scientific reports

Check for updates

OPEN

Genome-wide assessment of genetic diversity and selective signatures in sunflower (*Helianthus annuus* L.) using a 10 K SNP array

Reza Darvishzadeh[1]✉, Hadi Alipour[1]✉, Aras Türkoğlu[2], Kamil Haliloğlu[3] & Sima Fatanatvash[1]

Sunflower (*Helianthus annuus* L.) is a valuable oilseed crop with significant economic importance. Enhancing breeding strategies in this species heavily relies on genetic diversity and molecular tools. Among these tools, SNP-based genotyping has proven to be an effective method for exploring genomic variation and understanding population structure in plants. This study aimed to evaluate the genetic diversity, population structure, and identify genomic regions under selection within a panel of sunflower inbred lines. A 10 K SNP (single-nucleotide polymorphism) array was employed to examine genetic variation among 94 sunflower inbred lines. The results showed that SNPs were unevenly distributed across the 17 sunflower chromosomes, with a higher density observed in telomeric regions. The average transition-to-transversion ratio was 3.75, confirming the reliability and quality of the genotypic data. Analysis of population structure revealed two distinct subgroups, as supported by the results from STRUCTURE, PCoA, and WPGMA methods. AMOVA revealed that 16% of total genetic variance occurred among populations (Fst = 0.156), and overall genetic diversity was moderate to high (PPL = 92.94%). A total of 283 potential genes were linked to genomic areas that were under selection. Functional enrichment analysis revealed significant involvement of these genes in *proteasome* activity and *pyruvate metabolism* pathways. The findings validate the presence of considerable genetic diversity and moderate genetic differentiation within the studied sunflower germplasm. This study reports, for the first time, the use of the Fst statistic to identify pathways that contribute to population differentiation and are under selection pressure in a panel of sunflower inbred lines. These results provide valuable insights that could inform future sunflower breeding efforts through marker-assisted selection and offer insights into the genetic mechanisms underlying adaptation and population divergence in sunflower.

Keywords Population structure, Fst statistic, Molecular markers, Heterozygosity, *Helianthus annuus*

Sunflower (*Helianthus annuus* L.) is an oilseed crop native to North America and one of the most important oil seed crops globally due to its high nutritional value and industrial applications[1]. It is a highly cross-pollinated crop with 2n = 34 chromosomes, exhibiting remarkable yield potential and adaptability to diverse environmental conditions[2]. According to reports from the United States Department of Agriculture (USDA), sunflower oil ranks third globally among edible oils derived from oilseeds, after soybean and rapeseed (canola). It is also rich in essential fatty acids such as linoleic acid (omega-6) and is important for human health[3].

Sunflower hybrid varieties were first introduced in the early 1970s through the crossing of cytoplasmic male-sterile lines and fertility-restorer lines, marking a significant milestone in sunflower breeding[4]. Genetic diversity in plants, especially in sunflower, plays a key role in breeding and improving agronomic traits[5]. Genotypic selection requires diversity, and with increased genetic diversity in a population, the scope of both natural and artificial selection expands. In fact, genetic diversity is essential for farmers and breeders to develop new cultivars with high yield, adaptability to environmental stresses, and resistance to pests and pathogens[6]. A genetically diverse germplasm collection is essential for enhancing crop traits, as it promotes effective allele accumulation and streamlines the screening process[7].

[1]Department of Plant Production and Genetics, Faculty of Agriculture, Urmia University, Urmia, Iran. [2]Department of Field Crops, Faculty of Agriculture, Necmettin Erbakan University, 42310 Konya, Turkey. [3]Department of Field Crops, Faculty of Agriculture, Ataturk University, 25240 Erzurum, Turkey. ✉email: r.darvishzadeh@urmia.ac.ir; ha.alipour@urmia.ac.ir

Since the morphological characterization of plant genetic resources has been limited to a few phenotypic traits, which are strongly influenced by environmental factors and show little variation, especially for highly heritable traits[8]. Therefore, the use of molecular markers as a precise and stable tool for evaluating genetic diversity and identifying important traits in plants is essential[9].

A variety of molecular markers have been utilized to evaluate genetic diversity and identify potential heterotic groups in sunflower. These markers include restriction fragment length polymorphism (RFLP)[10], amplified fragment length polymorphism (AFLP)[11,12], random amplified polymorphic DNA (RAPD)[13], simple sequence repeats (SSR)[10,14], and single nucleotide polymorphism (SNP)[15,16]. The selection of each marker system depends on the specific application, the degree of polymorphism, available technical facilities and expertise, as well as financial limitations[17].

Single nucleotide polymorphisms (SNPs) are the most common sequence variations in plant genomes and are widely used for analyzing genetic variation, population structure, marker-trait association, genomic selection, QTL mapping, and other breeding applications requiring extensive genome coverage[18]. SNP arrays based on hybridization and genotyping-by-sequencing (GBS) are among the most widely used platforms for genotyping and have been effectively applied in crop species. While GBS tends to have a high rate of missing data due to its low sequencing depth, SNP arrays provide higher marker density[19].

Genotyping arrays serve as a practical alternative to whole-genome sequencing for acquiring genomic insights. Various arrays have been developed for SNP genotyping in sunflower, including 46 K[20], 7 K and 5 K[21], 1 K[22], 25 K[23].

Several studies have investigated the genetic diversity of sunflower. For example, a study developed an SNP array for sunflower, containing over 25 K markers. This array was successfully used for genotyping a panel of lines, hybrids, and mapping populations, accurately reflecting the genetic diversity of cultivated sunflower[23]. Filippi et al.[24] used genotyping-by-sequencing (GBS) and generated a matrix of 11,834 common SNPs to assess the genetic diversity in sunflower germplasm collections from INTA, INRA, and USDA-UBC. The results showed that the genetic diversity was generally moderate. Population structure analyses and linkage disequilibrium (LD) patterns revealed structural differences between the collections, with the highest LD observed on chromosomes Chr10, Chr17, Chr5, and Chr2. The results of Mandel et al.[25] using a 10 K Illumina SNP chip on the USDA-UBC accessions indicated that K = 3, supporting the existence of three genetically distinct clusters.

Given the importance of genetic diversity in sunflower germplasm and the limited application of GBS (Genotyping-by-Sequencing) in sunflower breeding programs, this study aimed to evaluate the utility of a 10 K SNP array for assessing genetic diversity, population structure, and identifying genomic regions under selection.

Materials and methods

Plant materials and SNP genotyping

In this study, 94 sunflower genotypes (*Helianthus annuus* L.) obtained from various research centers in France, Iran, and the USA were examined (Table 1). The genotypes were chosen to include lines with known resistance to diseases (e.g., Phoma, Sclerotinia) and tolerance to environmental stresses (e.g., drought, salinity, phosphorus deficiency).

Genomic DNA was extracted from young leaves of 15-day-old seedlings following the method described by Dellaporta et al.[26] with slight modifications. The SNP genotyping of genomic DNA was conducted by TraitGenetics, Gatersleben, Germany (http://www.traitgenetics.com), using the 10 K sunflower Genotyping Array[27], which contained 9,250 SNP markers. The variants were selected from the 25 K Infinium array and identified using the reference sequence of sunflower[16]. SNP markers with heterozygosity greater than 10%, minor allele frequency (MAF) less than 5%, and more than 10% missing data were filtered and removed to obtain more accurate results. Finally, 7,909 SNP markers remained and were used for further analysis. Their genomic distribution and density across 94 sunflower lines were visualized using the CMplot online tool at Bioinformatics.com.cn (https://www.bioinformatics.com.cn/plot_basic_SNP_density_by_CMplot_107_en).

Linkage disequilibrium (LD)

Linkage disequilibrium (LD) among SNP marker pairs was estimated using TASSEL version 5 software (https://www.maizegenetics.net/tassel). In this analysis, LD within the sunflower genome was calculated as a squared allele frequency correlation (r^2) for SNP pairs with P-values < 0.001. This threshold ensures that only significant SNP pairs are included in the LD analysis. Also, LD decay distance was determined by plotting r^2 values against genetic distance (in bp) using quantile regression with the quantreg package in R version 4.4.1 (https://CRAN.R-project.org/package=quantreg9:13).

Evaluation of population structure

To analyze the genetic structure and determine the most probable number of genetically distinct groups (K), the dataset was analyzed using the Bayesian model-based clustering algorithm in Structure version 2.3.4 software (https://www.stucture-software.org)[28]. A total of 20 runs were performed for each K value (K = 1–10) with 100,000 Markov chain Monte Carlo (MCMC) iterations and a burn-in period of 100,000 runs, based on an admixture model and correlated allele frequency[29]. The optimal number of subpopulations was identified using the Structure Selector program (https://lmme.ac.cn/StructureSelector/) by determining the best K value. Additionally, the 94 sunflower accessions were grouped hierarchically using the weighted pair group method with arithmetic mean (WPGMA) based on the similarity matrix[30] in DARwin version 6.010 software (https://darwin.cirad.fr)[31]. PCoA analyse was also performed in GenAlex version 6.501 (https://biology-assets.anu.edu.au/GenAlEx).

