



Plant Archives

Journal homepage: <http://www.plantarchives.org>DOI Url : <https://doi.org/10.51470/PLANTARCHIVES.2026.v26.no.2.058>

CHARACTERIZATION OF SOYBEAN GENETIC RESOURCES FOR CARBOHYDRATE-ASSOCIATED LOCI USING GENE-LINKED SSR MARKERS

Riya Mishra¹, Manoj Kumar Tripathi^{1,2*}, Prakash N. Tiwari³, Jagendra Singh², Sanjeev Sharma¹,
Kumar Sanu¹, Priya Arvind Giri¹ and Sejal Thakur¹

¹Department of Genetics and Plant Breeding, Rajmata Vijayaraje Scindia Agricultural University, Gwalior, 474002, M.P., India

²Zonal Agricultural Research Station, Rajmata Vijayaraje Scindia Agricultural University, Morena, 476001, M.P., India

³Department of Basic Sciences and Humanities, College of Horticulture and Forestry, Pasighat, Arunachal Pradesh,
Central Agricultural University, Imphal, India

*Corresponding author E-mail: drmanojtripathi64@gmail.com

ORCID: <https://orcid.org/0009-0004-1921-3046>

(Date of Receiving-25-05-2026; Date of Revision-29-07-2026; Date of Acceptance-18-08-2026)

ABSTRACT

While breeding pipelines have long chased protein and oil gains in soybean, the carbohydrate fraction, which accounts for nearly a third of seed dry matter and directly determines feed digestibility and food grade quality, has remained a genetically underexplored trait space. This study addresses that gap by profiling 153 diverse soybean genotypes employing gene-linked SSR markers anchored to carbohydrate biosynthesis loci. All three markers were found biallelic, with polymorphic information content ranging from 0.115 (Satt 324) to 0.345 (Satt 270 with an average value 0.265), identifying Satt 270 and Satt 598 as the more discriminating loci for this trait class. Bayesian clustering (STRUCTURE) resolved the panel into three subpopulations (K = 3; 41, 55 and 57 genotypes), with population 3 showing the highest allelic and gene diversity, marking it a priority source for carbohydrate focused in future breeding programmes. Despite this stratification, AMOVA attributed 99% of the total molecular variance to differences within populations and only 1% to divergence among populations, suggesting limited differentiation among the investigated populations, possibly reflecting common breeding backgrounds and extensive utilization of related germplasm. Principal Coordinate Analysis corroborated this pattern, with the first two axes capturing 90.46% of total variation, while UPGMA clustering separated a genetically distinct group of ten genotypes from a major cluster of 143 genotypes, flagging these divergent accessions as untapped parental material. Collectively, this study provides preliminary insights into genetic diversity associated with carbohydrate related loci among soybean germplasm, offering breeders a concrete allele-based rationale for parent selection and marker assisted improvement of seed carbohydrate composition in soybean.

Key words: Carbohydrate content; Genetic diversity; Marker assisted selection; Population structure; Soybean (*Glycine max*); SSR markers.

Introduction

Seed composition in soybean is determined by the partitioning of assimilated carbon among protein, oil and carbohydrate biosynthetic pathways during seed filling stage. The relative allocation of carbon among these storage compounds largely determines seed quality and economic value. This carbon partitioning process underpins the agronomical and economical importance

of soybean as a major global oilseed crop (Allen and Young, 2013; Li *et al.*, 2015; Chowdhury *et al.*, 2026). Global production has climbed to a record 425.8 million metric tons in the 2025-26 marketing year, with Brazil alone supplying close to 40% of that total and China absorbing over 112 million metric tons in imports each year. This global demand is driven by the consistent production of high-quality protein and oil in soybean seeds

(USDA, 2026). However, the continued improvement of soybean cultivars depends on the availability and effective utilization of diverse genetic resources that harbour useful allelic variation for complex quality traits (Gilman and Bilyeu, 2012; Zhuang *et al.*, 2022; Vargas-Almendra *et al.*, 2024; Mishra *et al.*, 2025a).

Despite its substantial contribution to seed composition and quality, the carbohydrate fraction has received comparatively less attention in soybean improvement programmes, which have historically emphasized protein and oil traits (Moretti *et al.*, 2020; Duan *et al.*, 2023; Mishra *et al.*, 2025b). Mature soybean seeds carry roughly 36-42 % protein and 18-22 % oil by dry weight, but carbohydrates account for nearly 26-35 % of seed dry matter, representing a major component of seed composition (Kumar *et al.*, 2025; Tripathi *et al.*, 2024; Mishra *et al.*, 2025c). Because protein, oil and carbohydrate synthesis all draw upon a common carbon pool during soybean seed development (Allen and Young, 2013; Song *et al.*, 2017; Tripathi *et al.*, 2024), an increase in one component often occurs at the expense of another, particularly given the well-documented inverse relationship between protein and oil and evidence that carbon reallocation during maturation also affects oligosaccharide accumulation (Patil *et al.*, 2017; Kambhampati *et al.*, 2020). Consequently, selection for protein and oil content may indirectly influence carbohydrate accumulation (Moretti *et al.*, 2020; Tamango *et al.*, 2022; Escamilla *et al.*, 2023). Therefore, understanding the genetic variation associated with carbohydrate accumulation is essential for developing soybean cultivars with improved nutritional and processing quality.

Soybean seed carbohydrates comprise several fractions, including sucrose, raffinose family oligosaccharides (RFOs), soluble sugars and other storage carbohydrates, which influence seed physiology, nutritional quality and end use properties. Sucrose contributes positively to seed palatability and nutritional quality (Elango *et al.*, 2022; Sanyal *et al.*, 2023). In contrast, RFOs have both beneficial and undesirable effects. Although RFOs contribute to desiccation tolerance and seed stability during storage, monogastric livestock are unable to digest them because of the absence of endogenous α -galactosidase activity (Zhawar *et al.*, 2011; Banti *et al.*, 2021; Rej *et al.*, 2024). Consequently, their presence is associated with reduced nutrient utilization, lower metabolizable energy and flatulence in poultry and swine. Therefore, carbohydrate composition influences not only seed physiology but also feed conversion efficiency and the nutritional quality of soy-based food products

(Valentine *et al.*, 2017; Rosso *et al.*, 2026).

Despite its importance, the genetic basis of carbohydrate accumulation remains less extensively characterized than that of protein and oil content. Molecular studies in soybean have predominantly focused on loci associated with protein and oil content, whereas comparatively fewer studies have investigated genetic variation associated with carbohydrate-related traits (Valliyodan *et al.*, 2016; Lee *et al.*, 2023). Information on genetic diversity, population structure and genetic differentiation in soybean germplasm based on carbohydrate-associated SSR markers remains limited. Assessing genetic diversity and population structure can facilitate the identification of genetically distinct germplasm, the selection of suitable parental combinations and an efficient utilization of genetic resources in soybean breeding programmes. SSR markers remain valuable tools for genetic diversity studies because of their co-dominant inheritance, high polymorphism, reproducibility and multi-allelic nature. Therefore, the present investigation was undertaken to address this knowledge gap. Using carbohydrate-associated SSR markers, the study aimed to assess genetic diversity, population structure and genetic differentiation among soybean genotypes and to generate molecular information that may support marker-assisted selection for improving carbohydrate-related seed quality traits in soybean in future.

Material and Methods

Experimental details

The study was conducted during *Kharif* 2024 and 2025 at Zonal Agricultural Research Station (ZARS), Morena, under Rajmata Vijayaraje Scindia Krishi Vishwa Vidyalaya (RVSKVV), Gwalior, Madhya Pradesh, India. The experimental site is located in the semi-arid agro-climatic region of northern Madhya Pradesh, characterized by variable rainfall distribution and temperature fluctuations. A total of 153 soybean (*Glycine max* (L.) Merrill) genotypes (Table 1), were evaluated in an augmented block design comprising five blocks. The experimental material was obtained from ZARS Morena, College of Agriculture, Sehore, RVSKVV, Gwalior and Jawaharlal Nehru Krishi Vishwa Vidyalaya, Jabalpur, India. Each genotype was planted in a single 3.0 m row with a row-to-row spacing of 45 cm and plant-to-plant spacing of 7-10 cm. The plot size for each genotype was 1.35 m² (3.0 × 0.45 m). Recommended agronomical practices for soybean cultivation were uniformly followed throughout the crop growth period. A basal fertilizer dose of 20:60:20:20 kg ha⁻¹ of N:P:K:S was applied at the time of sowing. For molecular characterization, Primers were

Table 1: List of genotypes used in experiment along with their pedigree/ source /origin.

