Supplement 1:

MultiGWAS report for Full GWAS model with filters

Table of Contents

MultiGWAS is a tool to do GWAS analysis in tetraploid organisms by executing in parallel and integrating the results from four existing GWAS software: two available for polyploids (GWASpoly and SHEsis) and two frequently used for diploids (GAPIT and TASSEL).

MultiGWAS executes two types of GWAS models: Full and Naive. The Full model with control for population structure and individual relatedness and the Naive model without any control. In both models, users can apply different control quality filters for the genomic data.

This report, created by the MultiGWAS tool, is a summary of the input parameters used to run the tool and its main outputs, including (1) Score tables with detailed information on the associations for each tool. (2) Venn diagrams of shared SNPs among the four tools. (3) Heatmaps of significative SNP profiles among the four tools. (4) Manhattan and QQ plots for the association found by each tool. And (5) Chord diagrams for the chromosomes vs. SNP by each tool.

1 Input Parameters

MultiGWAS uses as input a configuration file where the user specifies the input genomic data (genotype and phenotype files) along with different values for parameters used by the MultiGWAS tool and by the other four GWAS software (GWASpoly, SHEsis, GAPIT, and TASSEL).

The following table shows the current input parameters specified by the user in the configuration file:

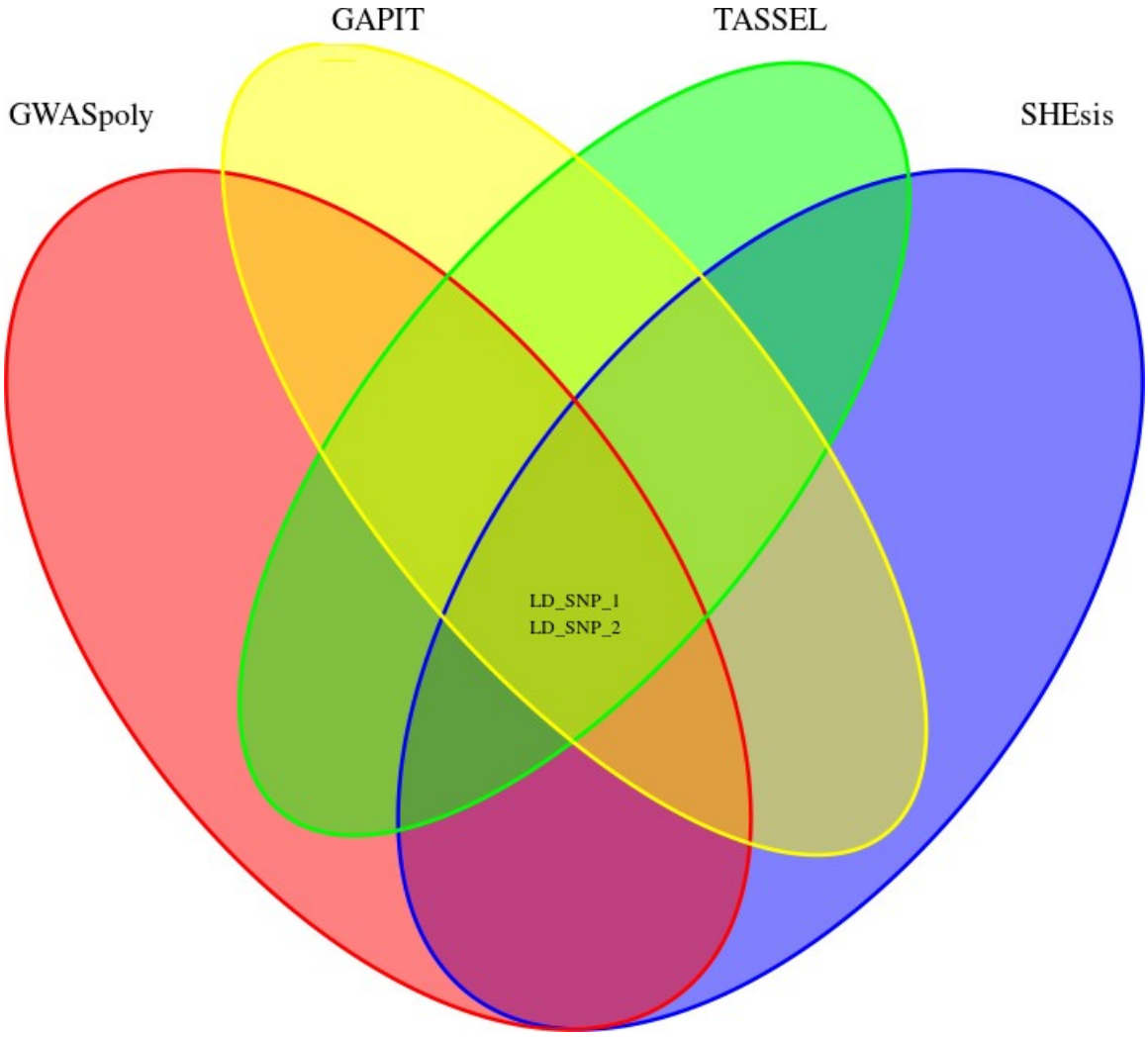
PARAMETER	VALUE
Genotype filename	example-genotype-tetra-gwaspoly-ACGT.csv
Phenotype filename	tuber_shape.csv
Genotype format (gwaspoly, matrix, vcf, updog, fitpoly)	gwaspoly
Map filename	
Ploidy (4 or 2)	4
Significance level (Genome-wide significance level)	0.05
Correction method (Bonferroni or FDR)	Bonferroni
GWAS model (Full or Naive)	full
Filtering (TRUE or FALSE)	TRUE
MIND Filter (Individual with missing genotype)	0.1
GENO Filter (SNPs with missing genotype)	0.1
MAF Filter (Minor allele frequency)	0.01
HWE Filter (Hardy-Weinberg test)	1E-10
R2 LD (Linkage disequilibrium threshold)	0.9
GWAS Tools	gwaspoly shesis gapit tassel
nBest (Number of top SNPs to be reported)	10

