From genes to metabolites: a multi-omics exploration of drought tolerance and seed oil quality in diverse Moringa

Dorcas Olufunmilayo Ojo

PhD

University of York

Biology

March 2025

Abstract

This is a comprehensive analysis of the Moringa plant, an understudied highly versatile tree the "drumstick tree" or "miracle tree". Firstly, transcriptome assemblies of 4 *Moringa stenopetala*, 1 *Moringa peregrinna* and 44 *Moringa oleifera* and a comparative analysis of these transcriptomes revealed valuable insights into the plant's population structure and genetic diversity. A comprehensive RNA-Seq revealed 19,325 differentially expressed genes (DEGs) across 49 co-expression modules, reflecting a complex transcriptional response to drought. Key genes related to drought stress, including SAG12 and Vignain, were identified alongside their varying expressions across plant tissues, suggesting tissue-specific responses. Co-expression analysis identified certain genes as prospective master regulators in the drought-responsive network of *Moringa oleifera*. Functional analysis via BLASTx revealed its top hit to *Carica papaya* gene LOC110809083, *Carica papaya* gene LOC110810283, CAK9329279.1 *Citrullus colocynthis*, XP_030952457.1 *Quercus lobata*, and LOC110821448 *Carica papaya* gene, respectively, indicating their possible role in ion transport, homeostasis, cell wall remodelling, glucose transport, and photoreactive repair through DNA and FAD binding activity during drought stress. Additionally, the study explored Moringa seed oil profiles and genetic markers (SNPs) linked to oil quality traits, providing insights into breeding for superior edible and industrial oil lines. The transcriptomic analysis provides 49 Moringa *de novo* assemblies, good foundational resources for understanding Moringa's adaptability to drought whilst contributing to food security. Overall, this research presents Moringa as a versatile crop for sustainable agriculture, with potential applications in food security, nutraceuticals and biofuel production.

Table of Contents

List of Tables

List of Figures

Supplementary Figures

Acknowledgements

This incredible journey would not have been possible without the unwavering grace, strength, and wisdom bestowed upon me by God. He has been my anchor, guiding me through every challenge, and to Him alone be all the glory.

I am most profoundly grateful to my parents, Pastor Emmanuel Olukokun Taiwo and Mrs Christianah Funke Olorunwa, whose dream was to witness me bag a PhD in their lifetime. Their relentless dedication and sacrifices laid the foundation for this achievement. Their investment in my education began early, going as far as creating a mini-laboratory in our family home, in collaboration with my science teacher, when I was just eight years old. That experience sparked my passion for science and the environment, ultimately shaping my academic journey from undergraduate studies to earning a PhD in Biology in the United Kingdom despite financial hardships and personal challenges.

Their words of encouragement instilled in me resilience and courage, enabling me to embark on this journey with my three wonderful children, Ayomide, Busola, and Favour Ojo, despite the uncertainties of single parenting alongside intense academic research. Their unwavering love and support have been my strength, providing me with the motivation to persevere. They have been a source of joy, encouragement, and inspiration, proving that together, we can overcome any obstacle. Love you, my fabulous trio!

I am deeply indebted to my primary sponsor, the Federal Government of Nigeria, TETFUND, for providing the necessary financial support that made this research programme possible.

A heartfelt thank you to my supervisors, Prof. Seth J. Davis and Dr Andrea Harper, whose mentorship, encouragement, and belief in my abilities extended far beyond academic guidance. Their support has been invaluable, fostering my growth not only as a researcher but also as an individual. Their kindness, patience, and wisdom made me feel at home, and I am forever grateful for the research skills, data analysis techniques, and professional development I have gained under their guidance.

My sincere appreciation also goes to my Thesis Advisory Panel (TAP) Chair, Prof. Daniel Jeffares, for his insightful advice and constructive feedback throughout this journey.

To the incredible members of the Davis Lab and Harper Lab, I am grateful for your support, collaboration, and intellectual exchanges. Special thanks to Jessica Downing for teaching me the use of Viking and Bash scripting, Sarah Ortega for her support with VCF analysis, Isaac Reynolds for his assistance with RNA-Seq analysis, and Kayla McCartney, Sarah Lock, and Ethan Redmond for their support and camaraderie.

To the CloudSpan Team, I extend my appreciation, especially to Emma Rand, who guided my first steps in R and Genomics. Your patience and expertise were instrumental in my learning. Thank you also for the scholarship support that eased this journey. I also wish to acknowledge the Data-Genomics and TF team, Sally James, Alastair Droop,

Fabiano, and Sandy McDonald, for their contributions and insights that enhanced my bioinformatics analysis.
To my wonderful siblings Tosin Olorunwa, Peace Ige, Femi Shadare, Funmilola Olorunwa, Anuoluwapo Olanrewaju, and Fisayo Oke, your unwavering financial and emotional support carried me through difficult times. Your encouragement kept me focused and determined, and I am blessed to have you all in my life.
To my spiritual mentors, Pastor Olufemi Oluwole, MFM, Nigeria; Pastor Ruth, MFM Leeds, UK; Pastor Adeleye; Pastors Ade and Ayo Sowole (RCCG KLT, York); Pastor Benjamin and Lady Pastor Assiffuah (Mustard Seed Chapel); and Pastors Jonathan and Nicki (York Vineyard Church), your prayers, encouragement, and words of wisdom provided me with the spiritual strength I needed to navigate this journey.

A special thank you to Prof. Sean Sweeney, for his welfare support through the Radhika V. Sreedhar Scholarship, and to the Student Support and Advice Team, whose assistance made my academic life smoother.
I am also grateful for the kindness and generosity of CAP, Anna Druzik, whose support was invaluable.
To Mr. and Mrs. Tolu Adesanya, words cannot express my gratitude for opening your home to me and my three children during the York housing crisis in September/October 2021. Your hospitality and generosity provided us with stability in one of our most challenging times.
To the wonderful Latinwo family, for their support morally, prayerfully and financially. I deeply appreciate Grace Latinwo for the time and effort devoted in reading and correcting my thesis for grammatical errors, ensuring clarity and accuracy.
To my friends who have been my pillars of strength, Dr Akin, Dr Toyin Ogunwale, Prof & Mrs Oluyitan, Titilayo Adebiyi, Samson Owoyokun, Adetola Ajulo, Arinze, and Modupe Laba, your encouragement, prayers, and unwavering belief in me meant the world.
Finally, to everyone, both mentioned and unmentioned, who has contributed in one way or another to this success, thank you. This PhD journey was never a solitary endeavour, but a collective triumph made possible by the love, sacrifices, and unwavering support of many. I am forever grateful.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references.

Signed 
.....
.....

Chapter 1 - Introduction

1.1 Why Moringa?

Food demand is predicted by Pardey *et al.*, (2014) to rise two-fold by 2025. World Hunger Facts (FAO, 2023) estimate that a major food crisis affects around 43 million people in 38 countries. While 3.1 billion people lack access to reasonably priced and nutritious food, almost 10% of the world's population goes hungry daily (Statista, 2022). Especially in Sub-Saharan Africa, where 85% of nations experience famine, food insecurity has reached hitherto unheard-of heights (Beyene, 2023; Roser, Ritchie and Rosado, 2023). Out of the 14,000 edible plant species, just 30 meet 95% of human caloric intake, according to Wellcome, 2024. There are three main crops rice, maize and wheat, which are popular and sufficiently meet the dietary needs of humans but will soon be inadequate for the expanding population; hence, comprehensive research on possible orphan species like *Moringa oleifera* is vital considering the significant effects of climate change on agricultural output and world food security. While guaranteeing enough nutritional output (Oderinde *et al.,* 2022; Dhakar *et al.,* 2011; Mallenakuppe *et al.,* 2019), crop variants should show tolerance to both biotic and abiotic challenges.

Lately, interest in the Moringa plant has grown from its use in local subsistence to its availability in global markets for food supplements, seed oils, and cosmetic products. Moringa cultivation now thrives in Asia, Africa and Latin America, with the value chains varying widely; leaves and powder are used for food, seed oil finds applications in cosmetics and biodiesel, and seed cake is utilised for animal feed and water purification. Nevertheless, the economic and social impacts are influenced by numerous factors such as market accessibility, traditional knowledge, and integration into local diets. Small-scale farmers are seen to dominate Moringa production, although industrial farming is gradually emerging in regions like India and West Africa, focusing on seed oil and leaf powder.

Despite the numerous potentials of Moringa for agriculture, research on it is still sparse. There are not many genomic resources, and there are not many studies that look at how different genotypes work in different environments. Likewise, there are not enough reference genomes and transcriptomes; it is therefore expedient to make genomic resources for the 13 species, perform population genomics and landscape genetic analyses to populate the phylogeographic data, set up co-ordinated field trials at multiple sites to augment genotype-environment data resources, carry out comparative transcriptomics and gene editing for functional characterisation of drought-related genes and standardise the rules for extraction and reporting oil composition.

The biology of Moringa, from its genome to its metabolites, remains poorly understood despite its growing importance. This study gives an overview of how combining phylogenomics, transcriptomics, metabolomics and Moringa accessions from diverse locations could help understand the genetic factors that affect oil yield, quality, and stress tolerance. This could lead to better crops and more widespread cultivation in different climates. There have been efforts to improve this versatile crop, as seen in recent research in Malaysia, where Muniandi *et al.,(2024)* evaluated *M. oleifera* genotypes for propagation and yield, indicating potential for clonal breeding. Studies on molecular diversity using SSR and ISSR markers show a lot of variation within populations, but there are no coordinated breeding programmes. Likewise, functional genomics, marker-assisted selection, and genome editing have not been systematically implemented. Setting up reference genomes for different Moringa species should also be a clear priority. Moringa's unique place in Brassicales, high nutritional and industrial value, and ability to survive in different environments make it a great example of how genomic diversity and metabolite function are connected. To make the most of its potential as a multipurpose crop, there is a need to expand comparative genomics and create coordinated breeding programmes.

1.2 Taxonomic classification and Growth habit of the *Moringa* genus

The Moringa genus consists of 13 species native to India and Africa, now cultivated in the tropical and subtropical regions of the world. The Moringa genus is the only genus and most phenotypically varied in the family Moringaceae (Boopathi and Abubakar, 2021). Moringa is in the order Brassicales, which also includes the families Brassicaceae and Capparaceae. (Mohammed and Hawar, 2022).

The species include the: *Moringa oleifera* Lam. the most widely known and often called the "drumstick tree", "horseradish tree" or "miracle tree"; *Moringa stenopetala* known as the African Moringa and native to East Africa (Kenya and Ethiopia); *Moringa peregrinna* Forssk. native to the Horn of Africa Southern Sinai, Egypt (Dadamouny *et al*., 2016; Dadamouny, n.d); *Moringa drouhardii*, native to southwestern Madagascar; *Moringa hildebrandtii* Engl indigenous to southwestern Madagascar*; Moringa ovalifola* Dinter & Berber indigenous to Namibia and Angola; *Moringa concanensis Nimmo* indigenous to northern India; *Moringa longituba* Engl indigenous to Ethiopia and Somalia*; Moringa arborea* Verdc. indigenous to Kenya*; Moringa borziana* Mattei indigenous to Somalia; *Moringa pygmaea* Verdc., indigenous to Somalia*, Moringa rivae* Chiov. native to Kenya and Ethiopia; *Moringa ruspoliana* Engl. indigenous to Ethiopia and *Moringa arborea* Verdc. indigenous to Kenya.

According to Verdcourt (1985), the Moringa genus can be classified into three clades based on molecular and morphological analyses: the Donalsdonia, Moringa, and Dysmoringa clades. The Donalsdonia clade, consisting of *M. stenopetala*, *M. hildebrandtii*, *M. ovalifolia*, and *M. drouhardii,* have bottle-shaped trunks with radially symmetric flowers. While the Moringa clade comprises slender trees, including *M. oleifera*, *M. concanensis*, and *M. peregrina,* identifiable by their irregular floral symmetry, perigynous flowers, and short receptacles; the tuberous trees: *M. borziana*, *M. longituba, M. pygmaea, and M. rivae;* and the Sarcorhrizal trees: *M. arborea* and *M. ruspoliana.* While the Dysmoringa clade contains the *M. longituba* species, which branches from the Moringa clade. The *M. longituba* is peculiar due to its long receptacle and red flowers (Olson, 2002). With the Donaldsonia clade generally recognised as the most basal within the family. The growing availability of DNA data is clarifying the evolutionary links across Moringa species (Olson, 2002). Among the thirteen species of the Moringa genus, *Moringa oleifera* is most well-known for its economic viability and dependability as a food source; medicine, oilseed crop and its great drought resilience help to explain this, according to Trigo, Castello and Ortola, 2022. The polyploid nature of the Moringa species plays an important role in the genus's adaptability to fit many climatic circumstances and diversity.

1.2.1 Morphology

The *Moringa* species possess a wide range of variability in their tree habit, from deciduous to evergreen, and in the shape of the tree canopy, from semi-spreading to upright. They show unique morphological variations adapted to their diverse habitats (Sutarno and Rosyida, 2020; Jattan *et al.,* 2021). The tree form, leaf structure, and flower characteristics vary among the 13 species. The *M. oleifera* has a slender tree form with a straight trunk. Devkota & Bhusal, 2020; Kumar *et al*., 2017; and Chang *et al*., 2022 found that this plant grows to reach 3 metres in roughly three months and 12 to 15 metres in a few years. *M. concanensis, M. arborea and M. peregrina* trees are slender, similar to *M. oleifera. M. peregrina* tree reaches up to 6-10 m tall. On the other hand, *M. drouhardii, M. stenopetala*, *M. hildebrandtii*, and *M. ovalifolia* are identifiable with their bottle-shaped trunk, which is adapted for water storage, and radially symmetric flowers (Olson, 2002), while *M. longituba* is shrub-like with distinctive red flowers and elongated floral tubes (Wikipedia). The *M. pygmaea, M. rivae, M. ruspoliana, M. longituba and M. borziana* are small shrubs ranging from 50cm to 6m in height with succulent tuberous roots.

Generally, leaves of the Moringaceae are generally bright green to dark green, soft, tender and glabrous when young. They may be bipinnate or tripinnate depending on the species and maturity. They are generally compound leaves arranged alternately on the

stem. The leaf length is about 45-60 cm long when mature. The leaflets are small and oval or obovate in shape, with smooth/entire margins and a rounded apex. They possess long and flexible drooping petioles. More specifically, there are variations among the species. The *M. oleifera* leaves are tripinnate and feathery; *M. stenopetala* leaves are bipinnate or pinnate with larger and thicker leaflets; *M. peregrinna* leaves are bipinnate, fewer but with larger leaflets and more succulent; *M. hildebrandtii* leaves are bipinnate with an umbrella-like leaf arrangement with some reddish pigmentation on the stem tip and leaf, while *M. longituba leaves* are bipinnate with fewer leaflets which are oblong to obovate rachis. They have some reddish pigmentation (Boopathi and Abubakar, 2021).

In the tropics, Moringa spp. flower year-round, and once yearly during the spring in the temperate regions. The flowers may be either bilaterally or radially symmetric, ranging in colour from white to cream to brown to bright red, with an exception with *M. longituba* rachis (Boopathi and Abubakar, 2021). Flowers of the *Moringa spp.* are hermaphroditic, containing both the male reproductive organs (stamens) and female reproductive organs (pistils) within a single flower. Their reproduction is mainly through cross-pollination. Agents of pollination are insects such as bees, the carpenter bees (Xylocopa) according to Jyothi, Atluri and Reddi (1990), honeybees (Sharma, 2019), thrips, *Apis sp.* (Bhattacharya and Mandal, 2004), and birds (Jyothi, *et al.*, 1990).

Almost all Moringa species exhibit the taproot system, but some species, such as the *M. peregrinna* and *M. hildebrandtii,* have succulent or thickened taproots. These features help to store water and enable them to survive in dry, arid climates. Moringa species are well-known for their resistance against environmental hazards connected to climate change. Their large taproot system and numerous small-sized leaves with waxy, shiny epidermis help them survive in drought (Tshabalala *et al.,* 2019).

1.2.2 Morphology of Moringa accessions studied.

The morphological characteristics of Moringa species, such as *Moringa oleifera*, *Moringa stenopetala*, and *Moringa peregrina*, exhibit notable variations in their leaves, seeds, roots, and shoots. The observed variations may affect species responses to environmental stressors, including drought conditions. *Moringa oleifera* exhibits glossy, ovate leaves that are vibrant green and contain three to nine leaflets. In contrast, *Moringa stenopetala* possesses larger, elongated leaves that are deeper green, thicker in texture, and consist of five to seven leaflets (Fig. 1.2). *Moringa peregrina* exhibits smaller, rounded leaves that are dark green and consist of three to five small leaflets, showing a less glossy appearance relative to the other two species. Variations in leaf structure as seen in Fig. 1.2 may influence plants' water retention capabilities and

photosynthetic efficiency during droughts, potentially resulting in species-specific adaptations to these conditions.

The seeds exhibit notable variations. *M. peregrina* produces the smallest and most triangular seeds, characterised by a dark brown, hard outer shell. In contrast, *M. oleifera* seeds are marginally larger, whereas *M. stenopetala* seeds are significantly larger and elongated, exhibiting a light brown to cream shell (Fig. 1.1). The traits of these seeds may affect germination rates and initial growth phases under conditions of limited water availability.

The root systems of the three species display variations that could influence drought tolerance; all possess a taproot system, although the roots of *M. peregrina* are particularly fleshy and edible (Fig.1.3). Changes in root architecture may affect plants' capacity to access deep water sources during drought conditions (Huang *et al*., 2024). The growth patterns and reproductive characteristics further differentiate the species. The pods of the three species display an elongated and dehiscent morphology, maturing to a cream colouration, whereas the flowers are small and cream-coloured. *M. oleifera* and *M. stenopetala* exhibit rapid growth, attaining heights of up to 10 metres, characterised by slender, robust green stems and a dense branching architecture. *M. peregrina* exhibits a slower growth rate, is significantly shorter, and demonstrates a sparser growth pattern. The differences in growth and morphology among species offer insights into how structural variations influence specific drought adaptation mechanisms, paving the way for future molecular-level investigations into these responses.

A B C

Figure 1.1 A. *M. oleifera* and *M. stenopetala* plants and seeds **B.** Samples of *Moringa* seeds studied **C.** *Moringa oleifera* flower

Figure 1.2 Leaves of the varying *Moringa* samples: **1.** R1 (WOT) **2.** *M peregrinna* Oman (OmMp1) **3.** R3 (SJ) **4.** R2 (BK) **5.** *M oleifera* Nigeria (MO56) **6.** *M oleifera* Nigeria (MO28) **7.** *M. oleifera,* Burkina Faso **8.** *M stenopetala* (KMs5)

Figure 1.3: Roots of **1.** *Moringa stenopetala* **2.** *Moringa oleifera* **3.** *Moringa peregrinna*

Figure 1.4 Flowers and fruit- pods containing seeds of *Moringa oleifera*

1.3 History and Global Distribution of *Moringa species.*

According to Senthilkumar *et al.,* 2018, ancient kings have used Moringa leaves and fruits for their alleged therapeutic properties with expectations of improving cognitive

performance and maintaining skin health since 150 BCE. Historical accounts indicated that many Roman Empire citizens had access to the health advantages the Greeks had found. As several health advantages of Moringa oil were discovered, the Jamaican chamber called a hearing in 1817 on its usage in salads and other cuisine. Human migration and commerce have made the Moringa tree highly sought after in the Philippines and other Southeast Asian nations to fight poverty, hunger, and food insecurity (Vélez-Gavilán, 2022; Mashamaite *et al.,* 2021).

Several authors recorded that Moringa originated from India/Arabia (Selvam, 2005; Godino *et al*., 2015; Lakshmidevamma *et al*., 2021), whereas Dalla Rosa (1993) said it originated from East Africa/Southeast Asia. Nevertheless, India as the native origin is the only common country out of the numerous claims. This plant has been carried by migrants and traders to various parts of the world, such as Africa, Southeast Asia and the Philippines in ancient times, and to the Americas, Europe and the Philippines by Spanish migrants in recent times (Velez-Gavilan, 2022). The means of movement and dispersal are by natural means whereby the winged seeds are dispersed over short distances by wind or pods floating on water; being disposed at dumpsites from seeds and cuttings from garden waste; or deliberately obtaining the seeds or stem cuttings for cultivation (Velez-Gavilan, 2022).

Moringa is widely cultivated due to its drought tolerance, fast growth, nutritional value, multipurpose use and low-input farming, requiring minimal fertiliser or pesticides. Only a few species are cultivated commercially; these include the *M. oleifera*, *M. stenopetala*, and *M. peregrinna*, while others are regionally endemic, wild, or used by local communities. There are up to six endemic Moringa species indigenous to Madagascar (Africa). Moringa is cultivated most extensively in the tropical and subtropical regions of the world, as it thrives well in warm climates. According to Olson (2019), it also grows as an annual in temperate regions. Regions of the world where Moringa is cultivated the most include India, Africa, Southeast Asia, Latin America and the Caribbean, the Middle East and the United States of America.

India has grown over 38,000 hectares and has an annual production of 2.2 million tonnes of tender fruits of Moringa, where they are used for food, Ayurvedic medicine and export, making India the country which cultivates Moringa the most (Pandey, 2011). India is the largest exporter of *M. oleifera,* with major production in Tamil Nadu, Andhra Pradesh, Karnataka and Odisha, exporting industrial outputs including leaf powder for supplements, teas and capsules; pods for fresh market and canning; and oil (ben oil) for cosmetics and nutraceuticals to the USA, Europe and the Middle East (Sekhar *et al.*, 2018).

Southeast Asia, including the Philippines, Indonesia and Thailand, grows them in home gardens and small farms with increasing commercial interest in leaf powder and oil for food, cosmetics and personal healthcare products such as soap, perfumes, shampoos and others (Palada, 2017). Moringa is locally known as “malunggay” in the Philippines. The Philippines has public-private partnerships promoting Moringa, hence cultivating commercial plantations for nutraceutical leaf powder and capsules, fortified food and beverages, and cosmetic products (Zion Market Research, 2023).

Nigeria is the major producer of Moringa among the African countries, although it is generally cultivated across West (Ghana, Senegal, Mali), East (Kenya, Ethiopia, Tanzania, Uganda), Central and Southern Africa. Moringa is in high demand in Niger, and about half of the Moringa consumed in Niger is imported from Nigeria (Pasternak *et al*., 2017). Moringa is grown in numerous family home gardens for local use as well as in large-scale by subsistent farmers supplying larger stakeholders for industrial purposes in Kano, Kaduna, Oyo, and Plateau states. These process Moringa for nutritional supplements, animal feed additives, domestic and export use (Adesina *et al*., 2013). Lately, there is a proposal for industrial scale Moringa cultivation of 15,000 hectares in Cross River state, Nigeria for biodiesel production and animal feed from the seedcake remnant (Agrictechmea, 2023). Also 500 hectares of *Moringa oleifera* is currently grown by Eden Moringa in Benin, Edo state Nigeria. Moringa is widely grown in Africa for food security, promotion of nutrition programmes, agroforestry and export in home gardens and farms.

It is significantly cultivated on a large scale in Kenya, focusing on the export of leaf powder, seed oil, value-added skincare and health products with much support from NGOs and women’s co-operatives. Kenya is among the countries with widespread hunger. Every household in Kenya grow and utilize *M. oleifera* and *M. stenopetala* primarily for food, as they include it in their daily diet, medicine and animal feed (Mwami *et al*., 2024; Kumssa *et al*., 2017).

Latin American and Caribbean countries such as Haiti, Mexico, Brazil, the Dominican Republic and Nicaragua cultivate Moringa in community agroforestry programmes for reforestation, prevention of malnutrition and as a business opportunity (Delphin *et al*., 2023). The Middle East, such as Yemen, Oman, and Saudi Arabia, cultivates *M. peregrinna* despite their dryland (Robiansyah *et al*., 2014). Moringa is also cultivated in Florida, California, Hawaii and Texas in the United States, mainly by small-scale organic farmers for speciality health products (Thurber and Fahey, 2009).

The rising demand for Moringa leaf powder, oil and capsules (fruits/pods) in Europe, North America and Asia is now influencing the commercial cultivation of Moringa.

Likewise, international organisations like the Food and Agriculture Organization (FAO) and some Non-Governmental Organisations (NGO) promote Moringa in climate adaptation and nutrition programmes due to its application in food and beverages, cosmetics and skincare products, bakery and confectionery, and the pharmaceutical industry (Boopathi and Raveendran, 2021; Zion Market Research, 2023).

Moringa has great economic worth; according to Zion Market Research (2023), “the global Moringa market size was worth USD 9.5 billion in 2022 and is predicted to grow to around USD 18.2 billion by 2030 with a compound annual growth rate (CAGR) of roughly 8.5% between 2023 and 2030.” The profitability of Moringa cultivation can be seen from a revenue per hectare generating ₦13,260,000 annually in the Nigerian market, as Ojiako *et al*. (2011) reported an estimated yield of 78 metric tonnes of fresh leaves per hectare per year and a dry weight yield of 13.26 metric tonnes per hectare. The export price of Moringa Leaf powder typically ranges from $1.40 to $1.60 per kilogram for bulk shipments. The wholesale price in India is approximately ₹160 to ₹250 per kilogram for dried leaf powder (ExportersIndia.com). Trials in Nicaragua with 1 million plants per hectare and 9 cuttings/year over 4 years gave an average fresh matter production of 580 metric tonnes/ha/year (230 long tonnes per acre), equivalent to about 174 metric tonnes of fresh leaves (Mendieta-Araica *et al*., 2012).

Given the many ways Moringa has travelled to other countries and its trade still ongoing worldwide, there is likely to be gene transfer both inside and across accessions. Though there are several known kinds, most Moringa trees found in West Africa, including Nigeria, are seedlings whose parentage or origin is unknown. Molecular analysis is required to identify the migratory pattern among Moringa accessions from various regions of the world since genomic knowledge of this plant is lacking. The best genotypes for various breeding environments can be discovered, the development and sources of these accessions may be clearer, and the Moringa plants can be properly categorised and labelled. Similarly, it offers information on the degree to which their genes have been preserved, which in turn offers hints about the plant's practices, adaptations, survival rate, and evolutionary history (Salgotra *et al.,* 2023).

Table 1.0 Summary of the native and global distribution, tree structure, leaf arrangement and root system of the 13 species of *Moringa*.

Species	**Native Region**	**Current Cultivation**	**Uses**	**Leaf Type**	**Leaf structure**	**Root System**	**References**
1. Moringa oleifera	Indian subcontinent	Global but mostly in Africa, Asia, Americas & the Caribbean	Grown for food, oil, medicine, agroforestry	Compound, tripinnate	Numerous small leaflets, long, slender petioles	Long, deep taproot, with some lateral roots; drought-resistant	Azza, S.M. (2014), Hassanein & Al-Soqeer (2018)

2. Moringa stenopetala	Southern Ethiopia, Northern Kenya	East Africa: Ethiopia, Kenya, Tanzania	uses are similar to *M. oleifera*	Compound, bipinnate	Larger, thicker leaflets; broader leaves	Deep, sturdy taproot; adapted to arid conditions	Azza, S.M. (2014), Hassanein & Al-Soqeer (2018)
3. Moringa peregrina	Arabian Peninsula, NE Africa	Oman, Yemen, Egypt, Israel, Jordan	Valued for its oil; grows in dry, rocky habitats	Compound, bipinnate	Large, succulent leaflets; fewer leaflets per stem	Deep taproot with a thick, succulent nature; highly drought-tolerant	Azza, S.M. (2014), Hassanein & Al-Soqeer (2018)
4. Moringa drouhardii	Madagascar (SW)	Local Use in Madagascar	local use, conservation concern	Compound, tripinnate	Smaller leaflets, shorter petioles	Prominent, swollen taproot, characteristic of dry regions	
5. Moringa hildebrandtii	Madagascar (W)	Rarely cultivated	critically endangered	Compound, tripinnate	Long petioles, larger leaves	Strong taproot, succulent system; adapted to semi-arid climates	
6. Moringa arborea	Northeastern Kenya	Semi-arid areas; wild and rarely cultivated	Unpopular drought-tolera nt	Compound, bipinnate	Smaller leaflets, denser clusters	Thin taproot; grows well in disturbed areas	
7. Moringa borziana	Kenya, Somalia	Grows in dry scrublands		Compound, bipinnate	Narrower leaflets, shrubby growth	Deep taproot, often in rocky or dry areas	
8. Moringa longituba	Kenya, Somalia, Ethiopia	Cultivated rarely, mostly wild		Compound, bipinnate	Long floral tubes, larger leaflets	Deep, slender taproot; resistant to arid conditions	
9. Moringa ruspoliana	Kenya, Ethiopia	Local home gardens, wild areas	cultivated as shade or ornamental	Compound, tripinnate	Broad leaflets, spreading leaves	Strong taproot, adapted to semi-arid zones	
10. Moringa concanensis	Western India (Konkan region)	Western Ghats and occasionally in home gardens	Close to *M. oleifera*, but less domesticated	Compound, tripinnate	Smaller leaflets, broader than *M. oleifera*	Moderate taproot with extensive lateral roots	Anbazhakan *et al.*, 2007
11. Moringa pygmaea	SW Madagascar	Grows in limestone, dry soils	Rare, endangered shrub	Compound, bipinnate	Small, dense clusters of leaves	Shorter taproot, grows in rocky, dry soils	
12. Moringa ovalifolia	Namibia, Angola	Semi-wild, found mostly in desert outcrops	occasionally protected	Compound, bipinnate	Oval-shaped, smaller leaflets	Shallow, thick taproot, suitable for dry conditions	Habtemariam, (2017)
13. Moringa rivae	NE Kenya, Ethiopia	Wild, used locally	Little-known species; few studies exist	Compound, bipinnate	Smaller, narrow leaflets	Thin taproot, grows well in drylands	

Agronomic challenges in Moringa cultivation and plantation systems

Moringa is generally propagated by seed and stem cutting. Seed quality and seed storage, which may cause seeds to be non-viable; limited access to better germplasm; planting systems, including spacing, pruning, water, salinity, pests and diseases; and lack of market regulation and inconsistent quality standards are among the factors that can affect the optimum yield of Moringa. Since most Moringa farmers cultivate mostly for leaf

production, it is necessary to control the storage of Moringa seeds to maintain their viability for a longer period. Usually, fresh seeds have a high germination rate of 60-90% (Reed *et al*., 2022), but it decreases with poor storage, such as in places with high temperature and humidity. Good spacing (Sutarno and Rosyida, 2020) and pruning are also vital for high biomass yield, seed production, flowering and ease of harvesting (Kavitha *et al*., 2024). Salinity has been reported to reduce biomass and nutrient content of Moringa (Nouman, 2013), raise some antioxidant compounds but usually reduce total oil yield.

