Received: 19 December 2024 | Accepted: 14 May 2025

DOI: 10.1002/csc2.70099

Crop Science

ORIGINAL ARTICLE

Crop Breeding & Genetics

Genome-wide association study of oil content and fatty acid composition in sesame (*Sesamum indicum* L.) under diverse environmental conditions

Mohammed Elsafy[1] | **Wafa Badawi**[2] | **Ahmed Ibrahim**[2] | **Amro B. Hassan**[3] | **Eu Sheng Wang**[1] | **Elamin Hafiz Baillo**[3] | **Tilal Sayed Abdelhalim**[4] | **Prabin Bajgain**[5] | **Mahbubjon Rahmatov**[1]

[1]Department of Plant Breeding, Swedish University of Agricultural Sciences (SLU), Alnarp, Sweden

[2]Agricultural Research Corporation (ARC), Wad Madani, Sudan

[3]Department of Food Science and Nutrition, College of Food and Agricultural Sciences, King Saud University, Riyadh, Saudi Arabia

[4]Biotechnology and Biosafety Research Center, Agricultural Research Corporation, Shambat, North Sudan

[5]USDA-ARS US Dairy Forage Research Center, Madison, Wisconsin, USA

Correspondence
Mohammed Elsafy and Mahbubjon Rahmatov, Department of Plant Breeding, Swedish University of Agricultural Sciences (SLU), PO Box 190, Alnarp, 234 22 Sweden.
Email: Mohammed.elsafy@slu.se and Mahbubjon.Rahmatov@slu.se

Assigned to Associate Editor Jode Edwards.

Funding information
Swedish Research Council (Vetenskapsrådet), Grant/Award Number: DR- 2020-04163

Abstract

Sesame (*Sesamum indicum* L.) is a valuable oilseed crop that is widely grown in tropical and subtropical regions because of its high oil content and favorable fatty acid profile. This study evaluated 200 genetically diverse genotypes in two distinct environments (Abu-Naama and Matuq) using an augmented block design. We employed three genome-wide association study (GWAS) models (fixed and random model circulating probability unification [FarmCPU], Bayesian information and linkage-disequilibrium iteratively nested keyway [BLINK], and multiple locus mixed model [MLMM]) to dissect the genetic basis of the oleic acid, linoleic acid, and oil content. Across environments, significant single nucleotide polymorphism (SNP) markers explained 3%–23% of the phenotypic variance, reflecting the quantitative nature of these traits. Notably, four SNPs (*Chr1_1693157*, *Chr3_23284702*, *Chr5_17024932*, and *Chr9_1711873*) were common across all three models, suggesting stable and robust associations between oleic acid and oil content. Candidate gene analysis revealed four notable sequences linked to these loci: a transcription repressor *OFP8* (*Sesamum alatum*), an *HVA22-like protein*, a *3-oxoacyl-[acyl-carrier-protein] synthase 3 A*, and a putative phospholipid *diacylglycerol acyltransferase 2* in (*S. indicum*), all of which may play key roles in oil biosynthesis and accumulation. Environment-specific loci have also emerged for linoleic acid, particularly on chromosomes 6, 9, and 13. These findings provide robust targets for marker-assisted selection and underscore the value of integrating multi-model GWAS and functional validation to develop elite sesame cultivars with improved oil quantity and quality.

Abbreviations: BLINK, Bayesian information and linkage-disequilibrium iteratively nested keyway; FarmCPU, fixed and random model circulating probability unification; GP, genomic prediction; GWAS, genome-wide association study; MLMM, multiple locus mixed model; QTL, quantitative trait loci; SNP, single nucleotide polymorphism.

Plain Language Summary

Sesame, known for its rich oils and nutrients, is a globally valued crop. Sudan, a major sesame producer, has unique varieties, yet little research has explored their potential. This study investigated the genetic factors behind Sudanese sesame's exceptional oil quality, focusing on healthy fats like oleic and linoleic acids. By examining 200 types of sesame in two environments, we identified key genetic markers (DNA "tags") that influence oil traits and revealed the significant role of the environment in shaping these qualities. Combining the right genes and growing conditions could produce sesame with superior oils. These findings offer valuable tools for farmers and breeders to improve sesame varieties, enhancing the crop's value for Sudanese producers and delivering better sesame oil to consumers worldwide.

1 | INTRODUCTION

Human well-being is closely tied to advancement in plant breeding, particularly in enhancing oleic and linoleic acids. Crops such as sesame (*Sesamum indicum* L.), soybean [*Glycine max* (L.) Merr.], canola (*Brassica napus*), and sunflower (*Helianthus annuus*) serve as valuable sources of these essential fatty acids and contributing significantly to nutrition and health (Velasco & Fernández-Martínez, 2002). Oleic acid, a monounsaturated fatty acid, is known for lowering low-density lipoprotein cholesterol and supporting cardiovascular health (Al-Madhagy et al., 2023). Linoleic acid, a polyunsaturated fatty acid, plays a key role in reducing chronic disease risk and maintaining healthy cell membranes (Kapoor et al., 2021; Mercola & D'Adamo, 2023). Thus, enhancing fatty acid content through plant breeding meets consumer demand for healthier foods, directly improving oil quality and nutritional value.

Sesame ($2n = 2x = 26$, SiSi) has a compact 350 Mb diploid genome and is one the earliest domesticated oilseed crops valued for its oil-rich seeds (Elsafy, 2023). Renowned for its nutritional and therapeutic qualities, sesame oil is rich in oleic and linoleic acids (He et al., 2020; Rauf et al., 2024). Although sesame is believed to have originated in Africa, where wild relatives persist, historical records link its early domestication to the Indian subcontinent, emphasizing its longstanding agricultural significance (Lim, 2012). Its adaptability to diverse environments has enabled its global expansion, particularly in marginal regions where other oilseeds struggle.

Advancements in plant breeding have been driven by genomics, phenomics, and gene–environment interactions (G × E), primarily through genome-wide association studies (GWASs). J. Wang et al. (2023) identified quantitative trait nucleotides associated with nine fatty acids in peanuts and detecting key single nucleotide polymorphisms (SNPs) across 15 chromosomes, particularly for oleic and linoleic acids. Similarly, Song et al. (2022) used GWAS and metabolomics to identify genetic markers linked to fatty acid variations, including palmitic and unsaturated fatty acids in sesame. These discoveries highlight the potential for marker-assisted breeding to enhance fatty acid content and quality. However, its application in predicting breeding value remains limited, underscoring the need for further research.

The genetic basis of oil content and fatty acid composition in sesame has been extensively explored, revealing key quantitative trait loci (QTL), SNPs, and candidate genes. Notably, W. Wei et al. (2013) identified significant marker-trait associations for oil content in diverse sesame germplasms, pinpointing loci related to lipid biosynthesis. Song et al. (2021) highlighted the *nsLTP* gene family, identifying specific genes (*SiLTPI.15* and *SiLTPVI.1*) that contribute to high sesame oil accumulation. Similarly, X. Wei et al. (2015) mapped genomic variations associated with oil traits, uncovering multiple lipid-related genes involved in the oil biosynthesis pathways.

In a comprehensive genome-sequencing effort, L. Wang, Yu, et al. (2014) discovered numerous SNPs linked to key enzymes in sesame's fatty acid biosynthetic pathway, such as stearoyl-ACP desaturase (SACPD) and FAD2, which regulate oleic and linoleic acid levels. Building on these findings, Zhou et al. (2022) integrated transcriptomic and QTL analyses to reveal genotype–phenotype associations for oil content identifying SNPs in regulatory regions that significantly impact seed lipid composition. These studies underscore the complexity of the sesame oil biosynthesis and highlight critical targets for marker-assisted selection, paving the way for developing cultivars with enhanced oil yield and improved fatty acid profiles.

Genomic prediction (GP) models leverage genome-wide genetic markers to predict breeding values, integrating all effects within a unified regression framework (Crossa et al., 2017). This approach has revolutionized plant breeding by enabling selection without direct phenotypic assessment, allowing breeders to make informed and efficient decisions

based on predicted values. GP is particularly valuable for sesame, where complex traits, such as fatty acid composition and oil content, are difficult to assess using traditional phenotypic methods (Bashir et al., 2023). Environmental factors significantly influence these traits, leading to trait variability across different growing conditions (Kurt et al., 2016). GP can enhance selection efficiency, improving resilience and nutritional quality. Understanding G × E is crucial for predicting trait expression and optimizing sesame genotypes in diverse climates (Hu et al., 2022). When combined, GWAS and GP serve as powerful, complementary tools, accelerating sesame breeding and the development of resilience, high-quality cultivars to meet rising consumer demand for nutrient-rich oilseed crops (Kole, 2019).

Global sesame production is dominated by Sudan, Myanmar, Tanzania, India, Nigeria, and China, collectively contributing70% of the world's harvest (FAOSTAT, 2022). Sudan stands out as a major producer and a primary center of origin for sesame, boasting rich genetic diversity (Sabiel et al., 2015). However, despite its leading role in sesame production, breeding advancements remain limited, with high yields attributed more to extensive cultivation than genetic improvements (Teklu et al., 2022).

Despite its economic significance, particularly in Sudan, its genetic basis and fatty acid composition remain poorly understood. Sudanese sesame genotypes exhibit high genetic diversity, emphasizing the need to explore their unique oil profiles and identify the genetic factors influencing fatty acid content. Enhancing nutritional value can improve marketability, generate value-added products, and create sustainable income for producers. This study represents the first genome-wide analysis of Sudanese sesame, aiming to uncover genetic factors associated with oil content and fatty acid composition in 200 genotypes. The objectives of this study were to (1) identify genetic loci associated with oil content and fatty acid composition through GWASs, (2) estimate breeding values for oil quality traits using GP models, and (3) analyze the genotype-by-environment (G × E) interactions influencing the expression of oil content and fatty acid traits.

Core Ideas

- Sudanese sesame is an underexplored genetic resource to identify novel alleles and enhance understanding of its role.
- Genome-wide association study (GWAS) and genomic prediction models identify loci and breeding values for oil quality, advancing sesame crop.
- Genotype-by-environment interactions influence oil content and fatty acid composition in sesame.