S. No	Germplasm	Pedigree/Source/origin
1.	JS-26	JNKVV, Jabalpur
2.	JS-93-05	Secondary selection from PS 73-22
3.	JS-97-52	PS 327 × L129
4.	JS-20-29	JS 97-52 × JS 95-56
5.	JS-20-34	JS 98-63 × PK 768
6.	JS-20-53	JS 97-52 × JS 20-02
7.	JS-20-69	JS 97-52 × SL 710
8.	JS-20-79	JS 97-52 × JS (1S) 90-05-12-1
9.	JS-20-89	JS 97-52 × JSM 286
10.	JS-20-94	JS 97-52 × JS 20-02
11.	JS-20-98	JS 97-52 × SL 710
12.	JS-20-116	JS 97-52 × JSM 120A
13.	JS-21-05	SL 738 × JS 95-60
14.	JS-21-07	JS 20-30 × JS 93-05
15.	JS-21-17	SL 738 × JS 95-60
16.	JS-21-72	SL 738 × JS 95-60
17.	JS-21-76	JS 20-30 × JS 93-05
18.	JS-21-77	JS 20-63 × JS 20-66
19.	JS-22-01	SL 738 × JS 95-60
20.	JS-22-08	JSM 240 × SL 517
21.	JS-22-12	JS 20-53 × JS 20-34
22.	JS-22-16	JS 20-88 × JS 20-34
23.	JS-23-03	JS 20-53 × JS 20-34
24.	JS-23-05	JS 20-63 × JS 20-35
25.	JS-23-09	JS 20-53 × JS 20-34
26.	JS-24-26	DS 3105 × JS 20-69
27.	JS-24-30	JS 20-53 × JS 20-34
28.	JS-24-34	JS 20-53 × JS 20-34
29.	JS-24-31	JS 20-53 × JS 20-24
30.	JS-25-03	JS 20-53 × JS 20-34
31.	JS-25-05	JS 20-89 × JS 20-34
32.	JS-25-06	JS 20-29 × JS 93-05
33.	JS-25-08	JS 20-71 × JS 20-87
34.	JS-25-02	JS 20-53 × JS 20-34
35.	JS-24-33	JS 20-53 × JS 20-34
36.	JS-25-50	JS 20-17 × PS 1092
37.	JS-25-52	JS 21-05 × JS 20-24
38.	JS-25-53	JSM 128 × PS 1556
39.	JS-25-55	DS 3105 × JS 20-69
40.	RVS-76	RVSKVV, Gwalior
41.	RVS-2001-4	JS 93-01 × EC 390981
42.	RVS-2011-10	JS 335 × PS 1042
43.	RVS-12-8	RVSKVV, Gwalior
44.	RVS-13-20	RVSKVV, Gwalior
45.	RVS-15-1	RVSKVV, Gwalior
46.	RVS-23-2	RVSKVV, Gwalior
47.	RVS-23-5	JS 97-52 × JS 93-05-5
48.	RVS-23-7	RVS 76 × JS 20-116-1
49.	RVS-23-8	RVS 76 × JS 20-116-2

50.	RVS-23-9	RSC 10-70 × JS 20-116-1
51.	RVS-23-10	RSC 10-70 × JS 20-116-2
52.	RVS-23-11	RSC 10-70 × JS 20-116-3
53.	RVS-23-12	JS 20-53 × JS 20-34-1
54.	RVS-23-13	AMS-MB-5-18 × RVS 76-1
55.	RVS-23-14	RVS 2009-4 × JS 93-05-1
56.	RVS-23-15	RVS 76-1 × JS 20-34-1
57.	RVS-23-16	RVS 2011-10 × JS 20-34-1
58.	RVS-23-17	RVS 20-29 × JS 20-22
59.	RVS-23-19	JS 20-53 × JS 20-34
60.	RVS-23-20	RVSKVV, Gwalior
61.	RVS-23-21	JS 20-31 × JS 95-60
62.	RVS-23-22	JS 335 × PK 1042
63.	RVS-23-23	JS 20-63 × JS 95-60
64.	RVS-23-25	JS 20-30 × JS 93-05
65.	RVS-23-26	RVSKVV, Gwalior
66.	NRC-130	EC 538828 × EC 390977
67.	NRC-138	NSRI, Indore
68.	NRC-142	NSRI, Indore
69.	NRC-150	NSRI, Indore
70.	NRC-152	NSRI, Indore
71.	NRC-166	NSRI, Indore
72.	NRC-192	NSRI, Indore
73.	NRC-201	NSRI, Indore
74.	NRC-254	NSRI, Indore
75.	NRC-255	NSRI, Indore
76.	NRC-258	NSRI, Indore
77.	NRC-259	NSRI, Indore
78.	NRC-264	NSRI, Indore
79.	NRC-265	NSRI, Indore
80.	NRC-266	NSRI, Indore
81.	NRC-267	NSRI, Indore
82.	NRC-269	NRC 128 × JS 95-60
83.	NRC-271	NSRI, Indore
84.	NRC-859	NSRI, Indore
85.	NRCSL-4	RAK, Sehore
86.	NRCSL-7	RAK, Sehore
87.	NRCSL-8	RAK, Sehore
88.	RSC-10-46	JS 335 × Bragg
89.	RSC-10-52	Raipur
90.	RSC-11-95	Raipur
91.	RSC-1165	Raipur
92.	ASB-85	RAK, Sehore
93.	ASB-93	RAK, Sehore
94.	ASB-101	RAK, Sehore
95.	DLSB-5	RAK, Sehore
96.	DLSB-40	RAK, Sehore
97.	PS 1569	RAK, Sehore
98.	PS-1696	RAK, Sehore
99.	Himso 1696	RAK, Sehore
100.	Himso 1695	RAK, Sehore

Continue ...1

101.	DS 1510	Delhi
102.	DS 1547	Delhi
103.	DS-1589	Delhi
104.	DS 1480	NRC 37 X JS 97-52
105.	RVSM-35	RAK, Sehore
106.	RVSM-12-21	RAK, Sehore
107.	RVSM-2011-35	RAK, Sehore
108.	RVSM-2012-4	RAK, Sehore
109.	KSS-213	RAK, Sehore
110.	KSS-225	RAK, Sehore
111.	AMS-2021-3	PDKV, Akola
112.	AMS-2021-4	PDKV, Akola
113.	AMS-100-39	PDKV, Akola
114.	AMS- 264	PDKV, Akola
115.	AMS 22-16	PDKV, Akola
116.	AMS-2022-1	AMS 100-39 x JS 93-05
117.	MAUS-124	RAK, Sehore
118.	MAUS-824	RAK, Sehore
119.	MAUS-791	RAK, Sehore
120.	MAUS-787	RAK, Sehore
121.	Rajsoya-24	RAK, Sehore
122.	CAUMS-3	RAK, Sehore
123.	MACS-1810	RAK, Sehore
124.	MACS-1834	RAK, Sehore
125.	MACS-824	RAK, Sehore
126.	TS-208	RAK, Sehore
127.	TS-156	RAK, Sehore
128.	TS-101	RAK, Sehore
129.	AUKS-212	RAK, Sehore
130.	AUKS-21-5	RAK, Sehore
131.	AS-26	RAK, Sehore
132.	AS-34	RAK, Sehore
133.	AS-55	RAK, Sehore
134.	Jhimka	Landrace
135.	Cat 492A	RAK, Sehore
136.	Cat 87	RAK, Sehore
137.	EC109563	Exotic collection
138.	KDS-1188	RAK, Sehore
139.	KDS-1193	RAK, Sehore
140.	KDS-1201	RAK, Sehore
141.	KDS-1203	RAK, Sehore
142.	KDSIS-1394	RAK, Sehore
143.	KBSL-23-36	RAK, Sehore
144.	VLS- 104	RAK, Sehore
145.	VLS-106	RAK, Sehore
146.	SL-311	PAU, Ludhiana
147.	SL-1315	PAU, Ludhiana
148.	THPS-6	RAK, Sehore
149.	BAUS(M)-6	Ranchi
150.	Lok soya 03	RAK, Sehore
151.	Pusa Sipani BS 8	RAK, Sehore
152.	Pusa Sipani BS 33	RAK, Sehore
153.	Pusa Sipani SPS 433	RAK, Sehore

synthesized by Bioserve Biotechnologies, India. PCR amplification and marker genotyping were performed following standard molecular protocols at Ankur Seeds Pvt. Ltd., Nagpur, Maharashtra, India.

