Multiple QTL mapping in autopolyploids with QTLpoly

Guilherme da Silva Pereira

g.pereira@cgiar.org

gdasilv@ncsu.edu

January 15, 2021

NC STATE UNIVERSITY

CIP INTERNATIONAL POTATO CENTER

Outline

1. Multiple QTL mapping
   - Components of variance
   - Random-effect QTL model
2. Applications
   - Potato: 'Atlantic' $\times$ B1829-5
   - Sweetpotato: 'Beauregard' $\times$ 'Tanzania'
3. Final considerations

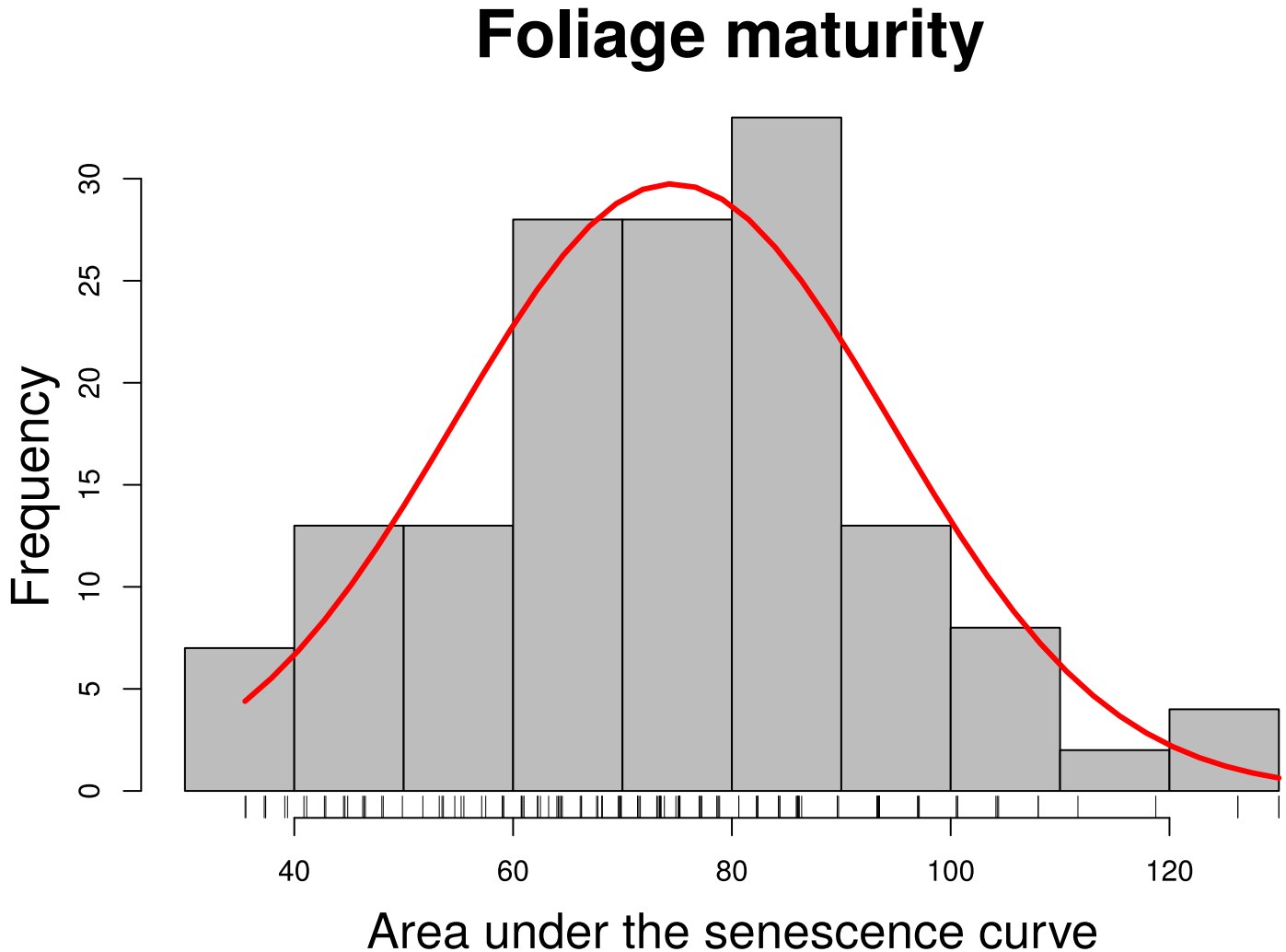
Outline

1. Multiple QTL mapping
   - Components of variance
   - Random-effect QTL model
2. Applications
   - Potato: 'Atlantic' $\times$ B1829-5
   - Sweetpotato: 'Beauregard' $\times$ 'Tanzania'
3. Final considerations

Outline

1. Multiple QTL mapping
   - Components of variance
   - Random-effect QTL model
2. Applications
   - Potato: 'Atlantic' × B1829-5
   - Sweetpotato: 'Beauregard' × 'Tanzania'
3. Final considerations

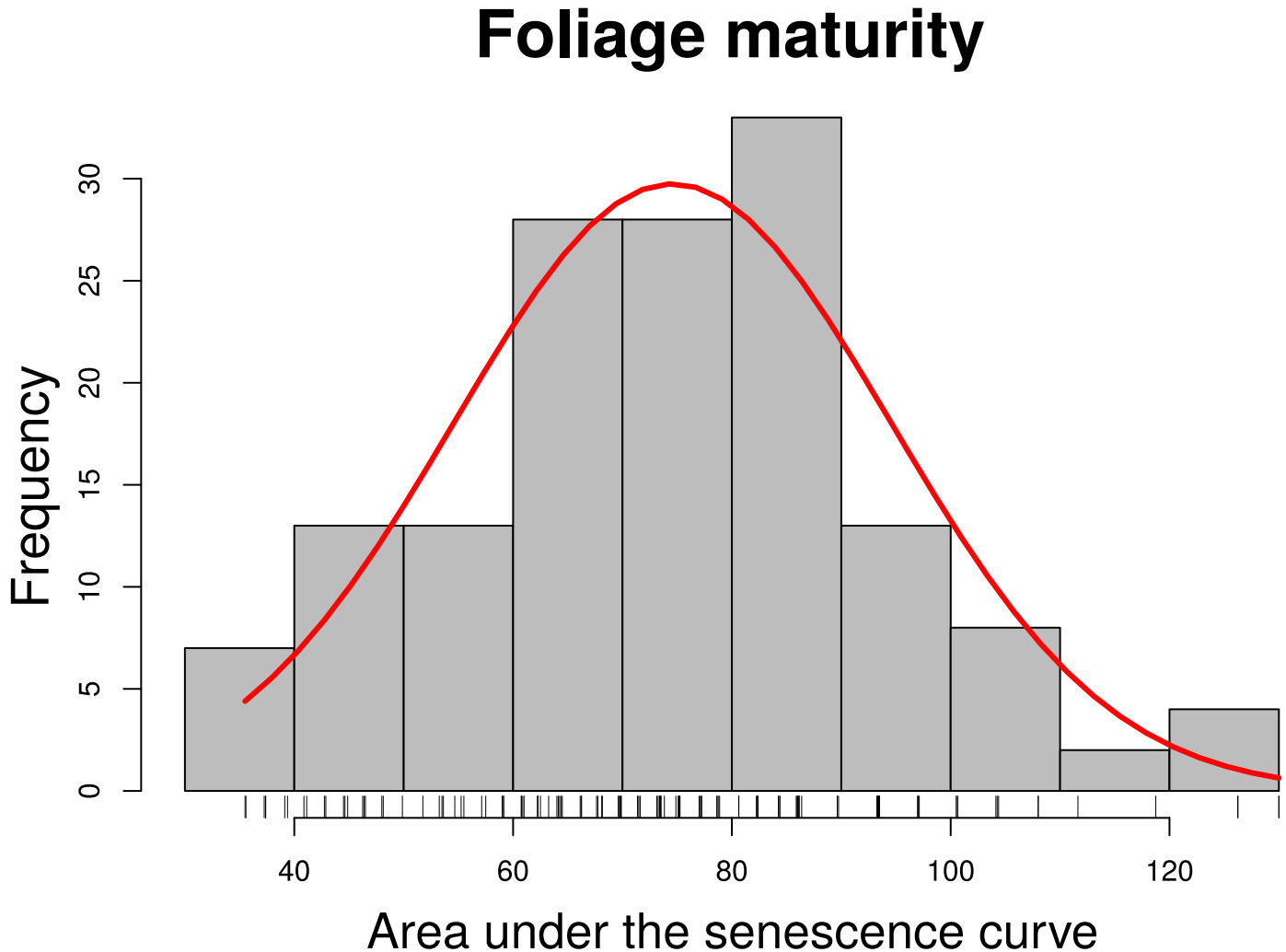
Potato: 'Atlantic' $\times$ B1829-5 ($N = 154$)



$$y_i \sim (\mu, \sigma_p^2)$$

Phenotypic data is usually described in terms of a mean and a variance

Potato: 'Atlantic' $\times$ B1829-5 ($N = 154$)



$$y_i \sim (74.57, 399.24)$$

Phenotypic data is usually described in terms of a mean and a variance

What makes potatoes different from each other?

1st case: all individuals are the same (e.g., clones of 'Atlantic')

$$\sigma_p^2 = \sigma_e^2$$

- $\sigma_e^2 = 399.24$: environmental component

2st case: individuals are an F_1 progeny ('Atlantic' $\times$ B1829-5)

$$\sigma_p^2 = \sigma_g^2 + \sigma_e^2$$

- $\sigma_g^2 = 183.65$: genetic component (polygenic effects)
- $\sigma_e^2 = 215.59$: environmental component

Phenotypic variance (σ_p^2) can be attributed to different sources

Genetic variance (σ_g^2)

Broad-sense heritability (H^2)

$$H^2 = \frac{\sigma_g^2}{\sigma_p^2}$$

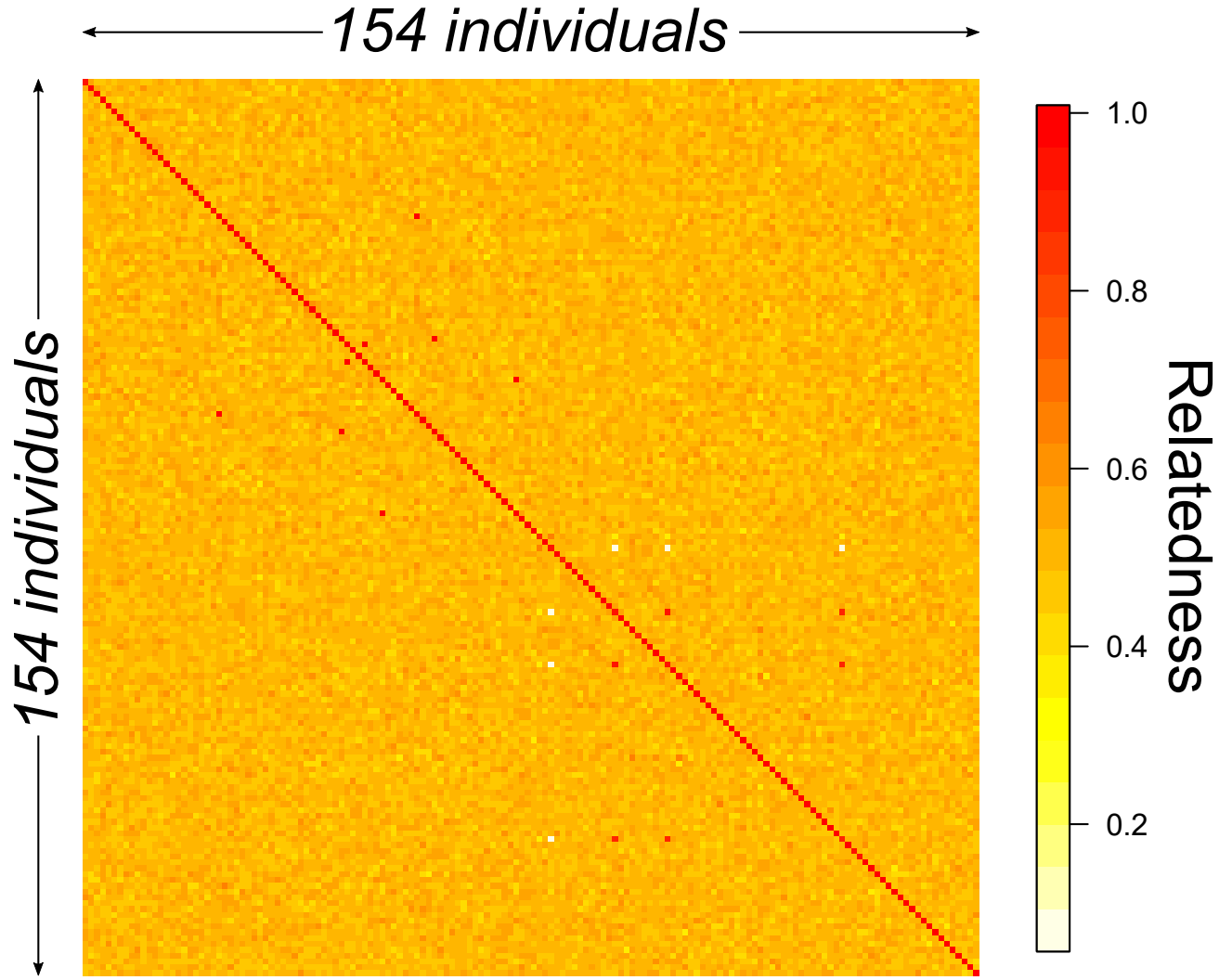
Additive genetic variance (σ_a^2)

$$\sigma_p^2 = \boldsymbol{A}^{(i,i')}\sigma_a^2 + \sigma_e^2$$

- $\boldsymbol{A}^{(i,i')}$ is the genome-wide relationship describing the average covariance between a sib-pair (i, i') if the covariance is caused by heritable effects
- Narrow-sense heritability: $h^2 = \frac{\sigma_a^2}{\sigma_p^2}$

Genetic variance (σ_g^2) can be further decomposed into additive (σ_a^2) and other variances (e.g., dominance, epistasis)

Realized relationship matrix G for 4,285 markers



Expected value for full-sibs: $\mathrm{E}(\boldsymbol{G}^{(i,i')}) = \boldsymbol{A}^{(i,i')} = \frac{1}{2}$ (the kinship matrix)

Are there any specific loci associated with the phenotype?

Quantitative trait loci (QTL)

Region of DNA which contributes to the variation of a particular trait

Variance due to a single QTL (σ_q^2)

$$\sigma_p^2 = \boldsymbol{G}_q^{(i,i')}\sigma_q^2 + \sigma_e^2$$

- $\boldsymbol{G}_q^{(i,i')}$ is the relationship describing the covariance between a sib-pair (i, i') if the covariance is caused by a QTL q

In the region of a gene influencing a trait of interest

- Correlated QTL effects among individuals are due to shared alleles identical-by-descent (IBD)
- Higher IBD allele sharing for a QTL is expected among individuals with similar phenotypic values

Diploid: outcrossing species

(2 gametes each parent)

$$ab \times cd$$

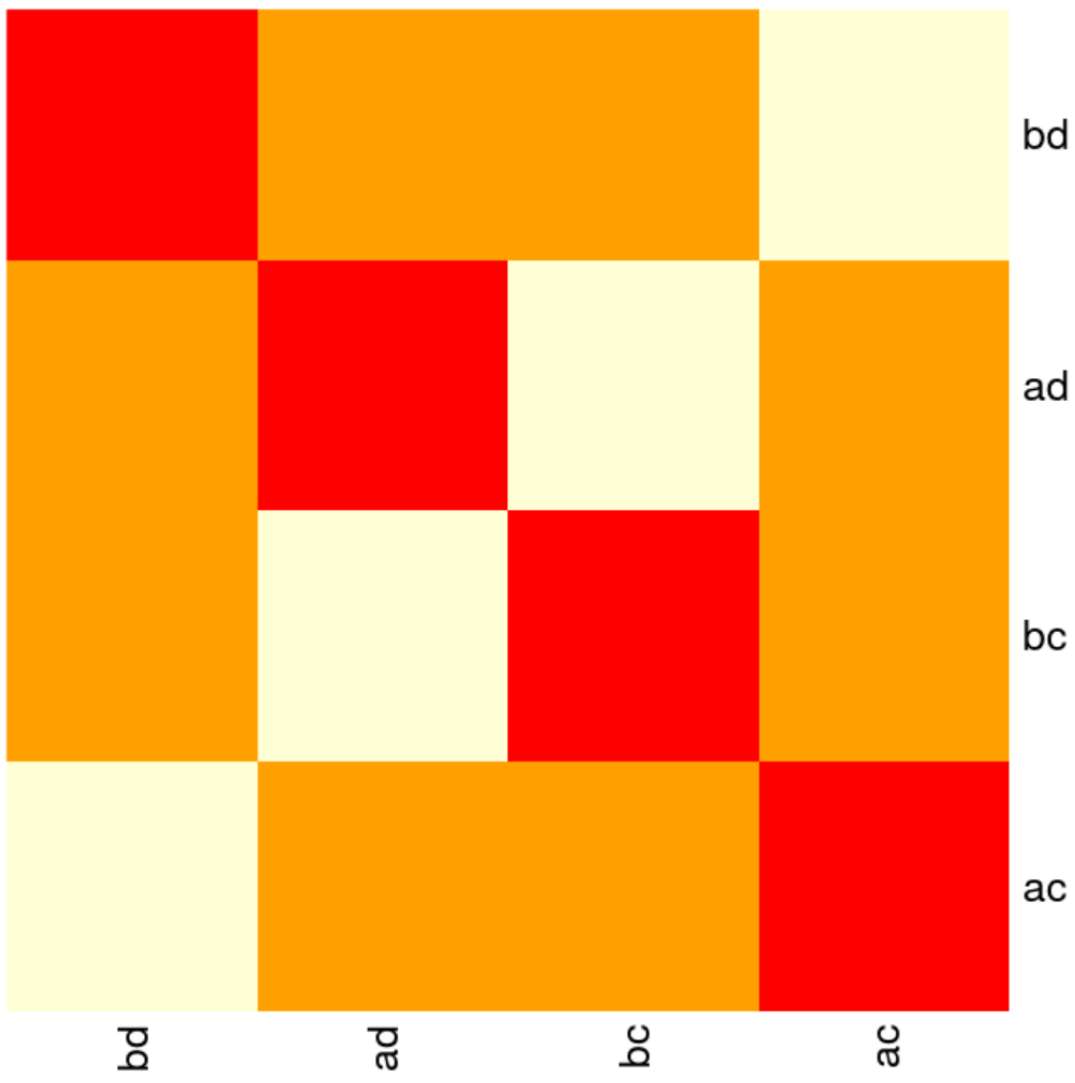
$$\downarrow$$

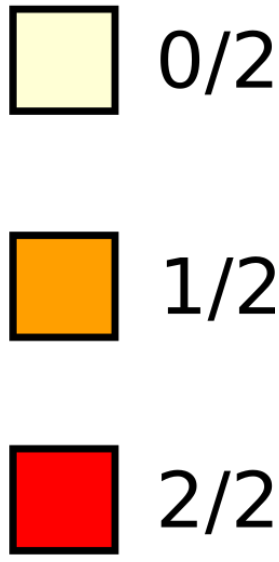
	c	d
a	ac	ad
b	bc	bd

(1 ac : 1 ad : 1 bc : 1 bd)

(4 genotypes)

Diploid: identity-by-descent





Autotetraploid: outcrossing species

(6 gametes each parent)

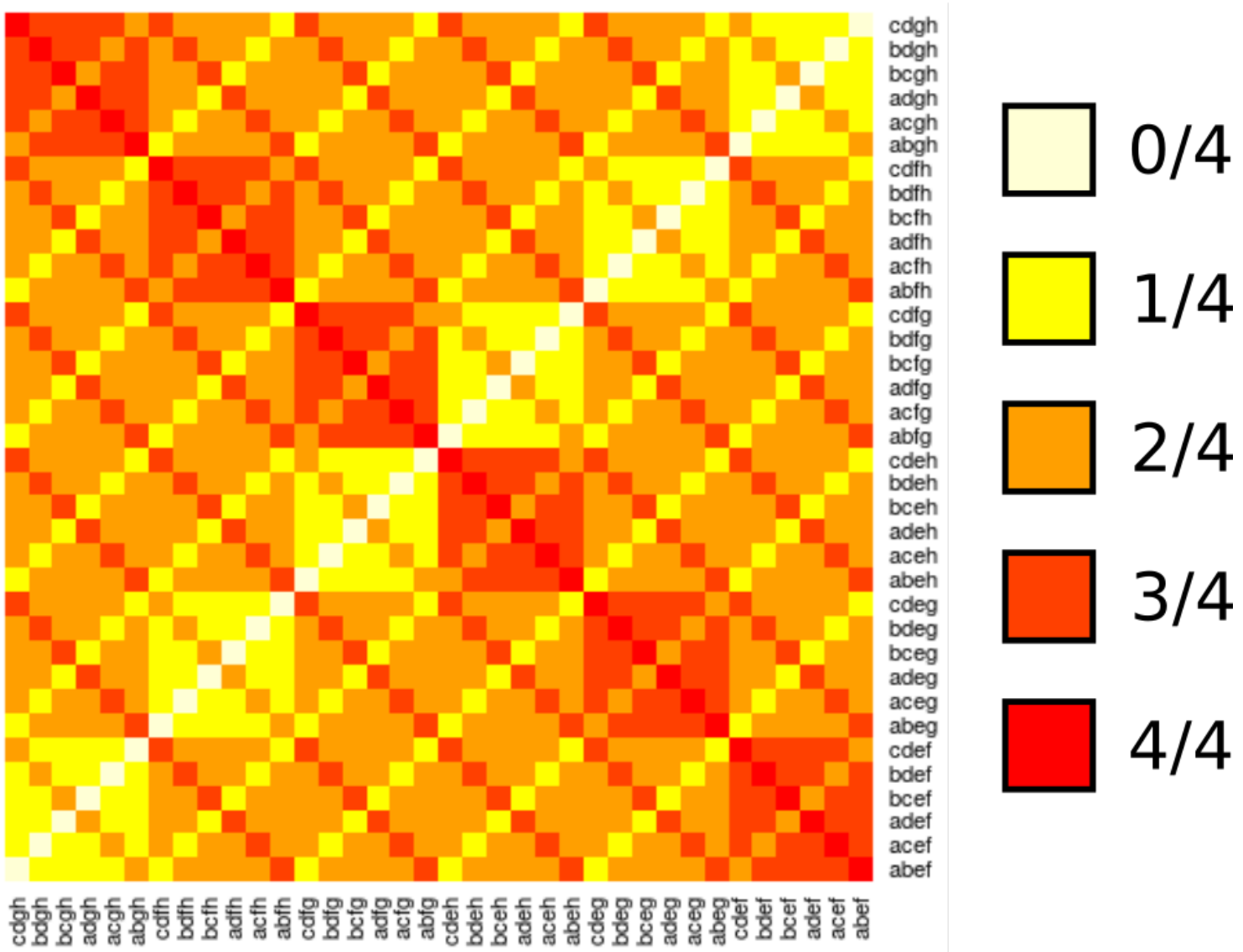
abcd $\times$ efgh

$\downarrow$

	ef	eg	eh	fg	fh	gh
ab	abef	abeg	abeh	abfg	abfh	abgh
ac	acef	aceg	aceh	acfg	acfh	acgh
ad	adef	adeg	adeh	adfg	adfh	adgh
bc	bcef	bceg	bceh	bcfg	bcfh	bcgh
bd	bdef	bdeg	bdeh	bdfg	bdfh	bdgh
cd	cdef	cdeg	cdeh	cdfg	cdfh	cdgh

