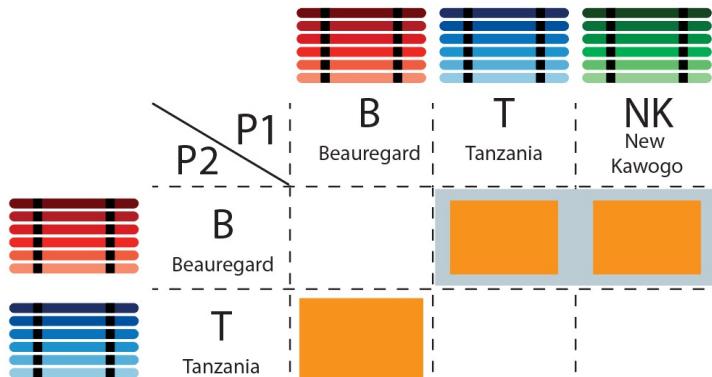
# Analysis in multiple inter-connected sweetpotato families



full-sib family + half-sib family (B)

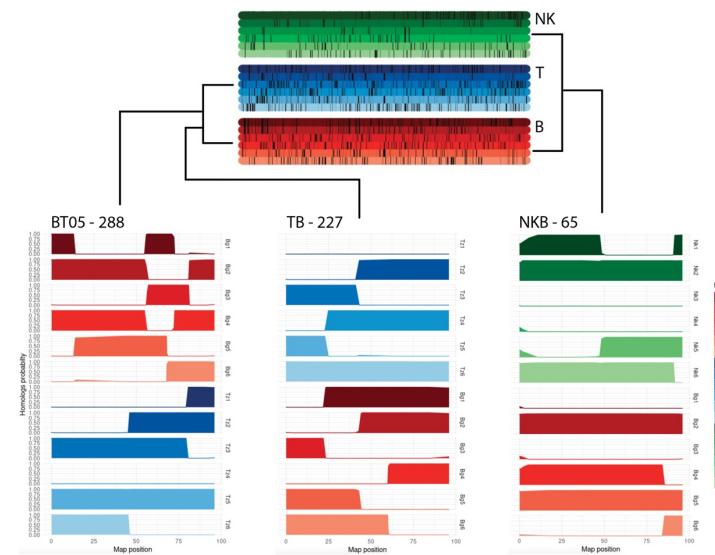
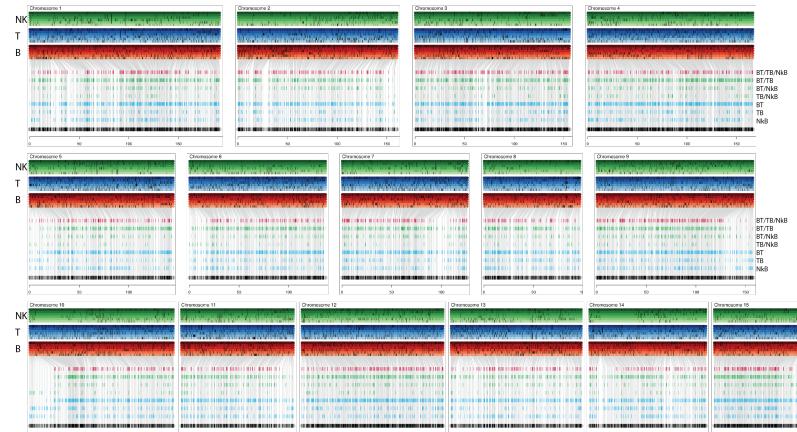
Gametes from  
both parents

