scientific reports

Check for updates

OPEN

Unraveling the genetic architecture of anti-nutritional factors in soybean (*Glycine max*.) for nutritional enhancement

Norberto Jose Palange[1,2,3]✉, Tonny Obua[1], Julius Pyton Sserumaga[4]✉, Enoch Wembabazi[5], Mildred Ochwo-Ssemakula[1], Emmanuel Amponsah Adjei[6], Isaac Onziga Dramadri[1,2], Richard Edema[1,2], Moses Matovu[4], Sharon Valerie Kweyu[2], Arfang Badji[2], Ephraim Nuwamanya[2,5] & Phinehas Tukamuhabwa[1]

Anti-nutritional factors (ANFs) can reduce nutrient bioavailability for monogastric animals. Therefore, this study aimed to understand the genetic architecture underlying ANF accumulation in soybean. Diversity arrays technology and a spectrophotometric method were employed to generate genotypic and phenotypic data, respectively, and gene mining was performed within 100-kb genomic window. A significant difference was found regarding ANFs content in the genotypes ($p < 0.001$). Significant SNP markers for phytate were identified on chromosomes 3, 4, 13, and 20 by FarmCPU, and for total trypsin inhibitors (TTI) on 6, 12, and 14 by CMLM models, whereas mrMLM model detected markers on chromosome 3, 12 and 15 for phytate, 4, 9, 13, 17 and 18 for TTI. Genes associated with phytate content include *Glyma.03G001600*, *Glyma.04G194600*, *Glyma.13G128200*, *Glyma.20G118700*, *Glyma.14G213400*, and *Glyma.16G126400*. For TTI, the genes are *Glyma.06G074700*, *Glyma.12G241600*, *Glyma.14G176700*, *Glyma.13G052700*, and *Glyma.18G050400*. These genes are primarily linked to plant defense and substrate interactions. Most promising SNP markers for marker-assisted selection aimed at reducing phytate levels include Soy_3_218818 (218,818 bp), Soy_3_241209 (241,209 bp), Soy_4_45462019 (45,462,019 bp), Soy_14_48672982 (48,672,982 bp), and Soy_6_5695090 (5,695,090 bp). For TTI, key markers include Soy_14_43649238 (43,649,238 bp), Soy_12_41339023 (41,339,023 bp), Soy_18_4301721 (4,301,721 bp), and Soy_13_14029215 (14,029,215 bp). These findings offer a valuable foundation for marker-assisted breeding aimed at improving soybean nutritional quality.

Keywords Phytate, Trypsin inhibitors, Soybean, SNP markers, GWAS

Soybean [(*Glycine max.*), 2n = 40)] is a preferred crop for addressing nutritional deficiencies in developing countries due to its rich content of protein (40–50%), lipids (20–30%), carbohydrates (26–30%)[1–4], and micronutrients[5]. In addition to these primary metabolites, soybeans produce various secondary metabolites known for their biological roles such as enhancing stress resilience[6], conferring disease resistance[7]. Despite their benefits in plants, secondary metabolites may exert anti-nutritional effects when consumed by monogastric animals depending on the concentrations. Anti-nutritional factors (ANFs) reduce soybean nutritional value by hindering nutrient digestion and absorption[8], thereby affecting human and animal growth[7]. The most important ANFs in legumes include phytate, proteinase inhibitors, tannins, saponins, oligosaccharides, and antigenic factors like oxalate. Among these, proteinase inhibitors (trypsin inhibitors), metal chelates (such as phytate),

[1]Department of Crop Science and Horticulture, College of Agricultural and Environmental Sciences (CAES), Makerere University, P.O. Box 7062, Kampala, Uganda. [2]Makerere Regional Centre for Crop Improvement, College of Agriculture and Environmental Science (CAES), Makerere University, P. O. Box 7062, Kampala, Uganda. [3]Faculty of Food and Agricultural Sciences, Rovuma University, P. O. Box 544, Nampula, Mozambique. [4]National Agricultural Research Organization, National Livestock Resources Research Institute (NaLIRRI), P.O. Box 5704, Kampala, Uganda. [5]National Agricultural Research Organization, National Crops Resources Research Institute (NaCRRI), P. O. Box 7084, Kampala, Uganda. [6]CSIR–Savanna Agricultural Research Institute, P.O. Box TL 52, Tamale, Ghana. ✉email: cheltonpalange@gmail.com; j.serumaga@gmail.com

oligosaccharides, and antigenic factors are typically the most abundant in soybean seeds[9]. Apart from the negative effect of ANFs, it has been reported reduction of nutrient intake and absorption may prevent development of certain diseases. For instance, chelating important cations for glucose transporters such Ca^{2+} ions, a co-factor of α-amylase, phytate (IP6) reduces the rate of starch digestion in humans and animals, preventing diabetes[10] and cancer[11]. The IP6 can also bind directly to starch or to proteins reducing its digestibility, bioavailability, and affect glycemic index value[10]. On the other hand, trypsin or chymotrypsin-inhibitors complexes with enzyme's active site inhibiting their catalytic activity[9], thus, preventing protein breakdown. The protease inhibitors reduce the function of all four classes of proteolytic enzymes, including, serine, cysteine, aspartyl, and metalloproteinases in the gastrointestinal tract of animals[12], affecting growth and triggering pancreas hypertrophy[13]. A study on gene regulatory network aiming to develop low and normal phytate soybean seeds revealed differentially expressed genes in the phytate biosynthetic pathways including *Glyma.11G238800*[14,15], *Glyma.01G016700*, *Glyma.09G206100, Glyma.11G218500, Glyma.18G038800, Glyma.11G218500, Glyma.18G038800*[15]. One QTL with a peak close to *Gm08_44814503* in chromosome 8 was identified using IciMapping analysis. A QTL located between single nucleotide polymorphisms (SNPs) *Gm08_44814503* and *Gm08_45270892* was reported to confer low Kunitz trypsin inhibitor (KTI) concentration in soybean[16].

Several processing methods have been employed to reduce or eliminate ANFs in crops due to their negative impact on animal nutrition. Among the methods, physical, chemical and enzymatic have largely been applied in soybean. Physical and chemical techniques include soaking, cooking, autoclaving, microwave cooking, extrusion, germination, irradiation[17], debranning[18] and dehulling[16], roasting, sprouting[10], whereas enzymatic methods involve fermentation and acetic acid—catalyzed processing[19]. These techniques may be used singly or in combination. Microwaving stands out as a quick, reliable, safe, effective, and environmentally friendly method of lowering ANFs. However, the intensity and length of microwave processing have a considerable impact on ANFs inactivation, and their use needs to be carefully considered[17]. Additionally, though these techniques have proved useful for long, they are costly, time-consuming[20], and some may require technical expertise or generate waste during processing[17]. To overcome these limitations, different breeding strategies are employed to develop soybean cultivars with low anti-nutritional content, including backcrossing[21], mutation breeding[22], molecular markers[23], and genome editing[24].

Traditional breeding systems are often time-consuming, lacks specificity, and ultimately delays variety release[25]. To accelerate genetic gains, a paradigm shift in breeding strategies was necessary. Over the years, morphological and biochemical markers have been widely employed to select genotypes based on traits including yield and quality traits[26–29]. Despite their utility, these markers often show instability due to environmental influences[26]. As a result, molecular markers have opened new avenues for more effective genotype selection. Molecular markers serve as powerful tools for tracking and manipulating genes in both plant and animal breeding[30,31]. More recently, marker-assisted selection (MAS) has gained prominence in soybean improvement programs, offering faster and more precise means of incorporating desirable traits. MAS has been successfully used to develop plants resistant to soybean cyst nematode[32], transfer disease resistance alleles among individuals, and pyramiding resistance alleles[33]. Additionally, MAS has proven useful in the genetic elimination of the Kunitz trypsin inhibitors (KTI) and lectin in soybean seeds[34]. Globally, MAS has been employed in soybean breeding for traits such as sucrose content[35], salt tolerance, insect resistance, agronomic characteristics[36], and pod shattering resistance[37]. Recent advances in gene editing have enabled the development of mutant alleles and molecular markers for *KTI1* and *KTI3* through CRISPR/Cas9-mediated mutagenesis, effectively reducing trypsin inhibitor content and activity in soybean seeds, with no observable difference regarding plant growth or maturity days of *kti1/3* transgenic and wild type plants[38]. Marker efficiency of discovering marker-trait associations has progressively improved from restriction fragment-length polymorphism (RFLP) to single-sequence repeat (SSR)[31]. SSR markers are relatively recent and they have been use to explore genetic diversity in soybean[35–37], and genotyping of Chinese cabbage varieties[13,39]. Though SSRs have contributed to progress in trait diversity and mapping studies, they are regarded to be numerous and polymorphic[9]. Therefore, high-throughput SNP marker genotyping technologies are being extensively adopted to provide genome-wide markers that increase the precision of mapping quantitative trait loci (QTL)[40]. Genome-wide association study (GWAS) has emerged as powerful tool for understanding the genetic basis of phenotypic variance and architecture in crops owing to its capability on the remarkable allele diversity present in natural populations and their historical recombination events. Historically recorded recombination events and rich allele diversity allow for better mapping resolution and causal gene discovery compared to genetic linkage mapping which relies on recent and artificial population with narrow gene pool and low recombination rate[41]. Single nucleotide polymorphism-based genome association study has helped to identify QTLs and genes linked to disease resistance[42,43] in Ugandan soybean accessions. However, no GWAS have been conducted to identify SNP markers linked to anti-nutritional factors (ANFs), despite their negative effect on soybean nutritional quality and contribution to high production costs of soybean meal. Against this background, the study aimed to understand the genetic architecture underlying ANF accumulation in 308 soybean accessions. Addressing this gap is crucial for developing molecular markers to support breeding programs for low-ANF soybean, thereby improving nutritional value and reducing processing costs for food and feed.

Results

Variability of phytate and total trypsin inhibitors

There was significant variation regarding phytate and total trypsin inhibitors content among the genotypes ($p < 0.001$). Mean phytate content was 1756.9 mg/kg [*min*. 14.8 mg/kg (BSPS 48A-6-3) and *max*. 6928.8 mg/kg (NGDT 2.15–7)]. For total trypsin inhibitors (TTI) was 850.3 mg/kg [*min*. 10.9 (DN 16_N); *max*. 1538.5 mg/kg (Duiker)] (Table 1). The observed variation in the genotypes reflect the broad genetic variability of the evaluated population and suggest a genetic control of anti-nutritional factors.

Trait	Min (mg/kg)	Max (mg/kg)	Mean	H^2	SD	CV (%)
Phytate	14.8	6928.8	1756.9	0.68	11.7	0.68
TTI	14.7	1538.5	850.3	0.84	5.4	0.34

Table 1. Summary statistics for phytate and total trypsin inhibitors. H^2, broad sense heritability; SD, standard deviation; CV, coefficient of variation; Min, minimum; Max, maximum; TTI, total trypsin inhibitors.