Entry No	Line name	Country	Group	Entry No	Lines name	Country	Group
1	H100A/83HR4	FRANCE	ASGROW	50	H205A/83HR4	FRANCE	ASGROW
2	H209A/LC1064	FRANCE	ASGROW	51	RHA265	FRANCE	ASGROW
3	H205A/H543R	FRANCE	ASGROW	52	PM1-3	USA	USDA
4	AS5306	FRANCE	ENSAT	53	RT948	FRANCE	RUSTICA
5	RHA858	USA	USDA	54	ENSAT-283	FRANCE	ENSAT
6	H209A/83HR4	FRANCE	ASGROW	55	QHP1	FRANCE	INRAMONT
7	AS3211	FRANCE	ENSAT	56	SDR19	USA	USDA
9	AS5304	FRANCE	ASGROW	57	HA33B	USA	USDA
10	1,009,329.2(100 K)	FRANCE	ENSAT	58	H100B	FRANCE	ASGROW
12	AS613	FRANCE	ASGROW	59	B454/03	HUNGARY	–
13	AF1 POPA	FRANCE	NOVARTIS	60	HA304	USA	USDA
14	OES	FRANCE	INRAMONT	61	RT931	FRANCE	RUSTICA
15	H100A/LC1064	FRANCE	ASGROW	62	HA335B	USA	USDA
16	RHA266	USA	USDA	63	NS-B5	FRANCE	NOVARTIS
17	PAC2	FRANCE	ENSAT	64	SDB3	USA	USDA
18	H157A/LC1064	FRANCE	ASGROW	65	LC1064C	FRANCE	ASGROW
19	5DES20QR	FRANCE	BRN	66	NS-R5	FRANCE	NOVARTIS
20	1,009,337 (100 K)	FRANCE	ENSAT	67	DM2	USA	USDA
21	AS3232	FRANCE	ENSAT	68	H156A/RHA274	FRANCE	ASGROW
22	12ASB3	FRANCE	ASGROW	69	SDB1	USA	USDA
23	8ASB2	FRANCE	ASGROW	71	AS5305	FRANCE	ASGROW
24	9CSAS	FRANCE	Caussade Semences	72	RHA274	USA	USDA
25	H049 + FSB	FRANCE	–	73	H158A/H543R-B	FRANCE	ASGROW
26	SSD-580	FRANCE	ASGROW	74	H100A/RHA274	FRANCE	ASGROW
27	5AS-F1/A2 × R2	FRANCE	ASGROW	75	H209A/H566R	FRANCE	ASGROW
28	7CR16 = PRH6	FRANCE	C.F	76	AS0-1-POP-A	FRANCE	ENSAT
29	ENSAT-699	FRANCE	ENSAT	77	AS6305	FRANCE	ENSAT
30	SSD-581	FRANCE	ASGROW	78	BF1 POPB	FRANCE	NOVARTIS
31	TMB-51	FRANCE	INRAMONT	79	D34	FRANCE	INRAMONT
32	11 × 12	IRAN	SPII	80	CAY	FRANCE	ENSAT
33	110	IRAN	SPII	81	346	IRAN	SPII
34	H603R	FRANCE	INRAMONT	82	NSF1 A5 × R5	FRANCE	NOVARTIS
35	4	IRAN	SPII	83	36	IRAN	SPII
36	703-Chlorina	FRANCE	ENSAT	85	SDB2	FRANCE	INRAMONT
37	NSF1 A4 × R5	FRANCE	NOVARTIS	86	H158A × LC1064C	FRANCE	ASGROW
38	28	IRAN	SPII	87	H156A/H543R	FRANCE	ASGROW
39	30	IRAN	SPII	90	15,038	FRANCE	ASGROW
40	F1250/03	HUNGARY	–	91	SF076	FRANCE	ENSAT
41	SDR18	USA	USDA	92	8A*LC1064C*	–	–
42	LP-CSYB	FRANCE	ENSAT	93	SF085	FRANCE	ENSAT
43	803–1	SERBIA	IFVC	94	SF092	FRANCE	ENSAT
44	1,009,370–3(100 K)	FRANCE	ENSAT	95	HC91	IRAN	SPII
45	CSWW2X	FRANCE	Caussade Semences	96	10–59	IRAN	SPII
46	1,009,370–1(100 K)	FRANCE	ENSAT	97	H100A/90R78	FRANCE	ASGROW
47	H158A/H543R	FRANCE	ASGROW	98	SF109	FRANCE	ENSAT
48	H100A	FRANCE	ASGROW	99	SF105	FRANCE	ENSAT
49	15,031	FRANCE	ASGROW	100	SF-023	FRANCE	ENSAT

Table 1. Name and origin of sunflower lines. Original germplasm codes (1–100) are retained. Six lines were excluded during quality filtering due to high missing data, resulting in 94 lines used in the final analysis.

Analysis of genetic diversity

Analysis of molecular variance (AMOVA) and genetic diversity parameters, including the number of alleles (Na), effective number of alleles (Ne), Shannon's index (I), observed heterozygosity (Ho), expected heterozygosity (He), and the percentage of polymorphic loci (PPL%), were calculated for the studied groups using the GenAlex version 6.5 software (https://biology-assets.anu.edu.au/GenAlEx)[32].Genetic diversity indices, including expected heterozygosity (Hs), observed heterozygosity (Ho), corrected heterozygosity for genetic diversity (Ht), average

heterozygosity within groups (Htp), genetic differentiation index (Dst), intergroup genetic differentiation coefficient (Fst), and inbreeding coefficient (Fis) were calculated using adegenet package in R version 4.4.1 software.

Identification of KEGG pathways

Fst is a measure of genetic difference between subpopulations. High Fst values (> 0.25) indicate that at some loci (gene regions) natural or artificial selection pressure has caused genetic differences in different populations[33]. In this study, the sequences of SNP loci with significant genetic divergence (Fst > 0.35) were aligned to the Helianthus annuus reference genome using the EnsemblPlants database (https://plants.ensembl.org/Multi/Tools/Blast). Genes with the lowest expected value (E-val) and the highest identity percentage were selected for gene ontology enrichment analysis through the DAVID database (https://davidbioinformatics.nih.gov/tools.jsp). Furthermore, the DAVID database was used to identify the associated KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways (www.kegg.jp/kegg/kegg1.html)[34,35].

Results

Evaluation SNP data

The data on SNP markers are summarized in Table 2. A total of 7,909 SNPs were identified across the sunflower genome using SNP calling (Table 2, Fig. 1). These SNPs were physically mapped to 17 chromosomes. The highest number of SNPs was detected on chromosome 9 (622 SNPs), while the lowest was found in unmapped or unknown genomic regions (5 SNPs). The marker density varied across the sunflower genome, ranging from 0 to over 50 SNPs/Mb (Fig. 1). In general, SNPs were more abundant in the telomeric regions of the chromosome arms compared to the pericentromeric regions (exact centromere location unknown). Of the identified SNPs, 6,245 were transition types (A ↔ G and T ↔ C), and 1,664 were transversion types (A ↔ C, A ↔ T, T ↔ G, and C ↔ G) (Table 2). The ratio of transitions to transversions (Ts/Tv) ranged from 1.50 to 5.61 across chromosomes, with an average value of 3.75. This indicates that transitions were approximately 3.75 times more frequent than transversions.

The highest Ts/Tv ratio was observed on chromosome 7 (5.61), while the lowest was found on unmapped regions (1.50). The SNP density plot (Fig. 1) reveals a non-uniform distribution of SNPs across the genome, with certain chromosomes (e.g., chromosomes 6, 7, 8, 9, 10 and 17) showing higher SNP density on one arm than the other. MAF values showed moderate variation across chromosomes, ranging from 0.2586 (in chromosome 14) to 0.3350 (in chromosome 2), with a genome-wide average of 0.2850. PIC values also varied among chromosomes, with the lowest value of 0.2769 observed on chromosome 14 and the highest value of 0.3234 on chromosome 2, resulting in an overall mean of 0.2948. The value of r^2, indicating linkage disequilibrium (LD) among SNP pairs, was used to calculate the relationships between SNP markers. A total of 358,995 marker pairs with an average $r^2 = 0.158$ were identified across the whole genome, of which 103,265 pairs exhibited significant linkage at $p < 0.001$. The r^2 values ranged from 0.1157 on chromosome 7 to 0.2387 on chromosome 2, suggesting stronger LD on chromosome 2. Additionally, Dist, representing the genetic distance between SNP pairs, varied from 6,433,051 base pairs on chromosome 8 to 12,983,604 base pairs on chromosome 4. The number of significant SNP pairs (NSSP) ranged from 2,220 on chromosome 7 to 9,305 on chromosome 10. These results indicate broad marker coverage, moderate linkage disequilibrium, and high marker resolution in this sunflower population (Table 3).

