

Figure S1. Distribution of agronomic traits in Biluochun tea plant varieties. Red characters represented that frequency distributions of agronomic traits approximately meet normal distributions. Black characters represented that frequency distributions of agronomic traits do not meet normal distributions.

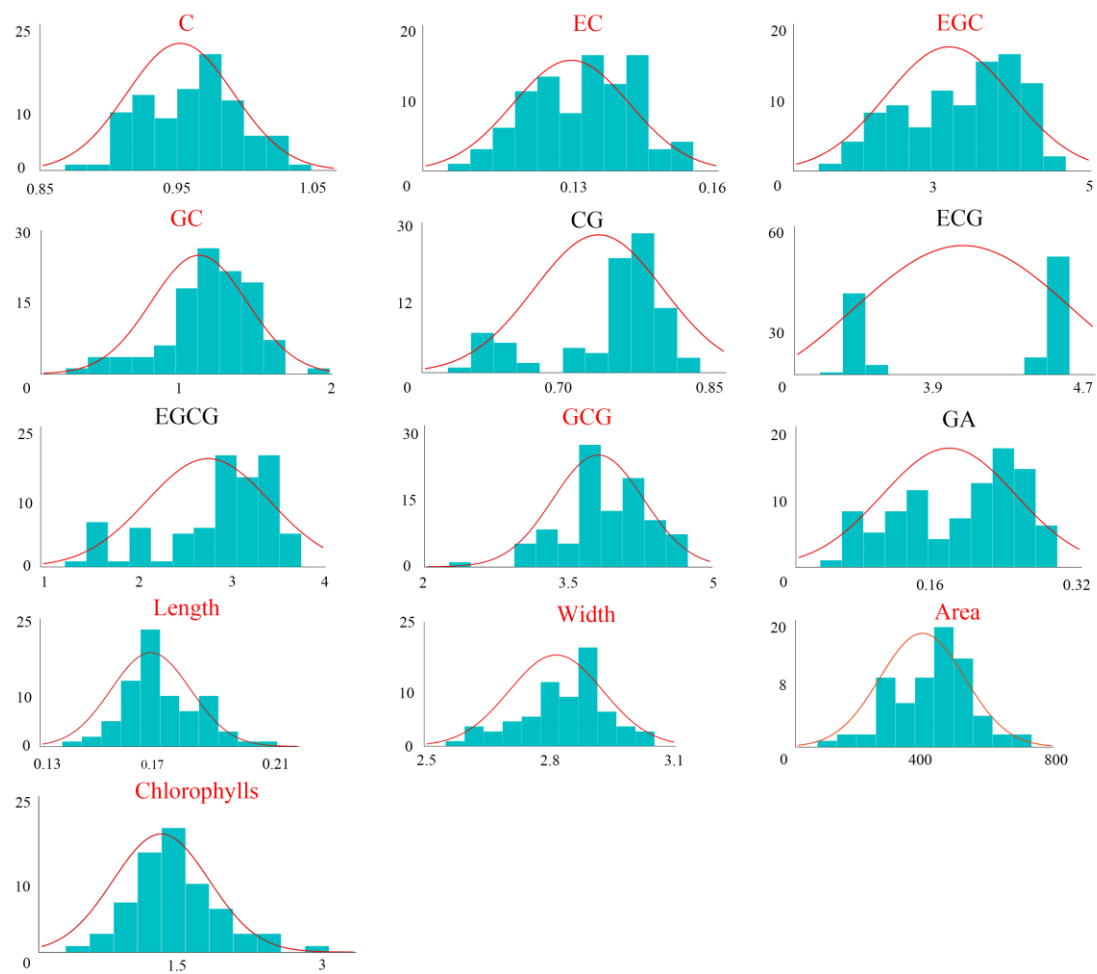


Figure S2. The correlation coefficient between pairwise metabolites, including SS, TP, TC, FAA, CAF, C, GC, GA, EC, EGC, ECG, CG, GCG, EGCG. Blue and positive numbers represent positive correlation, red and negative numbers represent negative correlation (* stands for p -value less than 0.05 and ** stands for p -value less than 0.01).

