

1

## **Supplementary Figures note**

2

3

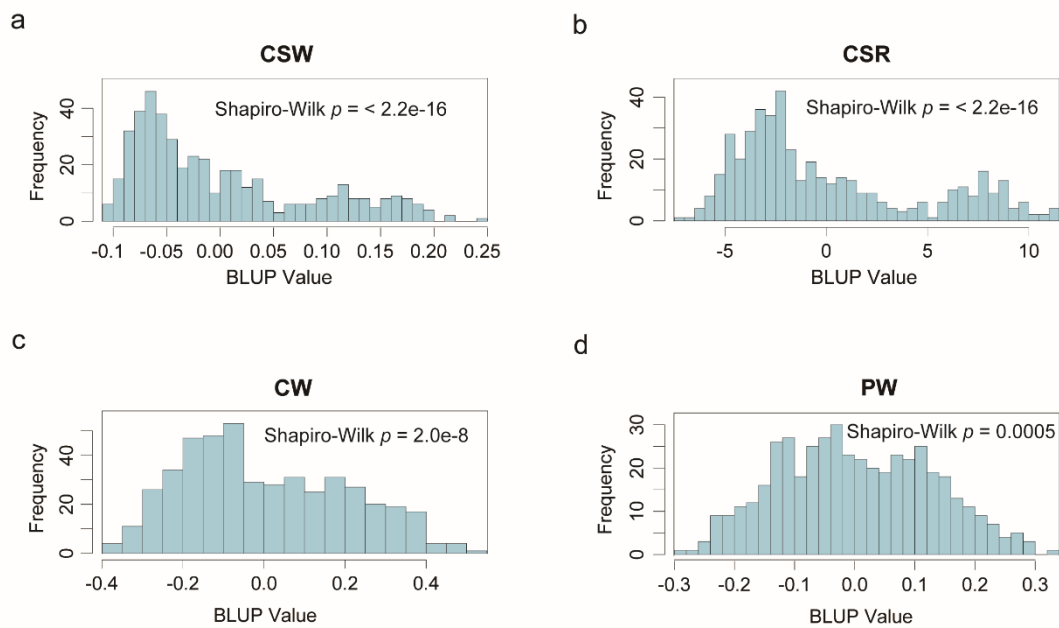
### **Rare Variant Analysis Reveals Divergent Genetic**

4

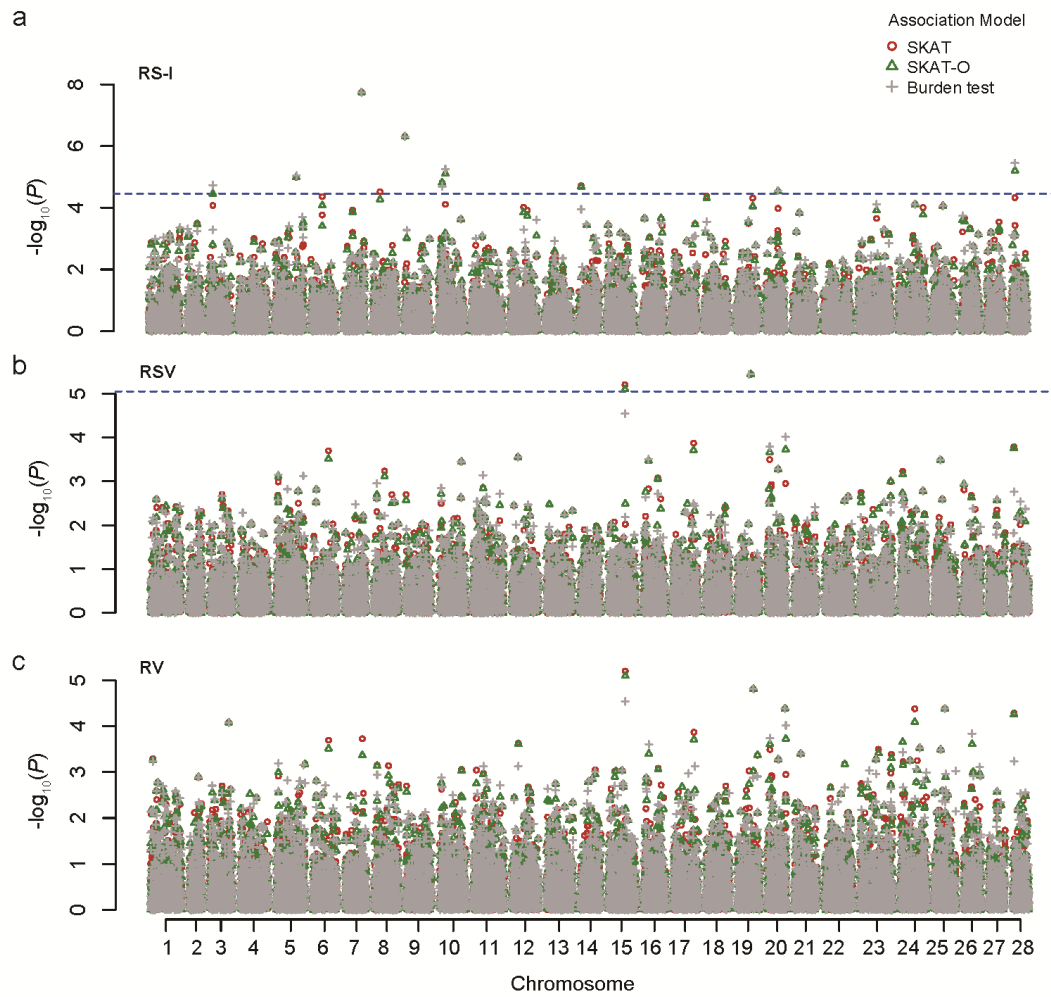
### **Architectures of Directly Selected and Correlated Silkworm**

