Genomic Challenges in Polyploid Crops: An Overview of Progress so Far

Marcelo Mollinari

mmollin@ncsu.edu

PAG – Yam Genomics Workshop – Jan 2024 – San Diego, CA

Polyploids

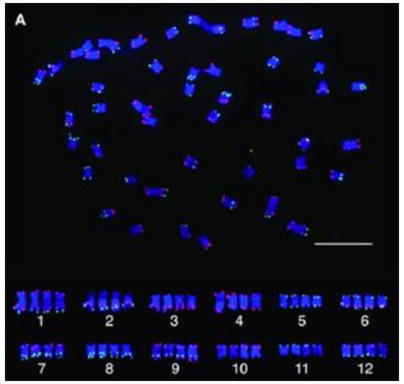
Polyploids are organisms that have multiple copies of the complete set of chromosomes. This multiplicity imposes numerous challenges to genomic analysis and subsequent applications in practical breeding.

Polyploid species

Autohexaploid

Multiple sets have the same origin

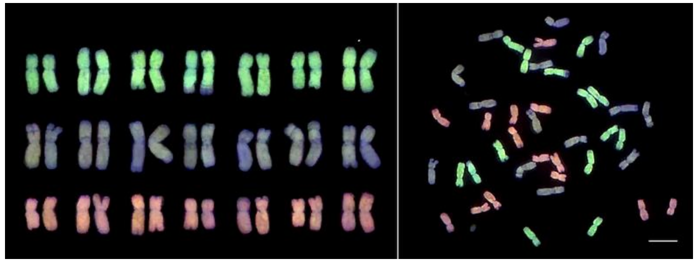
Autotetraploid potato



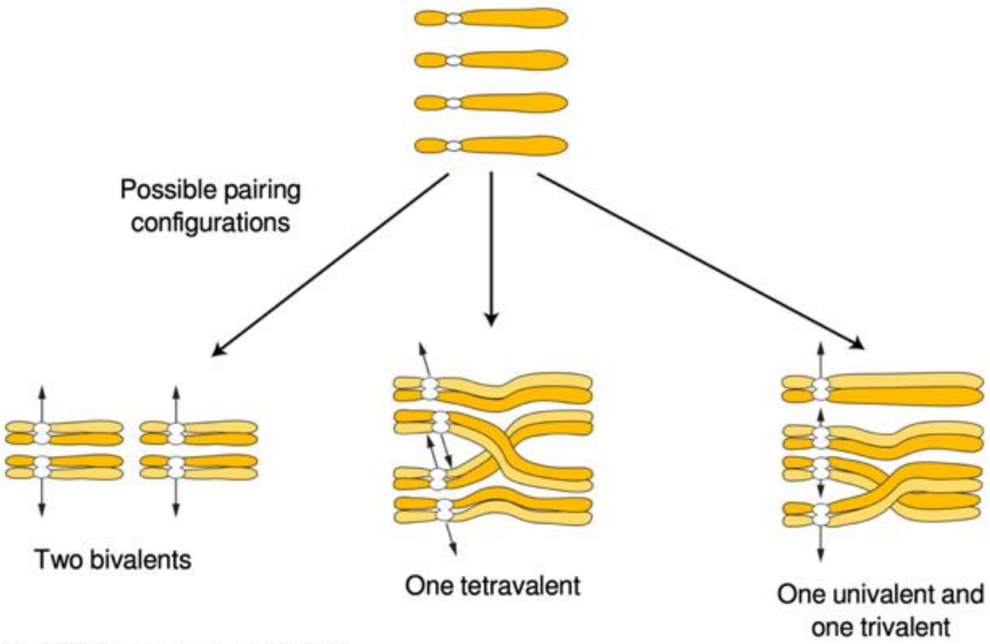
Braz et al., 2018

Allohexaploid

Multiple sets have different origins



Meiotic pairing in autotetraploids



Griffiths et al. (2004)

Segregation in polyploids*

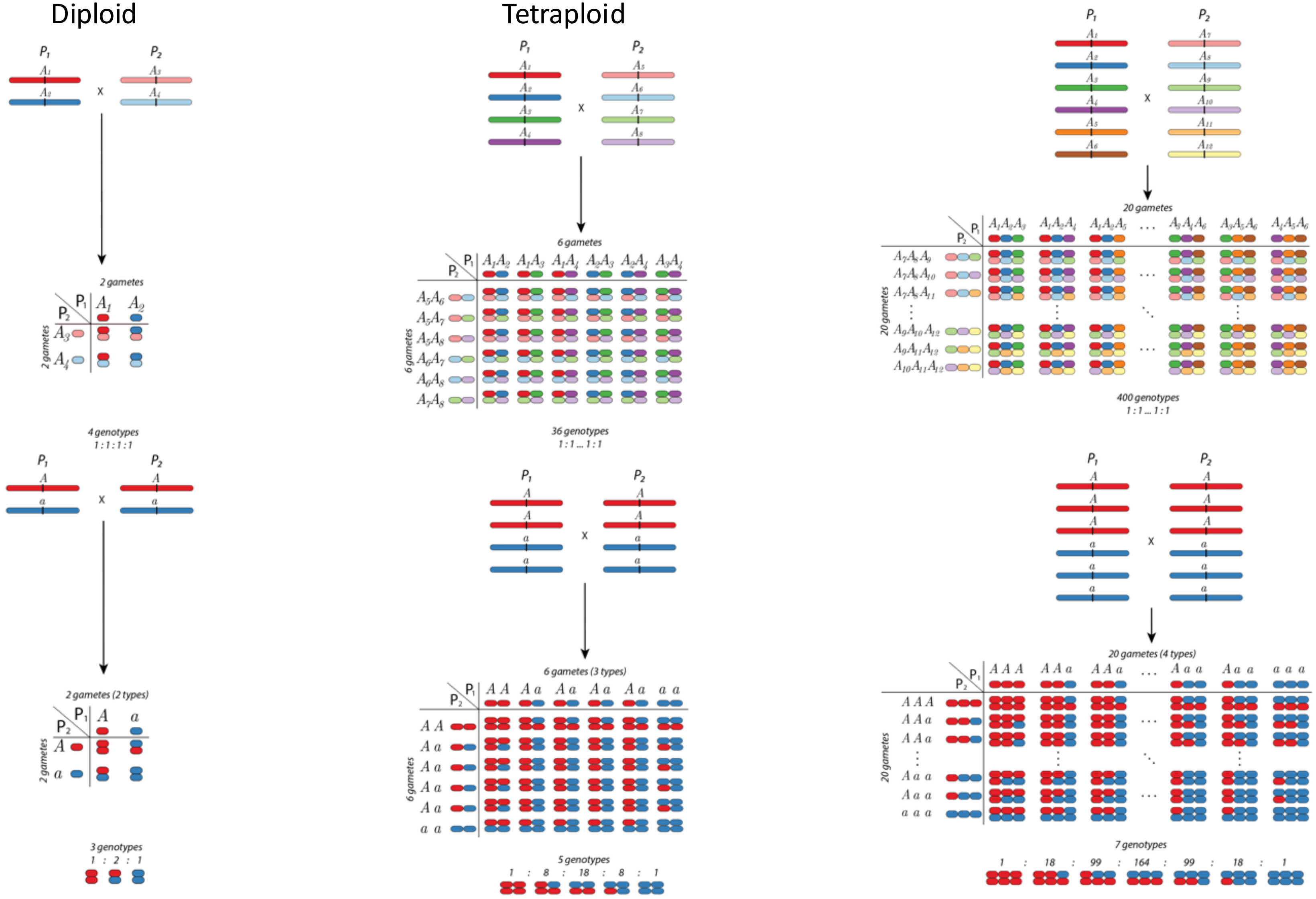
Diploid

Tetraploid

Hexaploid

Multiallelic

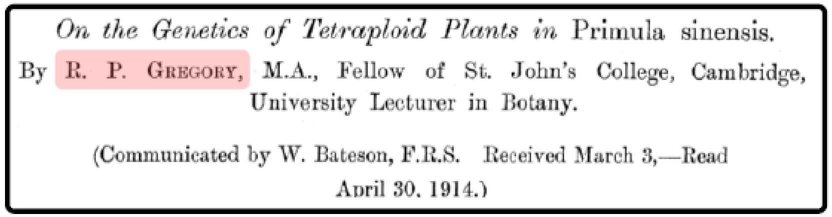
Biallelic



*random pairing and no double reduction

Linkage analysis

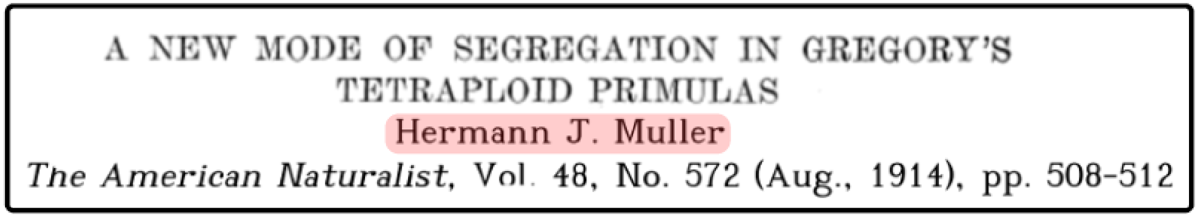
Theoretical foundations: 1914 - 1941



On the Genetics of Tetraploid Plants in Primula sinensis.

By R. P. Gregory, M.A., Fellow of St. John's College, Cambridge, University Lecturer in Botany.

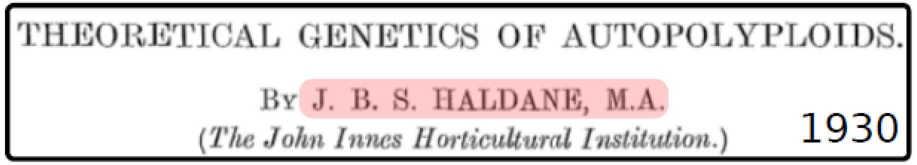
(Communicated by W. Bateson, F.R.S. Received March 3,—Read April 30, 1914.)



A NEW MODE OF SEGREGATION IN GREGORY'S TETRAPLOID PRIMULAS

Hermann J. Muller

The American Naturalist, Vol. 48, No. 572 (Aug., 1914), pp. 508-512

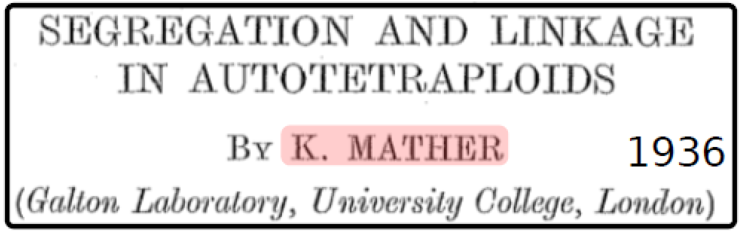


THEORETICAL GENETICS OF AUTOPOLYPLOIDS.

By J. B. S. HALDANE, M.A.

