



Genome-wide association studies reveal potential candidate genes associated with amino acid in tea plants

Qidi Wu^{1,2}, Li Song¹, Dingchen Bai³, Yihan Wang⁴, Yuting OuYang³, Kaixin Rao³, Qinfei Song³, Juanying Wang¹, Suzhen Niu^{1,2} and Yujie Ai³

¹ College of Life Sciences/Institute of Agro-Bioengineering, Guizhou University, Guiyang, Guizhou, China

² Key Laboratory of Plant Resource Conservation and Germplasm Innovation in Mountainous Region (Ministry of Education), Guizhou University, Guiyang, Guizhou, China

³ College of Tea Science, Guizhou University, Guiyang, Guizhou, China

⁴ Department of Tea Studies (Guizhou Duyun Maojian Tea College), Guizhou Vocational College of Economics and Business, Duyun, Guizhou, China

ABSTRACT

Background. Amino acids, as the main flavor substances of umami in tea, are also the primary components determining the taste of tea, which is positively correlated with the quality and grade of tea. The Guizhou Plateau is located in the core area of the origin of the tea plant and has abundant tea germplasm. However, there are relatively few studies using genome-wide association studies (GWAS) to mine genes related to amino acid content in tea plants in the Guizhou Plateau.

Results. In this study, 78,819 high-quality single nucleotide polymorphisms (SNPs) markers were identified from 212 tea accessions composed by our group in the previous study by genotyping sequencing technology (GBS), and the population structure, genetic diversity, and GWAS of 212 tea accessions resources of tea were analysed. Phylogenetic tree and population structure analysis divided all germplasm into four inferred groups (Q1, Q2, Q3, Q4). By analysing the eight SNPs associated with amino acids obtained by GWAS, four candidate genes that may be related to amino acids were identified. Reverse transcription quantitative polymerase chain reaction (RT-qPCR) was used to verify the expression levels of four candidate genes, suggesting that there may be a potential gene that is important for the accumulation of amino acid content.

Conclusions. This study provides new information for the in-depth analysis of the genetic mechanism of amino acid content in tea plants and provides important genetic resources for accelerating the cultivation of new tea varieties with suitable amino acid content.

Subjects Agricultural Science, Bioinformatics, Genetics, Plant Science

Keywords Tea plant, Amino acid, Genotyping-by-sequencing, Genetic diversity, Genome-wide association analysis

INTRODUCTION

The tea plant is a perennial, evergreen, woody plant, and tea beverages processed from the young leaves or shoots of tea plants have become one of the most popular beverages in the world after water (*Wu et al., 2020*). Tea contains 20 protein amino acids (PAAs) and six non-protein amino acids (PAAs), among which theanine (Thea), glutamic acid

Submitted 22 October 2024

Accepted 21 January 2025

Published 4 March 2025

Corresponding authors

Suzhen Niu, niusuzhen@163.com

Yujie Ai, cusboaiyujie@126.com

Academic editor

Diaa Abd El-Moneim

Additional Information and
Declarations can be found on
page 18

DOI 10.7717/peerj.18969

© Copyright

2025 Wu et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

(Glu), proline (Pro), arginine (Arg), and aspartic acid (Asp) are important amino acids that play an important role in tea quality (Guo et al., 2023; Zhang et al., 2024). For example, some amino acids exhibit certain good aromas while also giving tea characteristic umami profiles. The main component that produces the umami taste of tea is the content of theanine (Ruan, Haerdter & Gerendás, 2010). L-theanine imparts the tea its distinctive caramel aroma, while glutamic acid and phenylalanine add floral notes to the tea's aroma (Yu & Yang, 2020). In addition, amino acids are not only important nitrogen components in tea plants, but also the main form of nitrogen transport between different organs through xylem and phloem (Besnard et al., 2016). Therefore, mining genes related to amino acids in tea plants is of great significance for improving the quality of tea plants. The climate in Guizhou is very suitable for the growth of tea plants, which makes Guizhou rich in wild tea resources and local group varieties, which may contain a large number of resources with specific biochemical components (Niu et al., 2019). Although some studies have revealed the differences in amino acid content in Guizhou tea germplasm and the important factors regulating some amino acids (Yang et al., 2022b; Mi et al., 2023). However, the genetic structure and molecular basis of the natural variation of amino acid content in tea plants are still not clear. Guizhou is one of the origins of tea plants, and it is also the only province in China that has the characteristics of low latitude, high altitude, and less sunshine, which unique geographical location has created its superior natural environment (Lee, 2017). Previous studies have shown that within a certain altitude range, increases in altitude will enhance N metabolism in tea plants, which is conducive to the accumulation of amino acids and other compounds (Wang et al., 2022). At the same time, light intensity reduction is a common approach to improve the concentration of free amino acids in tea leaves (Chen et al., 2017). Therefore, the low latitude, high altitude, and low sunshine of Guizhou jointly affect the synthesis and accumulation of amino acids in tea plants by affecting the photosynthesis, temperature adaptability, and interaction with other environmental factors. Guizhou's unique environmental conditions provide the basis for the unique flavor and quality of Guizhou tea. But the mountains and rivers brought by the environment and the rugged roads have also brought obstacles to the local economic development. This also makes Guizhou still retain relatively complete tea germplasm. However, due to the fact that tea is a self-incompatible plant, the genome is highly heterozygous and difficult to assemble, and a large number of somatic mutations have accumulated due to long-term asexual reproduction (Liu et al., 2023b; Li et al., 2024). The protection and utilisation of tea germplasm resources in Guizhou Province is imminent.

Due to the lack of genetic resources for metabolic and genetic analysis, the current study mainly dissects the genetic mechanism of the target compounds in tea plants from the perspective of reverse genetics. In recent years, many genes involved in the amino acid pathway of tea plants have been revealed through reverse genetic studies, such as the CsPDX2.1 gene has the function of L-theanine hydrolase (Fu et al., 2020); CsMYB73 is a transcriptional repressor involved in L-theanine biosynthesis in tea plants (Wen et al., 2020), and CsLHT1 is able to transport a wide range of amino acids in tea plants (Li et al., 2021a). The above studies show that we can indirectly regulate amino acids through metabolic bypass pathways that affect amino acid synthesis. However, the potential

functional genes for the target traits affected by genetic and environmental factors are still largely unknown and unvalidated, which limits the prospects of tea breeding. Therefore, it is important to elucidate the genetic basis of amino acid pathways in tea plants and provide a useful resource for improving tea quality (Huang *et al.*, 2022b).

Forward genetic methods such as quantitative trait locus (QTL) mapping and association mapping are important tools for identifying regulatory sites of free amino acid content in higher plants. Zhong *et al.* (2011) performed phenotypic determination of protein and amino acid content in recombinant inbred line populations and identified 112 rice QTLs by QTL mapping. Xu *et al.* (2015) identified 17 QTLs associated with lysine, threonine, and methionine content in rapeseed meal through QTL mapping. In tea plants, QTL mapping has been used to identify important qualities and agronomic traits. Ma *et al.* (2018) identified 10 genetic variants related to theobromine and caffeine content and proportion in tea plants through QTL mapping. Huang *et al.* (2022b) identified four QTLs associated with the content of free amino acids (theanine, glutamic acid, glutamine, aspartic acid, and arginine) in tea plants through QTL mapping. Liu *et al.* (2023a) combined QTL mapping, GWAS, and transcriptomic analysis to identify a candidate gene associated with the timing of spring bud flush in tea plants. However, accurate QTL mapping based on linkage analysis first requires the construction of a specific large mapping population, which is a very time-consuming process for tea plants with long juvenile stages, and because of the use of parental hybridisation, the allele variation captured in QTL mapping becomes limited, so that the natural variation of natural populations cannot be fully understood (Wang *et al.*, 2019; Zhang *et al.*, 2021). Different from QTL mapping, genome-wide association studies (GWAS) have the advantages of being rapid, inexpensive, and high-resolution and have become a powerful tool for identifying multi-related candidate genes that regulate major crop traits. Based on the principle of linkage disequilibrium (LD), GWAS analyses the relationship between genetic variation and population sample characteristics and excavates genetic loci related to target traits, so as to provide a basis for the improvement of crop target traits and a new research path and ideas for molecular breeding. At present, GWAS has been widely used in the study of metabolite traits in many model plants and widely cultivated crops, such as Arabidopsis (Filiault & Maloof, 2012), maize (Wen *et al.*, 2014), and rice (Sun *et al.*, 2020). In recent years, many researchers have used GWAS to identify candidate genes associated with important agronomic traits in tea plants. Wang *et al.* (2019) used GWAS to identify favourable single nucleotide polymorphisms (SNPs) alleles as well as candidate genes controlling *C. sinensis* TBF. Chen *et al.* (2023) obtained SNP markers and candidate genes significantly associated with four leaf traits using GWAS analysis. Based on GWAS, Wang *et al.* (2024) identified four potential candidate genes associated with OTL and OtL in tea plants.

With the rapid development of next-generation sequencing (NGS) technology, whole-genome sequencing (WGS) data based on high-throughput genotyping and millions of SNPs have been widely used in various genetic studies (Chung *et al.*, 2017; Gaudin & Desnues, 2018; Zhao, Wang & Yuan, 2020). WGS is the most straightforward strategy to achieve GWAS, but it is still an expensive means for most researchers. Genotyping by sequencing (GBS) has distinct advantages over WGS, such as being fast, simple, and

low-cost. It uses restriction enzymes (RE) to digest DNA, perform high-throughput sequencing of sequences at both ends of the digestion fragment, and genotype the SNP information obtained by analysis (He *et al.*, 2014). GBS technology has been widely used in the analysis of genetic diversity, genetic mapping, agronomic trait localisation, and related gene mapping of wheat (Alipour *et al.*, 2017), soybean (Sonah *et al.*, 2015), apple (Lee *et al.*, 2017), coffee (Hamon *et al.*, 2017), tea plant (Huang *et al.*, 2022b; Huang *et al.*, 2022a), and other crops.

In this study, 212 tea accessions from the Guizhou Plateau were used as the research objects, and their population structure, linkage disequilibrium, and genetic diversity were analyzed. At the same time, we used the phenotype of the amino acid trait and the genotype of 212 tea accessions to determine the loci related to the amino acid content by GWAS technology. The study found that there may be a candidate gene associated with amino acid content. The results of this study provide useful information for the future study of the key genes of the amino acid metabolism pathway in tea plants and are of great significance for the use of molecular breeding technology to improve the amino acid traits of tea plants.

MATERIALS & METHODS

Plant materials and sample preparation

The 212 tea plant materials in this study were derived from the core germplasm population constructed by the research group using 415 tea germplasm resources in the early stage (Table S1) (Niu *et al.*, 2019). Among them, 210 materials were from Guizhou Province, and the remaining two materials were from Fujian Province and Zhejiang Province, respectively. The experimental materials were all planted in the teaching experimental base of Guizhou University (106°40'E and 26°20'N). Each plant material is vegetatively propagated and grown in a randomised complete block design with three replicates per plant set to ensure consistency and reliability. Protective rows were set around the experimental field, and the field management of soil fertility and irrigation in the experimental field was basically the same.

