

Haplotyping in F₁ populations

Marcelo Mollinari

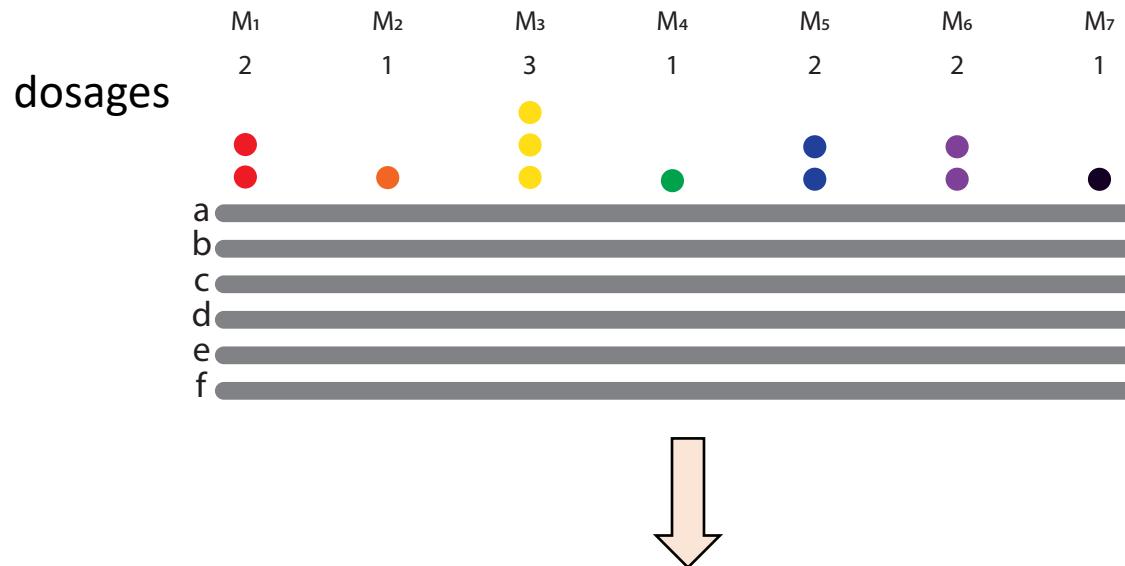
mmollin@ncsu.edu



BILL & MELINDA
GATES *foundation*

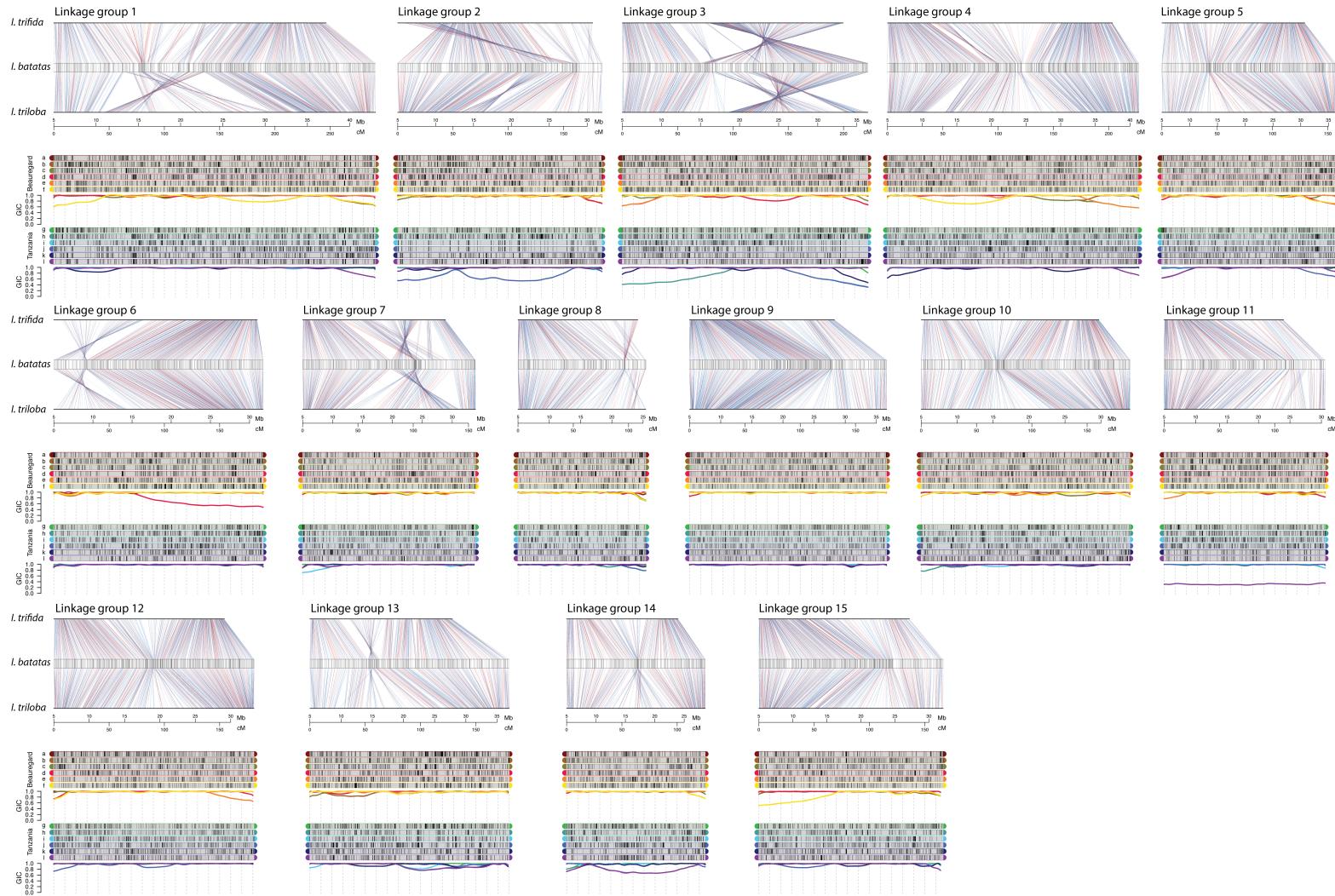
Haplotyping in polyploids

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



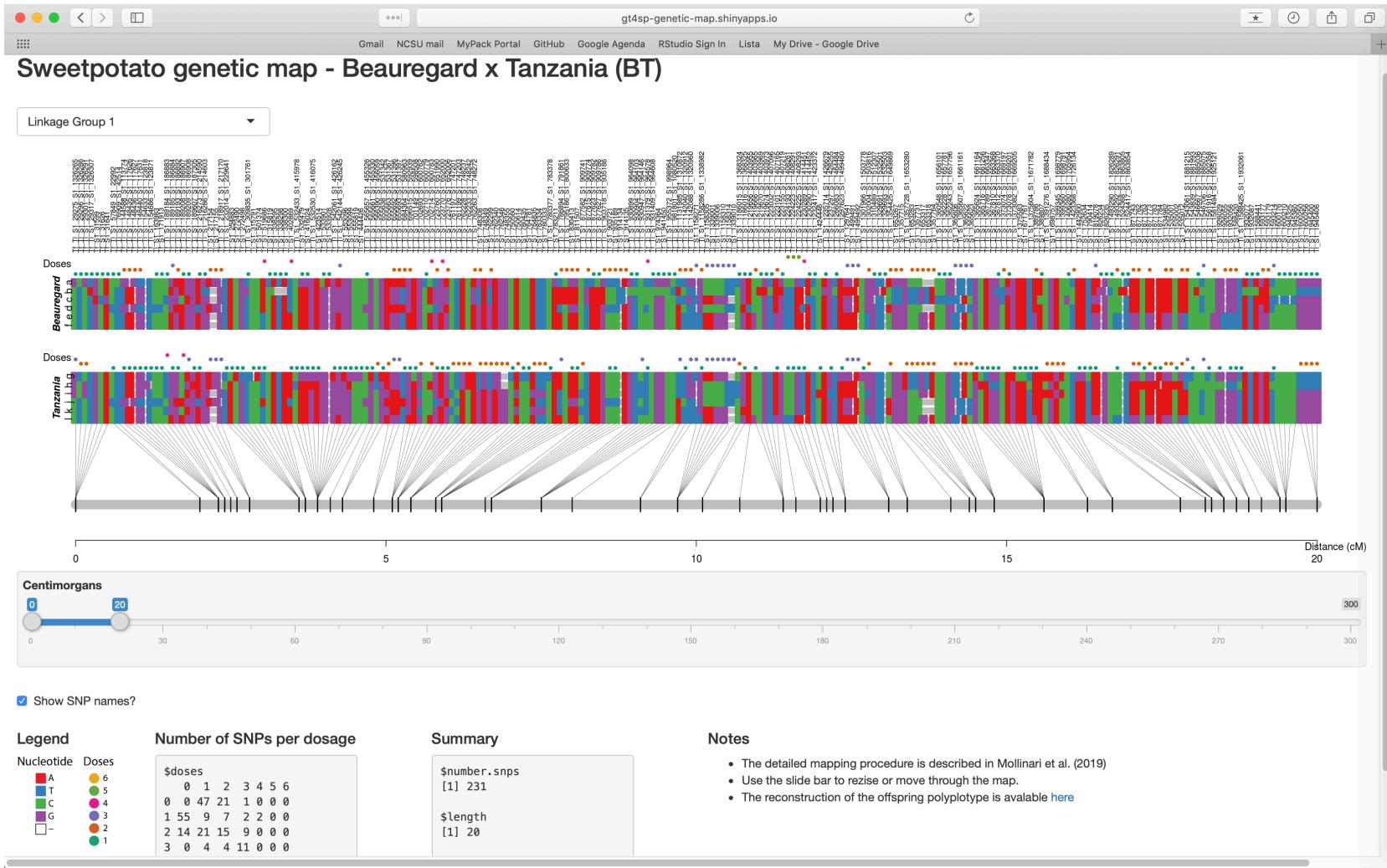
Example in a sweetpotato F_1 population

Parental haplotyping



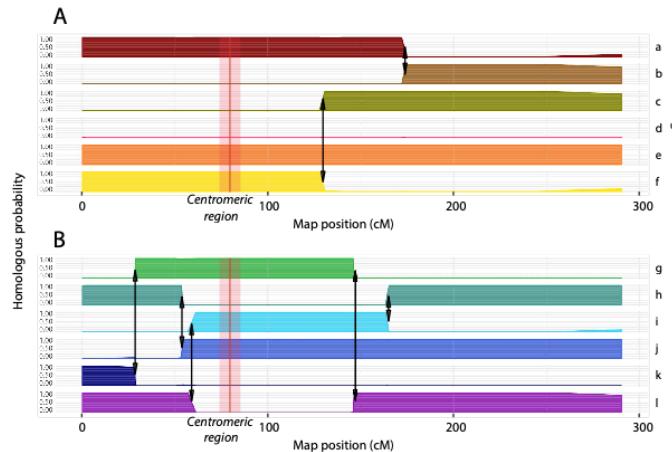
