Received: 2 June 2024 | Revised: 15 August 2024 | Accepted: 3 September 2024

DOI: 10.1111/ppl.14542

ORIGINAL RESEARCH

Physiologia Plantarum

Genome-wide association study reveals the genetic basis of amino acids contents variations in Peanut (*Arachis hypogaea* L.)

Muhammad J. Umer[1†] | **Qing Lu**[1†] | **Lu Huang**[1] | **Raufa Batool**[2] | **Hao Liu**[1] | **Haifen Li**[1] | **Runfeng Wang**[1] | **Yu Qianxia**[1] | **Rajeev K. Varshney**[3] | **Manish K. Pandey**[4] | **Yanbin Hong**[1] | **Xiaoping Chen**[1]

[1]Crops Research Institute, Guangdong Academy of Agricultural Sciences, Guangdong Provincial Key Laboratory of Crop Genetic Improvement, South China Peanut Sub-Centre of National Centre of Oilseed Crops Improvement, Guangzhou, China

[2]Institute of Plant Protection, Guangdong Academy of Agricultural Sciences, Key Laboratory of Green Prevention and Control on Fruits and Vegetables in South China Ministry of Agriculture and Rural Affairs, Guangdong Provincial Key Laboratory of High Technology for Plant Protection, Guangzhou, China

[3]WA State Agricultural Biotechnology Centre, Centre for Crop and Food Innovation, Food Futures Institute, Murdoch University, Murdoch, Western Australia, Australia

[4]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India

Correspondence
Qing Lu,
Email: luqing@gdaas.cn

Yanbin Hong,
Email: hongyanbin@gdaas.cn

Xiaoping Chen,
Email: chenxiaoping@gdaas.cn

Funding information
Guangzhou Basic and Applied Basic Research Foundation, Grant/Award Numbers: 202201010281, 2023A04J0776; Special Support Program of Guangdong Province, Grant/Award Number: 2021TX06N789; Special Fund for High Level Academy of Agriculture Science, Grant/Award Numbers: R2020PY-JG005, R2020PY-JX004, R2021PY-QY003, R2022YJ-YB3025, R2023PY-JG007; Guangdong Provincial Key Research and Development Program-Modern Seed Industry, Grant/Award Number: 2022B0202060004; Open Competition Program of Top Ten Critical Priorities of Agricultural Science and Technology Innovation for the 14th Five-Year Plan in Guangdong Province, Grant/Award Number: 2022SDZG05; Basic and Applied Basic Research Foundation of Guangdong Province, Grant/Award Number: 2023A1515010098; Guangdong Science and Technology Plan Project, Grant/Award Number: 2023B1212060038; National Key Research and Development Program of China,

Abstract

Peanut is a significant source of protein for human consumption. One of the primary objectives in peanut breeding is the development of new cultivars with enhanced nutritional values. To further this goal, a genome-wide association study (GWAS) was conducted to analyze seed amino acids contents in 390 diverse peanut accessions collected worldwide, mainly from China, India, and the United States, in 2017 and 2018. These accessions were assessed for their content of 10 different amino acids. Variations in amino acids contents were observed, and arginine (Arg) was found to have the highest average value among all the amino acids quantified. The geographical distribution of the accessions also revealed variations in amino acids contents. High and positive correlation coefficients were observed among the amino acids, suggesting linked metabolic pathways or genetic regulation. A total of 88 single nucleotide polymorphisms (SNPs) spanning various chromosomes were identified, each associated with different amino acids. By using a combination of GWAS, expression anlaysis, and genomic polymorphisim comparisions, the *Ahy_A09g041582* (*LAC15*) gene located on chromrosome A09 was identified as the key candidate which might be involved in plant growth and regulation and may alter amino acids levels. Expression analysis indicated that *Ahy_A09g041582* has higher expressions in the shells and seeds than other genes located in the candidate region. This study may help with marker-based breeding of peanuts with higher nutritional value and offers fresh insights into the genetic basis of the amino acids contents of peanuts.

† Muhammad Jawad Umer and Qing Lu contributed equally to this work.

Grant/Award Number: 2023YFD1202800; National Natural Science Foundation of China, Grant/Award Numbers: 32172051, 32301869; China Agriculture Research System of MOF and MARA, Grant/Award Number: CARS-13

Edited by A. Fernie

1 | INTRODUCTION

Peanut (*Arachis hypogaea* L.), a tetraploid oilseed crop, originated from a cross-fertilization event between its diploid ancestors, *Arachis duranensis* (AA) and *Arachis ipaensis* (BB), in South America (Lu et al., 2024). Peanuts contain around 45–56% oil and 24% protein, as well as fatty acids (FA) like oleic acid (OA). OA is a healthy unsaturated FA that makes up about 50% of peanut oil (Liu et al., 2018). Recent studies have highlighted the health benefits associated with a diet rich in monounsaturated oleic acid (OA, C18:1), including improved cholesterol levels, reduced risk of coronary heart diseases, and better management of inflammatory conditions (Vassiliou et al., 2009, Isleib et al., 2006).

The amino acid composition and quantity of seed storage proteins are related to the nutritional quality of seeds (Mandal and Mandal, 2000, Young and Pellett, 1994). Nevertheless, the widely planted peanut cultivars typically have insufficient levels of essential amino acids, such as lysine (Lys) and tryptophan (Try) (Misra et al., 1972). However, the genetic basis of natural variations in amino acids contents in peanut has not yet been clarified. To facilitate breeding for balanced amino acid composition, it is important to identify the genes controlling amino acid content in the peanut seeds.

Although more than 180 amino acids have been discovered in nature, only 20 amino acids constitute proteins. Many amino acids, such as homoserine, homocysteine, ornithine and citrulline, play important roles in growth and development (Dunlop et al., 2015), and defence against insect herbivores (Huang et al., 2011). Amino acids are also important signalling molecules regulating several signal pathways related to the growth and development of both animals and plants. Some studies have found that aspartate plays an important role in human cell proliferation (Birsoy et al., 2015, Sullivan et al., 2015). Proline could help maintain cellular osmotic homoeostasis, as well as redox balance and energy status (Krishnan et al., 2008). Proline may also function as a molecular chaperone to protect proteins from denaturation (Mishra and Dubey, 2006, Sharma and Dubey, 2005), an antioxidant to scavenge ROS, a singlet oxygen quencher (Matysik et al., 2002).

Forward genetic approaches, including quantitative trait loci (QTL) mapping and association mapping, are important for identifying loci regulating free amino acids contents in higher plants. The amino acids metabolic pathways, including biosynthesis, degradation and regulation, are well-studied in microorganisms (Miflin and Lea, 1977). Studies of the model plant *Arabidopsis thaliana* have focused on the roles of amino acids in nitrogen nutrition (Crawford and Forde, 2002), metabolism and regulation (Hell and Wirtz, 2011). Some key genes regulating free amino acids content have been identified in Arabidopsis (Angelovici et al., 2013), tobacco (Maloney et al., 2010), soybean (Ishimoto et al., 2010), rapeseed (Moulin et al., 2006), rice (Zhou et al., 2009) and maize (Wang et al., 2007). Opaque2 (O2) is an endosperm-specific transcription factor belonging to the bZIP family, whose mutation could increase free Lys levels and enhance the overall nutritional value of grain by reducing the 22- kD a- and b-zein transcripts and proteins in maize (Hunter et al., 2002). In *Glycine max*, Khandaker et al. (2015) identified 16 QTLs associated with seven amino acids traits through QTL mapping (Khandaker et al., 2015). Zhong et al. (2011) measured protein and amino acids content in an F_9 population of recombinant inbred line and identified 112 QTLs in rice by QTL mapping (Zhong et al., 2011). In addition, Xu et al. (2014) identified 17 QTLs associated with the content of Lys, threonine (Thr) and methionine (Met) in rapeseed meal by QTL mapping (Xu et al., 2015). Unlike QTL mapping, genome-wide association studies (GWASs) are quick, inexpensive, and high-resolution tools for identifying higher plant genetic variants. Currently, GWASs have been extensively used for the study of metabolite traits in many model plants and widely cultivated crops, such as *Arabidopsis thaliana* (Filiault and Maloof, 2012), *Oryza sativa* (Sun et al., 2020), *Zea mays* (Wen et al., 2014), *Triticum aestivum* (Chen et al., 2020), *Brassica napus* (Wang et al., 2017), *Glycine max* (Zhang et al., 2019a), and *Solanum lycopersicum* (Ye et al., 2017).

With the rapid development of next-generation sequencing technology, GWASs have been conducted to study several traits in peanut. Zhang et al. (2023) used 158 peanut accessions to dissect the molecular footprint of agronomic traits related to domestication using specific-locus amplified fragment sequencing (SLAF-seq method) (Zhang et al., 2023a). By carrying out GWAS analyses, 93 non-overlapping SNP peaks that are significantly associated with four yield-related traits in peanut were identified by (Wang et al., 2019). GWAS, with principal component analysis, identifies QTLs associated with main peanut flavor-related traits (Zhang et al., 2023a). A GWAS conducted on 103 peanut accessions from the US mini-core collection identified 90 significant SNPs linked to traits like lateral branch angle, main stem height, lateral branch length, extent radius, and plant type index (Li et al., 2022). A comprehensive GWAS was conducted to uncover the genetic basis of yield components and quality traits in peanuts, utilizing genotype re-sequencing and phenotypic analysis across four distinct environments (Guo et al., 2024). Genome-wide association studies (GWAS) were conducted to identify genetic loci associated with four pod-related traits in peanuts (Zhang et al., 2023b). Nevertheless, few studies have used GWAS to explore the genetic basis of natural

variation in the content of free amino acids in peanut. Identification of more favourable genes and increasing the understanding of the underlying amino acid biosynthetic pathways are the key steps for breeding peanut with high-quality protein.

To better understand the genetic components underlying the natural variations and metabolism of amino acids in peanuts, we quantified amino acids in mature peanut seeds using near-infrared spectroscopy technology in 2017 and 2018. We conducted a GWAS based on the genotype data of 390 peanut accessions to identify loci and candidate genes associated with amino acids. Furthermore, haplotype analysis was performed, and candidate genes were validated using available transcriptome data and RT-qPCR at different developmental stages. Overall, the data acquired in this study provide abundant genetic information for studying key genes involved in the amino acid metabolic pathway in peanut. These results provide new insights for understanding amino acids biosynthesis and thus enhancing the breeding of high-nutrition peanuts. Moreover, this study aims to clarify the genetic basis of amino acids variations to enhance peanut breeding strategies.

2 | MATERIALS AND METHODS

2.1 | Plant materials

The 390 accessions were gathered from prominent nations worldwide that cultivate peanuts, such as China, the USA, and India. They were then preserved at the Crops Research Institute of the Guangdong Academy of Agricultural Sciences, located in Guangzhou, China. We first cultivated all accessions in 2016 (early (March–July) and late (August–December) growth seasons) to guarantee the genetic purity of each accession, and we subsequently harvested pods of each accession from separate plants. The accessions were grown at the experimental site of the Guangdong Academy of Agricultural Sciences in Guangzhou, China (23° 7′ 44.3784″ N and 113° 15′ 11.7000″ E), throughout the two growing seasons of 2017 and 2018. To extract DNA, tender leaves from individual seedlings were immediately gathered and frozen in liquid nitrogen. Using the CTAB technique, 1.5 μg of total genomic DNA was extracted from each sample. Whole-genome sequencing libraries were generated using the TruseqNano DNAHT Sample Preparation Kit (Illumina), after which index codes were added to attribute sequences to distinct samples. The libraries were sequenced using the Illumina HiSeq X Ten platform, yielding a 150-bp read length and a total of 1.29×10^{13} bases (Lu et al., 2024).

