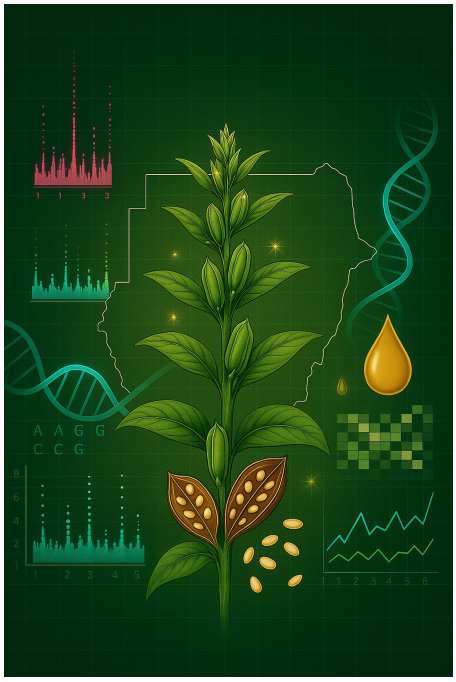
Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2025:50
Faculty of Landscape Architecture, Horticulture and Crop Production Science

Genomic insights for sesame improvement

Oil composition, seed coat pigmentation, shattering, and shelf life

Mohammed Elsafy



Genomic insights for sesame improvement

Oil composition, seed coat pigmentation, shattering, and shelf life

Mohammed Elsafy

Faculty of Landscape Architecture, Horticulture and Crop Production Sciences

Department of Plant Breeding

Alnarp

SLU

SWEDISH UNIVERSITY OF AGRICULTURAL SCIENCES

DOCTORAL THESIS

Alnarp 2025

Acta Universitatis Agriculturae Sueciae

2025:50

Cover: Integration of genomics and phenotyping in sesame improvement, highlighting key research elements such as GWAS peaks, DNA sequence motifs, oil content analysis, and trait variation in Sudanese germplasm. The central figure is a stylized sesame plant superimposed over the map outline of Sudan, reflecting the focus on native diversity and local breeding relevance. Copyright ChatGPT (OpenAI) in collaboration with Mohammed Elsafy, 2025 (No commercial reuse without permission from the creator).

ISSN 1652-6880
ISBN (print version) 978-91-8046-559-5
ISBN (electronic version) 978-91-8046-564-9
https://doi.org/10.54612/a.3b0u19hekl

Swedish University of Agricultural Sciences, Department of Plant Breeding, Alnarp, Sweden
 Print: SLU Grafisk service, Alnarp 2025

Genomic insights for sesame improvement

Abstract

Sesame is a drought-resilient, nutritionally rich oilseed crop widely grown in tropical and subtropical regions. Sudan is recognized as a major global producer and a center of genetic diversity for sesame. However, Sudanese germplasm remains largely underutilized in global breeding efforts. This thesis summarizes five studies evaluating key agronomic and biochemical traits in 200 diverse Sudanese sesame genotypes, assessed across multiple seasons and agroecological zones. A storage stability study (for 10 commonly cultivated genotypes) under accelerated aging conditions (55°C, 60% RH) demonstrated genotype-specific variation in oil stability and antioxidant retention. Landrace Abusundoug displayed exceptional oxidative stability, highlighting its value for breeding improved shelf-life cultivars. Comprehensive biochemical profiling revealed substantial variations in oil content (32.8–50.2%), oleic acid (41.3–47.6%) and linoleic acid (35.0–41.4%). Seed coat color showed extensive variation (CIELab parameters: L*, a*, b*), with limited correlation with oil traits. Black and dark brown seeds contained higher antioxidant-related compounds, while white-seeded genotypes typically exhibited a higher oil content. Genome-wide association studies (GWAS) and genomic prediction identified genetic loci underlying complex traits, including oil quality, seed pigmentation, and capsule-shattering resistance. GWAS revealed 21 significant SNP loci linked to seed coat pigmentation involving flavonoid biosynthesis genes (e.g., *WRKY, DOF zinc finger*). Capsule-shattering analyses identified five robust SNP markers, notably *Chr2_15649330* and *Chr8_31466064*, and candidate genes (*MKK5*, *RZF1*, *COR27*) which are homologous to pod-shattering genes from Arabidopsis and *Brassica napus*. These findings highlight the considerable potential of Sudanese sesame germplasm for developing climate-resilient, nutritionally enhanced, and commercially valuable sesame varieties through marker-assisted breeding.

Keywords: Gene analysis, Genetic resources utilization, Neglected crops, Marker-assisted selection, Plant breeding, Sudan

Genomic insights for sesame improvement

Sammanfattning

Sesam är en torktålig och näringsrik oljeväxt som odlas i tropiska och subtropiska regioner. Sudan erkänns globalt som en ledande producent och ett viktigt centrum för genetisk diversitet hos sesam. Trots detta förblir sudanesisk sesam till stor del underutnyttjad i globala förädlingsprogram. Denna avhandling sammanfattar resultat från fem studier där viktiga agronomiska och biokemiska egenskaper utvärderades hos 200 genetiskt varierande sudanesiska sesamgenotyper över flera säsonger och agroekologiska zoner. En lagringsstudie utförd på tio ofta odlade genotyper under lagring (55°C, 60 % relativ luftfuktighet) visade genotypspecifik variation i oljestabilitet och förekomst av antioxidanter. Den lokala sorten Abusundoug uppvisade exceptionellt hög oxidationsstabilitet, vilket understryker dess värde för utveckling av sorter med förbättrad lagringshållbarhet. Omfattande biokemiska analyser avslöjade stor variation i oljehalt (32,8–50,2 %), oljesyrahalt (41,3–47,6 %) och linolsyrahalt (35,0–41,4 %). Fröfärgen visade betydande variation, men korrelationen med oljeegenskaper var begränsad. Svarta och mörkbruna frön innehöll mer antioxidativa komponenter, medan vita frön generellt hade högre oljehalt. Genomstudier (GWAS) och genomisk prediktion identifierade genetiska regioner kopplade till komplexa egenskaper, som oljekvalitet, fröpigmentering och motståndskraft mot kapselsprickning. En GWAS-analys identifierade 21 signifikanta SNP-lokus associerade med fröfärg, vilket inkluderade kandidatgener i flavonoidbiosyntesen (t.ex. WRKY, *DOF-zinkfinger*). För kapselsprickning identifierades fem robusta SNP-markörer, där *Chr2_15649330* och *Chr8_31466064*, samt kandidatgener (*MKK5*, *RZF1*, *COR27*) var homologa med gener för skid- och kapselsprickning hos Arabidopsis och *Brassica napus*. Dessa resultat belyser den stora potentialen hos sudanesisk sesam för utveckling av klimattåliga, näringsinnehållande och kommersiellt värdefulla sorter genom markörstödd växtförädling.

Nyckelord: Genanalys, Utnyttjande av genetiska resurser, Försummade grödor, Markörstödd selektion, Växtförädling, Sudan

Dedication

To the loving memory of my dear parents, Samira and Ahmed, whose sacrifices and values shaped the foundation of my journey.

To my cherished siblings, Khalid, Habab, and Nassania, for their unwavering support and belief in me.

To my beloved wife, Eleni, and our precious children, Mareal and Lynn, whose love, patience, and joy give meaning to every step I take.

To my aunt Nadia and her family for their kindness and encouragement along the way.

To Moneim Fatih, a wellspring of wisdom whose clarity of thought, principled life, and quiet strength have profoundly shaped my own. From you, I have not only learned but understood.

To my comrade in struggle, Magdi Elgezoli, whose friendship and solidarity have been a source of strength throughout this journey.

To all those around the world who continue to struggle for dignity, justice, and liberation, this work stands in solidarity with your fight and hopes for a better future for all.

Contents

List of publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I. Elsafy, M., Ekholm, A., Elkhatim, K. A. S., Hamid, M. G., Othman, M. H., Abdelhalim, T. S., Rahmatov, M., Johansson, E. & Hassan, A. B. (2024). Tracking the storage stability in sesame (Sesamum indicum L.): impact of accelerated storage on storability characteristics, seed quality, phytochemical content, and fatty acids. Discover Agriculture, 2(1), 55. https://doi.org/10.1007/s44279-024-00077-4

II. Elsafy, M., Badawi, W., Ibrahim, A., Hafiz Baillo, E., Bajgain, P., Abdelhalim, T. S., & Rahmatov, M. (2025). Genome-wide association scan and candidate gene analysis for seed coat color in sesame (Sesamum indicum L.). Frontiers in Plant Science, 16, 1541656. https://doi.org/10.3389/fpls.2025.1541656

III. Elsafy, M., Badawi, W., Zakaria, A., Abdelhalim, T. S., Rahmatov, M., & Johansson, E. (2025). Exploring the Diversity in Oil Content, Fatty Acid Profiles, and Seed Coat Color in Sudanese Sesame Germplasm: Implications for Breeding and Crop Improvement. Plant-Environment Interactions, 6(2), e70051. https://doi.org/10.1002/pei3.70051

IV. 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 (Sesamumindicum L.) under diverse environmental conditions. Crop Science, 65, e70099. https://doi.org/10.1002/csc2.70099

V. Elsafy, M., Badawi, W., Ibrahim, A., Baillo, E.H., Abu Assar, A. H., Brhane, H., Mahmood, U., Bajgain, P., Abdelhalim, T.S., Rahmatov, M. (2025). Novel Alleles Regulating Capsule Shattering in Sesame Revealed Through Multi-Model Genomic Analysis and Field-Based Phenotyping. (Manuscript).

The contribution of Mohammed Elsafy to the papers included in this thesis was as follows:

I. Planned and designed the seed storage stability experiment, including accelerated aging, chemical assays, and colorimetric analysis. Analyzed the data, created the figures and tables, and drafted the manuscript with input from co-authors.

II. Responsible for designing the study, supervising seed coat color phenotyping, and performing GWAS, candidate gene analyses and statistical data interpretation. He wrote the first draft of the manuscript and coordinated revisions with co-authors.

III. Led the extraction and quantification of seed oil and fatty acid profiles from Sudanese sesame germplasm, performed the data analysis, interpreted the results, wrote the draft of the manuscript, and incorporated feedback from all authors.

IV. Carried out the GWAS and genomic prediction analyses for oil content and fatty acid composition, interpreted the results in collaboration with co-authors, created all visualizations, and wrote the manuscript draft.

V. Performed genome-wide association analyses for capsule-shattering traits using multiple GWAS models, interpreted the candidate gene results together with co-authors, led the integration of phenotypic and genomic data, created all visualizations, and wrote the manuscript draft.

List of tables

List of figures

Abbreviations

AOAC	Association of Official Analytical Collaboration
BLINK	Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway
BLUP	Best Linear Unbiased Prediction
BLUE	Best Linear Unbiased Estimate
CIELab	Commission Internationale de l'Éclairage color space
FarmCPU	Fixed and random model Circulating Probability Unification
GBS	Genotyping-by-Sequencing
GP	Genomic Prediction
GWAS	Genome-Wide Association Study
LOD	Logarithm of the Odds
MAS	Marker-Assisted Selection
MLMM	Multi-Locus Mixed Model
NCBI	National Center for Biotechnology Information
QTL	Quantitative Trait Locus
SNP	Single Nucleotide Polymorphism

1. Introduction

Sesame (*Sesamum indicum* L.) is a crucial crop cultivated predominantly in arid and semi-arid regions, serving as a vital cash crop for small-scale farmers (Elsafy 2023). Its cultivation often provides the essential financial resources for acquiring agricultural inputs critical to sustaining staple food production. Despite its significance, sesame has historically received limited attention in terms of research and breeding efforts, leading to stagnation in productivity and overall production efficiency. Exploiting sesame's rich genetic diversity through advanced molecular breeding techniques presents an opportunity to develop high-yielding varieties which can thrive under climate change-induced stresses and adverse environmental conditions. Successful breeding initiatives can significantly enhance sesame production, thereby directly improving the livelihoods and economic resilience of vulnerable smallholder farming communities.

2. Background

2.1 Economic Importance

Sesame plays a vital role in the rural economy of many tropical and subtropical countries. It is considered a high-value oilseed crop due to its adaptability to diverse agro-climatic conditions, including in drought-prone and marginal lands, where few other crops can thrive. It serves as a major income source for small-holder farmers, often requiring low input but offering a relatively high market value. According to (FAOSTAT 2023), major producers include Sudan, Myanmar, India, China, and Nigeria, with Sudan being a global leader in both production and export (Figure 1). However, the crop is still underutilized in terms of global research and development investment, which has resulted in relatively low average yields compared to other oilseeds. Despite being grown on over 12 million hectares, sesame production is not meeting its potential, yielding around 500 kg/ha on average, largely due to constraints such as capsule shattering, biotic and abiotic stressors, and limited mechanization (Myint *et al.* 2020).

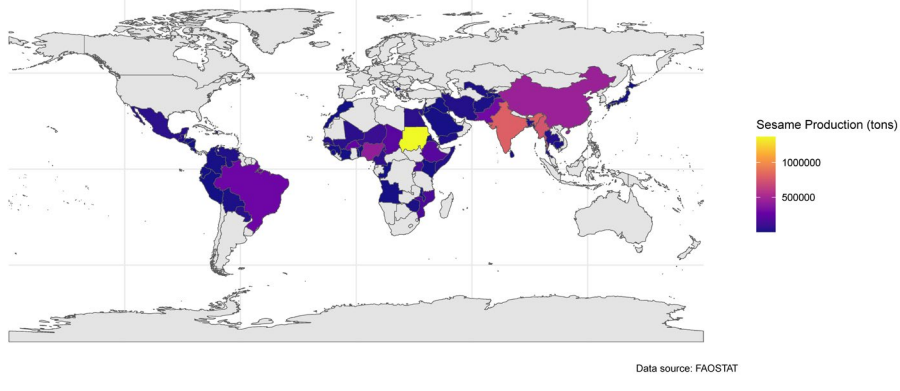


Figure 1. Global sesame production by country

2.2 Nutritional Value of Sesame

Sesame seeds are highly nutritious and serve as a valuable component of both human and animal diets. They are particularly rich in lipids (35–60%), predominantly unsaturated fatty acids such as oleic and linoleic acids, and contain significant amounts of protein (19–30%), making them an important protein supplement in many diets (Elleuch *et al.* 2007). The seeds are also rich in essential micronutrients, including calcium, magnesium, phosphorus, iron, and zinc. Moreover, sesame is a source of important bioactive compounds, such as lignans (e.g., sesamin, sesamolin), tocopherols (vitamin E), and phytosterols, which contribute to its antioxidant, anti-inflammatory, and cholesterol-lowering effects (Wu 2007; Pathak *et al.* 2017). In addition, the leaves of sesame, though underutilized, have high nutritional value and are consumed in several African communities during food shortages because of their high levels of ascorbic acid, carotenoids, and minerals (Latham 1979; Bennett 1998).

2.3 Sesame Domestication

Sesame is considered to be one of the oldest oilseed crops domesticated by humans, with archaeological evidence dating its cultivation back over 3,000 years. The origin of sesame has long been debated, with both Africa and the Indian subcontinent theorized as centers of domestication. The genus *Sesamum*, comprising over 30 species, shows high diversity in tropical Africa, where many wild relatives, such as *S. angustifolium*, *S. radiatum*, and *S. alatum,* are found (Bedigian 2003). However, *S. malabaricum*, found in India, is genetically closest to cultivated sesame, suggesting possible independent or secondary domestication in South Asia. This complex domestication history reflects sesame's diverse adaptation and usage across cultures, in which its seeds, oil, and leaves have been traditionally used for food, medicine, and religious rituals (Orsi *et al.* 2017; Elsafy 2023).

2.4 Genetic Resources

Despite its long cultivation history, sesame exhibits a relatively narrow genetic base in modern varieties, posing a challenge for breeders. Genetic erosion caused by habitat destruction, civil unrest, and the replacement of traditional landraces with modern cultivars has been particularly severe in areas like Sudan, which hosts diverse indigenous germplasm (Bedigian 2003;

Robinson 2005). To counteract this loss, initiatives such as the Eastern African Plant Genetic Resources Network (EAPGRN) and NordGen have promoted the conservation and documentation of genetic resources, supported by Sida (Demissie 2006). Core collections, gene banks, and systematic germplasm evaluations play a vital role in identifying key traits, such as disease resistance, stress tolerance, and oil quality, which are often retained in wild relatives and traditional landraces (Ashri 1998).

2.5 Sesame Morphology and Biology

Botanically, sesame (*Sesamum indicum* L.) is a diploid species (2n = 26) with an indeterminate growth habit and extensive phenotypic plasticity. The plant typically reaches 0.5 to 2.5 meters in height and has a taproot system that enables drought tolerance. Its flowers are zygomorphic and predominantly self-pollinated, although cross-pollination can occur via insects. Capsules vary widely in shape, size, and dehiscence level, often influencing yield stability. Cytogenetic studies have divided *Sesamum* species into three categories based on chromosome numbers: 2n = 26, 32, and 64. Techniques such as BAC-FISH and chromosome painting have enabled the mapping of sesame chromosomes and the identification of chromosomal markers (Zhao *et al.* 2018; Zhang *et al.* 2021a). This cytogenetic knowledge supports both classical and molecular breeding approaches.

2.6 Breeding Methods in Sesame

Conventional breeding approaches in sesame include mass selection, pure-line selection, and pedigree-based hybridization. However, a narrow genetic base in some regions often limits breeding gains. To overcome this, mutagenesis (using EMS, gamma rays, or space mutagenesis) has been applied to induce novel variation (Liu *et al.* 2004; Wang *et al.* 2017). Heterosis breeding has been explored, with hybrid cultivars developed to exploit vigor. However, hybrid seed production remains challenging due to a complex flower structure. Interspecific hybridization has also been attempted to introgress traits from wild relatives, though cross-ability barriers exist. Molecular breeding, including marker-assisted selection (MAS), is increasingly used, utilizing SSR, SNP, and InDel markers linked to desirable traits (Zhang *et al.* 2021b).

2.7 Important Traits of Sesame

Improving yield and harvest efficiency is a major breeding objective in sesame. Key yield-related traits include the number of capsules per plant, seed size (e.g., thousand-seed weight), and capsule-shattering resistance. Shattering is a major cause of yield loss and a barrier to mechanized harvesting. Indehiscent or semi-indehiscent capsule traits are under active selection, with the cl1 mutation being the first fully shatter-resistant variant described (Zhang *et al.* 2018b). Other important traits include seed coat color (linked to market preference and antioxidant content), branching habit, and plant height (Cui *et al.* 2021; Duan *et al.* 2021). Biotic stress resistance to pathogens like Phytophthora, Alternaria, and phyllody, and abiotic stress tolerance (e.g., drought, salinity) are also key breeding targets (Islam *et al.* 2016).

2.8 Sesame Genome

The construction of a high-quality reference genome for sesame evolved considerably between 2010 and 2015, with the primary objectives encompassing comprehensive genome sequencing, systematic annotation, and the development of functional genomic resources (Zhang et al. 2021b). The first reference genome for cultivated sesame was published by Wang *et al.* (2014b), based on Illumina sequencing of the *Zhongzhi* 13 variety. Although widely used in genetic studies, this assembly was fragmented and contained numerous gaps. Recently, a significant improvement was achieved using a multi-platform approach combining PacBio HiFi sequencing, Bionano optical mapping, and Hi-C scaffolding technologies. This effort resulted in a high-contiguity, chromosome-level assembly of the *Baizhima* variety (v3 assembly) with 97.54% of the genome anchored to 13 pseudochromosomes and a contig N50 of 13.48 Mb (Wang *et al.* 2022). The updated annotation identified 24,345 high-confidence protein-coding genes and revealed improved structural accuracy and completeness compared to earlier versions, thereby providing a robust genomic resource for sesame breeding and functional genomics. This integrative, multi-platform strategy enabled the construction of a high-quality chromosome-level genome assembly for cultivated sesame, marking a significant milestone in oilseed genomics. Furthermore, these methodological frameworks were subsequently applied to wild *Sesamum* species, enabling comprehensive comparative genomic analyses which elucidated evolutionary relationships within the genus. The resulting genomic resources have provided foundational

support for marker-assisted breeding strategies in sesame improvement programs, accelerating the development of improved cultivars with enhanced agronomic performance (Zhang et al. 2021b).

2.9 Genome-Wide Association and Quantitative Trait Locus (QTL) Mapping in Sesame

GWAS and QTL mapping are pivotal tools for identifying the genomic regions associated with complex traits. In sesame, GWAS using high-density SNP data has revealed loci linked to yield traits, capsule traits, and stress tolerance. For instance, studies have identified candidate genes, such as *SiLPT3* and *SiACS8,* involved in capsule development and seed set (Zhou *et al.* 2018). QTL mapping has elucidated loci for traits like plant height, male sterility, waterlogging resistance, and oil content (Zhang *et al.* 2014; Liu *et al.* 2015). These findings facilitate marker-assisted selection, genomic prediction, and the development of ideotypes tailored to climate-resilient agriculture.

2.10 Historical Development and Genetic Richness of Sesame in Sudan

Sudanese sesame germplasm encompasses substantial genetic diversity, reflected in numerous landraces and local varieties, distinguished by diverse morphological characteristics and seed coat pigmentation patterns (Zhang *et al.* 2021a). Traditional landraces are identified by regional nomenclature, including *Housh*, *Abdel-razig*, *Babinga*, *Bilia*, *Abu-sofa*, *Hurhairy*, and various *Gerabin* types (light and late), exhibiting considerable phenotypic variation in vegetative and reproductive traits. Seed coat coloration ranges from white to black with intermediate phenotypes, with white-seeded cultivars generally preferred for export markets, while pigmented varieties are predominantly utilized for oil extraction.

Indigenous landraces remain prevalent in cultivation systems, particularly throughout the western part of Sudan. These genetic resources were systematically collected during the 1950s and early 1960s, followed by phenotypic selection and purification programs focusing primarily on seed color characteristics (Osman 1985). This effort yielded several improved white-seeded cultivars, including *Zeraa-1*, *Zeraa-3*, *Zeraa-6*, and *Zeraa-7*, which were subsequently distributed to commercial and semi-commercial production

systems. Concurrent efforts to introduce non-shattering traits through introgression of the indehiscent gene (a recessive mutation identified in Venezuela in 1943) from American germplasm proved unsuccessful, with the resulting lines demonstrating inadequate agronomic performance.

A collaborative breeding initiative between the Agricultural Research Corporation (ARC) of Sudan and the University of California, Riverside, established in 1975 with funding from the United Nations Development Programme (UNDP), resulted in the development of the cultivar *Kenana-1* (Osman, 1985). Subsequent varietal releases in the mid-1980s included *Kenana-1*, *Zeraa-9*, and *Hurria-49*. *Kenana-1* exhibits superior yield potential, with large white seeds containing approximately 48% oil and 24% protein, bacterial blight resistance, and distinctive eight-locule capsules. However, susceptibility to webworm has been documented in *Kenana-1*. *Zeraa-9* is characterized by small white seeds and synchronous maturation (Osman 1985).

The 1990s breeding programs produced *Kenana-2*, *Kenana-3*, and *Kenana-4* cultivars. *Kenana-2* features large white seeds, early maturity, and enhanced drought tolerance. Additionally, the cultivar *Promo*, selected from Greek temperate germplasm, demonstrates profuse branching, a medium growth duration, uniform maturity, and delayed capsule dehiscence (Ahmed 2008). In 2003, cultivars *Um-Shagara* and *Gadarif-1* were released; *Gedarif-1*, derived from temperate-tropical hybridization, exhibits determinate growth habit, medium-late flowering, and vigorous growth characteristics, while *Um-Shagara*, selected from introduced-local crosses, features extensive branching, white seeds, and synchronized maturity (Ahmed *et al.* 2003).

More recently, the cultivar *Elgizouli* was developed and released in 2012. It originated through the segregation of populations of crosses between introduced materials sourced from International Development Research Centre projects and indigenous cultivars. This variety exhibits extensive branching, elongated capsules, white seeds, and uniform maturation characteristics (Khalafalla & Ahmed 2012).

3. Aims and Objectives of the Research

This PhD thesis aims to advance the genetic and biochemical understanding of key agronomic, nutritional, and postharvest traits in Sudanese sesame, with a focus on supporting the development of resilient, high-quality cultivars adapted to semi-arid environments. Through the integration of high-resolution field phenotyping, biochemical profiling, genome-wide association studies (GWAS), gene analysis, and genomic prediction, this research aims to elucidate the genetic architecture underlying key traits, including oil content and composition, seed coat pigmentation, capsule-shattering behavior, and seed storage stability. By exploiting the genetic diversity present in Sudanese germplasm, a largely underutilized yet globally significant resource, the study aims to identify molecular markers and candidate genes which can be applied in breeding programs to improve yield stability, nutritional quality, and postharvest performance in sesame.

The specific objectives of this research are:

- The assessment of genotypic variation in sesame seed quality and oil stability under accelerated storage conditions to identify germplasm with enhanced storability for hot climate production systems.
- Characterization of the variation in oil content, fatty acid profiles, and seed coat color in Sudanese sesame germplasm to identify elite genotypes for breeding programs, targeting enhanced nutritional and market value.
- Characterization of the genetic basis of seed coat color to facilitate breeding of market-preferred sesame varieties.
- Elucidation of genetic factors associated with oil content and fatty acid composition profiles.
- Identification of the genetic determinants of capsule-shattering traits to improve the yield of sesame.

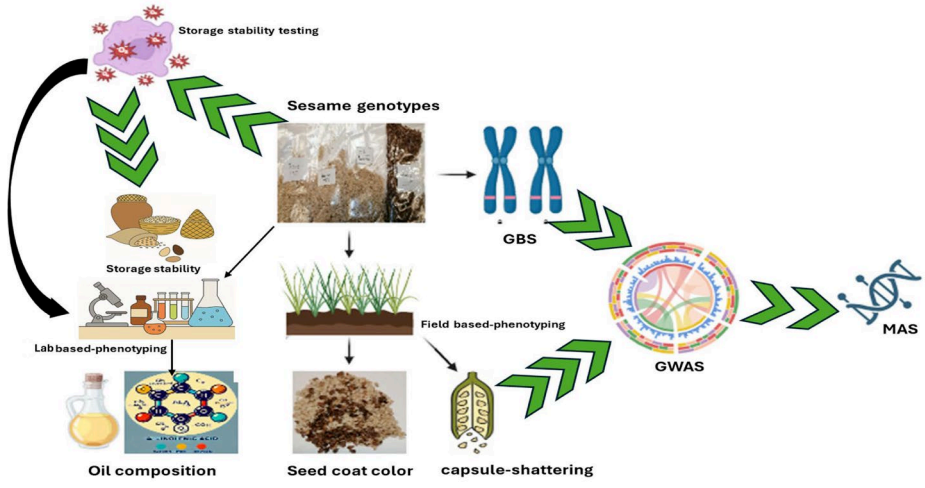


Figure 2. Graphical abstract of the study objectives

4. Materials and Methods

4.1 Plant Material

The studies involved diverse collections of sesame germplasm. Paper I employed ten Sudanese sesame genotypes, including widely cultivated cultivars (*Kenana-2* (KN), *Gadarif* (GF), *Eltayeb* (EB), *Tagarub* (TB), *Bromo* (BO)) and landraces (*Radoum* (RM), *Southern Kordofan* (SK), *Rufaee* (RE), *Hurhairy* (HR), and *Abusundoug* (AS), selected for their diverse market profiles and nutritional traits. Papers II, IV, and V utilized the same panel of 200 genetically diverse Sudanese sesame accessions, comprising genebank accessions, advanced breeding lines, and released varieties. Paper III (Oil Content and Fatty Acid Profiles) evaluated 87 accessions from the Agricultural Plant Genetic Resources Conservation and Research Centre (APGRC) representing nine regions of Sudan, with *Kenana-2*, *Promo*, and *Herheri* as checks.

4.2 Experimental Locations

Field experiments were conducted primarily at:

- Matuq Research Station (Gezira State; 14°11′10″N, 32°34′48″E) for Papers II, III, IV, and V.
- Abu Naama Research Station (Sennar State; 12°44′43″N, 34°07′21″E) for Paper IV.

Experiments employed augmented block designs (ABD) to evaluate genotypes, and agronomic practices were carefully standardized across experiments.

4.3 Phenotyping and Trait Evaluation

4.3.1 Storage Stability Assessment (Paper I)

Sesame seeds were exposed to accelerated storage conditions (55°C, 60% RH) for 0, 16, and 32 days, simulating commercial storage of 6–12 months (Taoukis *et al.* 1997). The seed quality attributes measured included germination, oil content (AOAC 1990), fatty acid profiles via GC–FID (Ivarson *et al.* 2017), fungal growth (Feldsine *et al.* 2002), water activity (aw), free fatty acids (AOAC 1990), and peroxide value (Østdal *et al.* 2000).

4.3.2 Phenolic Compounds and Antioxidant Activity (Paper I)

The total phenolic and flavonoid extraction was conducted according to Talhaoui *et al.* (2014). Phenolics were quantified using the Folin–Ciocalteu assay, as outlined by Waterhouse (2002), and flavonoid content was determined using the aluminum chloride colorimetric method (Kim *et al.* 2003). Regarding antioxidant activity assays, DPPH radical scavenging activity, ferric-reducing antioxidant power, and ABTS radical scavenging activity were measured according to (Re *et al.* 1999; Chang *et al.* 2001; Benzie & Devaki 2018).

4.3.3 Seed Coat Color Analysis (Papers I, II, and III)

Seed coat color parameters were quantified using a Minolta Chroma Meter CR-400 in CIELab space (L*, a*, b*), calibrated against a standard white plate. Each sample (~50 g) underwent three replicate measurements to ensure precision.

4.3.4 Oil Content and Fatty Acid Profiling (Papers III & IV)

Oil extraction was performed using Soxhlet extraction with hexane solvent. Fatty acid methyl esters (FAMEs) were analyzed using gas chromatography (GC–FID, Agilent Intuvo 9000) (Tesfaye *et al.* 2024).

4.3.5 Capsule Shattering Morphological Traits (Paper V)

Three capsule traits, bicarpellate capsule shape (BS), capsule beak type (TCB), and shattering type (ST), were visually assessed using standardized descriptors adapted from IBPGR (2004) (Table 1). Evaluations were conducted on five randomly selected mature plants per accession across three growing seasons (2021–2023).

Table 1. Qualitative traits and their categorical descriptions in sesame capsule morphology and shattering behavior.

Traits	Description
Bicarpellate capsule shape (BS)	(1) Tapered at apex
	(2) Narrow oblong
	(3) Broad oblong
	(4) Square
Type of capsule beak (TCB)	(1) Short
	(2) Long
	(3) Curved
	(4) Cleft
	(5) Other
Shattering type (ST)	(1) Non-shattering
	(2) Partially shattering
	(3) Completely shattering

4.4 Genotyping-by-sequencing

Young leaf tissues (~5 mm diameter) were collected from each genotype for genomic DNA extraction using the Qiagen BioSprint 96 DNA Plant Kit. Genotyping-by-sequencing libraries were prepared following the double-digest restriction enzyme protocol with PstI and MspI enzymes (Poland *et al.* 2012). Sequencing was performed using the Illumina NovaSeq 6000 platform at the University of Minnesota Genomics Center. Bioinformatics processing included alignment with BWA v0.7.4 to the sesame genome, SNP calling with SAMtools and bcftools, and filtering SNPs with MAF $\geq 3\%$ and $\leq 20\%$ missing data.

4.5 Candidate Gene Analysis (Papers II, IV, and V)

Candidate genes associated with significant SNP markers identified through GWAS were annotated using functional genomics databases, including the National Center for Biotechnology Information (NCBI), and comparative analyses with model species, such as *Arabidopsis thaliana* and *Brassica napus*. Genes linked to pigment biosynthesis pathways (Paper II), lipid metabolism

(Paper IV), and capsule-shattering mechanisms (Paper V) were identified and functionally characterized based on homology and domain predictions, facilitating targeted marker-assisted selection.

4.6 Statistical and Genomic Analyses

Statistical analyses employed ANOVA, broad-sense heritability calculations, and correlation assessments using R software. Genome-wide association studies (GWAS) utilized GAPIT software, employing BLINK, FarmCPU, and MLMM models to detect marker-trait associations. Genomic predictions were executed via *rrBLUP* models. Population structure and genetic diversity were assessed through PCA and ADMIXTURE analyses, evaluating linkage disequilibrium and relatedness among accessions.

Broad sense heritability (H) for oil composition (Paper IV) was calculated by applying 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.

In paper V, the genetic evaluation, BLUEs (Best Linear Unbiased Estimates), is obtained using the following model:

$$Y_{ijk} = \mu + G_i + E_j + (GxE)_{ij} + B_k + \epsilon_{ijk}$$

Where: $\boldsymbol{Y_{ijk}}$ is the trait value for genotype *i* in environment *j*, block *k,* $\boldsymbol{\mu}$ is the overall average trait value across all genotypes and environments, $\mathbf{G_i}$ is a fixed effect of genotype *i*, $\mathbf{E_j}$ is a fixed effect of environment, $\mathbf{(GxE)_{ij}}$ is a random G×E interaction, $\mathbf{B_k}$ is a random effect of block, and $\boldsymbol{\epsilon_{ijk}}$ is the residual error.

