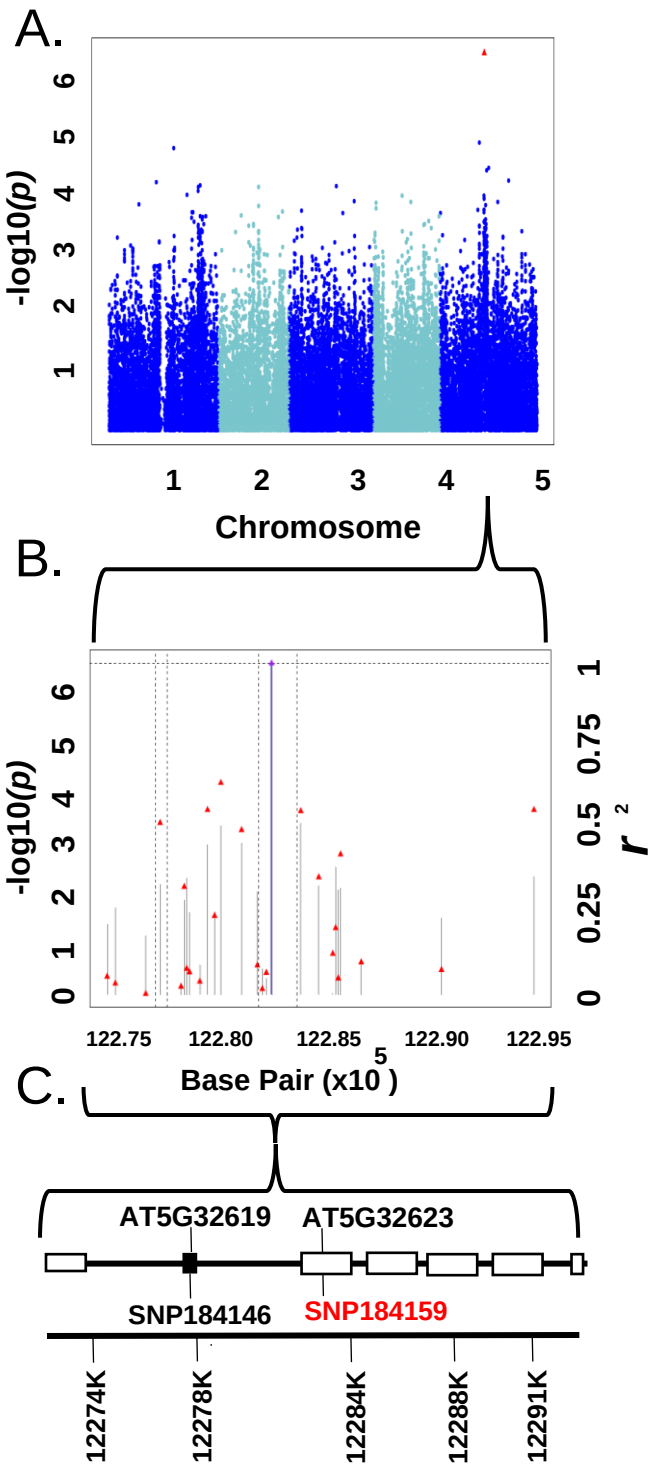
