



OPEN

Genetic architecture of tuber-bound free amino acids in potato and effect of growing environment on the amino acid content

Jeewan Pandey¹, Dalton Thompson², Madhumita Joshi², Douglas C. Scheuring¹, Jeffrey W. Koym³, Vijay Joshi^{1,2✉} & M. Isabel Vales^{1✉}

Free amino acids in potato tubers contribute to their nutritional value and processing quality. Exploring the natural variation in their accumulation in tubers across diverse genetic backgrounds is critical to potato breeding programs aiming to enhance or partition their distribution effectively. This study assessed variation in the tuber-bound free amino acids in a diversity panel of tetraploid potato clones developed and maintained by the Texas A&M Potato Breeding Program to explore their genetic basis and to obtain genomic-estimated breeding values for applied breeding purposes. Free amino acids content was evaluated in tubers of 217 tetraploid potato clones collected from Dalhart, Texas in 2019 and 2020, and Springlake, Texas in 2020. Most tuber amino acids were not affected by growing location, except histidine and proline, which were significantly lower (−59.0%) and higher (+129.0%), respectively, at Springlake, Texas (a location that regularly suffers from abiotic stresses, mainly high-temperature stress). Single nucleotide polymorphism markers were used for genome-wide association studies and genomic selection of clones based on amino acid content. Most amino acids showed significant variations among potato clones and moderate to high heritabilities. Principal component analysis separated fresh from processing potato market classes based on amino acids distribution patterns. Genome-wide association studies discovered 33 QTL associated with 13 free amino acids. Genomic-estimated breeding values were calculated and are recommended for practical potato breeding applications to select parents and advance clones with the desired free amino acid content.

Potato is a major dietary staple crop consumed worldwide. Potato tubers are a rich source of carbohydrates (mainly starch), dietary fiber, high-quality protein, vitamins, antioxidants, and minerals^{1,2}. The total protein content of potato tubers is relatively low (1.0–3.0% FW or 5.0–14.0% DW) but comparable to cereals on a dry basis and better than commonly consumed vegetables³. Potato protein has high biological activity due to its high free amino acid score (69.0–76.0%)^{4,5} and higher nutritional value than other commonly consumed plant sources due to its high content of essential amino acids⁶. Potato contains all nine essential amino acids³, qualitatively superior to many other plant-based proteins⁷. Nearly half of the total amino acids in potatoes are in the free form^{8,9}. Asparagine and glutamine amides dominate the free amino acids (34.0–90.0%)^{10,11}, whereas lysine, tyrosine, and the sulfur-containing amino acids methionine and cysteine are still limited in potatoes^{3,12}. Nevertheless, potato tubers contain higher amounts of lysine than cereals¹³. Enhancing the protein quality of potato tubers by traditional breeding or genetic manipulation of pathways involved in the metabolism of limiting amino acids holds the potential to improve potato's biological activity and nutritional value and thus enhance human health and nutrition.

Free amino acids react with reducing sugars during frying, resulting in a non-enzymatic browning reaction (Maillard) which affects the organoleptic and other quality traits of processed potato products^{14,15}. Asparagine reacts with reducing sugars to form a carcinogenic compound called acrylamide^{16,17}, a major concern in

¹Department of Horticultural Sciences, Texas A&M University, College Station, TX 77843, USA. ²Texas A&M AgriLife Research and Extension Center, Uvalde, TX 78801, USA. ³Texas A&M AgriLife Research and Extension Center, Lubbock, TX 79403, USA. ✉email: Vijay.Joshi@tamu.edu; Isabel.Vales@tamu.edu

processed potato products^{18–22}. Genetically modified (GM) potatoes produced by manipulating the expression of the *asparagine synthetase* gene using RNA interference have low asparagine content and reduced acrylamide levels after processing^{23,24}. Free asparagine and reducing sugars are precursors for acrylamide formation via deamination and decarboxylation processes^{16,17,25}. In potatoes, the precursor concentration plays a significant role in determining the acrylamide-forming potential^{18,20–22,26–28}. A study reported a significant correlation between free asparagine concentration and acrylamide-forming potential in French fry varieties^{18,28}. Despite the need, polyploidy and high heterozygosity make breeding and selection of low acrylamide-forming potential potato cultivars for healthy fries and chips challenging. Possibilities of using the Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) system to manipulate reducing sugars and asparagine metabolism in potatoes are already being considered²⁹. Another amino acid affecting flavor is methionine, a precursor of methional, a major volatile associated with the aroma of potato chips and french fries³⁰. Enhancing free methionine in tubers for nutritional and aroma benefits is of interest to the food industry and consumers. Transgenic approaches to improve free methionine levels in potato tubers have been used but had, in several cases, undesired effects, including morphological changes in tubers or plants^{31–37}. Given the uncertainty about the acceptance of GM potatoes, selecting low asparagine and high methionine in potato tubers through conventional breeding could be the way to develop healthy and flavorful potato varieties. Most studies addressing quality traits focus on increasing starch content^{38–40}, diminishing reducing sugars^{41,42}, and possessing glycoalkaloid levels below toxic limits^{43,44} in potato tubers. Increasing the protein content in potato tubers is also desirable. However, breeding for more free amino acids in potato tubers may favor indirect selection for low carbohydrate levels because of their inverse relationship with protein content⁴⁵.

Significant genetic variation in the relative amounts of free proteinogenic and non-proteinogenic amino acids has been observed in cultivated potatoes and landraces^{46–48}. Screening of a diploid potato population using targeted⁴⁹ and untargeted⁴⁶ metabolite analysis showed significant variation in amino acid content, explaining the underlying genetic variation and identifying metabolic quantitative trait loci (QTL). Due to technological advancements in genotyping, genome-wide association studies (GWAS) have evolved as a powerful tool for exploring the genetic basis of various traits in several crop species, including potatoes^{50–58}. A GWAS study performed using the metabolome of cooked potatoes of the Solanaceae Coordinated Agricultural Project (SolCAP) diversity panel⁵⁹ identified several metabolites, including five free amino acids, significantly associated with SNP markers⁶⁰. Evaluating phenotypic variation for free amino acid levels in diversity panels used by breeding programs should provide a baseline to plan selection strategies to improve potatoes' nutritional properties. Diversity panels can be used to understand the genetic basis of the traits using GWAS and generate genomic-estimated breeding values (GEBVs) of traits using genomic selection (GS) for practical marker-assisted selection in breeding programs. Losses in nitrogen and amino acids in potato tubers differ significantly according to cooking techniques; chips and canned potatoes show the maximum loss, followed by drum-dried and french-fried potatoes^{61–64}. Field production practices, growing environmental conditions, nitrogen-fertilization rate and application time, residual soil nitrogen, soil-moisture availability, and storage conditions also influence the amino acid content^{65–68}. The broader goal of this research would be to assist the breeding program in developing healthier and more nutritious potatoes using conventional breeding and guide the development of strategies for metabolic engineering. The specific objectives were to i) evaluate the phenotypic variation for free amino acids of raw potato tubers of a diversity panel of tetraploid clones and to assess the effect of field location and year ii) understand the genetic basis of free amino acids using GWAS, and iii) generate GEBVs via GS to facilitate the selection of parents and advancement of clones through the breeding pipeline.

Results

Phenotypic variation for free amino acids in potato. The diversity panel (217) of cultivated potato clones evaluated in Dalhart, Texas in 2019 and 2020, and Springlake, Texas in 2020 showed phenotypic variation for the relative amounts of all 19 free amino acids in potato tubers. The phenotypic distributions of amino acid content were normal or close to normal in most cases (Supplementary Fig. 1). Histidine, arginine, asparagine, aspartic acid, alanine, proline, valine, leucine, and phenylalanine distributions differed significantly from normal per Shapiro–Wilk test⁶⁹ ($p < 0.05$) (Supplementary Fig. 1). Hierarchical cluster analysis of free amino acids reflected the metabolic relationship between amino acids (Fig. 1).

For all traits evaluated, the analysis of variance indicated significant differences between clones (Supplementary Table 1). The interactions between clones and environments were significant for most amino acids, except arginine, asparagine, glutamine, threonine, glycine, and lysine. Analysis of variance from the two Texas locations (Dalhart and Springlake) in 2020 indicated that most tuber amino acids were not affected by growing location, except histidine and proline (Fig. 2 and Supplementary Table 2), which were significantly lower (–59.0%) and higher (+129.0%), respectively, at Springlake, TX (a location that regularly suffers from abiotic stresses, mainly high-temperature stress). The average histidine content in tubers at Springlake was 0.5%, whereas tubers harvested from the Dalhart location had an average of 1.2%. In relation to proline, the average in Springlake was 1.6%, whereas the value was 3.7% in Dalhart (Fig. 2). In Springlake, the lowest amount of histidine was observed in Russet Norkotah 112 (0.1%), and the highest was found in ATTX98453-3R (1.4%). In Dalhart, the lowest amount of histidine was observed in ATX08181-5Y/Y (0.1%) the highest was found in ATX9332-8Ru (2.2%) (Supplementary Table 2). Regarding proline, in Springlake, the lowest value was obtained in TX05249-11W (1.3%), and the highest value was obtained in clone NDTX7590-3R (8.1%) (Supplementary Table 2). In Dalhart, the lowest value of proline was found in ATX9332-8Ru (0.5%), and the highest value was observed in AORTX09037-1W/Y (4.6%) (Supplementary Table 2). Analysis of variance with data from 2 years (2019, 2020, Dalhart location) revealed no significant differences between years, but the interaction between clone and year was significant for histidine, serine, and proline (Supplementary Table 1).