The BLUPs (Best Linear Unbiased Predictions) were estimated using a model that treated genotype as a random effect.

$$Y_{ik} = \mu + G_i + B_k + \epsilon_{ijk}$$

Where: $\mathbf{G_i} \sim N(0, \sigma^2_G)$ is a random genetic effect, $\mathbf{B_k} \sim N(0, \sigma^2_B)$ is the random block effect, and $\boldsymbol{\epsilon_{ik}} \sim N(0, \sigma^2_e)$ is the residual error.
Broad-sense heritability was calculated using Cullis *et al.* (2006). The broad-sense heritability was calculated using the following:

$$H^2 = 1 - \frac{mean\ PEV}{2\sigma^2_G}$$

Where **PEV** is the average prediction error variance from the matrix of squared standard errors of pairwise differences between genotype predictions, $\boldsymbol{\sigma^2_G}$ is the genetic variance component (from the BLUP model), and the factor of **2** adjusts for the average variance of pairwise genotype differences.

5. Results and Discussion

5.1 Storage Stability Assessment (Paper I)

5.1.1 Seed Quality Characteristics

The sesame seeds maintained germination rates across all genotypes and storage durations (Table 1, Paper I), demonstrating high viability under accelerated storage conditions ($p > 0.05$). This observation aligns with prior studies indicating sesame seeds' robustness in retaining germination rates exceeding 80% post-storage (de Lima *et al.* 2014).

Significant variation was observed in seed color parameters (L*, a*, b*, ΔE), influenced by storage durations, cultivar types, and their interactions ($p < 0.001$). Accelerated storage for 32 days notably increased seed coloration parameters compared to 16-day storage (Table 1, Paper I), which initially reduced coloration. Color degradation at 16 days was likely caused by pigment loss and browning reactions, whereas prolonged storage (32 days) induced pigment and melanoidin formation from sustained heat-induced chemical reactions (Lattanzio 2003). Cultivars exhibited inherent color differences, with brown landrace genotypes (AS, HR) demonstrating higher stability in coloration compared to whitish-released cultivars. Previous reports associate such color changes with non-enzymatic browning connected with phenolic compounds and tannins (Reyes-Moreno *et al.* 2000).

Accelerated storage significantly impacted oil content, decreasing total oil and oleic acid levels while increasing linoleic acid content (Table 2, Paper I). This selective oxidation is attributed to the lower activation energy required by unsaturated fats to form lipid radicals compared to saturated fats, enhancing susceptibility to oxidative degradation (Choe & Min 2006). This is consistent with prior findings emphasizing the oxidative vulnerability of unsaturated fatty acids, particularly under high temperature and humidity conditions (Lee & Choe 2012).

5.1.2 Storability Characteristics

Storage conditions significantly influenced fungal growth, water activity (aw), free fatty acids (FFA), and peroxide values (PV), with all parameters exhibiting a notable increase during storage ($p < 0.001$) (Table 3, Paper I). Fungal growth and PV notably increased over time, particularly after prolonged storage periods, reflecting higher oxidative rancidity and microbial contamination under accelerated conditions (Kumar *et al.* 2019). Interestingly, the FFA content peaked at 16 days, likely due to decreased lipase enzyme activity caused by prolonged heat exposure (Zhou *et al.* 2002). The sesame seeds' oxidative stability, indicated by peroxide values, demonstrated rapid deterioration, emphasizing the need for optimal storage conditions to maintain seed quality, as also identified in a previous study (O'brien 2008). Genotypic differences were apparent, with some genotypes (EB, KN, HR) exhibiting higher fungal growth. In contrast, others (RM, TB, SK, BO, RE) showed resistance, indicating inherent genetic variations that affect storability traits.

5.1.3 Phytochemical content and antioxidant activity

Accelerated storage significantly reduced total phenolic compounds (TPC), flavonoid content (TFC), ferric-reducing antioxidant power (FRAP) and ABTS radical scavenging activities ($p < 0.001$), although DPPH activity remained unaffected (Table 4, Paper I). The observed reduction in phytochemicals and antioxidant activities aligns with reports from other crops under similar stress conditions, suggesting chemical degradation and oxidative reactions induced by high-temperature storage (Lü *et al.* 2010; Bragança *et al.* 2020). The retained DPPH activity suggests a persistent radical scavenging capacity even after the degradation of other antioxidants (Chang *et al.* 2001).

Genotypic variation was also significant, with the genotype *Abusundoug* (AS) notably exhibiting higher phytochemical content and antioxidant capacities, suggesting its potential for better storage stability and oxidative stress resistance compared to other genotypes.

5.1.4 Multivariate Analysis and Genotypic Implications

Principal Component Analysis (PCA) effectively differentiated genotypes based on phytochemical content, fatty acid profiles, and storage duration (Figure 1A, Paper I). High fatty acid content is inversely correlated with phytochemical concentrations and antioxidant capacities (Figure 1B, Paper I). Moreover, hierarchical clustering revealed that genotype AS demonstrated exceptional stability under storage conditions, correlating with its higher antioxidant capacity and lower fatty acid content (Figure 1C, Paper I). In contrast, the genotype HR showed susceptibility to oxidative deterioration, aligning with its higher fatty acid profile. These insights underscore significant genotype-dependent responses to storage conditions, highlighting the importance of identifying and utilizing genetic resources, such as genotype AS, to breed for enhanced storage stability (Laurentin & Karlovsky 2006; Zhou *et al.* 2022). This study provides critical insights into the storability characteristics of sesame, underscoring genotype-specific responses to accelerated storage conditions. The genotype *Abusundoug* (AS) emerged as a promising candidate for breeding programs due to its superior storage stability linked to its phytochemical richness and lower polyunsaturated fatty acid content. Future research should extend this evaluation to broader

genetic material, examine genetic factors underpinning these characteristics, and test stability across diverse environmental conditions.

5.2 Genetic Scanning of Seed Coat Color in Sesame (paper II & III)

5.2.1 Seed Coat Color Phenotyping (Paper II)

The analysis of CIELab parameters (L*, a*, and b*) across 200 Sudanese sesame accessions over two years showed high consistency ($r = 0.997$, $p < 0.001$), indicating strong genetic control over seed coat color traits. Lightness (L*) values showed the greatest range, confirming broad phenotypic variation (Paper II). Moderate correlations between L* and a* ($r = -0.42$), L* and b* ($r = 0.37$), and a* and b* ($r = 0.47$) indicate interconnected pigment profiles (Figure 1A, Paper II). Principal component analysis (PCA) accounted for 88.4% of the total variance, revealing three distinct phenotypic groups (high L*, high a*, high b*), with notable overlap between light and yellow seeds (Figure 1B, Paper II). Broad-sense heritability values were exceptionally high (0.9991 for L*, 0.9975 for a*, and 0.9974 for b*), suggesting minimal environmental influence. These findings align with previous QTL mapping studies (Wang *et al.* 2016; Du *et al.* 2019; Cui *et al.* 2021), which demonstrates that seed coat color is largely controlled by a limited number of stable loci, making it suitable for straightforward selection.

5.2.2 Seed Coat Color Variation and Its Implications (Paper III)

Seed coat color parameters (L*, a*, b*) exhibited significant variation. L* values ranged from 34.1 to 79.8, confirming broad pigmentation diversity (Supplementary data, Paper III). However, regression analyses revealed no significant association between oil content and seed coat color, contrasting with findings from Wang *et al.* (2020) and supporting the notion that seed pigmentation and oil biosynthesis may be independently regulated. Despite farmers' perceptions linking darker seeds to higher oil content (ETI 2023), our data does not support this association.

Nonetheless, dark seeds have been associated with higher lignan and antioxidant content in prior studies (Dossou *et al.* 2021; Comini *et al.* 2023) and may offer

health benefits, such as reduced oxidative stress and inflammation (Bolvig *et al.* 2017; Pilar *et al.* 2017). Thus, seed coat variation remains relevant for both nutrition and market preferences.

5.2.3 Genomic Diversity and Structure Analysis (Paper II)

Whole-genome SNP analysis showed heterogeneous variant distributions across the 13 chromosomes, which were particularly high for chromosomes 5, 7, and 10 (Figure 2a, Paper II). ADMIXTURE and DK methods identified two main genetic clusters among the sesame accessions (Figure 2b, Paper II). The kinship heatmap revealed clear blocks of relatedness (Figure 2c, Paper II), suggesting shared ancestry or breeding histories (Eynard *et al.* 2016). Linkage disequilibrium (LD) decay was rapid, with r^2 dropping below 0.1 at ~204 kb, confirming high recombination (Figure 2d, Paper II); this was consistent with findings in other sesame populations (Du *et al.* 2019; Wang *et al.* 2022) compared to earlier studies that used older genome assemblies (Wang *et al.* 2014b). This analysis, using the updated reference genome, allowed for a more accurate resolution of LD decay, supporting its utility for fine mapping and genome-assisted breeding.

5.2.4 GWAS and SNP-trait Associations (Paper II)

The GWAS identified significant SNPs for all three-color traits. For L*, key associations were found for chromosomes 3, 6, and 12. The strongest SNPs on Chr12 (*16523829* and 1*6523899* bp) explained over 6.5% of the phenotypic variance (Table 1, Paper II). This agrees with Wang *et al.* (2020) and Li *et al.* (2021), who linked similar regions to light pigmentation. For a*, SNPs on chromosomes 3 and 6 were significant, particularly *Chr6_27694080* (9.2% variance), corresponding with loci for red pigmentation and flavonoid pathways (Wang *et al.* 2023). For b*, novel associations were identified on chromosomes 9 and 13. Notably, Chr13 SNPs at positions 345249 and 345322 bp explained over 6% of phenotypic variance, highlighting the unique adaptation in Sudanese germplasm, which has not been captured previously by the use of Chinese or Indian panels (Cui *et al.* 2021; Dutta *et al.* 2022).

5.2.5 Candidate gene discovery (Paper II)

Candidate genes were annotated near significant SNPs based on the LD decay range (204,890 bp). *APMJ01001391* (*DOF3*.1 zinc finger protein) on Chr3 was linked to a* (Table 2, Paper II), which in previous studies has been found to potentially regulate flavonoid biosynthesis under light stimuli (Iorizzo *et al.* 2019). *APMJ01003628* on Chr6 have previously been shown to encode *histidine-containing phosphotransfer* protein (HPT), linked to cytokinin signaling and L* variation (Cortleven & Schmülling 2015). Furthermore, the *KAK4407764* and *APMJ01001731* (*STY8* kinases) modulate carotenoid metabolism and plastid biogenesis (Mazur *et al.* 2021). *WRKY23* (*APMJ01003151*) contributes to anthocyanin regulation under stress (Chen *et al.* 2019; Meraj *et al.* 2020). *APMJ01007050* (*SABP2*) and *APMJ01006505* (*SBP1*) were associated with b* and L*, respectively, implicating salicylic acid signaling and light-mediated gene regulation in pigmentation (Sánchez-Retuerta *et al.* 2018; Shaukat *et al.* 2022).

5.2.6 Implications for Breeding (Paper II)

The high heritability and strong SNP-trait associations of seed coat color traits underscore their value in breeding. The GWAS findings is supporting the use of marker-assisted selection for color traits that also relate to oil quality and stress resistance. The candidate genes identified here provide a foundation for functional validation and gene editing. Sudanese sesame, as a center of diversity, offers unique allelic variants with potential adaptation advantages. These insights promote targeted breeding strategies to meet both nutritional and market demands.

5.3 Diversity in Oil Content, Fatty Acid Profiles (Paper III and IV)

5.3.1 Variations in Oil Content and Fatty Acid Profiles (Paper III)

Sudanese sesame accessions showed broad diversity in oil content, ranging from 32.8% to 50.2%, with a mean of 41.5% (Supplementary data, Paper III). This range aligns with earlier studies by Mondal *et al.* (2010), Uzun *et al.* (2008), Were *et al.* (2006), and Kurt (2018), reinforcing the presence of high oil-yield

potential within Sudanese germplasm. Among the fatty acids, oleic acid (18:1) ranged from 41.3% to 47.6%, and linoleic acid (18:2) from 35.0% to 41.4%, dominating the unsaturated fraction. Stearic (18:0), palmitic (16:0), arachidic (20:0), and gadoleic (20:1) acids contributed to the saturated fraction, albeit in smaller proportions (Supplementary data, Paper III).

Notably, the high oleic acid levels observed in several accessions exceed ranges reported in Turkey, Kenya, Tanzania, and Uganda, and one commercial variety from India (Were *et al.* 2006; Uzun *et al.* 2008), underlining their potential for oxidative stability and nutritional enhancement (Bahkali *et al.* 1998; Kamdar *et al.* 2021). This trait is highly relevant to human health and functional food applications. Furthermore, the high triacylglycerol (TAG) and glyceride (Gly) content found in some accessions may influence shelf stability and processing quality (Zeb & Ahmad 2017; Saini *et al.* 2020).

5.3.2 Correlation and Multivariate Analyses

Strong positive correlations were found between oil content and total TAG ($r = 0.91$, $p < 0.01$) and between oleic and stearic acids (Figure 2, Paper III), while oleic and linoleic acids showed a strong negative correlation ($r = -0.88$, $p < 0.01$), consistent with findings from Kurt (2018) and Relina *et al.* (2022). This inverse relationship reflects the enzymatic desaturation pathway mediated by FAD2 (Wei et al., 2015). Minor fatty acids, such as linolenic acid (18:3), were positively associated with linoleic acid, supporting coordinated biosynthesis.

Principal Coordinate Analysis (PCoA) did not show distinct clustering by geographical origin, except for a loose cluster among South Kordofan accessions (Figure 3, Paper III). This suggests the historical movement and trade of sesame seeds across Sudan, which may have blurred region-specific differentiation (Basak *et al.* 2019). Accessions from North Kordofan and Gedarif displayed unique profiles, indicating possible outliers with breeding potential.

5.3.3 Breeding Perspectives (Paper III)

The study highlights the untapped potential of Sudanese sesame in enhancing oil quality and composition. Accessions with high oleic acid and desirable seed coat types offer targets for breeding programs. Moreover, understanding the genetic basis of these traits, including the possible roles of FAD2 and other genes involved in lipid metabolism, will enhance selection efficiency.

Given that traits like fatty acid composition are environment-sensitive (Zahran *et al.* 2020), multi-environment trials and genome-assisted selection should be prioritized. Sudan's germplasm, despite being underutilized globally, holds key alleles for resilience and nutritional traits. Their conservation and integration into breeding pipelines is crucial amid climate stress and shifting market demands. This study highlights the considerable diversity in oil content, fatty acid profiles, and seed coat color in Sudanese sesame. It demonstrates the potential for the genetic improvement of oil quality traits and debunks the presumed correlation between dark seed color and oil richness. The findings promote evidence-based breeding strategies and advocate for deeper studies of genetic and environmental interactions to harness the full potential of Sudanese sesame.

5.3.4 GWAS Study of Oil Content and Fatty Acid Composition (Paper IV)

5.3.4.1 Phenotypic Variation in Oil and Fatty Acid Composition

Significant variability was observed in oleic acid, linoleic acid, and total oil content among the 200 Sudanese sesame genotypes evaluated in two contrasting environments, Abu Naama and Matuq (Table 1, Paper IV). Oleic acid levels ranged from 3.07 mg/10 seeds (32%) to 9.00 mg/10 seeds (48.6%) at Abu Naama, with a mean of 5.34 mg/10 seeds (40.5%), whereas values at Matuq ranged from 3.10 mg/10 seeds (36%) to 8.35 mg/10 seeds (49%), averaging 5.78 mg/10 seeds (44%). Linoleic acid content also displayed considerable variation, ranging from 3.28 mg/10 seeds (32%) to 9.12 mg/10 seeds (51%) at Abu Naama (mean = 5.54 mg/10 seeds, 42%) and from 2.76 mg/10 seeds (33%) to 8.22 mg/10 seeds (46%) at Matuq (mean = 5.02 mg/10 seeds, 38.5%). The total oil content varied greatly across genotypes, with higher average values recorded at Abu Naama (mean = 44.45%; range = 33.14% to 62.05%) compared to Matuq (mean = 42.06%; range = 30.95% to 50.27%). These findings highlight the substantial genotypic diversity in fatty acid composition and oil accumulation under different environmental conditions. These ranges are consistent with prior studies (Uzun *et al.* 2008; Mondal *et al.* 2010; Wei *et al.* 2015) and underscore the potential of Sudanese germplasm to improve oil quality.

Despite considerable trait variability, correlation analysis revealed weak relationships between environments for oleic and linoleic acids (r = 0.03 and -0.06, respectively), emphasizing the significant role of genotype-by-environment

interactions. This result aligns with Kurt *et al.* (2016) and Hu *et al.* (2022), who reported environmental sensitivity in oil biosynthesis traits.

Broad-sense heritability was moderate for oil and oleic acid (H = 0.28) but negligible for linoleic acid, indicating a strong environmental effect. Similar findings have been observed in other oil crops (Holland *et al.* 2003; Uzun & Çağırgan 2006; Khan & Nawaz 2022), reinforcing the need for multi-environment testing and environment-aware selection strategies in sesame breeding.

5.3.4.2 Genetic Diversity and Linkage Disequilibrium

PCA and structure analyses revealed substantial population stratification, with PC1 capturing 23.3% of the total genetic variance and clustering separating GenBank accessions from other types (Figure 2, Paper IV). Structure analysis supported two subpopulations (K = 2) (Figure 3A-B, Paper IV), indicating historical selection and domestication patterns (Mohamed 2011; Parry & Hawkesford 2012).

The kinship matrix revealed diverse genetic relationships (Figure 4A, Paper IV), indicating suitability for genome-wide association studies (GWAS). LD decay (Figure 4B, Paper IV) was estimated at ~204,890 bp ($r^2 = 0.1$), longer than earlier studies (Wang *et al.* 2014a; Tesfaye *et al.* 2022; Seay *et al.* 2024), likely due to differences in marker density, genome annotations, and sample structure. Nevertheless, this LD decay supports resolutions sufficient for candidate gene identification.

5.3.4.3 GWAS Analysis and Genomic Prediction

Multi-model GWAS (FarmCPU, Blink, MLMM) identified numerous SNPs associated with oil content, oleic acid, and linoleic acid (Table 2, Paper IV). Notably, *Chr1_1693157* and *Chr5_17024932* were consistently linked to oleic acid in Abu Naama, while *Chr3_23284702* and *Chr9_1711873* were associated with oil content across environments. These loci correspond with previous findings in chromosomes 3, 5, and 9 (Wei *et al.* 2013; Li *et al.* 2014; Zhao *et al.* 2022).

GWAS signals showed modest effect sizes (R^2 = 3–23%) and low minor allele frequencies (MAF = 0.06–0.49), consistent with the polygenic and quantitative

nature of these traits. Linoleic acid loci were environment-specific, supporting gene-by-environment (G × E) interactions.

Genomic prediction using *rrBLUP* yielded moderate predictive abilities (Figure 6, Paper IV). Linoleic acid at Matuq and oil content at Abu Naama had the highest mean predictive abilities (0.21 and 0.22, respectively). Oleic acid had low predictability (0.07 at Abu Naama), suggesting trait and environment specific model performance. These results confirm findings by Crossa *et al.* (2017) and Chen *et al.* (2015), supporting genomic selection as a useful, though environment-sensitive, tool.

5.3.4.4 Candidate Genes Associated with Oil Traits

Candidate genes were identified within 409 kb windows of significant SNPs. For oleic acid (Table 3, Paper IV), *APMJ01003105* (*Chr5_17024932*) encodes 3-oxoacyl-(ACP) synthase 3A (KAS III), an enzyme involved in fatty acid elongation (Berg *et al.* 2015; Guo *et al.* 2019). *APMJ01000051* (*Chr1_1693157*) matched OFP8, a transcription repressor potentially regulating lipid biosynthesis. For oil content, *APMJ01001210* (*Chr3_23284702*) encoded *HVA22-like* protein A, implicated in membrane trafficking and stress response (Chen *et al.* 2002), while *APMJ01005016* (*Chr9_1711873*) corresponded to PDAT2, an enzyme catalyzing triacylglycerol (TAG) biosynthesis (Dahlqvist *et al.* 2000). These candidate genes align with pathways highlighted in multi-omics analyses of fatty acid regulation (Zhou *et al.* 2022).

5.4 Capsule-shattering Genomic Association (Paper V)

5.4.1 Phenotypic Variation and G×E

Analysis of 200 sesame genotypes over three years revealed distinct distributions in capsule-related traits: bicarpellate capsule shape (BS), type of capsule beak (TCB), and shattering type (ST). The narrow oblong BS was most common, while ST showed a significant temporal shift; complete shattering increased from 69 accessions in Year 1 to 129 in Year 3, indicating environmental sensitivity (Figure 1, Paper V). The non-shattering phenotype was absent, necessitating binary classification of ST. Principal Component Analysis (PCA) and Multiple Correspondence Analysis (MCA), which confirmed ST and TCB as primary contributors to phenotypic variation (Figure 2A-B, Paper V). ST exhibited a

significant genotype-by-environment (G×E) interaction, with pronounced crossover behavior among genotypes, indicating strong environmental influence (Figure 3 Paper V). In contrast, BS was stable, and TCB showed moderate year-to-year variation. These results are consistent with findings by Zhang *et al.* (2018a) and Teboul *et al.* (2022), who also described ST as an environmentally responsive trait with moderate heritability.

5.4.2 Genetic Structure and Linkage Disequilibrium

PCA on SNP genotypes revealed three partially overlapping populations, suggesting moderate stratification due to gene flow and shared ancestry (Figure 4A, Paper V). Kinship analysis confirmed subclusters of related genotypes (Figure 4B, Paper V), while LD decay dropped below $r^2 = 0.2$ at ~218 kb (Figure 4C, Paper V). This LD distance is longer than the 166 kb observed in the USDA collection (Seay *et al.* 2024), possibly due to the narrower genetic base of the Sudanese germplasm compared to the larger collection of the USDA. The moderately fast decay indicates the population is well-suited to high-resolution GWAS.

5.4.3 GWAS Identifies Key Loci for Shattering Traits

Three GWAS models (BLINK, FarmCPU, MLMM) identified significant SNP-trait associations (Table 2, Paper V). Five SNPs were significant across all models: *Chr2_15649330* and *Chr8_31466064* for ST, and *Chr1_19419575*, *Chr8_19392181* and *Chr8_30292484* for TCB. These markers explained up to 39% of phenotypic variance. *Chr8_31466064* was part of a tightly linked haplotype cluster, suggesting a potential QTL block. These findings build on and expand previously identified loci, like *S8_5062843* (Yol *et al.* 2021) and *SiNST1* (Ju *et al.* 2024), but also introduce novel loci not reported before, such as *Chr2_15649330* and *Chr1_19419575*.

Haplotype analysis of *Chr2_15649330* (Figure 6A, Paper V) and *Chr8_31466064* (Figure 6B, Paper V) showed clear phenotypic differentiation, with Tukey's HSD tests confirming significant differences among alleles ($p < 0.01$). These loci are strong candidates for marker-assisted selection (MAS).

5.4.4 Candidate Genes and Functional Annotation

From 273 genes within 437 kb windows surrounding the key SNPs, 25 Arabidopsis orthologs with functions linked to pod shattering were identified (Figure 7a, Paper V). These include regulators of cell wall modification (*PLL2*, *ARAF*), hormone signaling (*COR27*, *OST1*), programmed cell death (*ATG2*, APX3), and stress responses (*RZF1*, *FLZ3*, *MKK5*) (Figure 7b, Paper V). BLASTP and expression analysis in *Brassica napus* revealed that *FLZ3*, *RZF1*, *MKK5*, and *COR27* exhibit expression patterns associated with pod-shattering resistance (Figure 8, Paper V). In particular, *FLZ3* presented downregulation at late pod stages in resistant lines, supporting its involvement in delayed dehiscence. *RZF1* and *MKK5* displayed differential expression between resistant and susceptible genotypes, consistent with roles in strengthening pod structures or delaying shattering. These expression patterns corroborate the functional roles previously described in other crops (De Zélicourt *et al.* 2016; Afridi *et al.* 2022; Xiao *et al.* 2024).

6. Conclusions

Sudanese sesame germplasm exhibits exceptional genetic diversity with great potential for global breeding. By applying integrated genomic and multi-environment phenotypic approaches, this work revealed significant variations in key traits, including oil content, fatty acid composition, seed coat color, and capsule-shattering resistance. Genotype-by-environment interactions were notable for oil quality and shattering, while seed coat color showed high heritability. Multi-model GWAS identified robust SNPs and novel candidate genes, including *FLZ3*, *MKK5*, and *COR27*, validated through genomic prediction, haplotype analysis, and transcriptomics. Notably, landrace *Abusundoug* displayed superior storage stability, attributed to its low polyunsaturated fatty acid content and high phytochemical content. Despite Sudan's role as a center of sesame diversity, its germplasm remains underutilized in global breeding due to geopolitical constraints. Harnessing this resource is critical for developing climate-resilient, nutritionally enhanced sesame varieties. Future efforts should focus on trait validation, marker-assisted selection, and breeding for improved yield, quality, and shelf life.

7. Future Perspectives

This thesis highlighted the untapped potential of Sudanese sesame germplasm and demonstrated the power of integrating genomics, phenotyping, and environmental analysis to accelerate sesame improvement. Building on these findings, future research should pursue several key directions to deepen biological insights and translate discoveries into breeding impact.

7.1 Conservation and Valorization of Landraces

The superior performance of certain landraces, such as *Abusundoug*, highlights the importance of in-situ conservation and the targeted utilization of indigenous germplasm. These genetic resources are crucial for preserving adaptive traits and enhancing the nutritional and economic value of crops.

7.2 Expansion of Germplasm Panels and Environments

The current panel, though diverse, represents a fraction of the sesame diversity found in Sudan and globally. Expanding GWAS to include broader African, Asian, and introgressed germplasm, coupled with multi-location trials, would strengthen the resolution of marker-trait associations and reveal genotype-by-environment interactions critical for climate resilience.

7.3 Genomic Selection and Predictive Breeding

The integration of genomic prediction models with field performance data opens new opportunities for rapid, data-driven selection. Future breeding pipelines should incorporate genomic selection for complex traits, such as oil composition, antioxidant capacity, and shelf life, to shorten breeding cycles and increase genetic gain.

7.4 Development of Climate-resilient and Mechanization-ready Cultivars

Key discoveries, such as SNP loci associated with capsule shattering on chromosomes 1 and 2, offer a foundation on which to develop non-shattering, high-yielding varieties suitable for mechanized harvesting. Coupled with traits like drought tolerance and storage stability, these cultivars will better meet the demands of smallholder and commercial production systems under climate stress.

7.5 Validation and Deployment of SNPs as KASP for MAS

Significant SNPs identified through GWAS should be validated and converted into user-friendly markers, such as Kompetitive Allele-Specific PCR (KASP) assays, to move from association to application. These markers provide a cost-effective and high-throughput method for screening breeding populations. Future work should prioritize converting the most robust SNPs, such as those linked to capsule shattering, oil content and seed coat color, into KASP markers. Validation across independent populations and diverse environments will be crucial to confirming the reliability of the marker. Once validated, these markers can be integrated into marker-assisted selection pipelines to accelerate the development of improved sesame cultivars with targeted traits.

7.6 Functional Validation of Candidate Genes

Although candidate genes associated with traits such as capsule shattering, seed pigmentation, and oil stability have been identified, their biological functions remain to be experimentally confirmed. Future studies should employ gene expression profiling, knockout/knockdown approaches (e.g., CRISPR/Cas9), and transgenic validation to establish causal links between candidate genes and phenotypic traits.

7.7 Seed System Integration and Farmer-centered Innovation

Future efforts must bridge the gap between research and practice to maximize the impact of genetic improvements. Participatory breeding, seed system

development, and agronomic packages tailored to local production systems will be essential to ensure adoption and impact, particularly in underserved regions.

7.8 Market Segmentation and Trait-targeted Breeding

Future sesame improvement should be guided not only by agronomic performance but also by market-driven trait preferences. Consumer demand varies widely across regions; for example, white seeds are preferred for export and bakery use, while pigmented seeds are valued in health food markets for their antioxidant properties. Oil composition, flavor, shelf life, and seed uniformity influence market segmentation. Integrating these preferences into breeding programs will require close collaboration between breeders, processors, and market actors. Targeted breeding for specific market classes, such as high-oleic oil types, antioxidant-rich black sesame, or non-shattering, white-seed varieties for mechanical harvesting, will enhance value chain integration and commercial viability.

8. References

Afridi, M., Ahmad, K., Malik, S.S., Rehman, N., Yasin, M., Khan, S.M., Hussain, A. & Khan, M.R. (2022). Genome-wide identification, phylogeny, and expression profiling analysis of shattering genes in rapeseed and mustard plants. *Journal of Genetic Engineering and Biotechnology,* 20(1), 124.

Ahmed, M. (2008). Evaluation of new sesame (Sesamum indicum L.) genotypes for yield, yield components and stability. *University of Khartoum Journal of Agricultural Sciences (Sudan),* 16(3).

Ahmed, M., Taha, M. & Khalafalla, A. (2003). Assessment of new promising sesame cultivars for the central rain lands and irrigated systems. *Variety Release Committee of Sudan*.

AOAC (1990). Official methods of analysis. *Assoc Anal Chem,* 62, 2742-2744.

Ashri, A. (1998). Sesame breeding. *Plant breeding reviews,* 16, 179-228.

Bahkali, A., Hussain, M. & Basahy, A. (1998). Protein and oil composition of sesame seeds (Sesamum indicum, L.) grown in the Gizan area of Saudi Arabia. *International journal of food sciences and nutrition,* 49(6), 409-414.

Basak, M., Uzun, B. & Yol, E. (2019). Genetic diversity and population structure of the Mediterranean sesame core collection with use of genome-wide SNPs developed by double digest RAD-Seq. *PLoS One,* 14(10), e0223757.

Bedigian, D. (2003). Evolution of sesame revisited: domestication, diversity and prospects. *Genetic resources and crop evolution,* 50, 779-787.

Bennett, M. (1998). Sesame seed. *Handbook for Farmers and Investors*.

Benzie, I.F. & Devaki, M. (2018). The ferric reducing/antioxidant power (FRAP) assay for non-enzymatic antioxidant capacity: concepts, procedures, limitations and applications. *Measurement of antioxidant activity & capacity: Recent trends and applications*, 77-106.

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

Bolvig, A.K., Nørskov, N.P., Hedemann, M.S., Foldager, L., McCarthy-Sinclair, B., Marco, M.L., Lærke, H.N. & Bach Knudsen, K.E. (2017). Effect of antibiotics and diet on enterolactone concentration and metabolome studied by targeted and nontargeted LC–MS metabolomics. *Journal of Proteome Research,* 16(6), 2135-2150.

Bragança, G.C.M., Ziegler, V., Ávila, B.P., Monks, J.L.F., Peres, W. & Elias, M.C. (2020). Multivariate analysis of the conditions of temperature, moisture and storage time in the technological, chemical, nutritional parameters and phytochemical of green lentils. *Journal of stored products research,* 87, 101617.

Chang, S.-T., Wu, J.-H., Wang, S.-Y., Kang, P.-L., Yang, N.-S. & Shyur, L.-F. (2001). Antioxidant activity of extracts from Acacia confusa bark and heartwood. *Journal of agricultural and food chemistry,* 49(7), 3420-3424.

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.

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.

Chen, X., Li, C., Wang, H. & Guo, Z. (2019). WRKY transcription factors: evolution, binding, and action. *Phytopathology Research,* 1(1), 1-15.

Choe, E. & Min, D.B. (2006). Mechanisms and factors for edible oil oxidation. *Comprehensive reviews in food science and food safety,* 5(4), 169-186.

Comini, E., Rubiales, D. & Reveglia, P. (2023). Variability of fatty acid composition and lignan content in sesame Germplasm, and effect of roasting. *ACS Food Science & Technology,* 3(10), 1747-1758.

Cortleven, A. & Schmülling, T. (2015). Regulation of chloroplast development and function by cytokinin. *Journal of experimental botany,* 66(16), 4999-5013.

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.

Cui, C., Liu, Y., Liu, Y., Cui, X., Sun, Z., Du, Z., Wu, K., Jiang, X., Mei, H. & Zheng, Y. (2021). Genome-wide association study of seed coat color in sesame (Sesamum indicum L.). *PLoS One,* 16(5), e0251526.

Cullis, B.R., Smith, A.B. & Coombes, N.E. (2006). On the design of early generation variety trials with correlated data. *Journal of agricultural, biological, and environmental statistics,* 11, 381-393.

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.

de Lima, F.S., Kurozawa, L.E. & Ida, E.I. (2014). The effects of soybean soaking on grain properties and isoflavones loss. *LWT-Food Science and Technology,* 59(2), 1274-1282.

