Genetic Mapping in Polyploids: from genotyping to haplotype reconstruction

Marcelo Mollinari
mmollin@ncsu.edu

GT4SP GENOMIC TOOLS FOR SWEETPOTATO IMPROVEMENT

Sweet GAINS

FAPESP FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO

CNPq Conselho Nacional de Desenvolvimento Científico e Tecnológico

BILL & MELINDA GATES foundation

April 22, 2021

North Carolina State University

- Some Numbers (fall 2020)
  - 36,762 students enrolled (48.7% female, 51.3% male)
    - (71% undergrads, 29% grads)
  - 2,457 professors (40.6% female, 59.4% male)
  - 7,239 staff (47.9% female, 52.1% male)
- Motto: Think and do!
- Public, land-grant and research university
- Raleigh, NC
- 3 campuses: Main, centennial, biomedical centennial.
- Nickname: Wolfpack (Go pack!)

Main campus

Talley Students Union

NCSU – Carter-Finley Stadium

NCSU - PNC Arena

Centennial Campus

Centennial Campus – J. B. Hunt Library

Plant Sciences Initiative

Outline

- Introduction to linkage
- Genotype calling in polyploids
- Modeling gamete formation, linkage and phasing
- Genetic mapping in hexaploid sweetpotato
- Assessing meiotic configurations and preferential chromosome pairing
- Multi-parental analysis in complex polyploids

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)

GT4SP

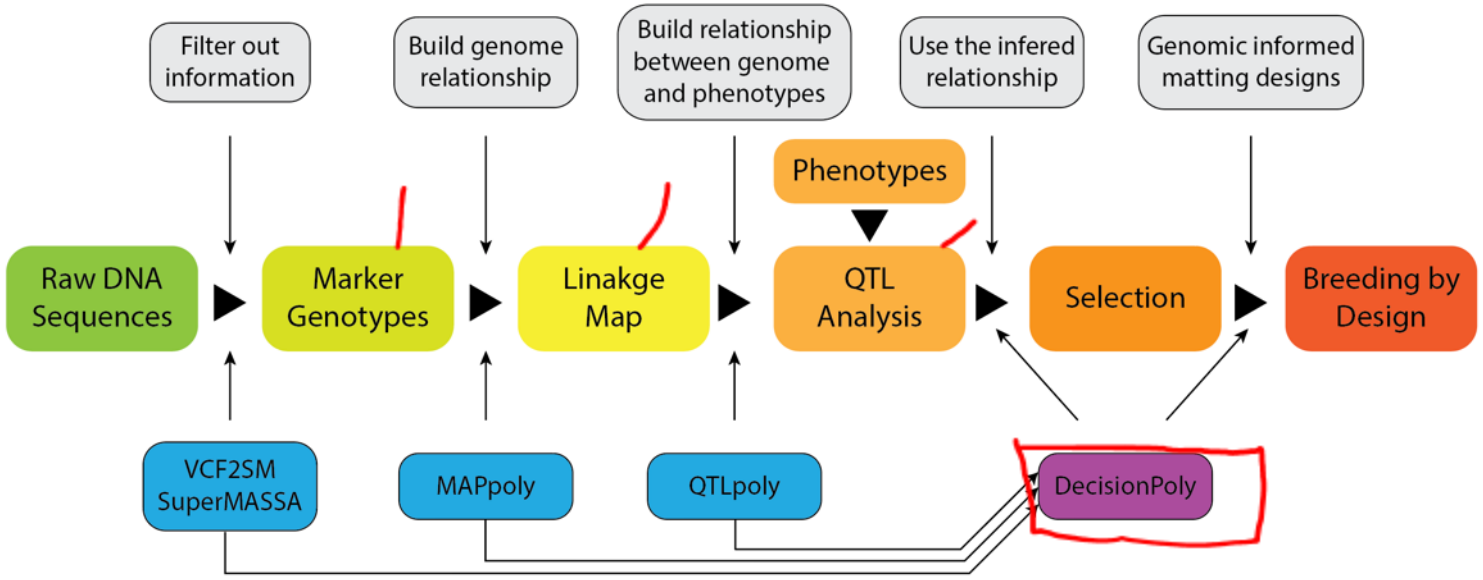
GENOMIC TOOLS FOR SWEETPOTATO IMPROVEMENT

Sweetpotato for profit and health initiative

Food security for Sub-Saharan Africa

GT4SP & SweetGAINS

- Develop computational tools for the whole pipeline data analysis

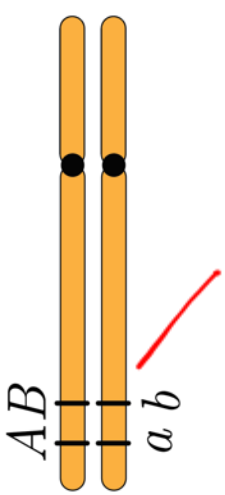


- VCF2SM: Python scripts processing DNA calls (VCF files) for SuperMASSA
- SuperMASSA: Call SNP dosage marker genotype probability distribution.
- **MAPpoly: An R package for constructing a complete linkage map for 2X, 4X, 6X, 8X**
- QTLpoly: An R package for QTL mapping in full-sib families for 2X, 4X, 6X, 8X
- DecisionPoly: 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.

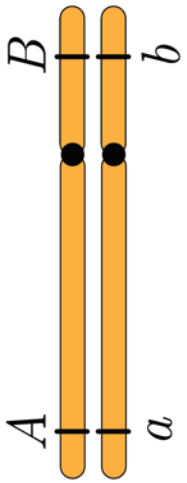
https://sweetpotatogenomics.cals.ncsu.edu/

Introduction

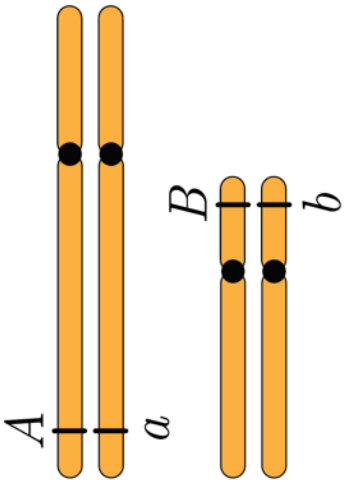
- Genetic linkage is the phenomenon where markers are likely to be inherited together.
- The closer the markers are, the lower the probability of crossing over events occur between them; consequently, the more likely they will be co-inherited.



Linked



Not linked



Not linked

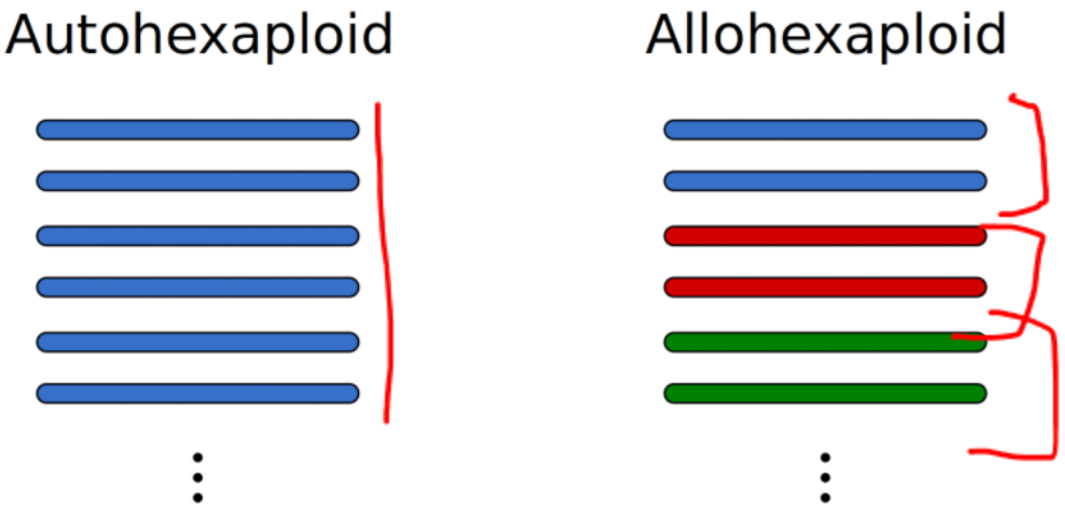
- How can we measure how likely A and B are co-inherited?

Linkage analysis

- We measure linkage using the *recombination frequency* (or fraction) in a segregation population.
- Genetic linkage is a concept applied to *at least two loci*.
- Recombination fraction 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 *mapping functions*. (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.
- First, let us address the *behavior o a single loci* when transited across generations.

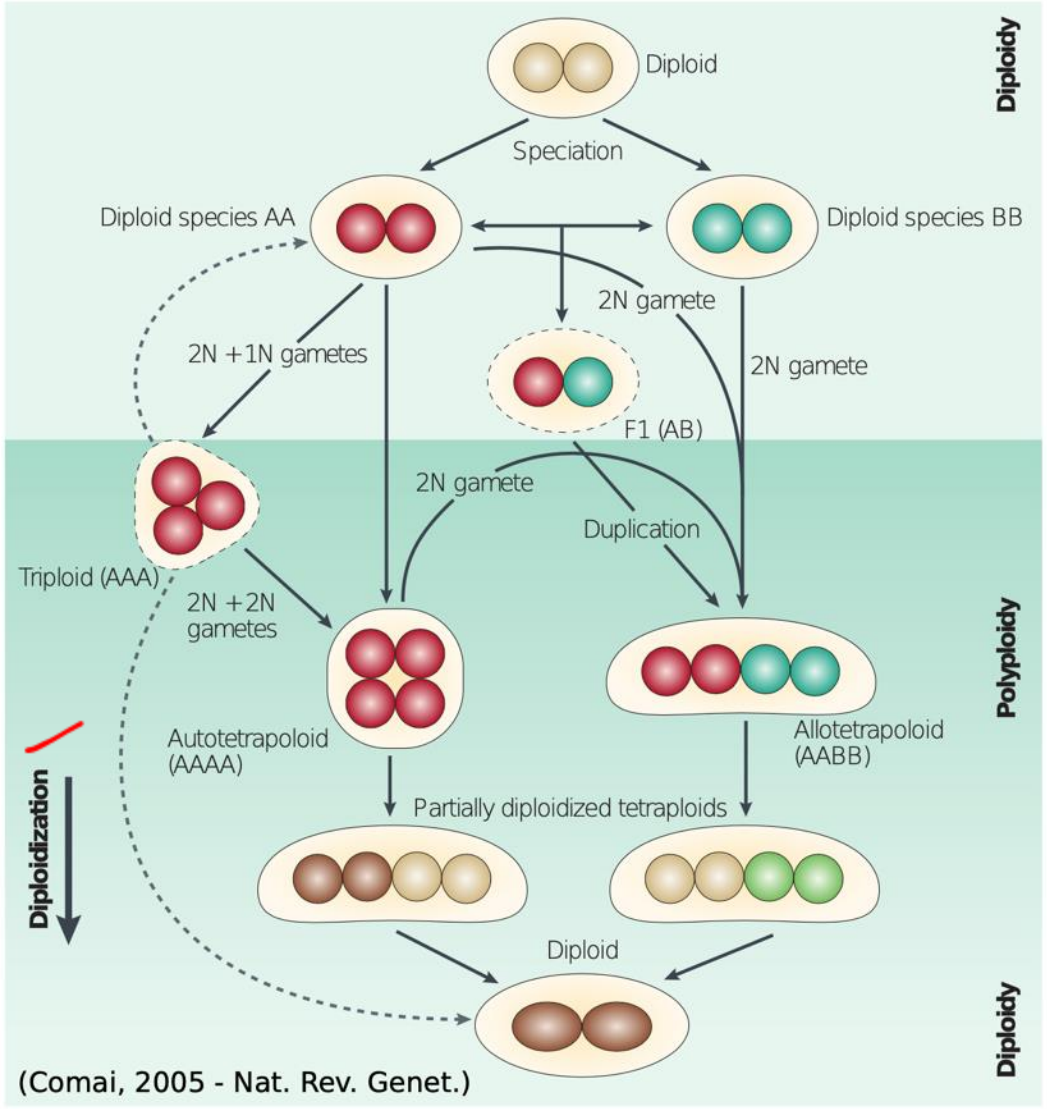
Polyploid Species

- Organisms with more than 2 complete sets of chromosomes

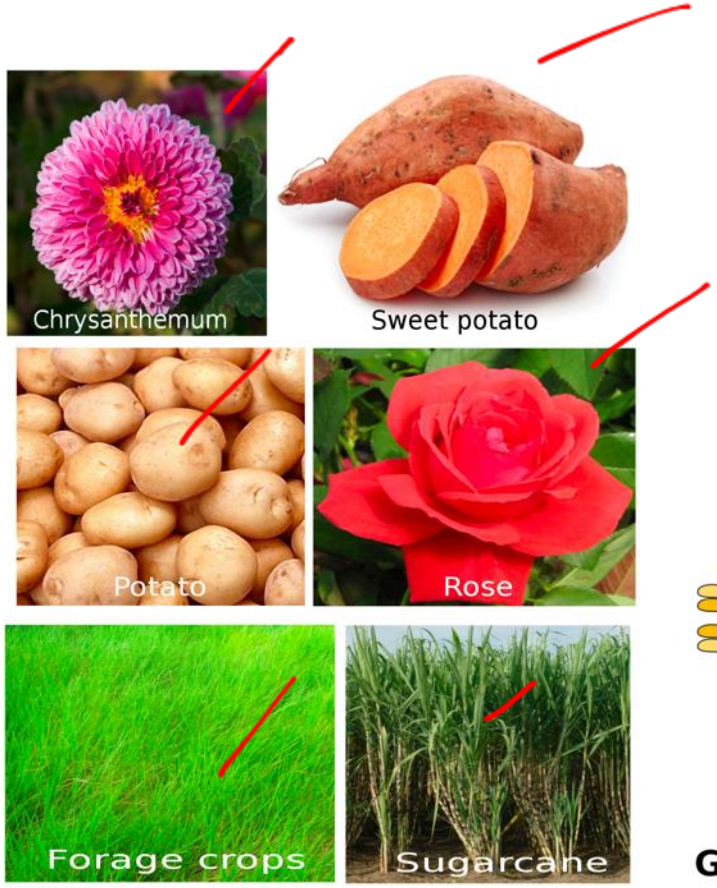


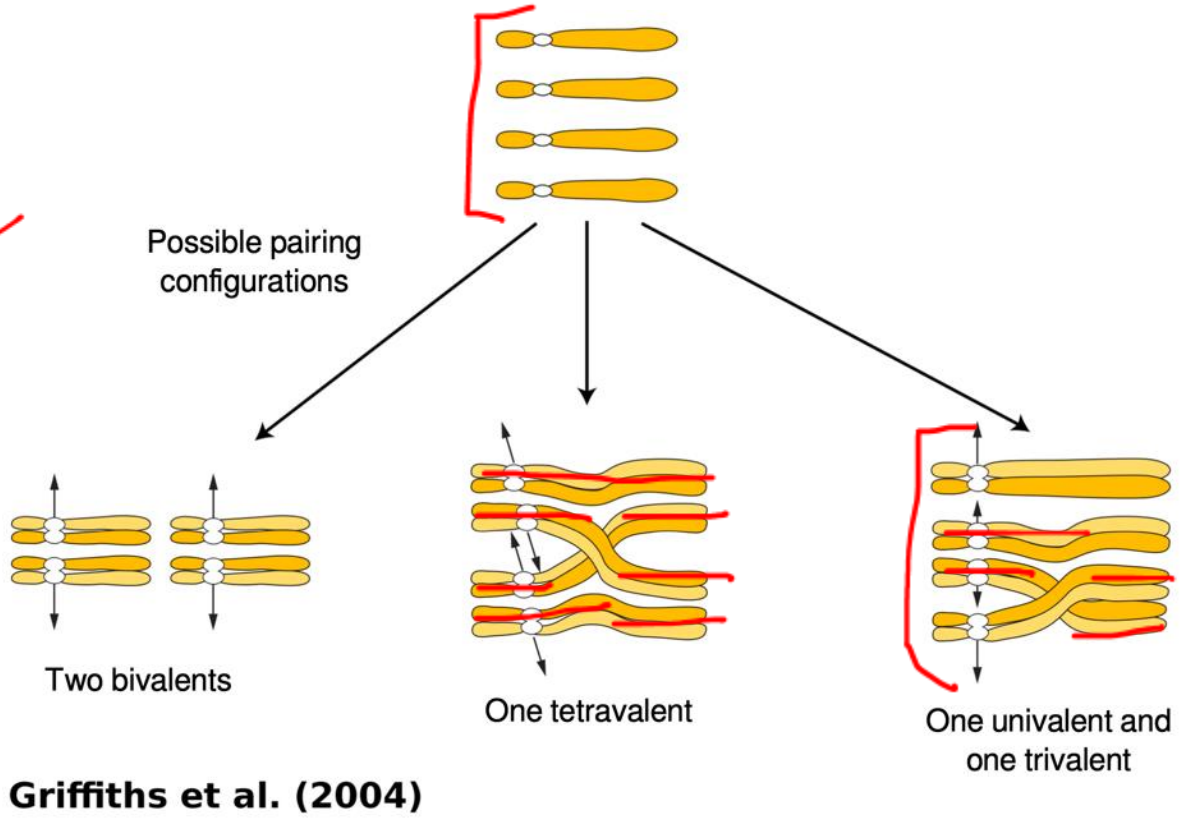
- Multiple sets have the same origin
- Multiple sets have different origins

How are polyploids formed



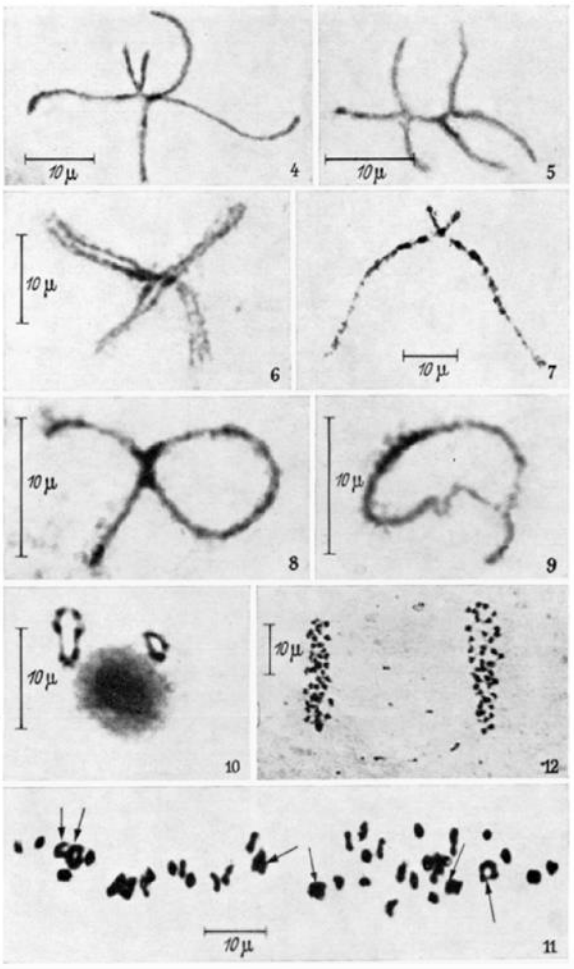
Autopolyploids and meiotic pairing





Meiotic Pairing in polyploids

Sweetpotato



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

Sugarcane

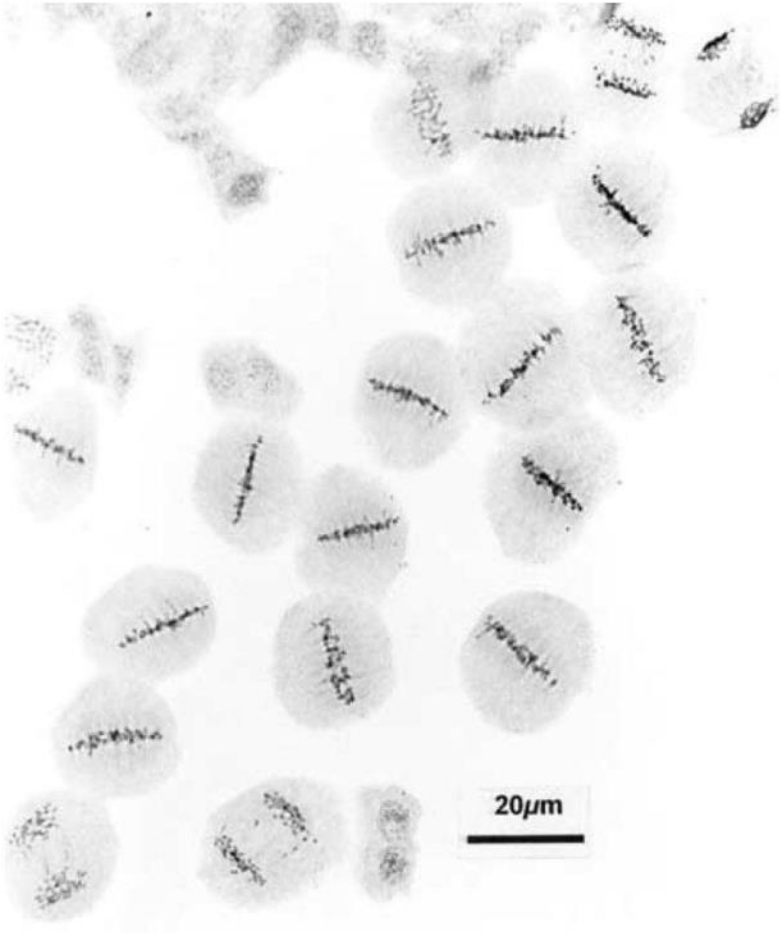


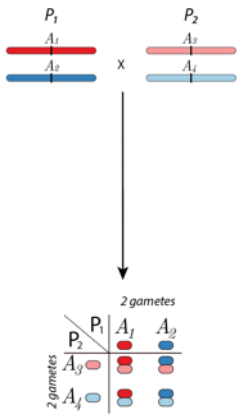
Figure 1. PMCs of *Saccharum* spp. hybrid clone 79N9059 at meiosis. As was the case in other clones, pairing was regular, bivalents generally formed. (Bielig *et al.* 2003)

