

## Electronic supplementary materials

**Title:** Cassava GWAS catalogue reveals new insights into the genetic architecture of defensive, morphological and quality-related traits

**Journal:** Plant Molecular Biology

**Author names:** Ismail Yusuf Rabbi<sup>1\*</sup>, Siraj Ismail Kayondo<sup>1</sup>, Guillaume Bauchet<sup>2</sup>, Muyideen Yusuf<sup>1</sup>, Cynthia Aghogho<sup>1</sup>, Kayode Ogunpaimo<sup>1</sup>, Ruth Uwugiaren<sup>1</sup>, Ikpan Andrew Smith<sup>1</sup>, Prasad Peteti<sup>1</sup>, Afolabi Agbona<sup>1</sup>, Elizabeth Parkes<sup>1</sup>, Ezenwaka Lydia<sup>3</sup>, Marnin Wolfe<sup>5</sup>, Jean-Luc Jannink<sup>5,6</sup>, Chiedozie Egesi<sup>1,3,4</sup>, and Peter Kulakow<sup>1</sup>

### Affiliations

<sup>1</sup> International Institute of Tropical Agriculture (IITA), Ibadan, Oyo State, 200001 Nigeria,

<sup>2</sup> Boyce Thompson Institute, Ithaca, New York 14853;

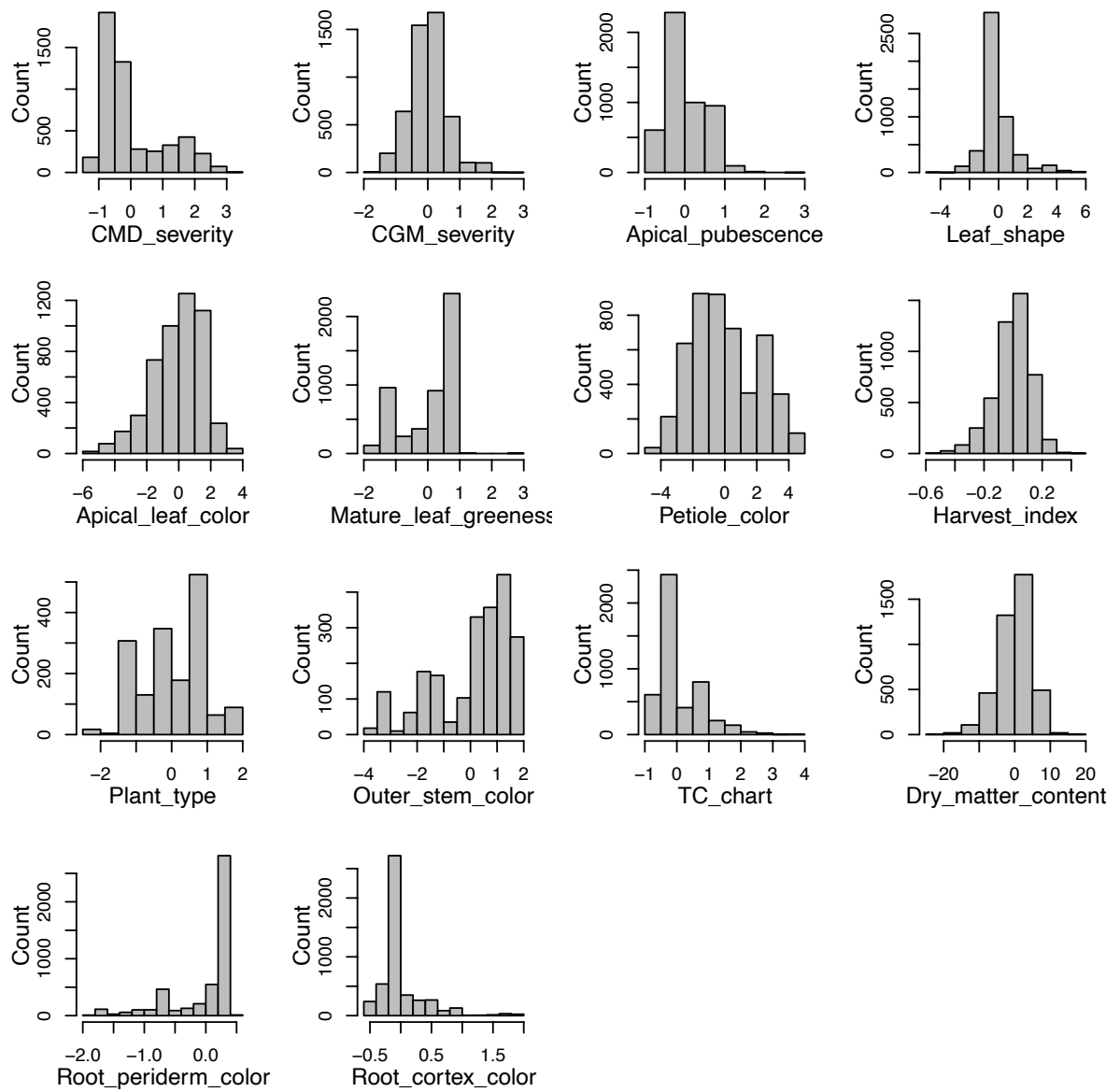
<sup>3</sup> National Root Crops Research Institute (NRCRI), PMB 7006, Umudike, 440221 Nigeria;

<sup>4</sup> Global Development Department, College of Agriculture and Life Sciences, Cornell Univ., Ithaca, New York 14850;

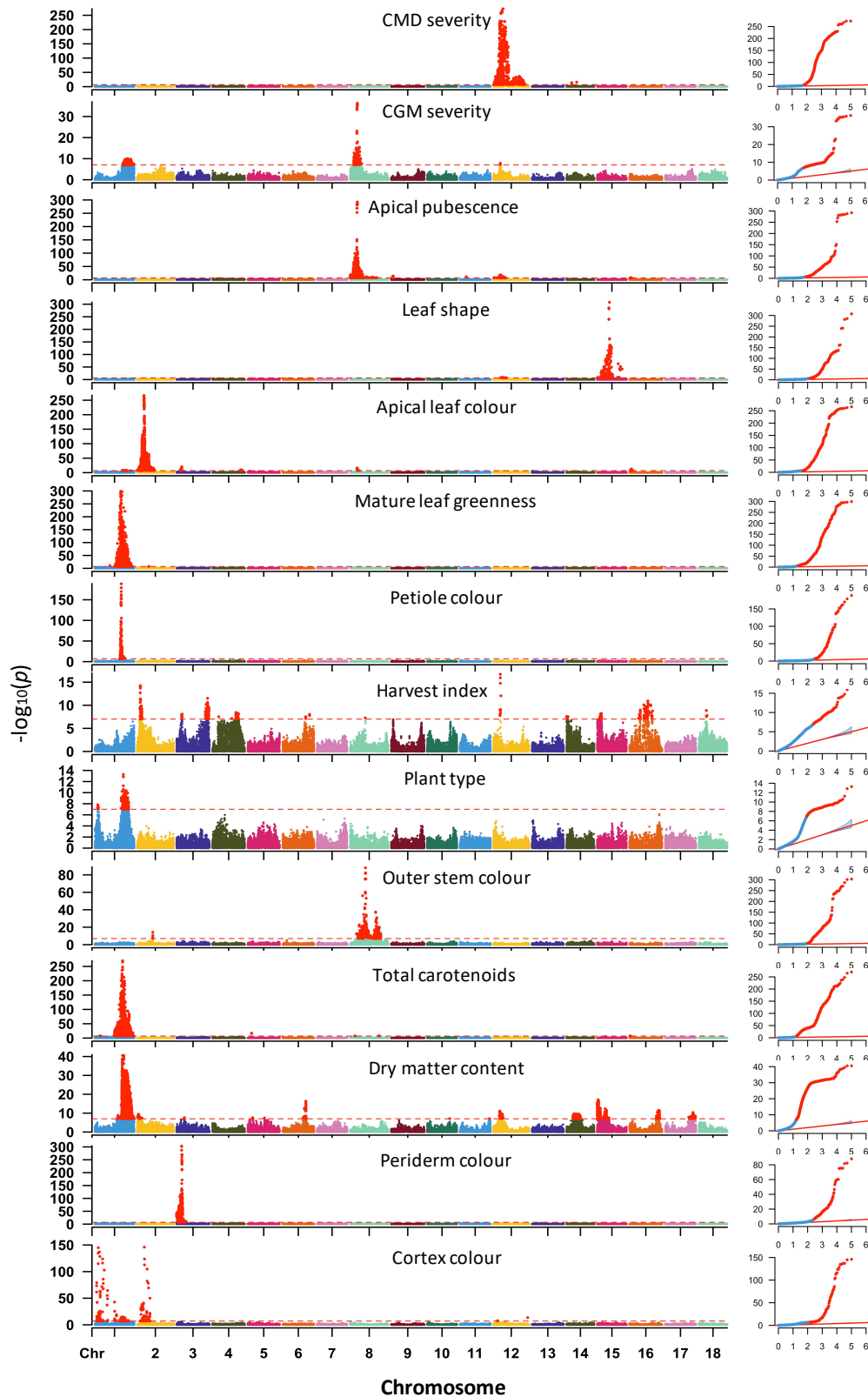
<sup>5</sup> Section on Plant Breeding and Genetics, School of Integrative Plant Sciences, Cornell Univ., Ithaca, NY 14850;

<sup>6</sup> United States Department of Agriculture - Agriculture Research Service, Ithaca, New York 14850;

\* Corresponding author: [I.Rabbi@cgiar.org](mailto:I.Rabbi@cgiar.org)

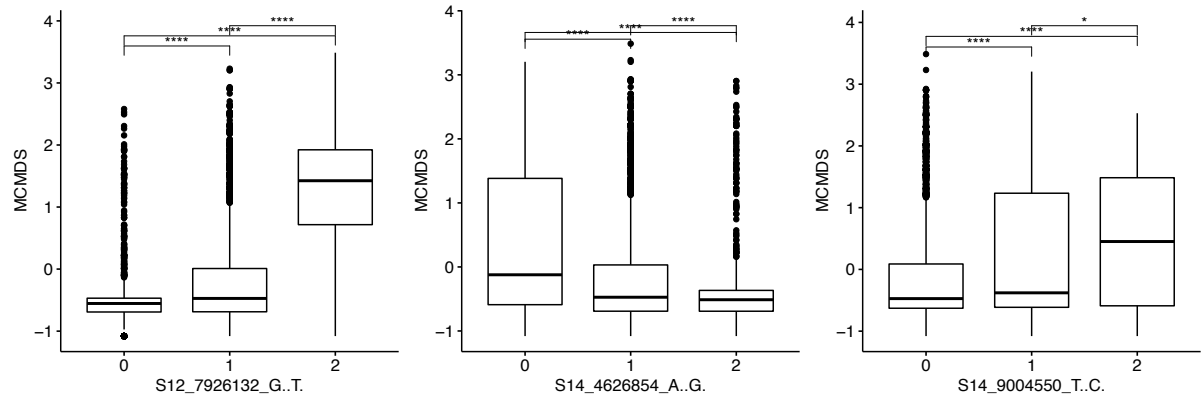


**Figure S1** Histogram showing the distribution of BLUPs for the 14 traits used in the GWAS analysis.

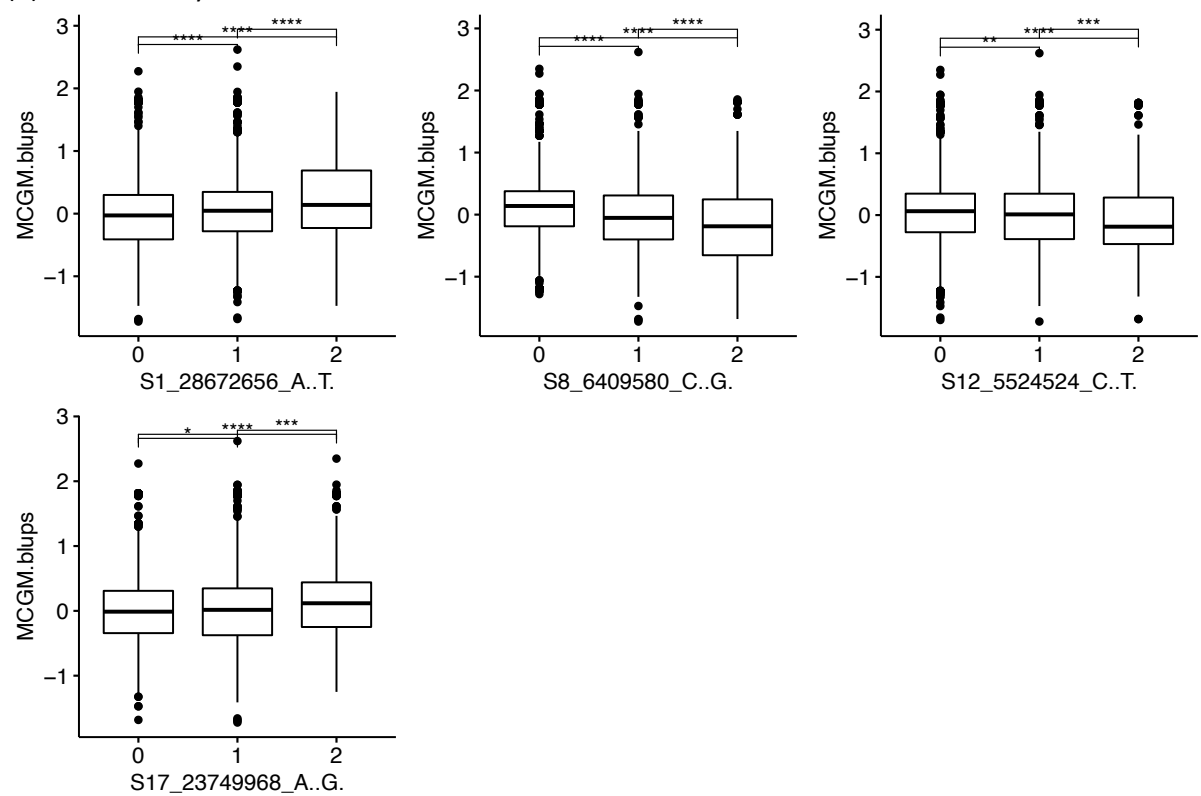


**Figure S2** Manhattan plots for GWAS for 14 traits of 5,130 cassava accessions using “MLMe” analysis approach. A total of 101,521 SNP markers were used for that GWAS analyses with the red horizontal line representing genome-wide significance threshold ( $\alpha=0.05/101521 = 4.93 \times 10^{-7}$ ). The QQ-plots inset - right with observed p-values on the y-axis and expected p-values on the x-axis.