2.2 | Determination of amino acids contents by near-infrared spectroscopy

Amino acids concentrations were assessed from three independent 25 g samples of whole seeds. This analysis utilized near-infrared spectroscopy technology, specifically the Perten NIR DA 7250 model from Perten Instruments (Zhang et al., 2018). The NIR device underwent calibration using the manufacturer's standard calibration module, which was developed based on an extensive set of 390 samples. Following the scanning process, amino acid concentrations were reported relative to the dry matter content of the samples. These values were subsequently standardized and expressed as grams per kilogram of total protein (g kg^{-1} TP) (Malle et al., 2020).

2.3 | Association studies

The genome-wide association study (GWAS) included an examination of the contents of 10 specific amino acids, alongside Best Linear Unbiased Prediction (BLUP) values, in relation to 2,564,993 single nucleotide polymorphisms (SNPs) that had been filtered to include only those with a minor allele frequency (MAF) greater than 0.05. BLUP values for each trait across various environments were calculated using the BLUP algorithm within the lme4 package, available on CRAN at https://cran.r-project.org/web/packages/lme4/, utilizing R software version 4.0.2. The association analysis was conducted using the Efficient Mixed-Model Association eXpedited (EMMAX) program (Kang et al., 2010). Already available resequencing data published recently by our group was used in the current study (Lu et al., 2024). A threshold for whole-genome significance was established at a level of $-\log_{10}(P) \geq 6$, which was derived using the Bonferroni correction method to adjust for multiple comparisons (Moran, 2003). Significant associated SNPs were subjected to an in-depth analysis process, beginning with the detection of GWAS-associated signals that met the established threshold. Linkage disequilibrium (LD) block analyses were utilized to delimit candidate association regions (Shin et al., 2006). The computation of the LD block was carried out using the LDheatmap package in R version 4.0.2. Further analysis focused on significant nonsynonymous SNPs, using annotated variations to inform the investigation. Additionally, each candidate gene that harboured a nonsynonymous significant SNP underwent functional annotation and homology analyses to elucidate potential biological significance and relationships.

2.4 | Expression analysis of candidate genes and validation via qRT–PCR

Already available transcriptome datasets published by our research group were utilized for the expression analysis of identified candidate genes (Lu et al., 2024). Expression dynamics of candidate genes were studied in shells and seeds of different peanut varieties and at different developmental stages to gain deeper insights into the role of identified genes. To confirm the expression levels of these candidate genes, a qRT–PCR analysis was conducted in the seeds at different developmental stages. Total RNA was extracted using a Plant RNA Extraction Kit (TIANGEN, DP432) and reverse transcribed into cDNA with the PrimeScript-RT Reagent Kit (Takara, KR116) according to the manufacturer's instructions. The qRT–PCR assay was performed in triplicate using SYBR Green Master Mix (Yeasen, 11203ES). Target gene expression levels were calculated using the comparative $2^{-\Delta\Delta Ct}$

method (Pfaffl, 2001). *Ah18S* was kept as the internal control. Primers were designed with Primer 3 (v4.1.0) (Untergasser et al., 2012) and are listed in Table S1.

2.5 | Statistical analyses

Statistical analyses were carried out using R software, version 4.2.0. The comparison of gene expression levels and phenotypic values between two distinct sample groups was executed using a two-tailed Student's *t*-test. This test was performed within the R package ggsignif (Ahlmann-Eltze and Patil, 2021), which provides functionalities for adding statistically significant comparisons directly to plots created with ggplot2 (Lu et al., 2024).

3 | RESULTS

3.1 | Variations among the amino acids across the years 2017 and 2018

A notable variation in amino acids contents was observed, reflecting seasonal and annual environmental impacts. In 2017, arginine (Arg) content in the spring (2017E) was significantly higher compared to the autumn sowing (2017 L). For methionine (Met), the levels were relatively low across all samples, yet there was a slight increase in the spring of 2018 (2018E). Histidine (His) showed more pronounced content in the 2017E sowing compared to 2017 L, with a decrease in the following year, especially in the autumn sowing (2018 L). Phenylalanine (Phe) levels were higher in the 2017E planting, decreased in 2017 L, and remained relatively stable in 2018. Isoleucine (Isl) content presented a decrease from spring to autumn in both years, with the highest level in 2017E and the lowest in 2018 L. Proline (Pro), known for its role in stress responses, showed an irregular pattern with a notable decrease in the autumn of 2017. Leucine (Leu) displayed a higher content in the spring sowings of both years compared to autumn, with 2017E having the most significant content. Threonine (Thr) showed a consistent decrease from spring to autumn in both years. Lysine (Lys), essential for human nutrition, had its highest content in the spring of 2017, with a decrease in autumn sowings. Lastly, valine (Val) exhibited a gradual decrease from spring to autumn across the two years, with the highest content in 2017E and the lowest in 2018 L (Figure 1). Frequency distributions are shown in Figure S1. Overall, the amino acids contents in peanuts exhibited a trend of higher levels in the spring than in autumn sowings; there were also variances between the years. This suggests that both the time of sowing and the annual climatic conditions significantly affect the nutritional profile of peanuts, especially in terms of amino acids composition.

3.2 | Country-wise climate variable trends for the years 2017 and 2018

Climate data from WorldClim (https://www.worldclim.org) was utilized to observe the variations among different climate variables across different countries (Figure S2 and File S1). In 2017, China, India, and Japan recorded the highest maximum temperatures, exceeding 30°C. This trend continued in 2018 with slight increases in maximum temperatures in some regions. China and Japan had some

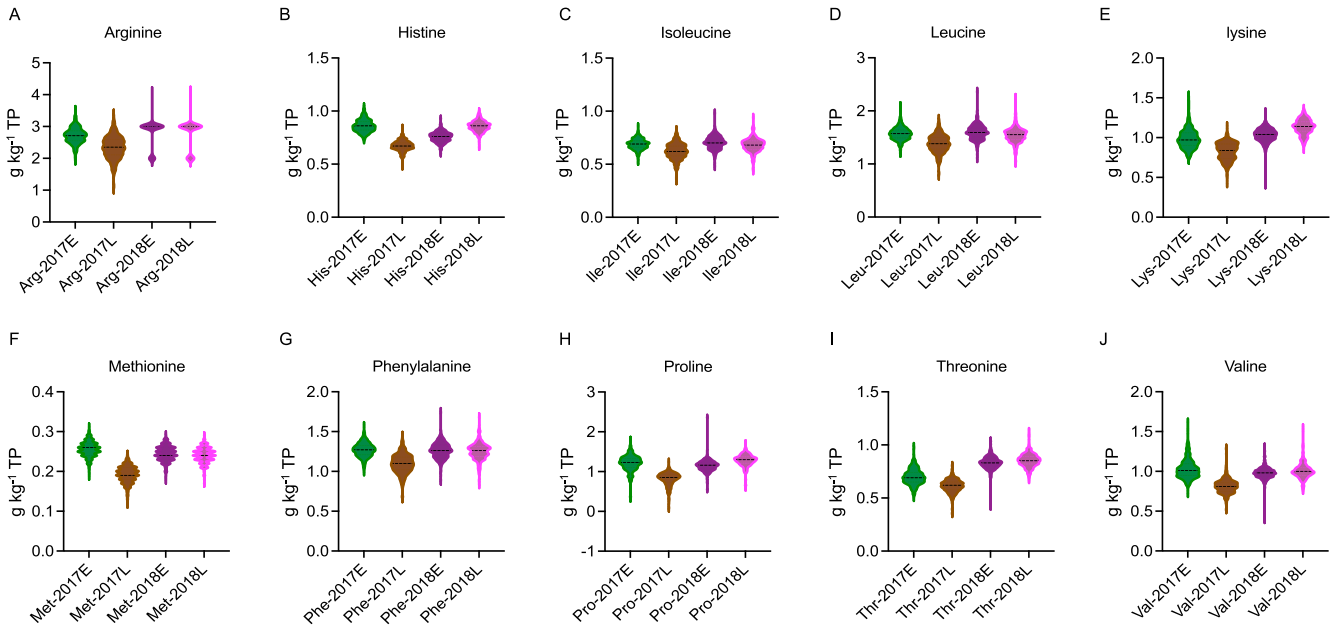

FIGURE 1 Seasonal and annual variations in amino acids contents of peanuts. This figure illustrates the amino acids contents (g kg^{-1} TP) in peanuts harvested from two sowing seasons, spring (March–July) (E) and autumn (August–December) (L), over two consecutive years, 2017 and 2018. Each bar represents the mean value of amino acid content (g kg^{-1} TP) for Arginine (Arg-), Methionine (Met-), Histidine (His-), Phenylalanine (Phe-), Isoleucine (Ile-), Proline (Pro-), Leucine (Leu-), Threonine (Thr-), Lysine (Lys-), and Valine (Val-) during the specified season and year.

of the lowest minimum temperatures in 2017, dropping below −20°C. In contrast, India and Indonesia maintained milder minimum temperatures, with lows above 0°C. These patterns persisted in 2018, with only slight changes in temperature. Precipitation levels varied across countries in 2017. Japan and Indonesia saw higher precipitation, while China and India had significant variability, ranging from low to high values. In 2018, most countries experienced increased precipitation, with Japan and Indonesia continuing to receive high levels. Relative humidity was highest in Indonesia in 2017, with values above 90%. China and India showed lower humidity levels with more fluctuations. These trends remained consistent in 2018, with slight increases in many regions. Across 2017 and 2018, China, India, and Japan consistently experienced extreme temperatures, both high and low, while Indonesia had more moderate and stable climates.

3.3 | Correlations among climate variables (maximum temperature, minimum temperature, precipitation, and humidity) and amino acids for the years 2017 and 2018

The analysis of the correlations between climate variables (maximum temperature at 2 m, minimum temperature at 2 m, total precipitation, and relative humidity at 2 m whereas "2 m" means the interpolated value from the sea level or land surface) and amino acid concentrations reveals distinct year-wise impacts, highlighting how different environmental conditions influenced amino acid levels across 2017 and 2018 (Figure 2).

