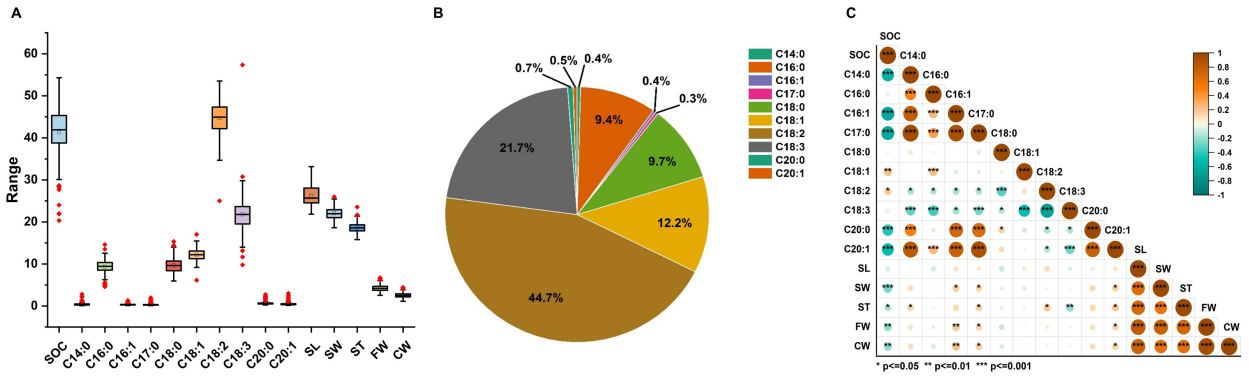


Fig. S1 | Phenotypic distribution, fatty acid (FA) composition profile, and correlation analysis of sixteen traits across 186 rubber tree accessions. A Boxplots showing the phenotypic distribution of seed oil content (SOC), FA composition traits—myristic acid (C14:0), palmitic acid (C16:0), palmitoleic acid (C16:1), margaric acid (C17:0), stearic acid (C18:0), oleic acid (C18:1), linoleic acid (C18:2), linolenic acid (C18:3), arachidic acid (C20:0), eicosenoic acid (C20:1), and seed size traits—seed length (SL), seed width (SW), seed thickness (ST), fruit weight (FW), and caryopsis weight (CW). Each box represents the interquartile range, with the horizontal line indicating the median, and individual points representing outliers. **B** Pie chart illustrating the relative proportional composition of individual FAs. **C** Pearson correlation heatmap showing pairwise correlation coefficients among SOC, FA composition, and seed size traits. Color scale ranges from deep brown (r = +1) to deep green (r = −1). Correlation coefficients are displayed within each cell, with statistically significant correlations indicated by asterisks (*P < 0.05, **P < 0.01, ***P < 0.001).

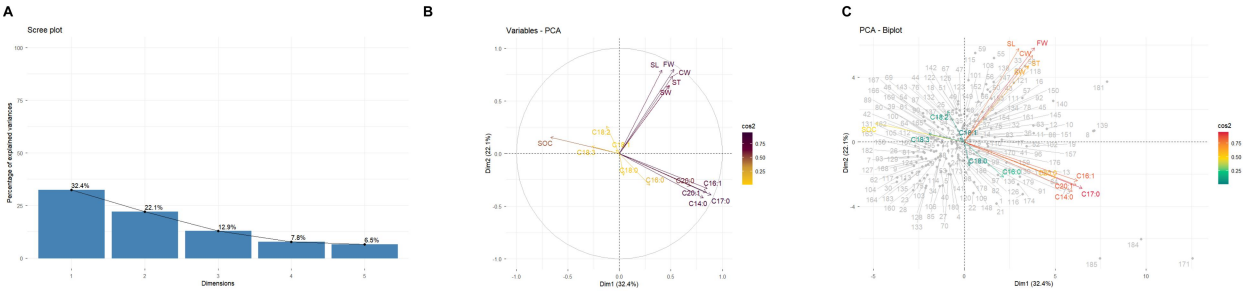
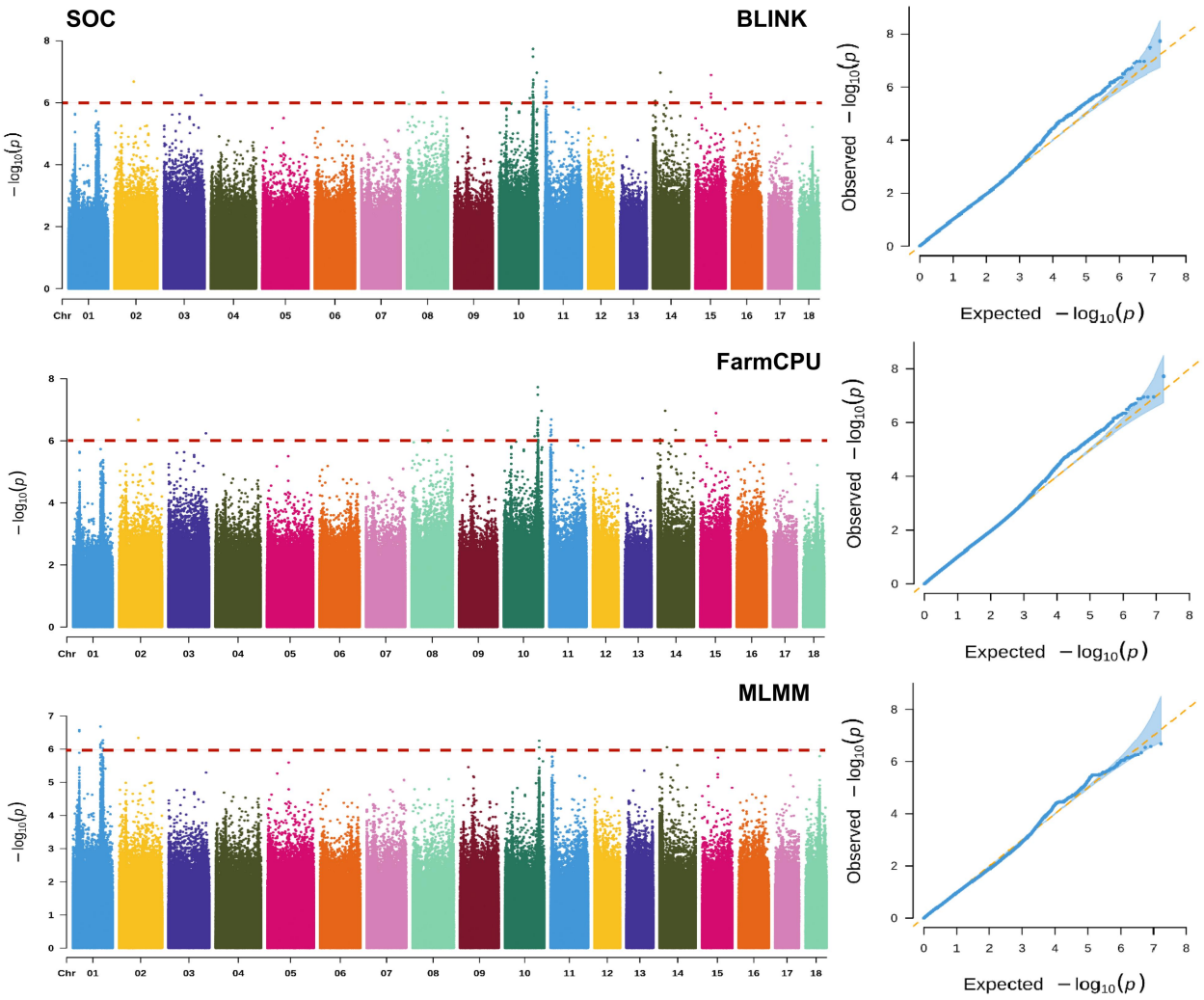