The r^2 values were plotted against the genetic distance across the genome to visualize the LD decay. A threshold of $r^2 = 0.1$ was applied to define the cut-off for LD decay. The analysis showed that LD in the sunflower genome begins to decay at approximately 1.2 million base pairs (bp). Within this distance, the genetic relationship between between pairs of markers decreases significantly, as illustrated in Fig. 2.

Findings on population structure

Based on structural analysis using SNP markers and Bayesian model in STRUCTURE software, two genetic subpopulations (K = 2) were identified among 94 sunflower inbred lines (Fig. 3). The proportions of genotype membership were 75.53% for Subgroup 1 (red) and 6.38% for Subgroup 2 (green), while 18.09% of the inbred

Chromosome	Chr1	Chr2	Chr3	Chr4	Chr5	Chr6	Chr7	Chr8	Chr9	Chr10	Chr11	Chr12	Chr13	Chr14	Chr15	Chr16	Chr17	ChrUn	Total
Total SNPs	461	344	490	434	548	379	251	571	622	531	402	441	427	469	383	581	570	5	7909
A/G	179	124	196	158	217	143	95	211	241	194	141	164	167	181	137	203	210	2	2963
T/C	179	139	205	179	230	151	118	233	253	225	183	190	166	179	166	256	229	1	3282
Total transitions	358	263	401	337	447	294	213	444	494	419	324	354	333	360	303	459	439	3	6245
Ts (%)	77.66	76.45	81.84	77.65	81.57	77.57	84.86	77.76	79.42	78.91	80.60	80.27	77.99	76.76	79.11	79.00	77.02	60.00	78.96
A/C	55	39	47	53	48	41	15	64	57	60	32	48	37	48	36	57	74	1	812
T/G	48	42	42	44	53	44	23	63	71	52	46	39	57	61	44	65	57	1	852
Total transversions	103	81	89	97	101	85	38	127	128	112	78	87	94	109	80	122	131	2	1664
Tv (%)	22.34	23.55	18.16	22.35	18.43	22.43	15.14	22.24	20.58	21.09	19.40	19.73	22.01	23.24	20.89	21.00	22.98	40.00	21.04
Ts/Tv ration	3.48	3.25	4.51	3.47	4.43	3.46	5.61	3.50	3.86	3.74	4.15	4.07	3.54	3.30	3.79	3.76	3.35	1.50	3.75

Table 2. Summary of SNP data in the studied sunflower panel. Chr Un: Represents unknown or unassigned genomic scaffolds that have not yet been placed onto any of the 17 assembled sunflower chromosomes.

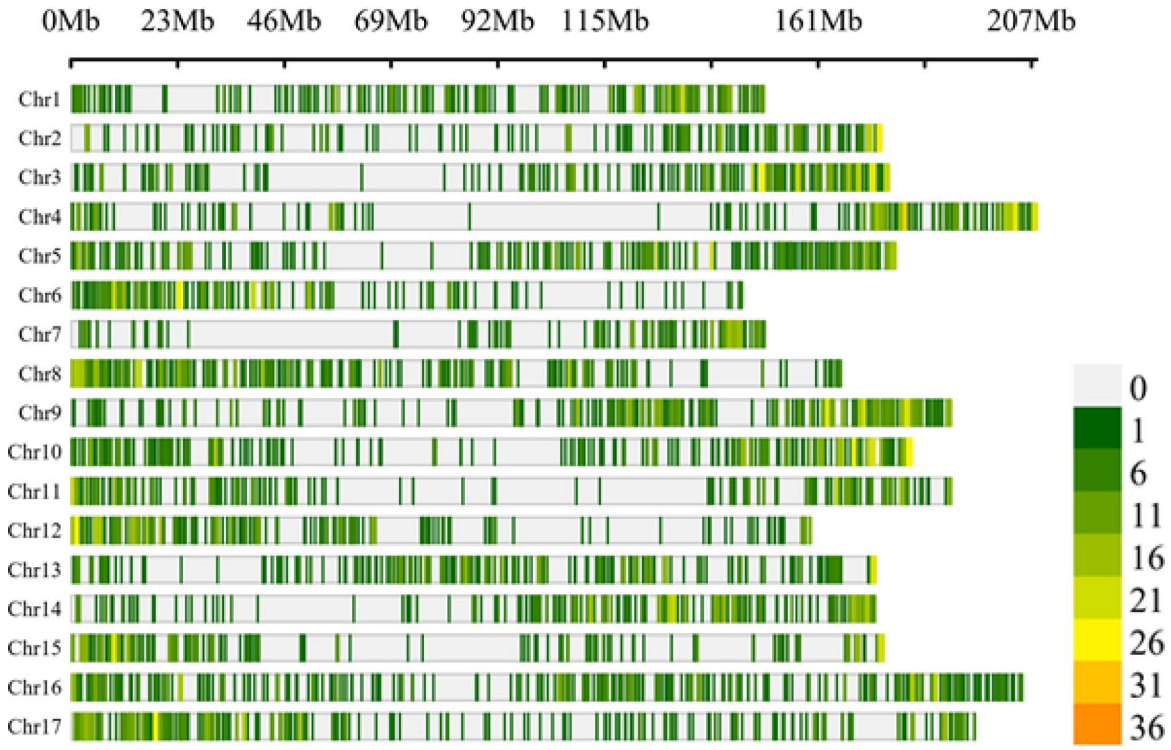

Fig. 1. Genomic distribution of 7,909 SNPs across 94 diverse sunflower lines and the corresponding SNP density, plotted using the CMplot online tool at Bioinformatics.com.cn (https://www.bioinformatics.com.cn/plot_basic_SNP_density_by_CMplot_107_en).

Chromosome	MAF	PIC	TNSP	r^2	Dist (bp)	NSSP
Chr1	0.2632	0.2817	20,460	0.1169	8,380,782	4083
Chr2	0.3350	0.3234	15,553	0.2387	12,926,698	6550
Chr3	0.2868	0.2949	22,108	0.1479	9,078,737	5984
Chr4	0.2972	0.2974	19,229	0.1374	12,983,604	4714
Chr5	0.3109	0.3070	25,253	0.1802	8,407,963	8611
Chr6	0.2652	0.2823	16,505	0.1382	8,034,330	3709
Chr7	0.3108	0.3048	10,681	0.1157	14,890,093	2220
Chr8	0.2695	0.2822	25,421	0.1766	6,433,051	8546
Chr9	0.2821	0.2959	29,896	0.1486	7,448,223	7980
Chr10	0.2774	0.2986	24,718	0.1927	8,257,981	9305
Chr11	0.2985	0.3021	19,074	0.1167	12,041,873	4000
Chr12	0.2875	0.3013	21,168	0.1386	9,297,652	5057
Chr13	0.2692	0.2837	19,017	0.1669	10,180,348	6095
Chr14	0.2586	0.2769	19,815	0.1640	8,924,969	5809
Chr15	0.2781	0.2960	17,978	0.1603	11,630,821	5104
Chr16	0.2876	0.2956	25,657	0.1938	8,427,885	9149
Chr17	0.2889	0.2976	26,447	0.1336	8,819,068	6349
ChrUn	0.2912	0.3076	15	0.0229	–	–
Whole genome	0.2850	0.2948	358,995	0.1581	9,394,877	103,265

Table 3. Polymorphism information content (PIC) and minor-allele frequency (MAF) calculated for each chromosome by using SNPs and a summary of observed linkage disequilibrium (r^2) among SNP pairs. *TNSP* total number of SNP pairs, *NSSP* Number of significant SNP pairs ($p < 0.001$), *Dist* genetic distance between SNP pairs.

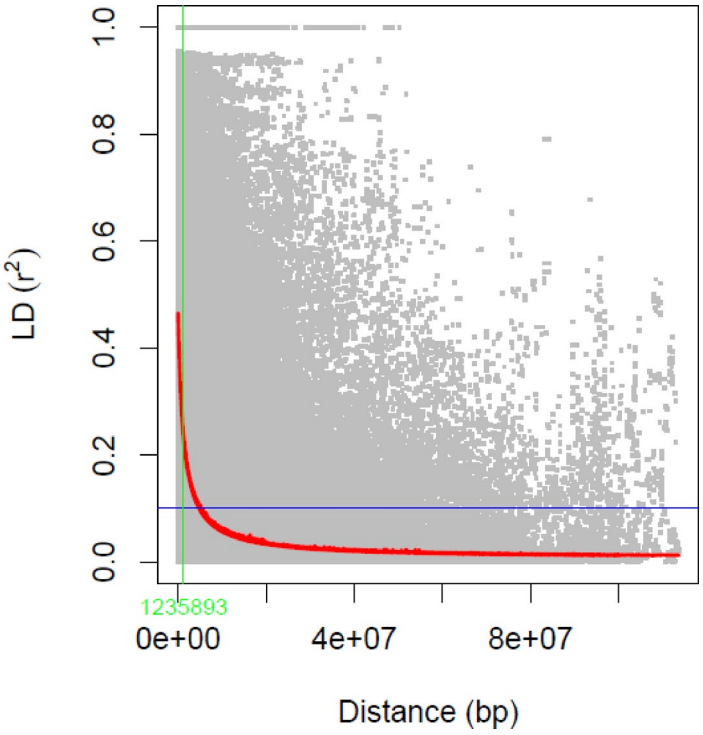

Fig. 2. Linkage disequilibrium (r^2) vs. physical distance (bp) for the sunflower genome. A cut-off line was plotted at $r^2 = 0.1$.