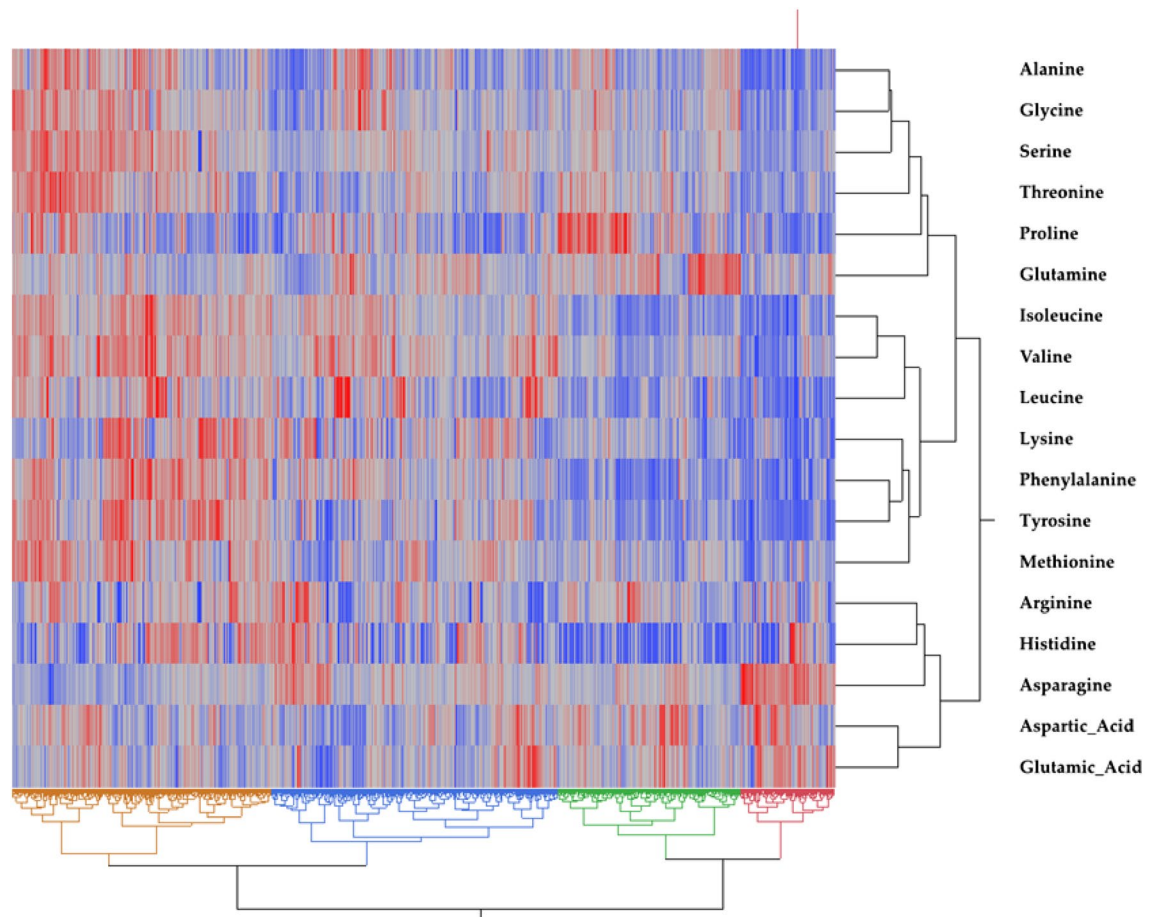
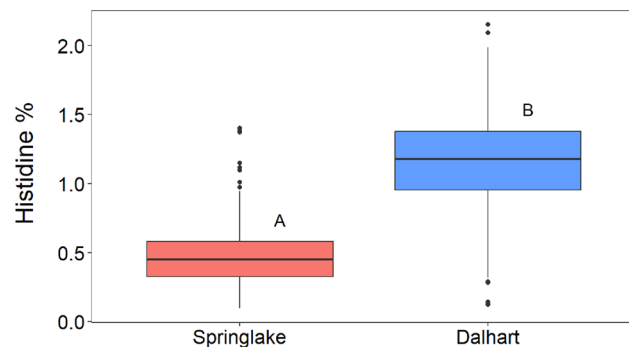


Figure 1. Hierarchical cluster TwoWay Analysis (Ward Method) of free amino acids across 217 tetraploid potato clones. Each clone is visualized as a single column. Each row represents an individual amino acid. Red denotes high amino acid content, whereas blue denotes a low free amino acid content.

A) Histidine



B) Proline

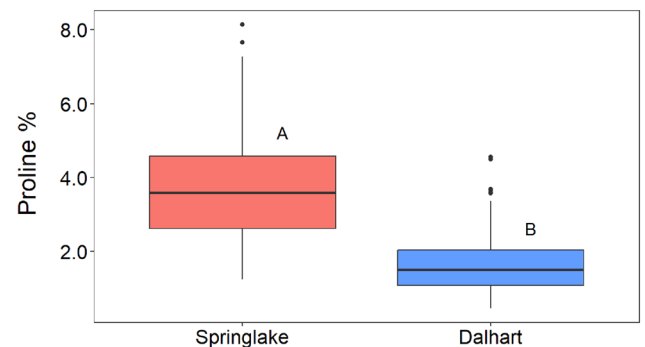


Figure 2. Box-plots for histidine (A) and proline (B) content in raw potato tubers diversity panel from the two locations (Springlake and Dalhart, TX). The amino acid means of tubers from the diversity panel (217 clones) at each location are represented by bold horizontal lines through the box-plot. Different letters above the boxplots denote significant differences between the means of the two locations based on Student t-test ($p < 0.05$).

Across the diversity panel, asparagine and glutamine were the most abundant amino acids (33.0%, and 21.0%, respectively). In contrast, histidine and glycine were the least abundant, with average relative compositions of (0.8%, and 0.9%, respectively) (Table 1). Asparagine was highest in the clone NDTX060700C-1W (55.0%) and lowest in the clone COTX94216-1R (21.0%) (Supplementary Table 3). The clone ATTX01178-1R had the highest concentration of glutamine (32.0%), whereas COTX08365F-3P/P had the highest concentration of valine (10.7%) (Supplementary Table 3). The highest proportion of methionine (2.5%) was found in the clone ATX9130-1Ru, whereas COTX03134-1Y/W had the highest concentration of lysine (5.6%) (Supplementary Table 3).

Free amino acids [†] (%)	Mean	Range		H ²
		Minimum	Maximum	
<i>Histidine</i>	0.80	0.02	1.91	0.15
Arginine	3.66	1.25	7.67	0.75
Asparagine	33.33	21.32	55.56	0.79
Glutamine	21.52	11.63	32.00	0.68
Serine	2.68	1.20	4.84	0.86
Glutamic acid	5.76	1.88	10.03	0.64
Aspartic acid	4.54	2.46	7.69	0.58
<i>Threonine</i>	2.12	0.63	3.56	0.81
Glycine	0.93	0.28	1.69	0.82
Alanine	2.92	0.77	7.51	0.89
Proline	2.66	1.15	5.28	0.61
<i>Lysine</i>	3.25	1.10	5.56	0.62
<i>Tyrosine</i>	1.20	0.26	2.19	0.62
<i>Methionine</i>	1.46	0.77	2.49	0.77
<i>Valine</i>	7.59	3.06	10.68	0.82
<i>Isoleucine</i>	2.50	1.15	3.82	0.69
<i>Leucine</i>	1.26	0.47	2.42	0.60
<i>Phenylalanine</i>	1.54	0.38	3.27	0.74

Table 1. Mean, range, and broad-sense heritability (H²) for free amino acids in 217 potato clones evaluated in Dalhart, Texas in 2019 and 2020, and Springlake, Texas in 2020. [†]Free amino acids in italics are essential amino acids.

Significant variation was observed between potato market groups for various free amino acids (Supplementary Table 4). Histidine was significantly higher in purples than in chippings and russets potatoes. The reds had a significantly lower asparagine (30.0%) than chippers (38.9%) and yellows (35.0%). Glutamine was significantly higher in chippers and phenylalanine in reds than in other market groups. Methionine was significantly higher in russets (1.6%) than in chippers (1.4%), yellows (1.4%), and purples (1.3%) (Supplementary Table 4). No significant differences were found between market groups for aspartic acid and lysine contents.

The diversity panel of clones was subjected to principal component analysis (PCA) based on free amino acid percentages and separation between fresh and processed (mainly chippers, since most of the russets included in the diversity panel belong to the fresh or dual-purpose) potato market classes. The first two principal components accounted for 47.9% of the total variance in the data, of which first principal component (PC1) explained 37.0% and second principal component (PC2) explained 10.9% of this variance (Fig. 3). Asparagine, isoleucine, valine, serine, alanine, phenylalanine, tyrosine, and glycine contributed the most to the PC1 and PC2 as observed by the intense red color in the contributions scale (Fig. 3, Supplementary Fig. 2).

The relationships among free amino acids in potato tubers were examined through a correlation network (Fig. 4). The correlation coefficient threshold for the network was set at 0.5, with *p*-values less than 0.05. The network consisted of 12 connected amino acids, and correlation coefficients were positive except with asparagine. Asparagine had the highest number of connections (nodal degree), followed by valine, tyrosine, valine, and serine, indicating its centrality to the network. The most abundant amino acid, glutamine, had a low level of connection. The connection between lysine, phenylalanine, and threonine was also poor. Valine had the highest positive and significant correlation with isoleucine ($r=0.80$). Significant negative correlations were found between asparagine and alanine ($r=-0.71$), valine ($r=-0.68$), and glycine ($r=-0.67$). Broad sense heritability values of the relative levels of free amino acids ranged from 0.58 to 0.89 for most amino acids except histidine, 0.15 (Table 1).

Average tuber weight per plant and average weight per tuber were significantly lower at Springlake, Texas compared to Dalhart, Texas. The correlations were examined and weak and/or moderate correlations were obtained. Interestingly, proline was negatively correlated with average tuber weight per plant and average weight per tuber (Table 2). Histidine was positively correlated with average tuber weight per plant and average weight per tuber (Table 2). There was no significant correlation between proline and average tuber number per plant. Also, no significant correlation was found between histidine and average tuber number per plant. The correlations of other amino acids with yield components were also examined. Negative correlation was observed for aspartic acid and proline with tuber weight (Supplementary Table 5). Likewise, positive correlations were observed for lysine, tyrosine, and methionine with tuber weight (Supplementary Table 5).

Linkage disequilibrium (LD). Using 10,106 SNPs for the full panel, pairwise LD was assessed using the squared-allele frequency correlations (r^2), and LD was found to be low even at zero distance (Supplementary Fig. 3). The function LD.plot was used to illustrate the amount of LD in the panel. The 5–10 Mb window size

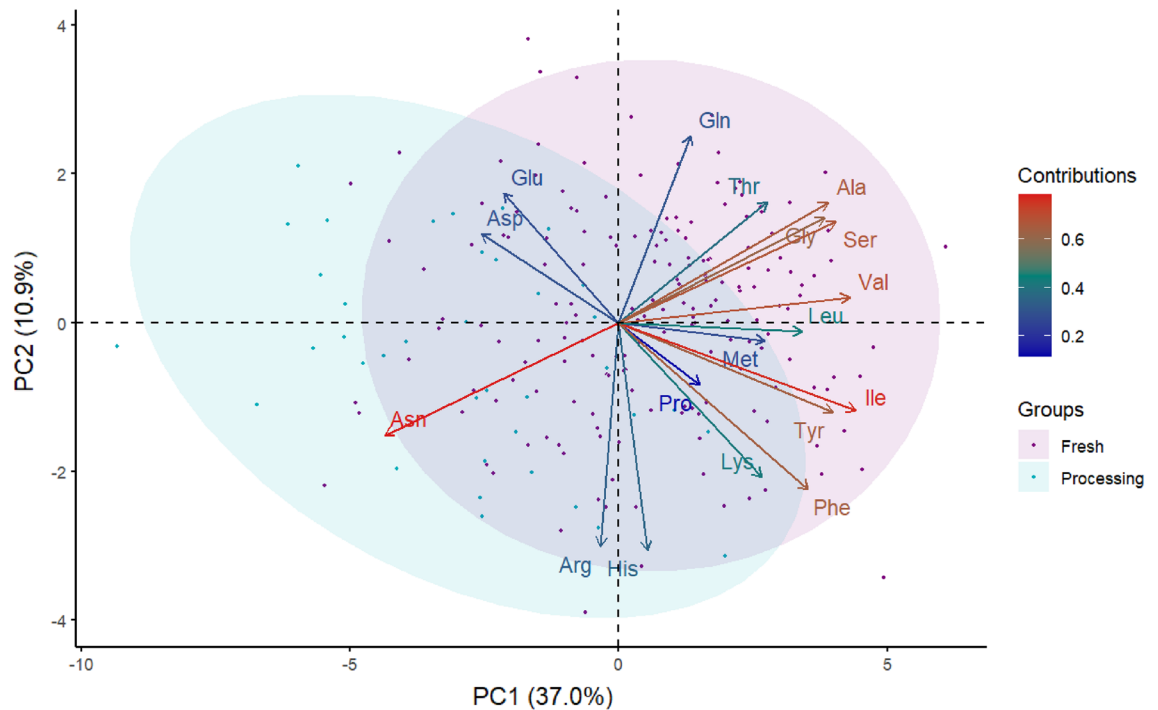


Figure 3. Principal component analysis separating fresh (specific gravity < 1.069) and processing (specific gravity > 1.070) potato market classes based on the free amino acid (His = histidine, Arg = arginine, Asn = asparagine, Gln = glutamine, Ser = serine, Glu = glutamic acid, Asp = aspartic acid, Thr = threonine, Gly = glycine, Ala = alanine, Pro = proline, Lys = lysine, Tyr = tyrosine, Met = methionine, Val = valine, Ile = isoleucine, Leu = leucine, Phe = phenylalanine) composition of 217 potato clones. Each point represents the projection of an individual potato clone in the PC1 and PC2 axes, and the dot colors represent the market group (purple = fresh and cyan = processing). The ellipses represent 95% confidence intervals around the centroid of each data cluster. The contributions (%) of free amino acids to PC1 and PC2 from low to high are indicated by blue to red color gradient, respectively.

seems appropriate to filter the output such that just the most significant marker inside a given window is returned based on the curve's shape.