In 2017, both spring (2017E) and autumn (2017 L) periods showed generally weak correlations between climate variables and amino acids. However, minimum temperature at 2 m consistently showed a positive correlation with several amino acids like lysine, threonine, and histidine. This suggests that in 2017, warmer minimum temperatures may have had a modest but consistent effect on enhancing the biosynthesis or accumulation of these amino acids. The correlation between total precipitation and amino acids like arginine and isoleucine was slightly positive in 2017E, indicating that higher rainfall might have supported the production of certain amino acids during this period. By 2017 L, relative humidity began to show some influence, with a weak positive correlation with histidine and a slight negative correlation with proline, hinting that humidity levels may have impacted amino acid synthesis differently as the year progressed. The year 2018 showed a noticeable shift in the impact of climate variables on amino acid levels. In spring 2018 (2018E), the minimum temperature at 2 m exhibited stronger positive correlations with amino acids such as isoleucine and leucine compared to 2017. This suggests that the warmer minimum temperatures in the early months of 2018 had a more pronounced effect on the levels of these

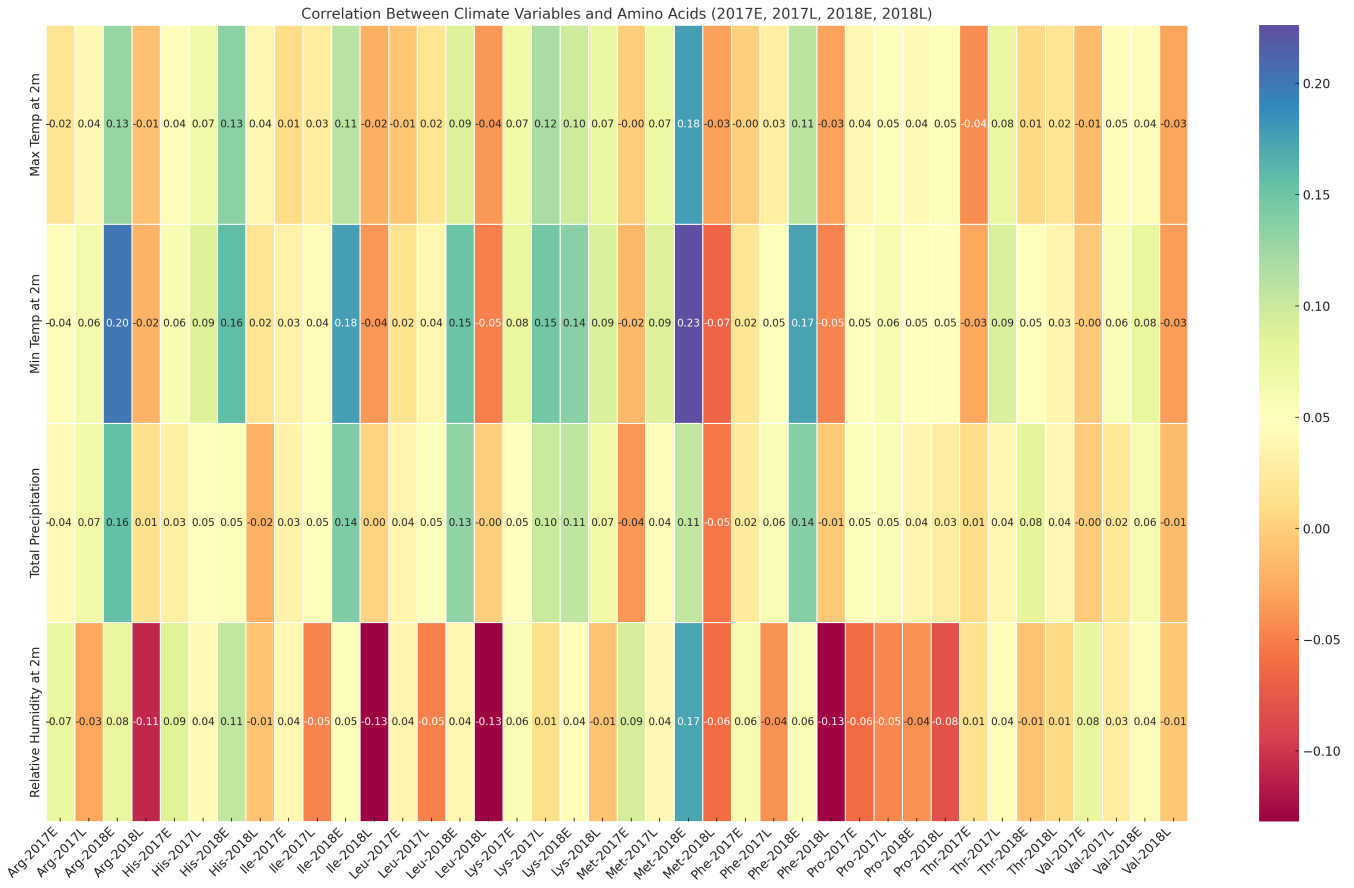

FIGURE 2 Correlation between climate variables and amino acids across 2017E, 2017 L, 2018E, and 2018 L.

amino acids, possibly due to different growing conditions or plant responses compared to the previous year. However, the maximum temperature at 2 m showed weaker correlations in 2018E, indicating that maximum temperatures were less influential during this period. By autumn 2018 (2018 L), the correlations shifted again. The positive influence of minimum temperature persisted but was less strong than in spring 2018, reflecting a diminishing impact as the year progressed. Total precipitation and relative humidity showed weak negative correlations with several amino acids, including proline and threonine. This suggests that in autumn 2018, higher moisture levels may have slightly suppressed the synthesis or stability of certain amino acids, potentially due to increased stress or changes in metabolic priorities in response to environmental conditions.

3.4 | Geographical distribution of peanut accession based on amino acids contents

The geographical distribution of ten essential amino acids was analyzed to check variations among the accessions collected from various regions for 2017 and 2018 during the spring and autumn seasons (Figure 3). The study revealed that valine and arginine typically have higher concentrations, suggesting robust synthesis and essential roles in plant growth and stress responses. In contrast, methionine and histidine generally show lower concentrations, likely due to environmental constraints. These findings underscore the significant impact of genetic and environmental interactions on amino acids production in peanuts. Overall, we observed that accessions belonging to northern China possess higher amino acids contents than accessions collected from southern China, particularly arginine and leucine. Such insights could guide targeted agricultural enhancements and further research to optimize nutritional outcomes.

3.5 | Descriptive statistics for the amino acids contents in 2017 and 2018

In analysing ten amino acids in peanuts over 2017 and 2018, arginine and leucine emerged as the amino acids with the highest average contents (Table S2). Arginine consistently showed the highest mean values, such as 2.71 g kg^{-1} TP in early 2017 and 2.73 g kg^{-1} TP in early 2018, indicative of its critical role in protein synthesis and nitrogen metabolism in peanuts. Leucine also exhibited high mean values, around 1.57 g kg^{-1} TP in early 2017 and 1.60 g kg^{-1} TP in early 2018, reflecting its importance in growth and nitrogen balance. On the lower end, methionine and histidine had the smallest mean values among the studied amino acids. Methionine recorded the lowest, with

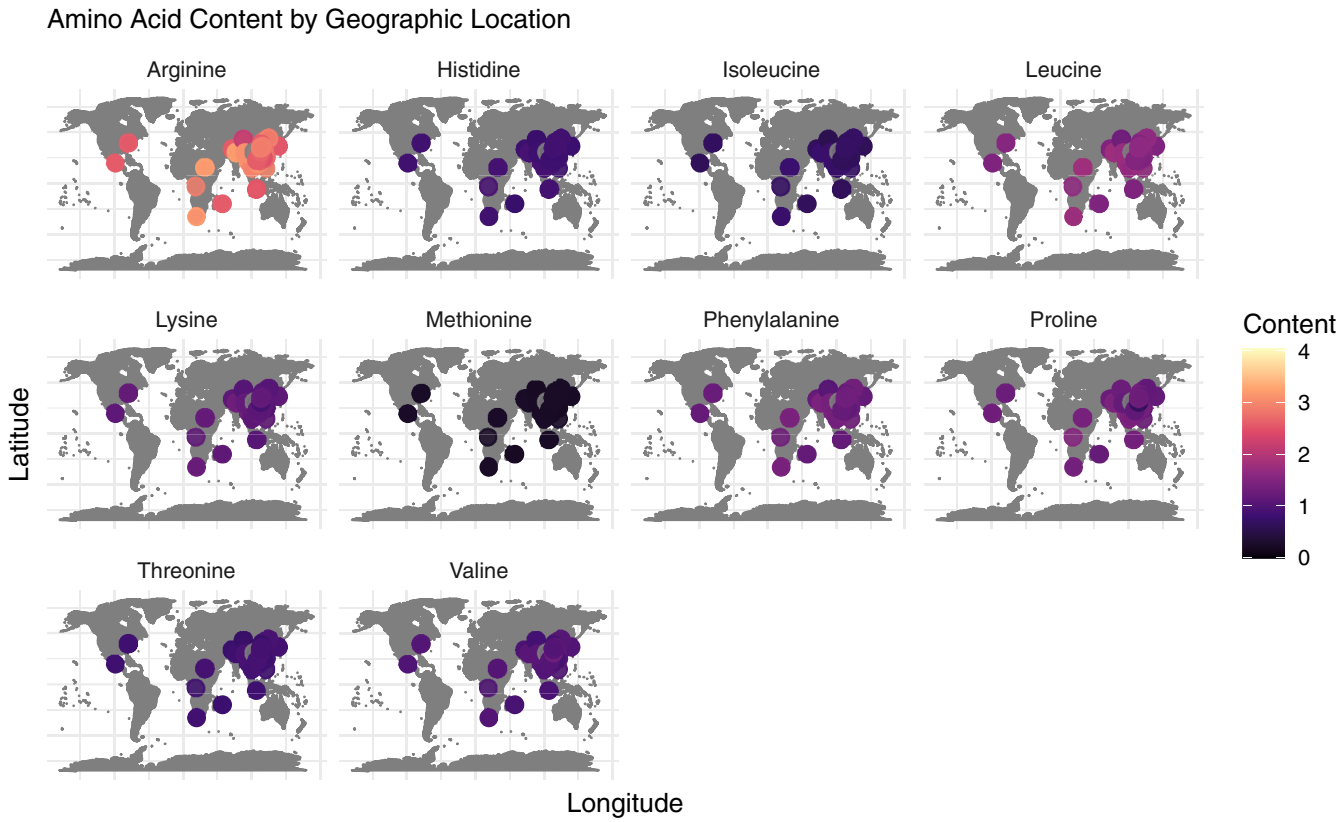

FIGURE 3 Geographic distribution of essential amino acids contents in peanuts. Each subplot represents the content of a specific amino acid, with the x-axis denoting longitude and the y-axis denoting latitude. The colour intensity of the circles corresponds to the concentration of the respective amino acids, as indicated by the colour bar on the right. Higher concentrations are represented by larger circles with warmer colours (yellow to orange), while lower concentrations are shown by smaller circles with cooler colours (purple to dark purple).

values like 0.26 g kg^{-1} TP in early 2017 and 0.24 g kg^{-1} TP in early 2018, possibly due to its metabolic cost and environmental factors affecting its synthesis. Histidine followed, with figures such as 0.86 g kg^{-1} TP in early 2017 and 0.76 g kg^{-1} TP in early 2018, reflecting lower synthesis rates or availability. This detailed profiling of amino acids helps underscore the nutritional variances in peanuts and can guide targeted enhancements in peanut breeding and cultivation to optimize amino acids contents.

3.6 | Descriptive statistics for the amino acids contents based on average values

In this study, we comprehensively analysed the average concentrations of ten amino acids (Table 1). Phenylalanine exhibited the highest mean concentration at 1.22 g kg^{-1} TP with a standard deviation of 0.13, while isoleucine had the lowest mean concentration at 0.67 g kg^{-1} TP) with a standard deviation of 0.07. Arginine displayed notable variability in concentration across samples, with the highest standard deviation of 0.34 and the largest range of 2.93. The skewness of the data indicated that methionine, proline, and isoleucine were predominantly skewed towards lower values, with skewness coefficients of −0.5973, −0.5943, and − 0.4822, respectively. Conversely, valine showed a positive skewness of 0.2952, suggesting a distribution leaning towards higher values. Notably, leucine and valine presented high kurtosis values of 1.4541 and 1.6551, respectively, indicating a leptokurtic distribution that suggests the presence of outliers or extreme values. These descriptive statistics elucidate the varying concentrations and distributions of amino acids, highlighting their diverse biological functionalities and metabolic significance.

3.7 | Correlation analysis and hierarchical clustering and multidimensional scaling for the average of 2017 and 2018

The correlation between various amino acids in peanuts across different seasons highlights both strong and moderate correlations (Figure 4A). Notably, methionine and phenylalanine show an exceptionally high correlation (0.98***), suggesting a significant co-variation in their seasonal profiles, which could indicate linked biosynthetic pathways. Similarly, valine and isoleucine also exhibit a very strong correlation (0.99***), reinforcing the idea of shared metabolic processes. Lysine and leucine, crucial for nutritional quality, correlate strongly (0.98***), hinting at their synchronized behaviour in response to environmental or genetic factors. On the other hand, proline and threonine show a moderately strong correlation (0.70***), suggesting lesser but significant co-dependence. However, negative correlations, especially involving histidine, might indicate competitive metabolic interactions with other amino acids. A detailed understanding of amino acids correlations is critical for developing peanut varieties with optimized nutritional profiles, considering the complex interplay of environmental, genetic, and agronomic factors. The dendrogram after K-means clustering displays the genotypes organized by similarity in amino acids composition, with the vertical axis indicating the clustering distance and four distinct clusters highlighted in different colours (blue, orange, yellow, red) (Figure 4B-C). The scatter plot on the right employs a dimensionality reduction technique to project the high-dimensional amino acids data onto two principal components, which explain 23.3 and 8.2% of the variance, respectively.