Segregation in polyploids*

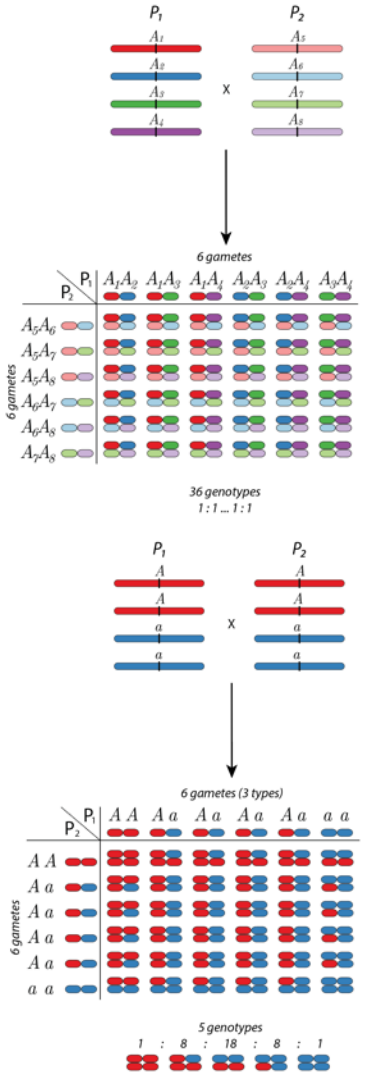
Multiallelic

Biallelic

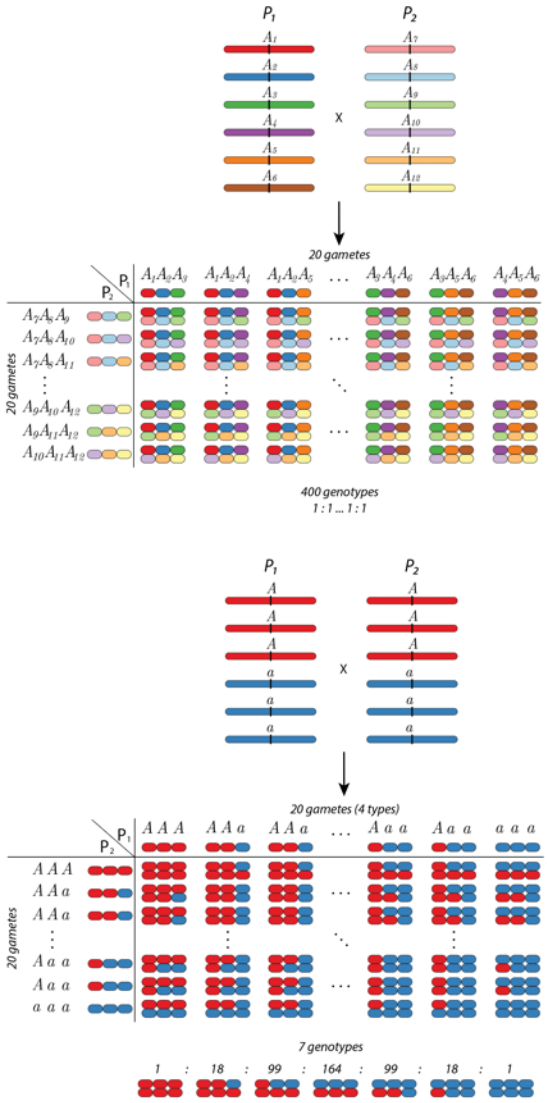
Diploid



Tetraploid

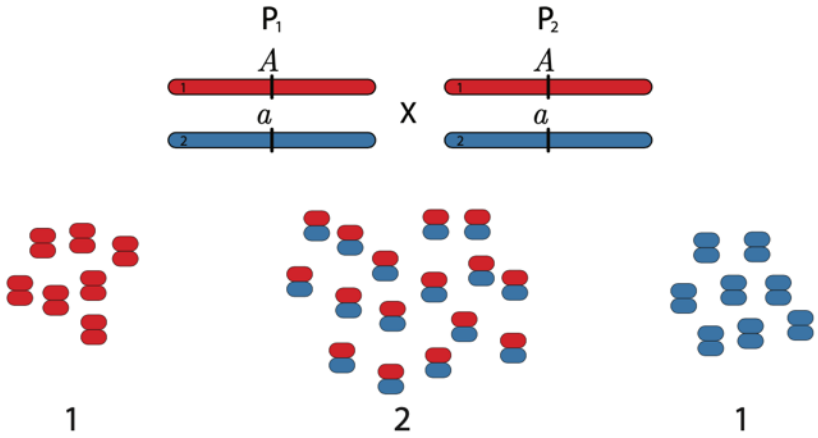


Hexaploid



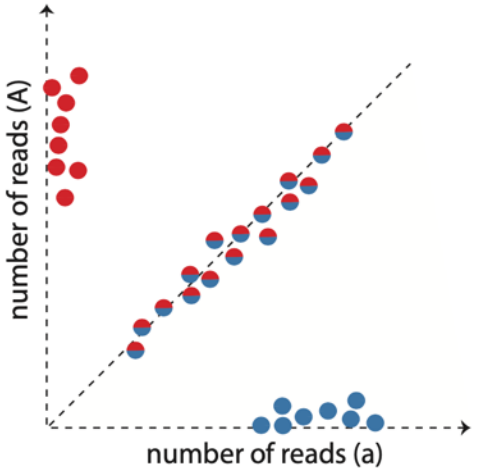
*random pairing and no double reduction

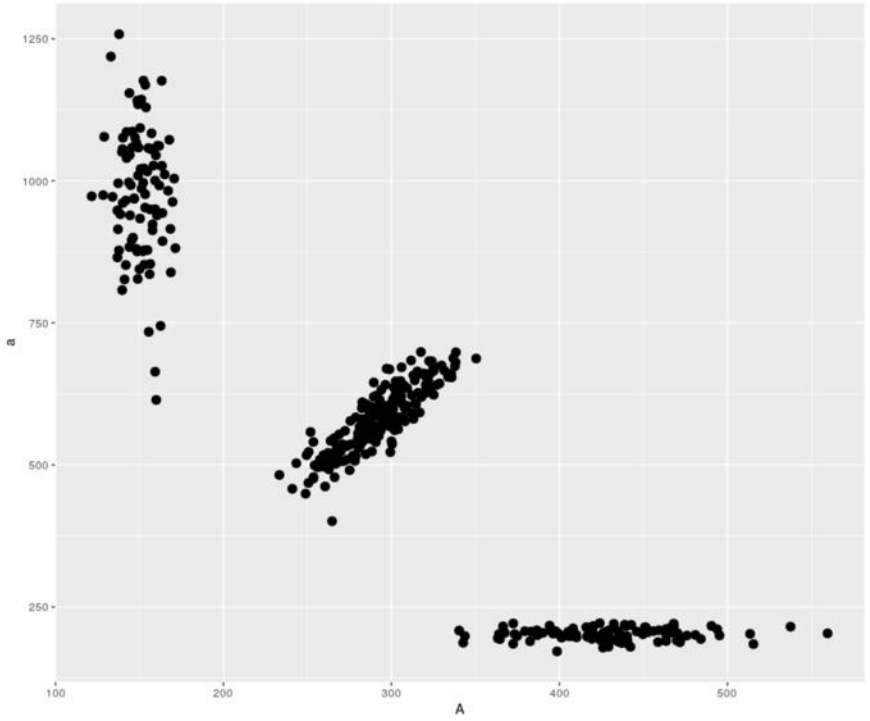
Genotype calling in diploids



GBS

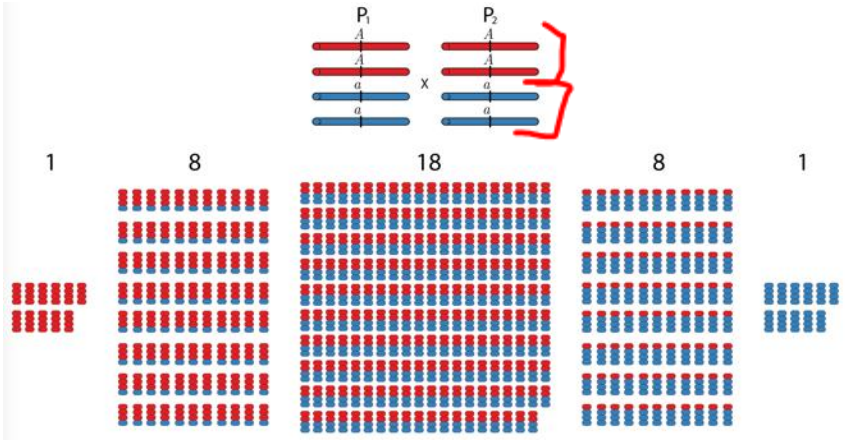
Read counts for each one of the allelic variants

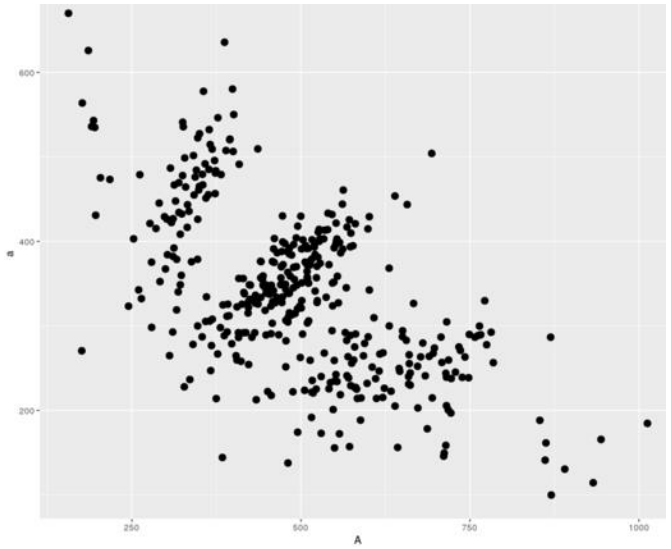




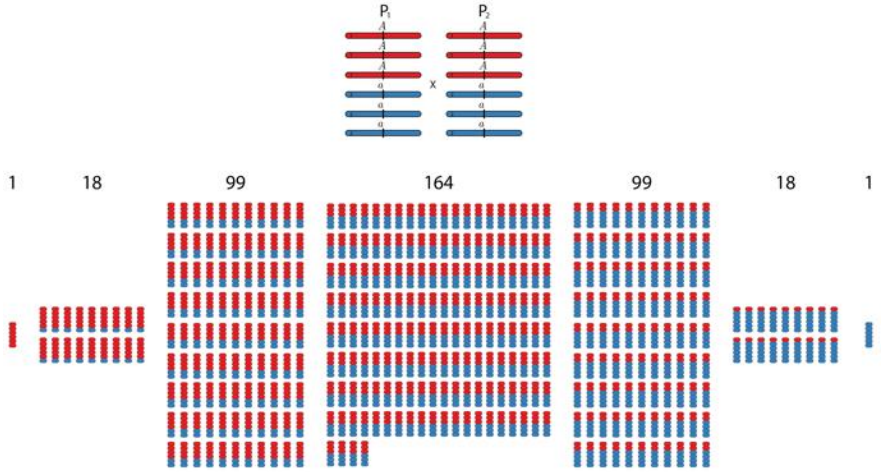
Genotype calling in polyploids

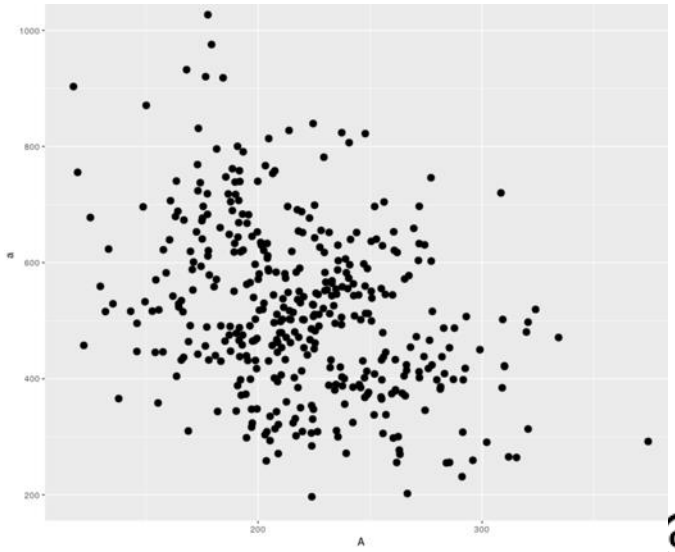
Tetraploid



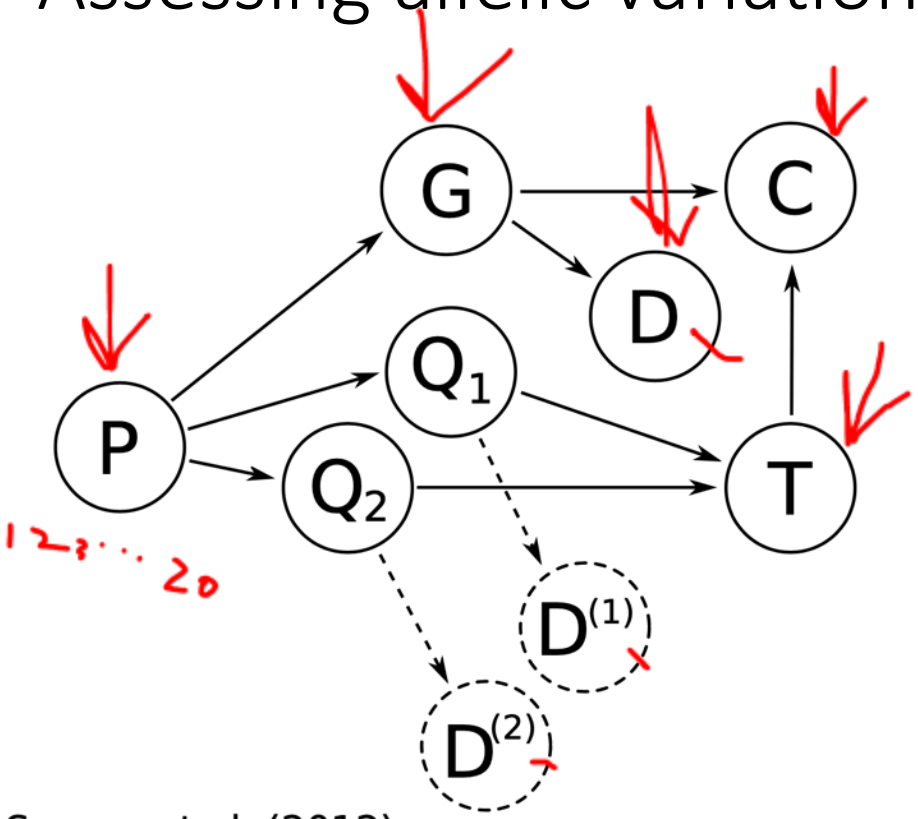


Hexaploid





Assessing allelic variation in polyploids



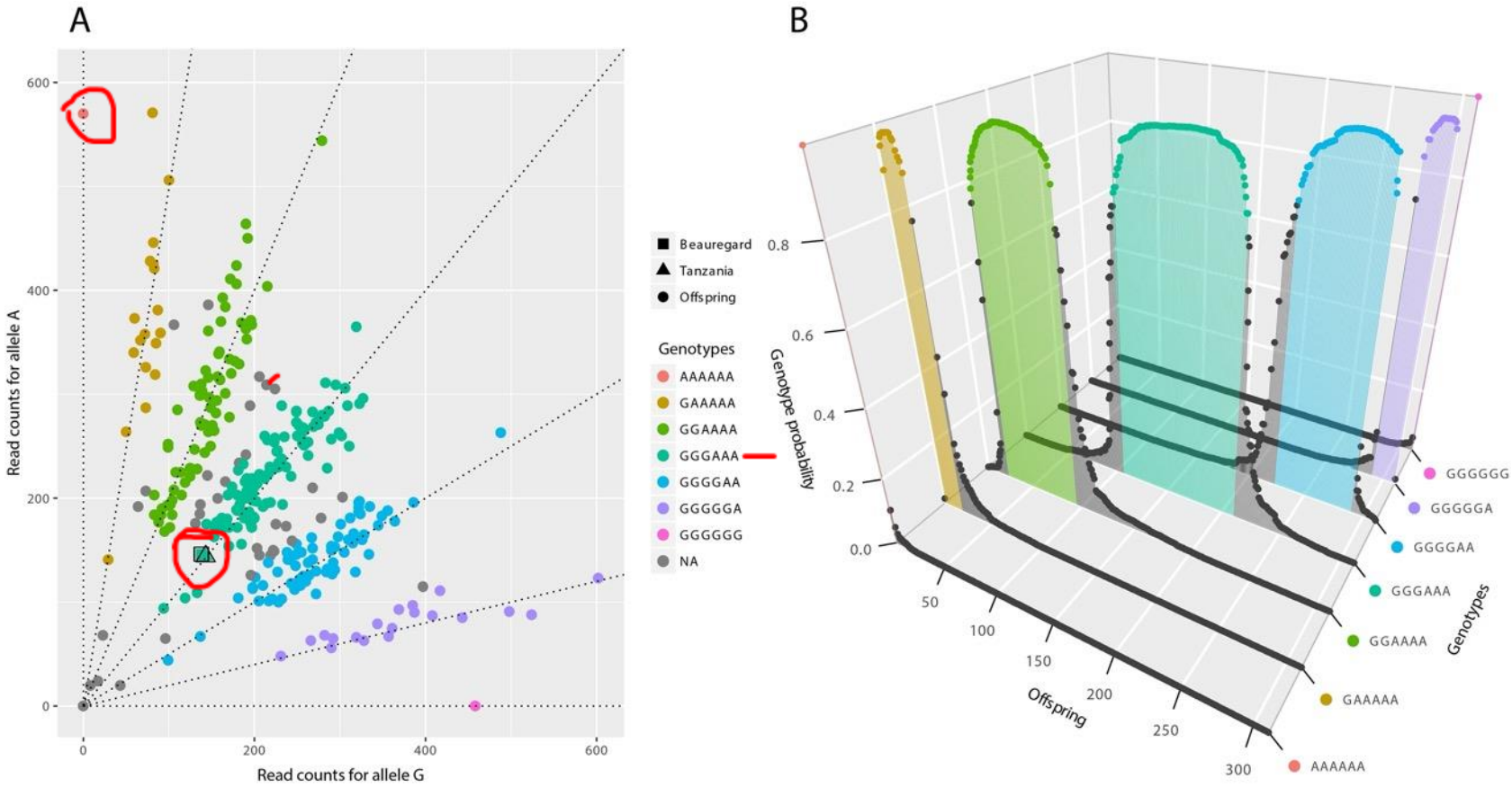
Serang et al. (2012)
Mollinari and Serang (2015)