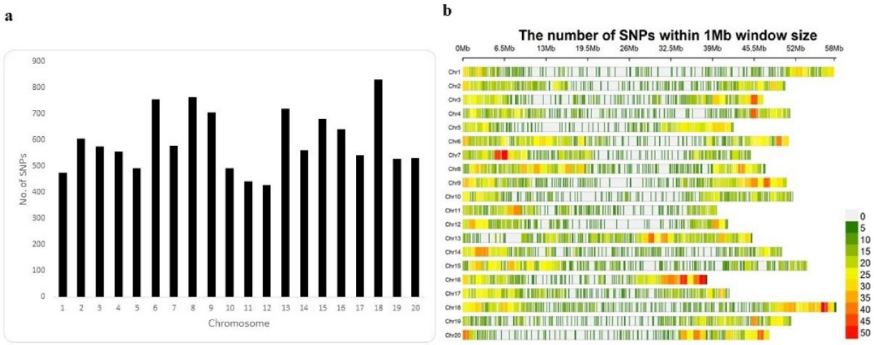

Fig. 1. (**a**): Number of SNPs per soybean chromosome. Chromosome 12 and 18 harbor the lowest and highest number of SNPs, respectively. Panel (**b**) shows the SNP density across soybean genome, where the vertical axis displays the chromosome number, horizontal axis displays chromosome length (1 Mb window), and the various colors represent SNP density or total number of SNPs per window. Chromosomes with high SNP density—such as Chr7, Chr9, Chr16, and Chr18—highlight regions of high genetic variation. These SNP-rich zones (in red) are useful for association mapping, diversity studies, and marker development. Conversely, SNP-poor chromosomes, including Chr2, Chr3, and Chr4, as well as relatively low-density regions on Chr1, Chr10, and Chr11 (green zones), suggest more conserved genomic segments. These regions may reflect low recombination or evolutionary conservation.

Marker distribution across chromosomes

The initial marker size was 17,300 SNPs. Upon SNP duplicates removal and filtering, 11,804 quality SNPs (68.2%) were retained for further analysis. SNPs were distributed fairly evenly along the 20 soybean chromosomes with chromosome 18 having the highest number of markers (824 SNPs) and 12, the lowest (422 SNPs). SNP markers across 20 chromosomes showed variable spacing, with average inter-marker distances ranging from ~ 40 kb (chromosome 16) to ~ 81 kb (chromosome 1). Maximum distances between adjacent SNPs ranged from ~ 951 kb to ~ 2.67 Mb. Chromosomes with greater number of SNPs, such as chromosomes 6, 8, 9, 13, and 18 in soybean, often reflect regions of higher historical recombination or genetic diversity. These regions are beneficial for GWAS as denser SNP coverage improves the ability to detect and fine-map trait-associated loci. In contrast, relatively SNP-poor regions (chromosomes 1, 5, 10, 11, 12, 17, 19 and 20) are often less informative for association studies, though can be biologically important due to their conserved nature (Fig. 1).

Linkage disequilibrium, principal component analysis (PCA) and population statistics

Pairwise correlation estimates between filtered SNPs were performed to assess the rate of linkage disequilibrium (LD) decay. Average LD peaked at $r^2 = 0.2$ and then decayed gradually below $r^2 = 0.1$ at a genetic distance of 50-kb (Fig. 2 a and b), suggesting moderate recombination and genetic diversity in Ugandan germplasm, or the genotypes may have shared common ancestry at some point in time.

The first two principal components (PC1 + PC2) cumulatively explain approximately 15% variation in the population, whereas the first 10 PCs explained up to 31.71% total variation (Fig. 3a). Hierarchical clustering analysis grouped the genotypes into four clusters, reflecting underlying genetic diversity within the soybean germplasm. Cluster 1 comprised 77 genotypes, cluster 2 had the highest representation with 107 genotypes, cluster 3 included 93 genotypes, and cluster 4 contained 31 genotypes. Genotype clusters show distinct geographic compositions. Cluster 1, is the smallest group with 77 genotypes, dominated by genotypes from the USA, accounting for 97.4%, with only one genotype each from Nigeria and Uganda. Cluster 2, the largest with 107 genotypes, is more diverse, including 33.6% from Uganda, 23.4% from Taiwan, 14% from Japan, 13.1% from Nigeria, 11.2% from Zimbabwe, and a small proportion (4.7%) from the USA. Cluster 3 consists mostly of Ugandan genotypes (90.3%), alongside small contributions from Nigeria, Japan, and Zimbabwe. Finally, cluster 4, with 31 genotypes, is primarily Ugandan (58.1%), followed by Nigerian genotypes at 25.8%, and minor representation from Japan and Taiwan (Fig. 3b). The population distribution reflects the genetic diversity and potential regional structuring within the germplasm, with some clusters dominated by specific sources, while others show more admixed origins, highlighting important considerations for breeding and conservation strategies.

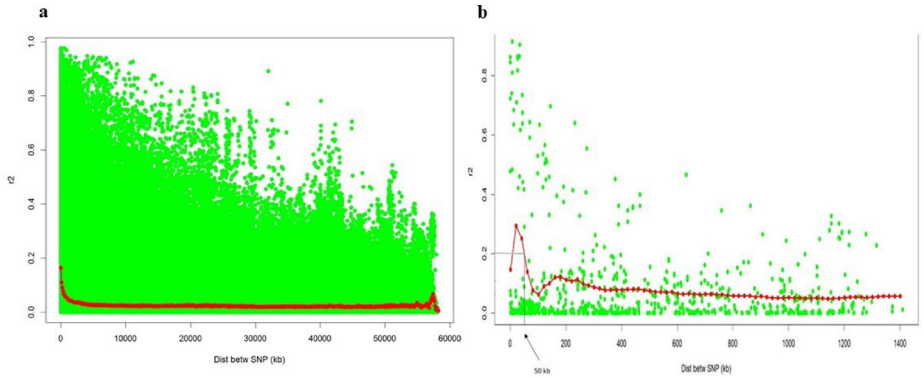

Fig. 2. (**a**)- Average linkage disequilibrium rate. The x-axis shows the distance (kilo base pairs) between SNPs, and the y-axis, the LD value (r^2). Panel (**b**) represents an amplified region from the averaged linkage disequilibrium (**a**) of ~ 1500 kb. LD decay is shown at around 50-kb at $r^2 = 0.2$ and the LD becomes obsolete at around 100-kb.

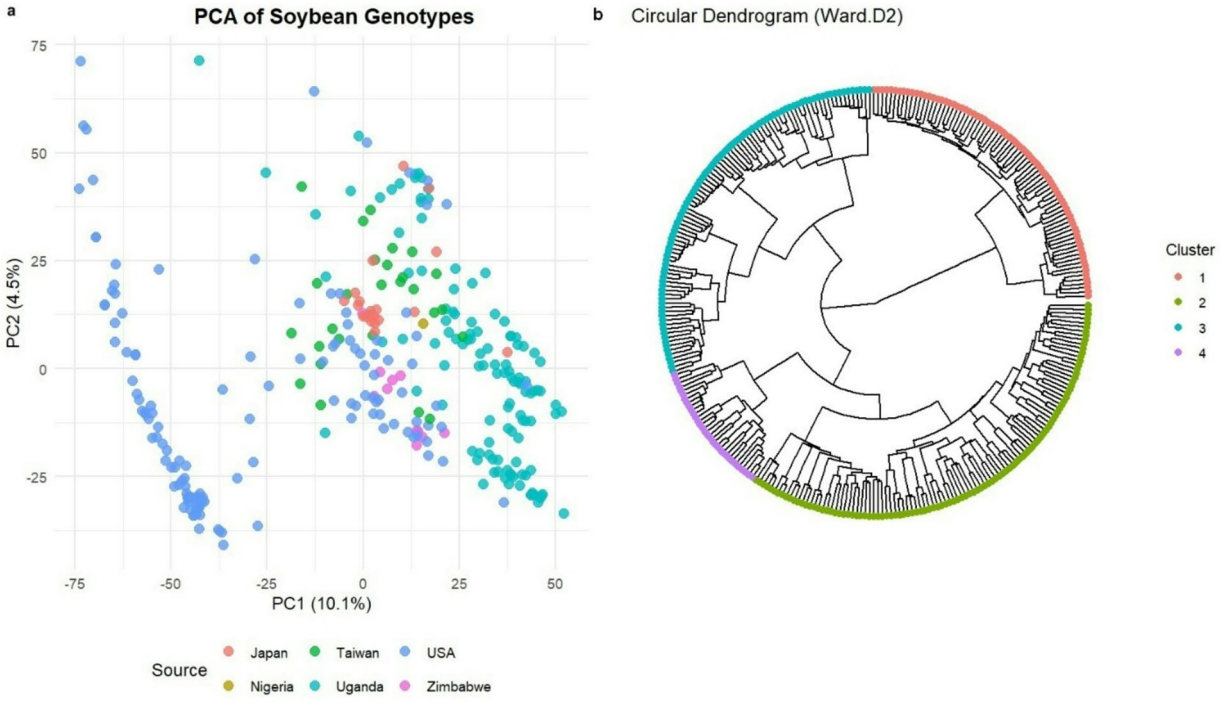

Fig. 3. Principal component analysis (PCA) showing trends of population distribution (**a**) and phylogenetic tree (**b**). The quadrants show a trend of stratification among the genotypes. Numbers 1, 2, 3 and 4 represent four distinct clusters in the population.

The means for genetic diversity (GD), polymorphism information content (PIC), minor allele frequency (MAF), observed heterozygosity (Ho) and inbreeding coefficient (F) were 0.3, 0.25, 0.21 and 0.18, respectively (Table 2). The Ugandan soybean population shows moderate genetic diversity, favorable for breeding and association studies. Moderate PIC and MAF values indicate that the markers are informative. The low observed heterozygosity (Ho) and positive inbreeding coefficient (F) are consistent with soybean's self-pollinating nature.

Marker-trait association

Manhattan plots show the significant SNPs associated with phytate and total trypsin inhibitors. The QQ plots reveal a good control of population parameters, and thus, minimum false positive and negative associations. SNPs above the threshold deviate significantly from the diagonal indicating true associations with the evaluated traits (Fig. 4a–d). Based on the FarmCPU model, phytate accumulation was found to be associated with SNPs located on chromosomes 3 (*pos* 218,818 bp), 4 (*pos* 45,462,019 bp), 13 (*pos* 23,167,455 bp, and 20 (*pos* 35,904,989 bp) (Fig. 4a). The CMLM model revealed SNPs significantly associated with total trypsin inhibitors

Pop stat[a]	Mean	Lower	Upper
GD[b]	0.3	0.09	0.5
PIC[c]	0.25	0.09	0.38
MAF[d]	0.21	0.05	0.5
Ho[e]	0.18	0.06	0.87
F[f]	0.4	− 1.88	0.79

Table 2. Summary of population statistics showing the means, lower and upper values. [a]Population statistics, [b]Genetic diversity; [c]Polymorphism information content; [d]Minor allele frequency; [e]observed heterozygosity; [f] inbreeding coefficient.