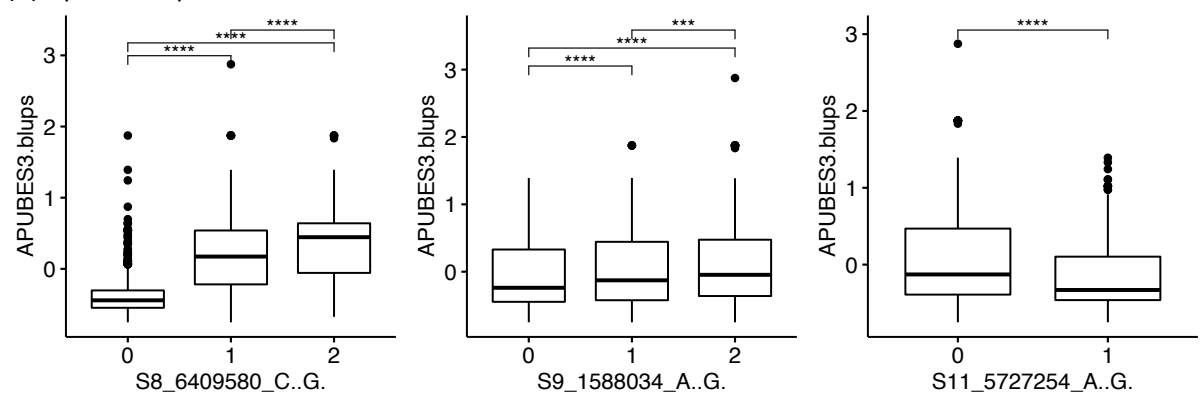
(A) CMD severity

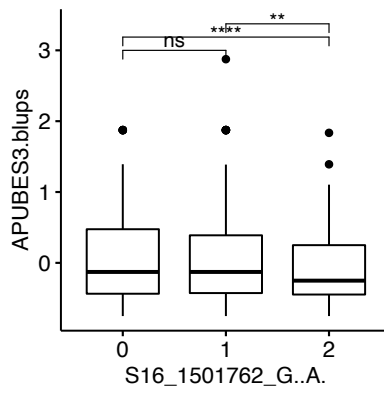
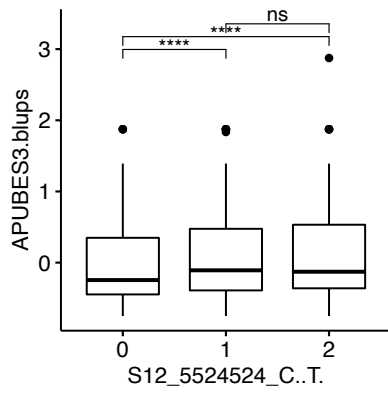


(B) CGM severity

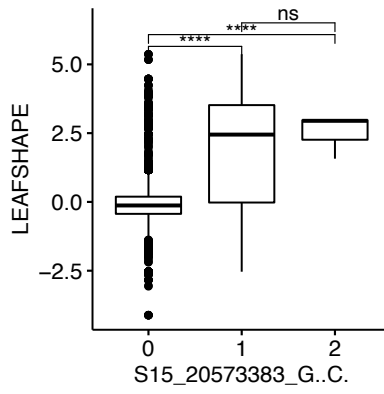
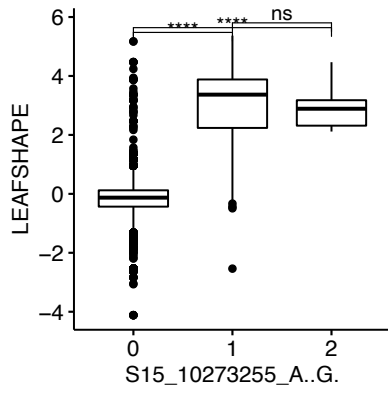


(C) Apical leaf pubescence

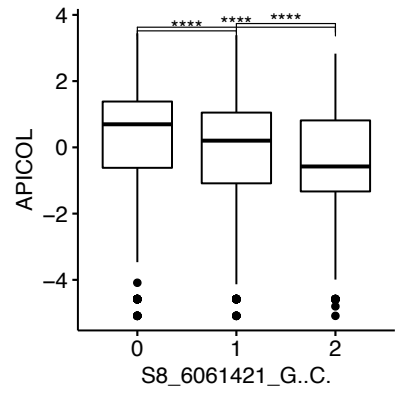
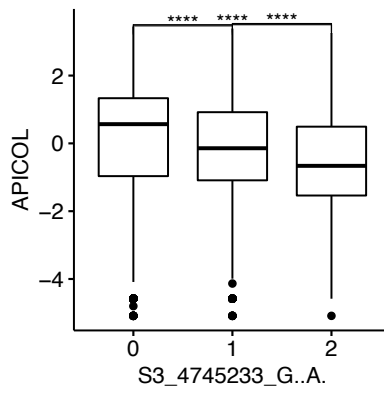
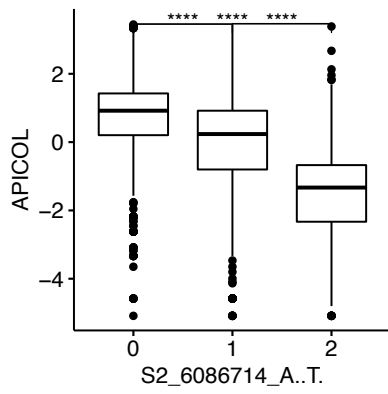