Molecular analysis

DNA extraction and quantification

Genomic DNA was isolated from fresh young leaf tissues collected from 2–3-week-old soybean plants using a modified cetyltrimethylammonium bromide (CTAB) extraction method described by Murray and Thompson (1980). Approximately 100 mg of fresh leaf samples were homogenized in CTAB extraction buffer supplemented with 0.2 % β -mercaptoethanol and incubated at 65°C to ensure an efficient cell lysis and release of genomic DNA. The lysate was subsequently purified using phenol: chloroform: isoamyl alcohol (25:24:1) extraction, followed by RNase treatment to remove residual RNA contaminants. DNA was precipitated using sodium acetate and ice-cold absolute ethanol, washed with 70 % ethanol, air dried and finally dissolved in TE (Tris-EDTA) buffer. The purity and concentration of the extracted DNA were determined spectrophotometrically by measuring absorbance at 260 and 280 nm. DNA samples exhibiting A_{260}/A_{280} ratio ranging from 1.7 to 1.9 were considered of acceptable quality for molecular analysis. The integrity of the isolated DNA was further confirmed through electrophoresis on 0.8% agarose gel and working DNA solutions were prepared by diluting the samples to a final concentration of 25–50 ng μL^{-1} for subsequent Polymerase chain reaction (PCR) amplification.

Polymerase chain reaction and SSR genotyping

Molecular characterization of soybean genotypes was performed using three trait-specific SSR markers previously reported to be associated with seed carbohydrate content (Table 2) and therefore employed for trait specific assessment of genetic diversity among soybean genotypes. PCR amplifications were carried out

Table 2: Primer sequences and details of carbohydrate associated SSR markers used for molecular characterization of soybean genotypes.

Marker	Forward sequence	Reverse sequence	References
Satt598	CGATTGTA ATATACTTA CCGTCTATA	CACAATAC CTGTGGCTG TTATACTAT	Wang <i>et al.</i> , (2014)
Satt324	GTTCCCAGG TCCCACCAT CTATG	GCGTTTCTT TTATACCT TCAAG	
Satt270	TGTGATGCC CCTTTTCT	GCGCAGTGCA TGGTTTCTCA	

in a total reaction volume of 10 µl reaction mixture consisted of 25-50 ng genomic DNA, 1 × PCR buffer, 0.2 mM dNTPs, 0.5 µM each primer and 1 U Taq DNA polymerase. The PCR reactions were conducted in a thermal cycler with an initial denaturation at 95°C for 5 minutes, followed by 35 amplification cycles consisting of denaturation at 95°C for 30 s, primer specific annealing at 55-62°C for 45 s and extension at 72°C for 1 min with a final extension at 72°C for 10 minutes, ensuring an efficient and specific amplification of the target loci. The amplified products were resolved by electrophoresis on 3 % agarose gel prepared in 1 × TBE buffer and subsequently visualized under ultraviolet light. Fragment sizes were scored based on banding patterns and allelic variation and the resulting data were utilized for subsequent genetic diversity and marker-based analysis.

Model based analysis of genotypic data

SSR alleles were scored on the basis of fragment size using GelAnalyzer version 19.1. The generated allelic data were subsequently utilized for the assessment of genetic diversity parameters using GenAlEx version 6.51b2 and PowerMarker version 3.25. Marker polymorphism and informativeness were evaluated by estimating major allele frequency, gene diversity and polymorphism information content (PIC). Population structure analysis was performed using STRUCTURE version 2.3.4 to infer the genetic stratification among soybean genotypes. Structure analysis was performed using K values ranging from 1 to 10 with 10 independent runs for each K. A burn-in period of 100,000 iterations followed by 100,000 Markov Chain Monte Carlo (MCMC) replications was employed under the admixture model with correlated allele frequencies. The optimal K value was determined using the Evanno method (Evanno *et al.*, 2005). Analysis of molecular variance (AMOVA) and Principal Coordinates Analysis (PCoA) were performed in GenAlEx version 6.51b2. Genetic distance matrices generated from SSR data were used for PCoA to visualize genetic relationships among genotypes. Significance of variance components was tested using 999 permutations. Furthermore, genetic relationships among genotypes were determined through cluster analysis based on Unweighted Pair Group Method with Arithmetic Mean (UPGMA) algorithm implemented in PowerMarker. The resulting dendrogram was visualized using MEGA version 4.0 to elucidate the genetic relatedness and clustering patterns among the studied genotypes (Tiwari *et al.*, 2023).

Results and Discussion

Genetic diversity analysis of SSR markers associated with carbohydrate content in soybean

Table 3: Genetic diversity parameters of SSR markers associated with carbohydrate content in soybean.

Marker	Major allele frequency	Number of alleles	Gene diversity	PIC
Satt 598	0.712	2	0.410	0.326
Satt 324	0.935	2	0.122	0.115
Satt 270	0.641	2	0.461	0.345
Mean	0.763	2	0.331	0.265

The genetic diversity parameters of the SSR markers associated with carbohydrate content in soybean are presented in Table 3. Three SSR markers viz., Satt 598, Satt 324 and Satt 270 were evaluated for their allelic diversity and marker informativeness across the soybean genotypes. Each marker produced two alleles, indicating the presence of polymorphism at all loci, although the extent of genetic variation varied among the markers. The major allele frequency ranged between 0.641 for the marker Satt 270 to 0.935 for Satt 324, with an overall mean of 0.763. The comparatively lower major allele frequency observed for the marker Satt 270 suggests a more a balanced distribution of alleles and consequently greater genetic variability at this locus. Conversely, the higher major allele frequency of Satt 324 indicates the predominance of a single allele among the evaluated genotypes, reflecting presence of comparatively lower polymorphism. Similar trends have also been reported by Hipparagi *et al.*, (2017), Kumar *et al.*, (2022), Rani *et al.*, (2023) in soybean diversity studies, where loci with lower major allele frequencies and higher PIC values generally exhibited greater genetic variability and superior discriminatory ability.

The gene diversity values varied between 0.122 for Satt 324 to 0.461 for Satt 270 with an average of 0.331. Among the three markers, Satt 270 exhibited the highest gene diversity followed by Satt 598 (0.410), indicating that these loci possessed greater allelic variation and were more effective in differentiating the soybean genotypes. In contrast, the low gene diversity recorded for Satt 324 suggests limited allelic variation at this locus. The moderate mean gene diversity obtained in the present investigation indicates the existence of appreciable molecular variation among the evaluated germplasm, although the diversity remains relatively moderate owing to the narrow genetic base of cultivated soybean (Kumar *et al.*, 2022; Saleem *et al.*, 2021). These observations are in agreement with the findings of Hyten *et al.*, (2006), Li *et al.*, (2013), Nawaz *et al.*, (2018) and Viana *et al.* (2022), who also reported reduced genetic diversity in cultivated soybean due to intensive selection and repeated utilization of elite parental lines in breeding programmes.

Table 4: Population wise genetic diversity statistics based on SSR markers linked to carbohydrate content in soybean.