(36 genotypes)

Autotetraploid: identity-by-descent



Autohexaploid: outcrossing species

(20 gametes each parent)

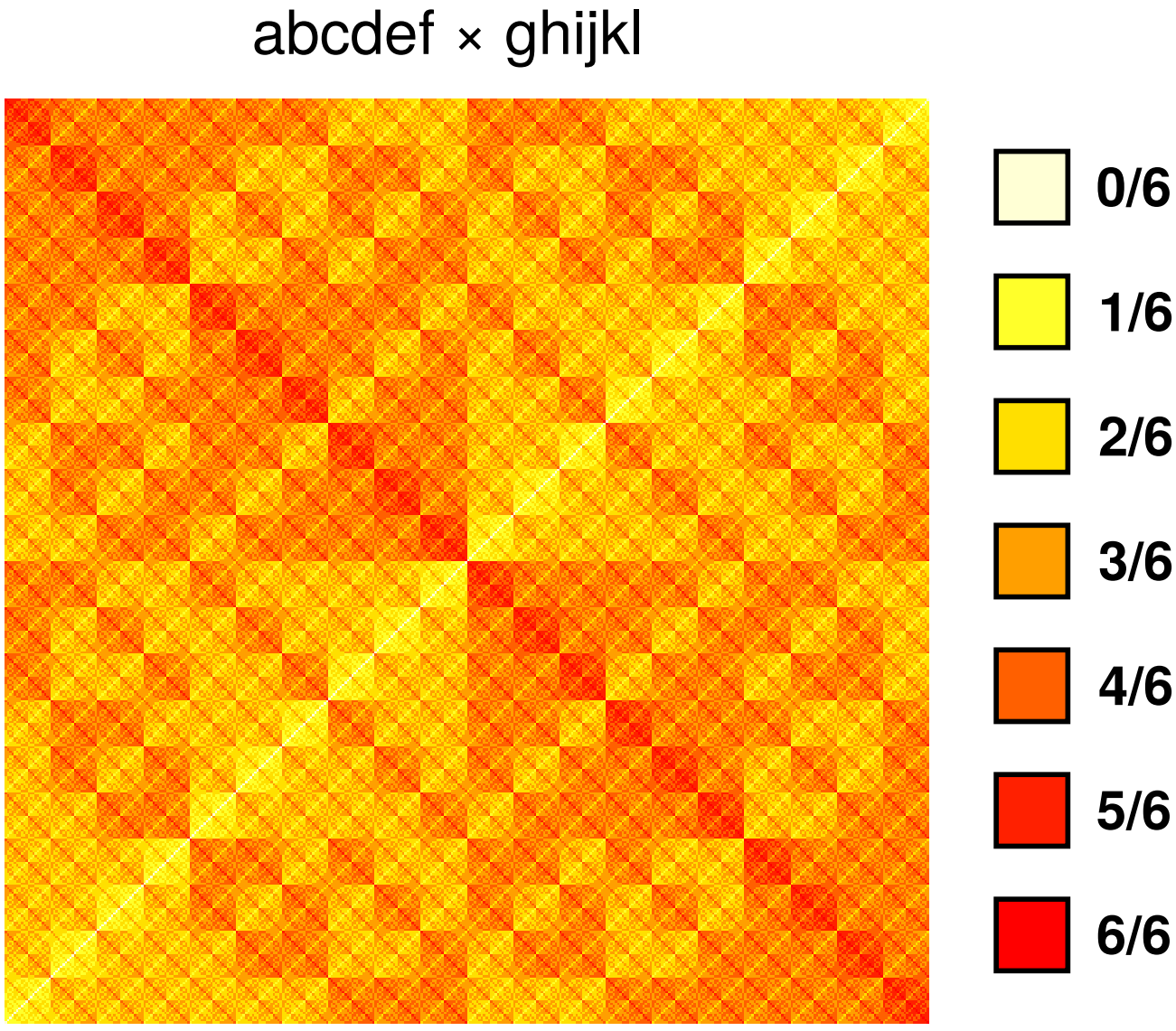
abcdef $\times$ ghijkl

$\downarrow$

	ghl	ghj	ghk	$\cdots$	ijl	ikl	jkl
abc	abcghl	abcghj	abcghk	$\cdots$	abcijl	abcikl	abcjkl
abd	abdghl	abdghj	abdghk	$\cdots$	abdijl	abdikl	abdjkl
abe	abeghl	abeghj	abeghk	$\cdots$	abeijl	abeikl	abejkl
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
cdf	cdfghl	cdfghj	cdfghk	$\cdots$	cdfijl	cdfikl	cdfjkl
cef	cefghl	cefghj	cefghk	$\cdots$	cefijl	cefikl	cefjkl
def	defghl	defghj	defghk	$\cdots$	defijl	defikl	defjkl

(400 genotypes)

Autohexaploid: identity-by-descent



Outline

1. Multiple QTL mapping
   - Components of variance
   - Random-effect QTL model
2. Applications
   - Potato: 'Atlantic' $\times$ B1829-5
   - Sweetpotato: 'Beauregard' $\times$ 'Tanzania'
3. Final considerations

Random-effect multiple QTL model

From single to multiple QTL analysis

- Random-effect multiple interval mapping (REMIM)

$$\sigma_p^2 = \sum_{q=1}^{Q} \boldsymbol{G}_q^{(i,i')} \sigma_q^2 + \sigma_e^2$$

- $\boldsymbol{G}_q^{(i,i')}$ is the relationship describing the covariance between a sib-pair (i, i') if the covariance is caused by a QTL q
- There are Q QTL contributing to the variance of a particular trait

QTL heritability (h_q^2)

$$h_q^2 = \frac{\sigma_q^2}{\sigma_p^2}$$

Random-effect multiple QTL model

$$\boldsymbol{y} = \mathbf{1}\mu + \sum_{q=1}^{Q} \boldsymbol{g}_q + \boldsymbol{e}$$

- $\boldsymbol{y}$ is the $n \times 1$ vector of phenotypic values
- μ is the fixed effect of the population mean
- $\boldsymbol{g}_q$ is the $n \times 1$ random vector of the genotype effects of QTL q $(q = 1, \ldots, Q)$ with $\boldsymbol{g}_q \sim \mathcal{N}(\mathbf{0}, \boldsymbol{G}_q \sigma_q^2)$
- $\boldsymbol{e}$ is the $n \times 1$ random vector of residual error with $\boldsymbol{e} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}\sigma_e^2)$
  - $\mathbf{1}$ and $\boldsymbol{I}$ are an $n \times 1$ vector of 1's and an $n \times n$ identity matrix, respectively
  - $\boldsymbol{G}_q$ is the $n \times n$ IBD-based relationship matrix at QTL q

Genetics, Vol. 215, 579–595 July 2020

HIGHLIGHTED ARTICLE
GENETICS | INVESTIGATION

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

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

Genome scanning

- Tests are performed every 1 cM along the whole genome
- Multiple-QTL model: QTLs are added consecutively to the model

Hypothesis testing: $H_0 : \sigma_q^2 = 0$; $H_1 : \sigma_q^2 > 0$

- Restricted maximum likelihood (REML): estimation
- Score statistics: significance of the variance component
- Resampling method: genome-wide significance level (α)

Model selection algorithm

1. A model starts with no QTL

$$\sigma_p^2 = \sigma_e^2$$

2. *Forward search* adds one QTL q at a time if significant ($\alpha = 0.20$)

$$\sigma_p^2 = \sum_{q=1}^{Q} \boldsymbol{G}_q^{(i,i')} \sigma_q^2 + \sigma_e^2$$

3. *Model optimization* re-evaluate every QTL r again conditional to all the other QTL in the model at a higher significance level ($\alpha = 0.05$)

$$\sigma_p^2 = \boldsymbol{G}_r^{(i,i')} \sigma_r^2 + \sum_{q \neq r} \boldsymbol{G}_q^{(i,i')} \sigma_q^2 + \sigma_e^2$$

   - *Position refining* updates QTL positions
   - *Backward elimination* excludes QTL showing no significant effect

QTL detection consists of a stepwise procedure

Outline

1. Multiple QTL mapping
   - Components of variance
   - Random-effect QTL model
2. Applications
   - Potato: ‘Atlantic’ $\times$ B1829-5
   - Sweetpotato: ‘Beauregard’ $\times$ ‘Tanzania’
3. Final considerations

Outline

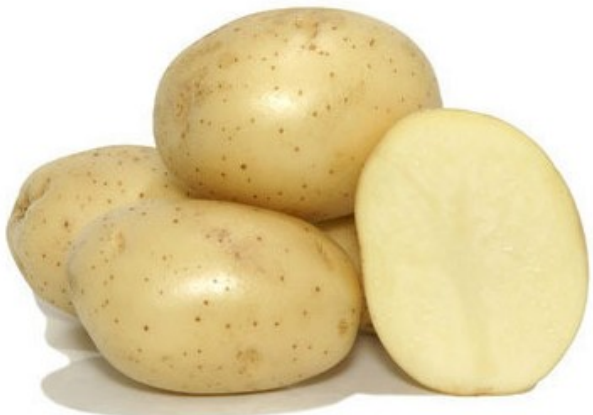
1. Multiple QTL mapping
   - Components of variance
   - Random-effect QTL model
2. Applications
   - Potato: 'Atlantic' $\times$ B1829-5
   - Sweetpotato: 'Beauregard' $\times$ 'Tanzania'
3. Final considerations

Potato: 'Atlantic' $\times$ B1829-5 ($N = 154$)

- *Solanum tuberosum* ($2n = 4x = 48$, $x = 12$)

Phenotypic data

- 4 years (2006, 2007, 2008, 2014)
- 7 traits
  1. Plant yield
  2. Internal heat necrosis (IHN) severity
  3. IHN incidence
  4. Specific Gravity
  5. Dry Matter
  6. Skin texture
  7. Foliage maturity



The recombination landscape and multiple QTL mapping in a *Solanum tuberosum* cv. 'Atlantic'-derived F_1 population

Guilherme Da Silva Pereira, Marcelo Mollinari, Mitchell J Schumann, Mark E Clough, Zhao-Bang Zeng, Craig Yencho

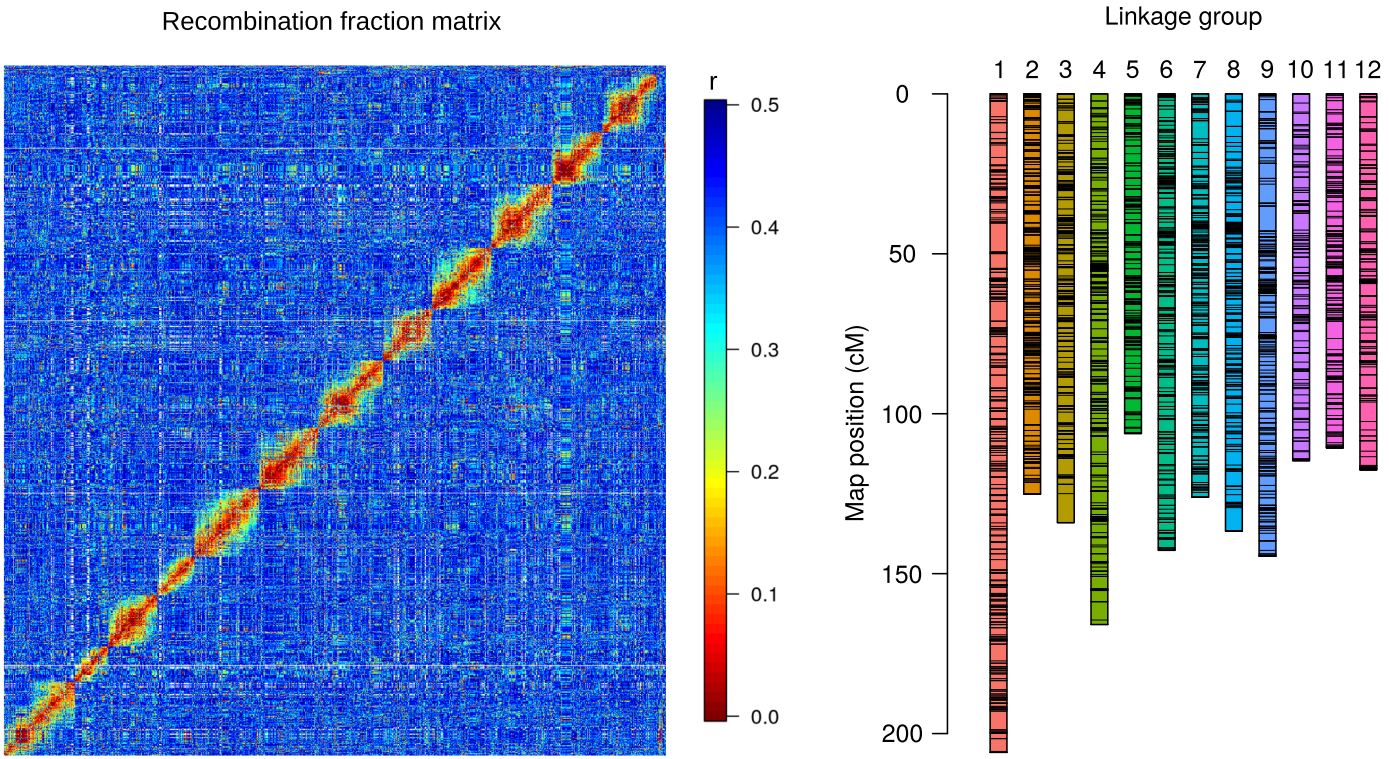
doi: https://doi.org/10.1101/2020.08.24.265397

This article is a preprint and has not been certified by peer review

Potato: 'Atlantic' × B1829-5

Genotypic data: 8,303 SNPs (Illumina chip array)

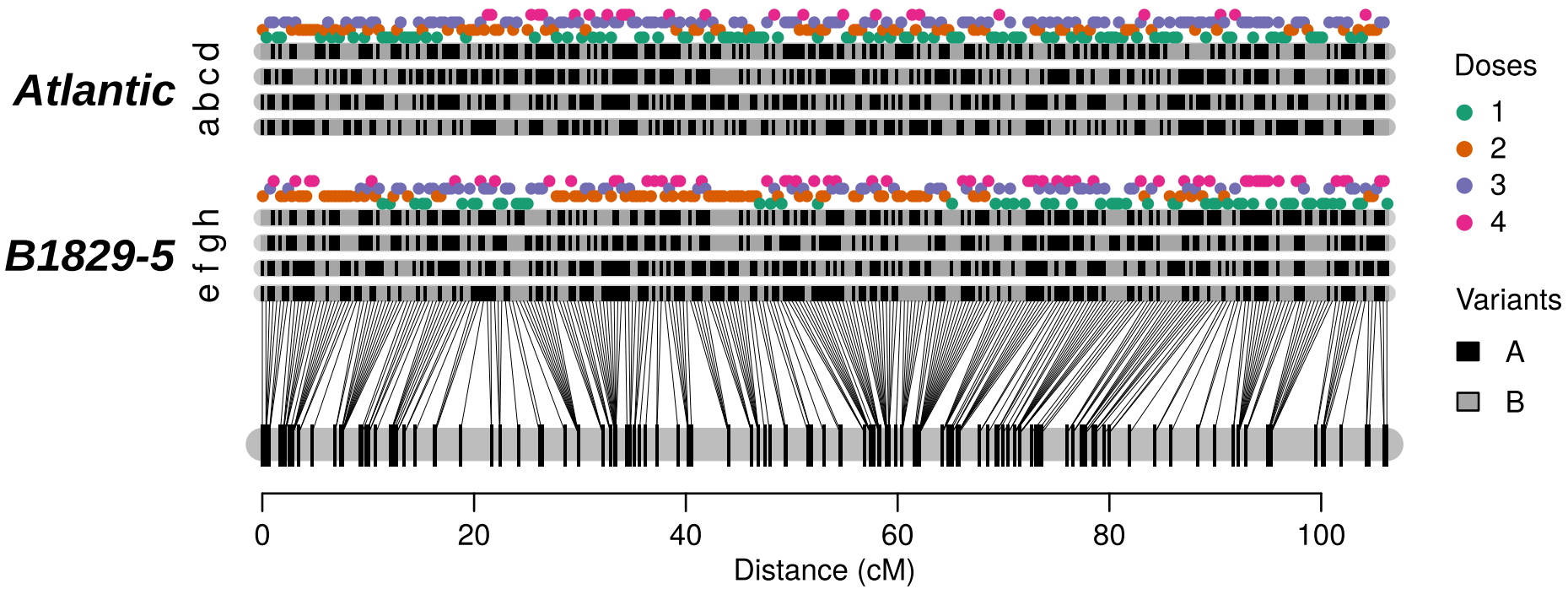
- 12 linkage groups: 4,285 SNPs (1,630 cM)
- Parental homologous haplotypes
- Inheritance of chromosomal segments from parents to progeny



Potato: 'Atlantic' × B1829-5

Genotypic data: 8,303 SNPs (Illumina chip array)

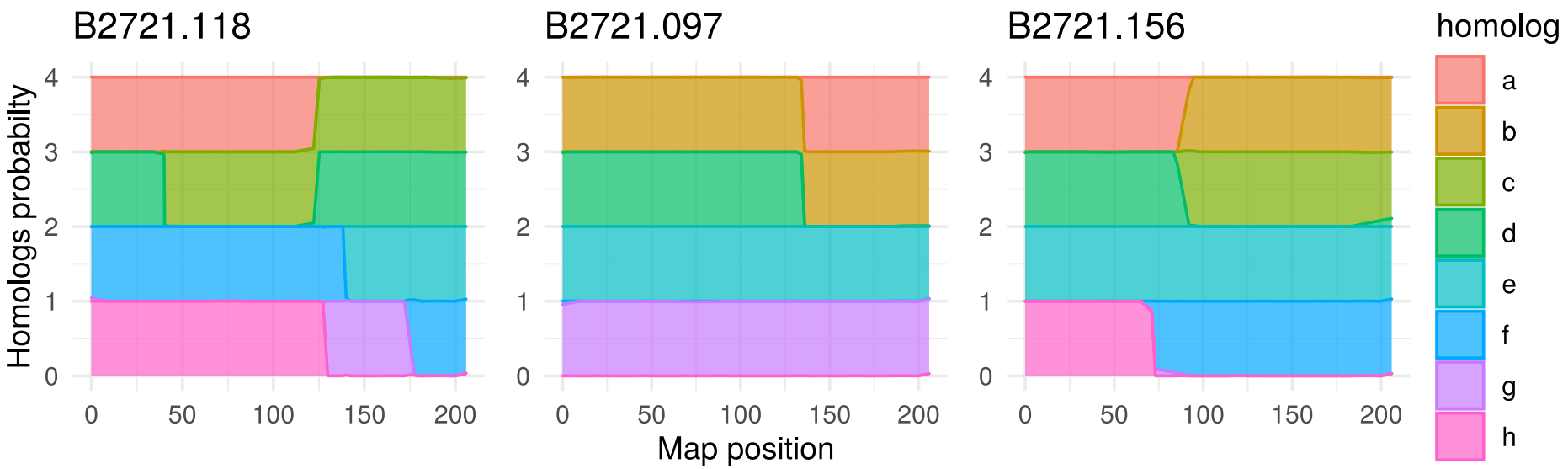
- 12 linkage groups: 4,285 SNPs (1,630 cM)
- Parental homologous haplotypes
- Inheritance of chromosomal segments from parents to progeny



Potato: 'Atlantic' × B1829-5

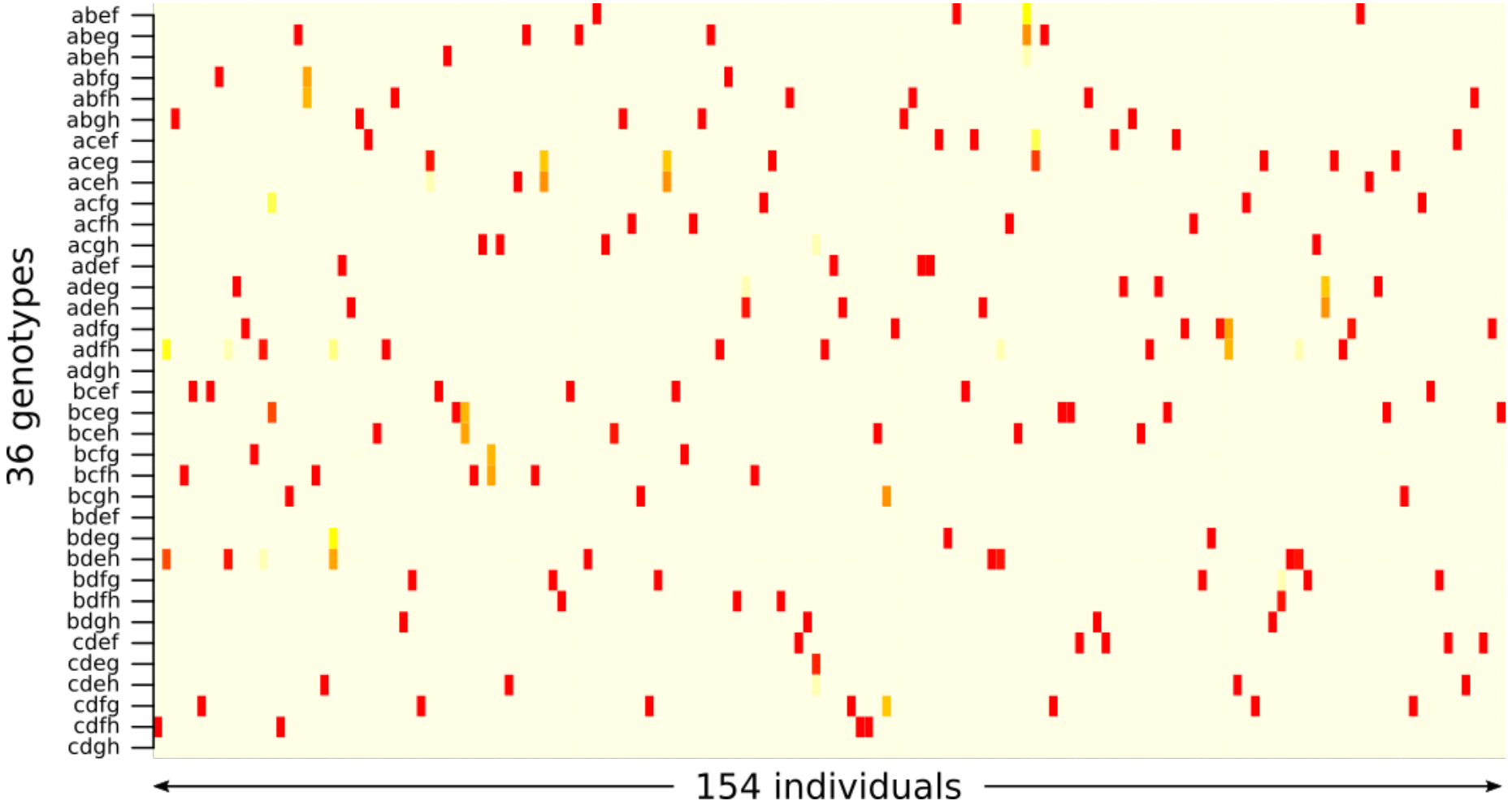
Genotypic data: 8,303 SNPs (Illumina chip array)

- 12 linkage groups: 4,285 SNPs (1,630 cM)
- Parental homologous haplotypes
- Inheritance of chromosomal segments from parents to progeny