2 SNPs in Linkage Disequilibrium above $R^2 = 0.9$

MultiGWAS reports a Venn diagram and a table for pairs of SNPs with squared correlation equal to or greater than the threshold R^2, where R^2 is defined by the user in the configuration file. MultiGWAS joins the N best associations found for each GWAS packages (SNPs with the lowest *p-value*), calculates for each pair of SNPs the R^2, and summarizes the results in a table with pairs of SNPs per row along with their calculated R^2.

Pairs of SNPs in LD are assigned a new ID (LD_SNP) and reported in a Venn diagram highlighting the shared SNPs in LD detected between the GWAS software. This view allows for quick identification of related SNPs with different names instead of a plain table, as most GWAS packages report their results.

LD_SNP	SNP1	SNP2	R2
LD_SNP_1	c1_8019	c2_25471	0.93
LD_SNP_2	c2_45611	c2_45606	1.00



2 Best-ranked SNPs

This section shows a table and a Venn diagram for the best ranked SNPs (For this report, N=10). The configuration file allows defining the parameter N.

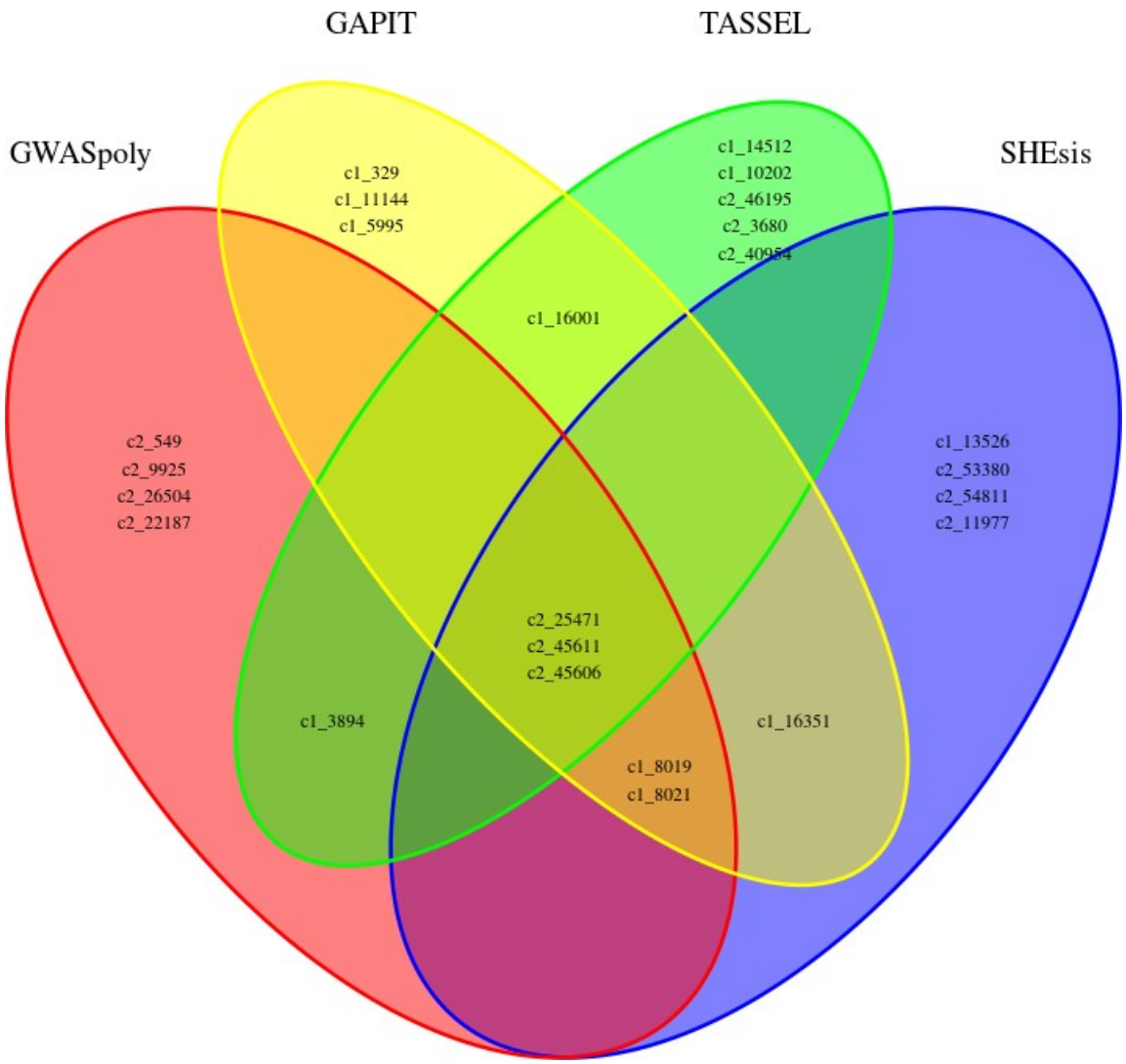
2.1 Table of best-ranked SNPs

The following table shows the associations between SNPs and the trait as transformed p-values to scores as -log 10 (p-value) sorted from the highest to the lowest scores based on each tool analysis and the model selected (Full or Naive). Each SNP has its associated information, such as the chromosome, position in the genome, and name. Also, based on the threshold calculated by a multiple testing correction (i.e., Bonferroni or FDR), the table marks the significance of each SNP as a TRUE or FALSE.