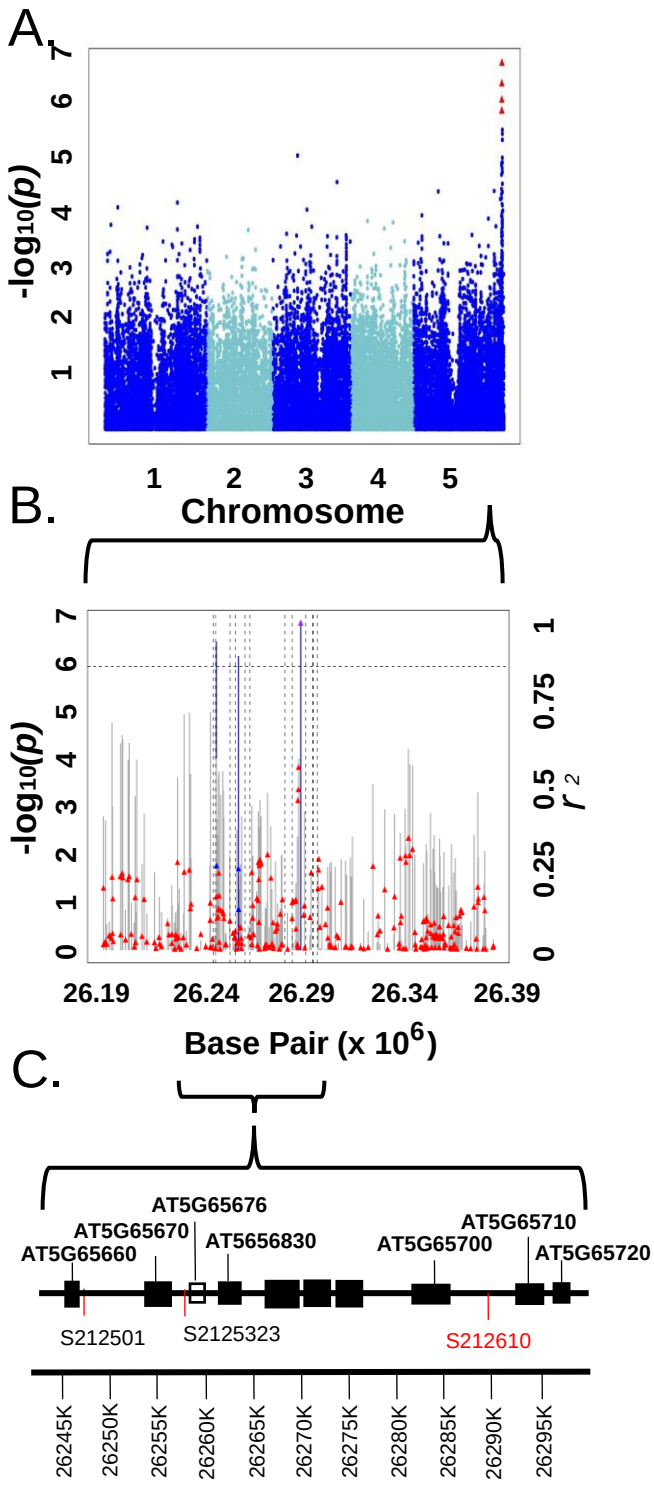
Supplemental Figure 1.