As much as Moringa adapts to a wide range of soils (Trigo *et al*., 2020), its seed development and oil yield are largely dependent on available soil nutrients. According to Atteya *et al*. (2021), soil acidity, low organic matter, and insufficient soil nutrients may cause yield variability across plantations. It grows very well in slightly acidic, well-drained loamy soil, suitable fertiliser and good spacing (Leone *et al*., 2016; Trigo *et al., 2020).* Although Moringa tolerates drought better than many annual crops, water availability during important reproductive stages influences seed yield and oil content. Although Moringa is generally reported to be pest-resistant, in practice Moringa could be susceptible to a number of pests, including thrips, pod flies and pathogens causing root rot, fungal fruit, seed rot, damage to leaves and poor quality of leaf yield.

The process and method of seed harvesting and storage determine oil quality and seed viability. Improved and standard methods of removing moisture from seeds after harvesting, controlled temperature for storage and quality testing, among others, are measures that could prevent seed rancidity and loss of oil quality.

Generally, the lack or scarcity of standardised post-harvest protocols for seed drying, storage, quality testing, registered varieties, resistant germplasm, co-ordinated breeding programmes, seed certification, and improved cultivars/germplasm with desirable traits such as lower VLCFA (Very Long Chain Fatty Acid) for edible oil and higher oleic acid content are challenges in Moringa plant breeding.

Opportunities for cultivation of other Moringa species

While *Moringa oleifera* has gained popularity in recent research, it is important to explore other species like *Moringa peregrina*, *Moringa stenopetala*, and the other 10 spp., as these offer additional avenues for crop improvement, adaptation to varied environmental conditions, andeconomic and ecological benefits of *Moringa spp..* cultivation (Table 1.0). *Moringa peregrina* is known to be well adapted for cultivation in arid environments, presenting opportunities for crop improvement and adaptation to desert conditions (Al-Khalifah & Shanavaskhan, 2017).

Crop improvement programmes for Moringa

To maximise the multipurpose benefits of the Moringa species, there is a need to improve its cultivation to move from landrace use to more stable commercialised cultivation. The following ideas would be necessary to be able to bridge the current gaps in Moringa research.

Establishing germplasm banks for all 13 species would be a good way of improving Moringa cultivation for high-value yield. The African Orphan Crops Consortium has been supporting Moringa sequencing. These seed banks would hold pieces of information on collection location, soil, and climate information which could be linked to genetic diversity with adaptive traits and help to characterise these collections using molecular markers and phenotypic traits. According to Stephenson and Fahey (2004) and Sarwar (2017), a number of *Moringa spp.* are in danger of extinction or regarded as endangered, including *M. peregrinna* (AbdAlla *et al*., 2023; Al Khateeb *et al*., 2013) and the species native to Africa, such as *M. stenopetala, M. rivae, M. borziana, M. longituba, M. arborea, M. ruspoliana. M. hildebrandtii* is extinct in the wild but found preserved by local communities in Madagascar (Letsera, Razafindrahaja and Razanatsoa, 2019).

It is important to breed Moringa for specific uses such as for edible oil, industrial use, drought stress adaptability and others. Inbreeding for oil quality for human edible use, expanding the reference genomes to cover each of the 13 species and pan-genome assemblies to capture diversity across the Moringa species. Developing genotyping platforms such as SNP panels and annotating key stress-resilience genes and oil biosynthesis, sterol, and tocopherol genes are some ways of developing genomic resources for Moringa research, breeding, and industrial use.

Creating dual breeding lines such as edible oil lines with high oleic, low behenic and sterol content for nutritional and antioxidant benefits and industrial lines with VLCFA for lubricants, cosmetics and biodiesel. Breeding for agronomic traits by selecting for high seed yield, leaf/biomass yield, uniform flowering, palatability for animals, climate resilience and pest and disease resilience

Developing marker-assisted selection for candidate genes, for instance, oil quality loci, and implementing genomic selection for polygenic traits like yield, biomass, and stress tolerance by linking GWAS to SNP markers would improve Moringa cultivation. Also improving methods of cultivation for quick multiplication of improved lines and developing certified seed systems that are easily accessible to farmers whilst establishing region-specific packages for Moringa. Harnessing biotechnological tools such as

CRISPR/Cas9 editing and applying transcriptomics and metabolomics for stress-response pathways and nutraceutical compound discovery.

1.4 Nutritional composition of Moringa

Moringa is categorised as a leafy vegetable, like the leaves of amaranth, hibiscus, sweet potato and others. Among 120 leafy vegetables, Moringa has been ranked the highest when tested for nutrient value, antioxidant properties, taste, and easy farming (Knez *et al., 2024).* Although the calcium concentration in Moringa leaves is one hundred times higher, their iron level is thirty times less than that of spinach leaves (Table 1.1). Moringa roots abound in magnesium, iron, calcium, zinc, sodium, and aluminium. Pasha *et al.,* (2020) found that whilst manganese and phosphorus are rather more common in leaves, potassium concentrations are higher in flowers and seeds. Moringa has not been found to naturally contain selenium (Table 1.1), but it can synthesize and accumulate it from selenium-rich soil (Lyons *et al*., 2017). All parts of the moringa plant, leaves, stems, roots, flowers, seeds, and pods are bursting in vital minerals, vitamins, and nutrients. Moringa oil nutrient composition vary according to seasons, weather and climate (Wiltshire *et al*., 2021), temperature of oil extraction process (Cervera-Chiner *et al*., 2024), tissue type (leave, seed) and developmental stage and location (Gomez-Martinez *et al* ., 2025; Shih *et al*., 2011).

Table 1.1 Nutritional value per 100 g of *Moringa oleifera* raw leaf and pod and leafy vegetables such as Amaranth and Brassicales such as Kale, Broccoli and Cabbage against the recommended daily nutritional requirement for growth and development.

Nutritional value per 100 g (3.5 oz)	**Recommended daily values 100 g based on a standard 2,000 kcal/day diet**	*Moringa oleifera* **leaf raw** Moringa oleifera -Wikiwand	*Moringa oleifera* **Pod raw** Moringa oleifera -Wikiwand	*Amaranthus caudatus* Spinach Leaves Amaranth-Wikiwand	**Kale (raw** *Brassica oleracea)* Kale -Wikiwand	**Broccoli raw** *Brassica oleracea* Broccoli - Wikiwand	**Cabbage raw** *Brassica oleracea* Cabbage -Wikiwand
Energy	**~200 kcal**	**270 kJ (64 kcal)**	**150 kJ (37 kcal)**	**33.47kJ (8 kcal)**	**207 kJ (49 kcal)**	**141 kJ (34 kcal)**	**103 kJ (25 kcal)**
Carbohydrates	**~30 g**	**8.28 g**	**8.53 g**	**1 g**	**8.8 g**	**6.64 g**	**5.8 g**
Sugars	**<5 g**				**2.3 g**	**1.7 g**	**3.2 g**
Dietary fiber	**~3 g**	**2.0 g**	**3.2 g**		**3.6 g**	**2.6 g**	**2.5 g**
Total Fat Saturated fat	**~7 g <2 g**	**1.40 g**	**0.20 g**		**0.9 g**	**0.37 g**	**0.1 g**
Protein	**~5 g**	**9.40 g**	**2.10 g**	**1.0 g**	**4.3 g**	**2.82 g**	**1.28 g**

Other constituents (Quantity) Water		**78.66 g**	**88.20 g**		**84.0 g**	**89.3 g**	**92 g**
Fluoride							**1 µg**
Vitamins							
Vitamin A equiv. β-Carotene Lutein zeaxanthin		**378 µg**	**4 µg**	**2996.35 µg**	**241 µg 6261 µg**	**31 µg 361 µg 1403 µg**	
Thiamine (B1)		**0.25 mg**	**0.05 mg**		**0.11 mg**	**0.07 mg**	**0.06 mg**
Riboflavin (B2)		**0.66 mg**	**0.07 mg**		**0.13 mg**	**0.12 mg**	**0.04 mg**
Niacin (B3)		**2.22 mg**	**0.62 mg**		**1.0 mg**	**0.64 mg**	**0.23 mg**
Pantothenic acid (Vitamin B5)		**0.13 mg**	**0.79 mg**		**0.9 mg**	**0.57 mg**	**0.21 mg**
Vitamin B6		**1.20 mg**	**0.12 mg**		**0.27 mg**	**0.18 mg**	**0.12 mg**
Folate B9		**40 µg**	**44 µg**	**29.75 µg**	**141 µg**	**63 µg**	**43 µg**
Choline					**0.8 mg**	**19 mg**	
Vitamin C	**~9 mg**	**51.7 mg**	**141.0 mg**	**15.16 mg**	**120 mg**	**89.2 mg**	**36.6 mg**
Vitamin E	**~1.5 mg**				**1.54 mg**	**0.78 mg**	
Vitamin K	**~12 µg**			**98 µg**	**390 µg**	**101.6 µg**	**76 µg**
Minerals							
Calcium	**~120 mg**	**185 mg**	**30 mg**	**75.25 mg**	**150 mg**	**47 mg**	**40 mg**
Iron	**~1.8 mg**	**4.00 mg**	**0.36 mg**	**0.81 mg**	**1.5 mg**	**0.73 mg**	**0.47 mg**
Magnesium	**~35 mg**	**147 mg**	**45 mg**		**47 mg**	**21 mg**	**12 mg**
Manganese		**0.36 mg**	**0.26 mg**	**0.31 mg**	**0.66 mg**	**66 mg**	**0.16 mg**
Phosphorus		**112 mg**	**50 mg**		**92 mg**	**316 mg**	**26 mg**
Potassium	**~340 mg**	**337 mg**	**461 mg**		**491 mg**	**33 mg**	**170 mg**
Sodium	**<230 mg**	**9 mg**	**42 mg**		**38 mg**	**0.41 mg**	**18 mg**
Zinc	**~1.1 mg**	**0.6 mg**	**0.45 mg**		**0.6 mg**	**0.41 mg**	**0.18 mg**
Selenium					**0.9 µg**		

1.5 Genomic approaches to Moringa plant diversity.

The versatility of *Moringa oleifera* ranging from its nutritional benefits to medicinal benefits which has gained high cognizance in recent years have encouraged research at the molecular level according to Trigo *et al*. (2020). Before genomic research came about, the genetic information of Moringa's many advantages were mostly unknown. *Moringa spp.* can be likened to model crops like rice, corn or tomatoes or model plant species like *Arabidopsis thaliana*, due to characteristics such as rapid growth rate reaching up to 10-15 feet tall within a year and simple genomes fit for manipulation and research. *Moringa oleifera* genome size being 315Mb (Tian *et al*., 2015) smaller than that

of *Oryza sativa* (rice) which is 430 million bp but greater than *Arabidopsis thaliana*'s which is 125 million bp (Brendel *et al*., 2002). Moringa can therefore access a wide range of genomic resources and technology, making it great for crop breeding techniques and climate resilience due to its great adaptability. This research has clarified the complex interactions among genes, regulatory areas, and responses connected with drought tolerance in *Moringa oleifera*, high yield, and the quality of *M. oleifera and M. stenopetala* seeds; as well as variation in the genome structure through transcriptomic analysis of the leaves, stem and root tissues of *M. oleifera, M. stenopetala and M. peregrinna*.

Depending on the species and assembly technique used, Moringa's genomes vary in size, from 236.4 Mb (Chang *et al.,* 2022) to 281 Mb (Shyamli *et al.,* 2021) and up to 315 Mb (Tian *et al.,* 2017). The first sequenced and published *M. oleifera* Lam.'s whole genome was in 2015 (Tian *et al.,* 2015). Chang *et al*. (2022) generated and published the AOCCv1 draft genome from domesticated *Moringa oleifera* strains. AOCCv2 was published in 2022 providing a chromosomal-level reference assembly with a thorough and methodically ordered gene complement. Except for the plastome (chloroplast genome) of the Egyptian wild *Moringa peregrina* (Abdalla *et al*. 2023), data on the exact molecular characterisation of Moringa is conspicuously lacking mostly due to insufficient genomic sequencing of wild species.

Despite the valuable resources of *Moringa spp.,* they have not been fully utilized, and its conservation is retarded because its seeds and other genetic material are not easily accessible for research and cultivation. (Popoola *et al.,* 2016). According to Azevedo *et al*. (2019), most landraces are still unknown, and our knowledge of the inter-relations across several species is both restricted and lacking. According to Beiki *et al*. (2019), thorough annotation and characterising recently identified transcripts and genes in the Moringa transcriptome are still lacking. Moreover, Moringa's extensive range and continuous international trade mean that gene flow among different accessions can happen both inside and between countries.

Based on the present knowledge of Moringa genome data, molecular studies are required to evaluate the degree of gene flow among accessions from different sites. This research enhances the understanding of the migration patterns and evolutionary history of these accessions. To find better genotypes capable of surviving under various breeding situations, a thorough review of the evolutionary background, population genetics, gene flow, and associations of this important plant is required (Salgotra *et al.,* 2023).

Though much has been done, knowledge of the genetic complexity of Moringa still lags, because of inadequate research on the genetic variety among Moringa's several

species, there is still much to learn about the uses of the plant in many spheres, including agriculture, nutrition, and medicine. While other members of the Moringa genus have been mainly neglected, recent genomic research have concentrated mostly on few Moringa species, especially *Moringa oleifera*. This neglect of their genetic variety and possible uses hinders their understanding. Generalising results from genomic studies of understudied Moringa species is difficult, rather there should be availability of genetic information on the other 12 species of Moringa to improve the knowledge of genetic variation across the species and possible uses of each Moringa species. Most especially clarifying the complexity of gene function and control, knowing the genetic diversity and comparing the genomes could be tasking resulting from limited dearth of knowledge of the functional genomics of different Moringa species (Manavalan *et al*. 2021; Gandji *et al*. 2018; Lakshmidevamma *et al*. 2021; Pirr'o *et al*. 2016).

Most studies of the Moringa genome have shown genes having diverse benefits such as the Superoxide dismutase (SOD), Catalase (CAT), Ascorbate peroxidase (APX) genes which encode antioxidant enzymes that scavenge reactive oxygen species (ROS), (Hajaji *et al*.,2024). Likewise, Glucosinolates and isothiocyanates pathway genes which have anti-inflammatory and anti-microbial compounds (Chang *et al*., 2022) and VTE1 which regulates vitamin biosynthesis and increases the levels of vitamins A, C and E making Moringa highly nutritious. Also, there are Dehydration - responsive element-binding (DREB) transcription factors which help the plant survive under water-deficit or heat stress (Vijayaragavan *et al*., 2025). Genes involved in secondary metabolite biosynthesis such as flavonoids which have protective effects against many infectious diseases and inducing death of certain, cancer cells have also been described according to Chang *et al*., 2022.

The current studies fall short insufficiently addressing how epigenetic elements affect Moringa genomes (Boopathi *et al.,* 2021). Earlier studies on Moringa have mostly focused on specific areas, especially those marked by significant Moringa farming (Gandji *et al.,* 2018; Mashamaite, 2021). As Davis-Turak *et al*. (2017) and Ejigu and Jung (2020) have shown, the lack of uniform genome annotations hampers the integration and comparison of research data. Cooperation and the building of a coherent framework for genomic annotation will help to better grasp the genetic terrain of Moringa. Crucially, multi-faceted solutions combining functional genomics with transcriptomics into a single framework are required.

1.6 Drought stress tolerance in Moringa

Moringa plants are a major agricultural benefit among climate change difficulties since they show unique features that allow their survival in arid environments (Chukwuebuka,

2015; Sahoo *et al.,* 2020; Emongor, 2011). The traits include a deep taproot system that enhances water availability over a large soil volume, tiny leaves with a waxy surface that boosts water retention in arid conditions, and the ability to turn on antioxidant enzymes protecting cells from damage.

There are evidences of genes found in Moringa that control the expression of stress-responsive genes (Islam *et al.,* 2023; Shyamli *et al.,* 2021) and protect cell structures under drought stress. Despite Moringa's drought-tolerant adaptability, it still experiences reduced leaf production, photosynthesis and yield under water shortage conditions. It is therefore important to study the mechanism of Moringa's adaptability for its application in the advancement of sustainable agriculture. An investigation of the genetic basis of drought resistance in Moringa will significantly increase world agricultural sustainability (Alavilli *et al.,* 2022). Analysing the techniques and genetic elements allowing Moringa to flourish in arid environments could help to create drought-resistant variants and support environmentally friendly farming methods. This research would improve food security, address water issues, support agricultural methods, and help to conserve resources (Gordon *et al.,* 2010). By means of important genes, processes, and regulatory networks connected to drought tolerance, one can maximise the benefits of plants and enhance the production of drought-resistant cultivars (Cattivelli *et al.,* 2008).

1.7 Benefits of Moringa seed oil

Derived from *M oleifera* seeds, moringa oil is becoming well-known for its nutritional value and possible health advantages (Anwar *et al.,* 2003). The nutritional profile of the fatty acid composition, tocopherol content, and phytosterol levels is good. Moringa seed oil could be considered better than olive oil (Ruttarattanamongkol *et al.,* 2015; Tsaknis *et al.,* 1999). While the considerable behenic acid concentration offers unique structural and stability properties favourable for many industrial uses, the high oleic acid content is connected to benefits in cardiovascular health. Together with significant levels of phytosterols, gamma and α-tocopherols improve its antioxidant and cholesterol-lowering effects. Moringa oil's unique health advantages and functional qualities guide its use in both food and industrial settings.

1.7.1 Phytosterols

Phytosterols are sterols found in great abundance in plants. The chemical structure of this plant sterols consist of a hydroxyl group at position 3 of the steroid structure, attached to a C-17 side chain and either one or more double bonds (Moreau *et al.,* 2018). Preserving the integrity and functionality of plant cell membranes depends critically on phytosterols (Rogowska, 2020).

Studies show that by lowering the absorption of low-density lipoprotein (LDL) cholesterol (Poli *et al.,* 2021), phytosterols may improve cardiovascular health, lower the risk of inflammatory diseases, and guard cells from oxidative damage. Like olive oil, moringa oil has substantial levels of β-sitosterol, campesterol, stigmasterol, and brassicasterol (Ferreira *et al.,* 2008; Stark & Madar, 2002). Moringa seed oil's phytosterol content has extraordinary qualities. Although olive and Moringa oils have similar phytosterol profiles, oils with higher phytosterol concentrations such as sunflower and canola may be more successful in controlling cholesterol levels (Abdulkarim *et al.,* 2005).

1.7.2 Fatty Acid Methyl Esters (FAMEs)

Oleic and palmitic fatty acids are the two most often occurring ones in Moringa oil. Comparable to olive oil, moringa oil comprises 9-10% palmitic acid (C16:0), 5-6% stearic acid (C18:0), and 65-75% oleic acid (C18:1), claims Tsaknis (1998) and Idris *et al*. (2020). Olive oil's primary quality is its fatty acid content. About 55–83% of the oil is oleic acid; palmitic acid constitutes about 7.5%–20%; stearic acid produces about 0.5%–5%. Milla *et al.,* 2021 reports less than 1% α-linolenic acid (C18:3) and 3.5– 21% linoleic acid (C18:2). Moringa oil is unique among other well-known edible oils in that it contains behenic acid (C22:0) at roughly 6–9%.

1.7.2.1 Implications of High Behenic Acid Content in Moringa oil

High behenic acid (C22:0) content makes Moringa oil unique and offers both positive and negative uses in the food, cosmetic, and industrial spheres. The significant oxidative stability of behenic acid's saturation improves the conditioning and efficacy of skin care products while extending their freshness. Behenic acid, a cholesterol-raising saturated fatty acid in humans is poorly absorbed (Cater and Denke, 2001) it is therefore important to exercise caution due to its poor absorption and potential to raise low-density lipoprotein (LDL) cholesterol levels. Furthermore, its strong qualities can make it inappropriate for food sector liquid oil use. Combining Moringa oil with other oils heavy in unsaturated fats, such as olive or canola oil, to maximise benefits and reduce health hazards. More study is required to fully evaluate the health hazards related to behenic acid and investigate improvements in its industrial uses.

1.7.3 Tocopherols

Acting as antioxidants, tocopherols protect oils from oxidation and improve human health by reducing cellular lipid peroxidation. Moringa oil is richer in α-tocopherol (134.42mg/kg), beta tocopherol (93.7mg/kg) and γ-tocopherol (48mg/kg) according to Anwar *et al*.,2003 compared with the record of Olive oil according to Tsaknis *et al*.,1998 where α-tocopherol content is 88.5mg/kg, beta tocopherol (9.9mg/kg) and γ-tocopherol(1.6mg/kg). Although, sunflower oil possesses higher alpha tocopherol level

of 613mg/kg (National Sunflower Association, 2001), than both Moringa oil and olive oil. Nadeem and Imran, 2016). Among the tocopherols composition in Moringa oil, the strong antioxidant γ-tocopherol has the highest value (Pluháčková *et al.,* 2023).

1.8 Potential for improvement of *Moringa oleifera*

Several techniques exist for the development of *M oleifera* to promote resilience and adaptability to different environments. Along with improving the plant's tolerance to environmental challenges such drought, salinity, and insect infestations (Muniandi *et al.,* 2024), this entails spotting and choosing genotypes that show disease resistance and great yield. Although some studies have used molecular markers such as Single nucleotide polymorphisms (SNPs) for the selection of high-quality crop varieties (Amiteye, 2021). Research on Moringa has mostly focused on a limited number of accessions from a limited range of areas or cultivars. A thorough knowledge of genetic variety calls for including both wild and cultivated types in the research. Furthermore, research on *M oleifera* around Genome Wide Association Studies (GWAS) where *M. oleifera*'s phenotypic characteristics are linked to genetic variants has not been adequately studied. Establishing a high-quality reference genomic data for other species such as *M. stenopetala* and *M. peregrinna* rather than just *M. oleifera* is crucial for improving GWAS and clarifying the genetic bases of important agronomic, nutritional, and medicinal features as genomic technologies including whole-genome sequencing and SNPs identification (Green and Sambroak, 2012).

Numerous genetic variations or genes connected to phenotypic variability in *M. oleifera* remains unidentified in absence of GWAS investigations. Knowing the molecular mechanisms with which certain unique traits are exhibited and linked to genetic and phenotypic variation requires thorough investigation (Wang *et al.,* 2021). This will help to enhance techniques of marker-assisted selection. Still mostly unknown is how transcriptome modifications affect the phenotypic variability of *M. oleifera*. Analysing the epigenetic and transcriptome characteristics of several *M. oleifera* accessions can help researchers better grasp the regulatory mechanisms influencing phenotypic diversity.

Unknown still is how farming techniques, soil, and temperature affect the phenotypic features of *M. oleifera*. Examining genotype by environment interactions can help one to grasp how this species adapts to different growing situations. Though more study on these topics is being done, it is still difficult to know the genetic processes controlling *M. oleifera*'s nutritional profile and drought stress tolerance. Different genetic and physiological elements affect Moringa's tolerance to drought which clearly connects with specific genes and genetic markers. There have been found genes controlling the activation of drought-responsive genes (Kumar *et al.,* 2023). Furthermore, recorded are

genes linked in synthesis and response to stress hormones (Singh *et al.,* 2023). Moreover, the ability of Moringa to control stomatal opening, retain water in its leaves, and withstand arid conditions is linked to genes associated with cell wall fortification, antioxidant action, and the synthesis of protective molecules which have been studied. Deeper knowledge of the molecular mechanisms and confirmation of these relationships could help to choose Moringa variants with improved drought resistance by use of genetic markers (Sharma *et al.,* 2024).

Many genetic and environmental elements affect the metabolic profile of Moringa oil and the various Moringa plant tissues such as the leaves, stem, roots, flowers, pods, and seeds. The genetic differences provide varying degrees of nutrients and health promoting components (Ogunsina *et al.,* 2011). Comparative genomic research and transcriptome analysis point to genes that synthesize and store vital nutrients like vitamin C (Ogbe *et al.,* 2011), carotenoids, and glucosinolates. The genetic pathways and regulating mechanisms controlling Moringa's nutritional composition are not yet well known. To improve knowledge of the genetics linked with Moringa's nutritional qualities, researchers now combine transcriptomic, genomic, and metabolomic data (Rurek and Smolibiwski, 2024). Though much progress has been made, more study is needed to clarify the nutritional profile of *M oleifera* and the genetic control of drought stress. As Singh *et al*. (2020) point out, further research on the genetic basis and enhancement of vital Moringa features calls for functional genomics, genetic mapping, and genome sequencing.

1.9 Improving Moringa by transcriptomics approach.

In a bid to improve Moringa, the transcriptome technology could provide benefits over the genome approach. This enables easy discovery of non-coding RNAs, study of regulatory networks, identification of splice variants and isoforms, and identification of differently expressed genes under certain treatments.

Transcriptomics is a technique used to examine RNA expression patterns in the Moringa plant to find genes with varied expression in response to stress, nutrient availability, or development stage. By means of gene identification that show either up- or down-regulation, Moringa's beneficial properties like stress resilience and biomass output can be improved to enable more efficient targeting of the pertinent pathways and mechanisms. Moringa's transcriptome analysis clarifies the complex regulating systems controlling gene expression patterns and the interactions among several transcripts, which is useful for guiding focused breeding or genetic modification projects meant to enhance particular Moringa plant features.

Transcriptomic methods give insights into splice variants, gene isoforms and differential gene expression in organisms. Genomic analysis might not sufficiently encompass this depth of information, as it emphasizes DNA sequence at the expense of the dynamic aspects of gene expression (Dai and Shen, 2022).

Recent transcriptome research has clarified the important roles non-coding RNAs, including long non-coding RNAs and microRNAs, have in the development and gene expression control of wheat plant, (Gupta *et al.,* 2021) suggested that targets for genetic changes meant to enhance desirable features could be regulatory RNAs. This idea can also be applied to Moringa plants.

The transcriptome approach provides a fast and exact assessment of plant reactions to environmental changes like temperature variations, and drought (Abdullah, Ab Rashid and Ghazalli, 2018). The results might be quite significant in producing Moringa varieties displaying improved resilience to environmental circumstances whilst offering a thorough investigation of Moringa's dynamic reactions, regulatory networks, and gene expression patterns. This information might help to create better strategies for enhancing this vital crop.

1.10 Survey of Moringa Variation by *de novo* Transcriptomics

Analysing and characterising genetic heterogeneity within Moringa populations with *de novo* transcriptomics identifies sequence variations including SNPs and INDELs among several Moringa accessions. Transcriptome sequence analysis of Moringa plants taken from diverse global locations would reveal notable species genetic variation. These findings help identify desirable traits and understand the genetic basis of traits which are molecular markers to improve cultivation and yield, population structure and germplasm research (Ondieki *et al*., 2023; Hölzer and Marz, 2019).

Annotating assembled transcripts reveals changes in gene expression, splicing patterns, or coding sequences, therefore clarifying the functional variation in the Moringa coding regions. This information helps direct efforts at genetic engineering meant to improve desired features or identify useful genetic variants in breeding projects. Comparative study of transcriptomes can help to identify common pathways and to regulate mechanisms among several Moringa species. This approach would help to identify the sources of metabolic pathway diversity, adaptability to many environments and phenotypic heterogeneity. *De novo* transcriptomics helps to find genes possibly controlling important features such as phytochemical production, stress tolerance and nutritional use efficiency (Wang *et al*., 2019). By examining differential expression patterns and co-expression networks within transcriptomes, it is possible to identify specific genes or gene modules likely involved in regulating essential characteristics (Zhao *et al*., 2019). These candidate

genes can be further studied through functional genomics, genetic engineering, or marker-assisted breeding to improve specific traits in Moringa.

Moringa tissues such as leaves, stem and roots subjected to drought stress could be collected to show tissue-specific transcriptional pattern. This information therefore clarifies the sites of gene activation by identifying specific genes and pathways involved in Moringa's growth, development and metabolite production of useful compounds, (Martin & Wang, 2011). Finding important genetic variations for the improvement of crops using genetic engineering or selective breeding has been made possible in great part by *de novo* transcriptomics

1.11 Differential gene expression in Moringa

Differential gene expression generally improves the knowledge of genetic control in response to stress, across several tissues, and during many phases of development. This approach uses multiple hypothesis testing and takes into consideration false discovery rate to be able to detect those genes that are differentially expressed in a given treatment (Jones *et al*., 2006; Jones *et al.,* 2009). This approach is applicable to identifying important genes and enzymes connected to drought stress reactions and adaptability, synthesis of valuable phytochemicals from Moringa plant tissues. Apart from supporting initiatives for breeding such as molecular cloning (Green and Sambrook, 2012), it clarifies genetic diversity. Martin & Wang, 2011's Next generation assembly technique would help to identify genes linked to important biological purposes. Integrating this data with other omics studies, according to Zhou *et al.,* (2014), improves our knowledge of Moringa's biology and helps to produce better cultivars and products. Integrating transcriptome data with metabolomic profiles, according to, improves our knowledge of metabolic pathway control and helps to devise strategies to maximise target chemical synthesis.

1.12 Weighted-Gene Co-expression Network Analysis (WGCNA)

WGCNA lets Moringa genes be arranged into modules based on their activity levels. Interaction of genes can influence important processes including stress response, development, and growth (Langfelder & Horvath, 2008; Zhang & Horvath, 2005). Beyond differential expression analysis, these modules reflect co-regulated gene networks or pathways that deepen our knowledge of plant biology at a systemic level. Emphasising the plant's reactions to drought stress, this study clarifies the links between Moringa characteristics and certain gene clusters, referred to as "hub genes". This study finds important genes with great connectivity that are necessary for the control of the module. Target genes for breeding or genetic modification might be these ones. By means of efficient integration of gene expression data with other omics data types, including

metabolomics, proteomics, and epigenomics, WGCNA helps to understand the complex traits of Moringa (Li *et al.,* 2018). Under different situations, comparing gene networks helps to find biomarkers and genetic targets for improving Moringa variants (Fuller *et al.,* 2007; Horvath *et al.,* 2006). WGCNA offers tools for gene network visualisation, thereby helping to comprehend and share results (Langfelder & Horvath, 2008). WGCNA data can be used by Moringa researchers to improve their knowledge of important gene regulating networks, identify critical genetic elements impacting important features, and create plans for focused genetic enhancement of the plant.