2 | MATERIALS AND METHODS

2.1 | Field experiment and plant materials

The field experiments were conducted at two locations in Sudan: Abu Naama Research Station (AN), Sennar State (12°44′43″ N, 34°07′21″ E) and Matuq (MT) Research Station, Gezira State (14°11′10″ N, 32°34′48″ E), to evaluate sesame genotypes under varying field conditions. At Abu Naama (Sennar State), the average high and low temperatures were 37.1°C and 25.9°C, respectively, with 39.99 mm average precipitation, 30.45% relative humidity, and 63 rainy days per year (≥1.0 mm rainfall). At Matuq (Gezira state), the average high and low temperatures were 37.43°C and 26.03°C, with 25.67 mm precipitation, 27.38% relative humidity, and 41 rain days per year (≥1.0 mm rainfall).

Two hundred genetically diverse sesame accessions, including gene bank accessions, landraces, released varieties, and breeding lines, were evaluated under field conditions using an augmented block design, with eight blocks per site. Each block contained 22 independent accessions and three replicated control checks, totaling 28 plots per block, with each plot measuring 4 m^2 one-row long. Standard agronomic practices, tailored to each location's environmental conditions, were applied throughout the growing season to ensure optimal growth and reliable results. Seeds were harvested and stored under optimal conditions prior to laboratory analysis to maintain data accuracy and consistency. This experimental design was carefully chosen to minimize environmental and positional effects and enhance comparative analysis robustness and reliability across different genotypes and locations.

2.2 | Gas chromatography

Total lipids were extracted as described by Tesfaye et al. (2024), with minor modifications. Briefly, 10 seeds per genotype were used, with three technical replicates. The seeds were homogenized in 1 mL of 0.15 M acetic acid and 3.75 mL of methanol/chloroform (2:1 v/v) using an IKA T18 ULTRA TURRAX homogenizer in a glass test tube. Chloroform (1.25 mL) and Millipore water (0.9 mL) were added, followed by vortexing for 10 s and centrifugation at 3000 rpm for 2 min.

After centrifugation, 200 μL of the lower chloroform phase was transferred to a clean screw-capped glass tube. The chloroform was evaporated at 70°C on a heated sand bed under a nitrogen gas stream. Once dry, the samples were reconstituted in 100 μL heptane and methylated by adding 2 mL of 2% H_2SO_4 in anhydrous methanol. The reaction proceeded at 90°C for 1 h in a sealed tube. After cooling, 1 mL Millipore water and 0.75 mL heptane were added, followed by vortexing (15 s) and centrifugation (3000 rpm for 2 min). Following centrifugation, 100 μL of the upper heptane phase

containing fatty acid methyl esters (FAMEs) was transferred to a gas chromatography (GC) for analysis.

Fatty acid profiles were analyzed using an Agilent 8860 gas chromatograph with a flame ionization detector. FAMEs were separated on a wall-coated open tubular fused-silica CP-wax 58 capillary column (50 m × 0.32 mm, Agilent) with a 10:1 split ratio. The oven temperature program started at 150°C for 0.2 min, increased by 4°C/min to 210°C, then by 10°C/min to 250°C, where it was held for 5 min. Fatty acids were determined by comparing retention times with a certified Me63 external standard (Larodan).

Fatty acid and oil content percentages were calculated using the following equations:

$$\text{Percentage of a fatty acid (\%FA)} = \frac{\text{Peak area of the FA}}{\sum \text{Peak areas of all FAs}} \times 100$$

To determine fatty acids mass, a known concentration of 17:0 artificial FA (Larodan) was used as an internal standard. The total amount of each free fatty acid was calculated based on their peak areas and molecular weights (Mw) in relation to the 17:0 internal standard according to the following equation:

$$\text{FA (mg)} = \left(\frac{\text{Peak area FA}}{\text{Peak area } 17:0}\right) \times \left(\frac{\text{Mw FA}}{\text{Mw } 17:0}\right) \times 17:0 \text{ control (mmol)} \times \text{Mw of FA species}$$

Fatty acid content (mg) was expressed relative to the weight of 10 seeds (mg/10 seeds). The total oil content was estimated based on the combined mass of all fatty acid species, together with the glycerol component of Triacylglycerol (TAG), using the following equation:

$$\text{Total oil content (mg)} = \sum \left[\text{FA mg} + \left(\left\{\frac{\text{FA } \mu\text{mol}}{3}\right\} \times 41 \text{ (Mw glycerol in ester form)}\right)\right]$$

The oil content was calculated using the following equation:

$$\text{Oil content \%} = \left(\frac{\text{Total oil content}}{\text{Mass of sample}}\right) \times 100$$

2.3 | Statistical analysis

Descriptive statistical for phenotypic traits across locations were computed using the "*psych*" package in R, while correlation analysis was performed with "*corrplot*". Broad-sense heritability (H^2) was estimated to determine the genetic contribution to oleic, linoleic, and oil content, using a mixed-effects model based on genotype means, resulting in the following estimation of H:

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_{GL}^2}{L} + \frac{\sigma_E^2}{L}}$$

Where σ_G^2 represents the genetic variance, σ_{GL}^2 denotes the genotype by location interaction variance, σ_E^2 is the residual from environmental variance, and L is the number of locations. The statistical analysis was conducted using R version 4.3.2, utilizing the "*lme4*" package to fit the mixed-effects models.

2.4 | Genotyping and data analysis

From each line, a circular section of young leaf tissue, approximately 5 mm in diameter, was harvested from each plant and placed in a 96-well plate designed for tissue collection. Genomic DNA was extracted using the Qiagen BioSprint 96 system with the Qiagen BioSprint DNA Plant kit (https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/dna-purification/genomic-dna/biosprint-96-dna-plant-kit/#orderinginformation). DNA was normalized to ng/μL, and sequencing libraries were prepared using a genotyping-by-sequencing protocol (Poland et al., 2012). Specifically, the restriction enzymes PstI and MspI were used to induce cuts at multiple sites in the genome, and the resulting pool was ligated with unique barcode adapters, multiplexed with 96 samples per lane, and sequenced on NovaSeq 6000 (Illumina). The DNA libraries were sequenced at the University of Minnesota Genomics Center.

Generated sequencing data were filtered for a minimum quality (Q) score of 30 and demultiplexed using "sabre" (https://github.com/najoshi/sabre) to sort separate reads corresponding to each sample. The reads were then aligned to the *S. indicum* updated genome assembly and annotations (M. Wang et al., 2022) using the Burrow–Wheelers Alignment tool version 0.7.4 (H. Li & Durbin, 2009). Genome-wide SNPs been identified using Samtools and bcftools (H. Li, 2011). The SNP markers were filtered to retain those with a minimum minor allele frequency (MAF) of 3% and a missing allelic proportion of 20% or less. This resulted in 3636 SNPs distributed among the 13 chromosomes and 17 high-confidence scaffolds.

2.5 | Population relatedness and linkage disequilibrium (LD)

The estimation of the genetic relatedness matrix among the genotypes was conducted in Tassel 5 utilizing the Centered_IBS method with default settings, followed by the generation of the kinship heatmap using the

"*kinship2*" R package. Moreover, to analyze the population structure and identify the optimal *K*, the software Structure 2.3.4 was employed (https://web.stanford.edu/group/pritchardlab/structure_software/release_versions/v2.3.4/html/structure.html), with the outcomes being compiled on the "StructureSelector" online platform (https://lmme.ac.cn/StructureSelector/). Furthermore, genetic variation across genotypes was assessed through principal component analysis (PCA) using the *prcomp* function in R. To estimate LD decay, Pairwise LD between markers was determined using Tassel 5 by applying a sliding window approach with 50 markers. LD decay was assessed using Tassel by employing a sliding window technique with a set of 50 markers. The calculated LD values, expressed as r^2, were plotted against the physical distances ascertained from the Sesame genome V.3.0 reference. To illustrate the LD decay pattern, a locally weighted scatterplot smoothing curve was applied, and the LD decay distance was estimated based on the approach outlined by Hill and Weir (1988).

2.6 | Association analysis and trait prediction

The SNPs marker identification was performed using three GWAS models: fixed and random model circulating probability unification (FarmCPU), Bayesian information and linkage-disequilibrium iteratively nested keyway (BLINK), and multiple locus mixed model (MLMM) on GAPIT 3 in R version 4.3.2. (Liu et al., 2016; J. Wang & Zhang, 2021). Significant SNPs were declared at the default Bonferroni thresholds in association analyses at $\alpha = 0.05$. For single SNP markers, at α/LD = 0.05/effective number of independent markers, the significance threshold corresponded to a *p*-value of approximately $-\log_{10}(0.0002451)$ or a logarithm of odds score equivalent of 3.6, and all the significant SNPs are displayed as Manhattan and *Q*–*Q* plots. A GP model was employed to predict oleic acid, linoleic acid, and oil contents across two locations using the "*rrBLUP*" R package (Endelman, 2011).

2.7 | Candidate gene search

In the pursuit of identifying potential genes that influence oil content and fatty acid composition, candidate genes were identified in areas proximate to only significant common SNP markers detected in the three GWAS models, involving the analysis of putative protein-coding sequences (Supporting Information) found within 409780 base pairs (bp) around significant genetic loci, utilizing data from a refined sesame genome assembly and annotation (M. Wang et al., 2022). However, this specific distance was selected based on the observed genome-wide average LD in sesame genotypes, which extended up to 204890 bp.

A protein BLAST search was conducted on the NCBI experimental clustered nr database platform (NCBI, 2019) to further refine this search, targeting *S. indicum* protein-coding sequences that exhibited more than 80% identity and an *E*-value of 1E-10 or lower. This process retained only the top three alignments for each *S. indicum* protein sequence for in-depth analysis. This subsequent phase of analysis involved filtering these alignments to identify candidate genes with known involvement in regulating the oil content and fatty acid profiles.

3 | RESULTS

3.1 | Phenotyping

Significant variability was observed in oleic, linoleic acid, and total oil content among the 200 sesame genotypes across the two sites, Abu Naama and Matuq (Table 1). Oleic acid levels averaged 5.34 mg/10 seeds (40.5%) at Abu Naama and 5.78 mg/10 seeds (44%) at Matuq, with individual accessions ranging from 3.07 (32%) to 9 mg/10 seeds (48.6%) and 3.1 (36%) to 8.35 mg/10 seeds (49%), respectively. Linoleic acid averaged 5.54 mg/10 seeds (42%) at Abu-Naama with a range of 3.28 (32%) to 9.12 mg/10 seeds (51%), and 5.02 mg/10 seeds (38.5%) at Matuq, within a 2.76 (33%) to 8.22 mg/10 seeds (46%) range.