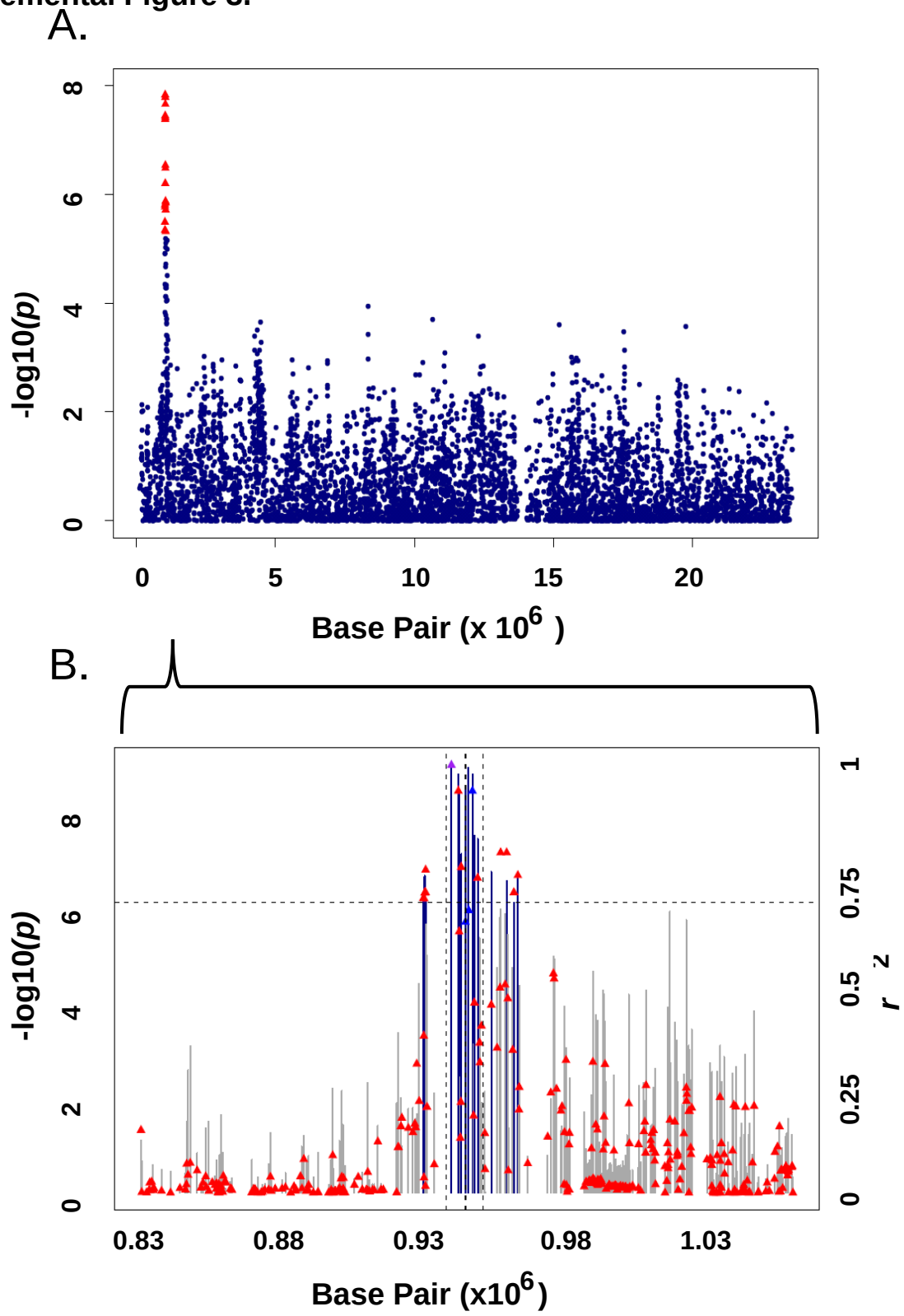
Supplemental Figure 1. GWAS summary of the Gly and Ser absolute sum. (A) A scatterplot of the association results from a unified mixed model analysis of the GS trait across the five *Arabidopsis* chromosomes. The negative log10-transformed *P*-values from the GWAS analysis are plotted against the genomic physical position. *P*-values for SNPs that are statistically significant for a trait at 5% FDR are in red. (B) A scatterplot of the association results from a unified mixed model analysis and pairwise LD estimates (*r*2) of SNP184159 across a 20 kb chromosomal interval centered on SNP184159 and plotted against chromosomal position. The blue lines are negative log10-transformed *P*-values for SNPs that are statistically significant for a trait at a 5% FDR level; grey lines are non-significant. The black vertical dashed lines indicate the position of At5G32619 and the pseudogene At5G32623 (left to right). Red triangles are the *r*2 values of each SNP relative to SNP82058 (indicated in purple). (C) Graphical representation of genes within the vicinity of SNP184159. Genes are represented by black boxes, and pseudogenes are represented by white boxes.