Example in a sweetpotato F_1 population

Parental haplotyping



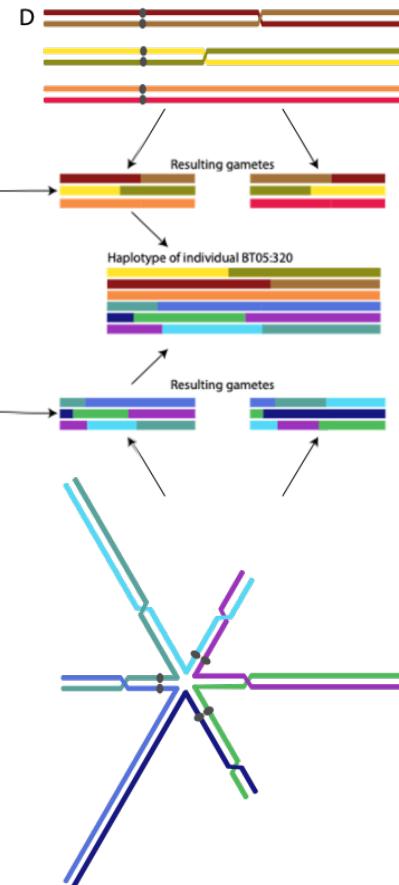
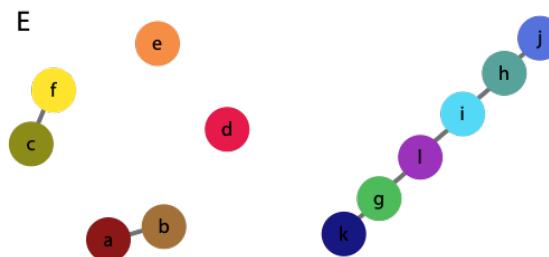
https://gt4sp-genetic-map.shinyapps.io/bt_map/

Haplotype reconstruction in the offspring



C

Parent	Homologous pair	Crossing-over position (cM)
Tanzania	k - g	28.6
Tanzania	h - j	53.8
Tanzania	i - l	60.1
Tanzania	g - l	146.5
Beauregard	c - f	130.1
Tanzania	h - i	164.5
Beauregard	a - b	173.5