TOOL	MODEL	GC	SNP	CHROM	POSITION	PVALUE	SCORE	THRESHOLD	SIGNIFICANCE
GWASpoly	additive	1.016	c1_8019	10	48863165	7E-06	5.18	4.65	TRUE
GWASpoly	additive	1.016	c2_25471	10	48808404	1.5E-05	4.82	4.65	TRUE
GWASpoly	additive	1.016	c2_45611	10	48203431	0.000132	3.88	4.65	FALSE
GWASpoly	additive	1.016	c2_549	9	16527499	0.000151	3.82	4.65	FALSE
GWASpoly	additive	1.016	c2_45606	10	48218826	0.000174	3.76	4.65	FALSE
GWASpoly	additive	1.016	c2_9925	1	81410256	0.000589	3.23	4.65	FALSE
GWASpoly	additive	1.016	c1_8021	10	48862950	0.000661	3.18	4.65	FALSE
GWASpoly	additive	1.016	c2_26504	9	21610960	0.001349	2.87	4.65	FALSE
GWASpoly	additive	1.016	c1_3894	1	71421420	0.001549	2.81	4.65	FALSE
GWASpoly	additive	1.016	c2_22187	11	40777537	0.001698	2.77	4.65	FALSE
SHEsis	additive	3.97	c1_8019	10	48863165	0	11.4559	5.3836	TRUE
SHEsis	additive	3.97	c1_13526	10	48020996	0	10.7986	5.3836	TRUE
SHEsis	additive	3.97	c2_53380	1	70371898	0	10.6073	5.3836	TRUE
SHEsis	additive	3.97	c2_25471	10	48808404	0	9.6253	5.3836	TRUE
SHEsis	additive	3.97	c2_54811	1	46270954	0	9.5467	5.3836	TRUE
SHEsis	additive	3.97	c2_45606	10	48218826	0	8.2013	5.3836	TRUE
SHEsis	additive	3.97	c2_11977	5	3515956	0	8.1656	5.3836	TRUE
SHEsis	additive	3.97	c1_8021	10	48862950	0	7.9872	5.3836	TRUE
SHEsis	additive	3.97	c1_16351	10	48761642	0	7.9666	5.3836	TRUE
SHEsis	additive	3.97	c2_45611	10	48203431	0	7.7878	5.3836	TRUE
GAPIT	additive	0.902	c2_25471	10	48808404	0.000346	3.4613	4.6847	FALSE
GAPIT	additive	0.902	c1_16001	10	47539878	0.000397	3.4007	4.6847	FALSE
GAPIT	additive	0.902	c2_45611	10	48203431	0.000408	3.3892	4.6847	FALSE
GAPIT	additive	0.902	c2_45606	10	48218826	0.000408	3.3892	4.6847	FALSE
GAPIT	additive	0.902	c1_8019	10	48863165	0.000955	3.0202	4.6847	FALSE
GAPIT	additive	0.902	c1_8021	10	48862950	0.00117	2.932	4.6847	FALSE
GAPIT	additive	0.902	c1_16351	10	48761642	0.001178	2.929	4.6847	FALSE
GAPIT	additive	0.902	c1_329	10	565197	0.001822	2.7394	4.6847	FALSE
GAPIT	additive	0.902	c1_11144	6	56135600	0.003267	2.4858	4.6847	FALSE
GAPIT	additive	0.902	c1_5995	7	55209778	0.003341	2.4762	4.6847	FALSE
TASSEL	additive	0.991	c1_14512	1	43409359	0.002967	2.5277	4.2695	FALSE
TASSEL	additive	0.991	c1_10202	4	70396887	0.005245	2.2803	4.2695	FALSE
TASSEL	additive	0.991	c2_46195	1	64259758	0.005269	2.2783	4.2695	FALSE
TASSEL	additive	0.991	c1_16001	10	47539878	0.005427	2.2655	4.2695	FALSE
TASSEL	additive	0.991	c2_45606	10	48218826	0.006296	2.2009	4.2695	FALSE
TASSEL	additive	0.991	c2_45611	10	48203431	0.006785	2.1685	4.2695	FALSE
TASSEL	additive	0.991	c2_3680	11	39908878	0.007068	2.1507	4.2695	FALSE
TASSEL	additive	0.991	c1_3894	1	71421420	0.008131	2.0898	4.2695	FALSE
TASSEL	additive	0.991	c2_40954	1	63756796	0.008774	2.0568	4.2695	FALSE
TASSEL	additive	0.991	c2_25471	10	48808404	0.008997	2.0459	4.2695	FALSE

2.2 Venn diagram of best-ranked SNPs

The Venn diagram shows the best-ranked SNPs with high scores that are either unique or shared by the four software. Shared SNPs appear in the intersection areas of the diagram.



3 Significative SNPs

This section shows a table and a Venn diagram for the signifcative SNPs (score is above the significance threshold for each tool).