Fig. S2 | Principal component analysis of phenotypic variation in seed oil content (SOC), fatty acid (FA) composition, and seed size traits across 186 *Hevea brasiliensis* accessions. A Scree plot showing the eigenvalues and cumulative percentage of total phenotypic variance explained by each principal component (PC) derived from correlation matrix decomposition of all sixteen evaluated traits. The horizontal dashed line indicates the eigenvalue threshold of 1, above which PCs are considered to contribute significantly to total variation. **B** PCA loading plot (variable plot) showing the contribution and direction of each trait variable along PC1 and PC2 axes. Arrow length indicates the relative contribution of each variable to the PCs, and arrow direction indicates positive or negative association with each axis. Variables with similar directions indicate positive co-associations, while opposing directions indicate negative associations. **C** PCA biplot overlaying individual accession scores and trait variable loadings along PC1 and PC2, illustrating the distribution of all accessions in multivariate trait space and their associations with specific trait clusters. Accession identifiers are labeled where relevant for interpretation. The percentage of total phenotypic variance explained by each PC is indicated in parentheses on the respective axis labels. C14:0, Myristic acid; C16:0, palmitic acid; C16:1, palmitoleic acid; C17:0, margaric acid; C18:0, stearic acid; C18:1, oleic acid; C18:2, linoleic acid; C18:3, linolenic acid; C20:0, arachidic acid; C20:1, eicosenoic acid; SL, seed length; SW, seed width; ST, seed thickness; FW, fruit weight; CW, caryopsis weight.