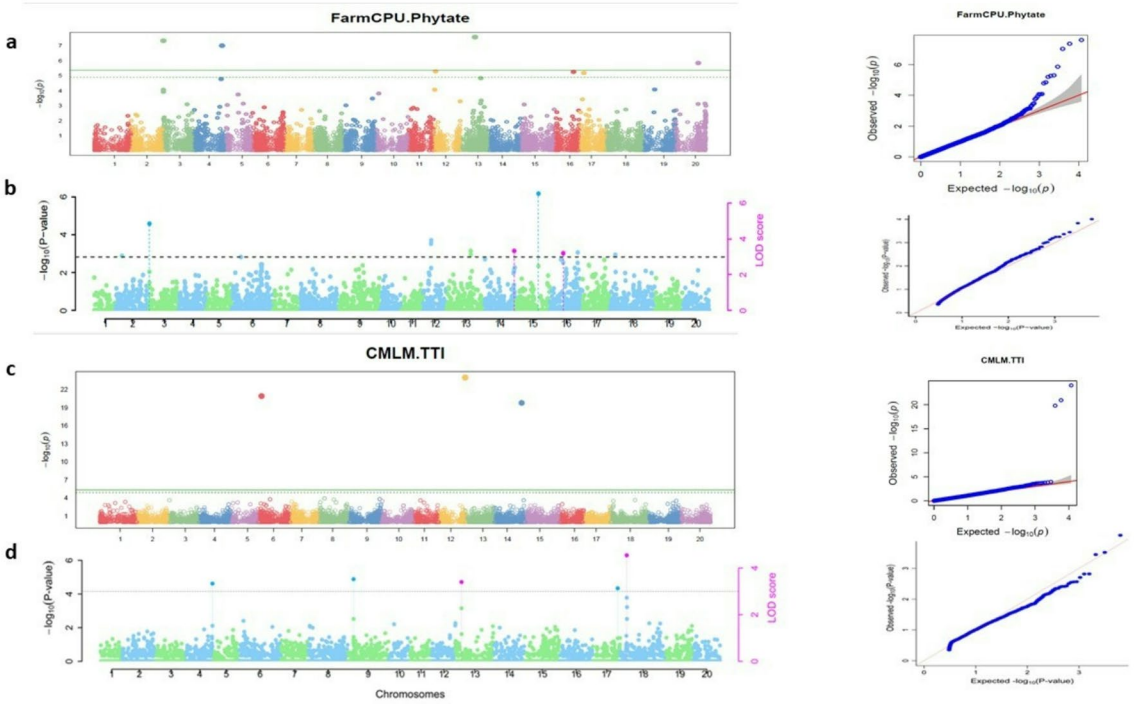

Fig. 4. Manhattan and QQ plots for phytate and total trypsin inhibitors. Significant SNPs have hit the threshold and respective QQ-plot depicts the distribution of observed versus expected *p*-values and the genetic associations (**a**–**d**). Among the models tested in GAPIT, FarmCPU and CMLM were the most effective in detecting significant SNP markers for phytate and TTI, respectively. No common markers were identified between GAPIT models. To assess marker detection power and consistency, six mrMLM methods were also applied to the same dataset. From an inter-model perspective, in general, no overlapping SNPs were detected between GAPIT and mrMLM outputs. However, an intra-model comparison revealed that two SNPs were consistently identified by multiple mrMLM methods (SNPs Soy_14_48672982 and Soy_16_26978144 for phytate; and Soy_13_14029215 and Soy_18_4301721 for TTI) suggesting a higher detection consistency and potential sensitivity of mrMLM methods in capturing trait-associated loci compared to the GAPIT models.

on chromosomes 6 (*pos* 5,695,090 bp), 12 (*pos* 41,339,023 bp), and 14 (*pos* 43,649,238 bp) (Fig. 4c). The SNP marker validation performed using mrMLM confirmed hit for phytate on chromosome 3 (*pos* 241,209 bp) (Fig. 4b). SNP markers located on chromosomes 14 and 16 were detected by at least two methods including mrMLM and FASTmrMLM for phytate. For total trypsin inhibitors, the mrMLM and FASTmrEMMA methods detected SNPs on chromosome 13; mrMLM, pLARmEB and ISIS EM-BLASSO, on chromosome 18 (Fig. 4d). Markers detected by at least two methods were ranked as significantly associated with the trait. Therefore, markers such as Soy_14_48672982 (methods 1 and 2), Soy_16_26978144 (methods 1 and 2), Soy_13_14029215 (methods 1 and 3) and Soy_18_4301721 (methods 1, 4, and 6) were considered most significant and used for gene annotation (Table 3).

Allelic effects of significant SNP markers on phytate and TTI expression

Contribution of phenotypic variation explained by significant SNP markers is illustrated in Figs. 5a–c and 6a–d. Marker–trait association analysis revealed that the expression of phytate and TTI is genotype-dependent. For phytate, SNP Soy_14_46872882 showed significant differences among genotypes ($F(2, 283) = 16.72$, $p < 0.0001$,

Trait	Method	SNP ID[a]	Chr[b]	Position (bp)	Alleles	*p*-value	Effect	PVE (%)[c]
Genome association and prediction integrated tool (GAPIT)								
Phytate	FarmCPU	Soy_3_218818	3	218,818	C/A	4.56E-08	− 5.53556	11.3461
		Soy_4_45462019	4	45,462,019	G/T	9.77E-08	− 3.755	24.9232
		Soy_13_23167455	13	23,167,455	C/A	2.61E-08	− 3.97445	9.0706
		Soy_20_35904989	20	35,904,989	C/A	1.40E-06	3.01311	0
TTI[d]	CMLM	Soy_6_5695090	6	5,695,090	C/T	1.14E-21	3.74803	1.3583
		Soy_12_41339023	12	41,339,023	G/A	9.58E-25	3.84849	7.1343
		Soy_14_43649238	14	43,649,238	A/G	1.56E-20	3.30279	0
Genome association and prediction integrated tool (GAPIT)								
Phytate	mrMLM, FASTmrMLM	Soy_14_48672982	14	48,672,982	G/A	4.0345	− 2.4226	5.7881
	mrMLM	Soy_15_21856770	15	21,856,770	T/A	7.3747	− 3.5413	7.9818
	mrMLM, FASTmrMLM	Soy_16_26978144	16	26,978,144	T/C	3.9594	3.1498	5.4889
	FASTmrMLM	Soy_3_241209	3	241,209	T/C	5.6341	− 1.9019	3.8632
TTI	mrMLM, FASTmrEMMA	Soy_13_14029215	13	14,029,215	C/T	4.4703	− 1.1927	5.0811
	mrMLM, pLARmEB, ISIS EM-BLASSO	Soy_18_4301721	18	4,301,721	A/G	5.6126	2.5593	7.2634
	FASTmrEMMA	Soy_9_4284965	9	4,284,965	C/T	4.249	1.8677	2.706
	pLARmEB	Soy_17_38803367	17	38,803,367	C/A	3.8335	− 0.7015	1.6504
	pLARmEB	Soy_4_50691237	4	50,691,237	AA	4.0484	− 0.9902	5.0311

Table 3. Significant SNPs identified for phytate and total trypsin inhibitors. #[a]Significant SNPs identified for phytate and total trypsin inhibitors; [b]Chromosome; [c]Phenotypic variation explained; [d] total trypsin inhibitors. Lower explained variation indicates that the SNPs play no significant role in determining the target trait. Negative SNP effects are more promising, as they suggest alleles associated with reduced ANF content. Based on GAPIT tool, SNPs with effects − 5.54 (Soy_3_218818), − 3.76 (Soy_4_45462019), and − 3.97 (Soy_13_23167455) for phytate and 3.30 (Soy_14_43649238) for TTI are particularly favorable. Similarly, for mrMLM results, SNPs with effects of − 3.54 (Soy_15_21856770), − 2.42 (Soy_14_48672982), and − 1.90 (Soy_3_241209) for phytate and − 0.7015 (Soy_17_38803367), − 0.9902 (Soy_4_50691237) stand out. These SNPs are ideal targets for marker-assisted selection in breeding programs.

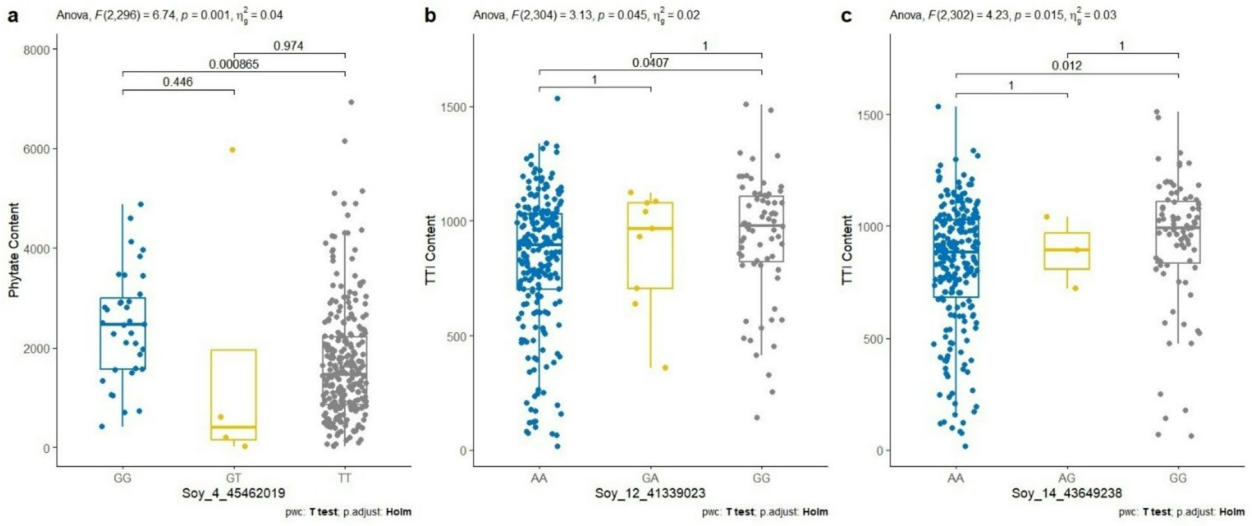

Fig. 5. Allelic effects on SNPs for phytate and TTI accumulation. Marker effect evaluated based on the genotypes of each marker exhibiting significant *p*-values, as identified through GAPIT models are presented in the boxplot "**a**" for phytate, "**b**" and "**c**" for TTI traits.

$\eta^2 = 0.11$), with GA genotypes exhibiting the highest levels and GG, lowest. SNP Soy_16_26978144 also showed a significant effect ($F(2, 293) = 8.27$, $p = 0.00032$, $\eta^2 = 0.05$), where TT genotypes controlling higher phytate content than CC. Additionally, Soy_4_45462019 was significant ($F(2, 296) = 6.74$, $p = 0.001$, $\eta^2 = 0.04$), with TC and TT genotypes exhibiting higher phytate control than CC.