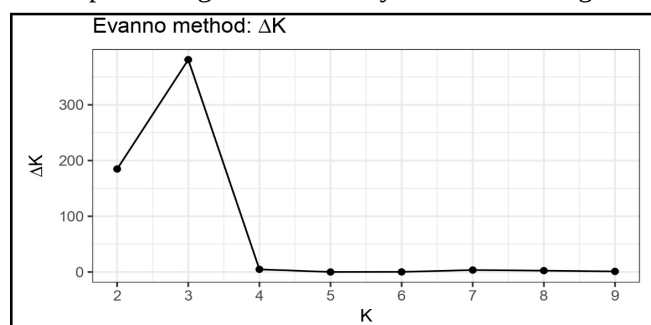
Population	N	Na	Ne	SI	He	uHe	FI
Pop1	41	2.000	1.520	0.460	0.302	0.306	1.000
Pop2	55	2.000	1.535	0.494	0.325	0.328	1.000
Pop3	57	2.000	1.569	0.504	0.335	0.338	1.000
Mean	51	2.000	1.541	0.486	0.321	0.324	1.000

Na: No. of Different Alleles; Ne: No. of Effective Alleles; SI = Shannon's Information Index; Ho = Observed Heterozygosity; He = Expected Heterozygosity; uHe = Unbiased Expected Heterozygosity; FI = Fixation Index

The polymorphic information content (PIC) values arrayed between 0.115 to 0.345, with a mean value of 0.265. The highest PIC value was recorded for the marker Satt 270 (0.345), closely followed by Satt 598 (0.326), demonstrating that these markers were moderately informative for assessing genetic diversity associated with carbohydrate content. In contrast, marker Satt324 exhibited the lowest PIC value (0.115), indicating limited discriminatory power among the soybean genotypes. According to the classification proposed by Botstein *et al.*, (1980), markers with PIC values below 0.25 are considered to possess low informativeness, whereas values between 0.25 and 0.50 indicate moderate informativeness. Therefore, Satt 270 and Satt 598 may be regarded as moderately informative SSR markers suitable for diversity analysis and marker assisted breeding, while Satt 324 showed relatively poor informativeness. Similar variation in PIC values among SSR loci has also been reported by Singh *et al.*, (2010), Zhang *et al.*, (2013), Bisen *et al.*, (2015) and Kumar *et al.*, (2022), emphasizing that marker informativeness depends on both genomic location and germplasm composition.

Population genetic diversity and structure analysis based on SSR markers associated with carbohydrate content

Population genetic diversity estimated using SSR

**Fig. 1:** Determination of the optimal number of genetic clusters (K) using the Evanno method (ΔK) derived from SSR marker data of soybean genotypes.

markers associated with carbohydrate content is presented in Table 4, while the Bayesian population structure is illustrated in Fig. 2-3. The Evanno ΔK method identified $K = 3$ as the most likely number of genetic clusters (Fig.1), indicating the presence of three genetically distinct sub-populations among the 153 soybean genotypes. The STRUCTURE bar plots (Fig. 2 and 3) further confirmed sub-divisions by assigning the genotypes in Pop1 (41 genotypes; red cluster), Pop2 (55 genotypes; green cluster) and Pop3 (57 genotypes; blue cluster) based on their membership coefficients.

Population 1 (Pop1) comprised 41 genotypes including JS 20-69 (7), JS 22-16 (22), NRC-258 (76), NRC-859 (84), NRC-7 (86), ASB-101 (94), DS-1510 (101), RVSM-2011-35 (107), AS-34 (132), KDS-1203 (141), JS 25-03 (30), RVS-15-1 (45), RVS-23-8 (49), NRC-267 (81), PS-1696 (98), TS-208 (126), EC109563 (137), Lok Soya-03 (150), JS 25-55 (39) and several other genotypes, which showed high membership coefficients and were predominantly assigned to the red genetic cluster. Population 2 (Pop2) consisted of 55 genotypes, including JS 20-29 (4), JS 20-53 (6), JS 21-17 (15), JS 23-03 (23), NRC-4 (85), RSC-11-95 (90), Himso-1696 (99), DS-1589 (103), DS-1480 (104), AUKS 21-5 (130), KDS-1201 (140), VLS-104 (144) and other related genotypes. While population 3 (Pop3) comprised 57 genotypes, including JS 25-53 (38), JS 21-72 (16), JS 22-08 (20), NRC-138 (67), RSC-1165 (91), ASB-85 (92), KSS-225 (110), TS-101 (128), KDS-1188 (138), JS 20-98 (11), JS 21-77 (18), JS 22-01 (19), RVS-76 (40), RVS 23-9 (50), RVS 23-23 (63), NRC-192 (72), NRC-255 (75), PS-1569 (97) and the remaining genotypes. Among 153 genotypes, 150 exhibited high membership coefficients (>0.80), indicating clear assignment to a single population, whereas 3 genotypes, viz., KDS-1188 (139), JS 23-09 (24) and JS 24-26 (26), showed mixed ancestry with genetic combinations from more than one cluster.

The population wise genetic diversity statistics (Table 4) revealed that the number of different alleles (N_a) was 2.000 across all populations. The effective number of alleles (N_e) ranged between 1.520 in Pop1 to 1.569 in Pop3, with an overall mean of 1.541, indicating a relatively more balanced allele frequency distribution in Pop3.

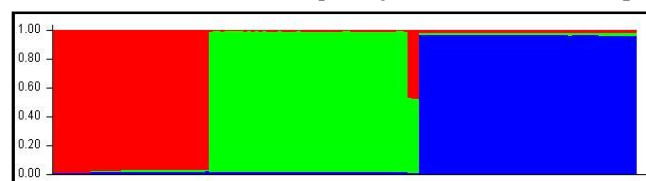
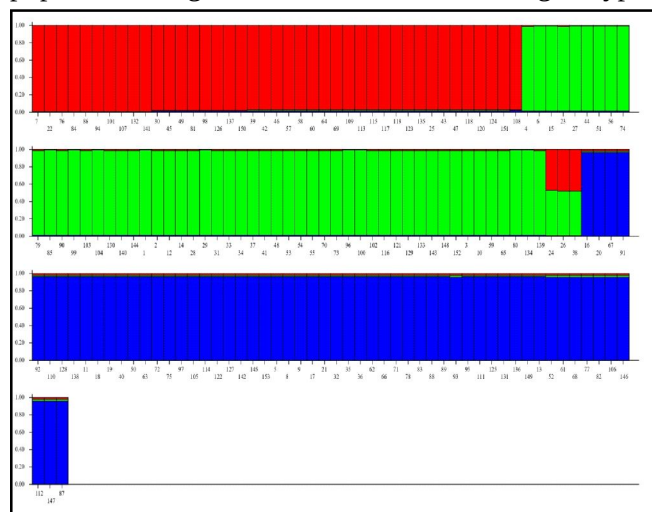
**Fig. 2:** Summary plot of population structure at $K = 3$ depicting proportional genetic membership of soybean genotypes into three distinct sub-populations.

Table 5: AMOVA among three soybean populations based on SSR markers associated with carbohydrate content.

Source of variation	df	SS	MS	Estimated variance	Variation (%)
Among populations	2	1.808	0.904	0.002	1
Among individuals within populations	150	103.055	0.687	0.344	99
Within individuals	153	0.000	0.000	0.000	0
Total	305	104.863	-	0.346	100

Similarly, Shannon's Information Index (I) ranged from 0.460 to 0.504 (mean = 0.486), while expected heterozygosity (He) varied from 0.302 to 0.335 (mean = 0.321) and unbiased expected heterozygosity (uHe) ranged between 0.306 to 0.338 (mean = 0.324). Among the three populations, population 3 consistently exhibited the highest Ne, I, He and uHe value, indicating comparatively greater genetic diversity, whereas Population 1 recorded the lowest diversity estimates. The fixation index ($F = 1.000$) for all three populations indicated complete homozygosity, which is consistent with the predominantly self-pollinated nature of soybean.

The clustering pattern obtained from the Bayesian STRUCTURE analysis was in close agreement with the diversity statistics, confirming the presence of moderate genetic differentiation among the evaluated soybean genotypes. The existence of genetically distinct populations together with a few admixed genotypes

**Fig. 3:** Population structure of 153 soybean genotypes inferred from SSR marker data showing membership coefficients for K=3 genetic clusters. Note: the genotype names and numbers can be verified from the Table 1.

suggests historical recombination and shared pedigree among the breeding materials (Silva *et al.*, 2025; Assfaw *et al.*, 2025). Similar levels of the population stratification and moderate molecular diversity have also been reported in soybean using SSR markers by Hyten *et al.*, (2006), Singh *et al.*, (2010), Gupta and Manjaya (2017), Zatybekov *et al.*, (2023) and Rani *et al.*, (2023). The comparatively higher genetic diversity observed in Pop3 indicates that genotypes belonging to this cluster serve as valuable parental resources for breeding programmes aimed at improving carbohydrate content through marker assisted selection and hybridization (Chiemeke *et al.*, 2024; Achina *et al.*, 2025).

Analysis of Molecular Variance (AMOVA) based on SSR markers associated with carbohydrate content

The analysis of molecular variance (AMOVA) based on SSR markers associated with carbohydrate content partitioned the total genetic variation among and within the three inferred soybean populations (Table 5, Fig. 4). The results revealed that 99% of the total molecular variation was attributed to differences among individuals within populations, whereas only 1% of the variation was observed among populations, indicating a very low level of genetic differentiation between the three STRUCTURE defined populations.