In the spring of 2020, samples were taken at the teaching experimental base of Guizhou University, and the sampling standard was one bud and two leaves. Fresh leaves harvested from each accession are immediately frozen in liquid nitrogen, freeze-dried, and stored in an ultra-low temperature freezer (−80 °C) for later use.

Phenotypic data analysis

In this study, an ultraviolet spectrophotometer was used to detect the amino acids in the samples according to the national standard method for the determination of the total amount of free amino acids in tea (GB/T8314-2013). The detection procedure is as follows: Accurately weigh 3 grams of ground dry sample, add 450 ml of boiling water, and immediately move it into a boiling water bath for 45 min, shaking every 10 min. It is then filtered under reduced pressure while hot, and the filtrate is placed in a 500 ml volumetric flask. The above extraction steps are repeated 2–3 times, and the filtrate is combined. After the filtrate is cooled, distilled water is added to set the volume to 500 ml for later use. Aspirate 1 ml of sample solution, inject it into a 25 ml volumetric flask, add 0.5 ml of

phosphate buffer with pH of 8 and 0.5 ml of 2% ninhydrin solution, heat it in a boiling water bath for 15 min, cool it to room temperature, and then add water to set the volume to 25 ml. After 10 min, the absorbance was measured at a wavelength of 570 nm with a one cm cuvette with distilled water as a blank. The amino acid content is obtained according to the calculation formula in the national standard method.

Genotype analysis

The sequencing data obtained in the previous research of our group were used (Niu *et al.*, 2019). The raw DNA reads were demultiplexed using a barcode, and the adapter was trimmed using a custom Perl script. Only the reads with quality values > 5 are kept as clean data and then aligned to the reference genome (<http://tpia.teaplants.cn/>) using BWA-MEM (v. 0.7.10) with default parameters (Xia *et al.*, 2020a; Jung & Han, 2022). GATK (v. 3.7.0) was used to call for SNPs (McKenna *et al.*, 2010). SNP filtering refers to previous research methods (Danecek *et al.*, 2011; Niu *et al.*, 2019). The specific filtering criteria are summarised as follows: (1) Biallelic SNPs must be selected. (2) “QUAL < 50.0 || QD < 2.0 || FS > 60.0 || MQ < 40.0 || Mapping Quality Rank Sum < -12.5 || Read Pos Rank Sum < -8.0” was used in variant filtration in GATK (v 3.7.0) to filter the SNPs. (3) VCFtools (v. 0.1.15) was used to filter SNPs with minor allele frequencies (MAF) < 0.05 or missing data rates > 20%. After filtering, 78,819 SNPs were retained from 212 tea accessions for further analysis.

Population structure and linkage disequilibrium

The .ped and .map files of the SNP genotype data were converted into .bed files using Plink software and used as input files. Using ADMIXTURE (v. 1.3.0) software for population structure analysis (Alexander, Novembre & Lange, 2009). The maximum likelihood estimation clustering method based on the Bayesian model is also applied. The number of clusters (K -values) is assessed by determining the cross-validation error (CV error), which is tested from two to nine and runs five iterations for each K -value, and the optimal K value was confirmed according to the minimum CV error estimated by ADMIXTURE (Liu *et al.*, 2020). We set the threshold to 0.8 to distinguish between pure and mixed groups. Then the principal component analysis was carried out using TASSEL (v. 5.2.72) software, and 3D scatter plots were drawn using R (v. 4.4.1) software; each individual and population was color-labeled, and the clustering relationship between them was observed (Glaubitz *et al.*, 2014). Then using MEGA (v. 10.2.4) software for Neighbor-Joining Tree mapping (Stecher, Tamura & Kumar, 2020). Finally, based on the allele frequency correlation (r^2), PopLDdecay (v. 3.30) calculates and plots the squared correlation coefficient (r^2) between genome-wide unpruned SNPs (Zhang *et al.*, 2019). The window size for LD decay analysis is 300 kb. The r^2 threshold was 0.21.

Genetic diversity analysis

Genetic diversity indicators included observed heterozygosity (H_o), expected heterozygosity (H_e), minor allele frequency (MAF), inbreeding coefficient (F_{is}), nucleotide diversity (P_i), Tajima's D , and genetic differentiation coefficient (F_{st}). Using Plink (v. 1.9), calculate the MAF, H_o , H_e , and F_{is} values for each inferred population (Chang *et al.*, 2015). Using

VCFtools (v. 0.1.15) to calculate the P_i and Tajima's D for each inferred population and the F_{st} value for the two pairs of the inferred population (Danecek et al., 2011). SPSS (v. 26, Armonk, NY, USA) statistical software counts significant differences between indicators. The larger the F_{is} , the closer the individuals in the group. When H_o is greater than H_e , the population follows the Hardy–Weinberg equilibrium. Failure to do so may indicate some form of selection, drift, migration, or non-random pairing in the population, leading to deviations in genotype frequency. The larger the P_i , the more polymorphic this locus is in the population, with a large amount of variation. The significance level of Tajima's D can tell whether a locus has undergone neutral evolution. The smaller the MAF , the worse the polymorphism at the SNP locus. The larger the F_{st} value, the greater the genetic differentiation between subgroups (Danecek et al., 2011).

Genome-wide association study

In order to ensure the accuracy of the results, SNPs with a minor allele frequency of less than 0.05 or a maximum deletion genotype frequency greater than 20% were screened for GWAS. Based on the 78,819 high-quality SNPs and 1 phenotypic trait obtained after GBS sequencing and filtering, six GWAS linear analysis models were performed by TASSEL (v. 5.2.72) software, which were (1) GLMQ = SNP + Q (2) GLMP = SNP + P (3) MLMQ + K = SNP + Q + K (4) MLMP + K = SNP + P + K (5) CMLMQ + K (6) CMLMP + K. These six models were compared to find the optimal model among the amino acid quality traits analysed (Bradbury et al., 2007). Among them, the general linear model (GLM) uses the population structure (Q) or principal component analysis (PCA) matrix as a fixed effect to control for possible false positives of traits. The mixed linear model (MLM) and the compressed mixed linear model (CMLM) use the kinship matrix (K) as a covariate to reduce the inequality correlation between genotypes (Yu et al., 2017). Comparing the Q–Q plots of the output of the six GWAS models, the best model fitting the curve of the expected value to the observed value was entered into the later analysis as the optimal model for the trait. Manhattan plots showed the correlation between SNP sites on each chromosome and this trait. $-\text{Log}_{10}(P) \geq 4.0$ was selected as the threshold to identify SNP sites closely associated with traits (Huang et al., 2021).

Identification and annotation of candidate genes

The best model for the trait was selected to screen for significant SNP loci and to search for candidate genes within 5 kb upstream and downstream of significant SNP loci. SNP marker sequences significantly associated with amino acids were used as probes, and BLAST alignment was performed in the NCBI (<https://www.ncbi.nlm.nih.gov/>) and TPIA (<http://tpia.teaplants.cn/>) databases to obtain candidate genes related to amino acids, and the candidate genes were functionally annotated (Zou et al., 2013).

Gene expression analyses using RT-qPCR

The nine tea materials (including three with high amino acid content, three with moderate amino acid content, and three with low amino acid content) were selected from 212 tea accessions for RT-qPCR. Firstly, the total RNA of nine samples of tea tree materials was extracted using the RNAPrep Pure Polysaccharide and Polyphenol Plant Total RNA

Extraction Kit of Beijing Tiangen Biochemical Technology Co., Ltd. Then, using RNA as a template, cDNA was synthesized using Hiscript reverse transcriptase (Novozan Biology, San Diego, CA, USA) under the following conditions: 42 °C water bath for 2 min, 37 °C for 15 min, and 85 °C for 5 s. Finally, the ChamQ Universal SYBR qPCR Master Mix Kit (Novozan Biology, San Diego, CA, USA) was used for RT-qPCR, that is, the expression levels of selected candidate genes in tea cultivars with different amino acid contents were detected using cDNA as a template. The specific reaction conditions were as follows: Pre-denaturing at 95 °C for 30 s; denaturation at 95 °C for 10 s, annealing at 60 °C for 30 s, 40 cycles. GAPDH (Forward primer: AGCTGCACAACCAACTGTTTG, Reverse primer: AGCTGCACAACCAACTGTTTG) was used as the reference gene for relative quantitative analysis. Finally, the results were analysed by the $2^{(-\Delta\Delta Ct)}$ method (Livak & Schmittgen, 2001). A total of three biological replicates and three technical replicates were performed. Statistical analysis was performed by one-way ANOVA.

Statistical analyses

Origin2021 software was used to create box plots of amino acid content to identify and exclude outliers. IBM SPSS Statistics (v. 26, Armonk, NY, USA) was used to perform descriptive statistical analyses of phenotypic traits, including range, mean, standard deviation (SD), and coefficient of variation (CV) for traits.

RESULTS

SNP marker mining and population structure

A total of 29,393,327 SNPs were obtained from 212 tea accessions, and after strict filtering conditions, 78,819 high-quality SNPs with a minor allele frequency (MAF) of >0.05 and deletion rate <20% were retained for subsequent analysis. In addition, we further investigated the distribution of high-quality SNPs. The results showed that all high-quality SNPs were randomly distributed on 15 chromosomes, with an average number of 4,811 SNPs per chromosome. Chromosome 1 contains the largest number of high-quality SNPs (7,596 SNPs). Chromosome 15 contains the least number of high-quality SNPs (3,273 SNPs) (Fig. S1 and Table S2). The lowest (24 SNPs/Mb) and highest (34 SNPs/Mb) SNP densities were detected on chromosomes 5 and 1, respectively (Table S3).

In order to further explore the genetic relationship of 212 tea accessions, a total of 78,819 high-quality SNPs were used for population structure analysis, principal component analysis, and phylogenetic analysis. The result revealed that when K increased sequentially from two to nine, the CV errors were found to increase gradually (Fig. 1A). The CV error values are 0.57840 and 0.57885 at $K = 2$ and $K = 3$, respectively (Fig. 1A). There is no significant difference between the two CV errors. In the process of the evolution of tea plants, due to long-term natural selection and artificial domestication, ancient landraces and modern landraces have gradually formed. In this study, it was observed that when $K = 2$, 212 tea accessions could be divided into ancient landraces group, wild tree group, and admixture group, while when $K = 3$, 212 tea accessions could be divided into ancient landraces group, modern landraces group, wild tree group, and admixture group (Tables S4 and S5). The above results showed that when $K = 3$, the landrace group

could be better subdivided into ancient and modern landrace groups. In order to perform more accurate grouping, we consider it possible to divide 212 tea accessions at $k = 3$. The genetic component (Q value) ≥ 0.8 indicated that the genetic background of the material was relatively single, and the genetic progenitor (Q value) < 0.8 meant that the genetic background was relatively complex and belonged to the mixed component. In this study, the population was divided into three pure groups and one mixed group (Fig. 1B and Fig. S2). The first pure group (referred to as the ‘Ancient landraces group’ or ‘Q1’ from now on) contained 77 tea accessions, including 74 *C. sinensis* (85.14% of which are ancient landraces) and three *C. remotiserrata*. The second pure group (referred to as the ‘Modern landraces group’ or ‘Q2’ from now on) consisted of 18 *C. sinensis* (66.67% are modern landraces), four *C. tachangensis*, and one *C. remotiserrata*. The third pure group (referred to as the ‘Wild tree group’ or ‘Q3’ from now on) contained seven *C. tachangensis* and two *C. remotiserrata*. The fourth admixture group (Q4) contained 103 tea accessions, including 46 *C. sinensis* and 32 *C. remotiserrata*, 22 wild *C. tachangensis*, and three near *C. taliensis* (Table S5). PCA was performed by the first three principal components, indicating that the tested accessions separated into four subgroups (Fig. 1C). Although each subgroup has a small number of joins at the intersection position. The phylogenetic tree divides the population into four subgroups by the same SNP set (Fig. 1D), which is consistent with the results of population structure and PCA analysis.