3.8 | Genome-wide association study for amino acids contents

A total of 88 single nucleotide polymorphisms (SNPs) spanning various chromosomes were identified across multiple environments, each correlating with different amino acids (Figure S3). Within the detected SNPs, those exhibiting the most significant association indicated by the lowest *p*-value within their respective genetic regions were designated as the putative causal SNPs for the corresponding amino acids. The pertinent details of these SNPs are compiled in Table 2.

The investigation revealed that SNP A09_3002690, was associated with the amino acid arginine. Additionally, three distinct SNPs, A05_95934413, A10_96970442, and B01_26348752 were

TABLE 1 Descriptive statistics for the average amino acids contents in the years 2017 and 2018.

Trait	Mean	SD	Median	Trimmed	Mad	Min	Max	Range	Skew	Kurtosis	Se
Phenylalanine	1.2191	0.1262	1.23	1.226	0.1186	0.68	1.74	1.06	−0.526	0.9508	0.0033
Methionine	0.2327	0.0314	0.24	0.2348	0.0297	0.12	0.31	0.19	−0.5973	−0.093	0.0008
Arginine	2.6112	0.3443	2.64	2.6271	0.3113	1.1	4.03	2.93	−0.4666	1.0084	0.009
Lysine	0.9929	0.1531	1	0.9976	0.1483	0.41	1.51	1.1	−0.2657	0.0608	0.004
Leucine	1.5204	0.165	1.53	1.5268	0.1483	0.8	2.36	1.56	−0.3192	1.4541	0.0043
Proline	1.1091	0.2481	1.16	1.1246	0.2372	0.08	2.36	2.28	−0.5943	0.9101	0.0065
Threonine	0.7488	0.1186	0.77	0.7512	0.1334	0.35	1.13	0.78	−0.1703	−0.6605	0.0031
Valine	0.9575	0.1313	0.97	0.9555	0.1038	0.38	1.6	1.22	0.2952	1.6551	0.0034
Isoleucine	0.6728	0.0703	0.68	0.676	0.0593	0.35	0.98	0.63	−0.4822	1.2124	0.0018
Histidine	0.7858	0.0939	0.79	0.7869	0.1038	0.47	1.04	0.57	−0.1226	−0.602	0.0024

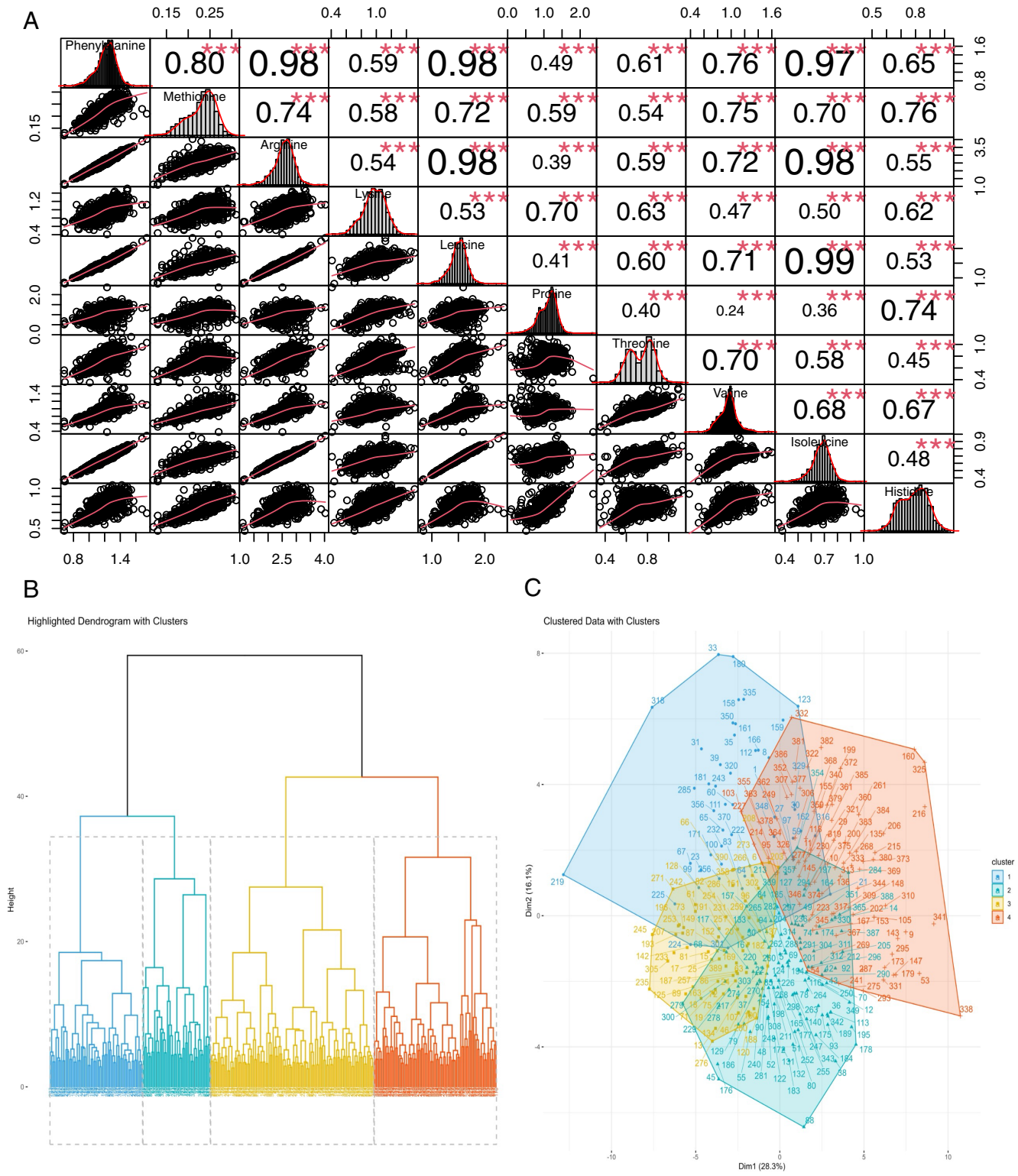

FIGURE 4 **A-** Seasonal correlation patterns of amino acids contents in peanut genotypes. **B-C-** The K-means clustering of amino acids contents in 390 peanut genotypes using a hierarchical clustering dendrogram and a scatter plot of clustered data.

each linked with histidine. Two SNPs, A09_3002690 and B06_2411068, found on chromosomes A09 and B06, respectively, were associated with isoleucine. The same SNP A09_3002690 was also associated with both leucine and phenylalanine. A trio of SNPs, A01_11846565, A01_11979142, and 09_23699956, across chromosomes A01 and B09 correlated with lysine. Methionine was associated with SNPs A08_47851652 and B05_136041542 on chromosomes A08 and B05, respectively.

TABLE 2 Significant SNPs linked to amino acids contents in the merged phenotype.

Trait	SNP	Chromosome	Position	$-\log_{10}(P)$
Arginine	3002690	A09	923614537	6.01857728
Histine	95934413	A05	577102692	6.59545709
	96970442	A10	1139748186	6.05837372
	26348752	B01	1185673334	6.34053518
Isoleucine	3002690	A09	923614537	6.61060555
	2411068	B06	1854155287	6.11998405
Leucine	3002690	A09	923614537	6.62614908
Lysine	11846565	A01	11846565	6.6963588
	11979142	A01	11979142	6.22108355
	23699956	B09	2220405569	6.21217573
Methionine	47851652	A08	918000863	6.49669171
	136041542	B05	1832230786	6.30317256
Phenylalanine	3002690	A09	923614537	6.43448573
Proline	23097364	A01	23097364	6.63368434
	44159780	A01	44159780	6.65473545
	44160492	A01	44160492	6.65851559
	62078394	A01	62078394	6.63784364
	62078488	A01	62078488	6.0018195
	62078492	A01	62078492	6.0018195
	62078493	A01	62078493	6.0018195
	105876532	A01	105876532	6.61363315
	5430835	A02	117320153	6.26955358
	114031025	A02	225920343	6.65333104
	103955842	A03	330213292	6.57661904
	110932354	A03	337189804	6.6361432
	121843395	A03	348100845	6.62169488
	121843409	A03	348100859	6.62800218
	11751219	A05	492919498	6.63831215
	14446421	A06	598026658	6.63227871
	89271861	A06	672852098	6.63655505
	97354186	A06	680934423	6.64544655
	97354212	A06	680934449	6.64228727
	97357479	A06	680937716	6.64039536
	38606279	A07	740594169	6.64182538
	51302216	A07	753290106	6.61386775
	51302241	A07	753290131	6.60403278
	129877213	A07	831865103	6.64039676
	24071558	A08	894220769	6.60155528
	24071561	A08	894220772	6.60155528
	18555264	A09	939167111	6.64039536
	18555274	A09	939167121	6.64039536
	47103224	A10	1089880968	6.10815193
	67245919	A10	1110023663	8.29191295
	67245923	A10	1110023667	7.04192614
	104920451	B01	1264245033	6.65217638
	104920914	B01	1264245496	6.63090074

(Continues)

TABLE 2 (Continued)

Trait	SNP	Chromosome	Position	$-\log_{10}(P)$
	104921027	B01	1264245609	6.64204955
	104941636	B01	1264266218	6.65279467
	137237947	B01	1296562529	6.62257053
	137239784	B01	1296564366	6.64092835
	137247597	B01	1296572179	6.63802254
	137337934	B01	1296662516	6.64565841
	137354000	B01	1296678582	6.6413597
	24576148	B02	1334676863	6.62938387
	24576160	B02	1334676875	6.62938387
	24576163	B02	1334676878	6.62938387
	33791822	B02	1343892537	6.65328537
	115234474	B03	1537509236	6.63635773
	20409426	B04	1589240277	6.60257378
	42017239	B06	1893761458	6.64039536
	42501661	B06	1894245880	6.61494468
	42501678	B06	1894245897	6.61494468
	42501685	B06	1894245904	6.61494468
	42649490	B06	1894393709	6.32596454
	42649497	B06	1894393716	6.31426208
	42652280	B06	1894396499	6.65941625
	42652282	B06	1894396501	6.65941625
	42652284	B06	1894396503	6.65941625
	42652289	B06	1894396508	6.65941625
	43080064	B06	1894824283	6.64167192
	34173323	B09	2230878936	6.62547327
	130116943	B09	2326822556	6.6322761
	22762189	B10	2382906726	6.64216162
	4933875	Scaffold2	2528013710	6.64544655
Threonine	37933248	A05	519101527	6.35915134
	4477786	B03	1426752548	6.79484194
	4656290	B03	1426931052	6.37906488
	4830958	B03	1427105720	7.42425531
	3963194	B05	1700152438	6.0646203
Valine	37933248	A05	519101527	7.04841828
	47232359	A05	528400638	6.54406375
	74804143	A05	555972422	6.19832985
	76155792	A05	557324071	6.24106127
	81906059	A05	563074338	6.45490627
	84531652	A05	565699931	6.00912276
	92136535	A05	573304814	6.62376905
	109324934	A06	692905171	6.11422206
	116553975	A07	818541865	7.25075018

Remarkably, sixty-one SNPs were linked to proline, distributed over a broad array of chromosomes: A01, A02, A03, A05, A06, A07, A08, A09, A10, B01, B02, B03, B04, B06, B10, and within a scaffold region. Five SNPs, A05_37933248, B03_4477786, B03_4656290, B03_4830958, and B05_3963194, spanning chromosomes A05, B03, and B05 were found to be associated with threonine. For valine, nine

TABLE 3 List of SNPs identified as common among different amino acids.