5

### **Traits**

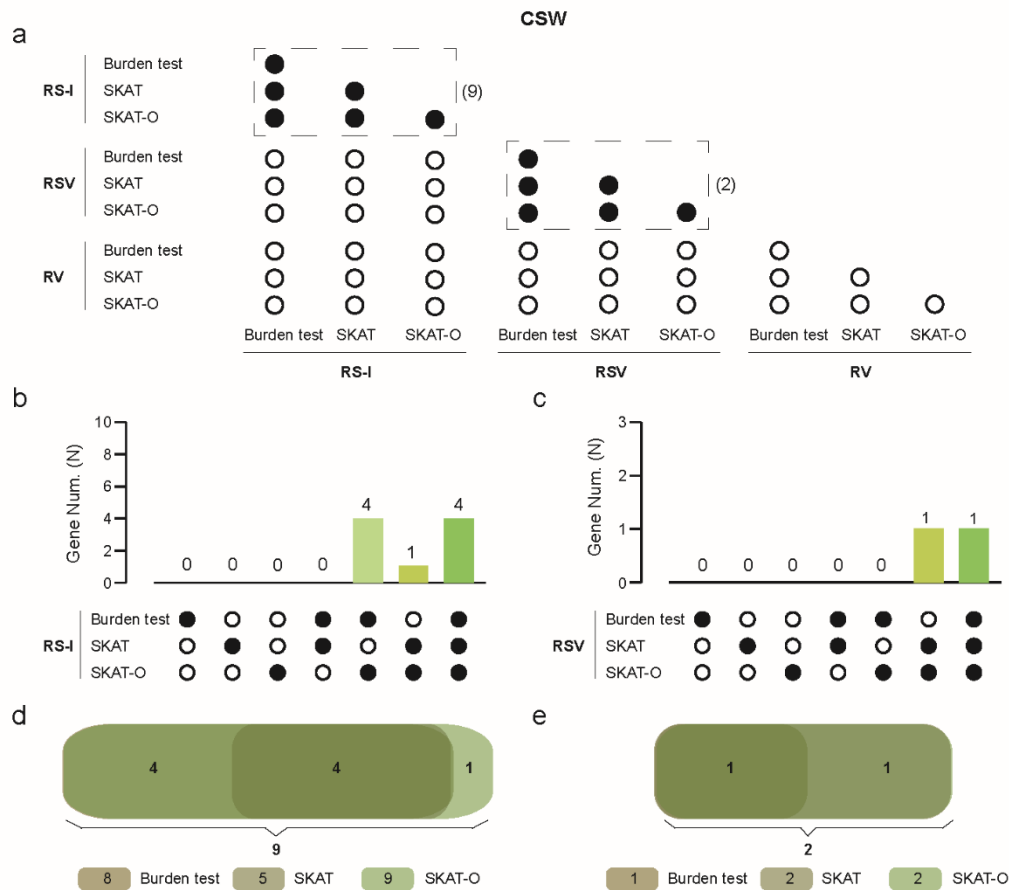


**Supplementary Fig. 1 |** Distribution and normality test of BLUP values for cocoon silk yield traits in the silkworm. Based on 459 germplasm resource samples, frequency distribution histograms of the BLUP values obtained from linear mixed model fitting are presented for four cocoon silk yield traits (**a**: CSW, **b**: CSR, **c**: CW, **d**: PW). The Shapiro-Wilk normality test indicates that the BLUP values for all traits significantly deviate from a normal distribution ( $p < 0.01$ ).