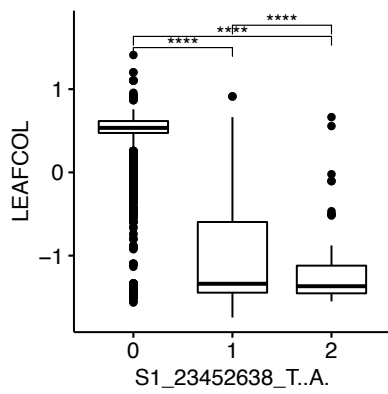
(D) Leaf shape



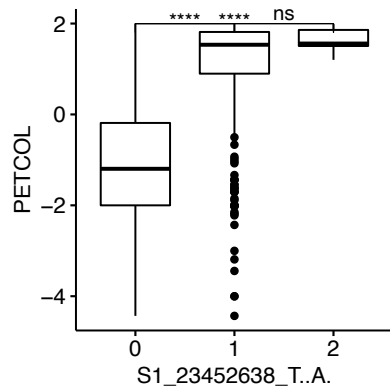
(E) Apical leaf colour



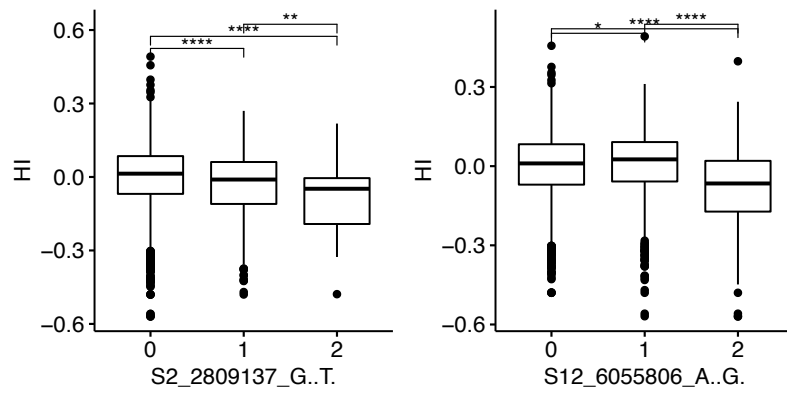
(F) Mature leaf greenness



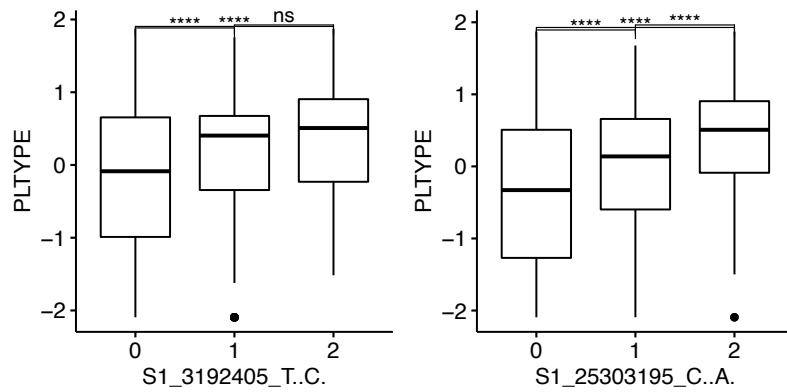
(G) Petiole colour



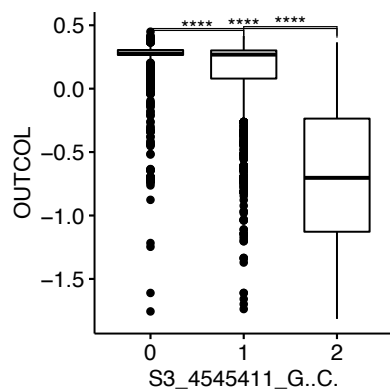
(H) Harvest index



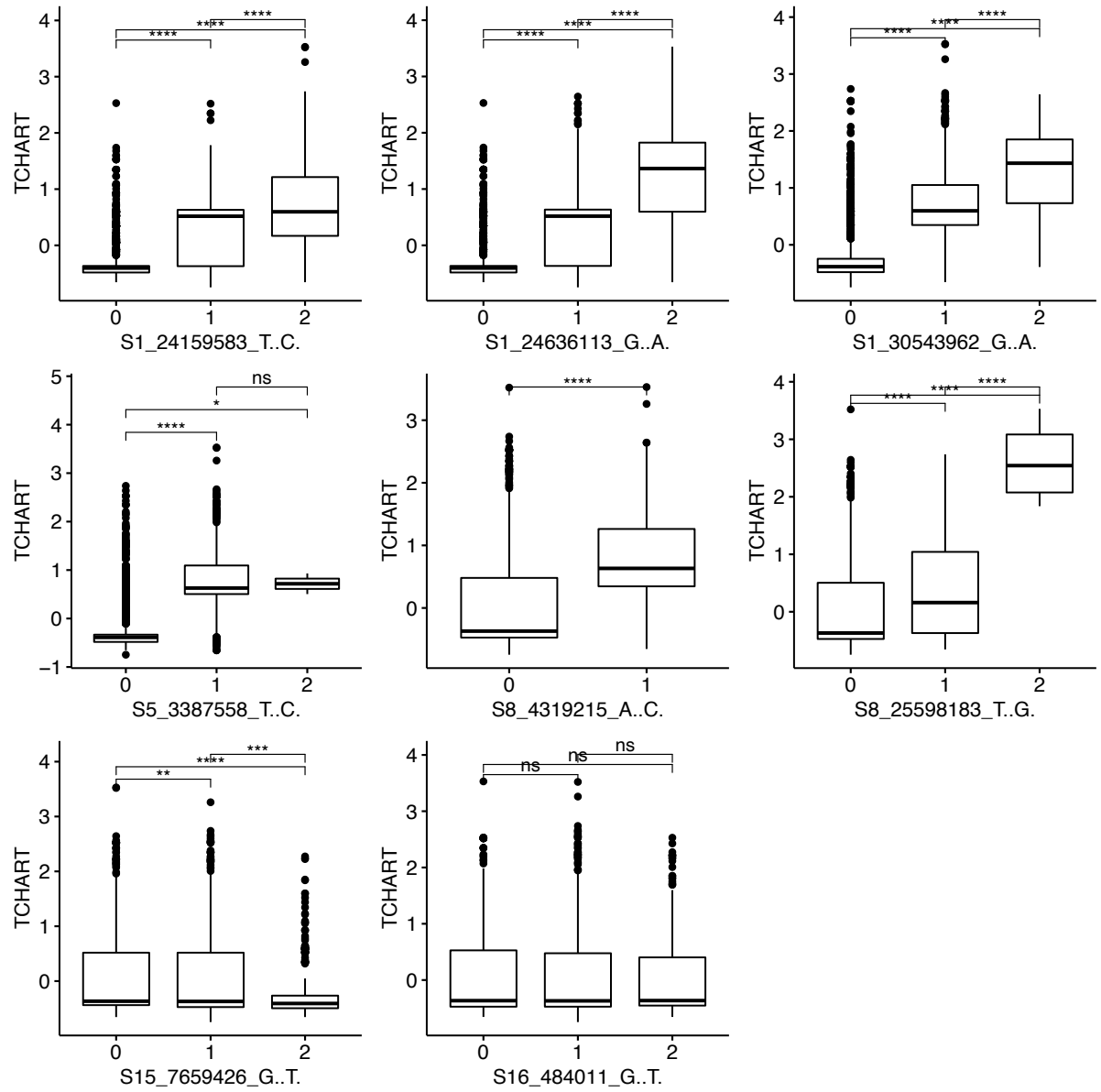
(I) Plant type



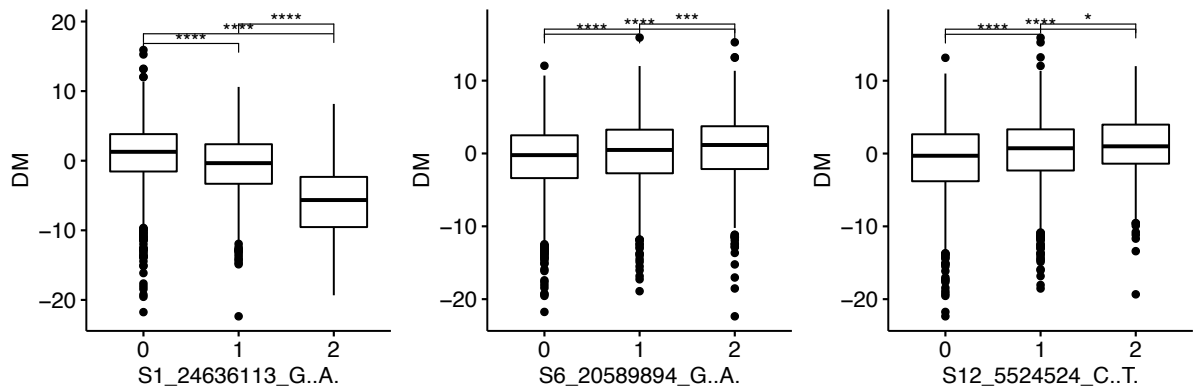
(J) Outer stem colour

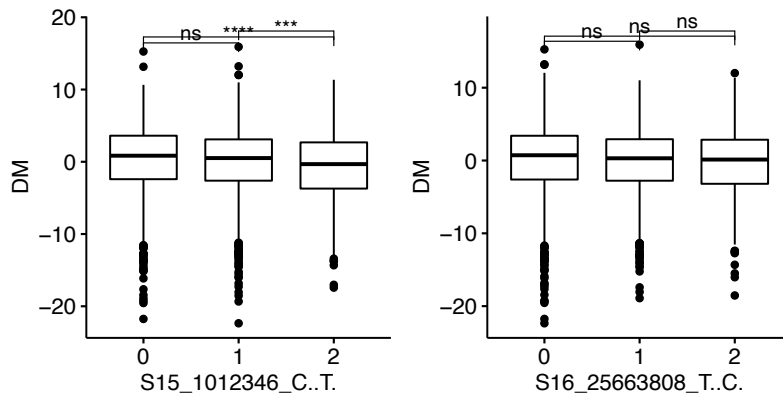


(K) Total carotenoids content

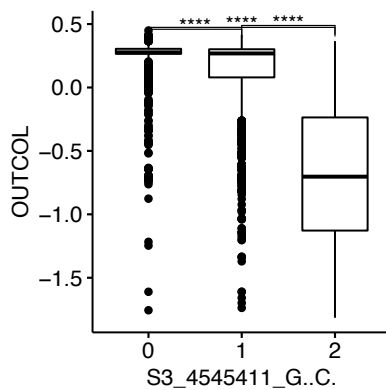


(L) Dry matter content

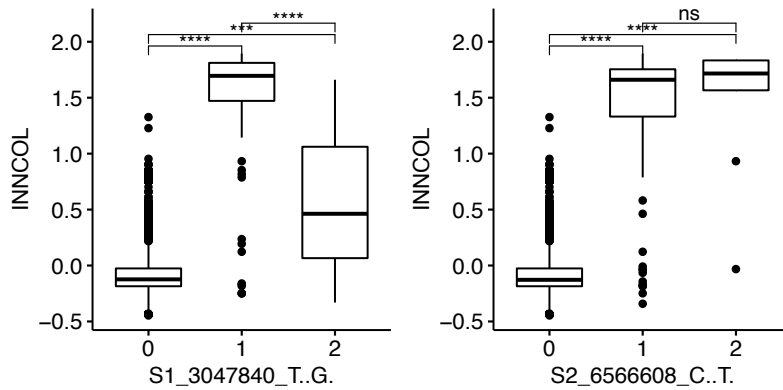




(M) Periderm colour

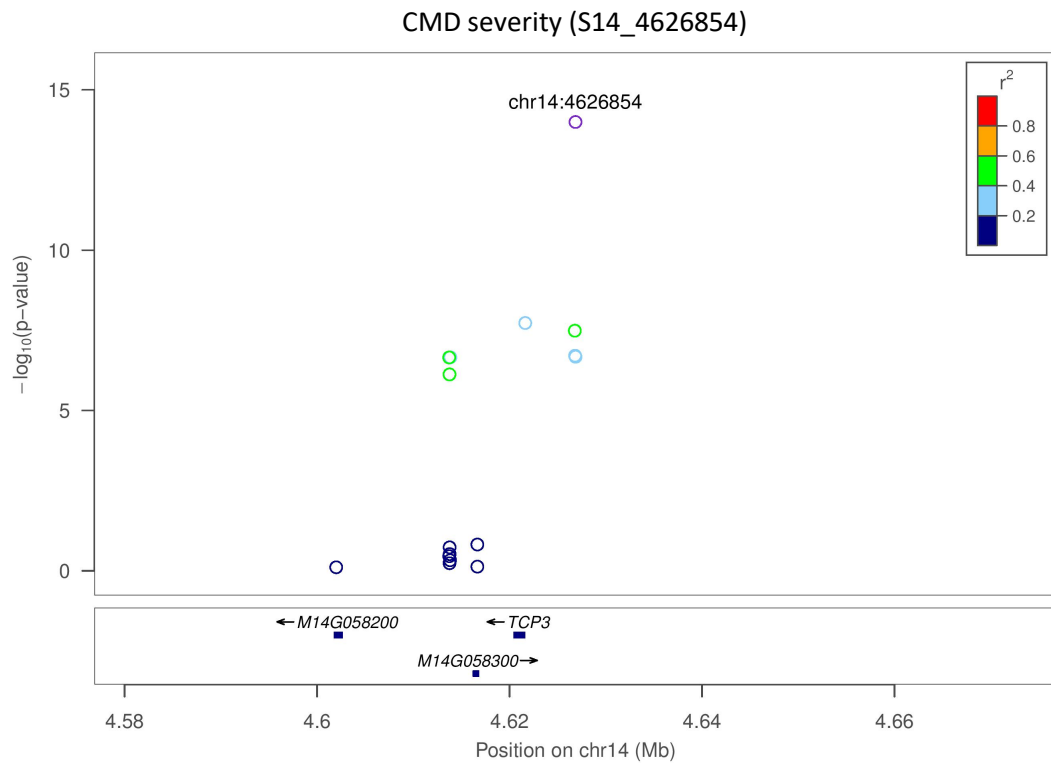
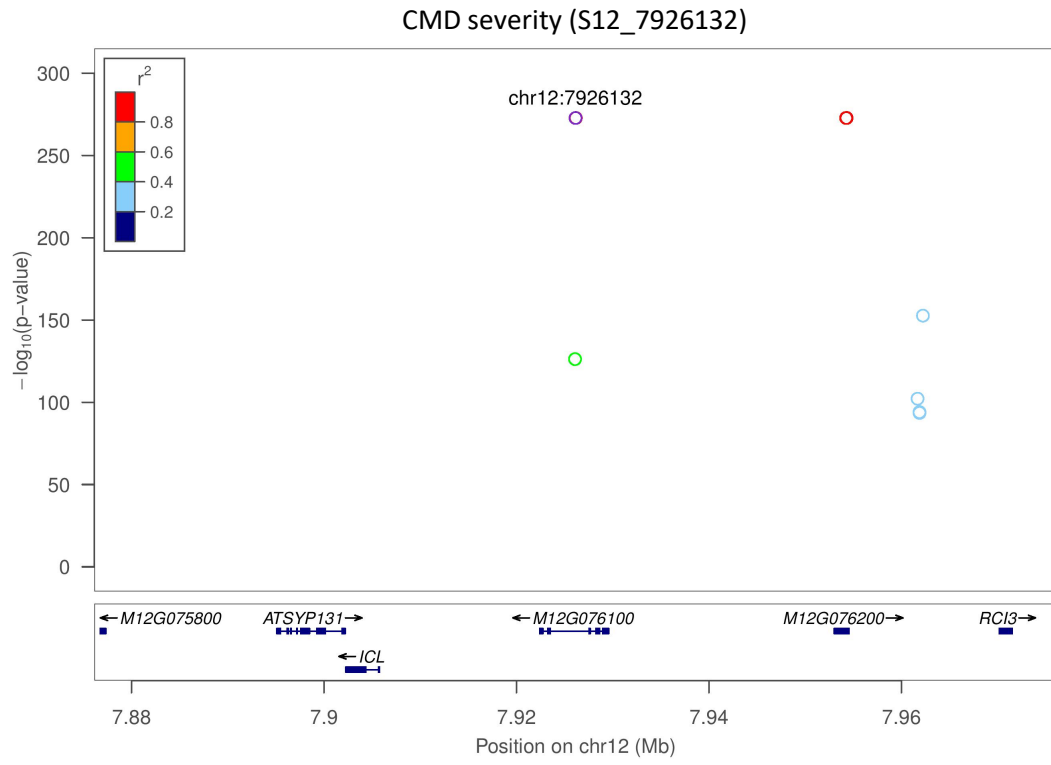


(N) Root cortex colour

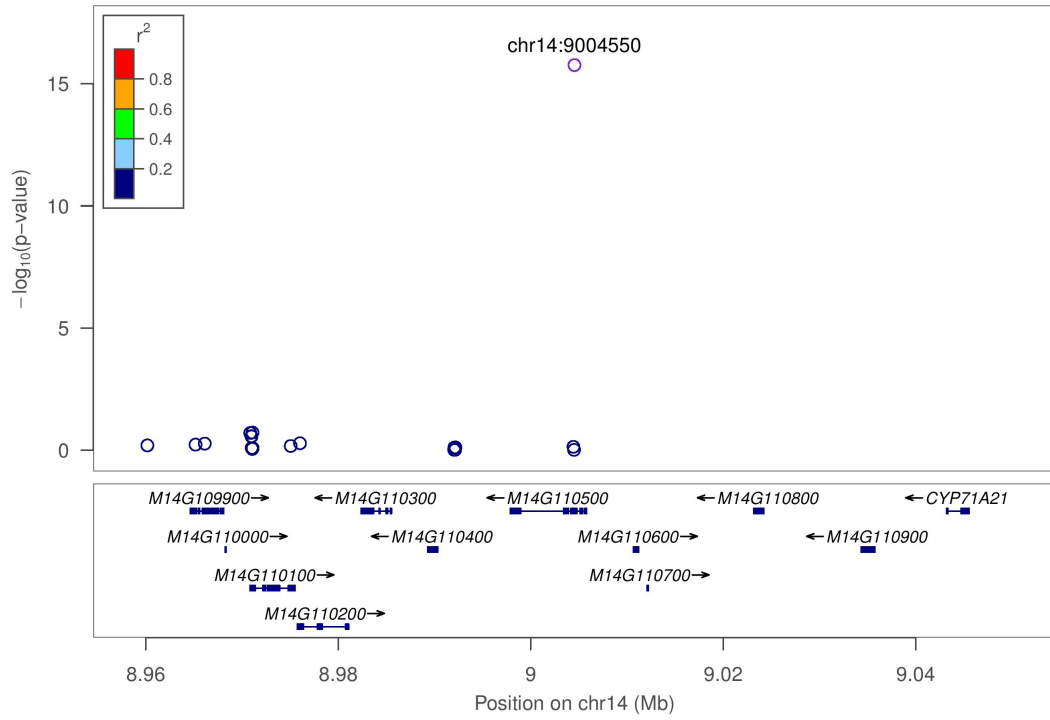


**Figure S3** Allelic substitution effects in the most significant SNP at each locus identified for each of the 14 studied traits. Trait BLUPs distribution on y-axis and SNP genotype status of the marker on x-axis. SNPs were converted to dosage format, where 0, 1, and 2 indicates the copies of the minor alleles. The first allele in the suffix of a SNP name denotes the allele being counted in the dosage coding. For example, dosage score of 2 in SNP S12\_7926132\_G..T. means homozygous for “G”, a score of 1 means heterozygote “GT”, and a score of 0 means homozygote alternate allele “T”. (A) CMD severity; (B) CGM severity; (C) Apical leaf pubescence; (D) Leaf shape(E); Apical leaf colour; (F) Mature leaf greenness; (G) Petiole colour; (H) Harvest index; (I) Plant type; (J) Outer stem colour; (K) Total carotenoids content; (L) Dry matter content; (M) Periderm colour; (N) Root cortex colour. Significance of pairwise t-test between the genotypic classes at each SNP presented. ns:  $p > 0.05$ ; \*:  $p \leq 0.05$ ; \*\*:  $p \leq 0.01$ ; \*\*\*:  $p \leq 0.001$ ; \*\*\*\*:  $p \leq 0.0001$ .



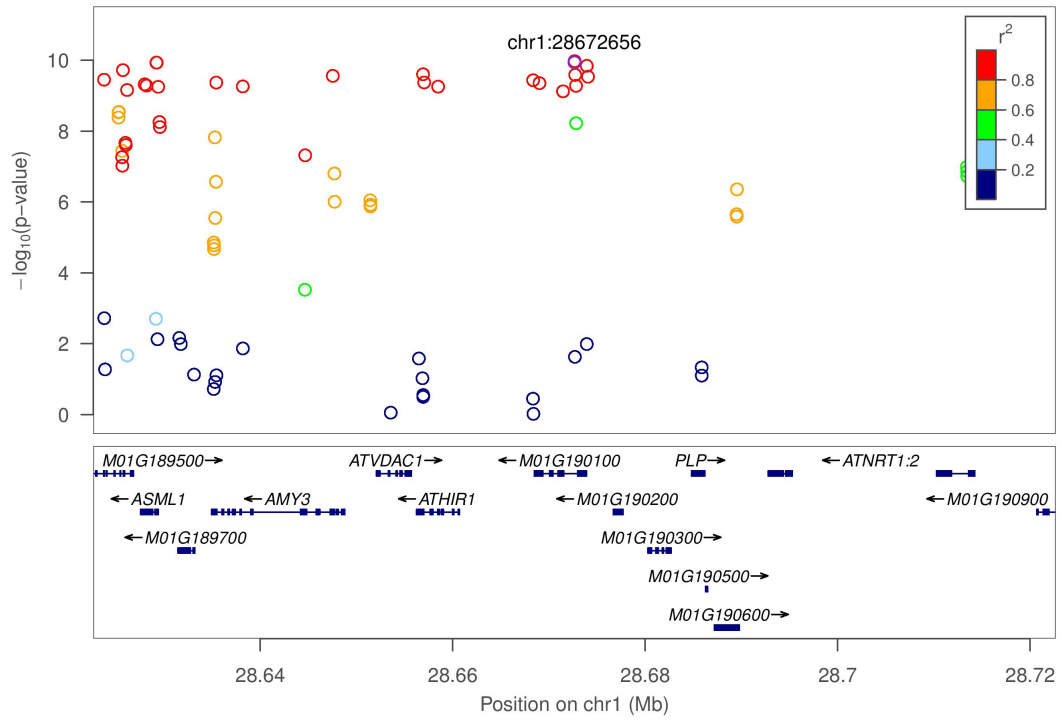
**A**

CMD severity (S14\_9004550)

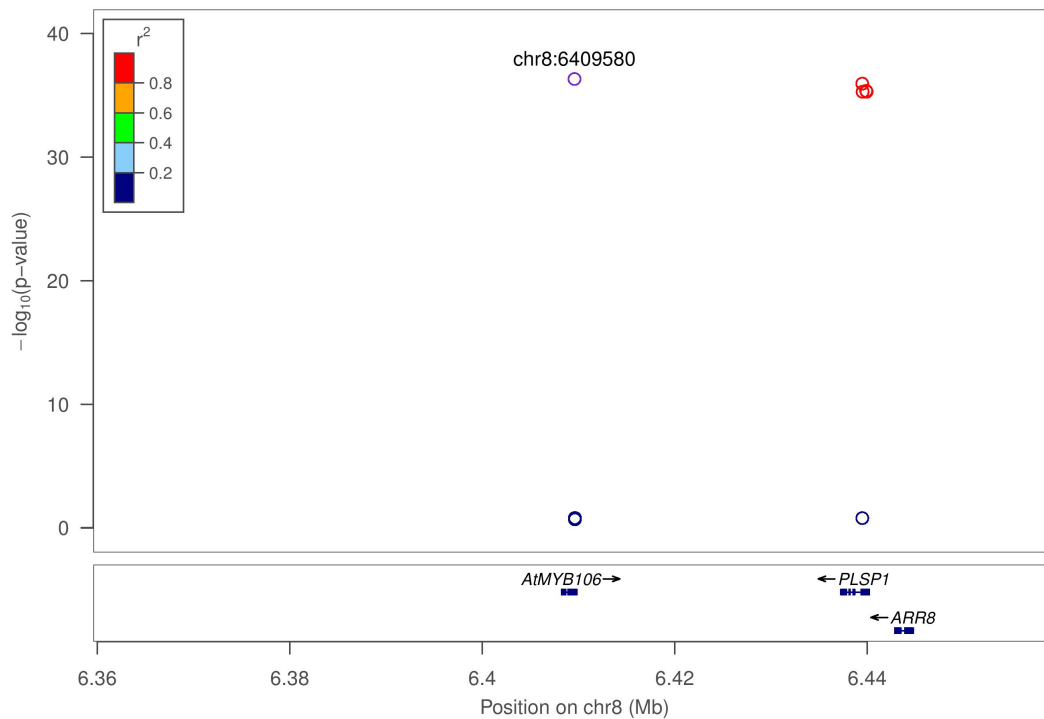


**B**

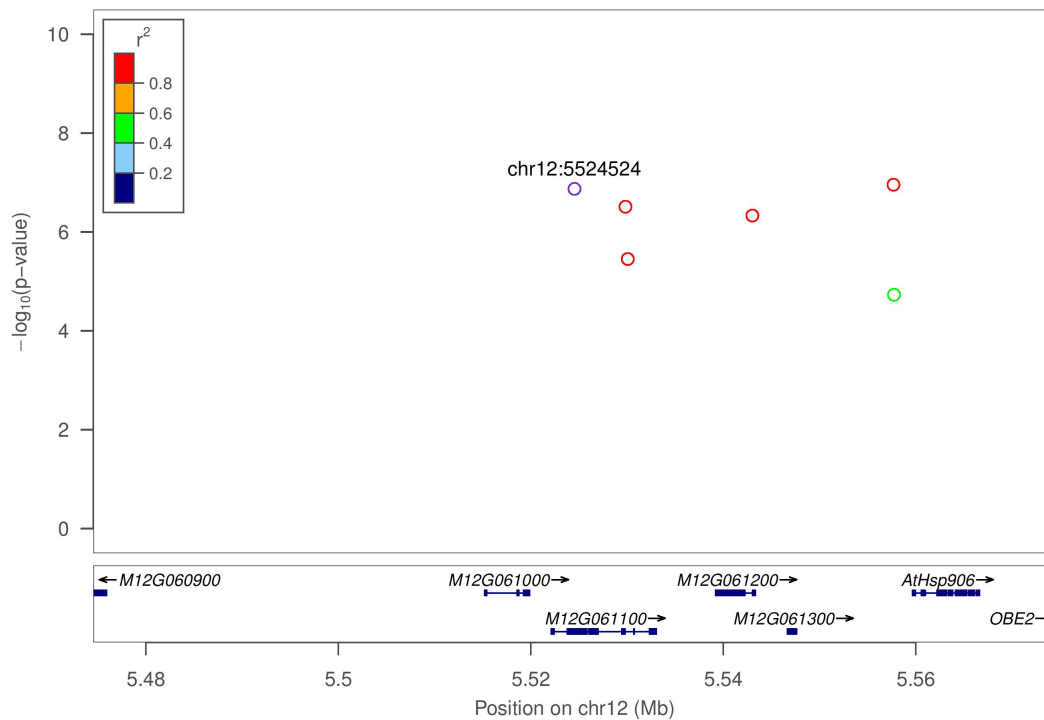
CGM severity (S1\_28672656)



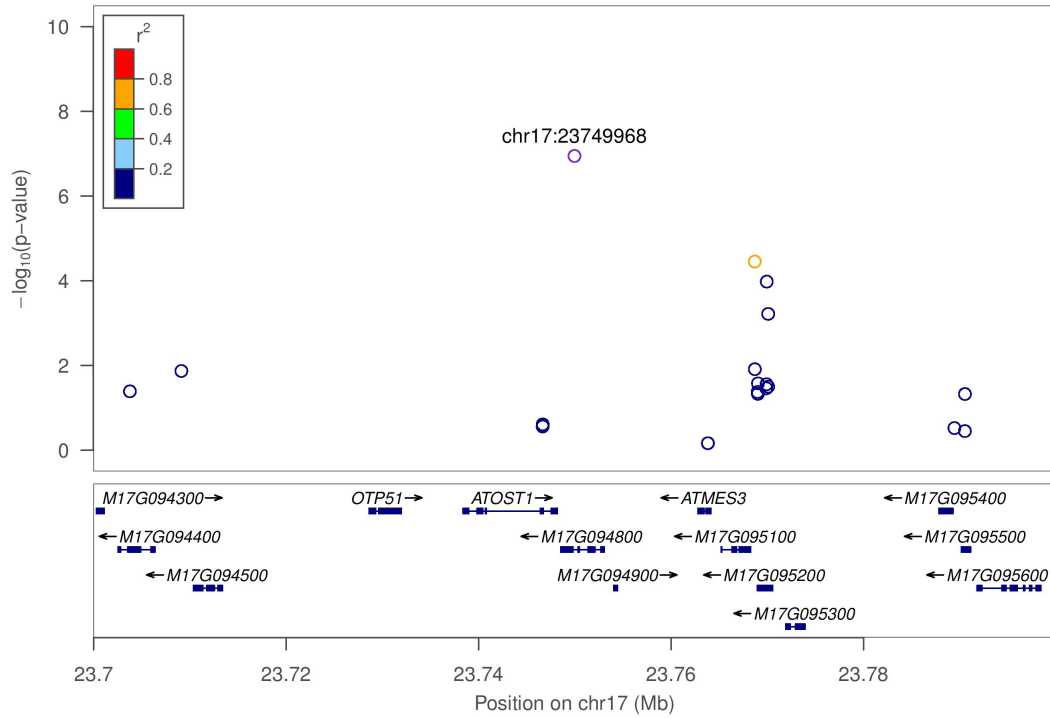
CGM severity (S8\_6409580)



CGM severity (S12\_5524524)