Potato: ‘Atlantic’ × B1829-5

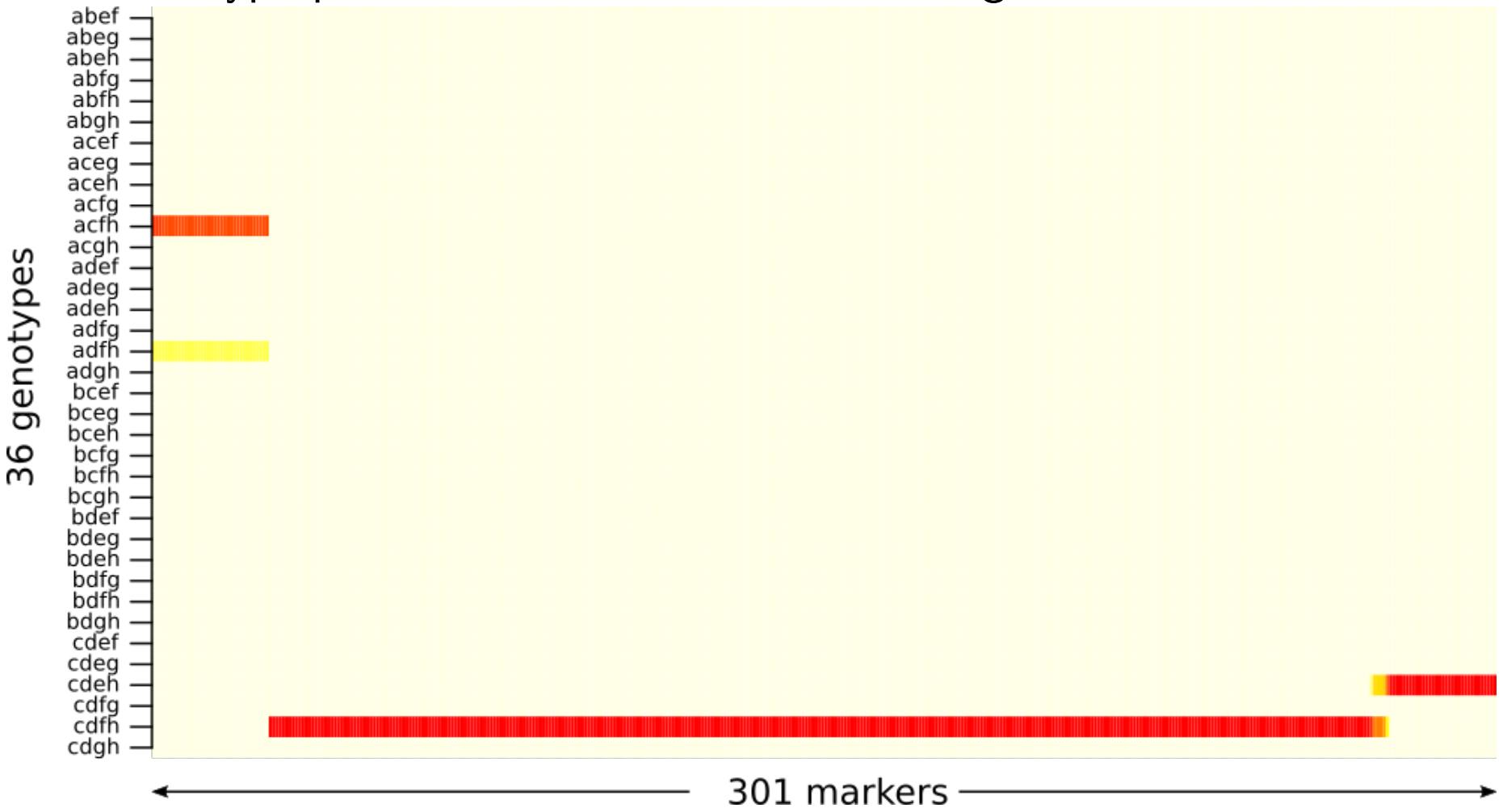
Genotype probabilities for all individuals on a marker



Individuals sharing the same alleles are more related

Potato: 'Atlantic' × B1829-5

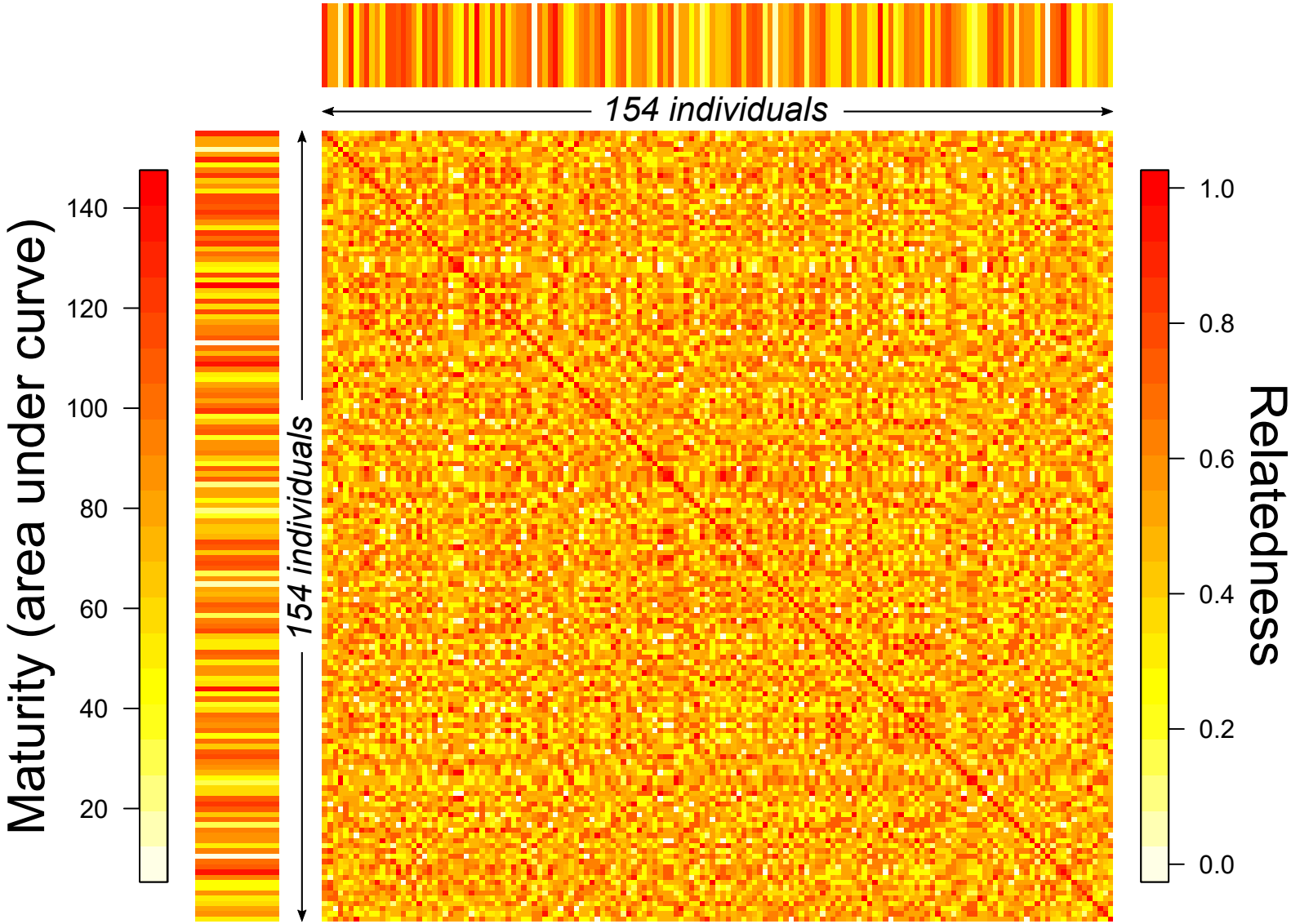
Genotype probabilities of an individual along the chromosome 5



Differences between individuals are due to recombination and segregation

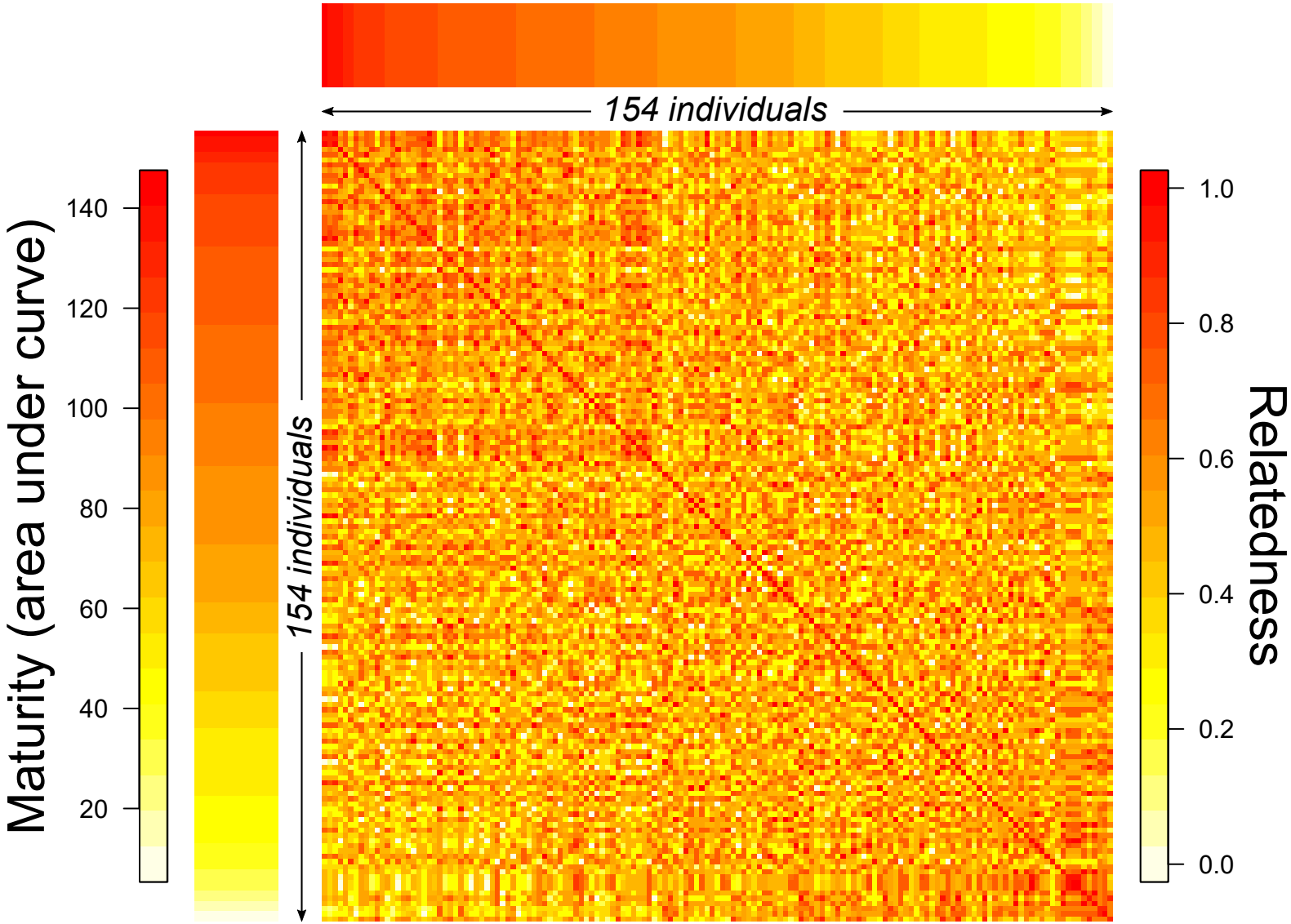
Potato: ‘Atlantic’ × B1829-5

- Unordered trait values

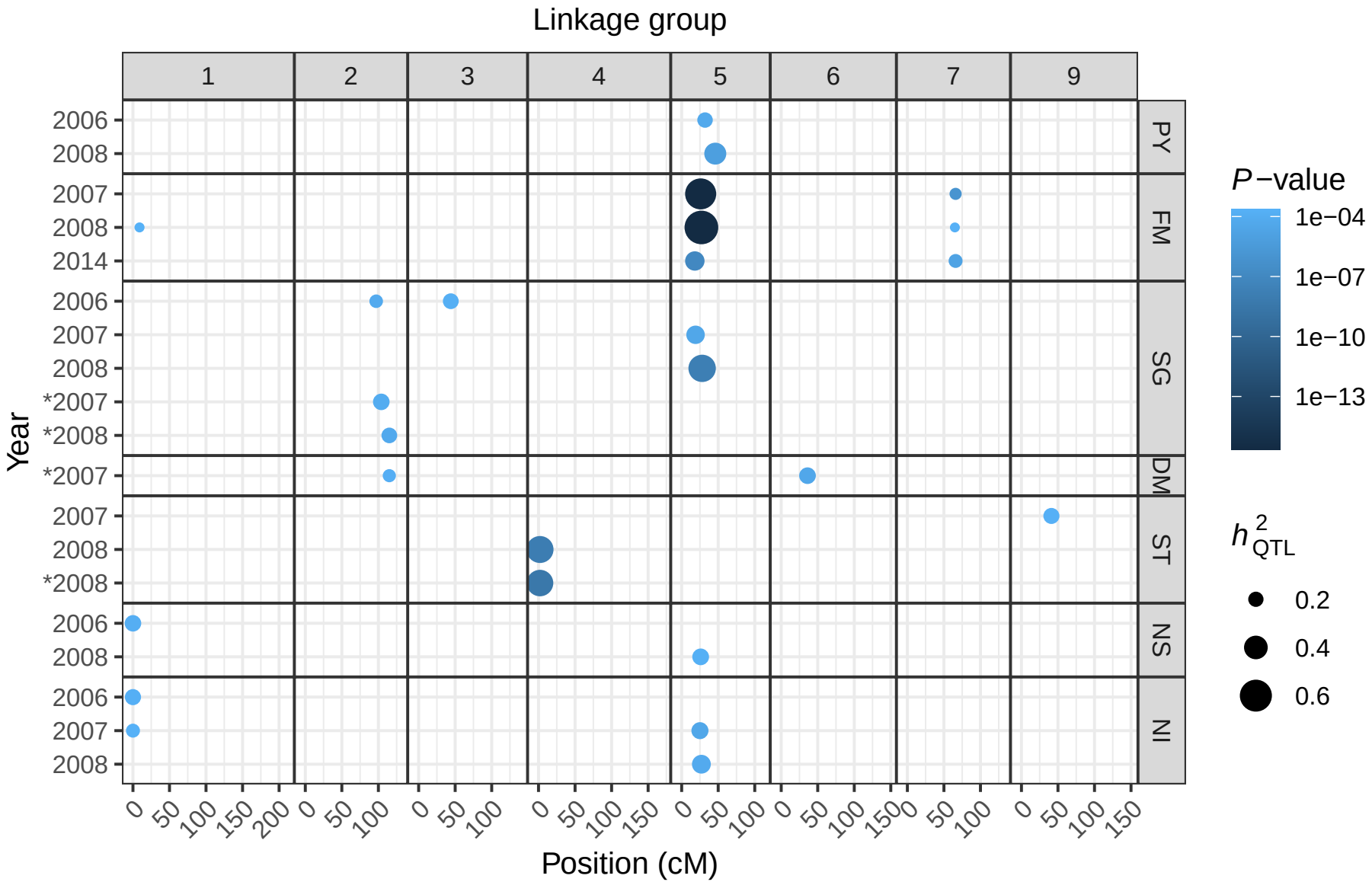


Potato: 'Atlantic' × B1829-5

- Ordered trait values



Potato: ‘Atlantic’ × B1829-5



Outline

1. Multiple QTL mapping
   - Components of variance
   - Random-effect QTL model
2. Applications
   - Potato: 'Atlantic' $\times$ B1829-5
   - Sweetpotato: 'Beauregard' $\times$ 'Tanzania'
3. Final considerations

Sweetpotato: Beauregard × Tanzania ($N = 315$)

- *Ipomoea batatas* ($2n = 6x = 90$, $x = 15$)

Phenotypic data

- 3 sites, 2 years
- 4 quality traits + 1
  1. DM – Dry Matter (%)
  2. Starch (%)
  3. BC – β-carotene (%)
  4. FC_P – Flesh color in Peru
  5. FC_U – Flesh color in Uganda

Genotypic data

- 15 linkage groups with 30,684 SNPs (2708.3 cM)

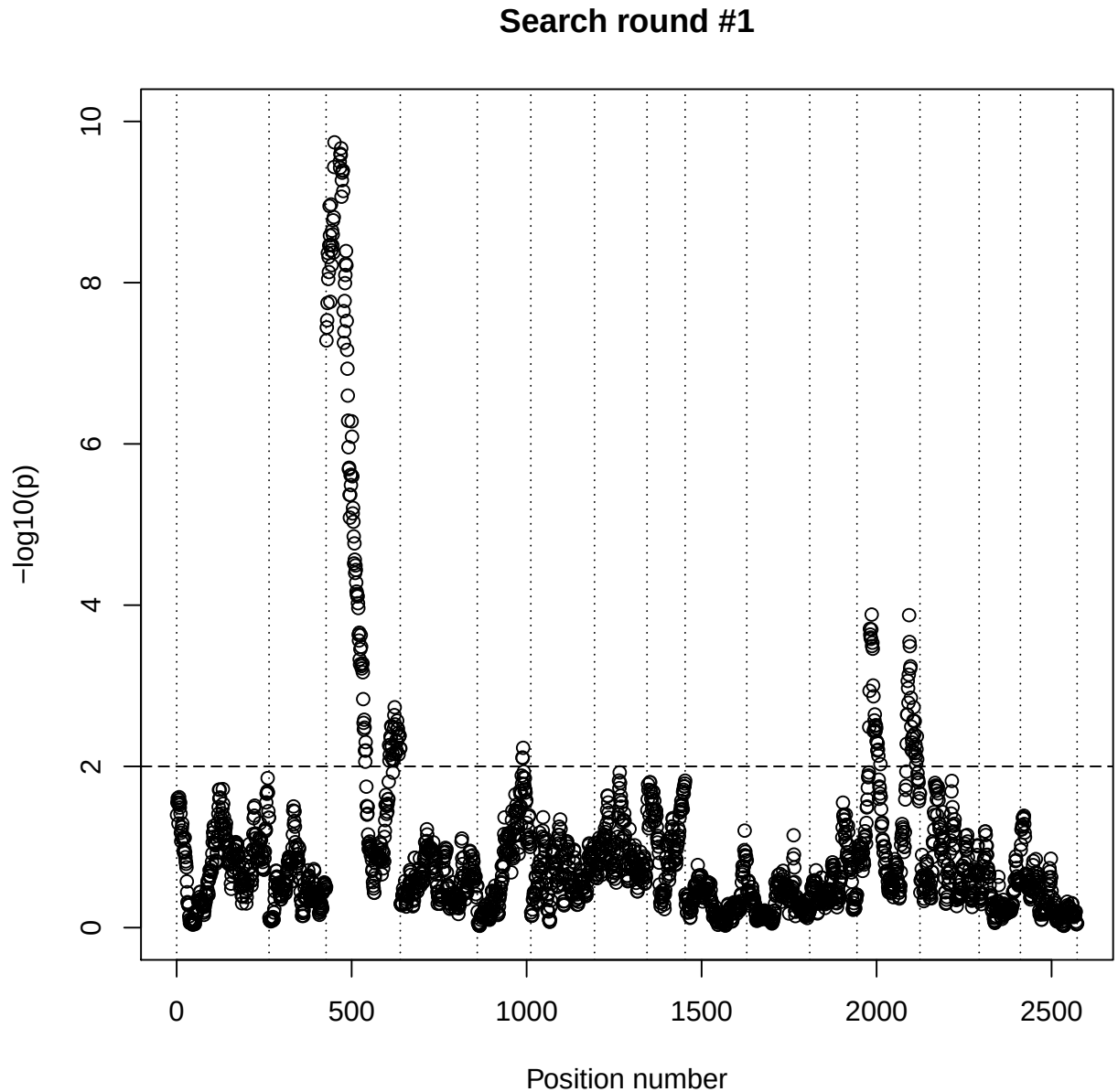
Theoretical and Applied Genetics (2020) 133:23–36
https://doi.org/10.1007/s00122-019-03437-7

ORIGINAL ARTICLE

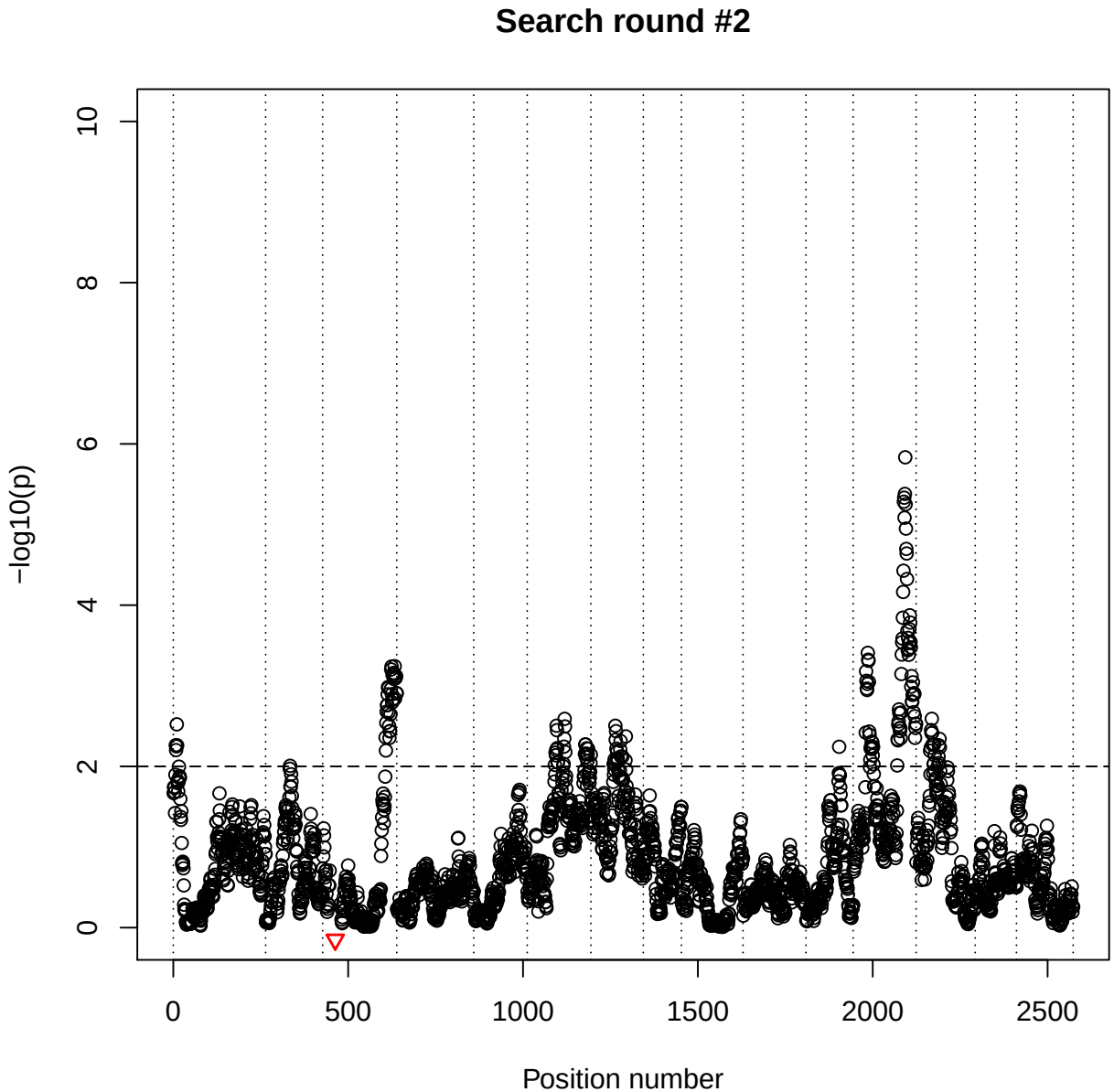
Quantitative trait loci and differential gene expression analyses reveal the genetic basis for negatively associated β-carotene and starch content in hexaploid sweetpotato [*Ipomoea batatas* (L.) Lam.]