SOC
BLINK
FarmCPU
MLMM
−log10(p)
Observed −log10(p)
Expected −log10(p)
Chr 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18

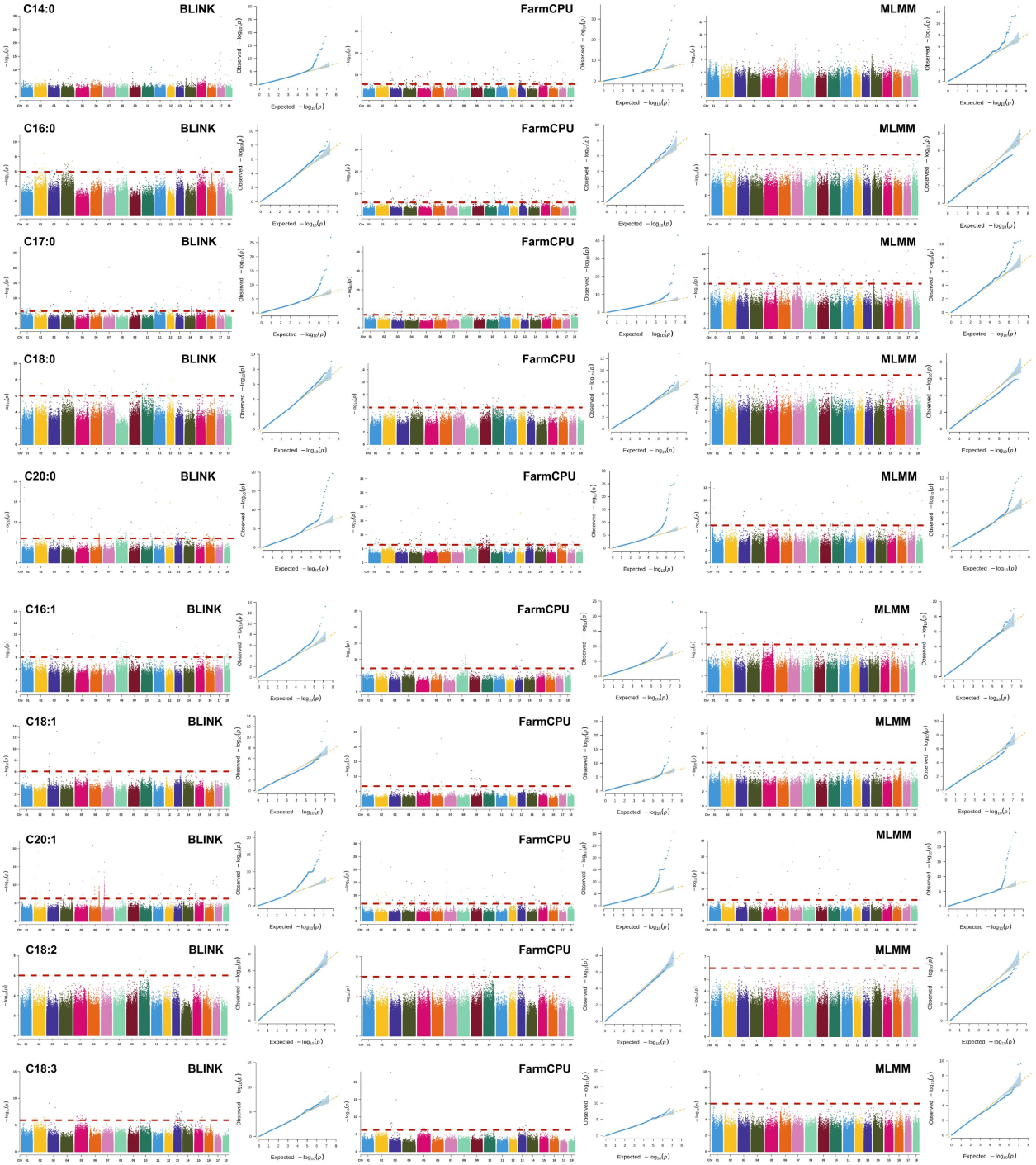


Fig. S3 | Manhattan plots and Quantile–Quantile (QQ) plots for seed oil content (SOC) and fatty acid (FA) composition traits across three GWAS models. Manhattan plots (left panels) and QQ plots (right panels) for SOC and FA composition traits—generated from GWAS models (BLINK, FarmCPU, and MLMM). In Manhattan plots, the x-axis represents chromosomal position (Mb) and the y-axis represents $-\log_{10}(P\text{-value})$. The red horizontal dashed line indicates the genome-wide significance threshold ($-\log_{10}(P) > 6$). Each point represents a single SNP, colored alternately by chromosome for visual clarity. In QQ plots, the x-axis represents expected $-\log_{10}(P\text{-values})$ under the null hypothesis, and the y-axis represents observed $-\log_{10}(P\text{-values})$ from the GWAS analysis. Deviation of points from the diagonal (red line) in the upper tail indicates genuine association signals. C14:0, Myristic acid; C16:0, palmitic acid; C16:1, palmitoleic acid; C17:0, margaric acid; C18:0, stearic acid; C18:1, oleic acid; C18:2, linoleic acid; C18:3, linolenic acid; C20:0, arachidic acid; C20:1, eicosenoic acid.