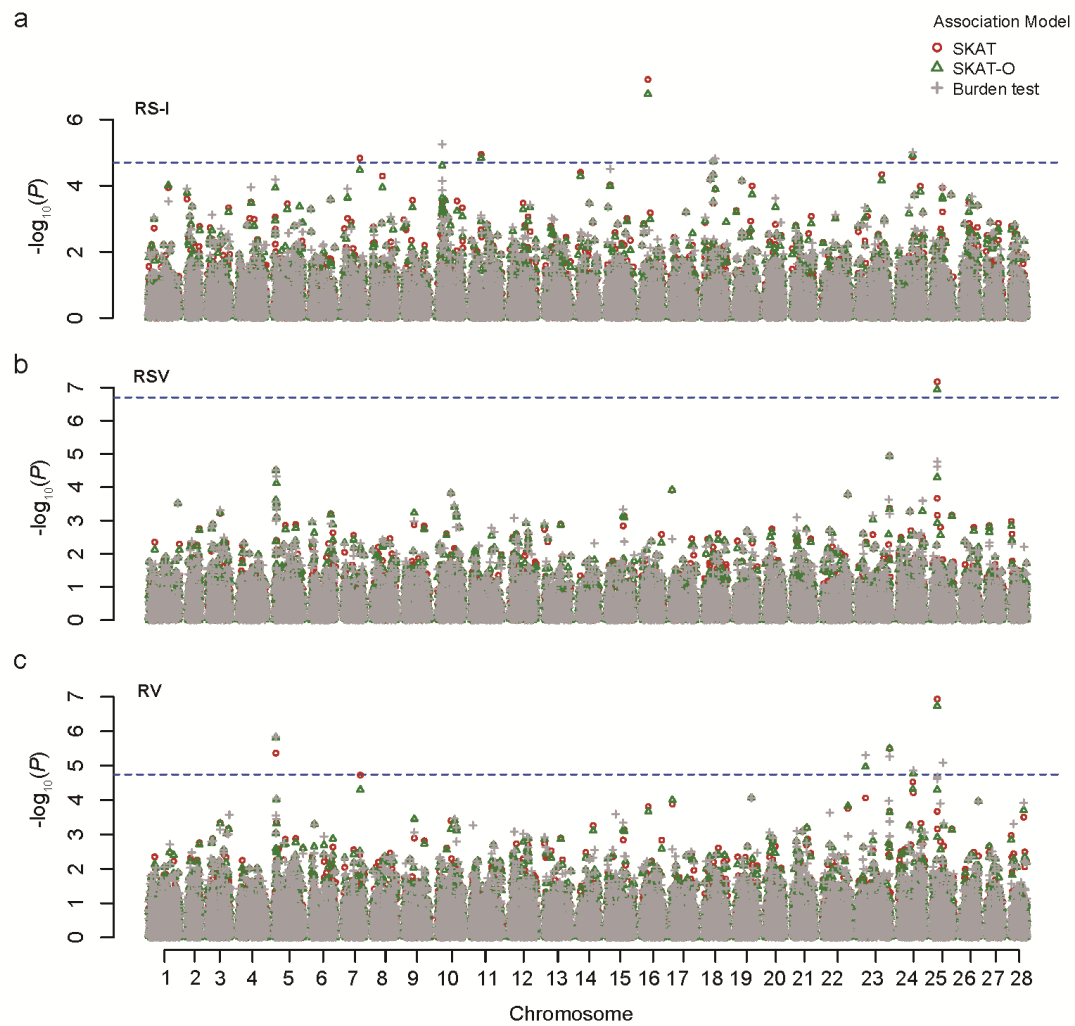


16

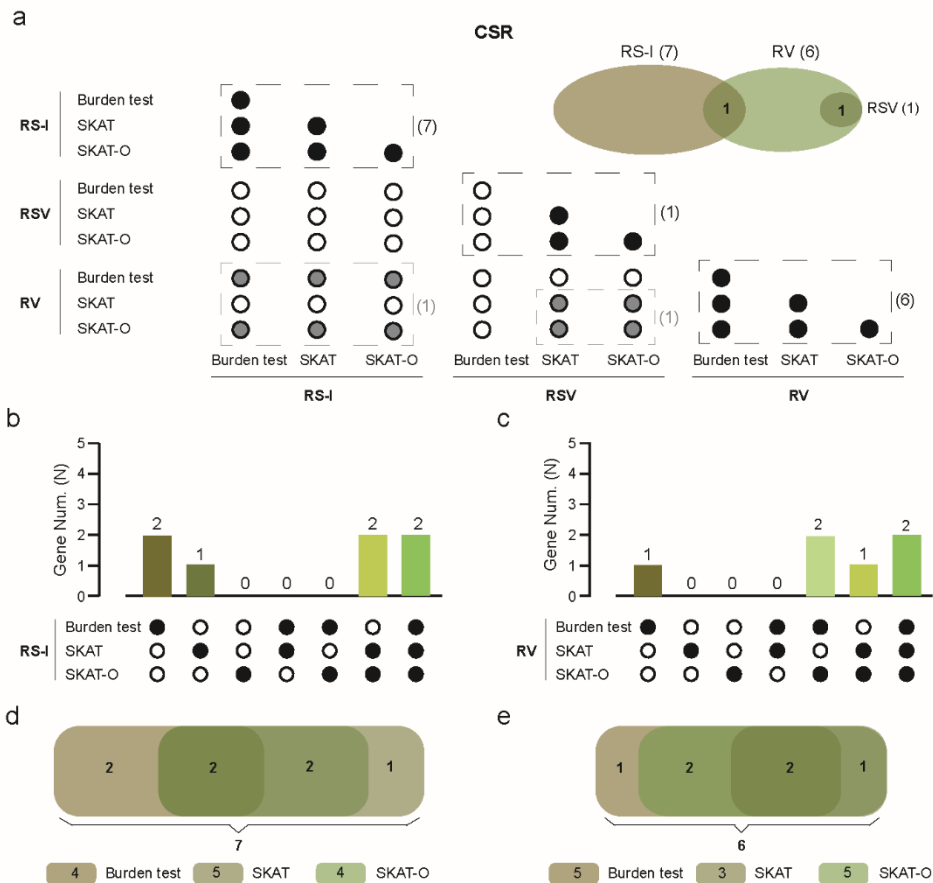
17 **Supplementary Fig. 2 |** Manhattan plots of rare variant association analysis (RVAS) for  
 18 the CSW trait based on three variant sets. Using the silkworm germplasm resource  
 19 population (RS-I: 459 samples; RSV and RV: 283 samples) and the analytical framework  
 20 from Fig. 2, RVAS was performed for the CSW trait using three variant sets: RS-I (a),  
 21 RSV (b), and RV (c), respectively. Each variant set was tested using three models:  
 22 Burden test, SKAT, and SKAT-O. The dashed line indicates the significance threshold of  
 23 FDR < 0.05 (determined based on the model corresponding to the maximum *P*-value in  
 24 each set). No dashed line is shown in panel c, indicating that no significant association  
 25 signals were detected in the RV set.



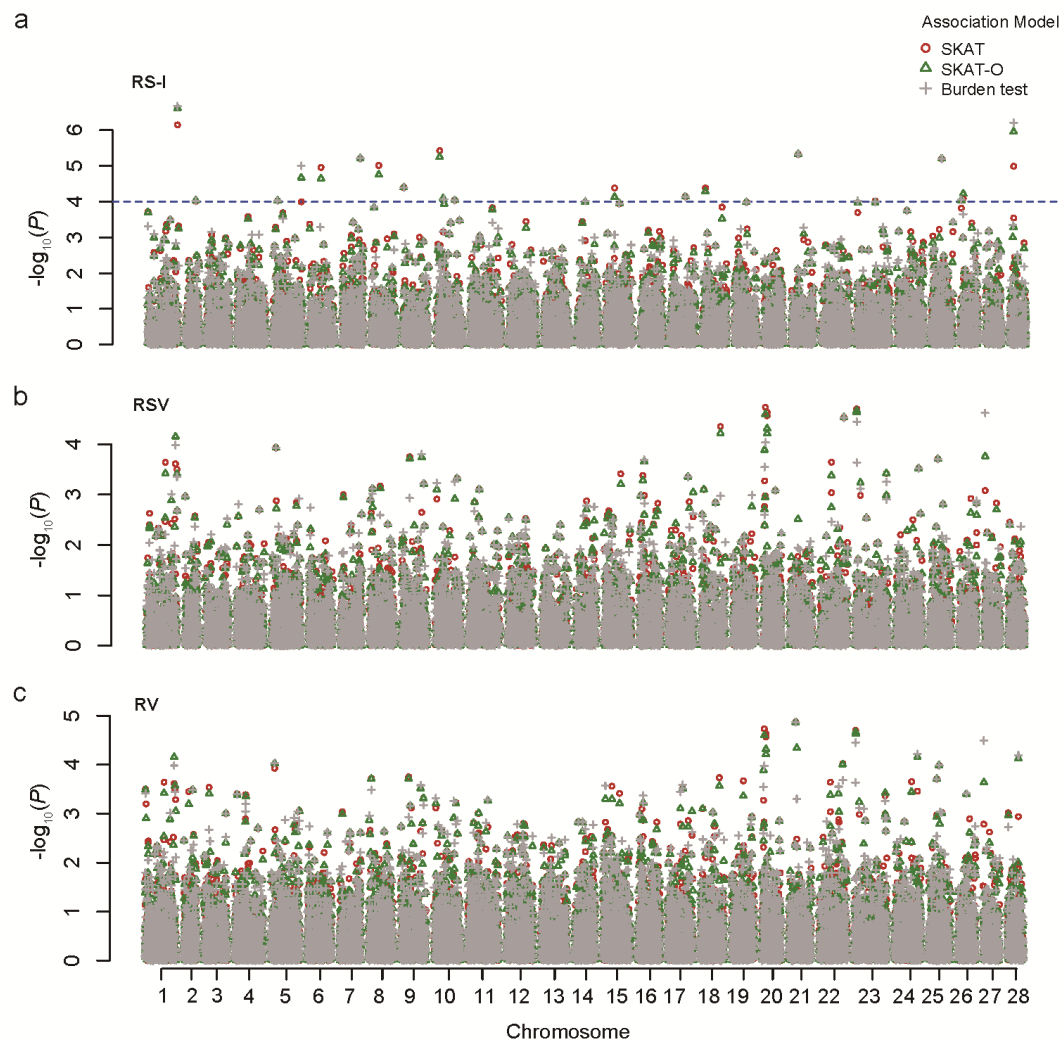
**Supplementary Fig. 3 | Statistical analysis of variant sets and test models for CSW trait association signals. a.** Detection of CSW-associated genes by the three variant sets (RS-I, RSV, and RV), with numbers indicating the total genes identified in each set. **b, c.** Distribution of the number of associated genes detected by different model combinations within the RS-I (b) and RSV (c) variant sets. **d, e.** Venn diagrams corresponding to panels b and c, showing the total number of genes detected by each of the three models and the overlap of genes between models for the RS-I (d) and RSV (e) variant sets.