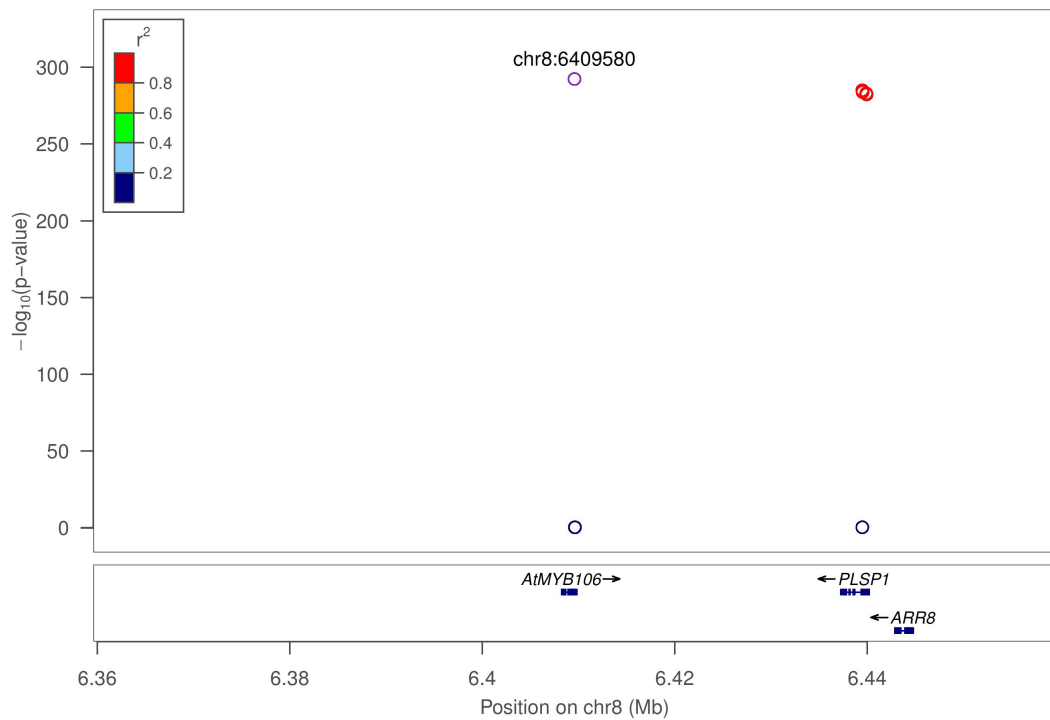


CGM severity (S17\_23749968)

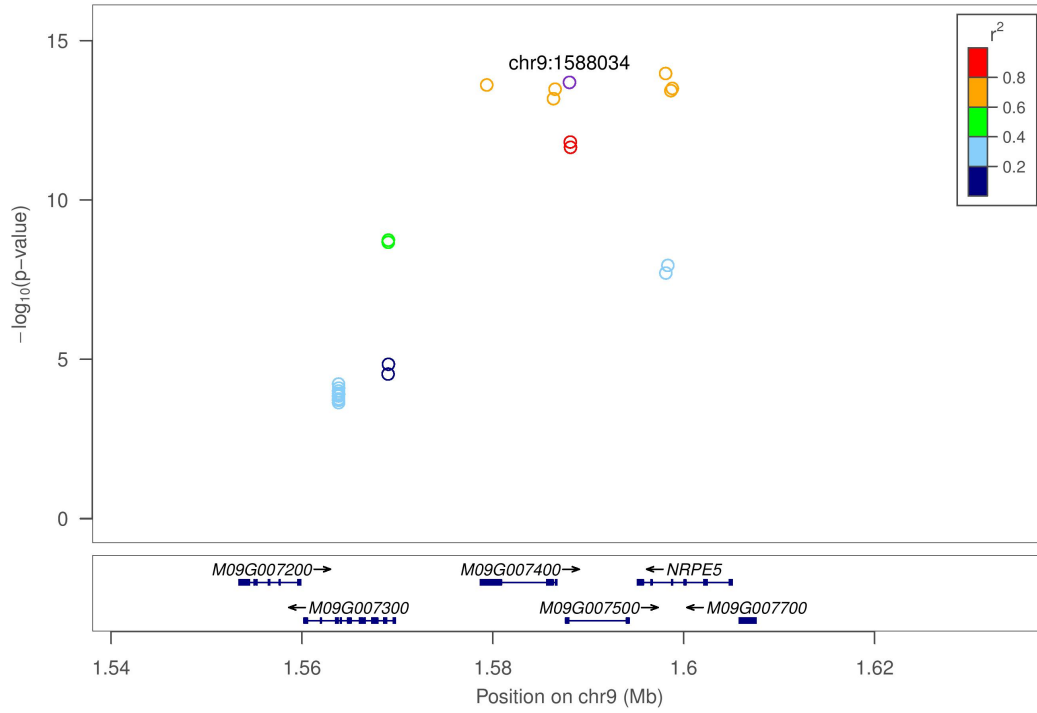


C

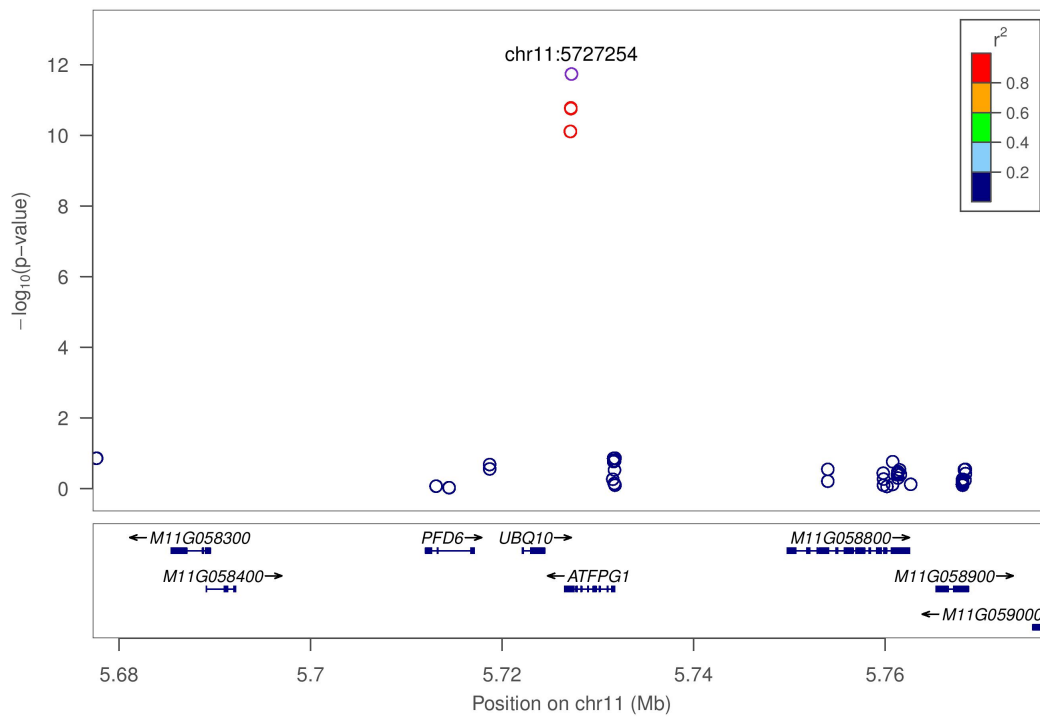
Apical pubescence (S8\_6409580)



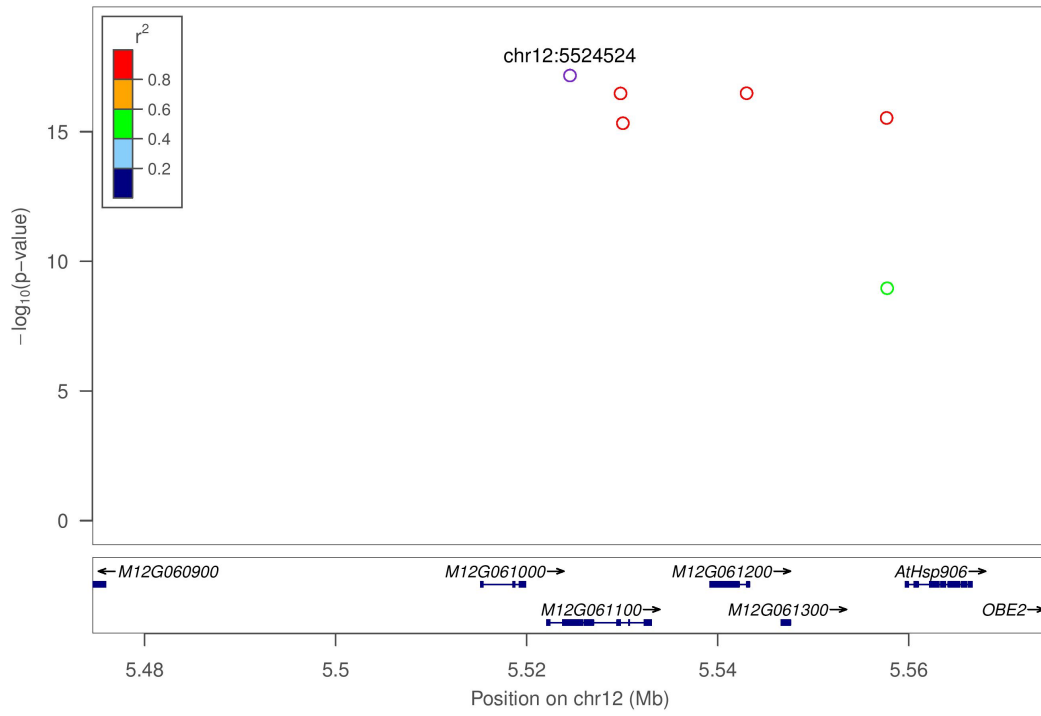
### Apical pubescence (S9\_1588034)



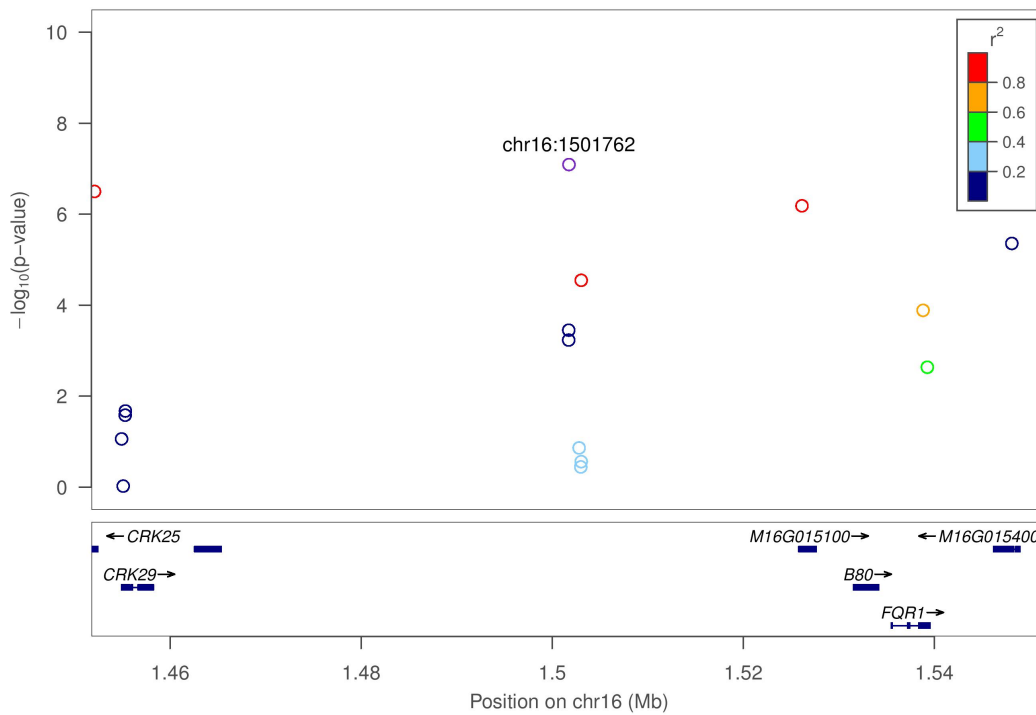
### Apical pubescence (S11\_5727254)