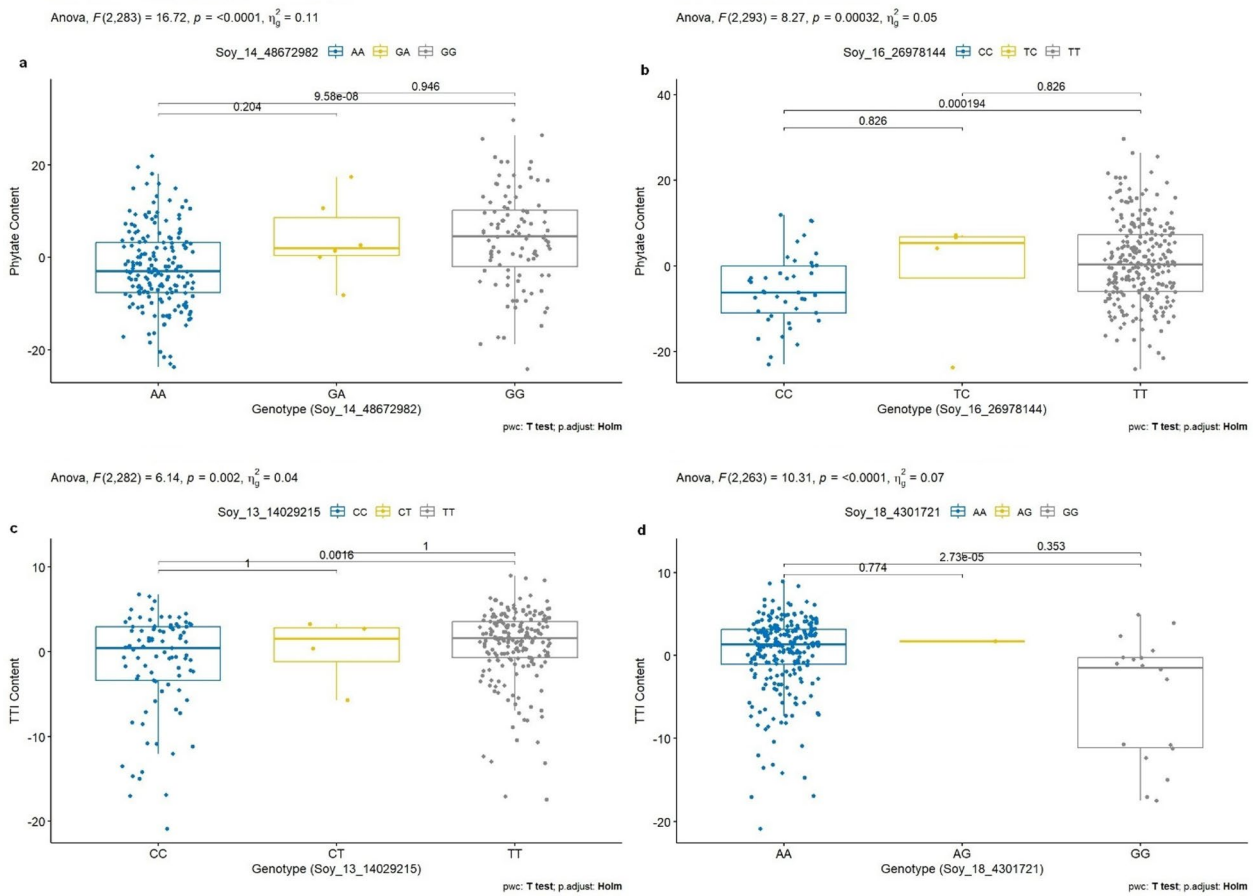

Fig. 6. Allelic effects on SNPs for phytate and TTI accumulation for each marker exhibiting significant *p*-values, as identified by mrMLM methods are presented in boxplots "**a**" and "**b**" for phytate, whereas "**c**" and "**d**" for TTI traits.

Another significant marker, Soy_14_43649238, showed a moderate genotype effect ($F(2, 302) = 4.23$, $p = 0.015$, $\eta^2 = 0.03$), where GG genotypes were associated with increased phytate compared to AA. Soy_12_41339023 also reached significance ($F(2, 304) = 0.045$, $p = 0.045$, $\eta^2 = 0.02$), with TC and TT genotypes having slightly higher phytate control than CC, suggesting a subtle allelic effect.

For TTI, SNP Soy_13_14025215 showed higher expression in CT genotypes compared to TT ($F(2, 282) = 6.14$, $p = 0.002$, $\eta^2 = 0.04$), and Soy_18_4301721 revealed increased TTI in GG over AA genotypes ($F(2, 263) = 10.31$, $p < 0.0001$, $\eta^2 = 0.07$). These findings confirm that allelic variation at specific SNP loci significantly influences phytate and TTI content in soybean.

Candidate genes identification

To investigate the genetic basis of phytate and trypsin inhibitors accumulation, significant SNP markers were identified through GWAS. Gene mining within the 100-kb genomic window revealed genes potentially linked to the targeted traits. The gene functions are classified into major categories including plant defense, gene regulation, substrate–substrate interactions. *Glyma.03G001600 is* potential candidate gene for SNP Soy_3_218818. The gene *Glyma.03G001600* codes for acid phosphatases, which in gene ontology (GO) is categorized as *molecular function*. This class of enzyme is involved in several enzymatic activities transferring phosphate between groups. Phosphate groups can be attached to inositol forming phytate (*myo*-inositol hexakisphosphate or inositol hexaphosphate (IP6)). Phytate (IP6) can act as a precursor in the biosynthetic pathway of diphosphoinositol polyphosphates, a reaction controlled by the gene *Glyma.14G213400* (GO: *molecular function*) which is linked to Soy_14_48672982. Diphosphoinositol is a precursor for phytate biosynthesis. The phytate six-carbon ring substrate can be supplied by hydrolysis of sugars mediated by glycosyl hydrolases family 38 C encoded by the gene *Glyma.16G126400* (GO: *molecular function*) linked to the SNP Soy_16_26978144. SNP Soy_4_45462019 is linked to the gene *Glyma.04G194600* coding for metallo-beta-lactamase superfamily (GO: *molecular function*) involved in hydrolysis of beta-lactam antibiotics. Both phytate and beta-lactamases play a role in plant health. On the other hand, gene *Glyma.18G050400* (GO: *molecular function*) linked to Soy_18_430172, codes for cation efflux proteins, which are found to increase tolerance to divalent metal ions such as cadmium, zinc, and cobalt. The gene *Glyma.13G128200* linked to Soy_13_23167455, codes for protein phosphorylation enzyme family (GO:

biological process) involved in substrate phosphorylation. In such reactions, phytate can act as phosphorus donor, implying a switch of protein function. SNP Soy_20_35904989 is linked to the gene *Glyma.20G118700* coding for glycerophosphodiester phosphodiesterase (GO: molecular function). These enzymes catalyze the hydrolysis of glycerophosphodiesters to produce free alcohol and glycerol 3-phosphate. Depending on the cellular state, glycerol 3-phosphate can be directed to inositol biosynthetic pathway. Trypsin inhibitors are proteins by nature. Gene *Glyma.06G074700* (GO: *molecular function*) linked to Soy_6_5695090 codes for serine protease inhibitor domain. The domain inhibits serine proteases activity by slicing peptide bonds in proteins. Translation initiation factor 2D (EIF2D) is encoded by the gene *Glyma.12G241600* (GO: *molecular function*) linked to the SNP Soy_12_41339023. The EIF2D is involved in the recruitment and delivery of aminoacyl-tRNAs to the P-site of the eukaryotic ribosome in a GTP-independent manner. Gene *Glyma.14G176700* (GO: *molecular function*) linked to Soy_14_43649238 codes for protein kinase domain. The domain contains the catalytic function of protein kinases involved in phosphorylation reactions. Phosphorylation plays a crucial role in regulating transcription, cell cycle progression, and apoptosis. Another transcription factor (TFs), K-box region encoded by *Glyma.13G052700* gene (GO: *molecular function*) linked to the SNP Soy_13_14029215, is a key component of gene regulation through protein–protein interactions. TFs control gene expression related to developmental processes, and the gene *Glyma.13G052700* is expressed in seeds during seeds development (Table 4).

Discussion

Anti-nutritional factors (ANFs) play crucial roles in plant defense[7]. These biologically active compounds are distributed across plant organs including grains, and nuts, leaves, roots, and fruits[79]. However, ANFs may have either positive or negative effect in monogastric animals including humans[80,81] depending on their concentrations in food. ANFs negatively affect nutrient digestibility and absorption by binding to proteins, carbohydrates, lipids and minerals. This study aimed to unveil the genetic basis of two ANFs including phytate and trypsin inhibitors in Ugandan soybean accessions. The observed variability ($p < 0.001$) in ANFs content among the genotypes can be attributed to differences in their genetic background and geographical origin, providing a broad genetic diversity, ideal for selection in breeding programs. Similar studies have reported differences in ANFs in self-pollinated plants at $F_{3:5}$, commercial germplasm, and food-grade soybean lines[9,82].

Higher recombination events were detected on chromosome 18, whereas chromosome 12 exhibited greater conservativeness. A study carried out to reveal the genetic basis for resistance to *Coniothyrium glycines* in the same population (~ 10% difference in size) reported similar recombination patterns[42]. Genetic regions with high recombination rates offer potential opportunity for studies to disclose more biological functions associated with the region. High rates of recombination break down linkage disequilibrium (LD). Thus, genetic architecture of complex traits are better investigated using LD analysis[83]. In this study, LD decay rate was estimated at $r^2 = 0.10$ for a genetic distance of 50-kb. For self-pollinated crops including soybean the LD decay rate is generally low at around $r^2 = 0.10$[84]. However, LD decay at $r^2 = 0.2$ within approximately 200-kb has also been reported[85]. The low LD in this study indicate a weak association between the markers, implying high recombination rate leading to independent segregation of alleles. This recombination rate may explain the separation of the population into distinct groups in the phylogenetic tree, which can be beneficial for breeding purposes.

Molecular markers are more powerful tools for assessing genetic diversity (GD) compared to phenotypic markers. The genetic diversity value of 0.3 and a polymorphism information content (PIC) value of 0.25 suggest moderate genetic variability and marker informativeness in the population. Similar GD and PIC values of 0.34 and 0.27, respectively, has been reported in soybean cultivars and advanced breeding lines from the U.S. and China[86]. Conversely, a slightly lower PIC of 0.22 was reported in a study using Korea, Japan, China, and the U.S populations[87], whereas, studies in soybean novel germplasm, advanced lines and cultivars released for commercial cultivation in Sub-Saharan Africa detected higher PIC = 0.38[40], GD of 0.70 and PIC of 0.71[88]. These differences may be explained by the population size and background, continuous selection for specific trait in breeding programs, along with the number and diversity index of markers, and geographic dispersion of accessions[87].

The smaller minor allele frequency (MAF) captured in this study suggest that most of the loci are nearly fixed, different from the higher MAF (MAF = 0.29) reported in the same population (with smaller size)[42].

Unlike in this study, quantitative trait loci (QTL) for Kunitz trypsin inhibitor (KTI) have been reported on chromosome 8[9], whereas SNP markers for phytate was identified on chromosomes 1, 9, 11, and 18 in soybean[15]. These discrepancies in SNP markers detection can be attributed to differences in phenotyping, population type and association tools.