Genome-wide association mapping—quantitative trait loci identification. We examined the inflation of the $-\log_{10}(p)$ using quantile–quantile (QQ) plots of the observed vs. expected values under the null hypothesis shown in Supplementary Fig. 4. Strongly associated SNPs deviated from the plot's diagonal at the upper-right end. A plot of the $-\log_{10}(p\text{-value})$ of the association statistic on the y-axis versus the chromosomal position of the SNP on the x-axis, commonly known as Manhattan plots, gave the visualization of the GWAS results (Fig. 5). Along the plot, regions with many highly linked SNPs in linkage disequilibrium emerged as the “Manhattan skyline”.

Allele dosage and different inheritance models were considered in GWASpoly⁵⁸. The additive inheritance model identified nine SNPs associated with eight free amino acids. The simplex dominant models identified seven (1-dom ref) and eight (1-dom alt) SNPs associated with eight and nine free amino acids, respectively (Table 3). The duplex dominant models identified six (2-dom ref) and five (2-dom alt) SNPs associated with six and five free amino acids, respectively (Table 3).

A total of 33 QTL were significantly associated with 13 free amino acids on chromosomes 1, 2, 3, 5, 6, 7, and 10 (Table 3). However, no QTL were detected above the threshold for five amino acids, which are glutamic acid, proline, lysine, tyrosine, and methionine. A QTL on chromosome 6 was identified for histidine in the additive and the dominant model. GWAS analysis of arginine content identified four QTL in chromosome 2. Five QTL were identified for asparagine content in chromosomes 2 and 7. Two QTL were identified for glutamine content in chromosomes 3 and 6. Four QTL were identified for serine content in chromosomes 2, 5, and 7. For aspartic acid content, the only QTL identified was in chromosome 1. For threonine content, two QTL were identified in chromosome 1. Three QTL were identified for glycine content in chromosome 2. Three QTL were identified for alanine content in chromosome 2. Four QTL were identified for valine content in chromosomes 2 and 3. One QTL was identified for isoleucine content in chromosome 2. A QTL was observed in chromosome 5 for leucine content, which explained 3.6% of the phenotypic variance. Two QTL were identified for phenylalanine content on chromosomes 2 and 10 under additive and dominant genetic models, respectively. QTL co-localization was discovered for a few amino acids regardless of the inheritance model tested. Asparagine, glycine, and serine showed two colocalizing QTL on chromosome 2 (SNP peak positions at 25.6 Mb and 26.6 Mb), and isoleucine

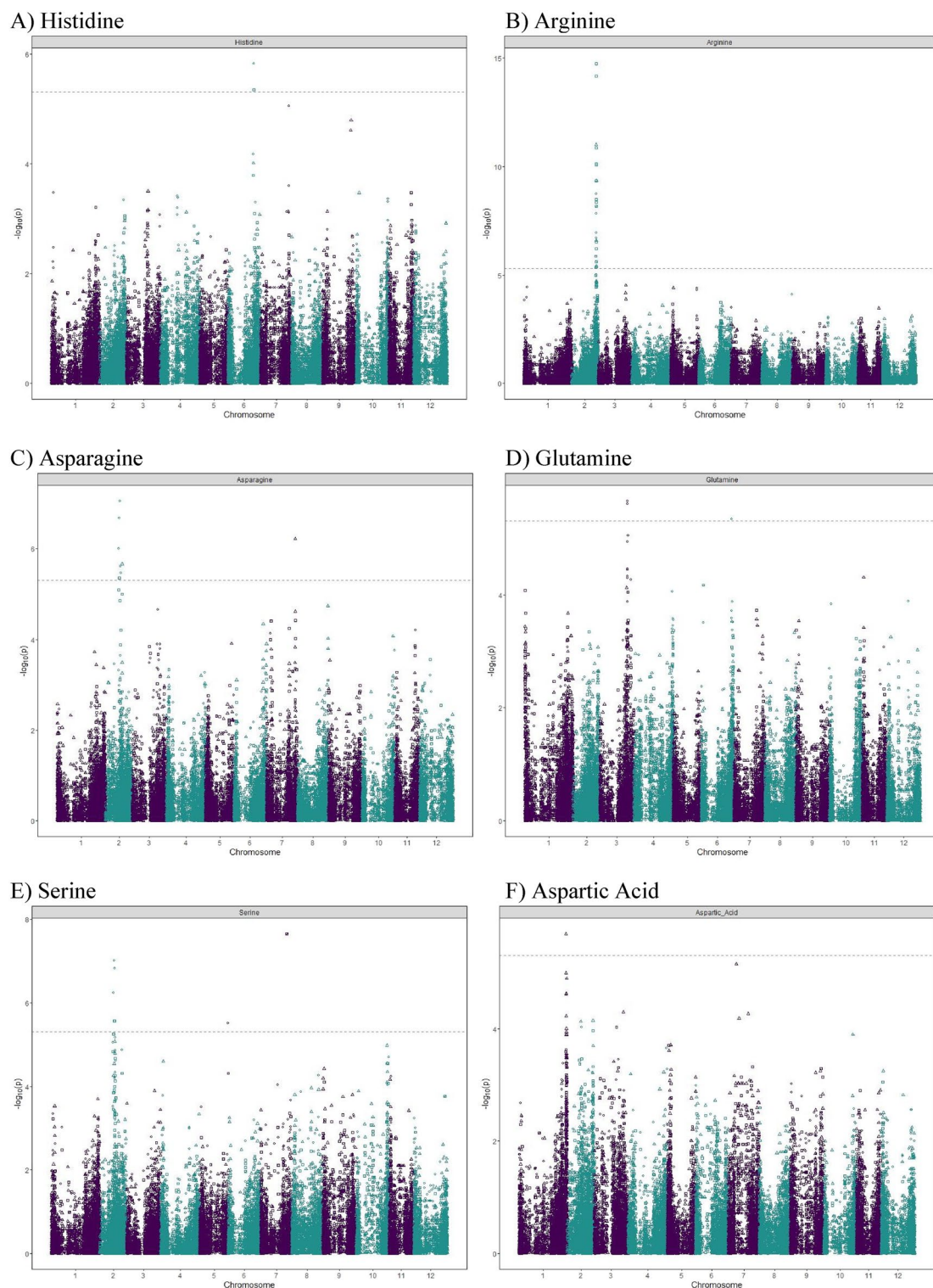
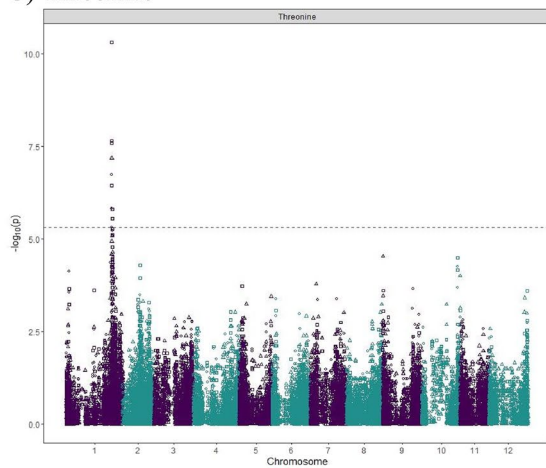
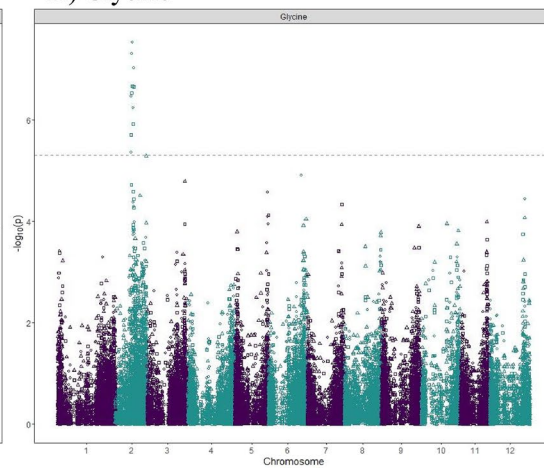


Figure 5. Manhattan plots displaying the marker–trait associations for Histidine (A), Arginine (B), Asparagine (C), Glutamine (D), Serine (E), Aspartic Acid (F), Threonine (G), Glycine (H), Alanine (I), Valine (J), Isoleucine (K), Leucine (L), Phenylalanine (M) by GWASpoly using the additive and dominant model in three combined environments. The horizontal axis indicates the chromosome number and the position of each SNP. The vertical axis indicates the negative logarithm of the p -value for each SNP. Each dot signifies an SNP. The broken line indicates the Bonferroni threshold level of 0.05.