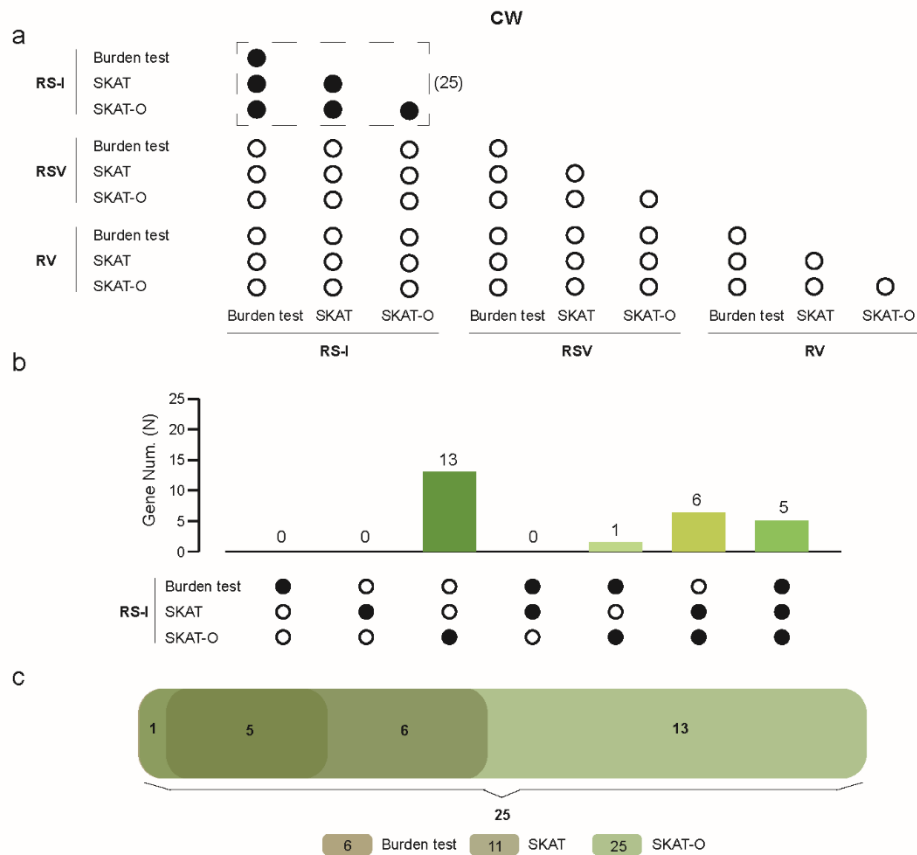
**Supplementary Fig. 4 |** Manhattan plots of RVAS for the CSR trait based on three variant sets. Using the silkworm germplasm resource population (RS-I: 459 samples; RSV and RV: 283 samples) and following the analytical framework of Fig. 2, RVAS was conducted for the CSR trait through three variant sets: RS-I (a), RSV (b), and RV (c). Each variant set was tested using three models: Burden test, SKAT, and SKAT-O. The dashed line indicates the significance threshold of  $FDR < 0.05$  (determined based on the model corresponding to the maximum  $P$ -value in each set).



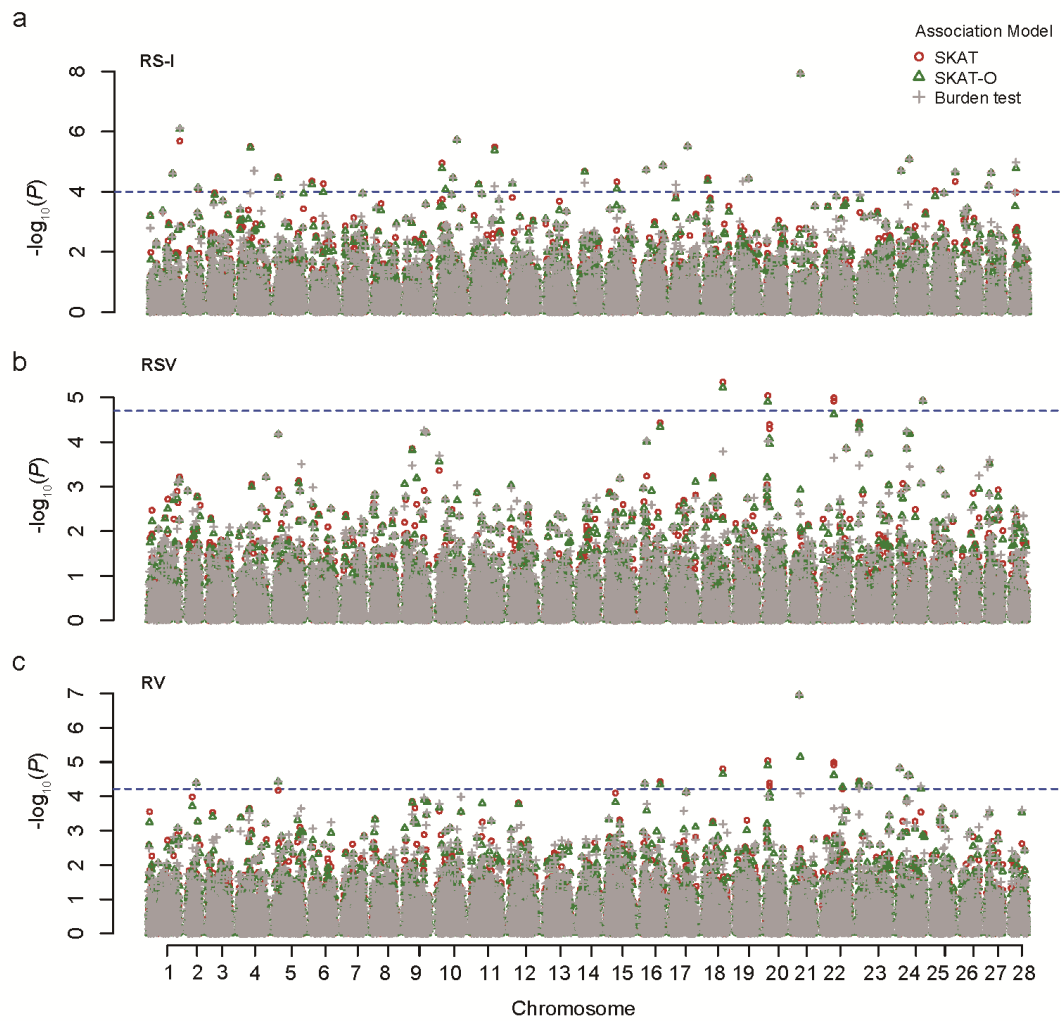
**Supplementary Fig. 5 | Statistical analysis of variant sets and test models for CSR trait association signals. a.** Detection of CSR-associated genes by the three variant sets (RS-I, RSV, and RV), with the numbers on the diagonal indicating the total genes identified in each set, and the gray off-diagonal areas displaying the counts of shared genes between sets; the results are simultaneously presented in the Venn diagram on the right. **b, c.** Distribution of the number of associated genes detected by different model combinations within the RS-I (b) and RV (c) variant sets. **d, e.** Venn diagrams corresponding to panels b and c, showing the total number of genes detected by each of the three models and the overlap of genes between models for the RS-I (d) and RSV (e) variant sets.