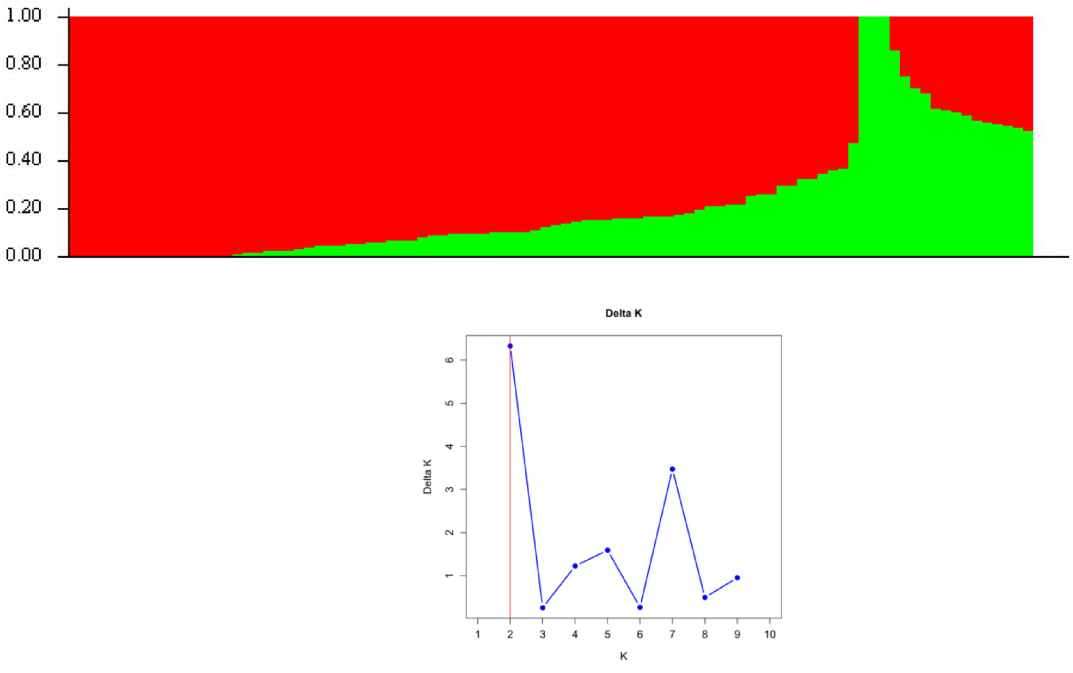

Fig. 3. A panel of 94 sunflower inbred lines genomic structure declared with SNP markers in Structure software applying Bayesian method. At K = 2, the highest ΔK value was observed.

lines exhibited admixed genetic backgrounds. The WPGMA dendrogram based on shared-allele genetic distances divided the sunflower lines into two main clusters The first and second cluster consisted of cluster A (green) and cluster B (red) sunflower lines, respectively. Most lines in the green cluster, including lines 1, 10, 15, 20, 26, 46, and 74, originated from French breeding programs, suggesting a common genetic background. In contrast, the red cluster contained several lines from U.S. (USDA) breeding programs as well as some Iranian

genotypes. The WPGMA clustering results were consistent with the STRUCTURE analysis, except for genotypes 49, 40, and 43, which were classified into different subgroups (Fig. 4).

The PCoA plot also supported the observed population structure (Fig. 5). Subgroup 1 (represented in red), showed a wider spread in the plot, indicating higher genetic diversity, whereas Subgroup 2 (represented in green) appeared more tightly clustered, reflecting closer genetic similarity among its members. The first two principal coordinates explained 9.64% and 7.22% of the total genetic variation, respectively. This result is expected in a panel of highly inbred lines genotyped with 7,909 SNP markers. The high level of inbreeding leads to reduced genetic diversity, resulting in a lower total variation being captured by the Principal Coordinates.

Genetic diversity

The analysis of molecular variance (AMOVA) showed that 16% of the total genetic variation was among populations, which was statistically significant (Fst = 0.156, P = 0.001). Additionally, 61% of the variation was attributed to differences among individuals within populations, and 24% was due to variation within individuals (Table 4). The analysis of genetic diversity between the two subgroups identified by structure analysis showed that Subgroup 1 exhibited higher genetic diversity with Na = 1.988, Ne = 1.647, and PPL = 98.81%, whereas Subgroup 2 showed lower genetic diversity with Na = 1.871, Ne = 1.360, and PPL = 87.06%. The overall mean for the number of observed alleles (Na = 1.929) and the number of effective alleles (Ne = 1.504) indicated a moderate level of genetic diversity in the population. Additionally, the high percentage of polymorphic loci (PPL = 92.94%) reflects a high potential for genetic improvement in this population (Table 5).

The genetic diversity parameters of the studied SNPs, summarized for each chromosome, are presented in Table 6. The level of variation in Ho, Hs, Ht, Fst, Fis, and Dest differed among the sunflower chromosomes. The highest values for Ho and Hs were observed for chromosomes 17 and 12 (0.10721 and 0.36733, respectively), which may be attributed to the presence of genes involved in maintaining genetic diversity in these regions. The Ht index was relatively uniform among chromosomes, with the lowest value (0.33085) recorded for chromosomes 14 and 13. The Fst values ranged from 0.03394 (chromosome 12) to 0.16564 (chromosome 4), indicating moderate to high genetic divergence among the studied genotypes across different chromosomes. Additionally, the Fis values, representing the level of inbreeding within populations, ranged from 0.61864 (chromosome 4) to 0.75483 (chromosome 12). These high Fis values indicate a strong excess of homozygosity, which is fully

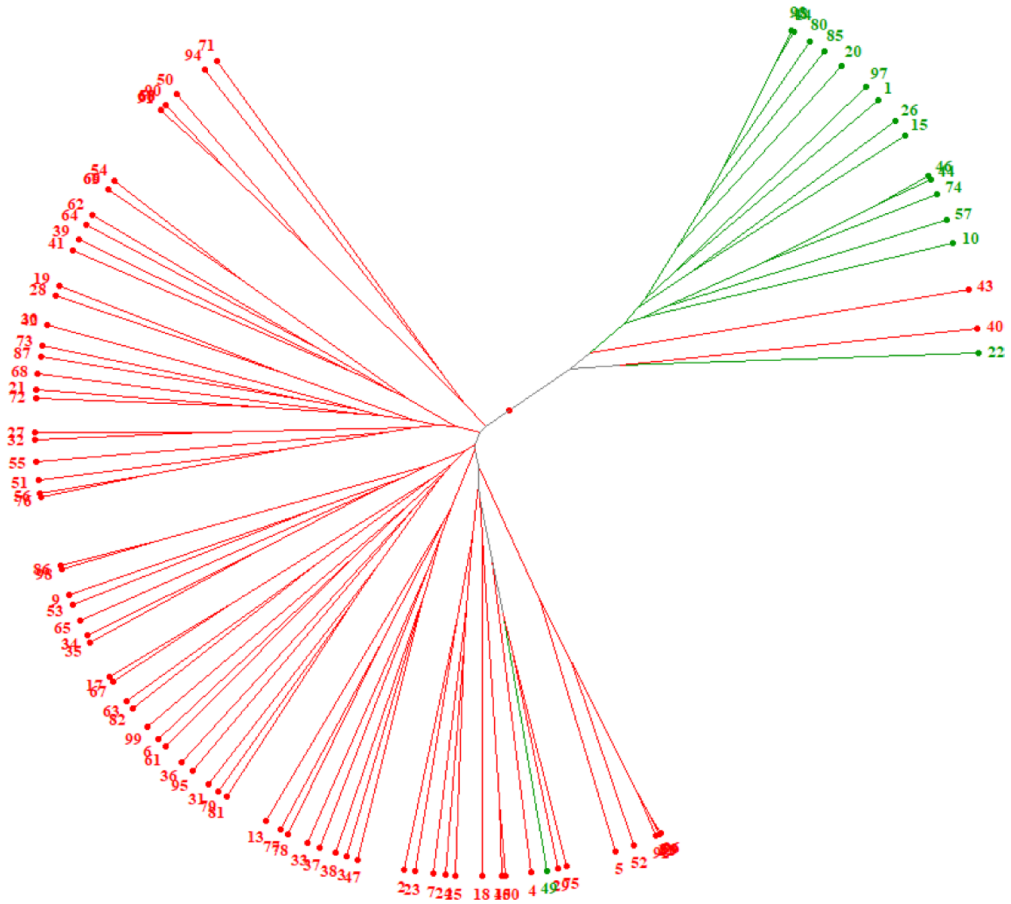

Fig. 4. Complete linkage clustering dendrogram generated using 7,909 SNPs and 94 sunflower inbred lines. Colors reflect groupings derived from structure analysis. The dendrogram was constructed using DARwin v6.0.10 (https://darwin.cirad.fr).