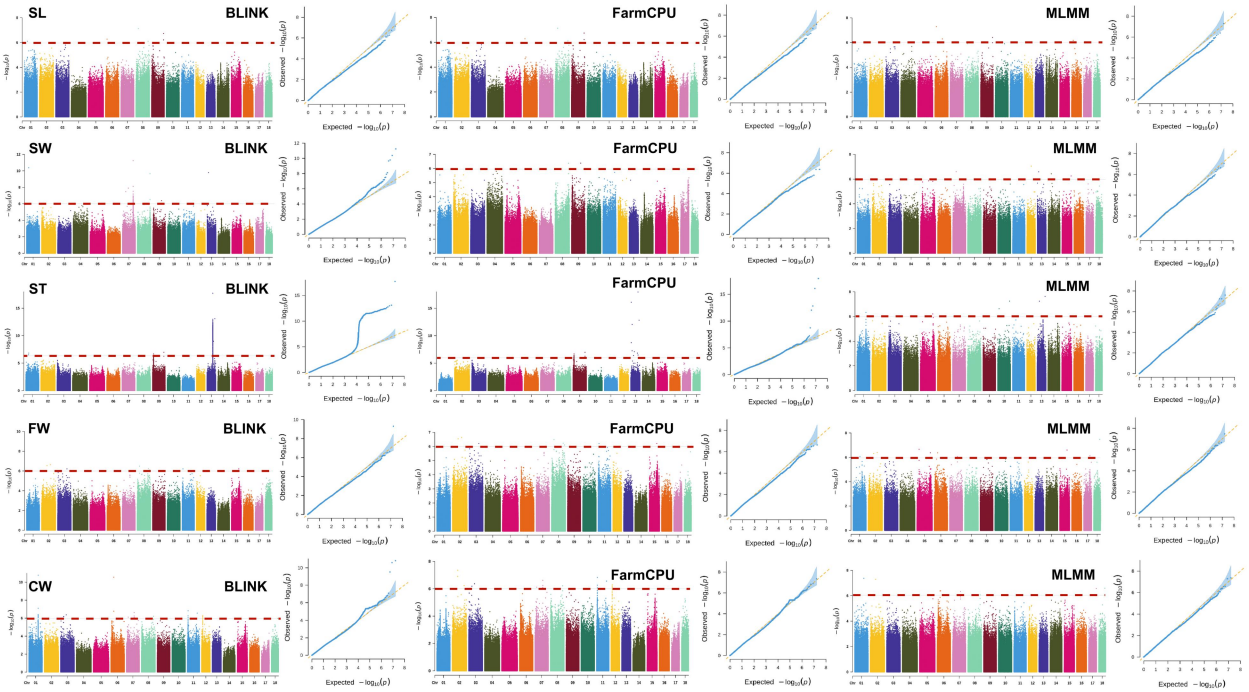


Fig. S4 | Manhattan plots and Quantile–Quantile (QQ) plots for seed size traits across three GWAS models. Manhattan plots (left panels) and QQ plots (right panels) for seed size traits generated from three GWAS models (BLINK, FarmCPU, and MLMM). In Manhattan plots, the x-axis represents chromosomal position (Mb) and the y-axis represents $-\log_{10}$(P-value). The red horizontal dashed line indicates the genome-wide significance threshold ($-\log_{10}(P) > 6$). Each point represents a single SNP, colored alternately by chromosome for visual clarity. In QQ plots, the x-axis represents expected $-\log_{10}$(*P*-values) under the null hypothesis, and the y-axis represents observed $-\log_{10}$(*P*-values) from the GWAS analysis. SL, seed length; SW, seed width; ST, seed thickness; FW, fruit weight; CW, caryopsis weight.