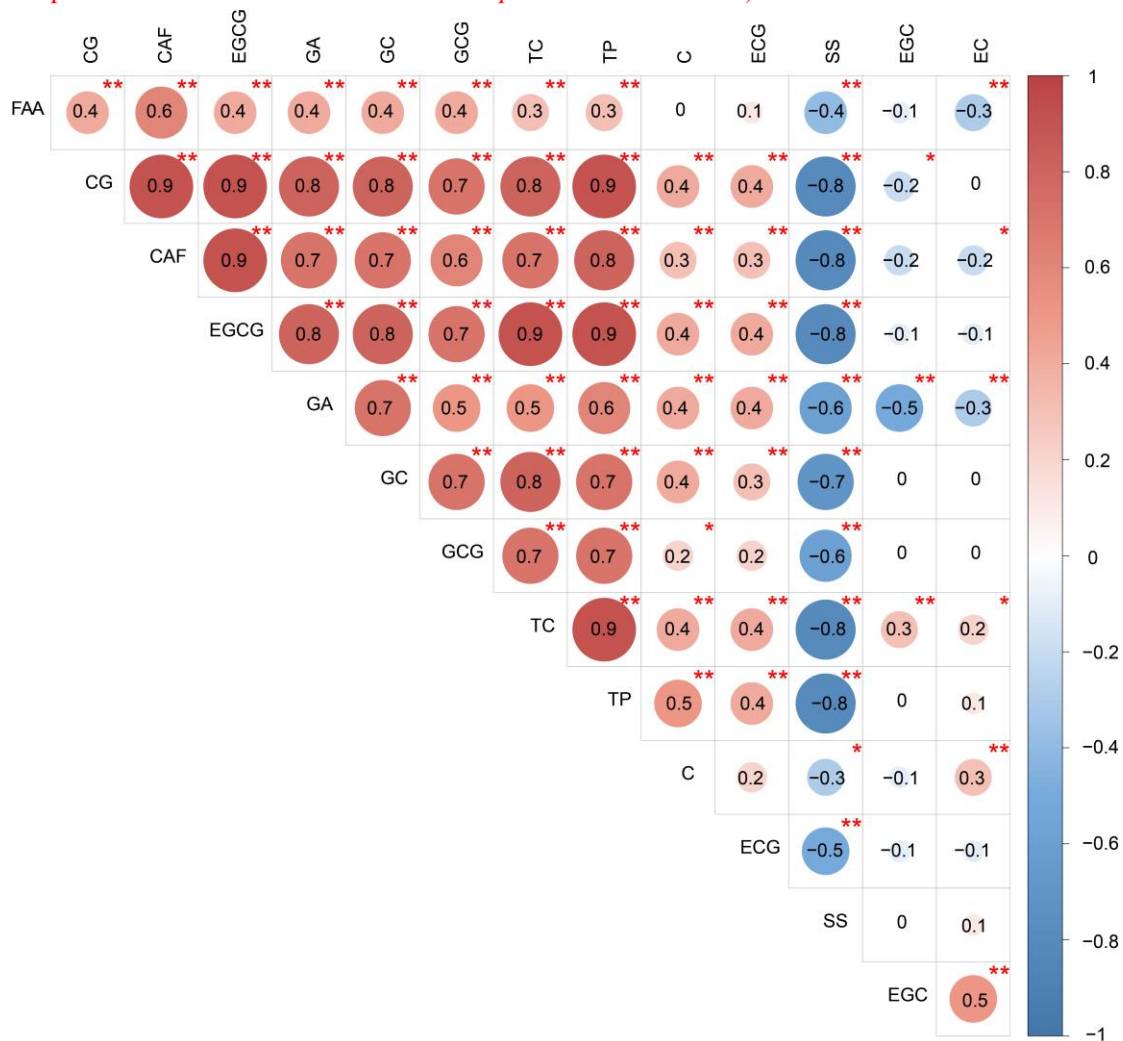


Figure S3. Population structure of 95 tea plant varieties. a. Cross validation of 95 Biluochun tea varieties from K=2 to K=9 clusters. b. Model-based Bayesian clustering of 95 Biluochun tea varieties.