Gene mining within the 100-kb range revealed 45 genes potentially linked to the targeted traits. The genes are associated with plant defense, gene regulation, and substrate–substrate interactions. For instance, phytate is linked to biosynthesis of abscisic acid (ABA) and gibberellins, two phytohormones involved in seed germination[89]. Resistance to disease has been associated with chromosome 13[42], resistance to abiotic stress to chromosome 1[90]. Genes on chromosome 19 has been associated with gene regulation[42], and PHD finger transcription factors were reported to be located on chromosome 12[91]. Transcription factors regulate gene expression during protein biosynthesis. Trypsin inhibitors, which are proteins by nature, play a critical role in plant defense and overall metabolism. Post-translational regulation can occur through phosphorylation or dephosphorylation of enzymes. The gene *Glyma.03G001600* encodes acid phosphatases involved in breaking down adenosine triphosphate (ATP) to release phosphate groups as sub-product[44,45]. Phosphate kinases (encoded by *Glyma.14G214700*) can be involved in phytate biosynthesis by incorporating phosphate groups into inositol in the *1L-myo-inositol-1,2,3,4,5,6-hexakis* (dihydrogen phosphate)[92], whereas the *Glyma.16G126400*, a glycosyl hydrolases gene, supply sugar backbone for *inositol* biosynthesis. These metabolic pathways are interconnected and together contribute to phytate biosynthesis in plants. On the other hand, metallo-beta-lactamase superfamily encoded by the gene *Glyma.04G194600*[49] is involved in plant immunological responses,

Trait	Method	SNP ID	Chr	Pos (Mb)	Gene ID	Functional annotation	PFAM	Reference
Genome association and prediction integrated tool (GAPIT)								
Phytate	FarmCPU	Soy_3_218818	Gm03	218,818	*Glyma.03G001600*	Acid phosphatases	PF03767	44,45
			Gm03		*Glyma.03G009000*	Phosphatidylinositol N-Acetylglucosaminyltransferase	PF12552	46
			Gm03		*Glyma.03G010200*	Plant phosphoribosyltransferase C-terminal (PRT_C)	PF08372	45
			Gm03		*Glyma.03G002200*	Protein kinases	PF00069	47
			Gm03		*Glyma.03G002000*	O-Glycosyl hydrolases	PF16499	48
		Soy_4_45462019	Gm04	45,462,019	*Glyma.04G194600*	Metallo-beta-lactamase superfamily	PF00753	49
			Gm04		*Glyma.04G195633*	Lycopene epsilon-cyclase	PF05834	50
			Gm04		*Glyma.04G206100*	Phosphatidylinositol 4-phosphate 5-kinase 1-related	PF01504	51
			Gm04		*Glyma.04G203600*	Protein Suppressor of Gene Silencing 3 (SGS3)	PF03468	52
			Gm4		*Glyma.04G205200*	Defensin-like protein 6 (DEFL6)	PF00304	53,54
			Gm4		*Glyma.04G196900*	F-box domain	PF00646	55
			Gm4		*Glyma.04G194600*	Beta-lactamase	PF00753	49
		Soy_13_23167455	Gm13	23,167,455	*Glyma.13G128200*	Protein phosphorylation	PF07714	56
			Gm13		*Glyma.13G127800*	Tetratricopeptide repeat	PF07719	57
			Gm13		*Glyma.13G132400*	Inorganic pyrophosphatase	PF00719	58
			Gm13		*Glyma.13G132700*	Phosphatidylinositol-bisphosphatase	PF03372	46
		Soy_20_35904989	Gm20	35,904,989	*Glyma.20G118700*	Glycerophosphodiester phosphodiesterase	PF00069	56
					Glyma.20G118000	Metallo-beta-lactamases	PF00753	49
					Glyma.20G119100	Protein phosphatase 2C 5-related	PF00481	47
TTI	CMLM	Soy_6_5695090	Gm06	5,695,090	*Glyma.06G074700*	Serine protease inhibitor domain	PF00067	59,60
					Glyma.06G074300	Kunitz/Bovine pancreatic trypsin inhibitor (BPTI) domain	PF13639	59
					Glyma.06G077400	WRKY transcription factor 11-related	PF10533	61
					Glyma.06G078100	Thioredoxin isoform B-related	PF07649	62
					Glyma.06G074300	Ring-Type Zinc finger	PF00010	63
		Soy_12_41339023	Gm12	41,339,023	*Glyma.12G241600*	Translation initiation factor 2D	PF01253	64
					Glyma.12G241600	Translation initiation factor SUI1	PF01253	64
					Glyma.12G240800	Ring finger domain	PF13920	65,66
		Soy_14_43649238	Gm14	43,649,238	*Glyma.14G176700*	Protein kinase domain	PF00069	47
					Glyma.14G178200	Ring finger domain	PF13639	66
					Glyma.14G185100	Helix-loop-helix DNA-binding domain	PF00010	63
	Multi-locus random-SNP-effect mixed linear model (mrMLM)							
Phytate	mrMLM, FASTmrMLM	Soy_14_48672982	Gm14	48,672,982	*Glyma.14G213400*	Diphosphoinositol Polyphosphate Phosphohydrolase	PF00293	56,67
					Glyma.14G214700	Protein kinase domain	PF00069	68
					Glyma.14G224300	Cyclin, N-terminal domain	PF00134	68,69
	mrMLM, FASTmrMLM	Soy_16_26978144	Gm16	26,978,144	*Glyma.16G126400*	Glycosyl hydrolases family 38 C-terminal domains	PF07748	70
					Glyma.16G128300	TSL-Kinase interacting protein 1	PF00249	71
TTI	mrMLM, FASTmrEMMA	Soy_13_14029215	Gm13	14,029,215	*Glyma.13G052700*	K-box region	PF01486	56
					Glyma.13G053600	Protein tyrosine kinase (Pkinase_Tyr)	PF07714	72
					Glyma.13G053733	U-Box domain-containing protein 50-related	NA	
					Glyma.13G055200	Eukaryotic translation initiation factor 3 subunit M	PF01399	73
					Glyma.13G055400	protein SPIRAL1 and related proteins (SPR1)	NA	
	mrMLM, pLARmEB, ISIS EM-BLASSO	Soy_18_4301721	Gm18	4,301,721	*Glyma.18G050400*	Cation efflux family	PF01545	74
					Glyma.18G051500	AP2 domain (transcription factors)	PF00847	75
					Glyma.18G051600	KIP1-like protein	PF07765	76
					Glyma.18G050300	Homodimerization region of STAR domain protein	PF16544	77
					Glyma.18G050600	Ribosomal protein L16p/L10e	PF00252	78

Table 4. Gene annotation for the significant SNPs linked to phytate, trypsin inhibitors in the soybean. Candidate genes along with their functions are detailed below. #SNP ID, single nucleotide polymorphism identifier; Chr, chromosome; Pos, position in *megabytes*; Gene ID, gene identifier; PFAM, protein families.

suggesting its connection with phytate role. Protein kinases encoded by *Glyma.13G128200 gene* are alternatively involved in phytate biosynthetic, whereas *Glyma.20G118700* encodes glycerophosphodiester phosphodiesterase involved in glycerol esters hydrolysis[56]. This pathway provides glycerol to the inositol biosynthesis[92]. Protein kinase domain encoded by *Glyma.14G176700* plays a pivotal role in cellular regulation of phosphorus[47]. Thus,

the interplay between phosphorylation and dephosphorylation determine the levels of phytate or inositol available in plants cells.

Protein inhibitors are classes of serine proteases encoded by the gene *Glyma.06G074700*. Protein biosynthesis can be regulated by K-box domain, encoded by *Glyma.13G052700* or translation initiation factor 2D encoded by *Glyma.12G241600*[64]. These domains ensure accurate tRNA placement during translation or post-translational modifications. Cation efflux family encoded by *Glyma.18G050400* export or redistribute positively charged ions across the cell membrane determining protein biosynthesis efficiency. Protease inhibitors block digestive enzyme activity by competing with substrates for the active site. This enzyme-inhibitor complex affects the ability of animals to break down ingested proteins. Undigested proteins are unable to be absorbed by the intestinal tract, thereby affecting animal growth.

Among the GAPIT models tested, FarmCPU[93] and CMLM[94] were most effective for detecting significant SNPs. The power of FarmCPU has previously been reported in a study comparing multiple GWAS models in soybean and maize. FarmCPU performed better than single-locus models by reducing false positives and false negatives[95]. The mrMLM methods showed higher detection consistency and potential sensitivity in capturing trait-associated loci compared to the GAPIT models. The potential of mrMLM methods was also reported applying 3VmrMLM (Three Variance components multi-locus random-SNP-effect Mixed Linear Model) aiming to dissect the genetic mechanism in rice[96].

To support these findings for MAS, further studies with larger populations across environments are needed to fully validate the discovered SNP markers. Following proteomic studies to validate the marker expression, the annotated genes can be used for achieving faster genetic progress for low anti-nutritional.

Conclusion

This study identified SNPs and candidate genes linked to phytate and total trypsin inhibitors (TTI) in soybean. Potential marker for low phytate content, include Soy_14_46872882, where the GG genotype consistently exhibited the lowest phytate levels. Likewise, Soy_16_26978144 and Soy_4_45462019 were associated with lower phytate accumulation in genotypes carrying the CC allele. Although Soy_12_41339023 showed a marginal effect, the CC genotype still demonstrated comparatively reduced phytate levels, and may offer value when combined with other markers in a selection strategy. For TTI, Soy_13_14025215 was linked to lower TTI levels in TT genotypes, while Soy_18_4301721 showed AA genotype as favorable effect in reducing TTI. Another promising marker, Soy_14_43649238, showed genotype-specific influence, with one genotype group presenting lower TTI levels. The identified markers and genotypes can be useful for marker-assisted selection (MAS) in breeding programs aiming to develop soybean varieties with reduced anti-nutritional content. However, validation across larger populations and environments, combined with proteomic studies, are essential to confirm marker effectiveness. By enhancing our understanding of the genetic basis of ANFs, this research paves the way for the development of nutritionally superior soybean cultivars that contribute to sustainable agriculture and food security.

Methods

Plant materials

A set of 308 soybean germplasm was obtained from the Makerere University Centre for Soybean Improvement and Development (MAKCSID) program. The collection is composed of lines sourced from Uganda (136), the United States (80), Taiwan (27), Japan (19), Zimbabwe (13), and Nigeria (33). To standardize the conditions and ensure consistent seed multiplication in this exploratory assay, genotypes were planted in 2023B at Makerere University Agricultural Research Institute Kabanyolo (MUARIK). Kabanyolo is geographically located in the Central Region of Uganda at the coordinates 0° 28′ N, 32° 36′ E, altitude of 1180 m above sea level. The mean annual temperature is 21.4 °C, and the mean annual rainfall is 1234 mm[97].The experiment was laid out in an augmented design with 31 blocks, each containing 10 plots consisting of three rows. Surfaces of young and apparently healthy leaves were cleaned with 70% ethanol and eight leaf discs obtained using a punch gun. Samples were incubated under 37 °C until they were sent to SEQART AFRICA located at International Livestock Research Institute in Nairobi for genotyping.