The variance component among populations was estimated at 0.002, with a sum of squares (SS) of 1.808 and a mean square (MS) of 0.904, accounting for only 1% of the total genetic variation. In contrast, the among individuals within populations component exhibited a substantially higher estimated variance (0.344), with an SS of 103.055 and an MS of 0.687, contributing 99% of the total molecular variation. No variation (0 %) was detected within individuals, resulting in an estimated variance of 0.000, which is expected for soybean because of its highly self-pollinated nature and consequently high level of homozygosity. The total estimated molecular variance across all sources was 0.346.

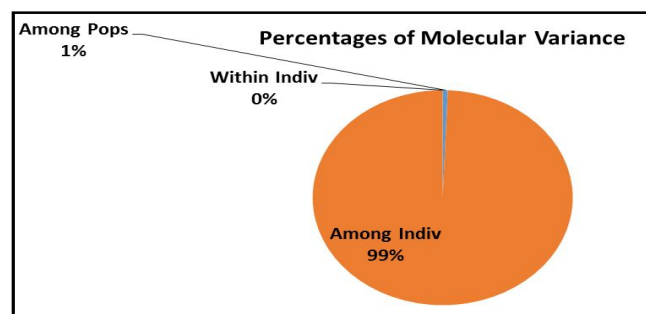
**Fig. 4:** Partitioning of molecular variance among and within soybean populations based on SSR markers linked to carbohydrate content.

Table 6: Summary of Principal Coordinate Analysis (PCoA) based on SSR marker data associated with carbohydrate content.

PCA	EV	PVE	CV
Axis 1	15.470	62.80	62.80
Axis 2	6.814	27.66	90.46
Axis 3	2.349	9.54	100.00
PCA: Principal coordinate axis; EV: Eigen value; PVE: Percentage variation explained (%); CV: Cumulative variation (%)			

The predominance of genetic variation within populations indicated that the soybean genotypes possess considerable molecular diversity despite exhibiting only weak population differentiation. This finding suggests that the three STRUCTURE derived populations share a substantial proportion of alleles and are not completely isolated, probably because of their common breeding history, frequent utilization of elite parental lines and exchange of germplasm during soybean improvement programmes (Chung *et al.*, 2014; Zhou *et al.*, 2015). Similar AMOVA patterns have also been widely reported in cultivated soybean by Fu *et al.*, (2021), Tsindi *et al.*, (2023) and Rani *et al.*, (2023), where most of the genetic variation resides within populations rather than among populations due to the relatively narrow genetic base of modern cultivars.

The AMOVA results are in close agreement with the Bayesian population structure analysis, which identified three genetic clusters with only limited differentiation and the presence of a few admixed genotypes. The very low proportion of variation among populations (1%) corroborates the moderate population structure inferred from STRUCTURE, while the predominance of variation within populations confirms the presence of sufficient allelic diversity to support effective selection within the existing germplasm (Wang *et al.*, 2015; Silue *et al.*, 2025).

Principal Coordinate Analysis (PCoA) based on SSR markers associated with carbohydrate content

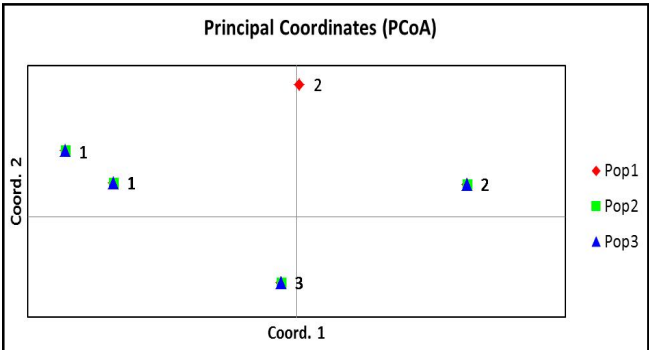


Fig. 5: PCoA depicting genetic relationships among 153 soybean genotypes based on SSR markers associated with carbohydrate content.

The genetic relationships among the 153 soybean genotypes were further investigated through Principal Coordinate Analysis (PCoA) using SSR markers associated with carbohydrate content (Table 6; Fig. 5). The PCoA effectively summarized the molecular variation by reducing the multidimensional marker data into a few principal coordinate axes, thereby facilitating the visualization of genetic relationships among the genotypes.

The first two coordinates explained 90.46% of the molecular variation, providing an adequate representation of genetic relationships. Axis 1 accounted for the largest proportion of variation, with an eigenvalue of 15.470, explaining 62.80% of the total genetic variation. Axis 2 contributed an additional 27.66% of the variation with an eigen value of 6.814, resulting in a cumulative variation of 90.46% for the first two axes. The third principal coordinate (Axis 3) had an eigenvalue of 2.349 and explained the remaining 9.54% of the variation, increasing the cumulative variation of 100%.

The predominance of the first principal coordinate indicates that the majority of molecular variation associated with the carbohydrate content is explained by a single major genetic dimension. Furthermore, the first two axes together explained more than 90% of the total variation, demonstrating that the two-dimensional PCoA plot provides an accurate representation of the genetic relationships among the soybean genotypes. The clustering pattern observed in the PCoA was consistent with the Bayesian population structure analysis, where the genotypes were grouped into three major genetic clusters with limited overlap and a few intermediate (admixed) genotypes. The high percentage of variation

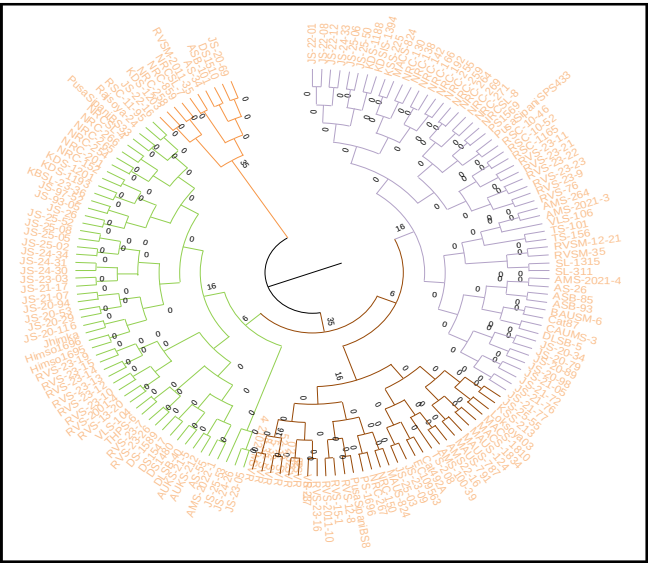


Fig. 6: UPGMA dendrogram illustrating genetic relationships among 153 soybean genotypes based on SSR marker polymorphism associated with carbohydrate content.

explained by the first two principal coordinate indicates that the SSR markers effectively discriminated the soybean genotypes according to their genetic similarity (Appiah-Kubi *et al.*, 2014; Gupta and Manjaya, 2017; Prysiachniuk *et al.*, 2025). Similar clustering patterns obtained from PCoA have also been reported in soybean by Kumar *et al.*, (2022), Rani *et al.*, (2023) and Zatybekov *et al.*, (2023), demonstrating the reliability of SSR markers for assessing genetic diversity and population structure. The agreement among PCoA, STRUCTURE analysis and AMOVA further confirms the presence of moderate genetic diversity with relatively low population differentiation among the evaluated soybean germplasm.

UPGMA cluster analysis based on SSR marker associated with carbohydrate content

The genetic relationships among the 153 soybean genotypes were further elucidated using Unweighted Pair Group Method with Arithmetic Mean (UPGMA) clustering based on SSR markers associated with carbohydrate content (Fig. 6). The dendrogram classified the genotypes into two major clusters, viz., Cluster I (minor cluster) and Cluster II (major cluster), indicating the presence of substantial genetic variation among the evaluated soybean germplasm.

Cluster I represented the minor cluster and consisted of 10 genotypes, which were clearly separated from the remaining genotypes, indicating presence of greater genetic divergence. This cluster was further divided into two sub-clusters. The major sub-cluster comprised six genotypes including JS 20-69, DS 1510, ASB-101, AS-34, RVSM 2011-35 and NRCSL-7, whereas the minor sub-cluster contained four genotypes *i.e.*, NRC-859, NRC-258, KDS-1203 and JS 22-16. This distinct grouping of these genotypes suggests the presence of unique allelic combinations at carbohydrate associated SSR loci, making them potential sources of novel genetic variation.