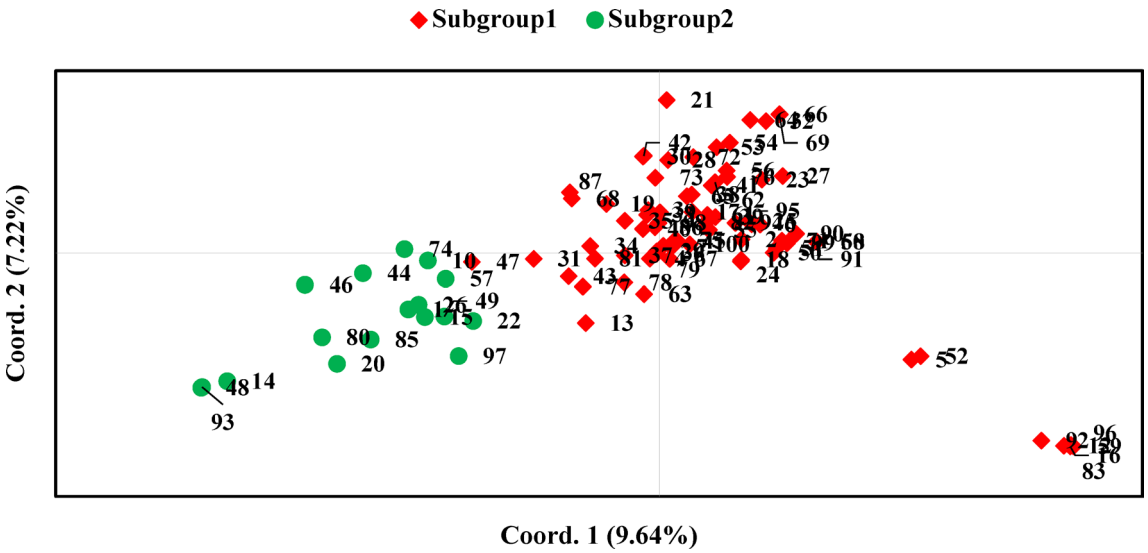

Fig. 5. Principal coordinate analysis (PCoA) of 94 sunflower inbred lines based on 7909 SNP markers. Colors reflect groupings derived from structure analysis.

Source of variation	df	MS	Estimated variance	%	F statistics (Fst)	Probability
Among pops	1	18,976.569	292.076	16	0.156	0.001
Among individuals	92	2707.319	1132.088	61		
Within individuals	94	443.144	443.144	24		
Total	187		1867.307	100		

Table 4. Analysis of molecular variance (AMOVA) result from the diversity panel of 94 sunflower inbred lines.

Subgroups	Na	Ne	I	Ho	He	uHe	PPL (%)
Subgroup1	1.988 + 0.001	1.647 + 0.003	0.547 + 0.002	0.125 + 0.001	0.371 + 0.001	0.373 + 0.001	98.81
Subgroup2	1.871 + 0.004	1.36 + 0.003	0.372 + 0.002	0.061 + 0.001	0.234 + 0.002	0.241 + 0.002	87.06
Average	1.929 + 0.002	1.504 + 0.003	0.459 + 0.002	0.093 + 0.001	0.302 + 0.001	0.307 + 0.001	92.94

Table 5. Genetic variation among two subgroups identified by structure analysis on a diversity panel of 94 sunflower inbred lines. Number of different alleles (Na); number of effective alleles (Ne); Shannon's information index (I); observed heterozygosity (Ho); expected heterozygosity (He); unbiased expected heterozygosity (uHe); percentage of polymorphic loci (PPL).

expected given that the materials used in this study are inbred lines. The Dest values, reflecting the true genetic differentiation among populations, ranged from 0.04364 (chromosome 12) to 0.20642 (chromosome 4). The high Dest value for chromosome 4 indicates that this chromosome may contain genomic regions with high genetic differentiation among populations.

Functional analysis of selected genomic regions

The process of selection can lead to changes in allele frequency within a population, resulting in genetic differentiation between populations in genomic regions under selection. Therefore, significant differences in allele frequency between populations can be considered as evidence of positive selection at specific genetic loci. Following selection, in addition to the allele associated with the target gene, neutral loci adjacent to the target allele are also retained in the genome, leading to an increase in the frequency of specific haplotypes in the selected region and a consequent reduction in haplotype diversity in that region. This shift in allele frequency and the frequency of specific haplotypes within a genomic region indicates the effect of natural selection[36]. To identify the regions under selection, Fst values between different subgroups were calculated (Fig. 6). In this study, we identified 285 genomic regions showing clear signatures of selection across the sunflower germplasm. These regions are associated with a total of 283 candidate genes (Table S1). Functional annotation of these genes using DAVID revealed three KEGG pathways, among which proteasome (P = 0.015) (Fig. 7) and pyruvate metabolism[34,35] (P = 0.020) (Fig. 8) were statistically significant based on p-values (< 0.05), suggesting their potential biological relevance in the context of selection.

Chromosome	Ho	Hs	Ht	Dst	Htp	Dstp	Fst	Fstp	Fis	Dest
Chr1	0.09364	0.30562	0.33829	0.03266	0.37095	0.06533	0.07778	0.13184	0.68661	0.10280
Chr2	0.09435	0.33770	0.38720	0.04950	0.43669	0.09899	0.11422	0.19006	0.71939	0.15283
Chr3	0.09366	0.29425	0.35001	0.05576	0.40576	0.11152	0.13373	0.21520	0.66865	0.16522
Chr4	0.10454	0.27457	0.34659	0.07201	0.41861	0.14403	0.16564	0.25614	0.61864	0.20642
Chr5	0.09872	0.33435	0.37598	0.04163	0.41762	0.08327	0.09385	0.16127	0.70999	0.13550
Chr6	0.10536	0.30704	0.33944	0.03240	0.37184	0.06480	0.07713	0.13211	0.65212	0.10277
Chr7	0.08829	0.32020	0.36356	0.04336	0.40692	0.08672	0.10045	0.17031	0.71156	0.13716
Chr8	0.09471	0.31200	0.33388	0.02187	0.35575	0.04375	0.05065	0.08735	0.69747	0.06996
Chr9	0.08002	0.29105	0.34142	0.05037	0.39178	0.10073	0.11904	0.19148	0.72459	0.14906
Chr10	0.09260	0.33310	0.36122	0.02812	0.38934	0.05624	0.06711	0.11643	0.71053	0.08975
Chr11	0.10275	0.33665	0.36701	0.03035	0.39736	0.06071	0.07086	0.12443	0.68710	0.09939
Chr12	0.09082	0.36733	0.38113	0.01380	0.39493	0.02760	0.03394	0.05846	0.75483	0.04364
Chr13	0.08884	0.30487	0.33085	0.02598	0.35683	0.05196	0.06474	0.11273	0.71125	0.08243
Chr14	0.08195	0.31397	0.33085	0.01688	0.34773	0.03377	0.04279	0.07418	0.72658	0.05352
Chr15	0.08252	0.32443	0.35264	0.02821	0.38085	0.05643	0.06642	0.11391	0.73957	0.08931
Chr16	0.08961	0.28390	0.35310	0.06920	0.42229	0.13839	0.15855	0.24929	0.67855	0.20428
Chr17	0.10721	0.29720	0.33706	0.03986	0.37692	0.07972	0.09904	0.16830	0.63438	0.12232
ChrUn	0.07756	0.33294	0.38670	0.05372	0.44044	0.10748	0.12282	0.20312	0.75820	0.16604
Whole Genome	0.09344	0.31249	0.35126	0.03877	0.39004	0.07755	0.09138	0.15157	0.69512	0.11902

Table 6. Diversity indices of 94 sunflower inbred lines using 7909 SNPs. Observed heterozygosity (Ho), expected heterozygosity (Hs), total heterozygosity (Ht), corrected Ht (HTp), gene diversity among samples (Dst), corrected Dst (Dstp), genetic differentiation (Fst), corrected Fst (Fstp), inbreeding coefficient (Fis), population differentiation (Dest).

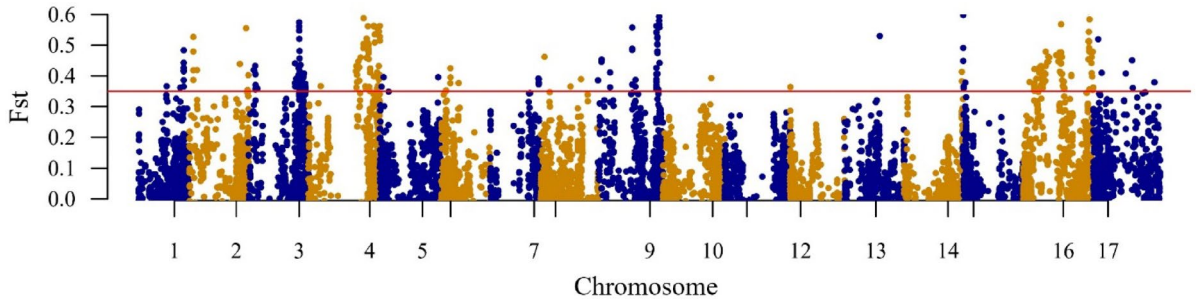

Fig. 6. Manhattan plot of distribution of genetic differentiation (Fst) values between identified subgroups.