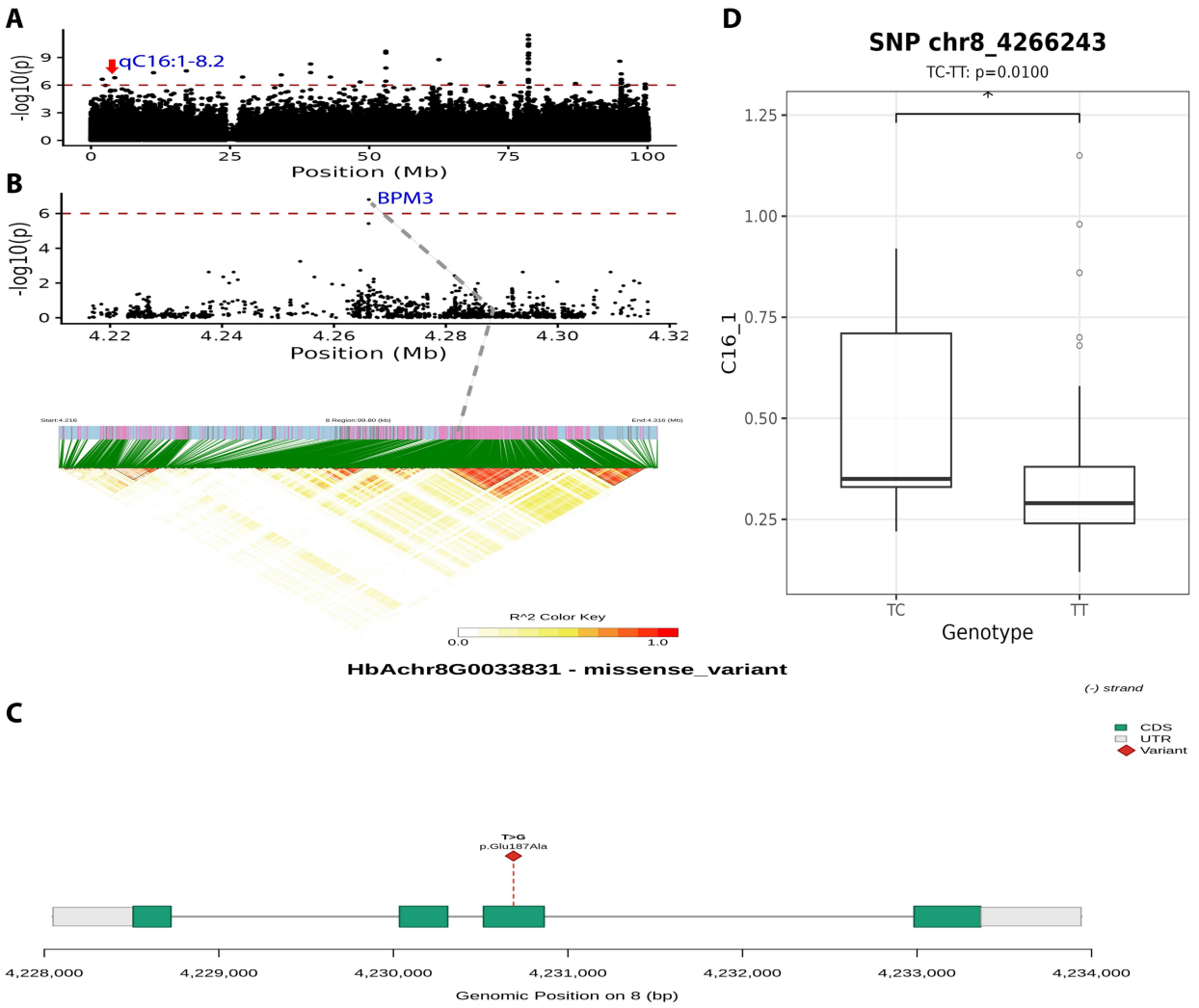


Fig. S5 | Fine mapping and candidate gene identification of HbAchr8G0033831 at locus (qC16:1-8.2). **A** Genome-wide Manhattan plot showing association signals for C16:1 across all chromosomes, with $-\log_{10}$(P-value) on the y-axis and chromosomal position (Mb) on the x-axis. The red horizontal dashed line indicates the genome-wide significance threshold ($-\log_{10}(P) > 6$). The peak association signal at qC16:1-8.2 on Chr08 is indicated by the red arrow. **B** Zoomed regional Manhattan plot of the qC16:1-8.2 interval (4.22–4.32 Mb) on Chr08, with the candidate gene BPM3 labeled in blue. The grey dashed lines connect the peak association signal in panel B to the LD heatmap. The LD heatmap below displays pairwise R^2 values among SNPs within the interval, color-coded from white/yellow according to the R^2 color key. **C** Gene structure of HbAchr8G0033831 on Chr08, showing five exons with CDS in teal and UTR in grey. One non-synonymous missense variant is annotated with a red diamond symbol at its position within the CDS: T>G (p.Glu187Ala) near 4,231,000 bp, indicated by a red dashed vertical line. Genomic coordinates (bp) are indicated on the x-axis (4,228,000–4,234,000 bp). **D** Boxplot showing the genotype–phenotype association of the lead SNP (8_4266243) with C16:1 content, comparing TC (heterozygous) and