Amino acid	SNP	Chromosome No.	Position	$-\log_{10}(P)$
Arginine	3002690	A09	923614537	6.018577281
Isoleucine	3002690	A09	923614537	6.610605546
Leucine	3002690	A09	923614537	6.626149084
Phenylalanine	3002690	A09	923614537	6.434485731
Threonine	37933248	A05	519101527	6.359151337
Valine	37933248	A05	519101527	7.048418284

SNPs, A05_37933248, A05_47232359, A05_74804143, A05_76155792, A05_81906059, A05_84531652, A05_92136535, A06_109324934, and A07_116553975, were significant and traced to chromosomes A05, A06, and A07. The results further suggested a commonality with SNP A09_3002690, which was found to be related to arginine, isoleucine, leucine, and phenylalanine on chromosome A09 (Table 3). Additionally, SNP A05_37933248 was another notable SNP common to both threonine and valine on chromosome A05. This comprehensive mapping of SNPs to specific amino acids paves the way for a deeper understanding of genetic influences on amino acids variations in peanuts.

3.9 | Candidate genes mining

To identify potential candidate genes in these loci, the linkage disequilibrium (LD) blocks around the significant SNPs were calculated for candidate genes. The results showed that the peak common SNP (A09_3002690) associated with arginine, isoleucine, leucine, and phenylalanine was mainly located in an LD block region from 2993205 to 3040091 bp (Figure 5A), containing four genes (*Ahy_A09g041582*, *Ahy_A09g041583*, *Ahy_A09g041585* and *Ahy_A09g041586*). To identify candidates, we functionally annotated the peanut genes using orthologs to first determine the likelihood that the gene could participate in pathways determining these traits. The results did not show any matches for *Ahy_A09g041583*, *Ahy_A09g041585*. Whereas *Ahy_A09g041582* was annotated as laccase-15-like "*LAC15*" and *Ahy_A09g041586* was annotated as geraniol 8-hydroxylase "C76B6".

3.10 | Expression profiling of identified candidate genes

In this study, we leveraged previously collected transcriptome data from our group to delve deeper into the expression profiles of selected candidate genes potentially involved in pod and shell development in peanuts. This dataset includes transcriptome-based expressions at 11 different developmental stages, as detailed in Chen et al. (2016) for the shell and Xiaoping Chen (unpublished) for the seed. Notably, the expression of *Ahy_A09g041582* (*LAC15*) increased significantly at later stages, with lower expression levels at the initial P0–P1 stages and notably higher levels during the P2SH-P6SH stages, suggesting its crucial role in pod swelling processes (Figure 6A). Furthermore, our investigation extended to the transcriptome data of two peanut cultivars, H176 (Kainong176) and L70 (Kainong70), characterized by varying oil contents in their mature seeds—over 70% in Kainong176 and less than 40% in Kainong70. Analysis of expression patterns at six seed developmental stages revealed that *Ahy_A09g041582* exhibits consistently higher expression levels than other candidate genes, reinforcing their potential involvement in the seed development (Figure 6B). Additionally, we analyzed RNA-seq data for seeds and shells of peanut varieties with large and small pods across five developmental stages (Haifen Li, unpublished). This analysis confirmed that *Ahy_A09g041582* manifest higher expression levels at various developmental stages when compared to other candidate genes (Figure 6C and D). Further, the qRT-PCR results in the seeds at different developmental stages (S0, S1, S2, S3, S4, and S5) also confirms that *Ahy_A09g041582* had higher expressions at later developmental stages (Figure 6E and F). Based on these observations, we hypothesize that this gene is the key contributor to peanut growth and development. Given the significant roles that *Ahy_A09g041582* appears to play, further research is essential.

3.11 | Exonic variations and haplotype analysis of *Ahy_A09g041582 (LAC15)*

Exon variations revealed that *AhLAC15* contained three non-synonymous SNPs, i.e. 3002787 (A/G), 3002808 (A/C) and 3002827 (G/C) (Figure 7A). Furthermore, analysis of haplotype frequencies revealed compelling insights. Three dominant haplotypes were identified, Hap1, Hap2 and Hap3 emerged as the dominant haplotypes with frequencies of 44.6, 36.2 and 9.8%, respectively (Figure 7B and C). Interestingly, accession with Hap3 has higher arginine, isoleucine, leucine, and phenylalanine contents (Figure 7D).

4 | DISCUSSION

Amino acids are widely used in the animal feed industry (Panthee et al., 2006, Fliege et al., 2022), with the poultry and swine industries consuming over 400,000 mt of Lys (Panthee et al., 2006) and spending approximately USD 100 million annually to supplement feed with synthetic Met (Panthee et al., 2004). Peanuts contain not only essential amino acids but also a large amount of unsaturated fatty acids; thus, they are widely consumed for health purposes. As a result, many

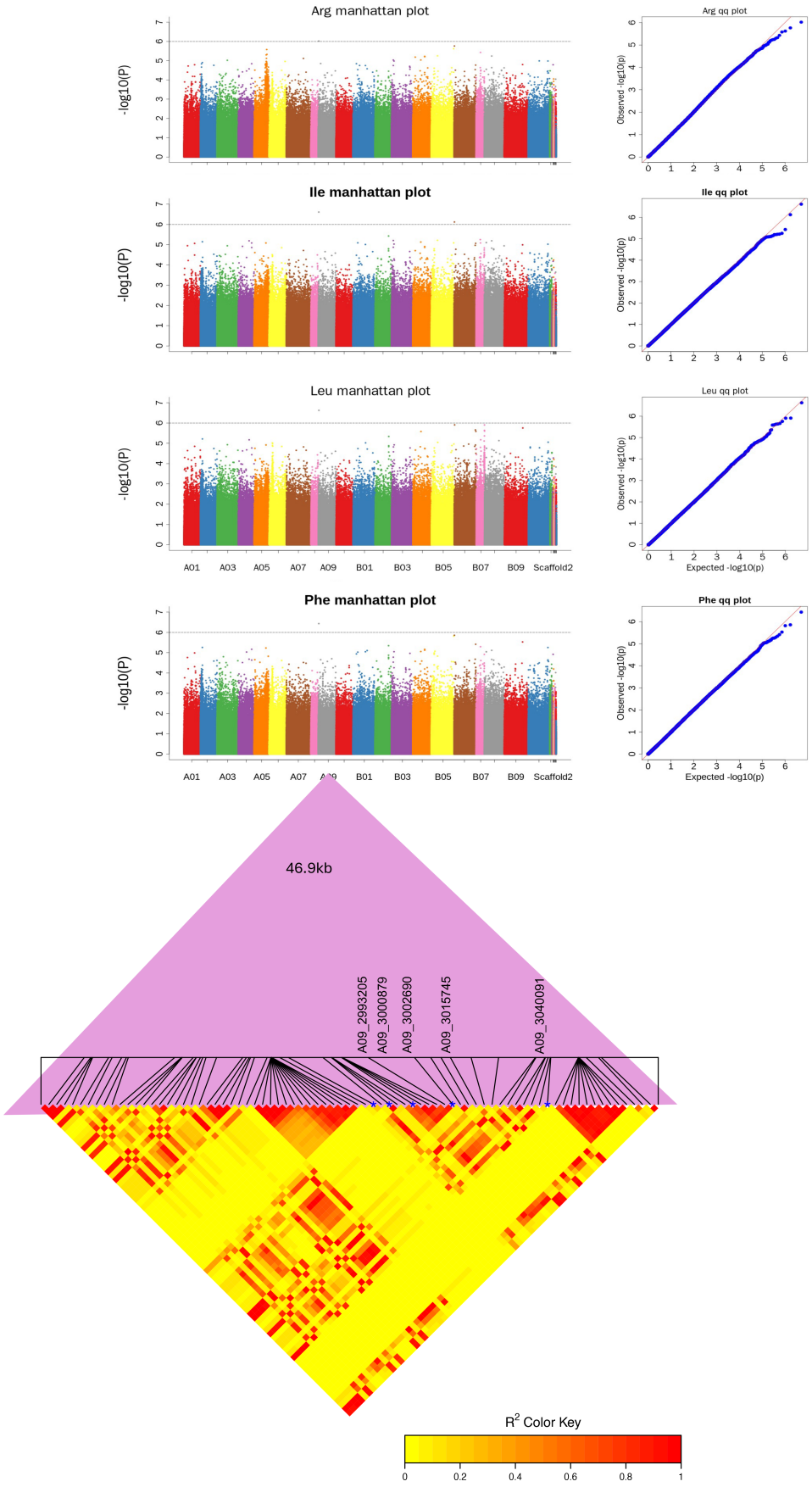

FIGURE 5 Local manhattan plots for amino acids (arginine, leucine, isoleucine and phenylalanine)-related genes on chromosome A09 and LD heatmap for the candidate region within the peak region of Ahy-3002690.