DNA extraction and diversity arrays technology "genotyping-by-sequencing" (DArTseq)

DNA extraction was performed using Nucleomag plant Kit, with concentrations of genomic DNA in the range of 50–100 ng/µl. DNA quality and quantity were checked on 0.8% gel agarose[98]. The DArTSeq complexity reduction method was used through the digestion of genomic DNA using a combination of PstI and MseI enzymes and ligation of barcoded adapters and common adapter followed by PCR amplification of adapter-ligated fragments[99]. Libraries were constructed through Single Read sequencing runs for 77 bases. Next-generation sequencing was carried out using Illumina the HiSeq 2500 platform (Illumina, Inc., Model HiSeq 2500, Rapid Run Mode). Genome profiling was conducted by Genotyping by Sequencing (GBS) DArTseq™ technology (Canberra, ACT, Australia). DArTseq markers scoring was achieved using DArTsoft14 which is an in-house marker scoring pipeline based on algorithms. Two types of DArTseq markers were scored, SilicoDArT markers and SNP markers which were both scored as binary 1 for presence and 0 absence of the restriction fragment with the marker sequence in the genomic representation of the sample. Both SilicoDArT markers and SNP markers were aligned to soybean- Wm82-a1-v4 reference genome to identify chromosome positions[99–101].

Biochemical analysis

Soybean seeds were ground using a food processor grinder (FOSS, Brook Crompton Laboratory mill, type 2-TDAB03J, England, 2014) at the Nutritional and Biochemical Laboratory of the National Crops Resources

Research Institute (NaCRRI) in Uganda. The fine ground samples were further used to determine phytate and total trypsin inhibitors content.

Phytate phenotyping

Phytate extraction followed an acidic method as described by Israel[102], with slight modifications on the sample initial volume and refrigeration during centrifugation. Briefly, 20 ml of 0.5 M HCl was added to 0.5 g of finely powdered soybean sample. The mixture was vortexed and shaken for 1 h at room temperature, then centrifuged at 4000 rpm for 30 min at 4 °C using a HERMLE 300 K centrifuge (Germany). For each sample, 5 ml of the supernatant containing the soluble fraction was transferred into three new 50 ml Falcon tubes, and the pH was adjusted by adding an equal volume of 0.5 M NaOH. To stop the reaction, equal volume 5 ml of NaCl was added to the system. Samples were filtered, and 2 ml of a chromogenic solution (ferric chloride III, $FeCl_3$) was added for spectrophotometric reading at 492 nm absorbance (Ledetect96 Microplate Reader, 2017, A-5020, EU). The optical densities (OD) from the readings were used to calculate the concentration of phytate in the samples, using a linear regression curve obtained from serial dilutions of standard solution of sodium phytate. The controls were similarly prepared, except addition of the analyte and used as subtraction term to obtain final concentration (mg/kg) of phytate in the sample.

Total trypsin inhibitor phenotyping

Total trypsin inhibitors (TTI) were extracted using a neutral method with slight modifications on the sample size, volume and refrigeration during centrifugation[103]. Briefly, 0.05 g of powdered soybean sample was placed in a clean centrifuge tube and homogenized with 5 ml of phosphate buffered saline (PBS). The mixture was vortexed and shaken on an orbital shaker for 1 h, then centrifuged at 10,000 rpm for 10 min at 4 °C using a HERMLE 300 K centrifuge (Germany). A 0.1 ml aliquot of the supernatant was transferred into micro centrifuge tubes and mixed with equal volume of 1 mg/ml trypsin solution, followed by incubation at 0 °C for 10 min. Then, 0.3 ml of 2% casein substrate was added, and mixture was incubated in a water bath for 20 min at 37 °C[103,104]. The reaction was stopped by adding 0.2 ml of 10% trichloroacetic acid (TCA), and centrifuged at 10,000 rpm for 5 min to remove undigested casein, larger inhibitor fragments, and enzyme protein. All extraction steps were performed in triplicate. Two control samples were set, one without any inhibitor addition (Blank 1), other with addition casein only (Blank 2). Sample readings were performed using a spectrophotometer at 410 nm absorbance (Ledetect96 Microplate Reader, 2017, A-5020, EU). Trypsin inhibition activity (TIA) was calculated[103] and the inhibitory activity percentage was converted into concentration of total trypsin inhibitor (TTI) expressed in mg/kg:

$$\%inhibition = \frac{test\ sample - blank2}{blank1 - blank2} * 100$$

Data analysis

Best Linear Unbiased predictions (BLUPs) were computed using *lme4* R package[105] considering genotype as fixed and block as random effects as follow: $Yij = \mu + Bi + Gj + G : Bij + \varepsilon ij$; where: where Yij = phenotypic observation for a trait, μ = grand mean, B = random effect of block (j), G = fixed effect of genotype (i), G:B = interaction effect between genotype (i), and ε ij = random residual term. The resulting BLUPs were then used for association analysis[106]. Heritability was calculated for phytate and trypsin inhibitors as follow: $H^2 = V_g/(V_g + V_e)$; where: H^2 is the broad sense heritability, V_g is the genotypic variance, and V_e is the error (residual) variance.

Linkage disequilibrium (LD)

The extent of linkage disequilibrium (LD) was determined as the pairwise correlations between each pair of SNPs using *LDcorSV* package in R version 4.3.0 (Zhang et al., 2023). Pairwise correlation coefficients among markers were then plotted against genetic distances in kilobases (kb). The genetic distance at which average LD decayed below $r^2 = 0.1$ was taken as the window for searching of putative genes within 100-kb genomic region of significant SNP markers on Phytozome database, version 13.0 (https://phytozome-next.jgi.doe.gov/) accessed on 11 June, 2024, using soybean-*Wm82-a4.v1* as reference genome. Gene functions were found using *InterPro* database hosted by the European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI) and National Center for Biotechnology Information (NCBI).

Principal component analysis

Principal Component Analysis (PCA) and hierarchical clustering were performed using *tidyverse*, *factoextra*, and *ggtree* R packages. SNP data were cleaned and imputed using the *dplyr* package, replacing missing values with marker-wise means. PCA was executed using *prcomp*, with scaling and centering applied. The proportion of variance explained by PC1 and PC2 was extracted and used for axis labeling. Euclidean distances among genotypes were calculated, and hierarchical clustering was conducted using the *Ward*.D2 method via *hclust*. A circular dendrogram was constructed using *ggtree*, and genotype clusters were colored accordingly. These visualizations revealed population structure and genetic diversity patterns.

GWAS analysis, linkage disequilibrium and candidate genes identification

Data filtration was performed with a threshold of 95% reproducibility and 95% call rate, 0.05 minor allele frequencies (MAF < 0.05) and imputation through k-nearest neighbor imputation (*knni*) using *snpReady* package in R version 4.3.0. Duplicates were removed from filtered SNP data using the *duplicated* function of *dplyr* package in R. Genome-wide association study (GWAS) was performed using Genome Association and

Prediction Integrated Tool (GAPIT) models, including Fixed and random model Circulating Probability Unification (FarmCPU) for phytate and compressed mixed linear model (CMLM)[93,94] for TTI. Furthermore, to assess models robustness in detecting SNP markers, multi-locus random-SNP-effect mixed linear model (mrMLM) methods including mrMLM[107], FASTmrMLM[108], FASTmrEMMA[109], pLARmEB[110], pKWmEB[111] and ISIS EM-BLASSO[112] were used for marker-trait association using *mrMLM* package in R[113]. These methods are reported to maintain computational advantage and increases statistical power. The GAPIT models were fitted with varying numbers of PCs and without any PC to teste for correction of spurious associations which could potentially arise due to population structure. Correction for kinship was performed using the *VanRaden* method, andManhattan and quantile–quantile (QQ) plots were generated to visualize outputs of the analysis. Boxplots were generated to visualize the allelic effects. The *ggpubr* package was used to display group comparisons, with Holm-adjusted *p*-values and 95% confidence intervals annotated on the plots.

Data availability

The phenotypic and genotypic datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request. Candidate genes were identified using publicly available information from the Phytozome database (https://phytozome-next.jgi.doe.gov/), NCBI https://www.ncbi.nlm.nih.gov/, and EMBL-EBI (https://www.ebi.ac.uk/).

Received: 6 September 2024; Accepted: 31 October 2025
Published online: 28 November 2025