De Zélicourt, A., Colcombet, J. & Hirt, H. (2016). The role of MAPK modules and ABA during abiotic stress signaling. *Trends in plant science,* 21(8), 677-685.

Demissie, A. (2006). Regional strategy for the ex situ conservation of plant genetic resources: Eastern Africa. Eastern Africa Plant Genetic Resources Network (EAPGRN), Entebbe, Uganda, July 2006. Abebe Demissie & Bjornstad, A. 1996. Phenotypic diversity of Ethiopian barleys in relation to geographical regions, altitudinal range and agroecological zones: as an aid to germplasm collection and conservation strategy. *Hereditas,* 124, 17-29.

Dossou, S.S.K., Xu, F., Cui, X., Sheng, C., Zhou, R., You, J., Tozo, K. & Wang, L. (2021). Comparative metabolomics analysis of different sesame (Sesamum indicum L.) tissues reveals a tissue-specific accumulation of metabolites. *BMC plant biology,* 21, 1-14.

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, 1-20.

Duan, Y., Ju, M., Miao, H. & Zhang, H. (2021). The Sesame Genome for Gene Discovery in Sesame. I: *The sesame genome*. Springer. 283-290.

Dutta, D., Harper, A. & Gangopadhyay, G. (2022). Transcriptomic analysis of high oil-yielding cultivated white sesame and low oil-yielding wild black sesame seeds reveal differentially expressed genes for oil and seed coat colour. *The Nucleus,* 65(2), 151-164.

Elleuch, M., Besbes, S., Roiseux, O., Blecker, C. & Attia, H. (2007). Quality characteristics of sesame seeds and by-products. *Food chemistry,* 103(2), 641-650.

Elsafy, M. (2023). Status of sesame breeding. *Introductory paper at the Faculty of Landscape Architecture, Horticulture and Crop Production Science*(2023: 2).

ETI (2023). *SESAME SEEDS IN SUDAN: Human rights risks linked to hummus and tahina*. https://ergonassociates.net/wp-content/uploads/2023/07/ETI-Sweden-Sesame-seed-report-FINAL-230511.pdf

Eynard, S.E., Windig, J.J., Hiemstra, S.J. & Calus, M.P. (2016). Whole-genome sequence data uncover loss of genetic diversity due to selection. *Genetics Selection Evolution,* 48, 1-13.

FAOSTAT (2023). Faostat. Food and Agriculture Organization of the United Nation

Feldsine, P., Abeyta, C. & Andrews, W.H. (2002). AOAC International methods committee guidelines for validation of qualitative and quantitative food microbiological official methods of analysis. *Journal of AOAC international,* 85(5), 1187-1200.

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.

Holland, J.B., Nyquist, W.E., Cervantes-Martínez, C.T. & Janick, J. (2003). Estimating and interpreting heritability for plant breeding: an update. *Plant breeding reviews,* 22.

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.

IBPGR (2004). Descriptors for sesame (Sesamum spp.). I: Resources, I.B.f.P.G. (red.).

Iorizzo, M., Cavagnaro, P.F., Bostan, H., Zhao, Y., Zhang, J. & Simon, P.W. (2019). A cluster of MYB transcription factors regulates anthocyanin biosynthesis in carrot (Daucus carota L.) root and petiole. *Frontiers in plant science,* 9, 1927.

Islam, F., Gill, R.A., Ali, B., Farooq, M.A., Xu, L., Najeeb, U. & Zhou, W. (2016). Sesame. I: *Breeding Oilseed Crops for Sustainable Production*. Elsevier. 135-147.

Ivarson, E., Leiva-Eriksson, N., Ahlman, A., Kanagarajan, S., Bülow, L. & Zhu, L.-H. (2017). Effects of overexpression of WRI1 and hemoglobin genes on the seed oil content of Lepidium campestre. *Frontiers in plant science,* 7, 2032.

Ju, M., Li, G., Tian, Q., Cao, H., Ma, Q., Duan, Y., Guo, H., Zhang, Z., Huang, Y. & Wang, H. (2024). Deletion of a 1,049 bp sequence from the 5′ UTR upstream of the SiHEC3 gene induces a seed non-shattering mutation in sesame. *Journal of Integrative Agriculture,* 23(8), 2589-2604.

Kamdar, J., Jasani, M., Ajay, B., Rani, K., Manivannan, N., Vasanthi, R., Dobariya, K., Pandey, M., Janila, P. & Radhakrishnan, T. (2021). Fatty acid desaturase-2 (ahFAD2) mutant alleles in peanut (Arachis hypogaea L.) pre-breeding lines: an insight into the source, features, discourse, and selection of novel pre-breeding lines. *Genetic resources and crop evolution,* 68, 529-549.

Khalafalla, A. & Ahmed, M. (2012). Proposal for the release of New Elite Sesame Genotypes for Rainfed. *A Paper Submitted to the National Variety Release Committee*.

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.

Kim, D.-O., Jeong, S.W. & Lee, C.Y. (2003). Antioxidant capacity of phenolic phytochemicals from various cultivars of plums. *Food chemistry,* 81(3), 321-326.

Kumar, M.V., Eevera, T., Ramesh, D. & Masilamani, P. (2019). Seed deterioration in long lived sesame seeds under accelerated ageing conditions. *Journal of Pharmacognosy and Phytochemistry,* 8(3), 706-709.

Kurt, C. (2018). Variation in oil content and fatty acid composition of sesame accessions from different origins. *Grasas y aceites,* 69(1), e241-e241.

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.

Latham, M.C. (1979). *Human nutrition in tropical Africa. A textbook for health workers, with special reference to community health problems in East Africa*.

Lattanzio, V. (2003). Bioactive polyphenols: Their role in quality and storability of fruit and vegetables. *Journal of Applied Botany,* 77(5/6), 128-146.

Laurentin, H.E. & Karlovsky, P. (2006). Genetic relationship and diversity in a sesame (Sesamum indicum L.) germplasm collection using amplified fragment length polymorphism (AFLP). *BMC genetics,* 7(1), 1-10.

Lee, E. & Choe, E. (2012). Changes in oxidation-derived off-flavor compounds of roasted sesame oil during accelerated storage in the dark. *Biocatalysis and Agricultural Biotechnology,* 1(1), 89-93.

Li, C., Duan, Y., Miao, H., Ju, M., Wei, L. & Zhang, H. (2021). Identification of candidate genes regulating the seed coat color trait in sesame (Sesamum indicum L.) using an integrated approach of QTL mapping and transcriptome analysis. *Frontiers in Genetics,* 12, 700469.

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.

Liu, H., Zhou, X., Wu, K., Yang, M. & Zhao, Y.J.M.b. (2015). Inheritance and molecular mapping of a novel dominant genic male-sterile gene in Sesamum indicum L. 35(1), 1-14.

Liu, L., Wang, J., Zhao, L., Guo, H., Zhao, S., Zheng, Q. & Yang, J.J.A.A.N.S. (2004). Research progress in mutational effects of aerospace on crop and ground simulation on aerospace environment factors. 18(4), 247-251.

Lü, J.M., Lin, P.H., Yao, Q. & Chen, C. (2010). Chemical and molecular mechanisms of antioxidants: experimental approaches and model systems. *Journal of cellular and molecular medicine,* 14(4), 840-860.

Mazur, R., Maszkowska, J., Anielska-Mazur, A., Garstka, M., Polkowska-Kowalczyk, L., Czajkowska, A., Zmienko, A., Dobrowolska, G. & Kulik, A. (2021). The SnRK2. 10 kinase mitigates the adverse effects of salinity by protecting photosynthetic machinery. *Plant Physiology,* 187(4), 2785-2802.

Meraj, T.A., Fu, J., Raza, M.A., Zhu, C., Shen, Q., Xu, D. & Wang, Q. (2020). Transcriptional factors regulate plant stress responses through mediating secondary metabolism. *Genes,* 11(4), 346.

Mohamed, I.A. (2011). The Sudanese Agriculture Development, History, Policies and Sustainability. *Agricultural and Natural Resource Economics Journal,* 3(82).

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.

Myint, D., Gilani, S.A., Kawase, M. & Watanabe, K.N. (2020). Sustainable sesame (Sesamum indicum L.) production through improved technology: An overview of production, challenges, and opportunities in Myanmar. *Sustainability,* 12(9), 3515.

O'brien, R.D. (2008). *Fats and oils: formulating and processing for applications*. CRC press.

Orsi, L., De Noni, I., Corsi, S. & Marchisio, L.V. (2017). The role of collective action in leveraging farmers' performances: Lessons from sesame seed farmers' collaboration in eastern Chad. *Journal of Rural Studies,* 51, 93-104.

Osman, H. (1985). Sesame growing in the Sudan. *sesame and safflower status and potential FAO plant production and protection paper,* 99.

Østdal, H., Andersen, H.J. & Nielsen, J.H. (2000). Antioxidative activity of urate in bovine milk. *Journal of agricultural and food chemistry,* 48(11), 5588-5592.

Parry, M.A. & Hawkesford, M.J. (2012). An integrated approach to crop genetic improvement F. *Journal of integrative plant biology,* 54(4), 250-259.

Pathak, K., Rahman, S.W., Bhagawati, S. & Gogoi, B. (2017). Sesame (Sesamum indicum L.), an underexploited oil seed crop: Current status, features and importance-A review. *Agricultural Reviews,* 38(3).

Pilar, B., Güllich, A., Oliveira, P., Ströher, D., Piccoli, J. & Manfredini, V. (2017). Protective role of flaxseed oil and flaxseed lignan secoisolariciresinol diglucoside against oxidative stress in rats with metabolic syndrome. *Journal of food science,* 82(12), 3029-3036.

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.

Re, R., Pellegrini, N., Proteggente, A., Pannala, A., Yang, M. & Rice-Evans, C. (1999). Antioxidant activity applying an improved ABTS radical cation decolorization assay. *Free radical biology and medicine,* 26(9-10), 1231-1237.

Relina, L., Suprun, O., Bohuslavskyi, R., Vecherska, L., Bezuhla, O., Kobyzeva, L., Vazhenina, O., Kolomatska, V., Ponurenko, S. & Ilchenko, N. (2022). Evaluation of genotype-environment interactions for non-polar lipids and fatty acids in chickpea (Cicer arietinum L.) seeds. *The Journal of VN Karazin Kharkiv National University. Series «Biology»,* 39, 20-34.

Reyes-Moreno, C., Okamura-Esparza, J., Armienta-Rodelo, E., Gomez-Garza, R. & Milán-Carrillo, J. (2000). Hard-to-cook phenomenon in chickpeas (Cicer arietinum L): Effect of accelerated storage on quality. *Plant Foods for Human Nutrition,* 55, 229-241.

Robinson, J. (2005). Desertification and disarray: the threats to plant genetic resources of southern Darfur, western Sudan. *Plant Genetic Resources,* 3(1), 3-11.

Saini, R.K., Rengasamy, K.R., Ko, E.-Y., Kim, J.-T. & Keum, Y.-S. (2020). Korean maize hybrids present significant diversity in fatty acid composition: an investigation to identify PUFA-rich hybrids for a healthy diet. *Frontiers in Nutrition,* 7, 578761.

Sánchez-Retuerta, C., Suaréz-López, P. & Henriques, R. (2018). Under a new light: regulation of light-dependent pathways by non-coding RNAs. *Frontiers in Plant Science,* 9, 962.

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.

Shaukat, K., Zahra, N., Hafeez, M.B., Naseer, R., Batool, A., Batool, H., Raza, A. & Wahid, A. (2022). Role of salicylic acid–induced abiotic stress tolerance and underlying mechanisms in plants. I: *Emerging plant growth regulators in agriculture*. Elsevier. 73-98.

Talhaoui, N., Gómez-Caravaca, A.M., León, L., De la Rosa, R., Segura-Carretero, A. & Fernández-Gutiérrez, A. (2014). Determination of phenolic compounds of 'Sikitita'olive leaves by HPLC-DAD-TOF-MS. Comparison with its parents 'Arbequina'and 'Picual'olive leaves. *LWT-Food Science and Technology,* 58(1), 28-34.

Taoukis, P.S., Labuza, T.P. & Saguy, I.S. (1997). Kinetics of food deterioration and shelf-life prediction. *The Editors,* 30.

Teboul, N., Magder, A., Zilberberg, M. & Peleg, Z. (2022). Elucidating the pleiotropic effects of sesame kanadi1 locus on leaf and capsule development. *The Plant Journal,* 110(1), 88-102.

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.

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.

Uzun, B. & Çağırgan, M.İ. (2006). Comparison of determinate and indeterminate lines of sesame for agronomic traits. *Field crops research,* 96(1), 13-18.

Wang, H., Zhang, H., Ma, Q., Wei, L., Ju, M., Li, C., Duan, Y. & Miao, H.J.J.o.H.A.S. (2017). Optimization of EMS mutagenesis condition and screening of mutants in sesame. 46(1), 36-41.

Wang, L., Dossou, S.S.K., Wei, X., Zhang, Y., Li, D., Yu, J. & Zhang, X. (2020). Transcriptome dynamics during black and white sesame (Sesamum indicum L.) seed development and identification of candidate genes associated with black pigmentation. *Genes,* 11(12), 1399.

Wang, L., Han, X., Zhang, Y., Li, D., Wei, X., Ding, X. & Zhang, X. (2014a). Deep resequencing reveals allelic variation in Sesamum indicum. *BMC plant biology,* 14, 1-11.

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, 1-13.

Wang, L., Yu, S., Tong, C., Zhao, Y., Liu, Y., Song, C., Zhang, Y., Zhang, X., Wang, Y. & Hua, W. (2014b). Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. *Genome biology,* 15, 1-13.

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.

Wang, X., Wang, M., Yan, G., Yang, H., Wei, G., Shen, T., Wan, Z., Zheng, W., Fang, S. & Wu, Z. (2023). Comparative analysis of drought stress-induced

physiological and transcriptional changes of two black sesame cultivars during anthesis. *Frontiers in plant science,* 14, 1117507.
Waterhouse, A.L. (2002). Determination of total phenolics. *Current protocols in food analytical chemistry,* 6(1), I1. 1.1-I1. 1.8.
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 (S esamum indicum L.) Germplasm. *Journal of integrative plant biology,* 55(8), 745-758.
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.
Were, B.A., Onkware, A.O., Gudu, S., Welander, M. & Carlsson, A.S. (2006). Seed oil content and fatty acid composition in East African sesame (Sesamum indicum L.) accessions evaluated over 3 years. *Field Crops Research,* 97(2-3), 254-260.
Wu, W.-H. (2007). The contents of lignans in commercial sesame oils of Taiwan and their changes during heating. *Food chemistry,* 104(1), 341-344.
Xiao, J., Zhou, Y., Xie, Y., Li, T., Su, X., He, J., Jiang, Y., Zhu, H. & Qu, H. (2024). ATP homeostasis and signaling in plants. *Plant Communications,* 5(4).
Yol, E., Basak, M., Kızıl, S., Lucas, S.J. & Uzun, B. (2021). A high-density SNP genetic map construction using ddRAD-seq and mapping of capsule shattering trait in sesame. *Frontiers in Plant Science,* 12, 679659.
Zahran, H.A., Abd-Elsaber, A. & Tawfeuk, H.Z. (2020). Genetic diversity, chemical composition and oil characteristics of six sesame genotypes. *OCL,* 27, 39.
Zeb, A. & Ahmad, S. (2017). Changes in acylglycerols composition, quality characteristics and in vivo effects of dietary pumpkin seed oil upon thermal oxidation. *Frontiers in chemistry,* 5, 55.
Zhang, H., Langham, D.R., Zhao, Y., Khalafalla, A. & Miao, H. (2021a). Traditional breeding in sesame. *The sesame genome*, 145-158.
Zhang, H., Miao, H., Wei, L., Li, C., Duan, Y., Xu, F., Qu, W., Zhao, R., Ju, M. & Chang, S. (2018a). Identification of a SiCL1 gene controlling leaf curling and capsule indehiscence in sesame via cross-population association mapping and genomic variants screening. *BMC plant biology,* 18, 1-12.
Zhang, H., Miao, H., Wei, L., Li, C., Duan, Y., Xu, F., Qu, W., Zhao, R., Ju, M. & Chang, S.J.B.p.b. (2018b). Identification of a SiCL1 gene controlling leaf curling and capsule indehiscence in sesame via cross-population association mapping and genomic variants screening. 18(1), 1-12.
Zhang, H., Zhao, Y., Wei, L., Zhou, T., Duan, Y. & Miao, H. (2021b). Molecular Mapping and Breeding in Sesame. *The sesame genome*, 159-177.
Zhang, Y., Wang, L., Li, D., Gao, Y., Lu, H. & Zhang, X.J.S.A.S. (2014). Mapping of sesame waterlogging tolerance QTL and identification of excellent waterlogging tolerant germplasm. 47, 422-430.
Zhao, M., Liu, S., Pei, Y., Jiang, X., Jaqueth, J.S., Li, B., Han, J., Jeffers, D., Wang, J. & Song, X. (2022). Identification of genetic loci associated with rough dwarf

disease resistance in maize by integrating GWAS and linkage mapping. *Plant Science,* 315, 111100.

Zhao, R., Miao, H., Song, W., Chen, C. & Zhang, H. (2018). Identification of sesame (Sesamum indicum L.) chromosomes using the BAC-FISH system. *Plant Biology,* 20(1), 85-92.

Zhou, R., Dossa, K., Li, D., Yu, J., You, J., Wei, X. & Zhang, X. (2018). Genome-wide association studies of 39 seed yield-related traits in sesame (Sesamum indicum L.). *International journal of molecular sciences,* 19(9), 2794.

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.

Zhou, Z., Robards, K., Helliwell, S. & Blanchard, C. (2002). Ageing of stored rice: changes in chemical and physical attributes. *Journal of Cereal Science,* 35(1), 65-78.

Popular Science Summary

Sesame is one of the world's oldest crops, long valued for its rich oil content, nutritional benefits, and ability to thrive in harsh environments. Yet, despite its importance, sesame has remained under-researched, especially in the countries where it is most widely grown. Sudan is a leading producer and a center of genetic diversity for sesame, but its unique landraces have been largely overlooked in modern breeding programs.

This research analyzed 200 genetically diverse sesame varieties collected from across Sudan. These accessions were grown and evaluated in multiple environments to understand the genetic and environmental factors affecting key traits, such as oil content, fatty acid composition, seed coat color, capsule shattering, and storage stability.

By combining traditional field observations with modern genomic tools like Genome-Wide Association Studies (GWAS) and genomic prediction, this research discovered valuable genetic markers and candidate genes linked to these traits. For example, novel genetic regions on chromosomes 1 and 2 influence capsule shattering, an issue that leads to seed loss during harvest. Identifying such markers allows breeders to develop varieties that are more suited for mechanical harvesting, creating better yield stability. Moreover, this work also explored the seed's biochemical profile. Some varieties, such as the local landrace *Abusundoug*, showed exceptional resistance to oil degradation under harsh storage conditions. This resilience is linked to higher phytochemical content and lower polyunsaturated fatty acids, essential for maintaining nutritional quality and shelf life.

Additionally, this study examined seed coat color, which ranges from white to black. Seed coat color is associated with antioxidant content and impacts consumer preference. GWAS applications identified genetic regions and candidate genes involved in pigment production, such as *WRKY* and *DOF zinc finger* transcription factors. These findings are not just of academic interest; they pave the way for marker-assisted breeding, a faster, more precise method for developing sesame varieties that are more nutritious, resilient to climate stress, and better suited for modern agriculture.

This research unlocks the potential of Sudanese sesame. It offers powerful tools to help transform this "neglected crop" into a climate-smart, nutritionally rich, and commercially valuable staple for farmers and consumers worldwide.

Populärvetenskaplig sammanfattning

Sesam är en av världens äldsta grödor och har länge uppskattats för sitt höga oljeinnehåll, sina näringsmässiga fördelar och sin förmåga att klara av tuffa miljöförhållanden. Trots sin betydelse är sesam fortfarande en underutforskad gröda, särskilt i de länder där den odlas mest. Sudan är en ledande producent och ett centrum för genetisk mångfald hos sesam, men dess unika lantsorter har till stor del förbisetts i moderna växtförädlingsprogram. Denna forskning analyserade 200 genetiskt varierande sesamvarianter insamlade från hela Sudan. Dessa accessions odlades och utvärderades i flera olika miljöer för att förstå de genetiska och miljömässiga faktorer som påverkar viktiga egenskaper, såsom oljehalt, fettsyresammansättning, fröfärg, kapselsprickning och lagringsstabilitet. Genom att kombinera traditionella fältobservationer med moderna genomiska verktyg som genome-wide association studies (GWAS) och genomisk prediktion, identifierades värdefulla genetiska markörer och kandidatgener kopplade till dessa egenskaper. Till exempel visade sig nya genetiska regioner på kromosom 1 och 2 påverka kapselsprickning – ett problem som leder till fröförlust vid skörd. Att identifiera dessa markörer hjälper förädlare att utveckla sorter som är bättre anpassade för mekanisk skörd och som ger stabilare avkastning. Forskningen undersökte även sesamfröets biokemiska profil. Vissa sorter, som den lokala lantsorten Abusundoug, visade exceptionell motståndskraft mot oljenedbrytning under svåra lagringsförhållanden. Denna hållbarhet är kopplad till ett högt innehåll av fytokemikalier och låga nivåer av fleromättade fettsyror – viktiga faktorer för att bevara näringskvalitet och hållbarhet. Studien granskade dessutom fröskalsfärg, som varierar från vitt till svart och påverkar både antioxidantinnehåll och konsumentpreferenser. Genom GWAS identifierades genetiska regioner och kandidatgener involverade i pigmentproduktion, såsom *WRKY*- och *DOF*-zinkfingertranskriptionsfaktorer. Dessa resultat är inte bara av akademiskt intresse. De banar väg för markörassisterad selektion, en snabbare och mer exakt metod för att utveckla sesamsorter som är mer näringsrika, klimatresistenta och bättre lämpade för modernt jordbruk.

Denna forskning låser upp potentialen hos sudanesisk sesam och erbjuder kraftfulla verktyg för att omvandla denna “förbisedda gröda” till en klimatklok, näringsrik och kommersiellt värdefull basgröda för odlare och konsumenter världen över.

Acknowledgments

I extend my deepest gratitude to those who have not only guided this research but also shaped the journey that led to it.

Prof. Eva Johansson, thank you for always being there. Your steady support, thoughtful guidance, and unwavering belief in me created a space where ideas could flourish and dreams could grow. Your calm encouragement has been a constant anchor throughout this work.

Dr. Mahbubjon Rahmatov, this has been more than supervision. It has been a journey spanning over 15 years, one built on mutual respect, friendship, growth, and an unspoken understanding. Words fall short of capturing the depth of my appreciation. Thank you for walking alongside me from the very beginning to this defining milestone.

Dr. Tilal Abdelhalim, your friendship and inspiration have been indispensable. Our endless brainstorming sessions, your ability to think beyond boundaries, and the clarity you bring to complex challenges made each day an adventure. Your presence transformed the challenging moments into shared momentum.

Dr. Amro Babiker, you lit the path through the intricate world of chemical analysis. Your sharp insight and patient mentorship illuminated methods and meanings I might never have reached alone. Thank you for making science not only precise but beautiful.

I would also like to sincerely thank all my colleagues at the **Department of Plant Breeding**, **Swedish University of Agricultural Sciences (SLU)**, especially the **Plant Product Quality group**, for their camaraderie, stimulating discussions, and the spirit of collaboration that enriched my PhD experience in countless ways.

This work was generously supported by the **Swedish Research Council (Vetenskapsrådet)**, whose funding made this research possible. I am deeply grateful for their investment in scientific inquiry and international collaboration.

I would also like to express my heartfelt appreciation to the **Agricultural Research Corporation (ARC), Sudan,** particularly the **Biotechnology and Biosafety Research Centre**, for their invaluable support and partnership. Their dedication and resilience, especially amid the ongoing challenges in Sudan, made it possible to conduct vital fieldwork and maintain the integrity of this research. This collaboration is a testament to the strength and spirit of Sudanese science.

Special thanks to the **Agricultural Plant Genetic Resources Conservation and Research Centre (APGRC) at ARC** for generously providing some of the

sesame accessions used in this study. Their commitment to preserving and sharing genetic resources was instrumental in enabling this research.
To every person and institution who stood by me, thank you. This thesis is not only a scientific document but a reflection of shared effort, enduring friendship, and unwavering belief in knowledge as a tool for change.

I

Discover Agriculture

Research

Tracking the storage stability in sesame (*Sesamum indicum* L.): impact of accelerated storage on storability characteristics, seed quality, phytochemical content, and fatty acids

Mohammed Elsafy[1] · Anders Ekholm[1] · Khitma A. Sir Elkhatim[2] · Manhal Gobara Hamid[2] · Mazahir H. Othman[2] · Tilal Sayed Abdelhalim[2] · Mahbubjon Rahmatov[1] · Eva Johansson[1] · Amro B. Hassan[3,4]

Received: 10 May 2024 / Accepted: 13 September 2024
Published online: 18 September 2024

Abstract
Storage stability under non-optimal conditions is an essential characteristic of Sudanese sesame. To understand opportunities to improve storage stability in sesame, seed quality, storability characteristics, content of fatty acids and phytochemicals, and antioxidant capacity were evaluated in ten Sudanese genotypes subjected to high temperature (55 °C) and humidity (60% RH) for 16 and 32 days. The accelerated storage increased seed color, linoleic acid, fungal growth, and peroxide value, while oil content, oleic acid, water activity, phytochemicals, and antioxidant capacity decreased ($P < 0.05$). The germination rate and content of saturated fatty acids were retained despite the storage ($P > 0.05$). The landrace Abusundoug showed better storage stability than the other genotypes due to generally low fatty acids and high phytochemical contents. The differences in storage stability in the Sudanese genotypes underscore the need for their further evaluation and use in breeding programs to improve sesame shelf life and quality.

Keywords Sesame · Cultivars · Oxidative stability · Accelerated storage conditions

1 Introduction

Sesame (*Sesamum indicum* L.) was domesticated over 3000 years ago and became one of the first oilseed crops utilized by humans [1]. As an oilseed crop suited for tropical regions, sesame continues to be an important commodity by yielding seeds with abundant levels of high-quality edible oils. In addition, sesame seeds contribute considerable amounts of protein and other essential nutrients to the human diet [2]. The sesame seed contains 35–60% oil, 3–19% protein, 13.5% carbohydrate, 5% ash, and a calorific value of 6355 kcal kg^{-1} [3], and it is also a rich source of nutrients such as copper, magnesium, manganese, iron, and vitamins E and B [4–6]. Sesame oil comprises 80% oleic and linoleic acids, while 18% are palmitic, stearic, and linolenic acids [7].

Several studies have indicated that sesame seed oil possesses significant antioxidant properties due to the presence of lignans, tocopherols, and Maillard reaction products [8]. A mixture of sesame and clove oil has also been shown to protect stored products against beetle infestations from, e.g., *Callosobruchus maculatus* [8, 9]. Due to the many positive effects of the

✉ Mohammed Elsafy, Mohammed.Elsafy@slu.se | [1]Department of Plant Breeding, Swedish University of Agricultural Sciences (SLU), P. O. Box 190, Alnarp, Sweden. [2]Biotechnology and Biosafety Research Center, Agricultural Research Corporation, Shambat, Khartoum North, Sudan. [3]Department of Food Science and Nutrition, College of Food and Agricultural Sciences, King Saud University, P. O. Box 2460, 11451 Riyadh, Saudi Arabia. [4]Environment and Natural Resource and Desertification Research Institute (ENDRI), National Center for Research, P. O. Box 6096, Khartoum, Sudan.

crop, sesame cultivation is expected to increase globally with a Compound Annual Growth Rate (CAGR) of 2.2% until the end of 2027 [10]. Thus, global sesame production is projected to reach 6.8 million tons shortly, with Sudan, Myanmar, Tanzania, India, Nigeria, and China ranked as the leading producing countries. [11].

Furthermore, sesame is expected to be increasingly recognized as a human health-contributing ingredient in food products [12]. As a result of Sudan being one of the centers of origin and a major producer of sesame, an enormous wealth of genetic diversity in sesame germplasm is present in the country [13]. Thereby, potentially important characters relevant to future sesame production should be present within the Sudanese germplasm.

Sesame, its oil, and other products are used locally in Sudan, but sesame is also Sudan's number one export commodity [14]. Therefore, sesame seeds and/or oil storage is necessary, and most traditional storage methods utilized are similar to those described for Ethiopia [15]. The climate in Sudan consists of desert conditions and temperatures around 45 °C (or even up to 50 °C) in the north and semi-desert or semi-arid conditions in the south [16]. Traditional storage under such a climate is often not beneficial for sesame seeds, which are high in oil and prone to fungal growth. In fact, storage under harsh conditions results in adverse effects on both sesame seeds and oils. Oxidation and fungal infestations during storage often deteriorate oil and seed quality, reduce germination potential, and increase levels of free fatty acids.

Additionally, long storage periods with adverse oxidation events result in color changes of the sesame seeds, destruction of essential fatty acids, and production of trans-acid and conjugated dienes, resulting in limitations of the use of the seeds for a range of agricultural and food applications [17]. Also, antioxidant and prooxidant properties were shown to be negatively affected by such storage conditions [18], which promoted the formation of polar compounds such as oxidized polymers, significantly and negatively affecting human health [19, 20]. Studies on Sudanese sesame have been limited until now, and the crop can be seen as a neglected crop in research. An increased understanding of the effect of harsh storage conditions on the quality of the sesame seeds is lacking. Thus, this study aimed to increase the understanding of how harsh storage conditions affect the sesame seeds' quality by comprehensively assessing color attributes, storability characteristics, phytochemical compound content, and fatty acid profiles concerning oxidative stability under accelerated storage conditions to improve the opportunities for storage stability. Additionally, the study aimed to predict further measures on how sesame seeds should be stored to improve the benefits of sesame for food, production, and export in Sudan and beyond.

2 Material and methods

2.1 Sample collection and preparation

A total of ten genotypes (cultivars landraces) native to Sudan were used in the present study, namely Kenana-2 (KN: whitish released cultivar), Gadarif (GF: whitish released cultivar), Eltayeb (EB: whitish released cultivar), Radoum (RM: whitish landraces), Tagarub (TB: whitish released cultivar), Southern Kordufan (SK: whitish landraces), Bromo (BO: whitish released cultivar), Rufaee (RE: whitish landraces), Hurhairy (HR: dark brown landraces), and Abusundoug (AS: light brown landraces). The genotypes were freshly harvested during the 2021–2022 season and exposed to accelerated storage conditions. Non-stored samples of each genotype were used as control samples and analyzed similarly to the stored samples described below.

The accelerated storage of sesame seeds was carried out over 16 and 32 days in a climate chamber with conditions set to 55 °C and 60% relative humidity (RH). The 55 °C/60% RH storage conditions result in a Q10 value of 3.4 for lipid oxidation, "the process by which lipids undergo oxidative degradation increases 3.4 times when the temperature is raised by 10 degrees Celsius" [21]. These conditions are considered a simulation of commercial storage at 25 °C for 6 and 12 months, which allows an evaluation of sesame seeds' changes in seed quality, storability characteristics, and antioxidant capacity.

2.2 Evaluation of seed quality

2.2.1 Seed color

A colorimeter (Chroma Meter CR 400, manufactured by Minolta, Japan) was used to determine color parameters (L*, a*, and b*) of stored and controlled sesame seeds. A standard white reflector plate was used to calibrate the equipment. The Petri dish attached to the device was filled with 50.0 g of sesame seeds for measurement. Color changes were estimated according to the following equation;

$$\Delta E = \sqrt{(\Delta L*)^2 + (\Delta a*)^2 + (\Delta b*)^2}$$

where; L* (luminosity); a* (negative = green and positive = red); b* (negative = blue and positive = yellow);. Δ*E* (the total color difference).

2.2.2 Germination rate

The germination rate was investigated utilizing 3 replicates of each cultivar for the different storage stages (control = non-stored samples of each cultivar, 16 days, and 32 days) using Petri dishes and filter paper. A total of 100 sesame seeds were used in every replicate. The germination rate of the sesame seeds was evaluated after 3 and 6 days of germination according to the International Seed Test Association [22].

2.2.3 Determination of oil content

Following [23] guidelines for oil extraction, the Soxhlet extraction method was used to determine the amount of oil in seeds. Seeds are ground into powder using a Soxhlet extractor (DWK Life Sciences Kimble™ KIMAX™ Soxhlet Apparatus with Allihn Condenser, Complete unit), and the oil is dissolved in a solvent. Then, the solvent is reheated, vaporized, and condensed over the seeds for several hours to dissolve the oil. Oil percentages were calculated by weighing the oil after it was extracted from the seed. The solvent was subsequently evaporated, resulting in the oil being left.