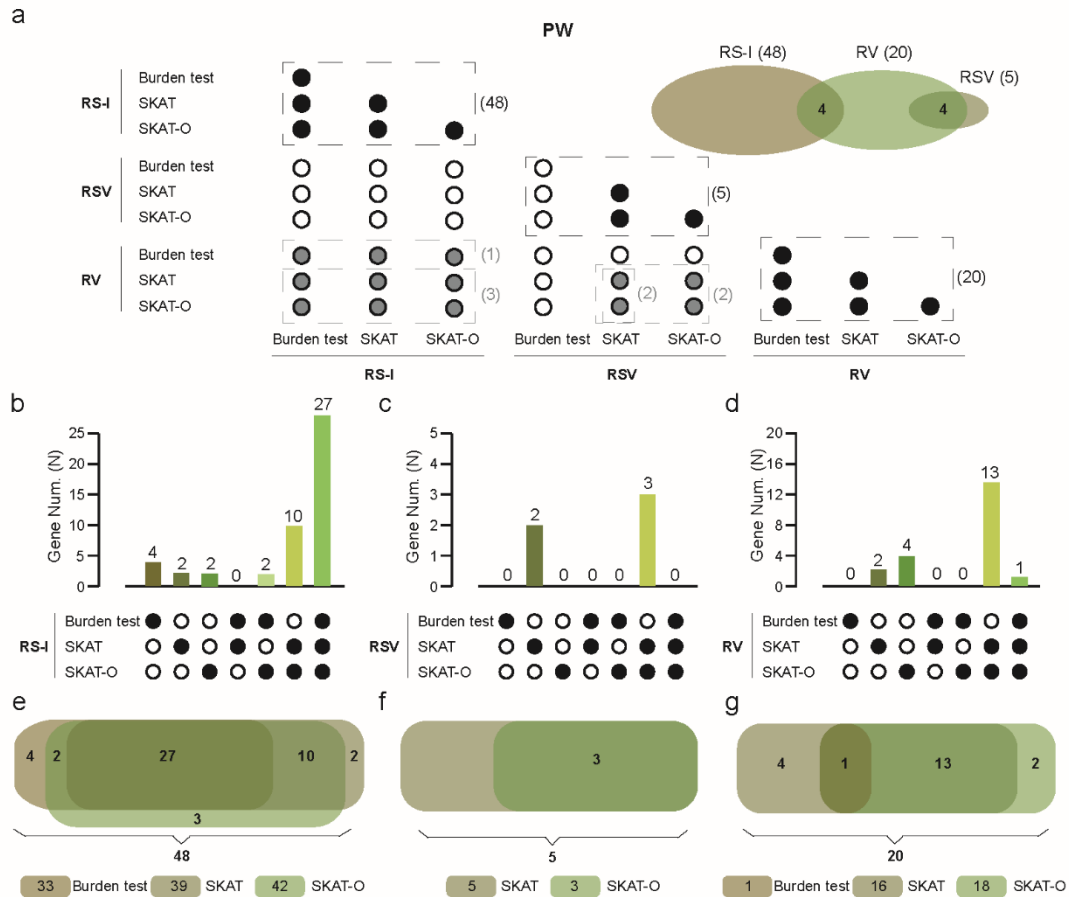
**Supplementary Fig. 6 |** Manhattan plots of RVAS for the CW trait based on three variant sets. Using the silkworm germplasm resource population (RS-I: 459 samples; RSV and RV: 283 samples) and following the analytical framework of Figure 2, RVAS was conducted for the CW trait through three variant sets: RS-I (a), RSV (b), and RV (c). Each variant set was tested using three models: Burden test, SKAT, and SKAT-O. The dashed line indicates the significance threshold of  $FDR < 0.05$  (determined based on the model corresponding to the maximum  $P$ -value in each set). No dashed lines are shown in panels b and c, indicating that no significant association signals were detected in either the RSV or RV sets.



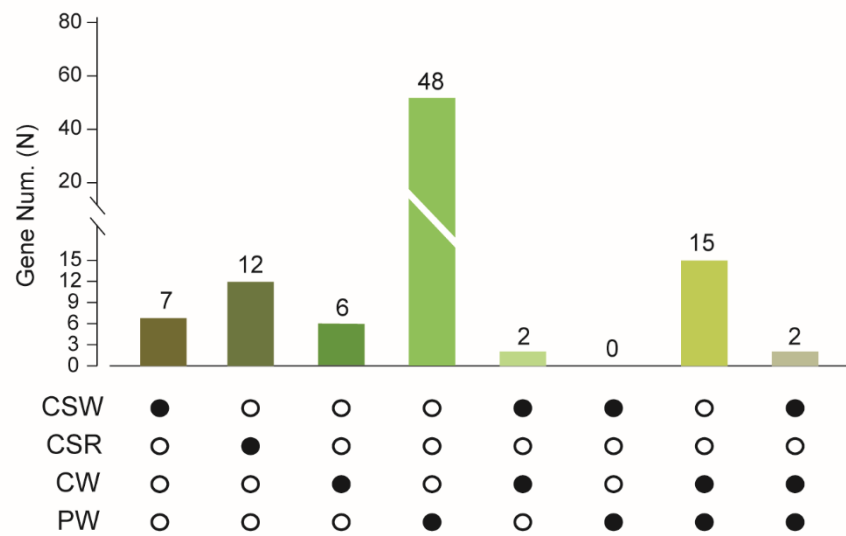
**Supplementary Fig. 7 | Statistical analysis of variant sets and test models for CW trait association signals. a.** Detection of CW-associated genes by the three variant sets (RS-I, RSV, and RV). Results indicate that only RS-I showed significant association signals, with the number representing the count of detected genes. **b.** Distribution of the number of associated genes detected by different model combinations within the RS-I variant set. **c.** Venn diagram for the RS-I variant set, showing the total number of genes detected by each of the three models and the overlap of genes between models.