Cluster II was the major cluster, comprising the remaining 143 genotypes, and was further sub divided into two principal sub-clusters, viz., green and brown groups. The green sub-cluster included 55 genotypes and was further partitioned into two sub-groups. The larger subgroup contained 52 genotypes, including Jhimka, RVS 23-19, DS 1589, AS-55, AMS 2022-1, DLSB-40, THPS-6, VLS-104, JS 20-53, JS 20-94, JS 24-30, JS 25-05, JS-26, KBSL 23-26, KDS-1201, Rajsoya-24, RSC 11-95, JS 93-05, NRC-152, JS 24-31, RVS 23-15, RVS 23-10, RVS 23-13 and RVSM 2001-4, among others, whereas the smaller subgroup consisted of only three genotypes namely JS 23-05, JS 24-26 and JS 25-53. The brown sub-cluster comprised 88 genotypes and was further differentiated into two major groups. The first subgroup

included 31 genotypes, such as RVSM 2012-4, RVS 23-8, Pusa Sipani BS-8, JS 23-09, MAUS-824, Cat492A, TS-208, MAUS-787, Lok Soya-03, JS 25-55, PS 1696, JS 25-03, NRC-267, RVS 15-1, NRC-150 and EC-109563, together with several closely related genotypes. The second subgroup represented by the purple cluster, consisted of 57 genotypes and formed the largest and most genetically homogenous group within the dendrogram. This group included JS 20-79, JS 20-98, SL-1315, DLSB-5, RVSM-35, AMS-264, RVS 23-23, AS-26, CAUMS-3, JS 20-34, JS 21-72, Cat-87, TS-101, TS-156, NRC-264, JS 22-01, JS 25-06, KDS-1188, KDSIS-1394, Pusa Sipani SPS-433, RSC 10-46, RSC-1165, NRC-130, MACS-824 and JS 24-33, among others, indicating their close genetic relationship among these genotypes.

The UPGMA clustering pattern revealed that the majority of soybean genotypes were grouped into the major cluster, suggesting a relatively narrow genetic base among cultivated soybean lines, whereas the small distinct Cluster I represented genetically divergent germplasm. The separation of Cluster I from the remaining genotypes indicated that these accessions possess comparatively unique genetic backgrounds and may therefore serve as valuable parental lines for broadening the genetic base in soybean improvement programmes targeting carbohydrate content (Khatab *et al.*, 2016; Kumar *et al.*, 2022). The present findings indicated that the genotypes grouped in the minor Cluster I and the smaller sub-clusters of Cluster II represent valuable sources of genetic diversity and could be utilized as parents in hybridization programmes to maximize recombination and enhance carbohydrate related traits. The clustering pattern observed in the present study is consistent with earlier reports, where soybean genotypes were grouped into two or more clusters based on SSR marker analysis, reflecting their genetic relationships and breeding history. Similar clustering has also been reported by Tantasawat *et al.*, (2011) (four clusters), Ghosh *et al.*, (2014) (two clusters with six sub-clusters), Chauhan *et al.* (2015) (two clusters), Hipparangi *et al.*, (2017) (three clusters) and Kumar *et al.*, (2022) (three clusters).

Conclusion

This study delivers the carbohydrate trait specific molecular diversity assessment of soybean germplasm, filling critical gap in a seed quality dimension long overshadowed by protein and oil-centric breeding efforts. Gene-linked SSR profiling of 153 genotypes identified Satt 270 and Satt 598 as moderately informative markers for capturing allelic variation at carbohydrate biosynthesis loci, while Bayesian clustering resolved three genetically

distinct subpopulations, with Population 3 emerging as the richest reservoir of the usable diversity. Although AMOVA confirmed that genetic differentiation among populations remains minimal, with variance overwhelmingly partitioned within populations, PCoA and UPGMA analysis converged in isolating a genetically divergent minor cluster of ten genotypes, marking these as high-value, previously untapped parental resources for carbohydrate trait improvement. From a transitional standpoint, these findings furnish breeders with a validated, marker-based framework for informed parental selection, targeted hybridization and marker assisted selection strategies with aim to optimize the sucrose-to-raffinose family oligosaccharide ratio, a trait axis with direct consequences for feed digestibility, monogastric animal nutrition and soy-food processing quality. The identified divergent accessions and carbohydrate associated loci provide useful preliminary resources for soybean improvement. Further validation using biochemical phenotyping and genome-wide marker systems will help confirm their contribution to carbohydrate related seed quality traits.

Declaration: All authors declare that no competing interest exists.

AI declaration: No data has been generated using AI in this manuscript. Only the paraphrasing tools have been used for enhancement of this manuscript.

Abbreviations: SSR: Simple Sequence Repeats; PIC: polymorphic information content; MCMC: Markov Chain Monte Carlo; AMOVA: Analysis of molecular variance; UPGMA: Unweighted Pair Group Method with Arithmetic Mean; JNKVV: Jawaharlal Nehru Krishi Vishwa Vidyalaya; RFOs: Raffinose family oligosaccharides; RVSKVV: Rajmata Vijayaraje Scindia Krishi Vishwa Vidyalaya; ZARS: Zonal Agricultural Research Station; PCR: Polymerase Chain Reaction; PCoA: Principal Coordinates Analysis.

References

- Achina, T., Amoako F.A., Amoah S., Karikari B., Acheampong S., Quain M.D. and Salifu S.P. (2025). Molecular characterization of soybean accessions using SSR markers and unveiling the genetic diversity. *Mol. Biol. Rep.* **52**, 563. <https://doi.org/10.1007/s11033-025-10652-7>
- Allen, D.K. and Young J.D. (2013). Carbon and nitrogen provisions alter the metabolic flux in developing soybean embryos. *Plant Physiol.* **161**, 1458-1475. <https://doi.org/10.1104/pp.112.203299>
- Appiah-Kubi, D., Asibuo J.Y., Quain M.D., Oppong A. and Akromah R. (2014). Diversity studies on soybean accessions from three countries. *Biocatal. Agric. Biotechnol.* **3**, 198-206. <https://doi.org/10.1016/j.bcab.2013.11.008>
- Assfaw, A.T., Bunmi O., Paterne A., Chigeza G., Mushoriwa H., Fowobaje K. and Abebe A.T. (2025). Genetic diversity and population structure analysis of soybean [*Glycine max* (L.) Merrill] genotypes based on agro-morphological traits and SNP markers. *PLoS One*, **20**(10), e0332895. <https://doi.org/10.1371/journal.pone.0332895>
- Banti, M. (2021) Raffinose family oligosaccharides, occurrence in food materials, nutritional implication and methods of analysis, a review. *World J. Food Sci. Technol.* **5**(3), 37-44. <https://doi.org/10.11648/j.wjfst.20210503.11>
- Bisen, A., Khare D., Nair P. and Tripathi N. (2015). SSR analysis of 38 genotypes of soybean (*Glycine max* (L.) Merr.) genetic diversity in India. *Physiol. Mol. Biol. Plants*, **21**(1), 109-115. <https://doi.org/10.1007/s12298-014-0269-8>
- Botstein, D., White R.L., Skolnick M. and Davis R.W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32**(3), 314-331.
- Chauhan, D.K., Bhat J.A., Thakur A.K., Kumari S., Hussain Z. and Satyawathi C.T. (2015). Molecular characterization and genetic diversity assessment in soybean (*Glycine max* (L.) Merr.) varieties using SSR markers. *Indian J. Biotechnol.* **14**(4), 504-510.
- Chiemeké, F.K., Olanmi B., Agre P.A., Mushoriwa H., Chigeza G. and Abebe A.T. (2024) Genetic diversity and population structure analysis of soybean [*Glycine max* (L.) Merrill] genotypes using agro-morphological traits and SNP markers. *Genes*, **15**(11), 1373. <https://doi.org/10.3390/genes15111373>
- Chowdhury, S., Kang B.H., Lee W.H., Shin S.Y., Kim D.Y., Kim W.J., Baek J.H., Kim S.H. and Ha B.K. (2026). Genome-wide identification of candidate loci associated with the regulation of the protein, oil, and carbohydrate content in soybean. *Plants*, **15**(6), 924. <https://doi.org/10.3390/plants15060924>
- Chung, W.H., Jeong N., Kim J., Lee W.K., Lee Y.G., Lee S.H., Yoon W., Kim J.H., Choi I.Y., Choi H.K., Moon J.K., Kim N. and Jeong S.C. (2014). Population structure and domestication revealed by high-depth resequencing of Korean cultivated and wild soybean genomes. *DNA Res.* **21**(2), 153-167. <https://doi.org/10.1093/dnares/dst047>
- Duan, Z., Li Q., Wang H., He X. and Zhang M. (2023). Genetic regulatory networks of soybean seed size, oil and protein contents. *Front. Plant Sci.* **14**, 1160418. <https://doi.org/10.3389/fpls.2023.1160418>
- Elango, D., Rajendran K., Van der Laan L., Sebastiar S., Raigne J., Thaiparambil N.A., El-Haddad N., Raja B., Wang W., Ferela A., Chiteri K.O., Thudi M., Varshney R.K., Chopra S., Singh A. and Singh A.K. (2022). Raffinose family oligosaccharides: Friend or foe for human and plant health? *Front. Plant Sci.*, **13**: 829118. <https://doi.org/10.3389/fpls.2022.829118>
- Escamilla, D.M., Xavier A., Vuong T.D., Nguyen H.T. and Rainey K.M. (2023). An assessment of the interaction between sucrose content and seed quality traits in soybeans. *Crop Sci.* **63**, 2650-2664. <https://doi.org/10.1002/csc2.21027>