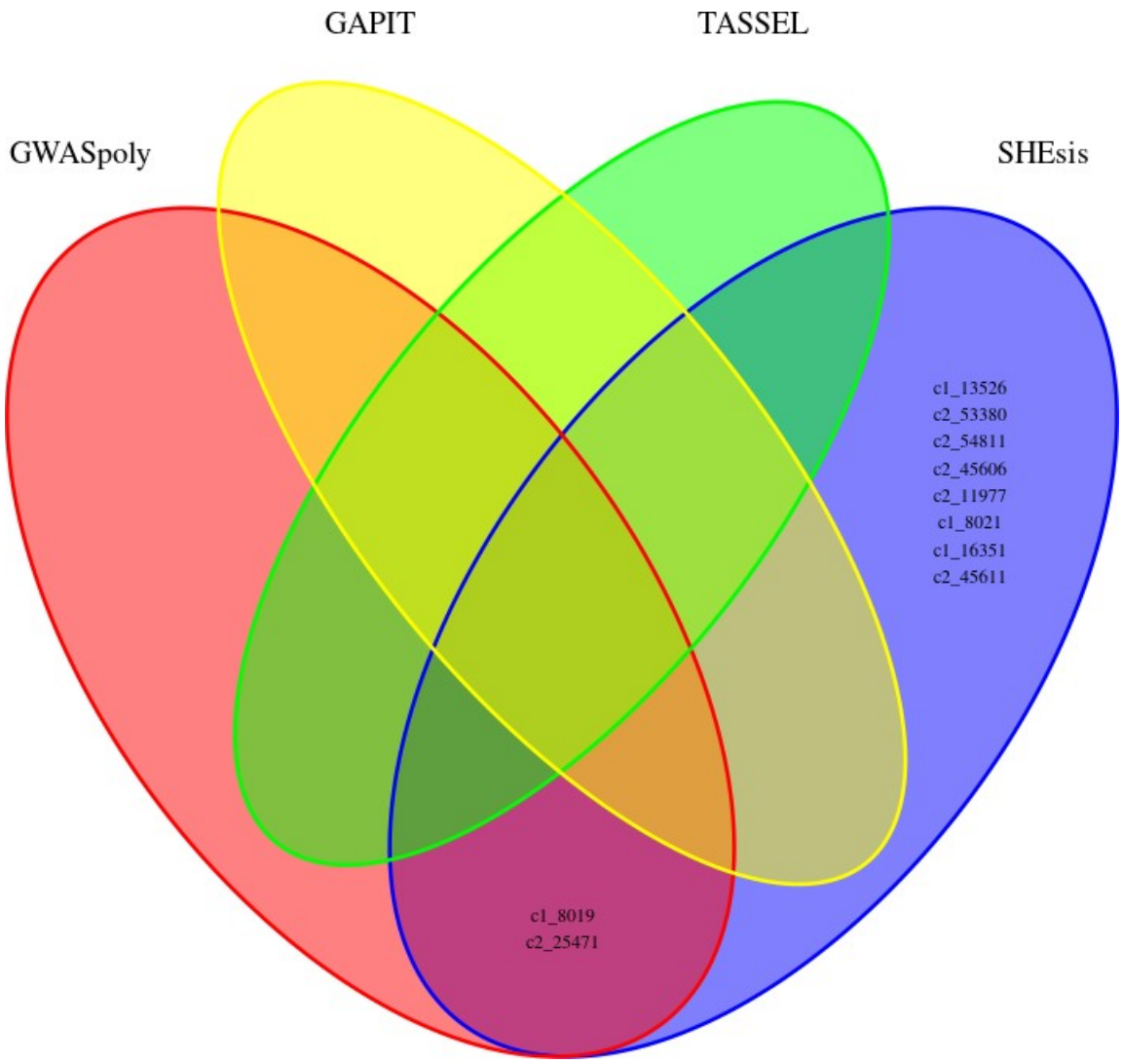
3.1 Table of significative SNPs

The following table shows the associations between SNPs and the trait as transformed p-values to scores as -log 10 (p-value) sorted from the highest to the lowest scores based on each tool analysis and the model selected (Full or Naive). Each SNP has its associated information, such as the chromosome, position in the genome, and name.

TOOL	MODEL	GC	SNP	CHROM	POSITION	PVALUE	SCORE	THRESHOLD	SIGNIFICANCE
GWASpoly	additive	1.016	c1_8019	10	48863165	7E-06	5.18	4.65	TRUE
GWASpoly	additive	1.016	c2_25471	10	48808404	1.5E-05	4.82	4.65	TRUE
SHEsis	additive	3.97	c1_8019	10	48863165	0	11.4559	5.3836	TRUE
SHEsis	additive	3.97	c1_13526	10	48020996	0	10.7986	5.3836	TRUE
SHEsis	additive	3.97	c2_53380	1	70371898	0	10.6073	5.3836	TRUE
SHEsis	additive	3.97	c2_25471	10	48808404	0	9.6253	5.3836	TRUE
SHEsis	additive	3.97	c2_54811	1	46270954	0	9.5467	5.3836	TRUE
SHEsis	additive	3.97	c2_45606	10	48218826	0	8.2013	5.3836	TRUE
SHEsis	additive	3.97	c2_11977	5	3515956	0	8.1656	5.3836	TRUE
SHEsis	additive	3.97	c1_8021	10	48862950	0	7.9872	5.3836	TRUE
SHEsis	additive	3.97	c1_16351	10	48761642	0	7.9666	5.3836	TRUE
SHEsis	additive	3.97	c2_45611	10	48203431	0	7.7878	5.3836	TRUE