Genetic diversity and differentiation

Based on obtaining the high-quality SNPs, we estimated the parameters of genetic diversity (P_i , H_o , H_e , MAF , and Fis) of 212 tea accessions and compared the genetic differences among four inferred populations. The results showed that the P_i , H_o , H_e , MAF , and Fis of 212 tea accessions were 0.2166, 14.1760, 6.2592, 0.1370, and 0.6646, respectively (Fig. 2A). Comparing the correlation coefficients among the four inferred populations, the results showed that the Q2 subgroup had the highest H_o and H_e among all populations, and the H_o of all inferred populations was greater than H_e , indicating that all populations had rich population genetic diversity (Zhang et al., 2018; Schmidt et al., 2021). The Q4 subgroup had the highest levels of P_i and MAF among all groups. The Q1 subgroup had the lowest H_o and H_e among all groups. The Q2 subgroup had the lowest MAF and Fis among all groups. The Q3 subgroup had the lowest P_i among all groups but the highest Fis , indicating that the individuals in this group were closely related, and the germplasm of the population was wild-type, which also confirmed the accuracy of the above results. Previous studies showed that Tajima’s $D > 0$ indicates the presence of a large number of medium-frequency alleles; this may be due to population bottleneck effects or balanced selection. MAF is commonly used in genetics to distinguish common polymorphisms in populations from rare variants (Nebert & Vesell, 2013). The MAF of the four tea populations was greater than 0.01 and less than 0.2, indicating that the four tea groups may be in the range of moderate frequency variation. At the same time, the Tajima’s D values for all four tea populations were positive, indicating that they had all experienced population bottlenecks or equilibrium selection (Fig. 2A) (Tandoh et al., 2021; Pandey et al., 2021). Among the four subgroups, the genetic differentiation coefficients (Fst) analysis showed that the Q1 and Q3 subgroups had the

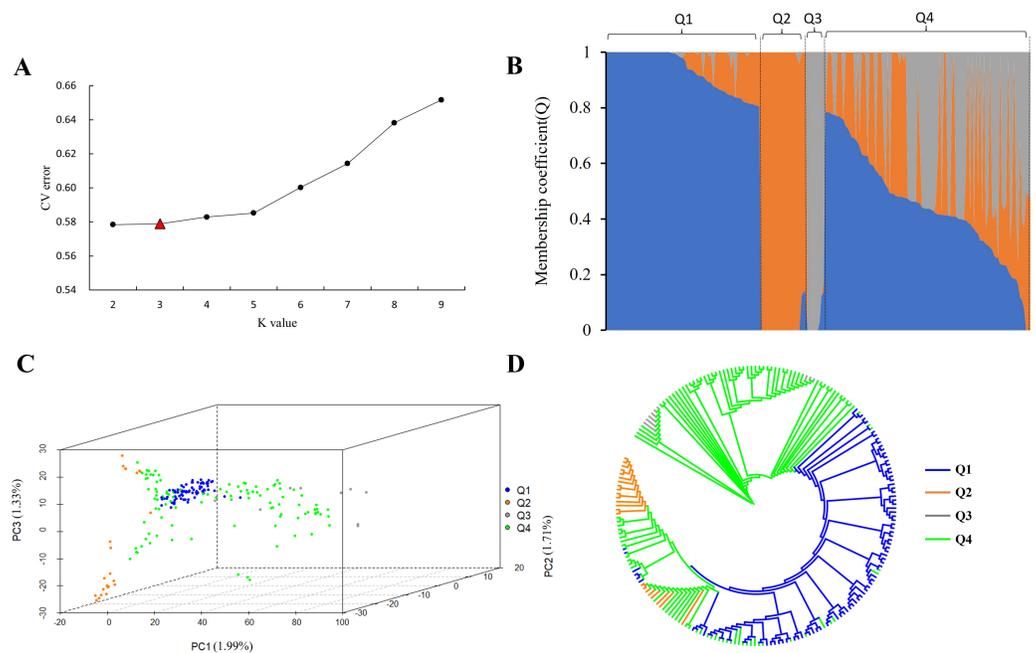


Figure 1 SNP genotyping data were used to analyze the population structure of 212 tea accessions. (A) The CV error is used to estimate the tea population, and the K range is 2~9. (B) Population structure based on $k = 2-9$. Q1, Q2, Q3, and Q4 represent subgroup I, subgroup II, subgroup III and subgroup IV, respectively, and blue, orange, and gray represent groups that may have come from three different ancestral groups, respectively. (C) PCA of 212 tea accessions. The first, second, and third coordinates indicate PC1, PC2, and PC3. Each dot represents an individual. Blue, orange, gray, and green, represent subgroup I (Q1), subgroup II (Q2), subgroup III (Q3), and subgroup IV (Q4), respectively; (D) Phylogenetic trees of 212 tea accessions (blue, orange, gray, and green represent subgroup I (Q1), subgroup II (Q2), subgroup III (Q3), and subgroup IV (Q4), respectively).

Full-size [DOI: 10.7717/peerj.18969/fig-1](https://doi.org/10.7717/peerj.18969/fig-1)

highest degree of differentiation, with the F_{st} value of 0.2021. The Q2 and Q4 subgroups were the least differentiated, with the F_{st} value of 0.0369 (Fig. 2A).

Linkage disequilibrium analyses

Linkage disequilibrium (LD) decay is caused by linkage imbalance; LD decay rate varies in different species or different subspecies. LD decay is usually used to determine the general distance as the attenuation distance of the population; if the LD decay is fast, more loci are needed to achieve a certain accuracy when performing GWAS analysis (Sallam & Martsch, 2015). Based on the high-quality SNPs of the tea genome, we used Pop LDdecay to calculate the LD, and the resulting linkage imbalance attenuation plot showed that the LD decreased significantly with the increase of physical distance (Fig. 2B), but the average distance when the LD dropped to half of the maximum value was very short, approximately 5 kb, which indicated that the genetic diversity of the tea population was high. Therefore, we consider that 5 kb can be used as a search range of candidate genes.

Phenotypic variation analysis of amino acid

Through the statistical analysis of the phenotypic data of 212 tea accessions in Guizhou, the results showed that the amino acid content was between 1.31% and 5.53%, the mean value

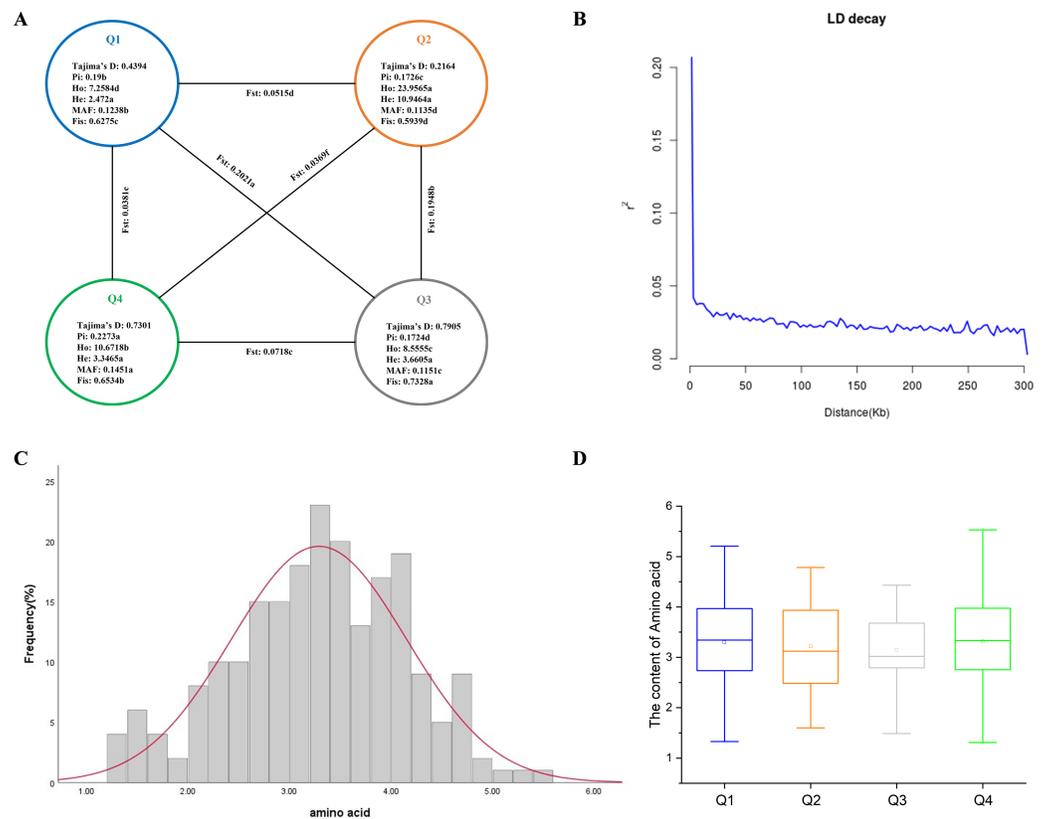


Figure 2 Genetic diversity, LD decay analysis, and amino acid phenotype analysis of 212 tea germplasm accessions. (A) Genetic diversity of four inferred populations of 212 tea accessions. *Pi* nucleotide diversity, *Ho* observed heterozygosity, *He* expected heterozygosity, MAF minor allele frequency, *Fis* inbreeding coefficient, *Fst* differentiation coefficient. The different letters indicate a significant difference in $p = 0.05$ levels by the *T*-test. (B) LD attenuation plots of 212 tea accessions. (C) Frequency distribution of amino acids in environment. (D) Box plot of amino acid contents in four inferred groups.

Full-size DOI: [10.7717/peerj.18969/fig-2](https://doi.org/10.7717/peerj.18969/fig-2)

was 3.29%, the standard deviation was 0.86, and the coefficient of variation was 26.24%. The above results showed that there was a wide variation of amino acids in the 212 tea accessions (Table 1 and Table S6). When the skewness *Z*-score and kurtosis *Z*-score values were within the range of ± 1.96 , the data in this group followed a normal distribution (Orcan, 2020). The skewness *Z*-score and kurtosis *Z*-score of amino acid content varied between ± 1.96 , indicating that the amino acid content was in a normal distribution (Fig. 2C), indicating that the amino acid content of 212 tea accessions was eligible for GWAS.