Oil content showed greater variation, averaging 44.45% at Abu Naama (ranging from 33.14% to 62.05%) and 42.06% at Matuq (with a range from 30.95% to 50.27%). Skewness and kurtosis in most traits indicated non-normal, asymmetric distributions among accessions. Pairwise correlation analysis showed a negligible linear relationship between the sites for oleic ($r = 0.03$) and linoleic acid (−0.06) (Figure 1), which support non-normal, asymmetric distributions among the genotypes, likely due to the complex interplay of genetic and environmental factors affecting these traits.

The broad-sense heritability analysis (H^2) revealed a small proportion of variance in oleic acid and oil content ($H^2 = 0.28$). On the other hand, for linoleic acid, it showed nearly zero heritability, suggesting minimal genetic influence under the study conditions.

3.2 | Population relatedness and LD

The study identified population structure among the sesame genotypes (Figure 2), with PCA showed that PC1 accounted for 33% of the genetic diversity. However, distinct groups can be distinguished based on the location of the genotype. Genotypes from the GenBank collection formed a relatively tight cluster, showing their genetic relatedness and likely a common ancestral background, whereas breeding, collection, landrace, and variety genotypes showed greater dispersed

TABLE 1 Variation in sesame oil and fatty acid composition traits across 200 accessions.

	Abu-Naama			Matuq		
Measure	**Oleic (mg/10 seeds)**	**Linoleic (mg/10 seeds)**	**Oil content%**	**Oleic (mg/10 seeds)**	**Linoleic (mg/10 seeds)**	**Oil content%**
Minimum	3.07	3.28	33.14	3.1	2.67	30.95
Maximum	9	9.12	62.05	8.35	8.22	50.27
Mean	5.43	5.54	44.45	5.78	5.02	42.06
SE	1.03	1.22	5.06	0.98	0.92	3.51
Median	5.07	5.1	43.42	5.21	4.52	40.58
Variance	4.36	5	261.1	7.79	5.97	385.46
Skewness	−1.17	−0.99	−1.72	−0.59	−0.53	−0.78
Kurtosis	0.83	0.47	1.64	−1.35	−1.29	−1.31

Note: The table provides a statistical summary of sesame oil and fatty acid composition traits, including oleic acid, linoleic acid, and total oil content, for 200 sesame accessions grown in Abu Naama and Matuq. The data included the minimum, maximum, mean, standard error, median, variance, skewness, and kurtosis values.
Abbreviation: SE, standard error.

patterns, exhibiting their higher genetic diversity and potential admixture. Structure analysis determined an optimal *K* value of 2 (Figure 3a), indicating the presence of two major subpopulations within the germplasm panel, which was corroborated by the population structure membership coefficients (Figure 3b), where most genotypes showed strong assignment to one of the two subpopulations. The kinship matrix heatmap (Figure 4a) highlighted varying degrees of genetic relatedness among the 200 sesame genotypes, based on genome-wide marker. LD decay showed $r^2 = 0.1$ at a physical distance of approximately 204890 bp (Figure 4b), indicating the extent of genetic recombination across the genome.

3.3 | Association analysis and trait prediction

Using three GWAS models (FarmCPU, Blink, and MLMM) across two environments (Abu-Naama and Matuq), several significant SNP markers were identified for oleic acid, linoleic acid, and oil content (Table 2; Figure 5a–c).

In Abu-Naama, FarmCPU detected five SNPs on chromosomes 1, 5, and 9 for oleic acid, two SNPs on chromosomes 9 and 13 for linoleic acid, and two SNPs on chromosome 3 for oil content, explaining 7%–9% of the phenotypic variance. Blink model identified oleic acid-associated SNPs on chromosomes 1, 5, and 7 ($R^2 = 11\%–16\%$) and an oil-content marker on chromosome 3 ($R^2 = 9\%$). MLMM pinpointed oleic acid SNPs on chromosomes 1 and 5 ($R^2 = 10\%–21\%$) and two oil content SNPs on chromosome 3.

In the Matuq, all three models consistently identified SNPs on chromosome 8 (*Chr8_31702733* and *Chr8_31825156*) and chromosome 11 (*Chr11_14710318*) for oleic acid and chromosome 9 (*Chr9_1711873*) for oil content (R^2 up to 23%). Notably, *Chr9_1711873* was consistently significant for oil content, through effect directions varied between models, highlighting how reference alleles and statistical frameworks influence marker effect estimation.

Across analyses, minor allele frequencies ranged from 0.06 to 0.49, underscoring substantial allelic diversity. While individual SNPs explained a modest portion of the variance, reflecting the polygenic nature of these traits, the consistent association of *Chr5_17024932* (oleic acid in Abu-Naama) and *Chr9_1711873* (oil content in Matuq) highlights potential key targets for further genetic dissection and breeding.

Across the three GWAS models, several common SNPs were identified in both environments (Abu-Naama and Matuq), indicating robust associations despite the model differences. In Abu-Naama, *Chr1_1693157* and *Chr5_17024932* were consistently associated with oleic acid, whereas *Chr3_23284702* was a shared marker for oil content across all three models. However, no significant SNPs were detected for linoleic acid in any model.

In the Matuq, *Chr9_1711873* consistently influenced the oil content across FarmCPU, Blink, and MLMM, underscoring its potential importance for oil trait improvement. However, no single SNP was universally identified for oleic or linoleic acid, likely reflecting their polygenic complexity.

The GP for oil content oleic and linoleic acid concentrations across two environments, Abu Naama and Matuq, indicated varying predictive abilities. Linoleic acid at Matuq was had the highest mean predictive ability (0.21), ranging from −0.24% to 0.56%, closely followed by oil content at Abu Naama (a mean prediction of 0.22 within the range of−0.10 to 0.56) (Figure 6). Oleic acid at Matuq showed a slightly lower mean prediction of 0.15, ranging from −0.34 to 0.52. Linoleic acid in Abu Naama had a mean predictive ability of 0.13, ranging from −0.24 to 0.40, while oleic acid in Abu Naama had the lowest mean prediction (0.07), with a broader range from −0.23 to 0.26. The weakest predictive ability was observed for oil content at Matuq, with a mean of −0.04 and a range of −0.42–0.25.

TABLE 2 Genome-wide identification of the genetic markers associated with oil content, oleic acid, and linoleic acid composition in 200 sesame accessions using the following models: fixed and random model circulating probability unification (FarmCPU), Bayesian information and linkage-disequilibrium iteratively nested keyway (BLINK), and multiple locus mixed model (MLMM).

Model	Trait/locations	SNP Marker	Chr	Position/bps	Alleles	MAF	LOD	R^2 (%)	Effect
FarmCPU	Abu-Naama								
	Oleic acid	Chr1_1693157*	1	1693157	C/T	0.45	3.21	7	0.4
	Oleic acid	Chr5_17024932*	5	17024932	C/T	0.1	3.22	7	−0.49
	Oleic acid	Chr5_17029387	5	17029387	T/A	0.11	2.95	7	0.45
	Oleic acid	Chr5_17034072	5	17034072	A/G	0.11	3.39	8	−0.5
	Oleic acid	Chr9_26574216	9	26574216	C/T	0.06	2.98	7	−0.89
	Linoleic acid	Chr13_13034832	13	13034832	A/T	0.14	3.03	7	0.48
	Linoleic acid	Chr9_3976219	9	3976219	G/A	0.07	3.01	7	−0.68
	Linoleic acid	Chr9_3976223	9	3976223	A/C	0.07	3.06	7	0.68
	Oil content	Chr3_23284702*	3	23284702	A/G	0.3	3.99	9	1.75
	Oil content	Chr3_23284761	3	23284761	T/A	0.31	3.76	8	−1.68
	Matuq								
	Oleic acid	Chr11_14710318	11	14710318	C/T	0.18	3.5	8	0.38
	Oleic acid	Chr8_31702733	8	31702733	G/A	0.06	3.61	8	0.63
	Oleic acid	Chr8_31825156	8	31825156	A/T	0.06	3.61	8	−0.63
	Oil content	Chr9_1711873*	9	1711873	A/G	0.09	4.07	8	−2.06
Blink	Abu-Naama								
	Oleic acid	Chr1_1693157*	1	1693157	C/T	0.49	4.28	11	−0.35
	Oleic acid	Chr5_17024932*	5	17024932	C/T	0.11	4.05	16	0.42
	Oleic acid	Chr7_15291065	7	15291065	G/A	0.08	3.22	12	−0.57
	Oil content	Chr3_23284702*	3	23284702	A/G	0.3	4.44	9	−1.38
	Matuq								
	Oleic acid	Chr11_14710318	11	14710318	C/T	0.25	3.3	3	−0.23
	Oleic acid	Chr6_25438022	6	25438022	G/T	0.46	3.45	20	−0.7
	Oleic acid	Chr8_31702733	8	31702733	G/A	0.06	3.68	18	0.43
	Oleic acid	Chr8_31825156	8	31825156	A/T	0.06	3.77	0	0.44
	Linoleic acid	Chr6_24421274	6	24421274	G/C	0.12	3.56	19	−0.29
	Oil content	Chr9_1711873*	9	1711873	A/G	0.11	4.32	23	1.22
MLMM	Abu-Naama								
	Oleic acid	Chr1_1693157*	1	1693157	C/T	0.49	3.48	10	−0.36
	Oleic acid	Chr5_17024932*	5	17024932	C/T	0.11	3.27	21	0.43
	Oil content	Chr3_23284702*	3	23284702	A/G	0.3	3.26	9	−1.38
	Oil content	Chr3_23284761	3	23284761	T/A	0.3	3.15	0	−1.35
	Matuq								
	Linoleic acid	Chr6_24421274	6	24421274	G/C	0.12	3.5	19	−0.36
	Linoleic acid	Chr6_24440370	6	24440370	G/A	0.12	3.15	0	−0.34
	Oil content	Chr9_1711873*	9	1711873	A/G	0.11	3.25	23	1.22

Note: This table provides details of single nucleotide polymorphism (SNP) markers significantly associated with key traits, including oleic acid, linoleic acid, and oil contents, in sesame accessions grown at two locations, Abu-Naama and Matuq. The table includes the SNP marker ID, chromosome (Chr) and physical position (position/bp) of the marker, alleles, minor allele frequency (MAF), logarithm of odds (LOD) score, proportion of phenotypic variance explained (R^2%), estimated effect size of the associated allele (effect), and the shared SNP markers across the three models (*).