TT (homozygous) genotypes. The y-axis represents C16:1 content. The horizontal line within each box indicates the median, box boundaries represent the 25th and 75th percentiles, and open circles represent outliers. The TC genotype showed significantly higher C16:1 content relative to TT (**P*=0.0100). BPM3, BTB/POZ domain and MATH domain-containing protein 3; C16:1, palmitoleic acid; CDS, coding sequence; UTR, untranslated region; LD, linkage disequilibrium; FA, fatty acid.

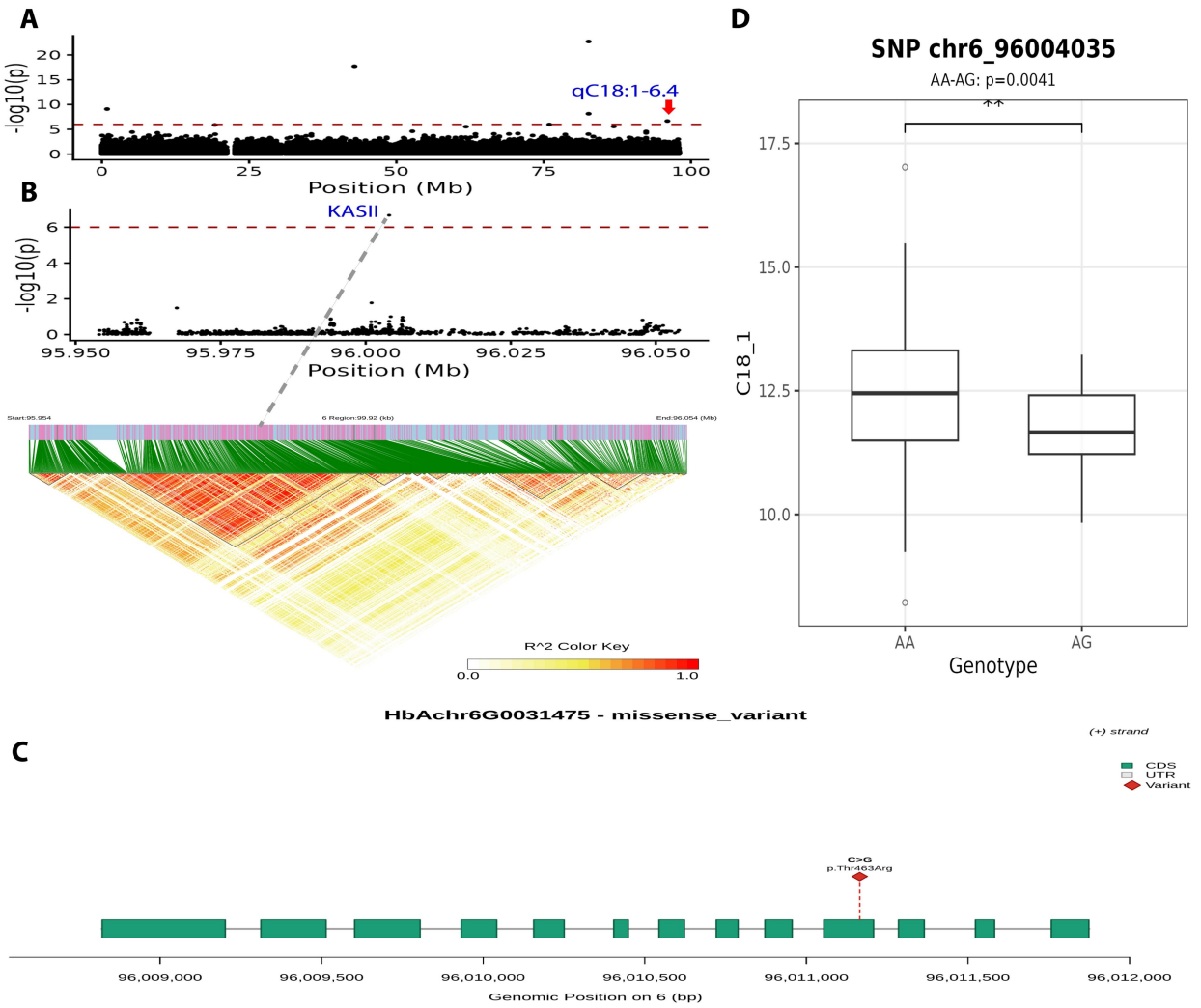


Fig. S6 | Fine mapping and candidate gene identification of HbAchr6G0031475 at locus (qC18:1-6.4). A Genome-wide Manhattan plot showing association signals for C18:1 across all chromosomes, with $-\log_{10}$(P-value) on the y-axis and chromosomal position (Mb) on the x-axis. The red horizontal dashed line indicates the genome-wide significance threshold ($-\log_{10}(P) > 6$). The peak association signal at qC18:1-6.4 on Chr06 is indicated by the red arrow, with the QTL name labeled in blue above the peak. **B** Zoomed regional Manhattan plot of the qC18:1-6.4 interval (95.95–96.05 Mb) on Chr06, with the candidate gene KASII labeled in blue above the regional plot. The grey dashed lines connect the peak association signal in panel B to the LD heatmap. The LD heatmap below the regional Manhattan plot displays pairwise R^2 values