(*The John Innes Horticultural Institution.*)

1930

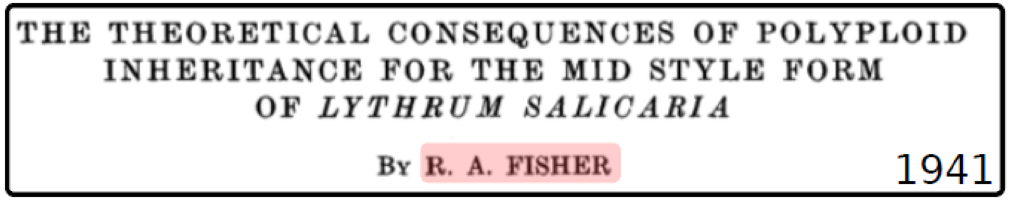


SEGREGATION AND LINKAGE IN AUTOTETRAPLOIDS

By K. MATHER

(*Galton Laboratory, University College, London*)

1936



THE THEORETICAL CONSEQUENCES OF POLYPLOID INHERITANCE FOR THE MID STYLE FORM OF *LYTHRUM SALICARIA*

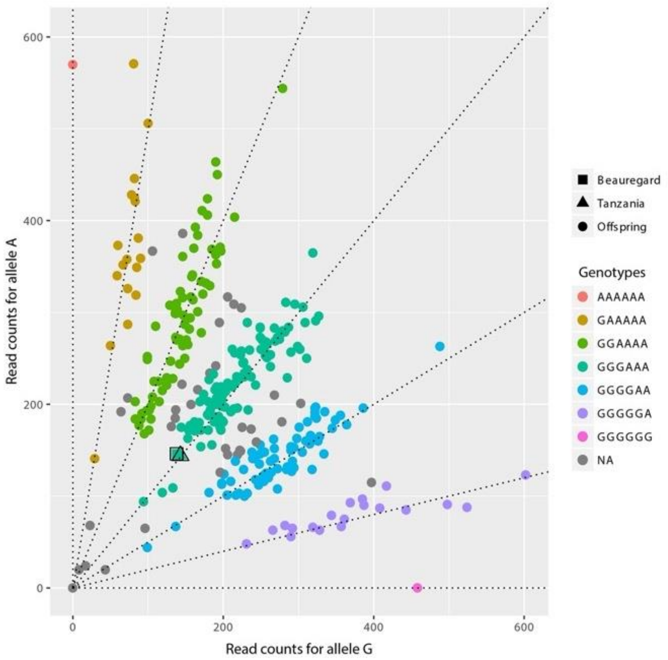
By R. A. FISHER

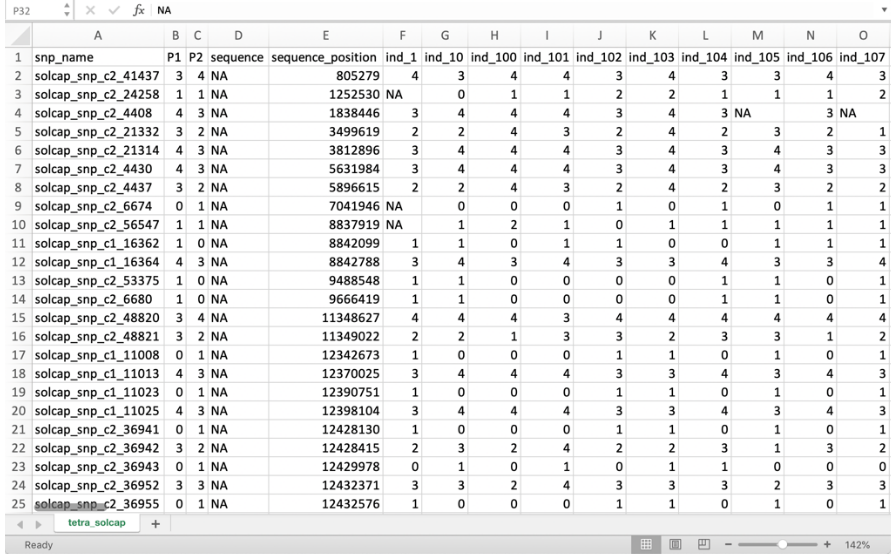
1941

Variant discovery and genotyping calling

- Process of finding variations in raw DNA sequences and estimating their dosages in all population individuals.

Raw DNA sequences or Genotyping Arrays

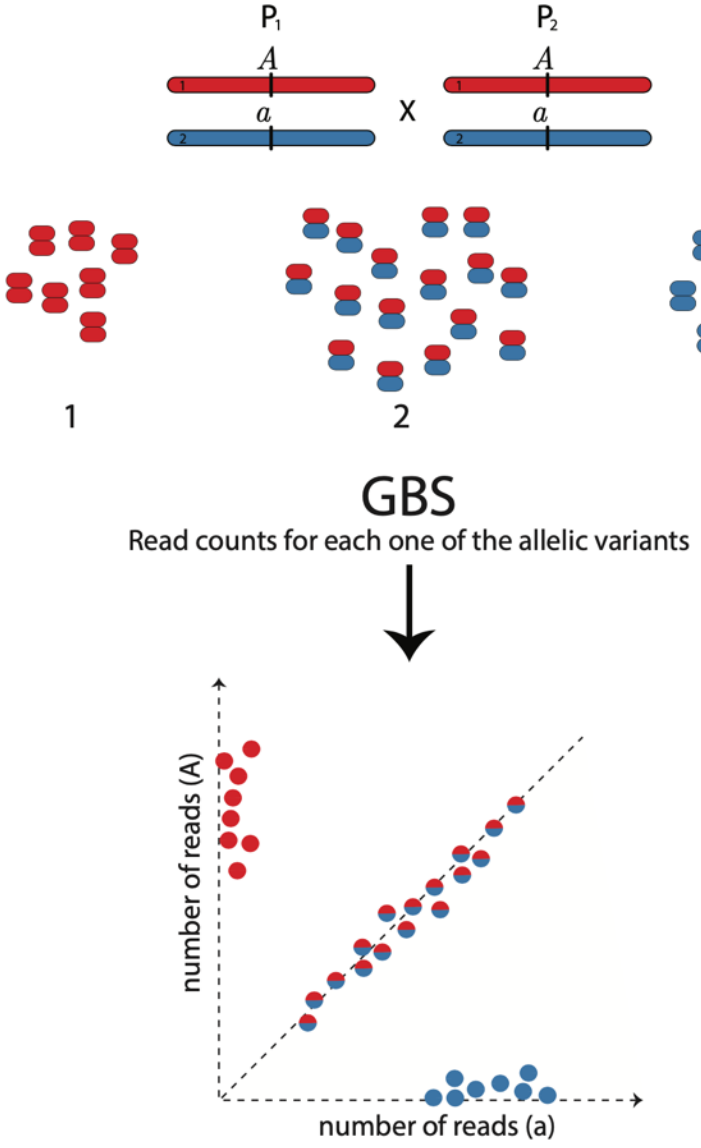


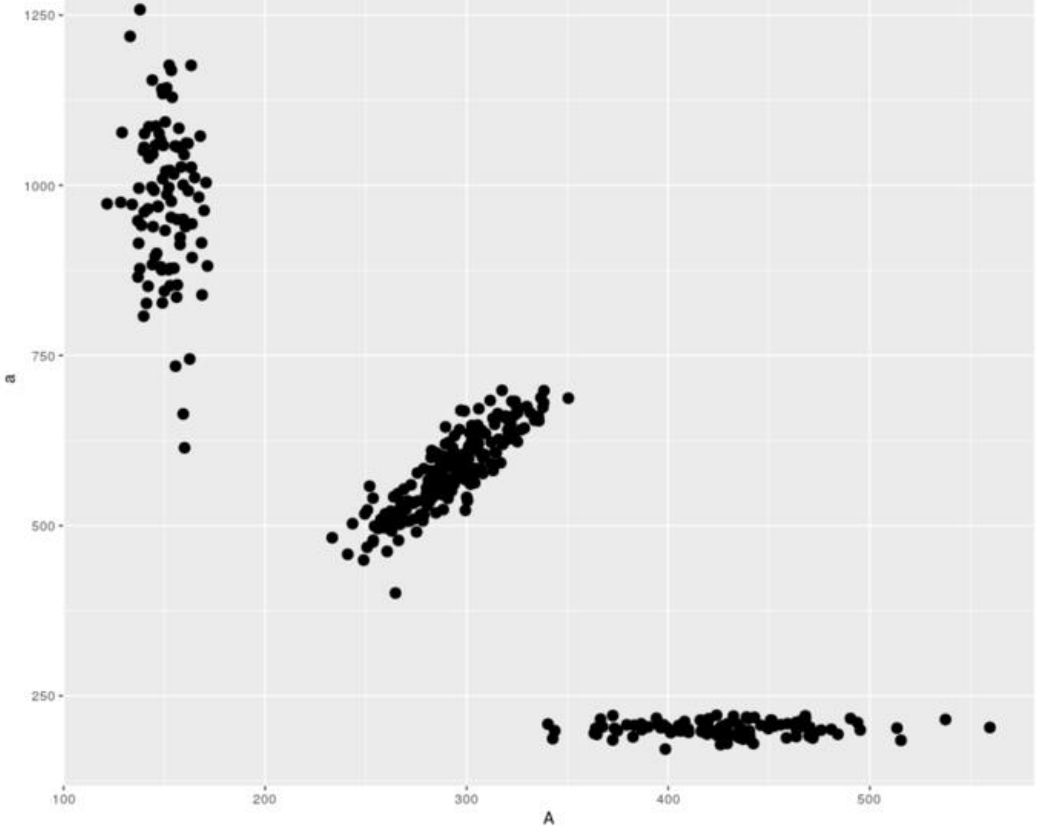


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	snp_name	P1	P2	sequence	sequence_position	ind_1	ind_10	ind_100	ind_101	ind_102	ind_103	ind_104	ind_105	ind_106	ind_107
2	solcap_snp_c2_41437	3	4	NA	805279	4	3	4	4	3	4	3	3	4	3
3	solcap_snp_c2_24258	1	1	NA	1252530	NA	0	1	1	2	2	1	1	1	2
4	solcap_snp_c2_4408	4	3	NA	1838446	3	4	4	4	3	4	3	NA	3	NA
5	solcap_snp_c2_21332	3	2	NA	3499619	2	2	4	3	2	4	2	3	2	1
6	solcap_snp_c2_21314	4	3	NA	3812896	3	4	4	4	3	4	3	4	3	3
7	solcap_snp_c2_4430	4	3	NA	5631984	3	4	4	4	3	4	3	4	3	3
8	solcap_snp_c2_4437	3	2	NA	5896615	2	2	4	3	2	4	2	3	2	2
9	solcap_snp_c2_6674	0	1	NA	7041946	NA	0	0	0	1	0	1	0	1	1
10	solcap_snp_c2_56547	1	1	NA	8837919	NA	1	2	1	0	1	1	1	1	1
11	solcap_snp_c1_16362	1	0	NA	8842099	1	1	0	1	1	0	0	1	1	1
12	solcap_snp_c1_16364	4	3	NA	8842788	3	4	3	4	3	3	4	3	3	4
13	solcap_snp_c2_53375	1	0	NA	9488548	1	1	0	0	0	0	1	1	0	1
14	solcap_snp_c2_6680	1	0	NA	9666419	1	1	0	0	0	0	1	1	0	1
15	solcap_snp_c2_48820	3	4	NA	11348627	4	4	4	3	4	4	4	4	4	4
16	solcap_snp_c2_48821	3	2	NA	11349022	2	2	1	3	3	2	3	3	1	2
17	solcap_snp_c1_11008	0	1	NA	12342673	1	0	0	0	1	1	0	1	0	1
18	solcap_snp_c1_11013	4	3	NA	12370025	3	4	4	4	3	3	4	3	4	3
19	solcap_snp_c1_11023	0	1	NA	12390751	1	0	0	0	1	1	0	1	0	1
20	solcap_snp_c1_11025	4	3	NA	12398104	3	4	4	4	3	3	4	3	4	3
21	solcap_snp_c2_36941	0	1	NA	12428130	1	0	0	0	1	1	0	1	0	1
22	solcap_snp_c2_36942	3	2	NA	12428415	2	3	2	4	2	2	3	1	3	2
23	solcap_snp_c2_36943	0	1	NA	12429978	0	1	0	1	0	1	1	0	0	0
24	solcap_snp_c2_36952	3	3	NA	12432371	3	3	2	4	3	3	3	2	3	3
25	solcap_snp_c2_36955	0	1	NA	12432576	1	0	0	0	1	1	0	1	0	1