Amino acid content was compared between the four inferred groups. The results showed that Q4 had the highest amino acid content among the four inferred groups, followed by Q1, Q2, and Q3 (Fig. 2D). Besides, the average of amino acid content of Q1 and Q4 is comparable, but Q4 has a wider range of amino acid content.

Table 1 Phenotypic statistical analysis of amino acids content.

Group	Mean (%)	Range (%)	SD	CV (%)	Skewness	Kurtosis
Q1	3.30	1.33–5.21	0.86	26.06	−0.219	−0.381
Q2	3.21	1.60–4.79	0.89	27.73	0	−0.648
Q3	3.15	1.49–4.43	0.91	28.89	−0.424	−0.066
Q4	3.31	1.31–5.53	0.87	26.28	−0.256	−0.059
All	3.29	1.31–5.53	0.86	26.14	−0.216	−0.307

Genome-wide association analysis

Six linear regression models (GLM-P, GLM-Q, MLM-P+K, MLM-Q+K, CMLM-P+K, and CMLM-Q+K) were used to evaluate the correlation between amino acid phenotypes and genotypes. QQ plots were used to assess the extent of accordance between the observed and expected P -values. Among the six models, the GLM-P was found to have the best curve fit between the expected value and the measured value, which shows that the model can be selected for subsequent analysis (Fig. 3).

For the six models, we chose $-\log_{10}(P) \geq 4.0$ as the threshold for screening important loci (Fig. 4) (Huang et al., 2021). Eight SNPs were found to be associated with amino acids by using the GLM-P model; the contribution rate (R^2) of SNP phenotypic variation was 8.82% to 11.36%. The GLM-Q analysis found 15 SNPs associating with amino acids, the R^2 value ranging from 9.36% to 14.15%. Three SNPs were identified as associating with amino acids by using the MLM-P+K model, with the R^2 value ranging from 9.71% to 11.99%. The MLM-Q+K analysis found four SNPs associated with amino acids, with the R^2 value ranging from 9.83% to 12.73%. The CMLM-P+K and CMLM-Q+K analyses found one SNP associated with amino acids, respectively. The contribution rate (R^2) of two SNPs phenotypic variations was 11.85% and 11.31% (Table 2 and Table S7).

Candidate gene screening and gene functional annotation

The genes detected in each significant locus, as well as the genes located upstream or downstream of the loci, were designated candidate genes, which were functionally annotated (Fang et al., 2021). Based on the LD results, 5 kb was selected as a reasonable range for searching genes. A total of four candidate genes were detected in GLM-P; all genes were subsequently functionally annotated using Blast in the TPIA (<http://tpia.teaplants.cn/index.html>). The *TEA022785.1* gene may have ribokinase (RBSK) activity and is involved in the synthesis of D-ribose-5-phosphate, which in turn produces ribose pyrophosphate (PRPP), which can subsequently be involved in the synthesis of histidine and tryptophan. The *TEA027937.1* gene may have L-tryptophan-pyruvate aminotransferase 1-like activity, which can catalyse the conversion of L-tryptophan and pyruvate to indole-3-pyruvate (IPA) and alanine, resulting in alanine accumulation. The *TEA013853.1* gene may have a DNA methylation 3-like factor, which may contain an XS domain, which contains a conserved aspartate residue, which may be functionally important. It may also contain an XH domain, which contains a conserved glutamate residue that may have an important function. And these two domains may interact with each other. The *TEA008491.1* gene may

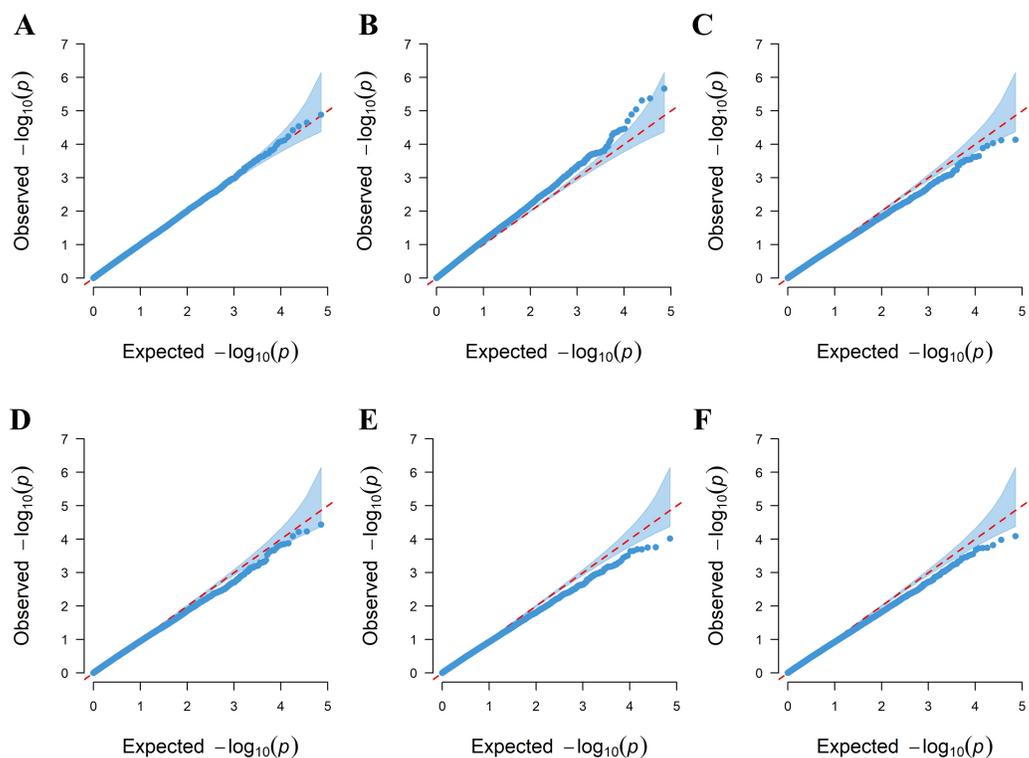


Figure 3 QQ plots for the amino acid traits of tea plant by six models. (A) GLM-P; (B) GLM-Q; (C) MLM-P+K; (D) MLM-Q+K; (E) CMLM-P+K; (F) CMLM-Q+K.

Full-size [DOI: 10.7717/peerj.18969/fig-3](https://doi.org/10.7717/peerj.18969/fig-3)

have xylan α -glucuronosyltransferase 1-like activity and may be involved in the substitution of the xylan backbone in stem glucuronide ketoglycan (Table 3).

Candidate gene validation

In order to verify whether the potential candidate genes of amino acids are involved in the accumulation of amino acid content in tea leaves, the expression levels of *TEA022785.1*, *TEA027937.1*, *TEA013853.1*, and *TEA008491.1* genes were determined by using the tea leaves of tea accessions with high amino acid content, moderate amino acid content, and low amino acid content by RT-qPCR. The results showed that the expression of the *TEA027937.1* gene was positively correlated with the amino acid content of the tea plant. The expression of the remaining *TEA022785.1*, *TEA013853.1*, and *TEA008491.1* genes was irregular (Fig. 5, Fig. S3, Tables S8 and S9). Therefore, the above *TEA027937.1* gene may be involved in regulating the accumulation of amino acid content in tea.

DISCUSSION

Genetic variation in tea plant populations

Tea plants originated in southwestern China and later spread around the world (Yamanishi, 1995). Natural selection and artificial domestication have yielded a wide variety of tea germplasm, including cultivars, landraces, and wild types (Fang et al., 2021). Population

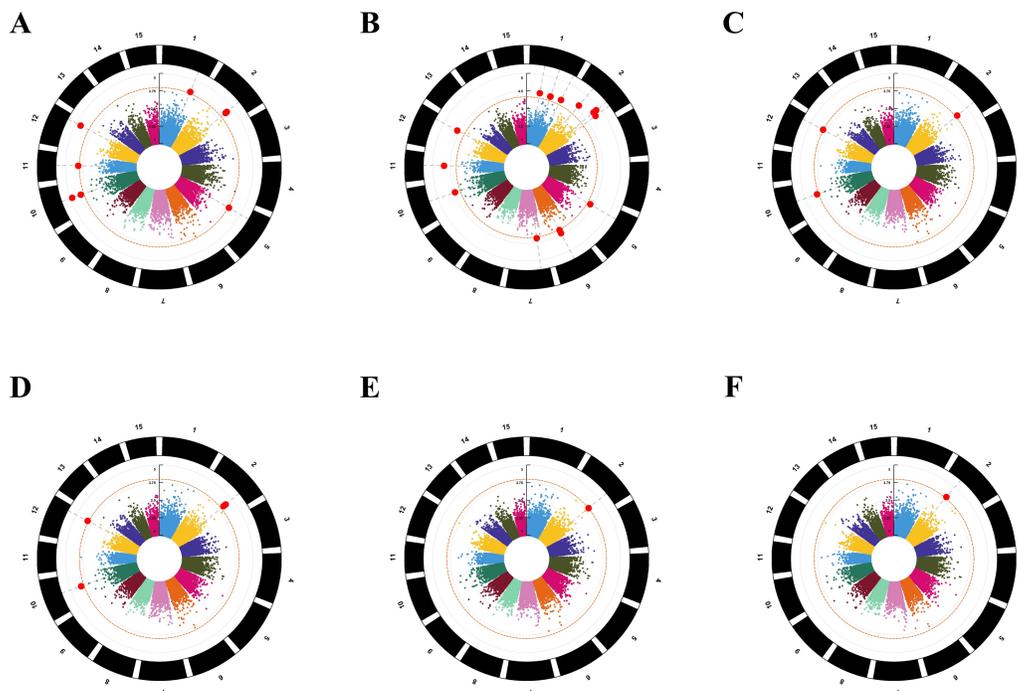


Figure 4 Manhattan plots for the amino acid traits of tea plant by six models. The orange dashed horizontal line indicates the significance threshold; the red solid line horizontal line indicates the extreme significance threshold. (A) GLM-P; (B) GLM-Q; (C) MLM-P+K; (D) MLM-Q+K; (E) CMLM-P+K; (F) CMLM-Q+K.

Full-size DOI: 10.7717/peerj.18969/fig-4

Table 2 The optimal model of amino acid of 212 tea accessions was used for SNPs analysis.

Traits	Model	SNPs	Chromosome	Position	$-\log_{10}p$	$R^2(\%)$	Allele
Aa	GLM(P)	S1_169435208	1	169435208	4.117	10.45%	G/A
		S2_151915980	2	151915980	4.534	11.36%	G/A
		S2_151915996	2	151915996	4.417	11.31%	T/C
		S5_46409247	5	46409247	4.057	8.82%	T/C
		S10_90937109	10	90937109	4.233	9.41%	G/C
		S10_90937133	10	90937133	4.880	10.49%	C/T
		S11_64920241	11	64920241	4.089	9.65%	C/T
		S12_134931733	12	134931733	4.647	10.49%	A/G

structure and genetic diversity analysis are of great significance for understanding the origin, evolution, and utilisation of crop germplasm. Meanwhile, it is also the basis for improving crop traits. Understanding the current level of crop diversity was one of the key factors for making the most effective use of crop germplasm for genetic improvement (Lavanya, Srivastava & Ranade, 2008; Xia et al., 2020a). Zhao et al. (2022) used 112,072 high-quality genotyping-by-sequencing to analyse the genetic diversity, principal components, phylogeny, population structure, and linkage disequilibrium and develop a core collection of 253 cultivated-type tea plant accessions from the Guizhou

Table 3 Functional annotation of four candidate genes related to the amino acid levels.