- Evanno, G., Regnaut S. and Goudet J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **14**(8), 2611-2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Fu, Y.B., Cober E.R., Morrison M.J., Marsolais F., Peterson G.W. and Horbach C. (2021). Patterns of genetic variation in a soybean germplasm collection as characterized with genotyping-by-sequencing. *Plants*, **10**(8), 1611. <https://doi.org/10.3390/plants10081611>
- Ghosh, J., Ghosh P.D. and Choudhury P.R. (2014). An assessment of genetic relatedness between soybean [*Glycine max* (L.) Merrill] cultivars using SSR markers. *Am. J. Plant Sci.* **5**(20), 3089-3096. <https://doi.org/10.4236/ajps.2014.520325>
- Gilman, J.D. and Bilyeu K.D. (2012). Genes and alleles for quality traits on the soybean genetic/physical map. In : *Designing Soybeans for 21st Century Markets*. Elsevier, 67-96. <https://doi.org/10.1016/B978-0-9830791-0-1.50009-1>
- Gupta, S. and Manjaya J.G. (2017). Genetic diversity and population structure of Indian soybean [*Glycine max* (L.) Merr.] revealed by simple sequence repeat markers. *J. Crop Sci. Biotechnol.* **20**(3), 221-231. <https://doi.org/10.1007/s12892-017-0023-0>
- Hipparagi, Y., Singh R., Choudhury D.R. and Gupta V. (2017). Genetic diversity and population structure analysis of Kala bhat (*Glycine max* (L.) Merrill) genotypes using SSR markers. *Hereditas*, **154**, 9. <https://doi.org/10.1186/s41065-017-0030-8>
- Hyten, D.L., Song Q., Zhu Y., Choi I.Y., Nelson R.L., Costa J.M., Specht J.E., Shoemaker R.C. and Cregan P.B. (2006). Impacts of genetic bottlenecks on soybean genome diversity. *Proc. Natl. Acad. Sci. U.S.A.* **103**(45), 16666-16671. <https://doi.org/10.1073/pnas.0604379103>
- Kambhampati, S., Aznar-Moreno J.A., Hostetler C., Caso T., Bailey S.R., Hubbard A.H., Durrett T.P. and Allen D.K. (2020). On the inverse correlation of protein and oil: Examining the effects of altered central carbon metabolism on seed composition using soybean fast neutron mutants. *Metabolites*, **10**(1), 18. <https://doi.org/10.3390/metabo10010018>
- Khatib, I.A., Morsy A.R., Fares W.M. and Kumamaru T. (2016). Genetic diversity of soybean genotypes revealed by agromorphological and SSR markers. *J. Fac. Agric. Kyushu Univ.* **61**(1), 79-87. <https://doi.org/10.5109/1564086>
- Kumar, R., Saini M., Taku M., Debbarma P., Kumar Mahto R., Sharma D., Rawat S., Ray P. and Talukdar A. (2025). Grain structure, composition, and its characteristics. In : *Soybeans*. Apple Academic Press, 15-32. <https://doi.org/10.1201/9781998511730-3>
- Kumar, S.P.J., Chowdhury S., Sripathy K.V., Agarwal D.K., Pal G., Singh A.N., Kumar S., Rai A.K. and Simal-Gandara J. (2022). Molecular characterization and genetic diversity studies of Indian soybean (*Glycine max* (L.) Merr.) cultivars using SSR markers. *Mol. Biol. Rep.* **49**(3), 2129-2140. <https://doi.org/10.1007/s11033-021-07030-4>
- Lee, D., Lara L., Moseley D., Vuong T.D., Shannon G., Xu D. and Nguyen H.T. (2023). Novel genetic resources associated with sucrose and stachyose content through genome-wide association study in soybean (*Glycine max* (L.) Merr.). *Front. Plant Sci.* **14**, 1294659. <https://doi.org/10.3389/fpls.2023.1294659>
- Li, L., Hur M., Lee J.Y., Zhou W., Song Z., Ransom N., Demirkale C.Y., Nettleton D., Westgate M., Arendsee Z., Iyer V., Shanks J., Nikolau B. and Wurtele E.S. (2015). A systems biology approach toward understanding seed composition in soybean. *BMC Genomics*, **16**(Suppl 3), S9. <https://doi.org/10.1186/1471-2164-16-S3-S9>
- Li, Y., Zhao S., Ma J., Li D., Yan L., Li J., Qi X., Guo X., Zhang L. and He W. et al. (2013). Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing. *BMC Genomics*, **14**, 579. <https://doi.org/10.1186/1471-2164-14-579>
- Mishra, R., Shrivastava M.K., Tripathi M.K., Amrate P.K., Singh Y., Solanki R., Sharma S. and Shukla S. (2025a). Unravelling soybean yield potential: Exploring trait synergy, impact pathways, multidimensional patterns and biochemical insights. *Plant Sci. Today*, **12**(2), 1-10. <https://doi.org/10.14719/pst.6401>
- Mishra, R., Yamini G., Goutam M., Sanjeev S., Singh J. and Tripathi N. et al. (2025b). Unlocking qualitative traits diversity in soybean (*Glycine max* (L.) Merrill) genotypes. *Plant Sci. Today*, **12**. <https://doi.org/10.14719/pst.8144>
- Mishra, R., Shrivastava M.K., Amrate P.K., Goutam J., Pandey A.K. and Sanjeev S. et al. (2025c). Resistance profiling of soybean RILs against yellow mosaic virus and charcoal rot under field conditions. *Plant Sci. Today*. <https://doi.org/10.14719/pst.9379>
- Moretti, A., Arias C.L., Mozzoni L.A., Chen P., McNeece B.T., Mian M.A.R., McHale L.K. and Alonso A.P. (2020). Workflow for the quantification of soluble and insoluble carbohydrates in soybean seed. *Molecules*, **25**(17), 3806. <https://doi.org/10.3390/molecules25173806>
- Murray, M.G. and Thompson W.F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* **8**(19), 4321-4326. <https://doi.org/10.1093/nar/8.19.4321>
- Nawaz, M.A., Yang S.H. and Chung G. (2018). Wild soybeans: An opportunistic resource for soybean improvement. In: *Rediscovery of Landraces as a Resource for the Future*. InTechOpen. <https://doi.org/10.5772/intechopen.74973>
- Patil, G., Mian R., Vuong T., Pantalone V., Song Q., Chen P., Shannon G.J., Carter T.C. and Nguyen H.T. (2017). Molecular mapping and genomics of soybean seed protein: A review and perspective for the future. *Theor. Appl. Genet.* **130**(10), 1975-1991. <https://doi.org/10.1007/s00122-017-2955-8>
- Prysiashniuk, L.M., Parii M.F., Iliuchenko A.O., Kozlova S.O., Shliakhtun I.S., Korol L.V. and Melnyk S.I. (2025). Analysis of soybean (*Glycine max* L.) breeding material using SSR markers. *Plant Var. Stud. Prot.* **21**(3), 112-120. <https://doi.org/10.21498/2518-1017.21.3.2025.339310>
- Rani, R., Raza G., Tung M.H., Rizwan M., Ashfaq H., Shimelis H., Razzaq M.K. and Arif M. (2023). Genetic diversity and population structure analysis in cultivated soybean (*Glycine max* [L.] Merr.) using SSR and EST-SSR markers. *PLoS One*, **18**(5), e0286099. <https://doi.org/10.1371/journal.pone.0286099>
- Rej, S., Banerjee S., Lal S.K. and Kumar P.R. (2024). Understanding the significance of raffinose family