- Discovery: Tassel, GATK
- Calling: fitPoly, SuperMASSA, polyRAD, updog, VCF2SM

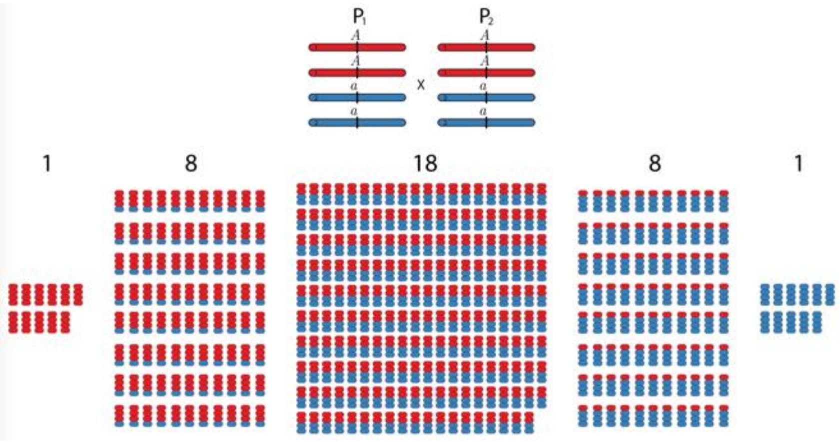
Genotype calling in diploids

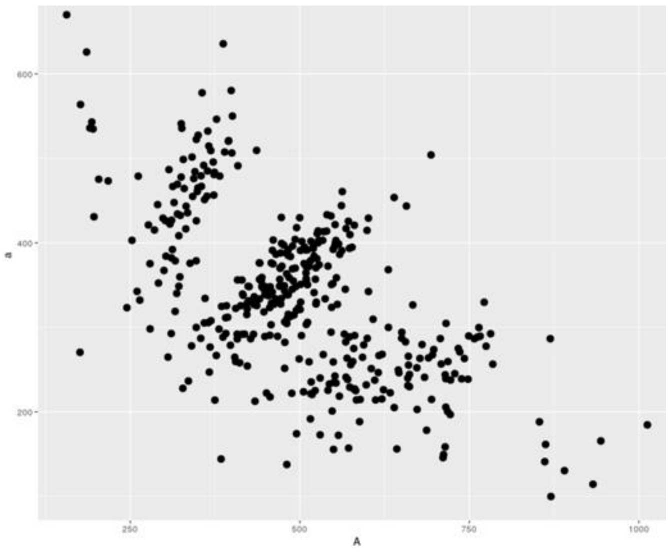




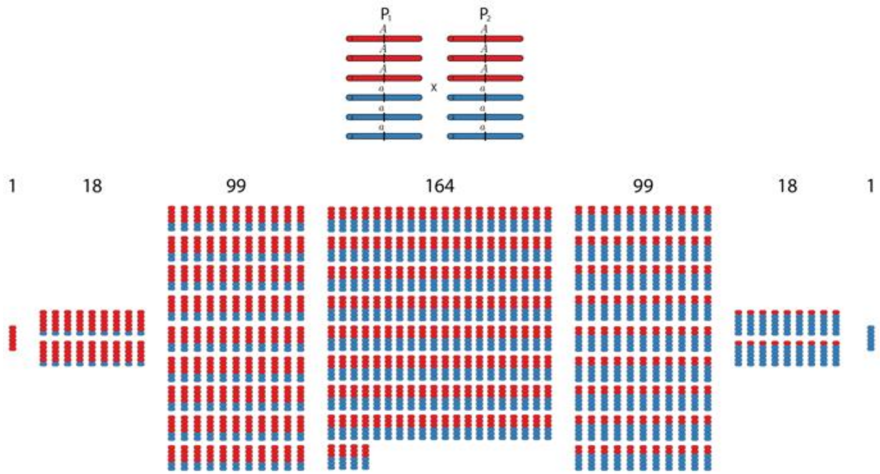
Genotype calling in polyploids

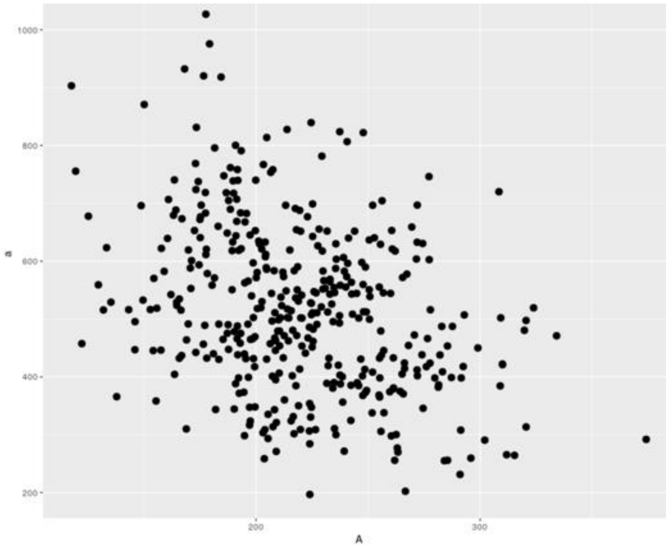
Tetraploid





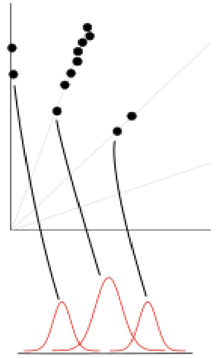
Hexaploid

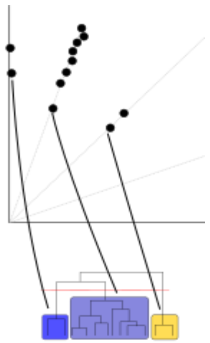


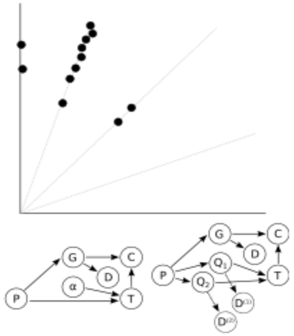


Genotype calling in polyploids

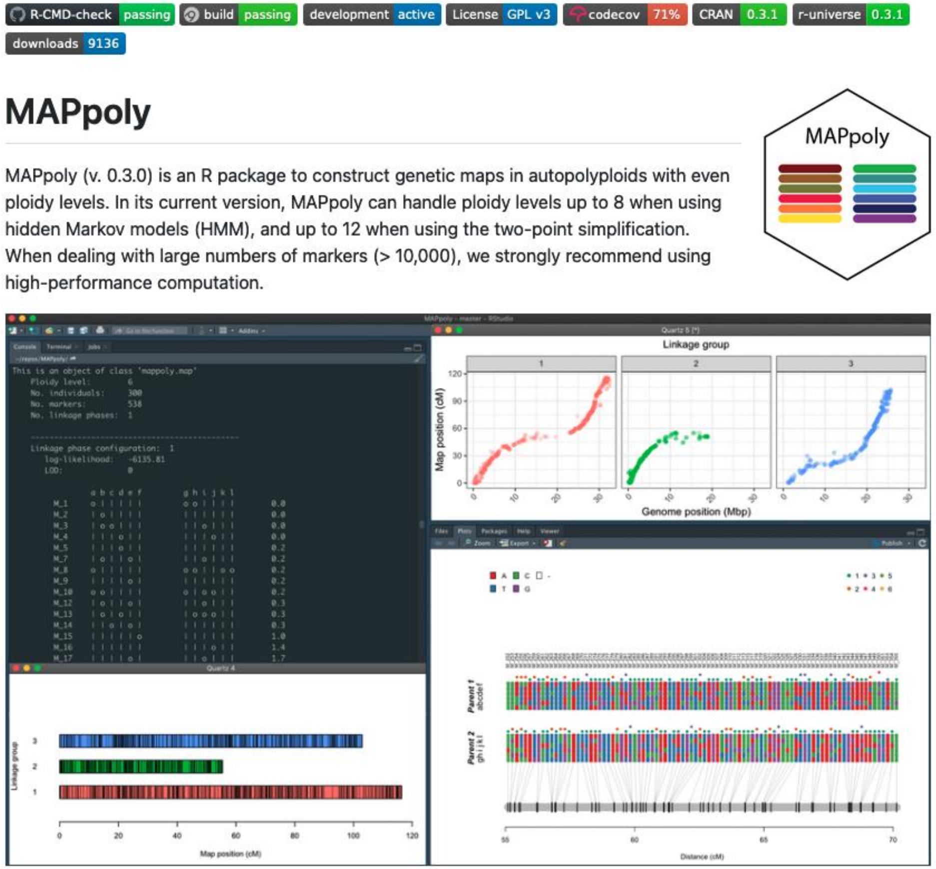
- fitTetra (tetraploids – array data):
  https://www.wur.nl/en/show/Software-fitTetra.htm
- ClusterCall (tetraploids):
  https://potatobreeding.cals.wisc.edu/software/
- SuperMASSA (any ploidy level):
  https://bitbucket.org/orserang/supermassa
- updog (any ploidy level, allows preferential pairing):
  https://github.com/dcgerard/updog
- polyRAD (any ploidy level, reads VCF, BAM, etc):
  https://github.com/lvclark/polyRAD







Genetic mapping and haplotype inference



MAPpoly

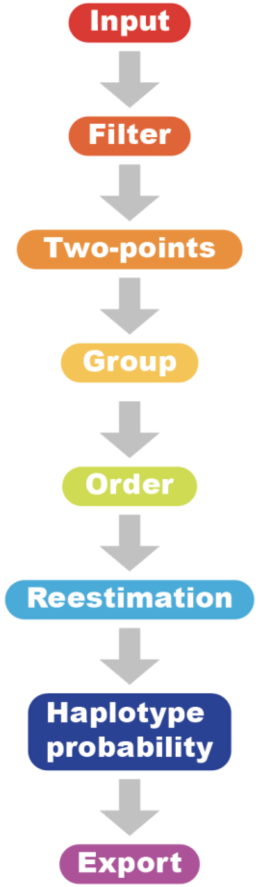
MAPpoly (v. 0.3.0) is an R package to construct genetic maps in autopolyploids with even ploidy levels. In its current version, MAPpoly can handle ploidy levels up to 8 when using hidden Markov models (HMM), and up to 12 when using the two-point simplification. When dealing with large numbers of markers (> 10,000), we strongly recommend using high-performance computation.