1.13 Thesis Outline

This thesis has five chapters. Chapter 1 addresses the fundamental questions of the research, which concentrated on the possibility of breeding crops with unique features and the survival mechanisms of this under-used crop in drought conditions. Emphasising the genes necessary for the plant's reaction to drought circumstances, Chapter 2 explores the genomic and transcriptome pathways allowing Moringa to adapt to arid conditions. The main genes and mechanisms of drought resistance are covered in this section. Chapter 3 looked at ways to improve the health advantages from growing Moringa seeds. Focussing on improving Moringa seed oil's medicinal and nutritional qualities, the study undertook thorough investigation on its characterisation. Examining the changes in phytochemical components related with various seed sources, this study aimed to increase the utility of Moringa seed oil. By means of a comparative transcriptome analysis, Chapter 4 provided understanding of the genetic variation among the investigated Moringa species and accessions. An integration and analysis of the outcomes of the previous chapters are given in Chapter 5. By compiling the facts from previous chapters and analysing their inferences, Chapter 5 gave a thorough scientific ending to the thesis. This thesis is to evaluate the drought-responsive genes and understand the pattern of adaptability to drought stress in Moringa, compare the metabolic profiles of diverse Moringa seeds in relation to their global distribution, genetic robustness, and phytochemical diversity to improve production techniques and widen uses in crop breeding.

Chapter 2

Drought-Responsive Gene Expression Patterns in *Moringa oleifera* through RNA-seq and WGCNA

Abstract

Moringa oleifera is widely acclaimed for its remarkable drought tolerance, yet the molecular mechanisms underlying this resilience remain largely unexplored. This study examined differential expression by comparing drought-stressed leaves and stem tissues to well-watered controls, aiming to identify significant changes in gene expression. Also, I investigated the complex interactions among genes in response to drought stress conditions. RNA sequencing (RNAseq), Differential Expression Sequencing (DESeq2) and Weighted Gene Co-expression Network Analysis (WGCNA) were employed to bridge this knowledge gap by profiling the gene expression patterns of 16-day old Moringa under drought stress, deprived of water for 5 days. An analysis of co-expression was performed to analyse the gene networks and modules identified. The results indicated genes exhibiting increased or reduced expression in either the leaf or stem tissue during drought conditions, implying alterations in the plant's transcriptional response to these stressors. Furthermore, the findings of this research included identification of upregulated tissue-specific drought-responsive genes, regulatory elements which include transcription factors, protective proteins, stress and signalling molecules important for adaptation to drought. Likewise, co-expression gene modules, revealing the regulatory architecture and hub genes that drive drought responses in Moringa were identified. Notably, Morol12g05670 (SAG12), a senescence-associated gene, and Morol01g26440 (ferritin-like domain) were upregulated under drought stress. Genes like Morol02g21300 (Peroxidase N1-like) were downregulated. Whereas, SAG12 and Vignain were significantly upregulated in the leaves, SAG12 and Gawky-like protein in the stem and KNAT2 and SBH1 increased in both stem and leaf under drought stress. Modules including cyan, blue, and green were enriched for stress-responsive genes related to autophagy, stomatal regulation, and oxidative stress. This transcriptional response of *M. oleifera* to drought stress shows a complex, tissue-specific mechanism involving stress signalling, senescence, protein turnover, and a developmental regulation. These findings hold significant implications for both fundamental plant biology and applied agricultural sciences, particularly in the context of increasing global water scarcity and climate change.

2.1 Introduction

Food security for the increasingly growing global human population coupled with the effect of climate change affecting agriculture requires looking into the benefits of such valuable and under-utilized plants as *Moringa oleifera (Hirokuyi, 2018).* Taking advantage of its drought-resilience, adaptability to diverse conditions, fast growth, pest and disease resilience and the valuable nutritional resources from its leaves, seeds, pods and flowers would be helpful in plant breeding (Vijayaragavan *et al*., 2025, Hodas *et al*., 2021). A number of researches into the Moringa plant drought response dwelt on identification of drought - stress related gene families and pathways including Heat shock factors (HSF), Transcription factors (TF) (Shyamli e*t al*.,2021), assessing recurring water stress (Rivas *et al*.,2013), and using the Moringa leaf extract as mitigation for drought stress in other plants such as Mung bean (Haris *et al*.,2024), rice quality (Khan *et al*., 2023). Drought has tremendous effect of agricultural processes and quality food provision in the long run, as it limits the quantity and quality of plant growth, it is therefore important to investigate the drought stress tolerance of *M. oleifera.* This will assist in sustainable agricultural practices in dry and semi-arid areas. Characterised by their special adaptations to arid conditions, including a deep taproot system that reaches up to 10-15 feet from the base of the tree enabling access to water and nutrients across a substantial soil volume, small leaves with a waxy surface that improves water retention in arid conditions, and the capacity to activate antioxidant enzymes that safeguard cells from damage. *M. oleifera* has genetic features protecting cellular components from drought by changing the expression of genes linked to stress responses (Shyamli *et al.,* 2021). *M. oleifera* plants show great agricultural potential to mitigate climate change (Trigo *et al.,* 2022). Promoting sustainable agriculture and addressing drought in *Moringa spp..* depend on an awareness of the genetic foundation of drought resistance in this plant. This work sought to identify the genes expressed in respect to drought stress tolerance in *Moringa oleifera*, so I explored the drought-stress response by asking these questions, which genes were significantly differently expressed under drought stress, what genes were co-expressed that is, identified the modules related to the transcriptional response of Moringa under drought? Under what genetic processes does Moringa species' drought - resilience arise? Which genes mostly control the drought - response transcription, that is the hub genes? With respect to other drought-tolerant crops, how does Moringa fair?

2.1.2 Variability among Moringa Species

Different species of Moringa show different physical traits that might influence their survival during a drought. Different root systems, seed morphologies, and leaf diameters help one distinguish the three species of Moringa: *M. oleifera*, *M. stenopetala,* and *M. peregrina*. *M oleifera* is much prized because of its huge taproot which allow it to obtain water from deeper soil levels and tiny, waxy leaves which allows it to conserve water during drought. *M. stenopetala*'s thicker stems and wider leaves could affect its water use efficiency. Realising adaptive strategies including drought resistance requires an awareness of the variations within the Moringa genus. Breeding projects and varietal choice can benefit from this understanding. This study focusses on the genetic adaptations of *M. oleifera* in reaction to drought stress.

2.1.3 Physiological Indicators of Drought Response

This study assesses several physiological markers of water stress to improve knowledge of Moringa's drought-adaptive capacity. By reducing water loss and preserving cell integrity, monitoring these elements helps one to grasp how plants control drought situations. The plant uses increased activity of antioxidant enzymes in times of dehydration to preserve homeostatic cell viability. As they relate to molecular alterations, traits like lowered stomatal conductance and higher root-to-shoot ratios are essential for connecting patterns of gene expression to physiological reactions.

2.1.4 Co-expression network

Co-expression networks identify genes that share similar patterns of expression. These are grouped into "modules" which often group functionally related genes. Within each module, genes (nodes) are connected with a strength based on their gene expression correlation. These connections (edges) between the genes form the gene network, with "hub" genes being those with the most connections. In a WGCNA, these are hypothesised to be genes which have a regulatory function within the module.

For instance, several genes are involved in photosynthesis; these include those controlling stomatal conductance, relative water content, canopy temperature, electrolyte leakage, hormone balance, and chlorophyll fluorescence, rather than looking at individual genes, this approach lets one identify modules in Moringa that concentrate on groups of genes controlling drought stress resistance. By controlling co-expression patterns among other genes, Langfelder and Horvath (2008) and Liu *et al*. (2019) showed that hub genes within modules preserve their stability and integrity. These genes could be involved in biotic and abiotic stress reactions as well as in responses to drought. Liu *et al*. (2019) propose that these genes most likely encode transcription factors essential for plant defense systems reducing oxidative stress damage in drought

circumstances. Eigengenes, which indicate expression profiles inside the modules, help one to assess the interaction between modules and their exterior characteristics. This summary captures the whole expression pattern of the module succinctly. By means of co-expression, the network shows the interactions among genes inside a module, therefore enabling the discovery of gene groups with comparable expression patterns.

With different patterns seen in stem and leaf tissues, this work intends to show that drought stress greatly modulates gene expression in *Moringa oleifera*. I hypothesised that some genes would show noticeably altered expression patterns in response to drought compared to well-watered controls, and that this investigation could identify hitherto unknown genes that are crucial for the adaptive response of the plant. This knowledge would help in understanding the transcriptional alterations linked to drought, signal transduction pathways, photosynthetic control, and responses to oxidative and water stress, all of which help to explain drought adaptation. Functional annotation was also done to identify the Moringa drought-responsive genes. This study sought to find mechanisms and identify gene networks that control drought tolerance which could be applicable to growing plants in arid areas. In the end, I sought to discover important regulating hub genes in Moringa that coordinate the transcriptional response to drought conditions, hence determining master regulators of drought stress.

2.1.6 Relevance of Weighted Gene Co-expression Network Analysis (WGCNA)

WGCNA identifies gene groups that interact and control drought responses by aggregating gene profiles into co-expression modules, therefore revealing gene networks influencing complex traits including stress adaption. This approach makes it easier to find "hub genes" enhancing drought resistance in agricultural breeding. Generally, drought-tolerance response in plants are complex and involve varying pathways which allows them respond to stress at any given time. They do this in several ways by regulating the rate of gas exchange and water loss through stomatal aperture, carotenoid degradation, anthocyanin accumulation, intervention of osmo-protectants such as sucrose, glycine and proline, ROS (reactive oxygen species) scavenging enzymes among others. These complex reactions to control drought in plants is also influenced by environmental factors as well as transcriptional factors such as DREB (dehydration-responsive element-binding protein, abscisic acid (ABA)-responsive element-binding factor (AREM) and NAM (no apical meristem) (Nakashima and Suenaga, 2017; Shrestha *et al.*, 2021). There exists some information about drought tolerant genes and network recorded in *Moringa oleifera, Oryza sativa* (rice) and *Zea mays* (maize).

Drought study on *Moringa oleifera* by Shyamli, 2021 revealed expanded families of heat-shock transcription factors (HSF) and highlighted drought-induced accumulation of osmolytes such as proline and polyphenols as the key survival mechanisms of Moringa to combat water stress. The research also reported an upregulation of ion transporters, stress-responsive proteins, antioxidant enzymes under stress. Similarly, Chang *et al*., 2022 mentioned that Moringa's successful strategy for coping with long duration of drought involved long term-investment of resources into multiple secondary metabolites such as isoprenoids and flavonoids which is a way of reducing fatal photo-inhibitory processes and oxidative damages. Furthermore, Chang *et al*. (2022) noted that model species such as *Arabidopsis thaliana* possess orthologs of the HOG0003249 orthogroup (AT5G04530.1, AT2G28630.1, and AT1G07720.2), encoding members of the 3-ketoacyl-CoA synthase family. These enzymes are essential for synthesizing very-long-chain fatty acids (VLCFAs), which contribute to cuticular wax production and thereby reduce non-stomatal water loss during drought.

Table 2.0 highlights the comparative overview of drought-responsive genes and associated mechanisms in *Oryza sativa* (rice), *Zea mays* (maize), *Arabidopsis thaliana* and *Moringa oleifera*. These include transcription factors, enzymes and regulatory elements which are essential for plant survival during drought.

Rice and maize primarily utilize abscisic acid (ABA)-dependent signalling pathways, including the PYR/PYL-PP2C-SnRK2 cascade and DREB/NAC regulons. Notably, rice shows a prominent involvement of AREB/ABF, WRKY, and ROS-ABA feedback mechanisms (Todaka *et al*., 2019). In contrast, *Moringa* displays increased expression of ABA-pathway genes, HSFs, DREB/NAC/WRKY-like transcription factors, and adaptive metabolic shifts, such as enhanced proline synthesis, antioxidant accumulation, and altered root architecture in response to drought (Vijayaragavan *et al*., 2025; Shafi *et al*., 2020).

Arabidopsis serves as a robust model for dissecting ABA-mediated gene regulation, with key players including SDIR1 E3 ligase, calcium-dependent protein kinases (CPKs), and NAC transcription factor-mediated feed-forward loops (e.g., NAC016-NAP→AREB1), as well as MYB44/ENAP-facilitated WRKY activity (Sakuraba *et al*., 2015; Mahmood *et al*., 2019; Nguyen *et al*., 2019;Yang *et al*., 2022).

Table 2.0 Comparison of drought-responsive genes in *Oryza sativa* (rice), *Zea mays* (maize), *Arabidopsis thaliana* and *Moringa oleifera*.

	Rice (*Oryza sativa*)	**Maize (*Zea Mays*)**	***Arabidopsis thaliana***	***Moringa oleifera***
ABA signalling	AREB/ABF bZIP TFs (e.g. OsAREB1); *NCED* and CPK6 activation under drought (Zhang *et al.*,2020; Zhou *et al.*, 2025)	PYR/PYL–PP2C–SnRK2 core; MAPK cascades (ZmPP2C84 → ZmMEK1/ZmSIMK1)	SDIR1 E3 ligase enhances ABA-related signalling; CPKs (e.g. CPK6) phosphorylate ABFs for stomatal control (Harb *et al.*, 2010)	Upregulates *NCED* and ABA-responsive CDPK/HSF genes under drought; genome reveals expanded HSF family (Pasha *et al.*, 2020; Chang *et al.*,2022)
DREB regulon	OsDREB1/2 induce osmotic and heat stress genes	ZmDREB2.7 activates osmotic-stress response genes	DREB family strongly co-induced under drought; ~2000 responsive genes, two-thirds ABA-dependent	RNA-seq detects DREB-like TFs among drought-induced transcripts
NAC TFs	OsNAC1 / OsNAC6 regulate protective genes	ZmNAC48/49/111/075 for drought resilience	NAC016/ NAP form feed-forward loop repressing AREB1; NAC019/55/72 (Nguyen *et al.*, 2019; Sakuraba *et al.*,2015)	Transcriptomic data shows upregulation of NAC family members under drought
WRKY TFs	OsWRKY13/45 link abiotic and biotic stress	WRKYs integrated in MAPK cascades	MYB44–ENAP interaction regulates histone acetylation and WRKY70 expression (Zhao *et al.*,2022)	Likely expressed; transcriptome shows stress-related WRKY induction (needs functional validation)
Hormonal crosstalk	ABA–GA balance; ROS–ABA via OsAAI1	Ethylene/auxin/ABA balance; root vs shoot allocation via lignin and hormone modulation	Extensive cross-talk: ABA dominant but JA, ethylene, auxin, cytokinin also modulate drought genes	Accumulates proline, phenolics, antioxidants; extensive root growth and stomatal regulation via ABA
Root adaptation	ABA and GA coordinate root growth under drought	Maintains root apical growth via ABA & IAA; boosts lignin biosynthesis	ABA regulation of root hydraulic conductance and aquaporin expression	Robust root system, stomatal regulation; transcriptomes enriched for water uptake and secondary metabolism pathways.

These compounds might be markers for the beginning and development of drought-induced ageing. By clarifying the molecular mechanisms controlling stress responses, research of hub genes within co-expression modules helps to guide methods to improve Moringa's drought tolerance. Studies show that *M oleifera* is rather resistant

to dry environments. According to Nouman (2013), this plant can maintain a good amount of its leaf area and photosynthetic activity even with low rainfall.
Stress from drought seems to improve Moringa's nutritional value. Studies show Moringa to be nutritionally robust and resilient. Whereas iron, phenolic compounds, and vitamin C are more common in cool and dry seasons, levels of beneficial components including vitamin A rise during hot and humid seasons (Zarina *et al.,* 2024; Yang *et al.,* 2006, 2007). Due to its drought resilience and improved nutritional profile in demanding conditions, *M oleifera* is a major plant for farming in such conditions.
Different patterns shown in the stem and leaf tissues of *M. oleifera* revealed how drought stress greatly influences gene expression. I hypothesised that in drought conditions some genes would show noticeably different expression patterns than in well-watered controls and that this study would identify new or previously uncharacterised genes necessary for the adaptive response of the plant. By use of WGCNA, co-expression modules linked with physiological features and stress indicators were found, thereby enabling an insight of the transcriptional changes brought about by drought. I reasoned those processes and signalling pathway including photosynthetic regulation, oxidative stress response, and water transport define drought adaptation in biological terms. Using functional annotation and network analysis, I examined the drought-responsive network connections in Moringa against those in other drought resistant plants. Comparative study of these two data sets will highlight both conserved mechanisms and unique paths, therefore improving the knowledge of new drought tolerance techniques (Ahmad *et al.,* 2018). The main goal was to find the core regulating hub genes controlling the transcriptional response to drought stress.

Aim and objectives of the experiment

The aims of this experiment were to examine the significant changes in gene expression associated with drought-stress in *Moringa oleifera,* compare differentially expressed genes (DEGs) between leaf and stem tissues under drought conditions, identify gene co-expression modules that correlate with physiological traits and drought-related indicators, to elucidate the biological processes and signalling pathways involved in drought adaptation in *M. oleifera*, identify key regulatory hub genes that coordinate the transcriptional response to drought stress and compare Moringa's drought-responsive pathways with those of other drought-tolerant species such as wheat and rice, as well as closely related taxa within the Brassicales such as broccoli, cabbage and kale. This comparison would give insight into novel or previously uncharacterised genes as well as novel drought tolerance strategies.

Reproducibility of Results

The data trends were reproducible across replicates and experiments. Raw data has been deposited in the Sequence Reads Archive (SRA) with accession reference :
PRJNA1197419. Scripts in Github https://github.com/andreaharper/HarperLabScripts

2.2 Materials and methods

2.2.1 Soil moisture content measurement

Standard untreated peat-based seed and cutting compost soil (F2+S) was collected. The soil sample underwent spontaneous air-drying over a two-week period. Twenty-four empty 8 cm pots were weighed, and a consistent dry soil sample of 100 grams was added to each pot. The moisture level was subsequently assessed with a moisture meter and documented. The soil saturation point was determined by evaluating the soil's porosity, which included measuring the volume of water necessary to fully saturate dry soil. Different water volumes (20ml, 40ml, 60ml, 80ml, 100ml, 120ml, and 140ml) were applied to three replicates of soil pots. The moisture content was assessed and recorded by inserting a probe into the soil with an ML3 Thetaprobe Soil Moisture sensor, paired with an HH2 Moisture Meter (γ-T Devices, Cambridge, United Kingdom). The mean soil moisture content percentage for 100 grams of soil was determined to be 38.87%, derived from the 80 millilitres of water needed for complete saturation. Both the control and

treatment groups received 80ml of water until the treatments were separated, at which point watering was halted for 5 days. By the fifth day, the soil exhibited complete dehydration, registering an average moisture content of 5.56%.

2.2.2 Sowing

Moringa oleifera seeds obtained from Ebay, UK (Mantig) were used for this experiment. The seeds underwent a pre-treatment process that involved cleansing with a solution consisting of 1 litre of distilled water and 200 millilitres of 4.5% diluted bleach to ensure purity and absence of contaminants such as germs and pests. The *Moringa oleifera* seeds of Ebay (Mantig) accession were placed into 50ml centrifuge tubes containing the bleach and water mixture, allowed to sit for 5 minutes, and then removed. The seeds underwent at least four washings with distilled water. The cleaned seeds were overnight submerged in water to improve seed germination. Sowing took place in the Rice block at the University of York Green House using illumination set to turn on at 5:00 AM and off at 5:00 PM. Daylight lasts twelve hours with supplementary illumination when light levels drop below 120 watts/m². Day heating set point at 28ºC and night heat set point at 24ºC.

For the RNASeq drought experiment (Chapter 2), 12 pots were filled with 100 grams of soil, and pre-treated seeds were positioned at a depth of 2 centimetres to prevent the sprouting seedling run out of energy before reaching the ground surface and start fast leaf development. Subsequently, 80 millilitres of water were added to each pot, which were then covered with lids to promote germination.

Similarly, for the transcriptome experiment requiring RNA extraction from leaves, stem and root tissues (Chapter 4). Seeds of 49 Moringa accessions were sown in seventy pots as earlier described. While most seeds germinated in five to seven days, few seeds of MO4, MO30, MO54, and MO63 accessions required about ten days to germinate.

On the seventh day post-sowing, 28 pots exhibited signs of germinated seeds. In the subsequent week, an additional 18 had germinated, resulting in a total of 2 weeks. The germination success rate is 46 out of 70. A subset failed to germinate and was discarded after the third week. While *M. oleifera,* MO8 accession was removed from the batch since it failed to germinate after fourteen days, the *M. oleifera* Burkina Faso accession, MO70 showed germination by the eleventh day. *M peregrinna* (Oman accession) germinated only after the removal of the seed coat and subsequent culturing.

The seeds were placed on MSO plates and stored in a refrigerator at 4℃ for a duration of three days. They were then moved to a chamber designed for plant development and kept 20℃ +/-2℃. The photoperiod was set for 14 hours (8:00–22:00 BST) and the light intensity was set at 120 µmol(m-2s-1PPFD) with Valoya L28 (NS12 Spectrum) LED lamps.

Usually, the relative humidity in the surroundings was 60% (±10%). Hypogeal germination (Cirlini *et al.,* 2022) is the phenomena wherein Moringa seeds' cotyledons remain beneath the soil surface during germination. The epicotyl had started to form on the seventh day, generating the first genuine leaves (***Fig****.2.0*).

Figure 2.0 Emerging cotyledon of *Moringa oleifera* at day 7

2.2.3 Experimental Design

Twelve *Moringa oleifera* seedlings were prepared in identical conditions and divided into 2 groups of 6 biological replicates each: control plants and 5-day drought-treated plants. 100mg of young leaves and stem tissues were taken quickly from the tip of the plants within a duration of 15-minutes to prevent RNA degradation at 1 PM (ZT8) from all plants at day 16 and at day 21, with the drought plants receiving no water during this time. These samples were snap-frozen in liquid nitrogen ready for RNA extraction. This experiment was aimed at identifying the changes in gene expression of Moringa plants under drought stress. RNA samples were stored at -80℃ before being sent for sequencing at Novogene. A total of 24 RNA samples were employed for the DESeq2 analysis (Barratt *et al.,* 2023). The choice of 1pm (ZT8) for the sampling was intended to capture the transcriptomic activity of drought regulation at the peak of daytime temperature and leaf physiological activity (Fig. 2.1).

RNA extraction was performed using the Omega EZNA Plant RNA Kit Protocol, following the manufacturer's instructions meticulously. The RNA concentration in ng/µl for each sample was measured with the Nanodrop ND-1000 Spectrophotometer, and the quality of the RNA was evaluated by analysing the samples on an Agilent 2100 Bioanalyser RNA 6000 Nano Chip. All samples demonstrated an average RNA Integrity score (RIN) quality value of 9.0 (LabRulez LCMS, 2024).

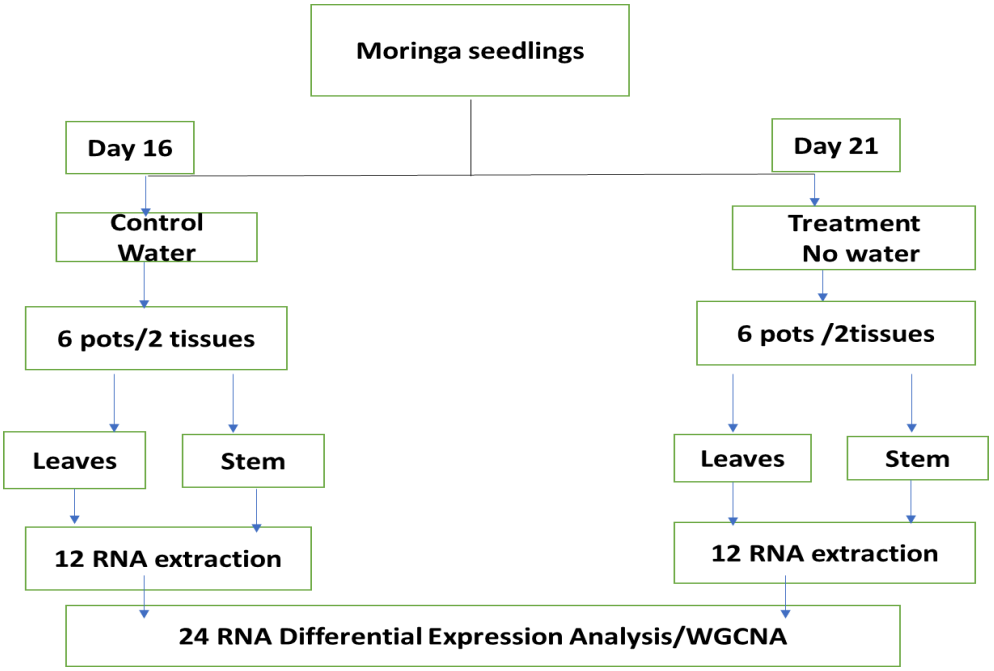


Figure 2.1 Experimental design of the RNA-Seq drought study in *Moringa oleifera*

Several steps were taken to maintain the integrity of my RNA samples and avoid RNA degradation by RNases prior to sequencing. This includes prompt collection of tissues while wearing protective gloves, utilizing sterilized equipment such as scissors thoroughly cleaned with RNAseZAP, rapidly freeze-drying the tissues in liquid nitrogen shortly after cutting, and conducting the extraction process as soon as possible after collection using a reliable RNA extraction kit and protocol that eliminates buffer residues after washing (Omega EZNA RNA kit). Every extraction session started with a thorough cleaning of the workstation with RNAseZAP and was consistently treated thereafter to maintain a sterile environment. RNA samples were promptly stored in a -80℃ freezer following the extraction of 5μL aliquots to minimize freeze-thaw cycles. The aliquot samples underwent 1% agarose gel electrophoresis for visualization of rRNA bands and were subsequently evaluated for quality using the Nanodrop RNA 6000 Nano assay. Additionally, RNA concentration and RIN quality measurements were conducted through capillary electrophoresis on the Agilent Bioanalyser 2100.

Weight of fresh biomass at the onset of drought experiment and on the 5th day of drought were taken:

Weight (fresh biomass before drought) = pot + soil + plant + water

Weight (fresh biomass after drought) = pot + soil + plant

2.2.4 Library Preparation and Sequencing

Library preparation and sequencing were conducted in accordance with the RNA Sequencing protocol established by Novogene. This process involves segmenting the sequences into 150 base pair lengths and subsequently attaching specific adapter sequences to the ends of the chosen fragments. Messenger RNA was isolated from total RNA utilizing poly-T oligo-attached magnetic beads. The library was constructed via end repair, A-tailing, adaptor ligation, size selection, amplification, and purification. The library was assessed via Qubit, quantified using real-time PCR, and its size distribution analysed with a bioanalyzer. The quantified libraries were pooled and sequenced on the Illumina paired-end Novaseq 6000 platform, according to effective library concentration and data volume (Novogene, 2022). A total of 122.9GB data of paired end reads measuring 150 base pairs was received, the minimum data for each library being 1,463,257KB. A thorough quality pre-sequencing assessment performed on all 24 samples utilizing 1% Agarose gel electrophoresis, revealed clear and distinct bands. The RNA Integrity Number (RIN) values exceeded 9 in all the samples.

2.3 Data Processing

The pre-processing steps comprised quality screening and trimming. Raw RNASeq reads underwent trimming including adapter removal utilizing TruSeq3-PE-2.fa

(Palomares *et al.,* 2019), and removal of low - quality sequences by trimming with Trimmomatic v0.39 (Bolger *et al.,* 2014) to improve alignment accuracy. The trimmed reads were aligned to the reference genome MoringaV2 (Chang *et al.,* 2022) using Salmon (Patro *et al.,* 2017). Salmon employs a quasi-mapping method to align raw reads in fastq format with a reference genome in Fasta format. The selection was based on its capacity to rapidly estimate read counts and abundance via a mapping-based approach that generates an index from the reference. Output files from Salmon, that is the normalized Count data including abundance (TPM), counts, and length, were produced for differential expression analysis utilizing the R (v4.1.2; R Core Team, 2021) package TxImport v1.24 (Soneson *et al.,* 2016).The count data underwent variance stabilizing transformations (VST) to yield converted data on the $\log_2$scale, thereby eliminating the variance-mean dependency (Love *et al.,* 2014). Transcript abundance (TPM) was utilized for WGCNA.

2.3.1 Statistical Analysis

The study employed a number of statistical tools to ensure the validity and reliability of the data collected.

2.3.1.1 Differential gene expression analysis

Differential expression analysis on *M. oleifera* plants using DESeq2 helped to identify genes that were either greatly up-regulated or down-regulated in respect to drought stress. This helped to identify important biological processes, cellular activity, and mechanisms either influenced by or necessary for life in arid environments or otherwise.

RNA was extracted from twenty-four samples in all, six duplicates of leaf and stem tissues of both drought-affected and control Moringa plants, then high throughput sequencing generated raw sequenced reads. By matching reads to a reference genome with an alignment tool like Salmon, they were precisely mapped to their corresponding genomic or transcriptomic sites. The normalized Count data were subject to statistical analysis utilizing a count table as the input data which underwent differential expression analysis DESeq2 R package as described by Love *et al*., (2014). The DESeq2 vignette filters out genes showing a count of 10 or less in the samples. This cutoff guarantees that the expression levels of the investigated genes satisfy a certain threshold to provide reliable statistical estimations. Low expression level genes were deleted from the dataset, therefore keeping only those with counts of 10 or above ($\geq$ 10).

The design had two group types, control and drought and two tissue types, leaf and stem. The Benjamini and Hochberg, False Discovery Rate (FDR) was applied to the data for multiple testing correction (Benjamini & Hochberg 1995) using the DESeq2 approach with significance threshold of $p < 0.05$ to identify genes with significant differences in

expression between treatment groups and tissue types, the count data underwent standardization using a variance-stabilizing transformation. The Tximport tool was used to import the transformed counts. Size factors and dispersions were determined with the DESeq tool before the model was fit to the data. For certain comparisons, drought against control groups and different tissue types among other things the filtering process produced pertinent genes. These DESeq2 standards guided gene selection: a p-value less than 0.05, an FDR adjusted value below 0.05, and a log2FoldChange value greater than 1.5 or less than -1.5 (Stephens, 2016). Ashr v2.2-63 was used to decrease the fold changes (LFCs) of genes showing notable dispersal, therefore improving interpretive accuracy and lowering false positives.

Different visualisation techniques were applied using Enhanced Volcano to provide a better knowledge of the interrelationships among the populations, including principal component analysis (PCA) (Blighe *et al.,* 2018). The variance stabilizing transformation (VST) was applied to the count data. Volcano graphs were used to visualize the significant genes based on their log-fold change and adjusted p-values. The variance in the most differentially expressed genes among the Moringa accessions was displayed using a heatmap generated in DESeq2. The expression levels of the genes across the leaf and stem tissues under drought stress were evaluated using ggplot2 (R Core team, 2014).