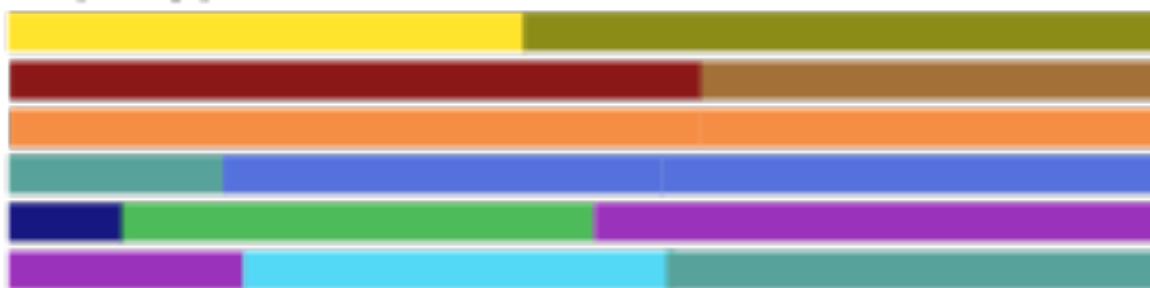


https://gt4sp-genetic-map.shinyapps.io/offspring_haplotype_BT_population/

Resulting gametes



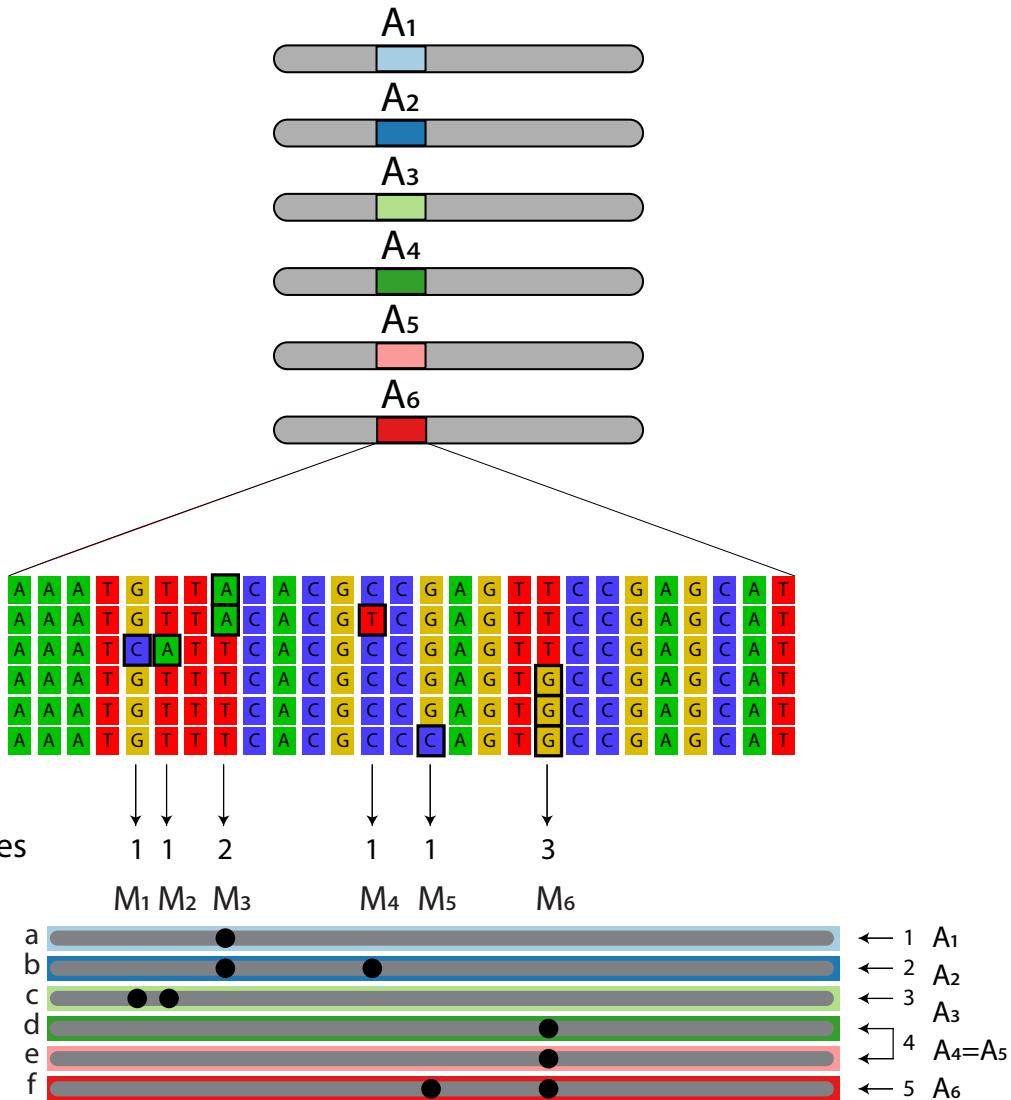
Haplotype of individual BT05:320



Resulting gametes

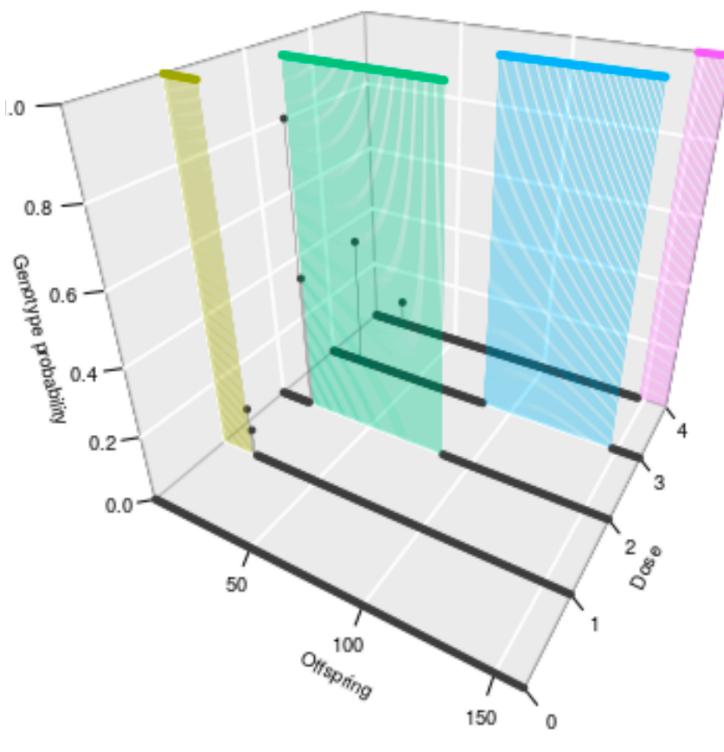


SNP calling and local haplotyping

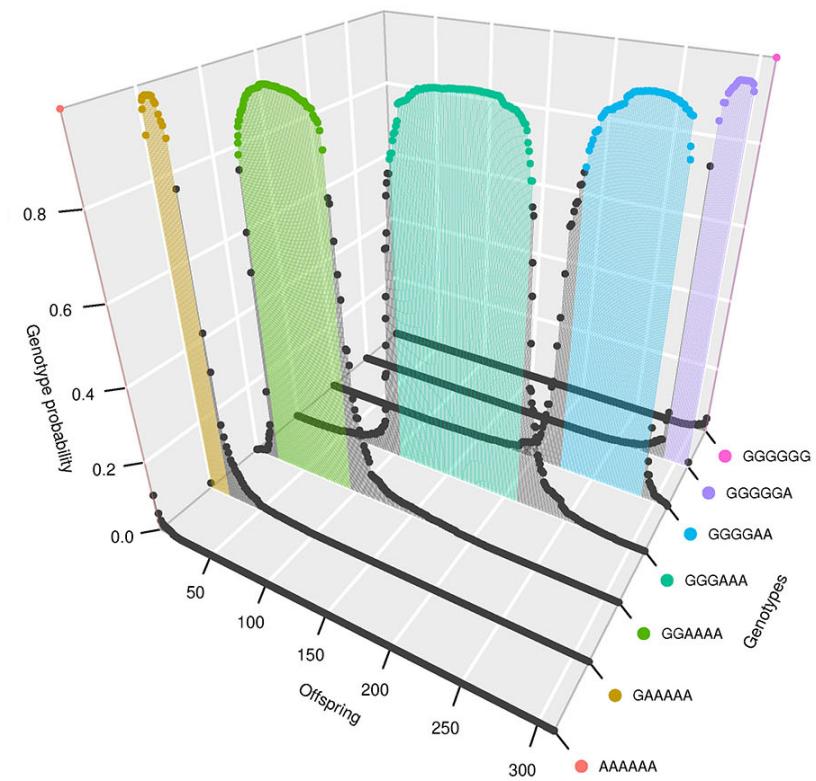


- The information is assessed at SNP level (essentially biallelic).
- It is not possible to distinguish more than two subsets of homologs.
- Combining information of several SNPs result in **haplotype-based markers**
- Mitigate **genotyping errors** (posterior state probabilities - tomorrow)

Comparison of genotype calling tetra vs. hexaploid

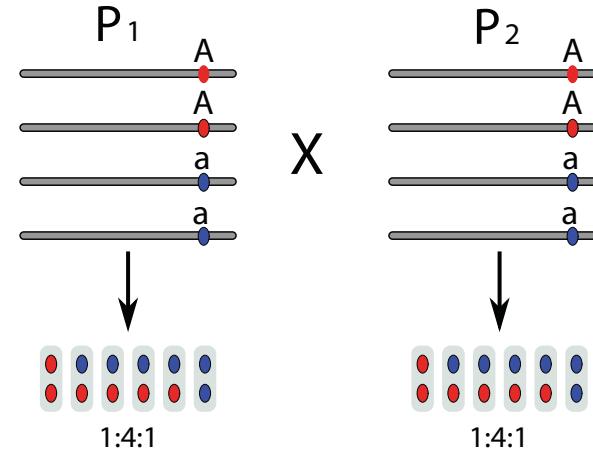
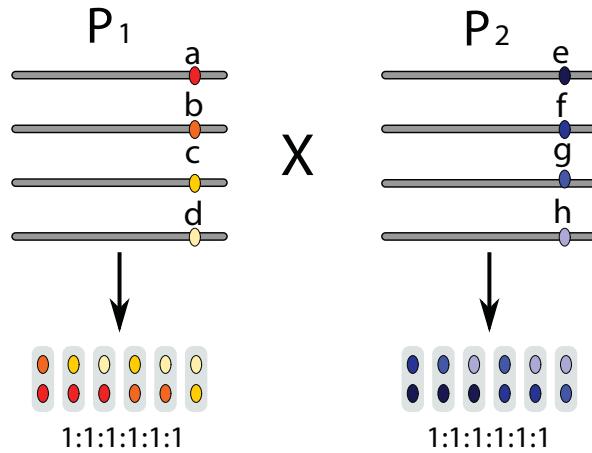


potato



sweetpotato

Inheritance pattern – Autotetraploid example



	ef	eg	eh	fg	fh	gh
ab	ab ef	ab eg	ab eh	ab fg	ab fh	ab gh
ac	ac ef	ac eg	ac eh	ac fg	ac fh	ac gh
ad	ad ef	ad eg	ad eh	ad fg	ad fh	ad gh
bc	bc ef	bc eg	bc eh	bc fg	bc fh	bc gh
bd	bd ef	bd eg	bd eh	bd fg	bd fh	bd gh
cd	cd ef	cd eg	cd eh	cd fg	cd fh	cd gh

Segregation 1 abef : 1 acef : ... : 1 bdhg : 1 cdgh

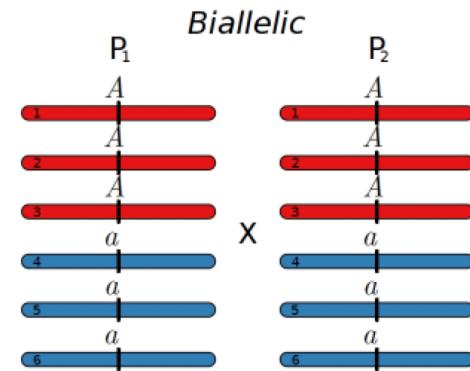
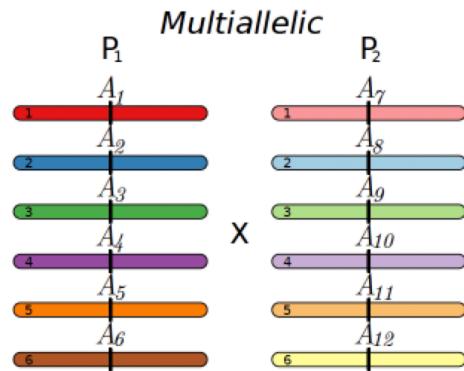
36 genotypes