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

- 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



Beauregard: 3 doses

Tanzania: 3 doses

Genotype calling

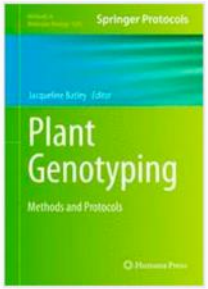
OPEN ACCESS Freely available online

PLoS one

Efficient Exact Maximum a Posteriori Computation for Bayesian SNP Genotyping in Polyploids

Oliver Serang[1,2]*, Marcelo Mollinari[3], Antonio Augusto Franco Garcia[3]

1 Department of Neurobiology, Harvard Medical School, Boston, Massachusetts, United States of America, **2** Department of Pathology, Children's Hospital Boston, Boston, Massachusetts, United States of America, **3** Department of Genetics, University of São Paulo/ESALQ, Piracicaba, São Paulo, Brazil



Plant Genotyping pp 215-241 | Cite as

Quantitative SNP Genotyping of Polyploids with MassARRAY and Other Platforms

Authors Authors and affiliations

Marcelo Mollinari, Oliver Serang

Genotype calling in polyploids

- fitTetra (tetraploids – array data):
https://www.wur.nl/en/show/Software-fitTetra.htm

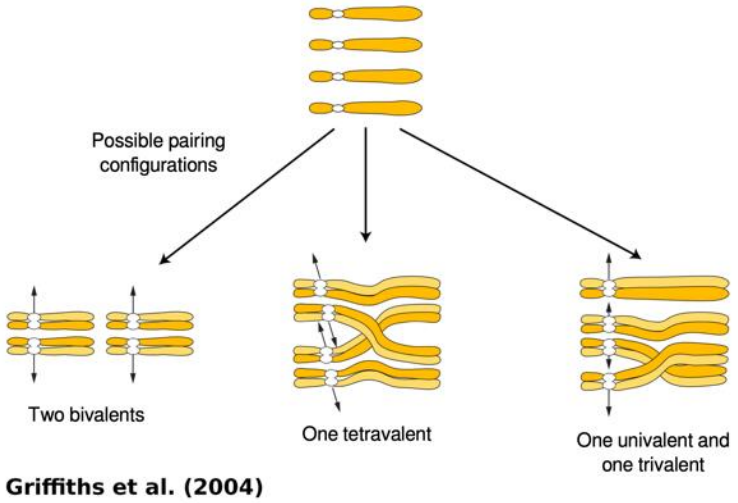
- ClusterCall (tetraploids):
https://potatobreeding.cals.wisc.edu/software/

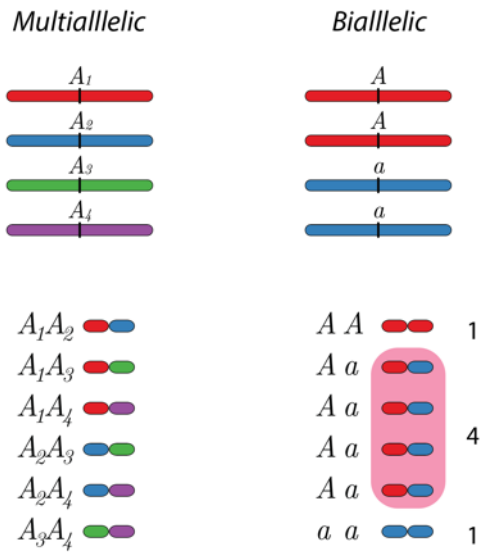
- SuperMASSA (any ploidy level):
https://bitbucket.org/orserang/supermassa

- updog (any ploidy level, allows preferential pairing):
https://github.com/dcgerard/updog

- polyRAD (any ploidy level, reads VCF, BAM, etc):
https://github.com/lvclark/polyRAD

Gamete formation in polyploids*



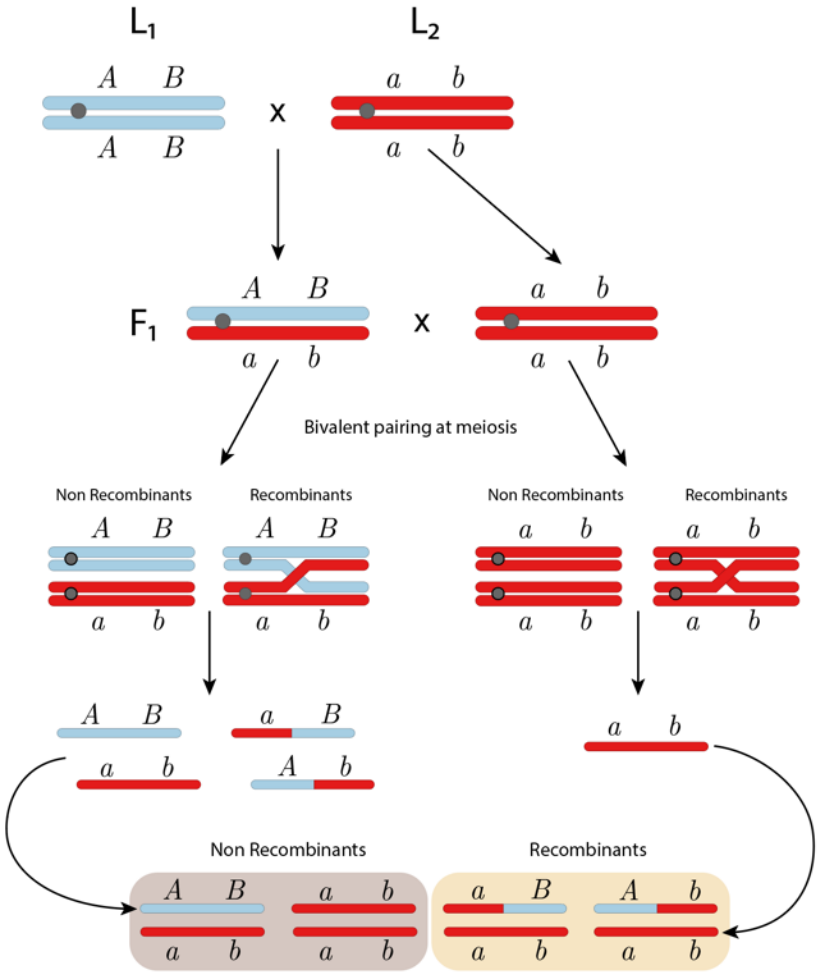


Number of possible gamtes considering one locus with no double-reduction in one and two parents

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

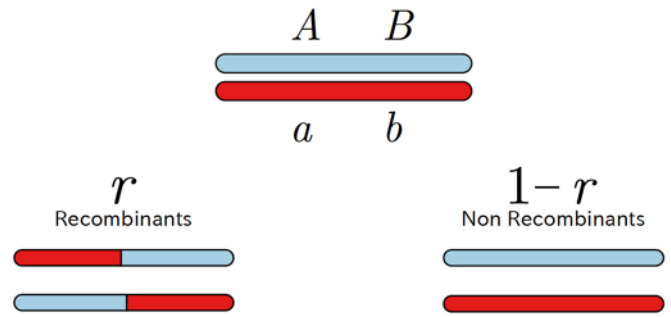
*random pairing and no double reduction

Recombination fraction in diploids



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

Recombination fraction in diploids - Likelihood



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

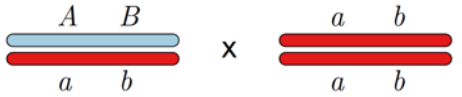
$$L(r) = \prod_n \Pr(G_A, G_B \mid r)$$

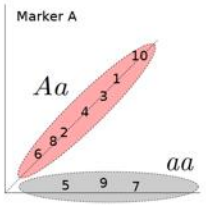
where n is the number of individuals. The maximum likelihood estimator of r is

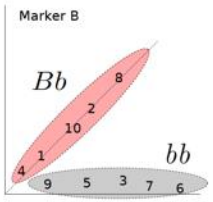
$$\hat{r} = \underset{r}{\operatorname{argmax}}\, L(r)$$

Recombination fraction in diploids

Toy example

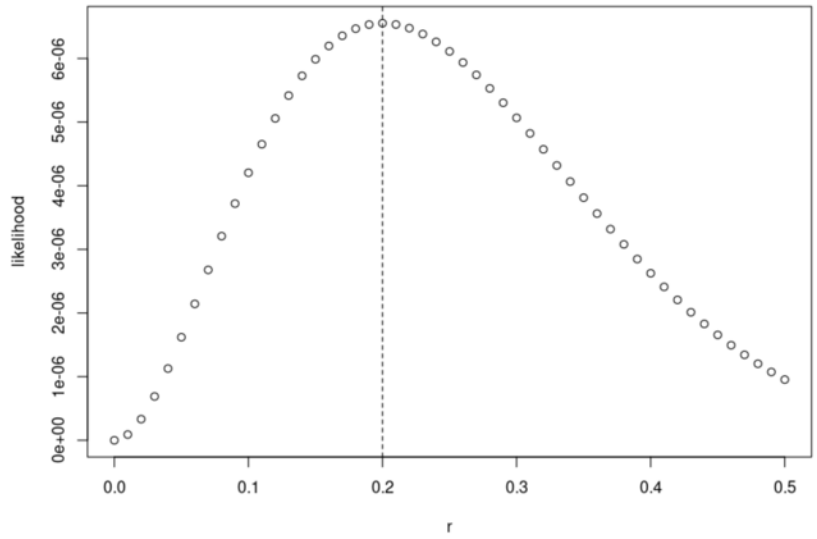






Individual	Obs. Gen.
1	$(A\,,B)$
2	$(A\,,B)$
3	$(A\,,b\,)$
4	$(A\,,B)$
5	$(a\,,b\,)$
6	$(A\,,b\,)$
7	$(a\,,b\,)$
8	$(A\,,B)$
9	$(a\,,b\,)$
10	$(A\,,B)$

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



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

Individual	Obs. Gen.	$\Pr(\mathrm{loc}_B,\mathrm{loc}_A)$
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	$(A\,,B)$	$\frac{1}{2}(1-r)$

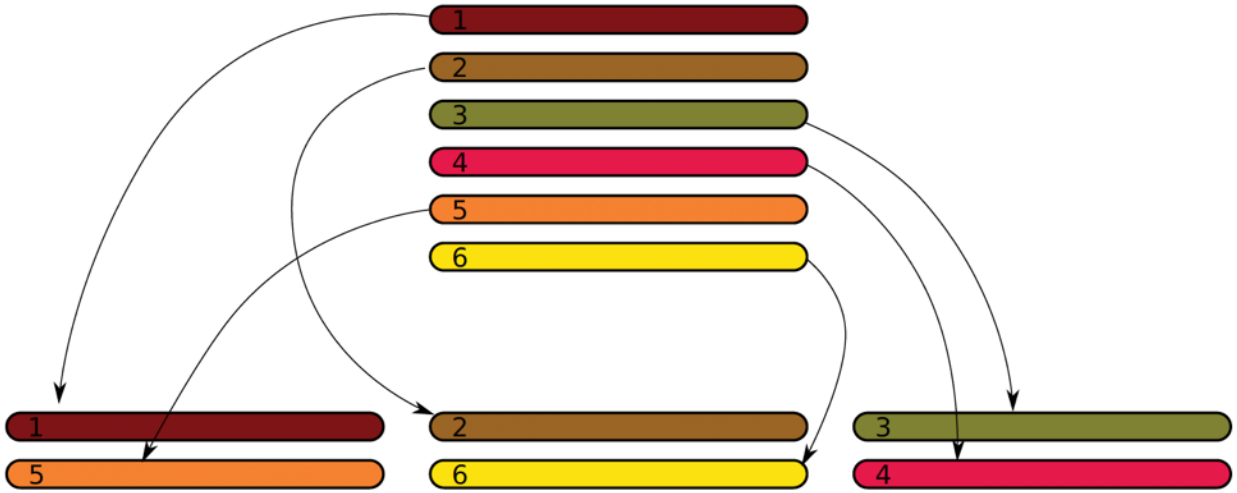
$$L = \left(\tfrac{r}{2}\right)^2 \left(\tfrac{1-r}{2}\right)^8$$

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

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*



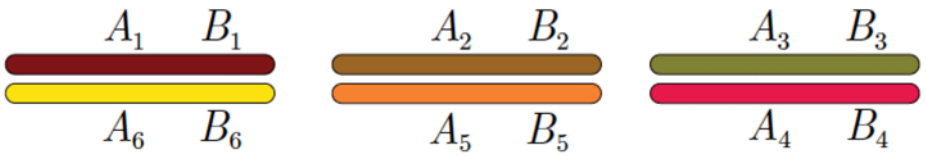
$$\psi_4$$

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

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

*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:

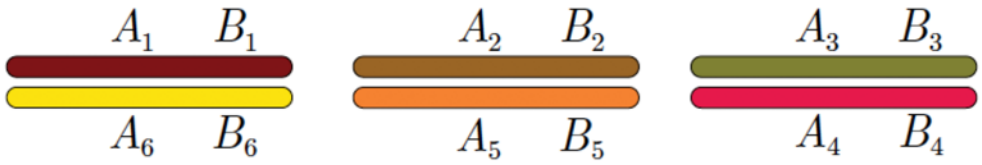
$$\mathbf{V}_1 \otimes \mathbf{V}_2 \otimes \cdots \otimes \mathbf{V}_{\frac{p}{2}}$$

All elements of this product are of the form

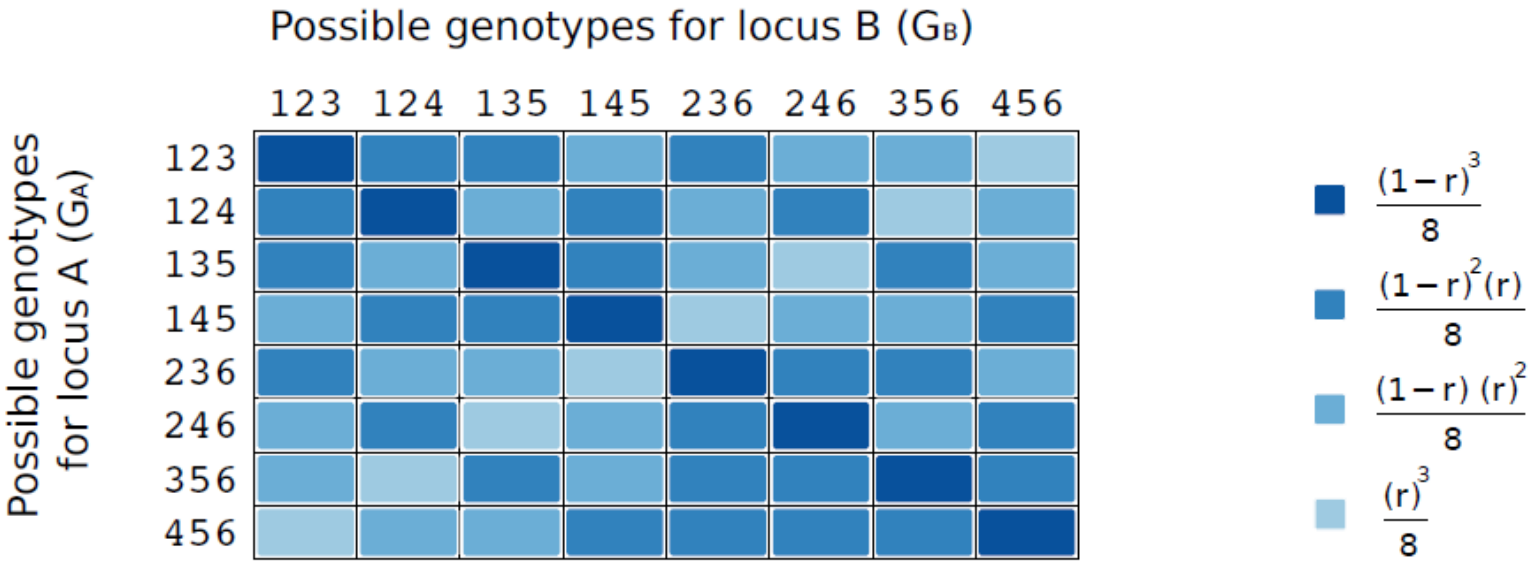
$$\Pr(G_A, G_B \mid \psi_j, r) = \frac{(1-r)^{(\frac{p}{2}-l)} p^l}{2^{\frac{p}{2}}}$$