2.3.1.2 Individual Gene analysis

The outcomes of the differential expression analyses, which were filtered for significant up-regulation or down-regulation using log_2fold change thresholds of +/- 1.5 and adjusted p-values of <0.05, were aligned with their corresponding annotations from the GO file to gather pertinent biological information. The notable annotated genes were submitted to the agriGO Singular Enrichment Analysis tool v2.0 (Du *et al*., 2010) for classification into four experimental design groups: Control vs Drought (CD), Stem vs Leaves (SL), Stem only (SD), and Leaf only (LD) under drought conditions. The expression levels of the two most prominent individual genes, which showed up-regulation and down-regulation across the four groups, were plotted using R to evaluate their expression under various conditions and to identify the most significant genes in each group by sorting the LFC column from highest to lowest in the data spreadsheet.

2.3.2 DEG gene ontology term enrichment analysis

To uncover biological pathways and functions associated with the observed expression changes, the relevant genes were subjected to Gene Ontology (GO) enrichment analysis using R, as described by Tomczak *et al*. (2018). Several R packages, including data_table, clusterProfiler, and GOfuncR, were used for the analysis. The MoringaV2.GO

EC.annot file (MoringaGO.txt) from Chang *et al*. (2022) is an annotation file that contains genes that have been annotated and their matching GO IDs. A DataFrame called desired_background_dataset was used to store the data. A combination of the GO annotations and the findings of the differential expression analyses were used to extract relevant biological information. The results were filtered for substantially increased or decreased expression using log_2fold change and adjusted p-value criteria. After that, the genes that were deemed significant were stored in text files so they could be analysed further. One method for this was to submit them to the agriGO Singular Enrichment Analysis v2.0 (Tian *et al.,*2017). A p-value threshold of less than 0.05, corrected for false discovery rate, was used to determine whether the genes were significantly enriched.

The GOfuncR package's get_names function was used to acquire GO term descriptions. In order to provide a complete dataset with gene annotations, this step was carried out independently. An extensive list was generated for analysis by extracting unique gene IDs from the drought and control groups, in addition to several tissue types from the background dataset. An annotated copy of the GO study, a DataFrame was established (go_gene), and to make enrichment analysis easier, relevant columns were chosen. The clusterProfiler package's (Yu *et al*., 2012) enricher function was used to conduct the GO enrichment study. False discovery rates and set significance parameters were accounted for at a p value threshold of 0.05.

2.3.3 Network Construction and Module detection

The analysis employed WGCNA v1.72-1 (Langfelder, 2008) and FlashClust Packages (Tian *et al.,* 2017) in the R environment (R Core Team, 2019). The input included the abundance table (TPM data) from Salmon or variance stabilizing transformation (vst) from DESeq2, derived from stem and leaf tissue samples collected prior to and following drought stress treatment. Initially, genes with zero counts were excluded, resulting in a total of 22,706 genes from 24 samples, which included stem and leaf tissue from 6 samples collected before and after drought stress. The samples were clustered (Maechler *et al.,* 2025) to identify outliers, leading to the formation of a sample tree. The pickSoftThreshold() function was used to determine the optimal soft-thresholding power for construction of the network. The power is important to building a network with scale-free topology, meaning that a few nodes are highly connected, and others only have a few connections thereby allowing me to identify the hub genes then blockwiseModules() automatically constructs the gene co-expression network and detects modules (clusters of highly co-expressed genes). The network was established, and the adjacency matrix and topological overlap matrix (TOM) were generated (Alvarezbuylla, *et al.,* 2006; Yip *et al*.,2006). Thresholds for TOMtype were established,

defining a minimum module size of 30 and a merge cut height of 0.25, resulting in the merging of modules with a distance below 0.25. The network was subsequently imported for visualization in Cytoscape v3.10.2 through the exportNetworkToCytosape() function (Shannon *et al.,* 2003).

The assessment of gene similarity was conducted through the calculation of TOM from the derived adjacency matrix, which relies on direct connections and shared neighbours. This provides a more reliable assessment of co-expression. TOM values span from 0 to 1, where elevated values signify increased similarity and connectivity between genes. The TOM was utilized for hierarchical clustering and is then converted into a dissimilarity measure (1-TOM) to enhance the clustering process.

2.3.4 Selection of modules of interest

GO terms of genes in each module were matched to GO terms of all genes in the co-expression network. Z test was used to determine whether the observed proportion of DEGs in each module was significantly greater than the predicted proportion. Modules were considered as significantly enriched in DEGs if the p value was less than 0.05 ($p<0.05$).

2.3.5 Hub Gene Identification

Hub genes were identified by calculating the degree scores for each gene within a module, which involved measuring the connections to and from each gene in the WGCNA edge file, utilizing the table() function in R. A list containing the edges which represent gene relationship and node attributes was collated and imported into Cytoscape which was employed for the module's visualization.

The genes in a module exhibiting the highest degree scores were identified as central hubs, with the most interconnected differentially expressed gene recognised as the hub gene within the networks. The R script utilized for calculating degree scores draws on the contributions of Isaac Reynolds and Harper Lab, as well as Barratt *et al*. (2023), especially for large modules that present challenges in loading, viewing, and analysing within Cytoscape. The DEG with the highest degree score was identified as the hub gene of the module due to its differential expression and strong connectivity within the module, thereby enhancing its potential role in regulating the transcriptional drought response compared to a well-connected non-DEG.

Cytohubba (Chin *et al*., 2014), a plugin for Cytoscape, was utilized to identify hub genes within the network through various algorithms that highlight significant connectivity. Cytohubba evaluated various centrality metrics, including degree centrality, betweenness, and closeness centrality. The top 10% of genes within each module, identified through degree centrality, are regarded as crucial regulators of drought

response in this investigation. The integration of these hub genes with GO enrichment analysis confirmed their biological relevance. The connections between modules and traits were evaluated through the application of Pearson correlation coefficients. The heatmap illustrated correlation values that span from -1 (blue) to 1 (red), with p-values indicated in parentheses. Relationships that are statistically significant ($p<0.05$) were represented with more vibrant hubs.

The identification of hub genes required the analysis of module-trait relationships to associate specific traits with co-expression modules (Farber, *et al.,*2009). Identifying hub genes is crucial for determining central regulators within the network that may play a significant role in the response to drought. Initially, the module eigengene was calculated for each co-expression module. The eigengene served as the main principal component that encapsulates the collective expression profile of the genes in the module. Principal component analysis helped to identify main expression patterns in Moringa tissues, therefore clarifying the link between many gene modules and drought response. R was utilized to perform this analysis. The heatmap uses colour gradients to emphasise modules strongly linked with drought, therefore enabling a quick evaluation of the overall patterns in module-trait correlations. The correlation coefficients and p-values for each module-trait combination were determined by associating module eigengenes with specific characteristics. The values indicated the statistical significance and strength of the correlations. Thresholds for correlation coefficients of ≥ 0.05 and p values of <0.05 were established to identify significant connections. This work sought to find regulatory modules connected to the drought response of the plant by looking at modules displaying notable correlations with relevant features.

By focussing on genes showing the highest connectivity within those modules, prioritising modules with strong linkages allowed the identification of hub genes. I identified the top 10% as a threshold for intra-modular connections and used connectivity measures to assess module membership by means of the examination of the link between gene expression and module eigengene, therefore determining connectivity.

Principal regulators were identified hub genes exhibiting the highest connectivity, possessing the potential to influence the expression of other genes within the network. The identified hub genes were confirmed by their co-expression with known stress responsive genes and their consistency of expression under several situations. This approach helped to identify candidates for master regulators of the transcriptional response to drought.

Several statistical techniques are available to evaluate hub genes' biological relevance, significance, and resilience. The correlation coefficient helps one to measure the

correlations between module eigengenes and traits on the module-trait heatmap. When their p-values are less than or equal to 0.05, correlations are judged statistically significant; their correlation coefficient is expressed by values between -0.05 and +0.05. False discovery rate (FDR) change helps to solve the multiple testing problem.

Analysing their biological activities and their participation in known regulatory circuits after discovery helped one to investigate the roles of hub genes. The module-trait connection analysis revealed hub genes with notable regulating potential in the drought responsive network.

2.4 Results

2.4.1 Drought stress effect

The growth of Moringa plants was significantly affected by drought when comparing control and stressed conditions (Fig.2.2). In the context of drought stress, several observable symptoms include leaf withering from dehydration, yellowing, abscission, and wilting of stem and leaf tips. The symptoms suggest that the plants are losing enough water and nutrients to drought stress causes higher transpiration and lower water uptake. Whereas the phenotypic changes were most noticeable in the drought treated group, the control plants had better leaves and a more solid overall structure. Leaf withering and chlorosis were measured using a standardised visual assessment of damaged plants. A score of 0 denoted the lack of symptoms; a score of five indicated extreme wilting or total discolouration. This approach enabled a consistent assessment of the drought stress degree among all samples.

Figure 2.2 Physical changes observed due to drought-stress “5-days No water treatment”

2.4.2 Transcriptome sequencing and reads Pre-processing

To investigate the transcriptional responses of many Moringa tissues under drought circumstances, I performed a transcriptome study. This work looked at changes in gene

expression following drought. Following sequencing, quality control of the raw data was performed using FastQC (Wai and Martti, 2023). The quality control assessment of this study verified the high quality of the sequencing data, producing a total output of 123.03GB. The average percentages of bases attaining q values of ≥20 and ≥30 were 97.60% and 93.52%, respectively, reflecting minimal sequencing error rates (0.03). The GC concentration of the reads varied between 44.13% and 45.97%. Transcriptome data were processed for quantification and mapping using Salmon v1.10.0 (Patro *et al.,* 2017), a tool recognised for its efficiency and accuracy in assessing transcript abundance. The overall mapping rate for all samples was 64%, indicating the presence of valuable data, as a substantial proportion of the sequenced reads aligned with the reference genome.

2.4.3 Identification of drought - responsive genes

A differential expression analysis was performed on 24 samples of Moringa leaves and stems to investigate the genes associated with drought stress, collected before and after drought treatment. The analysis utilized a $\log_2$fold change threshold greater than +1.5 or less than -1.5, in conjunction with an FDR-adjusted p-value of under 0.05 to determine significance. After excluding lowly expressed genes (count <10), 9,249 significant differentially expressed genes (DEGs) were identified from a total of 22,706 genes. Table 2.1 summarizes the distribution of these 9,249 significant genes below.

The Principal Component Analysis (PCA) Fig.2.3A, revealed how the samples group by treatment, Control vs Drought and Stem vs Leaf tissues under drought stress. There is a clear separation of the Leaf samples: control and drought along the PC1 except for 3D and particularly 5D leaf_drought which is separated from the Leaf drought cluster, these outliers might indicate distinct expression profile requiring further investigation. The stem samples clustered distinctly into control group and drought group towards PC2 as expected.

Among the differentially expressed genes (DEGs), 2,139 were categorised within the Control vs Drought significant group (CD_sig group), which includes 1,052 genes with increased expression (positive $\log_2$fold change greater than 1.5) and 1,087 genes with decreased expression (negative $\log_2$fold change less than -1.5) in response to drought. Figure 2.3B shows the heatmap of responses of Moringa tissues under drought vs control. The distribution of significant genes was widely dispersed across the dataset as seen in the Principal Component Analysis in Figure 2.3A, indicating a notable variance in gene expression between control and drought conditions, as well as among other tissues. Further investigation revealed gene expression patterns that are specific to tissue types in reaction to drought stress. A total of 2,935 significant genes were

identified in both stem and leaf tissues under drought stress (Stem and Leaf significant group, SL_sig), with 1,695 genes showing increased expression and 1,240 genes showing decreased expression. Additionally, 979 significant genes were exclusively identified in the stem tissues (Stem Droughted, SD_ sig), including 627 that were DE increased and 352 that were DE decreased in response to drought conditions. In contrast, 3,196 significant genes were identified exclusively in the leaf tissues (Leaf Droughted, LD_ sig), with 1,437 genes showing increased expression and 1,759 genes showing decreased expression under dry conditions (Figures 2.3 D, E, F). The results demonstrate distinct tissue-specific transcriptional responses to drought stress in Moringa, providing a foundation for further exploration of the molecular pathways involved. The identification of differentially expressed genes across tissues highlights the potential for targeted studies on the functional roles of these genes in drought adaptation.

Table 2.1 Summary of the 9,249 significant DEGs identified under droughted Moringa after removing lowly expressed genes from 22,706 genes identified.

Group of Treatments	Annotated Genes	Increased Annotated Genes	Decreased Annotated Genes	Significant DEGs	Increased DEGs	Decreased DEGs
Control vs Drought (CD)	6677	3092	3585	2139 (23.13%)	1052	1087
Stem and Leaf (SL)	9267	5086	4181	2935 (31.73%)	1695	1240
Stem Droughted (SD)	2913	1842	1071	979 (10.58%)	627	352
Leaf Droughted (LD)	9951	4367	5584	3196 (34.55%)	1437	1759

CD group is the comparison of all the DEGs of the Control (C) versus Drought (D).

SL group is the DEGs identified in the Stem (S) and the Leaves (L).

SD group is all the DEGS in the Stem (S) tissues under drought condition.

LD group is all the DEGS in the Leaf (L) tissues under drought condition.

Visualization of expression pattern of 5-day drought treatment on 16-day old *Moringa oleifera*.

A

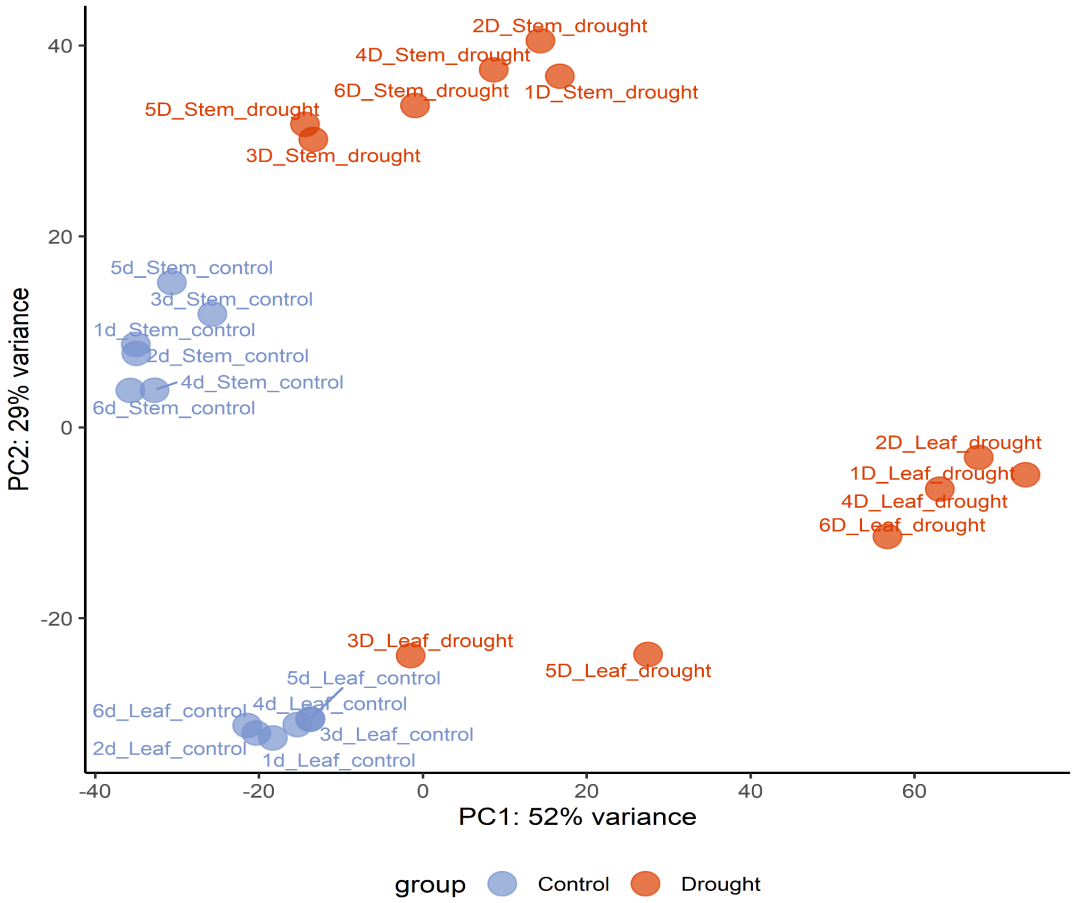
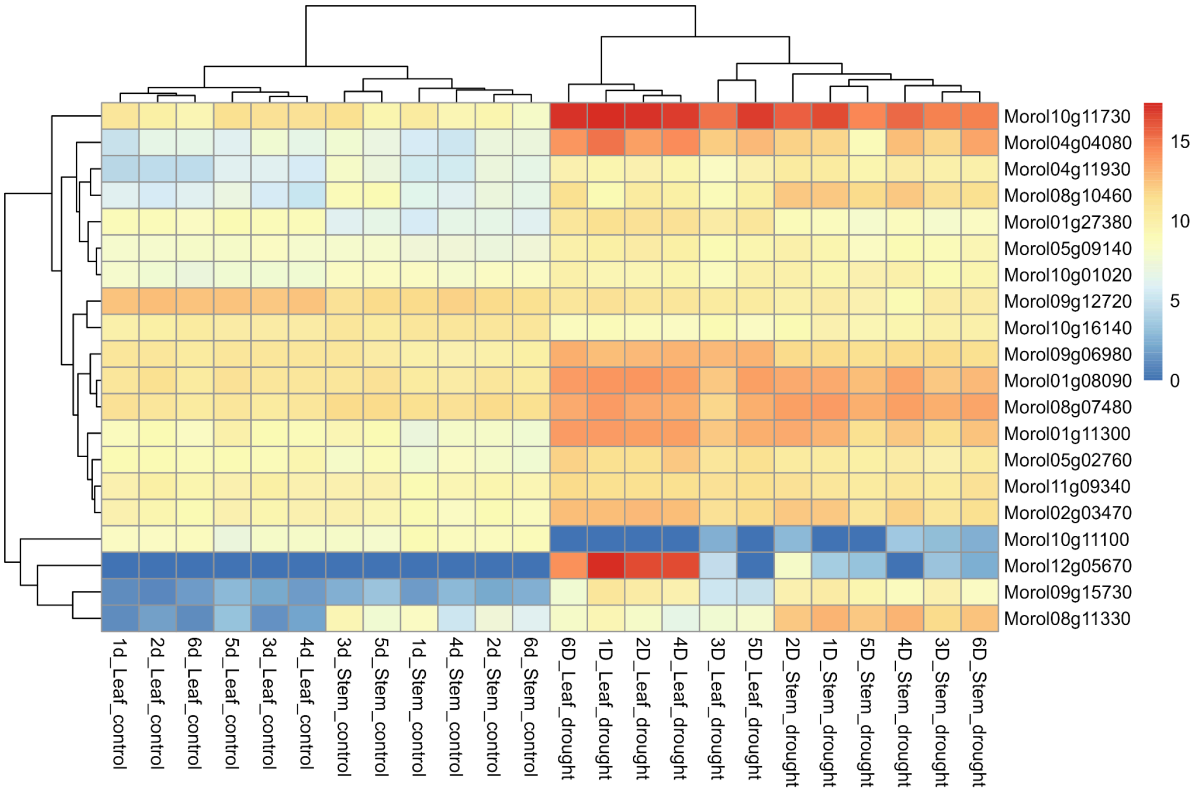


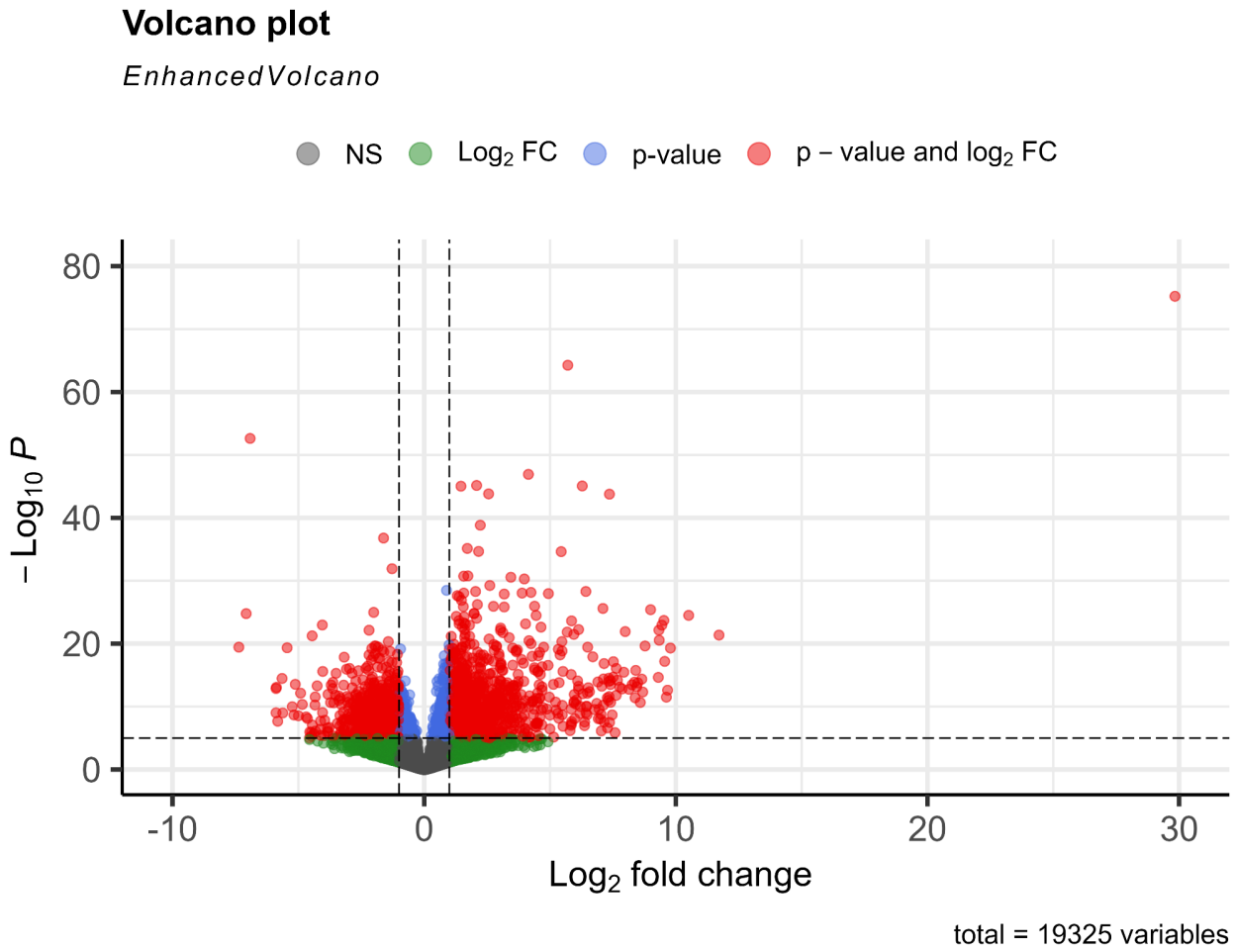
Figure 2.3 Five days drought treatment on 16-day old *Moringa oleifera* resulted in significant differences across the transcriptomic profiles.
A. Principal Component Analysis (PCA) of variance-stabilised transcript counts shows distinct clustering of 24 replicates of 6 samples based on treatment (control vs drought) and tissue type (stem vs leaf). This shows a clear separation of control and drought samples on PC1 and PC2 separates the stem and leaf samples. PC1 and PC2 accounted for 81% of the total variance. The control samples are labelled 1d, 2d, 3d, 4d, 5d, and 6d while the samples subjected to drought are labelled 1D, 2D, 3D, 4D, 5D and 6D. (R-DESeq2)

B. Heatmap



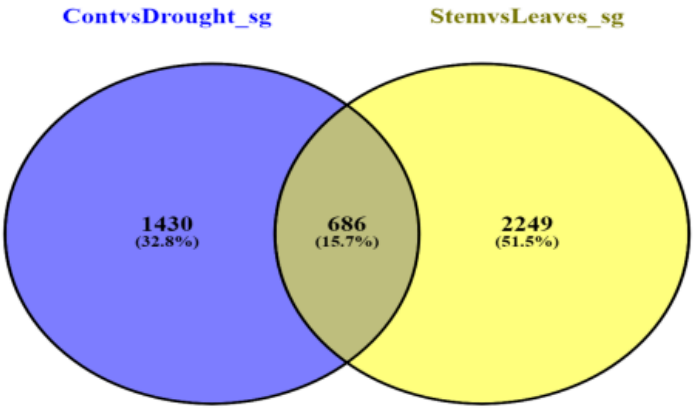
B. Heatmap of all samples based on drought versus control. This shows that the top 20 Moringa genes based on p-value show the strongest differentiation between the leaf drought samples versus other samples. This could suggest that the most strongly differentially expressed genes are present in the leaf under drought condition. (R-DESeq2)

C. Volcano Plot

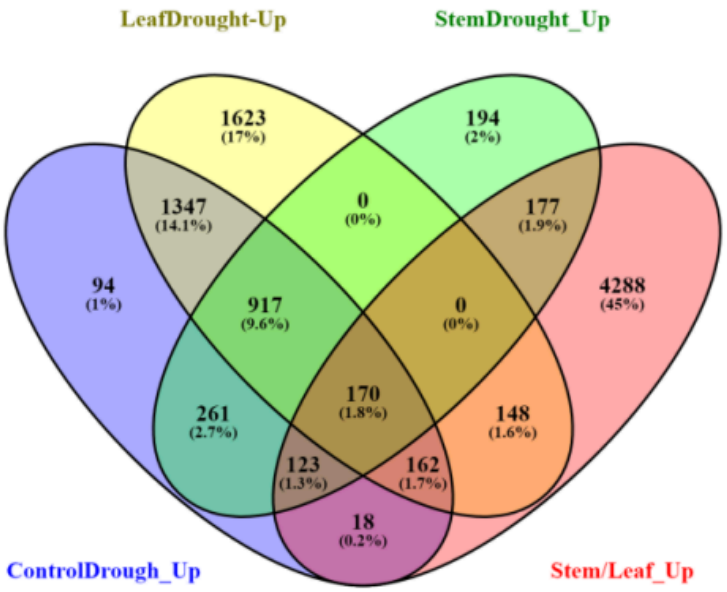


C. Volcano Plot of drought versus control in Moringa. This illustrates that more genes are increased than decreased. The largest fold changes, and most significant results are the increased genes. The extremely differentially expressed gene Morol12g05670 (GO) seems interesting and worth exploring.

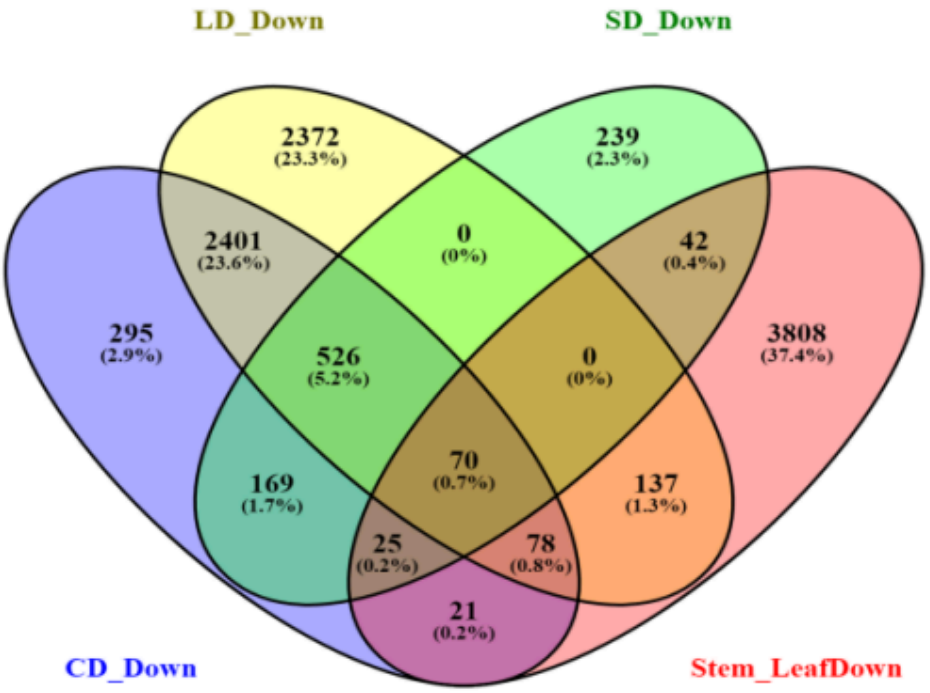
D. Venn Diagrams



D.Venn Diagram of drought vs control in Moringa (Venny 2.1.0, Oliveros,2007-2015)



E. Venn Diagram comparing 4 groups of data increased under drought condition: Leaf, Stem, Control vs Drought and Stem and Leaves.



F. Venn Diagram comparing 4 groups of data decreased under drought condition: Leaf, Stem, Control vs Drought and Stem and Leaves.

Analysis of gene expression variations based on a log_2fold change threshold greater than +/- 1.5 revealed that Moringa's stem tissues exhibited different responses to drought stress relative to its leaves, with a higher number of genes showing differential expression in the leaves compared to the stems. In drought conditions, the expression of 3,196 genes (34.55% of total significant DEGs, 9249) in leaf tissues was modified, with 1,437 genes demonstrating increased expression and 1,759 genes reflecting decreased expression. In contrast, the expression of 979 genes (10.58% of total significant DEGs, 9249) in the stem tissues, only 627genes exhibited an increase, whereas 352 genes showed a decrease in the stem tissues.

The analysis identified the ten most significantly differentially expressed genes, determined by the optimal p-value, as shown in the Heatmap (Fig.2.3B & Table 2.2). The descriptions of the GO terms were acquired using the get_names function from the GOfuncR package. A comprehensive search was performed to clarify the function of genes associated with drought response. The comprehensive analysis of differentially expressed genes alongside WGCNA provides a detailed view of the molecular pathways activated by drought stress. The ten most important genes (Table 2.2) were found to be associated with various processes connected to drought response, including senescence, metal ion binding, cellular protective mechanisms, and reactions to water and abiotic stress. The gene SAG12, linked to leaf senescence, was identified, while additional genes were associated with the binding of essential metal ions such as copper, iron, zinc, and magnesium.