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

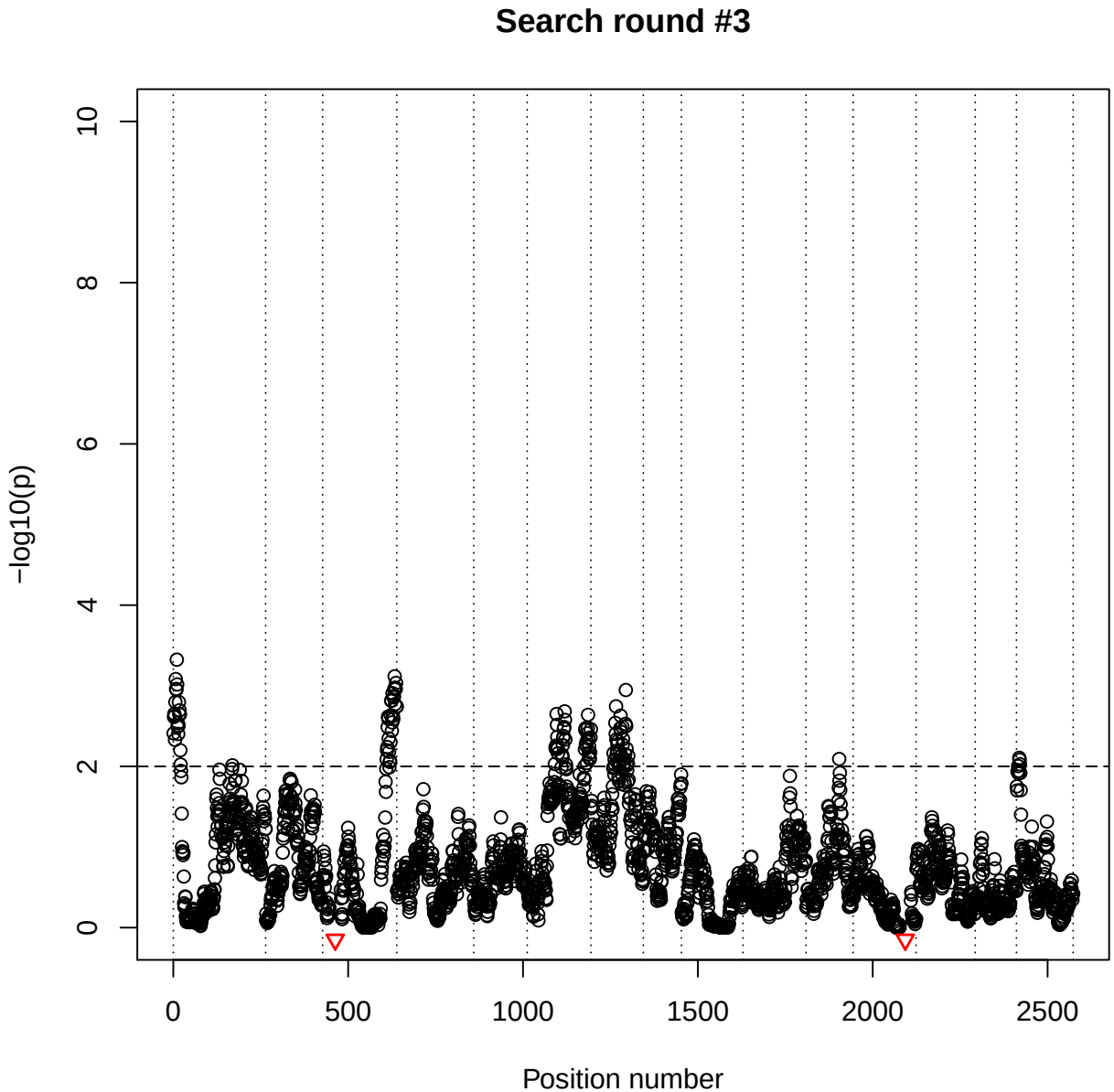
Dry matter – Forward search



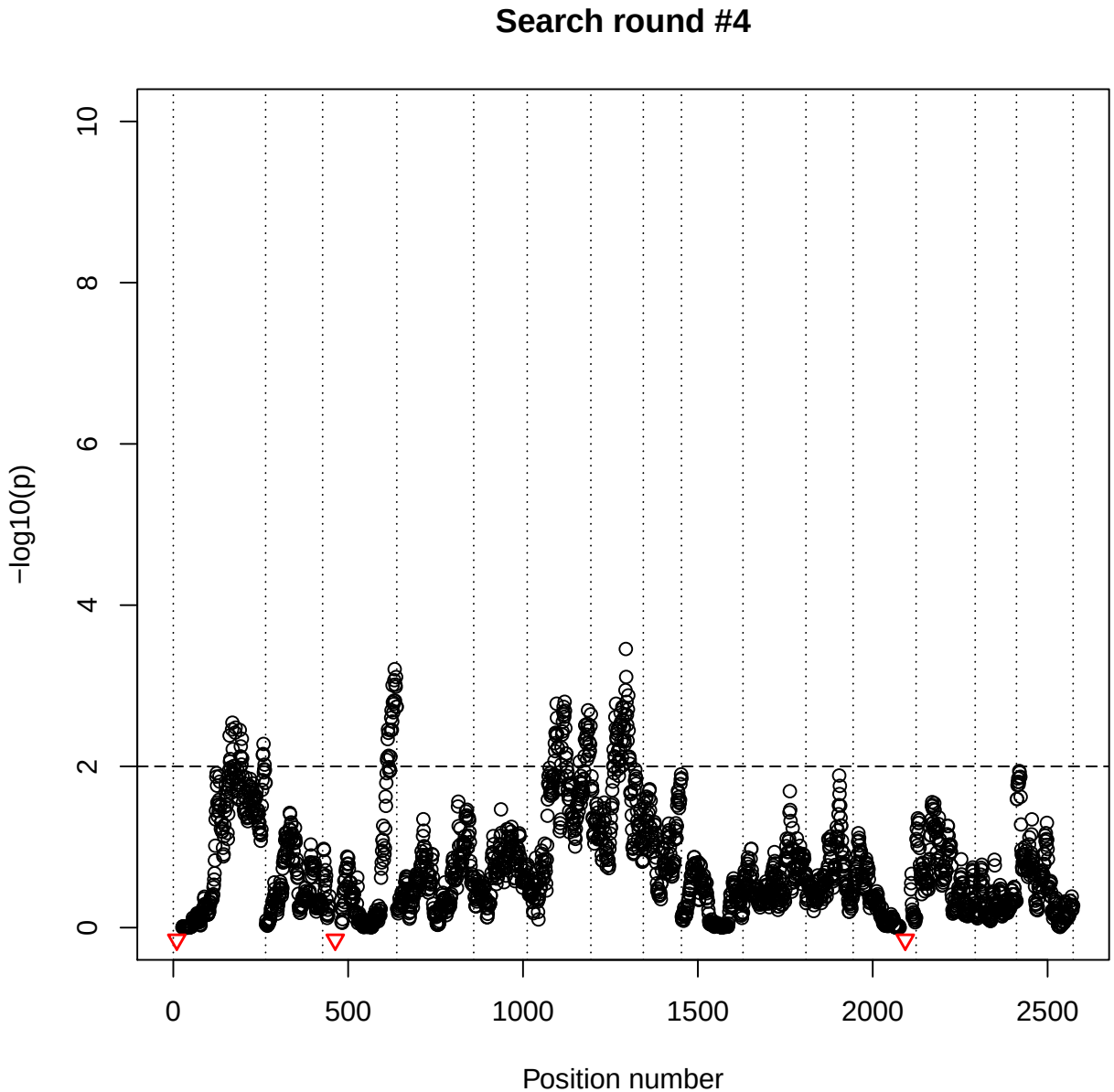
Dry matter – Forward search



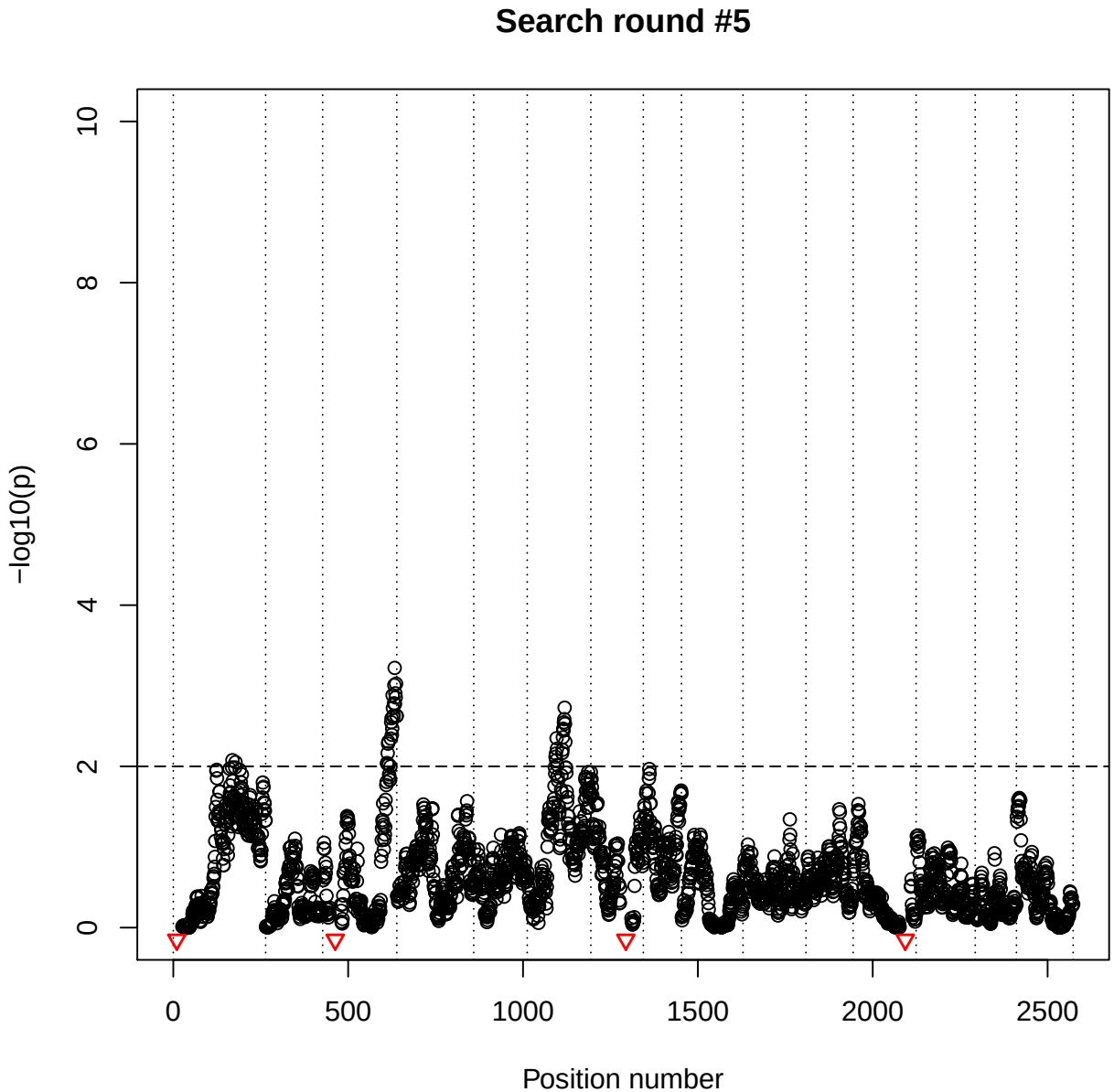
Dry matter – Forward search



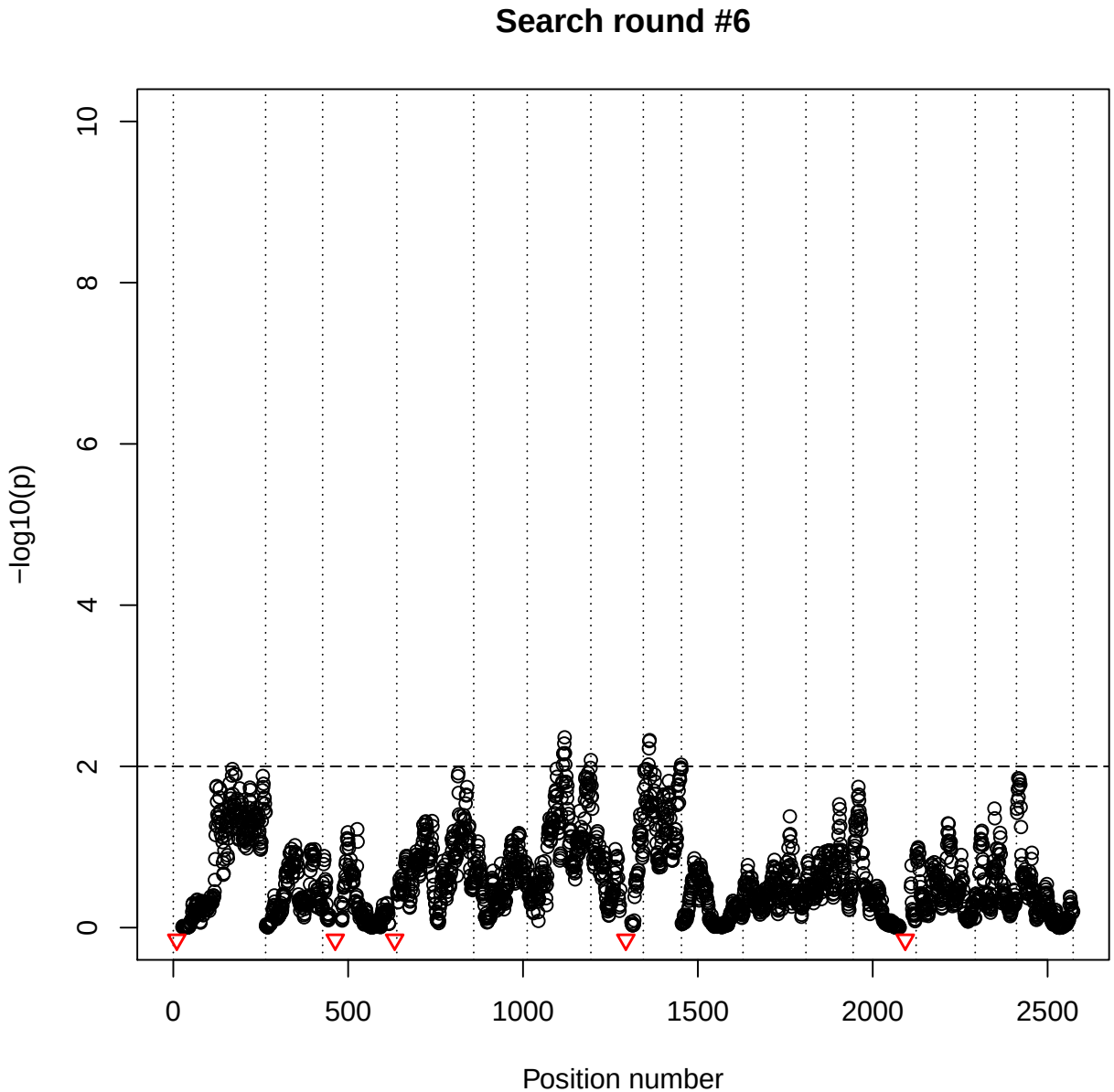
Dry matter – Forward search



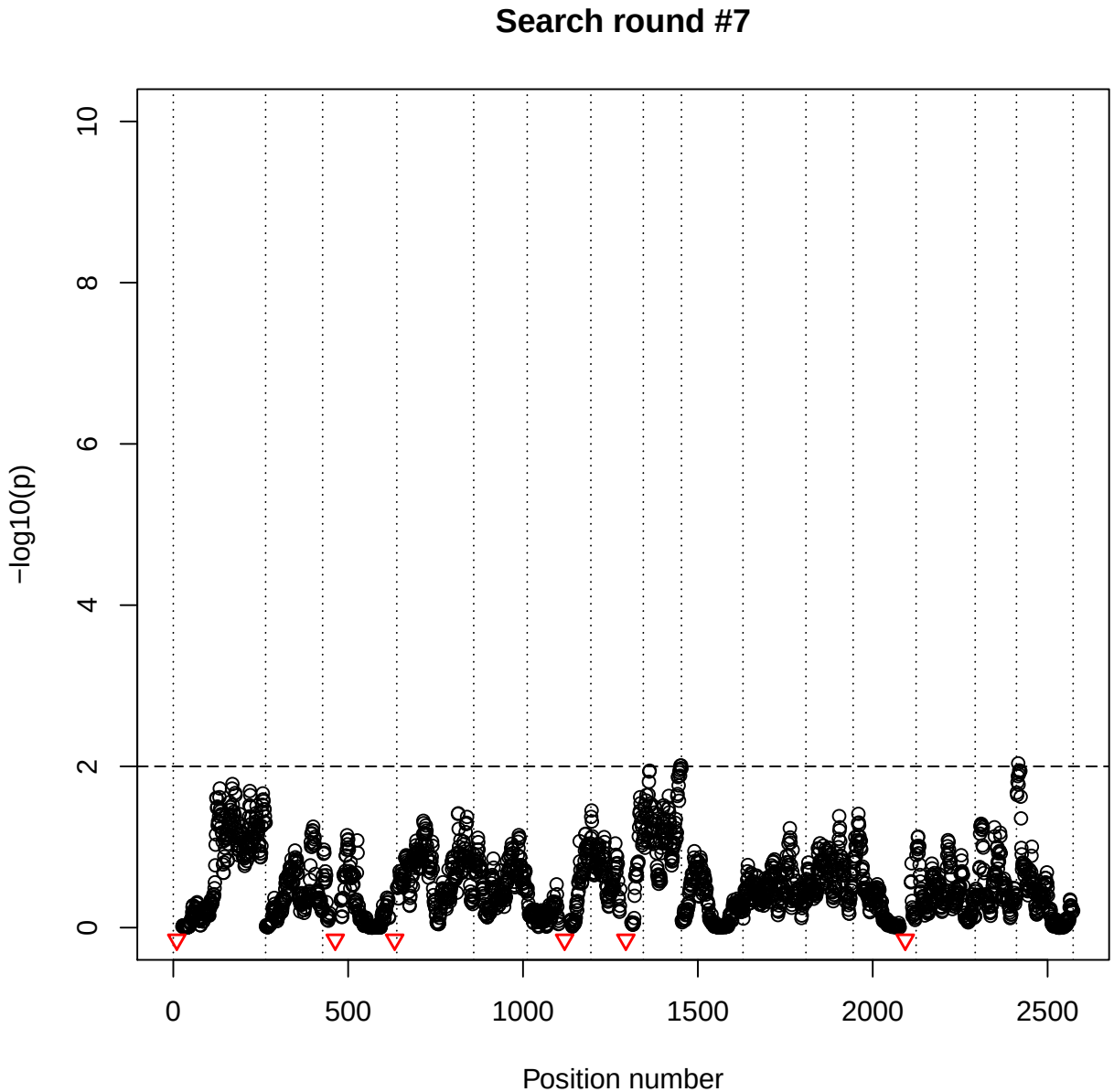
Dry matter – Forward search



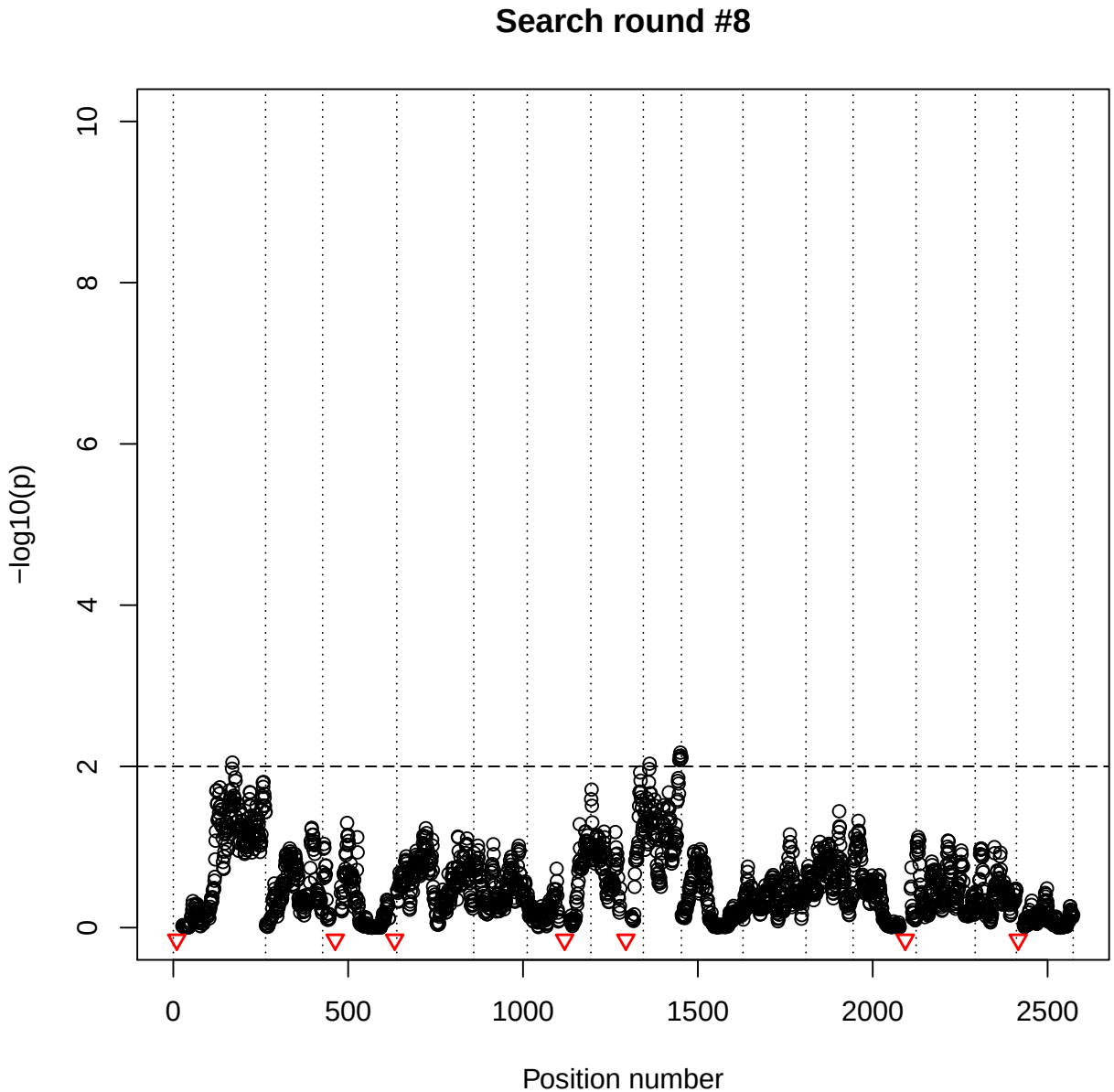
Dry matter – Forward search



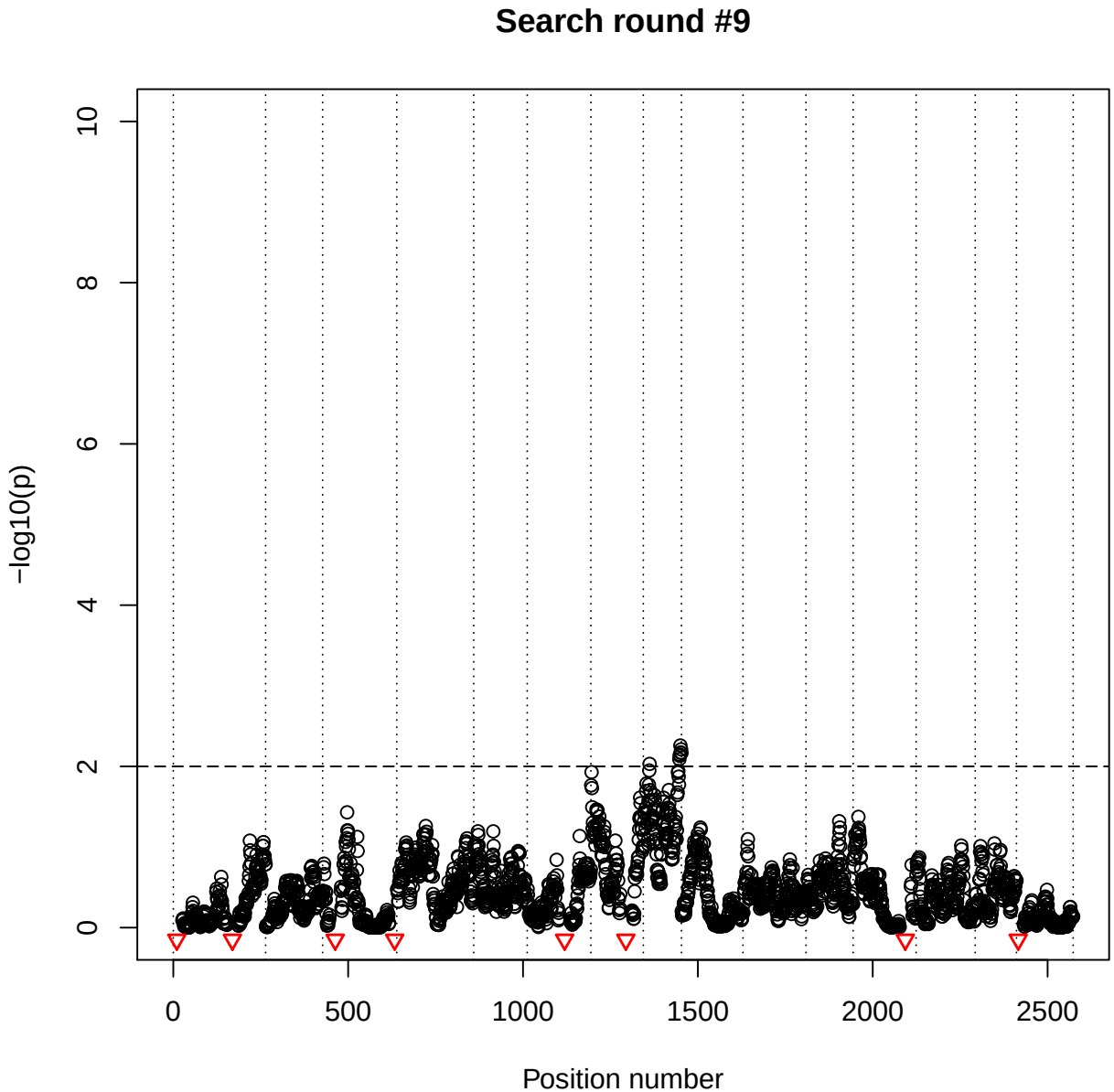
Dry matter – Forward search



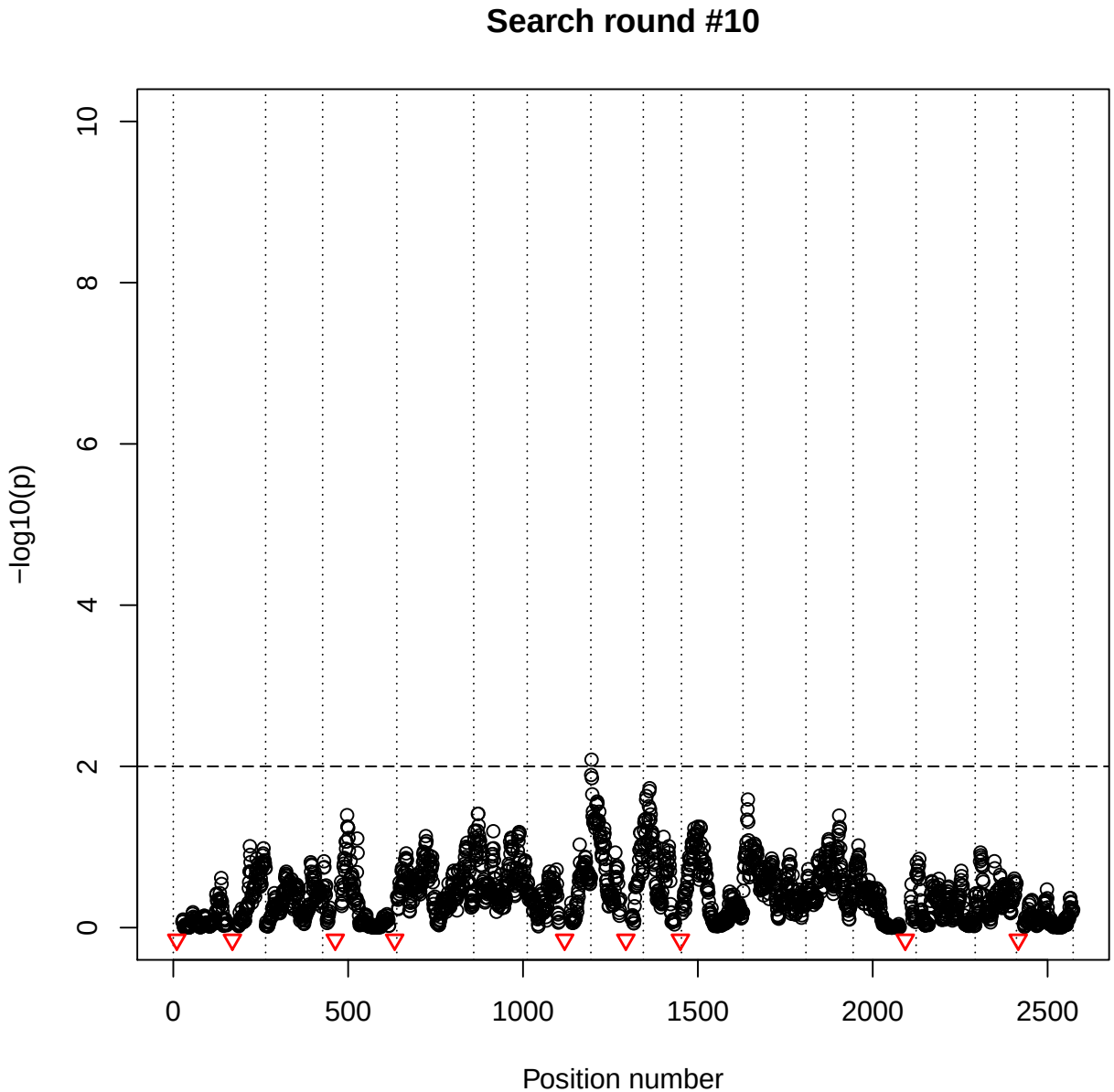
Dry matter – Forward search



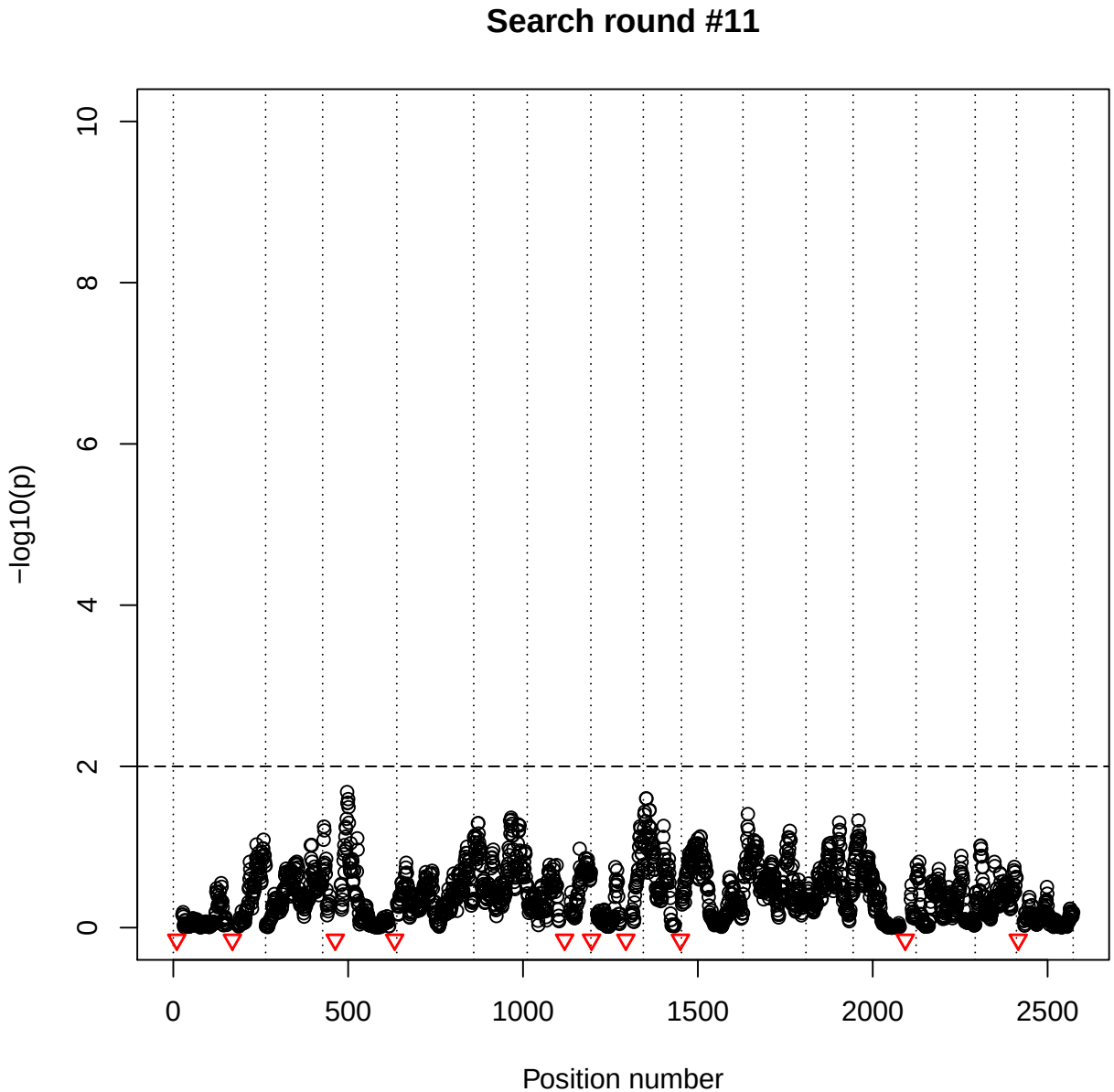
Dry matter – Forward search



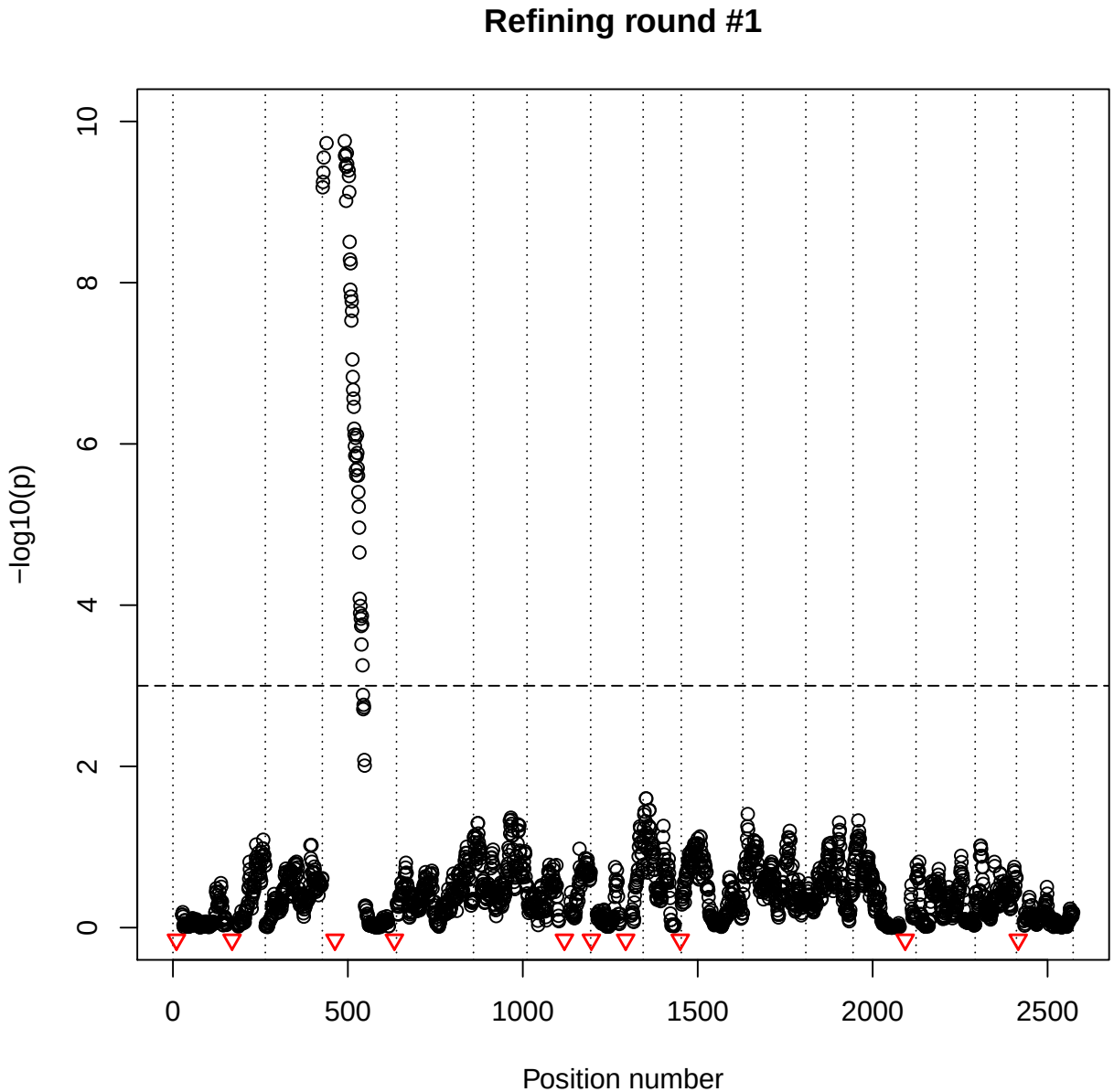
Dry matter – Forward search



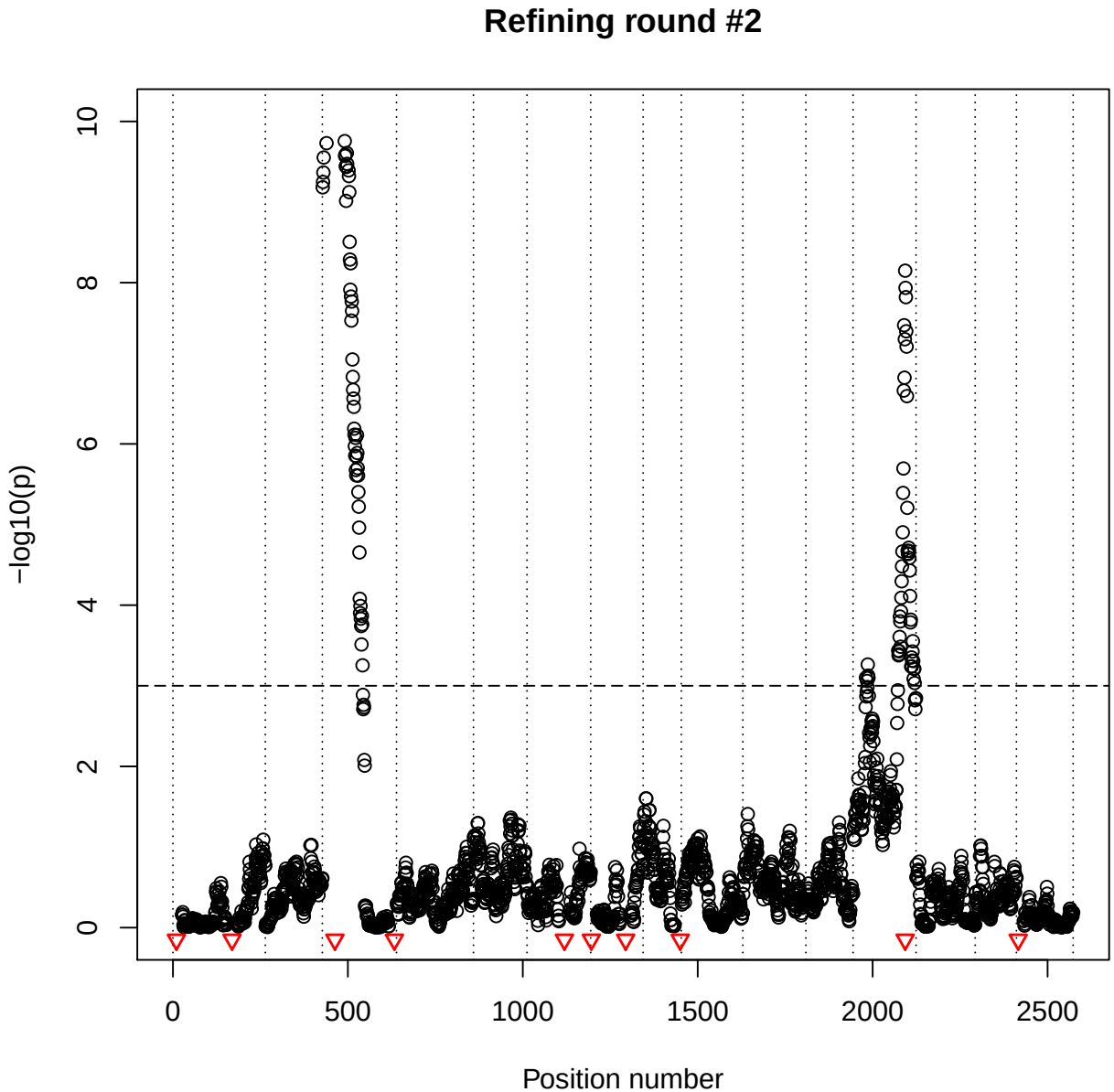
Dry matter – Forward search



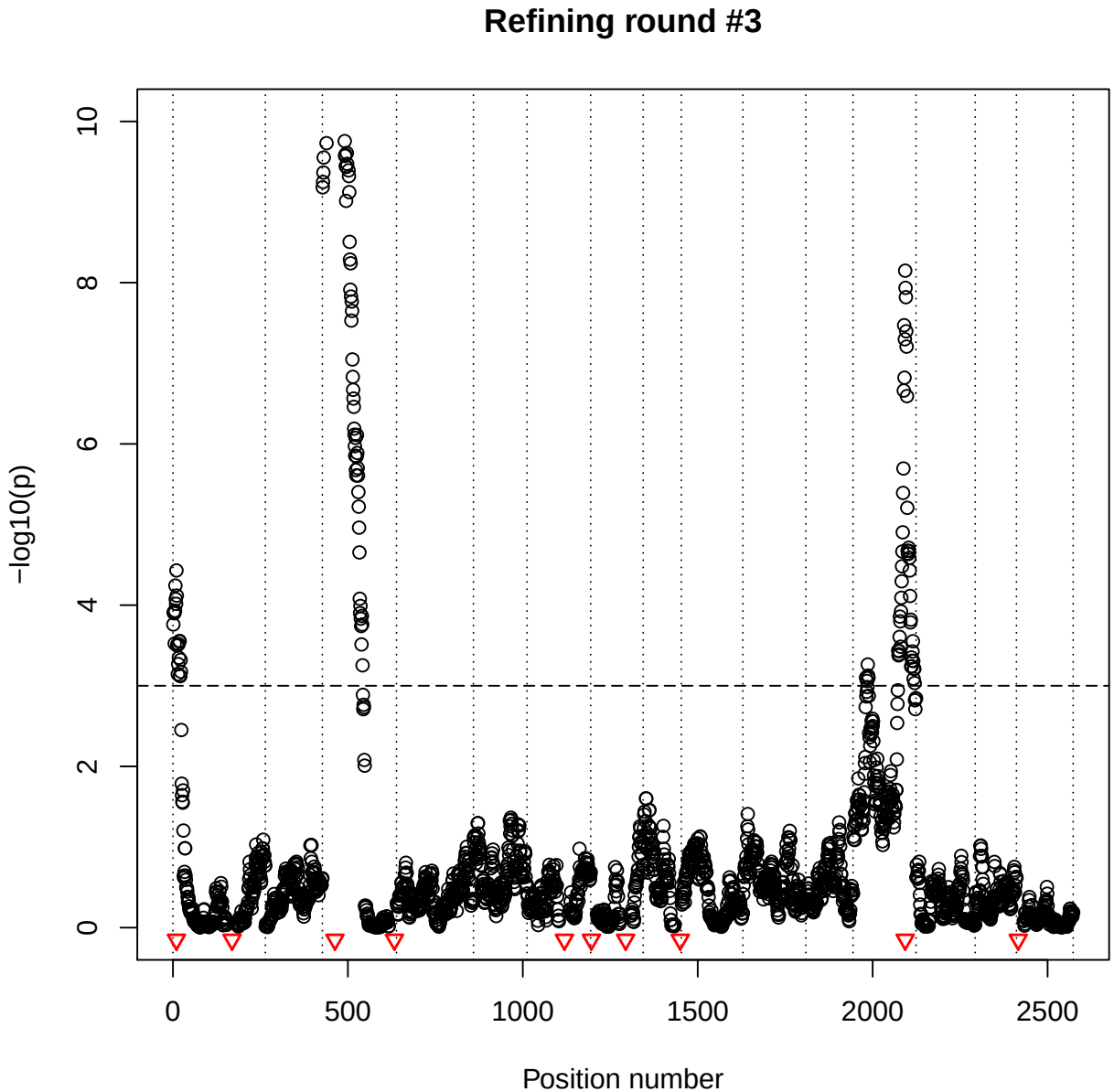
Dry matter – Model optimization



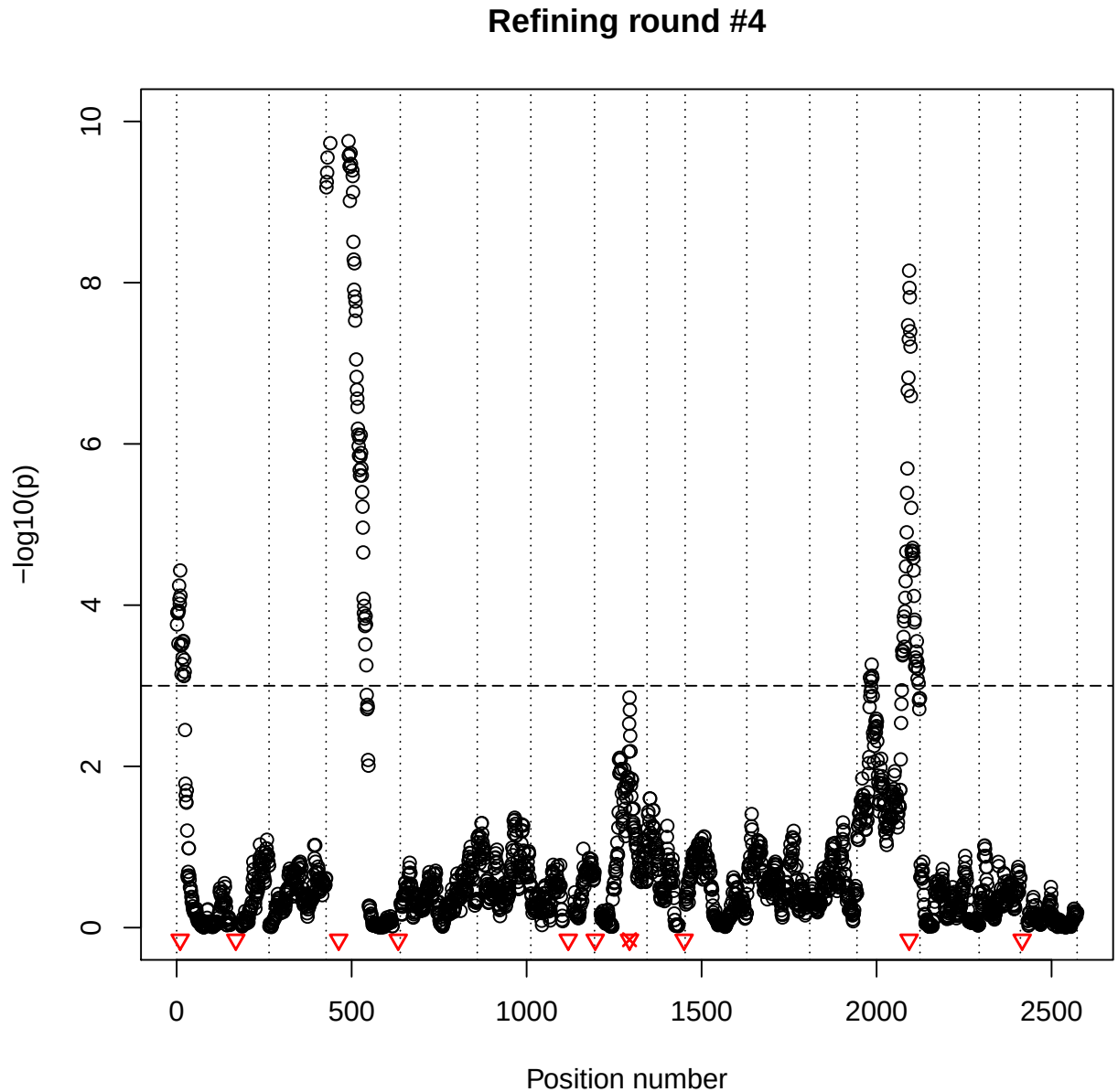
Dry matter – Model optimization



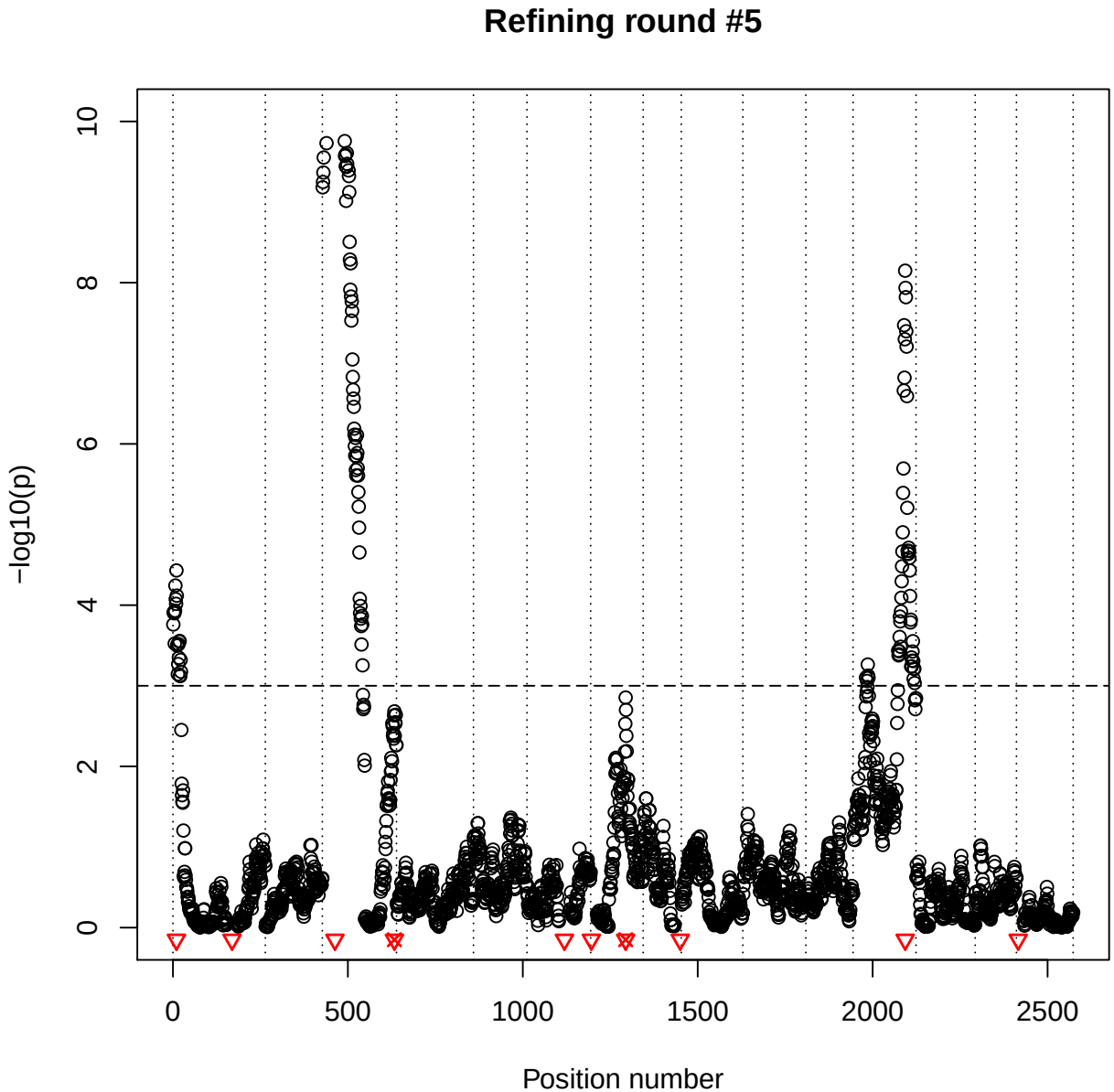
Dry matter – Model optimization



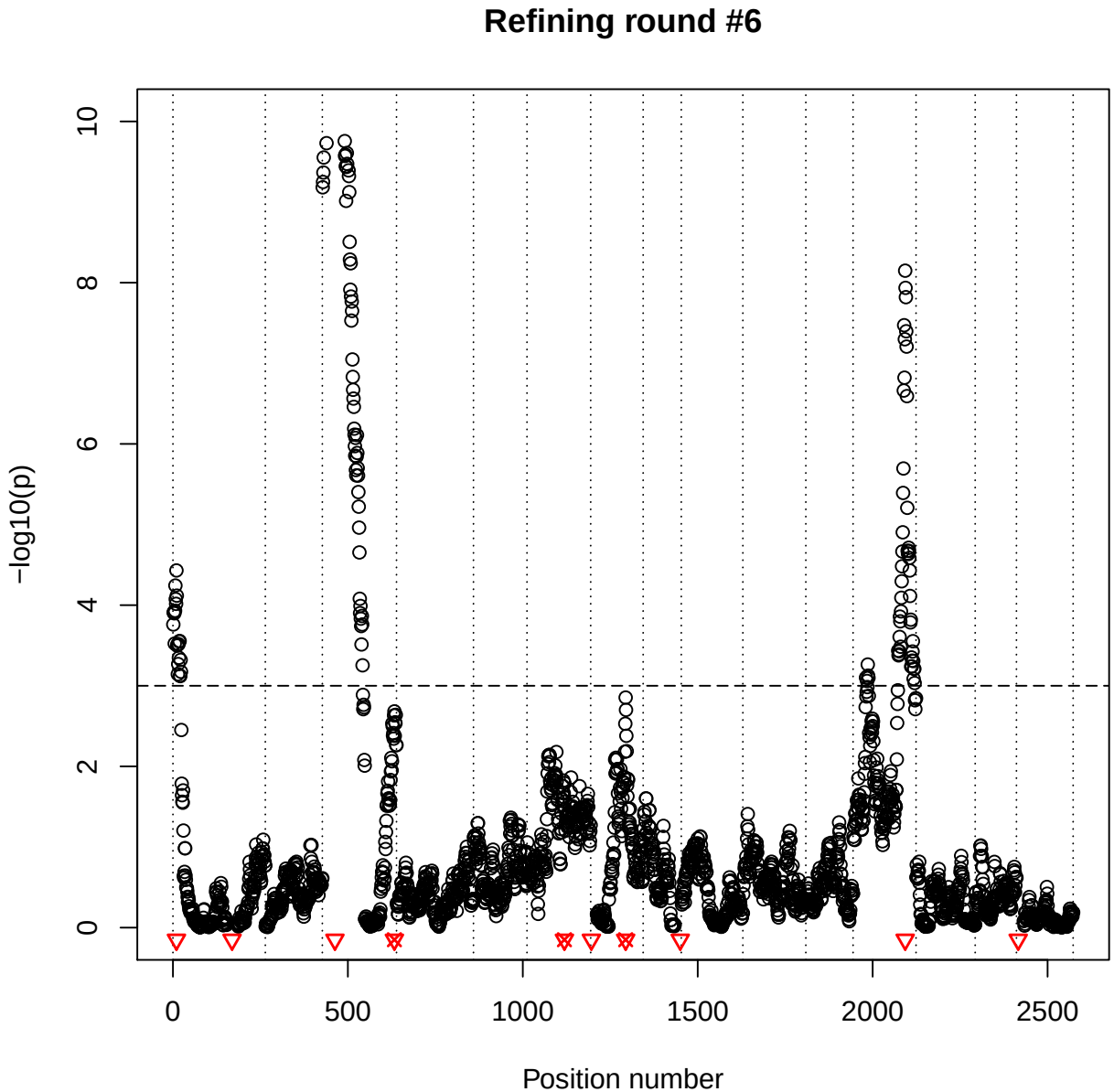
Dry matter – Model optimization



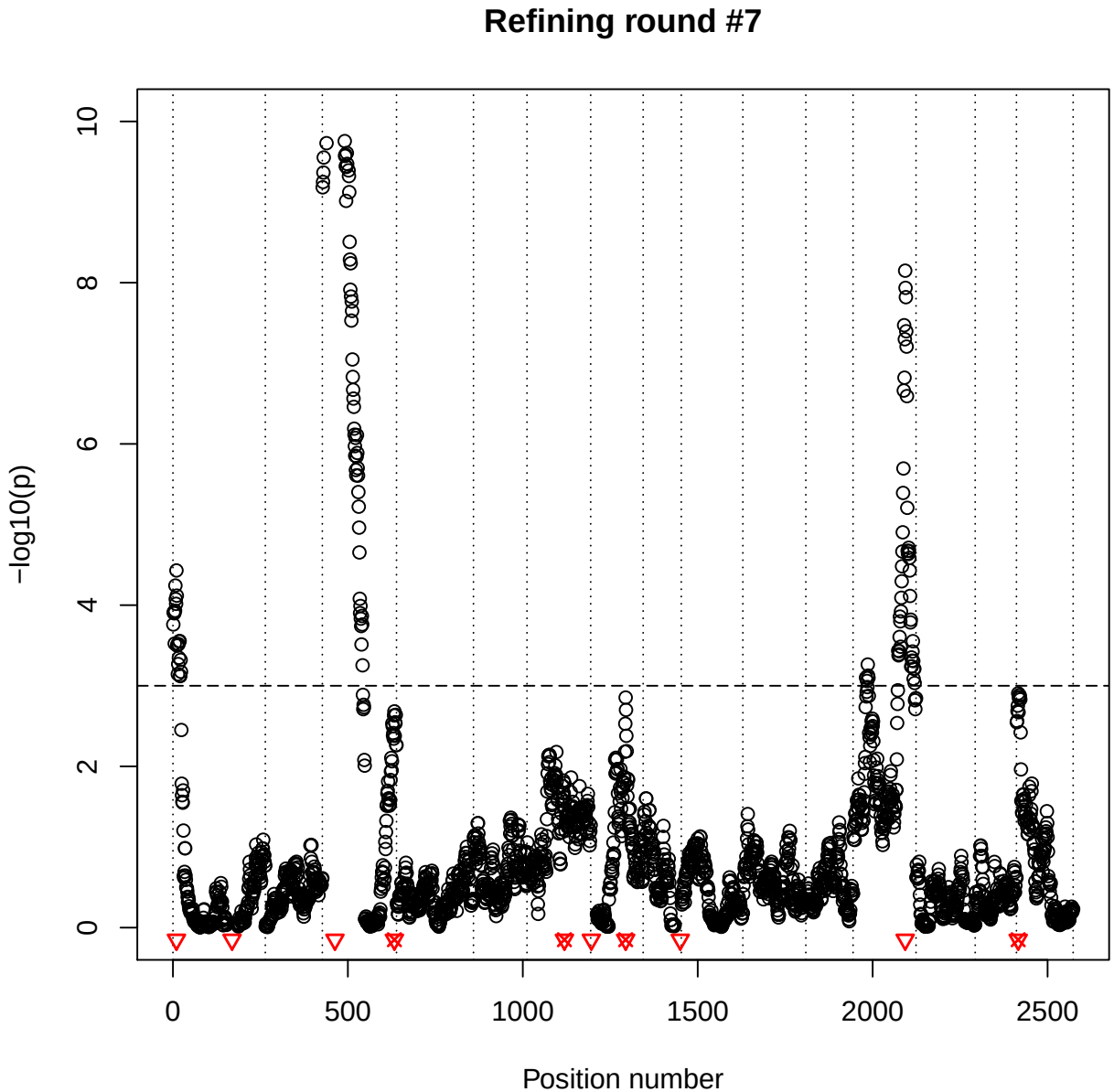
Dry matter – Model optimization



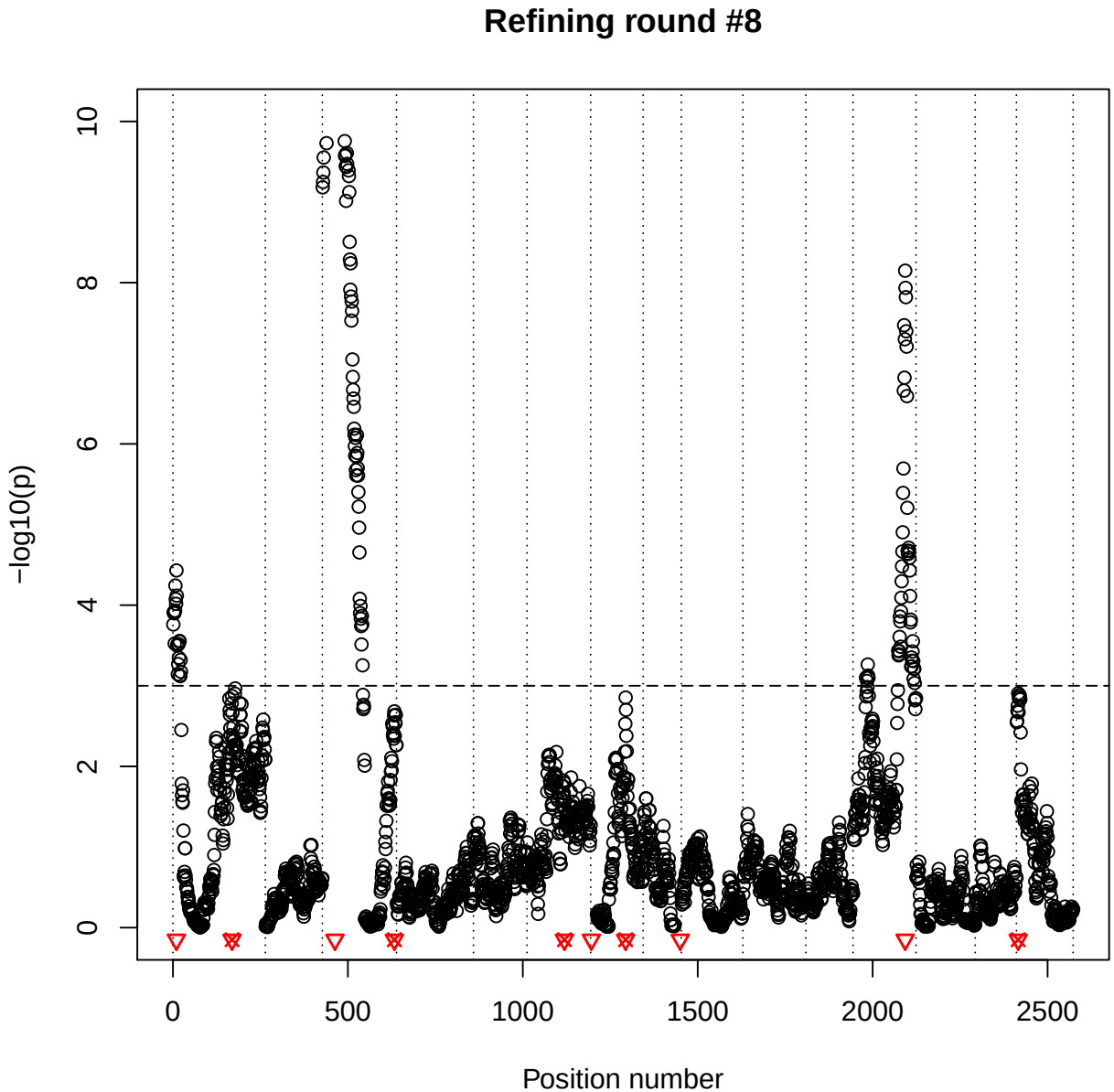
Dry matter – Model optimization