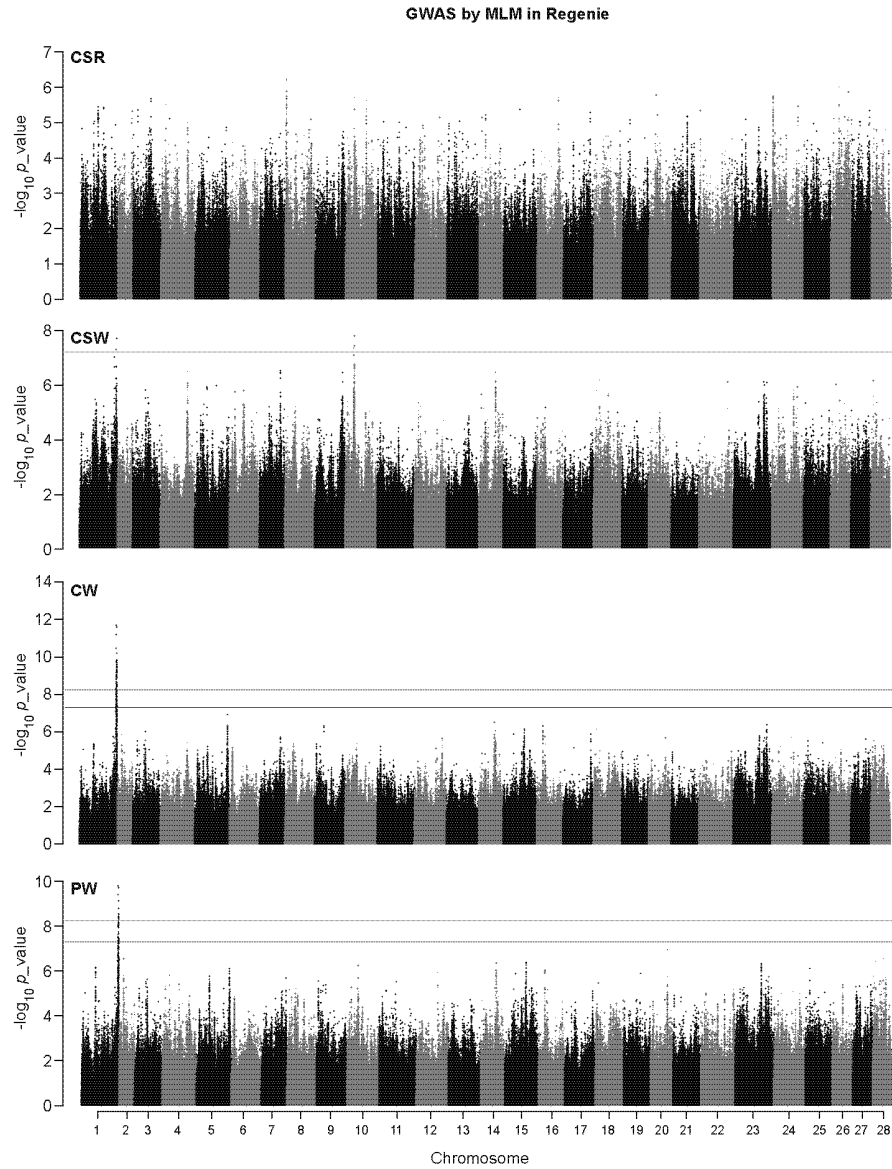
**Supplementary Fig. 8 |** Manhattan plots of RVAS for the PW trait based on three variant sets. Using the silkworm germplasm resource population (RS-I: 459 samples; RSV and RV: 283 samples) and following the analytical framework of Fig. 2, RVAS was conducted for the PW trait through three variant sets: RS-I (a), RSV (b), and RV (c). Each variant set was tested using three models: Burden Test, SKAT, and SKAT-O. The dashed line indicates the significance threshold of  $FDR < 0.05$  (determined based on the model corresponding to the maximum  $P$ -value in each set).