3.2 Venn diagram of significative SNPs

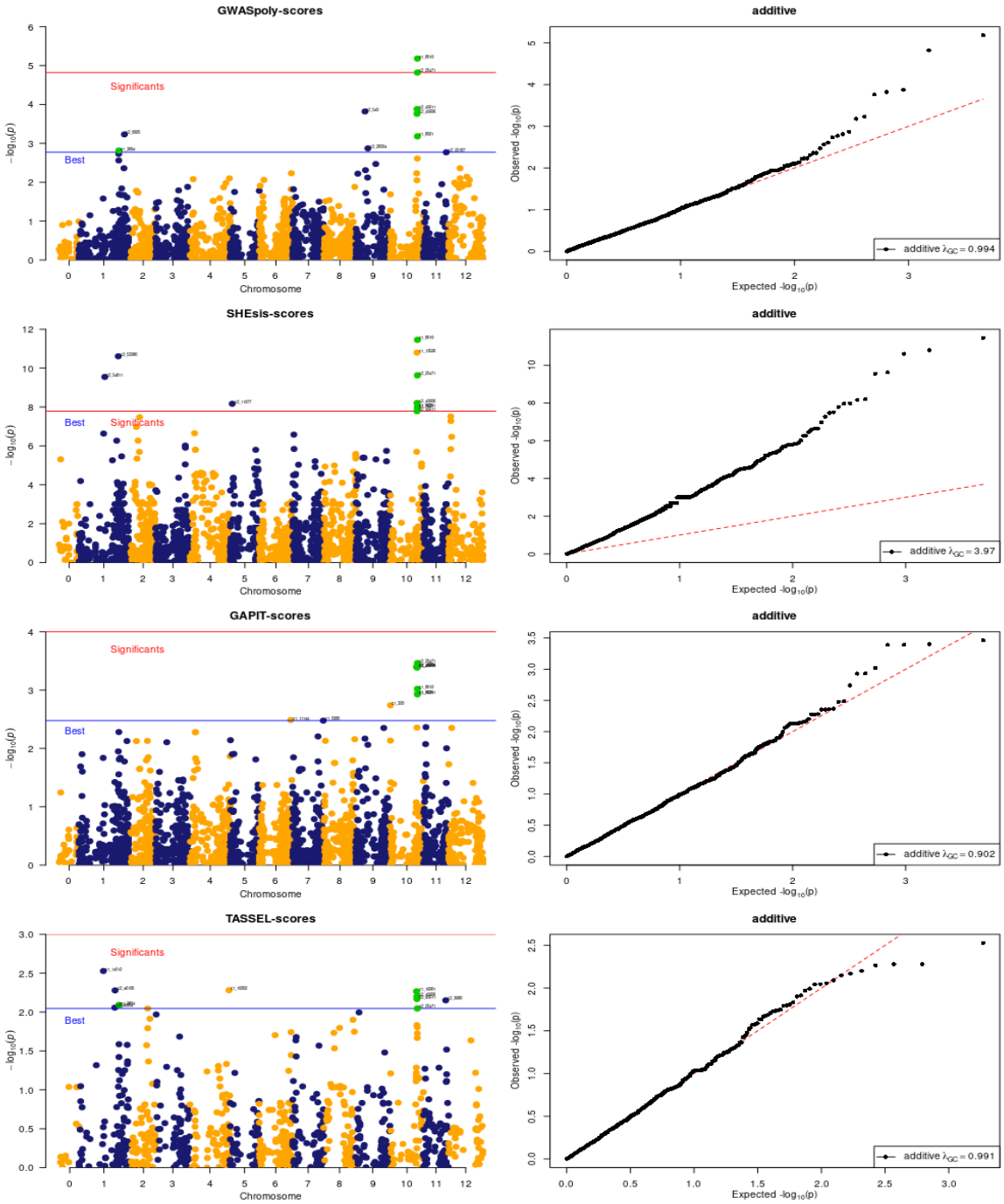
The Venn diagram shows the significative SNPs that are either unique or shared by the four software. Shared SNPs appear in the intersection areas of the diagram.



4 Manhattan and QQ plots

MultiGWAS uses classical Manhattan and Quantile–Quantile (QQ) plots to visually summarize GWAS results and identify both best-ranked and significative SNPs for each GWAS software. The Manhattan plot shows each SNPs distributed in the genetic location (x-axis) vs. the p-value (y-axis). The best-ranked SNPs are above the blue line and the significative SNPs above the blue line. In green are the SNPs that got high scores in more than one software.