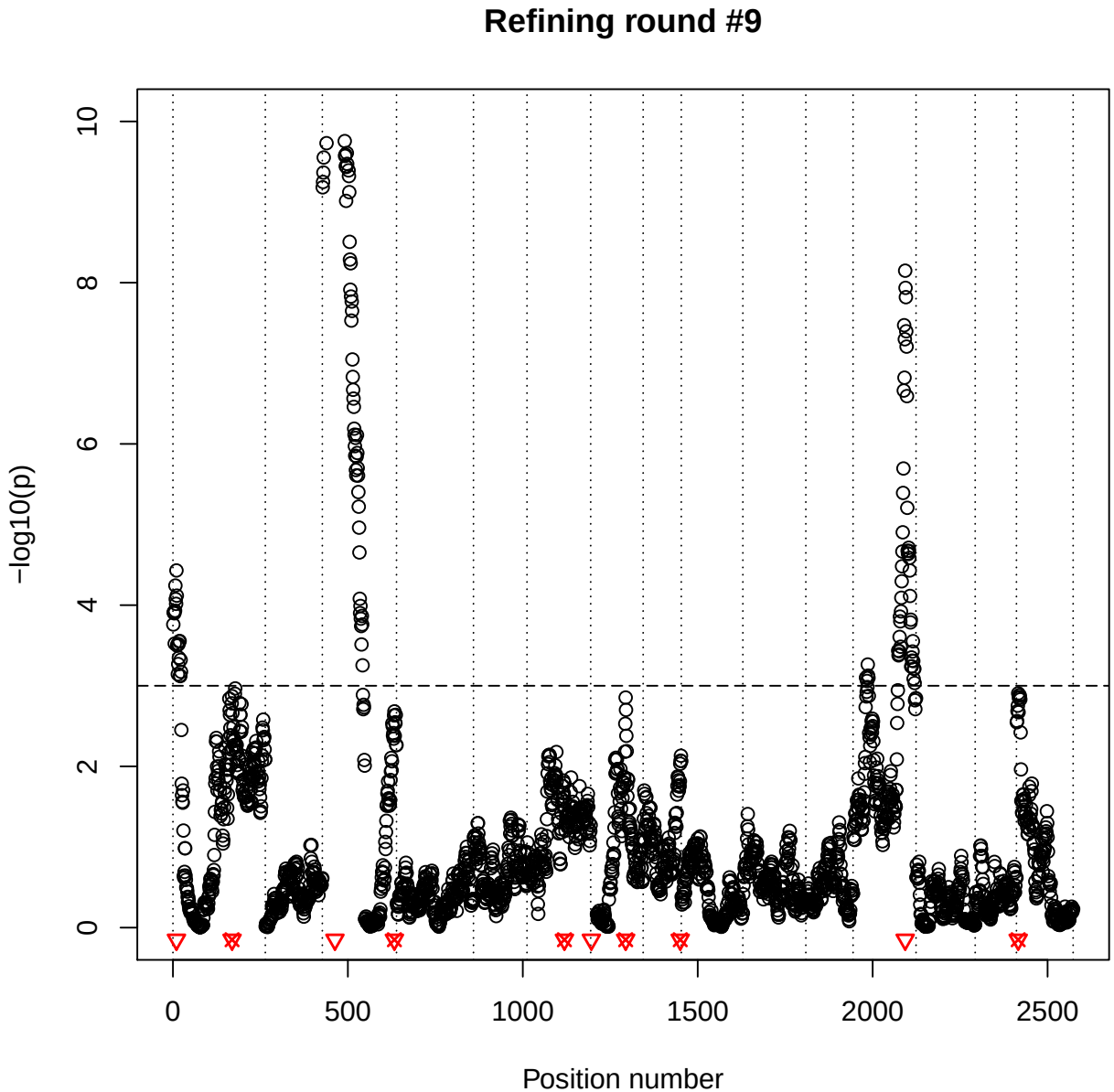
Dry matter – Model optimization



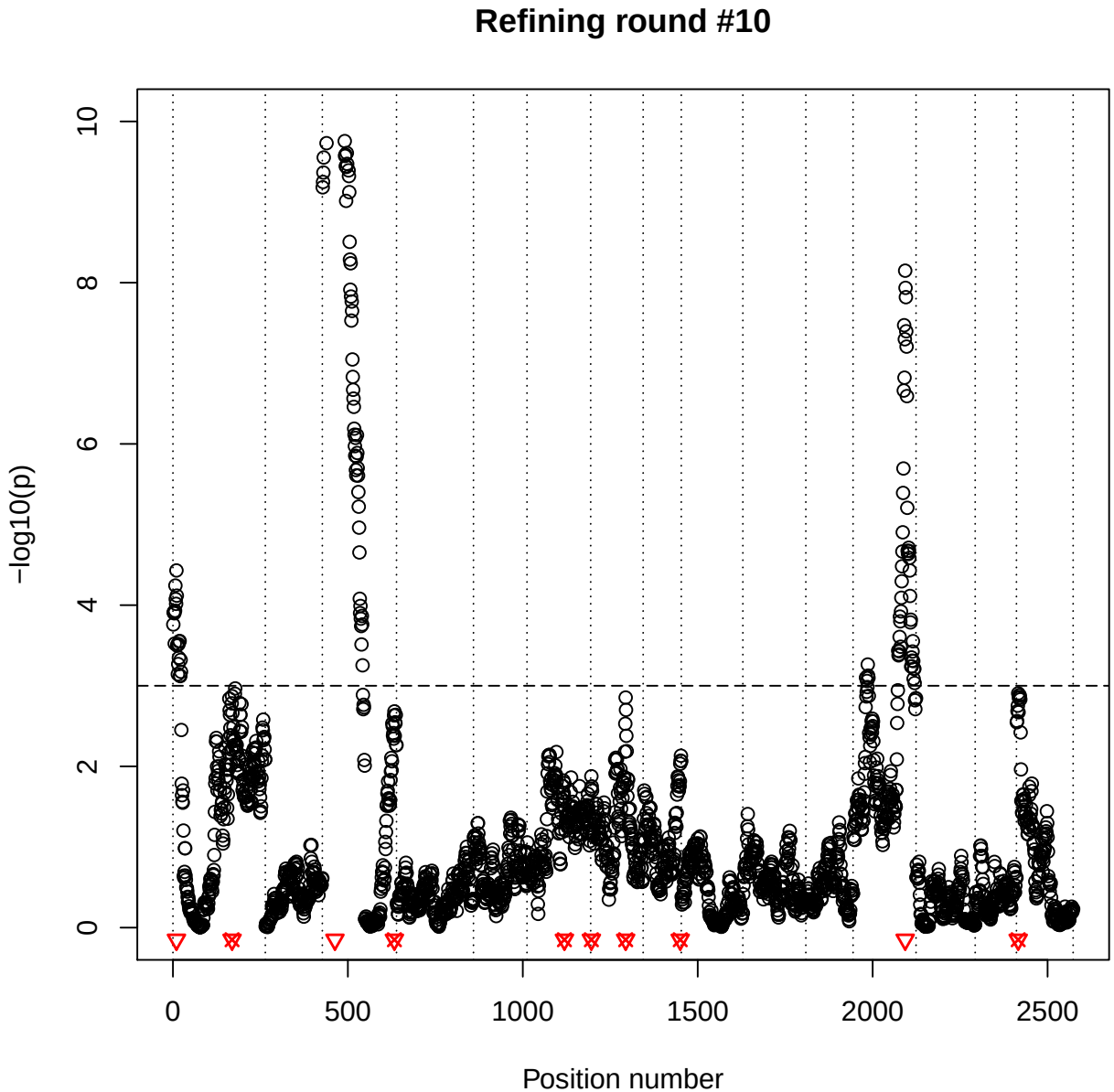
Dry matter – Model optimization



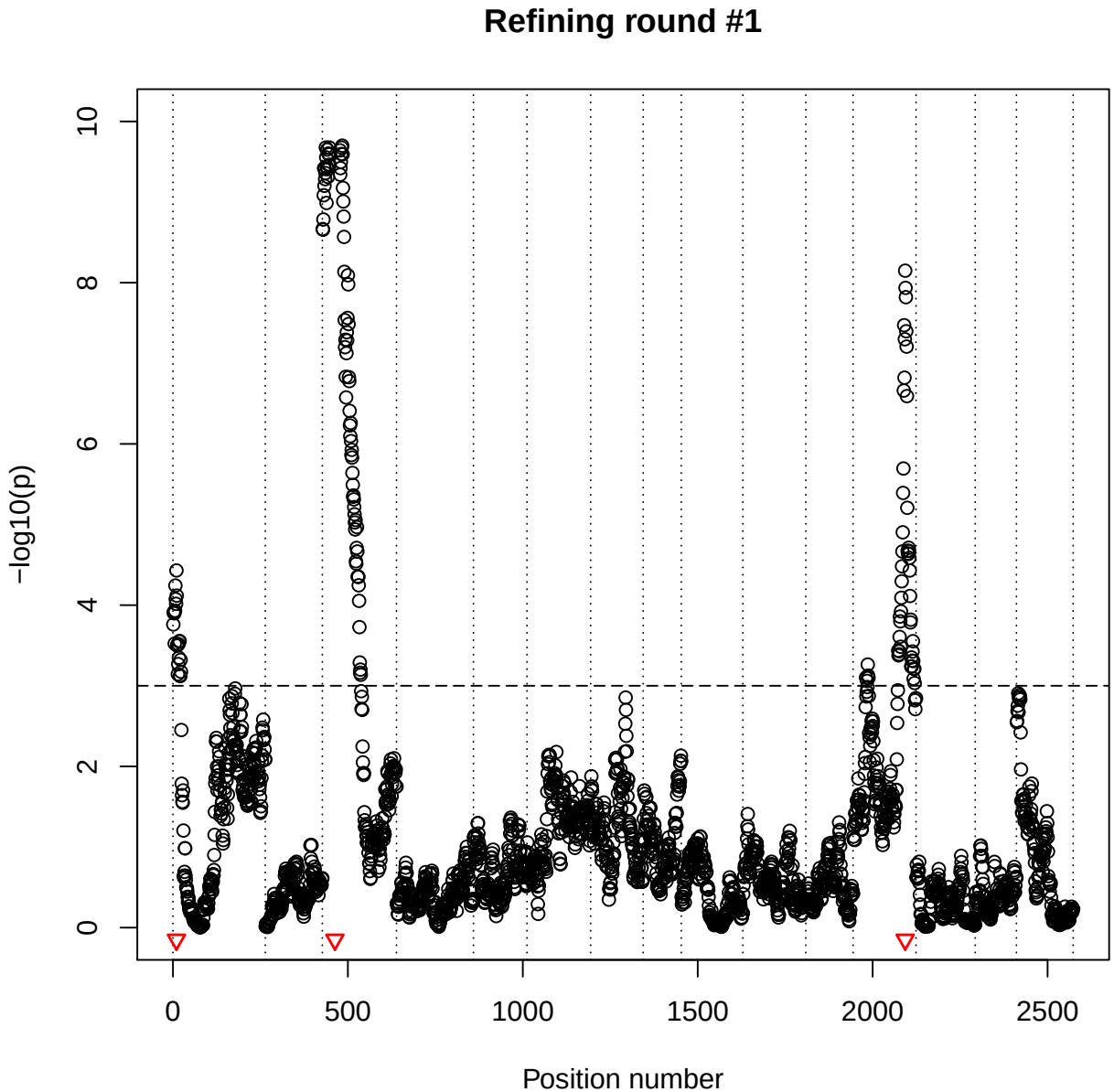
Dry matter – Model optimization



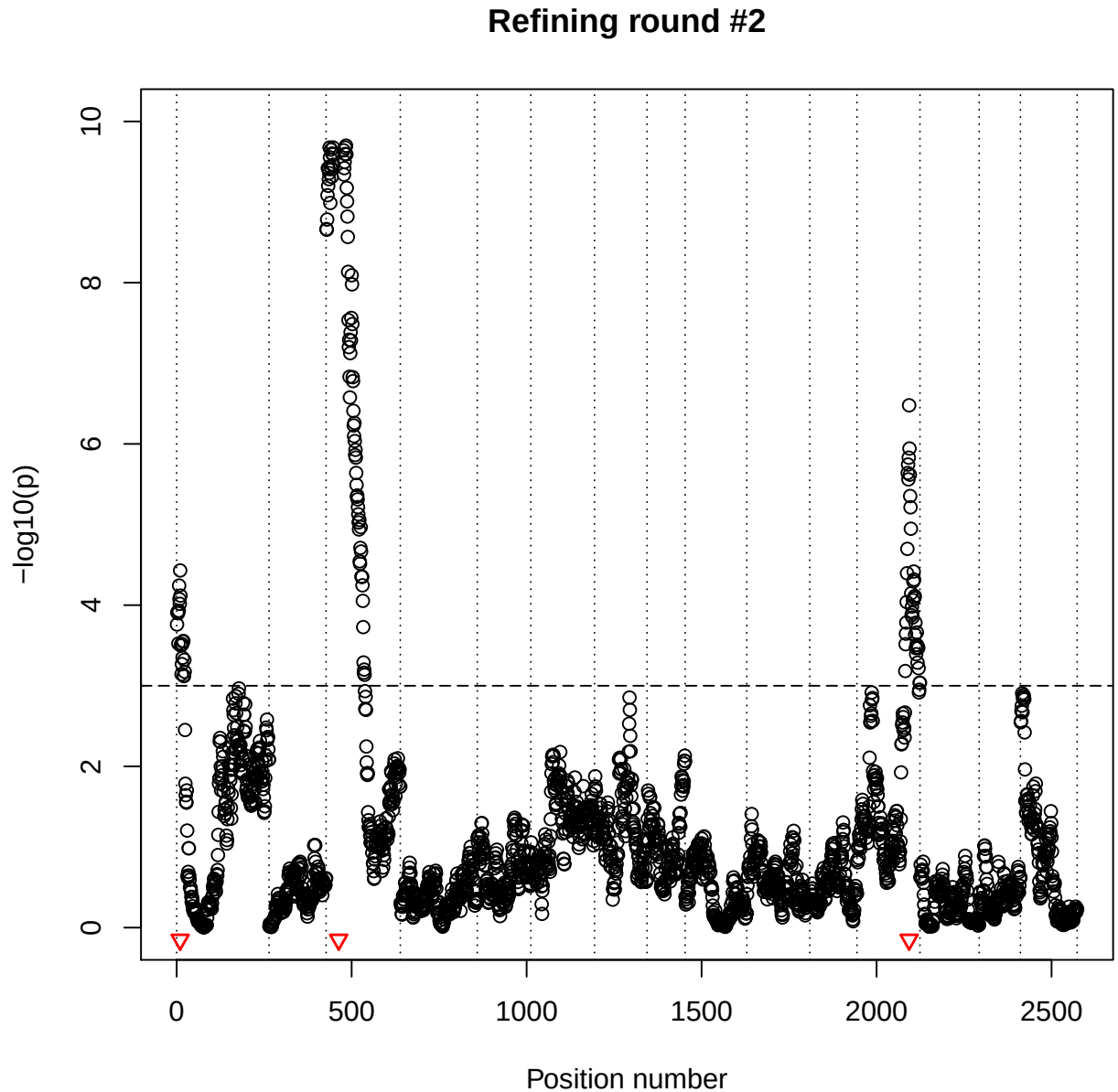
Dry matter – Model optimization



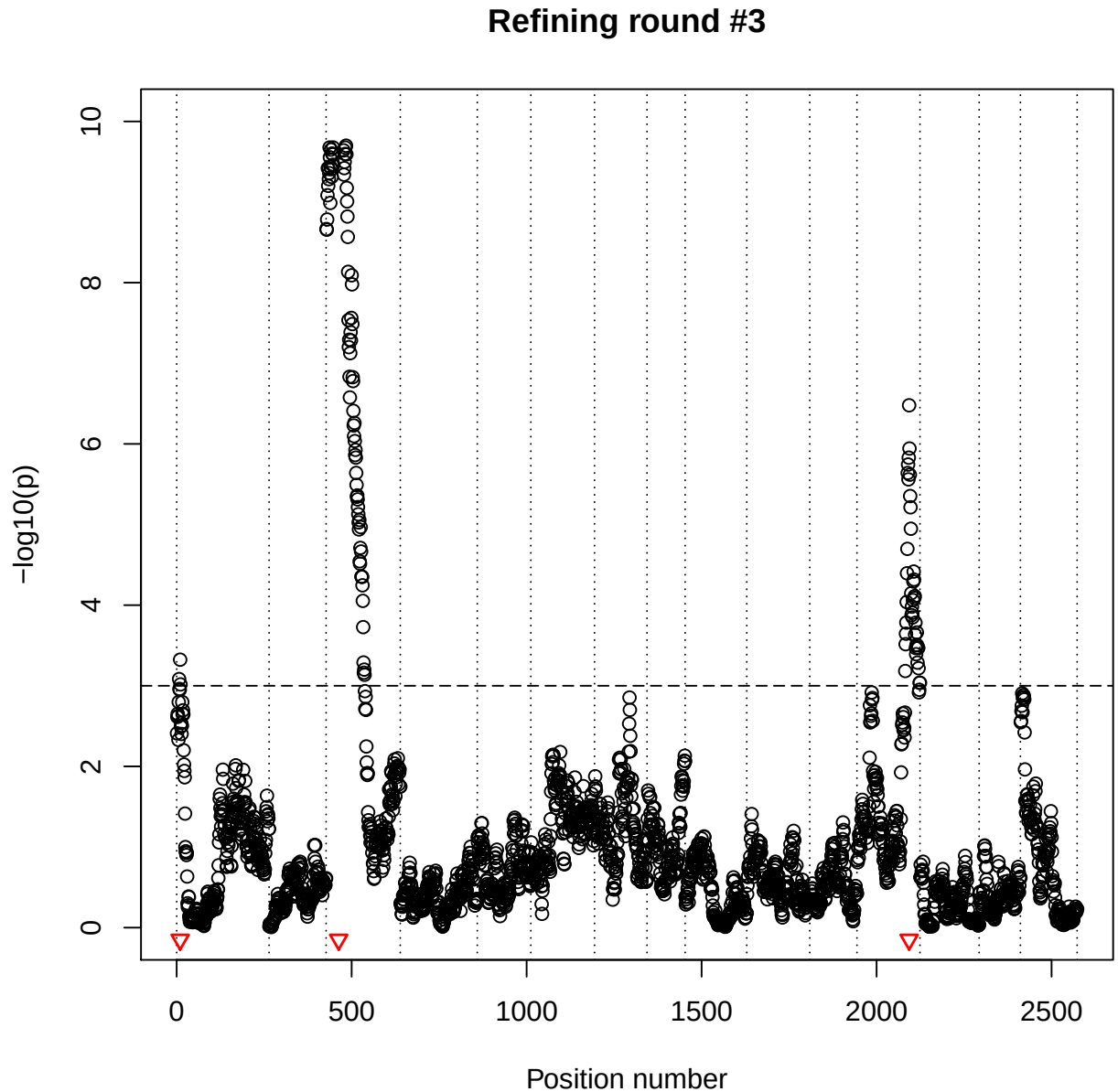
Dry matter – Model optimization



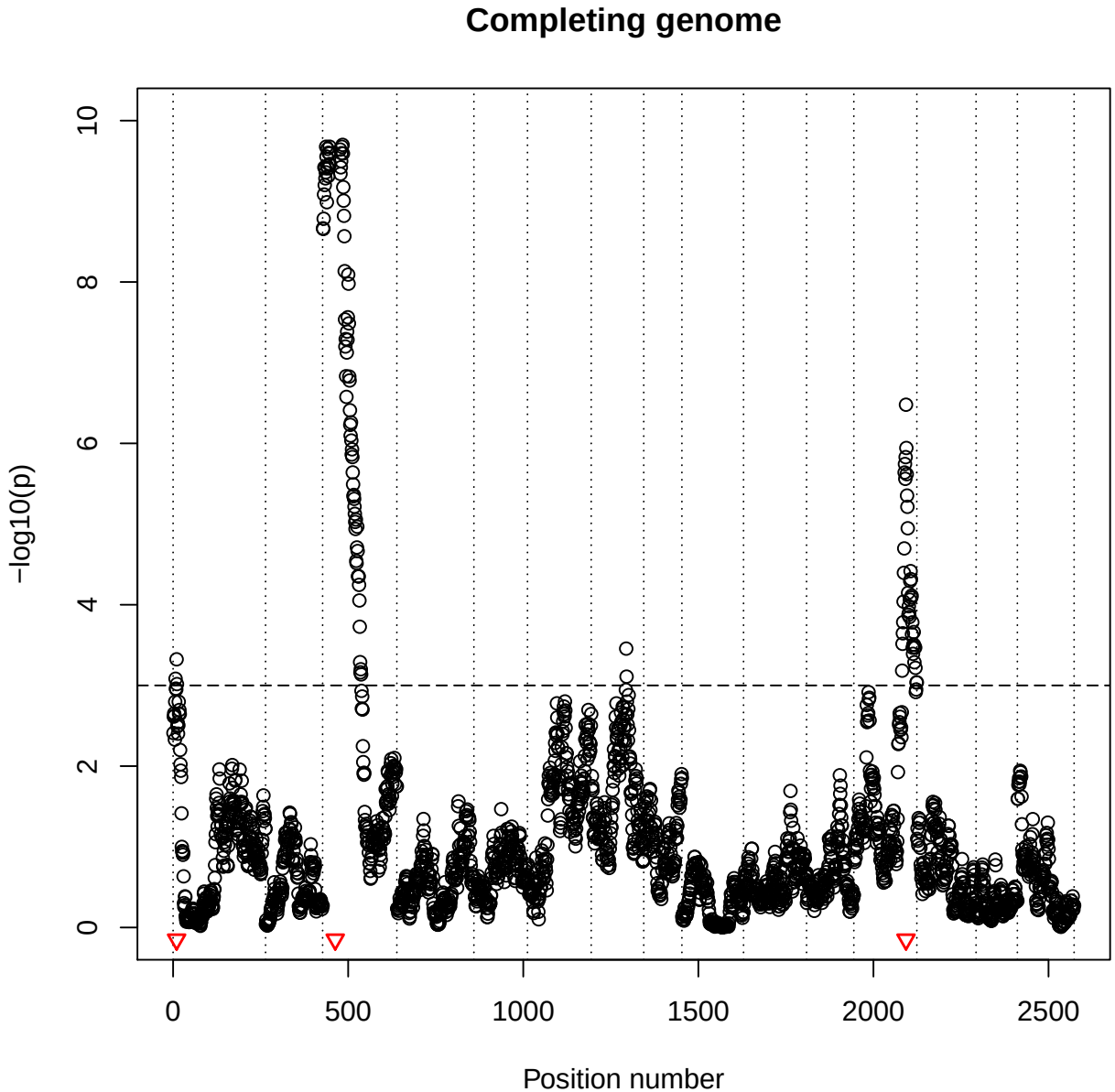
Dry matter – Model optimization



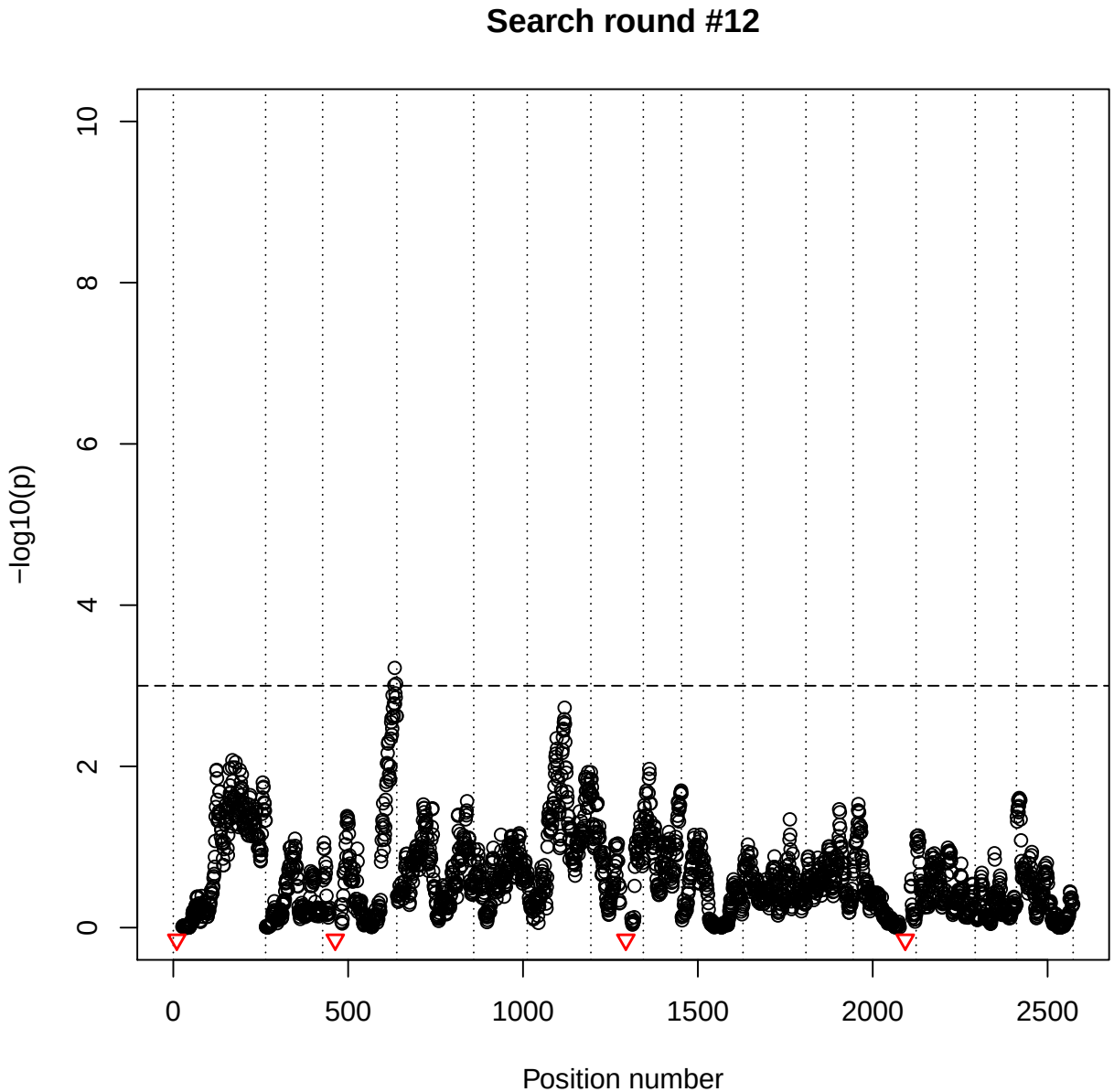
Dry matter – Model optimization



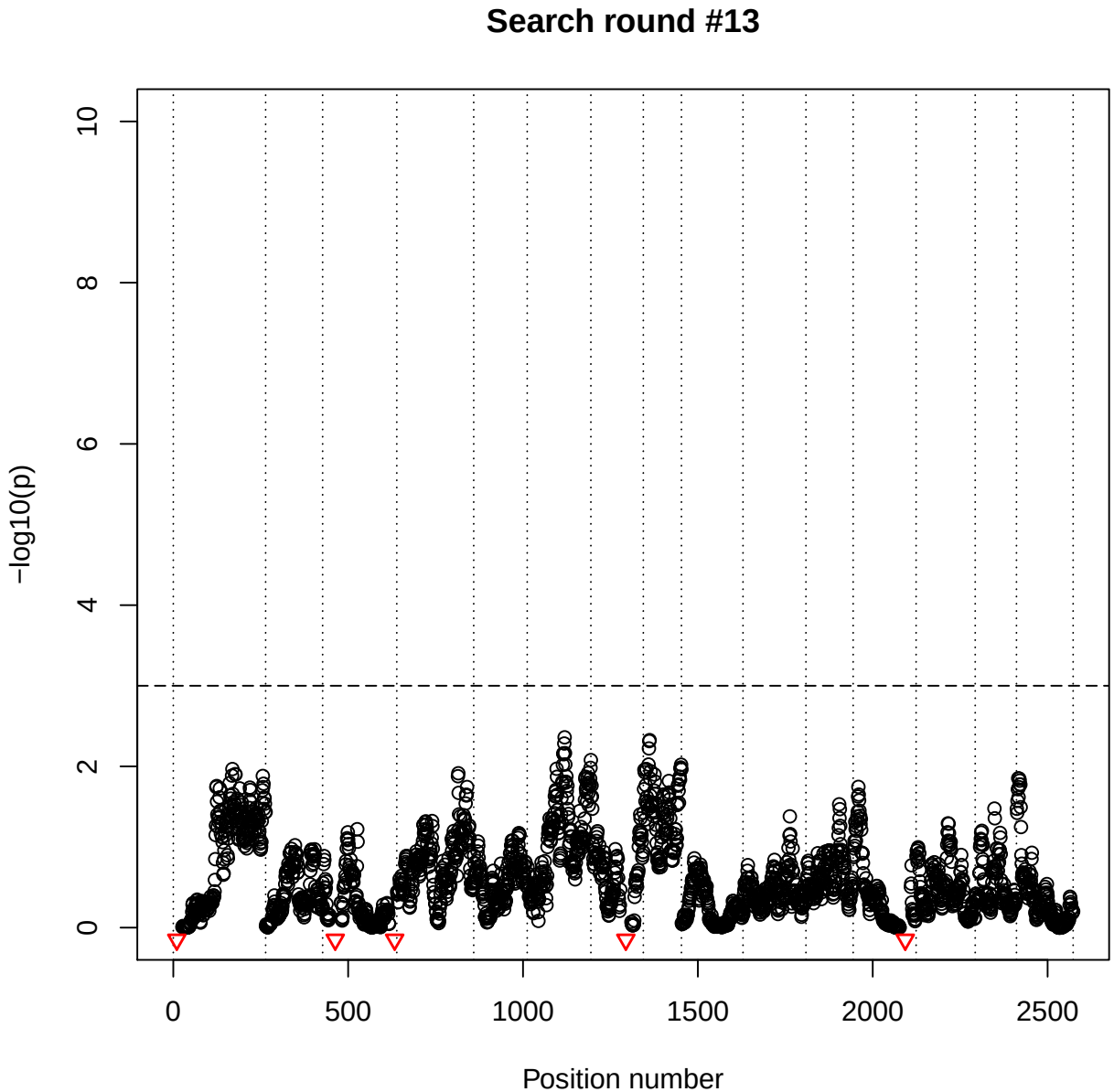
Dry matter – Model optimization



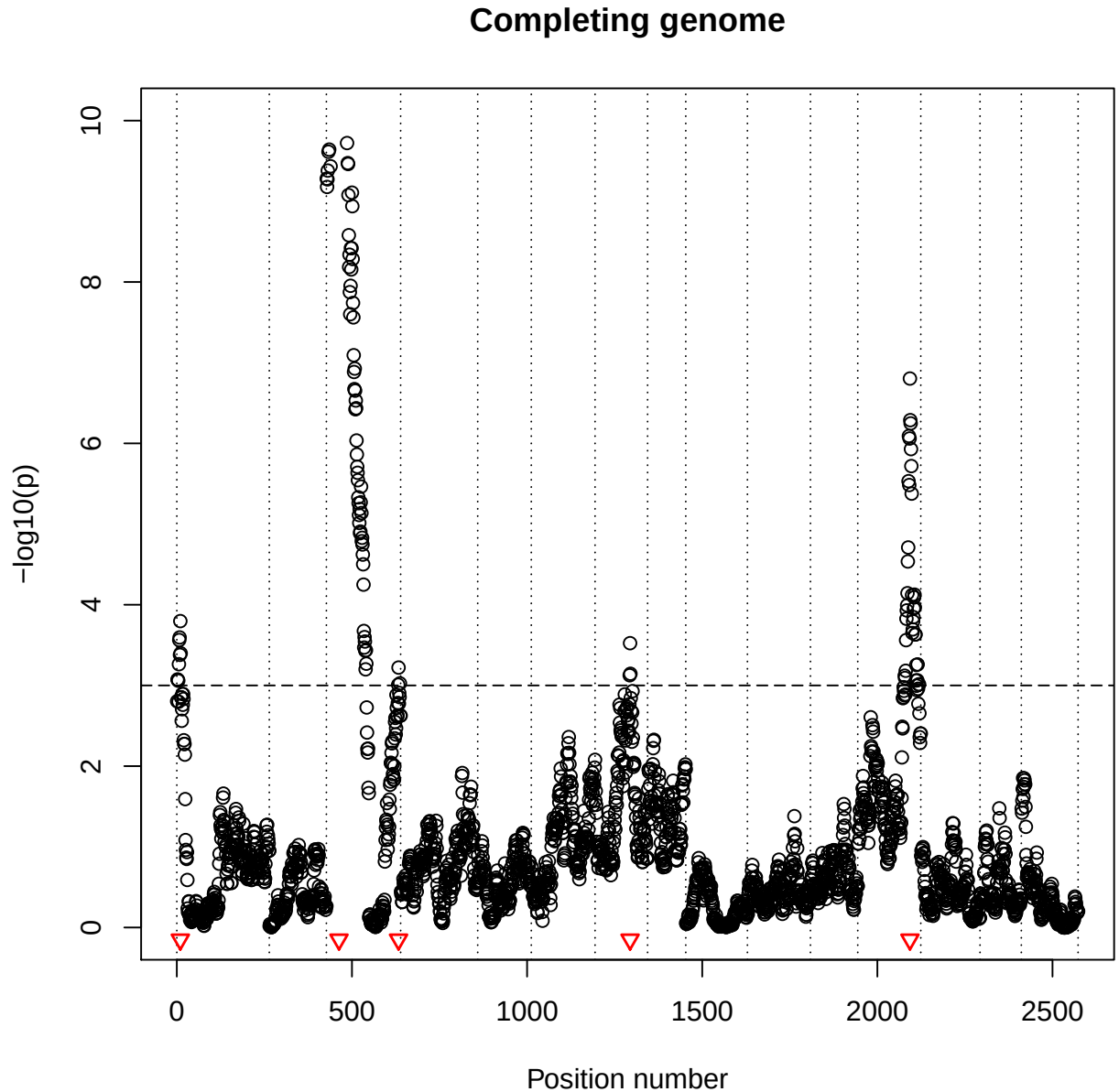
Dry matter – Forward search



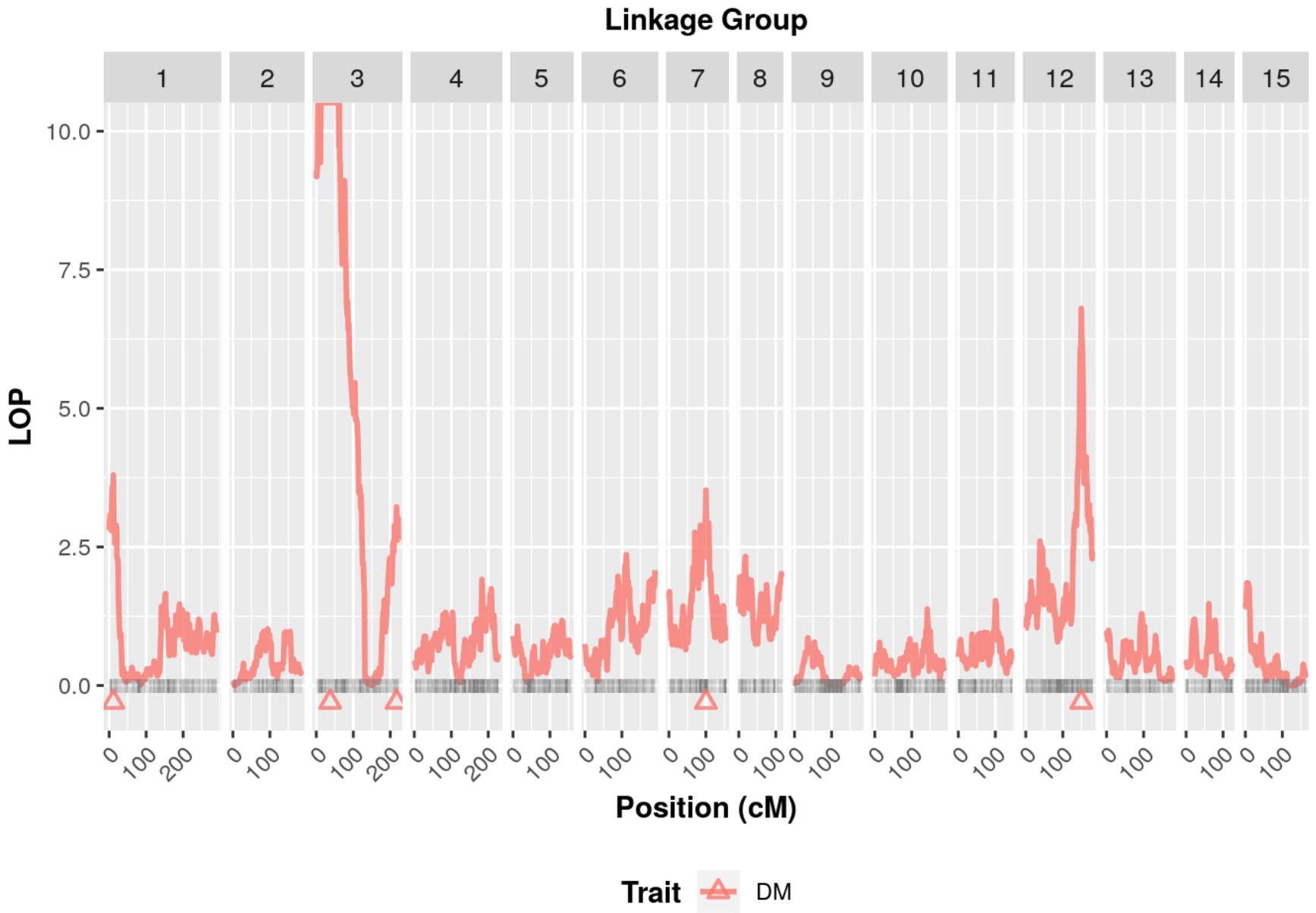
Dry matter – Forward search



Dry matter – Model optimization



Dry matter – Log of P-value (LOP) profile



Dry matter – Summary

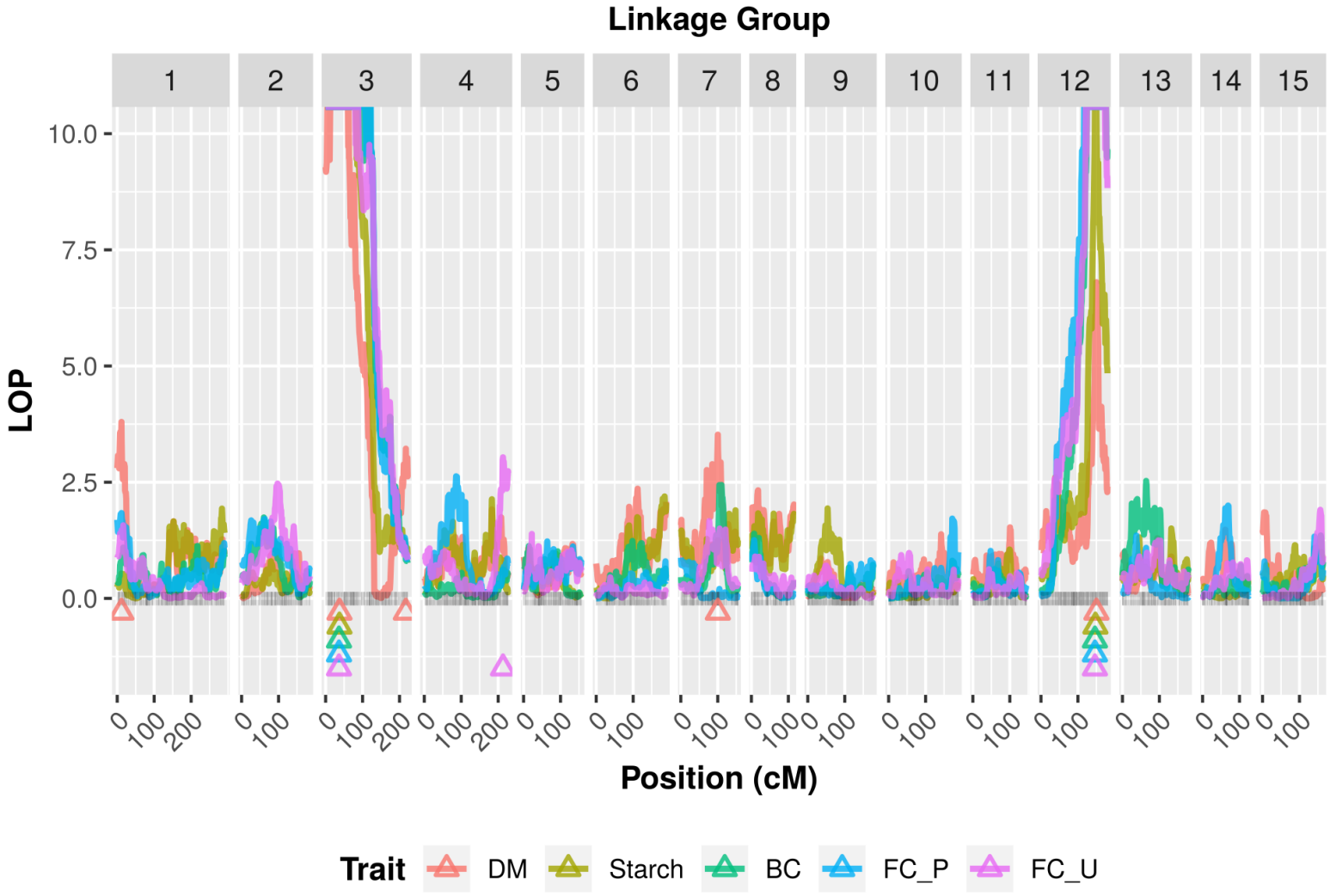
QTL	LG	Pos	Score	P-value	μ	σ_q^2	σ_e^2	h_q^2
1	1	11.41	158.93	1.59e-04		0.659047		0.0690
2	3	37.44	604.15	7.72e-16		3.571334		0.3738
3	3	217.13	133.93	6.00e-04		0.590448		0.0618
4	7	100.26	151.34	2.99e-04		0.570229		0.0597
5	12	150.05	264.05	1.57e-07		1.065668		0.1115
					27.33		3.096347	0.6759

- There are $Q = 5$ QTL contributing significantly to the variation of Dry Matter in B×T population:

$$\sigma_p^2 = \sum_{q=1}^{5} \boldsymbol{G}_q^{(i,i')} \sigma_q^2 + \sigma_e^2$$

- They explain a total of $h^2_{\text{total}} = 67.6\%$ of the phenotypic variance

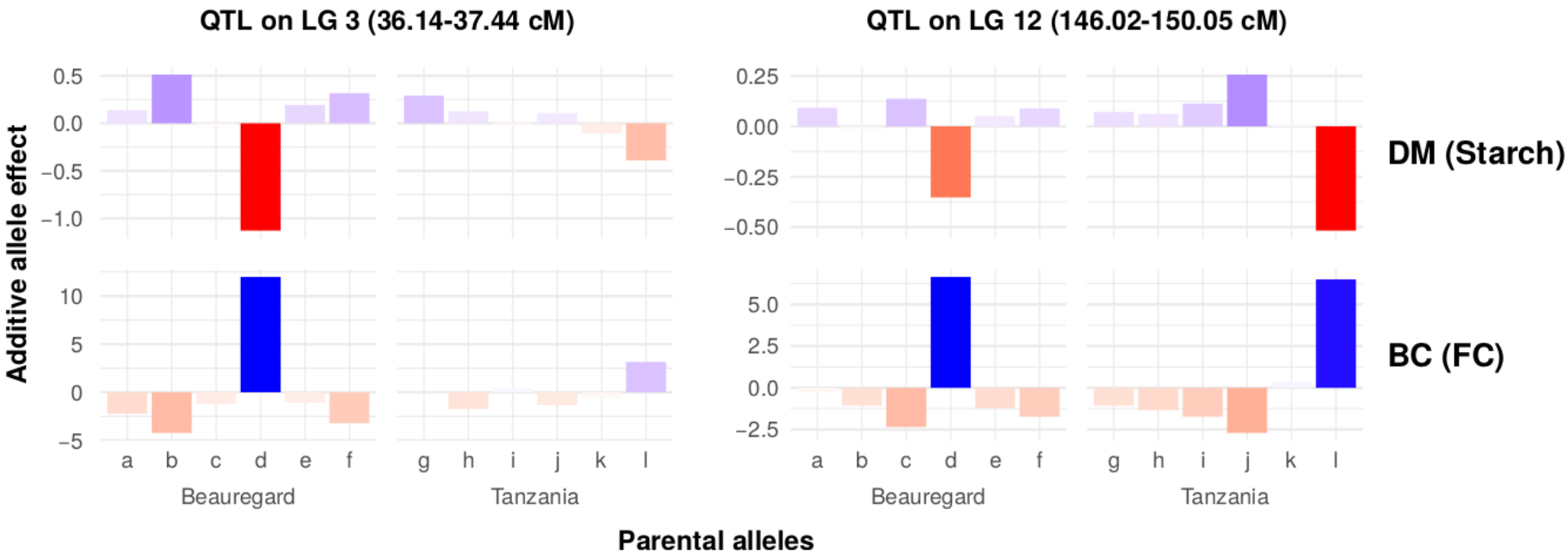
Quality-related traits – Log of P-value (LOP) profiles



Total QTL heritability (%) per trait

- DM = 67.6, Starch = 68.2, BC = 79.2, FC_P = 82.9, FC_U = 79.7

Quality-related traits – Allele effects



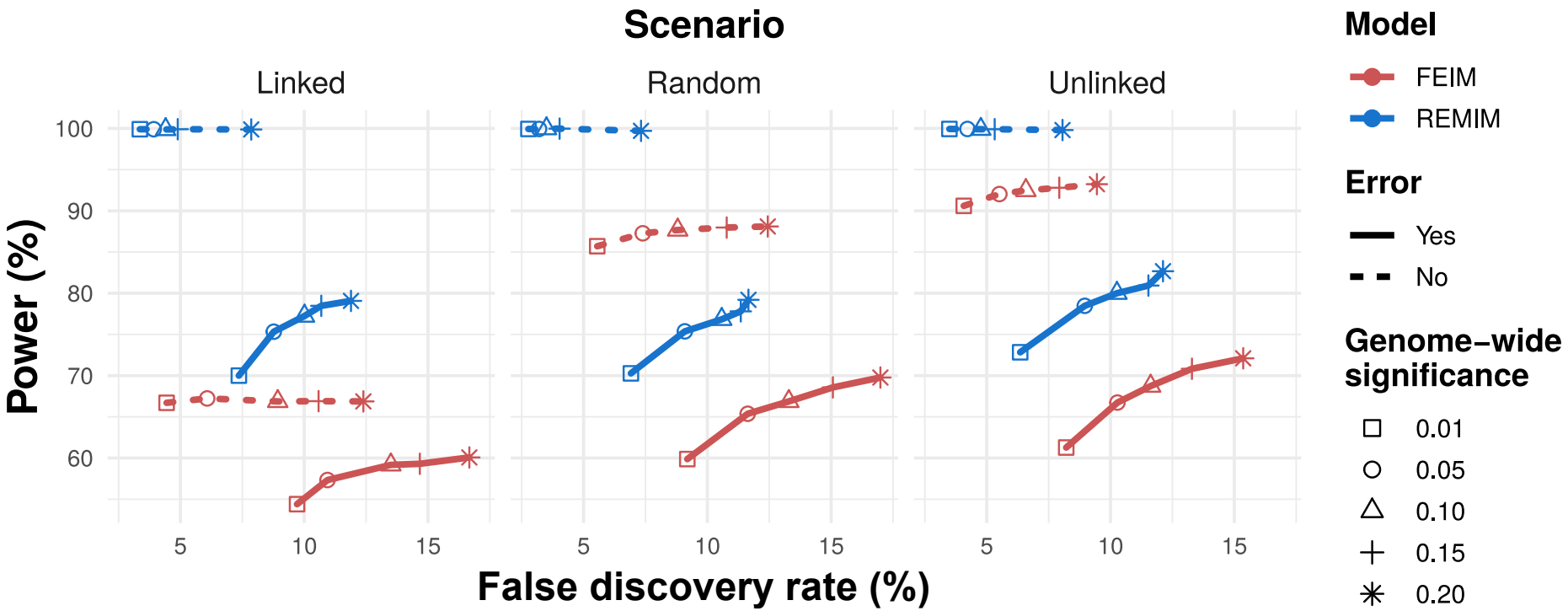
Genes: RNA-seq and *Ipomoea trifida* genome

- LG 3: *phytoene synthase* and *sucrose synthase* linked genes are involved in carotenoid and starch biosynthesis, respectively
- LG 12: *Orange* gene acts as a molecular switch for chromoplast biogenesis (carotenoid accumulation)

Fixed *vs.* random-effect models: simulated traits

1,000 simulations using 'Beauregard' × 'Tanzania' genetic map

- 3 QTL each simulation: $h^2_q = \{0.3, 0.2, 0.1\}$
- 3 scenarios: linked, random, unlinked QTL
- 5 genome-wide significance levels: $\alpha = \{0.20, 0.15, 0.10, 0.05, 0.01\}$
- Error: yes $(\sum_{q=1}^{3} \boldsymbol{G}_q^{(i,i')} \sigma^2_q + \sigma^2_e)$ or no $(\sum_{q=1}^{3} \boldsymbol{G}_q^{(i,i')} \sigma^2_q)$



Outline

1. Multiple QTL mapping
   - Components of variance
   - Random-effect QTL model
2. Applications