AA	Aa	Aa	Aa	Aa	Aa	aa	
AA	AAAA	AAAa	AAaA	AAaA	AAAa	AAAa	AAaa
Aa	AaAA	AaAa	AaAa	AaAa	AaAa	AaAa	Aaaa
Aa	AaAA	AaAa	AaAa	AaAa	AaAa	AaAa	Aaaa
Aa	AaAA	AaAa	AaAa	AaAa	AaAa	AaAa	Aaaa
aa	aa Aa	aaaa					

Segregation 1 AAAA : 8 AAAa : 18 AAaA : 8 Aaaa : 1 aaaa

5 genotypes

Inheritance pattern – Autohexaploid example



Gametes

	P ₁	1 A ₁ A ₂ A ₃	2 A ₁ A ₂ A ₄	3 A ₁ A ₂ A ₅	...	18 A ₃ A ₄ A ₆	19 A ₃ A ₅ A ₆	20 A ₄ A ₅ A ₆
P ₂		red blue green	red purple blue	red blue orange	...	green brown red	green brown orange	purple orange brown
1	A ₇ A ₈ A ₉	red blue green	red blue green	red blue green	...	red blue green	red blue green	red blue green
2	A ₇ A ₈ A ₁₀	red blue green	red purple blue	red blue orange	...	red blue green	red blue orange	red blue orange
3	A ₇ A ₈ A ₁₁	red blue green	red purple blue	red orange orange	...	red blue green	red blue orange	red blue orange
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
18	A ₉ A ₁₀ A ₁₂	green yellow yellow	red purple yellow	red blue orange	...	green yellow yellow	red purple yellow	red blue orange
19	A ₉ A ₁₁ A ₁₂	green yellow yellow	red purple yellow	red orange orange	...	green yellow yellow	red purple yellow	red orange orange
20	A ₁₀ A ₁₁ A ₁₂	purple yellow yellow	red purple yellow	red orange orange	...	purple yellow yellow	red purple yellow	red orange orange

400 possible genotypes

1 : 1 : ... : 1 : 1

	P ₁	1 A ₇ A ₈ A ₉	2 A ₁ A ₂ A ₄	3 A ₁ A ₂ A ₅	...	18 A ₃ A ₄ A ₆	19 A ₃ A ₅ A ₆	20 A ₄ A ₅ A ₆
P ₂		red red red	red red blue	red red blue	...	red red red	red red blue	red red blue
1	A ₇ A ₈ A ₉	red red red	red red red	red red red	...	red red red	red red red	red red red
2	A ₇ A ₈ A ₁₀	red red red	red red blue	red red blue	...	red red red	red red blue	red red blue
3	A ₇ A ₈ A ₁₁	red red red	red red blue	red red blue	...	red red red	red red blue	red red blue
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
18	A ₉ A ₁₀ A ₁₂	red red blue	red red blue	red red blue	...	red red blue	red red blue	red red blue
19	A ₉ A ₁₁ A ₁₂	red red blue	red red blue	red red blue	...	red red blue	red red blue	red red blue
20	A ₁₀ A ₁₁ A ₁₂	red red blue	red red blue	red red blue	...	red red blue	red red blue	red red blue

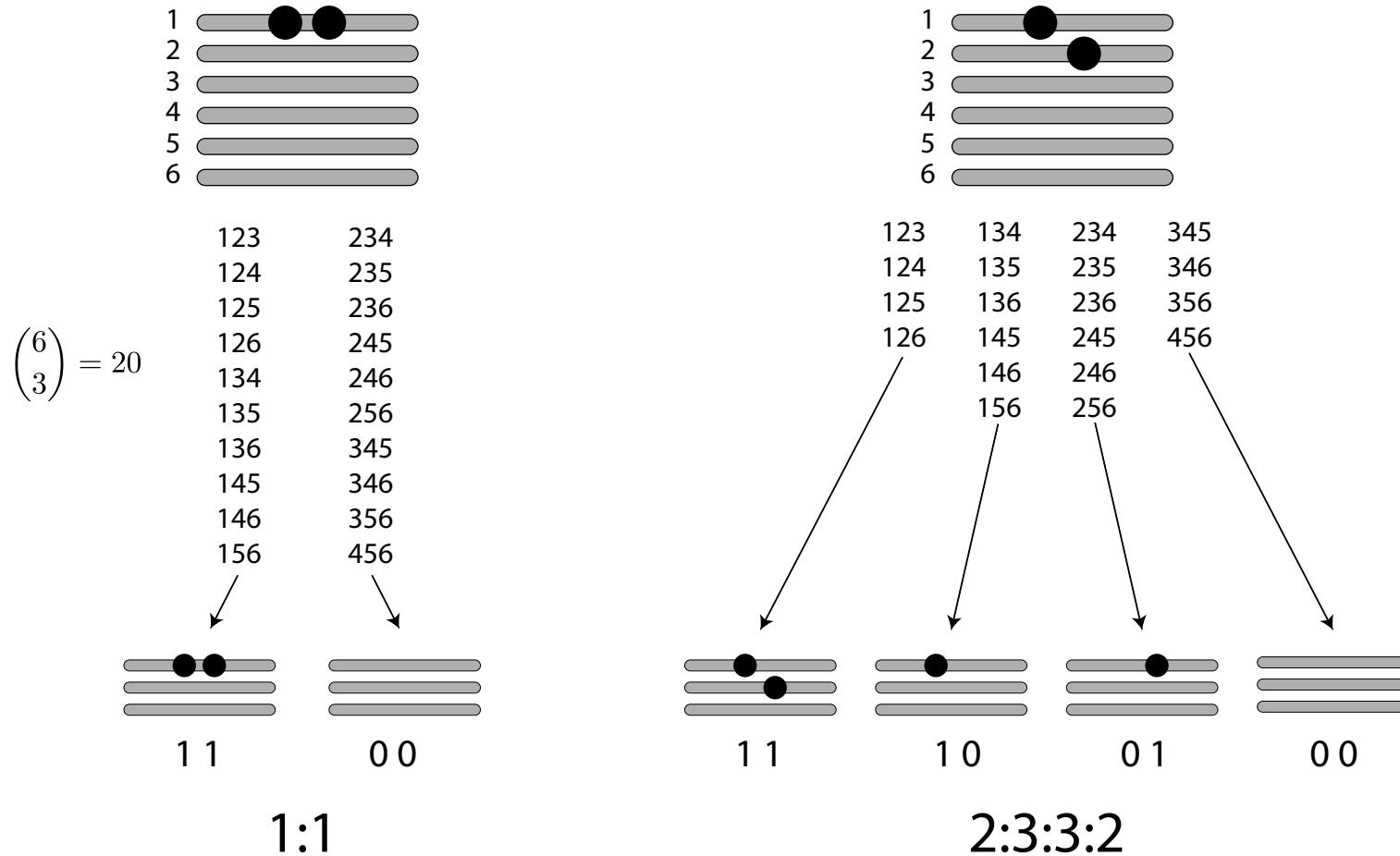
7 possible genotypes

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

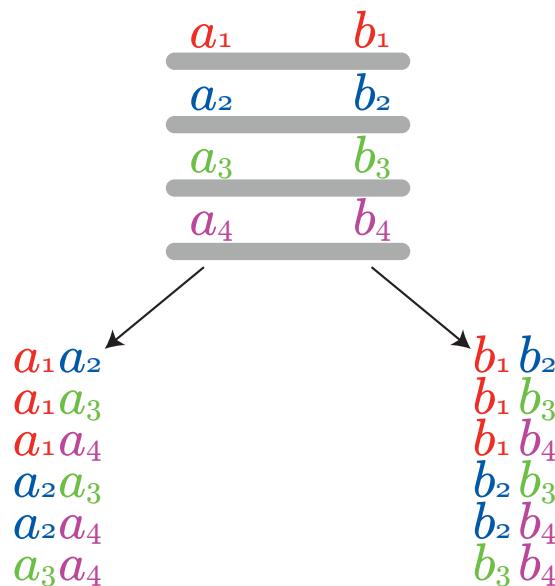


Phasing using population structure – local haplotyping

Complete linkage



Recombination Fraction - autotetraploid

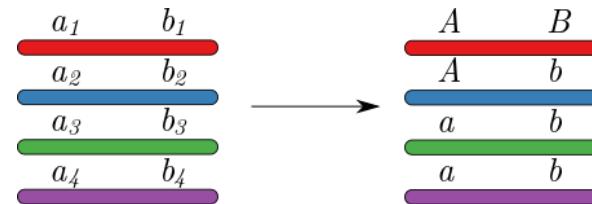


	$a_1 a_2$	$a_1 a_3$	$a_1 a_4$	$a_2 a_3$	$a_2 a_4$	$a_3 a_4$
$b_1 b_2$	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$
$b_1 b_3$	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$
$b_1 b_4$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$
$b_2 b_3$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$
$b_2 b_4$	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$
$b_3 b_4$	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$