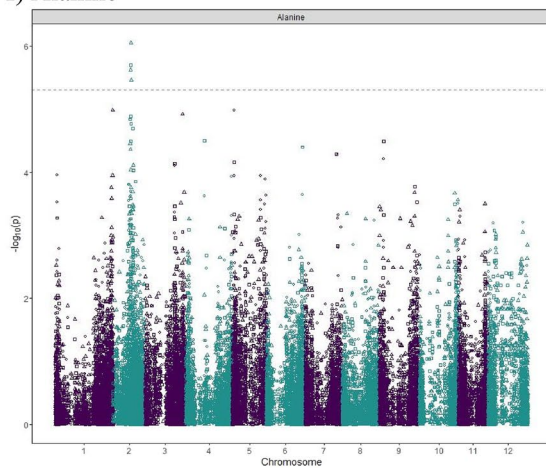
G) Threonine



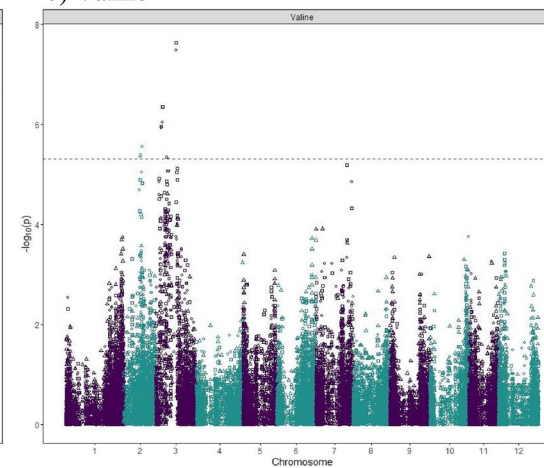
H) Glycine



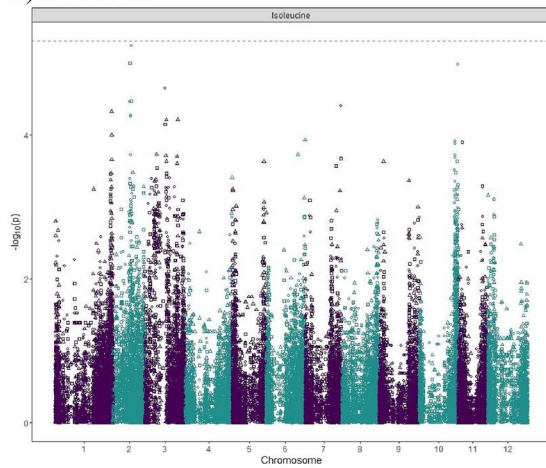
I) Alanine



J) Valine



K) Isoleucine



L) Leucine

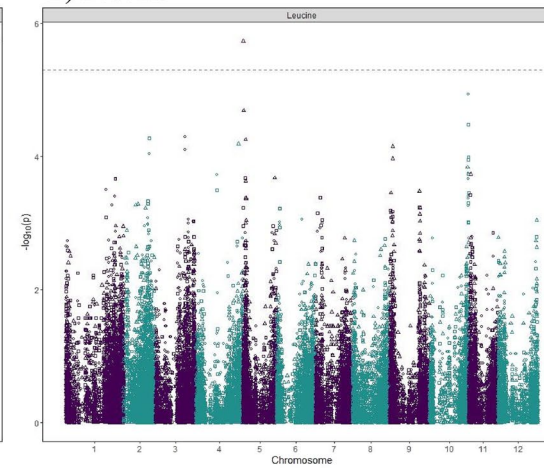


Figure 5. (continued)

M) Phenylalanine

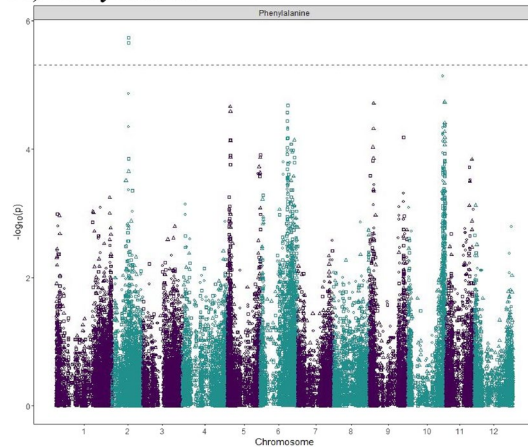


Figure 5. (continued)

histidine (Fig. 6). The GEBVs (Supplementary Table 7) showed that clone TX12474-1P/R ranked topmost for histidine. NDTX060700C-1W ranked topmost for asparagine and aspartic acid. Considering the GEBVs, the topmost was BTX1544-2W/Y for arginine, NDTX050169-1R for glutamine, and COTX00104-7R for serine content. BTX1544-2W/Y came out on top for arginine, whereas NDTX050169-1R came out on top for glutamine. COTX00104-7R had the highest GEBV for serine content, and AOTX97213-1Ru had the highest GEBV for glutamic acid. ATTX96746-1R came out on top for threonine, and Krantz came on top for glycine. COTX05211-4R ranked topmost for alanine, and NDTX081648CB-4W ranked topmost for proline. COTX04015-3W/Y had the highest GEBVs for lysine. COTX05261-1R/Y ranked topmost for tyrosine, and ATX9130-1Ru ranked topmost for methionine. Based on the GEBVs, ATTX98500-3P/Y came on top for valine. The topmost clone for isoleucine was ATTX98510-1R/Y, leucine was ATTX03516-2R, and phenylalanine was ATTX10265-4R/Y.

Weighted standardized multitrait selection indexes. Chipping clones with all other amino acids rich but with low asparagine was found using weighted normalized multitrait selection indexes (Supplementary Table 8). Three clones (NDTX059828-2W, ATTX95490-2W and AOTX95309-2W) were found to have the highest rankings (Z -score > 2) among the chipping potatoes (Supplementary Table 8).

Discussion

This study uncovered the natural variation for various free amino acids in a potato diversity panel from a potato breeding program in the USA⁵⁴ and linked phenotypic variation for several traits to genetic markers to understand the genetic basis of the traits and obtain genomic-estimated breeding values to improve breeding efficiencies. The panel of potato clones used included reference varieties for various market groups and advanced selections of cultivated tetraploid potatoes from the Texas A&M Potato Breeding Program, developed over 40 years. The outcome of this research should guide breeding efforts in potatoes to improve health, nutritional, and processing qualities.

Significant differences were found between tetraploid potato clones for the relative quantity of all 19 free amino acids evaluated. Variations in the amounts of free amino acids were found in potato tubers from different growing locations and years⁷⁰. Asparagine and glutamine were the most abundant amino acids, with average relative compositions of 33.0% and 21.0%, respectively in our potato diversity panel. Asparagine levels showed great variation in the panel, from 21.0 to 55.0%. The most prevalent amino acids in potato tubers were asparagine and glutamine, which together made up to 90.0% of the total free amino acid content¹⁰. In another study, asparagine accounted for the greatest proportion of free amino acids (up to 30.0–36.0%)⁷¹. Asparagine and reducing sugars form acrylamide, a carcinogenic chemical that is a concern, especially in processed potato products²⁸. Both asparagine and reducing sugars should be reduced to decrease acrylamide content in processing potatoes⁴². According to this study, targeted breeding efforts would be required to reduce the asparagine content of conventionally bred potatoes to the range required for acrylamide mitigation. Since potatoes had never been bred for low tuber asparagine content, the effectiveness of this strategy was questionable⁴². Our study found that the chipping (processing) market group had significantly higher asparagine content in raw tubers than other market groups (Supplementary Table 4). A previous study showed a negative correlation between tuber sucrose and amino acid content⁴⁵. It reported that tubers with a reduced sucrose transporter expression displayed significant increases in alanine, arginine, asparagine, glutamine, glycine, serine, threonine, and valine⁴⁵. In selecting chip potatoes with low-reducing sugars and cold-sweetening resistance, these breeding efforts may have favored indirect selection for higher accumulation of free amino acids including asparagine, but additional research is needed to confirm this relationship. Some research has shown that asparagine content in potatoes is larger and less variable than glucose and fructose, which helps to explain why reducing asparagine could be challenging²². On the other hand, researchers have shown that specific genotypes with low asparagine contents can be exploited in breeding efforts to lower tuber asparagine content^{41,72}.

Amino acid	Model	Threshold	SNP at QTL peak	Chr	Peak position (bp)	Score – log ₁₀ (p)	Effect	R ² (%)
Histidine	additive	5.3	PotVar0070127	6	46,406,899	5.3	–0.2	3.3
Histidine	1-dom-ref	5.1	PotVar0070127	6	46,406,899	5.8	–0.2	3.6
Arginine	additive	5.3	solcap_snp_c1_12771	2	44,810,486	14.8	–0.7	10.5
Arginine	1-dom-alt	5.0	solcap_snp_c2_43350	2	44,773,246	7.8	1.0	5.3
Arginine	1-dom-ref	5.1	solcap_snp_c1_12771	2	44,810,486	8.8	–0.9	6.0
Arginine	2-dom-alt	5.2	PotVar0001749	2	45,037,953	5.4	0.8	3.4
Arginine	2-dom-ref	5.2	PotVar0001550	2	44,561,226	11.0	–1.1	7.7
Asparagine	additive	5.3	solcap_snp_c2_52656	2	24,920,780	5.4	4.3	3.4
Asparagine	1-dom-alt	5.0	solcap_snp_c2_21654	2	25,560,378	7.1	6.4	4.5
Asparagine	1-dom-ref	5.1	PotVar0089318	2	26,584,883	5.6	–5.3	3.6
Asparagine	2-dom-alt	5.2	PotVar0044416	7	52,317,483	6.2	8.1	4.0
Asparagine	2-dom-ref	5.2	solcap_snp_c1_13929	2	29,929,027	5.7	–3.8	3.5
Glutamine	1-dom-alt	5.0	PotVar0041119	6	55,411,335	5.3	–4.3	3.3
Glutamine	1-dom-ref	5.1	solcap_snp_c1_439	3	50,959,945	5.7	3.3	3.6
Serine	additive	5.3	PotVar0089318	2	26,584,883	5.6	0.4	2.2
Serine	additive	5.3	solcap_snp_c2_38787	7	47,875,892	7.7	0.7	3.4
Serine	1-dom-alt	5.0	solcap_snp_c2_21654	2	25,560,378	7.0	–0.7	3.4
Serine	1-dom-alt	5.0	PotVar0034567	5	51,698,206	5.5	–0.4	2.0
Serine	1-dom-ref	5.1	PotVar0089318	2	26,584,883	6.8	0.6	2.6
Serine	1-dom-ref	5.1	solcap_snp_c2_38787	7	47,875,892	7.7	0.7	2.9
Aspartic acid	2-dom-alt	5.2	solcap_snp_c2_53088	1	85,861,165	5.7	–1.1	3.5
Threonine	additive	5.3	solcap_snp_c2_12087	1	71,176,726	10.3	–0.3	7.1
Threonine	1-dom-alt	5.0	solcap_snp_c2_12092	1	71,205,164	5.8	0.3	3.7
Threonine	1-dom-ref	5.1	solcap_snp_c2_12087	1	71,176,726	6.7	–0.4	4.4
Threonine	2-dom-ref	5.2	solcap_snp_c2_12087	1	71,176,726	7.2	–0.6	4.7
Glycine	additive	5.3	solcap_snp_c2_21654	2	25,560,378	6.7	–0.2	4.3
Glycine	1-dom-alt	5.0	solcap_snp_c2_21654	2	25,560,378	7.5	–0.3	4.9
Glycine	1-dom-ref	5.1	PotVar0089318	2	26,584,883	6.2	0.3	4.0
Glycine	2-dom-ref	5.2	PotVar0002461	2	46,623,091	5.3	–0.2	3.3
Alanine	additive	5.3	PotVar0089338	2	26,606,332	5.7	0.5	3.6
Alanine	2-dom-alt	5.2	PotVar0124308	2	27,089,538	6.1	–1.2	3.8
Alanine	2-dom-ref	5.2	PotVar0124132	2	27,325,817	5.5	–0.9	3.4
Valine	1-dom-alt	5.0	solcap_snp_c2_38935	2	27,860,245	5.6	–1.3	1.6
Valine	1-dom-alt	5.0	solcap_snp_c2_5742	3	10,383,457	6.0	–1.2	2.5
Valine	1-dom-ref	5.1	solcap_snp_c1_4775	3	31,056,417	7.5	1.5	1.2
Valine	2-dom-alt	5.2	solcap_snp_c2_36018	3	17,013,558	5.3	0.8	3.3
Isoleucine	1-dom-alt	5.0	solcap_snp_c2_38935	2	27,860,245	5.2	–0.5	3.3
Leucine	2-dom-ref	5.2	PotVar0048910	5	349,077	5.7	–0.7	3.6
Phenylalanine	additive	5.3	PotVar0117639	2	25,332,954	5.7	0.3	3.6
Phenylalanine	1-dom-ref	5.1	solcap_snp_c1_5010	10	55,210,592	5.1	0.3	3.2

Table 3. Significant marker-trait associations identified for the free amino acid samples obtained from three field experiments in Dalhart, Texas in 2019 and 2020, and Springlake, Texas in 2020.