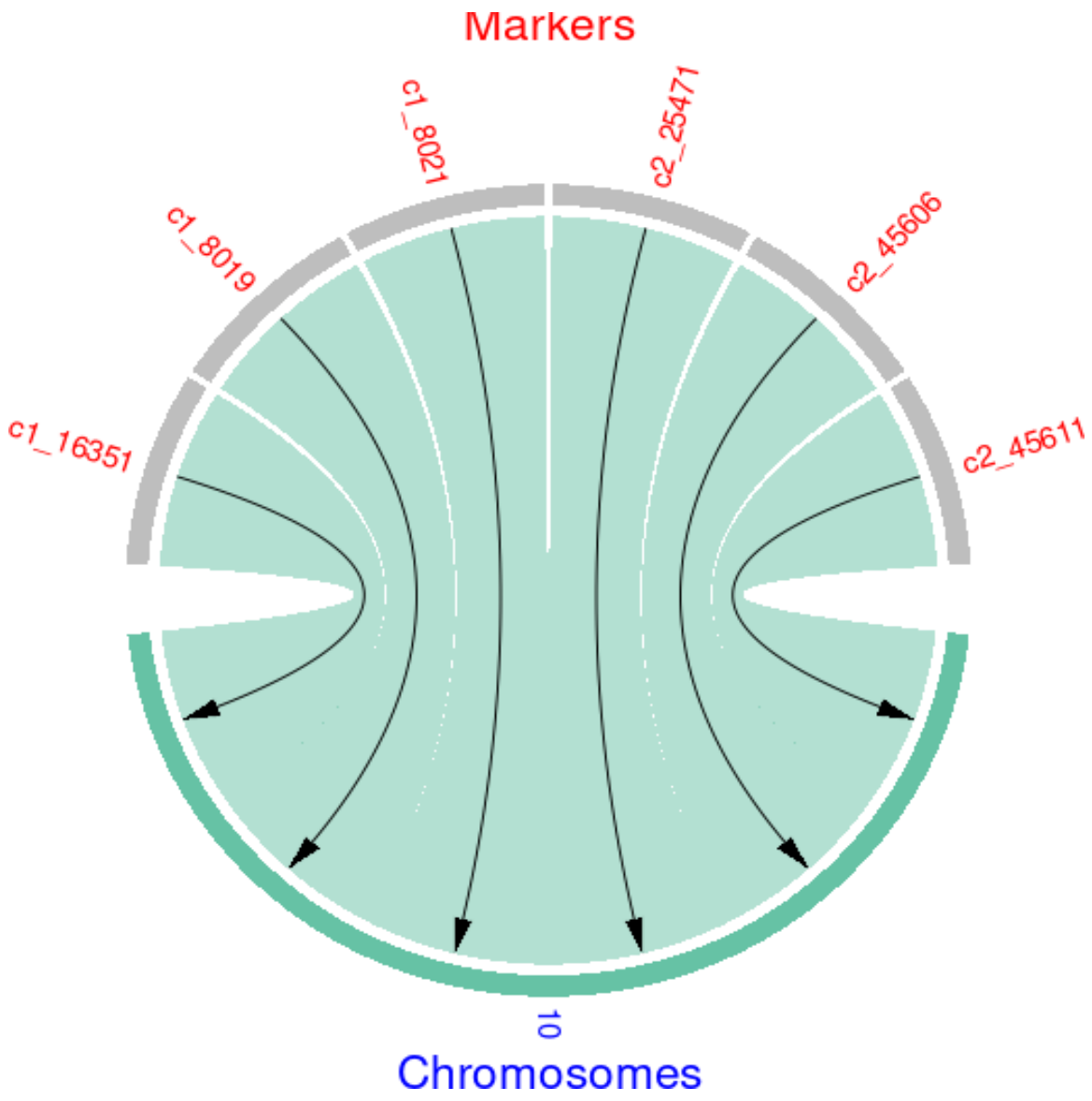
The QQ plot plots the observed (black dot lines) vs. expected -log 10 (p-value) (red line).



5 Chord diagrams for SNPs by chromosome

Chord diagrams are used in MultiGWAS to highlight nearby SNPs found on the same chromosome. These SNPs may have been identified simultaneously by multiple GWAS packets, which can be interpreted as a possible strong association, as in the Manhattan charts, where these associations form clean peaks with multiple SNPs displaying the same signal.

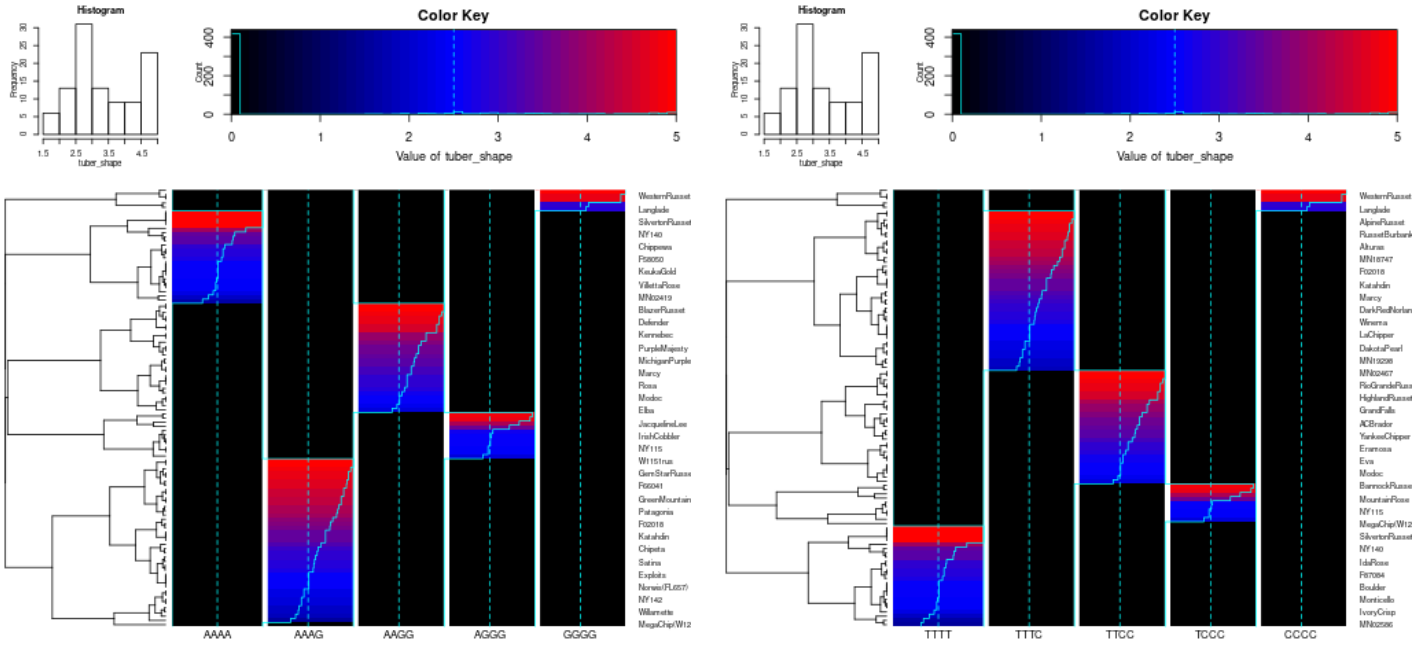
Next, we show the chord diagram for the shared SNPs shown in the Venn diagram of the best-ranked SNPs (Section 2.2).