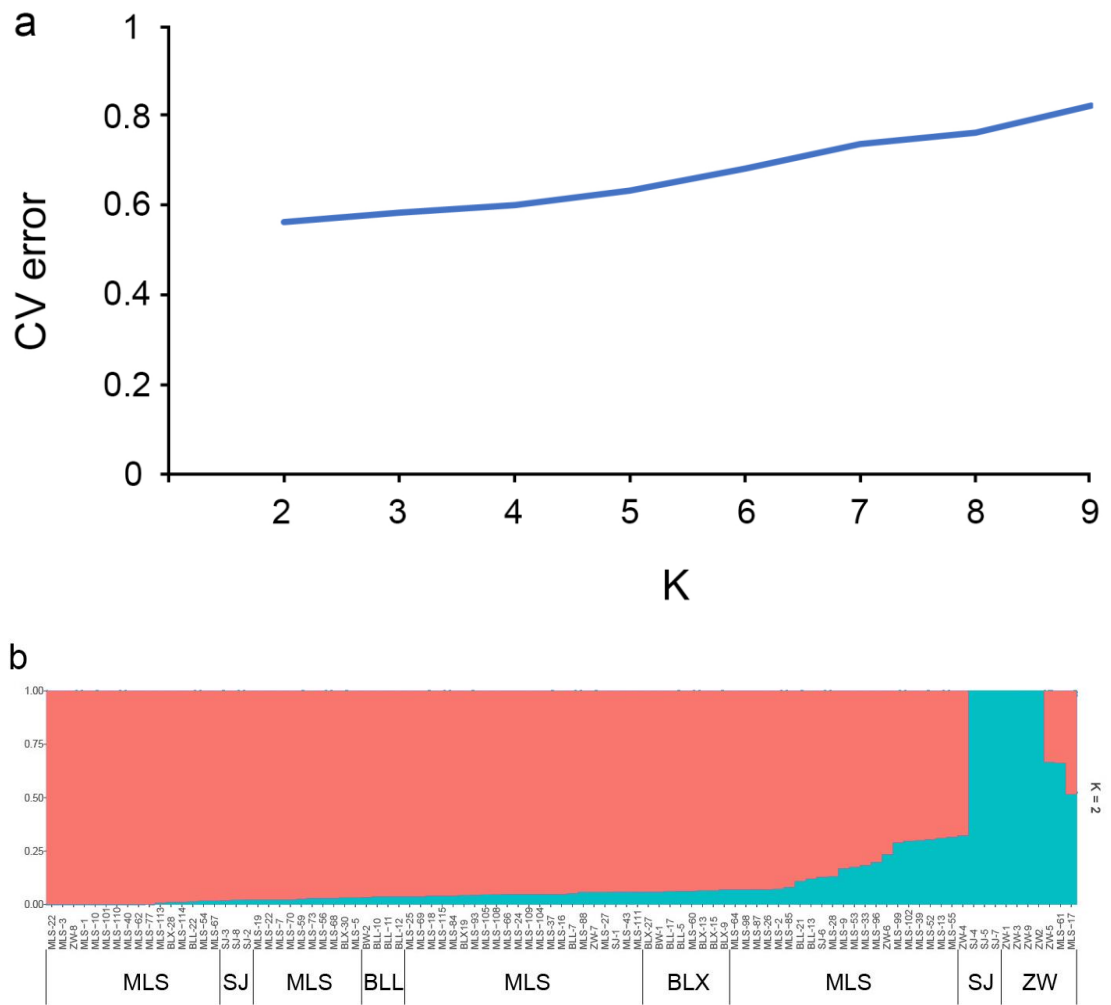


Figure S4. Distribution of SNP markers on chromosomes. The x-axis represents the number of SNPs in the 25 Mb window, and the y-axis represents 15 chromosomes of tea plant. Different colors represent different density of SNPs.

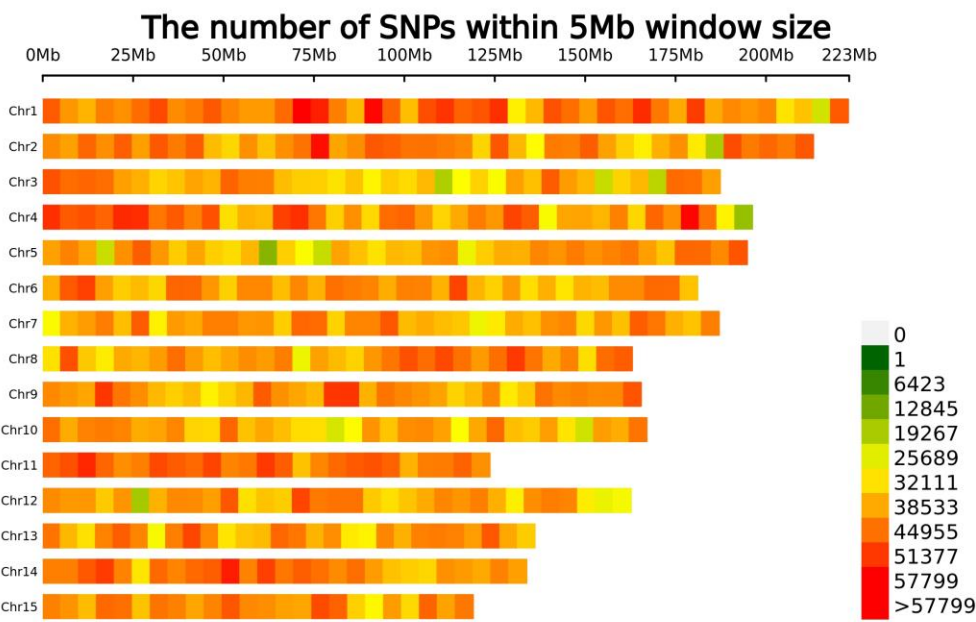
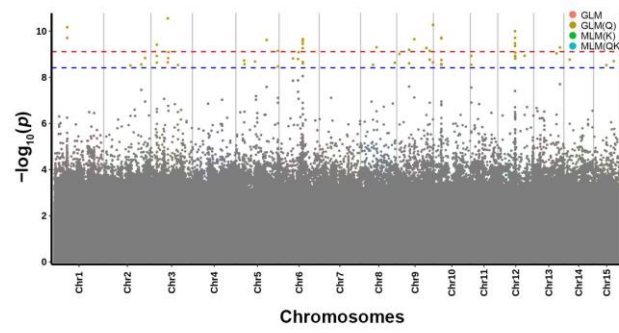
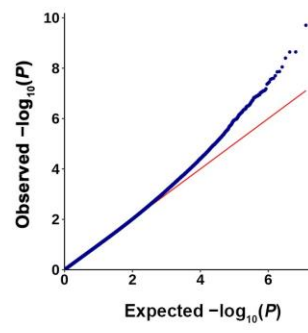