Discussion

Sunflower is an agronomically important oilseed crop and the application of molecular markers will improve the efficiency and accuracy of selection in sunflower breeding[20]. Among DNA markers, SNP are widely distributed across the genome and can be found in coding, non-coding, and intronic regions of genes[37]. They are effective tools for assessing genetic variation and improving breeding efficiency through marker-assisted selection (MAS) and genomic selection[18]. A few GBS approaches were reported recently for sunflower[20,22,38,39]. The use of GBS-SNP markers provides high-resolution insights into population structure and genetic diversity, enhancing breeding strategies and accelerating the development of improved sunflower cultivars. Expanding genetic diversity studies using SNP platforms can significantly improve sunflower breeding outcomes[40].

The results of this study provide valuable insights into the genetic diversity and population structure of sunflower germplasm based on SNP markers. The identification of 7909 SNPs across the sunflower genome demonstrates the resolution and depth of the SNP calling method used in this study. The highest number of SNPs on chromosome 9 (622 SNPs) and the lowest in unmapped regions (5 SNPs) indicate differences in marker density and recombination rate across the genome. The higher abundance of SNPs in telomeric regions compared to pericentromeric regions is consistent with previous findings in other plant species[19,41,42], which have attributed this pattern to increased recombination rate and gene density in telomeric regions.

The prevalence of transition mutations over transversion mutations, with a Ts/Tv ratio of 3.75, reflects the expected bias toward transitions in plant genomes and suggesting that the observed SNPs were not a product of sequencing and mapping errors[43]. Moreover, the Ts/Tv ratio observed in our study (3.75) was considerably higher compared to other plant species such as wheat (1.75;[18]), maize (2.05;[19]), and tall fescue aligned to wheat

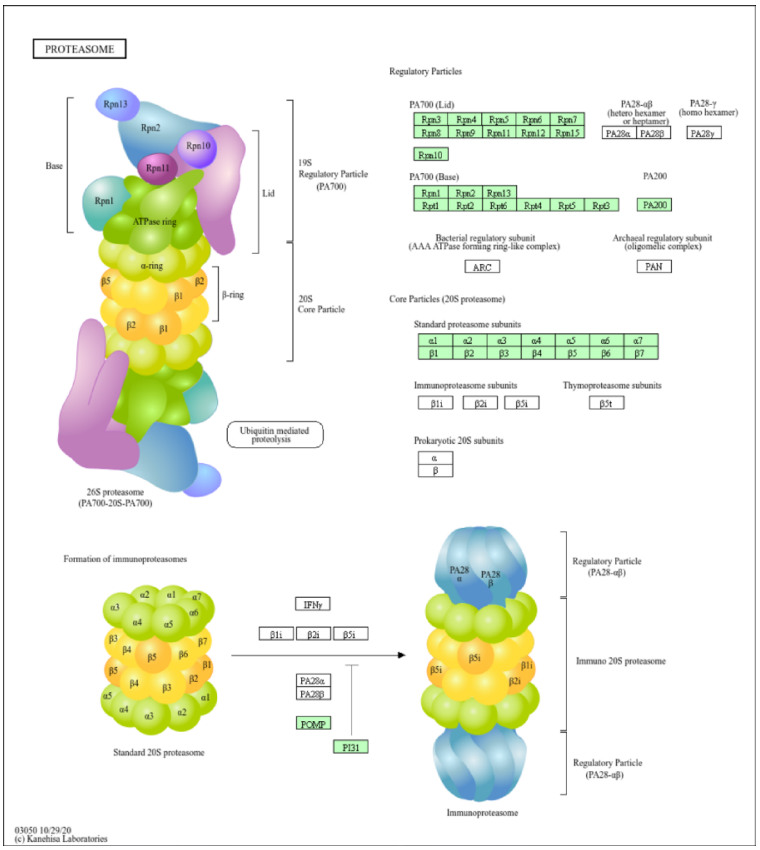

Fig. 7. KEGG pathway of proteasome.

chromosomes (1.73;[44]). Notably, the high Ts/Tv ratio on chromosome 7 (5.61) indicates increased stability or selection pressure in this region. The PIC values in the present study ranged from 0.2769 to 0.3234, with a genome-wide average of 0.2948, indicating a moderate level of polymorphism across the sunflower genome. These values are in line with previous sunflower studies. For example, Filippi et al.[45] reported an estimated PIC value of 0.232 using SNP markers. Similarly, Zeinalzadeh-Tabrizi et al.[46] observed similar levels of diversity (0.10 to 0.58) using SSR markers. This consistency supports the reliability of the genetic diversity detected in the current germplasm panel. The MAF values observed in the present study (0.2850) are consistent with those reported by, Filippi et al.[45], who showed MAF value greater than 0.1. MAF is a key indicator for identifying high-quality markers, with values greater than 0.5% indicating high polymorphism in the genome[47]. Higher MAF values enhance genetic variation analysis by increasing the detection of minor alleles and genetic diversity[48]. This emphasizes the importance of considering MAF in genome-wide mapping studies. It is worth mentioning that, the highest MAF (0.3350) and PIC (0.3234) on chromosome 2 suggest the presence of loci with high polymorphism and gene diversity, which may be useful for future breeding efforts.

In this study, the population structure analysis revealed two distinct subgroups (K = 2). Similarly, Rasoulzadeh Aghdam et al.[49] identified two possible subpopulations (K = 2) in the studied population panel. In contrast, Mandel et al. (2013) using a 10 K SNP chip on the USDA-UBC accessions indicated that K = 3, supporting the existence of three genetically distinct clusters. The clustering pattern observed in both PCoA and WPGMA trees supports the presence of genetic differentiation between these subgroups, which may be attributed to differences in breeding origin and selection history, reinforced by inbreeding and limited gene flow among sunflower breeding programs. In several studies, PCoA has been used to assess population structure and confirm genetic structure in sunflower[7,12,23,46]. This method can usually be used in conjunction with other population structure analysis methods to better understand and examine the results.

Although, the use of SNP markers is common in genetic research, studies that specifically investigate the genetic diversity of sunflower using these markers are still limited. Filippi et al.[24] analyzed 11,834 SNP markers using AMOVA and found that 4.58% of the total genetic variance was due to differences among breeding collections, indicating a significant but relatively low level of population differentiation. In contrast, Kholghi et al.[14], using SSR markers, reported that 86% of the total genetic variation was attributed to within-population diversity among Iranian confectionery sunflower landraces, reflecting an open and heterogeneous genetic structure. Similarly, Basirnia et al.[8], employing REMAP and IRAP markers, found that 94% of the genetic variation existed within groups, with only 6% attributed to between-group differences, further supporting the high intra-population diversity in sunflower germplasm. Additionally, in a study by Sahari et al.[50] using 30

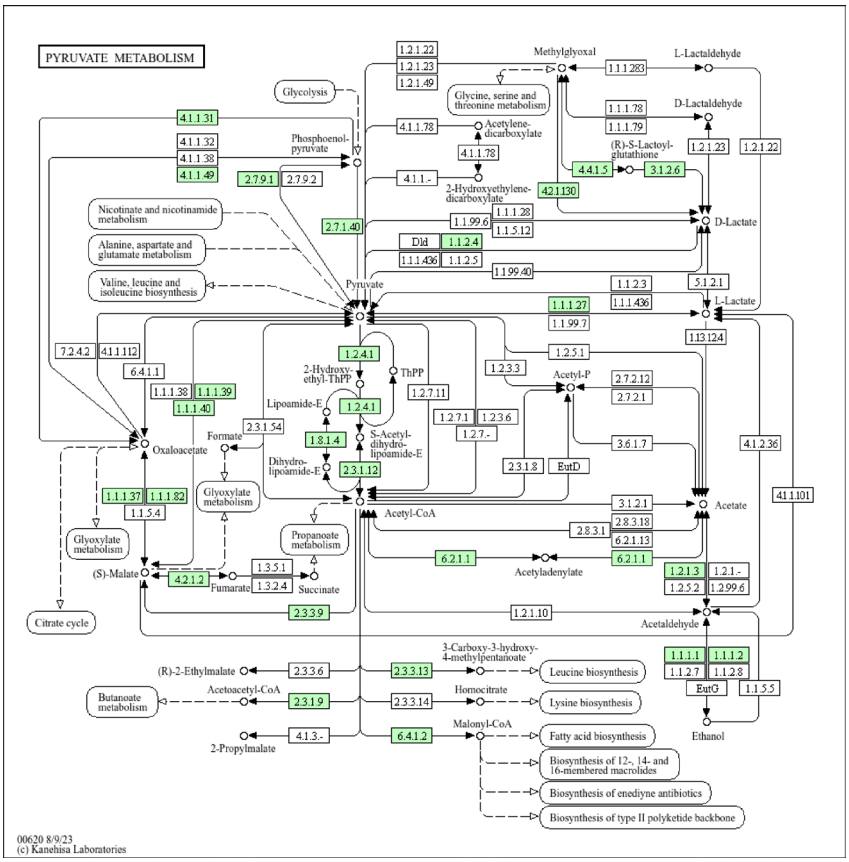

Fig. 8. KEGG pathway of pyruvate metabolism.