References

1. Gibbs, B. F., Zougman, A., Masse, R. & Mulligan, C. Production and characterization of bioactive peptides from soy hydrolysate and soy-fermented food. *Food Res. Int.* **37**, 123–131 (2004).
2. Murithi, H. M., Beed, F., Tukamuhabwa, P., Thomma, B. P. H. J. & Joosten, M. H. A. J. Soybean production in eastern and southern Africa and threat of yield loss due to soybean rust caused by *Phakopsora pachyrhizi*. *Plant Pathol.* **65**, 176–188 (2016).
3. Rui, X. et al. Optimization of soy solid-state fermentation with selected lactic acid bacteria and the effect on the anti-nutritional components. *J. Food Process. Preserv.* https://doi.org/10.1111/jfpp.13290 (2017).
4. Guo, B. et al. Soybean genetic resources contributing to sustainable protein production. *Theor. Appl. Genet.* **135**, 4095–4121 (2022).
5. Malle, S., Morrison, M. & Belzile, F. Identification of loci controlling mineral element concentration in soybean seeds. *BMC Plant Biol.* **20**, 1–14 (2020).
6. Kumar Sharma, R. Production of secondary metabolites in plants under abiotic stress: An overview. *Signif. Bioeng. Biosci.* **2**, 196–200 (2018).
7. Sinha, K., Khare, V., Scientist, J. & Bharti, L. Review on: Antinutritional factors in vegetable crops. *Pharma Innov. J.* **6**, 353–358 (2017).
8. Mohapatra, D., Patel, A. S., Kar, A., Deshpande, S. S. & Tripathi, M. K. Effect of different processing conditions on proximate composition, anti-oxidants, anti-nutrients and amino acid profile of grain sorghum. *Food Chem.* **271**, 129–135 (2019).
9. Rosso, M. L. et al. Development of breeder-friendly KASP markers for low concentration of kunitz trypsin inhibitor in soybean seeds. *Int. J. Mol. Sci* **2675**, 1–16 (2021).
10. Duraiswamy, A. et al. Genetic manipulation of anti- nutritional factors in major crops for a sustainable diet in future. *Front. Plant Sci.* **13**, 1–26 (2023).
11. Vucenik, I. & Shamsuddin, A. M. Protection against cancer by dietary IP6 and inositol. *Nutr. Cancer* **55**, 109–125 (2006).
12. Haq, S. K., Atif, S. M. & Khan, R. H. Protein proteinase inhibitor genes in combat against insects, pests, and pathogens: Natural and engineered phytoprotection. *Arch. Biochem. Biophys.* **431**, 145–159 (2004).
13. Miladinović, J., Burton, J. W., Tubić, S. B. & Miladinović, D. Soybean breeding: Comparison of the efficiency of different. *Turk J Agric* **35**, 469–480 (2011).
14. Jin, H., Yu, X., Yang, Q., Fu, X. & Yuan, F. Transcriptome analysis identifies differentially expressed genes in the progenies of a cross between two low phytic acid soybean mutants. *Sci. Rep.* **11**, 1–14 (2021).
15. DeMers, L. C., Raboy, V., Li, S. & Saghai Maroof, M. A. Network inference of transcriptional regulation in germinating low phytic acid soybean seeds. *Front. Plant Sci.* **12**, 1–17 (2021).
16. Zhong, Y., Wang, Z. & Zhao, Y. Impact of radio frequency, microwaving, and high hydrostatic pressure at elevated temperature on the nutritional and antinutritional components in black soybeans. *J. Food Sci.* **80**, C2732–C2739 (2015).
17. Suhag, R. et al. Microwave processing: A way to reduce the anti-nutritional factors (ANFs) in food grains. *LWT* **150**, 111960 (2021).
18. Samtiya, M., Aluko, R. E. & Dhewa, T. Plant food anti-nutritional factors and their reduction strategies: An overview. *Food Prod. Process. Nutr.* **5**, 1–14 (2020).
19. Huang, L. & Xu, Y. Effective reduction of antinutritional factors in soybean meal by acetic acid-catalyzed processing. *J. Food Process. Preserv.* **42**, 1–8 (2018).
20. Zubko, V. et al. Inactivation of anti-nutrients in soybeans via micronisation. *Res. Agric. Eng.* **68**, 157–167 (2022).
21. Kumar, V., Rani, A., Rawal, R. & Mourya, V. Marker assisted accelerated introgression of null allele of kunitz trypsin inhibitor in soybean. *Breed. Sci.* **65**, 447–452 (2015).
22. Clarke, E. J. & Wiseman, J. Developments in plant breeding for improved nutritional quality of soya beans II. Anti-nutritional factors. *J. Agric. Sci.* **134**, 125–136 (2000).
23. Kumar, V., Rani, A., Shukla, S. & Jha, P. Development of Kunitz Trypsin inhibitor free vegetable soybean genotypes through marker-assisted selection. *Int. J. Veg. Sci.* **27**, 364–377 (2021).
24. Songstad, D. D., Petolino, J. F., Voytas, D. F. & Reichert, N. A. Genome editing of plants. *CRC. Crit. Rev. Plant Sci.* **36**, 1–23 (2017).
25. Lamichhane, S. & Thapa, S. Advances from conventional to modern plant breeding methodologies. *Plant Breed. Biotech* **2022**, 1–14 (2022).
26. Tazeb, A. Molecular marker techniques and their novel applications in crop improvement: A review article. *Glob. J. Mol. Sci.* **13**, 1–16 (2018).
27. Sarker, A., Masuda, M. S., Mushrat, Z. & Khan, M. K. Selection of superior genotypes using morpho-biochemical traits and crossability among them in cherry tomato (*Solanum lycopersicum*). *Discov. Plants* https://doi.org/10.1007/s44372-025-00162-y (2025).
28. Ibrahim, E. A. et al. Morphological, biochemical, and molecular diversity assessment of Egyptian bottle gourd cultivars. *Genet. Res. (Camb.)* **2024**, 1–15 (2024).

29. Ghonaim, M. M., Attya, A. M., Aly, H. G., Mohamed, H. I. & Omran, A. A. Agro-morphological, biochemical, and molecular markers of barley genotypes grown under salinity stress conditions. *BMC Plant Biol.* **23**, 1–19 (2023).
30. Liu, Z. J. & Cordes, J. F. DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* **238**, 1–37 (2004).
31. Nadeem, M. A. et al. DNA molecular markers in plant breeding: Current status and recent advancements in genomic selection and genome editing. *Biotechnol. Biotechnol. Equip.* **32**, 261–285 (2018).
32. Santana, F. A., Freire, M., Kellen, J., Guimarães, F. & Flores, M. Marker-assisted selection strategies for developing resistant soybean plants to cyst nematode. *Crop Breed. Appl. Biotechnol.* **14**, 180–186 (2014).
33. Alzate-marin, A. L., Cervigni, G. D. L., Moreira, M. A. & Barros, E. G. Seleção Assistida por Marcadores Moleculares Visando ao Desenvolvimento de Plantas Resistentes a Doenças, com Ênfase em Feijoeiro e Soja. *Fitopatol* **30**, 333–342 (2005).
34. Maria, R. et al. Assisted selection by specific DNA markers for genetic elimination of the kunitz trypsin inhibitor and lectin in soybean seeds. *Euphytica* **149**, 221–226 (2006).
35. Clever, M. et al. Genetic diversity analysis among soybean genotypes using SSR markers in Uganda. *Afr. J. Biotechnol.* **19**, 439–448 (2020).
36. Guo, Y. et al. SSR marker development, linkage mapping, and QTL analysis for establishment rate in common bermudagrass. *Plant Genome* **10**, 1–11 (2017).
37. Fan, M., Gao, Y., Wu, Z. & Zhang, Q. Linkage map development by EST-SSR markers and QTL analysis for inflorescence and leaf traits in. *Plants* **9**, 1–15 (2020).
38. Wang, Z. et al. Development of new mutant alleles and markers for KTI1 and KTI3 via CRISPR/Cas9-mediated mutagenesis to reduce trypsin inhibitor content and activity in soybean seeds. *Frontiers (Boulder).* **14**, 1–13 (2023).
39. Shavrukov, Y., Hinrichsen, P. & Watanabe, S. Editorial: Plant genotyping: From traditional markers to modern technologies. *Front. Plant Sci.* **15**, 1–4 (2024).
40. Chander, S. et al. Genetic diversity and population structure of soybean lines adapted to sub-Saharan Africa using single nucleotide polymorphism (Snp) markers. *Agronomy* **11**, 604 (2021).
41. Mercier, R., Solier, V., Lian, Q. & Loudet, O. Enhanced recombination empowers the detection and mapping of Quantitative Trait Loci. *Commun. Biol.* https://doi.org/10.1038/s42003-024-06530-w (2024).
42. Lukanda, M. M. et al. Genome-wide association analysis for resistance to *Coniothyrium glycines* causing red leaf blotch disease in soybean. *Genes (Basel).* **14**, 1–23 (2023).
43. Msiska, U. M. et al. Biochemicals associated with Callosobruchus Chinensis resistance in soybean. *Int. J. Adv. Res.* **6**, 292–305 (2018).
44. Bull, H., Murray, P. G., Thomas, D., Fraser, A. M. & Nelson, P. N. Acid phosphatases. *J. Clin. Pathol. Mol. Pathol.* **55**, 65–72 (2002).
45. Plaxton, W. C. & Tran, H. T. Update on metabolic adaptations metabolic adaptations of phosphate-starved plants. *Plant Physiol.* **1**(156), 1006–1015 (2011).
46. Raboy, V. Myo-inositol-1, 2, 3, 4, 5, 6-hexakisphosphate. *Phytochemistry* **64**, 1033–1043 (2003).
47. Cohen, P. et al. Phosphorylation on the. *Trends Biochem. Sci.* **25**, 596–601 (2000).
48. Davies, G. & Henrissat, B. Structures and mechanisms of glycosyl hydrolases. *Structure* **3**, 853–859 (1995).
49. Palzkill, T. Metallo-β-lactamase structure and function. *Ann. N. Y. Acad. Sci.* **1277**, 91–104 (2013).
50. Francis, X. et al. Functional analysis of the B and E lycopene cyclase enzymes of arabidopsis reveals a mechanism for control of cyclic carotenoid formation. *Plant Cell* **8**, 1613–1626 (1996).
51. Loijens, J. C. & Anderson, R. A. Type I phosphatidylinositol-4-phosphate 5-kinases are distinct members of this novel lipid kinase family*. *J. Biol. Chem.* **271**, 32937–32943 (1996).
52. Zhang, D. & Zhang, D. homology between dUF784, dUF1278 domains and the plant prolamin superfamily typifies evolutionary changes of disulfide bonding patterns. *Cell Cycle* **8**, 3428–3430 (2009).
53. Hartung, W., Sauter, A. & Hose, E. Abscisic acid in the xylem: Where does it come from, where does it go to?. *J. Exp. Bot.* **53**, 27–32 (2002).
54. Sher, R. et al. Plant defensins: Types, mechanism of action and prospects of genetic engineering for enhanced disease resistance in plants. *Biotech* **9**, 1–12 (2019).
55. Skowyra, D., Craig, K. L., Tyers, M., Elledge, S. J. & Harper, J. W. F-box proteins are receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex. *Cell* **91**, 209–219 (1997).
56. Hanks, S. K. Genomic analysis of the eukaryotic protein kinase superfamily: A perspective. *Genome Biol.* **4**, 111 (2003).
57. Xie, H. et al. Large-scale protein annotation through gene ontology. *Genome Res.* **12**, 785–794 (2002).
58. Cooperman, B. S., Baykov, A. A. & Lahti, R. Evolutionary conservation of the active site of soluble inorganic pyrophosphatase. *Elsevier Sci.* **17**, 262 (1992).
59. Agarwal, P. K. & Jha, B. Transcription factors in plants and ABA dependent and independent abiotic stress signalling. *Biol. Plantarun* **54**, 201–212 (2010).
60. Gamsjaeger, R., Liew, C. K., Loughlin, F. E., Crossley, M. & Mackay, J. P. Sticky fingers: Zinc-fingers as protein-recognition motifs. *TRENDS Biochem. Sci.* **32**, 63 (2006).
61. Bakshi, M. & Oelmüller, R. Jack of many trades in plants WRKY transcription factors jack of many trades in plants. *Plant Signal. Behav.* **9**, 1–18 (2014).
62. Martins, L., Trujillo-hernandez, J. A. & Reichheld, J. Thiol based redox signaling in plant nucleus. *Front. Plant Sci.* **9**, 1–9 (2018).
63. Riechmann, V., Criichten, I. Van & Sablitzky, F. The expression pattern of Id4, a novel dominant negative helix-loop-helix protein, is distinct from Id1, 162 and Id3. **22**, (1994).
64. Fields, A. C. & D., M. fields1994 (1).pdf. *Biochem. Biophys. Res. Commun.* **198**, 288–291 (1994).
65. Dehghan, M., Akhtar-Danesh, N. & Merchant, A. T. Childhood obesity, prevalence and prevention. *Nutr. J.* **4**, 1–8 (2005).
66. Saurin, A. J., Borden, K. L. B., Boddy, M. N. & Freemont, P. S. Does this have a familiar RING?. *Elsevier Sci.* **0004**, 208–214 (1996).
67. Kupke, T., Caparrós-Martín, J. A., Malquichagua Salazar, K. J. & Culiáñez-Macià, F. A. Biochemical and physiological characterization of *Arabidopsis thaliana* AtCoAse: A Nudix CoA hydrolyzing protein that improves plant development. *Physiol. Plant.* **135**, 365–378 (2009).
68. Russo, A. A., Jeffrey, P. D. & Pavletich, N. P. © 199 6 Nature Publishing Group http://www.nature.com/nsmb. *Nature* **3**, 696–700 (1996).
69. Henrissat, B. Glycosidase families. *Biochem. Soc. Trans.* **26**, 153–156 (1998).
70. Ehsan, H., Reichheld, J. P., Durfee, T. & Roe, J. L. TOUSLED kinase activity oscillates during the cell cycle and interacts with chromatin regulators. *Plant Physiol.* **134**, 1488–1499 (2004).
71. Goto, K., Pi, T., Words, K., March, R. & Genetics, M. Function and regulation of the Arabidopsis floral homeotic gene PISTILLATA. *Genes Dev.* **8**, 1548–1560 (1994).
72. Khan, H. et al. Genome-wide identification and expression analysis of U-box gene family in *Juglans regia* L. *Genet. Resour. Crop Evol.* **70**, 2337–2352 (2023).
73. Aravind, L. & Ponting, C. P. Homologues of 26S proteasome subunits are regulators of transcription and translation. *Protein Sci.* **7**, 1250–1254 (1998).
74. Xiong, A. & Jayaswal, R. K. Molecular characterization of a chromosomal determinant conferring resistance to zinc and cobalt ions in Staphylococcus aureus. *J. Bacteriol.* **180**, 4024–4029 (1998).
75. Balaji, S., Madan Babu, M., Iyer, L. M. & Aravind, L. Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. *Nucleic Acids Res.* **33**, 3994–4006 (2005).