Table 2.2: Top DEGs (Control vs Drought) and their functions (w.r.t Heatmap, lowest p-value, the most significant). *Modules associated with the drought response identified in section 2.4.4 (Module column)*

Gene ID	GO Term	Gene	Function	Log Fold Change LFC	p value	Padj value	Module (WGCNA)
Morol12g05670	GO:0005615	Senescence-associated gene 12 (SAG12) [*Hibiscus trionum*]	Leaf & floral organ senescence	29.83	3.19 E-80	6.13 E-76	Yellow
Morol10g11730	GO:0005507	XP_031375161.1 metallothionein-like protein type 3	Copper ion & Zinc ion binding	5.71	5.81 E-69	5.59 E-65	Yellow
Morol04g04080	GO:0009507	XP_021908809.1 ferriti n-3, chloroplastic-like	Ferrous & Ferric ion binding	6.28	2.67 E-49	8.55 E-46	Yellow
Morol04g11930	GO:0006950	XP_021282404.1 remor in 4.1-like	Response to stress & organic substance	3.44	2.59 E-34	2.77 E-31	Green
Morol08g10460	-	Uncharacterised/ Unannotated wt REF Genome	-	3.98	5.11 E-34	5.18 E-31	Cyan
Morol01g27380	GO:0005524	XP_012092323.1 receptor-like cytosolicserine/ threonine - protein kinase RBK2	Protein phosphorylation; ATP binding	2.08	1.84 E-49	7.09 E-46	Yellow
Morol10g01020	GO:0005634	XP_028246970.1 ubiquitin-conjugating enzyme E2-17 kDa-like	ATP binding, protein polyubiquitination	1.57	1.63 E-34	1.85 E-31	Green

Morol03g14060	GO:0016021	Gawky-like	integral component of membrane, cellularcomponent	20.88	4E-06	1.3E-4	Yellow
Morol01g09210	GO:0051603	Vignain	Proteolysis involved in cellular protein catabolic process.	14.36	9.93E-05	1.30E-2	Yellow
Morol01g26440	GO:0016021	Ferritin-like	integral component of membrane, cellular component	11.72	4E-06	1.3E-4	Yellow
Morol05g15350	GO:0000978	KNAT2	Regulation of transcription by RNA polymerasel l	11.74	1.084 E-3	6.103 E-2	Red
Morol12g01460	GO:0006357	SBH1	Regulation of transcription by RNA polymerasel l	11.22	1.084 E-3	6.103 E-2	Red
Morol07g03510	GO:00015051 ; GO:0007186	LEA	regulation of neurotrans mitter levels; G protein-cou pled receptor signalling pathway	-8.15	4.26E -14	7.30E-12	Turquoise
Morol01g12840	-	Uncharacterised/ Unannotated wt REF Genome	-	-7.74	1.379	9.08E-08	Turquoise
Morol02g21300	GO:0098869	PRX	Peroxidase activity	-7.36	1.29E -05	6.97E-4	Light-gree n
Morol10g11100	GO:00458937	bZIP	positive regulation of transcription , DNA-templa ted	-7.91	6.26E -27	2.82E-24	Turquoise
Morol02g21290	GO:0006979 GO:0098869	CCP2	response to oxidative stress, cellular oxidant detoxificatio n	-7.07	3.63E -28	1.66E-25	Blue
Morol05g04740	GO:0016705	CYP82D	oxidoreduct ase activity	-8.573	8.38E -26	3.49E-24	Yellow

Morol09g12720	GO:0009926	XP_022735113.1 auxin efflux carrier component 3-like	ATP binding, protein polyubiquitination	-1.6155	9.40 E-41	1.65 E-37	Blue
Morol10g16140	GO:0016887	XP_021902819.1 probable phospholipid transporting ATPase 4	Magnesium binding Ion ATP & binding	-1.2785	9.46 E-36	1.21 E-32	Turquoise
Morol09g06980	GO:0009881	AML76715.1 putative LOV domain-containing protein	Response to Blue light, protein chromophore linkage	1.7133	4.52 E-39	7.25 E-36	Yellow

2.4.3.1 Individual Gene analysis

The genes with the most significant log fold change (LFC) are listed in Table 2.2, indicating an increase in expression, whereas those with the least LFC were classified as showing a decrease in expression as discussed in Figure 2.4 to Figure 2.11).

Genes with significantly elevated expression in the Stem under Drought-stress.

Under drought conditions, certain genes showed heightened expression in both leaf and stem tissues but more in the stem, such as SAG12 Morol12g05670, which had a log_2fold change (LFC) of 32.58, and Protein Gawky-like Morol03g14060, displaying an LFC of 20.88 (**Figure 2.4)**. The results demonstrate that molecular pathways are triggered in stem tissues when faced with water scarcity. Analysis of the network indicated that these genes are located within the yellow module (**Figure 2.14**).

A. **B.**

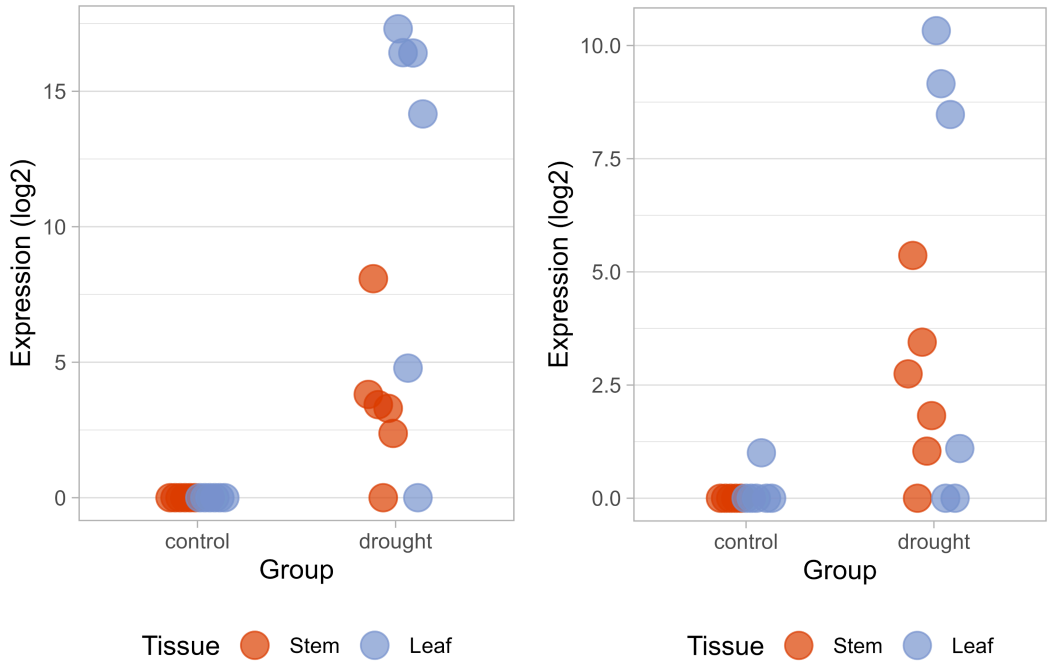


Figure 2.4 Genes with significantly elevated expression in the Stem under Drought-stress **A**) Morol12g05670 (LFC32.58) **B**) Morol03g14060 (LFC20.88)

Genes with significantly lowered expression in the Stem under Drought-stress

The genes exhibiting reduced expression in the stem under drought stress included Morol07g03510, with a $\log_2$ fold change (LFC) of -8.15, and Morol01g12840, which had an LFC of -7.74 (**Figure 2.5)**. Morol07g03510 is linked to the turquoise module within the co-expression network and encodes a late embryogenesis abundant (LEA) protein. This protein is situated in the nucleus and is involved in molecular processes associated with stress responses. Morol01g12840 is linked to the turquoise module **Figure 2.16**; however, it is uncharacterised in the reference genome and lacks functional annotation.

A. **B.**

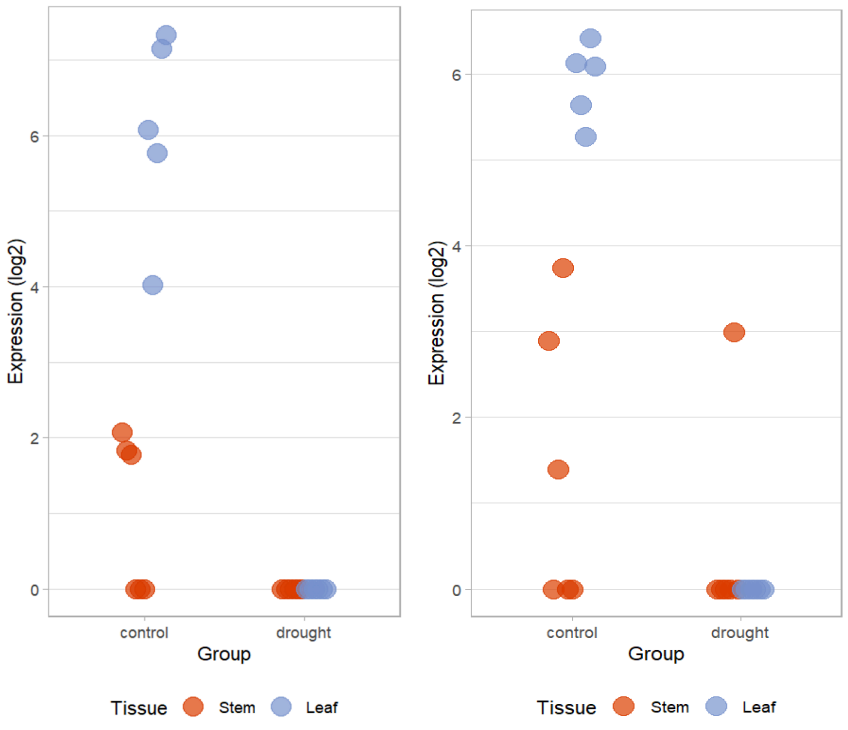


Figure 2.5 Genes with significantly lowered expression in the Stem under Drought-stress **A)** Morol07g03510 (LFC:-8.15) **B)** Morol01g12840(LFC:-7.74) Unannotated from REF Genome.

Genes with significantly elevated expression in the Leaf under Drought-stress.

Several genes exhibited increased expression in the leaf under drought stress, with prominent examples being Morol12g05670 (LFC 29.64) and Morol01g09210 (LFC 14.36) as seen in **Figure 2.6**. Morol12g05670 encodes the senescence-associated gene SAG12, a cysteine protease involved in processes related to leaf senescence. Morol01g09210 encodes the protein Vignain, which is associated with various cellular functions in plants. These 2 genes are associated with the Yellow module (**Figure 2.14)**

A. **B.**

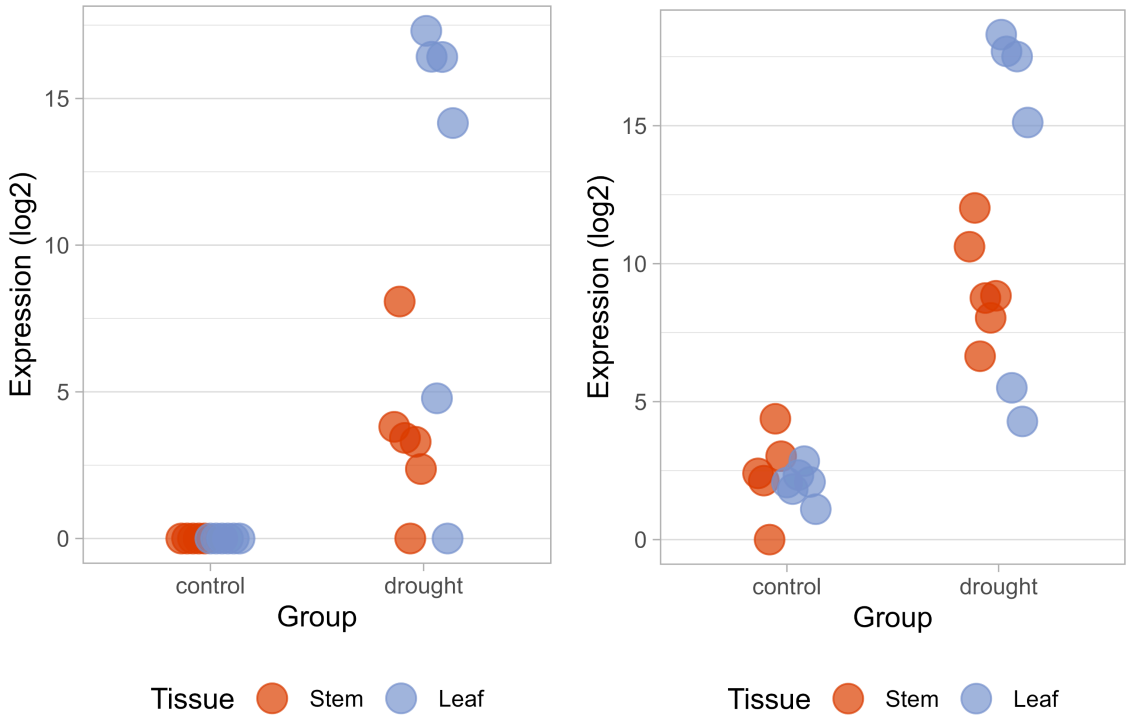


Figure 2.6 Genes with significantly elevated expression in the Leaf under Drought-stress. **A)** Morol12g05670 (LFC 29.64) SAG12; **B)** Morol01g09210 (LFC14.36) Vignain

Genes with significantly lowered expression in the Leaf under Drought-stress.

A considerable number of genes showed reduced expression in leaf tissue subjected to drought stress, notably Morol02g21300 with a $\log_2$fold change (LFC) of -9.46, and Morol10g11100 with an LFC of -7.91(**Figure 2.7)**. Morol02g21300 encodes a Peroxidase N1-like protein involved in oxidative stress responses and is linked to the Light-green module. Morol10g11100 encodes a basic Leucine zipper (bZIP) transcription factor implicated in gene regulation.

A. **B.**

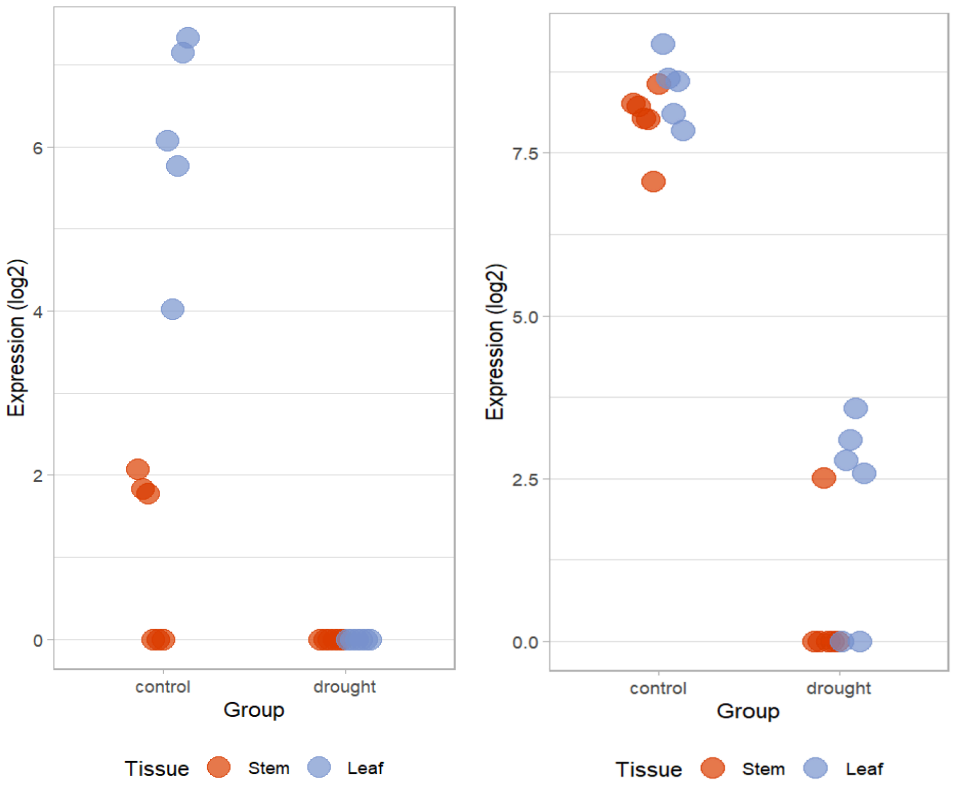


Figure 2.7 Genes with significantly lowered expression in the Leaf under Drought-stress. **A)** Morol02g21300 (LFC:-9.46) PeroxidaseN1-like; **B)** Morol10g11100 (LFC:-7.91) Basic Leucine zipper34

Genes with significantly elevated expression in all samples, Control vs Drought.

A notable proportion of genes exhibited heightened expression in all samples when contrasting control conditions with drought stress, especially Morol12g05670 and Morol01g26440 (**Figure 2.8)**. Morol12g05670, exhibiting a log_2fold change (LFC) of 29.83, encodes the senescence-associated gene SAG12, which is a cysteine protease that plays a role in leaf senescence. Morol01g26440 exhibits an LFC of 11.715 and encodes a protein featuring a ferritin-like domain, linked to iron storage and metabolism. Both genes are components of the yellow co-expression module network (**Figure 2.14).**

A. **B.**

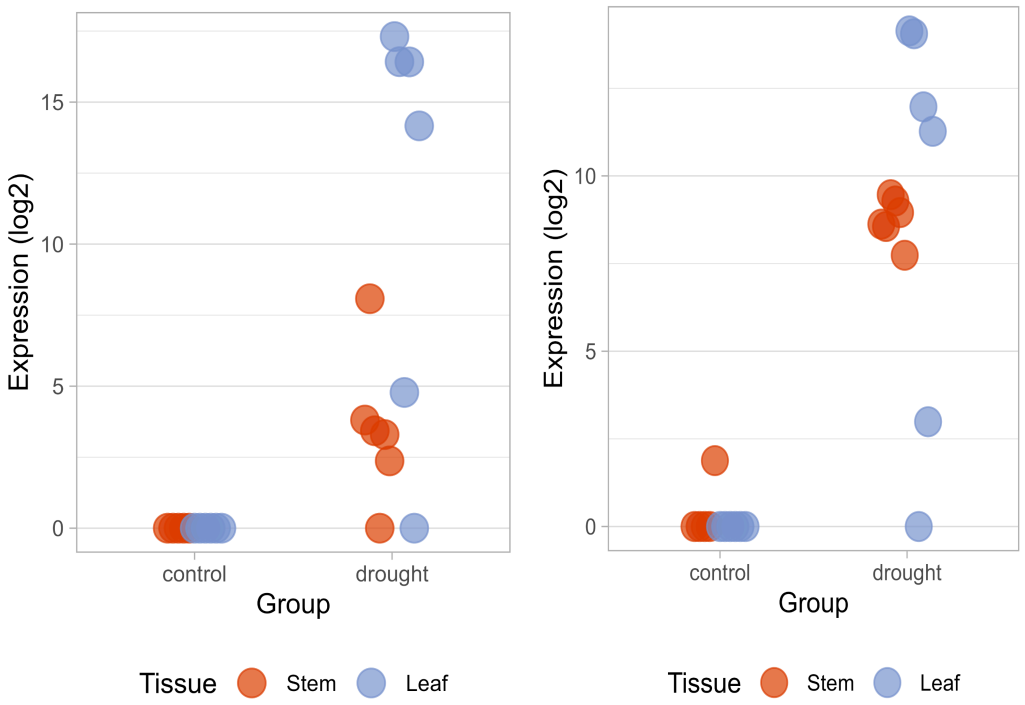


Figure 2.8 Genes with significantly elevated expression in both plant tissues, Control vs Drought. **A)** Morol12g05670 (LFC:29.83) SAG12. **B)** Morol01g26440 (LFC:11.715) Ferritin-like domain containing protein

Genes with significantly lowered expression in all tissues, Control vs Drought

When comparing control conditions to drought stress, the expression of several genes, including Morol02g21300 and Morol02g21290, decreased across the leaf and stem tissues (**Figure 2.9)**. Morol02g21300 encodes a Peroxidase N1-like protein (PRX) that is present in the plant-type cell wall, with a log_2fold change (LFC) of -7.36. The Light-Green module is associated with Morol02g21300. The CCP2 protein in this study is associated with the blue module co-expression network **Figure 2.17** and is encoded by Morol02g21290, which has an LFC of -7.07.

A. **B.**

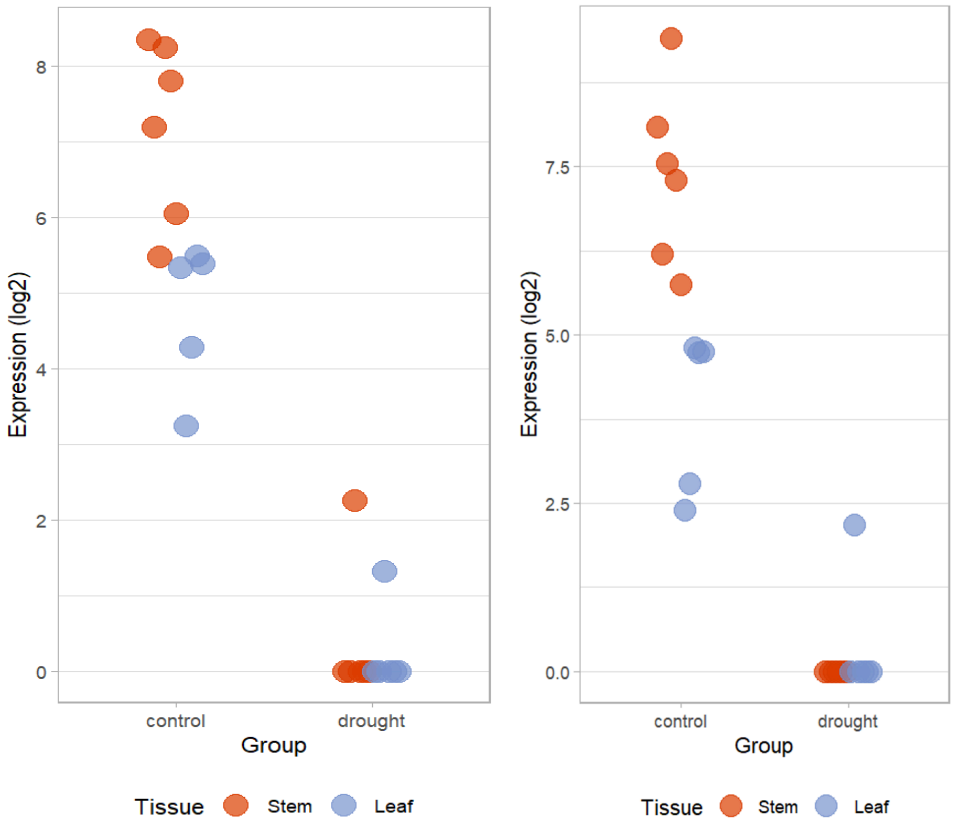


Figure 2.9 Genes with significantly lowered expression in all tissues, Control vs Drought. **A)** Morol02g21290 (LFC:-7.07) CCP2 **B)** Morol02g21300 Peroxidase N1-like protein (PRX).

Genes with significantly elevated expression in Stem & Leaf during Drought.

Several genes showed increased expression in both stem and leaf tissues under drought conditions, particularly Morol05g15350 and Morol12g01460. Both genes Morol05g15350 and Morol12g01460 were elevated in the stem of Moringa during drought (**Figure 2.10)**. Morol05g15350 exhibited a log_2fold change (LFC) of 11.736 and encodes the Homeobox protein knotted-1-like 2 (KNAT2), which plays a role in various regulatory functions in plant development and is situated in the red module. Morol12g01460 has an LFC of 11.215 and encodes the Homeobox protein SBH1, a gene associated with transcriptional control and a component of the red module.

A. **B.**

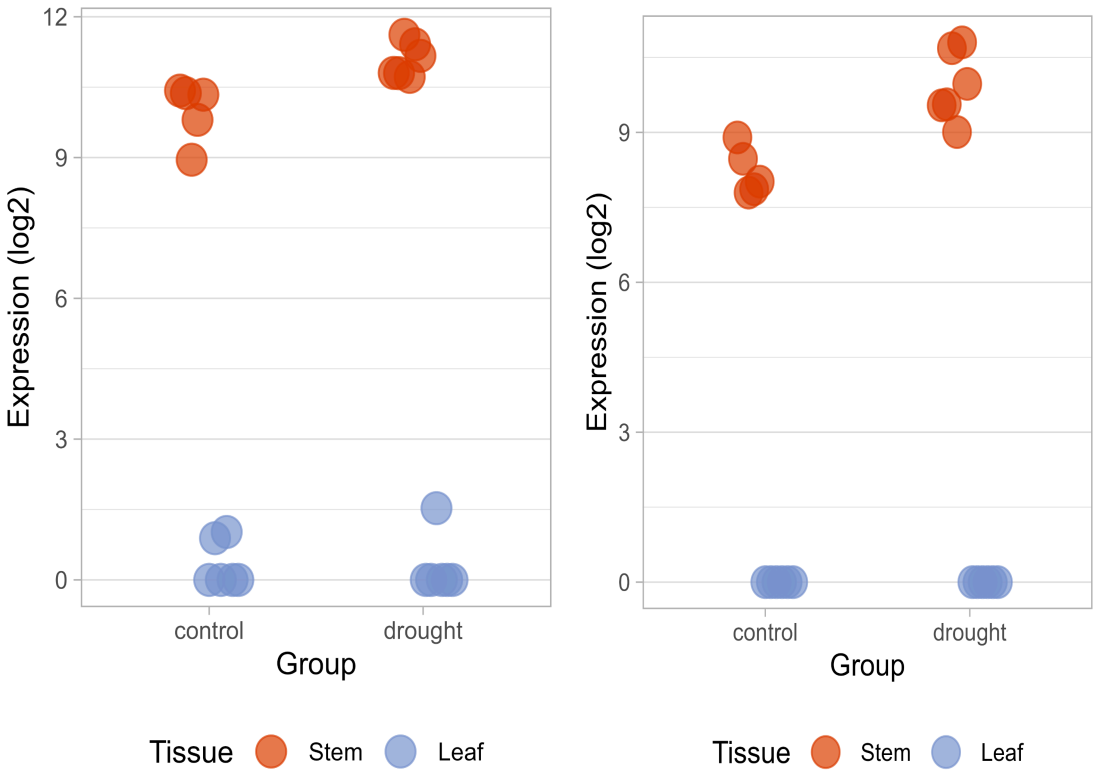


Figure 2.10 Genes with significantly elevated expression in Stem & Leaf during Drought.
A) Morol05g15350(LFC:11.736) KNAT2 **B)** Morol12g01460 (LFC:11.215) SBH1
Genes with significantly lowered expression in Stem & Leaf during Drought.
Differential expression analysis was conducted to compare gene expression levels in drought-stressed samples with those in well-watered controls across both tissues. Under drought conditions, the expression of multiple genes decreased in both stem and leaf tissues. Morol05g04740 and Morol12g05670 (LFC-8.466) SAG12 represent notable examples as seen in (**Figure 2.11)**.

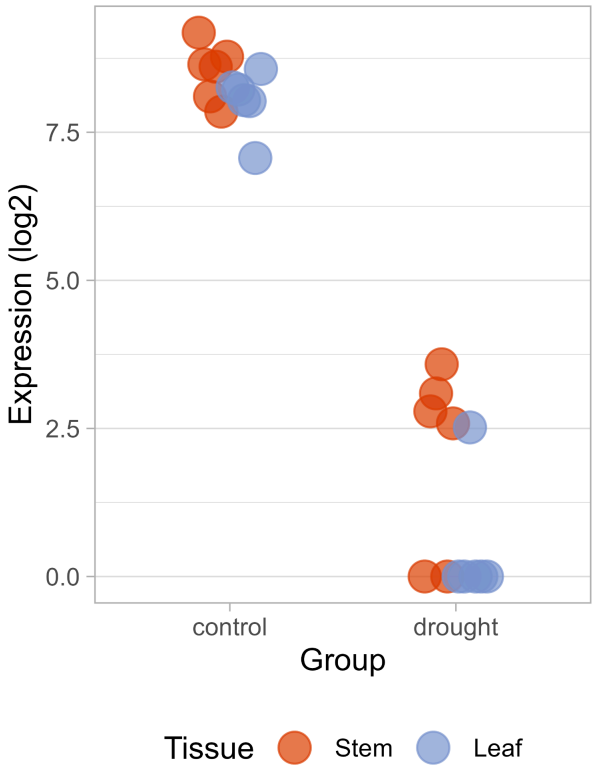


Figure 2.11 Genes with significantly lowered expression in Stem & Leaf during Drought. Morol05g04740(LFC-8.573) CYP82D

2.4.3.2 Functional analysis of drought -responsive genes

Gene Ontology (GO) enrichment analysis was employed to explore the functions of differentially expressed genes (DEGs) categorized into increased and decreased groups across several comparisons: Control vs Drought (CD), Leaf vs Drought (LD), Stem vs Drought (SD), and Stem & Leaf vs Drought (SL). The GOfuncR and clusterProfiler packages, along with data table, were utilized for the analysis of Gene Ontology term enrichment. The study focused on identifying common biological pathways among the differentially expressed genes, utilizing a false discovery rate threshold to maintain significance. The CD group showed 47 enriched GO terms for increased genes and 167 for decreased genes, while the LD group exhibited 73 enriched GO terms for increased genes and 148 for decreased genes. The SL group discovered 168 enriched GO terms associated with DE increased genes and 120 linked to DE decreased genes. The SD group identified with 68 enriched GOterms associated with DE increased genes and 42 linked to DE decreased genes (Table 2.3).

The analysis involved the use of Gene Ontology (GO) enrichment to classify these genes according to common biological processes, molecular functions, and cellular

components, utilizing the agriGO custom tool for Singular Enrichment Analysis. Key terms associated with Gene Ontology (GO) concerning drought stress response showed significant enrichment in the LD, CD, and SD increased gene groups. Key terms included “water transport” (GO:0006833, FDR= 0.036), "response to chemical" (GO:0042221, FDR = 0.041), “response to water channel activity” (GO:0015250, FDR = 0.013), and "response to oxidative stress" (GO:0006979, FDR = 0.0068). Furthermore, terms such as "response to reactive oxygen species (ROS) metabolic process" (GO:0072593, FDR = 0.02) and those associated with "transmembrane transport" (GO:0055085, FDR = 0.014; GO:0055085, 6.6e-06) were identified *(Supplementary Tables.).* The activity of transcription factors (GO:0003700, FDR = 0.032) showed significant enrichment in the increased genes, highlighting its role in regulating gene expression under drought stress conditions.