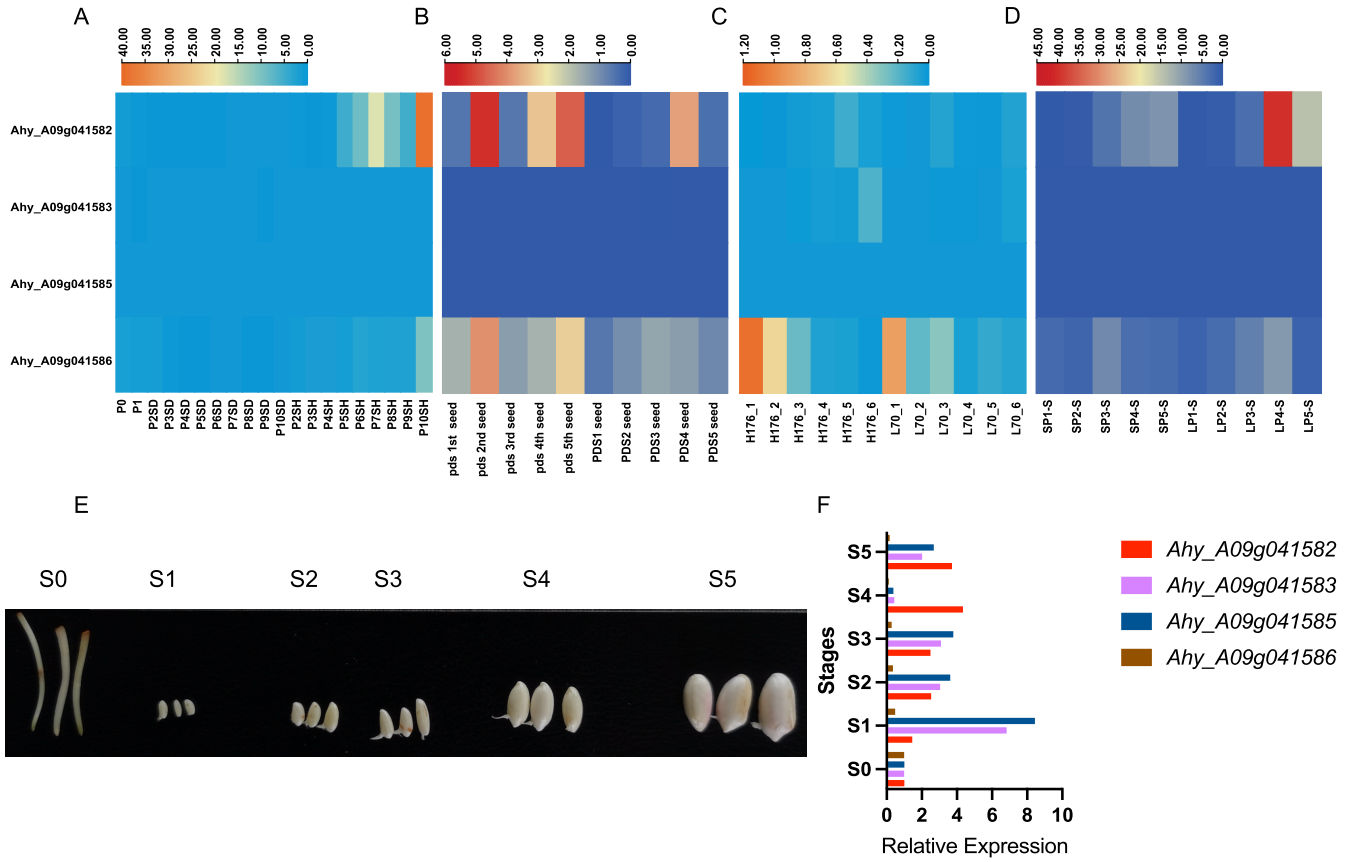

FIGURE 6 Expression profiling of candidate genes at different development stages of seed and shell. **A**- RNA-seq for shell (Chen et al., 2016) and seed (Lu et al., 2024) at 11 different development stages, respectively. **B**- Transcriptome based expression in high and low OA peanut varieties (Liu et al., 2019). **C-D**- RNA-seq for seed and shell of large- and small-pod verities at five different development stages, respectively (Haifen Li, unpublished) (Lu et al., 2024). **E-F-** Peanut seeds at different developmental stages and expression analysis of candidate genes via qRT-PCR. SD stands for seed stage, and SH stands for shell stage, with the numbering representing 11 developmental stages. Moreover, pds represents the small pod variety, and PDS represents the large pod variety (1, 2, 3, 4, and 5 indicate different developmental stages). H176 is a high-oleic cultivar, and L70 is a normal-oleic cultivar, both undergoing six seed developmental stages (1, 2, 3, 4, 5, 6). SP stands for small pod, and LP stands for large pod (1, 2, 3, 4, 5 denote different seed developmental stages).

studies have been conducted on QTLs involved in regulating protein and oil content (Yang et al., 2023, Sun et al., 2022, Zhang et al., 2021, Wilson et al., 2017, Sarvamangala et al., 2011, Pandey et al., 2014) but there are no comprehensive reports on the QTLs for amino acids. Herein, we conducted an extensive genome-wide association study for identifying QTLs and key candidates linked to amino acids content and regulation in peanut. The peanut accessions used in this study were collected from major global peanut-growing countries, including India, the USA and China (Lu et al., 2024) which can be utilized for peanut improvement by applying the GWAS to identify useful alleles.

In this study, we conducted GWAS for ten amino acids contents in four different plant environments, identifying dozens of significant associations on different chromosomes. In addition, our findings suggested that arginine is the major amino acid across the germplasm used in this research. The present study analyzed 390 peanut accessions grown for two years (2017 and 2018), spring and autumn, for their contents of ten amino acids. Ten constituent amino acids were analyzed in peanuts over 2017 and 2018, and results showed that arginine and leucine emerged as the amino acids with the highest average contents. Arginine consistently showed the highest mean values, such as 2.71 in early 2017 and 2.73 in early 2018, indicative of its critical role in protein synthesis and nitrogen metabolism in peanuts and it was found that the the correlation between each amino acid was significantly positive. Chotekajorn et al. (2021) analyzed the free amino acids from 316 wild soybean accessions and identified that Arg was the most abundant, while most of the amino acids were positively correlated with each other, similar to the results of this study (Chotekajorn et al., 2021).

In the GWAS results for the content of the ten amino acids, a common SNP A09_3002690 which was found to be related to arginine, isoleucine, leucine, and phenylalanine on chromosome A09. Additionally, SNP A05_37933248 is a noteworthy SNP shared between threonine and valine. We identified three unique SNPs associated with histidine A05_95934413, A10_96970442, and B01_26348752 located on chromosomes A05, A10 and B01, respectively. SNPs A09_3002690 and B06_2411068, located on chromosomes A09 and B06, respectively, were linked to isoleucine. Interestingly, SNP A09_3002690 also correlates with both leucine and phenylalanine. A set of three SNPs,

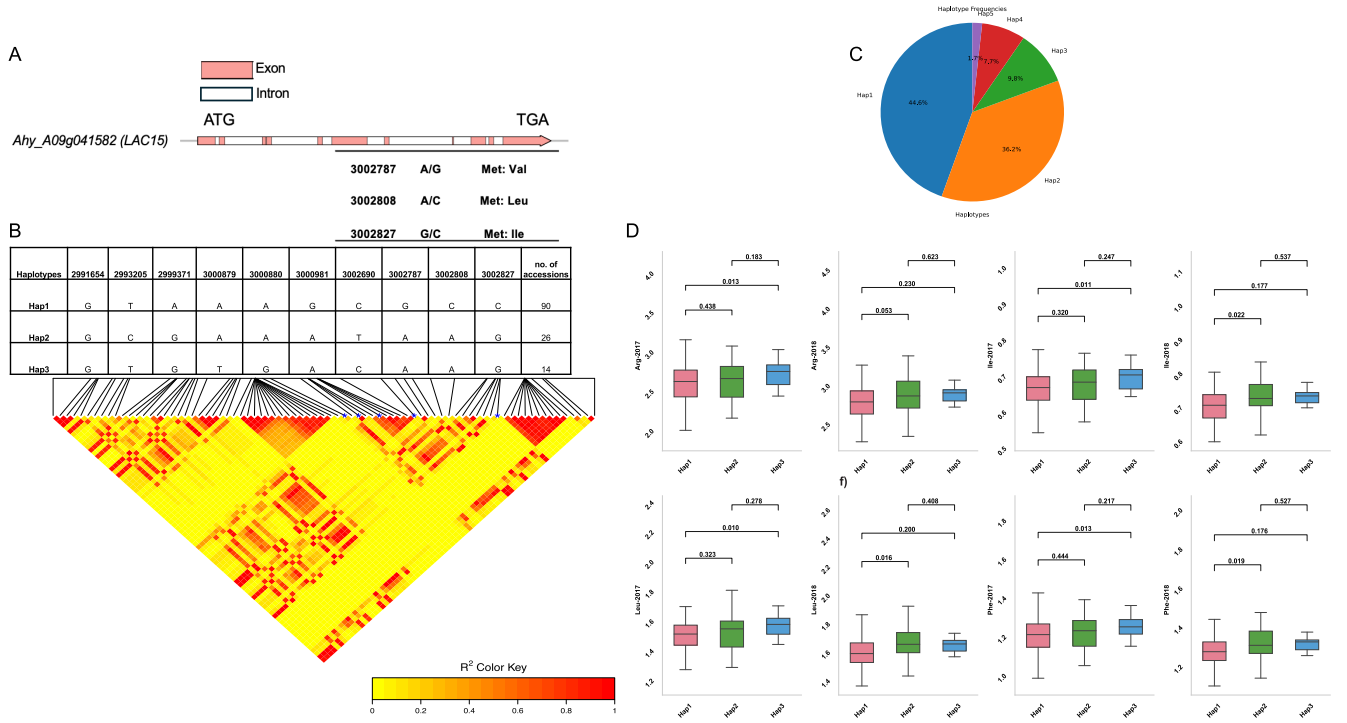

FIGURE 7 **A-** Gene structure analysis, **B-** Haplotype analysis, **C-** Frequency of haplotypes, **D-** frequency of *Ahy_A09g041582* (*LAC15*) haplotypes in peanut accessions based on amino acids contents.

A01_11846565, A01_11979142, and B09_23699956, spread across chromosomes A01 and B09, related to lysine. Methionine was linked to SNPs, A08_47851652 and B05_136041542 on chromosomes A08 and B05, respectively. Previously, Zhang et al. (2018), in their GWAS on amino acid concentrations in soybeans, identified 54 SNPs associated with 18 amino acids. They also reported that 38 of these 54 SNPs were linked to just one amino acid each, highlighting specific genetic influences on individual amino acid synthesis. Furthermore, 11 SNPs were found to be associated with multiple amino acids, ranging from 2 to 12, indicating their broader impact on amino acids metabolism in soybeans (Qin et al., 2019).

In a significant finding, sixty-one SNPs were associated with proline spanning a diverse set of chromosomes: A01, A02, A03, A05, A06, A07, A08, A09, A10, B01, B02, B03, B04, B06, B10, and within a scaffold region. Five SNPs, A05_37933248, B03_4477786, B03_4656290, B03_4830958, and B03_3963194, covering chromosomes A05, B03, and B05, were associated with threonine. For valine, nine SNPs, A05_37933248, A05_47232359, A05_74804143, A05_76155792, A05_81906059, A05_84531652, A05_92136535, A06_109324934, and A07_116553975, were significant and linked to chromosomes A05, A06, and A07. The current comprehensive analysis highlights a diverse array of SNPs across various chromosomes, demonstrating specific associations with key amino acids such as arginine, histidine, isoleucine, leucine, phenylalanine, lysine, methionine, threonine, proline, and valine. These findings not only enrich our understanding of the genetic bases of amino acid synthesis in plants but also open pathways for future research aimed at manipulating these genetic elements for agricultural improvement and enhanced nutritional outcomes.

To mine the candidate genes associated with the common SNP (A09_3002690) we performed LD block analysis and identified four genes (*Ahy_A09g041582*, *Ahy_A09g041583*, *Ahy_A09g041585* and *Ahy_A09g041586*). Conjoint analyses of gene annotation, homologous comparison and exon variation revealed that the *Ahy_A09g041582* (*AhLAC15*) annotated as laccase-15-like "*LAC15*" might be the target gene. In addition, amino acid synthesis involves several complex processes (de la Torre et al., 2014) and can be regulated by shikimate dehydrogenase, chorismate mutase, arogenate dehydratase, asparagine, and aminotransferase (Kim et al., 2023). This research on peanut development harnesses previously collected transcriptome data to analyze expression profiles of genes involved in pod and shell development. The gene *AhLAC15* showed a significant increase in expression during later developmental stages, which could indicate a crucial role in pod swelling (Chen et al., 2016, Lu et al., 2024). Furthermore, the *AhLAC15* showed consistently higher expression across different developmental stages in two peanut cultivars with varying oil contents, suggesting their potential influence on seed development (Liu et al., 2019). Previously, it has been reported that *LAC15* is involved in plant growth, development, and defence responses (Zhang et al., 2019b, Hu et al., 2018, Cheng et al., 2019). The *LAC* genes displayed a variety of temporal and spatial expression patterns (Figure 8), indicating potential functions for these genes in the formation of roots, flowers, and seeds. For instance, *AtLAC15* contributes to proanthocyanidin polymerization