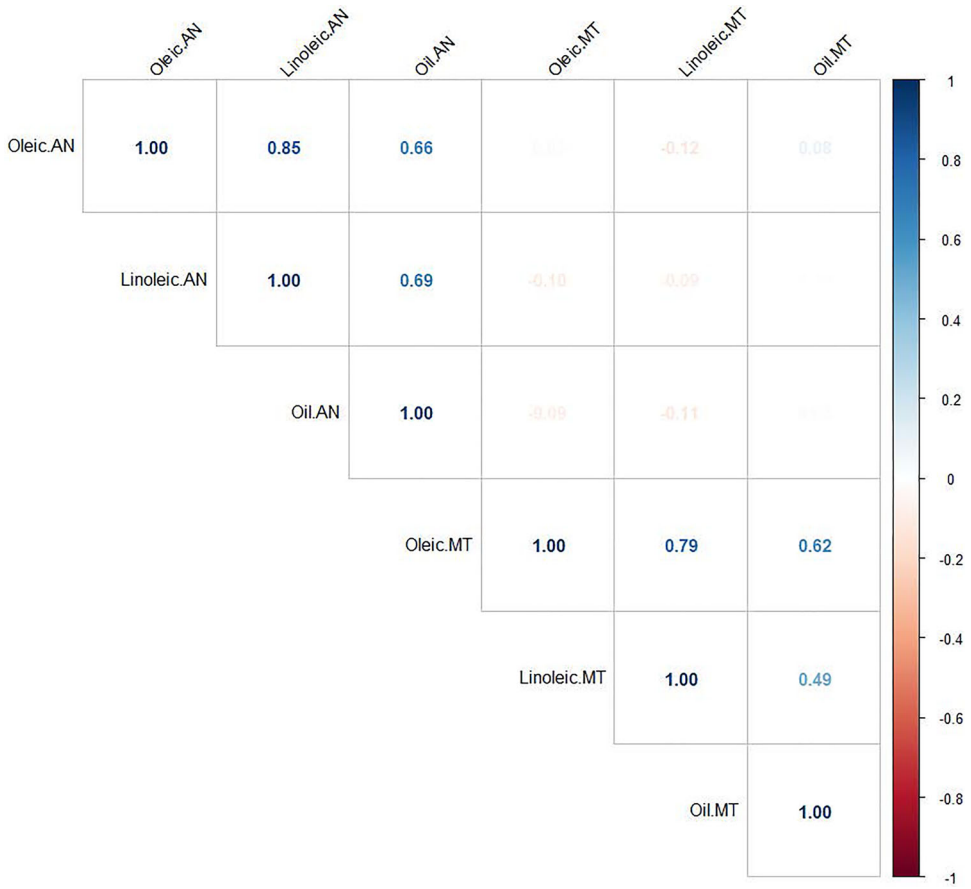

FIGURE 1 Correlation of sesame oil and fatty acid composition traits between locations. The heat map depicts the correlation between sesame oil and fatty acid composition traits, including oleic acid (oleic acid), linoleic acid (linoleic acid), and total oil content (oil), across the two locations, Abu Naama (AN) and Matuq (MT). The values represent the correlation coefficients, indicating the strength and direction of the relationship between the two locations for each trait.

TABLE 3 Candidate genes associated with fatty acid and oil content traits in sesame.

NCBI candidate genes	Trait	SNP	Species	Annotation	*E*-value	% identity
APMJ01000051	Oleic acid	Chr1_1693157	*S. alatum*	Transcription repressor OFP8	4.00E-90	81.40
APMJ01001210	Oil content	Chr3_23284702	*S. indicum*	HVA22-like protein a	2.00E-114	100
APMJ01003105	Oleic acid	Chr5_17024932	*S. indicum*	3-oxoacyl-[acyl-carrier-protein] synthase 3 A	0.0	99.70
APMJ01005016	Oil content	Chr9_1711873	*S. indicum*	putative phospholipid:diacylglycerol acyltransferase 2	0.0	100

Note: The table summarizes the key candidate genes linked to oleic acid and oil content traits in *Sesamum* species identified through significant SNPs. Gene annotations, *E*-values, and percentage identity scores indicate the strength of genetic association and homology to known proteins.
Abbreviation: SNP, single nucleotide polymorphism.

3.4 | Candidate gene search

Candidate gene analysis using the *S. indicum* v1.0 reference genome identified four protein-coding sequences with high homology (≥ 80% identity, *E*-value ≤ 1E-10) near significant SNP markers (Table 3; Supporting Information). *APMJ01000051*, associated with oleic acid at SNP *Chr1_1693157*, showed 81.4% identity (*E*-value: 4.00E-90) with a transcription repressor *OFP8* from *S. alatum*. For oil content, *APMJ01001210* (SNP *Chr3_23284702*) matched a *HVA22-like protein a* in *S. indicum* (100% identity, *E*-value: 2.00E-114). Another oleic acid candidate, *APMJ01003105* (SNP *Chr5_17024932*), had 99.7% identity (*E*-value: 0.0) to *3-oxoacyl-[acyl-carrier-protein] synthase*

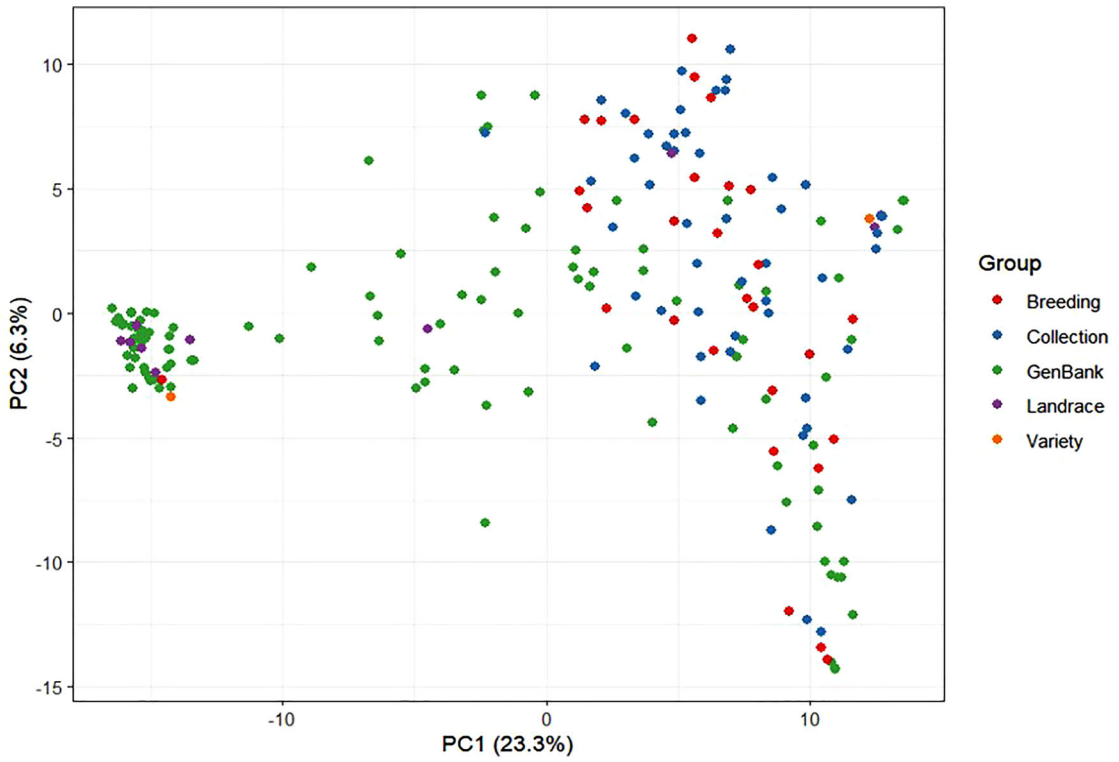

FIGURE 2 Genetic diversity and population structure of 200 sesame accessions. The principal component analysis (PCA) plot illustrates the genetic diversity and population structure among various sesame accession groups. Each point represents an individual accession, with colors distinguishing different groups, such as breeding lines, collections, GenBank accessions, landraces, and varieties. The *x*- and *y*-axes correspond to the first and second principal components, respectively, representing the primary dimensions of the genetic variation within the dataset.

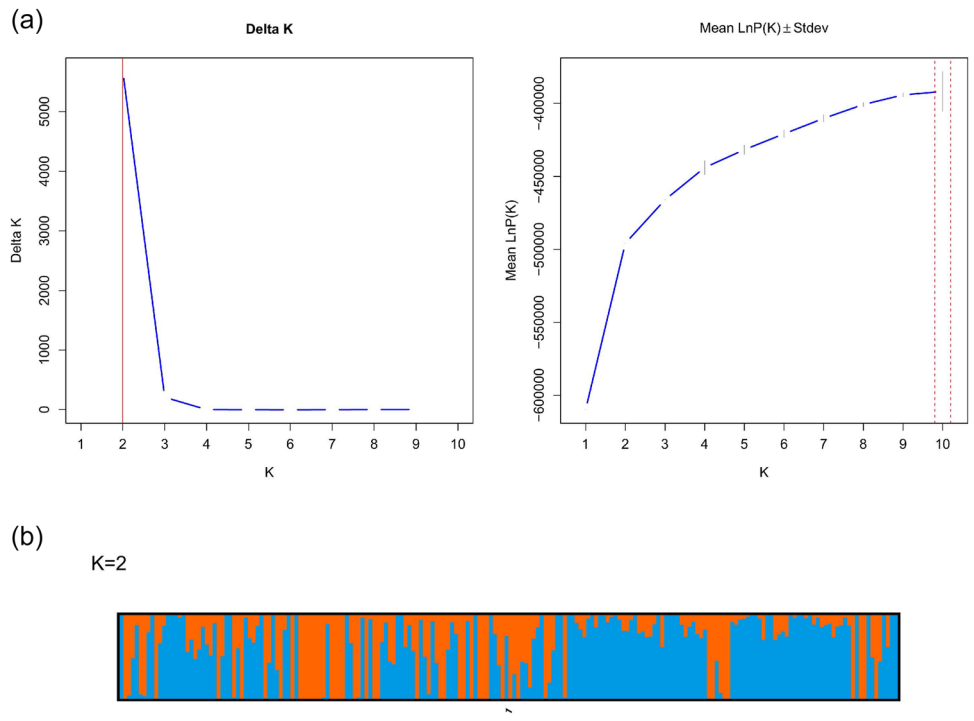

FIGURE 3 Population structure analysis: (a) estimated Delta *K* and Ln*P*(K) values for different *K* values. (b) Population structure of the 200 sesame accessions. The heatmap illustrates the subpopulations structure, with accessions grouped into two distinct subpopulations represented in red and green.