Table 2.3 Summary of significantly elevated and lowered gene expression (DEGs) and Enriched GO terms observed across the leaf and stem tissues under drought condition.

Group of Treatments	Significant DEGs	Increased DEGs	Enriched GO terms/ increased DEGs	Decreased DEGs	Enriched GO terms/ Decreased DEGs
Control vs Drought significant group(CD_sig)	2139	1052	47	1087	167
Stem and Leaf significant group SL_ sig	2935	1695	168	1240	120
Stem Droughted SD_ sig	979	627	68	352	42
Leaf Droughted LD_ sig	3196	1437	73	1759	148

The enriched GO terms observed in the drought-stressed Moringa plants indicate the numerous strategies these plants employ to adapt to drought conditions such as increasing the expression of water transport and channel genes, which allows maximum water uptake and retention; initiation of genes responding to oxidative stress to prevent damage caused by Reactive Oxygen species (ROS), thereby helping to maintain cellular function during drought; ability to sense and respond to chemical signals and adjusting physiological processes in real-time, and initiating gene networks required for drought resilience, allowing for a coordinated response to varying levels of water availability.
The GO terms that were significantly enriched among the genes that decreased in the SL group were related to photosynthetic processes. The identified terms included "photosynthesis" (GO:0015979, FDR = 8.7e-64; GO:0019684, FDR = 2.2e-35;

GO:0009765, FDR = 1.1e-17), "chloroplast" (GO:0009658, FDR = 0.00042), and "photosystem" (GO:0048564, FDR = 0.001). Enrichment was observed for terms associated with "electron transporter activity" (GO:0045156, FDR = 0.013) and "thylakoid" structures (GO:0009579, FDR = 7.2e-58; GO:0044436, FDR = 1.1e-57). The results establish the fact that drought conditions hinder photosynthesis.

A network analysis was performed to cross-reference the Gene Ontology enrichment results with supplementary datasets, aiming to integrate the GO data with additional network studies. This method seeks to enhance understanding of gene function in drought conditions and to validate the findings. The prevalence of these GO keywords among drought-responsive genes highlights the intricate mechanisms of drought adaptation in plants. Plants are able to combat drought stress through improved water transport, regulation of chemical responses, management of oxidative stress, and utilization of transcription factors.

2.4.4 WGCNA

A co-expression network analysis was conducted to identify gene clusters associated with drought response. The study identified a total of 46 modules comprising 9,249 co-expressed genes (DEGs). The module size varied from 37 to 5,370 genes, with an average size of 446 and a median size of 115. The selection of co-expression network analysis was based on its capacity to identify gene clusters with coordinated expression patterns, which may indicate functional relationships.

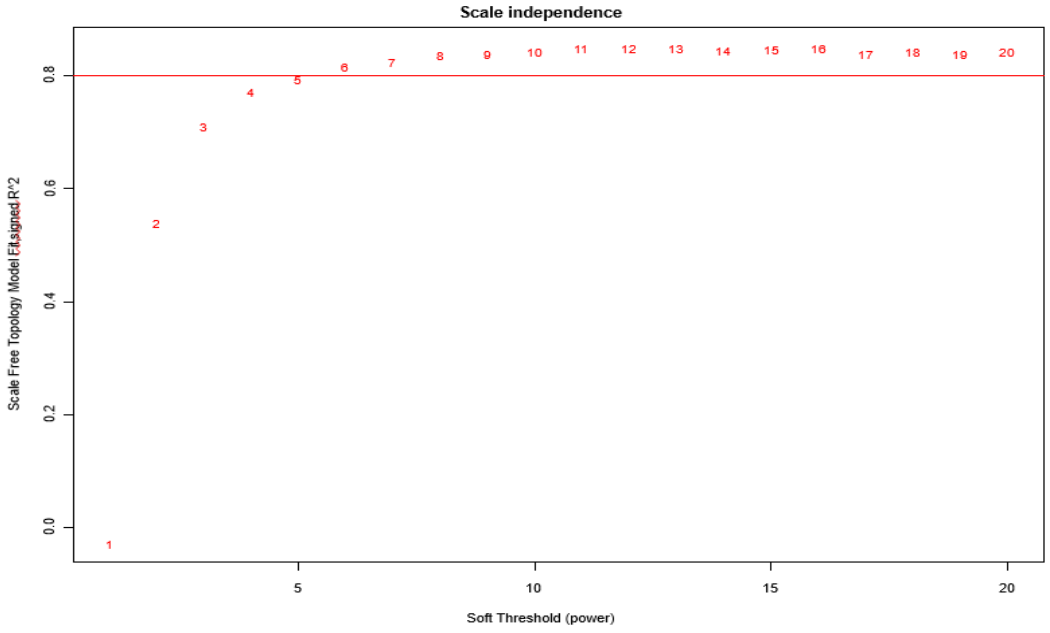


Figure 2.12 Scale Independence. The ‘x’axis is the Soft Threshold (power) while the ‘y’axis is the "Scale Free Topology Model Fit,signed R^2"

A pairwise correlation matrix was computed using Pearson to assess the relationships between gene expression profiles. This was then transformed into an adjacency matrix using a power transformation (soft thresholding) to enhance strong correlations. I selected a suitable soft-thresholding (β) power of 12 (Supp.Table2.2), as it is the closest power to the cutoff value of 0.9 and has a low enough mean.k value (between 100 and 10) making it appropriate. It is also the saturation point of the R^2 value, with no other power creating a more scale-free topology (**Fig. 2.12)**.

Gene Ontology enrichment analysis was performed on each module to determine associations with drought response, utilizing all genes in the network as the background reference. Thresholds for false discovery rates were utilized. Gene Ontology (GO) enrichment was utilized to identify common biological processes or molecular activities among genes within each module. To identify modules linked to drought response terms, I examined GO terms enriched in "response to stress" (GO: 0006950), "response to water" (GO: 0006833), “response to heat” (GO:0009408), “response to water deprivation” (GO:0009414),“ cellular response to water deprivation” (GO:0042631; GO:0042631) and "response to oxidative stress" (GO: 0034599). The findings revealed that 1 out of 46 modules showed significant enrichment in “stress”-related Gene Ontology categories, indicating a possible involvement in drought adaptation. Only the green-yellow module demonstrated significant enrichment for "response to stress," with (FDR) value of (3.12E-03).

2.4.5 Module-Trait Relationship

The significance of module-trait connections was determined with a threshold of $p<0.05$. Modules showing absolute correlation values of 0.05 or higher and p-values below 0.05 were considered significantly correlated with the respective attribute, indicating that their gene expression profiles were meaningfully linked to the experimental conditions. A variety of modules demonstrated strong and significant correlations with tissue type and drought treatment as seen in Fig. 2.13, highlighting their potential relevance in understanding Moringa's molecular response to drought.

The MEred module exhibited a strong positive correlation with tissue type ($r = 0.88$, $p<0.001$), suggesting that its genes were expressed at elevated levels in stem tissue. The MEgreen module showed notable correlations with tissue type ($r = 0.50$, $p < 0.001$) and drought treatment ($r = 0.66$, $p < 0.001$), indicating its possible involvement in drought response, particularly in stem tissue. The MEcyan module exhibited significant correlations with tissue type ($r = 0.47$, $p < 0.001$) and treatment ($r = 0.61$, $p < 0.001$), suggesting that its genes could play a role in mediating drought responses in the stem.

Conversely, the MElightcyan module demonstrated a strong negative correlation with tissue type ($r = -0.96$, $p<0.001$), indicating that its genes were predominantly expressed in leaf tissue. The MEbrown module exhibited a notable negative correlation with treatment ($r = -0.82$, $p<0.001$), suggesting a possible downregulation under drought conditions. The MEblue module showed inverse correlations with tissue type ($r = -0.73$, $p < 0.001$) and treatment ($r = -0.49$, $p < 0.001$), suggesting heightened expression in leaf tissue.

The key modules identified namely MEred, MElightcyan, MEgreen, MEcyan, MEbrown, and MEblue were analysed in greater detail to identify hub genes that may clarify the molecular pathways associated with drought response in Moringa. Two main criteria were employed to pinpoint these hub genes through WGCNA: i) Intra-modular connectivity (kWithin), where genes demonstrating the highest within-module connectivity values (evaluated in Cytoscape) were ranked as the most interconnected nodes within their respective modules, with the top 10 genes identified as potential hub genes; and ii) module membership (kME), where genes with kME values nearing 1 were considered the most central and significant members of the module.

Following the establishment of the network, CytoHubba, a plugin for Cytoscape, was utilized to pinpoint hub genes through various techniques, such as Degree, Betweenness, and Closeness centrality. These approaches assess genes based on their interconnections and significance within the network, highlighting essential regulatory nodes. The data were exported for further analysis, including functional enrichment studies on the identified hub genes to explore their biological significance and roles in various pathways. This thorough approach enhances the understanding of gene interactions and facilitates the identification of crucial genes linked to specific phenotypes.

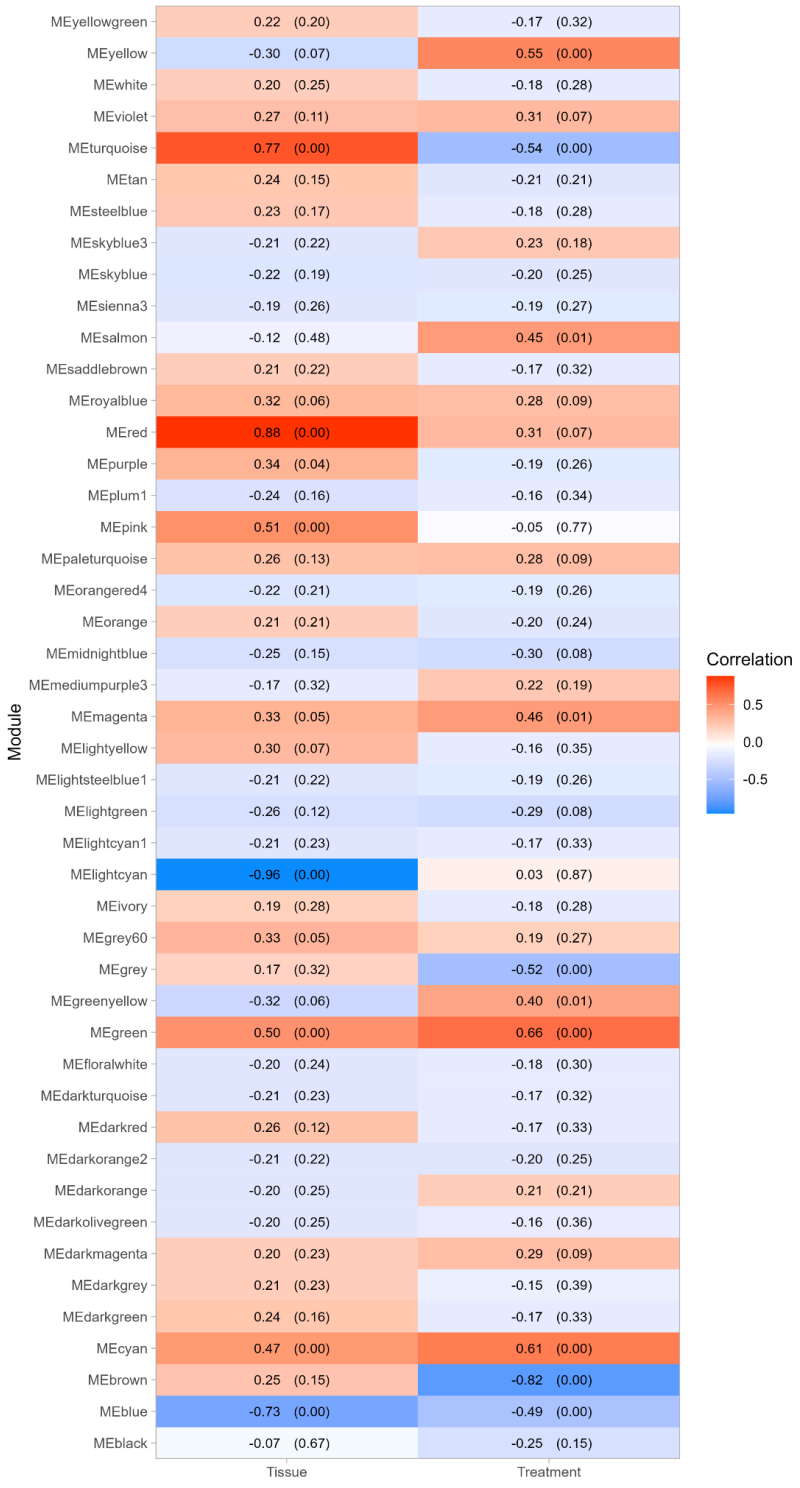


Figure 2.13 Module -Trait Relationship.

The heatmap of module trait associations for the various modules (labelled Module Eigengene (ME) followed by a colour) related to drought conditions in Moringa leaves and stem tissues.

Module detection

Out of 46 modules detected, the turquoise module contains most of the genes (5,370) associated with drought response but not significantly the most differentially expressed

(Figure 2.14). Only the green-yellow module demonstrated significant enrichment for "response to stress," with (FDR) value of (3.12E-03). The GO may not be a good way to decide the modules, therefore the DEG enrichment was more appropriate and was applied. From the analysis of differentially expressed genes (DEGs) performed to pinpoint the gene modules linked to the response to drought and to identify modules with a higher-than-expected proportion of differentially expressed genes (DEGs), the result revealed potential involvement in drought-related processes.

A total of 9,249 genes from the co-expression network were identified as differentially expressed genes. 7 out of the 46 modules exhibited a significantly higher proportion of differentially expressed genes (DEGs) than expected (Table 2.4), suggesting that these modules may be crucial in drought-responsive mechanisms. The modules exhibiting a greater proportion of DEGs than expected were considered to represent groups of co-expressed genes potentially impacting the drought response. The modules were subsequently highlighted for further analysis due to their potential roles in mediating transcriptional changes under drought conditions.

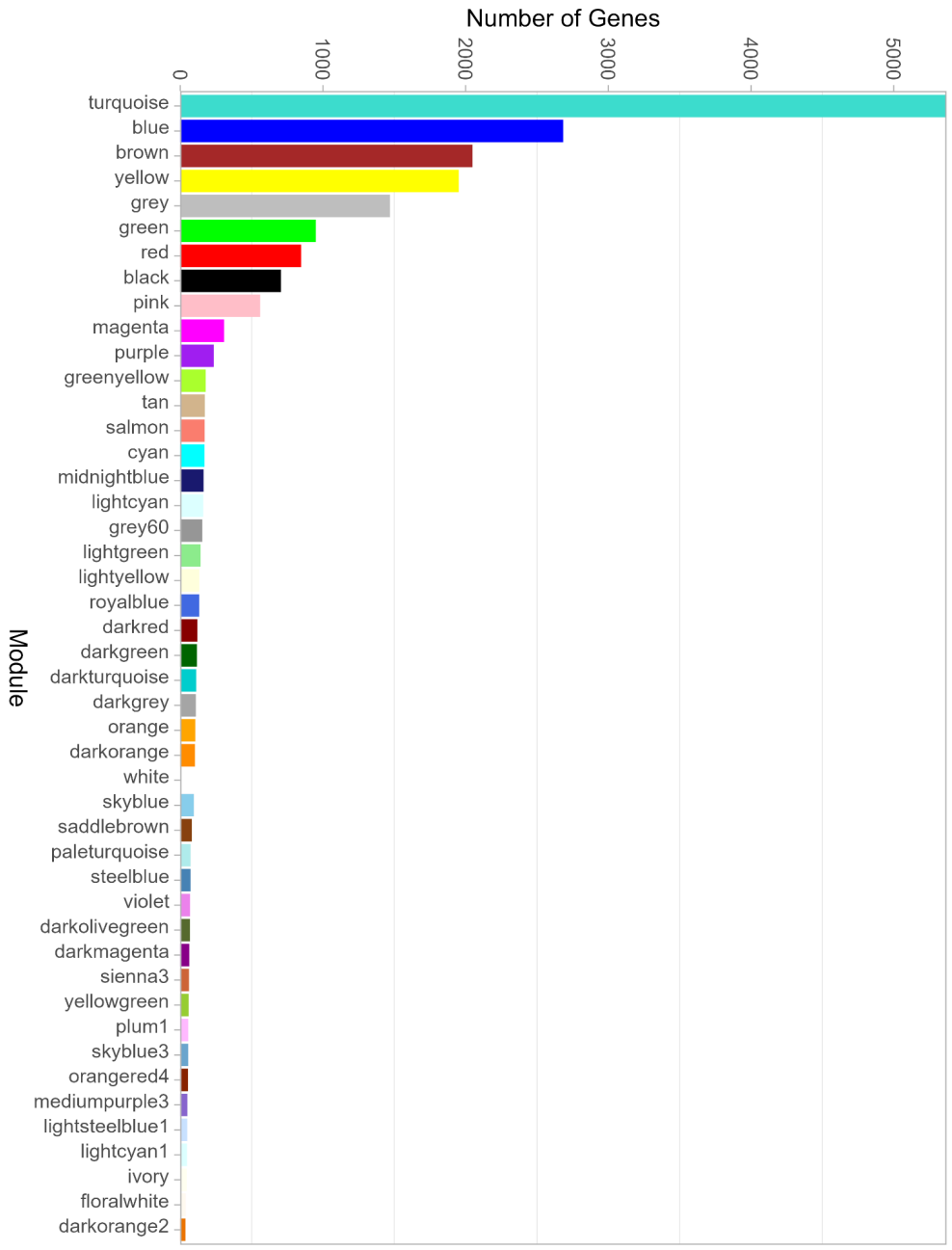


Figure 2.14 Module Summary of the network obtained from module membership.

The hub genes identified within the modules were regarded as potential key regulators of the transcriptional response to drought. The identification of these hub genes was predicted based on their significant connectivity that is higher degrees of connectivity within the co-expression network, depicting their essential function in modulating the expression of other genes in response to drought conditions.

The integration of DEG enrichment results with stress-related GO term enrichment facilitated the identification of modules specifically involved in drought-related biological processes. The analysis included Gene Ontology (GO) term enrichment, detailing all

thresholds applied to ascertain significant enrichment via the agriGO custom tool SEA. Simultaneous analyses of differentially expressed genes (DEG) and Gene Ontology (GO) word enrichment were conducted to validate findings and emphasize biologically significant modules. A validation of hub gene activity under drought conditions has not been conducted to date. The research identified specific gene modules as critical components in the transcriptional response to drought, with hub genes serving as potential master regulators.

Table 2.4: Over-representation of DEGs in individual modules.

Module	Module Size	Stem Leaf Drought DEGs	Observed % of DEGs	p-Value	Leaf Drought DEGs	Observed % of DEGs	p-Value	Stem Drought DEGs	Observed % of DEGs	p-Value
Yellow	1952	558	29	1.61 E-62	871	47	6.19 E-168	181	9	5.44 E-06
Cyan	170	98	58	1.85 E-54	53	31	9.16 E-05	117	68	8.29 E-228
Magenta	307	122	40	8.59 E-34	83	27	7.10 E-19	118	38	2.31 E-108
Salmon	171	78	46	3.73 E-29	77	45	5.12 E-17	52	30	4.36 E-35
Green	950	237	25	1.12 E-17	341	36	3.9 E-36	104	11	9.20E115
Green yellow	178	26	15	0.5724 816	44	25	0.0480 404	19	11	0.01910 284
Red	847	89	11	0.9999	128	15	0.9999	99	12	6.13 E-09

2.4.5 Hub genes

The identified prospective hub genes in 5 selected modules based on the highest degree scores (Table 2.5) including Morol02g05650 (turquoise), Morol01g14900 (cyan), Morol08g04710 (yellow), Morol06g06410 (green), and Morol01g26020 (blue) deemed to be associated with drought stress (Table 2.5) and may act as master-regulators of the transcriptional drought response as they are significantly co-expressed with many stress associated and /or drought-responsive genes. The rankings were established using the degree technique. To examine and interpret WGCNA networks using Cytoscape and CytoHubba for the identification of hub genes, the co-expression network data derived from WGCNA, which includes an edge list that outlines gene interactions and node

characteristics, such as module assignments was imported into Cytoscape, employing various layout methods to visually organize the network. The nodes were customized based on different characteristics, such as module membership or connections, which aided in understanding the network topology.

These hub genes likely play roles in stress related processes based on GO, such as response to water transport (Turquoise), response to heat (Cyan), response to water deprivation (Green, Yellow and Blue), Cellular response to water deprivation (Green, Yellow and Blue) and response to stress (Green). On the other hand, the coding sequences of the prospective hub genes in the drought-responsive network modules such as Morol02g05650 (turquoise), Morol01g14900 (cyan), Morol08g04710 (yellow), Morol06g06410 (green), and Morol01g26020 (blue) were queried using BLASTx (Altschul *et al.*, 1990) against the NCBI non-redundant protein sequence database and UniProt/Uniparc (Ahmad *et al*.,2025).

To provide functional insights to the Morol02g05650 (turquoise module), the coding sequence underwent BLASTx and UniParc searches against the NCBI database. The best hit was LOC110809083 from *Carica papaya*, which was annotated as dentin sialophosphoprotein-like isoform X2 (XP_021890487.1). The match had an e-value of 3×10^{-44} and a sequence identity of 68.15%. Domain analysis also showed similarity to the TIGR00927 family (Conserved Protein Domain family-2A1904 K^+-dependent Na^+/Ca^+ exchanger), which is involved in the transport and binding of cations and iron-carrying compounds. The findings suggest that Morol02g05650 may be involved in ion transport and homeostasis during drought stress conditions.

For Morol01g14900 (cyan module), the best hit was LOC110810283 from *Carica papaya,* annotated as probable polygalacturonase (XP_021892085.1). The match had an e-value of 0.0 and a sequence identity of 76.82%. Domain analysis also showed similarity to the cl26318 family (Conserved Protein Domain family-Glyco_hydro_28, polygalacturonase EC:3.2.1.15 as well as rhamnogalacturonase A (RGase A), EC:3.2.1), which is involved in cell wall metabolism. These findings infer Morol01g14900 may be involved in cell wall metabolism during drought stress condition.

The coding sequence of Morol08g04710 (yellow module) from the BLASTx and UniParc searches against the NCBI non-redundant protein database. The best hit was to an unidentified protein product from *Citrullus colocynthis* (CAK9329279.1), which had an e-value of 0.0 and a sequence identity of 88.29%. Domain analysis found PLN02287, which is the same as 3-ketoacyl-CoA thiolase.

Likewise, I used BLASTx to look up the coding sequence of Morol06g06410 (Green module). The best hit was for *Quercus lobata* putative plastidic glucose transporter 2

(XP_030952457.1), which had an e-value of 0.0 and an 83.16% sequence identity. The conserved domain analysis found cd17315 (MFS_GLUT_like), a group of glucose and other sugar transporters in the Major Facilitator Superfamily (MFS). The domain e-value was 6.29×10^{-115}.

Also, BLASTx was used to look at the coding sequence of Morol01g26020 (Blue module). The best hit was to the *Carica papaya* gene LOC110821448, which was identified as blue-light photoreceptor PHR2 (XP_021906987.1) and had 81.36% sequence identity. Functional annotation showed that DNA and FAD binding activity was linked to photoreactive repair. This was backed up by the presence of the FAD_binding_7 domain, which is a feature of DNA photolyases.

Table 2.5 Top genes based on degree scores analysed using Cytohubba/Cytoscape.

Turquoise	Degree Score	Cyan	Degree Score	Yellow	Degree Score	Green	Degree Score	Blue	Degree Score
Morol02g05650	1006	Morol01g14900	89	Morol08g04710	577	Morol06g06410	111	Morol01g26020	873
Morol12g09880	957	Morol10g09360	82	Morol08g03720	529	Morol03g07040	86	Morol11g00520	729
Morol08g14580	918	Morol08g06370	75	Morol09g07020	528	Morol12g09410	86	Morol01g23760	695
Morol09g15300	898	Morol05g06560	68	Morol06g09810	513	Morol01g10190	84	Morol01g10520	668
Morol01g02630	891	Morol07g04620	66	Morol06g13690	499	Morol08g10580	84	Morol03g14800	626
Morol05g00210	864	Morol13g03760	66	Morol05g14450	486	Morol03g20510	81	Morol11g05390	618
Morol09g06150	822	Morol09g03450	54	Morol01g31180	483	Morol03g05290	80	Morol06g06610	616
Morol04g04870	817	Morol12g08750	49	Morol05g06280	467	Morol07g02390	77	Morol03g07440	607
Morol02g16150	804	Morol08g10450	46	Morol11g04130	457	Morol09g08210	72	Morol01g09120	582
Morol06g03060	760	Morol01g04830	45	Morol10g15650	439	Morol10g11350	70	Morol06g15850	542
Morol11g04080	749	Morol04g02570	45	Morol01g30200	435	Morol09g04250	69	Morol10g08900	506

2.4.5.1 Yellow module network

The top 10 genes identified in the yellow co-expression network **Figure 2.15** using the Degree method for hub gene ranking shows the genes by their rank, name, and connectivity score, which indicates their importance within the network. The highest-ranked gene is Morol08g04710, with a score of 577, followed closely by Morol08g03720 at 529 and Morol09g07020 at 528. The remaining genes, also

demonstrate high connectivity scores, highlighting their potential significance in the network (Table 2.5). This ranking provides valuable insight into which genes may serve as key regulatory nodes in the biological processes under investigation. The top-ranked genes, such as Morol08g04710 (GO:0003988; GO:0005777) encodes the enzyme ketoacyl-CoA C-acyltransferase, peroxisomal is likely to a play central role in regulating biological processes.

Figure 2.15 Yellow Module Network
Summary of statistics of the yellow network analysis using Cytoscape/Cytohubba: Number of nodes: 752; Number of edges: 24190 Average number of neighbours: 64.335; Network diameter: 6; Network radius:3; Characteristic path length: 2.123; Clustering coefficient:0.845; Network density: 0.086; Network heterogeneity: 1.435; Network centralization:0.684; Connected components: 1; Analysis time (sec): 0.788**.**

2.4.5.2 Cyan module Network

The genes exhibiting the highest degree ratings in the Cyan module **Figure 2.16** are associated with and likely regulate other genes in response to arid conditions. The group includes the Polygalacturonase (Morol01g14900), Cytochrome P450 (Morol05g06560, GO: 0005506), and Probable glutathione S-transferase (Morol10g09360).

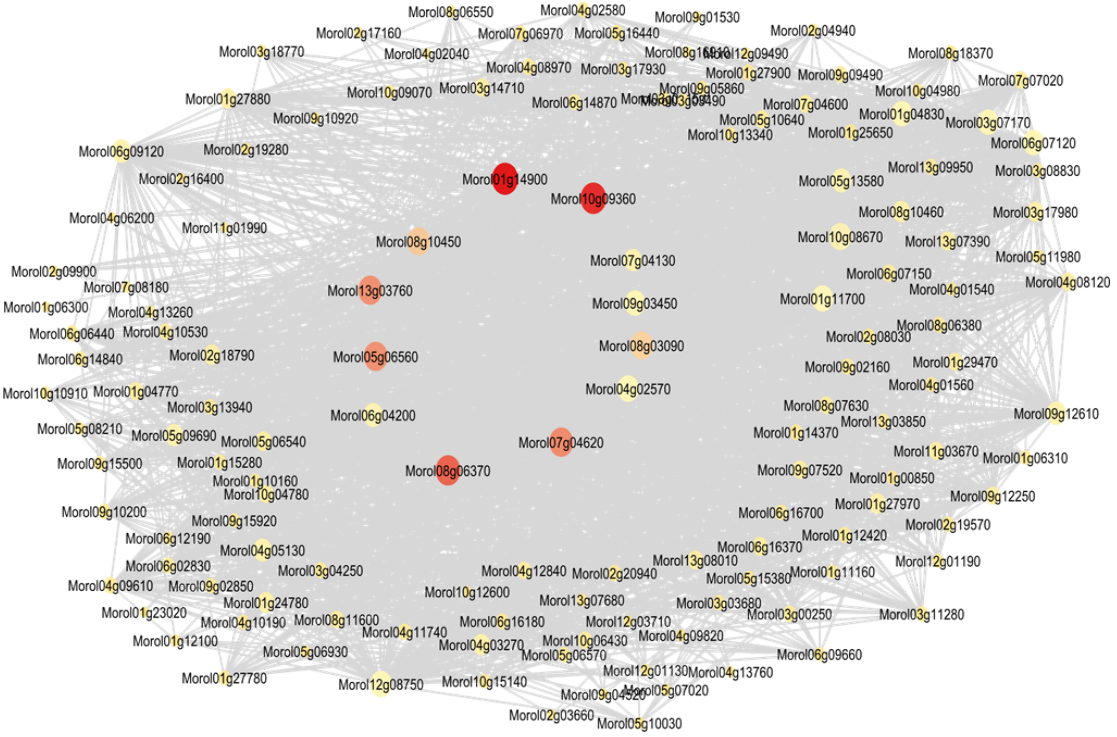


Figure 2.16 Cyan Module Network**.**

Summary of statistics of the cyan network analysis using Cytoscape/Cytohubba: Number of nodes: 141; Number of edges: 2632; Average number of neighbours: 37.333; Network diameter: 4; Network radius: 2 ; Characteristic path length: 1.773; Clustering coefficient:0.787; Network density: 0.267; Network heterogeneity: 0.768; Network centralization:0.628; Connected components: 1; Analysis time (sec): 0.204**.**

2.4.5.3 Turquoise module network

The Turquoise module network consists of numerous genes that affect the drought stress response in Moringa through their impact on gene expression, cell wall remodelling, and structural adaptations. The turquoise module network **Figure 2.17** comprises Morol02g05650 (MYB-like transcription factor), *Morol12g09880* (coatomer subunit delta), Morol08g14580 (xyloglucan glycosyltransferase 6), Morol02g16150 (coatomer subunit gamma), and Morol05g00210 (galactan β-1,4-galactosyltransferase GALS1). They interact in various ways to influence drought responses in Moringa plants, enhancing their capacity to withstand water-deficit conditions and stress.