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

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

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

https://github.com/mmollina/MAPpoly


Input

Filter

Two-points

Group

Order

Reestimation

Haplotype probability

Export

- MAPpoly can easily be used in conjunction with several genotype calling, mapping and QTL analysis software

Mollinari and Garcia (2019) – G3
doi:10.1534/g3.119.400378

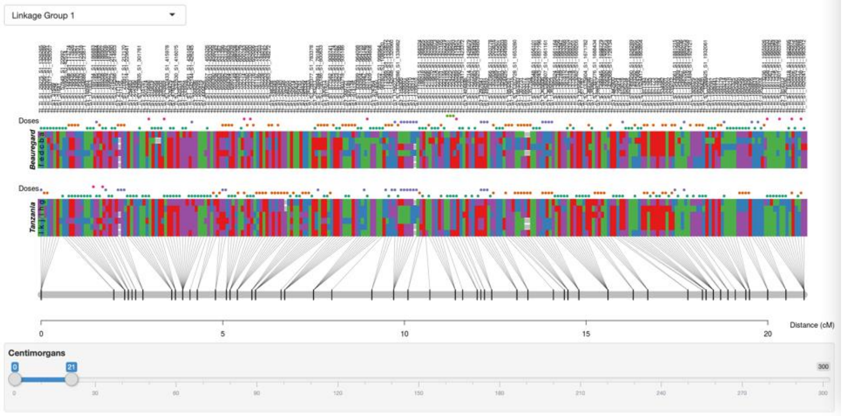
Genetic Mapping - Sweetpotato

Beauregard x **Tanzania**

Mollinari *et al.* (2020) – G3

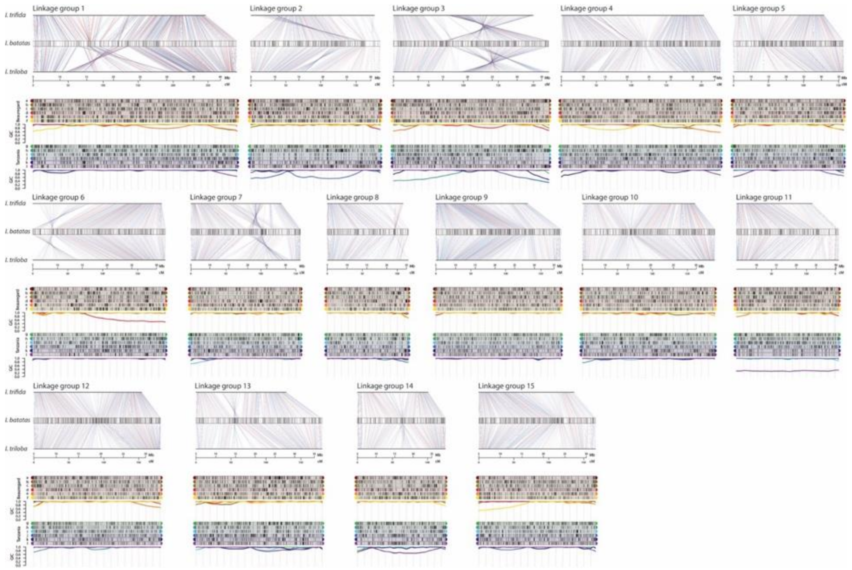
[doi:10.1534/g3.119.400620](https://doi.org/10.1534/g3.119.400620)