### Apical pubescence (S12\_5524524)

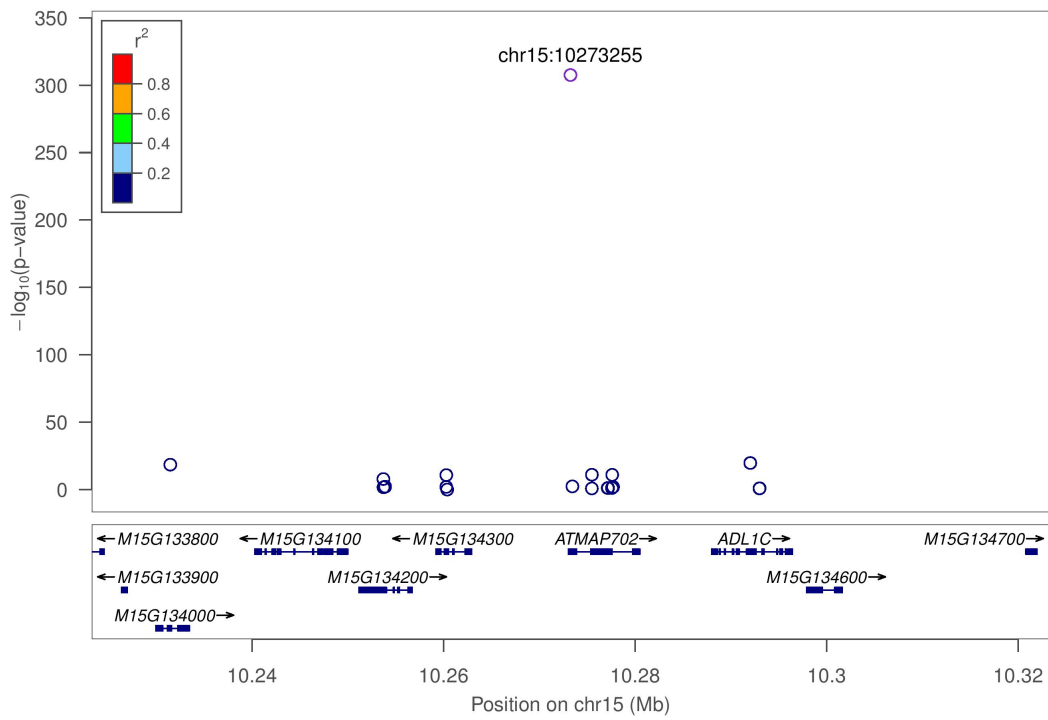


### Apical pubescence (S16\_1501762)

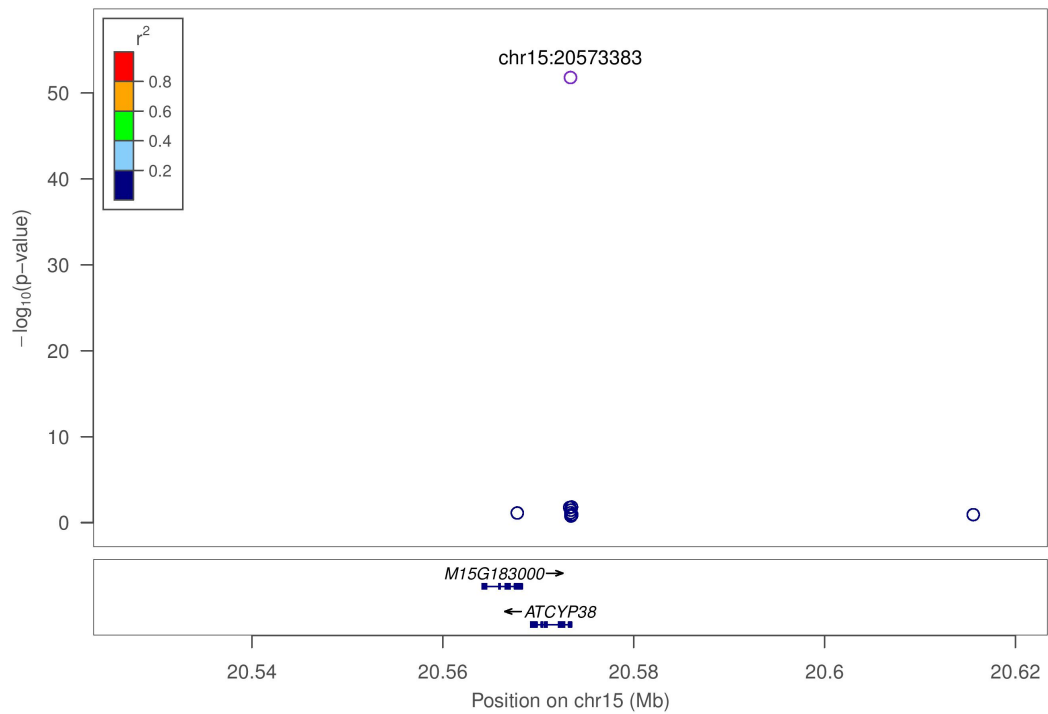


D

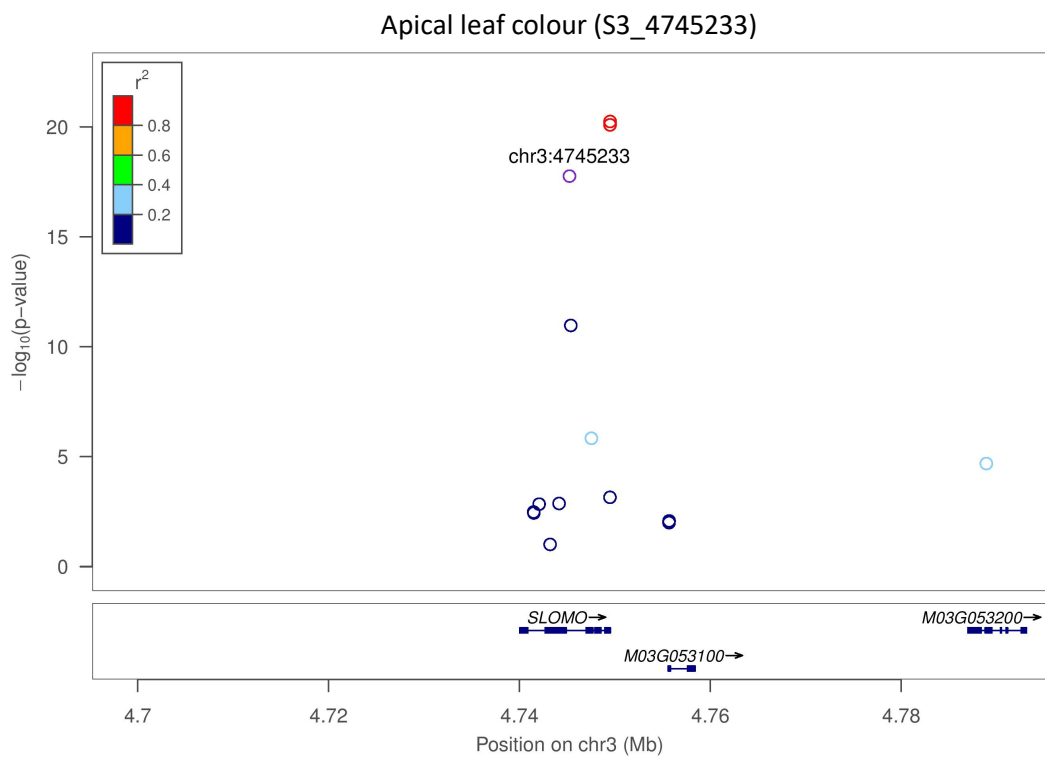
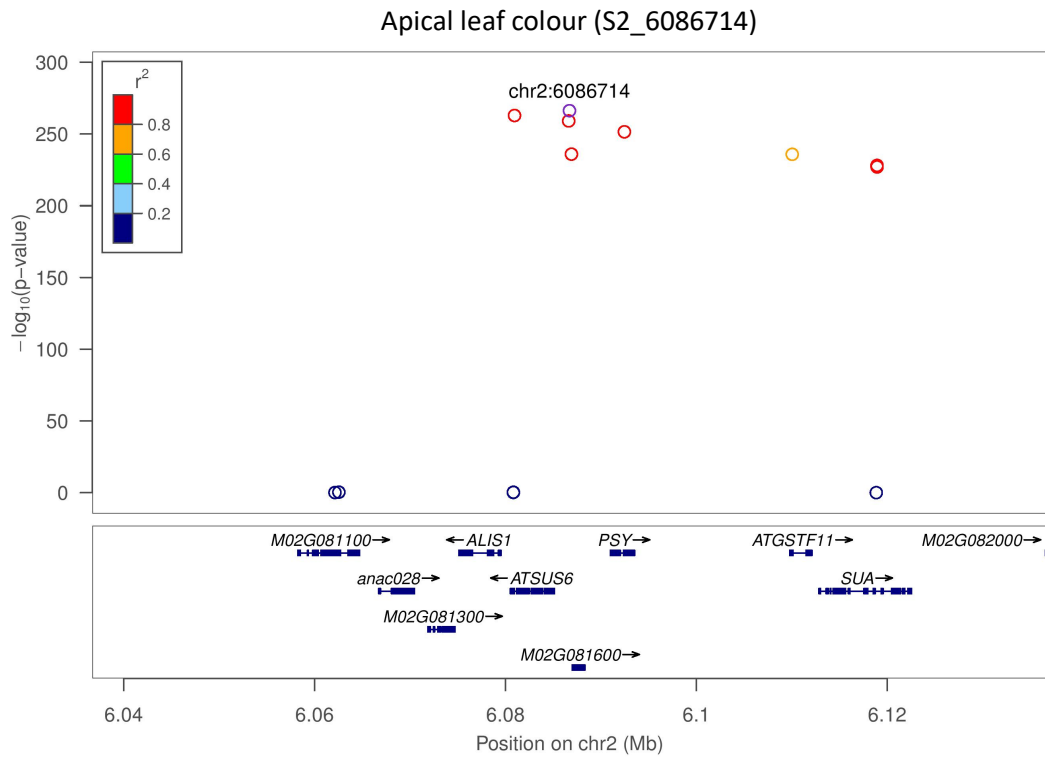
Leaf shape (S15\_10273255)



Leaf shape (S15\_20573383)

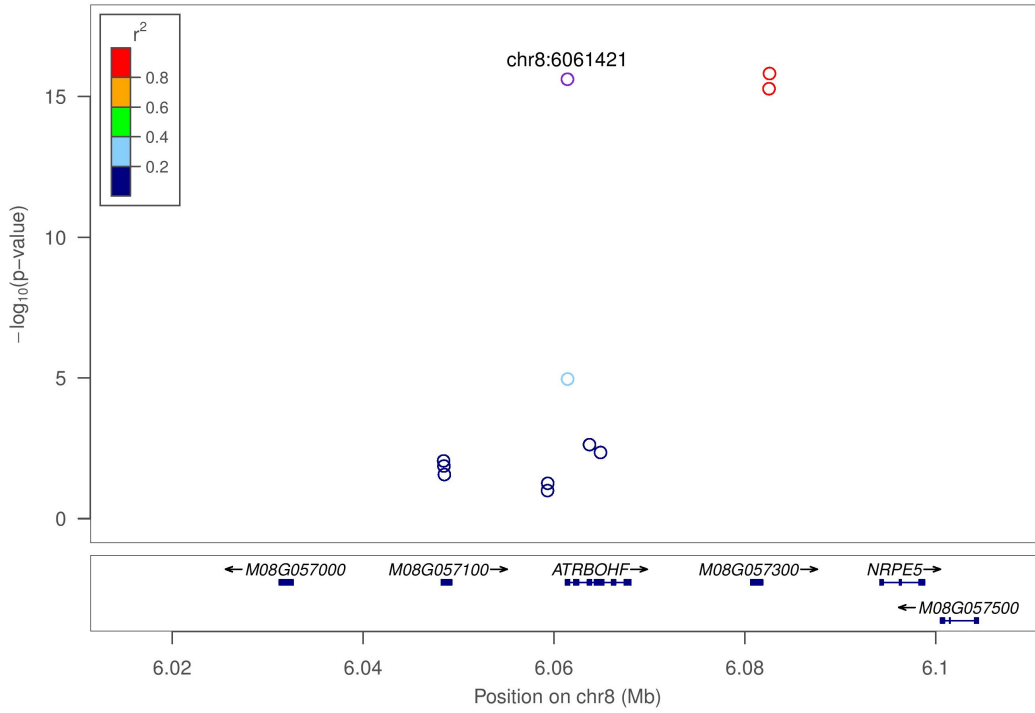


E



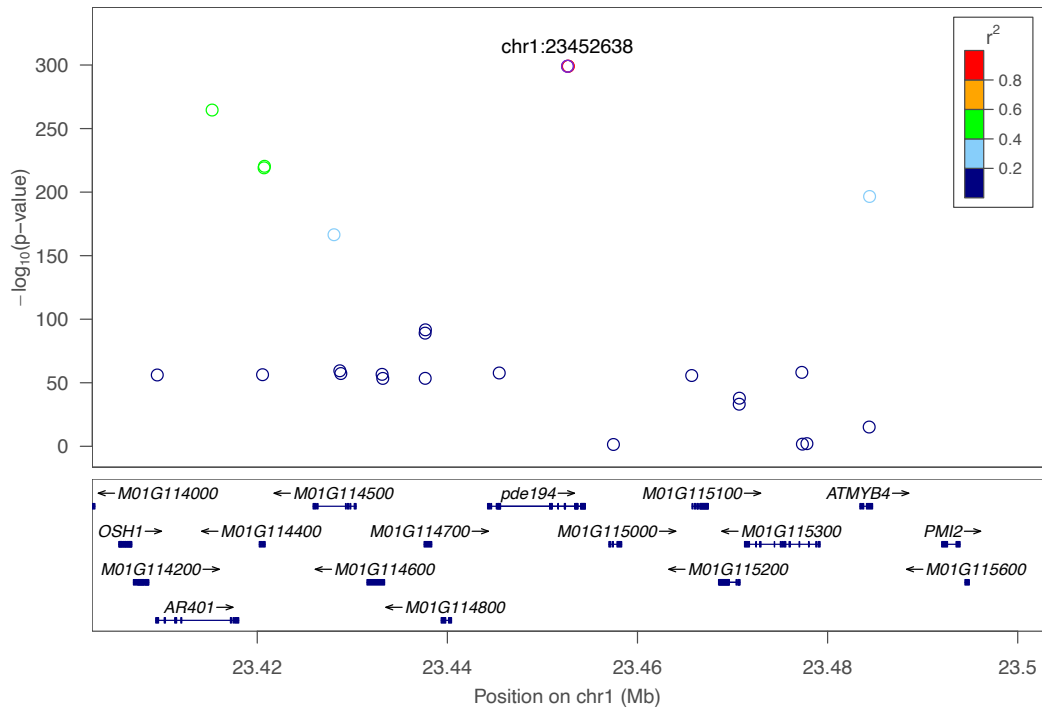


### Apical leaf colour (S8\_6061421)



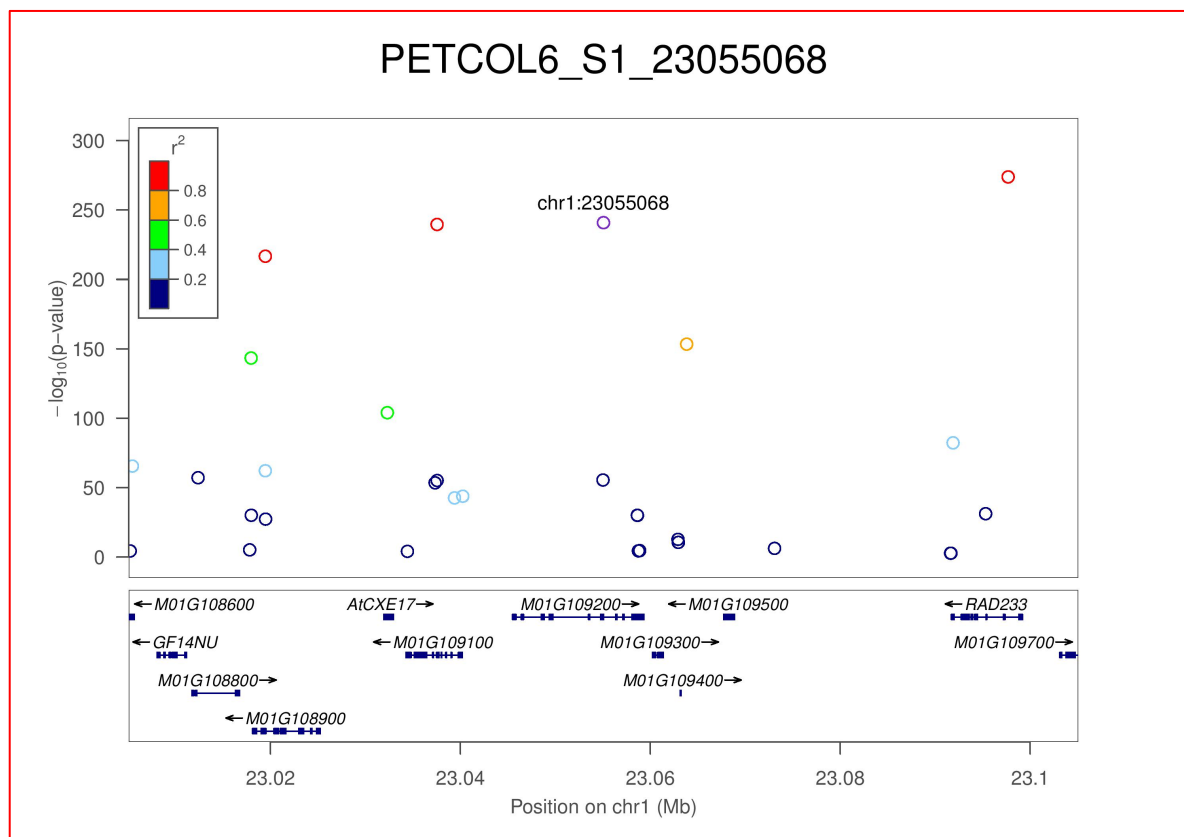
### F

### Mature leaf greenness (S1\_23452638)

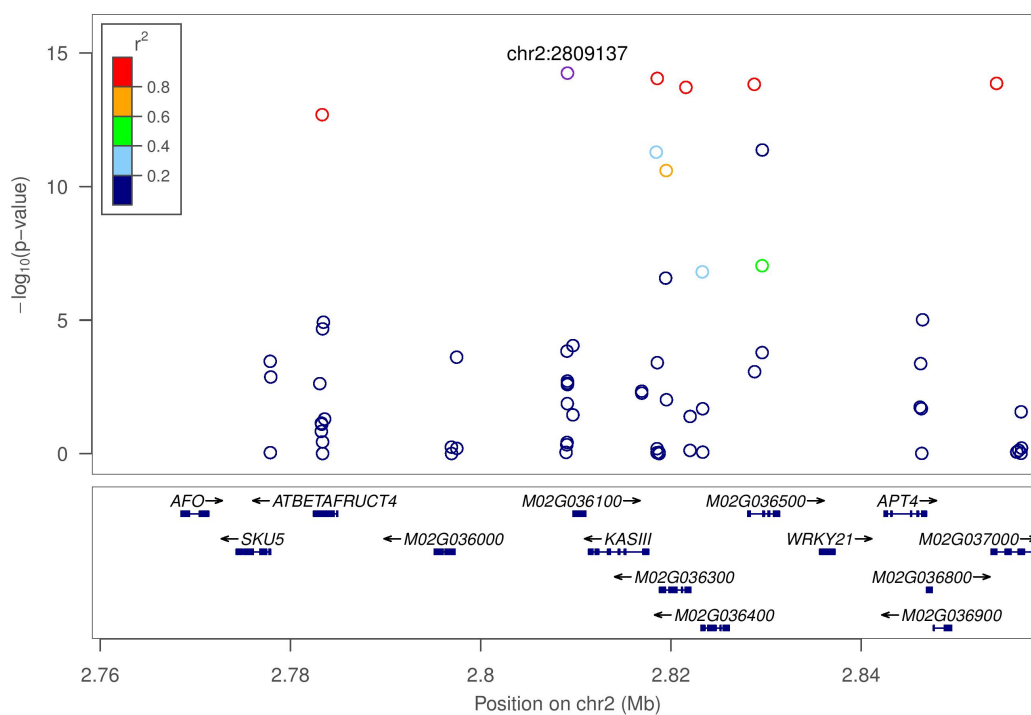


**G**

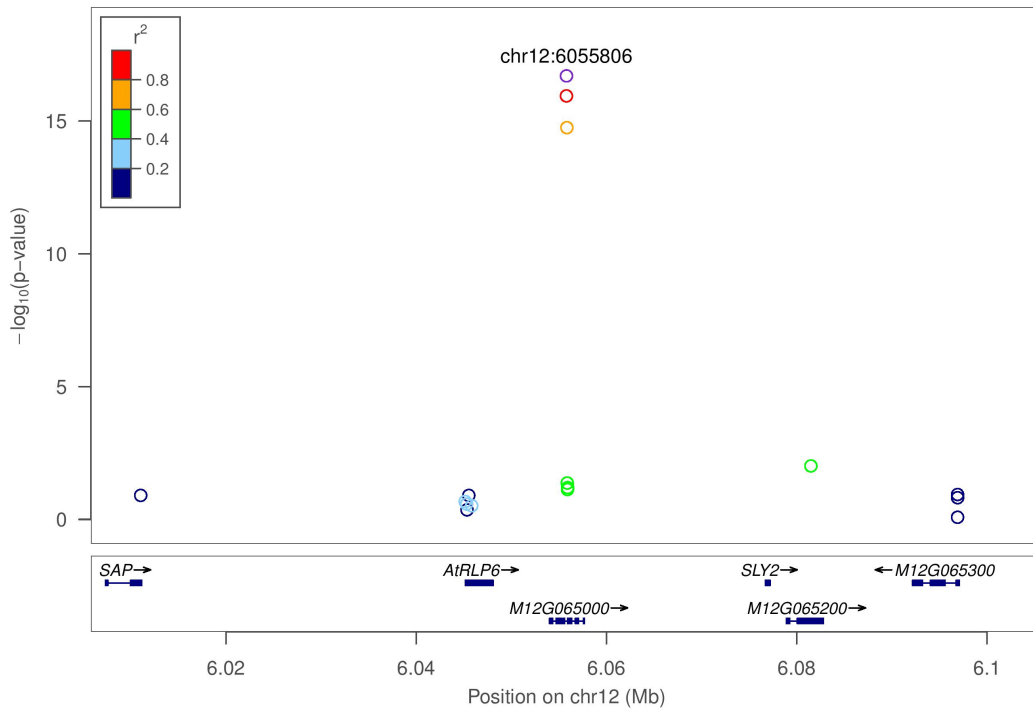
Petiole colour (S1\_23055068)