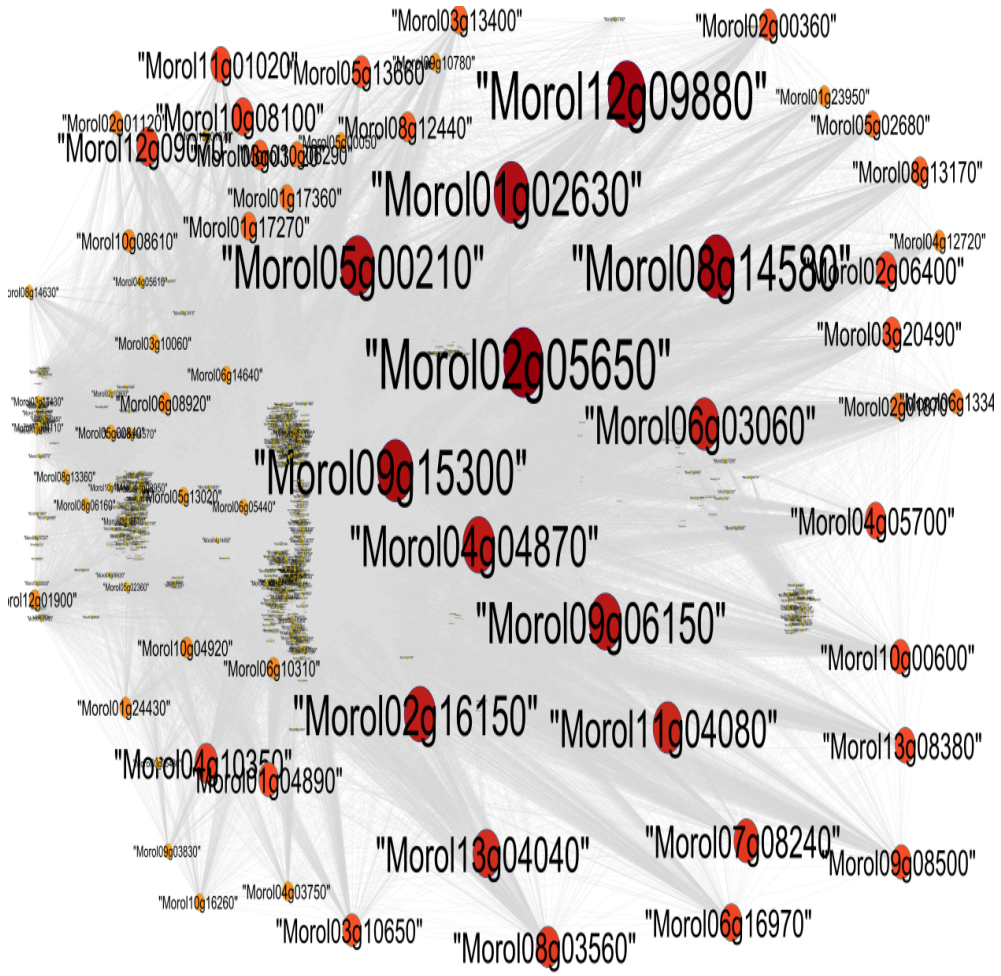


Figure 2.17 Turquoise Module Network
Summary of statistics of the turquoise network analysis using Cytoscape/Cytohubba: Number of nodes: 197; Number of edges: 877; Average number of neighbours:5.263; Network diameter: 6; Network radius: 3; Characteristic path length: 2.799; Clustering coefficient: 0.184; Network density: 0.037; Network heterogeneity: 1.810; Network centralization:0.395; Connected components: 2; Analysis time (sec): 0.031

2.4.5.4 Blue module Network

The Blue module **Figure 2.18** includes the ABCG gene Morol01g26020 (GO:0071949), situated in the cell membrane and participating in trans-membrane transport. The protein encoded by Morol01g26020 has the capability to bind ATP, hydrolyse it, and release energy, which is then employed to aid in the transport of molecules across cellular membranes.

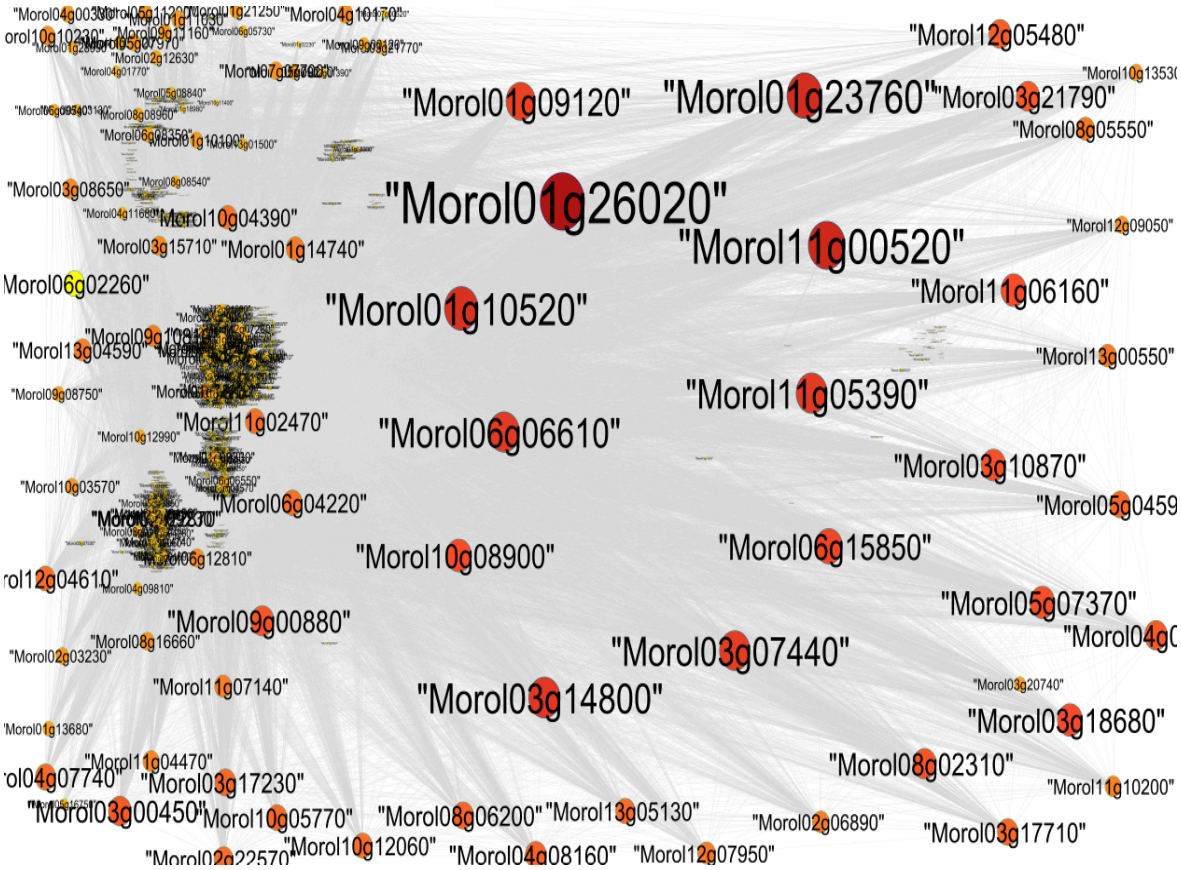


Figure 2.18 Blue Module Network

Summary of statistics of the blue network analysis using Cytoscape/Cytohubba: Number of nodes: 1402; Number of edges: 37823; Average number of neighbours: 54.031; Network diameter: 9; Network radius: 5; Characteristic path length: 2.561; Clustering coefficient: 0.806; Network density: 0.039; Network heterogeneity: 1.723; Network centralization:0.586; Connected components: 2; Analysis time (sec): 3.152

2.4.5.5 Green module Network

The AT1G67300 gene (Morol06g06410) is the hub gene/ master regulator here, other significant genes include Morol02g20210, Morol03g07040, Morol03g20510, Morol07g02390, Morol03g05290, Morol10g00070, Morol01g10190, Morol07g02110 and PP2CG1 (Morol05g00780), among others identified through co-expression network and cytohubba analysis (**Figure 2.19)**.

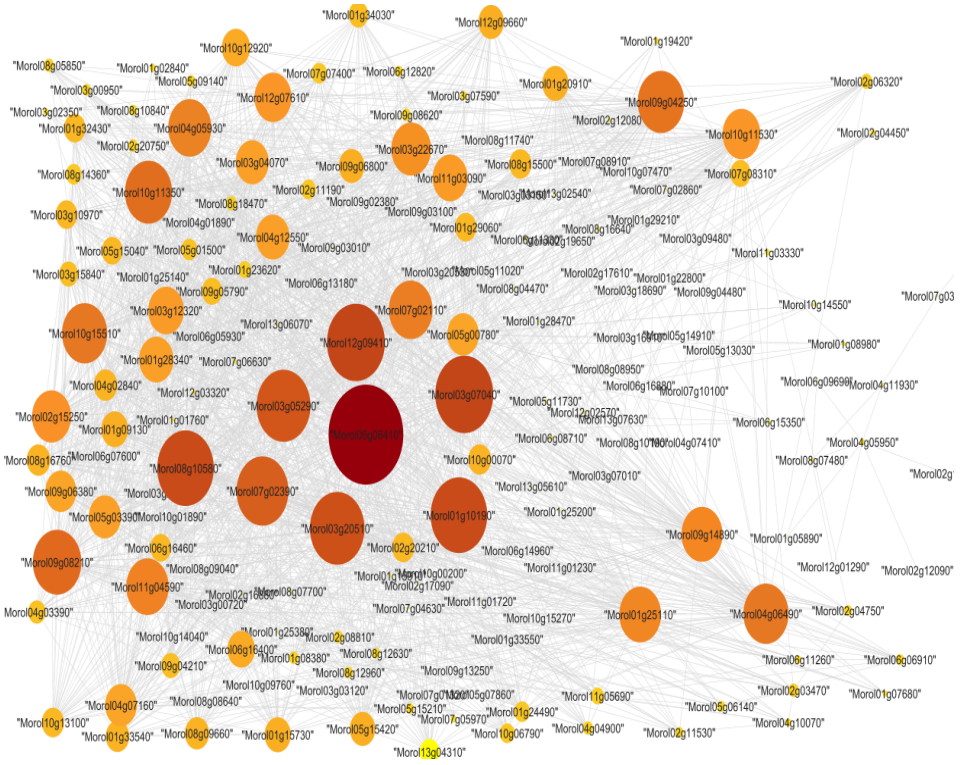


Figure 2.19 Green Module Network

Summary of statistics of the green network analysis using Cytoscape/Cytohubba: Number of nodes: 195; Number of edges: 1860; Average number of neighbours: 21.565; Network diameter: 5; Network radius: 3 ; Characteristic path length: 2.277; Clustering coefficient: 0.639; Network density:0.128; Network heterogeneity: 1.130; Network centralization: 0.536; Connected components: 5; Analysis time (sec):0.023

2.5 Discussion

2.5.1 Identification of significant drought - responsive genes

Plant reactions to drought conditions are related to changes in gene expression and physiological mechanisms. Drought stress, either activates or suppresses genes linked to water conservation, stress signalling, and metabolic changes meant to prevent or lessen consequences of dehydration and preserve cellular homeostasis. Changes seen at the whole-plant level point to molecular systems responding to drought being triggered. These systems can entail modifications in photosynthetic activity or control of stress-related genes. Water stress affects plant resistance and development; indications of a drought include major wilting, leaf chlorosis, and abscission. Analysing the ratio of leaf loss or visual damage index over time helps estimate the effects of drought quantitatively. This link internal molecular alterations with outward symptoms. Leveraging the heterogeneity in drought tolerance shown among Moringa species could allow one to find or create drought-tolerant cultivars. Examining the processes behind these network connections would help to create focused plans to increase drought resilience in Moringa.

The results corresponded with the expected outcomes, as VST generated the rows and columns of gene counts per sample in a transformed format. The measures implemented ensured the data's suitability for further analyses, including differential gene expression and co-expression network analysis. The obtained sequencing data is of high quality, and the effectively mapping demonstrates authentic genetic diversity (Chung *et al.,*2021). The sequence encompasses diverse coding and non-coding regions, facilitating the identification of specific loci linked to drought resistance. This establishes a robust basis for identifying genes that react to drought and comprehending the specific responses of tissues in Moringa. The findings from this transcriptome experiment will be essential for connecting observed physiological changes to their underlying molecular mechanisms.

Although, the overall mapping rate for all samples was 64% this is lower than the optimal mapping rates typically observed in prior research on *Arabidopsis thaliana*, which is between 92.4% and 99.5% (Schaarschmidt *et al.,* 2020), Barley demonstrates a 93% rate (Ost *et al.,* 2023), and *Zea mays* 91.14% (Gui *et al.,* 2022), surpasses 85%. Both instances are underpinned by well-annotated genomes and high-quality RNA samples. The observed mapping rate of 64% in this experiment on *Moringa oleifera* is comparable to the 61% mapping rate reported in Wheat (Barrat *et al.,* 2023). The 64% mapping rate may result from *Moringa oleifera* possessing a less defined genome relative to other extensively studied species, resulting in unaligned reads that suggest the presence of novel transcripts or variations not included in the existing genome assembly. This emphasises the necessity for improved genomic and transcriptomic characterisation to enhance mapping rates and elevate the quality of transcriptomic data.

The top genes identified to be drought – responsive are generally responsible for binding metal ions which play a crucial role in enabling photosynthesis, energy metabolism, root and shoot development, as well as providing antioxidant defense in arid conditions. Metal ions such as copper, iron, zinc, and magnesium play a crucial role in enhancing drought tolerance by interacting with specific proteins that are vital to key physiological processes. Copper-binding proteins play a crucial role in electron transport and enzyme activation, thereby influencing energy metabolism, particularly in scenarios involving water stress. Iron plays a crucial role in the production of chlorophyll by binding to porphyrin rings, thereby increasing chlorophyll levels to enhance light absorption in drought conditions. Zinc plays a vital role in the synthesis of auxin, promoting root development and improving water absorption. Additionally, it regulates the activity of guard cells, helping to minimize water loss through transpiration. Magnesium plays a crucial role in maintaining the integrity of chlorophyll, facilitating energy transfer, and

managing cellular structures in arid environments. The ten most significant genes identified, such as SAG12, metallothionein-like protein type 3, and ferritin-3, highlight numerous potential targets for enhancing drought resistance through genetic manipulation or breeding methods.

2.5.2 DEGs in the Leaf against Stem tissues under Drought-stress.

The difference in gene expression between Moringa's stem tissues and leaves indicates that the leaves may have a more significant role in responding to drought stress, potentially due to alterations in water utilization strategies or stress signalling pathways.

In the context of drought stress in *M. oleifera*, the gene Morol12g05670 demonstrated a significant increase in expression within stem and leaf tissues. The gene encodes the senescence-specific cysteine protease SAG12, located in the lysosome and participating in cysteine-type endopeptidase activity, proteolysis, protein catabolism and associated with nutrient recycling during drought stress. A $\log_2$FC value of 32.58 was associated with the differential expression of this gene in stem and leaf tissues (Fig.2.3C, Volcano plot). The SAG12 gene functions as a potential genetic marker for evaluating the onset and progression of drought-induced senescence in plants, given its strong association with the initiation of leaf senescence.

Identifying SAG12 as a drought-responsive gene is a crucial preliminary step in investigating the influence of aging on drought adaptation moving forward. Species that demonstrate drought resistance and contain SAG12 homologs include Soybean, *Arabidopsis thaliana*, *Zea mays* (maize), *Oryza sativa* (rice), *Triticum aestivum* (wheat), *Hordeum vulgare* (barley), and *Sorghum bicolor* (sorghum). Alterations to these homologs have resulted in elevated photosynthetic rates, improved biomass, and yield through the delay of senescence (Merewitz, *et al.*, 2011). *Os*SAG12 over-expression in rice moderates senescence progression (James *et al.*, 2018). Improved water utilization can be attained through their support, as they prolong photosynthetic activity in drought-affected crops and encourage the development of deeper root systems, allowing plants to access water from deeper soil layers. Modifying SAG12 gene could aid in controlling leaf senescence, while increasing the expression of auxin efflux carriers might improve root development and water uptake, hence enhancing plant adaptation to drought, increases yield, regulates stress hormone transmission, and boosts antioxidant production.

Gawky-like protein Increased in the Stem tissue under Drought-stress.

The protein Morol03g14060, resembling Gawky-like proteins also known as GW182 which plays a role in RNA-mediated gene silencing by micro-RNAs (miRNAs) (Rejwinkel, Ansmant, Gatfield and Izaurralde, 2005), decay and post transcriptional regulation. It

remains insufficiently characterised; nonetheless, its notable up-regulation indicates a potential role in stress response pathways. The increased expression of Gawky-like proteins suggests that *M. oleifera* may be giving preference to drought-responsive transcripts by degrading non-essential miRNAs in the stem, hence reallocating resources during drought stress. The increased expression of these genes is probably essential for preserving stem function and structural integrity in dry conditions. Functional validation studies are crucial for elucidating their distinct roles in drought adaptation.

LEA (Late Embryogenesis Abundant) Decreased in the Stem tissue under Drought-stress.

The diminished expression of Late Embryogenesis Abundant (LEA) proteins observed within the stem tissues of *Moringa oleifera* (Morol07g03510, LEA, LFC:-8.15) may signify several physiological adaptations. While LEA proteins are known for their significant role in helping cells combat loss of water and safeguarding cellular integrity during drought (Wise and Tunnacliffe, 2004). According to Magwanga *et al*., 2018 they were highly expressed in drought tolerant cultivars of cotton such as *Gossypium tomentosum* and lowly expressed in drought susceptible, *G. Hirsutum.* Their downregulation in *Moringa oleifera* stem tissues could indicate a tissue-specific adaptive strategy or degree of sensitivity to drought. This may involve allowing protection of water loss within the leaf tissues, while stems potentially make use of other structural or vascular modifications. Furthermore, it might suggest a transition from osmo-regulatory protective mechanisms to mechanical or hormonal adaptations within the stem. From an agricultural perspective, these findings imply that enhancing drought tolerance could be more effectively achieved by targeting LEA expression specifically within aerial tissues, such as leaves or reproductive organs, rather than through ubiquitous overexpression across all plant tissues.

Vignain-like gene upregulated in the Leaf tissue under Drought-stress.

Morol01g09210 (LFC 14.36) Vignain-like gene upregulated in leaves to maintain protein homeostasis. This cysteine protease involved in proteolysis also described as SAG39 (senescence-specific cysteine protease) *Zea mays* GRMZM2G028862, likely used a mechanism of degrading damaged or misfolded proteins and freed amino acids for protective molecule synthesis for maintaining protein quality during drought stress. The vignain-like gene was possibly involved in stomatal regulation to reduce transpiration (Buet *et al*.,2019).

Genes with significantly lowered expression in the Leaf tissue under Drought-stress.

The down-regulation of Morol02g21300 (LFC:-9.46) PeroxidaseN1-like and Morol10g11100 (LFC:-7.91) Basic Leucine zipper34 (bZIP34) in the leaves of *M. oleifera*

reveals Moringa's transcriptional strategy of selectively reducing energy-intensive processes under drought conditions to balance survival and metabolic cost.

PeroxidaseN1-like (PRX) enzyme which belongs to the large clasIII peroxidase family functions in hydrogen peroxide (H_2O_2) detoxification, cell wall reinforcement and lignification. They help manage oxidative damage caused by abiotic stresses such as drought by reducing reactive oxygen species (ROS) accumulation (Ahmad *et al*., 2021). The decreased expression in Moringa leaves suggests that Moringa might be using an alternative antioxidant defence mechanism which probably relies more on superoxide dismutases (SODs) or catalases which are more energy-proficient under prolonged stress (Prochazkova *et al*.,2001; Ahmad *et al*., 2018). It may imply also that Moringa is actively limiting oxidative metabolism and conserving energy by modifying the cell wall. The decreased peroxidase activity could also infer a controlled, lower metabolic state, thereby avoiding over-accumulation of lignin or excessive ROS signalling which might otherwise stimulate early senescence. Relating this to agricultural context implies that not all antioxidant enzymes require to be increased for a plant to be drought-tolerant, as some like Moringa might even be reduced to keep the plant's systems balanced. It will be more beneficial for plant breeding efforts to focus on creating plants with adaptable antioxidant systems, rather than just boosting all ROS-fighting genes. Different drought-tolerant plants show varied responses in their antioxidant systems depending on the tissue and how severe the drought is, highlighting the benefit of a specific, tailored response for long-term survival. In drought-resilient plant species such as wheat, millet and barley, diverse categories of antioxidants are differentially regulated, based on tissue type and the severity of the drought condition (Gupta *et al*., 2017; Chaves *et al*., 2003).

Basic Leucine zipper34 (bZIP34)

BLZ34, a constituent of the basic leucine zipper (bZIP) transcription factor family, functions in modulating response to abiotic stress, hormonal signal transduction, and the transcriptional activation of protective genes, especially those responsive to abscisic acid (ABA). Specific bZIP transcription factors are recognized for their capacity to activate drought-responsive elements (DREs) and to influence stomatal aperture regulation, osmolyte biosynthesis, and stress hormone signalling pathways (Hussain *et al*., 2021).

The observed downregulation of BLZ34 in *M. oleifera* under drought stress condition is contrary to evidences in rice bZIPs (*Osb*ZIP12, *Osb*ZIP71 and *Osb*ZIP46) and maize (Zea mays) *ZmbZIP4* whose overexpression increased ABA levels and subsequently improved abiotic stress tolerance (Yoon *et al*., 2020; Yang *et al*., 2019). However, the observed decrease in expression of BLZ34 in *M. oleifera* leaf tissues is in line with Wang *et al*., 2015 who reported *Mtb*ZIP34 in *Medicago truncatula* was down-regulated under

drought stress and homologous to ABFs (ABRE binding factors)/AREB (ABA-responsive element binding protein). This may signify several strategic adjustments. Firstly, it could represent a deliberate attenuation of ABA signalling, especially if sustained activation might hinder growth or photosynthetic efficiency. Secondly, it might indicate a mechanism to manoeuvre energy-intensive stress responses thereby preserving physiological homeostasis. Thirdly, it could suggest a compensatory function assumed by other transcription factor families, including NAC, DREB, or MYB, which may be preferentially utilized by Moringa under drought conditions.

From an agricultural perspective, these findings explains the principle of regulated transcriptional modulation, rather than a uniform upregulation of all drought-associated genes. Furthermore, the drought-intensity-specific regulation of bZIPs could facilitate the maintenance of crop growth during mild to moderate drought, which is crucial for sustaining agricultural yields. Gene editing methodologies, such as CRISPR, could be employed to investigate attenuated bZIP activity, potentially enabling a controlled stress response without compromising biomass accumulation. Comparatively, in Rice (*Oryza sativa)* the overexpression of OsbZIP23 has been demonstrated to enhance drought tolerance, although at the expense of growth retardation (Yang *et al*., 2019). Similarly, Ma, *et al*., 2018 reported overexpression of Maize (*Zea mays*) ZmbZIP4 with increased ABA levels while the ABA levels of zmbzip4 mutants showed opposite pattern. Altogether, these observations highlight the importance of dynamic gene expression and cross-talk with other biological pathways.

KNAT2 Increased in both Stem & Leaf during Drought.

According to Song *et al*., 2021 KNAT2 mediated changes in *Populus alba* plant structure in response to drought stress which led to reduced gibberellin levels, smaller leaves and shorter internodes. In Moringa, this may likely signal similar morphological adaptations, whereby providing good plant form and structure for drought resilience through hormonal regulation.

Genes with significantly lowered expression in Stem & Leaf during Drought.

CYP82D47

Morol05g04740 encodes Cytochrome P450 CYP82D47 $\log_2$fold change (LFC) of -8.573 was observed, an enzyme known to control cell division and cell expansion, vascular differentiation, root development among others. They also protect plants from drought, dehydration, and UV stress. They are classified as an oxidoreductase, involved in the biosynthesis of various secondary metabolites such as flavonoids, carotenoids and hormones as well as defensive compounds, sterols, fatty acids, and components of the plants cell wall (Minerdi, Savoi & Sabbatini, 2023).Although, CYP450 genes do not

directly regulate drought stress, they tend to regulate the levels of ABA, protect plants from damaging effects of oxidative stress (Pandian *et al*., 2020) by synthesising antioxidants like flavonoids and carotenoids, and also help plants cope with the challenges of drought. According to Xing *et al*., 2022, over expression of *Ib*CYP82D47 increased carotenoid contents in transgenic sweet potato (*Ipomoea batatas*) roots and influenced the expression of genes involved in carotenoid biosynthesis. It is expected to see an upregulation of this enzyme in Moringa during drought, but as CYP82D27 was observed to decrease in both stem and leaves of Moringa during drought stress, this may suggest that *M. oleifera* prioritizes energy conservation by attenuating secondary metabolic processes to cope with the absence of water, such that it redistributes its resources to other genes that are more important for survival under drought stress. While these compounds offer protective benefits, their production entails a substantial energetic cost. From an agricultural perspective, this phenomenon emphasises the utility of temporal regulation of secondary metabolism in breeding programmes, specifically its suspension during periods of severe environmental duress and its subsequent reactivation during recovery phases or under conditions of moderate stress.

SAG12

The senescence-associated gene SAG12, encoded by Morol12g05670, exhibiting an LFC of -8.466 was also seen to be upregulated. This dual regulation, such as an initial transient upregulation succeeded by subsequent suppression, may be indicative of temporal dynamics. This phenomenon could represent an initial activation of senescence mechanisms for the purpose of nutrient reallocation, subsequently followed by a suppression of these processes to facilitate the maintenance of vital tissues during stress. Understanding the stage-specific expression patterns of SAG12 could provide valuable insights for plant breeding programmes thereby enabling a more effective equilibrium between senescence and tissue viability to be achieved.

PRX

When comparing control conditions to drought stress, the expression of several genes, including Morol02g21300 and Morol02g21290, decreased across the leaf and stem tissues. The Morol02g21300 encoding a Peroxidase N1-like protein (PRX). This protein protects cells from ROS damage as it takes part in the catabolism of hydrogen peroxide and the response to oxidative stress by exhibiting peroxidase activity and being involved in heme binding. According to Su *et al*., 2023, overexpression of *Ta*PRX-2A enhanced the tolerance to drought stress of the transgenic lines than the wild type wheat plants.

CCP2

A Cationic Peroxidase 2-like protein (CCP2) likely shares structural and functional similarities with Cationic Peroxidase 2 (CCP2), a plant enzyme known for its positive charge and peroxidase activity being able to catalyse oxidation reactions using hydrogen peroxide as an electron acceptor (Nnamchi, *et al*., 2021). They are likely involved in plant defense by oxidizing various substrates like phenolic compounds to help strengthen the cell wall (Chai *et al.*, 2020). The results indicated that these genes exhibited consistent decreases in expression across samples, suggesting that there are significant transcriptional changes that occur in response to drought.

2.5.3 Biological processes and signalling pathways involved in drought adaptation in *M. oleifera*.

The interrelation of these genes plays a multi-layered drought-response strategy in as seen in the physiological processes such as stress signalling, metal homeostasis, and auxin transport, all of which significantly influence Moringa's ability to endure water scarcity. Furthermore, genes linked to metal detoxification and protection against oxidative stress, including metallothionein and ferritins, suggest possible pathways for improving resilience under drought conditions. Summarily, the integrated mechanisms comprising cellular protection and senescence (SAG12, Vignain), developmental modulation (KNAT2), post transcriptional regulation (Gawky-like protein), and membrane and signalling adjustments (SBH1) allows Moringa to prioritize survival, conserve resources and maintain tissue integrity during drought.

Additional studies are necessary to confirm these genes to enable the understanding of their roles in drought responses, exploring their relationship, and evaluating the impact of their modification on overall plant performance in water-limited environments.

2.5.4 Key regulatory hub genes that coordinate the transcriptional response to drought stress.

The key regulatory hub genes that coordinate the transcriptional response to drought stress in *Moringa oleifera* are seen in the module networks yellow (Fig.2.14), cyan (Fig.2.15), turquoise (Fig.2.16), blue (Fig.2.17), green (Fig.2.18) and discussed as follows:

Yellow module Network

Summary of the network analysis statistics reveals the network contains 752 nodes (genes) connected by 24190 edges (interactions), network is fully connected consisting of a single component (all nodes are reachable within the network). Each node shows relatively high connectivity across the network with an average of 64 genes. The network diameter is 6, and radius 3 which infers that the farthest three nodes can be connected through only 6 steps, and most nodes are within two steps of a central point. Path length is 2.12, which implies efficient connection and strong integration within the network. Clustering coefficient 0.845 is high, indicating co-regulated gene sets. Network density is

0.086 showing network is not fully saturated. The network heterogeneity (1.435) infers there is variability in node connectivity, meaning that some nodes such as the red/orange nodes act as hubs with many more connections than others. The centralization value of 0.684 also supports the presence of key hub nodes (genes) that regulate or dominate the network's structure.

The Morol08g04710 (GO:0005777) ketoacyl-CoA C-acyltransferase enzyme also known as ketoacyl CoA thiolase 2 is involved in fatty-acid breakdown which releases acetyl Co-A and synthesis of essential metabolites such as CoA ester of a fatty acid with 2 carbons shorter (Goudarzi, 2019; Szrok-Jurga *et al*., 2023). During drought stress, plants often rely on stored lipids as energy sources. By facilitating the breakdown of fatty acids, Morol08g04710 can help provide energy and substrates necessary for stress responses and likely contribute to the drought tolerance of Moringa plants. Its central position in the yellow module indicates its significant influence on the network of co-expressed genes, making it an important target for further research into enhancing drought resistance in this species.

Cyan Module Network

Summary of the network analysis statistics reveals the network contains contains 141 nodes (genes) connected by 2632 edges (interactions), network is fully connected consisting of a single component (all nodes are reachable within the network). Each node shows relatively high connectivity across the network with an average of 37 genes. The network diameter is 4, and radius 2 which infers that the farthest two nodes can be connected through only 4 steps, and most nodes are within two steps of a central point. Path length is 1.77, which implies efficient connection and strong integration within the network. Clustering coefficient 0.787 is high, indicating co-regulated gene sets. Network density is 0.267 showing network is not fully saturated. The network heterogeneity (0.768) infers there is variability in node connectivity, meaning that some nodes such as the red/orange nodes act as hubs with many more connections than others. The centralization value of 0.628 also supports the presence of key hub nodes(genes) that regulate /dominate the network's structure.