76. Skirpan, A. L. et al. Isolation and characterization of kinase interacting protein 1, a pollen protein that interacts with the kinase domain of PRK1, a receptor-like kinase of petunia. *Plant Physiol.* **126**, 1480–1492 (2001).
77. Beuck, C. et al. Structure of the GLD-1 homodimerization domain: Insights into STAR protein-mediated translational regulation. *Structure* **18**, 377–389 (2010).
78. Ramakrishnan, V. & Moore, P. B. Atomic structures at last: The ribosome in 2000. *Curr. Opin. Struct. Biol.* **11**, 144–154 (2001).
79. Salim, R. et al. A review on anti-nutritional factors: Unraveling the natural gateways to human health. *Front. Nutr.* **10**, 1215873 (2023).
80. Coelho, S. E. dos A. C., Vianna, R. P. de T., Segall-Correa, A. M., Perez-Escamilla, R. & Gubert, M. B. Insegurança alimentar entre adolescentes Brasileiros: Um estudo de validação da Escala Curta de Insegurança Alimentar. *Rev. Nutr.* **28**, 385–395 (2015).
81. Nieto-Veloza, A. et al. Lunasin protease inhibitor concentrate decreases pro-inflammatory cytokines and improves histopathological markers in dextran sodium sulfate-induced ulcerative colitis. *Food Sci. Hum. Wellness* **11**, 1508–1514 (2022).
82. Rosso, M. L., Shang, C., Correa, E. & Zhang, B. An efficient HPLC approach to quantify kunitz trypsin inhibitor in soybean seeds. *Crop Sci.* **58**, 1616–1623 (2018).
83. Zhang, R. et al. GWLD : An R package for genome-wide linkage disequilibrium analysis. *G3 Genes Genomes Genet.* **13**, 1–8 (2023).
84. Flint-garcia, S. A., Thornsberry, J. M. & Iv, E. S. B. *Tructure of.* https://doi.org/10.1146/annurev.arplant.54.031902.134907 (2003).
85. Kim, S., Tayade, R., Kang, B., Hahn, B. & Ha, B. Genome-Wide Association Studies of Seven Root Traits in Soybean (Glycine max L.) Landraces. (2023).
86. Liu, Z. et al. Comparison of genetic diversity between Chinese and american soybean (*Glycine max* (L.)) accessions revealed by. *Front. Plant Sci.* **8**, 1–13 (2017).
87. Jo, H. et al. Genetic diversity of soybeans (*Glycine max* (L.) Merr) with black seed coats and green cotyledons in Korean germplasm. *Agronomy* **11**, 581 (2021).
88. Yoon, M. S. et al. DNA profiling and genetic diversity of Korean soybean (*Glycine max* (L.) Merrill) landraces by SSR markers. *Euphytica* **165**(1), 69–77. https://doi.org/10.1007/s10681-008-9757-7 (2009).
89. Rao, V. S., Srinivas, K., Sujini, G. N. & Kumar, G. N. S. Protein-protein interaction detection: Methods and analysis. *Int. J. Proteom.* **2014**, 1–12 (2014).
90. Sharmin, R. A. et al. Genome-wide association study uncovers major genetic loci associated with seed flooding tolerance in soybean. *BMC Plant Biol.* **21**, 1–17 (2021).
91. Ravelombola, W. et al. Genome-wide association study and genomic selection for yield and related traits in soybean. *PLoS ONE* **16**, 1–21 (2021).
92. Martins, V., Ferrari, F. & White, P. J. Plant physiology and biochemistry phytic acid accumulation in plants: Biosynthesis pathway regulation and role in human diet. *Plant Physiol. Biochem.* **164**, 132–146 (2021).
93. Tang, Y. et al. GAPIT Version 2: An enhanced integrated tool for genomic association and prediction. *Plant Genome* **9**, (2016).
94. Li, M. et al. Enrichment of statistical power for genome-wide association studies. *BMC Biol.* **12**, 1–10 (2014).
95. Kaler, A. S., Gillman, J. D., Beissinger, T. & Purcell, L. C. Comparing different statistical models and multiple testing corrections for association mapping in soybean and maize. *Front. Plant Sci.* **10**, 1–13 (2020).
96. He, L., Wang, H., Sui, Y. & Miao, Y. Genome-wide association studies of five free amino acid levels in rice. *Front. Plant Sci.* **13**, 1–17 (2022).
97. Obua, T. et al. Yield stability of tropical soybean genotypes in selected agro-ecologies in Uganda. *S. Afr. J. Plant Soil* **37**, 168–173 (2020).
98. Macherey-Nagel. Genomic DNA from plant - User manual (NucleoSpin® Plant II, -Midi, -Maxi). 36 at (2018).
99. Kilian, A. et al. Diversity arrays technology: A generic genome profiling technology on open platforms. *Methods Mol. Biol.* **888**, 67–89 (2012).
100. Egea, L. A., Mérida-García, R., Kilian, A., Hernandez, P. & Dorado, G. Assessment of genetic diversity and structure of large garlic (*Allium sativum*) germplasm bank, by diversity arrays technology 'genotyping-by-sequencing' platform (DArTseq). *Front. Genet.* **8**, 1–9 (2017).
101. Baloch, F. S. et al. A whole genome DArTseq and SNP analysis for genetic diversity assessment in durum wheat from central fertile crescent. *PLoS ONE* **12**, 1–18 (2017).
102. Israel, D. W. Genetic variability for phytic acid phosphorus and inorgaic phosphorus in seeds of soybeans in maturity groups V, VI, and VII. **46**, 67–71 (2013).
103. Anozie, A. N., Salami, O. A., Babatunde, D. E. & Babatunde, O. E. Comparative evaluation of processes for production of soybean meal for poultry feed in Nigeria Evaluación comparativa de procesos para la producción de harina. *Anim. Sci.* **52**, 193–202 (2018).
104. Kakade, M. L., Rackis, J. J., Mcghee, J. E. & Pusk, G. Determination of trypsin inhibitor activity of soy products: A collaborative analysis of an improved procedure. *Cereal Chem.* **51**, 376–382 (1974).
105. Bates, D., Mächler, M., Bolker, B. M. & Walker, S. C. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
106. Henderson, C. R. Best linear unbiased estimation and prediction under a selection model published by: International biometric society stable. *Biometrics* **31**, 423–447 (1975).
107. Wang, S. B. et al. Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Sci. Rep.* **6**, 1–10 (2016).
108. Zhang, Y. et al. Multi-locus genome-wide association study reveals the genetic architecture of stalk lodging resistance-related traits in maize. *Front. Plant Sci.* **9**, 1–12 (2018).
109. Wen, Y. J. et al. Methodological implementation of mixed linear models in multi-locus genome-wide association studies. *Brief. Bioinform.* **19**, 700–712 (2018).
110. Zhang, J. et al. PLARmEB: Integration of least angle regression with empirical Bayes for multilocus genome-wide association studies. *Heredity (Edinb).* **118**, 517–524 (2017).
111. Ren, W. L., Wen, Y. J., Dunwell, J. M. & Zhang, Y. M. PKWmEB: Integration of Kruskal–Wallis test with empirical Bayes under polygenic background control for multi-locus genome-wide association study. *Heredity (Edinb).* **120**, 208–218 (2018).
112. Tamba, C. L., Ni, Y. L. & Zhang, Y. M. Iterative sure independence screening EM-Bayesian LASSO algorithm for multi-locus genome-wide association studies. *PLoS Comput. Biol.* **13**, 1–20 (2017).
113. Zhang, Y. W. et al. mrMLM v4.0.2: An R platform for multi-locus genome-wide association studies. *Genom. Proteom. Bioinform.* **18**, 481–487 (2020).

Acknowledgements

We acknowledge the Partnership for Applied Skills in Sciences, Engineering and Technology-Regional Scholarship and Innovation Fund (PASET-RSIF) and Carnegie Corporation of New York for funding this work. Gratitude to the Makerere University, Makerere Regional Centre for Crops Improvement (MaRCCI) for providing facilities and to Makerere University Centre for Soybean Improvement and Development (MAKCSID) for providing soybean materials used in this study. Extended gratitude goes to the National Agricultural Research Organization (NARO), and the technicians at the National Crops Resources Research Institute (NaCRRI) for their valuable support.

Author contributions

Conceptualized the experiment: N.J.P, T.O and P.T; Methodology: N.J.P, E.N and E.W. Project administration: M.O.S, I.O.D and R.E; Data collection: N.J.P; Data curation: N.J.P, EAA and E.W; Formal analysis: N.J.P, E.W, S.V.K; Software analysis: N.J.P and E.W; Visualization: N.J.P and E.W; Supervision: P.T, T.O, M.M, J.P.S and E.N; Writing - original draft: N.J.P; Writing - review & editing, validation of analyzed data: N.J.P, A.B, J.P.S, S.V.K.

Funding

This research was funded by the PASET-RSIF[grant number B8501G30218] and Carnegie Corporation of New York. The Genotyping cost was co-funded by the Bill and Melinda Gates Foundation [grant number OPP1093174] through the Integrated Genotyping Sequence Support (IGSS) project.

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval and consent to participate

The seeds used in the study are owned by the Makerere University Centre for Soybean Improvement and Development (MAKCSID) led by Senior Soybean Breeder in Uganda, Prof. Phinehas Tukamuhabwa. Therefore, the collection of the seeds used in the study complies with local or national guidelines with no need for further affirmation.

Consent for publication

All authors have read and agreed to the published version of the manuscript.

Additional information

Correspondence and requests for materials should be addressed to N.J.P. or J.P.S.