Chromosome 1

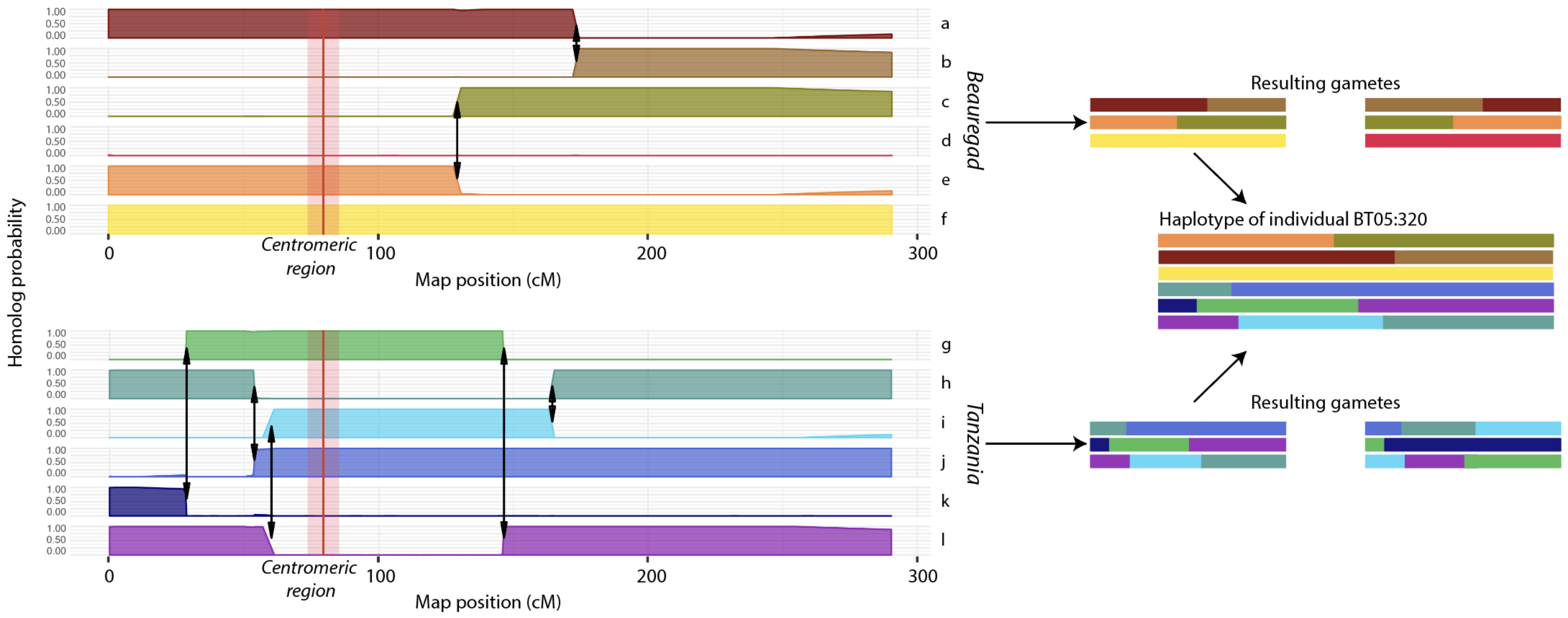


https://gt4sp-genetic-map.shinyapps.io/bt_map/

Final map – 15 chromosomes



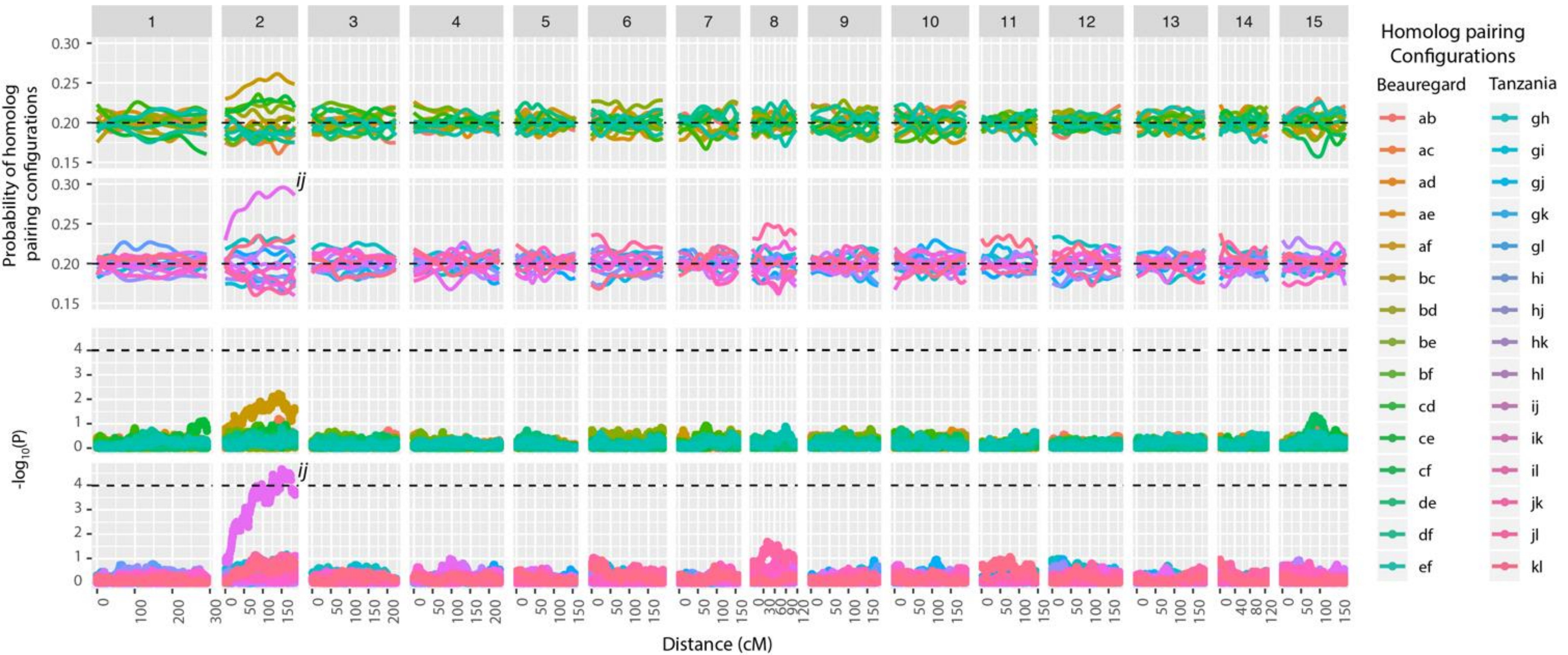
Haplotype reconstruction – BT offspring



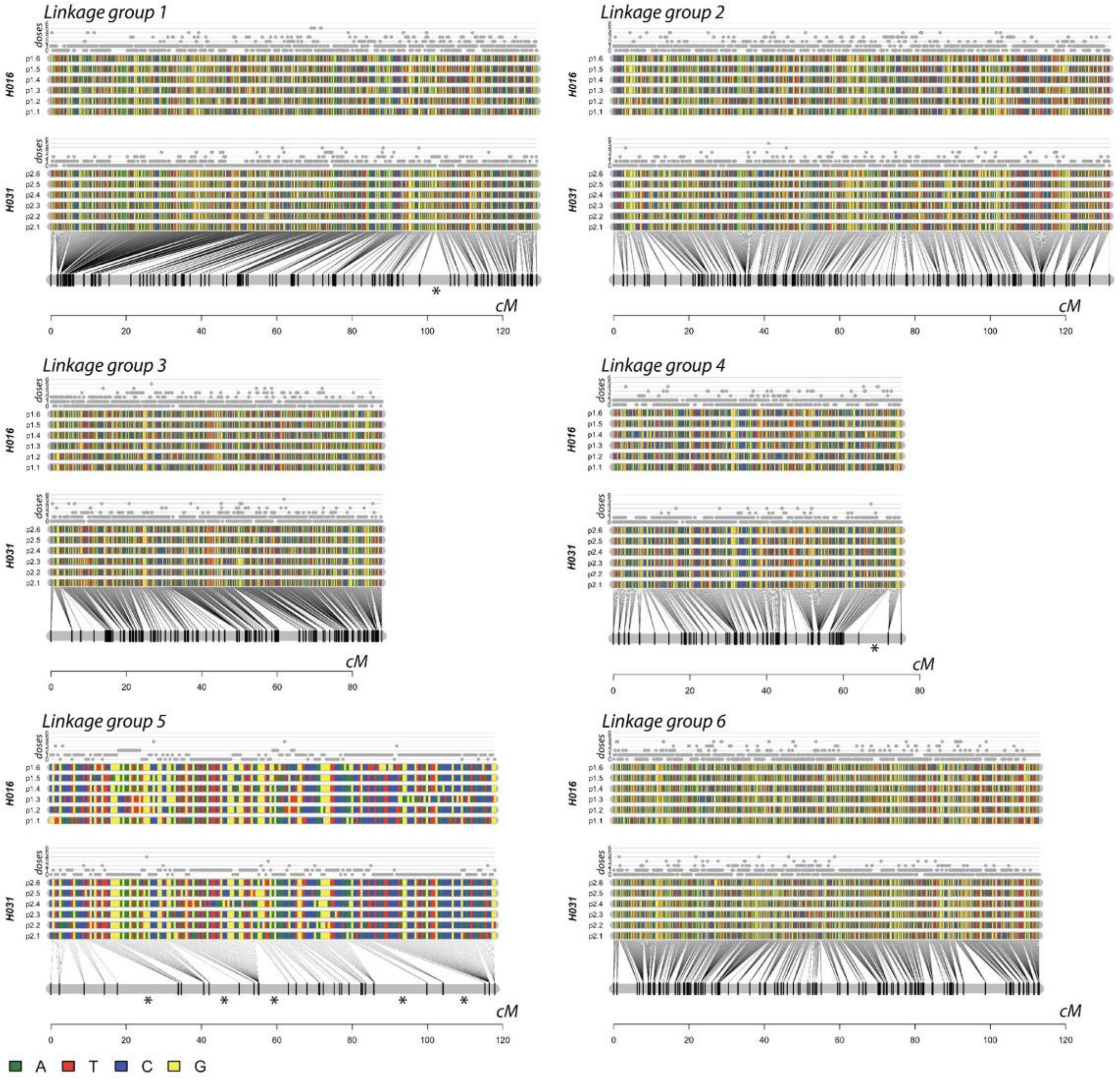
- MAPpoly reconstructs the genome of individuals in the offspring by tracing back their haplotypes to their founders.
- It also estimates degrees of preferential pairing and frequency of multivalent configurations.

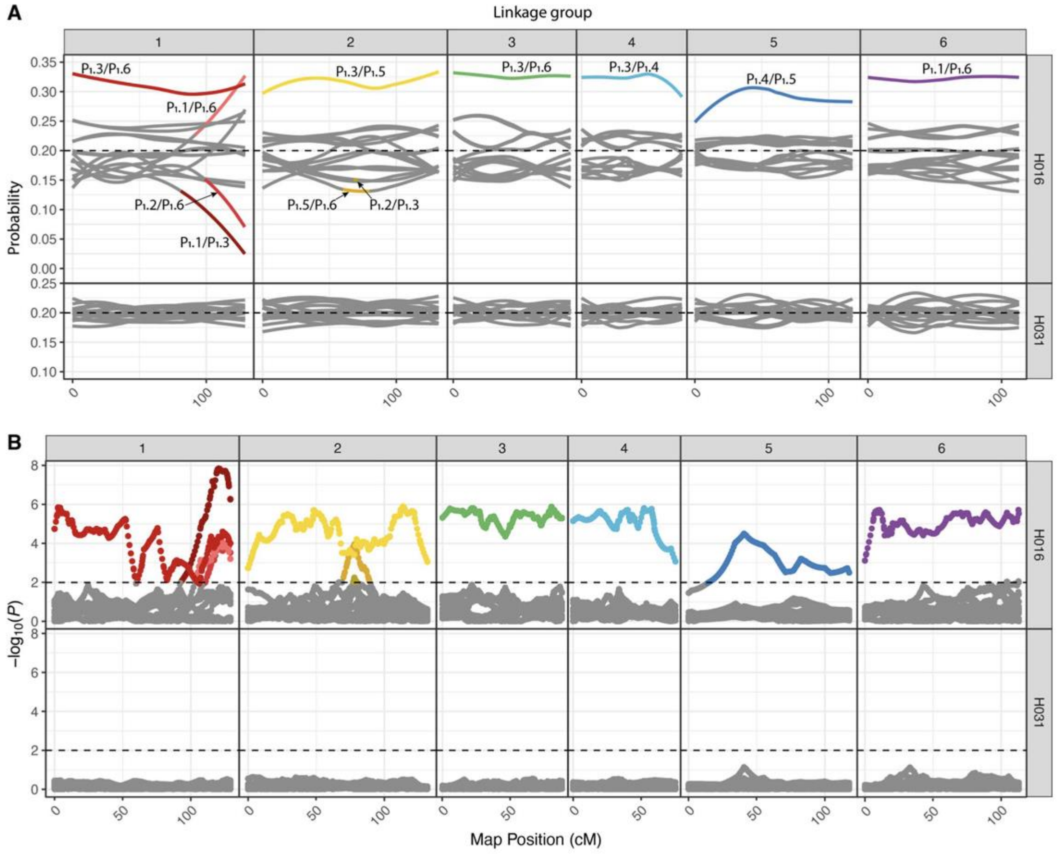
https://gt4sp-genetic-map.shinyapps.io/offspring_haplotype_BT_population/

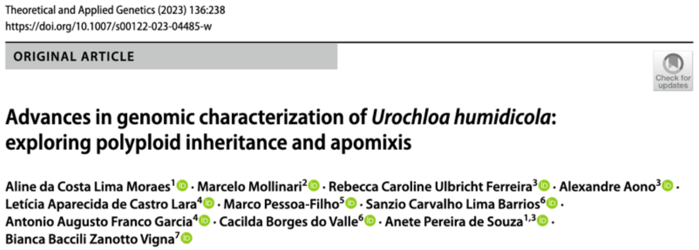
Preferential pairing profiles



Genomic characterization of *Urochloa humidicola*





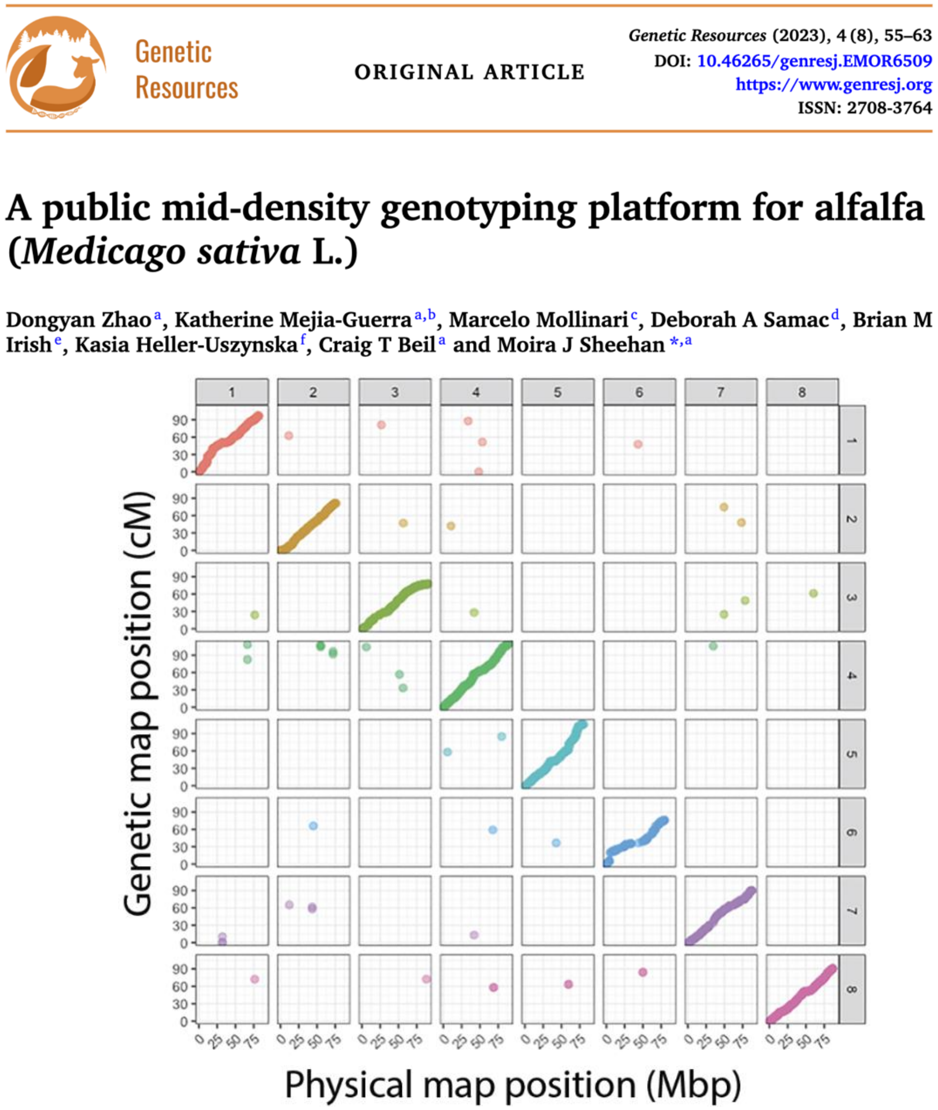

Theoretical and Applied Genetics (2023) 136:238
https://doi.org/10.1007/s00122-023-04485-w

ORIGINAL ARTICLE

Advances in genomic characterization of *Urochloa humidicola*: exploring polyploid inheritance and apomixis

Aline da Costa Lima Moraes[1] · Marcelo Mollinari[2] · Rebecca Caroline Ulbricht Ferreira[3] · Alexandre Aono[3] ·
Letícia Aparecida de Castro Lara[4] · Marco Pessoa-Filho[5] · Sanzio Carvalho Lima Barrios[6] ·
Antonio Augusto Franco Garcia[4] · Cacilda Borges do Valle[6] · Anete Pereira de Souza[1,3] ·
Bianca Baccili Zanotto Vigna[7]