Gametes from  
parent B



Schematic representation of three  
interconnected populations: BT, TB, and NKB

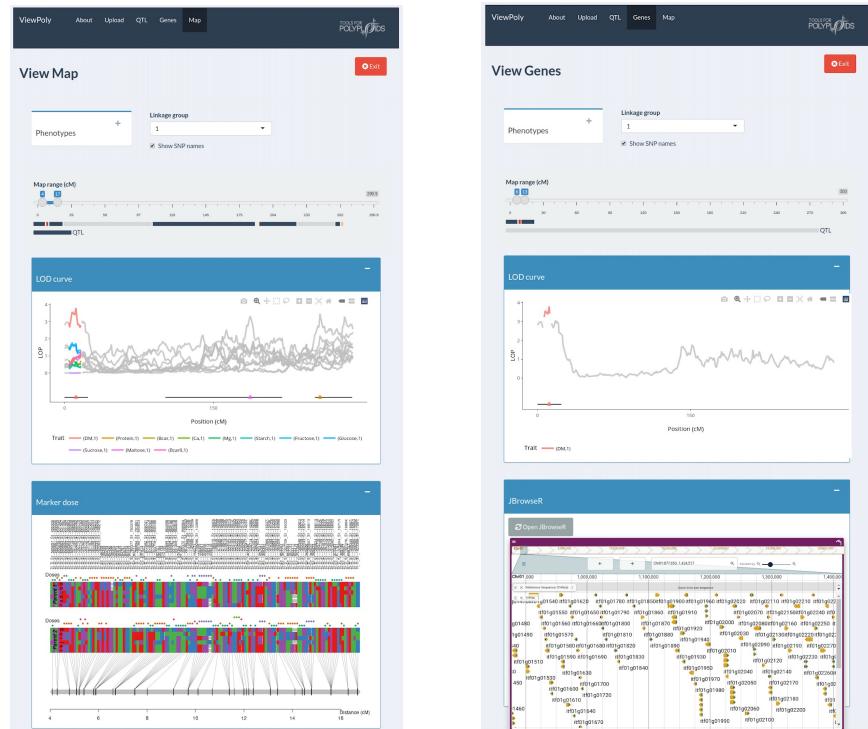
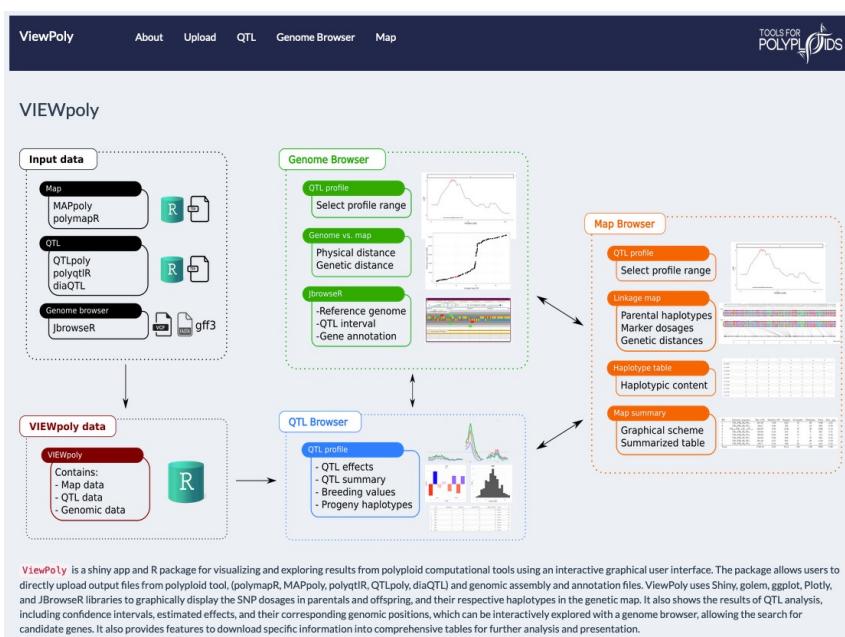
Complete genetic map for BT, TB, and NKB



Example of haplotyping in chromosome 15:  
Beauregard, Tanzania and New Kawogo



# User friendly tools



- **VIEWpoly:** Interactively explore and select QTL regions, find candidate genes, find favorable haplotypes, dosage effects, list breeding values, and QTL information.