Supplemental Figure 2.



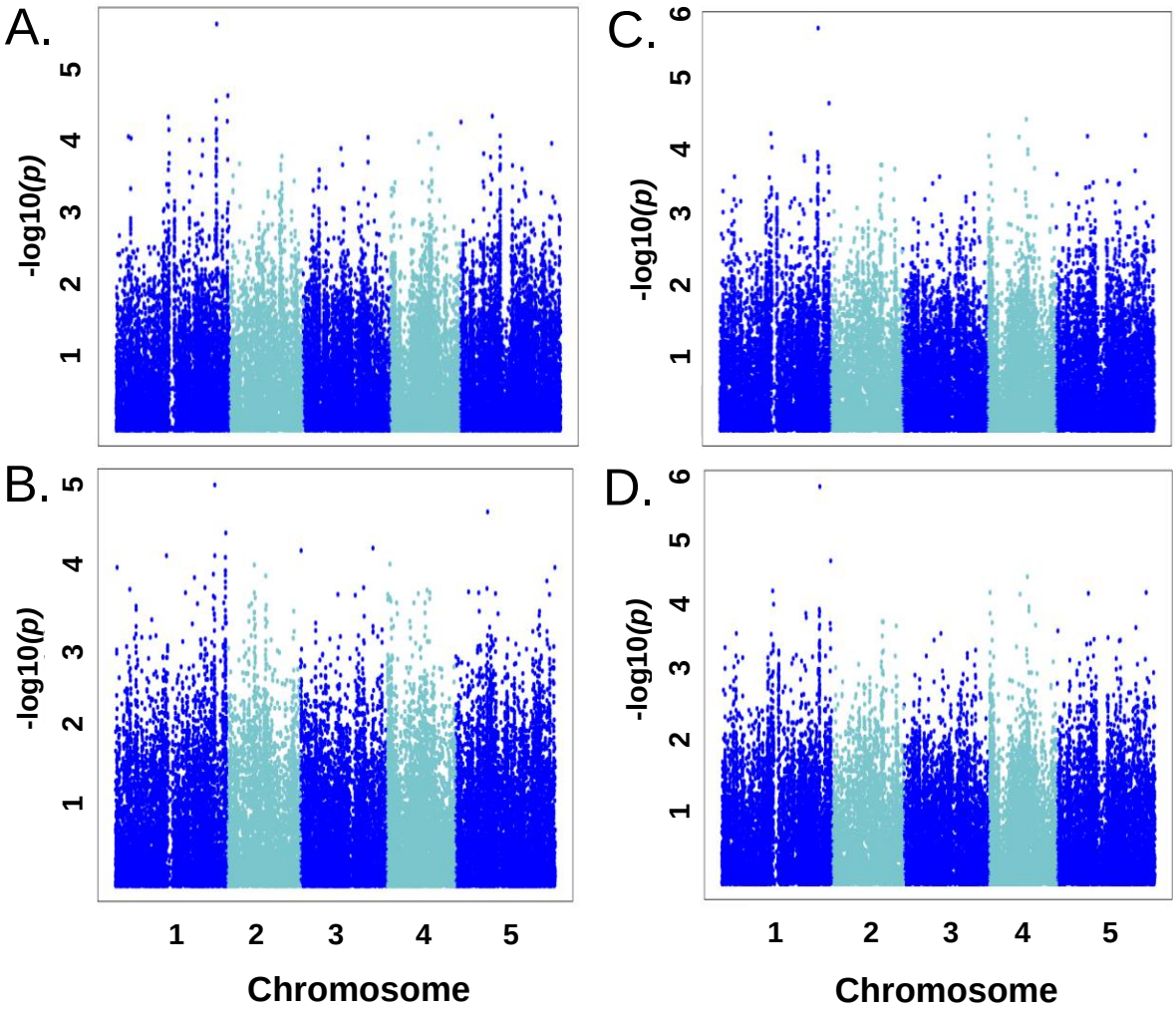
Supplemental Figure 2. GWAS summary of S/LIVKHSQR. (A) A scatterplot of the association results from a unified mixed model analysis of the S/LIVKHSQR trait across the five *Arabidopsis* chromosomes. The negative log10-transformed *P*-values from the GWAS analysis are plotted against the genomic physical position. *P*-values for SNPs that are statistically significant for a trait at 5% FDR are in red. (B) A scatterplot of the association results from a unified mixed model analysis and pairwise LD estimates (*r*2) of SNP212610 across a ±100 kb chromosomal interval centered on SNP212610 and plotted against chromosomal position. The blue lines are negative log10-transformed *P*-values for SNPs that are statistically significant at a 5% FDR level; grey lines are non-significant. Black vertical dashed lines indicate the positions of At5G65660, At5G65670, At5G65676, At5G65700, and At5G65710 (left to right). Red triangles are the *r2* values of each SNP relative to SNP212610 (indicated in purple), and blue triangles represent the *r2* values of the four significant SNPs. (C) Graphical representation of genes within the vicinity of SNP212610. Genes are represented by black boxes, and pseudogenes are represented by white boxes.

Supplemental Figure 3.