$$\begin{aligned}
 n_0 &= \#a_1 a_2 b_1 b_2 + \dots + \#a_3 a_4 b_3 b_4 \\
 n_1 &= \#a_1 a_2 b_1 b_3 + \dots + \#a_3 a_4 b_2 b_4 \\
 n_2 &= \#a_1 a_2 b_3 b_4 + \dots + \#a_3 a_4 b_1 b_2
 \end{aligned}$$

$$\begin{aligned}
 L &= \Pr(\text{model} | \text{data}) = \prod_n \Pr(\text{loc}_B, \text{loc}_A) \\
 &= \left[\frac{(1-r)^2}{6} \right]^{n_0} \left[\frac{(1-r)r}{12} \right]^{n_1} \left[\frac{r^2}{6} \right]^{n_2}
 \end{aligned}$$

Recombination Fraction – autotetraploid



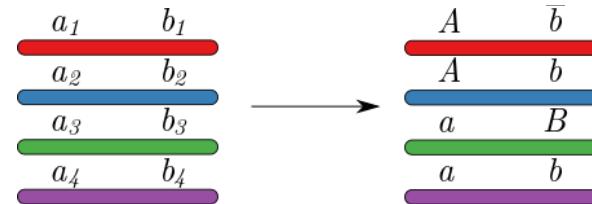
	A A	A a	A a	A a	A a	a a
B b	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$
B b	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$
B b	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$
b b	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$
b b	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$
b b	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$

$$B b \begin{bmatrix} A A & A a & a a \\ \frac{(1-r)}{6} & \frac{1}{3} & \frac{r}{6} \\ b b & \frac{r}{6} & \frac{1}{3} & \frac{(1-r)}{6} \end{bmatrix}$$

$$L_I = \left[\frac{r}{6} \right] \times \left[\frac{1}{3} \right] \times \left[\frac{(1-r)}{6} \right]$$

$$\#AAbb + \#aaBb \quad \#AaBb + \#Aabb \quad \#AABb + \#aabb$$

Recombination Fraction – autotetraploid



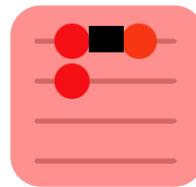
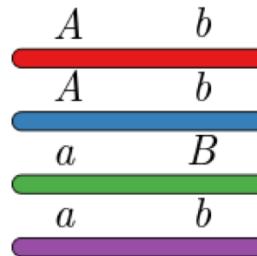
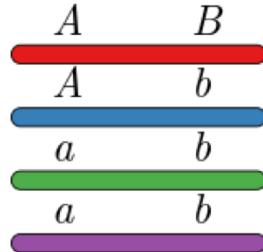
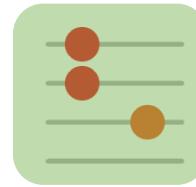
	$A\ A$					
bb	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$
bB	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$
BB	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$
Bb	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$
bb	$\frac{(1-r)r}{12}$	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$	$\frac{(1-r)r}{12}$
bB	$\frac{r^2}{6}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)r}{12}$	$\frac{(1-r)^2}{6}$

$$\begin{matrix} & \begin{matrix} A & A & A & a & a & a \end{matrix} \\ \begin{matrix} bb \\ Bb \end{matrix} & \left[\begin{matrix} \frac{(1-r)}{6} & \frac{1}{3} & \frac{r}{6} \\ \frac{r}{6} & \frac{1}{3} & \frac{(1-r)}{6} \end{matrix} \right] \end{matrix}$$

$$L_2 = \left[\frac{(1-r)}{6} \right] \times \left[\frac{1}{3} \right] \times \left[\frac{r}{6} \right]$$

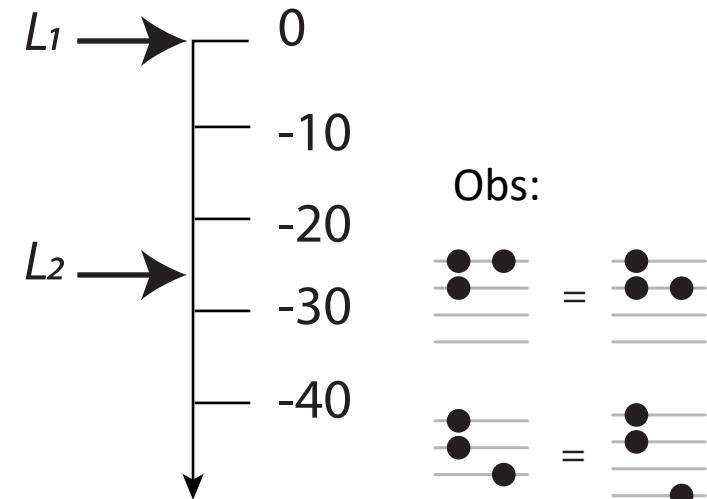
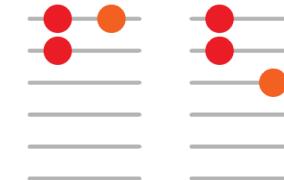
$$\#AAbb + \#aaBb \quad \#AaBb + \#Aabb \quad \#AABb + \#aabb$$

Evaluating linkage phase configurations using two-point analysis

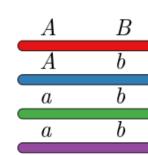
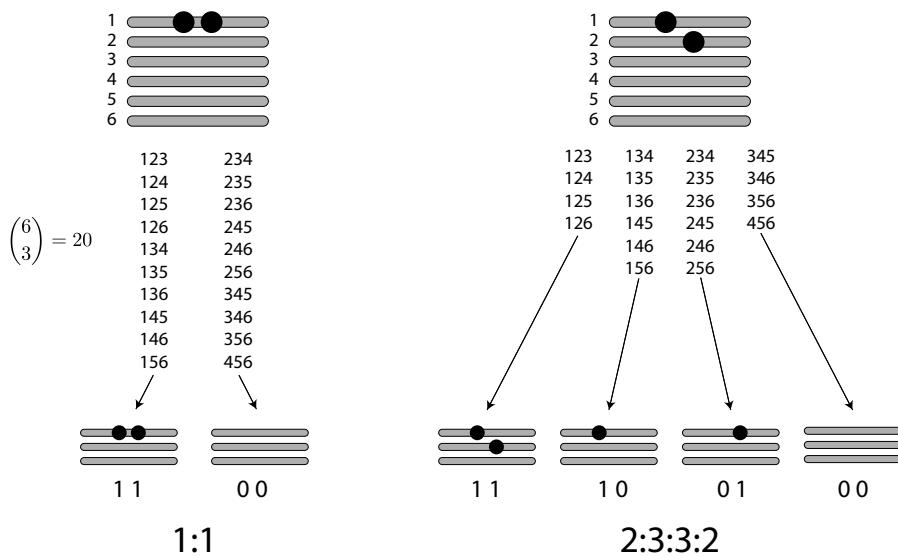
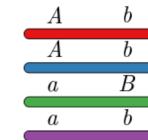
 L_1  L_2

$$LOD = \begin{cases} \log_{10} \left(\frac{L_2}{L_1} \right), & \text{if } L_1 > L_2 \\ \log_{10} \left(\frac{L_1}{L_2} \right), & \text{if } L_2 > L_1 \end{cases}$$

	LOD_ph	rf	LOD_rf
L_1	1	0.0	26.5
L_2	0	-26.5	0.5

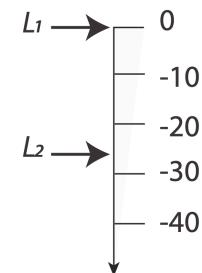
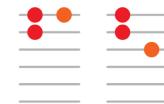