**H**

Harvest index (S2\_2809137)

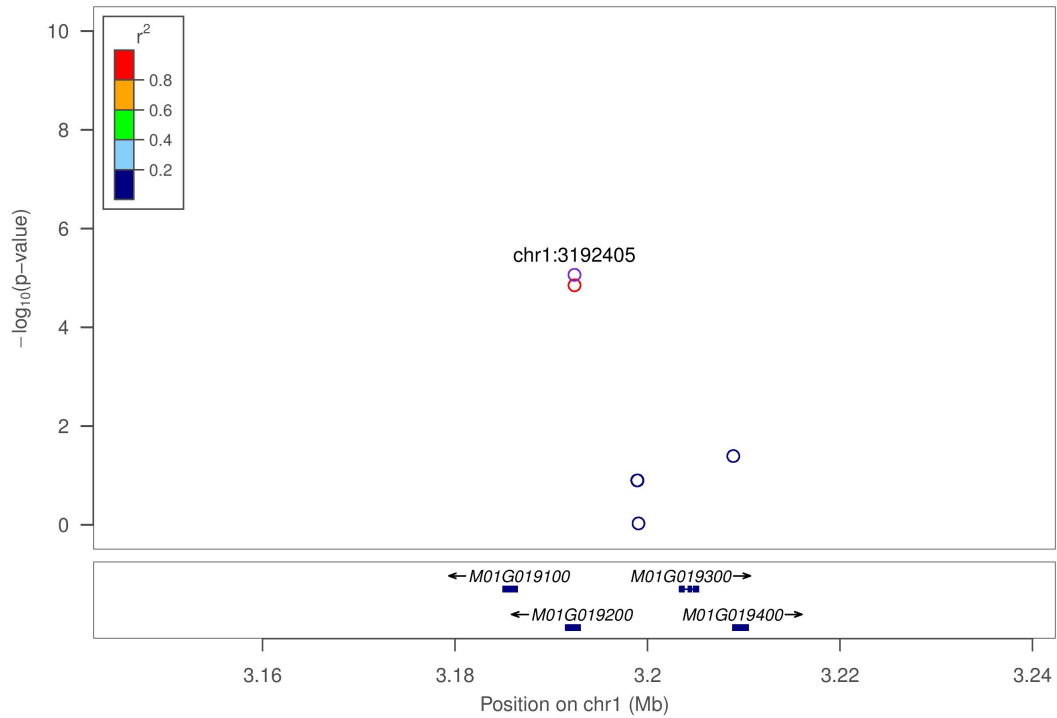


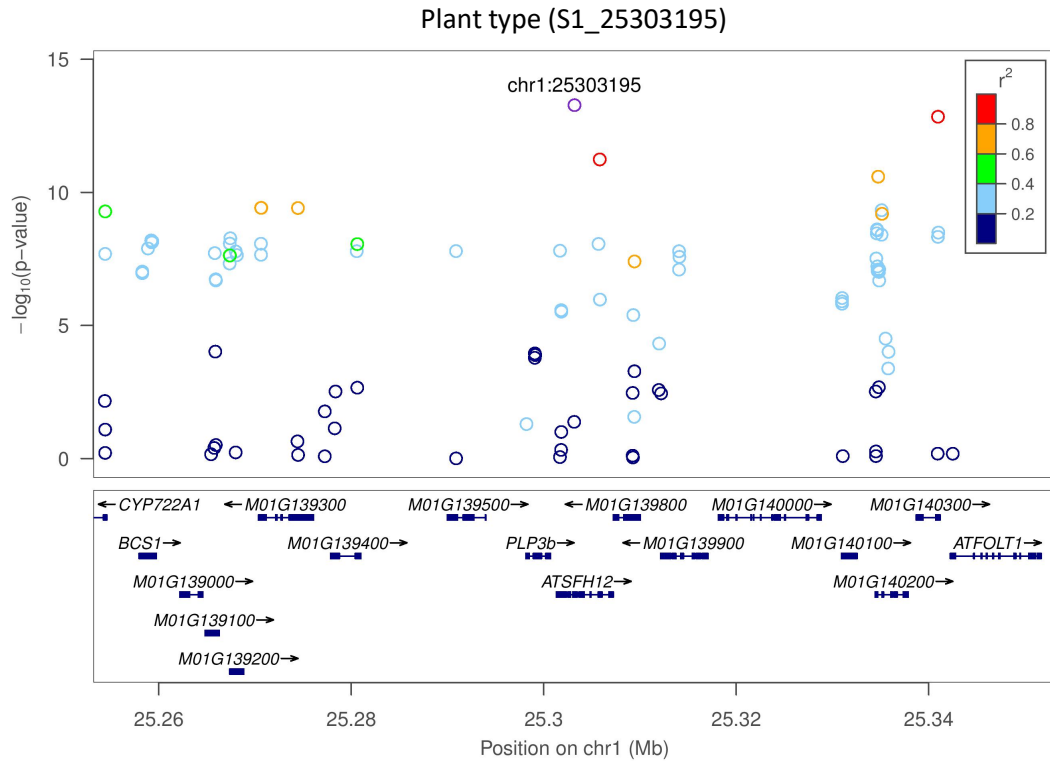
### Harvest index (S12\_6055806)



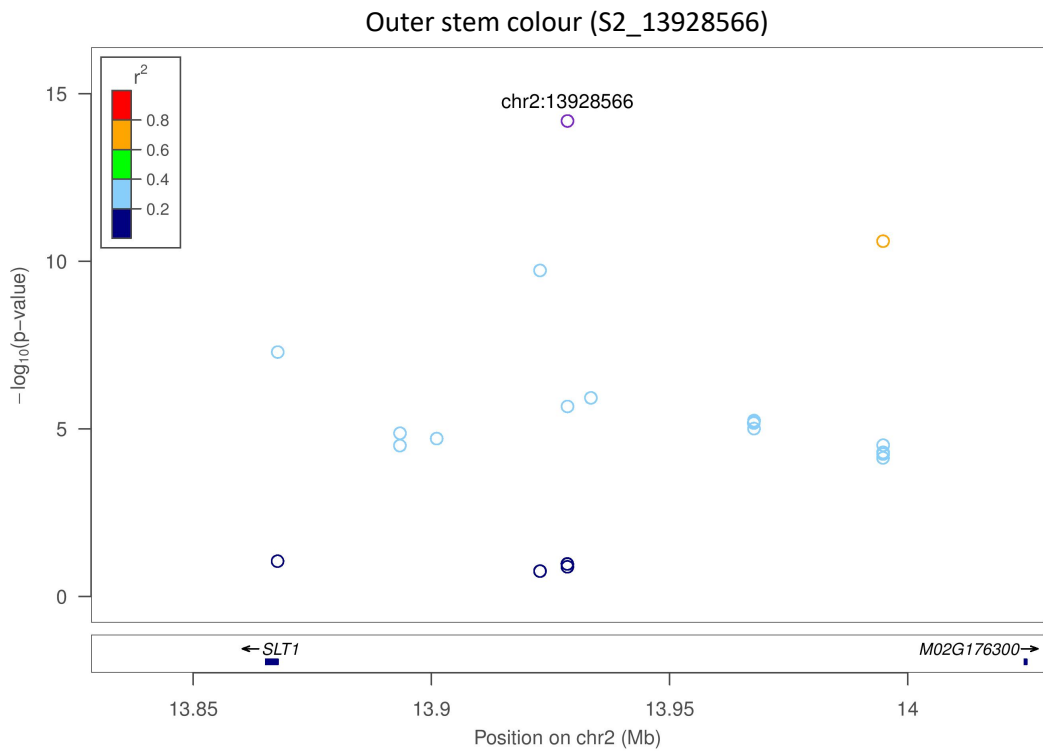
I

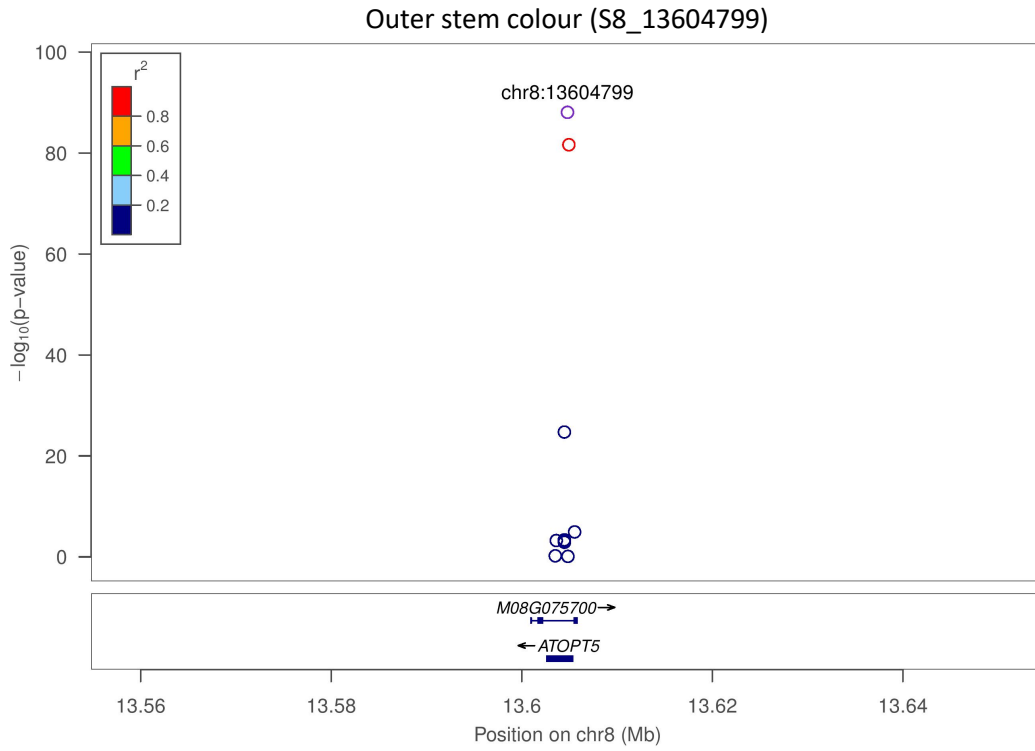
### Plant type (S1\_3192405)