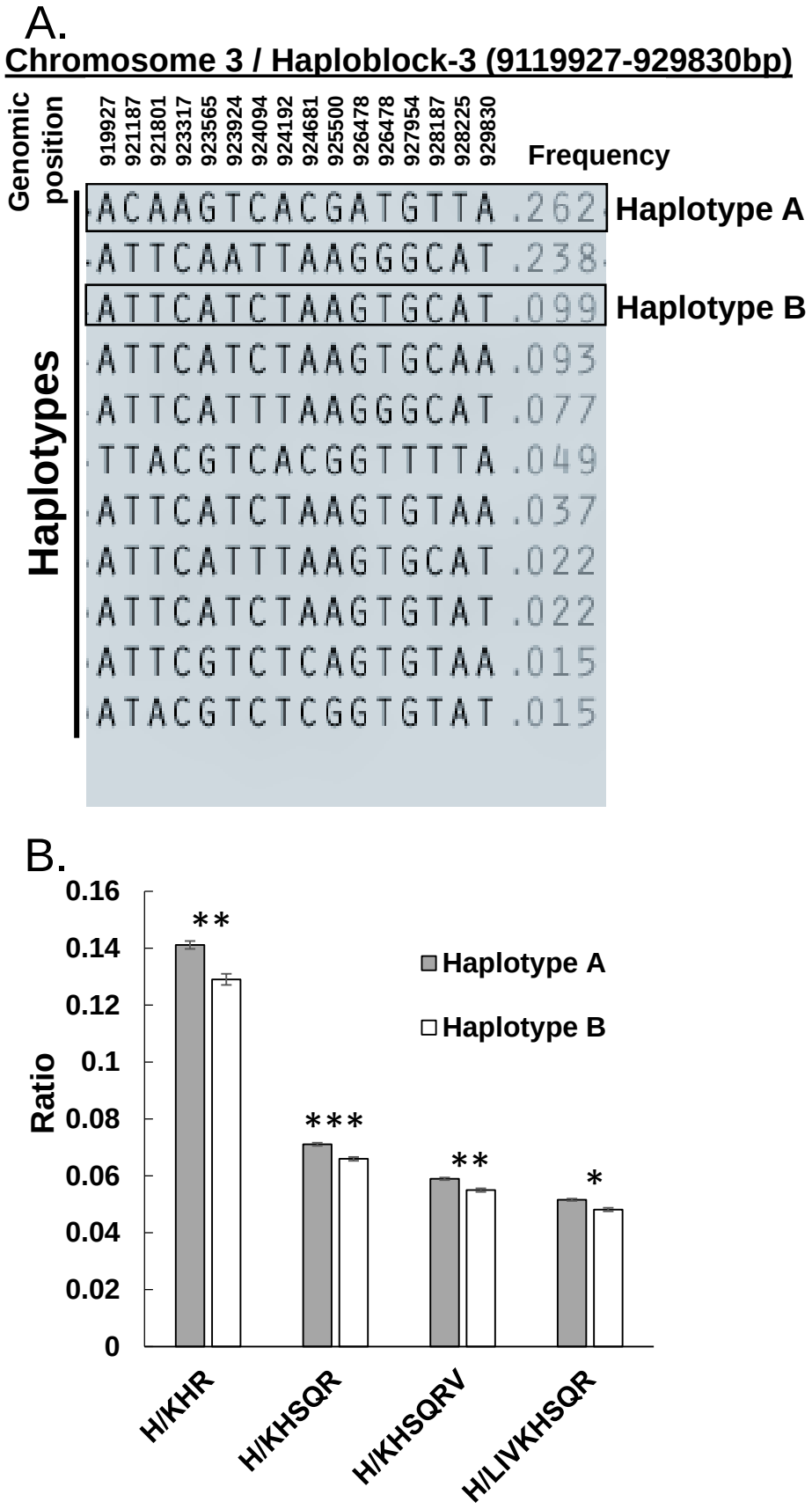
Supplemental Figure 3. GWAS of H/KHSQR. (A) A scatterplot of the association results from a unified mixed model analysis of H/KHSQR on chromosome 3 only. The negative $\log_{10}$-transformed *P* values from the GWAS analysis are plotted against the genomic physical position. *P*-values for SNPs that are statistically significant for a trait at 5% FDR are in red. (B) A scatterplot of the association results from a unified mixed model analysis and pairwise LD estimates (r^2) of SNP82058 across a ± 100 kb chromosomal interval centered on SNP82058 and plotted against chromosomal position. The blue lines are negative $\log_{10}$-transformed *P*-values for SNPs that are statistically significant at a 5% FDR level; grey lines are non-significant. The black vertical dashed lines indicate the positions of *RIF10* and *CAT4* (left to right). Red triangles are the r^2 values of each SNP relative to SNP82058 (indicated in purple), and blue triangles represent other very strongly significant SNPs.