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

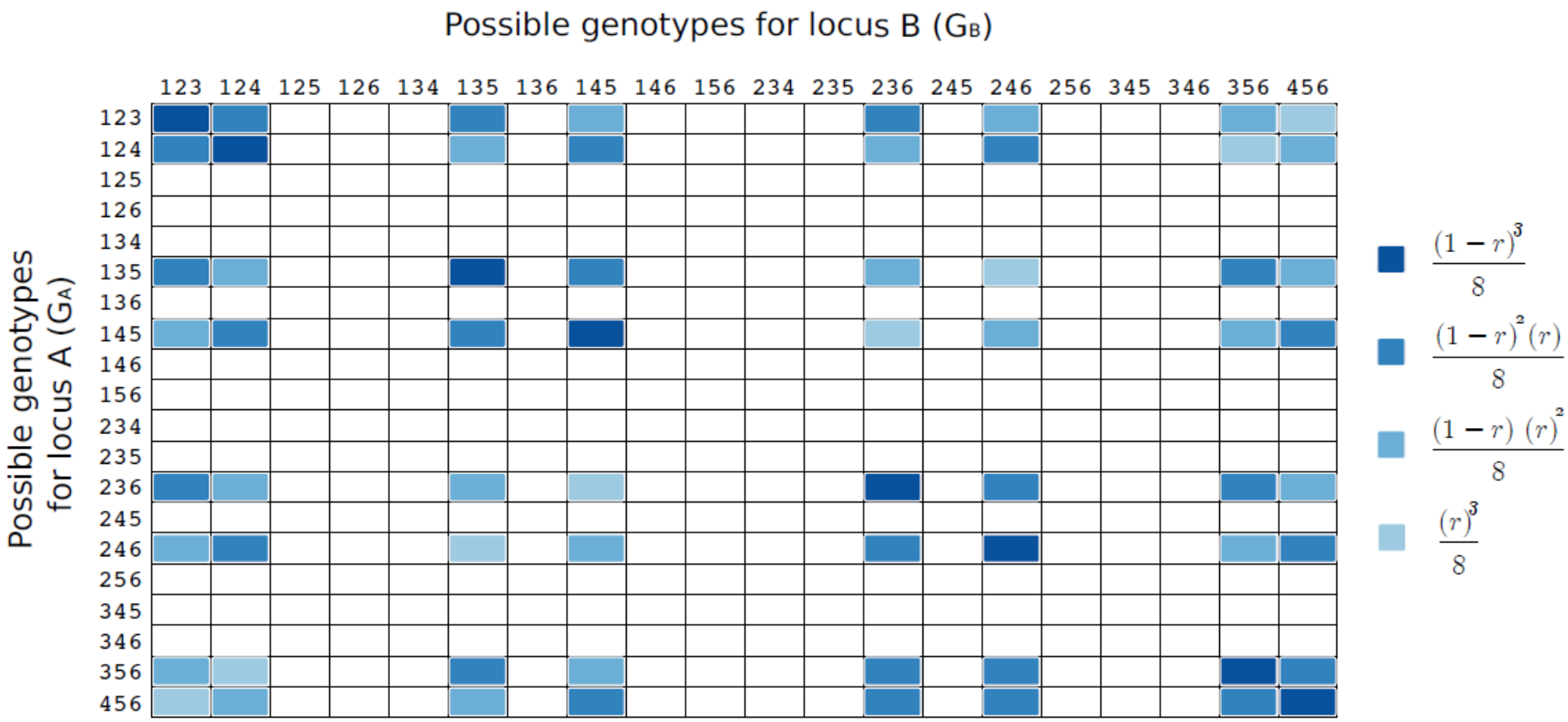
Gametic probability for ψ_1



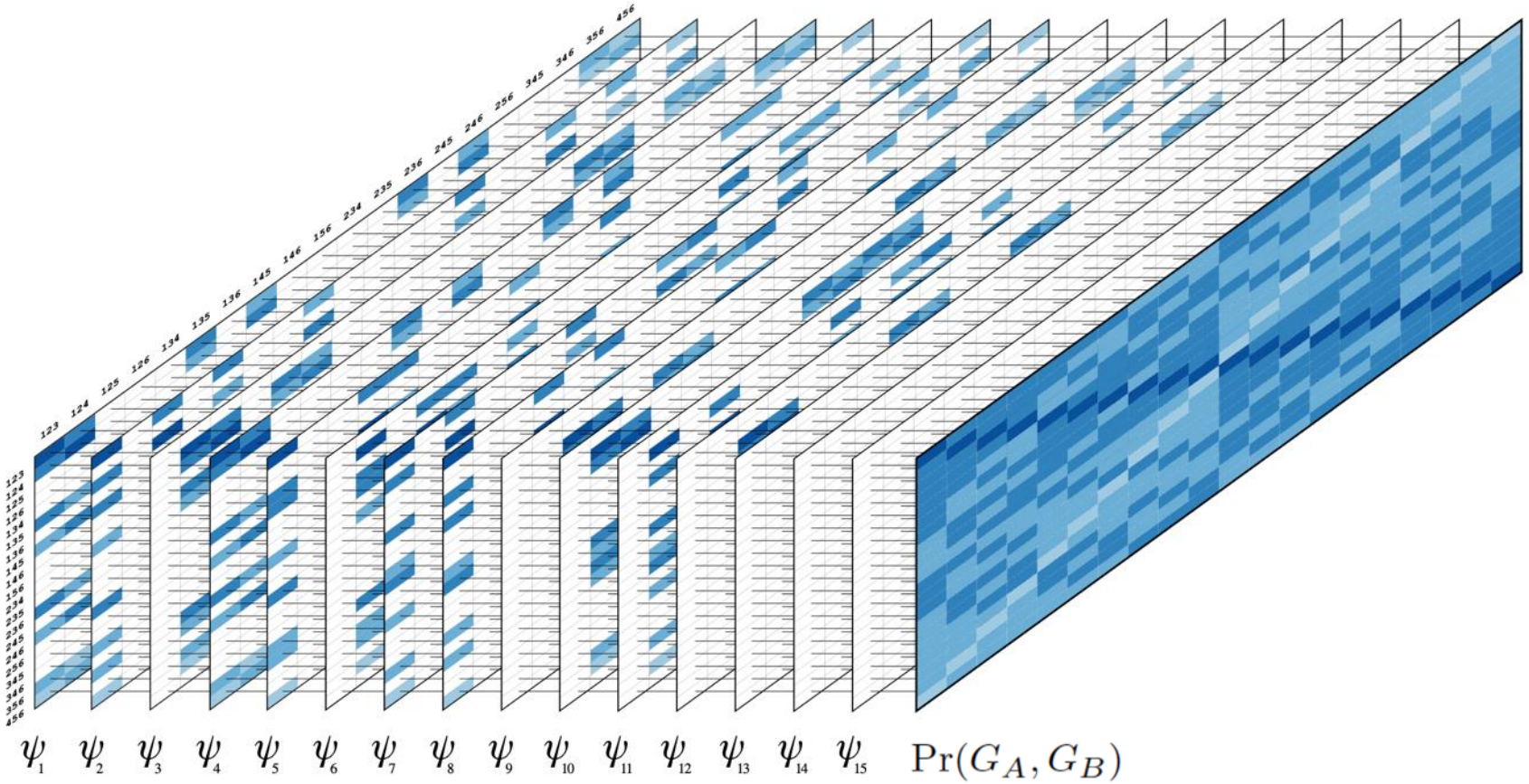
$$\begin{bmatrix} \frac{1-r}{2} & \frac{r}{2} \\ \frac{r}{2} & \frac{1-r}{2} \end{bmatrix} \otimes \begin{bmatrix} \frac{1-r}{2} & \frac{r}{2} \\ \frac{r}{2} & \frac{1-r}{2} \end{bmatrix} \otimes \begin{bmatrix} \frac{1-r}{2} & \frac{r}{2} \\ \frac{r}{2} & \frac{1-r}{2} \end{bmatrix}$$



Gametic probability for ψ_1



Unconditional gametic probability



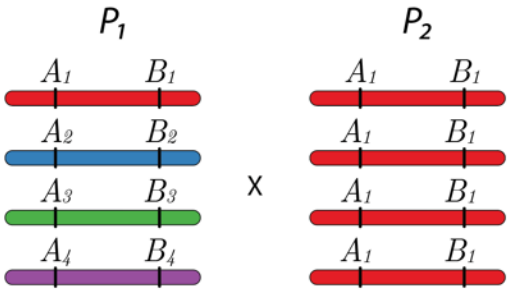
$$\Pr(G_A, G_B) = \sum_j \Pr(G_A, G_B \mid \psi_j) \Pr(\psi_j)$$

$$= \frac{l!\left(\frac{m}{2} - l\right)!}{w_m} \frac{(1 - r_k)^{\frac{m}{2} - l} (r_k)^l}{2^{\frac{m}{2}}}$$

Mollinari and Garcia (2019)

Recombination Fraction – Autotetraploid

Fully informative marker



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

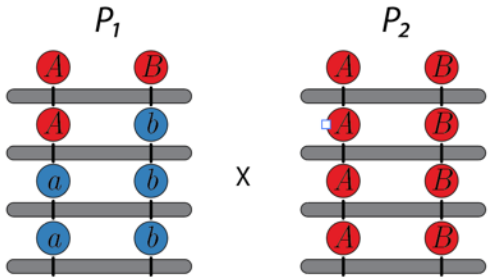
	B_1B_2	B_1B_3	B_1B_4	B_2B_3	B_2B_4	B_3B_4
A_1A_2	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$
A_1A_3	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$
A_1A_4	$\frac{(1-r)r}{12}$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

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

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

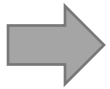
Recombination Fraction – autotetraploid

Partially informative marker – Duplex/simplex – Association



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

	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	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$

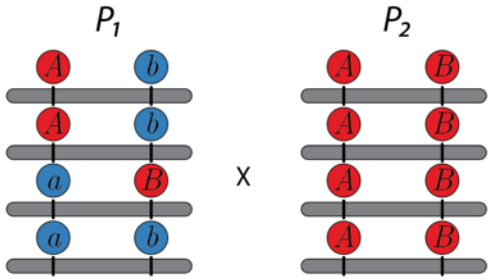


	Bb	bb
AA	$\frac{(1-r)}{6}$	$\frac{r}{6}$
Aa	$\frac{1}{3}$	$\frac{1}{3}$
a a	$\frac{r}{6}$	$\frac{(1-r)}{6}$

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

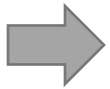
Recombination Fraction – autotetraploid

Partially informative marker – Duplex/simplex – Repulsion



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

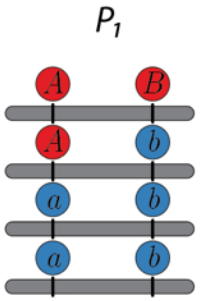
	bb	Bb	bb	Bb	bb	Bb
AA	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$
Aa	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$
Aa	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$
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	$\frac{1}{3}$	$\frac{1}{3}$
a a	$\frac{(1-r)}{6}$	$\frac{r}{6}$

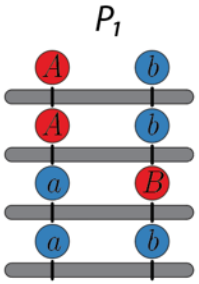
$$L_2(r) = \left[\frac{r}{6}\right]^{n_1} \left[\frac{1}{3}\right]^{n_2} \left[\frac{(1-r)}{6}\right]^{n_3}$$

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

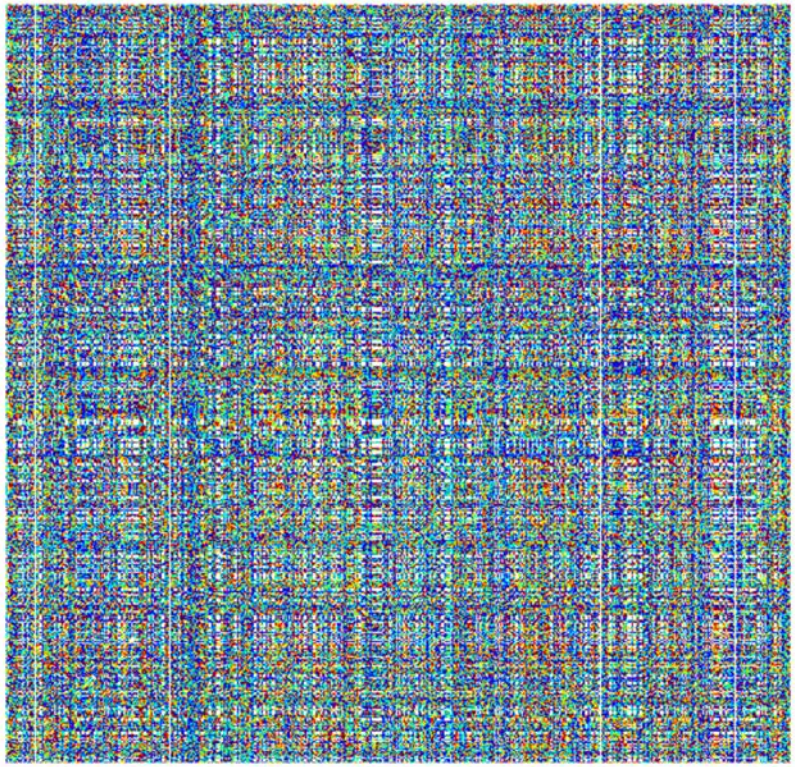


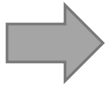
$$\hat{r_2} = \underset{r}{\operatorname{argmax}}\, L_2(r) \implies L_2(\hat{r_2})$$

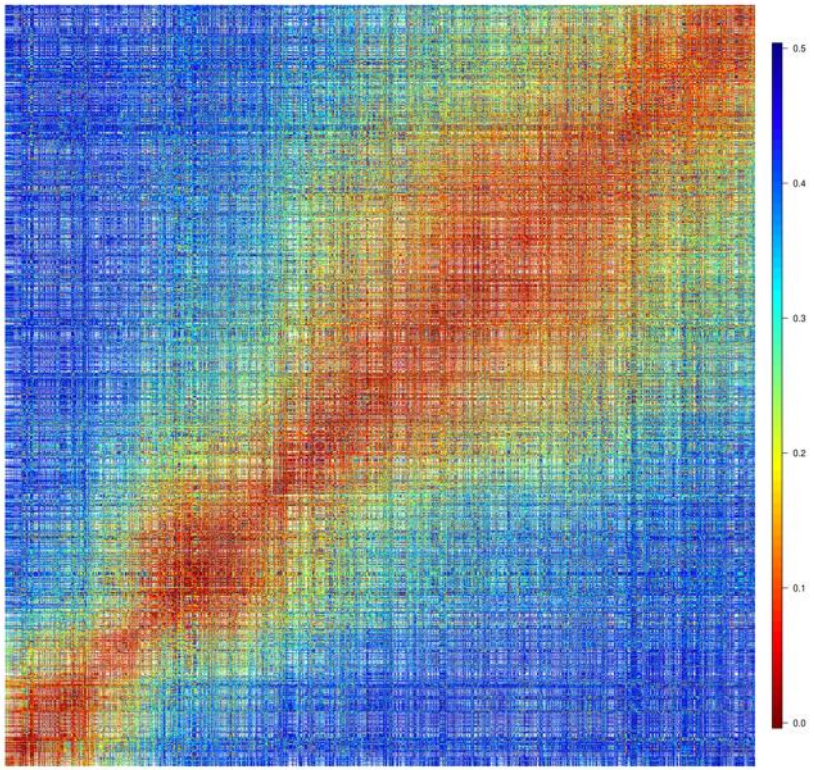
- Pairwise MLE of r are used to group markers into linkage groups and order markers within each linkage group using optimization algorithms such as MDS
- Given a sequence of ordered markers, it is possible to extend the idea of comparing likelihoods of competing linkage phases throughout multiple markers

Multidimensional Scaling Algorithm (MDS)

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





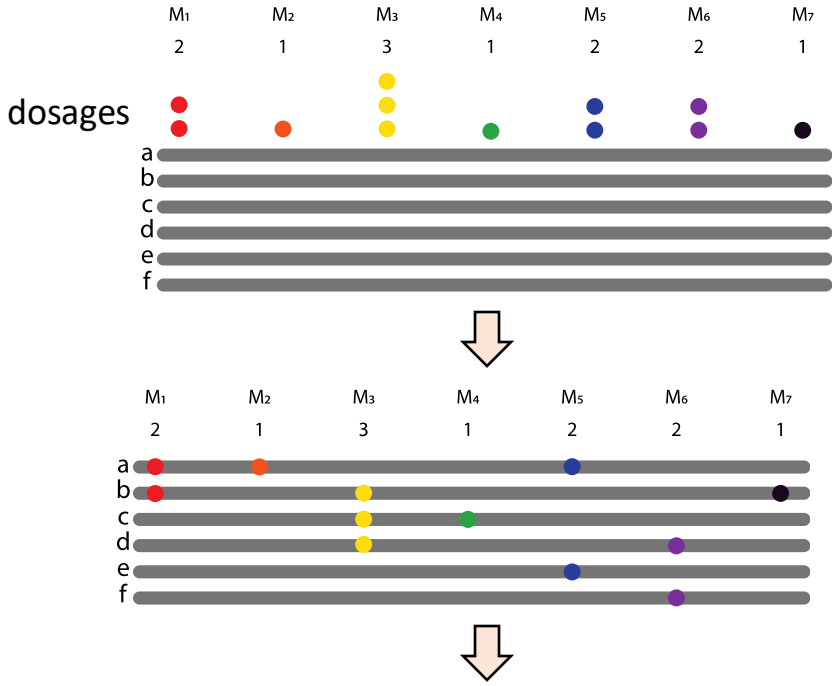


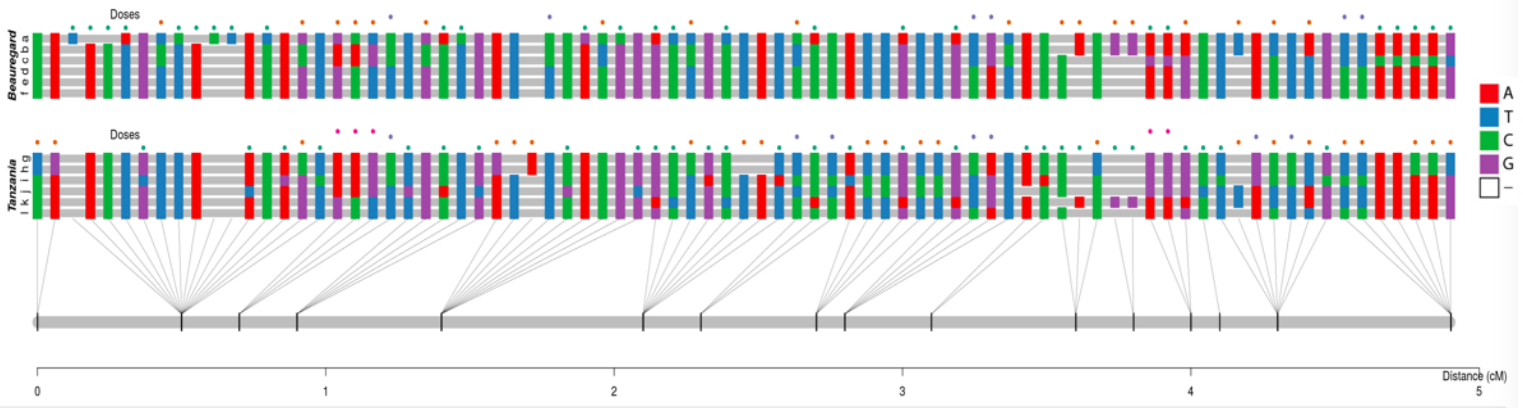
Sweetpotato linkage group 1: 2745 markers

Preedy and Hackett, 2016

Haplotyping in polyploids

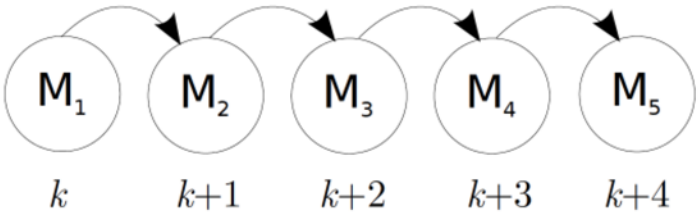
- Disposition of allelic variants in the homologs in a homology group





Multilocus linkage analysis in polyploids

Markov Model: conditional independence

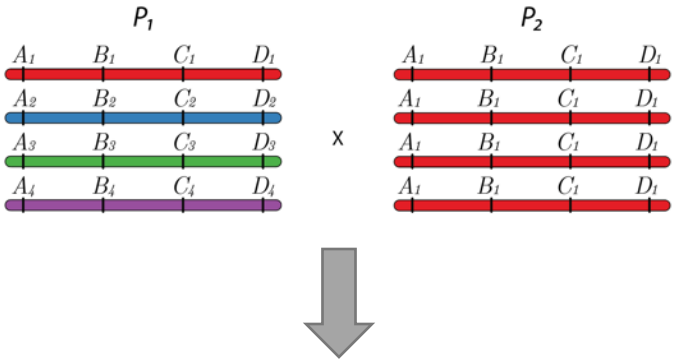


$$\Pr(G_{k+1}|G_k) = \frac{(1-r_k)^{\frac{p}{2}-l}(r_k)^l}{\binom{\frac{p}{2}}{l}}$$

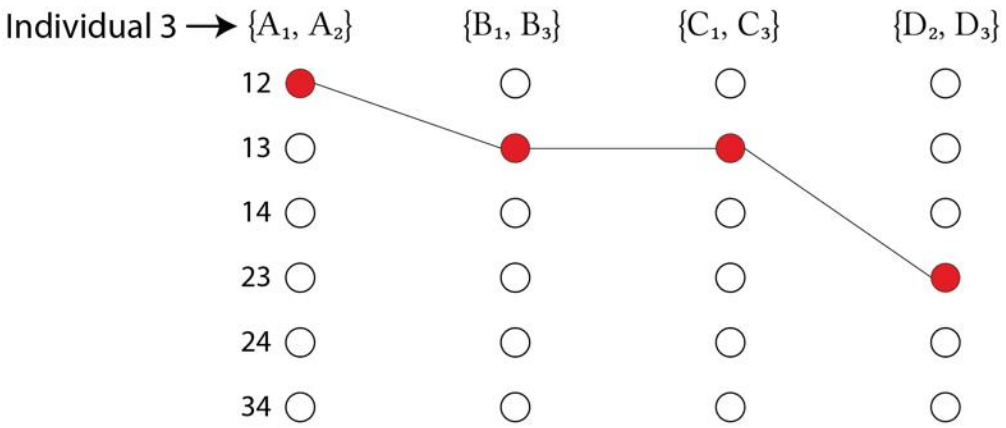
where r_k is the recombination frequency between loci k and k+1, p is the ploidy level and l is the number of recombinant events between k and k+1.