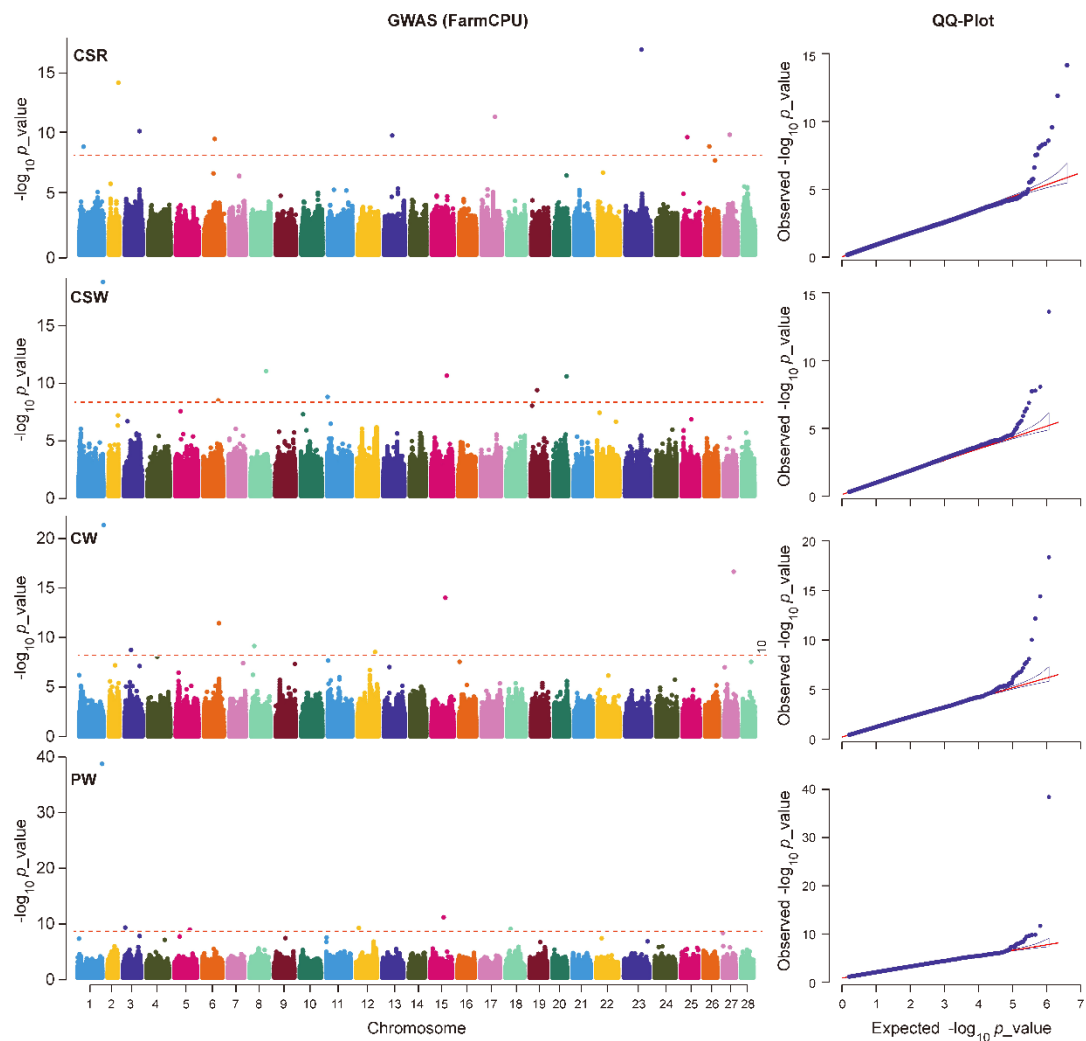
**Supplementary Fig. 9 | Statistical analysis of variant sets and test models for PW trait association signals. a.** Detection of PW-associated genes by the three variant sets (RS-I, RSV, and RV). The diagonal of the matrix shows the total number of genes detected in each variant set, while the off-diagonal gray areas indicate the number of shared genes between sets; these results are simultaneously presented in the Venn diagram on the right. **b, c, d.** Distribution of the number of associated genes detected by different model combinations within the RS-I (b), RSV (c), and RV (d) variant sets. **e, f, g.** Venn diagrams corresponding to panels b, c, and d, showing the total number of genes detected by each of the three models and the overlap of genes between models for the RS-I (e), RSV (f), and RV (g) variant sets.



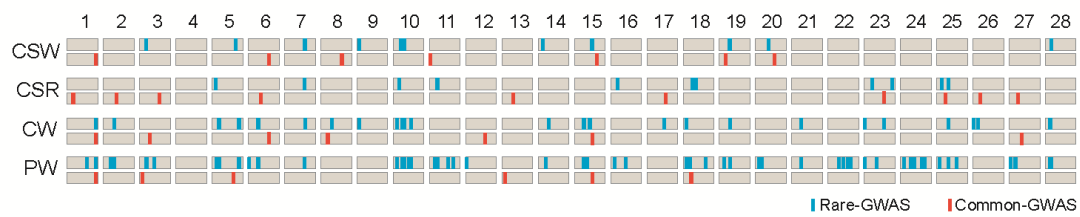
**Supplementary Fig. 10** | Shared status of genes associated with various cocoon silk yield traits. Depicts the unique and overlapping distribution of genes associated with the four cocoon silk yield traits (CSW, CSR, CW, PW).



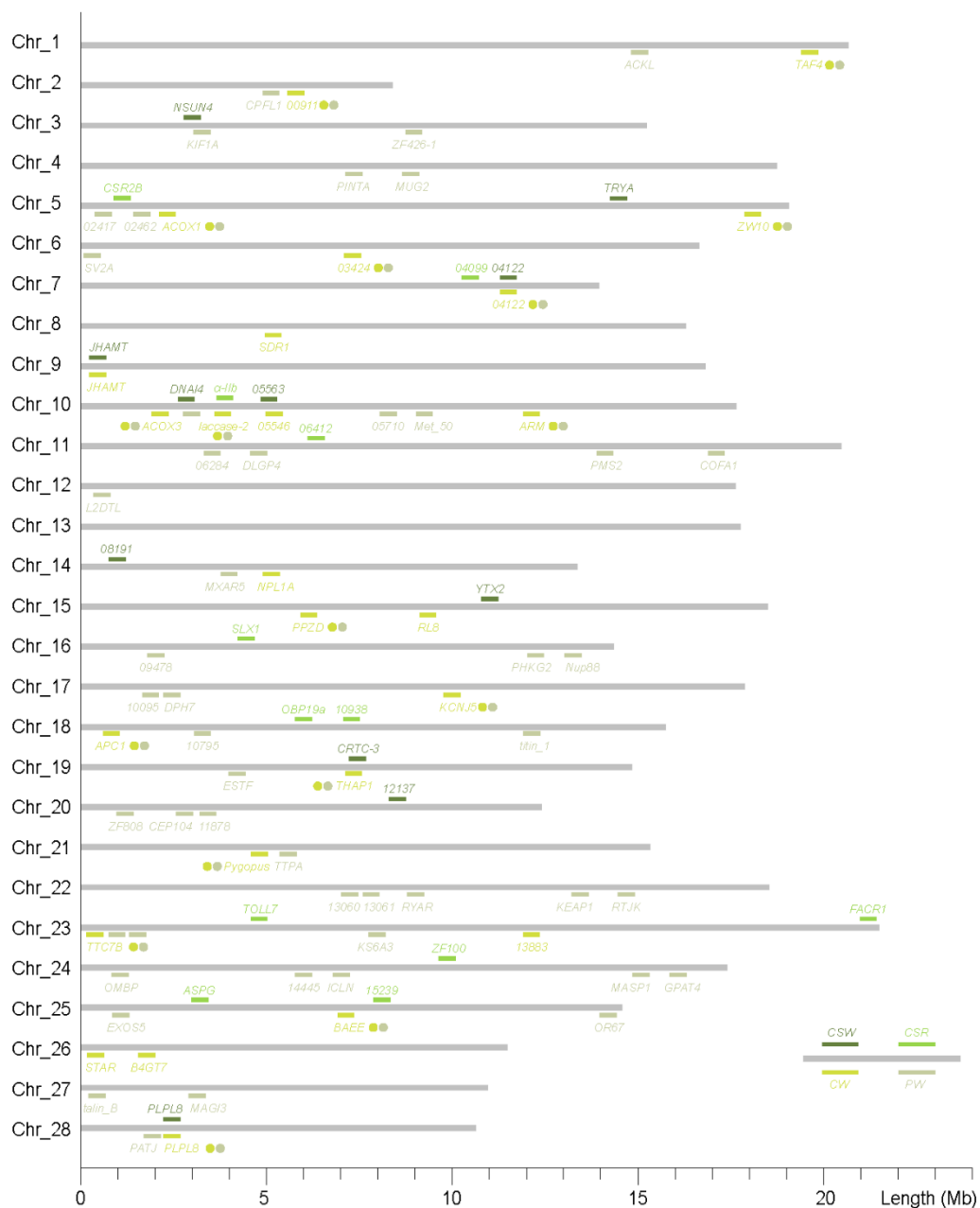
**Supplementary Fig. 11 |** Manhattan plots of GWAS for cocoon silk yield traits based on common variants. Using genome-wide common SNPs from 459 germplasm resources, a mixed linear model (MLM) was employed to perform GWAS for 4 silk yield traits (from top to bottom: CSR, CSW, CW, PW), with the top five principal components included as covariates to correct for population structure. The plot indicates the genome-wide Bonferroni correction significance threshold ( $P < 0.05/8,995,741$ ) and the suggested threshold ( $P = 5 \times 10^{-8}$ ).