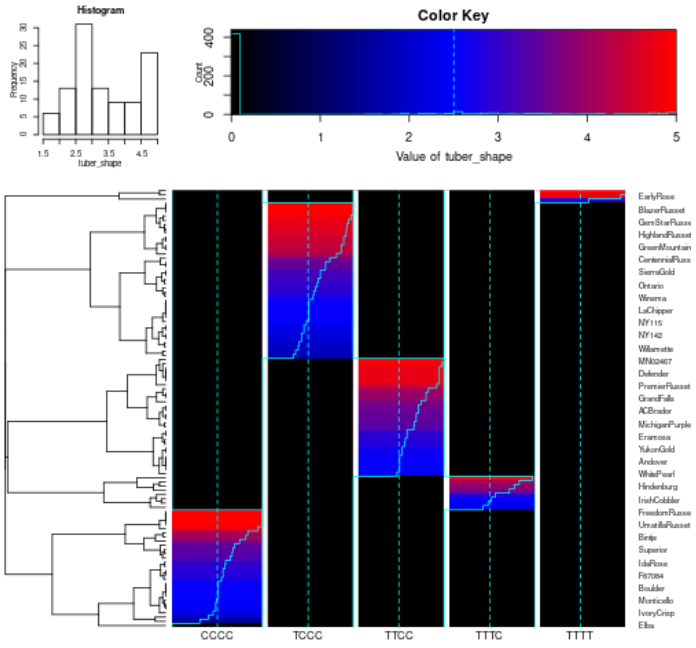
6 Profiles for common significative SNPs

For the common significative SNPs, MultiGWAS provides a figure to visualize each trait by individuals (rows) and genotypes (columns). At the left, the individuals are grouped in a dendrogram by their genotype. At the right, there is the name or ID of each individual. At the bottom, the genotypes are ordered from left to right, starting from the major to the minor allele (i.e., AAAA, AAAB, AABB, ABBB, BBBB). At the top, there is a description of the trait based on a histogram of frequency (top left) and by an assigned color for each numerical phenotype value using a grayscale (top right). Thus, each individual appears as a colored line by its phenotype value on its genotype column. For each column, there is a solid cyan line with the mean of each column and a broken cyan line that indicates how far the cell deviates from the mean.

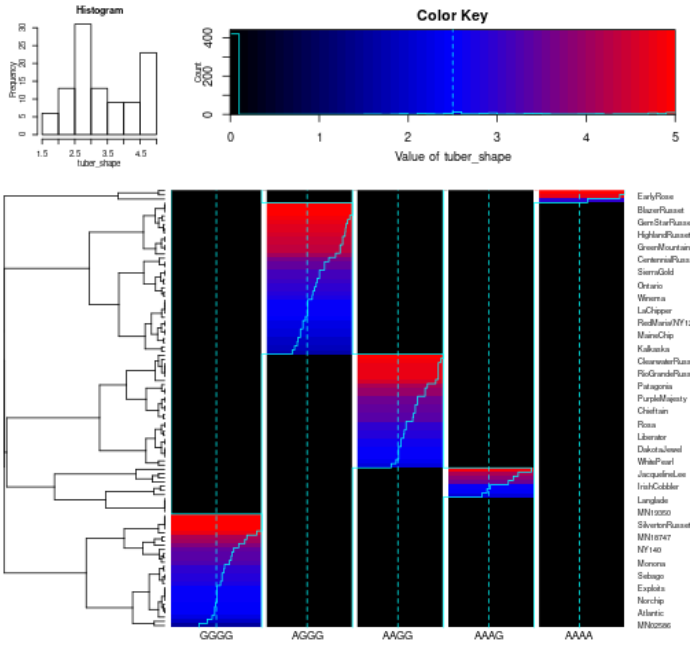
Because each multiGWAS report shows one specific trait at a time, the histogram and color key will remain the same for all the best-ranked SNPs.