Supplemental Figure 4.



Supplemental Figure 4. Conditional GWAS of the four most significant His-related traits. Scatterplots of the association results from a unified mixed model analysis for (A) H/LIVKHSQR, (B) H/KHR, (C) H/KHSQR, and (D) H/KHSQRV traits, including SNP80258 as a co-variant. The negative log10-transformed *P*-values from the GWAS analysis are plotted against the genomic physical position. No significant associations were detected. The inclusion of any one of the four most significant SNPs as co-variants yielded similar results (data not shown).

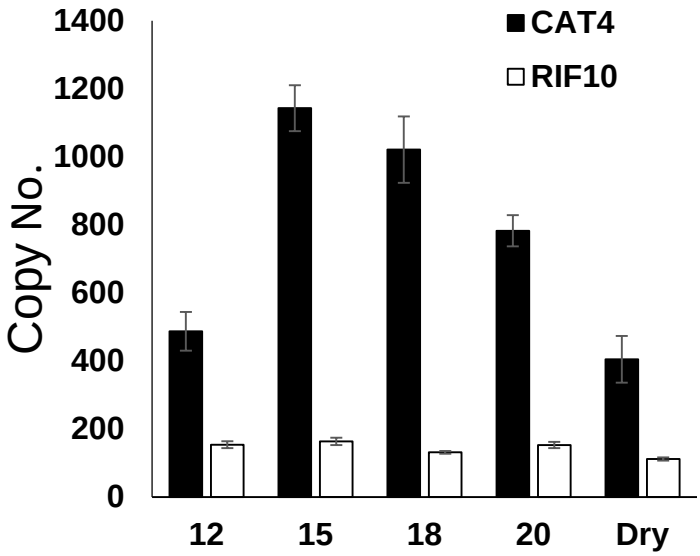
Supplemental Figure 5.



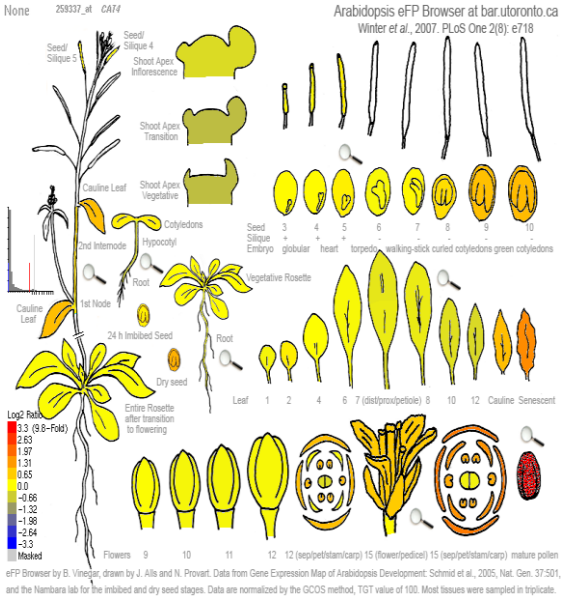
Supplemental Figure 5. Haplotype analysis of haploblock 3. (A) Graphical representation of the haplotypes in haploblock 3 and their frequencies in the association panel. The most significant haplotype pair is marked in black boxes and annotated as haplotypes A and B. (B) Average levels of H/KHR, H/KHSQR, H/KHSQRV, and H/LIVKHSQR traits from the backtransformed BLUPs calculated from three biological replicates of the population. Grey and white bars represent the trait average from accessions with haplotype A (higher levels) and haplotype B (lower levels). *P=1.21E-05; **P=2.6-4.6E-06; *** P=2.38E-07.