Traits	SNP locus	Chromosome	Candidate gene	Function annotation
Aa	S5_46409247	5	<i>TEA022785.1</i>	Ribokinase: involved in the pentose phosphate pathway
	S11_64916608	11	<i>TEA027937.1</i>	L-tryptophan–pyruvate aminotransferase: participate in IAA biosynthesis
	S11_64920241	11	<i>TEA013853.1</i>	DNA methylation factors: Methyltransferase activity
	S12_134931733	12	<i>TEA008491.1</i>	Xylan alpha-glucuronosyltransferase: may be involved in the substitutions of the xylan backbone in stem glucuronoxylan

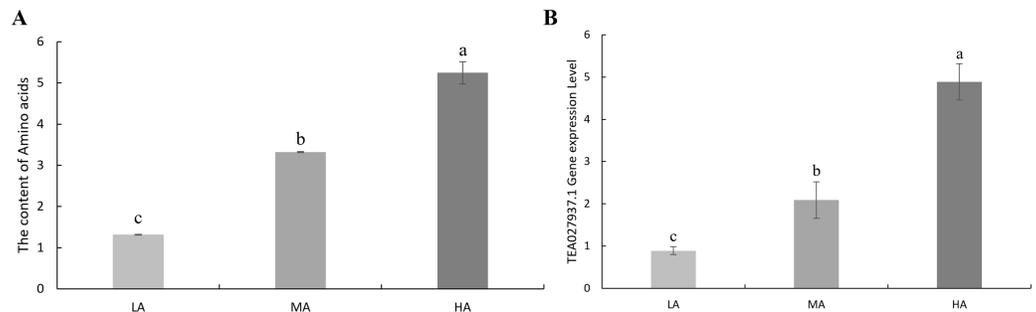


Figure 5 Amino acid content and expression levels of candidate genes. (A) The content of amino acids in tea plants. (B) The expression level of *TEA027937.1* gene related to amino acids in tea plants. HA, leaves with high amino acid content; MA, leaves with moderate amino acid content; LA, Leaves with low amino acid content.

Full-size DOI: 10.7717/peerj.18969/fig-5

Plateau. *He et al. (2023)* analysed the genetic diversity, population structure, principal components, phylogeny, and linkage disequilibrium of 159 wild tea trees from different altitudes and different geological types in the Guizhou Plateau, and the results revealed the genetic diversity and geographical distribution characteristics of wild tea trees in the Guizhou Plateau. In order to understand the population structure and genetic relationship of 212 tea accessions, the population structure, the phylogenetic tree, and the genetic differentiation analysis were carried out. In this study, 212 tea accessions were divided into four subgroups. The phylogenetic tree also divides into four categories. Q1 and Q2 were mainly cultivars; Q3 and Q4 were mainly wild-type. Previous studies showed that after the domestication of species, due to factors such as geography and human selection, reproductive isolation occurred, and different populations, such as cultivated species and wild species, were gradually formed. However, some degree of gene flow still occurs between geographically close populations (*Wang et al., 2017*). Some studies have shown that gene flow and reproductive isolation may jointly affect the structure and traits of tea populations. Tea populations in different regions communicate through gene flow, while unique tea types are formed in a given region due to reproductive isolation (*Seehausen et al., 2014; Rabosky, 2016; Xia et al., 2020b*). For example, the tea tree was first discovered in China, but it is now an important evergreen crop grown in more than 50 countries around the world. Tea types with unique local characteristics have also been developed in different regions of China, such as West Lake Longjing, Yunnan Pu'er, and Wuyi Rock Tea. This

dynamic influence not only promotes the diversification of tea germplasm but also has an impact on the quality and adaptability of tea.

Genome-wide association study of amino acids

Plant material in association studies should have a wide genetic diversity that maximises the capture of historical recombination events. In this study, the amino acid traits of 212 tea accessions were investigated, and the results showed that the amino acid traits of 212 tea plants had a wide range of variation. In addition, 78,819 high-quality SNPs from 212 tea accessions were retained after filtration, which was comparable to the number of 79,016 high-quality SNPs obtained from 415 Guizhou tea accessions reported previously, indicating that the 212 tea accessions from Guizhou had rich genetic diversity and could be used for GWAS ([Niu et al., 2019](#)). GWAS has emerged as a powerful approach for the identification of genetic regions underlying complex traits in crop plants ([Wu et al., 2017](#)). In sweet potato, a total of seven genes involved in sugar and starch metabolism in sweet potato were identified through genome-wide association ([Nie et al., 2023](#)). [Yang et al. \(2022a\)](#) presented a genome-wide association study revealing candidate genes related to stem diameter of cucumber (*Cucumis Sativus* Louis). [Li et al. \(2021b\)](#) revealed Four quantitative trait loci for germination rate were identified using a genome-wide association study during early germination in rice. Although GWAS has been extensively studied in major food crops, there are fewer studies on the use of GBS for GWAS for amino acid traits in tea compared to other crops. But the above research provided a good reference for the quality breeding of horticultural crops such as tea plants. As one of the birthplaces of tea, Guizhou is rich in wild tea germplasm resources. Tea is the most drunk beverage in the world and is processed from young leaves. As one of the main quality components of tea, amino acids play an important role in the flavor of tea. Mining candidate genes associated with amino acids can improve the abundance of key compounds associated with the flavor of tea, thereby improving the health benefits of tea plants. In this study, we employed genotyping sequencing-based GWAS to develop amino acid-related SNP loci and identify potential candidate genes. (GLM) (GLM-P and GLM-Q), (MLM) (MLM-P+K and MLM-Q+K), and (cMLM) (cMLM-P+K and cMLM-Q+K) were the six most common statistical models in GWAS. The Quantile–Quantile plots (Q–Q plots) represented the degree to which the model fits the data and were used to judge the deviation between the observed and expected values. The Manhattan diagram showed the correlation between the SNPs locus on each chromosome and the trait. We identified 17 amino acid-related SNPs under six GWAS models, of which eight ultimately significant SNPs were identified in one GWAS model determined by the best QQ plot. GLM is a general linear model, which has a wider range of applications and can screen more SNPs related to traits, but the accuracy is lower than that of MLM. MLM is a hybrid linear model. Compared with GLM, the model needs to add a genealogy matrix, *i.e.*, the kinship K matrix. On the basis of MLM, MLM is further optimised to obtain CMLM, which is called the compressed hybrid linear model. Since MLM correction is too strict and filters out some of the SNP markers that are really relevant, CMLM aims to redetect those false-negative SNP markers. It is important to note that all three models can add principal components (PCA) and population structure

(Q) to control for possible false positives for traits. In this study, GLM-P was selected according to the best QQ plot. Eight significant SNPs were associated with the GLM-P. Among the 8 significant SNPs, in addition to GLM-P, S2_151915980 was also found in GLM-Q, MLM-P+K, MLM-Q+K, and CMLM-P+K. S2_151915996 can also be found in GLM-Q and MLM-Q+K; S5_46409247 can also be found in GLM-Q; S10_90937133 can also be found in GLM-Q, MLM-P+K, and MLM-Q+K; S11_64920241 can also be found in GLM-Q; S12_134931733 can also be found in GLM-Q, MLM-P+K, and MLM-Q+K. Most of the SNPs screened by the model determined according to the best QQ diagram can coincide with the SNPs screened by other models, which proves that the selection of the best model is accurate.

Prediction of candidate genes

LD decay distance is used to identify regions of potential candidate genes in GWAS. To confirm the beneficial alleles on each SNP peak associated with the seedless phenotype, [Zhang et al. \(2017\)](#) considered the gene model of the LD decay distance genomic region upstream and downstream of each SNP peak in the reference grape genome as a seedless candidate gene. Our results show that the attenuation distance is about 5 kb when r^2 drops to half. Based on LD attenuation, genes located within the 5 kb region surrounding trait-related SNPs were identified as potential candidate genes. We found an interesting phenomenon that when we extended the region to 50 kb and 100 kb, we found that we got the same four candidate genes as in the 5 kb range. Quantitative real-time polymerase chain reaction (qPCR) or reverse transcription qPCR (RT-qPCR) is a commonly used tool for gene expression analysis, detection, and quantification ([Foley, Leonard & Engel, 1993](#)). [Mahmoudzadeh et al. \(2016\)](#) established a simple and rapid method for extracting RNA from beef matrix to reduce the error caused by matrix effects on the experimental results. [Dermić et al. \(2023\)](#) have developed a new method for RT-qPCR using a specially modified primer in the reverse transcription step. All of these studies demonstrate that reverse transcription qPCR (RT-qPCR) can be used as a reliable and sensitive method for studying gene expression. We chose a kit that is easy to operate and has a strong matrix removal effect for the test, and the test is carried out in strict accordance with the requirements of all kits, so as to minimise the impact of the external environment on the test results. The expression levels of the above four genes were verified by RT-qPCR, and only the expression level of the *TEA027937.1* gene was significantly positively correlated with the amino acid content, which showed that the gene could be used as a candidate gene related to amino acid traits, and the function of the gene was L-tryptophan-pyruvate aminotransferase. L-tryptophan aminotransferase has been shown to be involved in auxin (IAA) biosynthesis. In plants, hormones derived from amino acid metabolic pathways, such as auxin (L-tryptophan) and ethylene (L-methionine), alter development and growth rates optimised for varying environmental conditions ([Santner & Estelle, 2009](#); [Jaillais & Chory, 2010](#)). Indoleacetic acid (IAA) is the most studied naturally active auxin. There are two main pathways for the biosynthesis of IAA: the tryptophan-dependent pathway and the tryptophan-independent pathway ([Woodward & Bartel, 2005](#)). The indole-3-pyruvate (IPyA) pathway, a tryptophan (Trp)-dependent pathway, appears to be a major contributor to the formation of free

IAA (Zhao, 2012). L-tryptophan aminotransferase converts L-tryptophan and pyruvate to indole-3-pyruvic acid (IPA) and alanine (Mashiguchi et al., 2011; Won et al., 2011). This is the first step in the IPyA branch of the tryptophan-dependent pathway. Arabidopsis tryptophan aminotransferase (TAA1) is the first known major enzyme for the biosynthesis of indole-3-acetic acid (IAA) from L-tryptophan, which belongs to the TAA1-related (TAA1/TAR) tryptophan aminotransferase family and has been identified in several mutant screens (Stepanova et al., 2008; Tao et al., 2008; Yamada et al., 2009). Previous studies have revealed that TAA1/TAR protein transaminase not only plays an important role in IAA biosynthesis but also plays an important role in plant nitrogen cycling (Le Deunff et al., 2019). These reports confirm our finding that one potential candidate gene identified in this study may be related to amino acids in tea plants, which will be beneficial for the development of molecular markers and MAS breeding in the future.