SSR markers and sequencing of 7 genes, a core collection of 8 populations was identified, capturing 88% of the genetic diversity present in Tunisian sunflower landraces.

In the current investigation, AMOVA was employed to evaluate genetic differentiation among groups. The analysis revealed statistically significant differences between the examined groups ($p < 0.001$), providing strong evidence for genetic divergence among the studied populations. Specifically, 16% of the total genetic variation was attributed to differences among populations, with an Fst value of 0.156, indicating a moderate to high level of genetic differentiation. while the remaining 84% was distributed within and among individuals, which may have resulted from selective breeding and environmental adaptation. Although high within-group diversity was observed, it does not rule out the existence of significant genetic structure across populations, as evidenced by AMOVA and Fst results. This pattern is further supported by the STRUCTURE and PCoA results, which are consistent with the differentiation identified by AMOVA.

An analysis of genetic diversity parameters in the sunflower panel studied showed that observed heterozygosity (Ho), expected heterozygosity (Hs), and fixation index (Fst) differed in different chromosomes, indicating a non-uniform distribution of genetic diversity across the genome. Notably, the Fst values ranged from 0.03394 to 0.16564, highlighting moderate to high genetic differentiation among genotypes depending on chromosomal regions. The Fst values reported in the present study were generally higher than those observed by Mandel et al.[25,51], who estimated Fst values of approximately 0.049 (SNP markers) and 0.047 (SSR markers) in panel of sunflower lines. This difference may be due to the broader genomic coverage and greater genetic diversity present in the population analyzed here, as well as to differences in marker density, population structure, and germplasm composition used in the two studies. Also, the Ho value observed in the present study (0.093) was higher than those reported by Filippi et al.[24], where observed heterozygosity ranged from 0.007 to 0.030 across different breeding populations. In contrast, the Hs values obtained here (0.312) are lower than those reported by Filippi et al.[24] using a 11 K SNP markers on three breeding populations (0.454). The high Fis value observed across the whole genome (Fis = 0.695) indicates a substantial deficit of heterozygotes relative to Hardy–Weinberg expectations. This pattern is expected in sunflower inbred lines, which undergo multiple generations of selfing to produce genetically uniform and homozygous materials. Therefore, the elevated Fis reflects the intentional inbreeding used in line development instead of biological inbreeding, limited gene flow, or population structure typically discussed in natural populations.

Using the top 0.35 Fst values as a threshold[19,52,53], multiple genomic regions showing signs of selection were identified across the studied subpopulations. One of the most important advantages of identifying selection signs is that it can be done solely by relying on molecular data, even when phenotypic information is not available, to identify genes involved in the differentiation process[19]. This represents the first known application of the

Fst method for selection signature analysis in sunflower, specifically focusing on identifying genetic regions and KEGG pathways involved in population differentiation. Functional analysis of the identified candidate genes revealed that the genes under selection were mainly involved in the regulation of cell growth, protein degradation via the proteasome pathway, and energy production and management via pyruvate metabolism.

In sunflower, central metabolic pathways such as pyruvate metabolism are significantly enriched during seed development and oil accumulation, indicating their key role in energy processes and carbon flux underlying phenotypic diversity[54]. Similarly, components of the proteolytic system, particularly proteasome activity, are involved in the response to oxidative stress in sunflower leaves, suggesting their relevance to stress adaptation[55].

Although genes involved in the proteasome and pyruvate metabolism pathways have been previously identified and functionally characterized in sunflower—mainly in the context of growth and stress response—the results of the present study indicate that these pathways are also targets of selection and have contributed to genetic differentiation among sunflower inbred lines. Interestingly, similar functional categories were also reported in maize by Arzhang et al.[19], suggesting that conserved biological processes may underlie genetic differentiation across different crop species.

The present study provides a detailed analysis of genetic diversity in 94 sunflower inbred lines. Although increasing the number of samples in future studies could enhance the accuracy of the results, the current data still provide a good representation of the genetic diversity of sunflower inbred lines. Additionally, the use of a 10 K SNP array in this study has offered a sufficient coverage of the sunflower genome and has successfully identified selected genomic regions. However, the use of arrays with a higher number of markers could lead to more accurate assessments of the entire genome.

Conclusions

Genetic diversity is considered the main basis for selection and breeding in plant breeding programs. In this study, 94 sunflower inbred lines were genotyped using a 10 K SNP marker array to perform SNP calling and assess genetic diversity. The findings indicated the existence of significant genetic diversity among population (Fst = 0.156), as well as a heterogeneous distribution of genetic diversity across the genome. The use of Fst statistic and a threshold of 0.35, several genomic regions under selection were identified. Moreover, functional analysis of the candidate genes revealed that these genes are involved in critical biological processes such as cell growth regulation, degradation via the proteasome pathway, and pyruvate metabolism. These results provide valuable insights into the application of SNP markers for identifying genetic diversity and the underlying biological pathways of differentiation in sunflower. Additionally, they could be useful for future sunflower studies, which supported by marker-assisted selection and genomic selection.

Data availability

Data have been deposited in the European Variation Archive (EVA) under accession number PRJEB107724. The data is publicly available and can be accessed at: https://www.ebi.ac.uk/eva/?eva-study=PRJEB107724.

Received: 21 October 2025; Accepted: 12 February 2026
Published online: 17 February 2026

References

1. Adeleke, B. S. & Babalola, O. O. Oilseed crop sunflower (*Helianthus annuus*) as a source of food: Nutritional and health benefits. *Food Sci. Nutr.* **8**, 4666–4684 (2020).
2. Lavudya, S. et al. Assessing population structure and morpho-molecular characterization of sunflower (*Helianthus annuus* L.) for elite germplasm identification. *PeerJ* **12**, e18205 (2024).
3. USDA Foreign Agricultural Service. *Oilseeds: World Markets and Trade. USDA*, (2024).
4. Fick, G. N., Zimmer, D. E. Parental lines for production of confectionery sunflower hybrids, pp. 15–16 (1974).
5. Salgotra, R. K. & Chauhan, B. S. Genetic diversity, conservation, and utilization of plant genetic resources. *Genes* **14**, 174 (2023).
6. Aziz, M. A. & Masmoudi, K. Molecular breakthroughs in modern plant breeding techniques. *Hortic. Plant J.* **11**, 15–41 (2025).
7. Darvishzadeh, R. et al. Molecular characterization and similarity relationships among sunflower (*Helianthus annuus* L.) inbred lines using some mapped simple sequence repeats. *Afr. J. Biotechnol.* **9**, 7280–7288 (2010).
8. Basirnia, A., Darvishzadeh, R. & Abdollahi Mandoulakani, B. Retrotransposon insertional polymorphism in sunflower (*Helianthus annuus* L.) lines revealed by IRAP and REMAP markers. *Plant Biosyst.* **150**, 641–652 (2016).
9. Bidyananda, N. et al. Plant genetic diversity studies: Insights from DNA marker analyses. *Int. J. Plant Biol.* **15**, 607–640 (2024).
10. Cheres, M. T. & Knapp, S. J. Ancestral origins and genetic diversity of cultivated sunflower: Coancestry analysis of public germplasm. *Crop Sci.* **38**, 1476–1482 (1998).
11. Hongtrakul, V., Huestis, G. M. & Knapp, S. J. Amplified fragment length polymorphisms as a tool for DNA fingerprinting sunflower germplasm: Genetic diversity among oilseed inbred lines. *Theor. Appl. Genet.* **95**, 400–407 (1997).
12. Dong, G., Liu, G. & Li, K. Studying genetic diversity in the core germplasm of confectionary sunflower (*Helianthus annuus* L.) in China based on AFLP and morphological analysis. *Russ. J. Genet.* **43**, 627–635 (2007).
13. Iqbal, M. A., Sadaqat, H. A. & Khan, I. A. Estimation of genetic diversity among sunflower genotypes through random amplified polymorphic DNA analysis. *Genet. Mol. Res.* **7**, 1408–1413 (2008).
14. Kholghi, M., Darvishzadeh, R., Bernousi, I., Pirzad, A. & Laurentin, H. Assessment of genomic diversity among and within Iranian confectionary sunflower (*Helianthus annuus* L.) populations by using simple sequence repeat markers. *Acta Agric. Scand. B Soil Plant Sci.* **62**, 488–498 (2012).
15. Kolkman, J. M. et al. Single nucleotide polymorphisms and linkage disequilibrium in sunflower. *Genetics* **177**, 457–468 (2007).
16. Bachlava, E. et al. SNP discovery and development of a high-density genotyping array for sunflower. *PLoS ONE* **7**, e29814 (2012).
17. Holtz, Y. et al. Genotyping by sequencing using program assessed with SSR and SNP markers. *Theor. Appl. Genet.* **120**, 1289–1299. https://doi.org/10.1007/s00122-009-1256-2 (2016).
18. Alipour, H. et al. Genotyping-by-sequencing (GBS) revealed molecular genetic diversity of Iranian wheat landraces and cultivars. *Front. Plant Sci.* **8**, 1293 (2017).
19. Arzhang S, Darvishzadeh R, Alipour H, Maleki HH, Dezhsetan S 2024 Genetic variability of maize (*Zea mays*) germplasm from Iran: genotyping with a maize 600K SNP array and genome-wide scanning for selection signatures. Crop Pasture Sci **75** 23288