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	$\{A_1, A_2\}$	$\{B_1, B_3\}$	$\{C_1, C_3\}$	$\{D_2, D_3\}$
$\vdots$			$\vdots$	
Ind_n	$\{A_2, A_3\}$	$\{B_2, B_3\}$	$\{C_2, C_3\}$	$\{D_2, D_3\}$

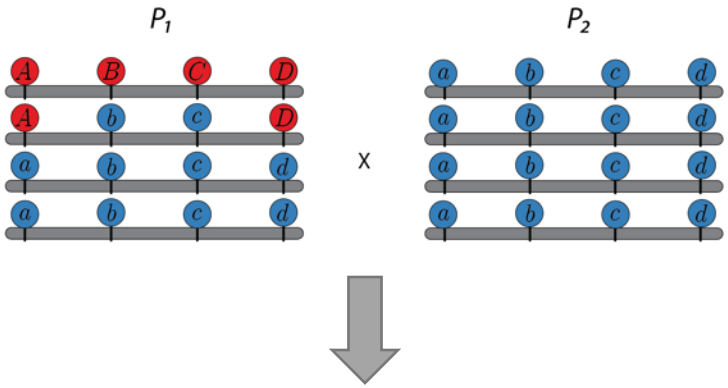


$$L(\mathbf{r}) = \prod_{n} \Pr(G_A, G_B, G_C, G_D \mid \mathbf{r})$$

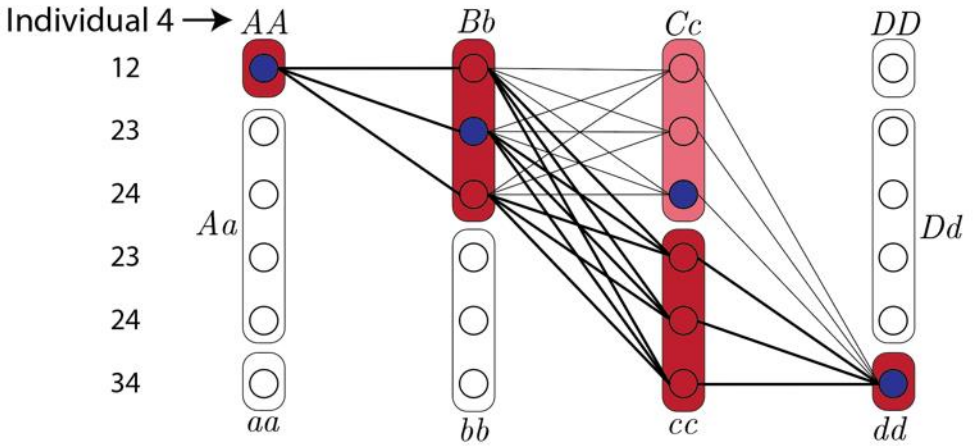
$$\hat{\mathbf{r}} = \underset{\mathbf{r}}{\operatorname{argmax}}\, L(\mathbf{r})$$

Hidden Markov Model - HMM

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



	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 0.8)	Dd
⋮			⋮	
Ind_n	Aa	Bb	Cc	Dd

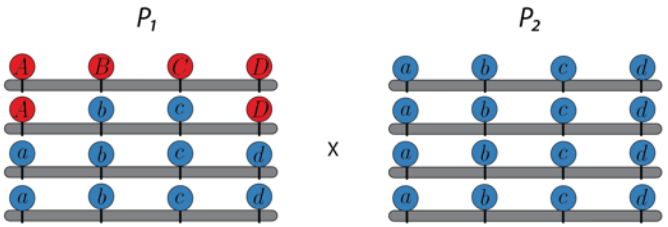


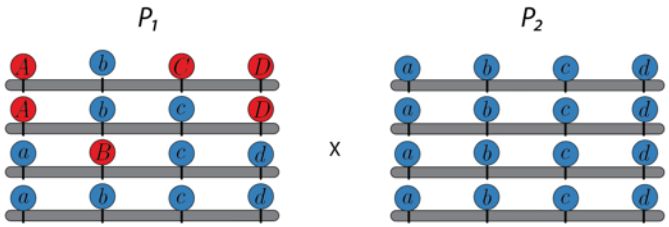
$$L(\mathbf{r}) = \prod_n \Pr(G_A, G_B, G_C, G_D \mid \mathbf{r})$$

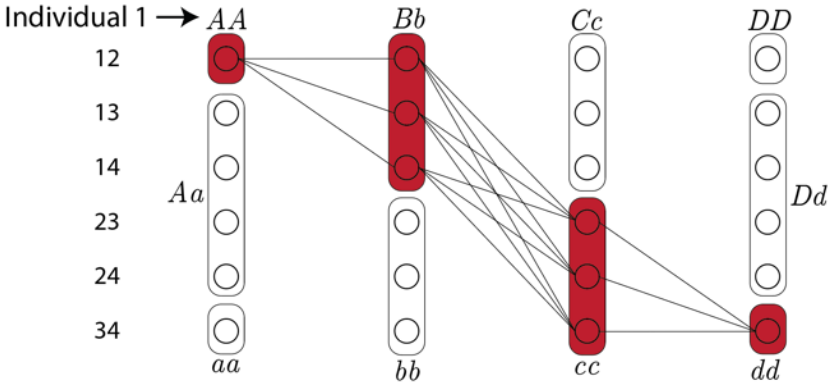
$$\hat{\mathbf{r}} = \underset{\mathbf{r}}{\operatorname{argmax}}\, L(\mathbf{r})$$

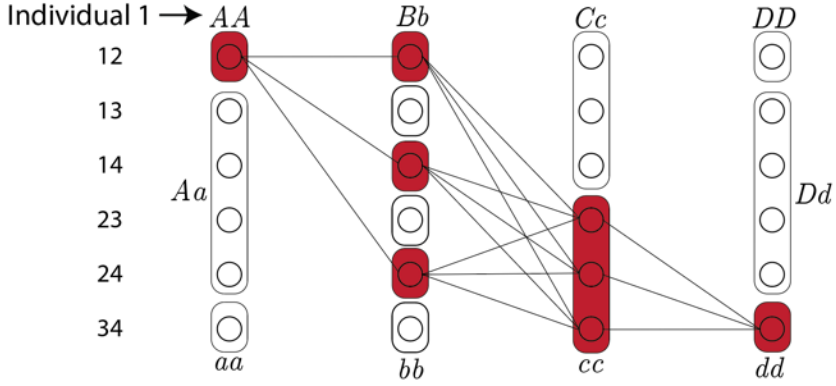
Hidden Markov Model - HMM

Assessing *different linkage phases* using multilocus analysis

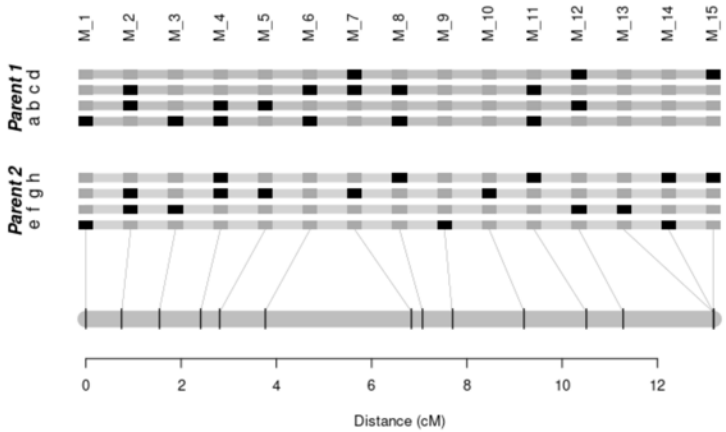






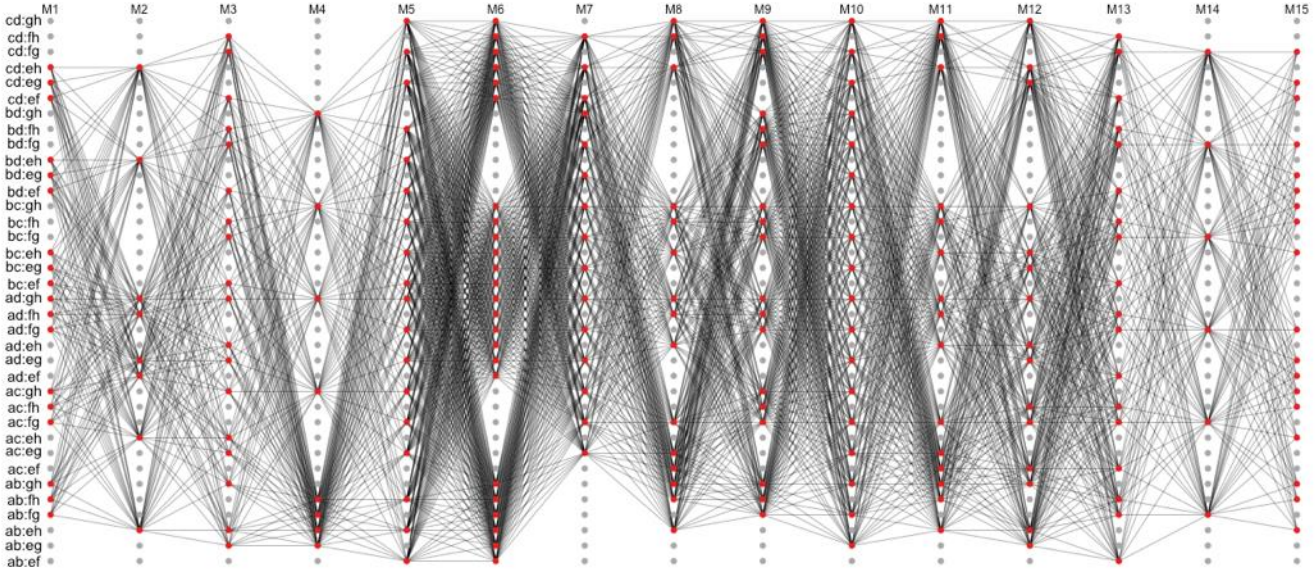


Hidden Markov Model - HMM



Individual 64:

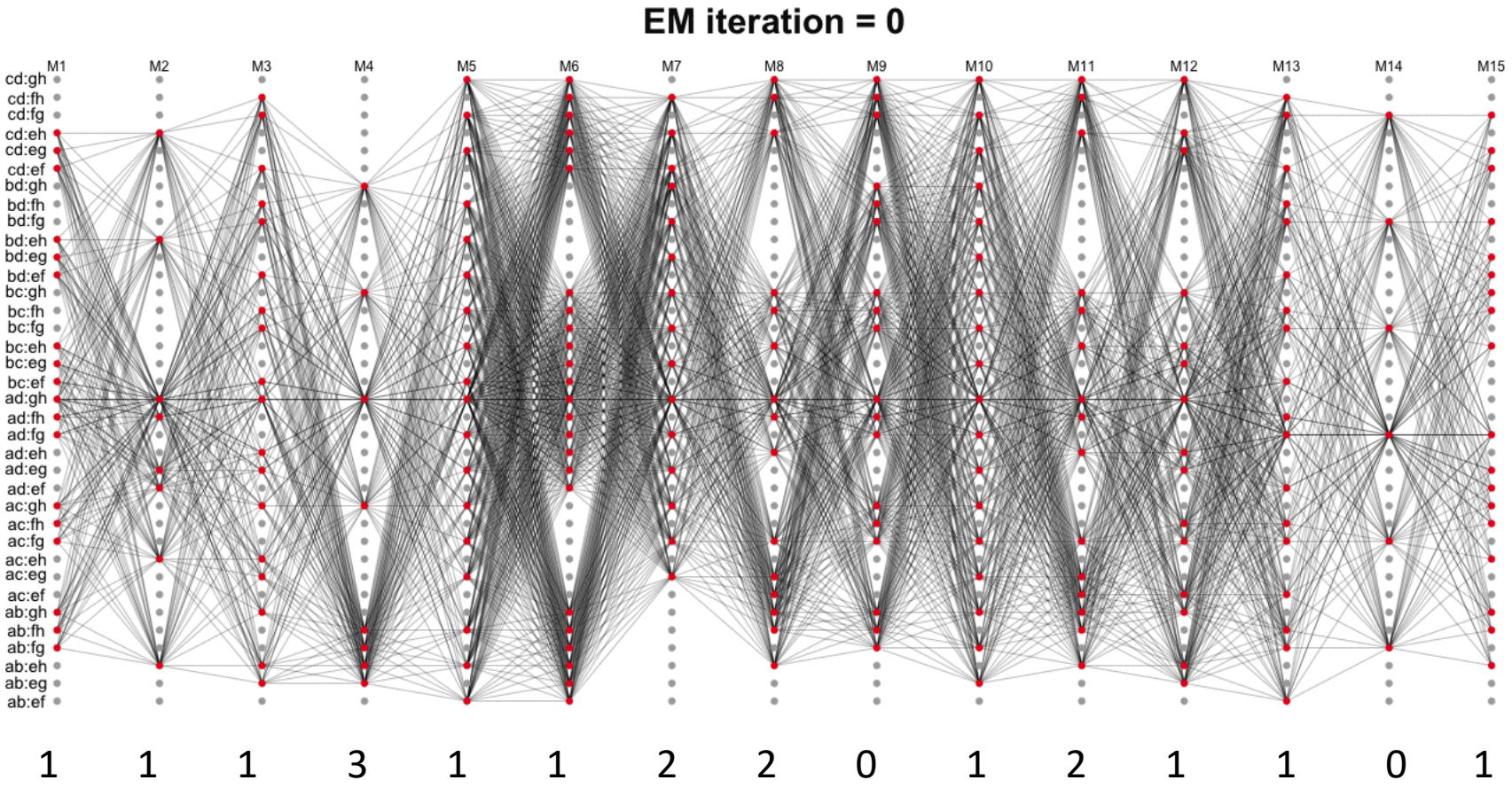
1	1	1	3	1	1	2	2	0	1	2	1	1	0	1



Rabiner (1989): https://doi.org/10.1109/5.18626

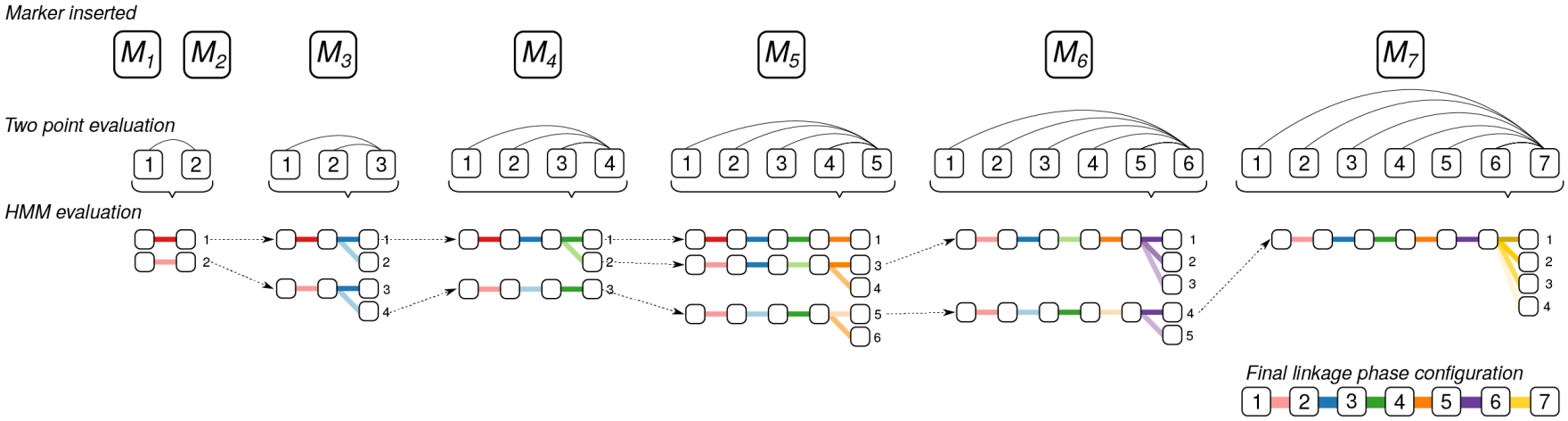
Hidden Markov Model - HMM

- Tetraploid example, one individual, 15 markers



Haplotype phasing – MAPpoly strategy

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



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

X

Tanzania

Biparental Population - BT

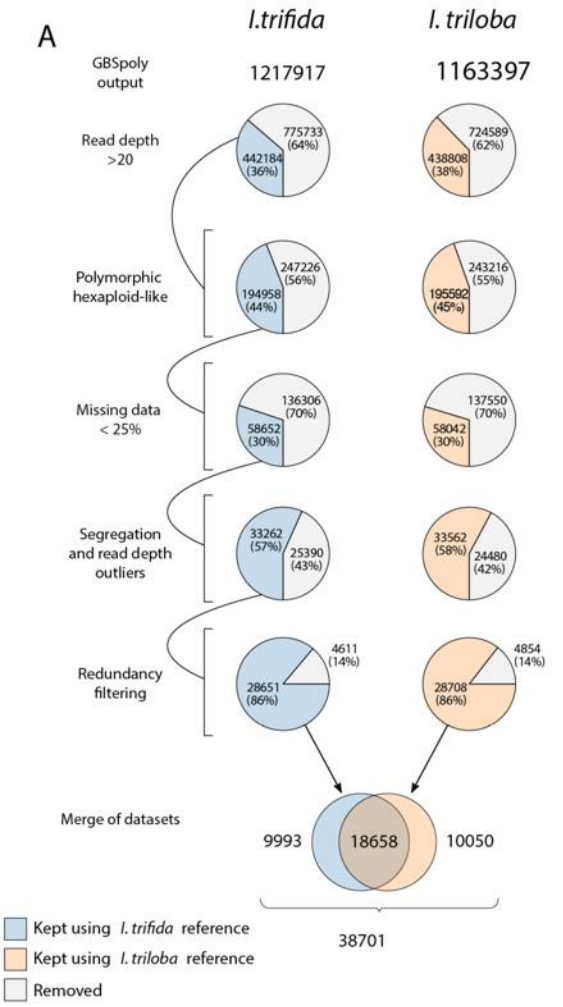
Beauregard

X

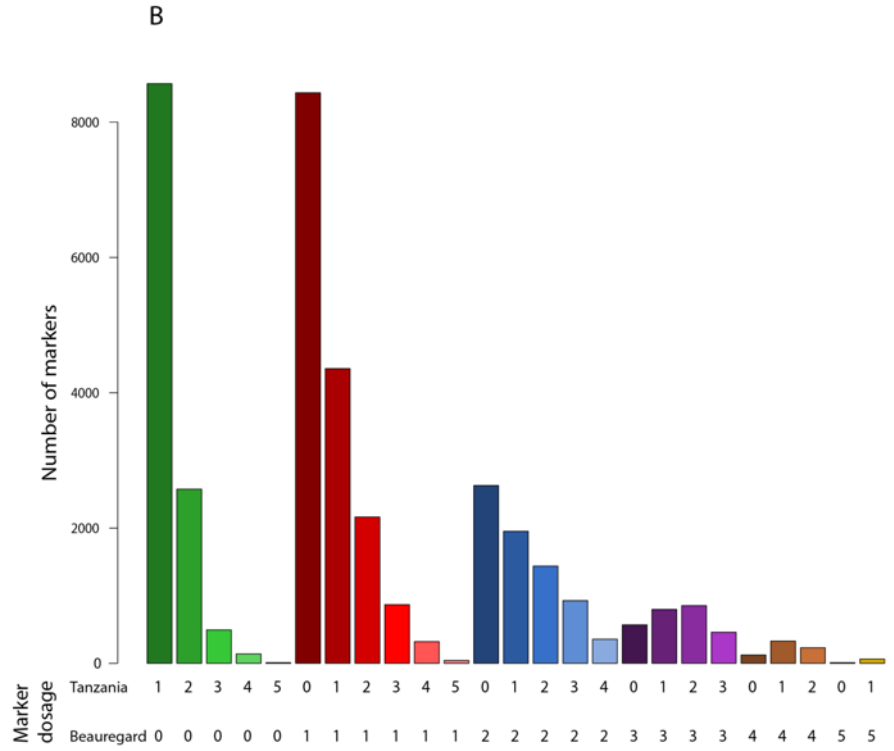
Tanzania

Results - Genotyping Calling - BT population

Filtering process

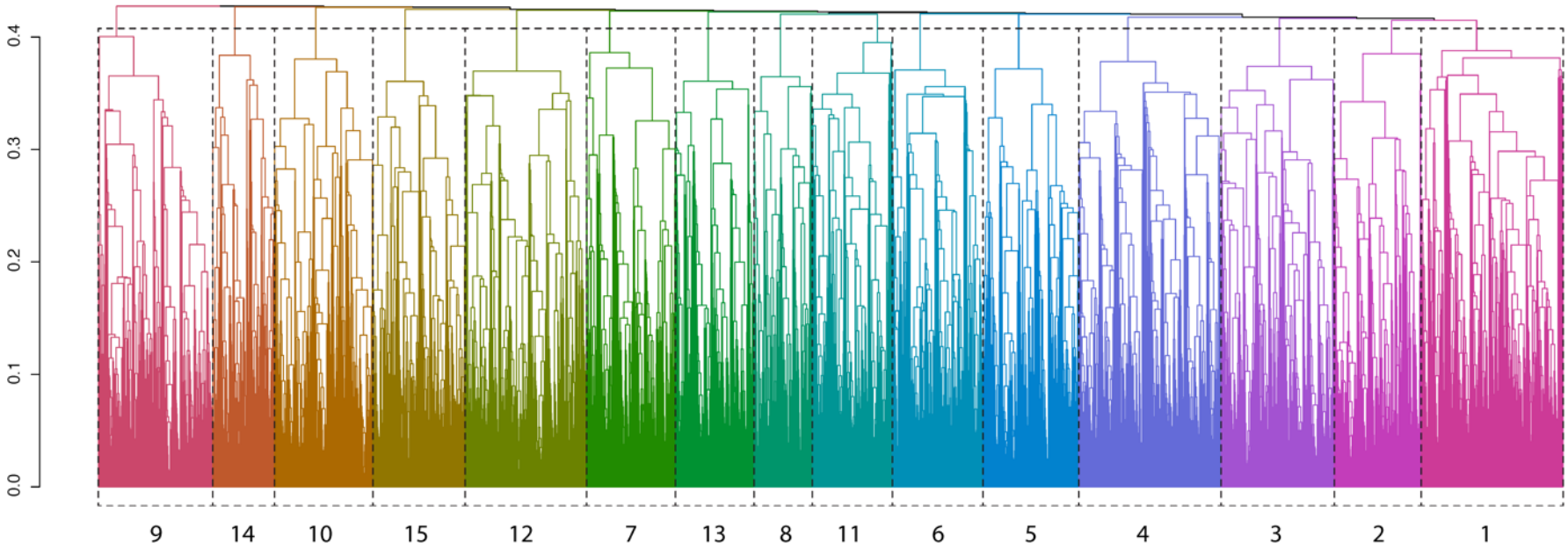


Distribution of the marker doses in *Beauregard* and *Tanzania* (38,701 SNPs).