Phasing using population structure

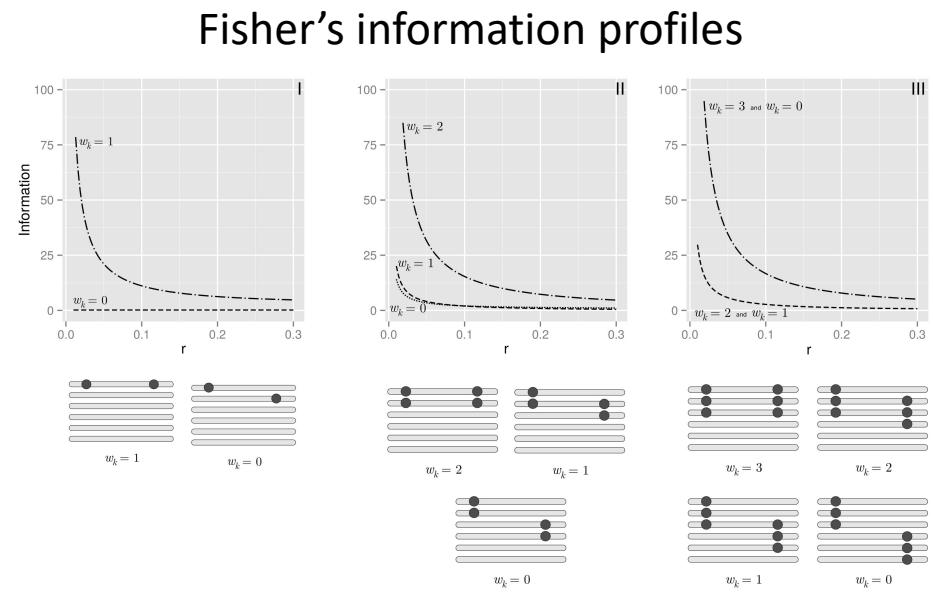
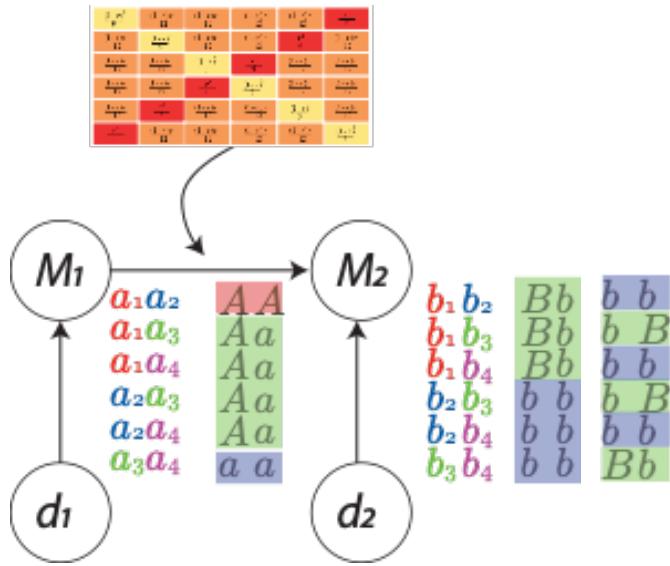
 L_1  L_2

$$LOD = \begin{cases} \log_{10}\left(\frac{L_2}{L_1}\right), & \text{if } L_1 > L_2 \\ \log_{10}\left(\frac{L_1}{L_2}\right), & \text{if } L_2 > L_1 \end{cases}$$

	LOD_ph	rf	LOD_rf
L_1	1	0.0	26.5
L_2	0	-26.5	0.5

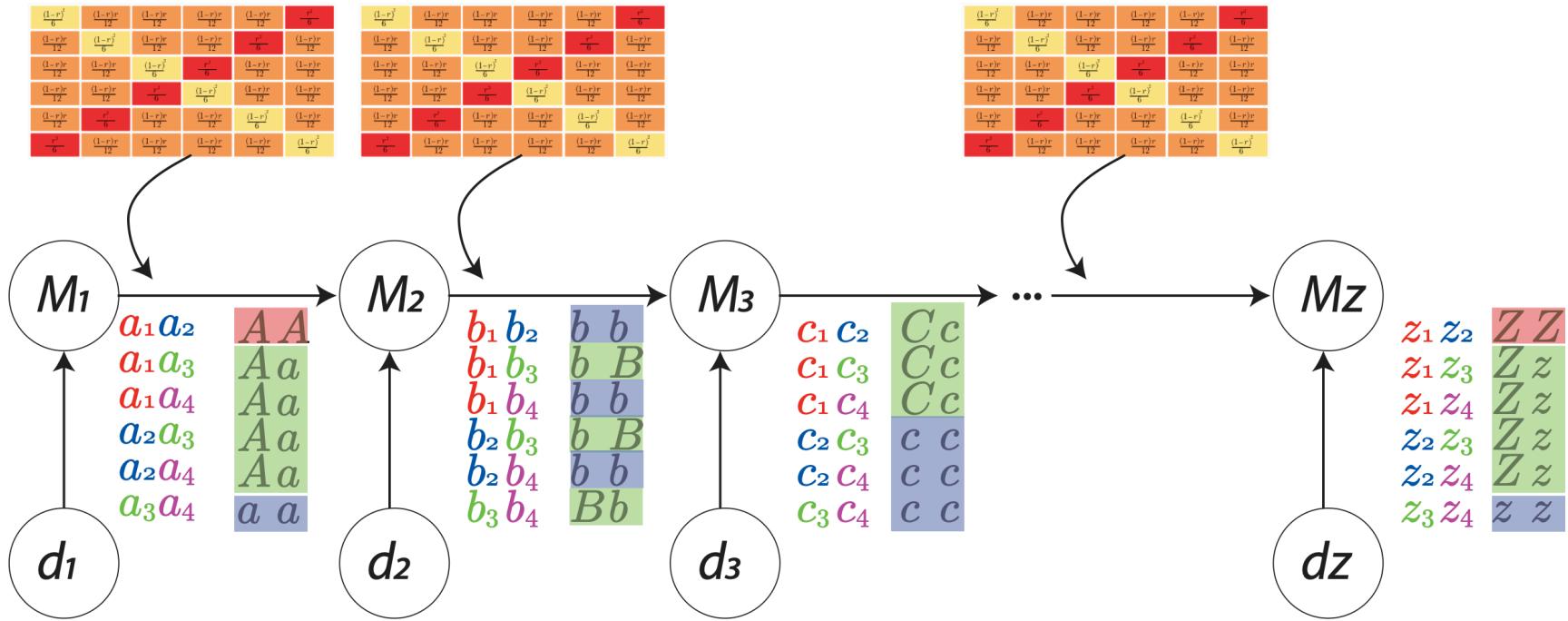


Graphical model for two-point analysis



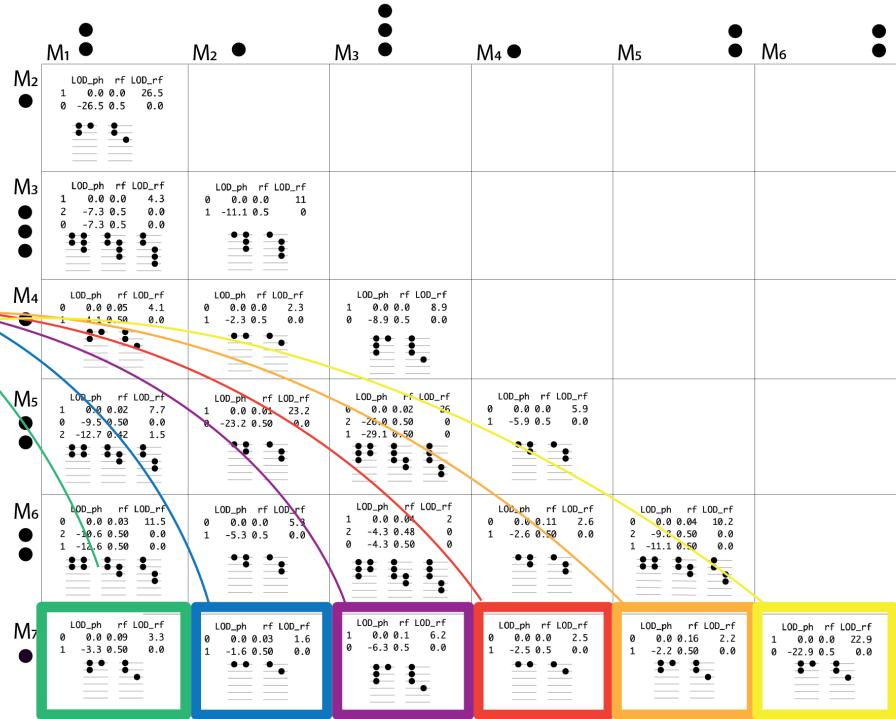
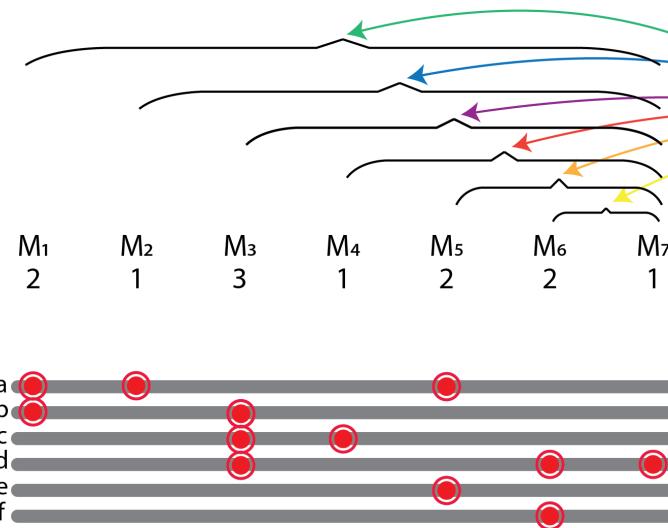
Sometimes, there is not enough information to judge each configuration we is the more likely just using two-point analysis