Figure S5. Manhattan and Quantile-quantile (QQ) plots of the GWAS for 14 characteristic traits. (A-N). Manhattan plots contain three phenotypic traits (leaf length, width and area); ten flavor-related metabolites (SS, FAA, TP, TC, C, EC, EGC, CG, ECG and EGCG) and a growth-related metabolite Chl. The x-axis displays the chromosome label, and the y-axis displays bonferroni correction threshold $-\log_{10}(p)$. The blue and red dashed lines show a significant correlation between SNPs and phenotypic values, with two thresholds $p\text{-value} < 3.89 \times 10^{-9}$ and $p\text{-value} < 7.78 \times 10^{-10}$, respectively. For quantile-quantile plots, the horizontal axis shows $-\log_{10}$ transformed expected p values, and the vertical axis indicates $-\log_{10}$ transformed observed p values. The red line represents the predicted value and the blue dot represents the observed value, which can show the difference between the predicted value and the observed value. Red, yellow, green, and cyan dots represent GLM, GLM(Q), MLM(K) and MLM(QK). (a-n) QQ plots represent the reliability of GLM analysis based on the above fourteen traits.

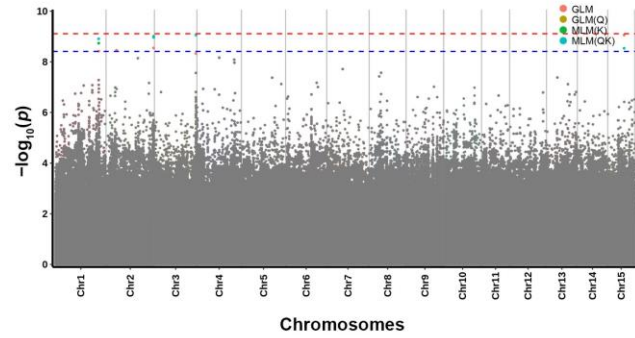
A Soluble sugar



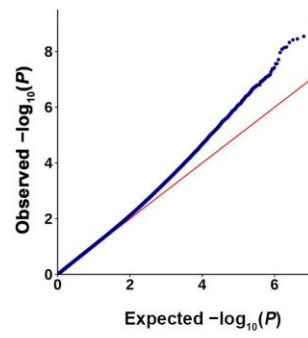
a



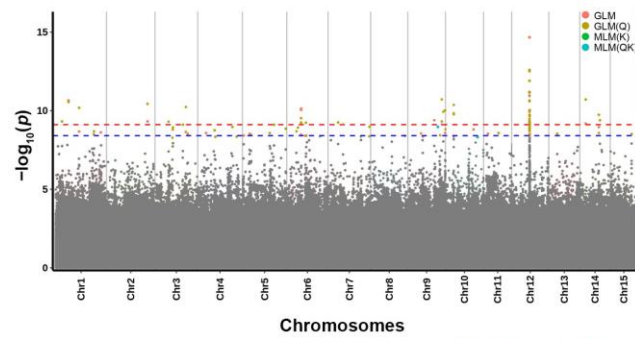
B Free amino acid



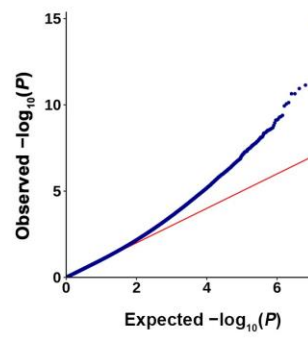
b



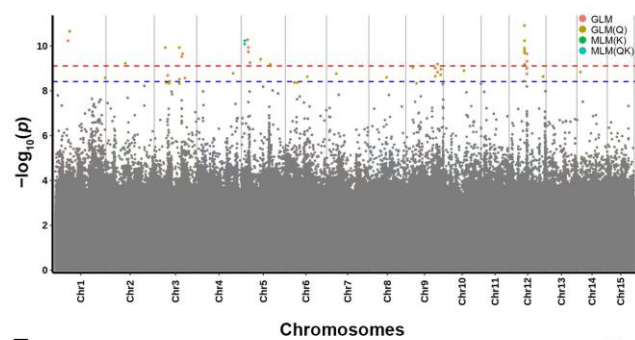
C Total polyphenols



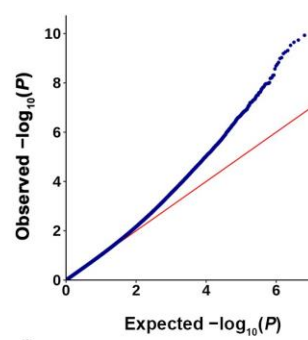
c



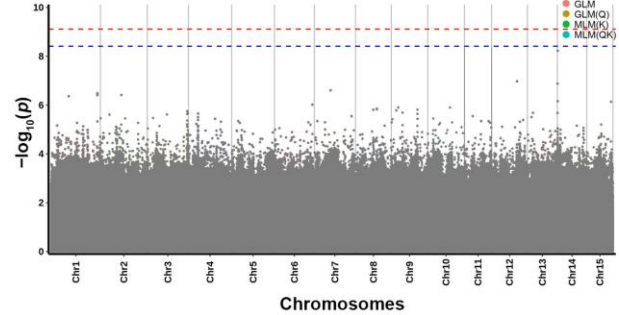
D Total catechins



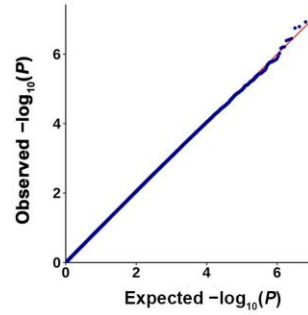
d

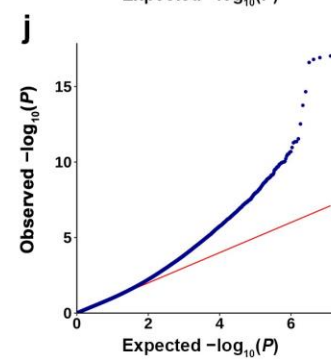
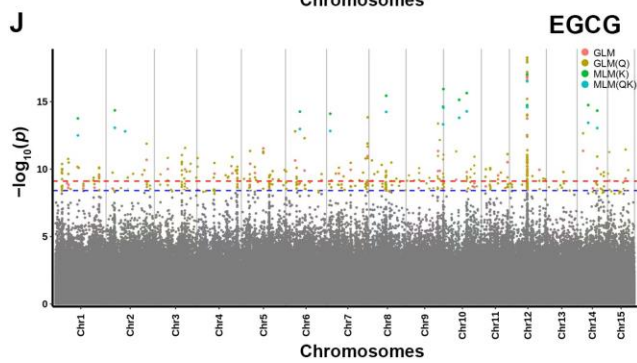
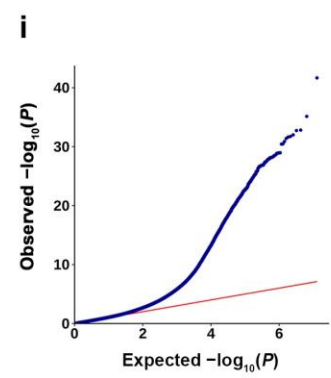
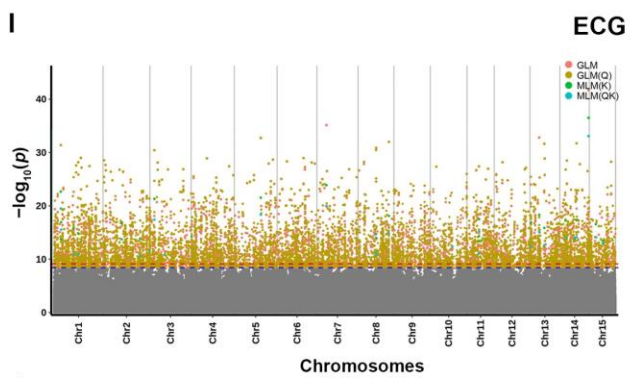
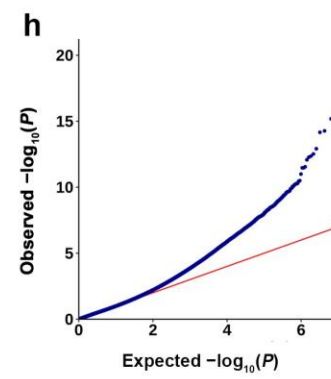
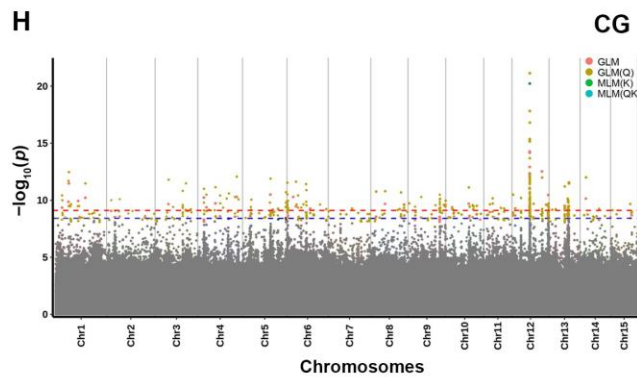
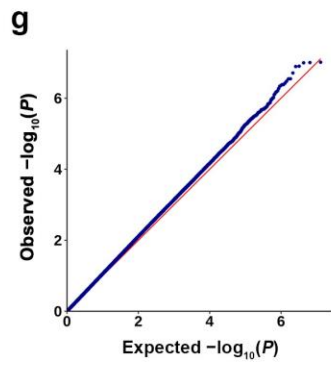
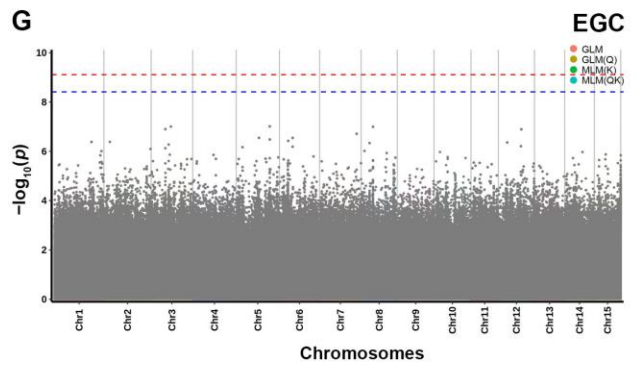
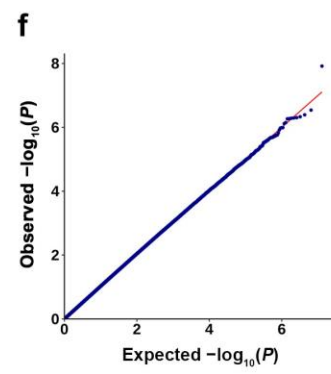
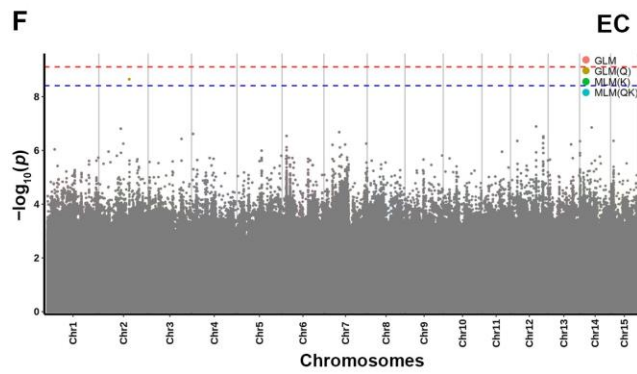