2.2.3.1 Fatty acid profile Oil extraction from sesame seeds was performed using a two-step solvent extraction process, followed by fatty acid analysis via gas chromatography (GC) as per the modified protocol [24]. The extraction was conducted in triplicate, with each replicate using ten seeds. First, non-polar lipids were extracted using hexane. Subsequently, the seed residue underwent a second extraction with a chloroform/methanol mixture (2:1 v/v) to recover any remaining polar lipids. The extracts from both steps were combined for each replicate to ensure complete lipid recovery. The combined extracts were then subjected to methylation at 90 °C before the n-Heptan partition to prepare fatty acid methyl esters (FAMEs) for GC analysis. An aliquot of 1 µl sample was injected into a GC–MS instrument with FID detector (Agilent Intuvo 9000 GC coupled with 5977B MSD detector and MassHunter workstation software.) operated with a split injection ratio of 10:1 (Gard chip: MS transfer line 250 °C, MS source 230 °C, and MS Quad 150°C). Analyte separation (250°C inlet temperature) utilized a 30 m DB-23 fused silica column (0.25 mm internal diameter and 0.25 µm film thickness) from Agilent. Ultra-high purity helium was used as carrier gas at a 2.5 ml/min programmed constant flow. The thermal ramping protocol for the GC oven began at an initial 150 °C held for 0.2 min, then increased at a gradient of 4 °C/min until reaching 210 °C. Subsequently, the temperature was further raised at 10 °C/min until the final 230 °C and maintained for 7 min. Chromatographic peaks were identified using mass spectral matching against reference signatures in the embedded NIST17 commercial library. Quantification relied on integrating total ion counts for each compound using MassHunter workstation software from Agilent.

2.3 Evaluation of storability characteristics

2.3.1 Fungal growth

Fungal growth in both the control and stored sesame seeds was assessed following the guidelines set by AOAC standards [25]. For each sample, one gram of material was homogenized in 10 ml of 0.1% peptone water using a vortex apparatus for 1 – 2 min, resulting in a final dilution of 1:10. Liquid samples were then serially diluted and plated on an appropriate medium. Specifically, selected dilutions (10^{-5}) of each sample (1 ml) were poured and plated on Petri dishes containing sterile Potato Dextrose Agar (PDA), followed by incubation for 5 days at 25°C. Colony formation was observed daily during the incubation period. Counts of colony-forming units per gram were recorded to quantify fungal growth (log cfu/g).

2.3.2 Water activity

The water activity (a_w) of the sesame seeds was assessed at room temperature using a hygrometer (Humimeter RH2, Schaller, Vienna, Austria) equipped with selectable sensors for measuring air humidity, material moisture, and water activity. The hygrometer was further enhanced with a temperature-controlled system to ensure a stable sampling environment according to the Official Methods of Analysis of the Association of Official Analytical Chemists (2005) described in [26].

2.3.3 Determination of free fatty acids

The free fatty acid (FFA) content in sesame seed was assayed according to [23]. The acidity of the sesame seed flour was calculated as mg KOH required to neutralize them in one-gram grain on a dry matter basis.

2.3.4 Determination of peroxide value

Measurement of peroxide values (PV) is critical for evaluating the progression of oxidation and rancidity development in oilseeds. Peroxide value was determined through the iron (II) thiocyanate assay as described by Østdal, Andersen [27] with brief modifications. Sesame seeds flour (1g) was mixed with methanol: chloroform solution (30 ml), vortexed at 1500 rpm for 30 s, and centrifuged to separate phases. The resulting lower chloroform phase was combined with iron (II) thiocyanate reagent prepared precisely according to documented compositions. After incubating for 5 min, absorbance was measured using a UV–Vis spectrophotometer (Shimadzu UV-1800, Thermo Scientific™ Evolution™ 300, or Agilent Cary 60) at 500 nm against a chloroform blank. Peroxide value was quantified using a standard conversion formula and expressed in milliequivalents (meq) of active oxygen per kilogram of oil ($meqO_2/kg$).

2.4 Determination of phenolic compounds and antioxidant activity

2.4.1 Preparation of phytochemical extracts

The samples were prepared following the procedures described by Talhaoui, Gómez-Caravaca [28]. First, the samples were mixed with absolute methanol (1g: 25 ml w/v) and stirred at room temperature for 24 h. The extraction of the phytochemicals was performed twice, and the extracted solution was then dried under a vacuum with a rotary evaporator to keep them dry for further analysis. All absorbance was measured using a UV–Vis spectrophotometer (Shimadzu UV-1800, Thermo Scientific™ Evolution™ 300, or Agilent Cary 60) at different wavelengths.

2.4.2 Determination of total phenolic content

The Folin-Ciocalteu Index (FCI) assay was used to analyze the total phenolics content (TPC) with slight modifications. Thus, according to [29], a solution of 20 µL (1:10 w/v) dried methanolic extract was mixed with 100 µL of Folin–Ciocalteu reagent, 300 µL of Na_2CO_3 (20%) and 1.58 ml H_2O. Then, the solution was incubated at 20 °C for 2 h. As a comparison to a blank solution, the absorbance was measured at 765 nm. The output was described as mg of gallic acid equivalents (GAE) per gram of sample (DW) based on a calibration curve using different concentrations of gallic acid ($R^2 = 0.99743$).

2.4.3 Determination of total flavonoid content

The Quantification of the samples' total flavonoid (TFC) content was determined as described previously [30]. The mixture of methanol extract (1 ml), 5% $NaNO_2$ solution (300 µL), 10% aluminum chloride (300 µL), and 1 mol/L sodium hydroxide (2 ml)was incubated at 25 °C for five minutes. Then, the solution was diluted to a volume of 10 mL with H_2O and vortexed thoroughly. The absorbance was recorded at 510 nm. A calibration curve was generated based on different concentrations of quercetin ($R^2 = 0.9761$). Quercetin equivalents (QEs) per gram of the sample (DW) were considered to calculate TFC.

2.4.4 Antioxidant activity

2.4.4.1 Radical scavenging activity Radical scavenging activity was measured using a 2-diphenyl-1-picrylhydrazyl (DPPH) assay as described by [31]. The DPPH was freshly produced by liquefying 3.9 mg DPPH in 100 ml of methanol to

obtain a 1.00 ± 0.01 unit of absorbance. After this, approximately 100 µl of sample extract or deionized water was mixed with Tris–HCl buffer (50 mM; pH 7.4; 900 µl) and 1000 µl of the DPPH solution, followed by incubation in darkness for 2 h. The absorbance of the mixture was measured at 517 nm. The amount of DPPH scavenging per gram sample is expressed as Trolox equivalents (mg TE/g).

2.4.4.2 Ferric reducing antioxidant power assay The extracts' Ferric Reducing Antioxidant Power (FRAP) was assessed using a UV/visible spectrophotometer at a wavelength of 593 nm, following the method described by Benzie and Devaki [32]. FRAP results were quantified and expressed as micromoles of Trolox equivalents per gram of the sample (mg Trolox/g).

2.4.4.3 ABTS free radical scavenging assay The ABTS radical cation decolorization assay was utilized to assess antioxidant capacity, following an established method by Re, Pellegrini [33] with modifications. The ABTS radical solution was prepared and diluted to an absorbance of 0.700 ± 0.02 at 734 nm and 30 °C before use. For the procedure, the sample was combined with the ABTS working reagent, and the decrease in absorbance was monitored at 1-min intervals for up to 6 min. Final absorbance was used to calculate Trolox equivalence against a standard curve ($R^2 = 0.9986$) and derive inhibition percentages. This colorimetric approach enables the measurement of radical scavenging activity, providing valuable standardized antioxidant capacity information through Trolox values (mg Trolox/g).

2.5 Statistical analysis

All analyses were carried out on three replicated samples, and mean values and standard deviations were calculated. To investigate the significance level of all the measurements across the three different storage conditions (non-stored and stored at 16 and 32 days, respectively) and the cultivar interaction, a two-way ANOVA was conducted using the General Linear Model (GLM) in the Minitab software version 19.2, which was also used to obtain the Pearson correlation matrix plot of highly correlated characteristics. [34]. Finally, the Principal Component Analysis (PCA) was carried out using the Factoextra-package in the R statistical program [35].

3 Results and discussion

3.1 Seed characteristics

The sesame seeds evaluated in this study showed no significant differences in germination rate, neither among cultivars nor due to the storage times applied here ($P > 0.05$; Table 1). Thus, similar to previous findings [36], which have shown a maintained germination rate of over 80% after accelerated storage, the tested sesame seeds in this study were hardy and highly viable after the storage.

Changes of color in sesame seeds while stored have been limitedly studied till now. The present study clearly showed a significant variation among storage times, cultivars, and their interaction for color parameters, including lightness (L*), redness (a*), yellowness (b*), and overall change (ΔE) (Table 1). Thus, the changes in the seed color of sesame were highly affected by the total time the seeds were stored under accelerated conditions. In fact, 32 days of accelerated storage resulted in a highly significant increase in the values of all color parameters (L*, a*, b*, and ΔE), while 16 days of storage instead reduced most color values (Table 1). The color degradation at 16 days may be attributable to the loss of pigments and browning reactions as bioactive compounds are utilized to counter storage stresses. The subsequent increase in coloration after 32 days may result from novel pigment and melanoidin formation from Maillard reactions and phenolic oxidation as constituents degrade further under sustained high temperatures [37, 38].

Among the cultivars evaluated here, the brown cultivars AS and HR (landraces) showed significantly lower L*, higher a*, and lower ΔE than the whitish cultivars. Additionally, the whitish landraces RM, RE, and SK tended to have lower L* but higher a* and lower ΔE compared to the released whitish cultivars KN, GF, EB, TB, and BO. These findings indicate that while color changed for all genotypes during accelerated storage, inherent color differences existed between brown, landrace whitish, and released whitish cultivars that persisted throughout storage. However, a relatively limited number of genotypes were evaluated here, and further research is needed to determine the specific compounds and genetic factors influencing color retention in sesame seeds during prolonged storage stresses. Considerable natural diversity exists in mature sesame seed coloration, ranging from black to white, including gray, brown, golden, yellow, and ivory.

Table 1 Germination and color characteristics of ten sesame genotypes under accelerated storage conditions

Genotypes	Germination (%)	Color characteristics			
		L *	**a ***	**b ***	**Δ*E***
KN	98.0 ± 0.4^{A}	72.3 ± 0.4^{C}	5.8 ± 0.1^{H}	16.5 ± 0.05^{G}	4.0 ± 0.4^{C}
GF	98.0 ± 0.4^{A}	71.6 ± 0.6^{C}	8.4 ± 0.02^{CD}	20.9 ± 0.2^{C}	8.0 ± 0.7^{A}
EB	97.0 ± 0.3^{A}	73.3 ± 0.7^{B}	7.8 ± 0.1^{DE}	20.4 ± 0.1^{D}	5.1 ± 0.7^{B}
RM	97.0 ± 0.4^{A}	69.6 ± 0.2^{D}	7.0 ± 0.04^{FG}	18.9 ± 0.02^{E}	1.6 ± 0.2^{DE}
TB	97.0 ± 0.4^{A}	74.3 ± 0.4^{A}	7.6 ± 0.1^{EF}	20.9 ± 0.2^{BC}	5.1 ± 0.4^{B}
SK	97.0 ± 0.5^{A}	71.8 ± 0.4^{C}	8.6 ± 0.2^{C}	21.2 ± 0.3^{B}	3.9 ± 0.4^{C}
BO	95.0 ± 0.5^{A}	74.7 ± 0.7^{A}	7.1 ± 0.02^{FG}	18.8 ± 0.1^{E}	5.2 ± 0.4^{B}
RE	95.0 ± 0.9^{A}	67.6 ± 0.2^{E}	6.8 ± 0.03^{G}	17.8 ± 0.1^{F}	2.0 ± 0.2^{D}
HR	93.0 ± 0.7^{A}	41.3 ± 0.1^{G}	11.7 ± 0.1^{A}	15.3 ± 0.01^{H}	0.9 ± 0.1^{E}
AS	92.0 ± 0.9^{A}	65.6 ± 0.3^{F}	10.1 ± 0.04^{B}	22.4 ± 0.1^{A}	3.8 ± 0.3^{C}
Storage time					
Control	96.0 ± 6.0^{A}	68.9 ± 10.0^{B}	8.0 ± 1.7^{B}	19.5 ± 2.6^{A}	0.0 ± 0.0^{C}
16	96.0 ± 6.0^{A}	66.3 ± 9.0^{C}	7.9 ± 1.8^{B}	18.9 ± 2.0^{B}	4.7 ± 3.0^{B}
32	96.0 ± 4.3^{A}	69.4 ± 11.5^{A}	8.4 ± 1.9^{A}	19.5 ± 2.8^{A}	7.1 ± 3.8^{A}
ANOVA					
F-Value, G	1.42^{NS}	4063.9***	146.5***	1429.9***	219.8***
F-Value, S	0.11^{NS}	375.2***	12.4***	134.4***	2214.0***
F-Value, G*S	0.8^{NS}	312.8***	8.7***	213.5***	101.9***
SE Fit	3.19	0.27	0.25	0.10	0.24

(L *) Luminosity, (a *) Negative = green and positive = red, (b *) Negative = blue and positive = yellow, (ΔE) total color difference, (G) Genotypes, (S) Storability, and values are means (±) SD of triplicate samples. Means in the same column that share the same superscript letters do not differ significantly ($P > 0.05$), ***; ($P < 0.001$), ^{NS}; no significant difference at ($P > 0.05$) as assessed by LSD

KN, Kenana-2; GF, Gadarif; EB, Eltayeb; RM, Radoum; TB, Tagarub; SK, Southern Kordufan; BO, Bromo; RE, Rufaee; HR, Hurhairy; AS, Abusundoug

This rich pigment variation provides opportunities to uncover genotype-specific resilience mechanisms that preserve color integrity under high temperature and humidity conditions. The elucidation of the biochemical basis for differential storage response among cultivars might contribute to guidance in breeding to enhance shelf life, which has also been suggested in previous studies [39]. In addition, previous research has reported that changes in the color of sesame seeds are associated with non-enzymatic browning, which results from polymerization reactions between phenolic compounds and tannins during storage conditions [40]. Thus, the change in sesame seed color reported here is likely the result of such a non-enzymatic browning due to temperature and storage period, similar to what has been reported for rice Park, Kim [41]. The differences found here in whitish color and color changes in landraces as compared to whitish released cultivars might then be the result of higher stability as related to non-enzymatic browning and, thereby, to polymerization between phenolic compounds and tannins in the landraces as compared to in the released cultivars.

Despite the fact that sesame seeds' nutritional value and shelf life depend heavily on oil retention and beneficial fatty acids, the effect of storage on these characteristics has been limitedly evaluated. The present study clearly showed that accelerated storage conditions significantly affected the total oil content and the content of oleic and linoleic acids (Table 2). Thus, corresponding with previous findings [42], the accelerated storage resulted in a decrease in total oil and oleic acid content, while it resulted in an increase in linoleic content and no significant change for the other evaluated fatty acids (Table 2). The fact that the accelerated storage resulted in differences in changes in the content of unsaturated (oleic and linoleic) and saturated (palmitic and stearic) fatty acids reflects the difference in the oxidation rates of these fats. It is well known that the low activation energy of unsaturated fats to form lipid radicals makes these more susceptible to oxidation than saturated fats [43]. Thereby, the oxidative instability of linoleic acid might also impair the shelf life of oil rich in linoleic acid, despite its known health benefits, such as e.g. a LDL cholesterol improver [44]. Accelerated storage conditions with high temperature and humidity are known to increase moisture content in sesame seeds, resulting in oxidation with a break-down of the oil and rancidity [45], which is the likely explanation for the decrease of the total oil content after storage. Ideal storage conditions for sesame seeds include cool, dry, and dark environments with low humidity, which will help preserve and prevent the oil content from spoilage.

Table 2 Oil content and content of fatty acids (mg/µl) of ten sesame genotypes under accelerated storage conditions

Genotypes	Oil content (%)	Content of fatty acids (mg/µl)						
		Palmitic	**Stearic**	**Oleic**	**Linoleic**	**Linolenic**	**Arachidic**	**Eicosenoic**
KN	**47.0 ± 0.1^{A}**	0.8 ± 0.01^{DE}	0.7 ± 0.01^{BC}	2.2 ± 0.1^{ABC}	3.9 ± 0.1^{BC}	0.03 ± 0.0^{AB}	0.10 ± 0.0^{BC}	0.02 ± 0.0^{AB}
GF	46.5 ± 0.2^{A}	0.9 ± 0.01^{CDE}	0.7 ± 0.01^{BC}	2.3 ± 0.1^{ABC}	4.5 ± 0.1^{ABC}	0.03 ± 0.0^{AB}	0.07 ± 0.0^{ABC}	0.02 ± 0.0^{AB}
EB	46.7 ± 0.1^{A}	1.0 ± 0.01^{ABCD}	0.8 ± 0.01^{AB}	2.9 ± 0.1^{ABC}	5.1 ± 0.1^{A}	0.03 ± 0.0^{AB}	0.09 ± 0.0^{AB}	0.02 ± 0.0^{AB}
RM	46.6 ± 0.1^{A}	0.8 ± 0.02^{E}	0.6 ± 0.01^{C}	1.7 ± 0.04^{BC}	3.5 ± 0.1^{C}	0.02 ± 0.0^{B}	0.07 ± 0.0^{CD}	0.01 ± 0.0^{AB}
TB	46.4 ± 0.2^{A}	1.0 ± 0.01 ^{BCD}	0.8 ± 0.01^{ABC}	2.7 ± 0.2^{ABC}	4.8 ± 0.1^{AB}	0.03 ± 0.0^{AB}	0.08 ± 0.0^{ABC}	0.02 ± 0.0^{AB}
SK	46.4 ± 0.1^{A}	0.9 ± 0.01^{BCDE}	0.7 ± 0.01^{ABC}	2.6 ± 0.1^{ABC}	4.0 ± 0.1^{BC}	0.02 ± 0.0^{B}	0.08 ± 0.0^{ABC}	0.02 ± 0.0^{AB}
BO	46.8 ± 0.1^{A}	1.0 ± 0.01^{ABC}	0.8 ± 0.01^{ABC}	3.0 ± 0.2^{ABC}	5.1 ± 0.1^{A}	0.03 ± 0.0^{AB}	0.09 ± 0.0^{ABC}	0.02 ± 0.0^{AB}
RE	46.4 ± 0.1^{A}	1.1 ± 0.01^{AB}	0.8 ± 0.01^{AB}	3.3 ± 0.3^{AB}	5.2 ± 0.1^{A}	0.10 ± 0.0^{AB}	0.09 ± 0.0^{A}	0.03 ± 0.0^{AB}
HR	46.4 ± 0.1^{A}	**1.2** ± 0.01^{A}	**0.9** ± 0.01^{A}	**4.3** ± 0.1^{A}	**5.4** ± 0.1^{A}	**0.07** ± 0.0^{A}	**0.10** ± 0.0^{A}	**0.04** ± 0.0^{A}
AS	46.4 ± 0.1^{A}	0.5 ± 0.01^{F}	0.4 ± 0.01^{D}	1.0 ± 0.1^{C}	2.2 ± 0.1^{D}	0.02 ± 0.0^{B}	0.04 ± 0.01^{D}	0.01 ± 0.0^{B}
Storage time								
Control	47.5 ± 0.7^{A}	0.9 ± 0.2^{A}	0.7 ± 0.1^{A}	3.3 ± 2.1^{A}	4.1 ± 1.2^{B}	0.04 ± 0.04^{A}	0.08 ± 0.02^{A}	0.03 ± 0.03^{A}
16	45.4 ± 0.6^{C}	0.9 ± 0.2^{A}	0.7 ± 0.1^{A}	2.1 ± 0.8^{B}	4.4 ± 1.0^{AB}	0.03 ± 0.01^{A}	0.08 ± 0.02^{A}	0.02 ± 0.00^{A}
32	46.7 ± 0.6^{B}	1.0 ± 0.2^{A}	0.7 ± 0.1^{A}	2.4 ± 1.0^{B}	4.6 ± 1.1^{A}	0.03 ± 0.01^{A}	0.08 ± 0.02^{A}	0.02 ± 0.01^{A}
ANOVA								
F-value, G	0.97^{NS}	22.5***	13.2***	3.96***	18.4***	2.72***	10.8***	1.80^{NS}
F-Value, S	67.4***	2.90^{NS}	1.90^{NS}	6.50***	3.55***	1.70^{NS}	0.40^{NS}	2.00^{NS}
F-Value, G*S	1.14^{NS}	0.30^{NS}	0.40^{NS}	0.24***	0.26^{NS}	0.83^{NS}	0.31^{NS}	0.70^{NS}
SE Fit	0.40	0.07	0.80	0.20	0.40	0.01	0.01	0.01

(G) Genotypes, (S) Storability, and values are means (±) SD of triplicate samples. Means in the same column that share the same superscript letters do not differ significantly ($P > 0.05$), ***; ($P < 0.001$), ^{NS}; no significant difference at ($P > 0.05$) as assessed by LSD

KN, Kenana-2; GF, Gadarif; EB, Eltayeb; RM, Radoum; TB, Tagarub; SK, Southern Kordufan; BO, Bromo; RE, Rufaee; HR, Hurhairy; AS, Abusundoug

However, differences in the content of the various fatty acids evaluated were noted among the cultivars, although the oil content did not differ significantly (Table 2). Generally, high levels of most fatty acids were recorded for the dark brown cultivar HR, while low levels were noted for the light brown cultivar AS (Table 2). Previous studies have indicated that genetic factors determine the oil content, fatty acid synthesis, and sesame compound retention during storage [46]. The present study included a relatively limited number of genotypes, and further elucidation of the genetic and biochemical basis for fatty acid stability during accelerated storage, including a higher number of genotypes, is a prerequisite if breeding for these characters should be possible.

3.2 Storability characteristics

Fungal growth, a_w, FFA, and PV exhibited significant differences both across storage conditions and among cultivars (Table 3). In general, fungal growth and PV increased with prolonged accelerated storage, while the significantly highest values for a_w and PV were recorded after 16 days of accelerated storage (Table 3). As shown by others [17], fungal infections are normally increasing if seeds are stored at high relative humidity levels, although a high FFA concentration also promotes fungal infection [47]. Additionally, a high FFA content contributes to rancidity and results in off-flavors and odors in oilseeds during storage and has, therefore, been associated with limitations in industrial uses of sesame [17]. However, in the present investigation, the FFA content was not significantly higher after 32 days of storage than in control samples. The FFA content was significantly higher at 16 days of storage, which might result from reduced lipase enzyme activity from heat damage during storage, as reported by others [48]. Oils and fats in sesame are commonly tested for oxidative rancidity with PV, a measure of the oxidation of hydroperoxides, which are unstable and readily decompose into volatile aldehydes [49]. As lipid oxidation decreases as storage time increases, PV is primarily useful as a flavor quality indicator in the early stages of lipid oxidation [50].

Similarly to previous studies by Abdelazim, Mahmoud [51], the PV increased gradually and rapidly as the storage period increased. The same author reported that sesame cake extract degraded the oxidation in soybean oil compared to other extracts. Thus, low PV values in sesame cultivars might be connected to a high antioxidant capacity in those

Table 3 Storability characteristics of sesame genotypes under accelerated storage conditions

Genotypes	Fungal growth (log cfu/g)	Water activity (a_w)	Free fatty acid (FFA; mgKOH/g)	Peroxide value (PV; meq O_2/g)
KN	5.7 ± 0.01^{B}	0.32 ± 0.005^{AB}	0.36 ± 0.01^{B}	1.48 ± 0.08^{D}
GF	2.8 ± 0.02^{C}	0.29 ± 0.004^{CD}	0.44 ± 0.04^{A}	1.35 ± 0.05^{F}
EB	5.9 ± 0.13^{A}	0.31 ± 0.005^{BCD}	0.38 ± 0.01^{AB}	1.27 ± 0.05^{G}
RM	0.0 ± 0.00^{D}	0.33 ± 0.001^{A}	0.34 ± 0.01^{B}	1.42 ± 0.07^{E}
TB	0.0 ± 0.00^{D}	0.30 ± 0.005^{BCD}	0.39 ± 0.02^{AB}	1.49 ± 0.08^{D}
SK	0.0 ± 0.00^{D}	0.29 ± 0.005^{D}	0.44 ± 0.01^{A}	1.33 ± 0.05^{F}
BO	0.0 ± 0.00^{D}	0.31 ± 0.001^{ABC}	0.33 ± 0.01^{B}	1.85 ± 0.09^{A}
RE	0.0 ± 0.00^{D}	0.32 ± 0.001^{AB}	0.36 ± 0.01^{B}	1.62 ± 0.07^{B}
HR	5.7 ± 0.01^{B}	0.32 ± 0.007^{AB}	0.36 ± 0.01^{B}	1.88 ± 0.09^{A}
AS	2.8 ± 0.01^{C}	0.29 ± 0.005^{CD}	0.39 ± 0.03^{AB}	1.56 ± 0.08^{C}
Storage time				
Control	2.1 ± 2.1^{C}	0.3 ± 0.04^{B}	0.3 ± 0.1^{B}	0.7 ± 0.1^{C}
16	2.2 ± 2.5^{B}	0.4 ± 0.02^{A}	0.5 ± 0.1^{A}	1.6 ± 0.4^{B}
32	2.5 ± 2.9^{A}	0.2 ± 0.01^{C}	0.3 ± 0.1^{B}	2.3 ± 0.3^{A}
ANOVA				
F-Value, G	55,830.1***	10.3***	2.0***	318.8***
F-Value, S	953.6***	447.5***	32.2***	16,051.1***
F-Value, G*S	484.7***	8.2***	4.5***	179.22***
SE Fit	0.02	0.01	0.04	0.02

(G) Genotypes, (S) Storability, and values are means (±) SD of triplicate samples. Means in the same column that share the same superscript letters do not differ significantly ($P > 0.05$), ***; ($P < 0.001$), ^{NS}; no significant difference at ($P < 0.05$) as assessed by LSD

KN, Kenana-2; GF, Gadarif; EB, Eltayeb; RM, Radoum; TB, Tagarub; SK, Southern Kordufan; BO, Bromo; RE, Rufaee; HR, Hurhairy; AS, Abusundoug

seeds. Water activity (a_w) is an essential measure of available moisture for chemical reactions, microbial growth, and shelf-life stability of products [52]. In the present study, the a_w increased significantly ($P < 0.0001$) to 0.4 after 16 days, while it was reduced back to 0.2 ($P < 0.0001$) after 32 days of storage. This reduction in a_w after prolonged storage is likely due to high temperature and duration, and proper aw maintenance has, in previous investigations, been found critical for seed viability over time [53].

Fungal growth was found to differ significantly among the studied genotypes, with the significantly highest fungal growth in the cultivar EB (5.9 log cfu/g), followed by the cultivars KN and HR (5.7 log cfu/g), while no fungal growth was observed in the RM, TB, SK, BO, and RE cultivars. In principle, some of the genotypes were likely infected at the beginning of the storage period, and on those genotypes, an increase in fungal growth was noted over the storage period. Although clear genotypic differences were noted for a_w, FFA, and PV, no clear pattern was observed regarding the cultivar type. Thus, breeding for these characters will need additional information regarding the genetic background of the differences observed.

3.3 Phytochemical compounds and antioxidant activity

The accelerated storage conditions clearly impacted TPC, TFC, FRAP, and ABTS, which all decreased significantly with the storage time (Table 4), while no change in DPPH was noted. Phenolic compounds and antioxidants play an essential protective role in the oil seeds by directly reacting with and neutralizing free radicals that cause oxidative damage [54]. Previous investigations on other crops, e.g., pistachios and soybeans [55], have also reported decreased phytochemicals and antioxidant activity with storage. Bragança, Ziegler [56] found that phytochemicals and antioxidant capacity reduction over storage periods might result from oxidation and chemical reactions during high-temperature stress. The DPPH assay is used to measure the radical scavenging activity, and lignans, as well as lignin glycosides in sesame cake, have been found to have sufficient radical scavenging activity [57]. The retained DPPH activity during the storage of sesame seeds shown in this work suggests that sesame antioxidants preserve free radical neutralization even after accelerated aging. In previous studies, sesame antioxidants have contributed to preventing oxidation associated with quality

Table 4 Phenolic compounds and antioxidant activity of sesame genotypes seeds under accelerated storage conditions

Genotypes	Phenolic compounds		Antioxidant capacity		
	TPC (mg GAE/g)	**TFC** (mg QE/g)	**DPPH** (mg Trolox/g)	**FRAP** (mg Trolox/g)	**ABTS** (mgTrolox/g)
KN	3.9 ± 0.3^{G}	46.5 ± 0.8^{E}	4.8 ± 0.3^{A}	3.6 ± 0.2^{FG}	12.0 ± 0.2^{B}
GF	4.9 ± 0.1^{D}	$\mathbf{103.4} \pm 3.3^{\mathbf{A}}$	4.9 ± 0.2^{A}	3.3 ± 0.3^{G}	11.4 ± 0.4^{C}
EB	2.9 ± 0.1^{H}	88.1 ± 5.5^{B}	4.9 ± 0.2^{A}	$\mathbf{2.3} \pm 0.1^{\mathbf{H}}$	9.9 ± 0.6^{D}
RM	5.6 ± 0.3^{C}	56.3 ± 0.6^{D}	4.8 ± 0.2^{A}	5.0 ± 0.3^{D}	8.7 ± 0.5^{E}
TB	7.5 ± 0.5^{B}	46.3 ± 0.9^{E}	4.9 ± 0.2^{A}	6.6 ± 0.2^{C}	6.2 ± 0.5^{G}
SK	4.2 ± 0.2^{F}	53.2 ± 1.0^{D}	4.8 ± 0.3^{A}	9.2 ± 0.6^{B}	11.2 ± 0.6^{C}
BO	$\mathbf{2.8} \pm 0.1^{\mathbf{H}}$	$\mathbf{42.1} \pm 2.4^{\mathbf{F}}$	4.9 ± 0.2^{A}	3.3 ± 0.2^{G}	5.1 ± 0.1^{H}
RE	4.6 ± 0.2^{E}	45.2 ± 1.4^{EF}	$4.7 \pm 0.3^{\mathbf{A}}$	4.2 ± 0.1^{EF}	$\mathbf{4.4} \pm 0.2^{\mathbf{I}}$
HR	4.3 ± 0.4^{EF}	43.8 ± 1.3^{EF}	4.9 ± 0.2^{A}	4.6 ± 0.3^{DE}	7.0 ± 0.3^{F}
AS	$\mathbf{10.2} \pm 0.3^{\mathbf{A}}$	63.3 ± 0.8^{C}	$\mathbf{5.0} \pm 0.2^{\mathbf{A}}$	$\mathbf{10.5} \pm 0.8^{\mathbf{A}}$	$\mathbf{16.9} \pm 0.3^{\mathbf{A}}$
Storage time					
Control	7.9 ± 3.1^{A}	79.1 ± 36.0^{A}	4.9 ± 0.4^{A}	8.7 ± 5.2^{A}	13.2 ± 4.6^{A}
16	4.5 ± 2.2^{B}	53.3 ± 13.6^{B}	4.9 ± 0.1^{A}	4.1 ± 1.9^{B}	8.7 ± 3.9^{B}
32	2.9 ± 1.8^{C}	44.1 ± 16.6^{C}	4.8 ± 0.5^{A}	3.0 ± 1.2^{C}	5.9 ± 3.4^{C}
ANOVA					
F-Value, G	1429.8***	653.7***	1.5^{NS}	397.4***	4006.4***
F-Value, S	6283.1***	1653.3***	1.8^{NS}	1657.6***	12,668.9***
F-Value, G*S	183.9***	144***	1.1^{NS}	118.7***	391***
SE Fit	0.1	1.4	0.01	0.2	0.1

(TPC) Total phenolic compounds, (TFC) Total flavonoid compounds, (DPPH) 2,2-diphenyl-1-picrylhydrazyl, (FRAP) Ferric Reducing Antioxidant Power, (ABTS) 2,2′-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid), (G) Genotypes, (S) Storability, and values are means (±) SD of triplicate samples. Means in the same column that share the same superscript letters do not differ significantly ($P > 0.05$), ***; ($P < 0.001$),^{NS}; no significant difference at ($P > 0.05$) as assessed by LSD

KN, Kenana-2; GF, Gadarif; EB, Eltayeb; RM, Radoum; TB, Tagarub; SK, Southern Kordufan; BO, Bromo; RE, Rufaee; HR, Hurhairy; AS, Abusundoug

deterioration in oils and oilseeds [58]. Thus, the present results suggest that bioactive compounds in sesame seeds might continue to effectively scavenge free radicals and mitigate oxidative damage during storage stress.

Differences in the content of phytochemicals and antioxidant capacity were noted among the genotypes in the present study (Table 4). In general, high content and capacity were found for the light brown landrace AS, except for TFC, for which the whitish-released cultivar GF showed the highest value (Table 4). However, no consistency was found in the variation of phytochemicals and antioxidant capacity among the evaluated genotypes, and additional studies incorporating a higher number of genotypes are needed to understand the genetic background for the variation.