Graphical model for HMM analysis



Phasing algorithm (two-points + HMM)

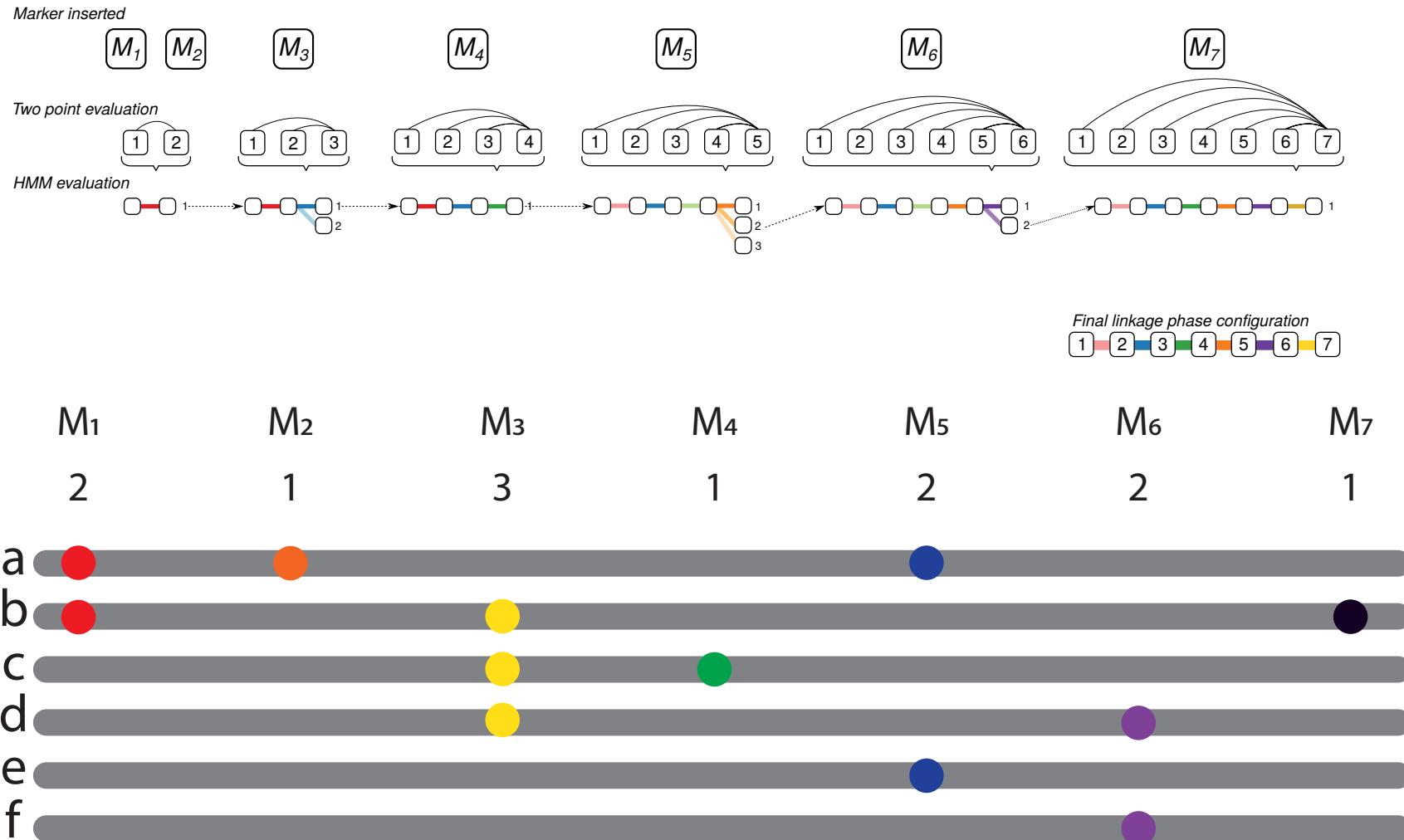
Two-point LOD threshold = 5.0



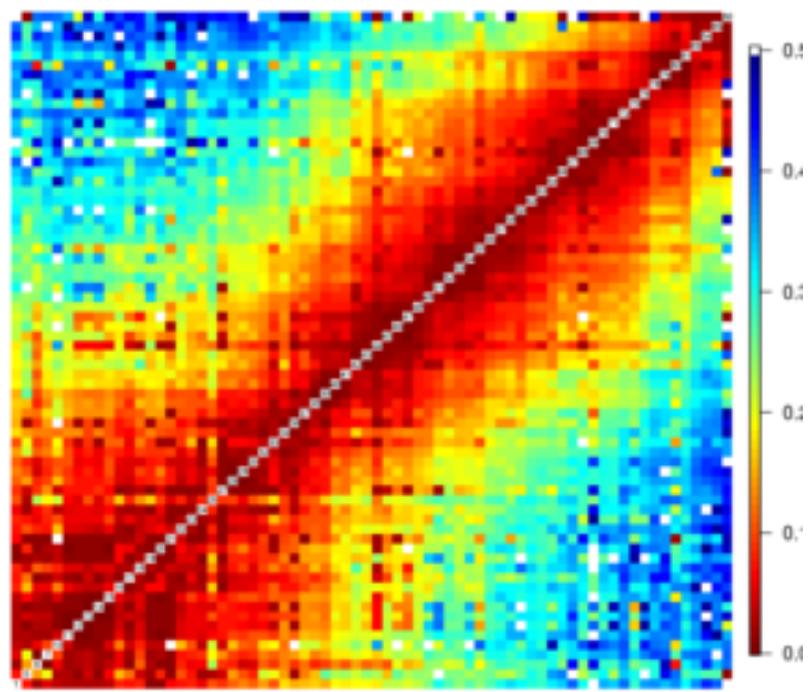
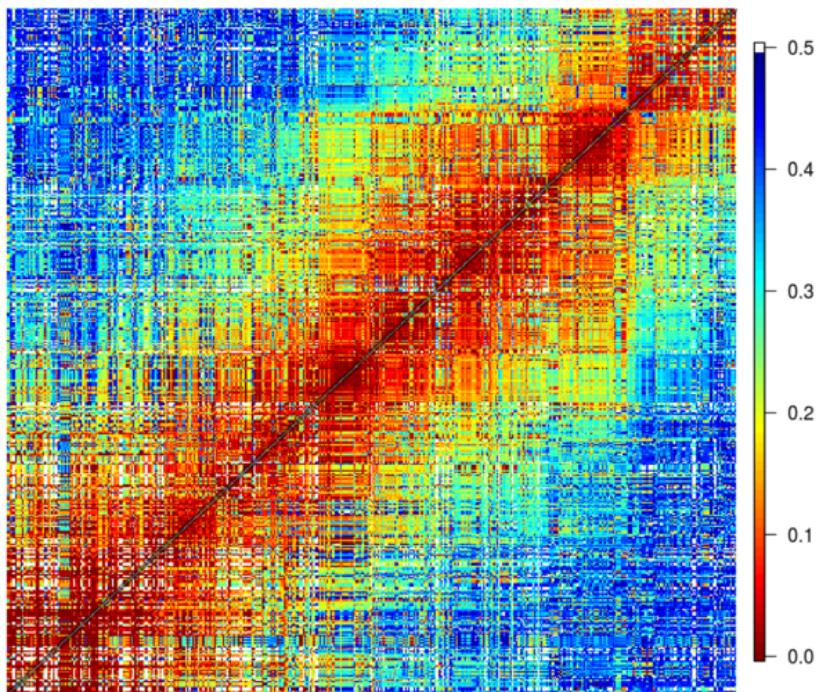
+