Ideally, breeders should aim for the identification and development of processing potato cultivars with low free-asparagine and reducing sugar as desirable characteristics for processing purposes. Recently, in a processing cultivar, FL-1533, a distinctive metabolite combination was found to have the least amounts of reducing sugar and asparagine compared to the other cultivars⁷³. The availability of such germplasm would make it possible to introduce the low sugar and low asparagine traits into varieties acceptable for broad use in the food industry. The implications for low asparagine on potato yield, disease susceptibility, starch properties etc. are limited but it would be worth exploring in the future. However, previous studies indicated that certain amino acids affect the physicochemical properties of potato starch⁷⁴. Charge-carrying amino acids (lysine, arginine, aspartic acid, and glutamic acid) can potentially alter the physical and chemical characteristics of starch-based products derived from potatoes⁷⁴. Increased amounts of charge-carrying amino acids during heating can lead to notable improvements in their nutritional qualities, achieved by reducing the swelling power, solubility, light transmittance, and gel strength of the potato starch⁷⁴. Another study found that amino acids, particularly glutamic acid and aspartic acid, can affect the gelatinization, pasting, and rheological properties of potato starches⁷⁵, potentially resulting in effects on food texture and mouthfeel, and influencing consumer acceptance.

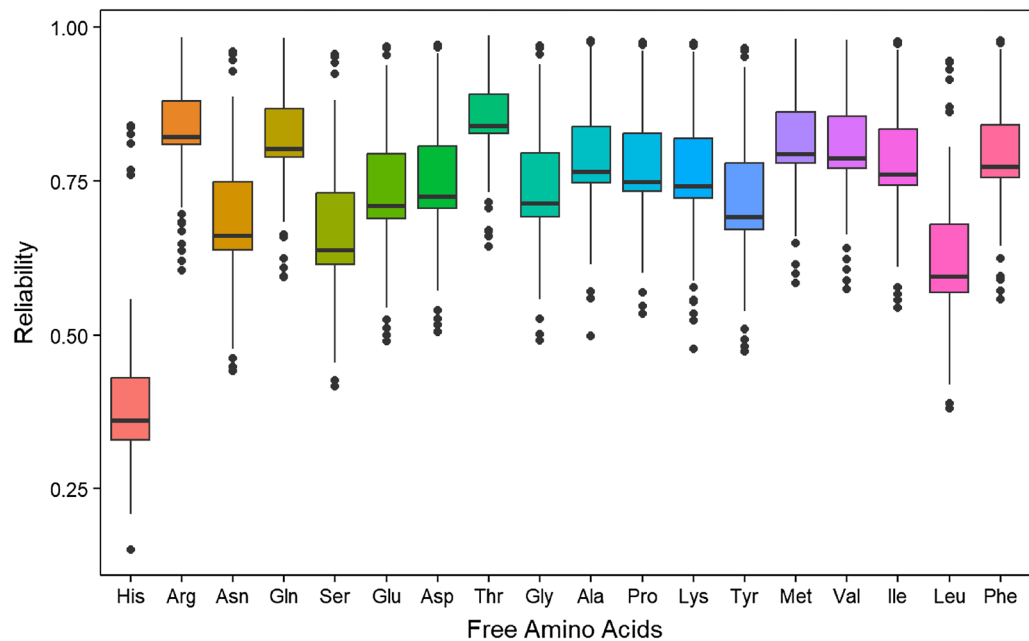


Figure 6. Box plot of the predicted reliabilities (the squared correlation between the true and predicted values) for free amino acids content (His = histidine, Arg = arginine, Asn = asparagine, Gln = glutamine, Ser = serine, Glu = glutamic acid, Asp = aspartic acid, Thr = threonine, Gly = glycine, Ala = alanine, Pro = proline, Lys = lysine, Tyr = tyrosine, Met = methionine, Val = valine, Ile = isoleucine, Leu = leucine, Phe = phenylalanine) in 217 potato clones.

Another essential amino acid, methionine, was significantly higher in russets and lowest in purples (Supplementary Table 4). The potato industry is interested in boosting the tuber methionine content because of the related nutritional and fragrance benefits³². Even though methionine is an important amino acid, no related QTL were found in the analysis. Increasing the sample size and studying more diverse populations could help capture a broader range of genetic variation and improve the chances of identifying relevant associations. If unearthing natural variation becomes a limitation, the biotechnological interventions to manipulate methionine metabolism is a parallel strategy. Lysine's average relative composition in the diversity panel was 3.3%, and the clone with the highest proportion of lysine (5.6%) was COTX03134-1Y/W. Potato is a good source of lysine and thus can be used as a supplement to cereals deficient in this amino acid⁷⁶.

Despite using similar production practices across the Texas potato field locations, the analysis of variance indicated significant environmental effects when comparing locations for histidine and proline content. On average, tubers produced at the Springlake location have lower histidine and higher proline content than those produced in Dalhart. Quantitative differences in the amino acid content between varieties based on location were previously reported⁷⁷. Proline was the amino acid that varied the greatest between the two regions which differed in rainfall and humidity⁷⁷. The Springlake and Dalhart locations differ for altitude, latitude, and longitude; day/night temperatures, photoperiod, radiation, type of soil, and biotic and abiotic stresses. The Springlake location regularly has higher night temperatures (often > 20 °C near harvest), resulting in lower marketable yield and more external and internal tuber defects as result of having more heat stress than the Dalhart location. Experiments under controlled greenhouse conditions showed that heat stress-induced chemical changes in potato tubers⁷⁸. Proline is crucial for plant growth and differentiation throughout the whole life cycle and also serves as a superior osmolyte and also performs three other crucial functions under stress, including metal chelation, antioxidant defense, and signaling⁷⁹. To protect plants from heat stress, proline functions as a signaling molecule⁸⁰. Our findings indicated that proline levels in tubers produced at Springlake, a location suffering abiotic stresses (mainly heat stress), were significantly higher (129% higher) than a more favorable location (Dalhart) that, despite having hot days during the day, has more cool nights that favor the accumulation of starch in tubers and less external and internal tuber defects. Proline is a precursor for other downstream metabolites, including polyamines and secondary metabolites and thus may affect tuber quality. Proline can be produced from arginine and glutamine/glutamate in the majority of animals, including humans and pigs, although rates of endogenous synthesis are insufficient for newborns, fish, and birds⁸¹, thus consuming proline-rich produce should be beneficial to contribute to the production of proteins in general, and collagen in particular, and thus help in the formation of skin, cartilages, and bones. Likewise, histidine is an essential amino acid for plant growth and development⁸². Our findings showed lower level of histidine in tubers in the stressful site Springlake. This result conflicts with the amount of free histidine present in canola cultivars, as histidine, serine, and cysteine were found to rise in the xylem sap of various canola cultivars during nickel toxicity⁸³. Further research is needed to fully understand if stress induces low histidine in tubers.

In our study, years did not significantly affect the content of free amino acids in tubers whereas location had a significant effect in histidine and proline. While breeding for high (or low) amino acids, multilocation evaluations would be more important than multiyear evaluations.

In this study, branched-chain amino acids (isoleucine and valine) displayed high correlation values ($r = 0.8$). Aromatic amino acids (i.e., tyrosine and phenylalanine) also highly correlated ($r = 0.7$). High correlations were anticipated due to linked pathways. Previous studies in potatoes reported $r > 0.7$ between branched-chain amino acids⁴⁶ and $r = 0.7$ between serine and glycine⁸⁴. However, in our research, unexpected correlations were also observed between amino acids that are not closely linked biosynthetically (e.g., phenylalanine and isoleucine, alanine and serine, and alanine and glycine). In previous metabolic profile studies of potato tubers, unexpected amino acid associations were also discovered^{46,84,85}. Asparagine and other amino acids were found to have significant negative correlations, suggesting that decreasing asparagine would probably increase the other amino acids involved in the correlations and vice versa.

In this study, proline content was negatively correlated and histidine content was positively correlated with yield components but the correlations were weak. The weaker correlations would likely make them unsuitable to use as markers for yield potential. The relationship between proline and histidine content and potato yield components are limited. However, the relationship may vary depending on several variables, including the type and degree of stress, the potato cultivar, and other environmental factors. In wheat, grain proline content during a drought was linked to greater thousand grain weight loss and does not seem to act positively in terms of productivity⁸⁶.

Thanks to recent developments in high-throughput metabolic profiling and sequencing technologies, GWAS has been used as a powerful strategy to unveil crop metabolism's genetic and biochemical underpinnings. GWASpoly is suited for conducting GWAS in polyploids as it investigates different inheritance models⁵⁸. For example, in this study, nine SNPs were shown to be linked to eight free amino acids with the additive inheritance model. Each additional copy of the allele can boost the trait's output. Using GWASpoly, we identified 33 QTL associated with 13 free amino acids in this study. In general, only a small portion (< 10%) of the phenotypic variance could be accounted by each QTL, except for arginine (10.5%). Co-localization of QTL for free amino acids was observed in a few cases independently of the model of inheritance investigated. It is not unusual to see colocalized QTL for biochemically related amino acids. For example, in the present study, isoleucine and valine had the strongest, most positive, and most significant correlation ($r = 0.80$), and shared a QTL in chromosome 2 (SNPs peak position at 27.9 Mb). Previous research on potatoes has shown a high interdependence of biochemically related amino acids⁴⁶. A shared QTL could come from closely linked independent regulators or a common upstream regulator⁴⁶ for amino acids that are not biochemically related. QTL linked to various amino acids can potentially improve multiple amino acid content in potatoes simultaneously, and they deserve further investigation.

By assigning potential candidate genes to GWAS outcomes based on metabolites, it is postulated that the candidate loci directly or indirectly encode intermediate products that would biochemically catalyze, modify, translocate, or regulate the associated metabolites— which thus warrants systemic evaluation to link them to end phenotypes. The success of GWAS in mapping genes is limited in its resolution due to the high level of LD in the breeding germplasm, annotation of the reference genome, and the number of high-quality SNPs. This method provides only indirect evidence for the association of the genomic region with individual amino acids. We anticipate continuing additional research to validate the functional role of these genes in using model species or transgenic approaches. Nevertheless, this study has established a foundation to analyze identified loci and their evolutionary significance in amino acid metabolism in potatoes. In our study, the genes within 100 kbp window size were identified according to the positions of the closest flanking significant SNPs. Several independent studies have used genomic space within ± 100 kb of each peak SNP proximity to identify candidate genes associated with varied traits using association mapping in potato^{87,88}. Expanding the search interval to 100 kbp increases the likelihood of capturing genetic variants in LD with the causal variant, improving the chances of identifying relevant candidate genes. In some instances, narrower or broader intervals may be considered, depending on the context and objectives of the study.