Amino acid content is a quantitative trait. Quantitative traits in plants are generally controlled by micro-effect polygenes. However, factors such as population size, recombination events, and genetic heterogeneity can affect the ability of QTLs to localise to identify loci or genes (Price, 2006). Omics techniques, including metabolomics, transcriptomics, proteomics, and ionomics, have contributed greatly to the identification and characterisation of genes, proteins, metabolites, and ions involved in the signalling pathways of interest (Budak et al., 2015). Transcriptome sequencing is an effective strategy to explore gene expression dynamics. In addition, metabolomics can be considered a complementary technique to transcriptomics (Lake et al., 2017). With the rapid development of bioinformatics and biostatistics tools, transcriptomics and metabolomics-based can correlate a large number of genes and metabolites to obtain a complete overview (Wu et al., 2019). For example, recent advances in metabolomics and transcriptomics techniques have greatly enhanced our understanding of flavonoid biosynthesis pathways and their regulatory networks (Wu et al., 2024). In tea plants, we can also combine multi-omics technology to conduct a comprehensive analysis of amino acid metabolism pathways in tea plants to further clarify the genetic basis of amino acids in tea plants.

CONCLUSIONS

This study demonstrates that it is feasible to generate large-scale SNP markers using the GBS strategy to analyse the genetic diversity and GWAS of amino acids in tea plants. Based on 78,819 SNPs developed by GBS, the genetic diversity of 212 tea accessions from Guizhou Province was analyzed. All 212 tea germplasms showed high genetic diversity. We used GWAS to identify a potential candidate gene associated with amino acid. It was found that this candidate gene was closely related to the auxin anabolic pathway. However, the auxin synthesis pathway is regulated by the amino acid metabolism pathway. In addition, the amino acid content of tea plants may be affected by external factors, so phenotypic data needs to be collected for several consecutive years to confirm this possibility. In conclusion, the results of this study can provide a reference for the future exploration of candidate genes for complex traits in tea plants, which is helpful for the breeding of excellent quality traits in tea plants.

ACKNOWLEDGEMENTS

We would like to thank all the participants for their help in completing this study together.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the National Science Foundation of China (32060700), National Guidance Foundation for Local Science and Technology Development of China ([2023] 009), Project on Guiyang Science and Technology Plan (Construction Technology Contract [2023] 48-21), Wangmo Eight Step Tea: Integration and Demonstration of Quality and Efficiency Enhancement Technologies (2021YFD1100307), the Science and Technology Plan Project of Guizhou Province, in RP China ([2023] General 480) and Youth Science and Technology Talent Project of Education Department of Guizhou Province (KY 2022-153). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

National Science Foundation of China: 32060700.

National Guidance Foundation for Local Science and Technology Development of China: [2023] 009.

Project on Guiyang Science and Technology Plan: Construction Technology Contract: [2023] 48-21.

Wangmo Eight Step Tea: Integration and Demonstration of Quality and Efficiency Enhancement Technologies: 2021YFD1100307.

Science and Technology Plan Project of Guizhou Province, in RP China: [2023] General 480.

Youth Science and Technology Talent Project of Education Department of Guizhou Province: KY 2022-153.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Qidi Wu conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Li Song conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Dingchen Bai analyzed the data, prepared figures and/or tables, and approved the final draft.
- Yihan Wang analyzed the data, prepared figures and/or tables, and approved the final draft.

- Yuting OuYang performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Kaixin Rao performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Qinfei Song conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Juanying Wang conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Suzhen Niu conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Yujie Ai conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The data is available in the [Supplemental Files](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.18969#supplemental-information>.

REFERENCES

- Alexander DH, Novembre J, Lange K. 2009.** Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**:1655–1664 DOI [10.1101/gr.094052.109](https://doi.org/10.1101/gr.094052.109).
- Alipour H, Bihamta MR, Mohammadi V, Peyghambari SA, Bai G, Zhang G. 2017.** Genotyping-by-Sequencing (GBS) revealed molecular genetic diversity of Iranian wheat landraces and cultivars. *Frontiers in Plant Science* **8**:1293 DOI [10.3389/fpls.2017.01293](https://doi.org/10.3389/fpls.2017.01293).
- Besnard J, Pratelli R, Zhao C, Sonawala U, Collakova E, Pilot G, Okumoto S. 2016.** UMAMIT14 is an amino acid exporter involved in phloem unloading in *Arabidopsis* roots. *Journal of Experimental Botany* **67**:6385–6397 DOI [10.1093/jxb/erw412](https://doi.org/10.1093/jxb/erw412).
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007.** TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**:2633–2635 DOI [10.1093/bioinformatics/btm308](https://doi.org/10.1093/bioinformatics/btm308).
- Budak H, Hussain B, Khan Z, Ozturk NZ, Ullah N. 2015.** From genetics to functional genomics: improvement in drought signaling and tolerance in wheat. *Frontiers in Plant Science* **6**:1012 DOI [10.3389/fpls.2015.01012](https://doi.org/10.3389/fpls.2015.01012).
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015.** Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**:7 DOI [10.1186/s13742-015-0047-8](https://doi.org/10.1186/s13742-015-0047-8).
- Chen Y, Fu X, Mei X, Zhou Y, Cheng S, Zeng L, Dong F, Yang Z. 2017.** Proteolysis of chloroplast proteins is responsible for accumulation of free amino acids in

- dark-treated tea (*Camellia sinensis*) leaves. *Journal of Proteomics* **157**:10–17 DOI [10.1016/j.jprot.2017.01.017](https://doi.org/10.1016/j.jprot.2017.01.017).
- Chen Y, Niu S, Deng X, Song Q, He L, Bai D, He Y. 2023.** Genome-wide association study of leaf-related traits in tea plant in Guizhou based on genotyping-by-sequencing. *BMC Plant Biology* **23**:196 DOI [10.1186/s12870-023-04192-0](https://doi.org/10.1186/s12870-023-04192-0).
- Chung YS, Choi SC, Jun T-H, Kim C. 2017.** Genotyping-by-sequencing: a promising tool for plant genetics research and breeding. *Horticulture, Environment, and Biotechnology* **58**:425–431 DOI [10.1007/s13580-017-0297-8](https://doi.org/10.1007/s13580-017-0297-8).
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group. 2011.** The variant call format and VCFtools. *Bioinformatics* **27**:2156–2158 DOI [10.1093/bioinformatics/btr330](https://doi.org/10.1093/bioinformatics/btr330).
- Dermić D, Ljubić S, Matulić M, Procino A, Feliciello MC, Ugarković Đ, Feliciello I. 2023.** Reverse transcription-quantitative PCR (RT-qPCR) without the need for prior removal of DNA. *Scientific Reports* **13**:11470 DOI [10.1038/s41598-023-38383-4](https://doi.org/10.1038/s41598-023-38383-4).
- Fang K, Xia Z, Li H, Jiang X, Qin D, Wang Q, Wang Q, Pan C, Li B, Wu H. 2021.** Genome-wide association analysis identified molecular markers associated with important tea flavor-related metabolites. *Horticulture Research* **8**:1–17 DOI [10.1038/s41438-021-00477-3](https://doi.org/10.1038/s41438-021-00477-3).
- Filialt DL, Maloof JN. 2012.** A genome-wide association study identifies variants underlying the *Arabidopsis thaliana* shade avoidance response. *PLOS Genetics* **8**:e1002589 DOI [10.1371/journal.pgen.1002589](https://doi.org/10.1371/journal.pgen.1002589).
- Foley KP, Leonard MW, Engel JD. 1993.** Quantitation of RNA using the polymerase chain reaction. *Trends in Genetics* **9**:380–385 DOI [10.1016/0168-9525\(93\)90137-7](https://doi.org/10.1016/0168-9525(93)90137-7).
- Fu X, Cheng S, Liao Y, Xu X, Wang X, Hao X, Xu P, Dong F, Yang Z. 2020.** Characterization of l-theanine hydrolase *in vitro* and subcellular distribution of its specific product ethylamine in tea (*Camellia sinensis*). *Journal of Agricultural and Food Chemistry* **68**(39):10842–10851 DOI [10.1021/acs.jafc.0c01796](https://doi.org/10.1021/acs.jafc.0c01796).
- Gaudin M, Desnues C. 2018.** Hybrid capture-based next generation sequencing and its application to human infectious diseases. *Frontiers in Microbiology* **9**:2924 DOI [10.3389/fmicb.2018.02924](https://doi.org/10.3389/fmicb.2018.02924).
- Glaubitx JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES. 2014.** TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLOS ONE* **9**:e90346 DOI [10.1371/journal.pone.0090346](https://doi.org/10.1371/journal.pone.0090346).
- Guo Y, Li D, Qiu H, Zhang X, Liu L, Zhao J, Jiang D. 2023.** Genome-wide association studies reveal the genetic basis of amino acid content variation in tea plants. *Journal of Integrative Agriculture* **22**:3364–3379 DOI [10.1016/j.jia.2023.10.002](https://doi.org/10.1016/j.jia.2023.10.002).
- Hamon P, Grover CE, Davis AP, Rakotomalala J-J, Raharimalala NE, Albert VA, Sreenath HL, Stoffelen P, Mitchell SE, Couturon E, Hamon S, De Kochko A, Crouzillat D, Rigoreau M, Sumirat U, Akaffou S, Guyot R. 2017.** Genotyping-by-sequencing provides the first well-resolved phylogeny for coffee (*Coffea*) and insights