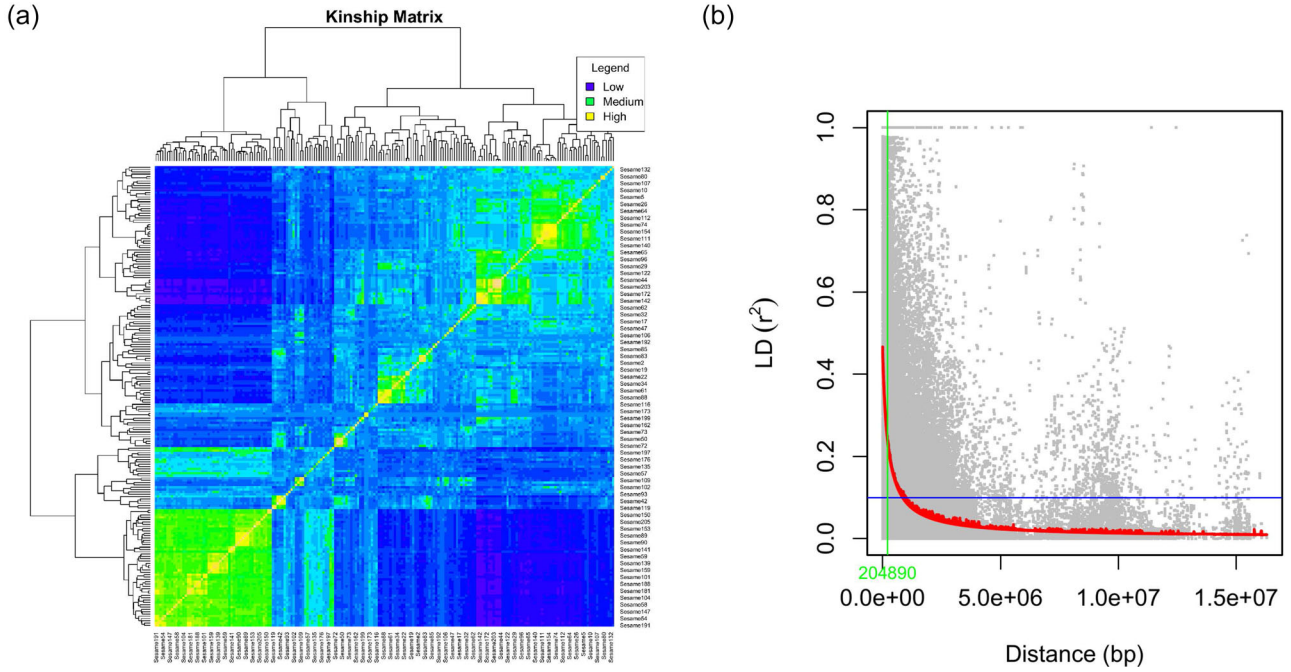

FIGURE 4 (a) Kinship heatmap showing the genetic relationships among 200 sesame accessions based on additive relationships. (b) Linkage disequilibrium (LD) patterns in the sesame genome. The figure shows the distribution of LD decay across genomic distances in the sesame populations. The *x*-axis represents the physical distance in base pairs (bp), whereas the *y*-axis represents the LD measure (r^2). Each data point corresponds to a pairwise comparison of the genetic markers. The red line depicts the overall trend of LD decay, indicating a decrease in genetic linkage between markers as the physical distance increases.

3 A, an enzyme in fatty acid elongation. *APMJ01005016* (SNP *Chr9_1711873*), linked to oil content, aligned with a *phospholipid*, *diacylglycerol acyltransferase 2*, in *S. indicum* (100% identity, *E*-value: 0.0), a key enzyme in triacylglycerol biosynthesis.

4 | DISCUSSION

In this study, 200 Sudanese sesame accessions were evaluated across two locations, Abu Namma and Matug Research farms of the Agricultural Research Corporation, Sudan, using an augmented design with replicated checks. The accessions represented a diverse collection, including gene bank accessions, landraces, advanced breeding materials, and released cultivars. Oil content and fatty acid composition, specifically oleic and linoleic acid, were analyzed using GC. Broad-sense heritability and population structure analyses were conducted to assess the genetic variation among the accessions. Three GWASs were performed using three models namely FarmCPU, BLINK, and MLMM to identify SNPs associated with oil content, oleic, and linoleic acid across both locations. A total of four SNPs were identified for oil content, and oleic acid across, while no significant SNPs were detected in case of linoleic acid content. To the best of our knowledge, this is the first study to utilize GWAS for identifying SNPs associated with oil content and fatty acid composition in Sudanese sesame germplasm

4.1 | Phenotypic traits

Evaluating 200 sesame genotypes across Abu Naama and Matuq locations revealed significant variations in fatty acid composition and oil content, shaped by genotype-by-environment (G × E) interactions. The wide range of oleic (32.8%–49.6%) and linoleic acids highlights substantial genetic diversity, offering valuable potential for breeding programs targeting health and industrial applications. Oleic acid is favored for its stability and health benefits, whereas linoleic acid, a polyunsaturated fatty acid, is essential for human health but reduces oil stability (Gunstone, 2011). Our finding aligns with previous reports including X. Wei et al. (2015), who documented oleic acid values of 32.08%–53.14% in a diverse sesame panel, while Mondal et al. (2010) reported a range of 36.7%–52.4% in Indian sesame and Uzun et al. (2008) reported 29.3%–41%. The observed variation confirms the potential of Sudanese sesame germplasm for genetic improvement to enhance oil quality, reinforcing its nutritional and economic value.

The significant variation in oil content, averaging of 44.45% in Abu Naama and 42.06% in Matuq, demonstrates

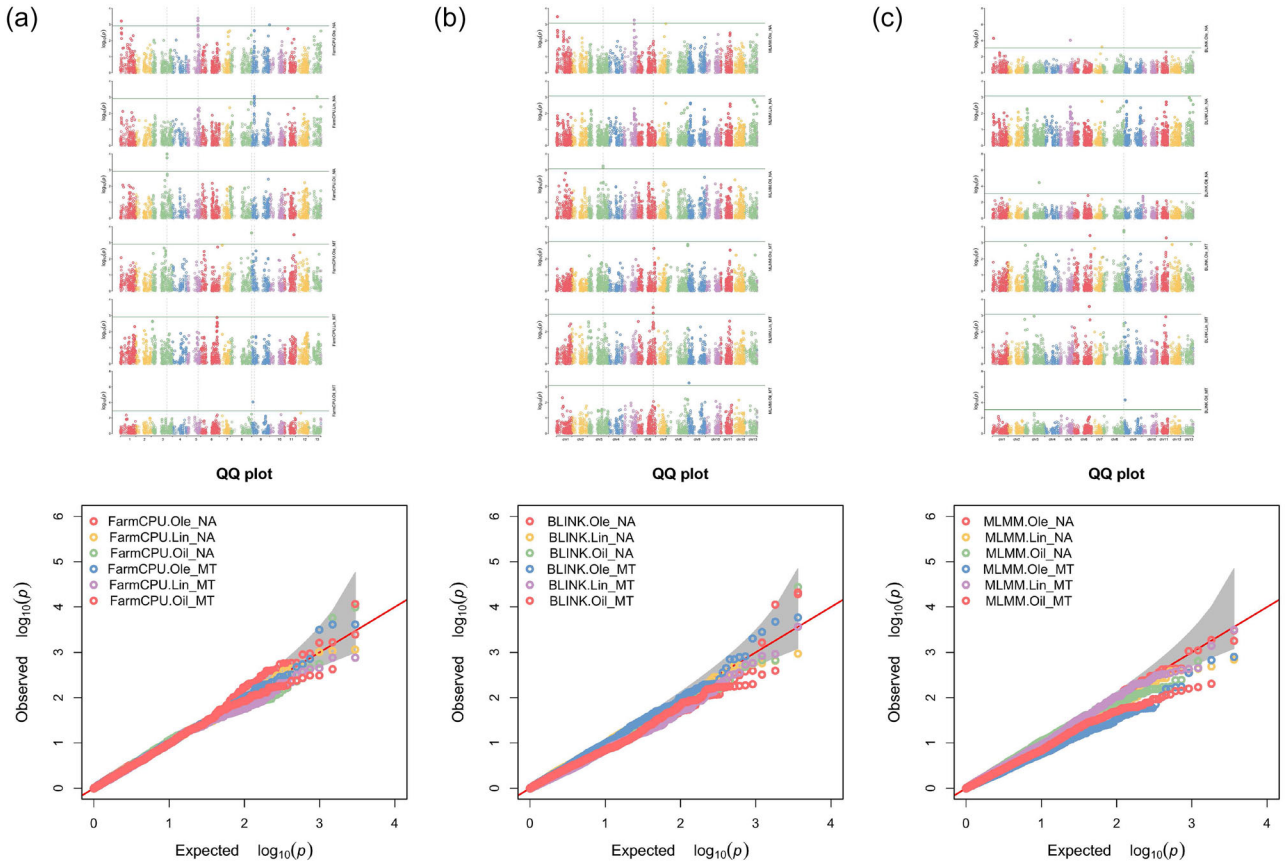

FIGURE 5 Manhattan and *Q*–*Q* plots based on three different GWAS models: (A) fixed and random model circulating probability unification (FarmCPU), (B) Bayesian information and linkage-disequilibrium iteratively nested keyway (BLINK), and (C) multiple locus mixed model (MLMM), identifying significant single nucleotide polymorphisms (SNPs) associated with oil content, oleic acid, and linoleic acid in 200 sesame accessions across two locations Abu Naama (AN) and Matuq (MT). The Manhattan plots show the distribution of $-\log_{10}(p\text{-values})$ for each SNP across the 13 sesame chromosomes, with the *x*-axis representing chromosomal position and the *y*-axis indicating the strength of association. Horizontal lines across the plots denote genome-wide significance thresholds. The *Q*–*Q* plots compare the expected versus observed *p*-values distributions, where deviations from the diagonal at the upper end suggests an enrichment of significantly associated SNPs beyond random expectation.

the potential for selecting genotype with higher oil yields. The minimal correlation between oleic and linoleic acid content across locations (with *r*-values near zero) indicated a weak linear relationship between these traits under the different environmental conditions. This aligns with previous findings that environmental factors, such as soil type and temperature, strongly influence oil content and fatty acid composition in sesame (Uzun et al., 2008).