E



e





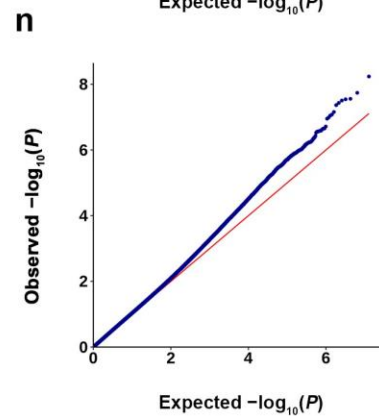
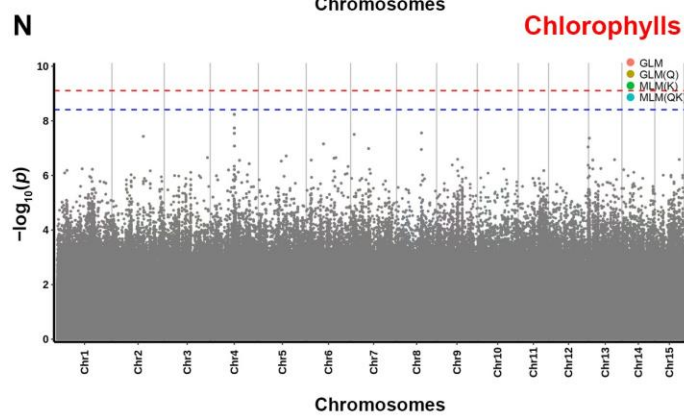
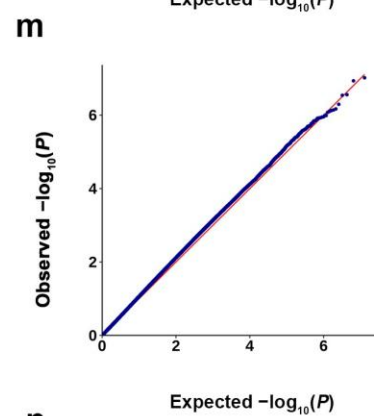
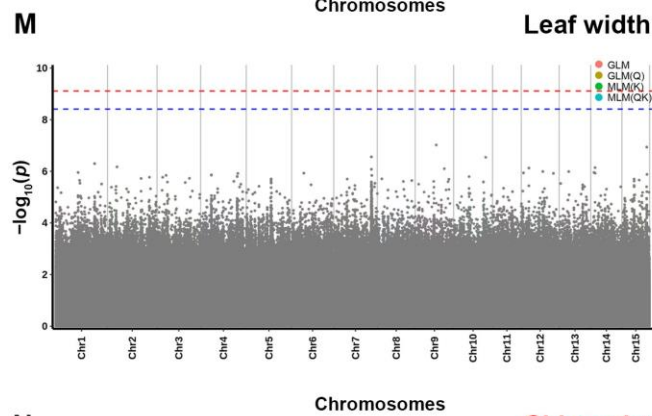
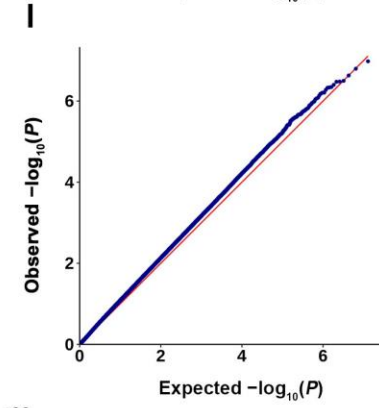
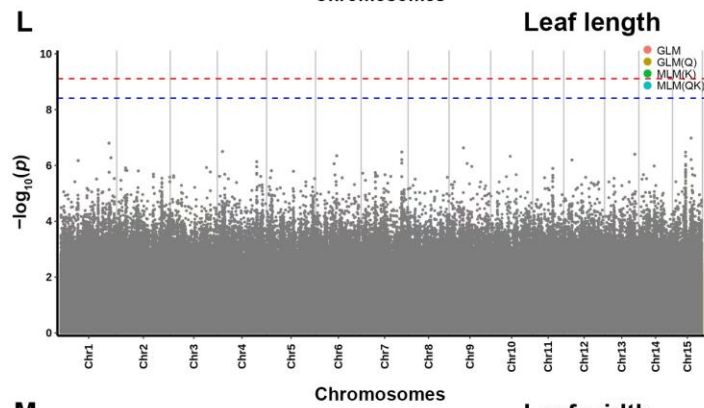
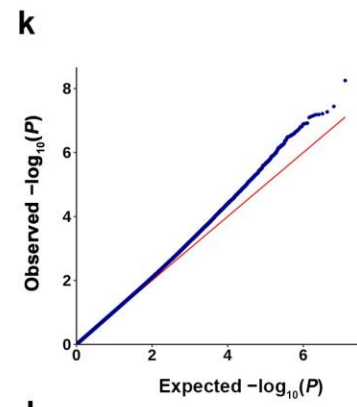