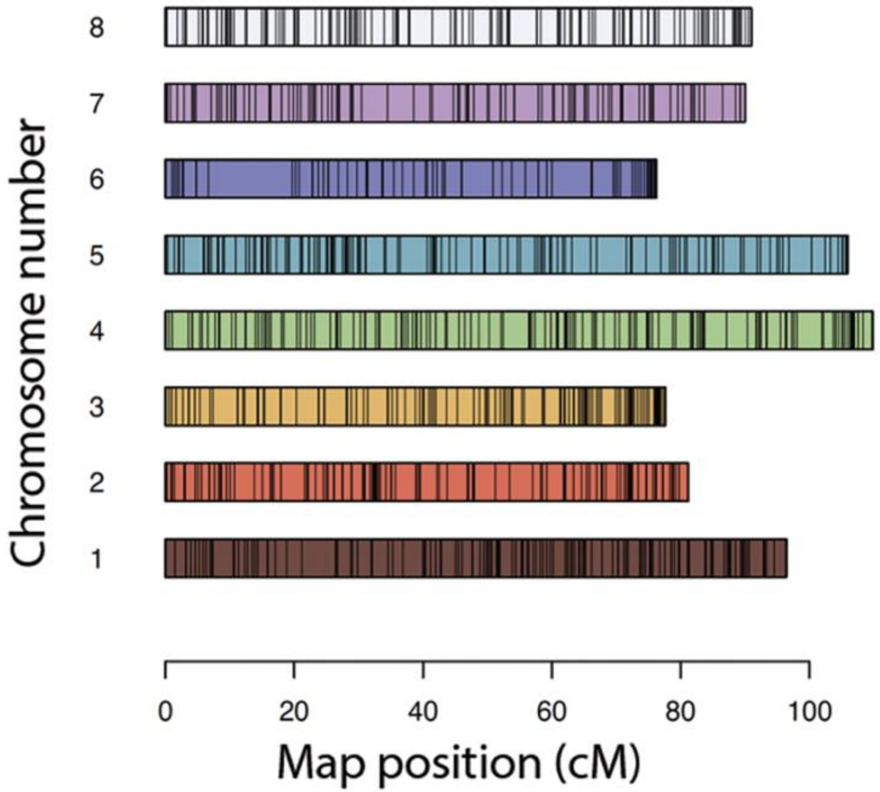
Public mid-density genotyping platform for alfalfa(*Medicago sativa* L.)



Genetic Resources
ORIGINAL ARTICLE
Genetic Resources (2023), 4 (8), 55–63
DOI: 10.46265/genresj.EMOR6509
https://www.genresj.org
ISSN: 2708-3764

A public mid-density genotyping platform for alfalfa (*Medicago sativa* L.)

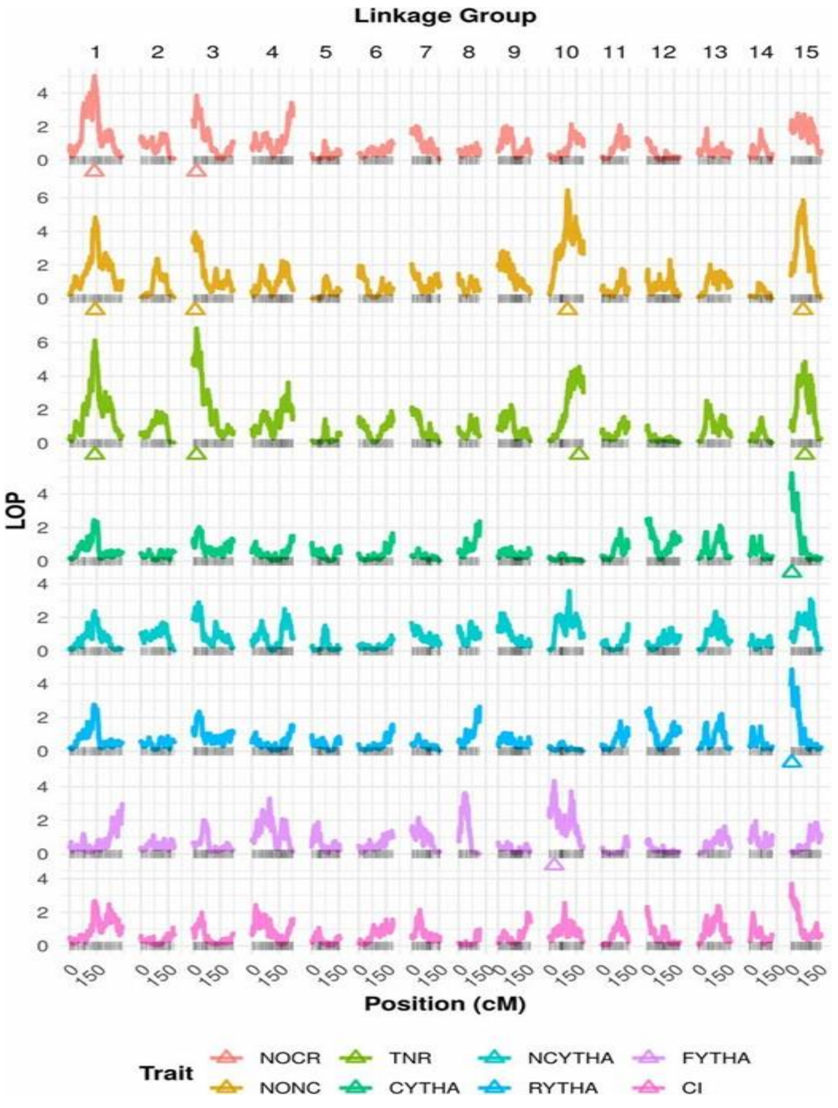
Dongyan Zhao[a], Katherine Mejia-Guerra[a,b], Marcelo Mollinari[c], Deborah A Samac[d], Brian M Irish[e], Kasia Heller-Uszynska[f], Craig T Beil[a] and Moira J Sheehan[*,a]



MAPpoly2 and alfalfa map construction

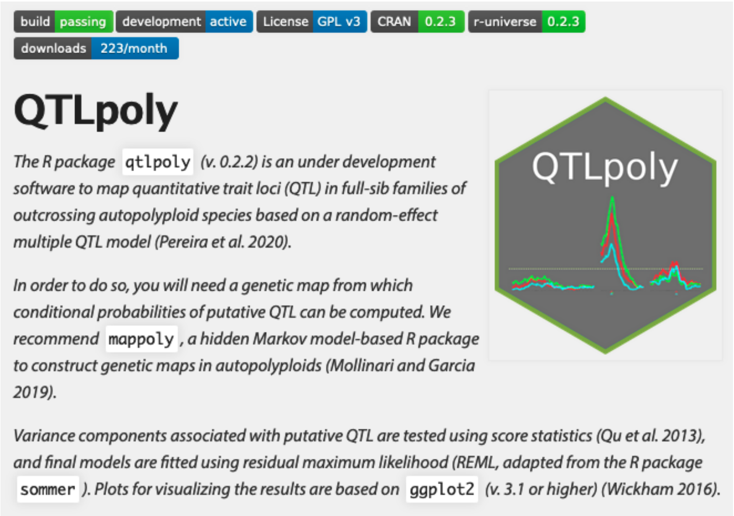
https://rpubs.com/mmollin/tutorial_mappoly2

QTL mapping



Pereira *et al.* (2020) - Genetics

[doi.org:10.1534/genetics.120.303080](https://doi.org/10.1534/genetics.120.303080)



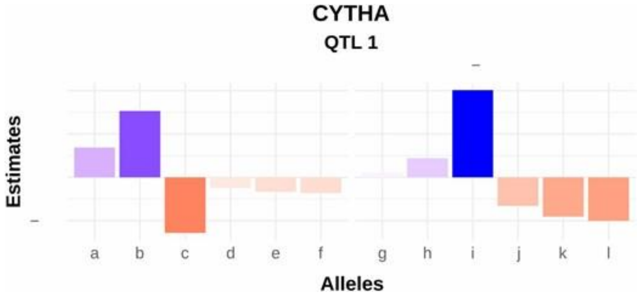
QTLpoly

The R package `qtlpoly` *(v. 0.2.2) is an under development software to map quantitative trait loci (QTL) in full-sib families of outcrossing autopolyploid species based on a random-effect multiple QTL model (Pereira et al. 2020).*

In order to do so, you will need a genetic map from which conditional probabilities of putative QTL can be computed. We recommend `mappoly`*, a hidden Markov model-based R package to construct genetic maps in autopolyploids (Mollinari and Garcia 2019).*

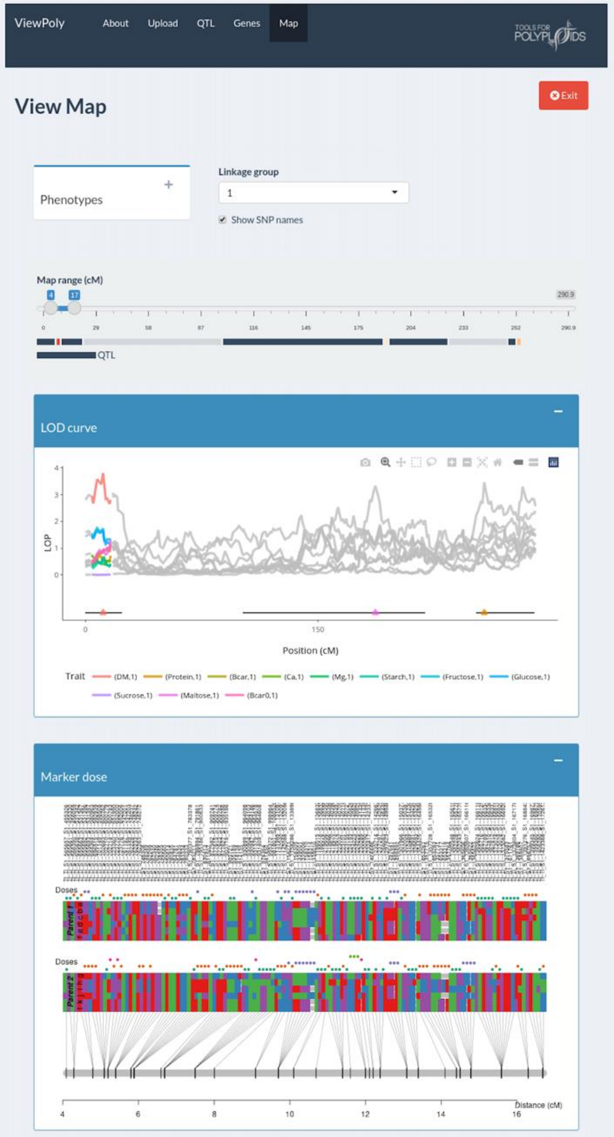
Variance components associated with putative QTL are tested using score statistics (Qu et al. 2013), and final models are fitted using residual maximum likelihood (REML, adapted from the R package `sommer`*). Plots for visualizing the results are based on* `ggplot2` *(v. 3.1 or higher) (Wickham 2016).*

	NOCR	NONC	TNR	CYTHA	NCYTHA	RYTHA	FYTHA	CI
σ_g^2	0.386	0.277	1.117	27.611	0.313	31.568	23.677	5.88×10^{-3}
σ_{ge}^2	0.272	0.213	0.571	17.028	0.310	18.538	34.451	2.30×10^{-3}
σ^2	0.686	0.559	1.462	32.271	1.082	35.098	50.836	5.23×10^{-3}
H^2 (%)	80.07	78.31	84.59	84.08	70.39	84.73	71.35	88.42

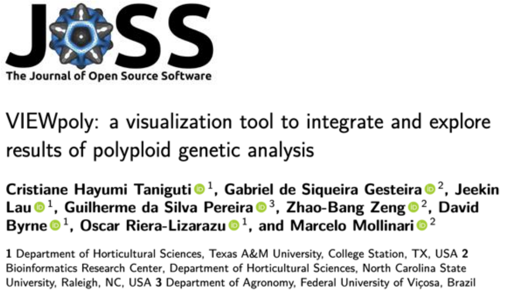


Additive allele effects from the decomposed BLUPs for the QTL 1 of commercial root yield