- oligosaccharides in seed physiology. *J. Adv. Biol. Biotechnol.* **27(9)**, 1463-1476. <https://doi.org/10.9734/jabb/2024/v27i91420>.
- Rosso, A., Castanedo M., Meloni R., Pelosin P., Corredig M., Martínez-Abad A. and Blandino M. (2026). Agronomic traits and nutritional features of importance in low-raffinose genotypes for added-value soybeans. *J. Agric. Food Res.* **26**, 102654. <https://doi.org/10.1016/j.jafr.2026.102654>
- Saleem, A., Muylle H., Aper J., Ruttink T., Wang J., Yu D. and Roldán-Ruiz I. (2021). A genome-wide genetic diversity scan reveals multiple signatures of selection in a European soybean collection compared to Chinese collections of wild and cultivated soybean accessions. *Front. Plant Sci.* **12**, 631767. <https://doi.org/10.3389/fpls.2021.631767>
- Sanyal, R., Kumar S., Pattanayak A., Kar A. and Bishi S.K. (2023). Optimizing raffinose family oligosaccharides content in plants: A tightrope walk. *Front. Plant Sci.* **14**, 1134754. <https://doi.org/10.3389/fpls.2023.1134754>
- Silue, T., Agre P.A., Olasanmi B., Adewumi A.S., Adejumbi I.I. and Abebe A.T. (2025). Genetic diversity and population structure of soybean (*Glycine max* (L.) Merrill) germplasm. *PLoS One*, **20(5)**, e0312079. <https://doi.org/10.1371/journal.pone.0312079>
- Silva, A.C. da, Silva D.C.G., Ferreira E.G.C., Abdelnoor R.V., Borém A., Arias C.A., Oliveira M.F., Oliveira M.E.F. and Marcelino-Guimarães F.C. (2025). Genetic diversity, population structure in a historical panel of Brazilian soybean cultivars. *PLoS One*, **20(1)**, e0313151. <https://doi.org/10.1371/journal.pone.0313151>
- Singh, R.K., Bhatia V.S., Bhat K.V., Mohapatra T., Singh N.K., Bansal K.C. and Koundal K.R. (2010). SSR and AFLP based genetic diversity of soybean germplasm differing in photoperiod sensitivity. *Genet. Mol. Biol.*, **33(2)**, 319-324. <https://doi.org/10.1590/S1415-47572010005000024>
- Song, Y., Wang X.D. and Rose R.J. (2017). Oil body biogenesis & biotechnology in legume seeds. *Plant Cell Rep.* **36(10)**, 1519-1532. <https://doi.org/10.1007/s00299-017-2201-5>
- Tamagno, S., Sadras V.O., Aznar-Moreno J.A., Durrett T.P. and Ciampitti I.A. (2022). Selection for yield shifted the proportion of oil and protein in favor of low-energy seed fractions in soybean. *Field Crops Res.* **279**, 108446. <https://doi.org/10.1016/j.fcr.2022.108446>
- Tiwari, P.N., Tiwari S., Sapre S., Babbar A., Tripathi N., Tiwari S. and Tripathi M.K. (2023). Prioritization of microsatellite markers linked with drought tolerance associated traits in chickpea (*Cicer arietinum* L.). *Legume Res.* **46(11)**, 1422-1430. <https://doi.org/10.18805/LR-5191>
- Tripathi, N., Tripathi M.K., Tiwari S. and Payasi D.K. (2022). Molecular breeding to overcome biotic stresses in soybean: update. *Plants*, **11**, 1967. <https://doi.org/10.3390/plants11151967>
- Tsindi, A., Eleblu J.S.Y., Gasura E., Mushoriwa H., Tongoona P., Danquah E.Y., Mwadzingeni L., Zikhali M., Ziramba E., Mabuyaye G. and Derera J. (2023). Analysis of population structure and genetic diversity in a Southern African soybean collection based on single nucleotide polymorphism markers. *CABI Agric. Biosci.* **4(1)**, 15. <https://doi.org/10.1186/s43170-023-00158-2>
- United States Department of Agriculture, Foreign Agricultural Service (USDA-FAS) (2026) Production—Soybeans. Washington, DC: United States Department of Agriculture, Foreign Agricultural Service. Available from: <https://www.fas.usda.gov/data/production/2222000>
- Valentine, M.F., De Tar J.R., Mookkan M., Firman J.D. and Zhang Z.J. (2017). Silencing of soybean raffinose synthase gene reduced raffinose family oligosaccharides and increased true metabolizable energy of poultry feed. *Front. Plant Sci.*, **8**, 692. <https://doi.org/10.3389/fpls.2017.00692>
- Valliyodan, B., Qiu D., Patil G., Zeng P., Huang J., Dai L., Chen C., Li Y., Joshi T., Song L., Vuong T.D., Musket T.A., Xu D., Shannon J.G., Chen S., Liu X. and Nguyen H.T. (2016). Landscape of genomic diversity and trait discovery in soybean. *Sci. Rep.* **6**, 23598. <https://doi.org/10.1038/srep23598>
- Vargas-Almendra, A., Ruiz-Medrano R., Núñez-Muñoz L.A., Ramírez-Pool J.A., Calderón-Pérez B. and Xoconostle-Cázares B. (2024). Advances in soybean genetic improvement. *Plants*, **13(21)**, 3073. <https://doi.org/10.3390/plants13213073>
- Viana, J.P.G., Fang Y., Avalos A., Song Q., Nelson R. and Hudson M.E. (2022). Impact of multiple selective breeding programs on genetic diversity in soybean germplasm. *Theor. Appl. Genet.* **135(5)**, 1591-1602. <https://doi.org/10.1007/s00122-022-04056-5>
- Wang, Y., Chen P. and Zhang B. (2014). Quantitative trait loci analysis of soluble sugar contents in soybean. *Plant Breed.* **133(4)**, 493-498. <https://doi.org/10.1111/pbr.12178>
- Wang, Y.H., Zhang X.J. and Fan S.J. (2015). Genetic diversity of wild soybean populations in Dongying, China, by simple sequence repeat analysis. *Genet. Mol. Res.* **14(3)**, 11613-11623. <https://doi.org/10.4238/2015.September.28.13>
- Zatybekov, A., Yermagambetova M., Genievskaya Y., Didorenko S. and Abugalieva S. (2023). Genetic diversity analysis of soybean collection using simple sequence repeat markers. *Plants*, **12(19)**, 3445. <https://doi.org/10.3390/plants12193445>
- Zhang, G., Xu S., Mao W., Hu Q. and Gong Y. (2013). Determination of the genetic diversity of vegetable soybean (*Glycine max* (L.) Merr.) using EST-SSR markers. *J. Zhejiang Univ. Sci. B*, **14(4)**, 279-288. <https://doi.org/10.1631/jzus.B1200243>
- Zhawar, V.K., Kaur N. and Gupta A.K. (2011). Phytic acid and raffinose series oligosaccharides metabolism in developing chickpea seeds. *Physiol. Mol. Biol. Plants*, **17(4)**, 355-362. <https://doi.org/10.1007/s12298-011-0080-8>
- Zhou, Z., Jiang Y., Wang Z., Gou Z., Lyu J., Li W., Yu Y., Shu L., Zhao Y., Ma Y., Fang C., Shen Y., Liu T., Li C., Li Q., Wu M., Wang M., Wu Y. and Dong Y. *et al.* (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* **33(4)**, 408-414. <https://doi.org/10.1038/nbt.3096>
- Zhuang, Y., Li X., Hu J., Xu R. and Zhang D. (2022). Expanding the gene pool for soybean improvement with its wild relatives. *aBIOTECH*, **3(2)**, 115-125. <https://doi.org/10.1007/s42994-022-00072-7>