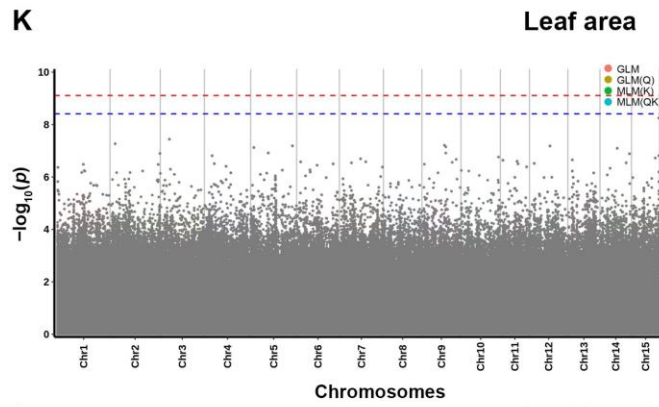
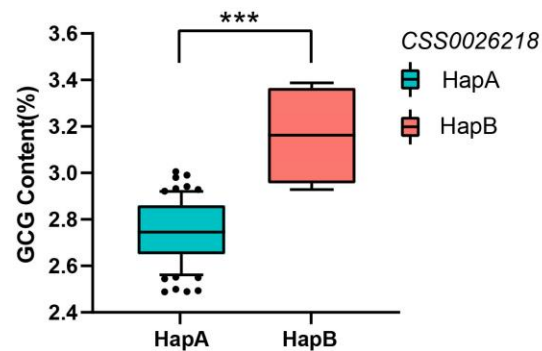


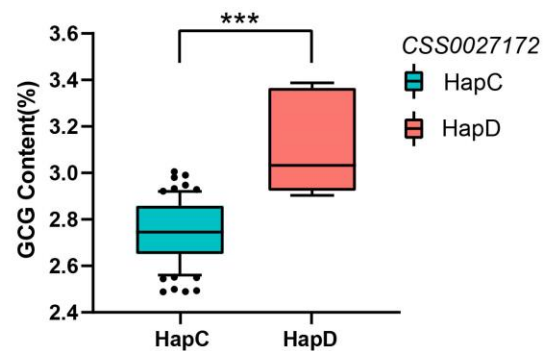
Figure S6. The sequence alignment of *CSS0039426*, *CSS0044862*, *CSS0026218* and *CSS0027172* related to GCG, GC and CAF and the distribution of nsSNPs in different haplotype types.