Viewpoly



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



The Journal of Open Source Software

VIEWpoly: a visualization tool to integrate and explore results of polyploid genetic analysis

Cristiane Hayumi Taniguti[1], **Gabriel de Siqueira Gesteira**[2], **Jeekin Lau**[1], **Guilherme da Silva Pereira**[3], **Zhao-Bang Zeng**[2], **David Byrne**[1], **Oscar Riera-Lizarazu**[1], **and Marcelo Mollinari**[2]

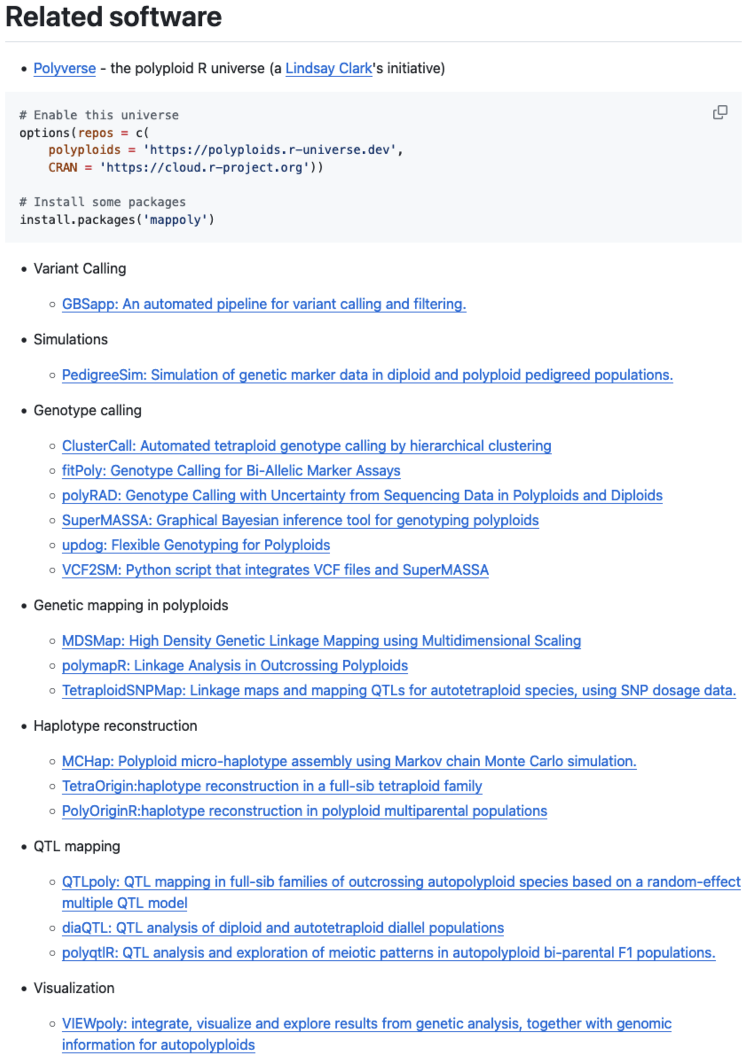
1 Department of Horticultural Sciences, Texas A&M University, College Station, TX, USA **2** Bioinformatics Research Center, Department of Horticultural Sciences, North Carolina State University, Raleigh, NC, USA **3** Department of Agronomy, Federal University of Viçosa, Brazil

Taniguti *et al.* (2022) - JOSS

doi:10.21105/joss.04242

Software

Genotype calling, mapping, QTL analysis and visualization



Related software

- Polyverse - the polyploid R universe (a Lindsay Clark's initiative)

```
# Enable this universe
options(repos = c(
    polyploids = 'https://polyploids.r-universe.dev',
    CRAN = 'https://cloud.r-project.org'))

# Install some packages
install.packages('mappoly')
```

- Variant Calling
  - GBSapp: An automated pipeline for variant calling and filtering.
- Simulations
  - PedigreeSim: Simulation of genetic marker data in diploid and polyploid pedigreed populations.
- Genotype calling
  - ClusterCall: Automated tetraploid genotype calling by hierarchical clustering
  - fitPoly: Genotype Calling for Bi-Allelic Marker Assays
  - polyRAD: Genotype Calling with Uncertainty from Sequencing Data in Polyploids and Diploids
  - SuperMASSA: Graphical Bayesian inference tool for genotyping polyploids
  - updog: Flexible Genotyping for Polyploids
  - VCF2SM: Python script that integrates VCF files and SuperMASSA
- Genetic mapping in polyploids
  - MDSMap: High Density Genetic Linkage Mapping using Multidimensional Scaling
  - polymapR: Linkage Analysis in Outcrossing Polyploids
  - TetraploidSNPMap: Linkage maps and mapping QTLs for autotetraploid species, using SNP dosage data.
- Haplotype reconstruction
  - MCHap: Polyploid micro-haplotype assembly using Markov chain Monte Carlo simulation.
  - TetraOrigin:haplotype reconstruction in a full-sib tetraploid family
  - PolyOriginR:haplotype reconstruction in polyploid multiparental populations
- QTL mapping
  - QTLpoly: QTL mapping in full-sib families of outcrossing autopolyploid species based on a random-effect multiple QTL model
  - diaQTL: QTL analysis of diploid and autotetraploid diallel populations
  - polyqtlR: QTL analysis and exploration of meiotic patterns in autopolyploid bi-parental F1 populations.
- Visualization
  - VIEWpoly: integrate, visualize and explore results from genetic analysis, together with genomic information for autopolyploids

https://github.com/mmollina/MAPpoly

Analysis in multiple inter-connected sweetpotato families

Beauregard

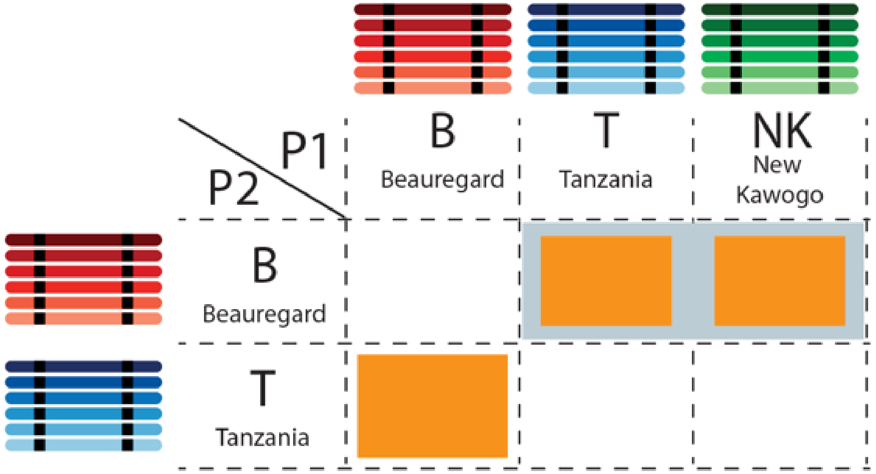
Tanzania

New Kawogo

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

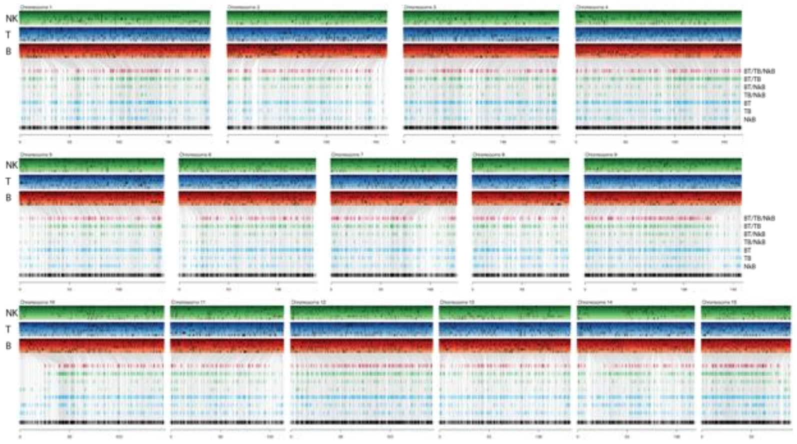
Gametes from both parents

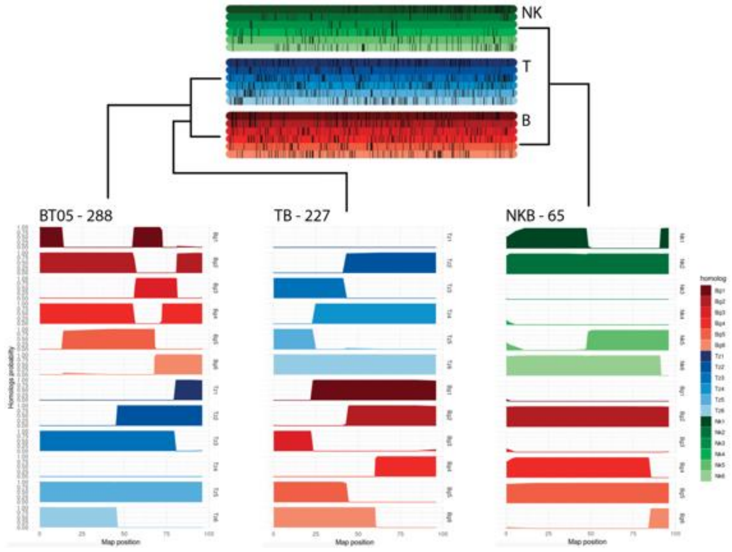
Gametes from parent B



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

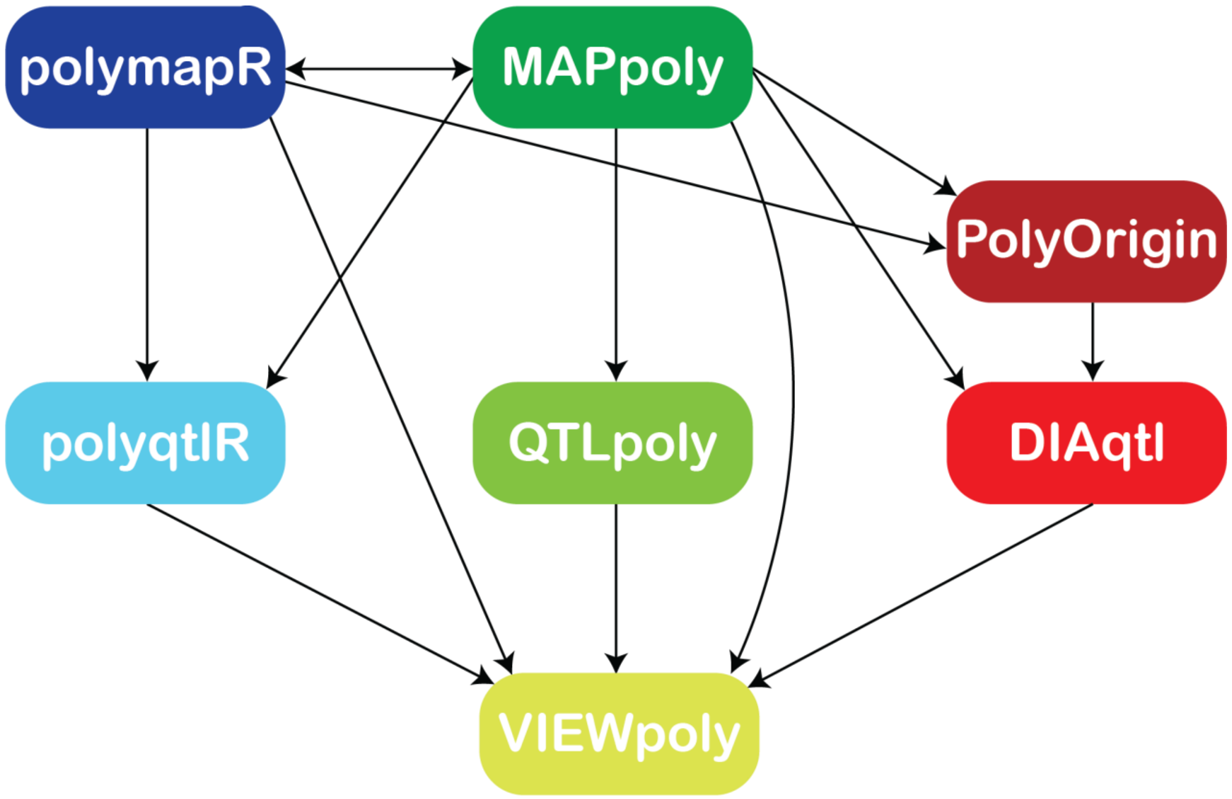
Complete genetic map for BT, TB, and NKB