The low broad-sense heritability ($H^2 = 0.28$) for oleic acid and oil content, along with near-zero heritability for linoleic acid, indicates minimal genetic variance, suggesting that the phenotypic differences are largely environmentally driven rather than genetic (Holland et al., 2003). This poses challenges for breeding programs, as selecting genotypes based on performance in one environment may not predict outcomes in different conditions. Similar findings in oil crops support this, with Uzun and Çağırgan (2006) demonstrating that environmental factors significantly impact fatty acid profiles in sesame. Likewise, Khan and Nawaz (2022) reported low heritability estimates for oil content and fatty acid composition, reinforcing the dominant role of environmental influences. Arslan et al. (2007) further reported significant genotype-by-environment interactions affecting oil content and fatty acid composition in sesame mutants, reinforcing the need for breeding strategies that integrate environmental assessments alongside genetic evaluations to improve the stability across diverse growing conditions.

4.2 | Population relatedness and LD

Population structure analysis revealed substantial genetic diversity, as indicated by PCA using 3636 SNP markers. The first principal component accounted for 33% of the genetic variation, surpassing the 27% reported by Sabag et al. (2021) in their study on flowering date and yield component

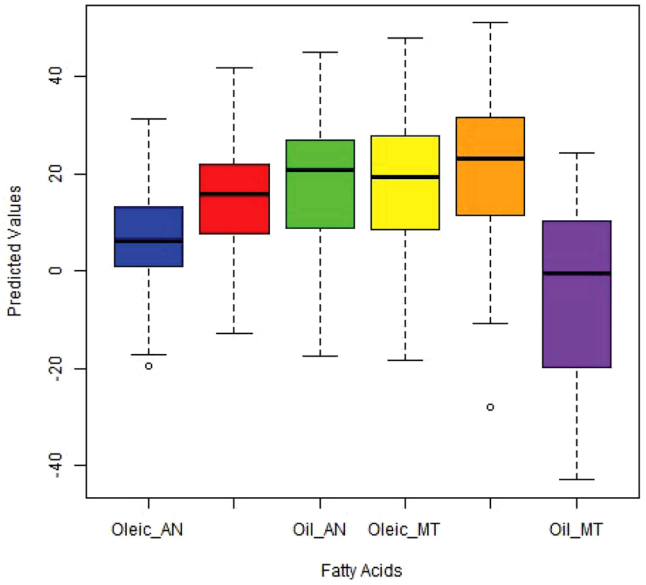

FIGURE 6 Predicted values for oil and fatty acid content in sesame accessions across two locations. The boxplots show the predicted distributions of oleic and linoleic acids and oil content in 200 sesame accessions grown at two locations: Abu Naama (AN) and Matuq (MT). The central line indicates the median, while the boxes represent the interquartile range (IQR). Whiskers extend to the non-outlier range, and individual points denote the outliers. Colors differentiate traits and locations: blue and yellow for oleic acid, red and orange for linoleic acid, and green and purple for oil content at AN and MT, respectively.

trade-offs in sesame using 20,294 SNP markers across 184 genotypes. This variation may stem from differences in the reference genome as their study used the L. Wang, Yu, et al. (2014) reference genome, whereas this study employed the updated annotation reference genome from M. Wang et al. (2022). The clustering of "GenBank" genotypes suggest close genetic relationship, likely due to shared ancestry and similar selection pressures.

The dispersed clustering of "breeding," "collection," "landraces," and "variety" genotypes reflects their broader genetic diversity, shaped by their geographical origins and distinct selection pressures. Sudan's long history of sesame cultivation (Mohamed, 2011) has contributed to this wide genetic variation, as confirmed by a structure analysis, which identified $K = 2$ as the optimal value, dividing the germplasm into two subpopulations. This classification provides valuable insights into genetic architecture, supporting the strategic integration of desirable traits for developing improved sesame varieties, aligning with Parry and Hawkesford (2012).

The kinship matrix illustrates genetic relationships within the sesame population, identifying diverse parent lines suitable for heterosis exploitation. A LD measure (r^2) of 0.1 at approximately 204890 bp indicates relatively rapid LD decay, differing from the 163930 bp reported by Tesfaye et al. (2022) using an older sesame reference genome version. Our LD decay is longer than 150 kb (L. Wang, Han, et al., 2014) and 166 kb (Seay et al., 2024) but shorter than 1639.3 kb reported by Tesfaye et al. (2022) in Ethiopian sesame accessions. This variation may stem from differences in population structure, as our Sudanese germplasm subset likely experienced greater recombination events or stronger bottlenecks, resulting in larger linkage blocks. In addition, differences in marker density and selection criteria could influence LD estimates; as our study utilized a denser SNP set and an updated reference genome, potentially detecting fewer recombination breakpoints and inflating LD block sizes.

Sample size and genetic diversity significantly influence LD decay; genetically uniform or those under strong selection tend to exhibit extended LD. In addition, methodological differences, such as defining LD thresholds ($r^2 = 0.1$ vs. $r^2 = 0.2$) and variations in analytical pipelines, contribute to discrepancies in LD estimates across studies. Despite this variation, the relatively high LD observed here supports fine mapping of agronomically important genes in sesame breeding. Our results suggest that recombination has fragmented linkage blocks over generations, and this rapid LD decay benefits association mapping by improving the resolution for identifying trait-controlled genes.

4.3 | Association analysis and trait prediction

GWAS has gained prominence in sesame research, identifying genomic associations for agronomic traits, such as 1000-seed weight, seed size, plant height, seed coat color (Du et al., 2019; L. Wang, Yu, et al., 2014; L. Wang et al., 2016). For quality traits, Wu et al. (2017) mapped QTLs linked to oil, protein, and sesamin content across sesame chromosomes. Zhou et al. (2022) expanded this by analyzing 14 fatty acids over 2 years using GC–mass spectrometry, identifying 249 QTLs associated with fatty acid composition. In addition, they detected 43 unique SNPs linked to key oil traits, including palmitic, stearic, oleic, linoleic, and arachidic acids. Shared loci across traits, such as those for linoleic acid, palmitic acid, and oil content on chromosome 11, highlight the genetic complexity of fatty acid biosynthesis. Further analysis of these SNP loci identified 671 genes within an 88 kb window, linked to metabolic, cellular, and singling pathways, providing critical targets for improving sesame oil quality through molecular breeding.

Despite progress in sesame genomics, few studies have examined the genetic associations of fatty acids and oil content, particularly in Africa. This study used GWAS to analyze oleic and linoleic acids and oil content in a set of 200 Sudanese sesame genotypes using GWAS across two distinct environments. These findings provide insight into the genetic architecture of sesame in Africa, where environmental conditions strongly influence trait expression. Similar to Zhou et al.

(2022), this study identified multiple SNP loci associated with key oil quality traits, enhancing the understanding of fatty acid composition and oil content. The results confirm the polygenic inheritance of seed oil traits in sesame, with multiple loci contributing to trait variability, consistent with findings in other oilseed crops (Pandey et al., 2014; Reinprecht et al., 2009; X. Wei et al., 2015).

Our GWAS results underscore the polygenic nature of sesame oil traits, revealing environment-dependent and model-specific SNP associations. The repeated detection of *Chr1_1693157* and *Chr5_17024932* across models and environments suggests a stable QTL for oleic acid, aligning with an earlier work mapping seed-quality traits to chromosome 5 (W. Wei et al., 2013). Similarly, *Chr3_23284702* and *Chr9_1711873* were consistently associated with oil content, supporting earlier findings that identified these chromosomes as oil-related QTL hotspots (C. Li et al., 2014).

A recent GWAS combined with transcriptome analysis identified novel loci and regulatory genes involved in fatty acid biosynthesis (Zhou et al., 2022). This underscore the importance of integrating functional validation and expression profiling, as single-marker GWAS may not fully capture regulatory genes influencing fatty acid composition. Combining expression data with marker-trait associations enables the identification of key transcription factors or enzymes with significant allelic effects under different environmental conditions.

The presence of low frequency of minor alleles (MAF as low as 0.06) with notable effect sizes, modest R^2 values (3-23%), and the environment-specific signals (linoleic acid SNPs on chromosome 9 in Abu-Naama vs. chromosome 6 in Matuq) support a quantitative inheritance model and G × E interaction for these traits. Recent transcriptomic studies show the importance of multi-omics approaches in dissecting complex pathways, confirming candidate genes and regulatory elements that GWAS alone overlooks. Future breeding efforts will benefit from validating these candidate genes at genomic and transcriptomic levels while integrating fine-mapping and marker-assisted selection to accelerate the development of high-oil, high-oleic sesame cultivars.

This study showed variations in predictive ability across traits and environments, highlighting the complex interplay between genetic and environmental factors in determining crop traits. The moderate predictive power observed for linoleic acid and oil content in specific environments suggests that genomic selection, particularly using models such as rrBLUP, can aid in breeding sesame varieties with improved oil content and fatty acid profiles. However, lower predictive abilities observed for oleic acid and oil content highlight the need to consider trait–environment interactions when applying genomic selection strategies. Moderate to high predictive abilities in some trait–environment combinations demonstrate genomic selection's potential for improving sesame oil quality, through further model optimization and integration of environmental covariates may improve accuracy, especially for traits with lower predictability (Crossa et al., 2017). Environmental factors significantly influence sesame oil content and fatty acid composition, leading to substantial trait variability (Kurt et al., 2016; J. Wang et al., 2023).

Hu et al. (2022) reported the significant influence of environmental and genetic diversity on sesame oil and fatty acid profiles, with oleic acid ranging from 39% to 54% and linoleic acid from 39% to 59%. The advancement of GP models has improved breeding value estimation, enabling breeders to make informed selections by considering multiple interacting factors. L. Chen et al. (2015) emphasized that genomic selection is particularly beneficial for complex traits like fatty acid composition, as it integrates environmental factors, improving traits predictability and crop performance under specific conditions. Combining genomic data with environmental variables can enhance the development of nutritionally rich, climate-resilient sesame varieties, promoting sustainable cultivation.

4.4 | Candidate gene search

Our candidate gene analysis identified four protein-coding sequences near the significant SNP markers associated with oleic acid and oil content. *APMJ01003105* on *Chr5_17024932* corresponds to *3-oxoacyl-[acyl-carrier-protein] synthase 3A* (KAS III), an enzyme involved in fatty acid synthesis by catalyzing the initial condensation of *acetyl-CoA* with *malonyl-ACP* (Guo et al., 2019). This elongation step is essential for seed oil composition (Berg et al., 2015). In addition, *APMJ01000051* on *Chr1_1693157* shares 81.4% identity with the transcription repressor *OFP8*, which may regulate key enzymatic genes influencing fatty acids accumulation in sesame seeds.