Our analysis did uncover a few candidate genes that are directly connected to major metabolic processes. Based on the KEGG pathway (SOT 102,589,057), S-adenosylmethionine decarboxylase proenzyme (SAMDC)—PGSC0003DMG400010051 was mapped to arginine metabolism. S-adenosylmethionine decarboxylase synthesizes polyamines putrescine, spermidine, and spermine across species⁸⁹. Arginine is the ornithine precursor, the substrate needed for polyamines biosynthesis. Ornithine decarboxylase converts ornithine to putrescine, which is further converted to spermidine and spermine by consuming decarboxylated S-adenosylmethionine⁹⁰. Transgenic potato plants expressing antisense SAMDC showed stunted phenotypes, reduced levels of SAMDC transcripts and enzyme activity, and polyamine content, suggesting the significance of polyamines in the tuber formation⁹¹. Additionally, GWAS of sorghum grain qualities identified S-adenosylmethionine decarboxylase proenzyme mapped to arginine, cysteine, methionine, proline, and starch metabolism⁹². It will be intriguing to perform additional genetic characterization to understand the role of this gene in polyamine synthesis or, like sorghum, its relevance in starch synthesis in potatoes. Another candidate gene, Proline oxidase/dehydrogenase (PGSC0003DMG400010050), catalyzes proline oxidation to Δ^1 -Pyrroline—5-Carboxylate Synthetase (P5C) in the mitochondria. Proline and arginine metabolic pathways are interconnected⁹³ as the second and last step in proline catabolism, P5C to glutamate is shared with arginine catabolism⁹⁴. Arginine can serve as a precursor and product of ornithine during the urea cycle via arginase, converting arginine to ornithine⁹⁵. Our study identified limited candidate genes directly associated with primary metabolic pathways, which is plausible given the complexity of their biosynthesis and catabolism involving several intermediates, their transport, and feedback regulations. Detailed characterization of identified genes using functional approaches such as cloning and

heterologous expression of genes in model species, tissue-specific knocking out genes using RNAi, antisense methods, or CRISPR-Cas would help understand their relevance in amino acid metabolism in potatoes.

GEBV was obtained for all members of the diversity panel (217 clones). GEBVs will allow faster and more effective selection of superior parents and/or the advancement of better clones. The predicted reliabilities in this study were encouraging, except for histidine. Prediction ability is influenced by the genetic make-up of the traits under investigation, the training population, and how similar the training population is to the validation population. Recently, software programs like StageWise⁹⁶ are being developed for polyploid crops to manage the numerous experimental designs, heritabilities, and spatial models for nongenetic variation. Individual GEBVs will likely not be used for selection since there are many important traits to consider in selection. To use GEBVs for selection, GEBVs for several traits might be merged into an overall index. In this study we utilized the standardized indexes for the chipping clones. Not all traits will have the same importance, and not all traits are measured using the same units or scales; thus, standardized scores and relative weight should be assigned to each trait to facilitate the selection of superior and inferior genotypes⁹⁷. Market demands, industry needs, and several other considerations drive the relative importance of traits. Using GEBVs to select potatoes with particular concentrations could significantly improve potatoes' health and nutritional value. GEBVs for traits such as yield, quality, disease/pest resistance, and tolerance to abiotic stresses should be part of the combined selection index to maximize the chances of selecting the best clones and make selection more efficient. For practical applications, GEBVs should be considered in the context of market groups and selection criteria.

Materials and methods

Plant materials and experimentation. A collection of 217 tetraploid potato clones was used for this study. The diversity panel represents varieties and advanced genotypes selected over four decades by the Texas A&M Potato Breeding Program. Details of the potato diversity panel used were described earlier⁵⁴. The clones included fresh and processing market classes with variations in tuber shape, skin type, skin and flesh color, biotic and abiotic stress tolerance, and quality traits⁵⁴.

Three field experiments were conducted at two locations in Texas in commercial potato grower fields. In 2019, the diversity panel was planted in Dalhart (35°58' N, 102°44' W), and in 2020 the clones were planted in Dalhart and Springlake (34°6' N, 102°19' W) in a 12-hill plot with two replications. Dalhart is approx. 268 km North of Springlake and 92 m higher. Both Texas locations experience high-temperature stress during the growing season, however, Dalhart yields often duplicate the yields in Springlake (<https://potato.tamu.edu/reports/>), likely due to cooler nights at Dalhart. Trials were planted in early May in Dalhart, Texas, and harvested in early September. At Springlake, trials were planted in late March to early April and harvested in early July. Vine desiccation was done 2–3 weeks before harvesting the trials. Standard production practices were used. Details about the field trials, including environmental conditions, spacing, irrigation, fertilization, weed, and insect control, can be obtained online for the corresponding years and locations from the 2019 and 2020 Texas A&M Potato Breeding Program reports (<https://potato.tamu.edu/reports/>).

DNA extraction and genotyping. Genomic DNA was extracted from 50 to 80 mg of fresh young potato leaves from tissue culture plantlets using the DNeasy Plant Pro Kit (Qiagen, Valencia, CA, USA) as described earlier⁵⁴. Samples were analyzed using the Infinium 22 K V3 Potato Array on the Illumina iScan (Illumina Inc., San Diego, CA, USA). 10,106 polymorphic SNPs were selected for analysis after filtering the marker dataset for minor allele frequency and polymorphism, as previously described⁵⁴.

Measurements of free amino acid content. Three tubers per plot were randomly selected and chopped into small cubes (4 mm per side). About 15 g of freshly chopped samples were collected into 50 mL tubes. Samples were stored in –80 °C freezer for a few days before freeze-drying them using a Labconco FreeZone 6-L Freeze-dry System (Labconco, Kansas City, MO, USA) at a collector temperature of –50 °C for 120 h at 22 Pa. Ceramic grinding cylinders (0.95 cm × 2.22 cm) with angle-cut ends (SPEX SamplePrep, Metuchen, NJ, USA) were used for homogenizing the dry tuber sample in 50 mL centrifuge tubes using SPEX Sample Prep 1600 MiniG tissue homogenizer (SPEX SamplePrep, Metuchen, NJ, USA) for 2 min set at a speed of 1500 strokes per minute. Powdered samples were stored at room temperature in tightly closed Minigrip® Red Line zip bags (Minigrip, Alpharetta, GA) until use.

Free amino acids from samples of potato tubers were extracted from two biological replications per environment (location-year combination) and three technical replicates per biological replication using an established protocol^{98,99}. The Kairos™ Amino Acid Kit was used for amino acid calibration (Waters Corp., Milford, MA, USA). Waters Acquity UPLC H-class system with a Waters Xevo TQ mass spectrometer with an electrospray ionization (ESI) probe was used to detect amino acids. Each amino acid's multiple reaction monitoring (MRM) transition, collision energy values, and cone voltage were optimized using IntelliStart software (Waters Corp., Milford, MA, USA). For instrument monitoring and data capture, Water's MassLynx™ software was used. The TargetLynx™ Application Manager (Waters Corp., Milford, MA, USA) was employed for data integration, calibration curves, and amino acid quantification. The relative composition of free amino acid levels (each as a percentage of total free amino acids, e.g., alanine /total) was determined.

Statistical analyses. Descriptive statistics (means, minimums, and maximums) were calculated for free amino acids using the software package META-R¹⁰⁰. Phenotypic distributions for the traits were generated and normality tests were conducted by the Shapiro–Wilk W-Test⁶⁹ using JMP Pro 16® statistical software (SAS Institute, Cary, NC). The free amino acids were separated into groups using a hierarchical clustering analysis using Euclidean distance metric and Ward's minimal variance method. The effect of location, years, and environment

(considered as location-year combination: Dalhart 2019, Dalhart 2020, and Springlake 2020), clone (217 clones), and the corresponding individual amino acid content of raw potato tubers were examined using analysis of variance using a mixed model in JMP Pro 16. Clones were considered fixed effects, while environment (location-year combination), replication within environments, and interactions were considered random.

Similarly, an analysis of variance was used to determine the effect of location (two field sites) on the traits (Dalhart and Springlake field trials conducted in 2020 were used). In this case, clones and locations were treated as fixed effects. Data from 2 years (2019, 2020) from one location (Dalhart) were used to ascertain the impact of years on the traits evaluated. Pearson's correlation coefficients (r) were obtained, and a correlation network was built to examine and illustrate the relationships among free amino acids. A correlation coefficient threshold was set at 0.5 with a p -value of less than 0.05 to visualize the stronger correlations between free amino acids. A significant correlation matrix was generated using the R program (version 4.1.2), and the network visualization was performed with Cytoscape 3.9.1¹⁰¹, using a previously reported method¹⁰². Also, principal component analysis (PCA) was completed using the function `prcomp` in the `stats` package and plotted using the function `ggbiplot`¹⁰³ and `factoextra` package (<https://github.com/kassambara/factoextra>). Confidence ellipses were included, representing 95% confidence intervals around the centroid value of each data cluster. The function `fviz_cos2` was used to visualise the quality of representation (\cos^2) of the variables in the principal components. Pearson correlations (r) between selected amino acid and yield-related traits¹⁰⁴ evaluated in Texas (Dalhart in 2019 and 2020, and Springlake in 2020) were also calculated.

Genome-wide association analysis (GWAS). Association analysis was performed for free amino acids with 10,116 SNPs using the GWASpoly Version 2 package in R⁵⁸. The leave-one-chromosome-out (LOCO) method¹⁰⁵ was used to account for population structure. For each trait, additive and dominant genetic models were tested. A LOD threshold for each trait, corresponding to a genome-wide false-positive rate of 5%, was established based on a Bonferroni test. Manhattan plots were generated using the GWASpoly package. GWASpoly estimated the proportion of phenotypic variance explained by significant SNPs at the peak of each QTL. Contextual sequences of SNPs at the peak of QTL were used in BLAST searches of DM1-3 pseudomolecules (Version 4.03) in the SpudDB database (<http://solanaceae.plantbiology.msu.edu/>) to identify putative candidate genes located in the 100 kbp (kilobase pair) flanking genomic region.

Genomics selection (GS)—Genomic-estimated breeding values. Genetic variance partitioning and genome-wide prediction were performed using allele dosage information^{96,97,104,106}. Standard errors and the best linear unbiased predictors (BLUP) were calculated using standard methods. For a generic random vector \mathbf{u} , $\text{BLUP}[\mathbf{u}] = \hat{\mathbf{u}} = \mathbf{C}\mathbf{P}\mathbf{y}$, where $\mathbf{C} = \text{cov}[\mathbf{u}, \mathbf{y}]$, $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$, \mathbf{X} is the incidence matrix for fixed effects, and \mathbf{V} is the variance–covariance matrix of the response variable¹⁰⁷. The reliability of \hat{u}_i is represented by the squared correlation with the true value which equals $r_i^2 = \frac{\text{var}[\hat{u}_i]}{\text{var}[u_i]}$, and $\text{var}[\hat{\mathbf{u}}] = \mathbf{C}\mathbf{P}\mathbf{C}'$.

Weighted standardized multitrait selection indexes. Standardized multi-trait selection indexes were calculated for all amino acids in the chipping class based on the Z values of GEBVs⁹⁷. The weighted multi-trait selection index for each clone was calculated assigning positive weight for all amino acids except asparagine. Asparagine was assigned a negative weight because lower values of asparagine are preferred by the processing industry.

Data availability

The genotype datasets used during the current study are available in our previous article in Scientific Reports, <https://www.nature.com/articles/s41598-021-87284-x#additional-information>. All other data generated or analyzed during this study are included in this published article.