HMM evaluation

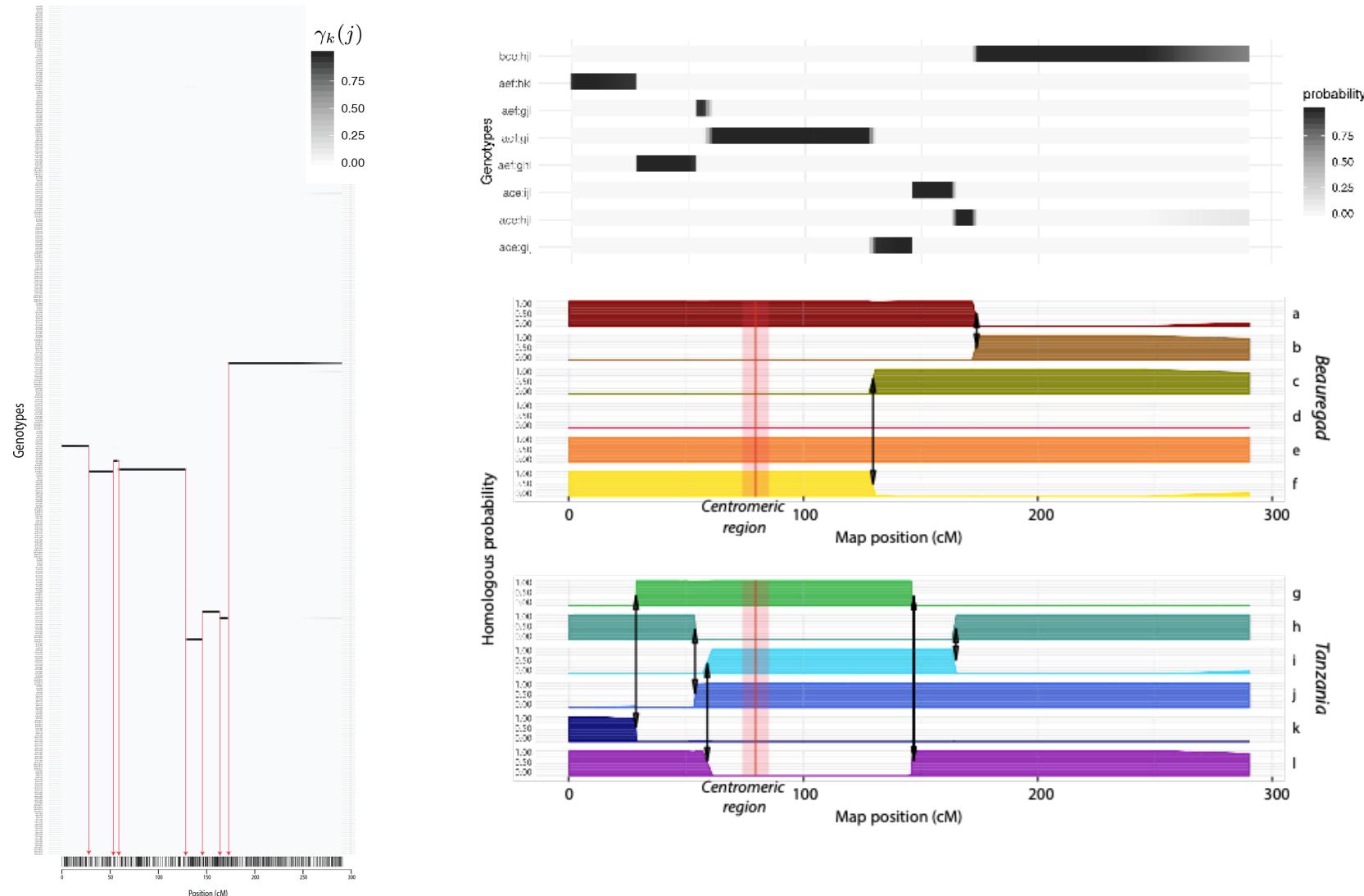
Phasing algorithm (two-points + HMM)



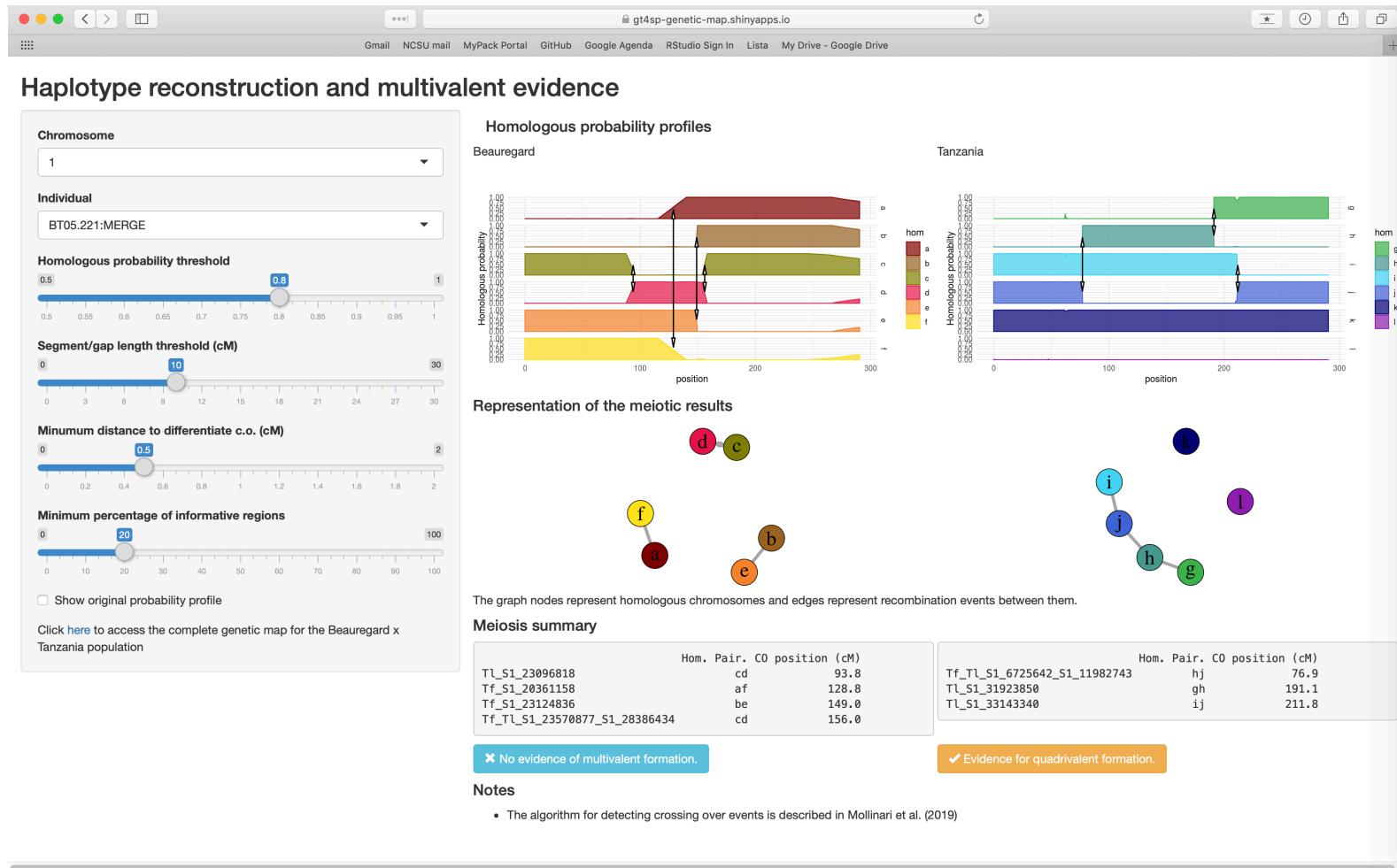
Effect of the information on the recombination fraction



Posterior state probability and offspring haplotyping



Haplotype reconstruction in the offspring



Conclusions

- There are two instances of haplotyping: **local** and **global**.
- **Local haplotyping** refers to assemble more informative markers in order to differentiate homologs at a specific locus. The more differentiated homologous, the more informative the marker.
- Dosage based markers allow us to integrate homologs in homology groups. However, the dosage information is a special case, an incomplete way to access the genotype.
- We do not need **fully informative markers** to improve analysis; **partially informative markers** also can improve it: (e.g. triallelic marker is more informative than biallelic)
- **Global haplotyping** help us to understand meiotic mechanisms, and consequently, segregation patterns
- Haplotyping can be used to **mitigate genotyping errors** (we will see tomorrow)

Acknowledgements

- Augusto Garcia
- Guilherme Pereira
- Bode Olukolu
- Dorcus Gemenet
- Craig Yencho
- Zhao-Bang Zeng
- FAPESP: São Paulo Research Foundation
- Bill and Melinda Gates Foundation