- into the evolution of caffeine content in its species. *Molecular Phylogenetics and Evolution* **109**:351–361 DOI [10.1016/j.ympev.2017.02.009](https://doi.org/10.1016/j.ympev.2017.02.009).
- He L, Luo J, Niu S, Bai D, Chen Y. 2023.** Population structure analysis to explore genetic diversity and geographical distribution characteristics of wild tea plant in Guizhou Plateau. *BMC Plant Biology* **23**:255 DOI [10.1186/s12870-023-04239-2](https://doi.org/10.1186/s12870-023-04239-2).
- He J, Zhao X, Laroche A, Lu Z-X, Liu H, Li Z. 2014.** Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in Plant Science* **5**:484 DOI [10.3389/fpls.2014.00484](https://doi.org/10.3389/fpls.2014.00484).
- Huang F, Duan J, Lei Y, Liu Z, Kang Y, Luo Y, Chen Y, Li Y, Liu S, Li S, Liu Z, Huang F, Duan J, Lei Y, Liu Z, Kang Y, Luo Y, Chen Y, Li Y, Liu S, Li S, Liu Z. 2022a.** Genetic diversity, population structure and core collection analysis of Hunan tea plant germplasm through genotyping-by-sequencing. *Beverage Plant Research* **2**:1–7 DOI [10.48130/BPR-2022-0005](https://doi.org/10.48130/BPR-2022-0005).
- Huang L, Min Y, Schiessl S, Xiong X, Jan HU, He X, Qian W, Guan C, Snowdon RJ, Hua W, Guan M, Qian L. 2021.** Integrative analysis of GWAS and transcriptome to reveal novel loci regulation flowering time in semi-winter rapeseed. *Plant Science* **310**:110980 DOI [10.1016/j.plantsci.2021.110980](https://doi.org/10.1016/j.plantsci.2021.110980).
- Huang R, Wang J-Y, Yao M-Z, Ma C-L, Chen L. 2022b.** Quantitative trait loci mapping for free amino acid content using an albino population and SNP markers provides insight into the genetic improvement of tea plants. *Horticulture Research* **9**:uhab029 DOI [10.1093/hr/uhab029](https://doi.org/10.1093/hr/uhab029).
- Jaillais Y, Chory J. 2010.** Unraveling the paradoxes of plant hormone signaling integration. *Nature Structural & Molecular Biology* **17**:642–645 DOI [10.1038/nsmb0610-642](https://doi.org/10.1038/nsmb0610-642).
- Jung Y, Han D. 2022.** BWA-MEME: BWA-MEM emulated with a machine learning approach. *Bioinformatics* **38**:2404–2413 DOI [10.1093/bioinformatics/btac137](https://doi.org/10.1093/bioinformatics/btac137).
- Lake BB, Codeluppi S, Yung YC, Gao D, Chun J, Kharchenko PV, Linnarsson S, Zhang K. 2017.** A comparative strategy for single-nucleus and single-cell transcriptomes confirms accuracy in predicted cell-type expression from nuclear RNA. *Scientific Reports* **7**:6031 DOI [10.1038/s41598-017-04426-w](https://doi.org/10.1038/s41598-017-04426-w).
- Lavanya GR, Srivastava J, Ranade SA. 2008.** Molecular assessment of genetic diversity in mung bean germplasm. *Journal of Genetics* **87**:65–74 DOI [10.1007/s12041-008-0009-3](https://doi.org/10.1007/s12041-008-0009-3).
- Le Deunff E, Beauclair P, Deleu C, Lecourt J. 2019.** Inhibition of aminotransferases by aminoethoxyvinylglycine triggers a nitrogen limitation condition and deregulation of histidine homeostasis that impact root and shoot development and nitrate uptake. *Frontiers in Plant Science* **10**:1387 DOI [10.3389/fpls.2019.01387](https://doi.org/10.3389/fpls.2019.01387).
- Lee J. 2017.** A study on Chinese Guizhou tea. *The Journal of Tea Culture & Industry Studies* **37**:99–134.
- Lee SJ, Ban SH, Kim GH, Kwon SI, Kim JH, Choi C. 2017.** Identification of potential gene-associated major traits using GBS-GWAS for Korean apple germplasm collections. *Plant Breeding* **136**:977–986 DOI [10.1111/pbr.12544](https://doi.org/10.1111/pbr.12544).

- Li F, Dong C, Yang T, Bao S, Fang W, Lucas WJ, Zhang Z. 2021a. The tea plant CsLHT1 and CsLHT6 transporters take up amino acids, as a nitrogen source, from the soil of organic tea plantations. *Horticulture Research* 8:1–12 DOI 10.1038/s41438-021-00615-x.
- Li W, Yang B, Xu J, Peng L, Sun S, Huang Z, Jiang X, He Y, Wang Z. 2021b. A genome-wide association study reveals that the 2-oxoglutarate/malate translocator mediates seed vigor in rice. *The Plant Journal* 108:478–491 DOI 10.1111/tpj.15455.
- Li F, Zhang S-R, Chen L, Chen J-D. 2024. Genome assembly of tea plants (*Camellia* spp.). In: Chen L, Chen J-D, eds. *The tea plant genome*. Singapore: Springer Nature, 195–212 DOI 10.1007/978-981-97-0680-8_9.
- Liu Y, Chen S, Jiang C, Liu H, Wang J, He W, Moon D, Chen J, Chen L, Ma J. 2023a. Combined QTL mapping, GWAS and transcriptomic analysis revealed a candidate gene associated with the timing of spring bud flush in tea plant (*Camellia sinensis*). *Horticulture Research* 10:uhad149 DOI 10.1093/hr/uhad149.
- Liu C-C, Shringarpure S, Lange K, Novembre J. 2020. Exploring population structure with admixture models and principal component analysis. *Methods in Molecular Biology* 2090:67–86 DOI 10.1007/978-1-0716-0199-0_4.
- Liu Z, Zhao Y, Yang P, Cheng Y, Huang F, Li S, Yang Y, Liu Z, Zhao Y, Yang P, Cheng Y, Huang F, Li S, Yang Y. 2023b. Population whole-genome resequencing reveals the phylogenetic relationships and population structure of four Hunan typical tea landraces. *Beverage Plant Research* 3:9 DOI 10.48130/BPR-2023-0009.
- Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* 25:402–408 DOI 10.1006/meth.2001.1262.
- Ma J-Q, Jin J-Q, Yao M-Z, Ma C-L, Xu Y-X, Hao W-J, Chen L. 2018. Quantitative trait loci mapping for theobromine and caffeine contents in tea plant (*Camellia sinensis*). *Journal of Agricultural and Food Chemistry* 66:13321–13327 DOI 10.1021/acs.jafc.8b05355.
- Mahmoudzadeh M, Hosseini H, Hedayati M, Mousavi Khanghah A, Djalma Chaves R, Azizkhani M. 2016. Establishment of a method for describing stx genes expression of *E. coli* O157:H7 in ground beef matrix during refrigerated storage. *Journal of Food Safety* 36:220–226 DOI 10.1111/jfs.12234.
- Mashiguchi K, Tanaka K, Sakai T, Sugawara S, Kawaide H, Natsume M, Hanada A, Yaeno T, Shirasu K, Yao H, McSteen P, Zhao Y, Hayashi K, Kamiya Y, Kasahara H. 2011. The main auxin biosynthesis pathway in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* 108:18512–18517 DOI 10.1073/pnas.1108434108.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20:1297–1303 DOI 10.1101/gr.107524.110.

- Mi X, Yang C, Qiao D, Tang M, Guo Y, Liang S, Li Y, Chen Z, Chen J. 2023. *De novo* full length transcriptome analysis of a naturally caffeine-free tea plant reveals specificity in secondary metabolic regulation. *Scientific Reports* **13**:6015 DOI [10.1038/s41598-023-32435-5](https://doi.org/10.1038/s41598-023-32435-5).
- Nebert DW, Vesell ES. 2013. Chapter 19—pharmacogenetics and pharmacogenomics. In: Rimoin D, Pyeritz R, Korf B, eds. *Emery and Rimoin's principles and practice of medical genetics*. Sixth Edition. Oxford: Academic Press, 1–27 DOI [10.1016/B978-0-12-383834-6.00023-9](https://doi.org/10.1016/B978-0-12-383834-6.00023-9).
- Nie H, Park H, Kim S, Kim D, Kim S, Kwon S-Y, Kim S-H. 2023. Genetic diversity assessment and genome-wide association study reveal candidate genes associated with component traits in sweet potato (*Ipomoea batatas* (L.) Lam). *Molecular Genetics and Genomics* **298**:653–667 DOI [10.1007/s00438-023-02007-3](https://doi.org/10.1007/s00438-023-02007-3).
- Niu S, Song Q, Koiwa H, Qiao D, Zhao D, Chen Z, Liu X, Wen X. 2019. Genetic diversity, linkage disequilibrium, and population structure analysis of the tea plant (*Camellia sinensis*) from an origin center, Guizhou plateau, using genome-wide SNPs developed by genotyping-by-sequencing. *BMC Plant Biology* **19**:328 DOI [10.1186/s12870-019-1917-5](https://doi.org/10.1186/s12870-019-1917-5).
- Orcan F. 2020. Parametric or non-parametric: skewness to test normality for mean comparison. *International Journal of Assessment Tools in Education* **7**:236–246 DOI [10.21449/ijate.656077](https://doi.org/10.21449/ijate.656077).
- Pandey J, Scheuring DC, Koym JW, Coombs J, Novy RG, Thompson AL, Holm DG, Douches DS, Miller JC, Vales MI. 2021. Genetic diversity and population structure of advanced clones selected over forty years by a potato breeding program in the USA. *Scientific Reports* **11**:8344 DOI [10.1038/s41598-021-87284-x](https://doi.org/10.1038/s41598-021-87284-x).
- Price AH. 2006. Believe it or not, QTLs are accurate!. *Trends in Plant Science* **11**:213–216 DOI [10.1016/j.tplants.2006.03.006](https://doi.org/10.1016/j.tplants.2006.03.006).
- Rabosky DL. 2016. Reproductive isolation and the causes of speciation rate variation in nature. *Biological Journal of the Linnean Society* **118**:13–25 DOI [10.1111/bij.12703](https://doi.org/10.1111/bij.12703).
- Ruan J, Haerdter R, Gerendás J. 2010. Impact of nitrogen supply on carbon/nitrogen allocation: a case study on amino acids and catechins in green tea (*Camellia sinensis* (L.) O. Kuntze) plants. *Plant Biology* **12**:724–734 DOI [10.1111/j.1438-8677.2009.00288.x](https://doi.org/10.1111/j.1438-8677.2009.00288.x).
- Sallam A, Martsch R. 2015. Association mapping for frost tolerance using multi-parent advanced generation inter-cross (MAGIC) population in faba bean (*Vicia faba* L.). *Genetica* **143**(4):501–514 DOI [10.1007/s10709-015-9848-z](https://doi.org/10.1007/s10709-015-9848-z).
- Santner A, Estelle M. 2009. Recent advances and emerging trends in plant hormone signalling. *Nature* **459**:1071–1078 DOI [10.1038/nature08122](https://doi.org/10.1038/nature08122).
- Schmidt TL, Jasper M-E, Weeks AR, Hoffmann AA. 2021. Unbiased population heterozygosity estimates from genome-wide sequence data. *Methods in Ecology and Evolution* **12**:1888–1898 DOI [10.1111/2041-210X.13659](https://doi.org/10.1111/2041-210X.13659).
- Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, Peichel CL, Saetre G-P, Bank C, Brännström Å, Brelsford A, Clarkson CS, Eroukhmanoff