among all SNPs within the displayed interval, color-coded from white/yellow according to the R^2 color key. The extensive LD block structure observed across the KASII gene interval reflects low local recombination and supports confident candidate gene localization. **C** Gene structure of HbAchr6G0031475 on Chr06, showing the multi-exon organization spanning approximately 96,009,000–96,012,000 bp. CDS are shown as teal blocks, UTR as grey blocks, and introns as connecting lines between exons. One non-synonymous missense variant is annotated with a red diamond symbol and red dashed vertical line at its position within the CDS: C>G (p.Thr463Arg) near 96,011,000 bp. Genomic coordinates are indicated on the x-axis (96,009,000–96,012,000 bp). **D** Boxplot showing the genotype–phenotype association of the lead SNP (6_96004035) with C18:1 content, comparing AA (homozygous) and AG (heterozygous) genotypes. The y-axis represents C18:1 content. The horizontal line within each box indicates the median, box boundaries represent the 25th and 75th percentiles, and open circles represent outliers. AA homozygous accessions exhibited significantly higher C18:1 content compared to AG heterozygous accessions (**P=0.0041). KASII, β-ketoacyl-ACP synthase II; C18:1, oleic acid; CDS, coding sequence; UTR, untranslated region; LD, linkage disequilibrium; FA, fatty acid.