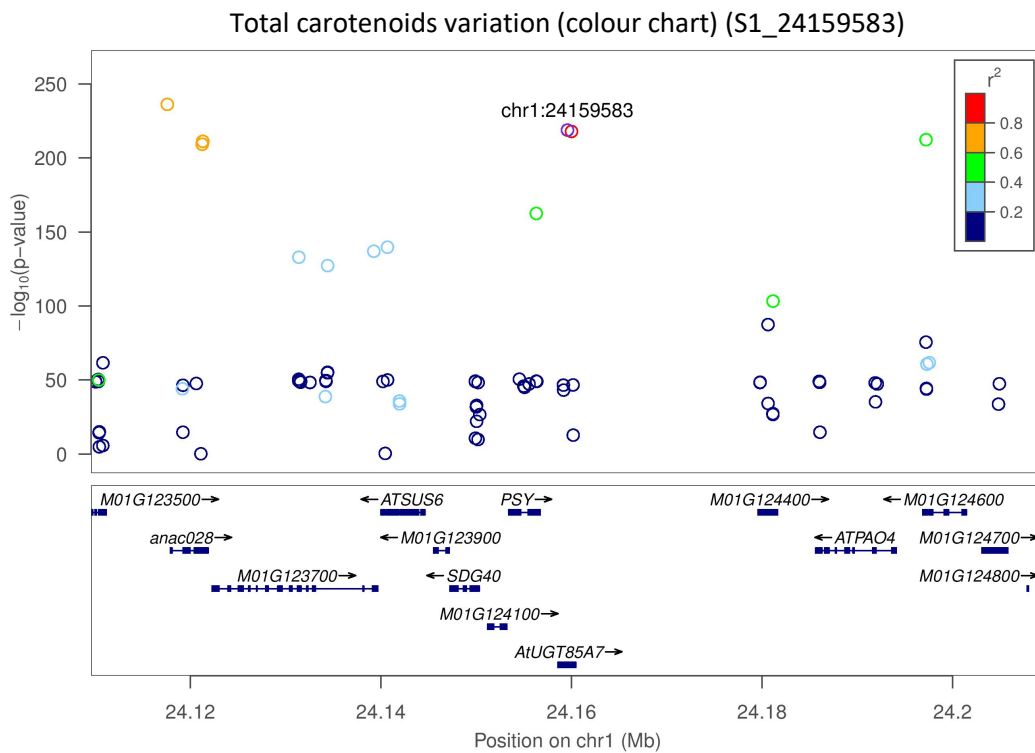


J

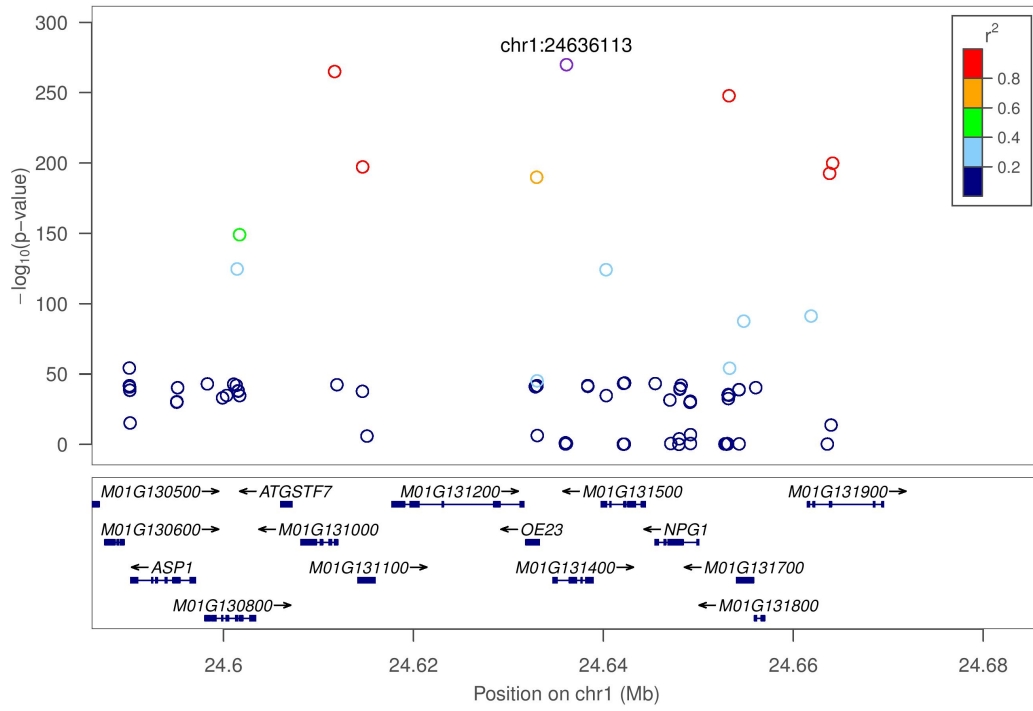




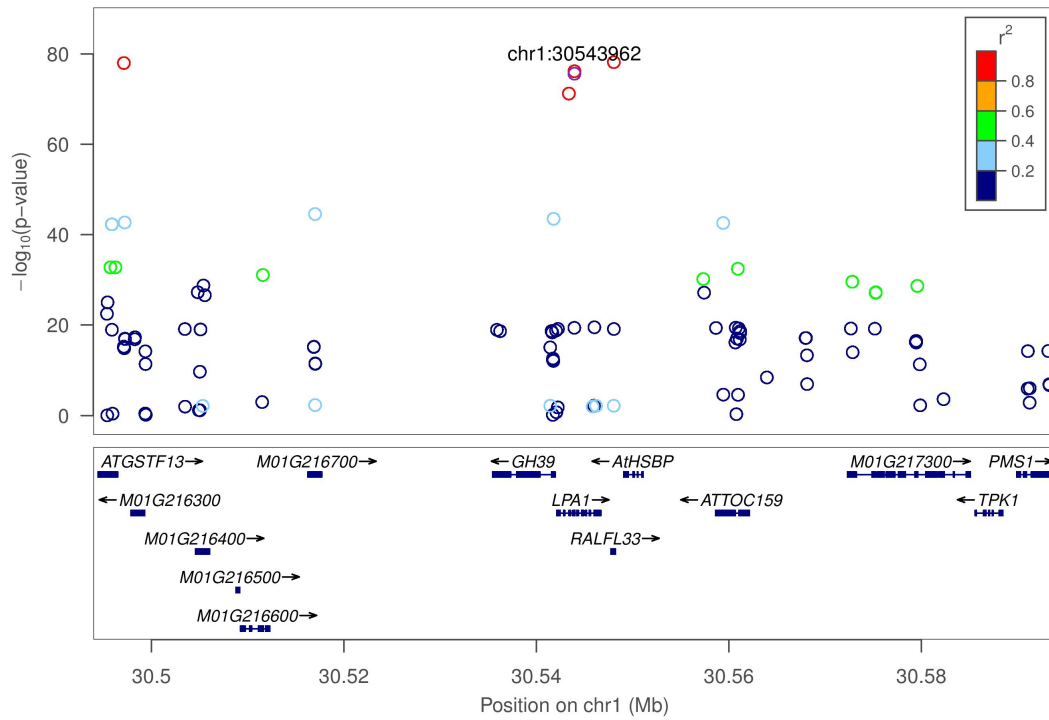
**K**



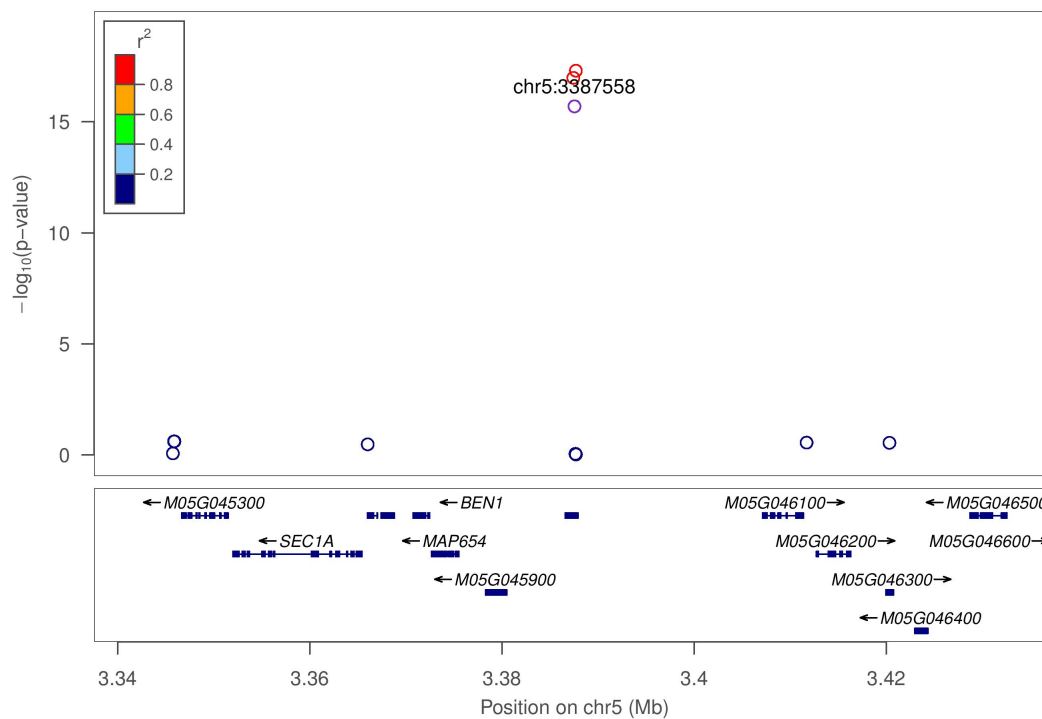
Total carotenoids variation (colour chart) (S1\_24636113)



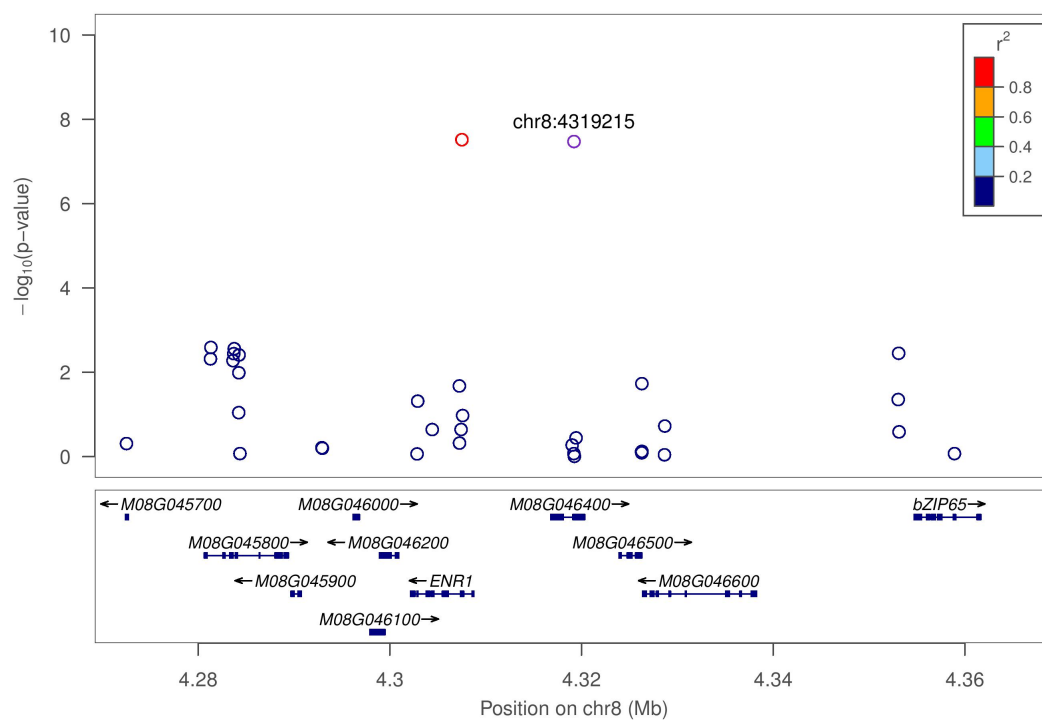
Total carotenoids variation (colour chart) (S1\_30543962)



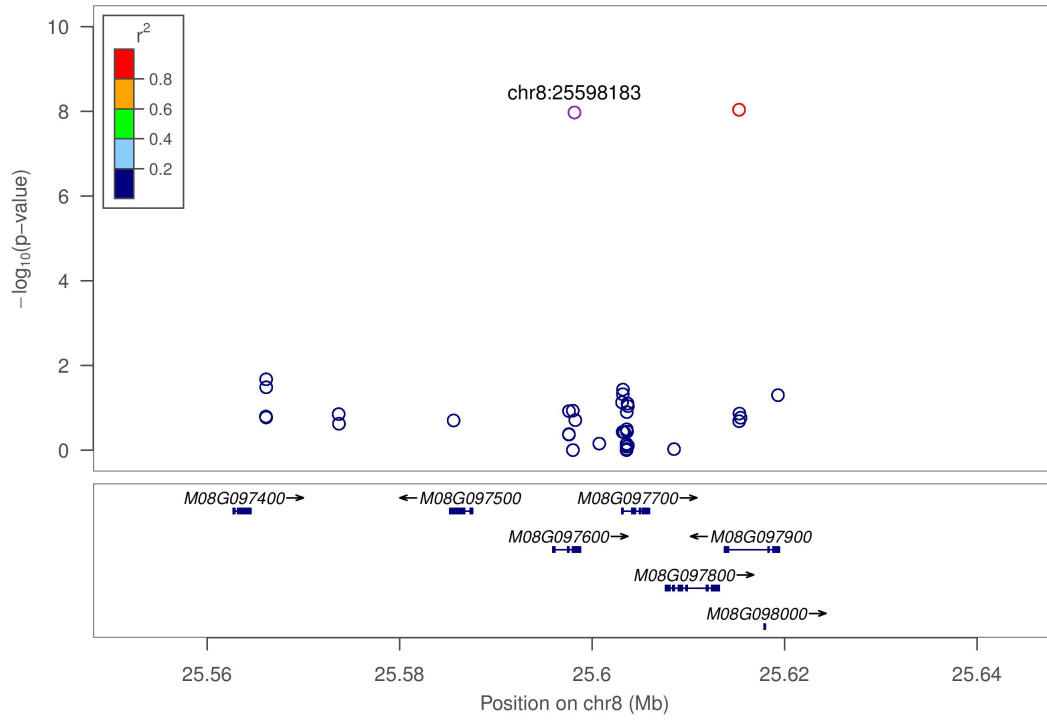
Total carotenoids variation (colour chart) (S5\_3387558)



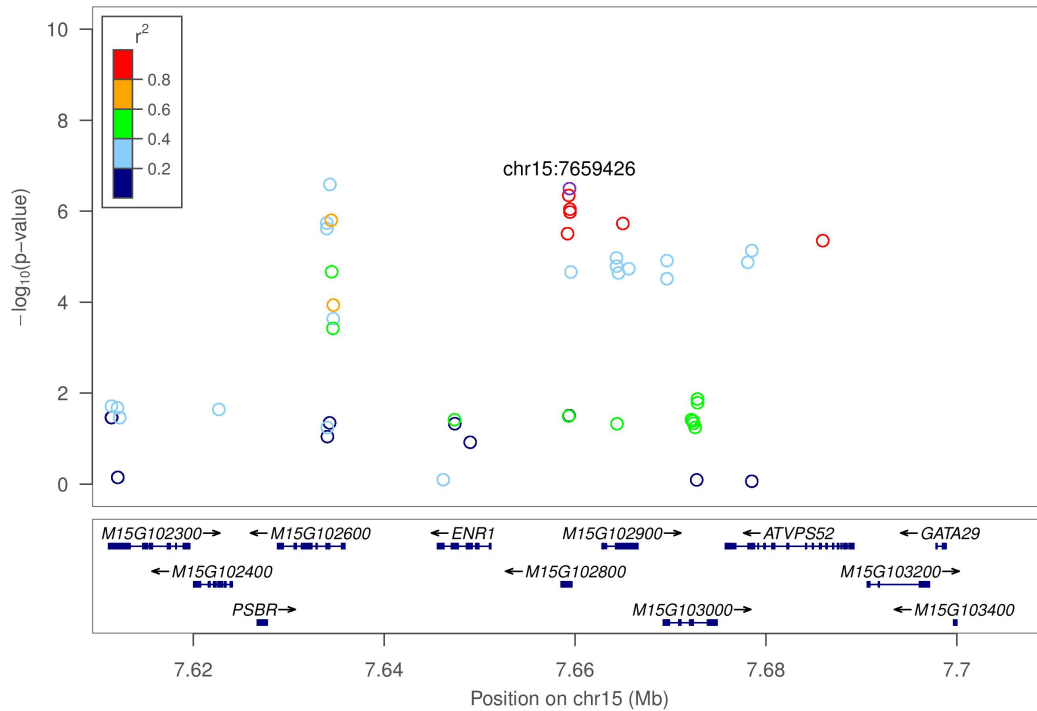
Total carotenoids (colour chart) (S8\_4319215)



Total carotenoids variation (colour chart) (S8\_25598183)

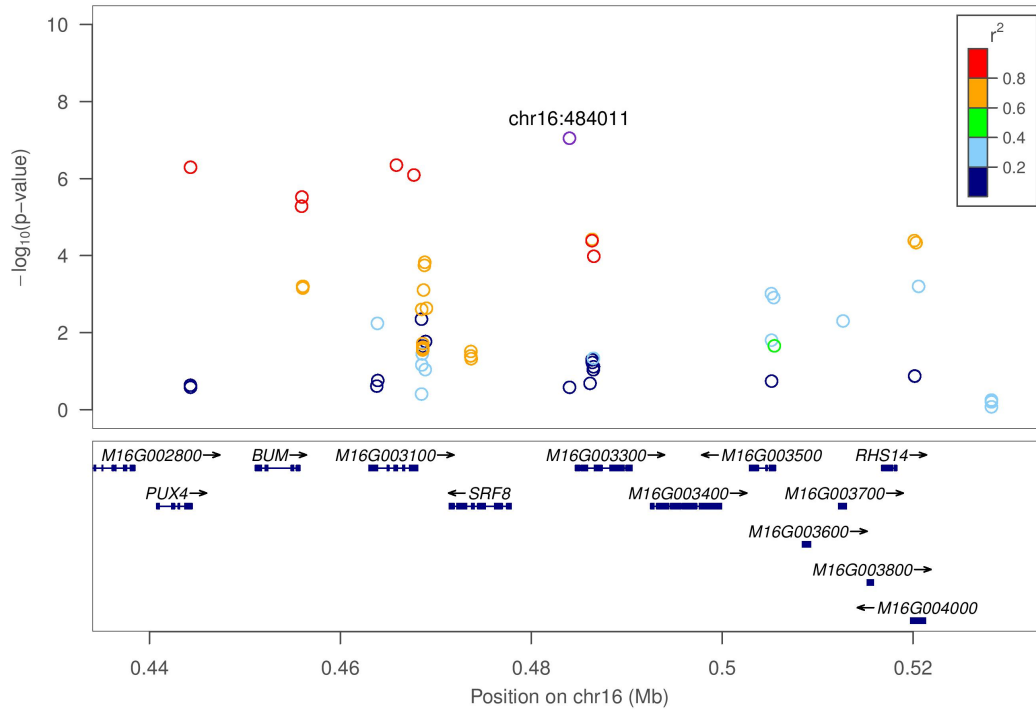


Total carotenoids variation (colour chart) (S15\_7659426)



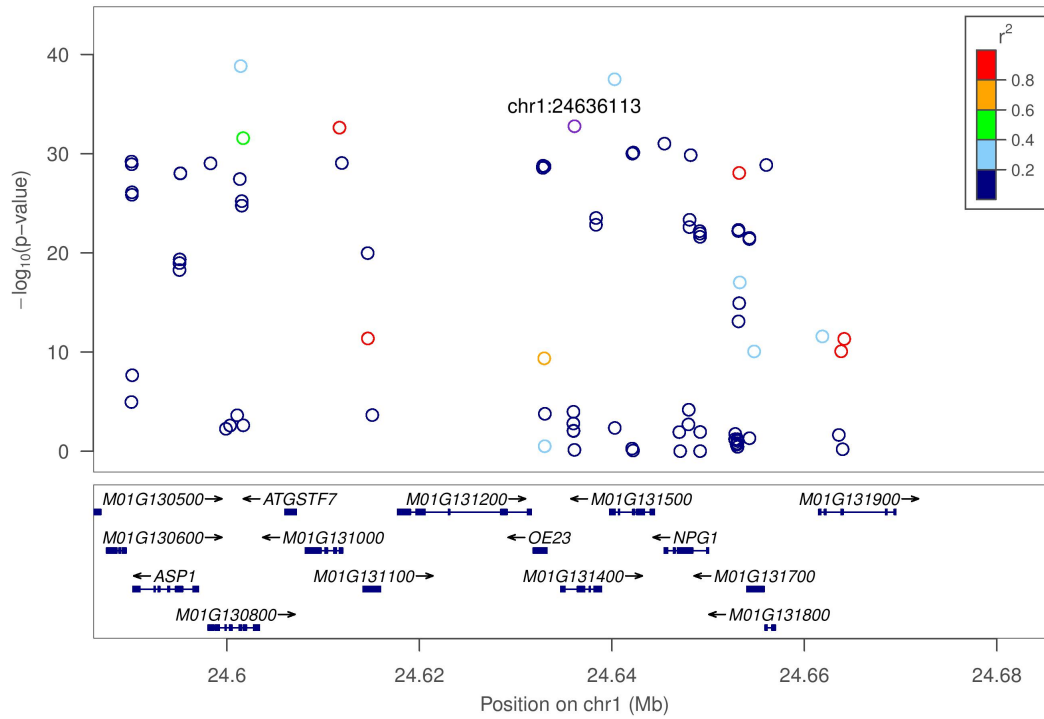


Total carotenoids variation (colour chart) (S16\_484011)