Two – point analysis and grouping

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

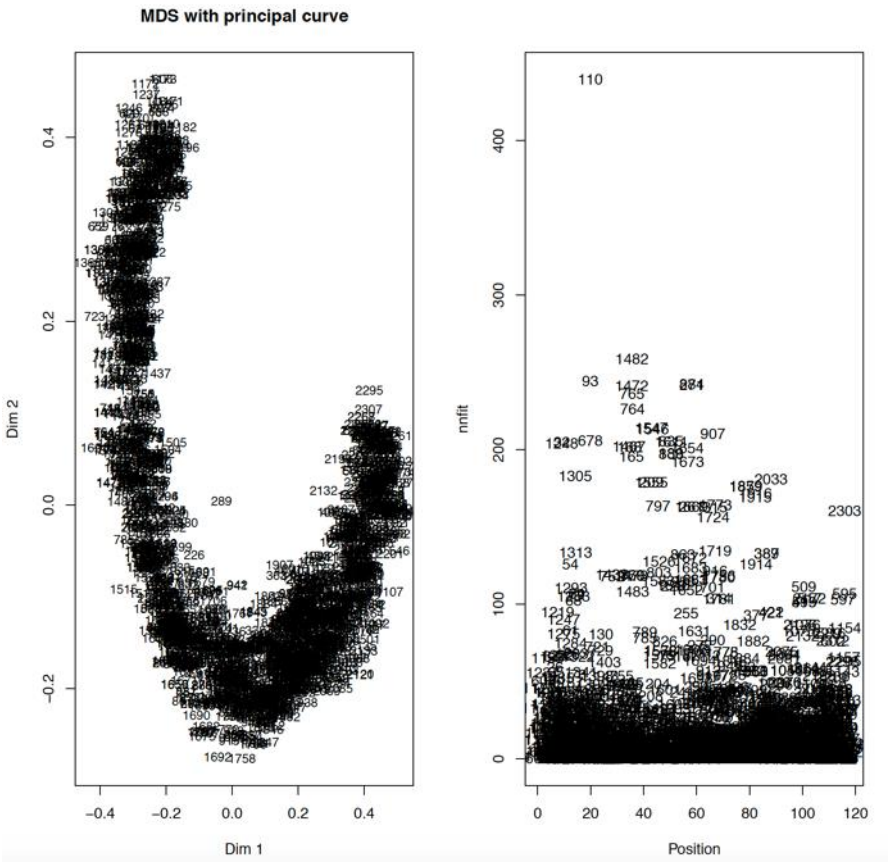


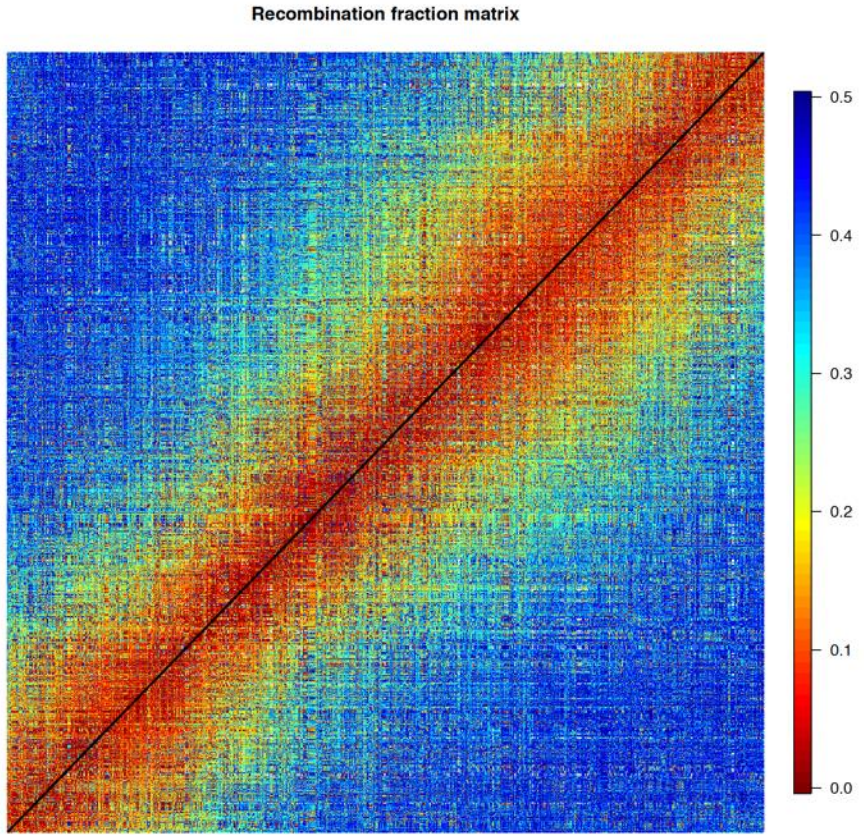
Numbers indicate the associated chromosomes in *I. trifida* and *I. triloba* reference genomes

Ordering - Multidimensional Scaling Algorithm (MDS)

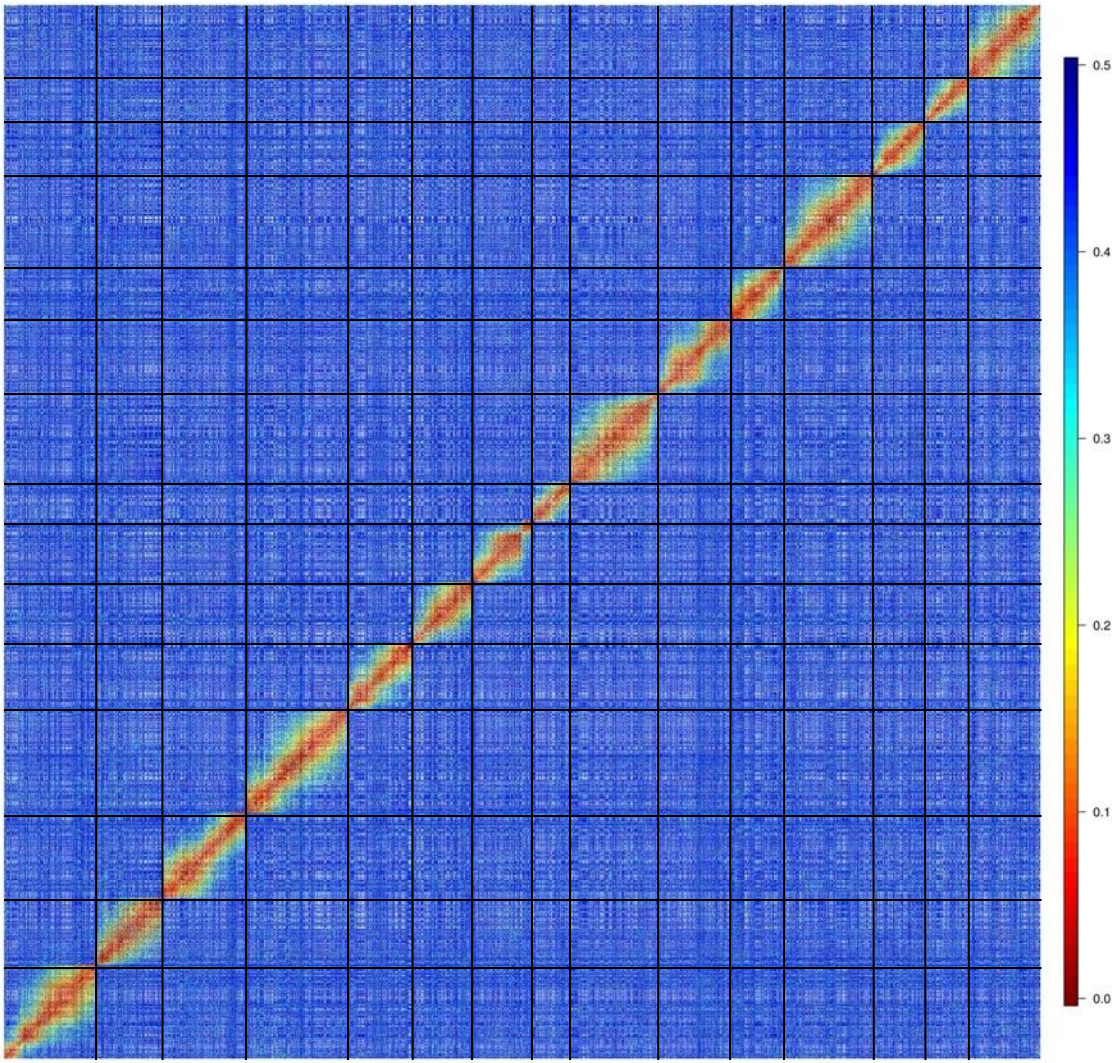
- Projects the information contained in the recombination fraction matrix in two (or more) dimensions. It provides a visual representation of the pattern of proximities among markers (Preedy and Hackett, 2016).

Linkage group 1: 2745 markers





Ordering with MDS – 15 linkage groups



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

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

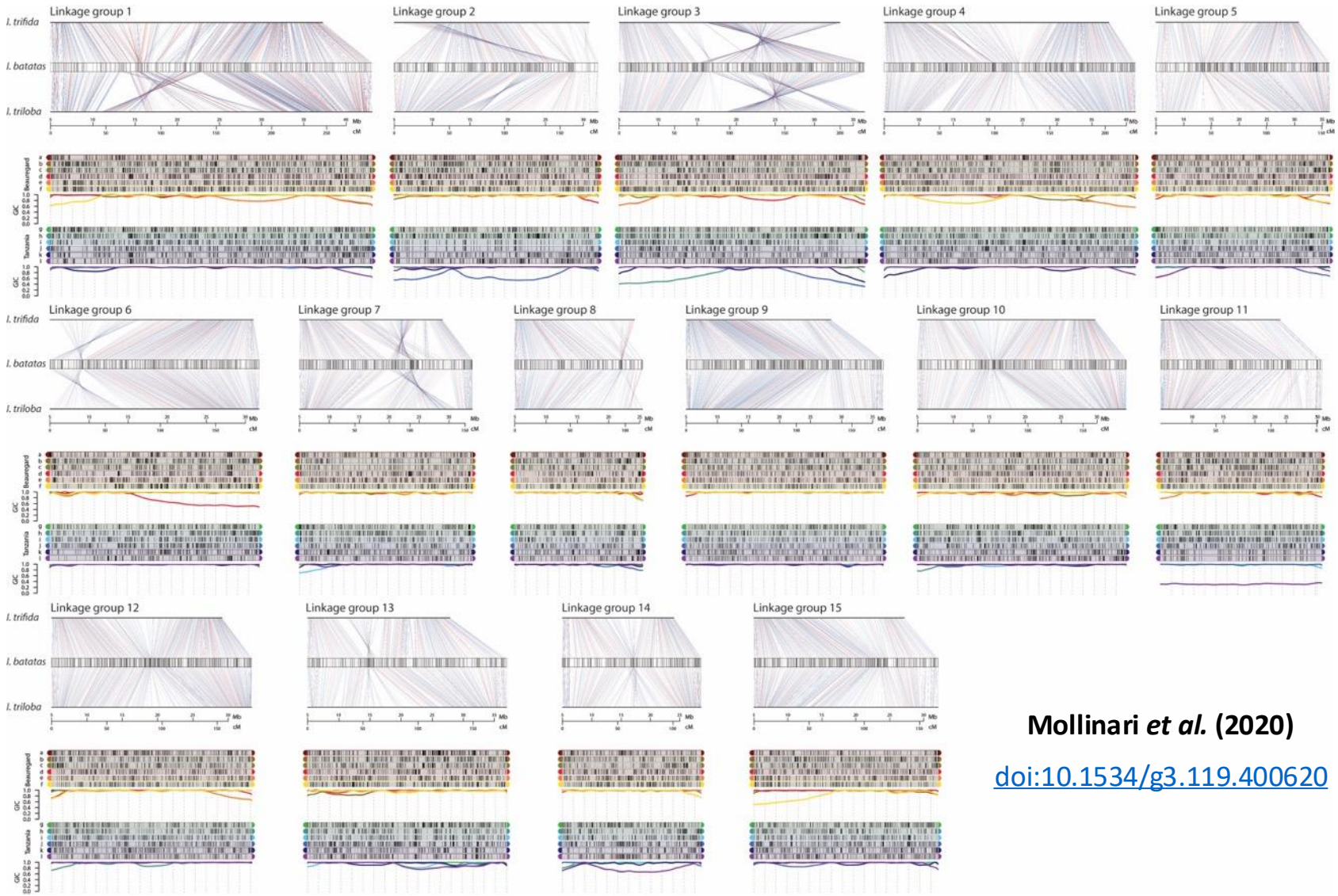
Unraveling the Hexaploid Sweetpotato Inheritance Using Ultra-Dense Multilocus Mapping

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

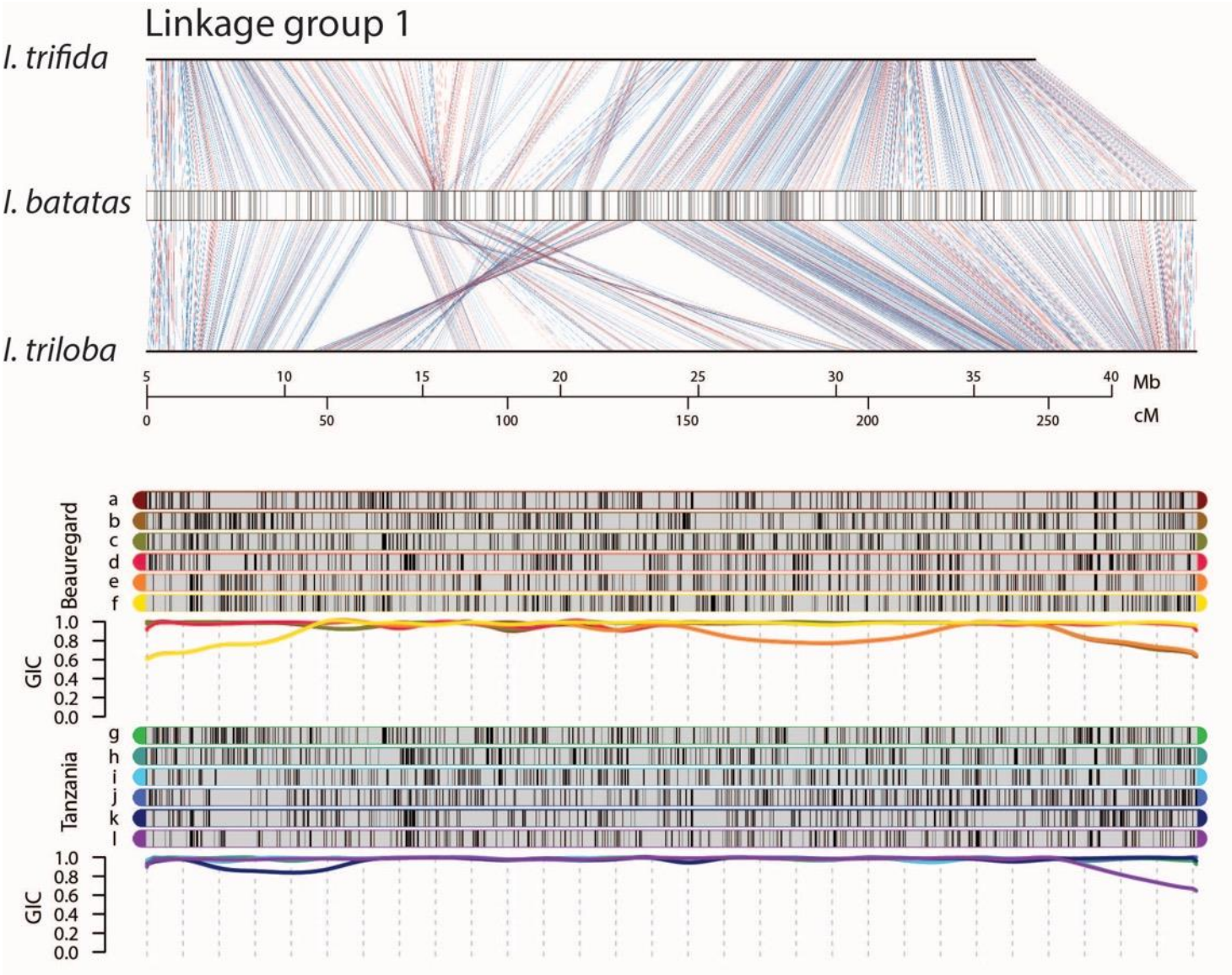
Sweetpotato genetic map



Mollinari *et al.* (2020)