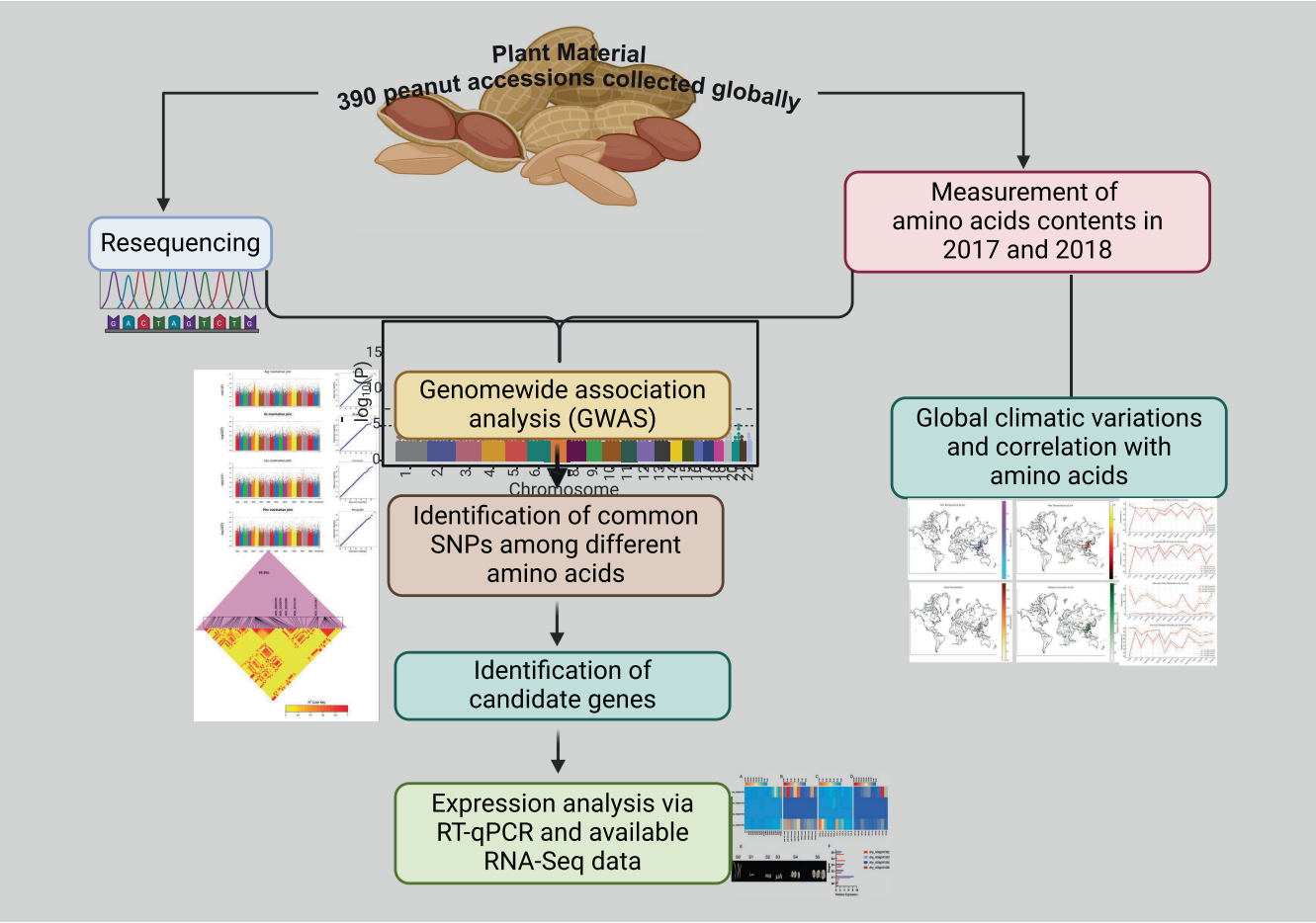

FIGURE 8 An illustrative figure for the current research work.

from its monomer epicatechin and root elongation (Cai et al., 2006). Moving forward, our research will focus on developing molecular markers, including PCR-based assays and targeted region sequencing, to validate this candidate gene within our association panel and other samples. Additionally, we plan to employ gene silencing techniques, such as CRISPR/Cas9, as a method to further confirm the functionality of the candidate gene. This information suggests that the candidate gene detected in the present study may directly or indirectly affect amino acids contents.

AUTHOR CONTRIBUTIONS

Writing—original draft preparation, M.J.U.; methodology, H.L., and M.J.U.; writing—review and editing, R.B.; L.Q.; H.L.; M.K.P.; and R.K.V.; resources, C.X.P.; L.H.; supervision, C.X.P.; Y.H.; and L.Q.; All authors have read and agreed to the published version of the manuscript.

FUNDING INFORMATION

This work was supported by the National Natural Science Foundation of China (32172051 and 32301869), the National Key R&D Program of China (2023YFD1202800), China Agriculture Research System of MOF and MARA (CARS-13), the Open Competition Program of Top Ten Critical Priorities of Agricultural Science and Technology Innovation for the 14th Five-Year Plan in Guangdong Province (2022SDZG05), Guangdong Provincial Key Research and Development Program-Modern Seed Industry (2022B0202060004), Guangdong Science and Technology Plan Project (2023B1212060038), Guangdong Basic and Applied Basic Research Foundation (2023A1515010098), the Special Support Program of Guangdong Province (2021TX06N789), Guangzhou Basic and Applied Basic Research Foundation (202201010281 and 2023A04J0776), Special Fund for High Level Academy of Agriculture Science (R2020PY-JX004, R2020PY-JG005, R2021PY-QY003,R2022YJ-YB3025 and R2023PY-JG007), the Foundation of Director of Crop Research Institute of Guangdong Academy of Agriculture Sciences (202101, 202201 and 202306), the Project of Collaborative Innovation Center of GDAAS(XTXM202203).

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

DATA AVAILABILITY STATEMENT

All the 390 genomic sequence data for GWAS analysis have been deposited in the National Center for Biotechnology Information

(NCBI) database under BioProject number PRJNA776707. The SNP and InDel genotypes have been deposited in Zenodo (Lu et al., 2024) (https://doi.org/10.5281/zenodo.10054109). The published transcriptomic datasets for candidate gene expression analysis can be downloaded from the NCBI Sequence Read Archive under accession numbers SRP167797 and SRP033292 mentioned in the corresponding original literature.

ORCID

Muhammad J. Umer https://orcid.org/0000-0003-2043-0684
Raufa Batool https://orcid.org/0000-0002-7514-9227

REFERENCES

Ahlmann-Eltze, C. & Patil, I. 2021. ggsignif: R Package for Displaying Significance Brackets for'ggplot2'.

Angelovici, R., Lipka, A. E., Deason, N., Gonzalez-Jorge, S., Lin, H., Cepela, J., Buell, R., Gore, M. A. & Dellapenna, D. 2013. Genome-wide analysis of branched-chain amino acid levels in Arabidopsis seeds. *The Plant Cell*, 25, 4827–4843.

Birsoy, K., Wang, T., Chen, W. W., Freinkman, E., Abu-Remaileh, M. & Sabatini, D. M. 2015. An essential role of the mitochondrial electron transport chain in cell proliferation is to enable aspartate synthesis. *Cell*, 162, 540–551.

Cai, X., Davis, E. J., Ballif, J., Liang, M., Bushman, E., Haroldsen, V., Torabinejad, J. & Wu, Y. 2006. Mutant identification and characterization of the laccase gene family in Arabidopsis. *Journal of Experimental Botany*, 57, 2563–2569.

Chen, J., Hu, X., Shi, T., Yin, H., Sun, D., Hao, Y., Xia, X., Luo, J., Fernie, A. R. & He, Z. 2020. Metabolite-based genome-wide association study enables dissection of the flavonoid decoration pathway of wheat kernels. *Plant Biotechnology Journal*, 18, 1722–1735.

Chen, X., Yang, Q., Li, H., Li, H., Hong, Y., Pan, L., Chen, N., Zhu, F., Chi, X. & Zhu, W. 2016. Transcriptome-wide sequencing provides insights into geocarpy in peanut (Arachis hypogaea L.). *Plant biotechnology journal*, 14, 1215–1224.

Cheng, X., Li, G., Ma, C., Abdullah, M., Zhang, J., Zhao, H., Jin, Q., Cai, Y. & Lin, Y. 2019. Comprehensive genome-wide analysis of the pear (Pyrus bretschneideri) laccase gene (PbLAC) family and functional identification of PbLAC1 involved in lignin biosynthesis. *PLoS one*, 14, e0210892.

Chotekajorn, A., Hashiguchi, T., Hashiguchi, M., Tanaka, H. & Akashi, R. 2021. Evaluation of seed amino acid content and its correlation network analysis in wild soybean (Glycine soja) germplasm in Japan. *Plant Genetic Resources*, 19, 35–43.

Crawford, N. M. & Forde, B. G. 2002. Molecular and developmental biology of inorganic nitrogen nutrition. *The Arabidopsis Book/American Society of Plant Biologists*, 1.

De La Torre, F., El-Azaz, J., Ávila, C. & Cánovas, F. M. 2014. Deciphering the role of aspartate and prephenate aminotransferase activities in plastid nitrogen metabolism. *Plant Physiology*, 164, 92–104.

Dunlop, R. A., Main, B. J. & Rodgers, K. J. 2015. The deleterious effects of non-protein amino acids from desert plants on human and animal health. *Journal of Arid Environments*, 112, 152–158.

Filiault, D. L. & Maloof, J. N. 2012. A genome-wide association study identifies variants underlying the Arabidopsis thaliana shade avoidance response. *PLoS genetics*, 8, e1002589.

Fliege, C. E., Ward, R. A., Vogel, P., Nguyen, H., Quach, T., Guo, M., Viana, J. P. G., Dos Santos, L. B., Specht, J. E. & Clemente, T. E. 2022. Fine mapping and cloning of the major seed protein quantitative trait loci on soybean chromosome 20. *The Plant Journal*, 110, 114–128.

Guo, M., Deng, L., Gu, J., Miao, J., Yin, J., Li, Y., Fang, Y., Huang, B., Sun, Z. & Qi, F. 2024. Genome-wide association study and development of molecular markers for yield and quality traits in peanut (Arachis hypogaea L.). *BMC Plant Biology*, 24, 244.

Hell, R. & Wirtz, M. 2011. Molecular biology, biochemistry and cellular physiology of cysteine metabolism in Arabidopsis thaliana. *The Arabidopsis book/American Society of Plant Biologists*, 9.

Hu, Q., Min, L., Yang, X., Jin, S., Zhang, L., Li, Y., Ma, Y., Qi, X., Li, D. & Liu, H. 2018. Laccase GhLac1 modulates broad-spectrum biotic stress tolerance via manipulating phenylpropanoid pathway and jasmonic acid synthesis. *Plant physiology*, 176, 1808–1823.

Huang, T., Jander, G. & De Vos, M. 2011. Non-protein amino acids in plant defense against insect herbivores: representative cases and opportunities for further functional analysis. *Phytochemistry*, 72, 1531–1537.

Hunter, B. G., Beatty, M. K., Singletary, G. W., Hamaker, B. R., Dilkes, B. P., Larkins, B. A. & Jung, R. 2002. Maize opaque endosperm mutations create extensive changes in patterns of gene expression. *The plant cell*, 14, 2591–2612.

Ishimoto, M., Rahman, S. M., Hanafy, M. S., Khalafalla, M. M., El-Shemy, H. A., Nakamoto, Y., Kita, Y., Takanashi, K., Matsuda, F. & Murano, Y. 2010. Evaluation of amino acid content and nutritional quality of transgenic soybean seeds with high-level tryptophan accumulation. *Molecular Breeding*, 25, 313–326.

Isleib, T. G., Pattee, H. E., Sanders, T. H., Hendrix, K. W. & Dean, L. O. 2006. Compositional and sensory comparisons between normal-and high-oleic peanuts. *Journal of agricultural and food chemistry*, 54, 1759–1763.

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S.-Y., Freimer, N. B., Sabatti, C. & Eskin, E. 2010. Variance component model to account for sample structure in genome-wide association studies. *Nature genetics*, 42, 348–354.