3.4 Multivariable analysis and remarks outlook

The first principal component (PC1; explaining 34.2% of the variation) of the Principal Component Analysis (PCA) separated mainly the content of fatty acids with negative PC1 values from the content of phytochemicals and antioxidant capacity, with positive PC1 values (Fig. 1A), which indicates that samples with high fatty acid content showed low phytochemical content and low antioxidant capacity, as well as vice versa, which was verified by a Pearson correlation analysis (Fig. 1B). The PCA also visualized that prolonged storage decreased phytochemicals and antioxidant capacity (Table 4), as the control samples were placed further to the positive side along the PC1 axes than the samples stored at 32 days (Fig. 1A). The second principal component (PC2, explaining 23.8% of the variation) separated the samples based on storage time, with the control samples showing negative PC2 values and the samples stored at 32 days showing the most positive PC2 values. The loadings contributing mainly to the separation of the samples according to storage were PV (peroxide value), DC (change in color), and OC (oil content), where control samples showed low PV (Table 3) and DC (Table 1), and low OC (Table 2), oppositely to samples stored during 32 days. Previous studies have shown that the storability characteristics of oilseeds are affected by a range of factors, including moisture content, oil content, and insect

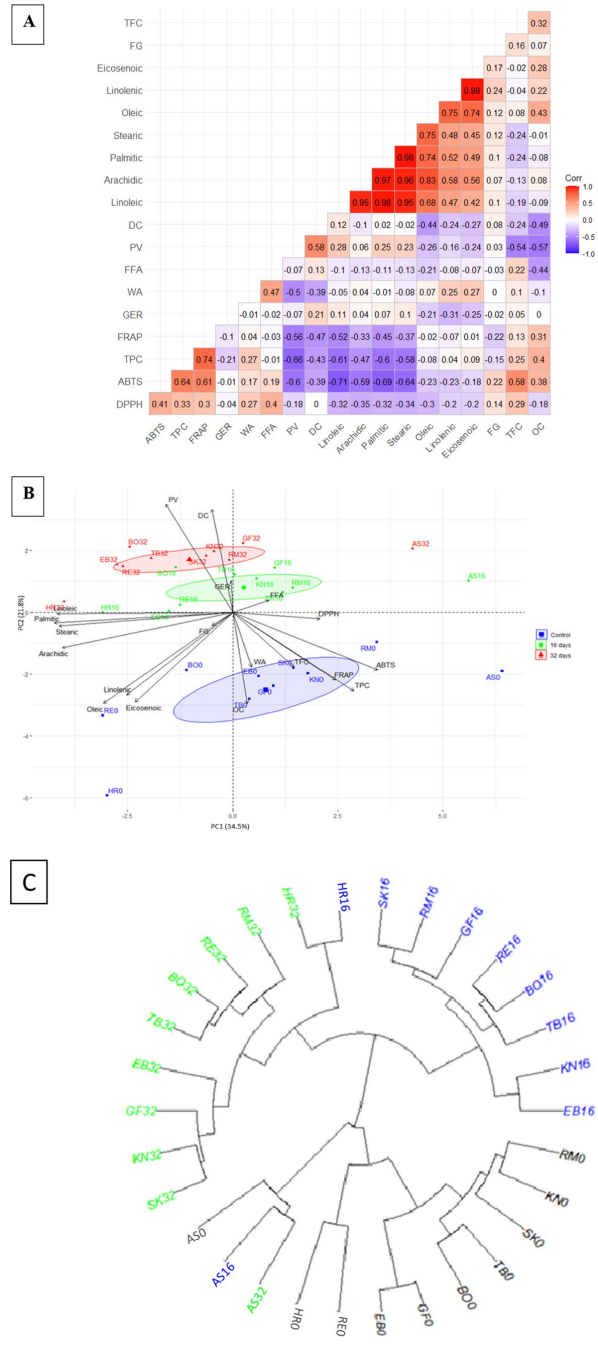


Fig. 1 Multivariate analyses comparing seed quality and storability characteristics, content of oil, fatty acids and phytochemicals, as well as antioxidant capacity in seeds of 10 sesame genotypes stored under control and accelerated condition; **A** biplot, from the principle component of analysis (PCA), combining the loading and score plot results, where Kenana-2 (KN), Gadarif (GF), Eltayeb (EB), Radoum (RM), Tagarub (TB), Southern Kordufan (SK), Bromo (BO), Rufaee (RE), Hurhairy (HR), Abusundoug (AS), Total color difference (DC), Total phenolic compounds (TPC), Total flavonoid compounds (TFC), 2,2-diphenyl-1-picrylhydrazyl (DPPH), Ferric Reducing Antioxidant Power (FRAP), 2,2′-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS), Water activity (WA), Germination (GER), Free fatty acids (FFA), Fungal growth (FG), and Oil content (OC), **B** correlation coefficients from a Pearson correlation analyses, and **C** a hierarchical dendrogram showing relatedness of different samples durin different length of storage

infestation [59]. Here, we could determine the importance and changes of various characters that impact the storability of sesame seeds. The important ones, e.g., PV, DC, OC, and content of phytochemicals and fatty acids, need further evaluation, and genotypic variation must be determined to be used in breeding for more storage-tolerant cultivars.

A hierarchical dendrogram (Fig. 1C), which explains the performance of the sesame cultivars across storage lengths, clustered the material into three main groups (0 days, 16 days, and 32 days) with a few exceptions. Thus, interestingly, all the AS (light brown landrace) samples (AS0 = control, AS16 and AS32 = stored 16 and 32 days) clustered together with the control (0) samples, indicating a high storage stability in that genotype. The AS genotype was found to have low levels of fatty acids (Table 2) and high levels of phytochemicals and antioxidant potential (Table 4), placing it with the highest positive PC1 values (Fig. 1A) among the genotypes. Additionally, HR16 was clustered with HR32 (dark brown landrace) in the cluster with accelerated storage in 32 days, indicating low storage stability in this genotype. The HR genotype was found to have a high content of fatty acids (Table 2), and HR16 and HR32 were placed with the most negative PC1 values. Therefore, the present results contribute a basis for understanding essential characteristics to evaluate storage stability in sesame. The AS and the HR are Sudanese landraces, and none have been previously evaluated concerning storage quality. Previous studies have indicated the importance of incorporating genes from landraces into adapted cultivars to improve the quality traits for long shelf-life and storage [60]. Thus, specifically, AS should be further evaluated for storage stability characteristics and genes behind such characters, including an evaluation of performance over additional seasons.

4 Conclusions

The seed quality, storability characteristics, phytochemicals content, and antioxidant potential were significantly affected by accelerated storage and the length of this storage in sesame. In principle, seed color increased, oil content and oleic acid decreased, while linoleic acid increased, and fungal growth and peroxide values increased. In contrast, water activity decreased, and phytochemicals and antioxidant potential decreased as correlated with the length of the accelerated storage time. Seed color change, peroxide value, and oil content were found to be the most important characteristics of sesame, correlating with the accelerated storage time. Among the ten evaluated genotypes, one landrace (Abusundoug, light brown) exhibited superior storage stability compared to the others, which enhanced stability correlated with lower polyunsaturated fatty acids and higher levels of phytochemicals, resulting in increased antioxidant potential. Given these favorable characteristics, it is recommended that Abusundoug be further evaluated for its potential incorporation into sesame breeding programs to improve seed storage stability and nutritional quality. The findings reported here indicate a potential to improve the storage stability of Sudanese sesame, which is extremely important in this high-value commodity crop for Sudan. Under current circumstances, the produced sesame seeds in Sudan are subjected to harsh storage conditions (especially high temperatures), which mostly have a negative impact on the quality of the seeds and the oil. Improving the storage stability will contribute to income opportunities for the country and its farmers, especially as the interest in sesame oil is growing in the world market. Thus, to achieve improved storage stability in Sudanese sesame, a more extensive range of genetic material should be analyzed for fatty acid composition, phytochemical content, and antioxidant capacity to assess the variability of these components more fully across different sesame accessions. Furthermore, the genes behind the content of these characters need to be evaluated. Additionally, accelerated storage experiments should be carried out on genotypes with exceptionally high and low levels of the identified components and grown in various environments (years and localities) to increase the understanding of storage stability.

Acknowledgements The authors acknowledge the sponsorship of The Swedish Research Council (VR) fund (DR—2020-04163).

Author contributions ME: Investigation and writing – original draft Preparation.AE, KS, MH, and MO: Lab Investigation. TA: Supervision and visualization. MR: Supervision, visualization, and funding acquisition. EJ: Supervision, review, editing, and funding acquisition. AH: Supervision, conceptualization, review, and editing.

Funding Open access funding provided by Swedish University of Agricultural Sciences.

Data availability Data will be made available upon request.

Declarations

Competing interests The authors declare no competing interests.

References

1. Pamei I, Makandar R. Association of 16SrII-D phytoplasma with phyllody disease in sesame (*Sesamum indicum*) in Telangana from Southern India. Plant Dis. 2016;100(8):1774–1774.
2. Najeeb U, et al. Sesame. In: Technological innovations in major world oil crops, vol. 1. New York: Springer; 2012. p. 131–45.
3. Elsafy M. Status of sesame breeding. Science. 2023: p. 2.
4. Elleuch M, et al. Quality characteristics of sesame seeds and by-products. Food Chem. 2007;103(2):641–50.
5. Nimmakayala P, et al. Sesamum. In: Kole C, editor., et al., Wild crop relatives: genomic and breeding resources. Berlin: Springer; 2011. p. 261–73.
6. Tunde-Akintunde T, Akintunde B. Some physical properties of sesame seed. Biosys Eng. 2004;88(1):127–9.
7. Kurt C. Variation in oil content and fatty acid composition of sesame accessions from different origins. Grasas Aceites. 2018;69(1):e241–e241.
8. Wan Y, et al. The relationship of antioxidant components and antioxidant activity of sesame seed oil. J Sci Food Agric. 2015;95(13):2571–8.
9. Soe TN, Ngampongsai A, Sittichaya W. Synergistic effect of sesame oil and clove oil on toxicity against the pulse beetle, *Callosobruchus maculatus* (Fabricius) (Coleoptera: Chysomelidae). Khon Kaen Agric J. 2019;47(1):2562.
10. Intelligence M. Sesame seeds market—growth, trends, COVID-19 impact, and forecasts (2022–2027). 2021. https://www.mordorintelligence.com/industry-reports/sesame-seeds-market. Accessed 12 May 2024.
11. FAOSTAT. Crops and livestock products. 2020. https://www.fao.org/faostat/en/#data/QCL. Accessed 20 May 2023.
12. Oboulbiga EB, et al. Physicochemical, potential nutritional, antioxidant and health properties of sesame seed oil: a review. Front Nutr. 2023;10:1127926.
13. Laurentin HE, Karlovsky P. Genetic relationship and diversity in a sesame (*Sesamum indicum* L.) germplasm collection using amplified fragment length polymorphism (AFLP). BMC Genet. 2006;7(1):1–10.
14. Ibrahim AH, Purnomo EP, Malawani AD. The most important agricultural products that sudan exports and the mechanisms to develop. Asian J Agric Ext Econ Sociol. 2020;38(08):121–33.
15. Alemayehu S, et al. Effects of storage duration and structures on sesame seed germination, mold growth, and mycotoxin accumulation. Toxins. 2023;15(1):39.
16. Iizumi T, et al. Rising temperatures and increasing demand challenge wheat supply in Sudan. Nat Food. 2021;2(1):19–27.
17. Kumar MV, et al. Seed deterioration in long lived sesame seeds under accelerated ageing conditions. J Pharmacogn Phytochem. 2019;8:706–9.
18. Choe E, Min DB. Mechanisms and factors for edible oil oxidation. Comp Rev Food Sci Food Saf. 2006;5(4):169–86.
19. Chung J, Choe E. Effects of sesame oil on thermooxidative stability of soybean oil. Food Sci Biotechnol. 2001;10(4):446–50.
20. Kim I, Choe E. Effects of bleaching on the properties of roasted sesame oil. J Food Sci. 2005;70(1):C48–52.
21. Taoukis PS, Labuza TP, Saguy IS. Kinetics of food deterioration and shelf-life prediction. The Editors. 1997;30.
22. International Seed Testing Association. International rules for seed testing. Rules 1999. 1999.
23. AOAC (Association of Official Analytical Chemists). Official methods of analysis. Assoc Anal Chem. 1990;62:2742.
24. Ivarson E, et al. Effects of overexpression of WRI1 and hemoglobin genes on the seed oil content of *Lepidium campestre*. Front Plant Sci. 2017;7:2032.
25. Feldsine P, Abeyta C, Andrews WH. AOAC International methods committee guidelines for validation of qualitative and quantitative food microbiological official methods of analysis. J AOAC Int. 2002;85(5):1187–200.
26. Hassan A. Effect of UV-C treatments followed by storage on the storability characteristics of sesame (*Sesamum indicum*) seeds. Res Sq. 2022. https://doi.org/10.21203/rs.3.rs-1636067/v1.
27. Østdal H, Andersen HJ, Nielsen JH. Antioxidative activity of urate in bovine milk. J Agric Food Chem. 2000;48(11):5588–92.
28. Talhaoui N, et al. Determination of phenolic compounds of 'Sikitita' olive leaves by HPLC-DAD-TOF-MS. Comparison with its parents 'Arbequina' and 'Picual' olive leaves. LWT-Food Sci Technol. 2014;58(1):28–34.
29. Waterhouse AL. Determination of total phenolics. Curr Prot Food Analy Chem. 2002;6(1):I1-1.
30. Kim D-O, Jeong SW, Lee CY. Antioxidant capacity of phenolic phytochemicals from various cultivars of plums. Food Chem. 2003;81(3):321–6.
31. Chang S-T, et al. Antioxidant activity of extracts from *Acacia confusa* bark and heartwood. J Agric Food Chem. 2001;49(7):3420–4.
32. Benzie IF, Devaki M. The ferric reducing/antioxidant power (FRAP) assay for non-enzymatic antioxidant capacity: concepts, procedures, limitations and applications. Measur Antioxid Act Capacit Rec Trends Appl. 2018;11:77–106.
33. Re R, et al. Antioxidant activity applying an improved ABTS radical cation decolorization assay. Free Radic Biol Med. 1999;26(9–10):1231–7.
34. Ryan BF, Joiner BL, Cryer JD. MINITAB handbook: update for release. Boston: Cengage Learning; 2012.
35. Kassambara A, Mundt F. Package 'factoextra'. Extract and visualize the results of multivariate data analyses. R Pack Vers. 2017;76(2):1.
36. de Lima FS, Kurozawa LE, Ida EI. The effects of soybean soaking on grain properties and isoflavones loss. LWT-Food Sci Technol. 2014;59(2):1274–82.

37. Han Z, et al. Food polyphenols and Maillard reaction: regulation effect and chemical mechanism. Crit Rev Food Sci Nutr. 2022. https://doi.org/10.1080/10408398.2022.2146653.
38. Lattanzio V. Bioactive polyphenols: their role in quality and storability of fruit and vegetables. J Appl Bot. 2003;77(5/6):128–46.
39. Zhang H, et al. Genetic analysis and QTL mapping of seed coat color in sesame (*Sesamum indicum* L.). PLoS ONE. 2013;8(5): e63898.
40. Reyes-Moreno C, et al. Hard-to-cook phenomenon in chickpeas (Cicer arietinum L): Effect of accelerated storage on quality. Plant Foods Hum Nutr. 2000;55:229–41.
41. Park C-E, et al. Changes in physicochemical characteristics of rice during storage at different temperatures. J Stored Prod Res. 2012;48:25–9.
42. Gulla S, Waghray K. Effect of storage on physico-chemical characteristics and fatty acid composition of selected oil blends. J Life Sci. 2011;3(1):35–46.
43. Lee E, Choe E. Changes in oxidation-derived off-flavor compounds of roasted sesame oil during accelerated storage in the dark. Biocatal Agric Biotechnol. 2012;1(1):89–93.
44. Rivellese AA, et al. Effects of dietary saturated, monounsaturated and n-3 fatty acids on fasting lipoproteins, LDL size and post-prandial lipid metabolism in healthy subjects. Atherosclerosis. 2003;167(1):149–58.
45. Akinoso R, Aboaba S, Olayanju T. Effects of moisture content and heat treatment on peroxide value and oxidative stability of un-refined sesame oil. Afr J Food Agric Nutr Dev. 2010;10(10): e62908.
46. Zhou W, et al. Genome-wide association analysis and transcriptome reveal novel loci and a candidate regulatory gene of fatty acid biosynthesis in sesame (*Sesamum indicum* L.). Plant Physiol Biochem. 2022;186:220–31.
47. Jain P. Microbial degradation of grains, oil seeds, textiles, wood, corrosion of metals and bioleaching of mineral ores. 2008.
48. Zhou Z, et al. Ageing of stored rice: changes in chemical and physical attributes. J Cereal Sci. 2002;35(1):65–78.
49. Ramadan MF, et al. Fat-soluble bioactives, fatty acid profile and radical scavenging activity of *Semecarpus anacardium* seed oil. J Am Oil Chem Soc. 2010;87:885–94.
50. O'brien RD. Fats and oils: formulating and processing for applications. Baco Raton: CRC Press; 2008.
51. Abdelazim AA, Mahmoud A, Ramadan-Hassanien MF. Oxidative stability of vegetable oils as affected by sesame extracts during accelerated oxidative storage. J Food Sci Technol. 2013;50:868–78.
52. Bell LN. Moisture effects on food's chemical stability. In: Barbosa-Cánovas GV, Fontana AJ, Schmidt SJ, Labuza TP, editors. Water activity in foods. Hoboken: Wiley; 2020. p. 227–53.
53. Demir I, et al. Effects of ambient storage temperature and seed moisture content on seed longevity of lettuce (*Lactuca sativa*). Am J Exp Agric. 2016;12(3):1–5.
54. Lü JM, et al. Chemical and molecular mechanisms of antioxidants: experimental approaches and model systems. J Cell Mol Med. 2010;14(4):840–60.
55. Crapiste GH, Brevedan MI, Carelli AA. Oxidation of sunflower oil during storage. J Am Oil Chem Soc. 1999;76(12):1437.
56. Bragança GCM, et al. Multivariate analysis of the conditions of temperature, moisture and storage time in the technological, chemical, nutritional parameters and phytochemical of green lentils. J Stored Prod Res. 2020;87: 101617.
57. Suja KP, Jayalekshmy A, Arumughan C. Free radical scavenging behavior of antioxidant compounds of sesame (*Sesamum indicum* L.) in DPPH• system. J Agric Food Chem. 2004;52(4):912–5.
58. Rashed K. Phytochemical and biological effects of *Sesamum indicum* L.—a review. Plantae Sci. 2022;5(1):8–11.
59. Terefe G, et al. Sesame production manual. Ethiopia: Ethiopian Institute of Agricultural Research Embassy of the Kingdom of the Netherlands; 2012.
60. Carvalho LJCB, et al. A genomic assisted breeding program for cassava to improve nutritional quality and industrial traits of storage root. Crop Breed Appl Biotechnol. 2011;11:289–96.

II

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 28 January 2025
DOI 10.3389/fpls.2025.1541656

OPEN ACCESS

EDITED BY
Muhammad Yasir,
Zhejiang Agriculture and Forestry University, China

REVIEWED BY
Feng Gao,
BASF, Canada
Muhammad Shahid Iqbal,
Chinese Academy of Agricultural Sciences, China

*CORRESPONDENCE
Mohammed Elsafy
Mohammed.Elsafy@slu.se
Mahbubjon Rahmatov
Mahbubjon.Rahmatov@slu.se

RECEIVED 08 December 2024
ACCEPTED 02 January 2025
PUBLISHED 28 January 2025

CITATION
Elsafy M, Badawi W, Ibrahim A, Hafiz Baillo E, Bajgain P, Abdelhalim TS and Rahmatov M (2025) Genome-wide association scan and candidate gene analysis for seed coat color in sesame (*Sesamum indicum* L.). *Front. Plant Sci.* 16:1541656. doi: 10.3389/fpls.2025.1541656

Genome-wide association scan and candidate gene analysis for seed coat color in sesame (*Sesamum indicum* L.)

Mohammed Elsafy[1*], Wafa Badawi[2], Ahmed Ibrahim[2], Elamin Hafiz Baillo[2], Prabin Bajgain[3], Tilal Sayed Abdelhalim[4] and Mahbubjon Rahmatov[1*]

[1]Department of Plant Breeding, Swedish University of Agricultural Sciences (SLU), Alnarp, Sweden, [2]Agricultural Research Corporation (ARC), Ministry of Agriculture, Wad Madani, Sudan, [3]Department of Agronomy and Plant Genetics, University of Minnesota, SaintPaul, MN, United States, [4]Biotechnology and Biosafety Research Center, Agricultural Research Corporation, Shambat, Khartoum North, Sudan

Introduction: Seed coat color in sesame is a crucial trait for breeding programs as it is closely associated with important characteristics such as oil content, protein levels, and disease resistance, which directly influence seed quality and market value.

Methods: This study investigates the genetic basis of seed coat color in 200 Sudanese sesame genotypes grown for two consecutive years through comprehensive phenotyping, genomic diversity analysis, genome-wide association studies (GWAS), and candidate gene discovery.

Results and discussion: Phenotypic analysis across two growing seasons revealed high heritability and significant correlations among color parameters (L*, a*, and b*), indicating strong genetic control over seed coat color. The genomic analysis identified distinct clusters among sesame accessions, with rapid linkage disequilibrium decay suggesting a high level of recombination. GWAS identified significant SNPs associated with seed coat color traits, revealing key genomic regions on chromosomes 3, 6, 9, 12, and 13. Candidate gene analysis highlighted several genes, including *DOF* zinc finger proteins and *WRKY* transcription factors, which may play essential roles in pigment biosynthesis pathways. These findings provide valuable insights for breeding programs to enhance desirable seed coat color traits in sesame.

KEYWORDS

biosynthesis, encoding proteins, GWAS, pigmentation, Sudan

1 Introduction

Sesame (*Sesamum indicum* L.) is a globally important oilseed crop cultivated in over 70 countries with a production of 6.8 million tons (FAOSTAT, 2022). Sesame is known as the 'queen of oil crops,' it is valued for its high oil content (up to 60%) and rich composition of proteins, fatty acids, and antioxidants like sesamin and sesamolin (Dar et al., 2019). Sudan, the largest producer of sesame, is considered a center of origin for this crop, making it a crucial location for genetic diversity studies and breeding efforts (Idris et al., 2023).

Among the various traits of interest in sesame breeding, seed coat color has emerged as a critical characteristic influencing consumer preferences and potential health benefits. However, sesame seeds exhibit a wide range of colors, including white, yellow, brown, and black, primarily determined by the accumulation of pigments such as chlorophyll, carotenoids, and various phenolic compounds in the seed coat (Wang et al., 2016). Recent studies have highlighted the complex relationship between seed coat color and agronomically important traits in sesame. For instance, black sesame seeds contain significantly higher levels of phenolic compounds and exhibit greater antioxidant activity than white seeds (Mi et al., 2022). Darker seeds also have higher concentrations of lignans, particularly sesamin and sesamolin, which are known for their health-promoting properties (Abbas et al., 2022). Regarding disease resistance, pigmented seed coats are associated with enhanced protection against pathogens, such as increased resistance to Fusarium wilt (Dutta et al., 2022). Additionally, seed coat color influences oil content and fatty acid composition, as white-seed varieties generally have higher oil content, while black-seed varieties often exhibit a more favorable fatty acid profile (Uzun et al., 2008; Wei et al., 2015).

The genetic basis of key traits in sesame has become easier to study due to its relatively small diploid genome ($2n = 26$), estimated at 357 Mb (Wang et al., 2016). Sequencing of the sesame genome (Wang et al., 2014) and its improved assembly and annotation (Wang et al., 2022) have greatly advanced genetic research. Recent advances in genomic technologies, particularly genome-wide association studies (GWAS), have identified genetic loci associated with traits such as oil content, fatty acid composition, and disease resistance (Wei et al., 2015; Zhao et al., 2022; Zhou et al., 2022). A previous GWAS on seed coat color in sesame identified 13 significant single nucleotide polymorphisms (SNPs) associated with this trait, including a major locus on LG4 harboring the *PPO* gene involved in melanin biosynthesis (Wei et al., 2015). Cui et al. (2021) identified 197 significant SNPs associated with seed coat color, including 30 detected across six environments and 92 candidate genes located near four of these SNPs. However, due to the complexity of seed coat color and its links to various biochemical and agronomic traits, further research is needed to understand its genetic architecture.

Understanding the genetic control of seed coat color is essential for developing cultivars with desired characteristics to meet diverse market demands. White-seed sesame is widely preferred in many markets for its perceived quality and value (Uzun et al., 2008), while black sesame is gaining popularity due to its higher antioxidant content and potential health benefits (Abbas et al., 2022). The CIELab color space, measuring lightness (L*), redness-greenness (a*), and yellowness-blueness (b*), provides a standardized method for quantifying seed coat color (Pathare et al., 2013). It has been widely recognized for its effectiveness in seed color characterization, developing a device-independent method that provides superior perception accuracy compared to Red, Green, and Blue (RGB) models (Ivanova et al., 2022). However, RGB frequently lacks the precision needed for subtle variations, which is essential in agricultural applications (Dong et al., 2018). It has been documented that CIELab is reliable in evaluating seed quality and vigor in various crop species (Armonienė et al., 2018). In addition, CIELab has been validated in other contexts, including food science and medical imaging (Dong et al., 2018).

In this study, we build upon previous GWAS efforts by leveraging a more comprehensive and genetically diverse sesame panel, combined with high-density SNP markers, to elucidate the genetic architecture underlying seed coat color. Advanced phenotyping approaches were employed to measure CIELab color parameters (L*, a*, and b*), providing a detailed assessment of seed coat color variation and enabling comprehensive analysis of its genetic contributors. The findings from this study will contribute to our understanding of pigment biosynthesis in sesame and provide valuable information for breeding programs aimed at tailoring seed coat color to meet market and nutritional demands. This study will enhance our understanding of sesame genetic diversity in Sudan, a major center of its origin, and may identify new alleles associated with seed coat color and other important traits.

2 Materials and methods

2.1 Experiment setup

The field trial was conducted over two consecutive growing seasons (2021 and 2022) at the Matuq Research Station in Gaziera State, Sudan (14°11'10"N, 32°34'48"E). A total of 200 genetically diverse sesame accessions, along with 3 control checks, were evaluated using an augmented block design. The experimental layout consisted of 8 blocks, comprising 28 plots. Within each block, 25 distinct accessions were randomly assigned, and the 3 check varieties were replicated across all blocks to provide a measure of inter-block variability.

Standard agronomic practices tailored to local conditions were carefully followed throughout the growing season. These practices included appropriate land preparation, timely sowing, optimal irrigation scheduling, and recommended fertilizer application rates. Pest and disease management were carried out as needed to ensure healthy plant growth and development. Upon reaching physiological maturity, the sesame plants were harvested manually, and seeds were carefully extracted, cleaned, and dried to a uniform moisture content. Subsequently, the seed samples were stored under controlled environmental conditions to maintain seed quality and viability until laboratory analysis could be performed.

2.2 Seed coat color measurement and data analysis

Seed coat color parameters were quantified using (Chroma Meter CR 400, manufactured by Minolta, Japan). This device measures color in the CIELab color space, where L* represents lightness (0 = black, 100 = white), a* indicates the red-green spectrum (-60 = green, +60 = red), and b* denotes the blue-yellow spectrum (-60 = blue, +60 = yellow). Before measurements, the colorimeter was calibrated using a standard white reflector plate (Y = 93.7, x = 0.3160, y = 0.3323) to ensure accuracy. For each sample, 50.0 g of sesame seeds were carefully placed into a clean, dry Petri dish attached to the colorimeter, ensuring a uniform layer with complete coverage of the measurement area. Three replicate measurements were taken for each sample, rotating the Petri dish 120° between readings to account for any potential heterogeneity in seed color distribution (Elsafy et al., 2024).

2.3 Statistical analyses

The relationships and distributions of color traits across two years (2021 and 2022) were examined using correlation analysis and scatter plots in the R ' *GGally*' package (Schloerke et al., 2021). The Principal Component Analysis (PCA) biplot for the seed coat color attributes was created using '*factoextra*' R package (Kassambara and Mundt, 2016).

The broad-sense heritability (H) for sesame seed coat color was calculated using:

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

Where σ_G^2 is represents the genetic variance, σ_{GL}^2 is denotes the genotype by year interaction variance, σ_E^2 is the residual from environmental variance, and L is the number of years.

2.4 Genetic material preparation and sequencing

From each line, a circular section of young leaf tissue, approximately 5 mm in diameter, was harvested from each plant and placed into a 96-well plate designed for tissue collection. Genomic DNA was extracted using the Qiagen BioSprint 96 system alongside the Qiagen BioSprint DNA Plant kit. DNA was normalized to ng/μL concentration, and sequencing libraries were prepared using a genotyping-by-sequencing (GBS) 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, sequenced on a NovaSeq 6000 (Illumina, San Diego, CA, USA). Sequencing of the DNA libraries was done at the University of Minnesota Genomics Center (St. Paul, MN, USA).

Generated sequencing data was 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 *Sesamum indicum* updated genome assembly and annotations (Wang et al., 2022) with the Burrow-Wheelers Alignment (BWA) tool version 0.7.4 (Li and Durbin, 2009). Genome-wide SNPs were identified using Samtools and bcftools (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 3,636 SNPs distributed among the 13 chromosomes and 17 high-confidence scaffolds.

2.5 Genetic diversity, population structure, and marker density analysis

Genetic relationships among accessions were assessed using TASSEL v5.2.60 (Bradbury et al., 2007). The genetic similarity matrix was computed using the identity-by-state (IBS) algorithm with 10,000 bootstraps. The resulting matrix was visualized as a heatmap using the '*heatmap*' package in R (Kolde and Kolde, 2015). The Bayesian clustering approach was implemented to elucidate population structure using ADMIXTURE v1.3.0 (Alexander et al., 2009). The optimal number of ancestral populations (K) was determined by running the analysis for K values ranging from 1 to 10, with 10 independent runs for each K. The best K was selected based on the lowest cross-validation error. Results were visualized using the '*pophelper*' R package (Francis, 2017). To evaluate LD decay, pairwise linkage disequilibrium between markers was calculated using Tassel 5, utilizing a sliding window method with 50 markers. The LD values, represented as r^2, were graphed against physical distances derived from the Sesame genome V.3.0 reference. A locally weighted scatterplot smoothing (LOWESS) curve was used to visualize the LD decay pattern, and the LD decay distance was estimated following the method described by Hill and Weir (1988), and the SNP linkage disequilibrium (LD) heatmap physical length and the number of SNPs within 1Mb were estimated using the SRPlot interface (Tang et al., 2023).

2.6 Seed coat color traits association and candidate gene search

GWAS was conducted using GAPIT 3 in R 4.3.2, employing the Fixed and random model Circulating Probability Unification (FarmCPU) method (Liu et al., 2016; Wang and Zhang, 2021). Significant markers were identified based on the Bonferroni-corrected threshold (α = 0.01). For individual SNPs, this corresponded to a *p*-value of roughly determined using a *cutoff* calculated as the total number of markers (3636) divided by 1000. This yielded a threshold corresponding to a logarithm of the odds (LOD) score of approximately 3, which is presented as Manhattan and QQ plots.

Searching for candidate genes was conducted by examining the regions surrounding significant SNP markers to identify genes

10.3389/fpls.2025.1541656

potentially influencing seed coat color. We analyzed protein-coding sequences within 409,780 bp of significant loci based on the average linkage disequilibrium in sesame (204,890 bp). Using a refined sesame genome assembly (Wang et al., 2022), we conducted a protein BLAST search on the NCBI clustered nr database (Coordinators, 2015). We focused on *Sesamum indicum* sequences with >80% identity and E-values ≤1E-10, retaining the top three alignments for each sequence. These were then filtered to identify candidate genes that regulate seed coat color.

3 Results

3.1 Seed coat color phenotyping

Analysis of sesame seed colorimetric parameters (L*, a*, and b*) across two growing seasons (2021 and 2022) revealed exceptionally high consistency between years ($r = 0.997$, $p < 0.001$), indicating solid genetic control over seed coat color (Figure 1A). The lightness (L*) values demonstrated the most comprehensive range, suggesting significant variability in seed coat brightness across accessions. Importantly, we observed moderate negative correlations between L* and a* values (r = -0.42, $p < 0.001$) and moderate positive correlations between L* and b* values ($r = 0.37$, $p < 0.001$), which indicates that lighter seeds tend to be less red but more yellow. The a* and b* values showed moderate positive correlations ($r = 0.47$, $p < 0.001$), indicating that reddish seeds also tend to be more yellow. On the other hand, the PCA showed significant insights into seed coat color variation, where the first two principal components account for a substantial 88.4% of the total variance of Dim1: 47.2%, Dim2: 41.2% (Figure 1B). Three distinct groups emerge a high b* group characterized by greater yellowness, a high a* group showing increased redness, and a high L* group representing lighter seeds. Interestingly, considerable overlap exists between the high L* and high b* groups, showing a positive correlation between seed lightness and yellowness. The broad sense heritability estimates for the seed coat color traits were remarkably high, with H values of 0.9991, 0.9975, and 0.9974 for L*, a*, and b*, respectively.