<p>*****</p> <p><i>CSS0026218</i> CSS</p> <p><i>CSS0026218</i> HapA with high GCG</p> <p><i>CSS0026218</i> HapB with low GCG</p> <p>Consensus</p>	<p>TAA G T T T G A G G A G T T G C A A A A G G C C A C A G G G T T C T T C G G C 660</p> <p>TAA G T T T G A G G A G T T G C A A A A G G C C A C A G G G T T C T T C G G C 660</p> <p>TAA G T T T G G G G A G T T G C A A A A G G C C A C A G G G T T C T T C G G C 660</p> <p>TAA G T T T G G A G T T G C A A A A G G C C A C A G G G T T C T T C G G C 660</p>	<i>CSS0026218</i>
<p>*****</p> <p><i>CSS0027172</i> CSS</p> <p><i>CSS0027172</i> HapC with high GCG</p> <p><i>CSS0027172</i> HapD with low GCG</p> <p>Consensus</p>	<p>T T T A C C T T C A C C T A C T C C A C T A C A C A C T C C A C T T C A C C T A 900</p> <p>T T T A C C T T C A C C T A C T C C A C T A C A C A C T C C A C T T C A C C T A 900</p> <p>T T T A C C T T C A C C T A C T C C A C T A C A C T C T C C A C T T C A C C T A 900</p> <p>T T T A C C T T C A C C T A C T C C A C T A C A C C T C C A C T T C A C C T A 900</p>	<i>CSS0027172</i>
<p>*****</p> <p><i>CSS0044862</i> CSS</p> <p><i>CSS0044862</i> HapG with high GC</p> <p><i>CSS0044862</i> HapH with low GC</p> <p>Consensus</p>	<p>G G T A A C C C C C A T G A T T A T C A T A T C C T C G A T G T T T T A C A C A 231</p> <p>G G T A A C C C C C A T G A T T A T C A T A T C C T C G A T G T T T T A C A C A 231</p> <p>G G T A A C C C C C A T G A T T G T C A T A T C C T C G A T G T T T T A C A C A 231</p> <p>G G T A A C C C C C A T G A T T T C A T A T C C T C G A T G T T T T A C A C A 231</p>	<i>CSS0044862</i>
<p>*****</p> <p><i>CSS0039426</i> CSS</p> <p><i>CSS0039426</i> HapE with high CAF</p> <p><i>CSS0039426</i> HapF with low CAF</p> <p>Consensus</p>	<p>G A C T C G G C G A C C A A T C T C C G T T C A G G C T T C C T C T G C T A T G 189</p> <p>G A C T C G G C G A C C A A T C T C C G T T C A G G C T T C C T C T G C T A T G 189</p> <p>G A C T C G G C G A C C A A T C T C C G T T C A G G G T T C C T C T G C T A T G 189</p> <p>G A C T C G G C G A C C A A T C T C C G T T C A G G T T C C T C T G C T A T G 189</p>	<i>CSS0039426</i>

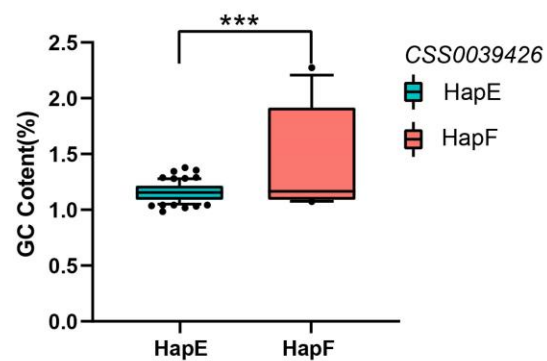
Figure S7. Box plot of GCG, GC and CAF content distribution about four pairs of haplotypes of *CSS0039426*, *CSS0044862*, *CSS0026218* and *CSS0027172*. The x-axis represents four pairs of haplotypes of *CSS0039426*, *CSS0044862*, *CSS0026218* and *CSS0027172* and the y-axis represents GCG, GC and CAF content. The below table is the detailed information of four pairs of haplotypes.



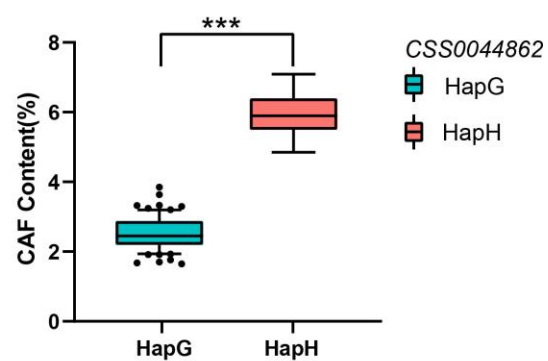
Haplotype	Position/bp	Chromosome	Mutation
HapA	164,656,928	9	A(Glu)
HapB			G(Gly)



Haplotype	Position/bp	Chromosome	Mutation
HapC	140,146,183	9	A(Thr)
HapD			T(Ser)



Haplotype	Position/bp	Chromosome	Mutation
HapE	141,418,417	9	A(Tyr)
HapF			G(Cys)



Haplotype	Position/bp	Chromosome	Mutation
HapG	144,566,106	9	C(Ala)
HapH			G(Gly)