# Perspectives

## *Specific Objectives*

- Continue to **bridge the gap** between the complex polyploid genetics and practical breeding
- Implementing **user friendly** computational tools to help breeders make the best use of their data.
- Such tool can assist them to make the short and long-term **breeding decisions** based on the collected and learnt **information** about their breeding populations for different breeding objectives
- Some objectives: which individuals to **select**, who to **mate** to who; and long-term **breeding strategy**, which may be informed through targeted **forward simulation** exercises.
- In order to improve this objectives, which sometimes are **crop-specific**, we need to have a direct communication with **polyploid breeders** and discuss their needs in terms of computational tools

## *Overarching objective*

*Use the inference of genome inheritance (haplotype transmission) through pedigrees, elucidating the genetic relationship between genome segregations and trait phenotypes in terms of QTL of founders' alleles.*

# Acknowledgements



Gabriel Gesteira, Zhao-Bang Zeng, Craig Yencho



Bode Olukolu



2/22/22



Guilherme Pereira



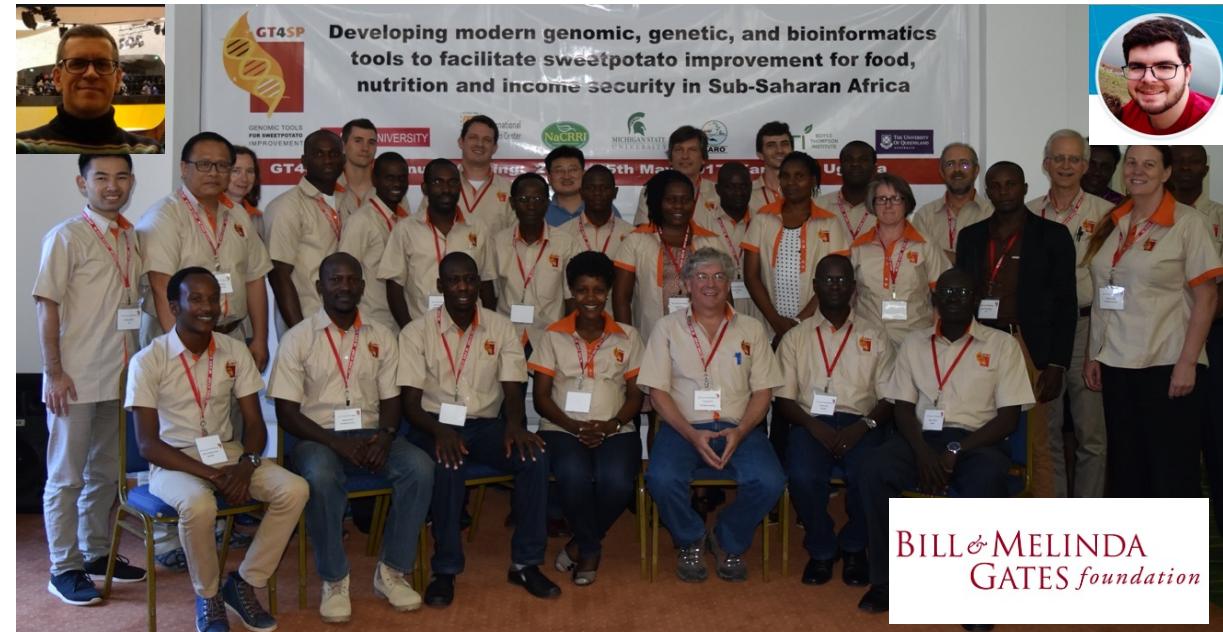
Shan Wu, Zhangjun Fei



Antonio Augusto Franco Garcia



Dorcus C. Gemenet



# References

- Haldane, J. Theoretical Genetics of Autopolyploids. *J. Genet.* **22**, 359–372 (1930).
- Mather, K. Reductive 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).
- Lander, E. S. & Green, P. Construction of multilocus genetic linkage maps in humans. *Proc. Natl. Acad. Sci. U. S. A.* **84**, 2363–2367 (1987).
- Rabiner, L. R. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* **77**, 257–286 (1989).
- Ripoll, 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* **9**, 3297–3314 (2019).