Khandaker, L., Akond, M., Liu, S., Kantartzi, S. K., Meksem, K., Bellaloui, N., Lightfoot, D. A. & Kassem, M. A. 2015. Mapping of QTL associated with seed amino acids content in "MD96-5722" by "Spencer" RIL population of soybean using SNP markers. *Food and Nutrition Sciences*, 6, 974.

Kim, W. J., Kang, B. H., Kang, S., Shin, S., Chowdhury, S., Jeong, S.-C., Choi, M.-S., Park, S.-K., Moon, J.-K. & Ryu, J. 2023. A Genome-Wide Association Study of Protein, Oil, and Amino Acid Content in Wild Soybean (Glycine soja). *Plants*, 12, 1665.

Krishnan, N., Dickman, M. B. & Becker, D. F. 2008. Proline modulates the intracellular redox environment and protects mammalian cells against oxidative stress. *Free Radical Biology and Medicine*, 44, 671–681.

Li, L., Cui, S., Dang, P., Yang, X., Wei, X., Chen, K., Liu, L. & Chen, C. Y. 2022. GWAS and bulked segregant analysis reveal the Loci controlling growth habit-related traits in cultivated Peanut (Arachis hypogaea L.). *BMC genomics*, 23, 403.

Liu, H., Gu, J., Lu, Q., Li, H., Hong, Y., Chen, X., Ren, L., Deng, L. & Liang, X. 2019. Transcriptomic analysis reveals the high-oleic acid feedback regulating the homologous gene expression of stearoyl-ACP desaturase 2 (SAD2) in peanuts. *International Journal of Molecular Sciences*, 20, 3091.

Liu, H., Li, H., Gu, J., Deng, L., Ren, L., Hong, Y., Lu, Q., Chen, X. & Liang, X. 2018. Identification of the candidate proteins related to oleic acid accumulation during peanut (Arachis hypogaea L.) seed development through comparative proteome analysis. *International journal of molecular sciences*, 19, 1235.

Lu, Q., Huang, L., Liu, H., Garg, V., Gangurde, S. S., Li, H., Chitikineni, A., Guo, D., Pandey, M. K. & Li, S. 2024. A genomic variation map provides insights into peanut diversity in China and associations with 28 agronomic traits. *Nature genetics*, 1–11.

Malle, S., Eskandari, M., Morrison, M. & Belzile, F. 2020. Genome-wide association identifies several QTLs controlling cysteine and methionine content in soybean seed including some promising candidate genes. *Scientific reports*, 10, 21812.

Maloney, G. S., Kochevenko, A., Tieman, D. M., Tohge, T., Krieger, U., Zamir, D., Taylor, M. G., Fernie, A. R. & Klee, H. J. 2010.

Characterization of the branched-chain amino acid aminotransferase enzyme family in tomato. *Plant physiology*, 153, 925–936.

Mandal, S. & Mandal, R. 2000. Seed storage proteins and approaches for improvement of their nutritional quality by genetic engineering. *Current Science*, 576–589.

Matysik, J., Alia, Bhalu, B. & Mohanty, P. 2002. Molecular mechanisms of quenching of reactive oxygen species by proline under stress in plants. *Current Science*, 525–532.

Miflin, B. & Lea, P. 1977. Amino acid metabolism. *Annual Review of Plant Physiology*, 28, 299–329.

Mishra, S. & Dubey, R. S. 2006. Inhibition of ribonuclease and protease activities in arsenic exposed rice seedlings: role of proline as enzyme protectant. *Journal of plant physiology*, 163, 927–936.

Misra, P. S., Jambunathan, R., Mertz, E. T., Glover, D. V., Barbosa, H. M. & Mcwhirter, K. S. 1972. Endosperm protein synthesis in maize mutants with increased lysine content. *Science*, 176, 1425–1427.

Moran, M. D. 2003. Arguments for rejecting the sequential Bonferroni in ecological studies. *Oikos*, 100, 403–405.

Moulin, M., Deleu, C., Larher, F. & Bouchereau, A. 2006. The lysine-ketoglutarate reductase–saccharopine dehydrogenase is involved in the osmo-induced synthesis of pipecolic acid in rapeseed leaf tissues. *Plant Physiology and Biochemistry*, 44, 474–482.

Pandey, M. K., Wang, M. L., Qiao, L., Feng, S., Khera, P., Wang, H., Tonnis, B., Barkley, N. A., Wang, J. & Holbrook, C. C. 2014. Identification of QTLs associated with oil content and mapping FAD2 genes and their relative contribution to oil quality in peanut (Arachis hypogaea L.). *BMC genetics*, 15, 1–14.

Panthee, D., Kwanyuen, P., Sams, C., West, D., Saxton, A. & Pantalone, V. 2004. Quantitative trait loci for β-conglycinin (7S) and glycinin (11S) fractions of soybean storage protein. *Journal of the American Oil Chemists' Society*, 81, 1005–1012.

Panthee, D., Pantalone, V., Saxton, A., West, D. & Sams, C. 2006. Genomic regions associated with amino acid composition in soybean. *Molecular Breeding*, 17, 79–89.

Pfaffl, M. W. 2001. A new mathematical model for relative quantification in real-time RT–PCR. *Nucleic acids research*, 29, e45-e45.

Qin, J., Shi, A., Song, Q., Li, S., Wang, F., Cao, Y., Ravelombola, W., Song, Q., Yang, C. & Zhang, M. 2019. Genome wide association study and genomic selection of amino acid concentrations in soybean seeds. *Frontiers in Plant Science*, 10, 1445.

Sarvamangala, C., Gowda, M. & Varshney, R. 2011. Identification of quantitative trait loci for protein content, oil content and oil quality for groundnut (Arachis hypogaea L.). *Field crops research*, 122, 49–59.

Sharma, P. & Dubey, R. S. 2005. Modulation of nitrate reductase activity in rice seedlings under aluminium toxicity and water stress: role of osmolytes as enzyme protectant. *Journal of plant physiology*, 162, 854–864.

Shin, J.-H., Blay, S., Mcneney, B. & Graham, J. 2006. LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *Journal of statistical software*, 16, 1–9.

Sullivan, L. B., Gui, D. Y., Hosios, A. M., Bush, L. N., Freinkman, E. & Vander Heiden, M. G. 2015. Supporting aspartate biosynthesis is an essential function of respiration in proliferating cells. *Cell*, 162, 552–563.

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.

Sun, Z., Qi, F., Liu, H., Qin, L., Xu, J., Shi, L., Zhang, Z., Miao, L., Huang, B. & Dong, W. 2022. QTL mapping of quality traits in peanut using whole-genome resequencing. *The crop journal*, 10, 177–184.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M. & Rozen, S. G. 2012. Primer3—new capabilities and interfaces. *Nucleic acids research*, 40, e115-e115.

Vassiliou, E. K., Gonzalez, A., Garcia, C., Tadros, J. H., Chakraborty, G. & Toney, J. H. 2009. Oleic acid and peanut oil high in oleic acid reverse the inhibitory effect of insulin production of the inflammatory cytokine TNF-α both in vitro and in vivo systems. *Lipids in Health and Disease*, 8, 1–10.

Wang, J., Xian, X., Xu, X., Qu, C., Lu, K., Li, J. & Liu, L. 2017. Genome-wide association mapping of seed coat color in Brassica napus. *Journal of agricultural and food chemistry*, 65, 5229–5237.

Wang, J., Yan, C., Li, Y., Li, C., Zhao, X., Yuan, C., Sun, Q. & Shan, S. 2019. GWAS discovery of candidate genes for yield-related traits in peanut and support from earlier QTL mapping studies. *Genes*, 10, 803.

Wang, X., Lopez-Valenzuela, J. A., Gibbon, B. C., Gakiere, B., Galili, G. & Larkins, B. A. 2007. Characterization of monofunctional aspartate kinase genes in maize and their relationship with free amino acid content in the endosperm. *Journal of experimental botany*, 58, 2653–2660.

Wen, W., Li, D., Li, X., Gao, Y., Li, W., Li, H., Liu, J., Liu, H., Chen, W. & Luo, J. 2014. Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nature communications*, 5, 3438.

Wilson, J. N., Chopra, R., Baring, M. R., Selvaraj, M. G., Simpson, C. E., Chagoya, J. & Burow, M. D. 2017. Advanced backcross quantitative trait loci (QTL) analysis of oil concentration and oil quality traits in peanut (Arachis hypogaea L.). *Tropical plant biology*, 10, 1–17.

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.

Yang, Y., Li, Y., Cheng, Z., Su, Q., Jin, X., Song, Y. & Wang, J. 2023. Genetic analysis and exploration of major effect QTLs underlying oil content in peanut. *Theoretical and Applied Genetics*, 136, 97.

Ye, J., Wang, X., Hu, T., Zhang, F., Wang, B., Li, C., Yang, T., Li, H., Lu, Y. & Giovannoni, J. J. 2017. An InDel in the promoter of Al-ACTIVATED MALATE TRANSPORTER9 selected during tomato domestication determines fruit malate contents and aluminum tolerance. *The Plant Cell*, 29, 2249–2268.

Young, V. R. & Pellett, P. L. 1994. Plant proteins in relation to human protein and amino acid nutrition. *The American journal of clinical nutrition*, 59, 1203S-1212S.

Zhang, D., Zhang, H., Hu, Z., Chu, S., Yu, K., Lv, L., Yang, Y., Zhang, X., Chen, X. & Kan, G. 2019a. Artificial selection on GmOLEO1 contributes to the increase in seed oil during soybean domestication. *PLoS Genetics*, 15, e1008267.

Zhang, H., Dean, L., Wang, M. L., Dang, P., Lamb, M. & Chen, C. 2023a. GWAS with principal component analysis identify QTLs associated with main peanut flavor-related traits. *Frontiers in Plant Science*, 14, 1204415.

Zhang, H., Wang, M. L., Dang, P., Jiang, T., Zhao, S., Lamb, M. & Chen, C. 2021. Identification of potential QTLs and genes associated with seed composition traits in peanut (Arachis hypogaea L.) using GWAS and RNA-Seq analysis. *Gene*, 769, 145215.

Zhang, J., Wang, X., Lu, Y., Bhusal, S. J., Song, Q., Cregan, P. B., Yen, Y., Brown, M. & Jiang, G.-L. 2018. Genome-wide scan for seed composition provides insights into soybean quality improvement and the impacts of domestication and breeding. *Molecular Plant*, 11, 460–472.

Zhang, X., Zhu, L., Ren, M., Xiang, C., Tang, X., Xia, Y., Song, D. & Li, F. 2023b. Genome-Wide Association Studies Revealed the Genetic Loci and Candidate Genes of Pod-Related Traits in Peanut (Arachis hypogaea L.). *Agronomy*, 13, 1863.

Zhang, Y., Wu, L., Wang, X., Chen, B., Zhao, J., Cui, J., Li, Z., Yang, J., Wu, L. & Wu, J. 2019b. The cotton laccase gene GhLAC15 enhances Verticillium wilt resistance via an increase in defence-induced lignification and lignin components in the cell walls of plants. *Molecular plant pathology*, 20, 309–322.

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

Zhou, Y., Cai, H., Xiao, J., Li, X., Zhang, Q. & Lian, X. 2009. Over-expression of aspartate aminotransferase genes in rice resulted in altered nitrogen metabolism and increased amino acid content in seeds. *Theoretical and Applied Genetics*, 118, 1381–1390.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Umer, M.J., Lu, Q., Huang, L., Batool, R., Liu, H., Li, H. et al. (2024) Genome-wide association study reveals the genetic basis of amino acids contents variations in Peanut (*Arachis hypogaea* L.). *Physiologia Plantarum*, 176(5), e14542. Available from: https://doi.org/10.1111/ppl.14542