20. Celik, I., Bodur, S., Frary, A. & Doganlar, S. Genome-wide SNP discovery and genetic linkage map construction in sunflower (*Helianthus annuus* L.) using a genotyping by sequencing (GBS) approach. *Mol. Breed.* **36**, 1–9 (2016).
21. Gubaev, R. et al. QTL mapping of oleic acid content in modern VNIIMK sunflower (*Helianthus annuus* L.) lines by using GBS-based SNP map. *PLoS ONE* **18**, e0288772 (2023).
22. Talukder, Z. I., Seiler, G. J., Song, Q., Ma, G. & Qi, L. SNP discovery and QTL mapping of Sclerotinia basal stalk rot resistance in sunflower using genotyping-by-sequencing. *Plant Genome* **9**, plantgenome2016-03 (2016).
23. Livaja, M. et al. Diversity analysis and genomic prediction of Sclerotinia resistance in sunflower using a new 25 K SNP genotyping array. *Theor. Appl. Genet.* **129**, 317–329 (2016).
24. Filippi, C. V. et al. Genetic diversity, population structure and linkage disequilibrium assessment among international sunflower breeding collections. *Genes* **11**, 283 (2020).
25. Mandel, J. R. et al. Association mapping and the genomic consequences of selection in sunflower. *PLoS Genet.* **9**, e1003378 (2013).
26. Dellaporta, S. L., Wood, J. & Hicks, J. B. A plant DNA minipreparation: Version II. *Plant Mol. Biol. Rep.* **1**, 19–21 (1983).
27. Bowers, J. E. et al. Development of an ultra-dense genetic map of the sunflower genome based on single-feature polymorphisms. *PLoS ONE* **7**, e51360 (2012).
28. Novembre, J., Pritchard, S., Stephens, M. & Donnelly, P. On population structure. *Genetics* **204**, 391–393. https://doi.org/10.1534/genetics.116.195164 (2016).
29. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).
30. Sokal, R. & Sneath, P. H. *Principles of numerical taxonomy* (WH Freeman and Company, 1963).
31. Perrier, X., Jacquemoud-Collet, J. *DARwin: dissimilarity analysis and representation for Windows.* Version 5, p. 158. CIRAD, Montpellier, France. Available at: http://darwin.cirad.fr/darwin (2006).
32. Peakall, R. O. D. & Smouse, P. E. GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* **6**, 288–295 (2006).
33. Cavanagh, C. R. et al. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 8057–8062 (2013).
34. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**(1), 27–30 (2000).
35. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. & Ishiguro-Watanabe, M. KEGG: Biological systems database as a model of the real world. *Nucleic Acids Res.* **53**, D672–D677 (2025).
36. Smith, J. M. & Haigh, J. The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23–35 (1974).
37. Deng, N., Zhou, H., Fan, H. & Yuan, Y. Single nucleotide polymorphisms and cancer susceptibility. *Oncotarget* **8**, 110635–110649. https://doi.org/10.18632/oncotarget.22372 (2017).
38. Mondon, A., Owens, G. L., Poverene, M., Cantamutto, M. & Rieseberg, L. H. Gene flow in Argentinian sunflowers as revealed by genotyping-by-sequencing data. *Evol. Appl.* **11**, 193–204 (2018).
39. Ma, X. F. et al. High resolution genetic mapping by genome sequencing reveals genome duplication and tetraploid genetic structure of the diploid *Miscanthus sinensis*. *PLoS ONE* **7**, e33821 (2012).
40. Dimitrijevic, A. & Horn, R. Sunflower hybrid breeding: From markers to genomic selection. *Front. Plant Sci.* **8**, 2238. https://doi.org/10.3389/fpls.2017.02238 (2018).
41. Yan, M. et al. Genotyping-by-sequencing application on diploid rose and a resulting high-density SNP-based consensus map. *Hortic. Res.* **5**, 17. https://doi.org/10.1038/s41438-018-0021-6 (2018).
42. Reddy, S. S. et al. Genome-wide association mapping of genomic regions associated with drought stress tolerance at seedling and reproductive stages in bread wheat. *Front. Plant Sci.* **14**, 1166439. https://doi.org/10.3389/fpls.2023.1166439 (2023).
43. Lorenc, M. T. et al. Discovery of single nucleotide polymorphisms in complex genomes using SGSautoSNP. *Biology* **1**, 370–382 (2012).
44. Shahabzadeh, Z., Darvishzadeh, R., Mohammadi, R., Jafari, M. & Alipour, H. High-throughput single nucleotide polymorphism genotyping reveals population structure and genetic diversity of tall fescue (*Festuca arundinacea*) populations. *Crop Pasture Sci.* **73**, 1070–1084 (2022).
45. Filippi, C. V. et al. Population structure and genetic diversity characterization of a sunflower association mapping population using SSR and SNP markers. *BMC Plant Biol.* **15**, 52 (2015).
46. Zeinalzadeh-Tabrizi, H., Haliloglu, K., Ghaffari, M. & Hosseinpour, A. Assessment of genetic diversity among sunflower genotypes using microsatellite markers. *Mol. Biol. Res. Commun.* **7**, 143 (2018).
47. Mammadov, J. A. et al. Development of highly polymorphic SNP markers from the complexity reduced portion of maize (*Zea mays* L.) genome for use in marker-assisted breeding. *Theor. Appl. Genet.* **121**, 577–588 (2010).
48. Blair, M. W. et al. A high-throughput SNP marker system for parental polymorphism screening, and diversity analysis in common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* **126**, 535–548 (2013).
49. Rasoulzadeh Aghdam, M., Reza, D., Sepehr, E. & Alipour, H. Association analysis of agronomic traits of oilseed sunflower (*Helianthus annuus* L.) lines with REMAP and IRAP markers under optimum and phosphorus deficit stress. *Iran. J. Field Crop Sci.* **52**, 179–196 (2021).
50. Sahari, K. et al. Genetic diversity and core collection constitution for subsequent creation of new sunflower varieties in Tunisia. *Helia* **39**, 123–137 (2016).
51. Mandel, J. R., Dechaine, J. M., Marek, L. F. & Burke, J. M. Genetic diversity and population structure in cultivated sunflower and a comparison to its wild progenitor, *Helianthus annuus* L. *Theor. Appl. Genet.* **123**, 693–704 (2011).
52. Serba, D. D. et al. Genetic diversity, population structure, and linkage disequilibrium of pearl millet. *Plant Genome* **12**, 180091 (2019).
53. Li, T. et al. Identification of ear morphology genes in maize (*Zea mays* L.) using selective sweeps and association mapping. *Front. Genet.* **11**, 747 (2020).
54. Mu, Y. et al. Transcriptome analysis reveals metabolic pathways and key genes involved in oleic acid formation of sunflower (*Helianthus annuus* L.). *Int. J. Mol. Sci.* **26**(14), 6757 (2025).
55. Pena, L. B., Pasquini, L. A., Tomaro, M. L. & Gallego, S. M. Proteolytic system in sunflower (*Helianthus annuus* L.) leaves under cadmium stress. *Plant Sci.* **171**(4), 531–537 (2006).

Acknowledgements

We thank the anonymous reviewers and editors for their constructive comments on this manuscript. Additionally, we acknowledged the support of the "Urmia University, Office of Vice Chancellor for Research" (Project number: 30840).

Author contributions

H.A. and R.D. proposed the idea and helped to provide the plant materials. S.F. performed the experiment and wrote a draft version of the manuscript. A.T. and K.H. helped with editing and improving the manuscript. All authors have read and provided their approval for the final version of the manuscript.

Funding
This research did not receive any specific funding.

Declarations

Competing Interests
The authors declare no competing interests.

Ethics approval and consent to participate
All the experiments done on wheat are in compliance with relevant institutional, national, and international guidelines and legislation.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-40372-2.

Correspondence and requests for materials should be addressed to R.D. or H.A.

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

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

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

© The Author(s) 2026