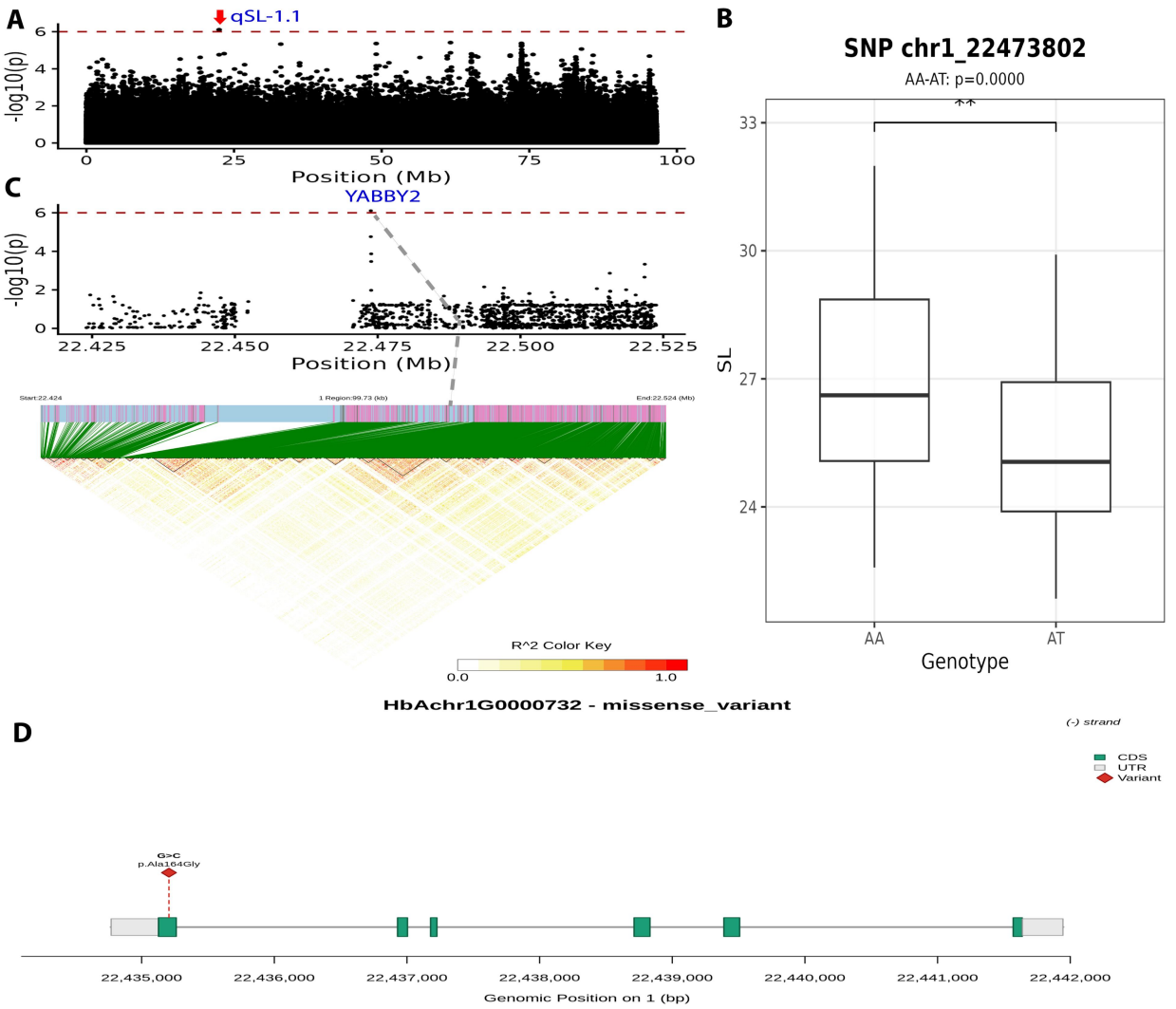


Fig. S7 | Fine mapping and candidate gene identification of HbAchr1G0000732 at locus (qSL-1.1). A Genome-wide Manhattan plot showing association signals for SL across all chromosomes, with $-\log_{10}$(P-value) on the y-axis and chromosomal position (Mb) on the x-axis. The red horizontal dashed line indicates the genome-wide significance threshold ($P < 6$). The peak association signal on Chr01 at qSL-1.1 is indicated by the red arrow, labeled with the QTL name. **B** Zoomed regional Manhattan plot of the qSL-1.1 interval (22.37–22.57 Mb) on Chr01, with the candidate gene YABBY2 labeled in blue. The grey dashed line connects the genome-wide peak in panel B to the LD heatmap. The LD heatmap below the regional plot displays pairwise R^2 values among SNPs within the interval, color-coded from white/yellow according to the R^2 color key. **C** Gene structure of HbAchr1G0000732 on Chr01, showing exon–intron organization with CDS in teal, UTR in grey, and introns represented as connecting lines. One non-synonymous variant is annotated with a red diamond symbol at its position within the CDS: G>C (p.Ala164Gly) near 22,435,000 bp. The gene is transcribed from the (−) strand. Genomic coordinates are indicated on the x-axis (22,435,000–22,442,000 bp). **D** Boxplot showing the genotype–phenotype association of the lead SNP (1_22473802) with SL, comparing AA (homozygous) and AT (heterozygous) genotypes. The y-axis represents SL. The horizontal line within each box indicates the median, box boundaries represent the 25th and 75th percentiles. A highly significant difference between genotype classes is indicated above the plot (***P<0.0001). YABBY2, YABBY transcription factor 2; SL, seed length; CDS, coding sequence; UTR, untranslated region; LD, linkage disequilibrium.