Received: 23 February 2023; Accepted: 17 August 2023

Published online: 25 August 2023

References

1. Beals, K. A. Potatoes, nutrition and health. *Am. J. Potato Res.* **96**, 102–110 (2019).
2. Camire, M. E., Kubow, S. & Donnelly, D. J. Potatoes and human health. *Crit. Rev. Food Sci. Nutr.* **49**, 823–840 (2009).
3. Woolfe, J. A. *The Potato in the Human Diet* (Cambridge University Press, 1987).
4. Muller, C., Pretorius, B. & Schönfeldt, H. C. Protein quality of South African potatoes to inform dietary choices. *Potato Res.* <https://doi.org/10.1007/s11540-021-09538-5> (2022).
5. Hussain, M. *et al.* Potato protein: An emerging source of high quality and allergy free protein, and its possible future based products. *Food Res. Int.* **148**, 110583 (2021).
6. McGill, C. R., Kurilich, A. C. & Davignon, J. The role of potatoes and potato components in cardiometabolic health: A review. *Ann. Med.* **45**, 467–473 (2013).
7. Gorissen, S. H. M. *et al.* Protein content and amino acid composition of commercially available plant-based protein isolates. *Amino Acids* **50**, 1685–1695 (2018).
8. Wegener, C. B., Jansen, G. & Jürgens, H.-U. Bioactive compounds in potatoes: Accumulation under drought stress conditions. *Funct. Foods Health Dis.* **5**, 108–116 (2015).
9. Koch, W. *et al.* Reduced amino acid content in transgenic potato tubers due to antisense inhibition of the leaf H⁺/amino acid symporter StAAP1. *Plant J. Cell Mol. Biol.* **33**, 211–220 (2003).
10. Brierley, E. R., Bonner, P. L. R. & Cobb, A. H. Aspects of amino acid metabolism in stored potato tubers (cv. Pentland Dell). *Plant Sci.* **127**, 17–24 (1997).

11. Brierley, E. R., Bonner, P. L. R. & Cobb, A. H. Factors influencing the free amino acid content of potato (*Solanum tuberosum* L.) tubers during prolonged storage. *J. Sci. Food Agric.* **70**, 515–525 (1996).
12. Jaynes, J. M., Yang, M. S., Espinoza, N. & Dodds, J. H. Plant protein improvement by genetic engineering: Use of synthetic genes. *Trends Biotechnol.* **4**, 314–320 (1986).
13. Waglay, A., Karboune, S. & Alli, I. Potato protein isolates: Recovery and characterization of their properties. *Food Chem.* **142**, 373–382 (2014).
14. Mottram, D. S. The Maillard reaction: Source of flavour in thermally processed foods. In *Flavours and Fragrances* (ed. Berger, R. G.) 269–283 (Springer Berlin Heidelberg, 2007). https://doi.org/10.1007/978-3-540-49339-6_12.
15. Nursten, H. E. *Maillard reaction* (The Royal Society of Chemistry, 2005). <https://doi.org/10.1039/9781847552570>.
16. Mottram, D. S., Wedzicha, B. L. & Dodson, A. T. Acrylamide is formed in the Maillard reaction. *Nature* **419**, 448–449 (2002).
17. Stadler, R. H. *et al.* Acrylamide from Maillard reaction products. *Nature* **419**, 449–450 (2002).
18. Muttucumaru, N. *et al.* Evidence for the complex relationship between free amino acid and sugar concentrations and acrylamide-forming potential in potato. *Ann. Appl. Biol.* **164**, 286–300 (2014).
19. Parker, J. K. *et al.* Kinetic model for the formation of acrylamide during the finish-frying of commercial french fries. *J. Agric. Food Chem.* **60**, 9321–9331 (2012).
20. Elmore, J. S. *et al.* Changes in free amino acids and sugars in potatoes due to sulfate fertilization and the effect on acrylamide formation. *J. Agric. Food Chem.* **55**, 5363–5366 (2007).
21. Becalski, A. *et al.* Acrylamide in french fries: influence of free amino acids and sugars. *J. Agric. Food Chem.* **52**, 3801–3806 (2004).
22. Amrein, T. M. *et al.* Potential of acrylamide formation, sugars, and free asparagine in potatoes: A comparison of cultivars and farming systems. *J. Agric. Food Chem.* **51**, 5556–5560 (2003).
23. Chawla, R., Shakya, R. & Rommens, C. M. Tuber-specific silencing of asparagine synthetase-1 reduces the acrylamide-forming potential of potatoes grown in the field without affecting tuber shape and yield. *Plant Biotechnol. J.* **10**, 913–924 (2012).
24. Rommens, C. M., Yan, H., Swords, K., Richael, C. & Ye, J. Low-acrylamide french fries and potato chips. *Plant Biotechnol. J.* **6**, 843–853 (2008).
25. Zyzak, D. V. *et al.* Acrylamide formation mechanism in heated foods. *J. Agric. Food Chem.* **51**, 4782–4787 (2003).
26. Elmore, J. S., Dodson, A. T., Bridson, A., Halford, N. G. & Mottram, D. S. The effects of storage on the formation of aroma and acrylamide in heated potato. in *Controlling Maillard Pathways To Generate Flavors* vol. 1042 95–109 (American Chemical Society, 2010).
27. De Wilde, T. *et al.* Influence of storage practices on acrylamide formation during potato frying. *J. Agric. Food Chem.* **53**, 6550–6557 (2005).
28. Halford, N. G. *et al.* Concentrations of free amino acids and sugars in nine potato varieties: Effects of storage and relationship with acrylamide formation. *J. Agric. Food Chem.* **60**, 12044–12055 (2012).
29. Ly, D. N. P., Iqbal, S., Fosu-Nyarko, J., Milroy, S. & Jones, M. G. K. Multiplex CRISPR-Cas9 gene-editing can deliver potato cultivars with reduced browning and acrylamide. *Plants* **12**, 379 (2023).
30. Fennema, O. R. *Food Chemistry* 3rd edn. (CRC Press, 1996).
31. Kumar, P. & Jander, G. Concurrent overexpression of *Arabidopsis thaliana* cystathionine γ -synthase and silencing of endogenous methionine γ -lyase enhance tuber methionine content in *Solanum tuberosum*. *J. Agric. Food Chem.* **65**, 2737–2742 (2017).
32. Huang, T., Joshi, V. & Jander, G. The catabolic enzyme methionine gamma-lyase limits methionine accumulation in potato tubers. *Plant Biotechnol. J.* **12**, 883–893 (2014).
33. Rinder, J., Casazza, A. P., Hoefgen, R. & Hesse, H. Regulation of aspartate-derived amino acid homeostasis in potato plants (*Solanum tuberosum* L.) by expression of *E. coli* homoserine kinase. *Amino Acids* **34**, 213–222 (2008).
34. Dancs, G., Kondrák, M. & Bánfalvi, Z. The effects of enhanced methionine synthesis on amino acid and anthocyanin content of potato tubers. *BMC Plant Biol.* **8**, 65 (2008).
35. Kreft, O., Hoefgen, R. & Hesse, H. Functional analysis of cystathionine γ -synthase in genetically engineered potato plants. *Plant Physiol.* **131**, 1843–1854 (2003).
36. Di, R. *et al.* Enhancement of the primary flavor compound methional in potato by increasing the level of soluble methionine. *J. Agric. Food Chem.* **51**, 5695–5702 (2003).
37. Zeh, M. *et al.* Antisense inhibition of threonine synthase leads to high methionine content in transgenic potato plants. *Plant Physiol.* **127**, 792–802 (2001).
38. Sverrisdóttir, E. *et al.* Genomic prediction of starch content and chipping quality in tetraploid potato using genotyping-by-sequencing. *Theor. Appl. Genet.* **130**, 2091–2108 (2017).
39. Ahmed, S., Zhou, X., Pang, Y., Jin, L. & Bao, J. Improving starch-related traits in potato crops: Achievements and future challenges. *Starch - Stärke* **70**, 1700113 (2018).
40. Carrillo, L. *et al.* Ectopic expression of the AtCDF1 transcription factor in potato enhances tuber starch and amino acid contents and yield under open field conditions. *Front. Plant Sci.* **14**, 1010669 (2023).
41. Shepherd, L. V. T. *et al.* Variation in acrylamide producing potential in potato: Segregation of the trait in a breeding population. *Food Chem.* **123**, 568–573 (2010).
42. Bethke, P. C. Progress and successes of the specialty crop research initiative on acrylamide reduction in processed potato products. *Am. J. Potato Res.* **95**, 328–337 (2018).
43. Jansky, S., Navarre, R. & Bamberg, J. Introduction to the special issue on the nutritional value of potato. *Am. J. Potato Res.* **96**, 95–97 (2019).
44. Friedman, M. Potato glycoalkaloids and metabolites: Roles in the plant and in the diet. *J. Agric. Food Chem.* **54**, 8655–8681 (2006).
45. Roessner-Tunali, U. *et al.* De novo amino acid biosynthesis in potato tubers is regulated by sucrose levels. *Plant Physiol.* **133**, 683–692 (2003).
46. Carreno-Quintero, N. *et al.* Cross-platform comparative analyses of genetic variation in amino acid content in potato tubers. *Metabolomics* **10**, 1239–1257 (2014).
47. Galdón, B. R., Mesa, D. R., Rodríguez, E. M. R. & Romero, C. D. Amino acid content in traditional potato cultivars from the Canary Islands. *J. Food Compos. Anal.* **23**, 148–153 (2010).
48. Syngé, R. L. M. Free amino acids of potato tubers: A survey of published results set out according to potato variety. *Potato Res.* **20**, 1–7 (1977).
49. Kloosterman, B. *et al.* From QTL to candidate gene: Genetical genomics of simple and complex traits in potato using a pooling strategy. *BMC Genom.* **11**, 158 (2010).
50. Zhang, F., Qu, L., Gu, Y., Xu, Z.-H. & Xue, H.-W. Resequencing and genome-wide association studies of autotetraploid potato. *Mol. Hortic.* **2**, 6 (2022).
51. Wilson, S. *et al.* Understanding the effectiveness of genomic prediction in tetraploid potato. *Front. Plant Sci.* **12**, 672417 (2021).
52. Wang, F., Zou, M., Zhao, L., Xia, Z. & Wang, J. Genome-wide association mapping of late blight tolerance trait in potato (*Solanum tuberosum* L.). *Front. Genet.* **12**, 714575 (2021).
53. Díaz, P., Sarmiento, F., Mathew, B., Ballvora, A. & Vásquez, T. M. Genomic regions associated with physiological, biochemical and yield-related responses under water deficit in diploid potato at the tuber initiation stage revealed by GWAS. *PLoS ONE* **16**, e0259690 (2021).