   - Potato: 'Atlantic' $\times$ B1829-5
   - Sweetpotato: 'Beauregard' $\times$ 'Tanzania'
3. Final considerations

Final considerations

QTLpoly R package

- R package for multiple mapping QTL in autopolyploids by using a random-effect model approach
  - Repository: `https://github.com/guilherme-pereira/qtlpoly`
  - Tutorial: `https://guilherme-pereira.github.io/QTLpoly/1-tutorial`
  - Issues: `https://github.com/guilherme-pereira/QTLpoly/issues`
  - Paper: `https://doi.org/10.1534/genetics.120.303080`

Main functions

- `read_data()` creates object structure with genotypes and phenotypes
- `remim()` performs REMIM in a automatic way
- `fit_model()` fits multiple QTL model with selected regions
- `qtl_effects()` estimates allele effects for each QTL

Final considerations

Other QTLpoly features

- It includes functions for fixed-effect model + permutations
- It allows for phenotype weights (more interesting for adjusted means derived from very unbalanced experiments)
- It is flexible such that one can build/modify the QTL model

To-do list

- Allow for genotypes originated from double-reduction events
- Implement multiple-trait/multiple-environment models
- Increase flexibility to analyze multi-parental populations
- Add experimental design effects (one-step analysis)
- ...

Thank you!

North Carolina State University (USA)

- Prof. Zhao-Bang Zeng
- Dr. Marcelo Mollinari
- Prof. Craig Yencho
- Prof. Bode Olukolu (currently at University of Tennessee Knoxville)

International Potato Center (Peru & Kenya)

- Dr. Dorcus Gemenet (currently at CIMMYT)

BILL & MELINDA GATES *foundation*

CIP is a research-for-development organization with a focus on potato, sweetpotato and Andean roots and tubers. It delivers innovative science-based solutions to enhance access to affordable nutritious food, foster inclusive sustainable business and employment growth, and drive the climate resilience of root and tuber agri-food systems. Headquartered in Lima, Peru, CIP has a research presence in more than 20 countries in Africa, Asia and Latin America.
www.cipotato.org

This work was undertaken as part of the CGIAR Research Program on Roots, Tubers and Bananas (RTB). Funding support for this work was provided by BMGF (through SweetGAINS) and the CGIAR Trust Fund Contributors.

CIP thanks all donors and organizations that globally support its work through their contributions to the CGIAR Trust Fund: www.cgiar.org/funders

This publication is copyrighted by the International Potato Center (CIP). It is licensed for use under the Creative Commons Attribution 4.0 International License