3.2 Genetic diversity and structure analysis

The genomic landscape of the sesame accessions is characterized by a heterogeneous distribution of genetic variants across the 13 chromosomes (Figure 2A). Chromosomes 5, 7, and 10 exhibit regions of high variant density, as indicated by the red and orange bands. Population structure analysis (Figure 2B) reveals a significant division of sesame accessions into two distinct clusters, as determined by the ΔK method, with the optimal K = 2 indicating clear genetic differentiation. Admixture analysis further supports this, showing distinct proportions of genotype patterns between the two subpopulations. The kinship heatmap (Figure 2C) illustrates the genetic relatedness among the sesame accessions, with distinct blocks indicating varying degrees of relatedness, highlighting the presence of genetically similar groups within the sesame population. Linkage disequilibrium (LD) decay analysis (Figure 2D) demonstrates a rapid decline in LD with increasing physical distance, with an r^2 value of 0.1 at a physical distance of approximately 0.204 Mb.

3.3 Seed coat color traits association

GWAS identified several SNPs associated with the seed coat color traits L*, a*, and b* (Table 1) and Manhattan plot (Figure 3A). For the L* trait, significant associations were detected on chromosomes 3, 6, and 12. The most relevant markers were located at positions 16,523,829 bp and 16,523,899 bp on chromosome 12 (allele G/A), with a p-value of 0.0009, explaining 6.51% of the phenotypic variance. For a* trait, significant SNPs were predominantly located on chromosome 3, including the marker at position 15960455 bp (allele C/A) with a p-value of 0.0004, explaining 7.17% of the phenotypic variance. Additionally, a highly significant marker

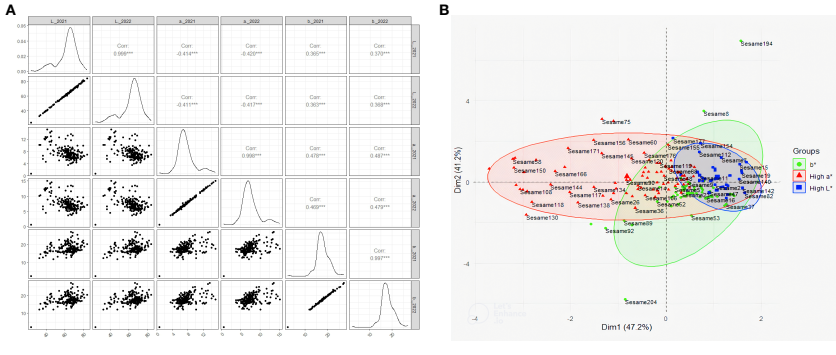


FIGURE 1
(A) Correlation of seed coat color traits of the sesame accessions across two consecutive seasons, 2021 and 2022. **(B)** Principle Component Analysis (PCA) of seed coat color traits of the sesame accessions.

10.3389/fpls.2025.1541656

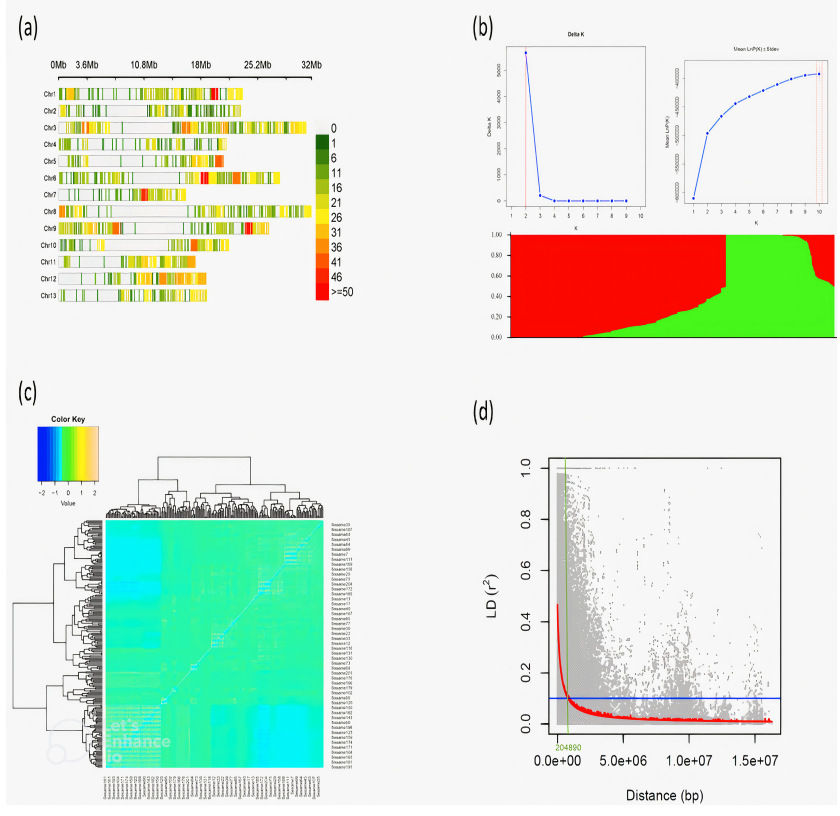


FIGURE 2
(A) Marker density across the 13 chromosomes of the sesame accessions. **(B)** Population structure analysis estimated by Delta K and LnP(K) values and heatmap shows the structure of subpopulations in sesame accessions. **(C)** Kinship heatmap showing the population relationships among the sesame accessions based on additive relationships. **(D)** linkage disequilibrium (LD) decay across genomic distance in a population of sesame accessions.

was identified on chromosome 6 at position 27694080 bp (allele T/G) with a p-value of less than 0.0001, accounting for 9.20% of the phenotypic variance. For the b* trait, significant associations were identified on chromosomes 9 and 13. The SNPs at positions 345,249 bp (allele G/A) and 345,322 bp (allele T/G) on chromosome 13 showed strong associations with p-values of 0.0013 and 0.0010, respectively, explaining over 6% of the phenotypic variance. This genetic association for seed coat color traits was confirmed by the QQ plot (Figure 3B), which considered population structure and quality control factors.

3.4 Searching for candidate genes

The candidate gene searching analysis identified several key genes associated with seed coat color space traits in this study (Table 2). *APMJ01001391*, encoding the *DOF3.1* zinc finger protein (80.7% identity), and *APMJ01001731*, encoding STY8-like isoform X2 (100% identity), were annotated on chromosome 3, both highly associated with trait a*. On chromosome 6, two genes were identified: *APMJ01003628*, encoding histidine-containing phosphotransfer protein 4-like isoform X2 (100% identity), associated with trait L*, and *KAK4407764*, encoding serine/threonine-protein kinase *STY8* (98.2% identity), associated with trait a*. As a result, all genes identified were found in *Sesamum indicium* except *KAK4407764*, which was found in *Sesamum angolense*. Chromosome 6 also contains *APMJ01003151*, encoding *WRKY* transcription factor 23 (96% identity). Further, *APMJ01007050*, encoding salicylic acid-binding protein 2 (98.5% identity), on chromosome 9 was associated with trait b*, while *APMJ01006505*, encoding Squamosa promoter-binding protein 1 (99.2% identity), on chromosome 12, was linked to trait L*.

10.3389/fpls.2025.1541656

TABLE 1 Genome-wide detection of genetic markers associated with seed coat color traits in 200 sesame accessions.

Trait	SNP marker	Chr	Pos (bps)	Alleles	p-value	LOD	MAF	R^2 (%)	Allelic effect
L*	*Chr12_16523829*	12	16523829	G/A	0.0009	3.0225	0.34	6.51	-3.80
L*	*Chr12_16523891*	12	16523891	C/T	0.0013	2.9002	0.33	6.26	3.73
L*	*Chr12_16523899*	12	16523899	G/A	0.0009	3.0225	0.34	6.51	-3.80
L*	*Chr3_16244425*	3	16244425	A/G	0.0010	3.0098	0.13	6.48	5.11
L*	*Chr6_6974622*	6	6974622	G/A	0.0010	3.0175	0.10	6.50	5.58
a*	*Chr3_15951803*	3	15951803	T/C	0.0007	3.1481	0.06	6.76	1.23
a*	*Chr3_15960455*	3	15960455	C/A	0.0004	3.3560	0.05	7.17	1.36
a*	*Chr3_15984070*	3	15984070	A/G	0.0004	3.3909	0.06	7.24	-1.27
a*	*Chr3_15984721*	3	15984721	A/G	0.0004	3.3909	0.06	7.24	-1.27
a*	*Chr3_15984975*	3	15984975	A/G	0.0004	3.3909	0.06	7.24	-1.27
a*	*Chr3_16249093*	3	16249093	G/C	0.0011	2.9629	0.06	6.39	1.13
a*	*Chr3_16593829*	3	16593829	C/A	0.0001	3.9078	0.13	8.26	-0.98
a*	*Chr3_26242291*	3	26242291	C/T	0.0002	3.6894	0.25	7.83	-0.74
a*	*Chr4_3654235*	4	3654235	A/C	0.0004	3.4379	0.18	7.34	1.03
a*	*Chr4_3654271*	4	3654271	G/A	0.0011	2.9605	0.18	6.38	-0.96
a*	*Chr4_3654307*	4	3654307	T/A	0.0002	3.7045	0.17	7.86	-1.10
a*	*Chr6_20862610*	6	20862610	A/G	0.0002	3.7350	0.19	7.92	0.76
a*	*Chr6_27694080*	6	27694080	T/G	0.0040	4.4023	0.04	9.20	-1.66
b*	*Chr13_345249*	13	345249	G/A	0.0013	2.8810	0.02	6.22	-4.59
b*	*Chr13_345322*	13	345322	T/G	0.0010	2.9834	0.06	6.43	-2.60
b*	*Chr9_23287055*	9	23287055	G/A	0.0011	2.9439	0.35	6.35	2.24

This table shows the details of single nucleotide polymorphism (SNP) markers significantly associated with key traits, including L*, a*, and b*, in sesame accessions. The information provided includes the SNP marker ID, (Chr) chromosome and (Position/bps) physical position of the marker, the alleles, (MAF)minor allele frequency, (LOD) logarithm of odds score, (R^2%) proportion of phenotypic variance explained, and the estimated effect size of the associated allele.

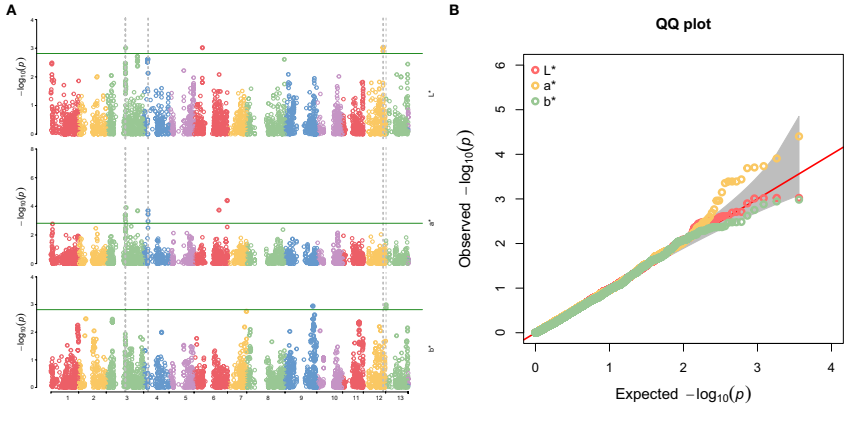


FIGURE 3
(A) Manhattan Plot for SNPs associations with seed coat color traits (L*, a*, and b*) in sesame accessions. **(B)** QQ plot showing deviation from the null hypothesis of associated seed coat color traits.

TABLE 2 Identified genes associated with L*, a*, and b*in sesame accessions.

NCBI candidate genes	Trait	SNP	Species	Annotation	E-value	% identity
APMJ01001391	a*	*Chr3_15984975*	*Sesamum indicum*	dof zinc finger protein DOF3.1-like	3e-123	80.7%
APMJ01001731	a*	*Chr3_26242291*	*Sesamum indicum*	serine/threonine-protein kinase STY8-like isoform X2	0.0	100%
APMJ01003628	L*	*Chr6_6974622*	*Sesamum indicum*	histidine-containing phosphotransfer protein 4-like isoform X2	2e-99	100%
KAK4407764	a*	*Chr6_20862610*	*Sesamum angolense*	Serine/threonine-protein kinase STY8	0.0	98-2%
APMJ01003151	a*	*Chr6_27694080*	*Sesamum indicum*	probable WRKY transcription factor 23	2e71	96%
APMJ01007050	b*	*Chr9_23287055*	*Sesamum indicum*	salicylic acid-binding protein 2	2e-89	98.5%
APMJ01006505	L*	*Chr12_16523829*	*Sesamum indicum*	Squamosa promoter-binding protein 1	1e-88	99.2%

4 Discussion

This study underlines the significant genetic basis of sesame seed coat color, as demonstrated by high heritability estimates for traits L*, a*, and b*. These findings confirm that seed coat color is stable across environments, with minimal influence from external factors, making it predominantly controlled by genetic factors. Such stability is essential for breeding programs, ensuring consistency in expressing traits across various cultivation conditions. Significant SNPs associated with the L* trait on chromosomes 3, 6, and 12, and with the b* trait on chromosomes 9 and 13, mark genomic regions of interest for further research. These loci, which explain a substantial proportion of phenotypic variance, provide a foundation for breeding strategies to tailor seed coat colors to specific market demands.

4.1 Seed coat color phenotyping

Our study demonstrated high year-to-year consistency in sesame seed coat color ($r = 0.997$, $p < 0.001$), indicating that the trait is stable, heritable, and minimally influenced by the environment. This finding aligns with previous research identifying key quantitative trait loci (QTLs) associated with seed coat color in sesame. Du et al. (2019) developed a high-density genetic map and found that seed coat color is primarily influenced by a few major genes and several QTLs, which significantly contribute to its heritability. Because sesame domestication has led to lighter seed colors, which are largely determined by genetic loci (Wei et al., 2016). Results demonstrate that the first two principal components account for 88.4% of the total variance in sesame seed coat color among Sudanese genotypes, indicating a high level of genetic variation within these traits. Based on the substantial variance explained by these components, a limited number of genetic factors are responsible for seed coat color, and distinct phenotypic groups can be formed, explaining genetic differentiation between genotypes. Our study revealed exceptionally high heritability (H) for L*, a*, and b* values, indicating that genetic factors rather than environmental influences predominantly control seed coat color traits. This finding aligns with previous studies, such as Wang et al. (2016), which reported on the fine mapping of plant height and seed coat color quantitative trait loci (QTLs) using a new high-density genetic map, and Du et al. (2019), who constructed a high-density genetic map using specific length amplified fragment (SLAF) sequencing and conducted QTL mapping of seed-related traits in sesame. Building on this, Sabag et al. (2021) emphasized that straightforward selection strategies can significantly enhance the genetic architecture of sesame seed coat color. Furthermore, since these traits are primarily governed by additive genetic variance, Cui et al. (2021) suggested that selection for these traits in breeding programs could be highly effective.

4.2 Genomic diversity and structure analysis

This study highlighted the heterogeneous distribution of genetic variants across the sesame chromosomes, particularly the high variant density on chromosomes 5, 7, and 10, provided a valuable framework for understanding the genetic basis of seed coat color and other agronomic traits using improved assembly and annotation of the sesame genome. The findings from various studies underscore the potential for utilizing this genetic information in breeding programs to enhance sesame quality and yield, and future research should continue to explore the functional implications of these genetic variants and their interactions with environmental factors to optimize sesame cultivation (Wu et al., 2014; Zhao et al., 2022; Zhou et al., 2022).

Our study revealed significant separation into two distinct clusters among sesame accessions, as determined by the ΔK method, which provides valuable insights into the genetic diversity and evolutionary history of sesame where hybridization and gene flow between populations can occur. Understanding the population structure and genetic differentiation among sesame accessions is critical to improving specific traits through breeding programs, such as seed coat color (Pandey et al., 2013). Strategically selecting parental lines representing diverse genetic backgrounds by identifying distinct genetic clusters can enhance hybrid vigor and trait improvement. Further, population structure can be used to

conserve genetic resources, making it easier for future breeding efforts to preserve various genetic materials (Dossa et al., 2016). The kinship heatmap result in this study showed a visual representation of the genetic relationships among the sesame accessions, where distinct blocks indicate varying degrees of genetic similarity. Such a heatmap is instrumental in identifying clusters of closely related accessions, which can indicate shared ancestry or common breeding practices (Eynard et al., 2016).

This study found that LD decay was rapid with increasing physical distance, which provides insight into sesame's genetic architecture. LD describes the non-random association of alleles at different loci, and its decay over distance can provide insight into recombination rates and the dynamics of historical populations. The observed r^2 value of 0.1 at a physical distance of approximately 0.204 Mb in this study indicates a relatively short-lived linkage disequilibrium (LD) in sesame, suggesting a high level of recombination within the genome. Comparatively, previous studies using the first version of the sesame genome (Wang et al., 2014) have reported varying patterns of LD in different sesame populations compared to our study that utilized the updated version of the sesame genome (Wang et al., 2022). For instance, Wang et al. (2014) found that LD in sesame decayed to an r^2 of 0.15 over a distance of around 150 kb, indicating the same trend of rapid decay but over a longer distance. According to Du et al. (2019), LD in a diverse set of sesame accessions exhibits a significant decline, with r^2 values dropping to around 0.1 at distances exceeding 100 kb. This finding is consistent with the short-term nature of LD in this study.

According to these comparisons, LD decay in sesame populations can vary. However, the trend still indicates a high level of recombination, and this will benefit breeding programs because it facilitates the introduction of genetic diversity and the selection of desirable traits. The LD decay results have important implications for breeding and genetic studies in sesame. For instance, Francia et al. (2005) and Dubcovsky (2004) suggested that rapid LD decay marker-assisted selection (MAS) could be effectively employed in breeding programs, as the genetic markers associated with desirable traits are likely to be closely linked to the target genes. However, The LDs observed in this study were relatively short, supporting the potential for efficient marker-assisted selection strategies in sesame breeding.

4.3 SNPs-trait association analysis

Our study identified significant associations for the L* trait on chromosomes 3, 6, and 12, with the strongest markers on chromosome 12 at positions 16523829 bp and 16523899 bp. These findings partially align with previous reports, such as Wang et al. (2020), where transcriptome analysis identified genes associated with flavonoid biosynthesis pathways involved in light pigmentation. Similarly, Li et al. (2021) identified QTL hotspots for L* on chromosome 12 using an F_2 population derived from Chinese accessions. Another study by Cui et al. (2021) focused on brown seed coat traits primarily associated significant SNPs with chromosome 6, while this study identifies both chromosome 6 and 12 loci for the L* trait. This difference might be attributable to the genetic diversity of the Sudanese sesame germplasm and the different reference genome used in this study (Wang et al., 2022). Moreover, the Sudanese sesame accessions, known for their adaptation to arid climates, may harbor unique alleles shaped by local selective pressures.

The GWAS results in this study identified significant SNPs for a* trait on chromosome 3, with a highly significant marker on chromosome 6 (position 27694080 bp) explaining 9.20% of the phenotypic variance. This finding aligns with the work of Wang et al. (2023), where major-effect QTLs for red pigmentation traits were also mapped to a 1.19 Mb interval on chromosome 6 (qBSCchr6). Additionally, genes associated with anthocyanin biosynthesis pathways, such as MYB and bHLH transcription factors, were identified in several studies as contributors to red pigmentation. While chromosome 3 emerged as a key region in this study, it is less frequently highlighted in Wang et al. (2023) and Dutta et al. (2022) using Chinese or Indian germplasm. Moreover, the SNPs on chromosome 6 in this study account for higher phenotypic variance than those reported in Cui et al. (2021) and Li et al. (2021), and this could be attributed to the distinct geographic origin of the Sudanese accessions likely plays a role. Different selective pressures, such as high UV exposure, may affect anthocyanin pigmentation in Sudanese sesame, leading to a stronger association on chromosome 3. Additionally, differences in linkage disequilibrium (LD) patterns between populations may result in unique SNP-trait associations in this study compared to those in Asian germplasm.

For the b* trait, our study revealed significant associations on chromosomes 9 and 13 were identified, with markers on chromosome 13 (positions 345249 bp and 345322 bp) showing strong associations. These findings are consistent with prior studies, such as Cui et al. (2021), which identified regions on chromosome 9 linked to seed coat pigmentation. Transcriptomic studies have further implicated pathways like carotenoid biosynthesis in yellow pigmentation traits. However, this study identified chromosome 13 markers as key contributors to b*, which is unique compared to previous studies (Cui et al., 2021; Du et al., 2019). Most studies highlight chromosomes 6 and 9 as major contributors to yellow pigmentation in non-Sudanese sesame. However, The prominence of chromosome 13 in our study may reflect unique genetic adaptations of Sudanese accessions. Environmental stressors such as drought and heat in Sudan may have shaped seed coat characteristics, favoring alleles in less prominent loci. The local cultural preferences for seed coat colors could also indirectly influence breeding practices, leading to unique allele frequencies.

4.4 Searching for candidate genes

This study identified candidate genes likely to play important roles in modulating pigment-related traits via their involvement in

regulatory and signaling pathways. Their functional characterization could uncover the molecular basis of pigmentation in *Sesamum indicum* and related species. The *DOF Zinc Finger* Protein encoded by *APMJ01001391* has been implicated in regulating light-responsive genes. Iorizzo et al. (2019) stated that the effects of light on flavonoid biosynthesis are crucial as light triggers the production of anthocyanins. As a result of its association with trait a*, this gene might be involved in modulating pigmentation by activating flavonoid pathway genes.

Phosphorylation-dependent signaling pathways require Serine/Threonine Kinases identified in *APMJ01001731* and *KAK4407764*, and Kinases such as *STY8* control carotenoid metabolism and plastid biogenesis (Mazur et al., 2021). This strong correlation indicates that these genes may regulate carotenoid synthesis or other pigment-related processes. Regarding *WRKY*, transcription factors, such as *WRKY23* encoded by *APMJ01003151*, play a critical role in secondary metabolism regulation under stress conditions (Meraj et al., 2020). As anthocyanin serves as a protective pigment, *WRKY23* may modulate anthocyanin biosynthesis under environmental stress. According to Chen et al. (2019), *WRKY* transcription factors are critical for stress-induced pigment accumulation in different species.

Histidine-containing phosphotransfer proteins (*HPTs*) are essential in the cytokinin signaling pathway. Cytokinin can affect pigment biosynthesis by affecting plasid development and secondary metabolite pathways (Cortleven and Schmülling, 2015). It also aligns with findings from other studies, where cytokinin signaling indirectly affects pigment accumulation through developmental cues (Wu et al., 2021). Furthermore, *SABP2*, encoded by *APMJ01007050*, is implicated in signal transduction in salicylic acid pathways, influencing flavonoid biosynthesis during biotic and abiotic stresses (Shaukat et al., 2022). This trait may mediate stress-mediated pigment accumulation due to its association with trait b*. One of the most notable discoveries is the identification of Squamosa Promoter-Binding Protein 1 (*SBP1*) encoded by *APMJ01006505*, which is linked to trait L*. According to Sánchez-Retuerta et al. (2018), *SBP* proteins regulate gene expression via light-mediated pigment pathways, and because of their high identity and significant annotation, *SBP1* may directly regulate genes involved in carotenoid biosynthesis.

4.5 Implications on breeding

The findings from this study have significant implications for plant breeding, particularly for improving desirable agronomic characteristics. The identification of high heritability and strong genetic correlations among color parameters (L*, a*, and b*) suggests that seed coat color can be effectively selected for breeding programs, facilitating the improvement of oil content and disease resistance traits that are closely associated with color. GWAS results identified significant SNPs linked to seed coat color traits, which can be used for marker-assisted selection to accelerate breeding. Moreover, discovering genes involved in pigment biosynthesis pathways, such as DOF zinc finger proteins and WRKY transcription factors, creates new opportunities for genetic manipulation and targeted breeding methods to optimize seed quality. The findings enhance our understanding of the genetic architecture of seed coat color and provide a strong framework for developing sesame varieties with improved market value and agronomic performance.

5 Conclusions

The results of this study underscore the importance of genetic factors in determining seed coat color in sesame, with high heritability estimates confirming the stability of these traits across environments. Identifying significant SNPs associated with color traits offers potential markers for marker-assisted selection in breeding programs. Furthermore, the discovery of candidate genes involved in pigment biosynthesis pathways provides a foundation for future functional studies to elucidate the molecular mechanisms underlying seed coat color variation. This study contributes to understanding genetic diversity in sesame and highlights the potential for targeted breeding strategies to improve seed quality and marketability based on color traits. Future studies should focus on the functional characterization of the identified candidate genes and their interactions with the environment to optimize sesame cultivation and enhance its agronomic value.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA1184775.

Author contributions

ME: Conceptualization, Formal Analysis, Investigation, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. WB: Data curation, Writing – review & editing. AI: Data curation, Writing – review & editing. EB: Data curation, Writing – review & editing. PB: Supervision, Visualization, Writing – review & editing. TA: Funding acquisition, Supervision, Visualization, Writing – review & editing. MR: Funding acquisition, Supervision, Visualization, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. Swedish Research Council (Vetenskapsrådet) grant number (DR-2020-04163).

Acknowledgments

Our sincere thanks go to the Agricultural Research Corporation-Sudan for providing research facilities and to the Swedish University of Agricultural Sciences for laboratory support and covering the publication fee.

Conflict of interest

WB, AI, EB and TA were employed by Agricultural Research Corporation.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abbas, S., Sharif, M. K., Sibt-e-Abbas, M., Fikre Teferra, T., Sultan, M. T., and Anwar, M. J. (2022). Nutritional and therapeutic potential of sesame seeds. *J. Food Qual.* 2022, 6163753. doi: 10.1155/2022/6163753

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109

Armonienė, R., Odilbekov, F., Vivekanand, V., and Chawade, A. (2018). Affordable imaging lab for noninvasive analysis of biomass and early vigor in cereal crops. *BioMed. Res. Int.* 2018 (1), 5713158. doi: 10.1155/2018/5713158

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308

Chen, X., Li, C., Wang, H., and Guo, Z. (2019). WRKY transcription factors: evolution, binding, and action. *Phytopathol. Res.* 1, 1–15. doi: 10.1186/s42483-019-0022-x

Coordinators, N. R. (2015). Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 43 (D1), D6–D17. doi: 10.1093/nar/gkv1290

Cortleven, A., and Schmülling, T. (2015). Regulation of chloroplast development and function by cytokinin. *J. Exp. Bot.* 66, 4999–5013. doi: 10.1093/jxb/erv132

Cui, C., Liu, Y., Liu, Y., Cui, X., Sun, Z., Du, Z., et al. (2021). Genome-wide association study of seed coat color in sesame (*Sesamum indicum* L.). *PloS One* 16, e0251526. doi: 10.1371/journal.pone.0251526

Dar, A. A., Kancharla, P. K., Chandra, K., Sodhi, Y. S., and Arumugam, N. (2019). Assessment of variability in lignan and fatty acid content in the germplasm of *Sesamum indicum* L. *J. Food Sci. Technol.* 56, 976–986. doi: 10.1007/s13197-018-03564-x

Dong, C., Liang, G., Hu, B., Yuan, H., Jiang, Y., Zhu, H., et al. (2018). Prediction of congou black tea fermentation quality indices from color features using non-linear regression methods. *Sci. Rep.* 8, 10535. doi: 10.1038/s41598-018-28767-2

Dossa, K., Wei, X., Zhang, Y., Fonceka, D., Yang, W., Diouf, D., et al. (2016). Analysis of genetic diversity and population structure of sesame accessions from Africa and Asia as major centers of its cultivation. *Genes* 7, 14. doi: 10.3390/genes7040014

Du, H., Zhang, H., Wei, L., Li, C., Duan, Y., and 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 Biol.* 19, 1–20. doi: 10.1186/s12870-019-2172-5

Dubcovsky, J. (2004). Marker-assisted selection in public breeding programs: the wheat experience. *Crop Sci.* 44, 1895–1898. doi: 10.2135/cropsci2004.1895

Dutta, D., Harper, A., and Gangopadhyay, G. (2022). Transcriptomic analysis of high oil-yielding cultivated white sesame and low oil-yielding wild black sesame seeds reveal differentially expressed genes for oil and seed coat color. *Nucleus* 65, 151–164. doi: 10.1007/s13237-022-00389-0

Elsafy, M., Ekholm, A., Elkhatim, K. A. S., Hamid, M. G., Othman, M. H., Abdelhalim, T. S., et al. (2024). Tracking the storage stability in sesame (*Sesamum indicum* L.): impact of accelerated storage on storability characteristics, seed quality, phytochemical content, and fatty acids. *Discov. Agric.* 2, 55. doi: 10.1007/s44279-024-00077-4

Eynard, S. E., Windig, J. J., Hiemstra, S. J., and Calus, M. P. (2016). Whole-genome sequence data uncover loss of genetic diversity due to selection. *Genet. Select. Evol.* 48, 1–13. doi: 10.1186/s12711-016-0210-4

FAOSTAT (2022). *Crops and livestock products*. (The Food and Agriculture Organization (FAO). Available online at: https://www.fao.org/faostat/en/#data/QCL

Francia, E., Tacconi, G., Crosatti, C., Barabaschi, D., Bulgarelli, D., Dall'Aglio, E., et al. (2005). Marker assisted selection in crop plants. *Plant Cell Tissue Organ Cult.* 82, 317–342. doi: 10.1007/s11240-005-2387-z

Francis, R. (2017). pophelper: an R package and web app to analyze and visualize population structure. *Mol. Ecol. Resour.* 17, 27–32. doi: 10.1111/men.2017.17.issue-1

Hill, W., and Weir, B. (1988). Variances and covariances of squared linkage disequilibria in finite populations. *Theor. Popul. Biol.* 33, 54–78. doi: 10.1016/0040-5809(88)90004-4

Idris, A. E., Abuali, A. I., Siddig, M. A. E., Abdelsalam, A. K., Osman, B. M. E., and Omer, O. A. M. (2023). Evaluation of genetic diversity of sesame (*Sesamum indicum* L.) genotypes, using agro morphological and molecular markers (RAPD). *Asian J. Res. Crop Sci.* 8, 51–61. doi: 10.9734/ajrcs/2023/v8i3168

Iorizzo, M., Cavagnaro, P. F., Bostan, H., Zhao, Y., Zhang, J., and Simon, P. W. (2019). A cluster of MYB transcription factors regulates anthocyanin biosynthesis in carrot (Daucus carota L.) root and petiole. *Front. Plant Sci.* 9, 1927. doi: 10.3389/fpls.2018.01927

Ivanova, M., Myroshnychenko, M., Olkhovsky, V., Grygorian, E., Fedulenkova, Y., Khara, G., et al. (2022). Analysis of color properties of raster images of histological microspecimens: own research experience. *Медичні перспективи* 27, 9–15. doi: 10.26641/2307-0404.2022.1.254314

Kassambara, A., and Mundt, F. (2016). *Factoextra: extract and visualize the results of multivariate data analyses, R package version 1.0. 7. 2020*. In *R Package Version*. Available online at: https://cir.nii.ac.jp/crid/1370004235968325765.

Kolde, R., and Kolde, M. R. (2015). Package 'pheatmap'. *R Package* 1, 790. Available online at: https://cran.ms.unimelb.edu.au/web/packages/pheatmap/pheatmap.pdf

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

Li, C., Duan, Y., Miao, H., Ju, M., Wei, L., and Zhang, H. (2021). Identification of candidate genes regulating the seed coat color trait in sesame (*Sesamum indicum* L.) using an integrated approach of QTL mapping and transcriptome analysis. *Front. Genet.* 12, 700469. doi: 10.3389/fgene.2021.700469

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

Liu, J., Yang, C., Shi, X., Li, C., Huang, J., Zhao, H., et al. (2016). Analyzing association mapping in pedigree-based GWAS using a penalized multitrait mixed model. *Genet. Epidemiol.* 40, 382–393. doi: 10.1002/gepi.2016.40.issue-5

Mazur, R., Maszkowska, J., Anielska-Mazur, A., Garstka, M., Polkowska-Kowalczyk, L., Czajkowska, A., et al. (2021). The SnRK2. 10 kinase mitigates the adverse effects of salinity by protecting photosynthetic machinery. *Plant Physiol.* 187, 2785–2802. doi: 10.1093/plphys/kiab438

Meraj, T. A., Fu, J., Raza, M. A., Zhu, C., Shen, Q., Xu, D., et al. (2020). Transcriptional factors regulate plant stress responses through mediating secondary metabolism. *Genes* 11, 346. doi: 10.3390/genes11040346

Mi, S., Wang, Y., Zhang, X., Sang, Y., and Wang, X. (2022). Discrimination of black and white sesame seeds based on targeted and non-targeted platforms with chemometrics: from profiling towards identification of chemical markers. *Foods* 11, 2042. doi: 10.3390/foods11142042

Pandey, S. K., Das, A., and Dasgupta, T. (2013). Genetics of seed coat color in sesame (*Sesamum indicum* L.). *Afr. J. Biotechnol.* 12 (42), 6061-6067. doi: 10.5897/AJB2013.13055

Pathare, P. B., Opara, U. L., and Al-Said, F. A.-J. (2013). Color measurement and analysis in fresh and processed foods: A review. *Food Bioprocess Technol.* 6, 36–60. doi: 10.1007/s11947-012-0867-9

Poland, J. A., Brown, P. J., Sorrells, M. E., and 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, e32253. doi: 10.1371/journal.pone.0032253

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

Sánchez-Retuerta, C., Suaréz-López, P., and Henriques, R. (2018). Under a new light: regulation of light-dependent pathways by non-coding RNAs. *Front. Plant Sci.* 9, 962. doi: 10.3389/fpls.2018.00962

Schloerke, B., Cook, D., Larmarange, J., Briatte, F., Marbach, M., Thoen, E., et al (2021). *GGally: Extension to'ggplot2'*. Available online at: https://shs.hal.science/halshs-03354562/.