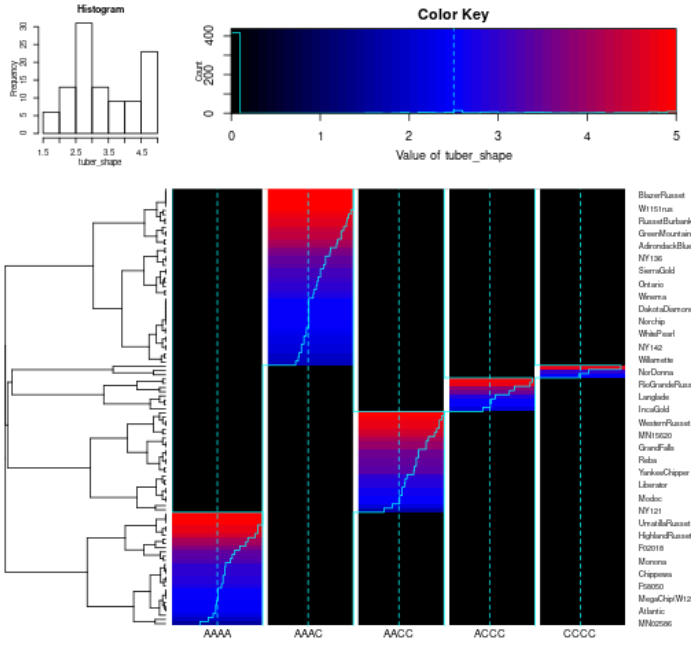
c1_8019 c2_25471



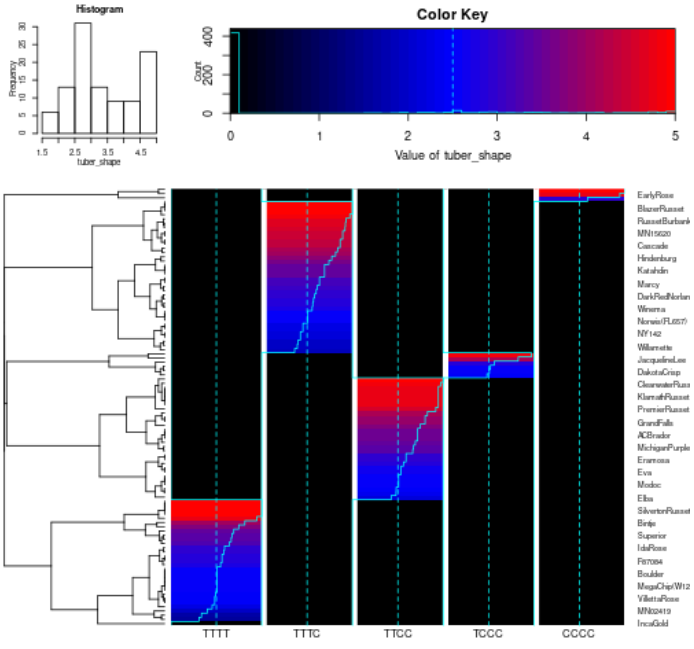
c2_45606



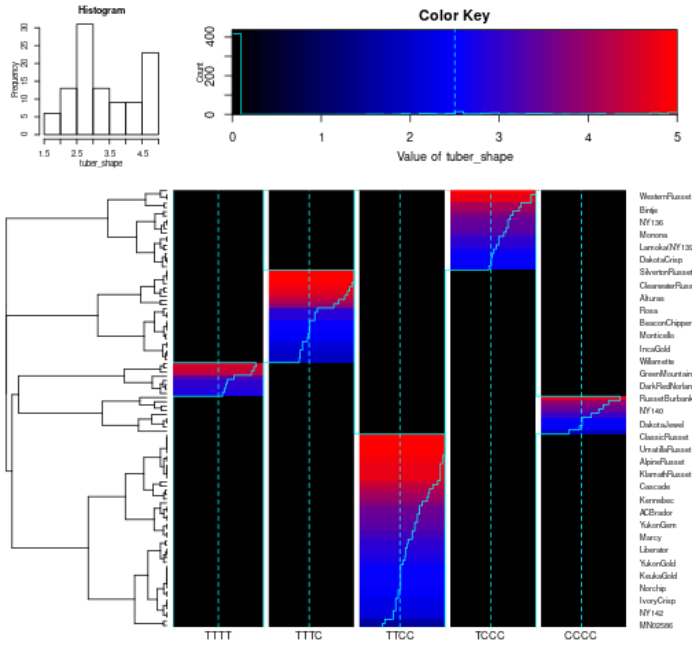
c2_45611



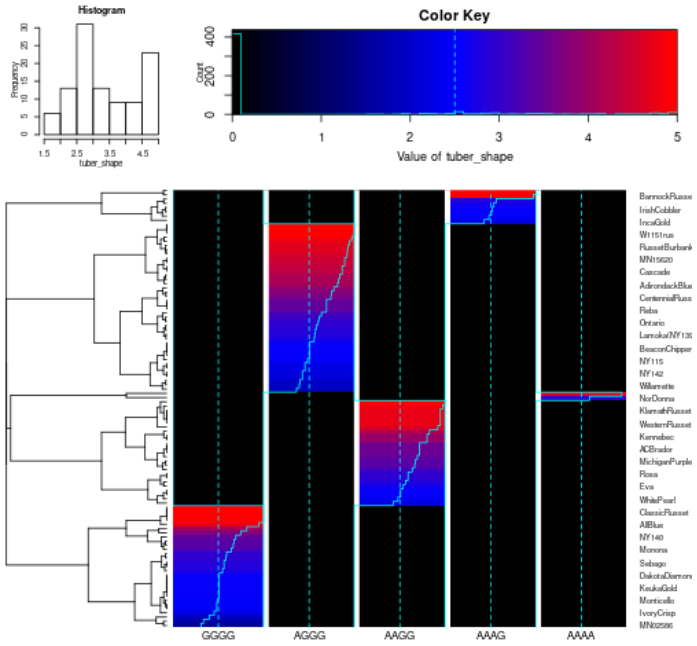
c1_16001



c1_16351



c1_3894



c1_8021