doi:10.1534/g3.119.400620

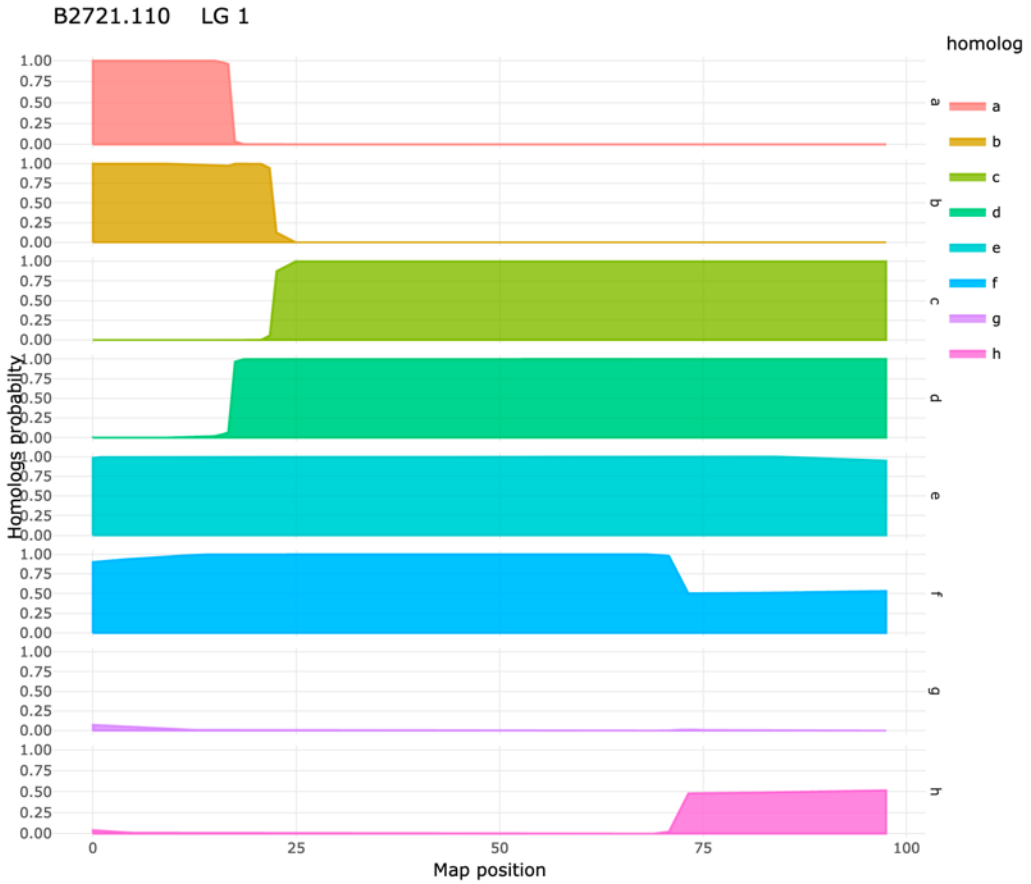
Sweetpotato genetic map



Probabilistic haplotype reconstruction

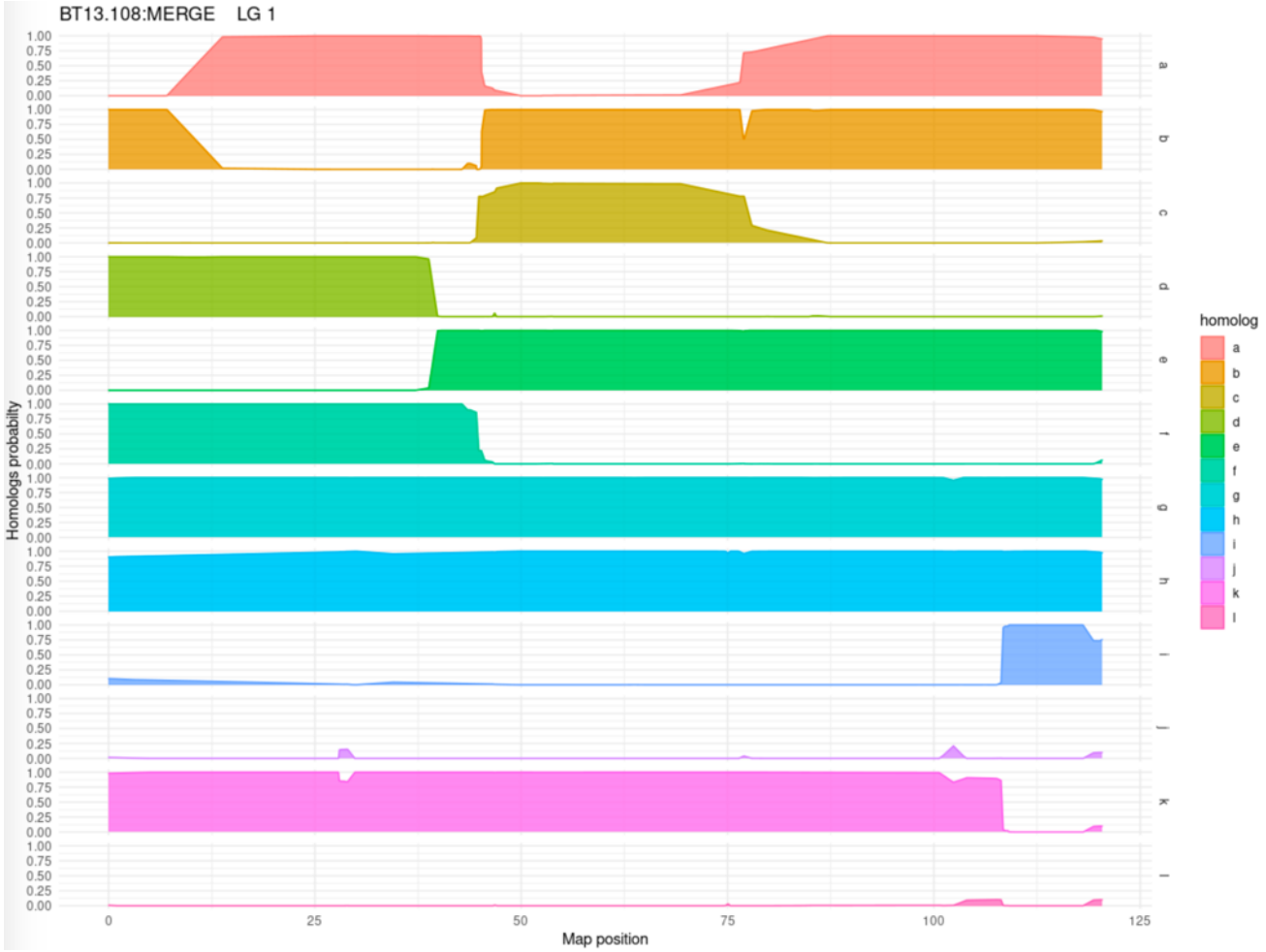
- When assuming a *prior probability* distribution of the genotypes, multilocus strategies can improve the quality of the inferred haplotypes

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

MAPpoly

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 Mollinari 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 (Mollinari 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 Git Hub. Within R, you need to install and load the package `devtools`:

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

To install MAPpoly from Git Hub use

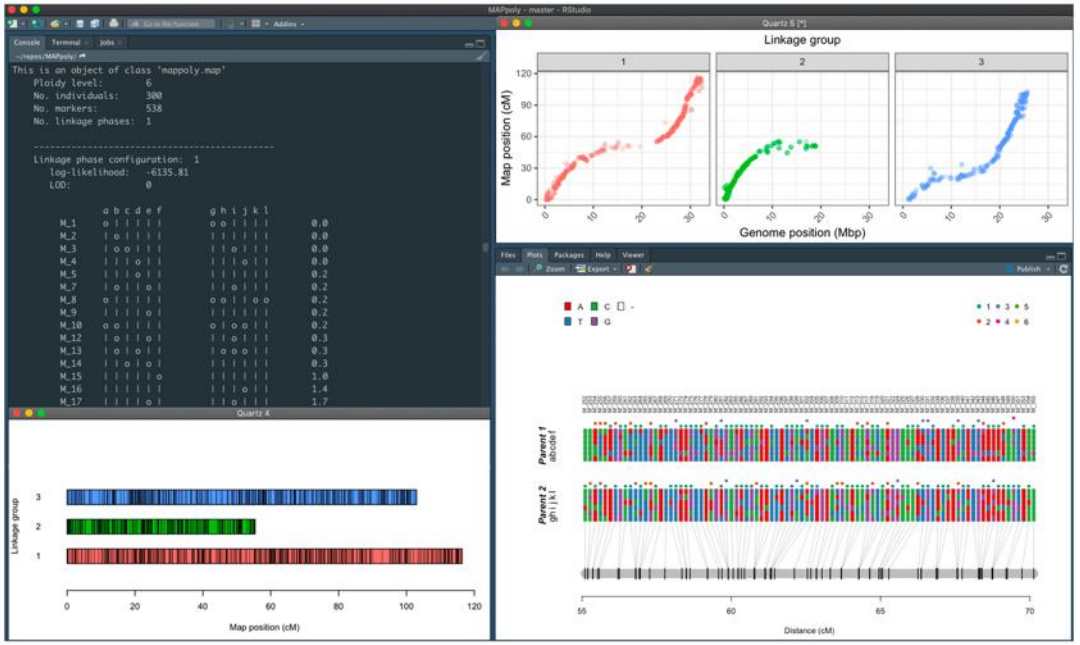
```
devtools::install_github("mmollina/mappoly")
```

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

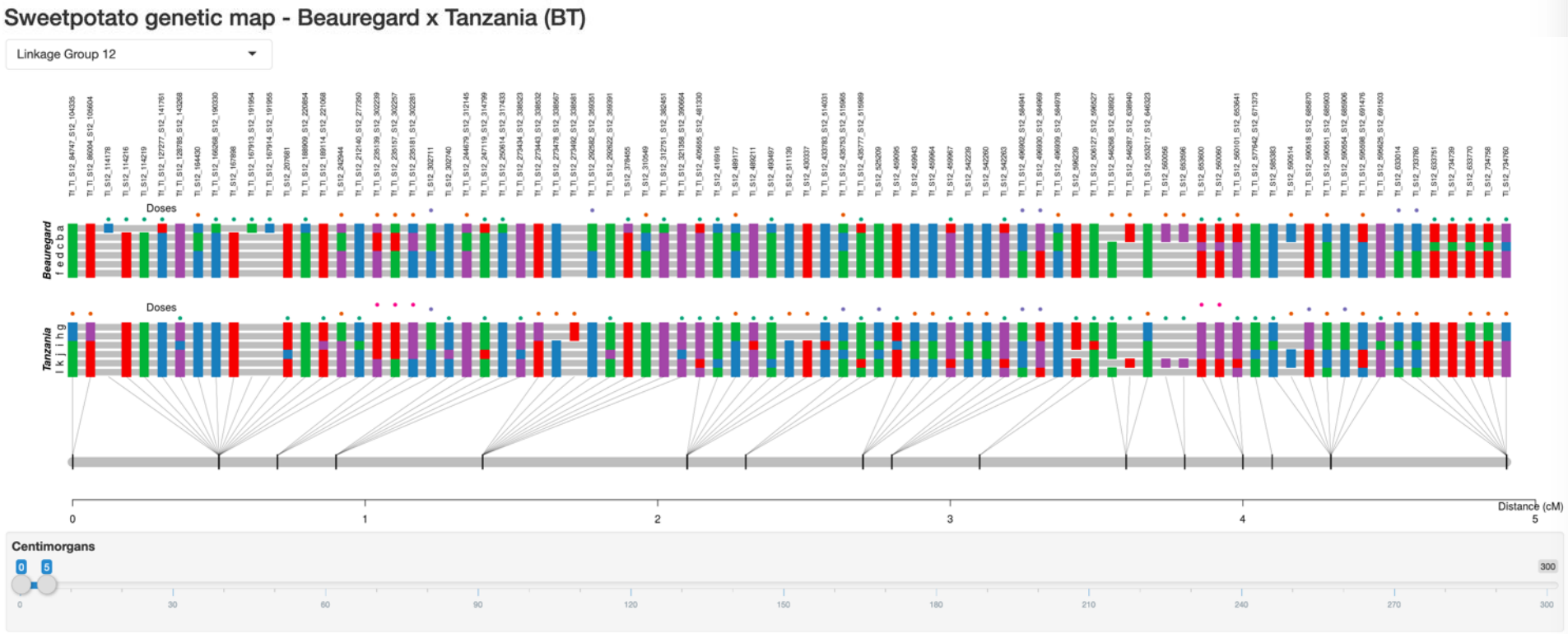
This package has been developed as part of the Genomic Tools for Sweetpotato Improvement project (GT4SP), funded by Bill & Melinda Gates Foundation.



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

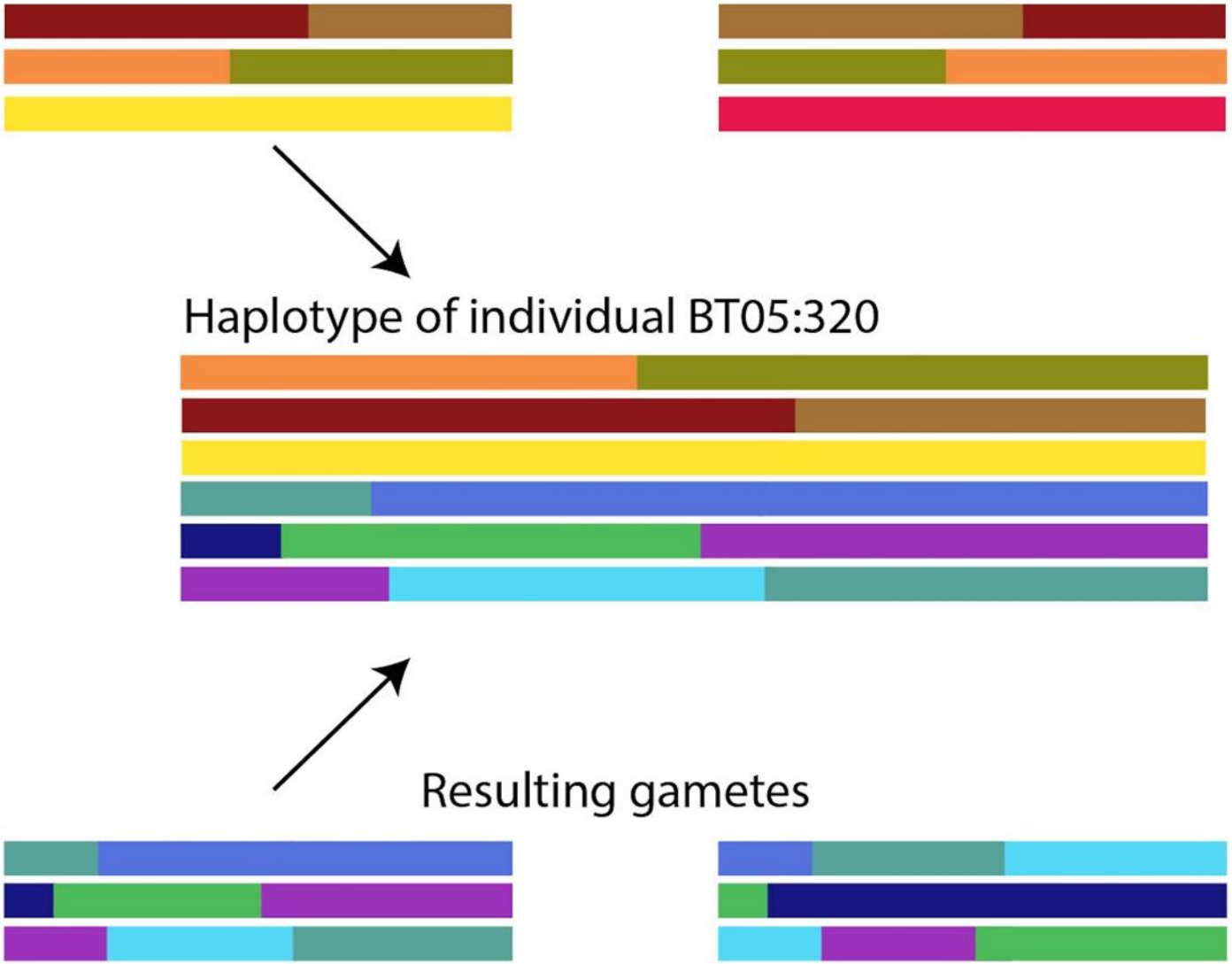
development: https://github.com/mmollina/MAPpoly

Genetic mapping – Linkage group 12 – 2661 SNPs



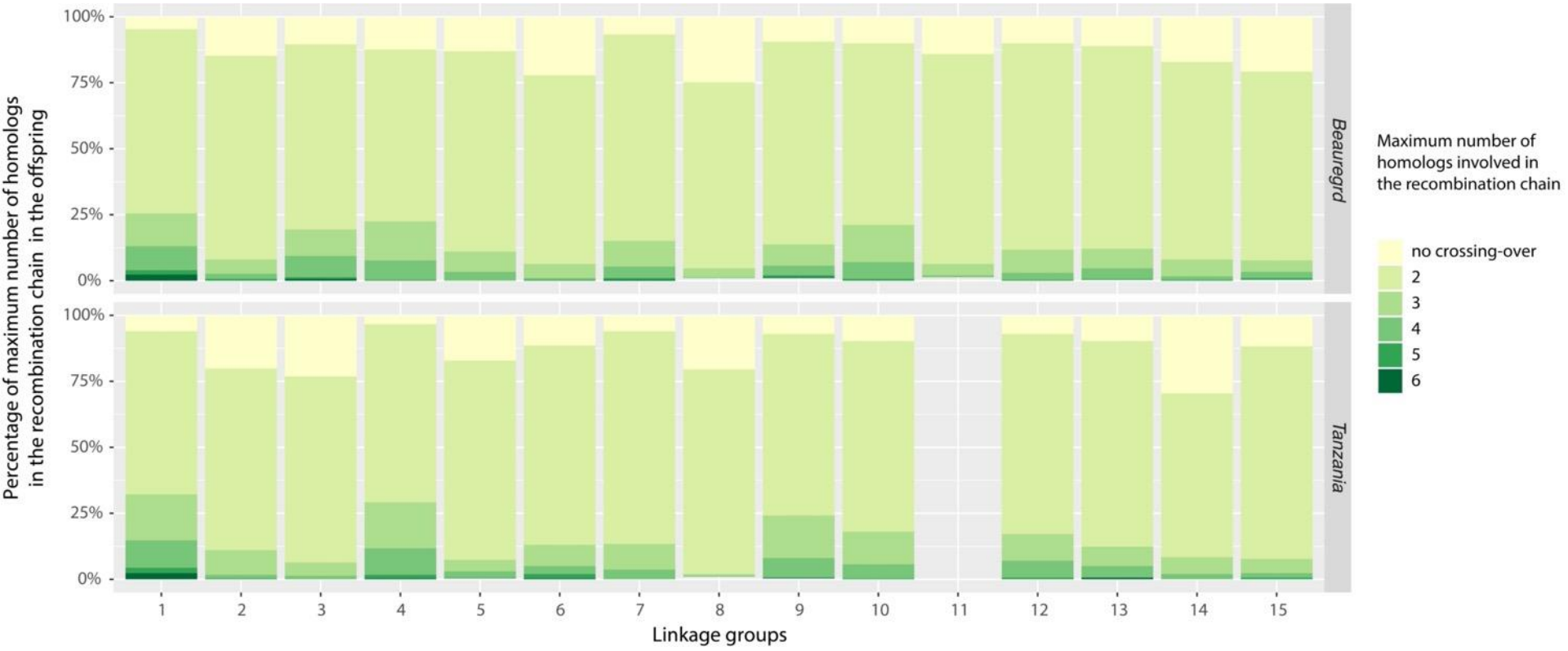
Interactive version: https://gt4sp-genetic-map.shinyapps.io/bt_map/

Haplotype reconstruction in the offspring



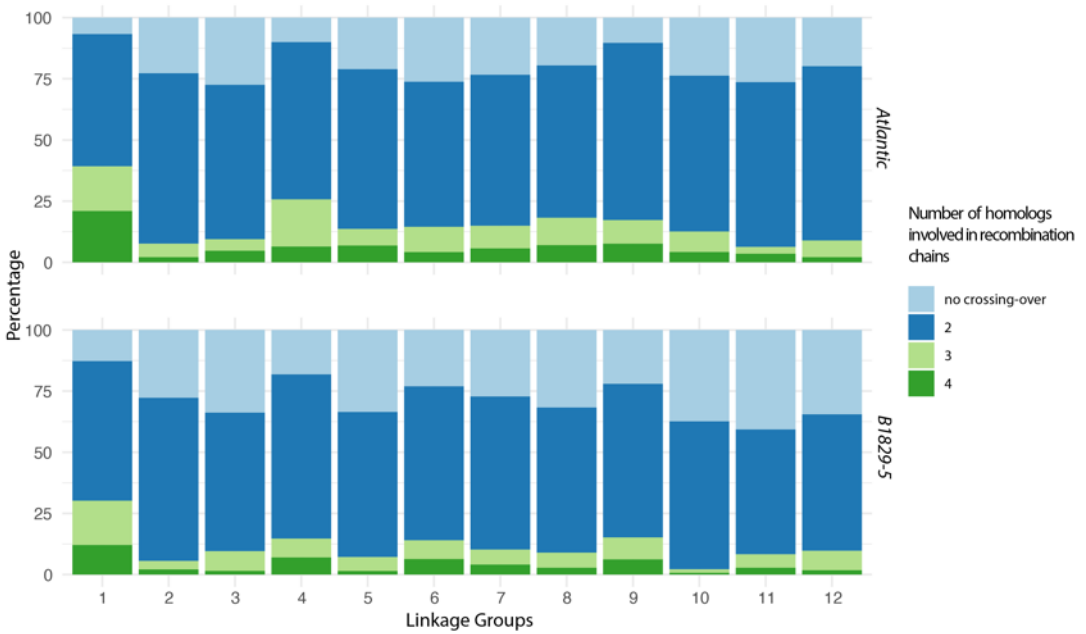
https://gt4sp-genetic-map.shinyapps.io/offspring_haplotype_BT_population/