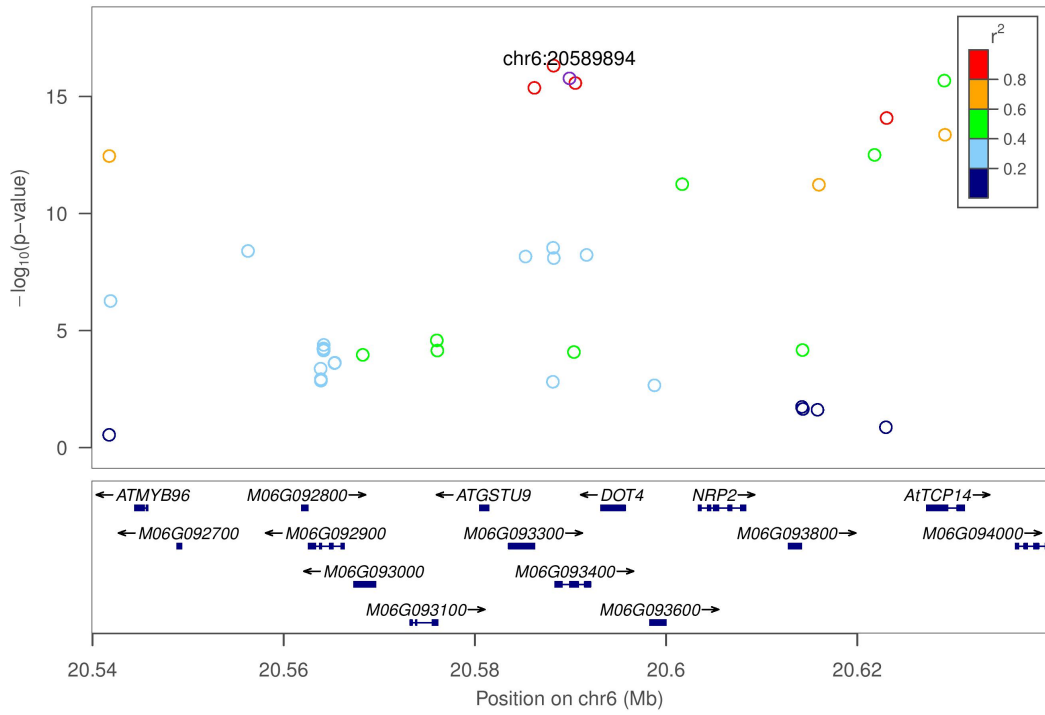


L

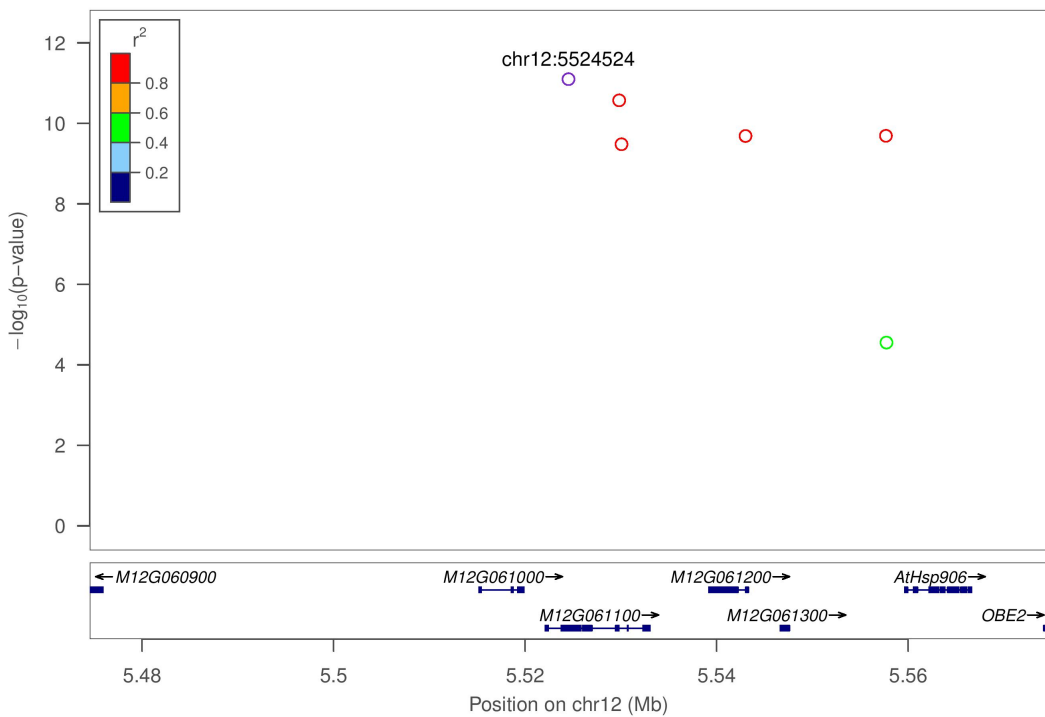
Dry matter content (S1\_24636113)



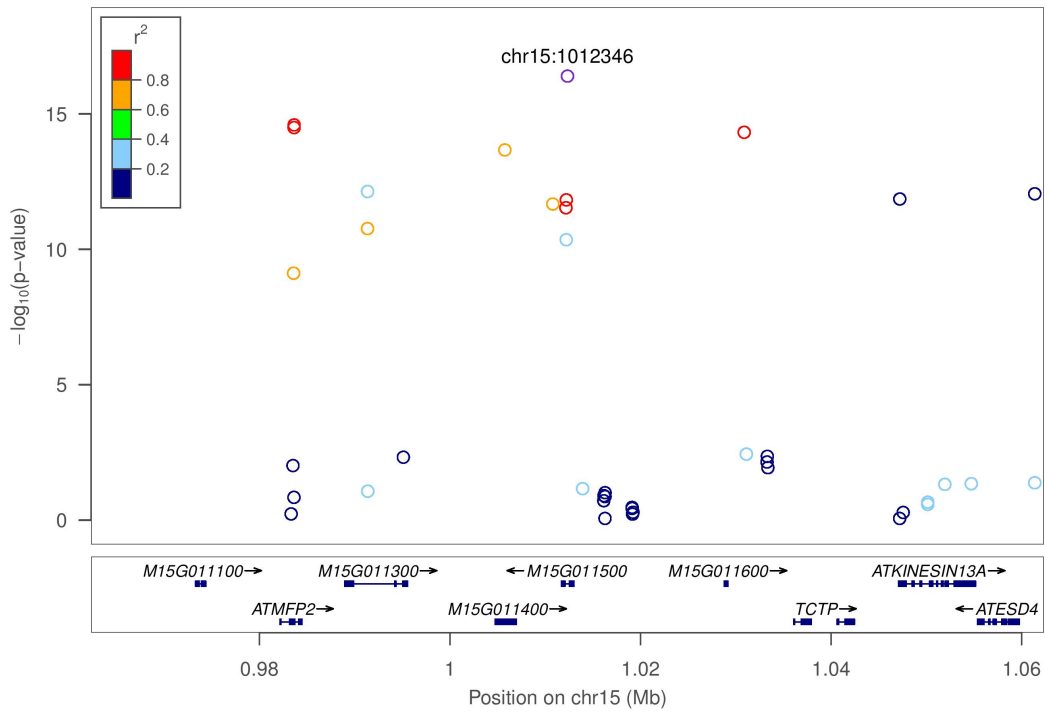
### Dry matter content (S6\_20589894)



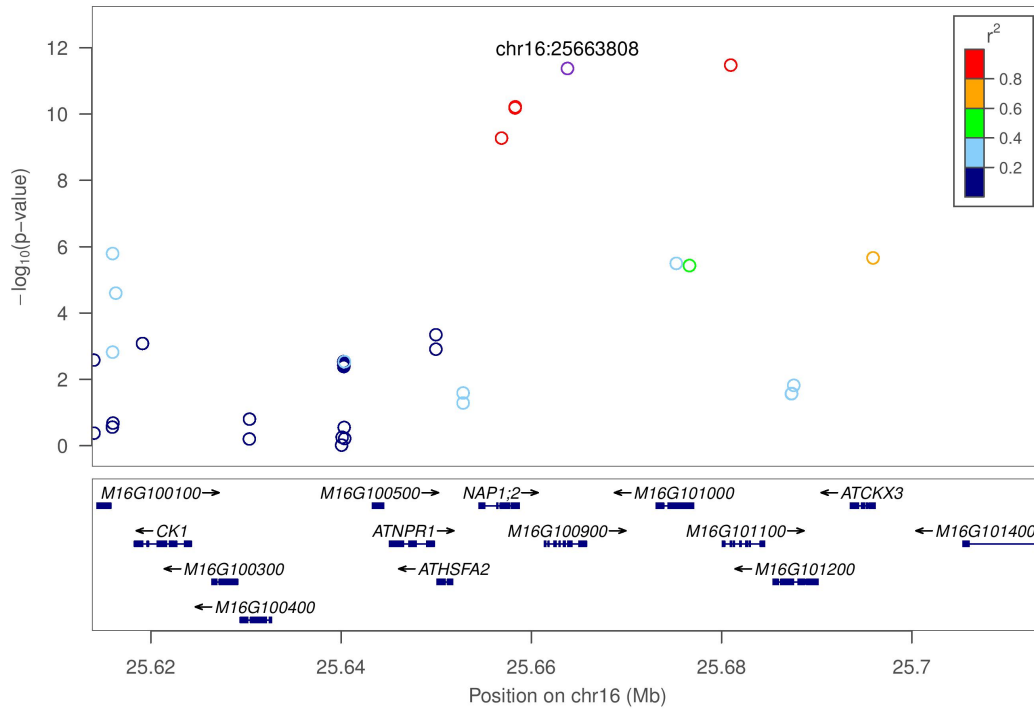
### Dry matter content (S12\_5524524)



### Dry matter content (S15\_1012346)

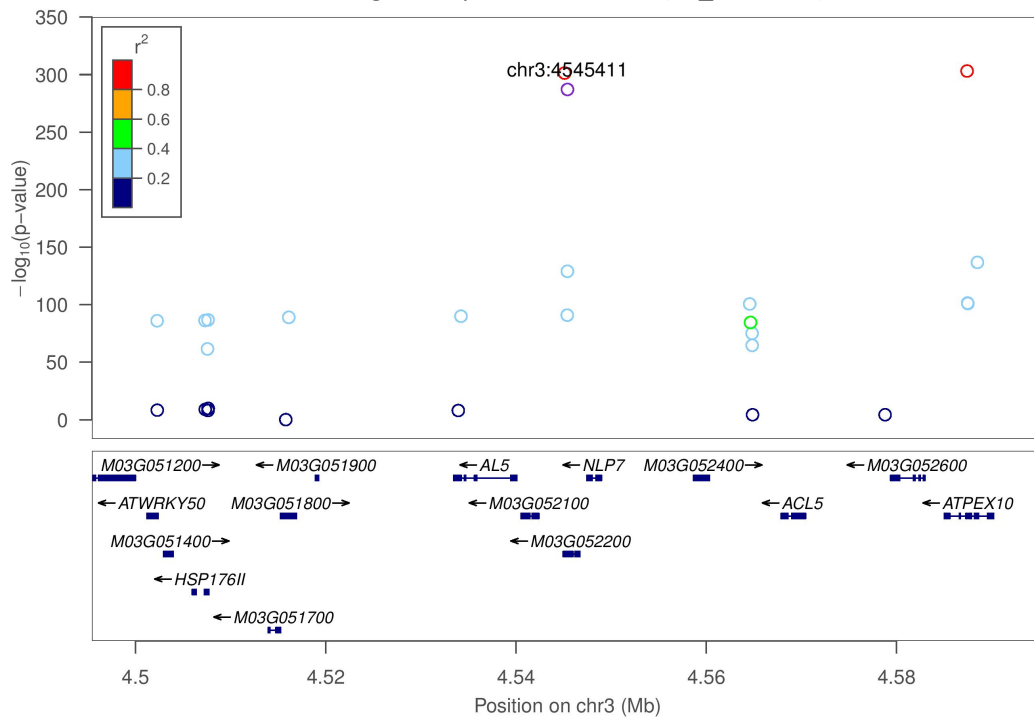


### Dry matter content (S16\_25663808)



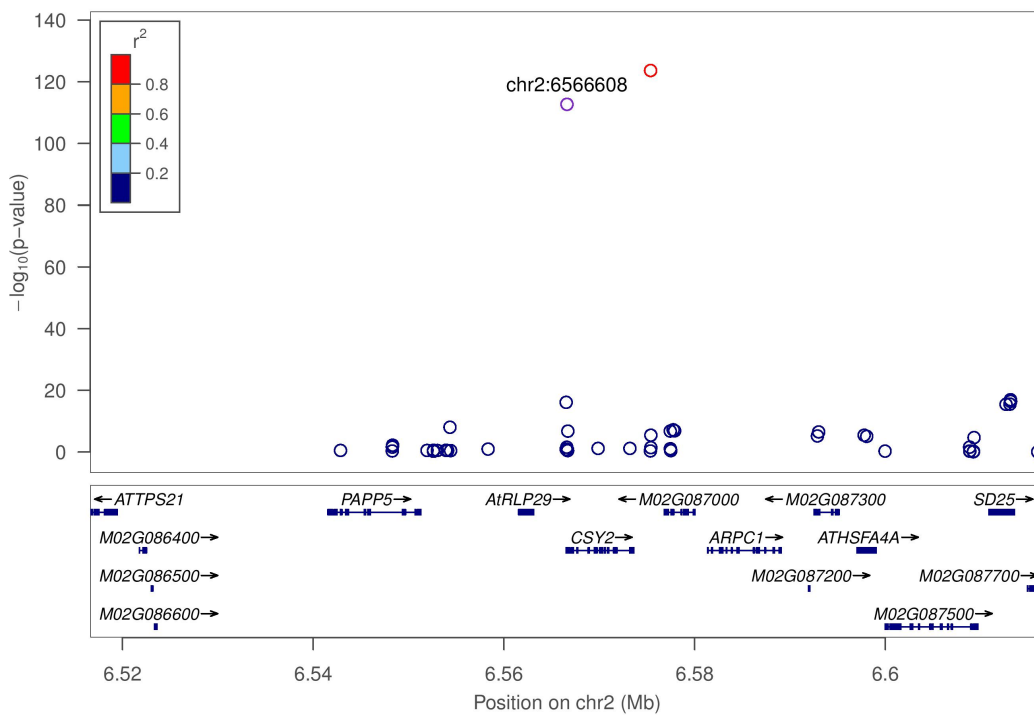
**M**

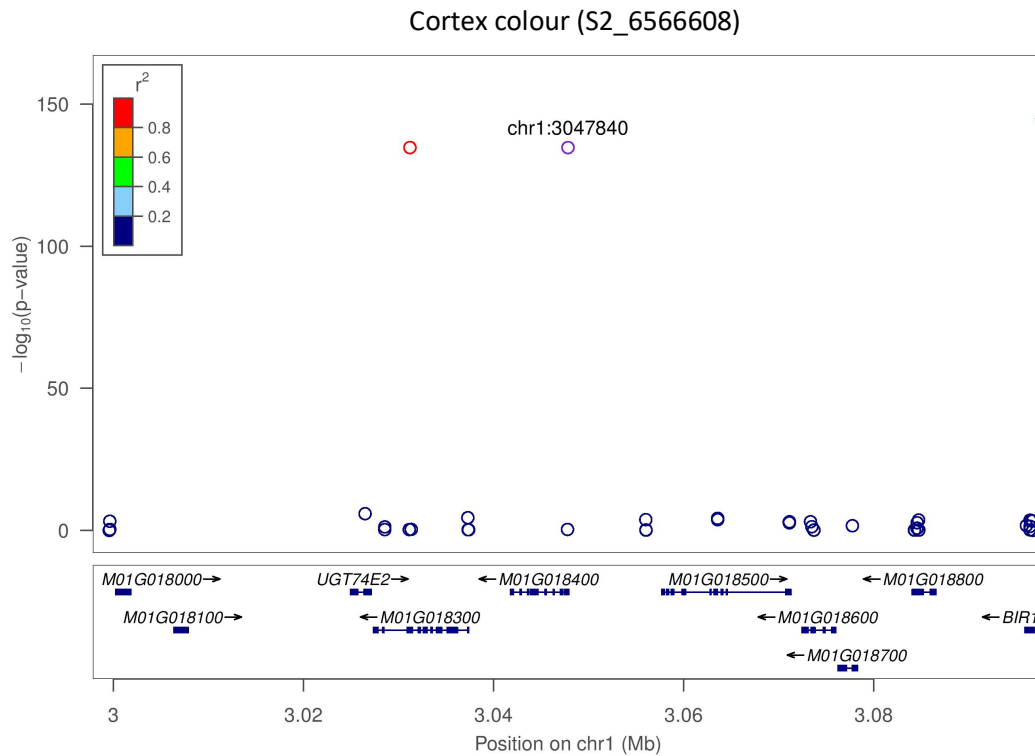
Storage root periderm colour (S3\_4545411)



**N**

Cortex colour (S1\_3047840)





**Figure S4** Regional Manhattan plots for each locus-trait combination. Plots include candidate genes within 100 Kb of the top SNP marker (50 kb upstream, 50 kb downstream) with some adjustments based on the extent of local linkage disequilibrium with the candidate marker. SNPs are color coded based on linkage disequilibrium with the top marker. (A) CMD severity; (B) CGM severity; (C) Apical leaf pubescence; (D) Leaf shape; (E) Apical leaf colour; (F) Mature leaf greenness; (G) Petiole colour; (H) Harvest index; (I) Plant type; (J) Outer stem colour; (K) Total carotenoids content; (L) Dry matter content; (M) Periderm colour; (N) Root cortex colour.