For oil content, *APMJ01001210* on *Chr3_23284702* encodes an *HVA22-like protein*, which is involved in stress response and membrane trafficking (W. Chen et al., 2002), although its direct role in oil biosynthesis remains unclear. Its strong alignment indicates a potential link between seed development and lipid metabolism. Another candidate, *APMJ01005016* on *Chr9_1711873*, corresponds to phospholipid *diacylglycerol acyltransferase 2* (PDAT2), an enzymes that facilitates TAG production by transferring *acyl* groups from phospholipids to diacylglycerol (Dahlqvist et al., 2000). Enhanced PDAT2 activity can boost TAG levels, ultimately increasing sesame oil content.

Zhou et al. (2022) identified three candidate genes (*SINPZ1100015*, *SINPZ1201700*, and *SINPZ1201748)* linked to key loci for fatty acid and oil content across diverse global accessions. While our study analyzed fewer accessions across two environmental conditions, it leveraged the broad

genetic diversity of Sudanese sesame germplasm, a center of origin for the crop. Integrating these findings with large-scale GWAS and transcriptomic studies (Dossa et al., 2019; Zhou et al., 2022) could enhance our understanding of the genetic mechanisms governing oil traits. Fine-mapping major QTLs across multiple environments is crucial for accurately capturing genotype–environment interactions.

The candidate genes highlight the key molecular regulators of fatty acid and oil biosynthesis, with potential links to stress response pathways that may influence oil yield and composition. Future studies should focus on functional validation through gene expression analysis, knockout/knockdown approaches, and overexpression studies. Marker-assisted selection targeting these genes could enhance breeding efficiency for high-oleic, high-oil sesame cultivars, thereby meeting growers, consumers, and industrial demands.

While this study does not capture the full genetic diversity of Sudanese sesame, there remains considerable untapped variation in wild relatives, landraces, and underutilized cultivars. Exploring these resources could reveal lines with enhanced oleic acid content. Molecular breeding approaches, including marker-assisted selection and CRISPR/Cas genome editing of key fatty acid biosynthesis genes, such as *FAD2* (Rauf et al., 2024), could accelerate genetic improvement. Introgression from related species and optimized agronomic practices may further expand the oil content and oleic acid range, enhancing sesame's nutritional and market value.

5 | CONCLUSION

This comprehensive GWAS analysis provides insight into the complex genetic factors influencing oil content and fatty acid composition (oleic and linoleic acids) in Sudanese sesame genotypes, identifying the key loci responsible for these traits. The observed variability across environments demonstrates the significant impact of environmental factors on trait expression, underscoring the need for environment-specific breeding strategies. The identification of candidate genes associated with these traits offers new opportunities for molecular breeding to enhance the nutritional quality of sesame oil.

GP also showed promise for specific traits in distinct environments, particularly for oleic acid in Abu Naama and oil content in Matuq, suggesting that genomic selection could improve these traits and optimize sesame breeding programs. This study enhances the genetic understanding of sesame and supports the development of improved varieties in both local and global markets. Future work should emphasize validating gene function (expression assays, knockouts, and overexpression) and using marker-assisted selection to boost breeding efficiency for high-oleic, high-oil sesame, meeting both market and consumer needs.

AUTHOR CONTRIBUTIONS

Mohammed Elsafy: Conceptualization; formal analysis; investigation; visualization; writing—original draft. **Wafa Badawi**: Data curation. **Ahmed Ibrahim**: Data curation. **Amro B. Hassan**: Writing—review and editing. **Eu Sheng Wang**: Investigation. **Elamin Hafiz Baillo**: Data curation; writing—review and editing. **Tilal Sayed Abdelhalim**: Funding acquisition; supervision; writing—review and editing. **Prabin Bajgain**: Supervision; visualization; writing—review and editing. **Mahbubjon Rahmatov**: Funding acquisition; supervision; visualization; writing—review and editing.

ACKNOWLEDGMENTS

We thank the Swedish Research Council (Vetenskapsrådet) (grant DR-2020-04163) for supporting this work. Furthermore, we extend our gratitude to the Agricultural Research Corporation—Sudan for the research facilities and the Swedish University of Agricultural Sciences for their laboratory facilities and publication fees.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Sequenced reads for the sesame lines used in this study are available under NCBI Bioproject PRJNA1184775.

ORCID

Mohammed Elsafy https://orcid.org/0000-0002-2008-5182

Prabin Bajgain https://orcid.org/0000-0002-3093-5582

REFERENCES

Al-Madhagy, S., Ashmawy, N. S., Mamdouh, A., Eldahshan, O. A., & Farag, M. A. (2023). A comprehensive review of the health benefits of flaxseed oil in relation to its chemical composition and comparison with other omega-3-rich oils. *European Journal of Medical Research*, *28*(1), 240.

Arslan, Ç., Uzun, B., Ülger, S., & İlhan Çağırgan, M. (2007). Determination of oil content and fatty acid composition of sesame mutants suited for intensive management conditions. *Journal of the American Oil Chemists' Society*, *84*, 917–920. https://doi.org/10.1007/s11746-007-1125-6

Bashir, H., Khalid, M., Ajmad, I., Ullah, F., & Ammar, A. (2023). Integrating genomics and biotechnological approaches to enhance abiotic stress tolerance in sesame (*Sesamum indicum* L.). *Bulletin of Biological and Allied Sciences Research*, *2023*(1), 37–37. https://doi.org/10.54112/bbasr.v2023i1.37

Berg, J., Tymoczko, J., Gatto, G., Jr., & Stryer, L. (2015). Protein synthesis. *Biochemistry*, *8*, 893–924.

Chen, L., Ekine-Dzivenu, C., Vinsky, M., Basarab, J., Aalhus, J., Dugan, M. E., Fitzsimmons, C., Stothard, P., & Li, C. (2015). Genome-wide association and genomic prediction of breeding values for fatty acid composition in subcutaneous adipose and longissimus lumborum muscle of beef cattle. *BMC Genetics*, *16*, 1–15. https://doi.org/10.1186/s12863-015-0290-0

Chen, W., Provart, N. J., Glazebrook, J., Katagiri, F., Chang, H.-S., Eulgem, T., Mauch, F., Luan, S., Zou, G., & Whitham, S. A. (2002). Expression profile matrix of *Arabidopsis* transcription factor genes suggests their putative functions in response to environmental stresses. *The Plant Cell*, *14*(3), 559–574. https://doi.org/10.1105/tpc.010410

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., De Los Campos, G., Burgueño, J., González-Camacho, J. M., Pérez-Elizalde, S., & Beyene, Y. (2017). Genomic selection in plant breeding: Methods, models, and perspectives. *Trends in Plant Science*, *22*(11), 961–975. https://doi.org/10.1016/j.tplants.2017.08.011

Dahlqvist, A., Ståhl, U., Lenman, M., Banas, A., Lee, M., Sandager, L., Ronne, H., & Stymne, S. (2000). Phospholipid:diacylglycerol acyltransferase: An enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants. *Proceedings of the National Academy of Sciences*, *97*(12), 6487–6492. https://doi.org/10.1073/pnas.120067297

Dossa, K., Li, D., Zhou, R., Yu, J., Wang, L., Zhang, Y., You, J., Liu, A., Mmadi, M. A., & Fonceka, D. (2019). The genetic basis of drought tolerance in the high oil crop *Sesamum indicum*. *Plant Biotechnology Journal*, *17*(9), 1788–1803. https://doi.org/10.1111/pbi.13100

Du, H., Zhang, H., Wei, L., Li, C., Duan, Y., & Wang, H. (2019). A high-density genetic map constructed using specific length amplified fragment (SLAF) sequencing and QTL mapping of seed-related traits in sesame (*Sesamum indicum* L.). *BMC Plant Biology*, *19*(1), 1–20. https://doi.org/10.1186/s12870-019-2172-5

Elsafy, M. (2023). *Status of sesame breeding* [Paper presentation]. Faculty of Landscape Architecture, Horticulture and Crop Production Science, Alnarp, Sweden. https://pub.epsilon.slu.se/30295/1/elsafy-m-2023-02-17.pdf

Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome*, *4*(3), 250–255. https://doi.org/10.3835/plantgenome2011.08.0024

FAOSTAT. (2022). *Crops and livestock products*. https://www.fao.org/faostat/en/#data/QCL

Gunstone, F. D. (2011). Production and trade of vegetable oils. In F. D. Gunstone (Ed.), *Vegetable oils in food technology: Composition, properties and uses* (pp. 1–24). John Wiley & Sons. https://doi.org/10.1002/9781444339925

Guo, Q.-Q., Zhang, W.-B., Zhang, C., Song, Y.-L., Liao, Y.-L., Ma, J.-C., Yu, Y.-H., & Wang, H.-H. (2019). Characterization of 3-oxacyl-acyl carrier protein reductase homolog genes in *Pseudomonas aeruginosa* PAO1. *Frontiers in Microbiology*, *10*, 442548. https://doi.org/10.3389/fmicb.2019.01028

He, Q., Lee, T.-R., Yu, J., Oo, W. H., Yoon, M.-Y., Min, M.-H., Chu, S.-H., Kim, K.-W., Lee, Y.-S., & Park, Y.-J. (2020). Genotypic variation in fatty acids in whole grain sesame (fatty acids in whole grain sesame). *Journal of Crop Science and Biotechnology*, *23*, 9–20. https://doi.org/10.1007/s12892-019-0195-0

Hill, W., & Weir, B. (1988). Variances and covariances of squared linkage disequilibria in finite populations. *Theoretical Population Biology*, *33*(1), 54–78. https://doi.org/10.1016/0040-5809(88)90004-4

Holland, J. B., Nyquist, W. E., Cervantes-Martínez, C. T., & Janick, J. (2003). Estimating and interpreting heritability for plant breeding: An update. In J. B. Holland, W. E. Nyquist, & C. T. Cervantes-Martínez (Eds.), *Plant breeding reviews* (vol. 22, pp. 9–112). John Wiley & Sons.

Hu, Y., Han, Z., Shen, W., Jia, Y., He, L., Si, Z., Wang, Q., Fang, L., Du, X., & Zhang, T. (2022). Identification of candidate genes in cotton associated with specific seed traits and their initial functional characterization in *Arabidopsis*. *The Plant Journal*, *112*(3), 800–811. https://doi.org/10.1111/tpj.15982

Kapoor, B., Kapoor, D., Gautam, S., Singh, R., & Bhardwaj, S. (2021). Dietary polyunsaturated fatty acids (PUFAs): Uses and potential health benefits. *Current Nutrition Reports*, *10*, 232–242. https://doi.org/10.1007/s13668-021-00363-3

Khan, M. A., & Nawaz, N. (2022). Devising selection strategy for increase in sesame yield based on variability heritability and genetic advance studies. *Pure and Applied Biology*, *11*(1), 226–231. https://doi.org/10.19045/bspab.2022.110024

Kole, C. (2019). *Genomic designing of climate-smart oilseed crops*. Springer. https://doi.org/10.1007/978-3-319-93536-2

Kurt, C., Arioglu, H., Erdem, T., Akkaya, M. R., Sabagh, A., & Isalam, M. (2016). A comparative study of fatty acid extraction methods of sesame (*Sesamum indicum* L.) varieties grown under mediterranean environment. *Journal of Experimental Biology and Agricultural Sciences*, *4*, 588–593.