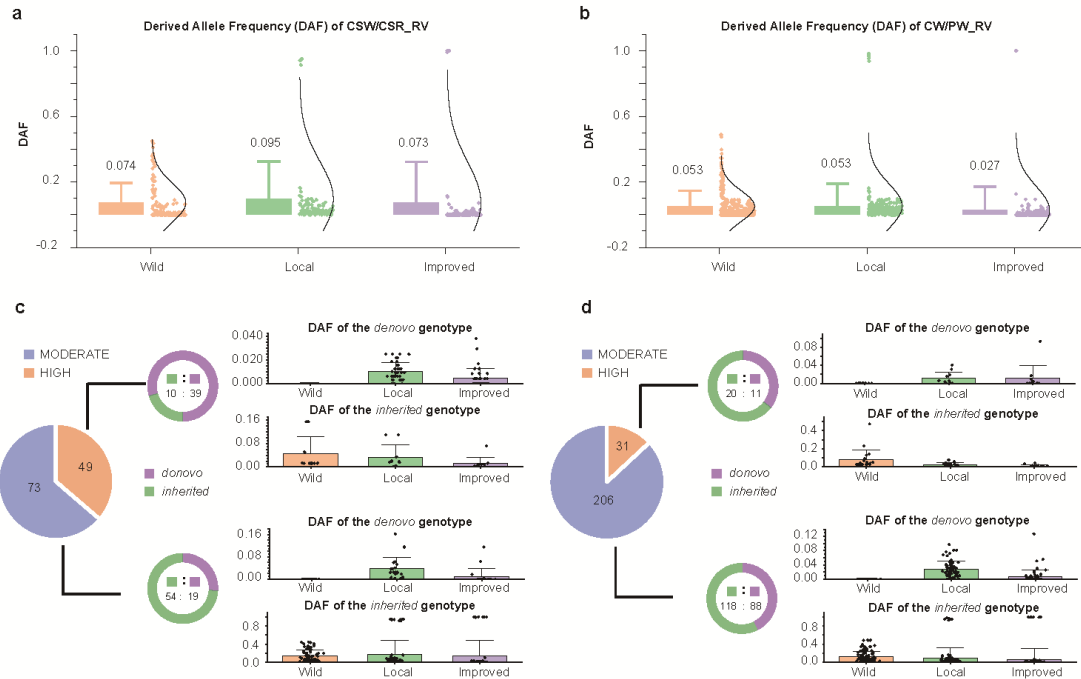
**Supplementary Fig. 12 |** Manhattan plots and QQ-plots of GWAS for cocoon silk yield traits based on common SNPs. Using genome-wide common SNPs from 459 germplasm resources, the FarmCPU model was employed to perform GWAS for 4 silk yield traits (from top to bottom: CSW, CSR, CW, PW), with the top five principal components included as covariates to correct for population structure. The plots indicate the genome-wide Bonferroni correction significance threshold ( $P < 0.05/8,995,741$ ). The corresponding QQ-plots on the right are used to evaluate the calibration of the association tests.



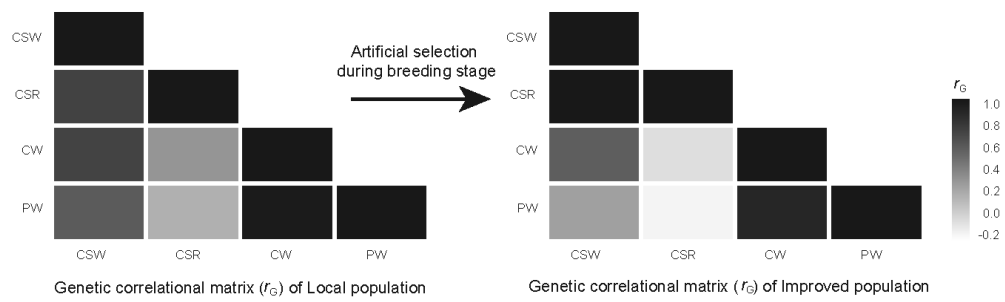
**Supplementary Fig. 13 |** Comparison of association signals between rare variants and common SNPs. The figure illustrates the distribution across chromosomes of significant association signals from common SNP-based GWAS (red) and genes associated with rare variants (blue). The results demonstrate no overlap between the two types of signals in the genome.



**Supplementary Fig. 14 |** Genomic distribution of genes associated with cocoon silk yield traits. Genes associated with each trait are marked in different colors (as shown in the legend in the lower right corner) at their corresponding chromosomal positions, represented by abbreviations or identifiers. Double dots indicate genes shared by the CW and PW traits. Complete gene information can be found in Supplementary Table 3.



**Supplementary Fig. 15 |** Analysis of the frequency distribution in different silkworm populations and functional annotation for rare variant of genes associated with cocoon silk yield traits. Panels a and b display the derived allele frequency (DAF) distribution of rare variants in genes associated with CSW/CSR (a) and CW/PW (b) across Wild (n = 39 biologically independent accessions), Local (n = 160 biologically independent accessions), and Improved (n = 118 biologically independent accessions) silkworms, respectively. Panels c and d correspond to CSW/CSR (c) and CW/PW (d) genes, showing: the proportional distribution of HIGH and MODERATE impact variants (left pie chart), the DAF of 'denovo' and 'inherited' genotype of RVs across the three populations (box plots).



**Supplementary Fig. 16 |** The genetic correlation of paired silk yield traits in different population. The left and right panel show the genetic correlational matrix ( $r_G$ ) of Local ( $n = 348$  biologically independent accessions) and Improved ( $n = 107$  biologically independent accessions) population respectively.