54. Pandey, J. *et al.* Genetic diversity and population structure of advanced clones selected over 40 years by a potato breeding program in the USA. *Sci. Rep.* **11**, 8344 (2021).
55. Khlestkin, V. K., Erst, T. V., Rozanova, I. V., Efimov, V. M. & Khlestkina, E. K. Genetic loci determining potato starch yield and granule morphology revealed by genome-wide association study (GWAS). *PeerJ* **8**, e10286 (2020).
56. Klaassen, M. T. *et al.* Genome-wide association analysis in tetraploid potato reveals four QTLs for protein content. *Mol. Breed.* **39**, 151 (2019).
57. Sharma, S. K. *et al.* Linkage disequilibrium and evaluation of genome-wide association mapping models in tetraploid potato. *G3 Genes Genomes Genetics* **8**, 3185–3202 (2018).
58. Rosyara, U. R., De Jong, W. S., Douches, D. S. & Endelman, J. B. Software for genome-wide association studies in autopolyploids and its application to potato. *Plant Genome* **9**, 1–10 (2016).
59. Hirsch, C. N. *et al.* Retrospective view of north american potato (*Solanum tuberosum* L.) breeding in the 20th and 21st centuries. *G3 Genes Genomes Genetics* **3**, 1003–1013 (2013).
60. Levina, A. V., Hoekenga, O., Gordin, M., Broeckling, C. & De Jong, W. S. Genetic analysis of potato tuber metabolite composition: Genome-wide association studies applied to a nontargeted metabolome. *Crop Sci.* **61**, 591–603 (2021).
61. Jayanty, S. S., Diganta, K. & Raven, B. Effects of cooking methods on nutritional content in potato tubers. *Am. J. Potato Res.* **96**, 183–194 (2019).
62. Talley, E. A., Toma, R. B. & Orr, P. H. Composition of raw and cooked potato peel and flesh: Amino acid content. *J. Food Sci.* **48**, 1360–1361 (1983).
63. Golan-Goldhirsh, A. Effect of the add-back process on the free amino acid pool of potatoes. *Z. Für Lebensm.-Unters Forsch.* **182**, 29–32 (1986).
64. Jaswal, A. S. Effects of various processing methods on free and bound amino acid contents of potatoes. *Am. Potato J.* **50**, 86–95 (1973).
65. Drapal, M. *et al.* Identification of metabolites associated with water stress responses in *Solanum tuberosum* L. clones. *Phytochemistry* **135**, 24–33 (2017).
66. Choi, S.-H., Kozukue, N., Kim, H.-J. & Friedman, M. Analysis of protein amino acids, non-protein amino acids and metabolites, dietary protein, glucose, fructose, sucrose, phenolic, and flavonoid content and antioxidative properties of potato tubers, peels, and cortexes (pulp). *J. Food Compos. Anal.* **50**, 77–87 (2016).
67. Muttucumar, N., Powers, S. J., Elmore, J. S., Mottram, D. S. & Halford, N. G. Effects of water availability on free amino acids, sugars, and acrylamide-forming potential in potato. *J. Agric. Food Chem.* **63**, 2566–2575 (2015).
68. Eppendorfer, W. H. & Eggum, B. O. Effects of sulphur, nitrogen, phosphorus, potassium, and water stress on dietary fibre fractions, starch, amino acids and on the biological value of potato protein. *Plant Foods Hum. Nutr.* **45**, 299–313 (1994).
69. Shapiro, S. S. & Wilk, M. B. An analysis of variance test for normality (complete samples). *Biometrika* **52**, 591–611 (1965).
70. Talley, E. A. & Porter, W. L. Chemical composition of potatoes. VII Relationship of the free amino acid concentrations to specific gravity and storage time. *Am. Potato J.* **47**, 214–224 (1970).
71. Pęksa, A., Miedzianka, J., Nemš, A. & Rytel, E. The free-amino-acid content in six potatoes cultivars through storage. *Molecules* **26**, 1322 (2021).
72. Bethke, P. C. & Bussan, A. J. Acrylamide in processed potato products. *Am. J. Potato Res.* **90**, 403–424 (2013).
73. Dahir, S. S. *et al.* Cold storage reveals distinct metabolic perturbations in processing and non-processing cultivars of potato (*Solanum tuberosum* L.). *Sci. Rep.* **10**, 6268 (2020).
74. Cui, M., Fang, L., Zhou, H. & Yang, H. Effects of amino acids on the physicochemical properties of potato starch. *Food Chem.* **151**, 162–167 (2014).
75. Wang, W., Chen, W., Yang, H. & Cui, M. Textural and rheological properties of potato starch as affected by amino acids. *Int. J. Food Prop.* **20**, S3123–S3134 (2017).
76. Kowalczewski, P. Ł. *et al.* The nutritional value and biological activity of concentrated protein fraction of potato juice. *Nutrients* **11**, 1523 (2019).
77. Davies, A. M. C. The free amino acids of tubers of potato varieties grown in England and Ireland. *Potato Res.* **20**, 9–21 (1977).
78. Gautam, S. *et al.* Raman spectroscopy detects chemical differences between potato tubers produced under normal and heat stress growing conditions. *Front. Plant Sci.* **14**, 1105603 (2023).
79. Hayat, S. *et al.* Role of proline under changing environments. *Plant Signal. Behav.* **7**, 1456–1466 (2012).
80. Iqbal, N., Fatma, M., Khan, N. A. & Umar, S. Chapter 28 - Regulatory role of proline in heat stress tolerance: modulation by salicylic acid. In *Plant Signaling Molecules* (eds Khan, M. I. R. *et al.*) 437–448 (Woodhead Publishing, 2019).
81. Wu, G. *et al.* Proline and hydroxyproline metabolism: implications for animal and human nutrition. *Amino Acids* **40**, 1053–1063 (2011).
82. Ingle, R. A. Histidine biosynthesis. *Arab. Book Am. Soc. Plant Biol.* **9**, (2011).
83. Ali, M. A., Ashraf, M. & Athar, H. R. Influence of nickel stress on growth and some important physiological/biochemical attributes in some diverse canola (*Brassica napus* L.) cultivars. *J. Hazard. Mater.* **172**, 964–969 (2009).
84. Dobson, G. *et al.* Phytochemical diversity in tubers of potato cultivars and landraces using a GC–MS metabolomics approach. *J. Agric. Food Chem.* **56**, 10280–10291 (2008).
85. Roessner, U. *et al.* Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell* **13**, 11–29 (2001).
86. Griffiths, C. A., Reynolds, M. P. & Paul, M. J. Combining yield potential and drought resilience in a spring wheat diversity panel. *Food Energy Secur.* **9**, e241 (2020).
87. Parra-Galindo, M.-A., Piñeros-Niño, C., Soto-Sedano, J. C. & Mosquera-Vasquez, T. Chromosomes I and X harbor consistent genetic factors associated with the anthocyanin variation in potato. *Agronomy* **9**, 366 (2019).
88. Jo, K. R., Choi, J.-G., Kwon, D.-H., Park, Y.-E. & Kim, S.-J. Revealing genetic variations associated with chip-processing properties in potato (*Solanum tuberosum* L.). *Agronomy* **13**, 642 (2023).
89. Pegg, A. E., Xiong, H., Feith, D. J. & Shantz, L. M. S-Adenosylmethionine decarboxylase: Structure, function and regulation by polyamines. *Biochem. Soc. Trans.* **26**, 580–586 (1998).
90. Pegg, A. E. Mammalian polyamine metabolism and function. *IUBMB Life* **61**, 880–894 (2009).
91. Kumar, A., Taylor, M. A., Arif, S. A. M. & Davies, H. V. Potato plants expressing antisense and sense S-adenosylmethionine decarboxylase (SAMDC) transgenes show altered levels of polyamines and ethylene: antisense plants display abnormal phenotypes. *Plant J.* **9**, 147–158 (1996).
92. Kimani, W., Zhang, L.-M., Wu, X.-Y., Hao, H.-Q. & Jing, H.-C. Genome-wide association study reveals that different pathways contribute to grain quality variation in sorghum (*Sorghum bicolor*). *BMC Genom.* **21**, 112 (2020).
93. Anwar, A., She, M., Wang, K., Riaz, B. & Ye, X. Biological roles of ornithine aminotransferase (OAT) in plant stress tolerance: Present progress and future perspectives. *Int. J. Mol. Sci.* **19**, 3681 (2018).
94. Forlani, G., Bertazzini, M., Zarattini, M. & Funck, D. Functional characterization and expression analysis of rice δ 1-pyrroline-5-carboxylate dehydrogenase provide new insight into the regulation of proline and arginine catabolism. *Front. Plant Sci.* **6**, 591 (2015).
95. Verslues, P. E. & Sharma, S. Proline metabolism and its implications for plant-environment interaction. *Arab. Book Am. Soc. Plant Biol.* **8**, e0140 (2010).

96. Endelman, J. B. Fully efficient, two-stage analysis of multi-environment trials with directional dominance and multi-trait genomic selection. *Theor. Appl. Genet.* **136**, 65 (2023).
97. Pandey, J., Scheuring, D. C., Koym, J. W., Endelman, J. B. & Vales, M. I. Genomic selection and genome-wide association studies in tetraploid chipping potatoes. *Plant Genome* **16**, e20297 (2023).
98. Joshi, V. *et al.* Genome-wide association study and population structure analysis of seed-bound amino acids and total protein in watermelon. *PeerJ* **9**, e12343 (2021).
99. Song, Q., Joshi, M., DiPiazza, J. & Joshi, V. Functional relevance of citrulline in the vegetative tissues of watermelon during abiotic stresses. *Front. Plant Sci.* **11**, 512 (2020).
100. Alvarado, G. *et al.* META-R: A software to analyze data from multi-environment plant breeding trials. *Crop J.* **8**, 745–756 (2020).
101. Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
102. Batushansky, A., Toubiana, D. & Fait, A. Correlation-based network generation, visualization, and analysis as a powerful tool in biological studies: A case study in cancer cell metabolism. *BioMed. Res. Int.* **2016**, e8313272 (2016).
103. Wickham, H. Data analysis. In *ggplot2: Elegant Graphics for Data Analysis* (ed. Wickham, H.) 189–201 (Springer International Publishing, 2016).
104. Pandey, J., Scheuring, D. C., Koym, J. W. & Vales, M. I. Genomic regions associated with tuber traits in tetraploid potatoes and identification of superior clones for breeding purposes. *Front. Plant Sci.* **13**, 952263 (2022).
105. Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M. & Price, A. L. Advantages and pitfalls in the application of mixed model association methods. *Nat. Genet.* **46**, 100–106 (2014).
106. Endelman, J. B. *et al.* Genetic variance partitioning and genome-wide prediction with allele dosage information in autotetraploid potato. *Genetics* **209**, 77–87 (2018).
107. Searle, S. R., Casella, G. & McCulloch, C. E. *Variance Components* (Wiley, 2009).

Acknowledgements

The USDA-NIFA (Grant No. 2019-03814), Hatch Project TEX09647, and Texas A&M AgriLife Vegetable Strategic Initiative provided funding for this research project. We acknowledge potato breeding programs at USDA-ARS Centers (Beltsville, Maryland; Aberdeen, Idaho; and Prosser, Washington) and State Universities (University of Wisconsin, Colorado State University, University of Minnesota, and North Dakota State University) for contributing unselected potato families to the Texas A&M University Potato Breeding Program. Additionally, we thank Sanjeev Gautam, Nina Rau, Taylor Swoboda, Mike Jensen, Angel Chappell, and Mythreyi Jamadagni for technical assistance.

Author contributions

J.P.: conceptualization, original draft, data collection, analysis, review, and editing; D.T. and M.J.: extraction and method development, analysis of amino acids; D.C.S. and J.W.K.: data collection, review, and editing; V.J.: conceptualization, funding acquisition, quantification of amino acid, analysis, review, and editing; M.I.V.: conceptualization, funding acquisition, data collection, analysis, review, and editing; All authors approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-40880-5>.

Correspondence and requests for materials should be addressed to V.J. or M.I.V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023