Shaukat, K., Zahra, N., Hafeez, M. B., Naseer, R., Batool, A., Batool, H., et al. (2022). "Role of salicylic acid–induced abiotic stress tolerance and underlying mechanisms in plants," in *In Emerging plant growth regulators in agriculture* (Elsevier), 73–98. doi: 10.1016/B978-0-323-91005-7.00008-4

Tang, D., Chen, M., Huang, X., Zhang, G., Zeng, L., Zhang, G., et al. (2023). SRplot: A free online platform for data visualization and graphing. *PloS One* 18, e0294236. doi: 10.1371/journal.pone.0294236

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

Wang, H., Cui, C., Liu, Y., Zheng, Y., Zhao, Y., Chen, X., et al. (2023). Genetic mapping of QTLs controlling brown seed coat traits by genome resequencing in sesame (*Sesamum indicum* L.). *Front. Plant Sci.* 14, 1131975. doi: 10.3389/fpls.2023.1131975

Wang, L., Dossou, S. S. K., Wei, X., Zhang, Y., Li, D., Yu, J., et al. (2020). Transcriptome dynamics during black and white sesame (*Sesamum indicum* L.) seed development and identification of candidate genes associated with black pigmentation. *Genes* 11 (12), 1399. doi: 10.3390/genes11121399

Wang, M., Huang, J., Liu, S., Liu, X., Li, R., Luo, J., et al. (2022). Improved assembly and annotation of the sesame genome. *DNA Res.* 29, dsac041. doi: 10.1093/dnares/dsac041

Wang, L., Xia, Q., Zhang, Y., Zhu, X., Zhu, X., Li, D., et al. (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, 1–13. doi: 10.1186/s12864-015-2316-4

Wang, L., Yu, S., Tong, C., Zhao, Y., Liu, Y., Song, C., et al. (2014). Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. *Genome Biol.* 15, 1–13. doi: 10.1186/gb-2014-15-2-r39

Wang, J., and Zhang, Z. (2021). GAPIT version 3: boosting power and accuracy for genomic association and prediction. *Genom. Proteomics Bioinf.* 19, 629–640. doi: 10.1016/j.gpb.2021.08.005

Wei, X., Liu, K., Zhang, Y., Feng, Q., Wang, L., Zhao, Y., et al. (2015). Genetic discovery for oil production and quality in sesame. *Nat. Commun.* 6, 8609. doi: 10.1038/ncomms9609

Wei, X., Zhu, X., Yu, J., Wang, L., Zhang, Y., Li, D., et al. (2016). Identification of sesame genomic variations from genome comparison of landrace and variety. *Front. Plant Sci.* 7, 1169. doi: 10.3389/fpls.2016.01169

Wu, W., Du, K., Kang, X., and Wei, H. (2021). The diverse roles of cytokinins in regulating leaf development. *Hortic. Res.* 8, 118. doi: 10.1038/s41438-021-00558-3

Wu, K., Yang, M., Liu, H., Tao, Y., Mei, J., and Zhao, Y. (2014). Genetic analysis and molecular characterization of Chinese sesame (*Sesamum indicum* L.) cultivars using Insertion-Deletion (InDel) and Simple Sequence Repeat (SSR) markers. *BMC Genet.* 15, 1–15. doi: 10.1186/1471-2156-15-35

Zhao, M., Liu, S., Pei, Y., Jiang, X., Jaqueth, J. S., Li, B., et al. (2022). Identification of genetic loci associated with rough dwarf disease resistance in maize by integrating GWAS and linkage mapping. *Plant Sci.* 315, 111100. doi: 10.1016/j.plantsci.2021.111100

Zhou, W., Song, S., Dossou, S. S. K., Zhou, R., Wei, X., Wang, Z., et al. (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 Physiol. Biochem.* 186, 220–231. doi: 10.1016/j.plaphy.2022.07.023

III

Plant-Environment Interactions

Plant-Environment Interactions

RESEARCH ARTICLE OPEN ACCESS

Exploring the Diversity in Oil Content, Fatty Acid Profiles, and Seed Coat Color in Sudanese Sesame Germplasm: Implications for Breeding and Crop Improvement

Mohammed Elsafy[1] | Wafa Badawi[2] | Ali Zakaria[3] | Tilal Sayed Abdelhalim[4] | Mahbubjon Rahmatov[1] | Eva Johansson[1]

[1]Department of Plant Breeding, Swedish University of Agricultural Sciences (SLU), Uppsala, Sweden | [2]Matuq Research Station, Agricultural Research Corporation, Matuq, Sudan | [3]Agricultural Plant Genetic Resources Conservation and Research Center, Agricultural Research Corporation, Wad Medani, Sudan | [4]Biotechnology and Biosafety Research Center, Agricultural Research Corporation, Shambat, Khartoum North, Sudan

Correspondence: Mohammed Elsafy (mohammed.elsafy@slu.se) | Eva Johansson (eva.johansson@slu.se)

Received: 24 January 2025 | **Revised:** 17 March 2025 | **Accepted:** 25 March 2025

Funding: This work was supported by Vetenskapsrådet (DR-2020-04163).

ABSTRACT

Sesame, a key oilseed crop, thrives in arid environments and offers high-quality oils. Sudan, a major producer and center of sesame genetic diversity, remains underutilized in breeding efforts. This study analyzed 87 Sudanese sesame accessions, revealing significant variations in oil content, fatty acid composition, and seed coat color. The findings highlight the potential of Sudanese germplasm for improving oil quality and broadening trait diversity in breeding programs. Oil content ranged from 32.8% to 50.2%, with oleic acid (41.3%–47.6%) and linoleic acid (35.0%–41.4%) as the predominant fatty acids, consistent with other regions. Some samples showed exceptionally high oleic acid levels. Seed coat color varied significantly, particularly in lightness (L^*), but it showed no correlation with oil content or fatty acid composition. Its potential link to bioactive compounds warrants further study. Principal coordinates analysis showed no link between oil levels, fatty acid profiles, and the original collection sites. The findings highlight the breeding potential of Sudanese sesame germplasm, particularly for developing varieties with high unsaturated fatty acids, such as oleic acid, and diverse seed coat colors. Further studies across environments and genetic investigations are needed to ensure trait stability and optimize their use.

1 | Introduction

Sesame (*Sesamum indicum* L.) is an ancient oilseed crop of considerable economic and nutritional significance. Sesame belongs to the genus *Sesamum* in the Pedaliaceae family, which comprises approximately 38 species that thrive in tropical and subtropical regions (Kapoor et al. 2015). Although it is primarily cultivated in regions north of the equator, sesame is adaptable across latitudes between 40° N and 40° S (Dossa et al. 2016).

Archaeological evidence suggests that sesame's domestication dates back to approximately 5000 B.C. in India, with findings from the Harappa civilization supporting this timeline (Kalaiyarasi et al. 2019). By 2000 B.C., sesame had spread to Mesopotamia and the Mediterranean, becoming a crucial crop during the Bronze Age (Zech-Matterne et al. 2015). However, identifying the exact origin of sesame is often challenging because of the complex relationship between the centers of origin and diversity. While the Indian subcontinent is acknowledged for its early domestication, East Africa, particularly Sudan, and Ethiopia, is recognized as a critical region of genetic diversity for sesame, which is crucial for improving and adapting the crop to various environments (Dossa et al. 2016; Negash et al. 2020).

Sudan, one of the leading sesame producers globally, relies on sesame as a significant agricultural and economic resource. It supports local economies, particularly in rural areas, and contributes to national export revenue (Karim and Ismai 2020).

The adaptability of sesame to semi-arid conditions enhances its food security and agricultural sustainability in regions where other crops may fail (Sabiel et al. 2015). Sesame is a valuable commodity in domestic and international markets because it has high oil content and beneficial nutritional properties. Recent efforts have been focused on improving production techniques and genetic traits to enhance yield, oil content, and overall crop resilience in Sudan (Mahmoud and Khalil 2019).

Sesame oil is known for its high content of unsaturated fatty acids, particularly oleic and linoleic acids, which contribute significant health benefits, including cardiovascular protection and anti-inflammatory properties (Bhunia et al. 2016; Pathak et al. 2017). Some studies have indicated that the seed coat color might be correlated to the fatty acid content and composition in sesame seeds (Wang et al. 2020). The genetic variation underlying sesame seed coat color is primarily controlled by two major genes and several quantitative trait loci (QTLs), which modulate the biochemical pathways responsible for pigmentation. Studies have demonstrated that the inheritance of seed coat color is multifactorial, involving both additive and epistatic interactions between multiple genes (Du et al. 2019; Wang et al. 2020). Some genes are critical in the flavonoid biosynthesis pathway, which plays a critical role in determining pigmentation. Genetic variations within these pathways give rise to distinct seed coat phenotypes, such as black and white seeds (Wang et al. 2020). Black sesame seeds are particularly rich in antioxidants and valued for medicinal use, whereas white seeds are preferred for culinary applications (Cui et al. 2021). Breeding programs aim to increase the marketability of the crop and its nutritional value by optimizing the seed coat color (Li et al. 2021).

Despite the economic importance of sesame, its genetic diversity is threatened by conflicts and environmental challenges, particularly in Africa. The prolonged civil war in Sudan led to the loss of vital agricultural land and unique sesame germplasm (Bedigian and Harlan 1983). Several initiatives, including the establishment of gene banks and collaboration through networks such as the Eastern African Plant Genetic Resources Network (EAPGRN), have been crucial for preserving the genetic diversity of sesame (Abebe 2006). However, ongoing conflicts threaten these efforts, underscoring the urgent need for conservation, genetic resource management, and characterization of stored genotypes.

While studies have extensively examined sesame in other regions, evaluations of the Sudanese sesame germplasm (center of diversity) remain limited. This study aimed to analyze a diverse collection of Sudanese sesame gene bank accessions for oil content, fatty acid composition, and seed coat color to characterize stored accessions, evaluate present diversity within the crop in Sudan, and identify genotypes suitable for future breeding programs. Additionally, this study aimed to contribute knowledge of the importance of enhancing nutritional and market value while contributing to global conservation, documentation, and crop improvement efforts. To our knowledge, this is the first report on the fatty acid composition of sesame from Sudan and the diversity present for this trait within Sudanese germplasm.

2 | Material and Methods

2.1 | Plant Materials

A total of 87 Sudanese sesame accessions, preserved through long-term conservation at the Genebank of the Agricultural Plant Genetic Resources Conservation and Research Centre (APGRC) in Sudan, were evaluated in this study. The geographical distribution and associated passport data for these accessions are detailed in Table S1 and are available on Genesys (https://www.genesys-pgr.org/a/overview/v2AB7lOJABP).

Three widely grown Sudanese sesame cultivars, namely, Kenana-2, Promo, and a farmer-preferred landrace, Herheri, were included as controls for comparison, allowing for the estimation of variations in oil composition and seed coat color (Table S2). Kenana-2, released in 1990, is an early-maturing, drought-tolerant sesame cultivar characterized by its large seed size and white color (Ahmed 1997). Promo, a high-yielding, medium-maturing cultivar released in 1998, is known for its high branching and delayed shattering ability (Ahmed 2008; Ahmed et al. 2003). From farmers' knowledge, Herheri is a high-yielding, early-maturing landrace characterized by its dark brown seed coat, high oil content, and drought tolerance.

2.2 | Plant Materials and Experimental Setup

This study used 87 sesame genbank accessions collected from 9 regions across Sudan, representing a wide geographical range. The highest number of accessions came from North Kordofan, with 23 accessions, followed by Gedarif (16), West Darfur (13), Blue Nile (10), South Kordofan (8), North Darfur (5), Kassala (4), Central Darfur (4), and South Darfur (4) (Figure 1 and Table S3). According to the characterization data from the Genbank of these accessions, the seed coat color of accessions varied from white, light brown, reddish brown, brown, and dark brown to gray and black. In addition, the 1000 seed weight varied from 2 to 6 g, and the days to 50% flowering ranged from 32 to 76 days (Table S1).

The selected accessions for this study were grown in the field during the 2022–2023 season at the Matouq Research Station (latitude 14°11′10″ N, longitude 32°34′48″ E) of the Agricultural Research Corporation (ARC), Sudan. The soil at the site is a vertisol, characterized by deep cracking and high clay content.

The 87 sesame accessions were sown using an Augmented Block Design (ABD) across five blocks, incorporating three standard varieties, Kenana-2, Bromo, and Herheri, as national checks. Each block consisted of 17 entries and three checks, which were randomized within each block. Each accession was sown in a 5-m row, spaced 0.7 m apart. Seeds were drilled at a rate of 3 kg/ha. Three weeks after sowing, the crop was thinned to one plant per hole with an inter-row spacing of 7.5 cm.

At maturity, three random plants from each experimental plot were selected for colorimetric measurement, oil content analysis,

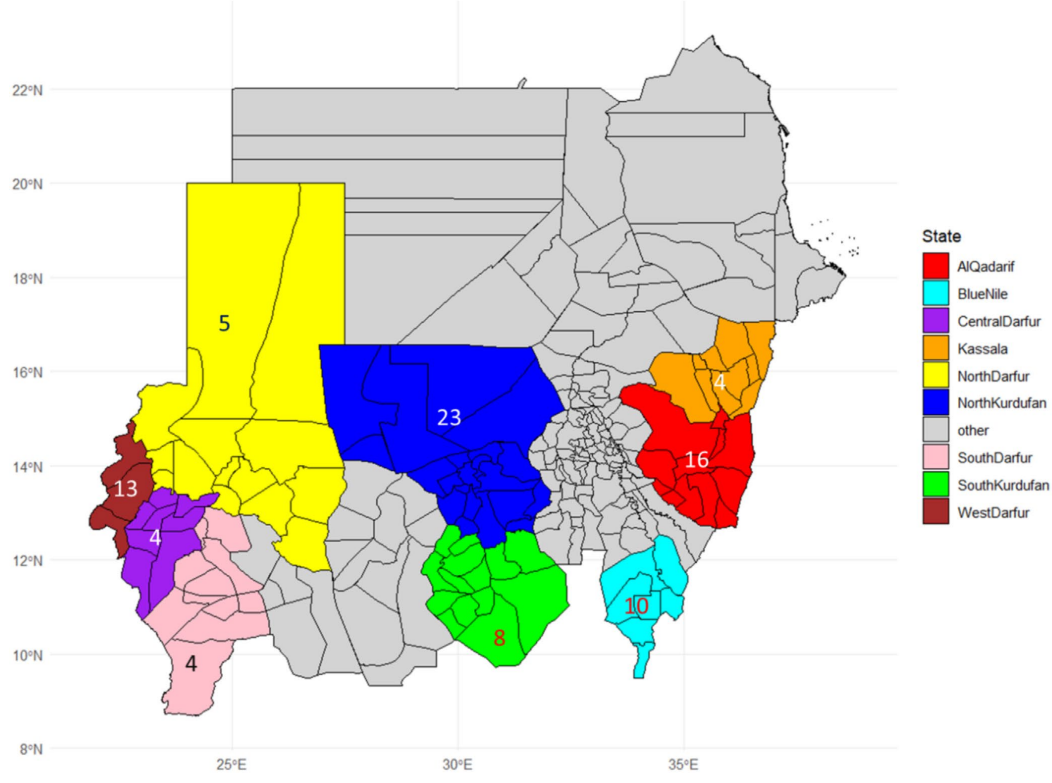


FIGURE 1 | Sudan map with the number of accessions collected from different states.

and fatty acid composition profiling. All accessions were carefully harvested, cleared of debris, winnowed, and stored at room temperature until laboratory analysis.

2.3 | Extraction of Lipids and Gas Chromatography

Total lipids were extracted and analyzed following the method described by Tesfaye et al. (2024), with minor modifications. Lipids were extracted from 10 seeds per genotype, with two technical replicates for each sample. The seeds were homogenized in 1 mL of 0.15 M acetic acid and 3.75 mL of a 2:1 (v/v) methanol/chloroform mixture using an IKA T18 basic ULTRA TURRAX homogenizer in a glass test tube. Subsequently, 1.25 mL of chloroform and 0.9 mL of Millipore-grade H_2O were added, and the samples were vortexed for 10 s. The mixture was then centrifuged at 3000 rpm for 2 min.

After centrifugation, 200 μL of the lower chloroform phase was transferred into a clean glass screw-capped tube. The chloroform was evaporated by placing the tubes on a heated sand bed (70°C) under a stream of nitrogen gas. Once evaporation was complete, the samples were reconstituted in 100 μL heptane and methylated by adding 2 mL of methylation solution (2% concentrated H_2SO_4 in anhydrous methanol). The methylation reaction was performed in a sealed tube at 90°C for 1 h. After the reaction, the samples were cooled to room temperature, and 1 mL of Millipore-grade H_2O and 0.75 mL of heptane were added. The samples were vortexed for 15 s and centrifuged at 3000 rpm for 2 min. After centrifugation, 100 μL of the upper heptane phase, containing the fatty acid methyl esters (FAMEs), was transferred to a gas chromatography (GC) vial with a glass insert for GC analysis.

Fatty acid profiles were generated for each sample using an Agilent Model 886 gas chromatography instrument, and data were collected using a flame ionization detector (FID). Separation of the FAMEs was achieved using a WCOT fused-silica CP-wax 58 capillary column (50 m × 0.32 mm, Agilent) with a split ratio of 10:1. The oven temperature was set to 150°C for 0.2 min, then increased by 4°C/min to 210°C, followed by an increment of 10°C/min to 250°C. The temperature was maintained at 250°C for 5 min to ensure full elution of the samples.

The identity of each fatty acid was determined by comparing the retention times of its respective FAMEs to those a certified Me63 external standard (Larodan, Sweden). The oil content, total triacylglycerols (total TAG), total glyceride (total Gly), and percentages of fatty acidwere calculated using the following equations:

$$\text{Oil Content}\,(\%) = \left(\frac{\text{Total Weight of Fatty Acids}}{\text{Initial Weight of Seeds}}\right) \times 100$$

$$\text{Total TAG (mg)} = \sum \text{Concentration of individual TAGs (mg)}$$

$$\text{Total Glyceride (mg)} = \left(\text{Sum of TAG, DAG, MAG (mg)} \times \frac{\text{Molecular Weight of Glycerol}}{\text{Molecular Weight of TAG and DAG}}\right)$$

$$\text{Percentage of Fatty Acid (FA)} = \left(\frac{\text{Peak Area of FA}}{\sum \text{Peak Areas of all FAs}}\right) \text{x100}$$

This calculation of oil content ensures that the total fatty acids measured are expressed as a percentage of the initial seed weight, providing an accurate and precise measure of the oil content in the seeds.

2.4 | Seed Color Measurement

A colorimeter (Chroma Meter CR 400, Minolta, Japan) was used to measure the color of the stored sesame seeds, following the method described by Hassan et al. (2021). The color parameters measured were L^* (lightness), a^* (red-green), and b^* (yellow-blue). The device was calibrated using a standard white reflector plate. Two technical replicates were performed for each measurement using a Petri dish filled with 50.0 g of sesame seeds. The same seeds from these replicates were used for lipid analysis.

2.5 | Statistical Analysis

All the statistics analyses were conducted using the open-source R environment, version 4.3.2. Pearson's correlation coefficients were calculated to examine relationships between variables (Sedgwick 2012). The correlations were computed using the "cor" function in R, and their significance was assessed using the "cor.test" function. The resulting correlation matrix and p-values provided insight into the strength and direction of linear relationships between the variables (Ornella et al. 2014). Principal Coordinates Analysis (PCoA) was employed to explore the similarity or dissimilarity among samples based on a distance matrix. The "cmdscale" function in R was used to perform classical multidimensional scaling on the distance matrix obtained from the "dist" function. The PCoA plot presents the samples in a reduced dimensional space, where the axes represent the principal coordinates that capture the maximum variation in the data.

Individual linear regression analyses were performed to assess the relationship between oil content and color parameters (L^*, a^*, and b^*). The regression models were constructed using Oil Content (%) as the predictor variable and each color parameter as the response variable. The models were fitted using the "lm" function from the *stats* package in R, and model diagnostics were conducted using the "broom" package (Robinson 2014). The regression equations are as follows:

$$L^* = \beta_0 + \beta_1 \times \text{oil content}\,(\%) + \epsilon$$

$$a^* = \beta_0 + \beta_1 \times \text{oil content}\,(\%) + \epsilon$$

$$b^* = \beta_0 + \beta_1 \times \text{oil content}\,(\%) + \epsilon$$

where β_0 is the intercept, β_1 is the coefficient for (L^*, a^*, b^*), respectively, and ϵ is the error term.

3 | Results

3.1 | Oil Content and Fatty Acid Composition Profiling

The oil content ranged from 32.8% to 50.2%, with a mean of 41.5% (Table S3). The total TAG varied from 7.4 to 18.3 mg, with an average of 13 mg, while total Gly content varied from 0.3 to 0.8 mg, with an overall mean of 0.6 mg. The oil composition differed among the evaluated samples (Table S3). The palmitic acid (16:0) ranged from 10.8% to 8.94% for the saturated fatty acids, with an overall average of 10.0%. Stearic acid (18:0) concentrations ranged from 6.25% to 7.80%, with a mean of 6.80%, while the arachidic acid content (20:0) ranged from 0.73% to 0.49%, with a mean value of 0.60%. For unsaturated fatty acid compositions, oleic acid (18:1) varied from 41.3% to 47.6%, with an overall mean of 44.4%. For linoleic acid (18:2), the content ranged from 41.4% to 35.0%, with a mean of 37.8%. The linolenic (18:3) acid content ranged from 0.31% to 0.14%, with a mean of 0.20%. Gadoleic acid (20:1) content varied from 0.13% to 0.03%, with a mean of 0.08%.

The range of oil content (32.8%–50.2%) in the Sudanese germplasm corresponds well with results from previous studies, including Mondal (2010) (36.3%–52.7%), Uzun (2008) (43.2%–59.0%), Were (2006) (41.7%–55.5%), and Kurt (2018) (42.6%–57.8%) (Table 1). For palmitic acid (16:0), levels ranged from 8.94% to 10.8% in our study, consistent with Kurt (2018) (8.19%–10.26%) and Uzun (2008) (8.0%–10.3%) but slightly narrower than Mondal (2010) (9.1%–14.8%). Stearic acid (18:0) ranged from 6.25% to 7.80% in this study, higher than Uzun (2008) (2.1%–4.8%) and closer to Kurt (2018) (4.63%–6.35%). Oleic acid (18:1) ranged from 41.3% to 47.6% in our study, exceeding the ranges reported by Uzun (2008) (29.3%–41.4%) and Were (2006) (35.1%–42.1%) but aligned with Kurt (2018) (36.13%–43.63%) and Mondal (2010) (36.7%–52.4%). Linoleic acid (18:2) ranged from 35.0% to 41.4% in our results, and it is within the ranges reported by Uzun (2008) (40.7%–49.3%), Were (2006) (41.9%–48.0%), and Kurt (2018) (39.13%–46.38%). Linolenic acid (18:3) ranged from 0.14% to 0.30% in this study, which is in line with Kurt (2018) (0.28%–0.40%). Arachidic acid (20:0) and gadoleic acid (20:1) were detected at 0.49%–0.7% and 0.03%–0.10% in this study but were missing in the other studies (Table 1).

3.2 | Colorimetric Measurement

The color parameter L^* ranged from 34.05 ± 0.50 to 79.8 ± 0.42, with an overall mean of 59.6 (Table S2). a^* values varied from 4.55 ± 0.07 to 15.1 ± 0.00, with an overall mean of 8.4. Similarly, the b^* values ranged from 12.2 ± 0.14 to 24.9 ± 0.99, with an overall mean of 18.2.

3.3 | Correlations and Regressions Among the Fatty Acid Composition and Colorimetric Traits

The total oil content exhibited a significant positive correlation with triacylglycerols (TAG) ($r = 0.91$, $p < 0.01$). A significant

TABLE 1 | A range comparison of oil content and fatty acid composition for this study and the previous studies.

Traits	This study (%)	Mondal (2010) (%)	Uzun (2008) (%)	Were (2006) (%)	Kurt (2018) (%)
Oil content	32.8–50.2	36.3–52.7	43.2–59.0	41.7–55.5	42.6–57.8
Palmitic (16:0)	8.94–10.8	9.1–14.8	8.0–10.3	8.4–10.5	8.19–10.26
Stearic (18:0)	6.25–7.8	Not specified	2.1–4.8	4.5–6.0	4.63–6.35
Oleic (18:1)	41.3–47.6	36.7–52.4	29.3–41.4	35.1–42.1	36.13–43.63
Linoleic (18:2)	35.0–41.4	30.4–51.6	40.7–49.3	41.9–48.0	39.13–46.38
Linolenic (18:3)	0.14–0.3	NS	NS	NS	0.28–0.40
Arachidic (20:0)	0.49–0.7	NS	NS	NS	NS
Gadoleic (20:1)	0.03–0.1	NS	NS	NS	NS

TABLE 2 | Linear regression analysis of oil content and seed coat color parameters in Sudanese sesame gene bank collection samples.

Response_Variable	Intercept	Slope	R^2	p_value_Intercept	p_value_Slope
L^*	46.27	0.32	0.01	7.71E-06	0.18
a^*	11.98	−0.09	0.02	3.59E-09	0.06
b^*	18.58	−0.01	0.00	3.64E-14	0.87

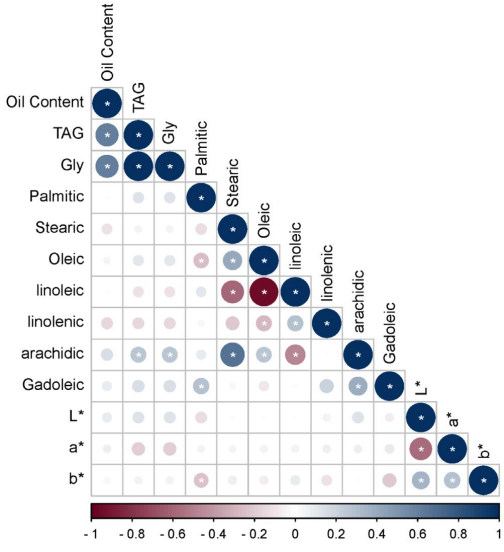


FIGURE 2 | Correlation heatmap of oil content, fatty acid composition, and color parameters in sesame accessions. * indicated significance at $p < 0.05$.

negative correlation was observed between oleic acid (18:1) and linoleic acid (18:2) ($r = -0.88$, $p < 0.01$) (Figure 2). Palmitic acid (16:0) displayed a significant positive correlation with stearic acid (18:0) ($r = 0.62$, $p < 0.05$). Stearic acid (18:0) also showed a significant positive correlation with arachidic acid (20:0) ($r = 0.54$, $p < 0.05$). Among minor fatty acids, linolenic acid (18:3) was positively correlated with linoleic acid (18:2) ($r = 0.46$, $p < 0.05$). No significant correlations ($p < 0.05$) were found among oil and fatty acids content and composition with seed coat color parameters (Figure 2).

The regression analysis showed that the intercepts for all color parameters (L^*, a^*, and b^*) were highly significant ($p < 0.001$), indicating their strong baseline values (Table 2). However, the slopes were largely non-significant, with minimal variation explained ($R^2 \leq 0.02$), suggesting weak or negligible relationships between the predictors and the response variables.

The Principal Coordinate Analysis (PCoA) revealed limited clustering patterns among sesame accessions based on their geographic origin. However, the accessions from South Kordofan formed a loose cluster, with most accessions showing positive values on PCoA1, indicating a possible similarity among the accessions from this region. The accessions from other states, for example, from North Kordofan, Kassala, and Blue Nile, were more dispersed across the plot, reflecting great variability of accessions from these regions (Figure 3). Notably, an overlap was observed between accessions from different states. However, some accessions from North Kordofan and Gedarif were positioned far from the main clusters, indicating their possible unique set of oil content, fatty acids composition, and seed coat color characteristics.

4 | Discussion

The present study, investigating the range of oil content, fatty acids composition, and seed coat color in Sudanese sesame, revealed a highly significant variation in these traits and the presence of high levels, especially in oleic acid content. As these traits are essential for sesame breeding and crop improvement strategies, the Sudanese germplasms are of high importance to be utilized by plant breeders in Sudan and beyond.

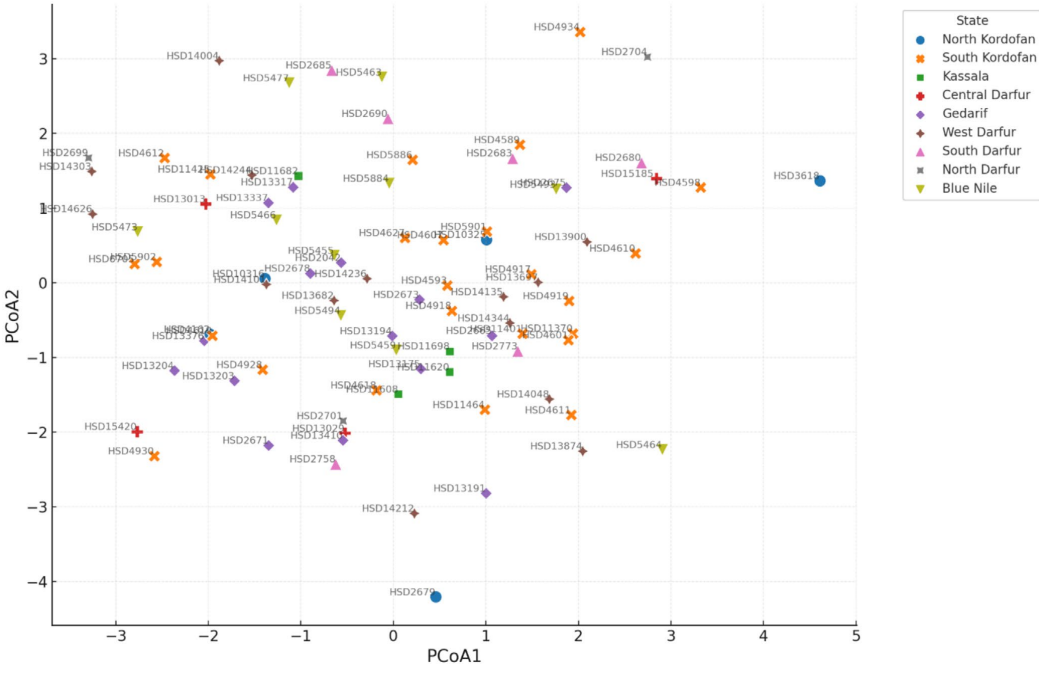


FIGURE 3 | Principal coordinates analysis cluster plot based on oil content, fatty acids composition, and Sudanese sesame accessions. Collection sites of the various accessions are displayed by different colors, as explained with the symbols.

The present study revealed that Sudanese sesame holds a significant range in total oil, total Gly, and total TAG contents. The Gly and TAG levels in sesame are important concerning human health and nutrition, as these compounds contain essential fatty acids supporting metabolic functions and overall health (Carino and Vital 2022). Previous studies have indicated that enhancing the TAG content in crops can increase beneficial fatty acids essential for preventing chronic diseases (Saini et al. 2020). Furthermore, TAG composition affects the shelf life of oils as well as their suitability for cooking (Zeb and Ahmad 2017). Additional studies are recommended to understand the performance of Gly and TAG contents in Sudanese sesame across diverse environments to assess their stability and adaptability. In addition, genomic analyses, including genome-wide association studies (GWAS) and quantitative trait loci (QTL) mapping, can potentially elucidate the genetic basis of oil content and connect desirable agronomic traits.