Number of homologs involved in recombination chains



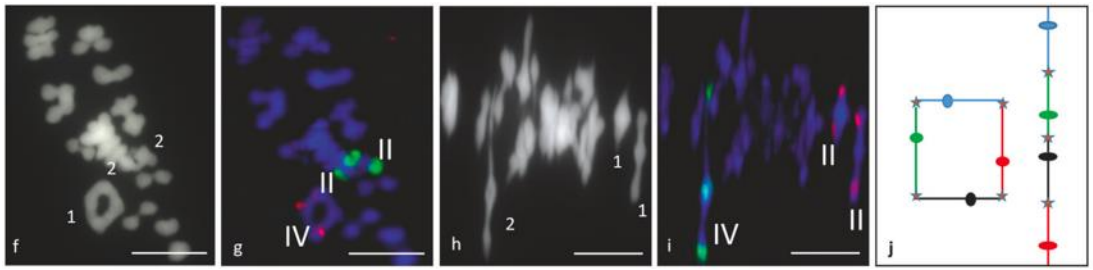
Homologs in recombination chains – potato

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

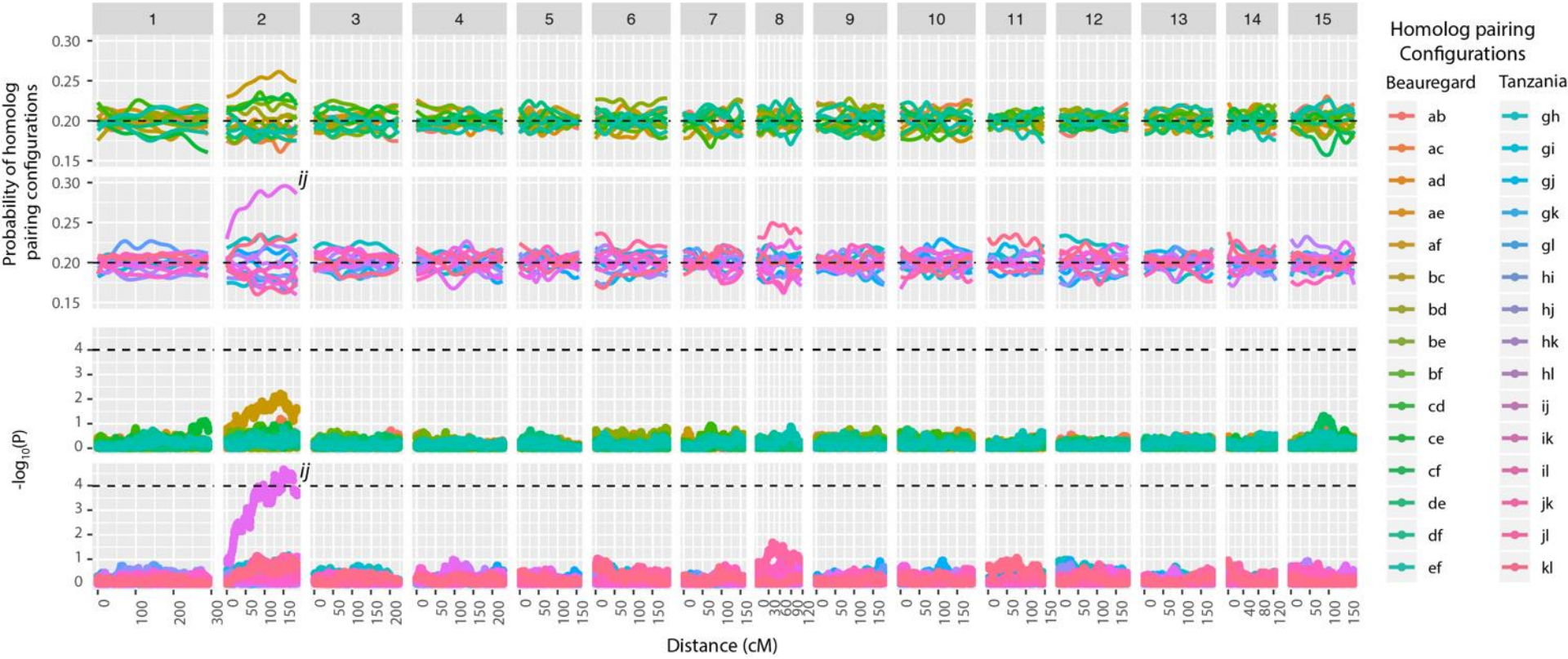


	Choudhary *et al.* (2020)	Pereira *et al.* (2020)
bivalents	predominant	62.3 %
multivalents	7~48%	2.2~39.2%

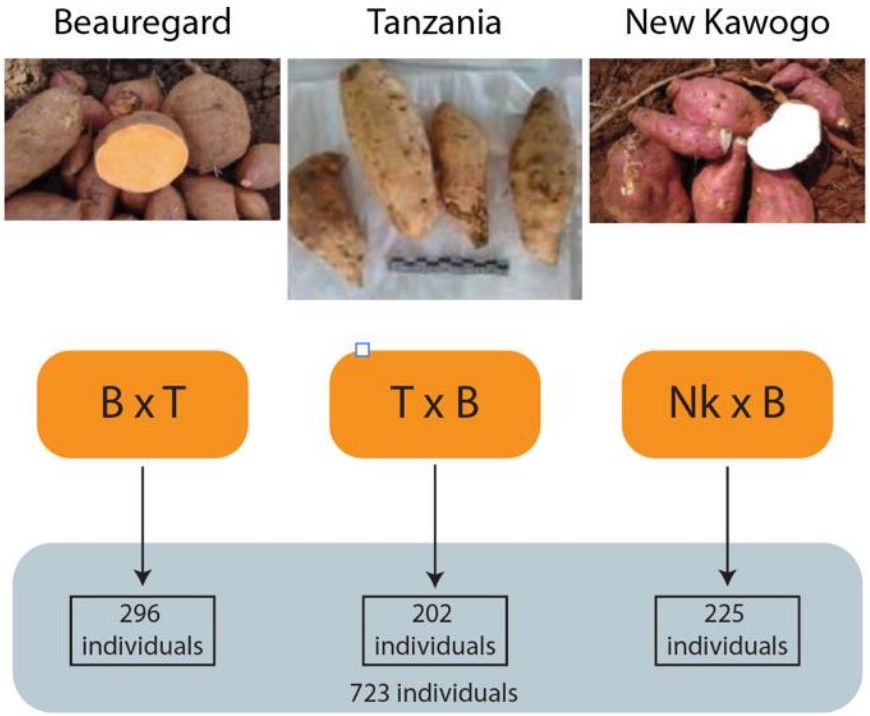
Choudhary *et al.* (2020) – Using fluorescence *in situ* hybridization (FISH)
5S rDNA probe (red) and 45S rDNA probe (green)

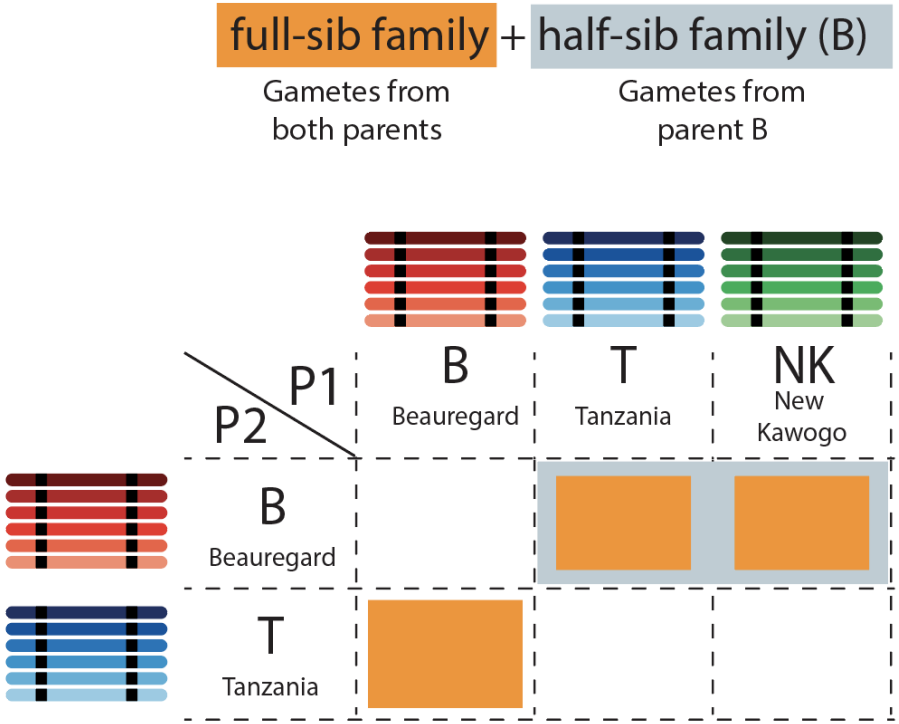


Preferential pairing profiles: Sweetpotato is vastly autohexaploid

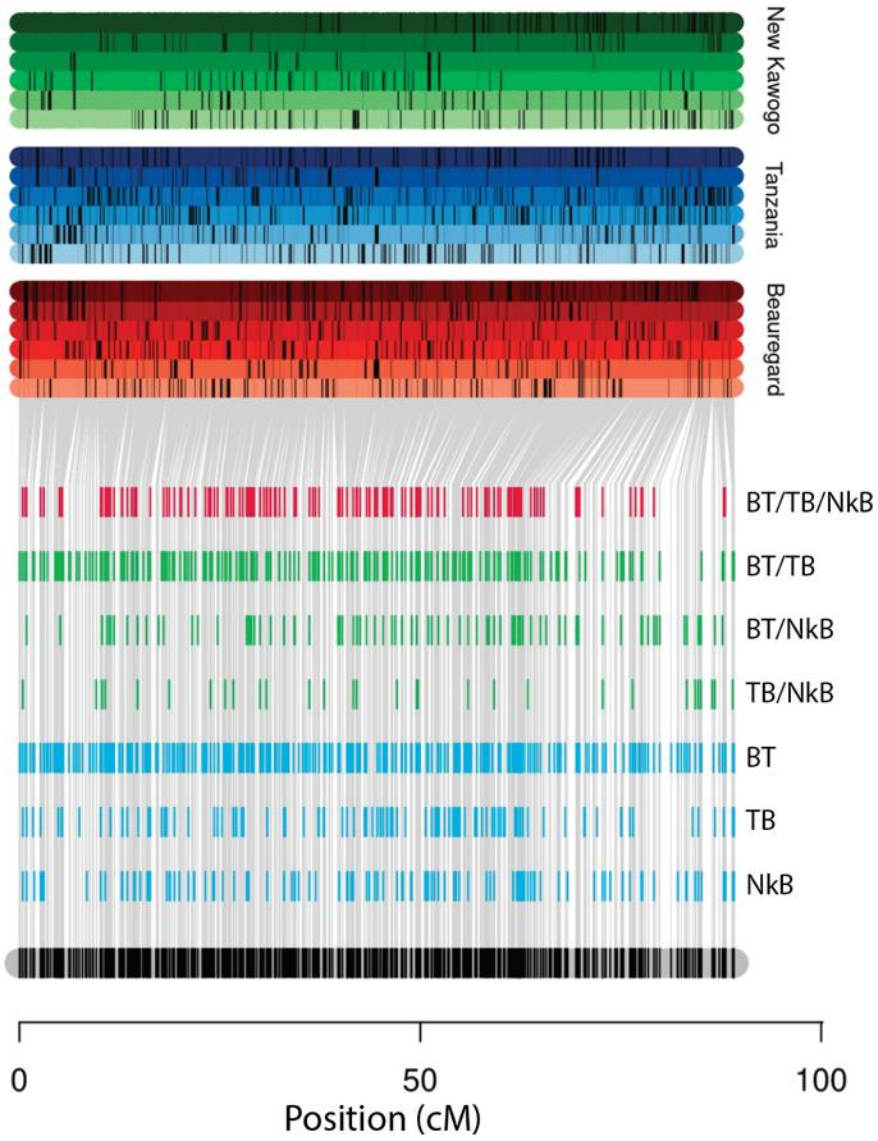


Linkage analysis in multiple inter-connected families

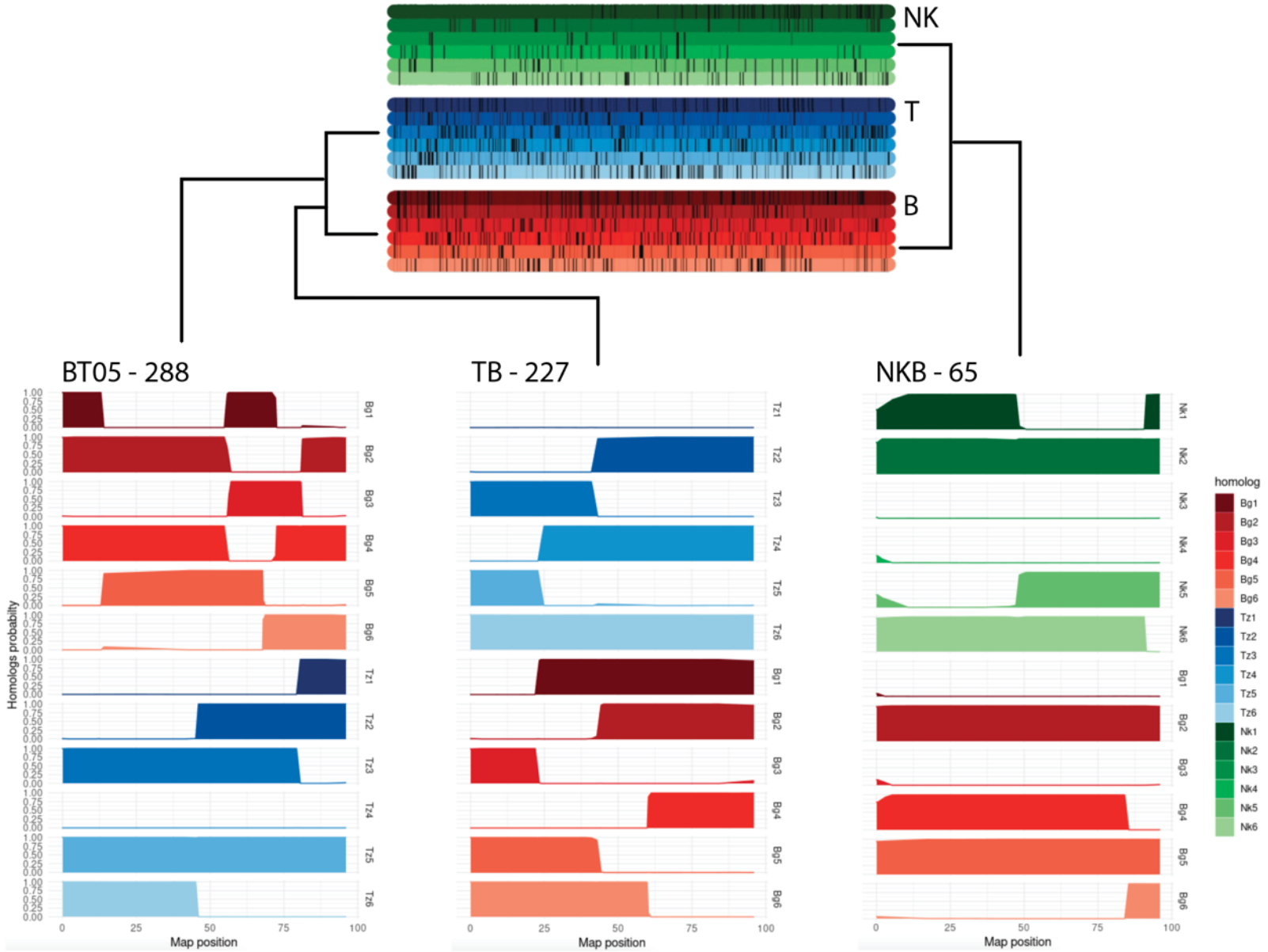




Multi-parental map BT-TB-NKB – Chromosome 15

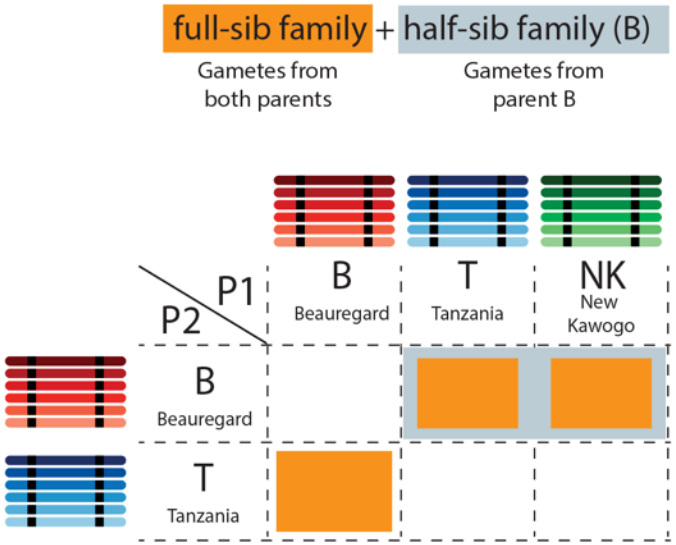


Haplotyping in BT-TB-NKB – Chromosome 15

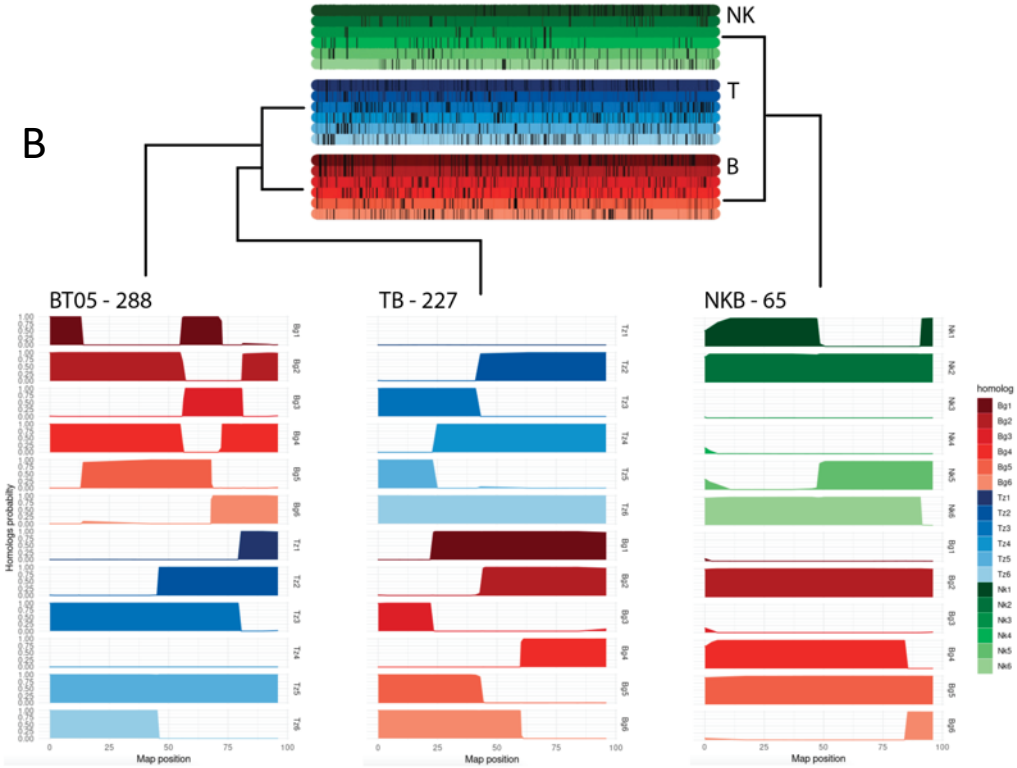


Haplotyping in BT-TB-NKB – Chromosome 15

A



B



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).
- 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).
- 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* **9**, 3297–3314 (2019).