- F, Feder JL, Fischer MC, Foote AD, Franchini P, Jiggins CD, Jones FC, Lindholm AK, Lucek K, Maan ME, Marques DA, Martin SH, Matthews B, Meier JI, Möst M, Nachman MW, Nonaka E, Rennison DJ, Schwarzer J, Watson ET, Westram AM, Widmer A. 2014. Genomics and the origin of species. *Nature Reviews Genetics* 15:176–192 DOI 10.1038/nrg3644.
- Sonah H, O'Donoghue L, Cober E, Rajcan I, Belzile F. 2015. Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. *Plant Biotechnology Journal* 13:211–221 DOI 10.1111/pbi.12249.
- Stecher G, Tamura K, Kumar S. 2020. Molecular evolutionary genetics analysis (MEGA) for macOS. *Molecular Biology and Evolution* 37:1237–1239 DOI 10.1093/molbev/msz312.
- Stepanova AN, Robertson-Hoyt J, Yun J, Benavente LM, Xie D-Y, Doležal K, Schlereth A, Jürgens G, Alonso JM. 2008. TAA1-mediated auxin biosynthesis is essential for hormone crosstalk and plant development. *Cell* 133:177–191 DOI 10.1016/j.cell.2008.01.047.
- Sun Y, Shi Y, Liu G, Yao F, Zhang Y, Yang C, Guo H, Liu X, Jin C, Luo J. 2020. Natural variation in the OsbZIP18 promoter contributes to branched-chain amino acid levels in rice. *New Phytologist* 228:1548–1558 DOI 10.1111/nph.16800.
- Tandoh KZ, Amenga-Etego L, Quashie NB, Awandare G, Wilson M, Duah-Quashie NO. 2021. Plasmodium falciparum malaria parasites in Ghana show signatures of balancing selection at artemisinin resistance predisposing background genes. *Evolutionary Bioinformatics Online* 17:1176934321999640 DOI 10.1177/1176934321999640.
- Tao Y, Ferrer J-L, Ljung K, Pojer F, Hong F, Long JA, Li L, Moreno JE, Bowman ME, Ivans LJ, Cheng Y, Lim J, Zhao Y, Ballaré CL, Sandberg G, Noel JP, Chory J. 2008. Rapid synthesis of auxin via a new tryptophan-dependent pathway is required for shade avoidance in plants. *Cell* 133:164–176 DOI 10.1016/j.cell.2008.01.049.
- Wang RJ, Gao XF, Yang J, Kong XR. 2019. Genome-wide association study to identify favorable SNP allelic variations and candidate genes that control the timing of spring bud flush of tea (*Camellia sinensis*) using SLAF-seq. *Journal of Agricultural and Food Chemistry* 67:10380–10391 DOI 10.1021/acs.jafc.9b03330.
- Wang M, Li J, Liu X, Liu C, Qian J, Yang J, Zhou X, Jia Y, Tang J, Zeng L. 2022. Characterization of key odorants in lingtong dancong oolong tea and their differences induced by environmental conditions from different altitudes. *Metabolites* 12:1063 DOI 10.3390/metabo12111063.
- Wang Y, Niu S, Deng X, Bai D, Chen Z, Deng X, Huang D. 2024. Genome-wide association study, population structure, and genetic diversity of the tea plant in Guizhou Plateau. *BMC Plant Biology* 24:79 DOI 10.1186/s12870-024-04761-x.
- Wang H, Vieira FG, Crawford JE, Chu C, Nielsen R. 2017. Asian wild rice is a hybrid swarm with extensive gene flow and feralization from domesticated rice. *Genome Research* 27:1029–1038 DOI 10.1101/gr.204800.116.

- Wen W, Li D, Li X, Gao Y, Li W, Li H, Liu J, Liu H, Chen W, Luo J, Yan J. 2014. Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nature Communications* 5:3438 DOI 10.1038/ncomms4438.
- Wen B, Luo Y, Liu D, Zhang X, Peng Z, Wang K, Li J, Huang J, Liu Z. 2020. The R2R3-MYB transcription factor CsMYB73 negatively regulates l-Theanine biosynthesis in tea plants (*Camellia sinensis* L.). *Plant Science* 298:110546 DOI 10.1016/j.plantsci.2020.110546.
- Won C, Shen X, Mashiguchi K, Zheng Z, Dai X, Cheng Y, Kasahara H, Kamiya Y, Chory J, Zhao Y. 2011. Conversion of tryptophan to indole-3-acetic acid by TRYPTOPHAN AMINOTRANSFERASES OF ARABIDOPSIS and YUCCAs in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* 108:18518–18523 DOI 10.1073/pnas.1108436108.
- Woodward AW, Bartel B. 2005. Auxin: regulation, action, and interaction. *Annals of Botany* 95:707–735 DOI 10.1093/aob/mci083.
- Wu L-Y, Lv Y-Q, Ye Y, Liang Y-R, Ye J-H. 2020. Transcriptomic and translaticomic analyses reveal insights into the developmental regulation of secondary metabolism in the young shoots of tea plants (*Camellia sinensis* L.). *Journal of Agricultural and Food Chemistry* 68:10750–10762 DOI 10.1021/acs.jafc.0c03341.
- Wu H, Naren G, Han C, Elsheery NI, Zhang L. 2024. Exploring the flavonoid biosynthesis pathway of two ecotypes of *leymus chinensis* using transcriptomic and metabolomic analysis. *Agronomy* 14:1839 DOI 10.3390/agronomy14081839.
- Wu B, Xiao X, Li S, Zuo G. 2019. Transcriptomics and metabonomics of the anti-aging properties of total flavones of *Epimedium* in relation to lipid metabolism. *Journal of Ethnopharmacology* 229:73–80 DOI 10.1016/j.jep.2018.09.039.
- Wu X, Xu P, Wu X, Wang B, Lu Z, Li G. 2017. Genome-wide association analysis of free glutamate content, a key factor conferring umami taste in the bottle gourd [*Lagenaria siceraria* (Mol.) Standl.]. *Scientia Horticulturae* 225:795–801 DOI 10.1016/j.scienta.2017.08.015.
- Xia E, Tong W, Hou Y, An Y, Chen L, Wu Q, Liu Y, Yu J, Li F, Li R, Li P, Zhao H, Ge R, Huang J, Mallano AI, Zhang Y, Liu S, Deng W, Song C, Zhang Z, Zhao J, Wei S, Zhang Z, Xia T, Wei C, Wan X. 2020a. The reference genome of tea plant and resequencing of 81 diverse accessions provide insights into its genome evolution and adaptation. *Molecular Plant* 13:1013–1026 DOI 10.1016/j.molp.2020.04.010.
- Xia E-H, Tong W, Wu Q, Wei S, Zhao J, Zhang Z-Z, Wei C-L, Wan X-C. 2020b. Tea plant genomics: achievements, challenges and perspectives. *Horticulture Research* 7:1–19 DOI 10.1038/s41438-019-0225-4.
- Xu J, Long Y, Wu J, Xu H, Wen J, Meng J, Shi C. 2015. QTL mapping and analysis of the embryo and maternal plant for three limiting amino acids in rapeseed meal. *European Food Research and Technology* 240:147–158 DOI 10.1007/s00217-014-2316-7.
- Yamada M, Greenham K, Prigge MJ, Jensen PJ, Estelle M. 2009. The TRANSPORT INHIBITOR RESPONSE2 gene is required for auxin synthesis and diverse aspects of plant development. *Plant Physiology* 151:168–179 DOI 10.1104/pp.109.138859.

- Yamanishi T. 1995.** VIII. Flavor of tea. *Food Reviews International* **11**:477–525
DOI [10.1080/87559129509541056](https://doi.org/10.1080/87559129509541056).
- Yang Y, Dong S, Miao H, Liu X, Dai Z, Li X, Gu X, Zhang S. 2022a.** Genome-wide association studies reveal candidate genes related to stem diameter in cucumber (*Cucumis sativus* L.). *Genes* **13**:1095 DOI [10.3390/genes13061095](https://doi.org/10.3390/genes13061095).
- Yang C, Qiao D, Guo Y, Liang S, Lin K, Chen Z. 2022b.** Analysis of amino acids and theanine from 115 tea tree resources in Guizhou Province and screening of specific resources. *Journal of Zhejiang Agricultural Sciences* **34**:1351–1360.
- Yu Z, Yang Z. 2020.** Understanding different regulatory mechanisms of proteinaceous and non-proteinaceous amino acid formation in tea (*Camellia sinensis*) provides new insights into the safe and effective alteration of tea flavor and function. *Critical Reviews in Food Science and Nutrition* **60**:844–858 DOI [10.1080/10408398.2018.1552245](https://doi.org/10.1080/10408398.2018.1552245).
- Yu L-X, Zheng P, Zhang T, Rodringuez J, Main D. 2017.** Genotyping-by-sequencing-based genome-wide association studies on Verticillium wilt resistance in autotetraploid alfalfa (*Medicago sativa* L.). *Molecular Plant Pathology* **18**:187–194 DOI [10.1111/mpp.12389](https://doi.org/10.1111/mpp.12389).
- Zhang C, Dong S-S, Xu J-Y, He W-M, Yang T-L. 2019.** PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **35**:1786–1788 DOI [10.1093/bioinformatics/bty875](https://doi.org/10.1093/bioinformatics/bty875).
- Zhang H, Fan X, Zhang Y, Jiang J, Liu C. 2017.** Identification of favorable SNP alleles and candidate genes for seedlessness in *Vitis vinifera* L. using genome-wide association mapping. *Euphytica* **213**:136 DOI [10.1007/s10681-017-1919-z](https://doi.org/10.1007/s10681-017-1919-z).
- Zhang M, Han W, Tang H, Li G, Zhang M, Xu R, Liu Y, Yang T, Li W, Zou J, Wu K. 2018.** Genomic diversity dynamics in conserved chicken populations are revealed by genome-wide SNPs. *BMC Genomics* **19**:598 DOI [10.1186/s12864-018-4973-6](https://doi.org/10.1186/s12864-018-4973-6).
- Zhang C, Li Y, Mei P, Gong Y, Liu D, Ye Y, Wen W, Yao M, Ma C. 2024.** Developmental-specific regulation promotes the free amino acids accumulation in chlorotic tea plants (*Camellia sinensis*). *Journal of Plant Physiology* **303**:154371 DOI [10.1016/j.jplph.2024.154371](https://doi.org/10.1016/j.jplph.2024.154371).
- Zhang J, Zhang D, Fan Y, Li C, Xu P, Li W, Sun Q, Huang X, Zhang C, Wu L, Yang H, Wang S, Su X, Li X, Song Y, Wu M, Lian X, Li Y. 2021.** The identification of grain size genes by RapMap reveals directional selection during rice domestication. *Nature Communications* **12**:5673 DOI [10.1038/s41467-021-25961-1](https://doi.org/10.1038/s41467-021-25961-1).
- Zhao Y. 2012.** Auxin biosynthesis: a simple two-step pathway converts tryptophan to indole-3-acetic acid in plants. *Molecular Plant* **5**:334–338 DOI [10.1093/mp/ssr104](https://doi.org/10.1093/mp/ssr104).
- Zhao Z, Song Q, Bai D, Niu S, He Y, Qiao D, Chen Z, Li C, Luo J, Li F. 2022.** Population structure analysis to explore genetic diversity and geographical distribution characteristics of cultivated-type tea plant in Guizhou Plateau. *BMC Plant Biology* **22**:55 DOI [10.1186/s12870-022-03438-7](https://doi.org/10.1186/s12870-022-03438-7).
- Zhao H, Wang S, Yuan X. 2020.** Detection of pathogenic microbe composition using next-generation sequencing data. *Frontiers in Genetics* **11**:603093 DOI [10.3389/fgene.2020.603093](https://doi.org/10.3389/fgene.2020.603093).

Zhong M, Wang L, Yuan D, Luo L, Xu C, He Y. 2011. Identification of QTL affecting protein and amino acid contents in rice. *Rice Science* **18**:187–195

[DOI 10.1016/S1672-6308\(11\)60026-7](https://doi.org/10.1016/S1672-6308(11)60026-7).

Zou B, Jia Z, Tian S, Wang X, Gou Z, Beibei L, Dong H. 2013. AtMYB44 positively modulates disease resistance to *Pseudomonas syringae* through the salicylic acid signalling pathway in *Arabidopsis*. *Functional Plant Biology* **40**:304–313

[DOI 10.1071/FP12253](https://doi.org/10.1071/FP12253).