The large range of content of unsaturated fatty acids, i.e., oleic, linoleic, and α-linolenic acids in the Sudanese sesame, and specifically, the high levels of oleic acid (up to 46.7%) contribute significantly to their specificity and potential to be used in plant breeding. These fatty acids are essential for human health and biofortification; several studies have highlighted their importance. Thus, oleic acid has been shown to reduce low-density lipoprotein (LDL) cholesterol levels and the risk of cardiovascular diseases (Kamdar et al. 2021), but it is also useful in breeding for oxidative stability (Bahkali et al. 1998). Linoleic acid is essential for maintaining skin health and regulating immune function (Kim et al. 2014). Additionally, it has been demonstrated that alanine is an omega-3 fatty acid that lowers blood pressure and improves lipid profiles, helping to reduce the risk of heart disease (Han et al. 2019; Wu et al. 2020). Thus, enriching crops with these fatty acids contributes to improved nutritional quality and health benefits Siahbalaci et al. (2021), and for this, the incorporation of high-content genotypes in plant breeding programs might be an important tool. Previous studies have shown that the fatty acid profile of sesame oil exhibits significant variation connected to the environmental difference in production (Zahran et al. 2020), which indicates the importance of environmental and genetic factors in determining the fatty acid composition in sesame. Specific genes, such as the FAD2, are involved in synthesizing fatty acids determining the oleic-to-linoleic acid ratios (Wei et al. 2015). A comprehensive understanding of the genetic basis underlying the high oleic acid content observed in some of the accessions evaluated in this study, along with insights into the stability of this trait across different environmental conditions, is crucial for leveraging these accessions in sesame breeding programs. Such knowledge would enhance efforts to develop improved sesame varieties with superior oil quality, benefiting breeding initiatives in Sudan and potentially contributing to global sesame improvement. Also, for the saturated fatty acids, a high variability was found in the present study, in particular for palmitic, stearic, and arachidic acids. Saturated fatty acids are known to impact human health negatively, contributing to increased cholesterol levels and cardiovascular disease risk (Sultan 2018). However, the

saturated fatty acids also play a significant role in the oxidative stability of the oil, thereby enhancing oil shelf life, and making them critical factors in oil formulation and processing (Hassanien et al. 2014). Among the saturated fatty acids in sesame, arachidic acid is of specific interest for certain food and cosmetics products (Cerone and Smith 2021). Arachidic acid was detected in relatively low concentrations (less than 1%) in sesame seeds analyzed in this study. Previous research has indicated that arachidic acid has a less significant negative effect on human health as compared to other saturated fatty acids, such as palmitic and stearic acids (Mthana et al. 2022). Genetic factors and environmental conditions have been shown to influence the levels of arachidic acid (Abd-Elhafeez et al. 2020) indicating a need to further understand the mechanisms of synthesis of this fatty acid.

The sesame accessions in the present study showed a large variability in seed coat color. Sesame seed coat color has in previous studies been linked to (i) taste and qualitative aspects when used in food formulations, where often light coat color is preferred (Cui et al. 2021; Pandey et al. 2013), (ii) biochemical functions that influence oil metabolism and disease resistance (Zhang et al. 2013), and (iii) nutritional aspects including the content of bioactive components, where often dark colored types hold a higher content (Dossou et al. 2021). This study did not identify any correlation between seed coat color and oil content, fatty acid content, or fatty acid composition, as has been reported in some previous studies (Uzun and Çağırgan 2006). Additionally, in Sudan, a prevalent belief among sesame farmers associates dark brown seed color with higher oil content (ETI 2023). Therefore, it is crucial to implement Farmer Field Schools (FFS) focused on sesame agronomy and seed characteristics to address such misconceptions. The variation in seed coat color observed among the Sudanese accessions presents new opportunities for breeding programs. For example, dark-colored seeds, potentially associated with higher lignin content, could be incorporated into breeding efforts to target specific traits of interest. Previous studies have indicated that dark seed coats are associated with a high lignan content, which enhances the oil antioxidant properties of sesame (Comini et al. 2023). Lignans are known to act as a *secoisolariciresinol diglucoside*, which enhances the activity of antioxidant enzymes such as superoxide dismutase and catalase (Pilar et al. 2017; Rajesha et al. 2006), thereby reducing inflammation and oxidative stress.

Additionally, lignans may contribute to cardiovascular health by decreasing C-reactive protein (CRP), a marker of inflammation (Bolvig et al. 2017). However, the connection between seed coat color and lignan content in the Sudanese accessions has to be further elucidated, as such studies have not been performed on Sudanese sesame. Also, in some ancient sesame germplasm, lignan content has been found to correlate with oil content, with white-seeded varieties showing higher lignan levels compared to dark-seeded germplasm (Kancharla and Arumugam 2020).

Consistent with previous studies (Kurt 2018), the present study found a significant negative correlation between oleic and linoleic acid. This relationship is attributed to the desaturation of oleic acid into linoleic acid, as Relina et al. (2022) reported.

The PCoA cluster plot indicated a large variation among the collected Sudanese sesame accessions, although it did not show any clear clustering pattern based on the regions of their collection However, the majority of the accessions from South Kordofan were found to have positive PCoA values, thereby forming a loose cluster. This indicates that the samples collected from this region might vary more similarly regarding oil, fatty acid, and composition than accessions from other regions. Also, the PCoA1 was found to explain a high degree of the variation, indicating that the spread of the accessions along this axis described the majority of the variation among the samples. It is well known that historical seed migrations and trade have shaped the genetic landscape of sesame, with accessions originating from Europe, Asia, and North America forming different groups when analyzed by double-digest site-associated DNA sequencing (Basak et al. 2019). The lack of clustering by accessions from different regions of Ethiopia indicates that sesame seeds have been traded across the country and that there are significant variations in oil content and fatty acid composition of origin other than regional. Thus, an increased understanding of the background for the variation in Sudanese sesame as regards oil content, fatty acids content, and composition, similar to what has been studied previously, sequencing a wide collection of more than 700 accessions collected globally (although only one representing the Sudanese gene pool; Wei et al. 2015), will contribute to improved opportunities in breeding high-quality sesame varieties for Sudan and beyond.

5 | Conclusions

Sudanese sesame accessions exhibit extensive variation in oil content, fatty acid composition, and seed coat color traits that are essential for breeding programs to improve oil quality and nutritional value in sesame. The observed diversity in fatty acid profiles, particularly the presence of high oleic acid content in certain accessions, provides valuable opportunities for breeders to enhance the nutritional attributes of sesame oil. Incorporating multiple traits into breeding efforts is critical to optimizing oil yield and quality. The significant variation observed within Sudanese sesame highlights the importance of utilizing these resources in breeding programs, not only in Sudan but also in global initiatives. Despite Sudan's status as a center of sesame diversity, the Sudanese gene pool remains underutilized in international breeding due to the country's unstable political and economic conditions in recent decades. However, with the increasing pressures of climate change and the need to feed a growing global population, every available genetic resource must be considered for breeding efforts. Preserving and utilizing the genetic diversity of Sudanese sesame is essential for developing high-quality, resilient cultivars capable of thriving in diverse environmental conditions. These efforts will contribute to food security and the sustainability of sesame production on both local and global scales.

Acknowledgments

We extend our appreciation to the Swedish Research Council (VR) for funding this work. In addition, we would like to thank the Agricultural Plant Genetic Resources Conservation and Research Center at Agricultural Research Corporation (ARC)—Sudan for providing the sesame accessions and their related geographical information. Special

appreciation goes to Dr. Eltahir Ibrahim, a former head of the Plant Genetic Resources Conservation and Research Center, for supporting this work.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

References

Abd-Elhafeez, E., S. A. Abou-El-Hawa, M. Rashwan, and B. Ramadan. 2020. "Influence of Roasting Treatments on Oil Quality Properties of Two Sesame Seed Varieties." *Assiut Journal of Agricultural Sciences* 51: 92–101.

Abebe, D. 2006. Regional Strategy for the Ex Situ Conservation of Plant Genetic Resources: Eastern Africa Eastern Africa Plant Genetic Resources Network (EAPGRN), Entabe, Uganda.

Ahmed, M. 1997. *Adaptability of New Varieties of Sesame to the Central Rain Lands of the Sudan*. Variety Release Committee of Sudan.

Ahmed, M. 2008. "Evaluation of New Sesame (*Sesamum indicum* L.) Genotypes for Yield, Yield Components and Stability." *University of Khartoum Journal of Agricultural Sciences (Sudan)* 16: 380–394.

Ahmed, M., M. Taha, and A. Khalafalla. 2003. *Assessment of New Promising Sesame Cultivars for the Central Rain Lands and Irrigated Systems. Variety Release Committee of Sudan*.

Bahkali, A., M. Hussain, and A. Basahy. 1998. "Protein and Oil Composition of Sesame Seeds (Sesamum Indicum, L.) Grown in the Gizan Area of Saudi Arabia." *International Journal of Food Sciences and Nutrition* 49: 409–414.

Basak, M., B. Uzun, and E. Yol. 2019. "Genetic Diversity and Population Structure of the Mediterranean Sesame Core Collection With Use of Genome-Wide SNPs Developed by Double Digest RAD-Seq." *PLoS One* 14, no. 10: e0223757.

Bedigian, D., and J. R. Harlan. 1983. "Nuba Agriculture and Ethnobotany, With Particular Reference to Sesame and Sorghum." *Economic Botany* 37: 384–395.

Bhunia, R. K., R. Kaur, and M. K. Maiti. 2016. "Metabolic Engineering of Fatty Acid Biosynthetic Pathway in Sesame (*Sesamum indicum* L.): Assembling Tools to Develop Nutritionally Desirable Sesame Seed Oil." *Phytochemistry Reviews* 15: 799–811.

Bolvig, A. K., N. P. Nørskov, M. S. Hedemann, et al. 2017. "Effect of Antibiotics and Diet on Enterolactone Concentration and Metabolome Studied by Targeted and Nontargeted LC–MS Metabolomics." *Journal of Proteome Research* 16, no. 6: 2135–2150.

Cerone, M., and T. K. Smith. 2021. "A Brief Journey Into the History of and Future Sources and Uses of Fatty Acids." *Frontiers in Nutrition* 8: 570401.

Carino, J. D., and P. G. Vital. 2022. "Characterization of Isolated UV-C-Irradiated Mutants of Microalga *Chlorella vulgaris* for Future Biofuel Application." *Environment, Development and Sustainability*: 1–18.

Comini, E., D. Rubiales, and P. Reveglia. 2023. "Variability of Fatty Acid Composition and Lignan Content in Sesame Germplasm, and Effect of Roasting." *ACS Food Science & Technology* 3: 1747–1758.

Cui, C., Y. Liu, Y. Liu, et al. 2021. "Genome-Wide Association Study of Seed Coat Color in Sesame (*Sesamum indicum* L.)." *PLoS One* 16: e0251526.

Dossa, K., X. Wei, Y. Zhang, et al. 2016. "Analysis of Genetic Diversity and Population Structure of Sesame Accessions From Africa and Asia as Major Centers of Its Cultivation." *Genes* 7: 14.

Dossou, S. S. K., F. Xu, X. Cui, et al. 2021. "Comparative Metabolomics Analysis of Different Sesame (*Sesamum indicum* L.) Tissues Reveals a Tissue-Specific Accumulation of Metabolites." *BMC Plant Biology* 21: 1–14.

Du, H., H. Zhang, L. Wei, C. Li, Y. Duan, and H. Wang. 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: 1–20.

ETI. 2023. *SESAME SEEDS IN Sudan: Human Rights Risks Linked to Hummus and Tahina*. Ergon Associates Ltd and Commissioned.

Han, T., G. Yang, X. Cao, H. Li, H. Pei, and Z. Zhang. 2019. "Preparative and Scaled-Up Separation of High-Purity α-Linolenic Acid From Perilla Seed Oil by Conventional and pH-Zone Refining Counter Current Chromatography." *Journal of Separation Science* 42: 2360–2370.

Hassan, A. B., E. Pawelzik, and D. Von Hoersten. 2021. "Effect of Microwave Heating on the Physiochemical Characteristics, Colour and Pasting Properties of Corn (*Zea mays* L.) Grain." *LWT* 138: 110703.

Hassanien, M. M., A. G. Abdel-Razek, M. Rudzińska, A. Siger, K. Ratusz, and R. Przybylski. 2014. "Phytochemical Contents and Oxidative Stability of Oils From Non-Traditional Sources." *European Journal of Lipid Science and Technology* 116: 1563–1571.

Kalaiyarasi, R., R. Rajasekar, K. Lokeshkumar, A. Priyadharshini, and M. Mohanraj. 2019. "Correlation and Path Analysis for Yield and Yield Traits in Sesame (*Sesamum indicum* L.) Genotypes." *International Journal of Current Microbiology and Applied Sciences* 8: 1251–1257.

Kapoor, S., S. S. Parmar, M. Yadav, et al. 2015. "Sesame (*Sesamum indicum* L.)." *Agrobacterium Protocols* 2: 37–45.

Karim, I. E. E. A., and I. S. M. Ismai. 2020. "Sesame Production and Export Performance in Sudan: A Comparative Advantage Analysis." *University of Khartoum Journal of Agricultural Sciences* 16: 230\–240.

Kamdar, J. H., M. D. Jasani, B. C. Ajay, et al. 2021. "Fatty Acid Desaturase-2 (ahFAD2) Mutant Alleles in Peanut (*Arachis hypogaea* L.) Pre-Breeding Lines: An Insight into the Source, Features, Discourse, and Selection of Novel Pre-Breeding Lines." *Genetic Resources and Crop Evolution* 68: 529–549.

Kancharla, P. K., and N. Arumugam. 2020. "Variation of Oil, Sesamin, and Sesamolin Content in the Germplasm of the Ancient Oilseed Crop *Sesamum indicum* L." *Journal of the American Oil Chemists' Society* 97, no. 5: 475–483.

Kim, M. C., M. G. Kim, Y. S. Jo, H. S. Song, T. I. Eom, and S. S. Sim. 2014. "Effects of C18 Fatty Acids on Intracellular Ca^{2+} Mobilization and Histamine Release in RBL-2H3 Cells." *Korean Journal of Physiology & Pharmacology: Official Journal of the Korean Physiological Society and the Korean Society of Pharmacology* 18, no. 3: 241.

Kurt, C. E. M. A. L. 2018. "Variation in Oil Content and Fatty Acid Composition of Sesame Accessions From Different Origins." *Grasas y Aceites* 69, no. 1: e241–e241.

Li, C., Y. Duan, H. Miao, M. Ju, L. Wei, and H. Zhang. 2021. "Identification of Candidate Genes Regulating the Seed Coat Color Trait in Sesame (*Sesamum indicum* L.) Using an Integrated Approach of QTL Mapping and Transcriptome Analysis." *Frontiers in Genetics* 12: 700469.

Mahmoud, M., and M. R. Khalil. 2019. "Breeding for Developing New Indeterminate Lines of Tomato (*Solanum lycopersicum* L.) by Selection." *Menoufia Journal of Plant Production* 4: 233–245.

Mthana, M. S., C. S. Gajana, B. Moyo, A. Hugo, N. Makhamba, and F. Thabethe. 2022. "Response in Growth Performance, Physico-Chemical

Properties, and Fatty Acid Composition of Broiler Meat to Different Levels of Mucuna Pruriens Seed Meal."

Negash, T. T., K. Tesfaye, G. K. Wakeyo, and C. Ziyomo. 2020. "Genetic Diversity, Population Structure Analysis Using Ultra-High Throughput Diversity array Technology (DArTseq) in Different Origin Sesame (*Sesamum indicum* L)." *Research Square*. https://doi.org/10.21203/rs.3.rs-103763/v1.

Ornella, L., P. Pérez, E. Tapia, et al. 2014. "Genomic-Enabled Prediction With Classification Algorithms." *Heredity* 112: 616–626.

Pandey, S. K., A. Das, and T. Dasgupta. 2013. "Genetics of Seed Coat Color in Sesame (*Sesamum indicum* L.)." *African Journal of Biotechnology* 12: 6061\–6067.

Pathak, K., S. W. Rahman, S. Bhagawati, and B. Gogoi. 2017. "Sesame (*Sesamum indicum* L.), an Underexploited Oil Seed Crop: Current Status, Features and Importance—A Review." *Agricultural Reviews* 38: 223–227.

Pilar, B., A. Güllich, P. Oliveira, D. Ströher, J. Piccoli, and V. Manfredini. 2017. "Protective Role of Flaxseed Oil and Flaxseed Lignan Secoisolariciresinol Diglucoside Against Oxidative Stress in Rats With Metabolic Syndrome." *Journal of Food Science* 82: 3029–3036.

Rajesha, J., K. N. C. Murthy, M. K. Kumar, B. Madhusudhan, and G. A. Ravishankar. 2006. "Antioxidant Potentials of Flaxseed by In Vivo Model." *Journal of Agricultural and Food Chemistry* 54: 3794–3799.

Relina, L., O. Suprun, R. Bohuslavskyi, et al. 2022. "Evaluation of Genotype-Environment Interactions for Non-Polar Lipids and Fatty Acids in Chickpea (*Cicer arietinum* L.) Seeds." *Journal of VN Karazin Kharkiv National University. Series Biology* 39: 20–34.

Robinson, D. 2014. "Broom: An R Package for Converting Statistical Analysis Objects Into Tidy Data Frames." *arXiv Preprint arXiv:1412.3565*.

Sabiel, S., M. Ismail, E. Abdalla, and A. Osman. 2015. "Genetic Variation in Sesame Genotypes (*Sesamum indicum* L.) Grown in the Semi-Arid Zone of The Sudan." *SABRAO Journal of Breeding & Genetics* 47: 214\–220.

Saini, R. K., K. R. Rengasamy, E. Y. Ko, J. T. Kim, and Y. S. Keum. 2020. "Korean Maize Hybrids Present Significant Diversity in Fatty Acid Composition: An Investigation to Identify PUFA-Rich Hybrids for a Healthy Diet." *Frontiers in Nutrition* 7: 578761.

Sedgwick, P. 2012. "Pearson's Correlation Coefficient." *BMJ* 345: e4483. https://doi.org/10.1136/bmj.e4483.

Siahbalaei, R., G. Kavoosi, and M. Noroozi. 2021. "Manipulation of *Chlorella vulgaris* Polyunsaturated ω-3 Fatty Acid Profile by Supplementation With Vegetable Amino Acids and Fatty Acids." *Phycological Research* 69, no. 2: 116–123.

Sultan, F. I. 2018. "Chromatographic Separation and Identification of Many Fatty Acids and Phenolic Compounds From Flowers of *Celosia cristata* L. and Its Inhibitory Effect on Some Pathogenic Bacteria." *Australian Journal of Basic and Applied Sciences* 12: 25–31.

Tesfaye, M., T. Feyissa, T. Hailesilassie, E. S. Wang, S. Kanagarajan, and L.-H. Zhu. 2024. "Rapid and Non-Destructive Determination of Fatty Acid Profile and Oil Content in Diverse Brassica Carinata Germplasm Using Fourier-Transform Near-Infrared Spectroscopy." *PRO* 12: 244.

Uzun, B., and M. İ. Çağırgan. 2006. "Comparison of Determinate and Indeterminate Lines of Sesame for Agronomic Traits." *Field Crops Research* 96: 13–18.

Wang, L., S. S. K. Dossou, X. Wei, et al. 2020. "Transcriptome Dynamics During Black and White Sesame (*Sesamum indicum* L.) Seed Development and Identification of Candidate Genes Associated With Black Pigmentation." *Genes* 11: 1399.

Wei, X., K. Liu, Y. Zhang, et al. 2015. "Genetic Discovery for Oil Production and Quality in Sesame." *Nature Communications* 6: 8609.

Wu, G., Y. Shen, R. Nie, et al. 2020. "The Bioactive Compounds and Cellular Antioxidant Activity of Herbaceous Peony (Paeonia Lactiflora Pall) Seed Oil From China." *Journal of Food Science* 85: 3815–3822.

Zahran, H. A., A. Abd-Elsaber, and H. Z. Tawfeuk. 2020. "Genetic Diversity, Chemical Composition and Oil Characteristics of Six Sesame Genotypes." *OCL* 27: 39.

Zeb, A., and S. Ahmad. 2017. "Changes in Acylglycerols Composition, Quality Characteristics and In Vivo Effects of Dietary Pumpkin Seed Oil Upon Thermal Oxidation." *Frontiers in Chemistry* 5: 55.

Zech-Matterne, V., M. Tengberg, and W. Van Andringa. 2015. "*Sesamum indicum* L. (Sesame) in 2nd Century BC Pompeii, Southwest Italy, and a Review of Early Sesame Finds in Asia and Europe." *Vegetation History and Archaeobotany* 24: 673–681.

Zhang, H., H. Miao, L. Wei, C. Li, R. Zhao, and C. Wang. 2013. "Genetic Analysis and QTL Mapping of Seed Coat Color in Sesame (*Sesamum indicum* L.)." *PLoS One* 8: e63898.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

IV

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 (*).

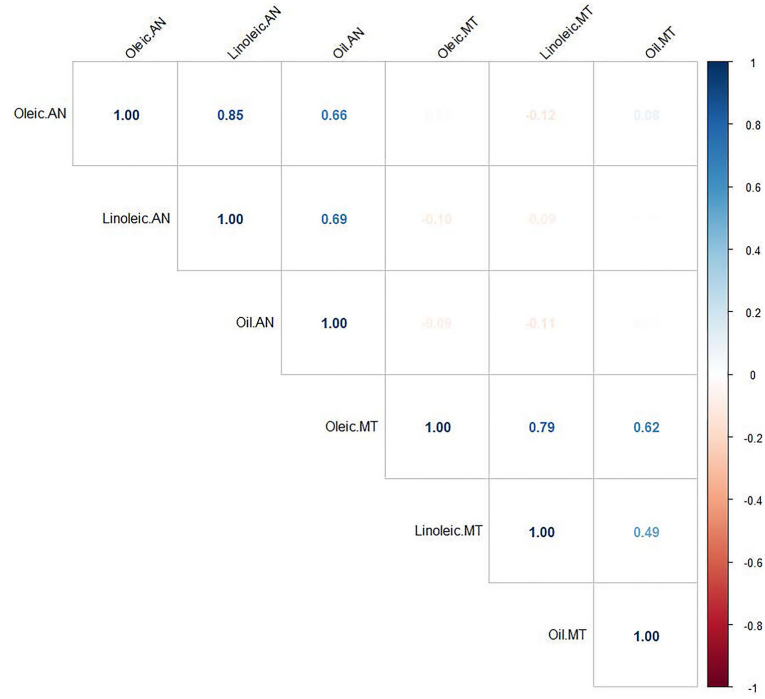


FIGURE 1 Correlation of sesame oil and fatty acid composition traits between locations. The heat map depicts the correlation between sesame oil and fatty acid composition traits, including oleic acid (oleic acid), linoleic acid (linoleic acid), and total oil content (oil), across the two locations, Abu 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*

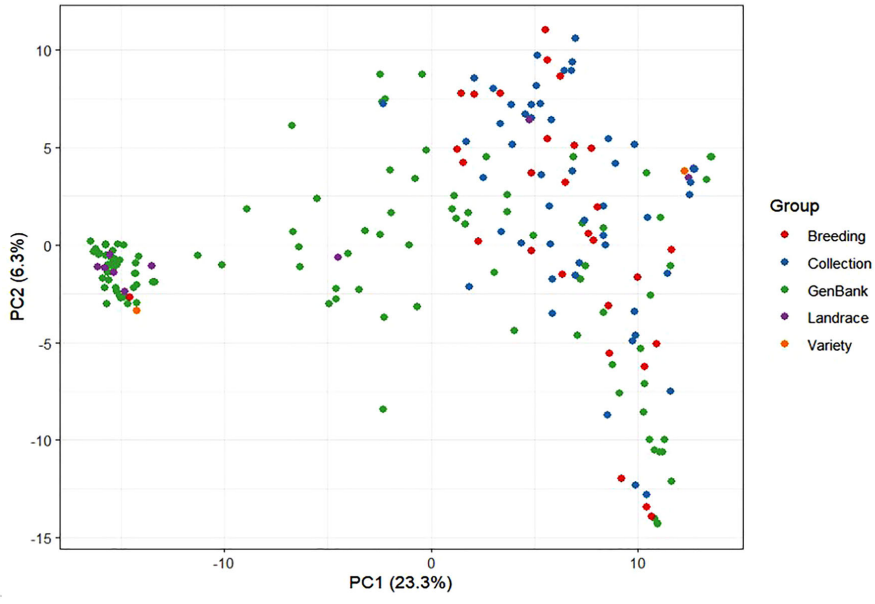


FIGURE 2 Genetic diversity and population structure of 200 sesame accessions. The principal component analysis (PCA) plot illustrates the genetic diversity and population structure among various sesame accession groups. Each point represents an individual accession, with colors distinguishing different 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.

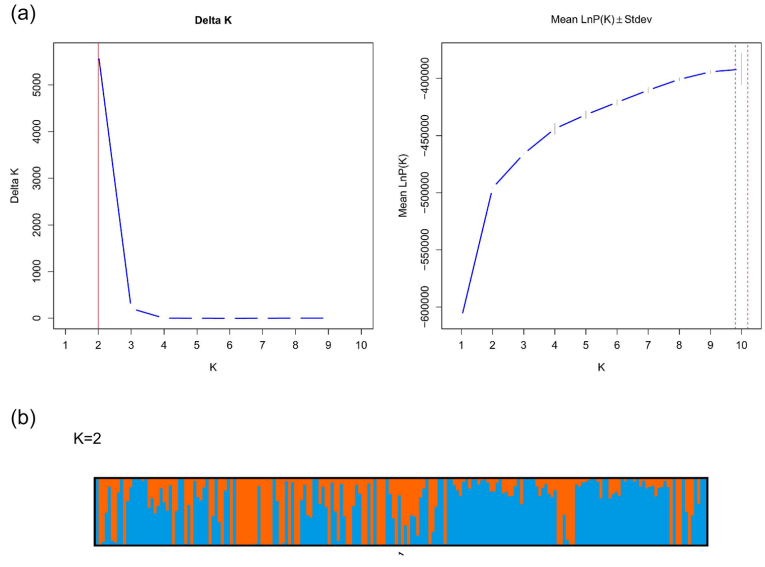


FIGURE 3 Population structure analysis: (a) estimated Delta *K* and Ln*P*(K) values for different *K* values. (b) Population structure of the 200 sesame accessions. The heatmap illustrates the subpopulations structure, with accessions grouped into two distinct subpopulations represented in red and green.

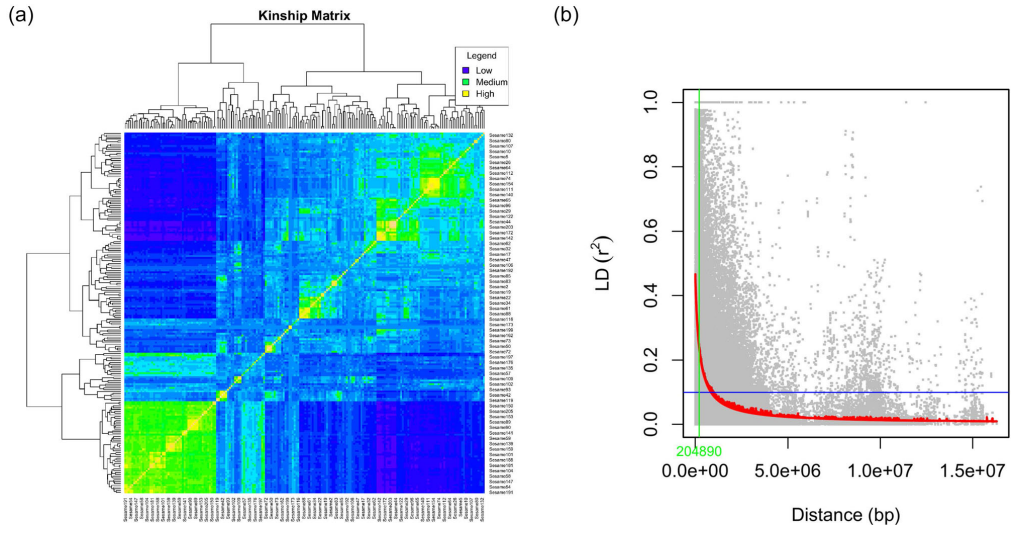


FIGURE 4 (a) Kinship heatmap showing the genetic relationships among 200 sesame accessions based on additive relationships. (b) Linkage disequilibrium (LD) patterns in the sesame genome. The figure shows the distribution of LD decay across genomic distances in the sesame populations. The *x*-axis represents 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

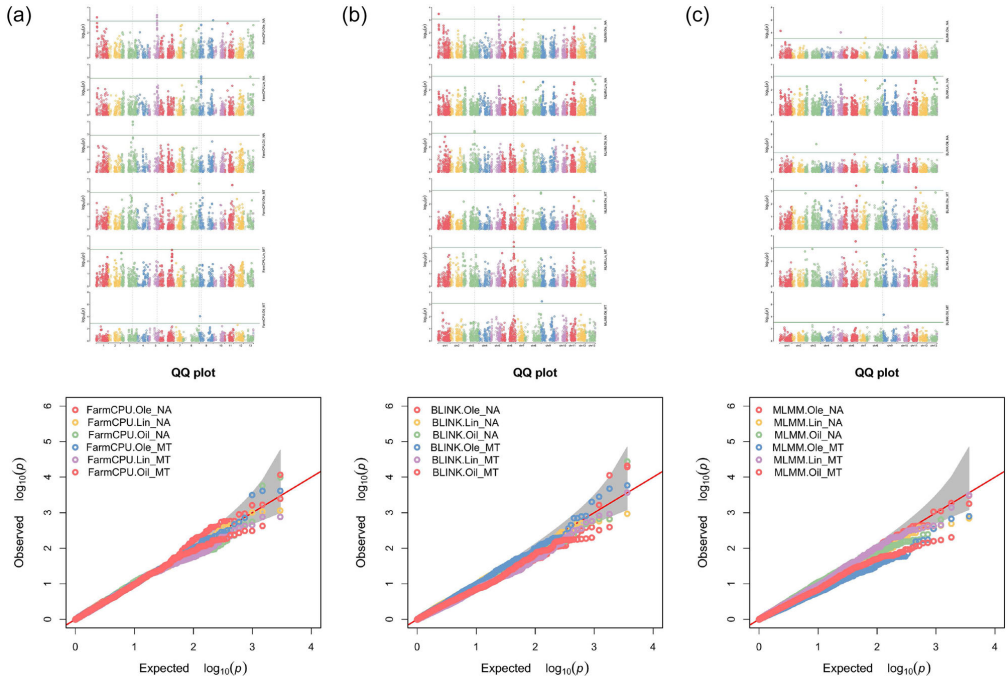


FIGURE 5 Manhattan and *Q*–*Q* plots based on three different GWAS models: (A) fixed and random model circulating probability unification (FarmCPU), (B) Bayesian information and linkage-disequilibrium iteratively nested keyway (BLINK), and (C) multiple locus mixed model (MLMM), identifying significant 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

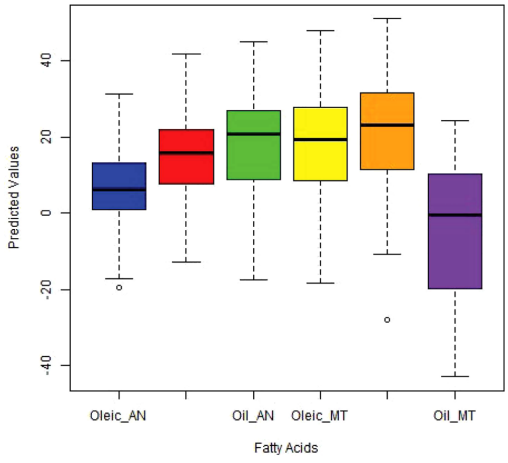


FIGURE 6 Predicted values for oil and fatty acid content in sesame accessions across two locations. The boxplots show the predicted distributions of oleic and linoleic acids and oil content in 200 sesame accessions grown at two locations: Abu Naama (AN) and Matuq (MT). The central line indicates the 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

Acta Universitatis agriculturae Sueciae
Doctoral Thesis No. 2025:50

This thesis evaluated agronomic, biochemical, and genetic traits of 200 Sudanese sesame genotypes across multiple environments. Comprehensive profiling revealed significant variability in oil content, fatty acid composition, antioxidant retention, and seed pigmentation. Genome-wide association studies identified key genetic loci and candidate genes for seed coat color, oil composition, and capsule shattering, underscoring the untapped potential of Sudanese sesame germplasm for breeding nutritionally enhanced, climate-resilient varieties.

Mohammed Elsafy received his doctoral and master's degrees from the Department of Plant Breeding at the Swedish University of Agricultural Sciences (SLU), Sweden, and his bachelor's degree from the Faculty of Agriculture at Nile Valley University, Sudan.

Acta Universitatis agriculturae Sueciae presents doctoral theses from the Swedish University of Agricultural Sciences (SLU).

SLU generates knowledge for the sustainable use of biological natural resources. Research, education, extension, as well as environmental monitoring and assessment are used to achieve this goal.

ISSN 1652-6880
ISBN (print version) 978-91-8046-559-5
ISBN (electronic version) 978-91-8046-564-9