Li, C., Miao, H., Wei, L., Zhang, T., Han, X., & Zhang, H. (2014). Association mapping of seed oil and protein content in *Sesamum indicum* L. using SSR markers. *PLoS One*, *9*(8), e105757. https://doi.org/10.1371/journal.pone.0105757

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Lim, T. K. (2012). *Edible medicinal and non-medicinal plants* (vol. 1). Springer.

Liu, J., Yang, C., Shi, X., Li, C., Huang, J., Zhao, H., & Ma, S. (2016). Analyzing association mapping in pedigree-based GWAS using a penalized multitrait mixed model. *Genetic Epidemiology*, *40*(5), 382–393. https://doi.org/10.1002/gepi.21975

Mercola, J., & D'Adamo, C. R. (2023). Linoleic acid: A narrative review of the effects of increased intake in the standard American diet and associations with chronic disease. *Nutrients*, *15*(14), 3129. https://doi.org/10.3390/nu15143129

Mohamed, I. A. W. (2011). The Sudanese agriculture development, history, policies and sustainability. *Agricultural and Natural Resource Economics Journal*, *3*(82). https://ssrn.com/abstract=1829469

Mondal, N., Bhat, K., & Srivastava, P. (2010). Variation in fatty acid composition in Indian germplasm of sesame. *Journal of the American Oil Chemists' Society*, *87*(11), 1263–1269. https://doi.org/10.1007/s11746-010-1615-9

National Center for Biotechnology Information (NCBI). (2019). *Protein BLAST: Search protein databases using a protein query*. NCBI.

Pandey, M. K., Wang, M. L., Qiao, L., Feng, S., Khera, P., Wang, H., Tonnis, B., Barkley, N. A., Wang, J., & Holbrook, C. C. (2014). Identification of QTLs associated with oil content and mapping FAD2 genes and their relative contribution to oil quality in peanut (*Arachis hypogaea* L.). *BMC Genetics*, *15*, 1–14. https://doi.org/10.1186/s12863-014-0133-4

Parry, M. A., & Hawkesford, M. J. (2012). An integrated approach to crop genetic improvement. *Journal of Integrative Plant Biology*, *54*(4), 250–259. https://doi.org/10.1111/j.1744-7909.2012.01109.x

Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J.-L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One*, *7*(2), e32253. https://doi.org/10.1371/journal.pone.0032253

Rauf, S., Basharat, T., Gebeyehu, A., Elsafy, M., Rahmatov, M., Ortiz, R., & Kaya, Y. (2024). Sesame, an underutilized oil seed crop: Breeding achievements and future challenges. *Plants*, *13*(18), 2662. https://doi.org/10.3390/plants13182662

Reinprecht, Y., Luk-Labey, S. Y., Larsen, J., Poysa, V., Yu, K., Rajcan, I., Ablett, G., & Pauls, K. (2009). Molecular basis of the low linolenic acid trait in soybean EMS mutant line RG10. *Plant Breeding*, *128*(3), 253–258. https://doi.org/10.1111/j.1439-0523.2008.01571.x

Sabag, I., Morota, G., & Peleg, Z. (2021). Genome-wide association analysis uncovers the genetic architecture of tradeoff between flowering date and yield components in sesame. *BMC Plant Biology*, *21*, 1–14. https://doi.org/10.1186/s12870-021-03328-4

Sabiel, S., Ismail, M., Abdalla, E., & Osman, A. (2015). Genetic variation in sesame genotypes (*Sesamum indicum* L.) grown in the semi-arid zone of the Sudan. *SABRAO Journal of Breeding & Genetics*, *47*(3), 214–220.

Seay, D., Szczepanek, A., De La Fuente, G. N., Votava, E., & Abdel-Haleem, H. (2024). Genetic diversity and population structure of a large USDA sesame collection. *Plants*, *13*(13), 1765. https://doi.org/10.3390/plants13131765

Song, S., You, J., Shi, L., Sheng, C., Zhou, W., Dossou, S. S. K., Dossa, K., Wang, L., & Zhang, X. (2021). Genome-wide analysis of nsLTP gene family and identification of SiLTPs contributing to high oil accumulation in sesame (*Sesamum indicum* L.). *International Journal of Molecular Sciences*, *22*(10), 5291. https://doi.org/10.3390/ijms22105291

Song, S., Zhang, L., Zhao, Y., Sheng, C., Zhou, W., Dossou, S. S. K., Wang, L., You, J., Zhou, R., & Wei, X. (2022). Metabolome genome-wide association study provides biochemical and genetic insights into natural variation of primary metabolites in sesame. *The Plant Journal*, *112*(4), 1051–1069. https://doi.org/10.1111/tpj.15995

Teklu, D. H., Shimelis, H., & Abady, S. (2022). Genetic improvement in sesame (*Sesamum indicum* L.): Progress and outlook: A review. *Agronomy*, *12*(9), 2144. https://doi.org/10.3390/agronomy12092144

Tesfaye, M., Feyissa, T., Hailesilassie, T., Wang, E. S., Kanagarajan, S., & Zhu, L. H. (2024). Rapid and non-destructive determination of fatty acid profile and oil content in diverse Brassica carinata germplasm using Fourier-transform near-infrared spectroscopy. *Processes*, *12*(2), 244.

Tesfaye, T., Tesfaye, K., Keneni, G., Alemu, T., & Alemu, A. (2022). Genome-wide association study for yield-related traits in sesame (*Sesamum indicum*). *Plant Breeding*, *141*(2), 246–256. https://doi.org/10.1111/pbr.13004

Uzun, B., Arslan, Ç., & Furat, Ş. (2008). Variation in fatty acid compositions, oil content and oil yield in a germplasm collection of sesame (*Sesamum indicum* L.). *Journal of the American Oil Chemists' Society*, *85*(12), 1135–1142. https://doi.org/10.1007/s11746-008-1304-0

Uzun, B., & Çağırgan, M. İ. (2006). Comparison of determinate and indeterminate lines of sesame for agronomic traits. *Field Crops Research*, *96*(1), 13–18. https://doi.org/10.1016/j.fcr.2005.04.017

Velasco, L., & Fernández-Martínez, J. M. (2002). Breeding oilseed crops for improved oil quality. *Journal of Crop Production*, *5*(1–2), 309–344. https://doi.org/10.1300/J144v05n01_13

Wang, J., Chen, H., Li, Y., Shi, D., Wang, W., Yan, C., Yuan, M., Sun, Q., Chen, J., & Mou, Y. (2023). Identification of quantitative trait nucleotides and development of diagnostic markers for nine fatty acids in the peanut. *Plants*, *13*(1), 16. https://doi.org/10.3390/plants13010016

Wang, J., & Zhang, Z. (2021). GAPIT version 3: Boosting power and accuracy for genomic association and prediction. *Genomics, Proteomics & Bioinformatics*, *19*(4), 629–640.

Wang, L., Han, X., Zhang, Y., Li, D., Wei, X., Ding, X., & Zhang, X. (2014). Deep resequencing reveals allelic variation in *Sesamum indicum*. *BMC Plant Biology*, *14*, 1–11. https://doi.org/10.1186/s12870-014-0225-3

Wang, L., Xia, Q., Zhang, Y., Zhu, X., Zhu, X., Li, D., Ni, X., Gao, Y., Xiang, H., & Wei, X. (2016). Updated sesame genome assembly and fine mapping of plant height and seed coat color QTLs using a new high-density genetic map. *BMC Genomics*, *17*(1), 1–13.

Wang, L., Yu, S., Tong, C., Zhao, Y., Liu, Y., Song, C., Zhang, Y., Zhang, X., Wang, Y., & Hua, W. (2014). Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. *Genome Biology*, *15*, 1–13. https://doi.org/10.1186/gb-2014-15-2-r39

Wang, M., Huang, J., Liu, S., Liu, X., Li, R., Luo, J., & Fu, Z. (2022). Improved assembly and annotation of the sesame genome. *DNA Research*, *29*(6), dsac041. https://doi.org/10.1093/dnares/dsac041

Wei, W., Zhang, Y., Lü, H., Li, D., Wang, L., & Zhang, X. (2013). Association analysis for quality traits in a diverse panel of Chinese sesame (*Sesamum indicum* L.) germplasm. *Journal of Integrative Plant Biology*, *55*(8), 745–758. https://doi.org/10.1111/jipb.12049

Wei, X., Liu, K., Zhang, Y., Feng, Q., Wang, L., Zhao, Y., Li, D., Zhao, Q., Zhu, X., & Zhu, X. (2015). Genetic discovery for oil production and quality in sesame. *Nature Communications*, *6*(1), 8609. https://doi.org/10.1038/ncomms9609

Wu, K., Wu, W., Yang, M., Liu, H., Hao, G., & Zhao, Y. (2017). QTL mapping for oil, protein and sesamin contents in seeds of white sesame. *Acta Agronomica Sinica*, *43*(7), 1003–1011. https://doi.org/10.3724/SP.J.1006.2017.01003

Zhou, W., Song, S., Dossou, S. S. K., Zhou, R., Wei, X., Wang, Z., Sheng, C., Zhang, Y., You, J., & Wang, L. (2022). Genome-wide association analysis and transcriptome reveal novel loci and a candidate regulatory gene of fatty acid biosynthesis in sesame (*Sesamum indicum* L.). *Plant Physiology and Biochemistry*, *186*, 220–231. https://doi.org/10.1016/j.plaphy.2022.07.023

SUPPORTING INFORMATION

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

How to cite this article: Elsafy, M., Badawi, W., Ibrahim, A., Hassan, A. B., Wang, E. S., Baillo, E. H., Abdelhalim, T. S., Bajgain, P., & Rahmatov, M. (2025). Genome-wide association study of oil content and fatty acid composition in sesame (*Sesamum indicum* L.) under diverse environmental conditions. *Crop Science*, *65*, e70099. https://doi.org/10.1002/csc2.70099