Supplemental figure 6.

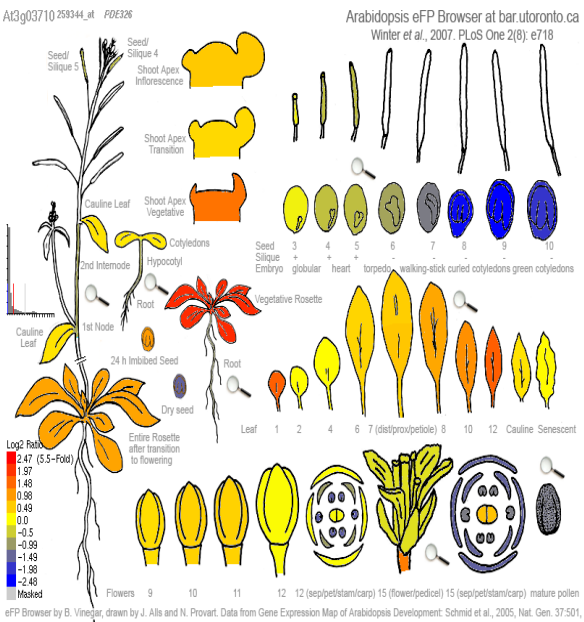
A.



B. CAT4 (AT3G03720)
Relative levels

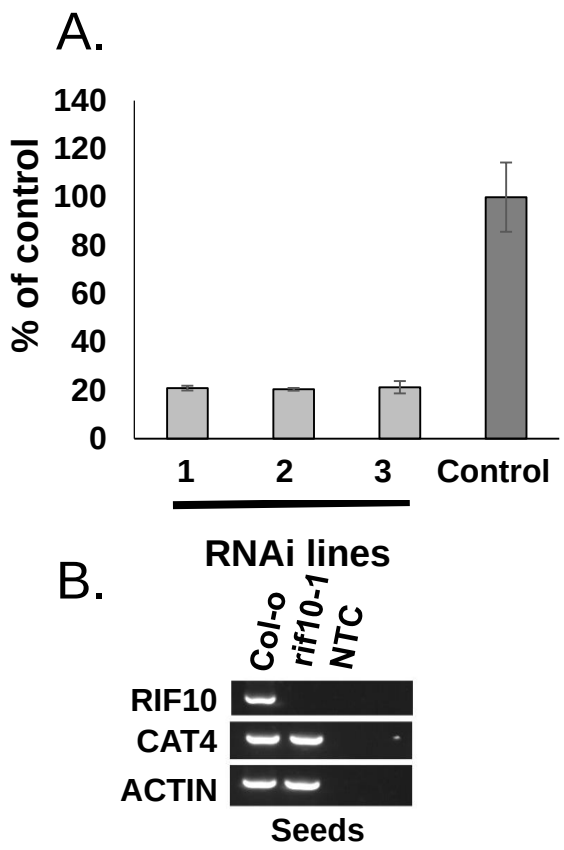


C. RIF10 (AT3G03710)
Relative levels



Supplemental Figure 6. Expression levels of CAT4 and RIF10. (A) Transcript levels of CAT4 (black) and RIF10 (white) were measured from a Col-0 accession at five time points of seed maturation and desiccation using quantitative PCR with Actin 2 mRNA as an internal control. Error bars represent standard errors (of two technical repeats of three independent biological replicates; n=6). Relative expression levels for (B) CAT4 and (C) RIF10, as reported at the *Arabidopsis* eFP Browser.

Supplemental Figure 7.



Supplemental Figure 7. *CAT4* and *RIF10* transcript levels in the mutant lines. Relative transcript levels of (A) *CAT4* transcripts levels were measured from the dry seeds of three RNAi lines and a Col-0 control containing an empty vector using Actin 2 mRNA as an internal control. Error bars represent standard errors (independent biological repeat; n=4-5). (B) RT-PCR analysis of total RNA from seeds of the WT (Col-0) and *rif10*. NTC – no template control.