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

Analytical environment for linkage and QTL analysis in polyploids



polymapR

MAPpoly

PolyOrigin

polyqtlR

QTLpoly

DIAqtl

VIEWpoly

TOOLS FOR
POLYPLOIDS

Tools for Polyploids
PLUS

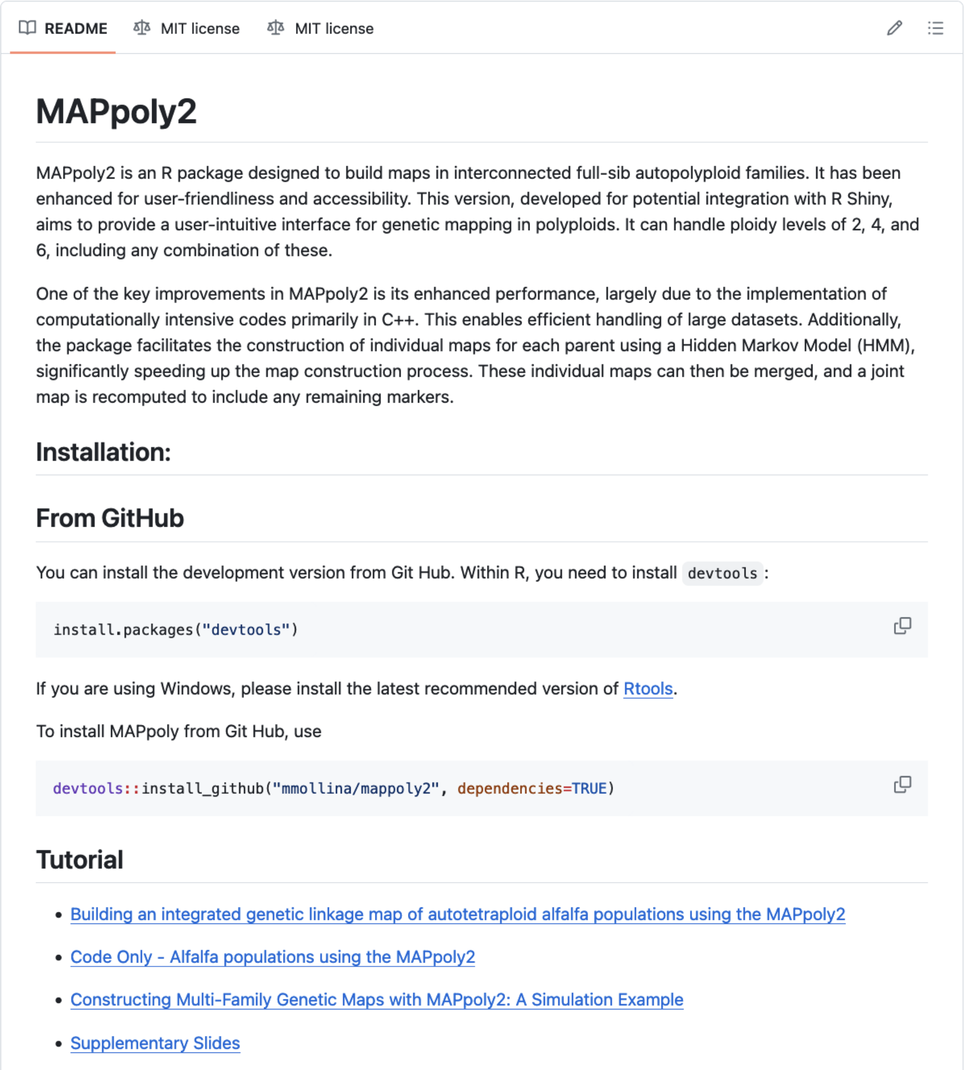
MAPpoly

QTLpoly

VIEWpoly

https://www.polyploids.org/

MAPpoly2: Constructing maps and integrating them in multiple F1 populations



MAPpoly2

MAPpoly2 is an R package designed to build maps in interconnected full-sib autopolyploid families. It has been enhanced for user-friendliness and accessibility. This version, developed for potential integration with R Shiny, aims to provide a user-intuitive interface for genetic mapping in polyploids. It can handle ploidy levels of 2, 4, and 6, including any combination of these.

One of the key improvements in MAPpoly2 is its enhanced performance, largely due to the implementation of computationally intensive codes primarily in C++. This enables efficient handling of large datasets. Additionally, the package facilitates the construction of individual maps for each parent using a Hidden Markov Model (HMM), significantly speeding up the map construction process. These individual maps can then be merged, and a joint map is recomputed to include any remaining markers.

Installation:

From GitHub

You can install the development version from Git Hub. Within R, you need to install `devtools`:

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

If you are using Windows, please install the latest recommended version of Rtools.

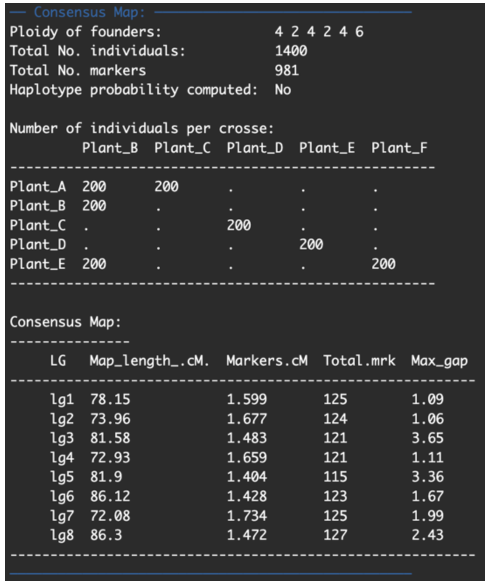
To install MAPpoly from Git Hub, use

```
devtools::install_github("mmollina/mappoly2", dependencies=TRUE)
```

Tutorial

- Building an integrated genetic linkage map of autotetraploid alfalfa populations using the MAPpoly2
- Code Only - Alfalfa populations using the MAPpoly2
- Constructing Multi-Family Genetic Maps with MAPpoly2: A Simulation Example
- Supplementary Slides

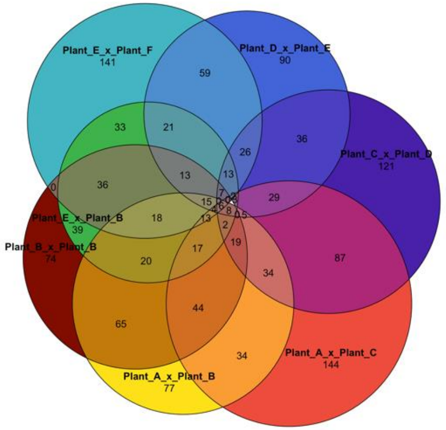
https://github.com/mmollina/mappoly2

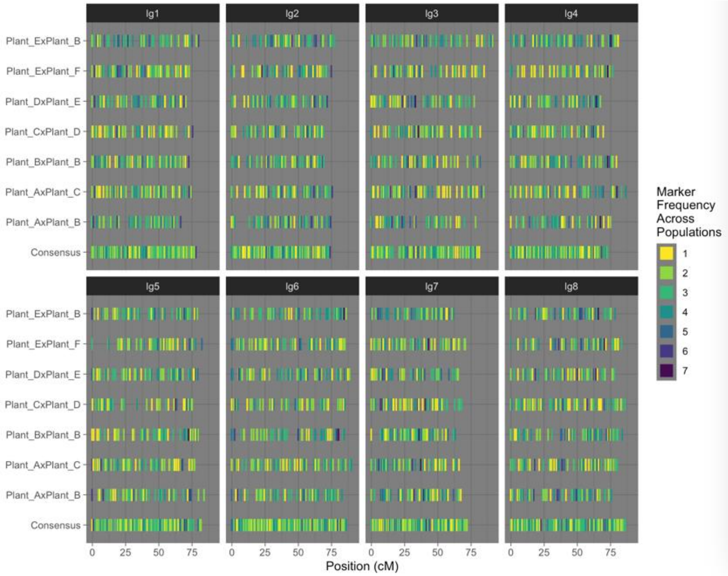


```
── Consensus Map: ──────────────────────────────
Ploidy of founders:             4 2 4 2 4 6
Total No. individuals:          1400
Total No. markers               981
Haplotype probability computed:  No

Number of individuals per crosse:
         Plant_B  Plant_C  Plant_D  Plant_E  Plant_F
-----------------------------------------------------
Plant_A  200      200      .        .        .
Plant_B  200      .        .        .        .
Plant_C  .        .        200      .        .
Plant_D  .        .        .        200      .
Plant_E  200      .        .        .        200
-----------------------------------------------------


Consensus Map:
---------------
     LG   Map_length_.cM.  Markers.cM  Total.mrk  Max_gap
----------------------------------------------------------
     lg1  78.15            1.599       125        1.09
     lg2  73.96            1.677       124        1.06
     lg3  81.58            1.483       121        3.65
     lg4  72.93            1.659       121        1.11
     lg5  81.9             1.404       115        3.36
     lg6  86.12            1.428       123        1.67
     lg7  72.08            1.734       125        1.99
     lg8  86.3             1.472       127        2.43
----------------------------------------------------------
──────────────────────────────────────────────────
```





Acknowledgements

- Gabriel S. Gesteira
- Cristiane Hayumi Taniguti
- Guilherme Da Silva Pereira
- Dongyan Zhao
- Shan Wu
- Aline da Costa Lima Moraes
- Bianca B. Z. Vigna
- Anete P Souza
- Antonio A. F. Garcia
- Zhangjun Fei
- Moira Sheehan
- David Byrne
- Oscar Riera-Lizarazu
- Craig Yencho
- Zhao-Bang Zeng