The Morol01g14900, Polygalacturonase, deemed to be the top master regulator of drought-responsive genes in the cyan module is notable for cell-wall metabolism, degrading and plant defense (Chaudhary and Amin, 2022). Berberine bridge enzyme-like13 (BBE-like 13) (Morol12g08750) are flavin adenine dinucleotide (FAD) - linked oxidases reported to be upregulated during stress conditions such as salt and

osmotic stress and *At*BBE-like 28 found present only in Brassicaceae family (Daniel *et al*., 2016; 2015). The CBFA/NF-YB transcription factor also play an important role in Moringa's drought resistance by regulating the expression of stress-responsive genes and improving protective mechanisms. Integrating diverse stress signals and coordinating a comprehensive response is crucial for a plant's adaptability to water scarce environments. The CBFA/NF-YB transcription factor (Morol10g08670, GO: 0016602) essential for transcriptional regulation have been reported to activate genes in response to dehydration (Sato *et al*., 2019). It likely enhances the plant's overall response to drought conditions hence enabling Moringa plants to withstand water scarcity. Other notable genes in the cyan module such as Cytochrome P450 (CYPs) (Morol05g06560, GO: 0005506) which is associated with various metabolic pathways and stress responses as reported by Xing *et al.*,2023, where overexpression of (*Lolium perenne*) *CYP* gene, *LpCYP72A15*, was identified to significantly enhance drought tolerance in both Arabidopsis and perennial ryegrass. Whereas Probable glutathione S-transferase GST (Morol10g09360) plays a role in defence against oxidative stress by neutralising harmful substances and ROS (Kumar *et al*., 2018). The metabolic pathways related to glucose (Morol01g14900, GO: 0004650) and glutathione (Morol10g09360, GO: 0004364) metabolism demonstrate a complex network of reactions that function to sustain cellular homeostasis during drought conditions.

Turquoise module network

Summary of the turquoise network analysis statistics reveals the network contains 197 nodes (genes) connected by 877 edges (interactions), network is fully connected consisting of a single component (all nodes are reachable within the network). Each node shows relatively high connectivity across the network with an average of 5 genes. The network diameter is 6, and radius 3 which infers that the farthest three nodes can be connected through only 6 steps, and most nodes are within 3 steps of a central point. Path length is 2.79, which implies efficient connection and strong integration within the network. Clustering coefficient 0.184 is low. Network density is 0.037 showing network is not fully saturated. The network heterogeneity (1.810) infers there is variability in node connectivity, meaning that some nodes such as the red/orange nodes act as hubs with many more connections than others. The centralization value of 0.395 also supports the presence of key hub nodes(genes) that regulate or dominate the network’s structure.

Morol02g05650 dentin sialophosphoprotein-like isoform X2 [*Carica papaya*] TIGR00927 is a K+-dependent Na+/Ca+ exchanger and significant transport and binding protein, as well as a carrier of cations and iron carrying compounds, considered a potential master regulator of drought-responsive genes in the turquoise module.

The MYB-like transcription factor plays a crucial role in regulating stomatal opening responsible for water-loss as seen in AtMYB60 and AtMYB96 in Arabidopsis (Li *et al*., 2019) and controlling gene expression related to the production of protective metabolites, including flavonoids, wax, and cutin (Wang *et al*., 2021). These metabolites are recognised for their role in improving drought tolerance through the reduction of water loss and the protection of cellular structures during stress conditions. The delta (Morol12g09880) and gamma (Morol02g16150) coatomer subunits are essential components of the COPI complex, playing a significant role in vesicular transport within the Endoplasmic Reticulum-Golgi pathway in the cell (Ahn, *et al.,* 2015). Under drought stress conditions, coatomer proteins are essential for the transport of proteins and membranes required for sustaining cellular integrity and function, thus aiding the plant in adapting to water scarcity. The GALS1 enzyme synthesises β-1,4- galactan which is a component of pectin in plant cell walls (Liwang *et al*., 2012) may not enhance drought tolerance directly but Xyloglucan endotransglucosylase/hydrolase (XTH) genes / Xyloglucan glycosyltransferase 6, like the wheat *Ta*XTH12.5a, positively contribute to drought tolerance (Qiao *et al*., 2022). They are crucial for maintaining cell wall composition and structural integrity during drought conditions, thereby aiding in the preservation of cell turgor pressure to avert wilting (Bi *et al*., 2024). Xyloglucan glycosyltransferase 6 is crucial for the biosynthesis of xyloglucan, an important hemicellulose that enhances cell wall rigidity, enabling the plant to endure mechanical stress. The genes work together to improve root growth and development, allowing the plant to reach deeper water sources in times of water scarcity. The roles of these genes in drought response mechanisms in stomatal closure, cell wall maintenance and the modulation of drought tolerance traits, indicate their potential as significant targets for enhancing drought resilience in Moringa.

Blue Module Network

Summary of the blue network analysis statistics reveals the network contains 1402 nodes (genes) connected by 37823 edges (interactions), network is fully connected consisting of a single component (all nodes are reachable within the network). Each node shows relatively high connectivity across the network with an average of 54 genes. The network diameter is 9, and radius 5 which infers that the farthest 5 nodes can be connected through only 9 steps, and most nodes are within 5 steps of a central point. Path length is 2.79, which implies efficient connection and strong integration within the network. Clustering coefficient 0.184 is low. Network density is 0.037 showing network is not fully saturated. The network heterogeneity (1.810) infers there is variability in node connectivity, meaning that some nodes such as the red/orange nodes act as hubs with many more

connections than others. The centralization value of 0.395 also supports the presence of key hub nodes (genes) that regulate or dominate the network's structure.

The ABCG transporters Morol01g26020 gene enables the function of ABC-type transporters, including ATP binding, ATP hydrolysis, and Flavin Adenine Dinucleotide (FAD) binding. The ABC transporter utilizes energy from ATP hydrolysis to facilitate the translocation of various molecules across the cell membrane (Rees, Johnson and Lewinson, 2009). Kumori *et al*., 2010 reports that overexpression of this transporter reduced water loss from leaves as it enables the delivery of ABA to guard cells. The gene plays a crucial role in the transfer of osmo-protectants, such as sugars and amino acids, highlighting its function in regulating water balance and protecting cellular structures under dry conditions. Additionally, Morol01g26020 may play a role in the transfer of phytohormones, including abscisic acid (ABA), which are critical regulators of responses to drought stress (Yadav *et al*., 2025; Jarzyniak and Jasinski, 2014). The gene's ability to bind FAD may play a role in redox based signalling pathways activated during water stress, facilitating the plant's recognition and response to drought-related environmental challenges. ABCG transporters Morol01g26020 facilitate the transport of cuticular lipids to reduce water loss, nutrients, hormones, and protective chemicals, contributing to a coordinated response to drought stress and enhancing plant resilience under such conditions. These mechanisms can inform breeding and biotechnology strategies aimed at improving drought resistance in crops.

Green module Network

Summary of the green network analysis statistics reveals the network contains 195 nodes (genes) connected by 1860 edges (interactions), network is fully connected consisting of a single component (all nodes are reachable within the network). Each node shows relatively high connectivity across the network with an average of 21 genes. The network diameter is 5, and radius 3 which infers that the farthest three nodes can be connected through only 5 steps, and most nodes are within3 steps of a central point. Path length is 2.277, which implies efficient connection and strong integration within the network. Clustering coefficient 0.639 is high, indicating co-regulated gene sets. Network density is 0.128 showing network is not fully saturated. The network heterogeneity (1.130) infers there is variability in node connectivity, meaning that some nodes such as the red/orange nodes act as hubs with many more connections than others. The centralization value of 0.536 also supports the presence of key hub nodes (genes) that regulate or dominate the network's structure.

The AT1G67300 gene (Morol06g06410) is situated in the plasma membrane and is responsible for the transmembrane transport of carbohydrates. This process facilitates

the passage of carbohydrates across the membrane, potentially improving the plant's capacity to adapt to drought conditions. Likewise, PP2CG1 (Morol05g00780) is a significant gene in this network, located in the cytoplasm, chloroplast, and nucleus. It functions within or upstream of pathways that are responsive to salinity stress, water scarcity, and abscisic acid (ABA). The PP2CG1 gene encodes an enzyme that has myosin phosphatase activity and may be involved in the regulation of a variety of stress response systems.

In response to water-deficit conditions, AT1G67300 and PP2CG1 function as negative regulators of ABA signalling, interacting with additional drought-responsive pathways to improve the plant's coordination. It is anticipated that the PP2C enzymes produced by these genes will interact with additional stress-responsive signalling pathways, including the Mitogen-Activated Protein Kinase (MAPK) pathways, thereby enabling the plant to precisely regulate its response to drought and other abiotic stresses.

2.5.5 Comparison with Other Drought-Resistant Crops

Moringa's drought response has both similarities and differences when compared to Arabidopsis, wheat and maize, which highlight the unique adaptations each species has developed to cope with drought, reflecting their evolutionary backgrounds and ecological niches Supp Table 2.1. In terms of these plants' drought - response mechanisms, Moringa exhibits extensive root systems for enhanced water uptake, efficient water use through stomatal regulation, and accumulation of osmo-protectants like proline. It also activates antioxidant enzymes to mitigate oxidative stress (Vijayaragavan *et al.,* 2025, Hajaji *et al.,* 2024). Arabidopsis responds to drought through early accumulation of abscisic acid (ABA), which regulates stomatal closure to reduce water loss. It also involves jasmonate signalling for drought resistance (Yao *et al.,*2021). Wheat's drought response includes osmotic adjustment through the accumulation of osmo-protectors, maintaining cell turgor, and reducing stomatal conductance to conserve water (Nardino *et al*., 2022; Ahmad, *et al.,* 2018). Maize responds to drought by activating ABA-mediated signalling pathways, osmotic protective substance synthesis, and protein folding responses. It also shows tissue-specific responses to drought stress (Singh *et al.,* 2023; Wang *et al.,* 2019).

Their adaptations to drought are described as follows: Moringa's drought tolerance is supported by morphological, physiological, biochemical, and molecular adaptations, making it highly resilient in arid conditions, while Arabidopsis uses a combination of early priming, acclimatization, and new homeostasis stages to manage drought stress. This includes changes in gene expression related to cell wall expansion and stress. Wheat shows physiological changes like increased root-to-shoot ratio, early maturity, and

enhanced antioxidant activity. Modern breeding techniques aim to improve drought tolerance by selecting for these traits' response. Maize's drought tolerance involves changes in photosynthesis, carbohydrate metabolism, and hormonal signalling. Advanced breeding and genome editing techniques are used to enhance drought tolerance in maize.

Moringa shows similar mechanisms as Arabidopsis, Wheat and Maize by utilizing ABA signalling and antioxidant mechanisms to manage drought stress. They also show physiological and biochemical changes to conserve water and protect cellular integrity. Whereas Moringa's extensive root system and high resilience in arid conditions set it apart. Arabidopsis relies heavily on early priming and acclimation stages, while wheat focuses on osmotic adjustment and early maturity. Maize exhibits tissue-specific responses and benefits from advanced breeding techniques for drought tolerance.

Guo *et al*., 2015 reported good levels of drought tolerance in *Brassica rapa* as GO terms related with sulphur utilization and regulation, indole glucosinolate synthesis, peroxisomal organization cellular response to glucose, ethylene metabolism and response to water stimulus were upregulated. Drought responses were induced by hormone signalling pathways such as Salicylic Acid (SA) and Jasmonic Acid as well as quick response networks to systemic acquired resistance, SA and MAPK cascade. Guo *et al*., 2015 also identified variations in drought tolerance among 10 accessions of *B. rapa* and *B. juncea* as some are more drought tolerant than the other and generally show reduced growth and productivity under drought stress. While *Moringa spp*. are particularly known for how they have evolved to survive in dry conditions, their adaptation to drought has also been linked to their robust cellular structure and ability to lower water potentials and maintain turgor pressure. Similarly to Moringa, Brassica utilizes osmotic adjustment, antioxidant defense and hormone regulation to combat the effects of drought (Mohan, MacDonald & Abbey, 2025)

2.6 Conclusion

The research findings offer a thorough understanding of the transcriptional response of Moringa species to arid stress. This study identified differentially expressed genes (DEGs) and categorized them into co-expression modules to clarify the molecular mechanisms that allow Moringa to endure water-deficit conditions. The analysis identified specific genes, gene families, and biological processes significant in drought response, indicating both established pathways and novel mechanisms. This understanding is required for developing strategies that improve drought tolerance in cereals by highlighting analogous genetic and molecular processes.

The differential expression analysis identified 9,249 differentially expressed genes distributed among 49 co-expression modules, reflecting a complex transcriptional response to drought. These, 9,249 significant genes were detected on applying LFC >+1.5 or <-1.5 and FDR/padj value of less than 0.5 on the initial total number of genes 19,325 in the analysis. The study demonstrated that genes Morol12g05670 (SAG12) and Morol01g26440 showed significant overexpression among 2,319 genes when comparing Control and Drought conditions, while genes such as Morol02g21300 (Peroxidase N1-like) and Morol02g21290 (Cationic Peroxidase 2-like) demonstrated notable downregulation. 3,196 genes were significantly differentially expressed in the leaf tissues, 979 genes in the stem tissues and 2,935 genes expressed when both stem and leaves were combined.

The results describe the dynamic characteristics of Moringa's gene expression in response to drought, showing diverse patterns for genes linked to specific tissues and stress responses. For example, Morol12g05670 (SAG12) and Morol03g14060 (Gawky-like protein) exhibited a significant increase in stem tissue, while genes such as Morol07g03510 (Late embryogenesis abundant protein-related) showed a decrease in the same tissue type. The Morol12g05670 (SAG12) and Morol01g09210 (Vignain) among others showed significant increase in the leaf tissue, while expression of genes like the Morol02g21300 (Peroxidase N1-like) and Morol10g11100 (basic leucine zipper 34) decreased. This indicates that each tissue exhibits a unique response to drought conditions.

The identification of uncharacterised genes, including Morol08g10460, suggests the potential existence of drought-responsive genes in Moringa, as evidenced by its pronounced differential expression with a highly significant corrected p-value (5.18E-31). The Cyan module, enriched with genes associated with oxidative stress defence and transcriptional regulation, is where the novel gene was identified, suggesting its potential role in these pathways. The insights may uncover novel gene functions or regulatory mechanisms, facilitating the development of new research pathways into the genetic foundations of drought tolerance.

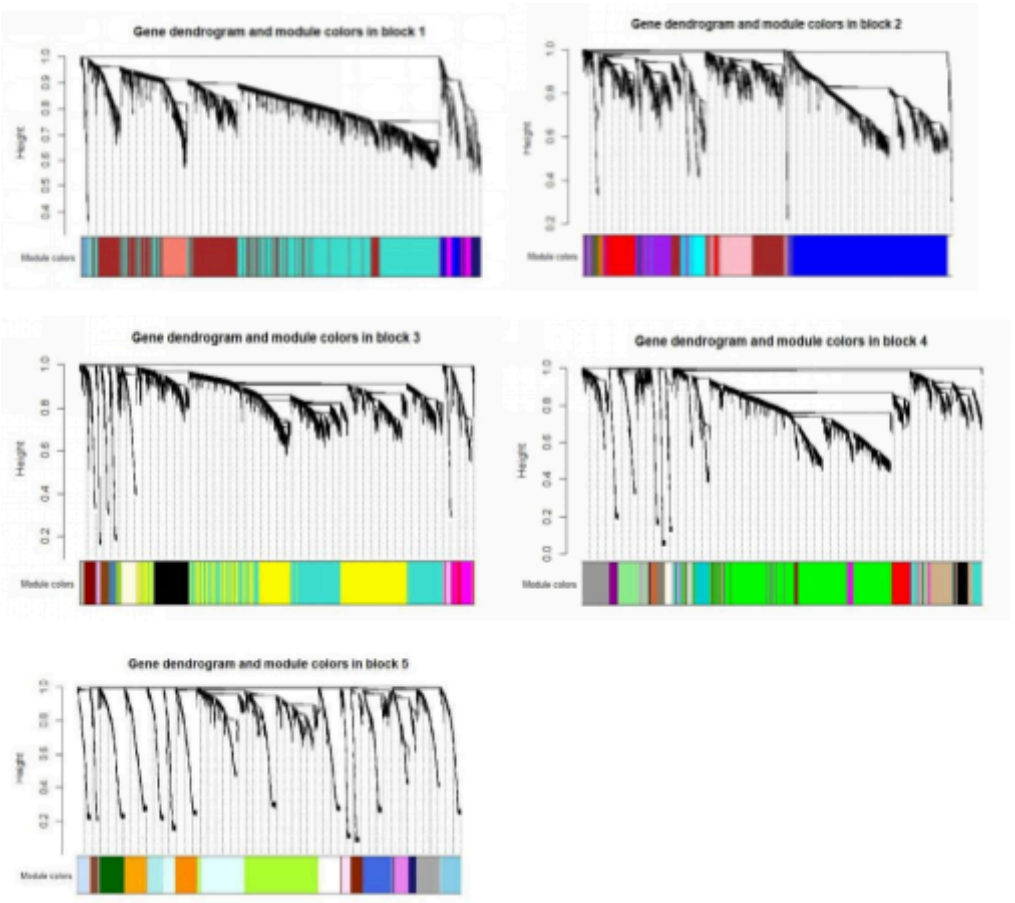
Several modules, such as Red, Light Cyan, Green, Cyan, Brown, and Blue, exhibited a notable correlation with the drought-stress response. The Cyan module was linked to oxidative stress defense and transcriptional regulation, with each module demonstrating distinct functional relationships. This module includes essential genes, such as the CBFA/NFYB transcription factor (Morol10g08670), which may affect the expression of drought-responsive genes related to signal transduction, osmo-protectant production, and stomatal regulation. The Turquoise module is linked to genes that play a role in

stomatal regulation and cell wall remodeling, including the MYB-like transcription factor and GALS1 (Galactan synthase 1). This association may assist the plant in preserving cell integrity under drought conditions. The presence of ABC transporters in the blue module suggests their involvement in the trans-membrane transfer of phytohormones and Osmo-protectants. The Green module contained negative regulators of ABA signalling, including AT1G67300 and PP2CG1, which coordinated the plant's response to drought. The association of the red module with autophagy-related processes indicates that cellular recycling systems may be enhanced during drought conditions to conserve resources.

This study delineates the transcriptional landscape of Moringa under drought stress, highlighting essential genes, modules, and pathways implicated in the response. The investigation of genetic elements, including the functional characterization of hub genes and the validation of novel genes, may facilitate the development of drought-resistant Moringa cultivars and comprehensive agricultural solutions to address water scarcity.

Appendix

Supplementary Figures



Supp. Figure 2.1. Gene dendogram and module colours for 5 network blocks using TOM dissimilarity.

Supplementary Tables

Supp. Table 2.1 Comparison of Drought resistance mechanisms of *Moringa oleifera* with rice, maize and *Arabidopsis thaliana*.

	Rice (*Oryza sativa*)	**Maize (*Zea Mays*)**	***Arabidopsis thaliana***	***Moringa oleifera***
ABA signalling	AREB/ABF bZIP TFs (e.g. OsAREB1); *NCED* and CPK6 activation under drought (Zhang *et al*., 2020; Zhou *et al*., 2025; Zha, He & Gong, 2025)	PYR/PYL–PP2C–SnRK2 core; MAPK cascades (ZmPP2C84 →ZmMEK1/ZmSIMK1) (Wang *et al*., 2018)	SDIR1 E3 ligase enhances ABA-related signalling; CPKs (e.g. CPK6) phosphorylate ABFs for stomatal control (Harb *et al*., 2010)	Upregulates *NCED* and ABA-responsive CDPK/HSF genes under drought; genome reveals expanded HSF family (Pasha *et al*., 2020; Chang *et al*.,2022)
DREB regulon	OsDREB1/2 induce osmotic and heat stress genes (Cui *et al*., 2011)	ZmDREB2.7 activates osmotic-stress response genes (Liu *et al*., 2013)	DREB family strongly co-induced under drought; ~2000 responsive genes, two-thirds ABA-dependent	NIL
NAC TFs	SNAC1 / OsNAC6 regulate protective genes (Wairich *et al*., 2023)	ZmNAC for drought resilience (Liu et al., 2023)	NAC016/ NAP form feed-forward loop repressing AREB1; NAC (Nguyen *et al*., 2019; Sakuraba *et al*.,2015)	NIL
WRKY TFs	OsWRKY13/45 link abiotic and biotic stress (Xiao *et al*., 2013)	WRKYs integrated in MAPK cascades (Wang *et al*., 2010)	AtWRKY57 shows tolerance to Drought (Jiang *et al*., 2012). MYB44–ENAP interaction regulates histone acetylation and WRKY70 expression (Zhao *et al*., 2022)	Likely expressed; transcriptome shows stress-related WRKY induction (needs functional validation) Zhang *et al*., 2019
Hormonal crosstalk	ABA–GA balance; ROS–ABA via OsAAI1(Long *et al*., 2023; Adzigbe *et al*., 2025)	Ethylene responsive factor/ABA/JA balance (Sheoran *et al*., 2022).	Extensive cross-talk: ABA dominant but JA, ethylene, auxin, cytokinin also modulate drought genes (Thonglim *et al*., 2023)	Accumulates proline, phenolics, antioxidants (Bekka et al., 2022); extensive root growth and stomatal regulation via ABA.
Root adaptation	ABA and GA coordinate root growth under drought (Kim *et al*., 2020).	Maintains root apical growth via ABA & IAA; boosts lignin biosynthesis (Feng *et al*., 2022).	Brassinosteroids signalling regulates root development / drought stress /suppresses PLT1, PLT2 expression in *Arabidopsis thaliana* (Zhao *et al*., 20230	Robust root system, stomatal regulation; transcriptomes enriched for water uptake and secondary metabolism pathways (Vijayaragavan *et al*., 2025) .

Supp. Table 2.2 Network topology analysis for choosing the soft threshold power. Soft threshold power of 12 picked for the TOM

```
> save(expression.data, file = "moringa_drought_network.RData") # ESSENTIAL
> # Choose a set of soft-thresholding powers
> powers <- c(1:20)
> # Call the network topology analysis function.
> # This will tell us the soft threshold power we need to get a
> # scale-free topology, which is one of the requirements of WGCNA.
> sft <- pickSoftThreshold(expression.data, powerVector = powers, verbose = 5,
+                          networkType = "signed hybrid")
pickSoftThreshold: will use block size 2178.
 pickSoftThreshold: calculating connectivity for given powers...
   ..working on genes 1 through 2178 of 20538
   ..working on genes 2179 through 4356 of 20538
   ..working on genes 4357 through 6534 of 20538
   ..working on genes 6535 through 8712 of 20538
   ..working on genes 8713 through 10890 of 20538
   ..working on genes 10891 through 13068 of 20538
   ..working on genes 13069 through 15246 of 20538
   ..working on genes 15247 through 17424 of 20538
   ..working on genes 17425 through 19602 of 20538
   ..working on genes 19603 through 20538 of 20538
   Power SFT.R.sq  slope truncated.R.sq mean.k. median.k. max.k.
1      1   0.0263  0.126          0.674  3430.0   3180.00   6530
2      2   0.5420 -0.659          0.863  1600.0   1440.00   4110
3      3   0.7120 -0.924          0.930   903.0    769.00   2910
4      4   0.7730 -1.080          0.958   568.0    444.00   2190
5      5   0.7950 -1.210          0.965   384.0    271.00   1720
6      6   0.8190 -1.300          0.972   273.0    173.00   1390
7      7   0.8270 -1.370          0.972   201.0    114.00   1150
8      8   0.8390 -1.410          0.977   153.0     78.70    961
9      9   0.8400 -1.450          0.975   119.0     55.50    816
10    10   0.8450 -1.490          0.976    94.5     39.90    699
11    11   0.8500 -1.510          0.977    76.3     29.80    607
12    12   0.8510 -1.530          0.977    62.5     22.70    531
13    13   0.8500 -1.550          0.976    51.8     18.00    468
14    14   0.8470 -1.560          0.973    43.4     14.20    415
15    15   0.8480 -1.570          0.973    36.7     11.10    369
16    16   0.8510 -1.570          0.973    31.4      8.65    331
17    17   0.8410 -1.580          0.966    27.0      6.77    298
18    18   0.8440 -1.570          0.965    23.4      5.34    269
19    19   0.8400 -1.570          0.957    20.4      4.26    244
20    20   0.8450 -1.560          0.957    17.9      3.41    222
```

Chapter 3

Comparative Analysis of Fatty Acid, Phytosterol, and Tocopherol Profiles in Moringa Seeds from Diverse Geographical Areas for Breeding and Utilization

Abstract

This chapter investigates the compositional diversity and ecological modulation of fatty acids, phytosterols and tocopherols in seed oil from multiple accessions of *Moringa oleifera* and *Moringa stenopetala* sourced across 48 distinct geographic regions through Genome Wide Association Studies (GWAS). Quantitative profiling revealed significant intra - and inter-accession variability, particularly in β-sitosterol, Δ^5-avenasterol, and Δ^7-stigmastenol, with notable outliers such as IMo9 (India) and WOT (commercial line) exhibiting extreme values. Δ^5-avenasterol concentrations varied widely in some Kenyan accessions KMo4, while others showed biochemical uniformity, suggesting differential genetic backgrounds and environmental stability. Δ^7- stigmastenol and stigmasterol distributions were similarly influenced by genotype–environment interactions. Notably, cholesterol and cholestanol, traditionally minor sterols in plant systems were present at detectable levels, especially in arid-zone accessions. Their occurrence implies potential roles in membrane stabilization under abiotic stress and oxidative conditions. Stigmasterol, with its established membrane and nutraceutical functions, also varied regionally, possibly reflecting metabolic adaptations to environmental pressures. Despite numerical differences in other minor sterols (e.g., campestanol, citrostadienol), statistical analyses did not support strong geographic influence, likely due to local variability. The overall sterol composition aligns Moringa oil with high-quality vegetable oils, underscoring its oxidative stability and health-promoting potential. These findings deepen our understanding of the biochemical diversity within Moringa species, provide valuable information for distinguishing different types based on their chemical profiles, and open up exciting opportunities for targeted breeding and industrial applications focused on improving sterol content. The Moringa accessions with high ratio of unsaturated fatty acids to saturated fatty acids were identified which could serve as markers for crop breeders desiring to cultivate Moringa for heart-health benefit, dietary and pharmaceutical purpose. While those accessions with high behenic acid was identified and would be useful for industrial purposes such as for lubcricants, biofuel and others. SNPs associated with the oil traits, fatty acids, phytosterols and tocopherols identified at unique loci linked to specific biosynthetic pathways, provides insight into genetic regulation and potential breeding targets for optimised seed oil quality.

Overall, this study would serve as markers for crop breeders desiring to cultivate Moringa for seed oil purpose.

3.1 Moringa oil

The Moringa species, especially *Moringa oleifera* and *Moringa stenopetala*, have become popular lately for their various applications in oil extraction, medicine, and nutrition (Anwar *et al*., 2006). Moringa seed oil, or Ben oil, is a multifunctional and nutrient-rich oil with similarities to other beneficial oils. Moringa oil is characterised by its high phytosterol content, varied tocopherol profile, and significant oleic acid concentration, leading to a nutritional composition comparable to that of olive oil (Leone *et al*., 2016). This study emphasises the influence of geographical variations on the composition of Moringa oil, noting its significant unsaturated fatty acid content, comparable to olive oil, along with its sterol and tocopherol profiles. Moringa oil is a versatile oil suitable for industrial, cosmetic, and culinary uses owing to its various properties. The characteristics include a high smoke point of 472°F (244°C) (Rashid *et al*., 2008), making it suitable for high-heat cooking methods without combustion; a rich fatty acid profile, particularly mono-unsaturated fatty acids akin to those in olive oil and avocado oil (Palafox-Carlos, 2011); an appealing mild nutty flavour (Nadeem and Imran, 2016); clarity and purity; storage stability; and oxidative stability, which is advantageous for biofuel and industrial lubricant applications (Oluduro, 2012; Rashid *et al*., 2008). It also serves as a good moisturiser highly effective for the epidermis, due to its significant levels of oleic, behenic, and linoleic acids (Ana Clara Sans Salomão Brunow Ventura *et al*., 2021**;** Tsaknis *et al*., 1998). As such, it is a viable alternative or complementary component in pharmaceutical, cosmetic, and food applications.

3.1.1 Geographical Variations in *Moringa spp*. Seed oil Composition

According to researches on variations in seed composition across various geographical regions, there is a relationship between the genetic diversity among Moringa species and their adaptation to local environmental conditions (Foild *et al*., 2001). Anwar & Bhanger, 2003; Leone *et al*., 2016, report that the composition of *Moringa* seed oil is strongly shaped by environmental factors such as temperature, soil type, and rainfall, which influence both oil yield and fatty acid profiles For example, soils richer in organic matter or minerals like zinc and magnesium may enhance oil quality and antioxidant content (Saini *et al*., 2024). Soil type, environment and soil health influences crop productivity (Xing,Wang and Mustafa, 2025). Likewise, fluctuations in rainfall can impact the growing conditions of *Moringa*, particularly under drought stress, which can shift the fatty acid balance often increasing oleic acid content to improve oil oxidative stability (Tsaknis *et al*., 1999).

M. oleifera and *M. stenopetala* are drought-tolerant, multipurpose trees valued for nutrition and seed oil. *M. oleifera* seed oil called “ben oil” is naturally high in oleic acid and

tocopherols and resists oxidative rancidity, traits attractive for edible and cosmetic uses and for breeding stable, health-oriented oils. *M. stenopetala*, endemic to East Africa, is likewise prized for food security in arid regions, making comparative analysis across oils from diverse accessions and species is relevant to both adaptation and utilization. Therefore, sampling 48 geographically and climatically distinct locations including Nigeria, Oman, the Philippines, Kenya, India, Mali, Burkina Faso, and samples provided by commercial companies was considered because metabolite profiles in seeds are shaped by both genotype and environment.

The concentration and composition of seed oil is usually influenced by both genetic predisposition and various environmental influences including, temperature, soil type, rainfall and other broad climate conditions according to Velasco *et al*., 2014. These factors are known to alter tocopherols and sterols during seed development, while local soils and farming practices can further influence the composition as seen in sunflower seeds (Roche *et al*., 2010; 2009). Fatty acids and sterol contents at harvest were increased in Sunflower oil under water stress and high temperature (Ferfuia *et al*., 2015). In soybean plant likewise, both genetics and planting location significantly affected total phytosterol concentrations up to 2.5-fold, greater percentage of campesterol and lower percent stigmasterol and β-sitosterol were observed at higher temperatures (Vlahakis and Hazebroek, 2000).

Seeds grown in warmer climates generally exhibit higher levels of saturated fatty acids, while those cultivated in cooler climates tend to have higher levels of polyunsaturated fatty acids. The unsaturation index of Moringa oils is notably influenced by temperatures in tropical and subtropical regions. Oleic acid levels fluctuate, indicating that the tree possesses an adaptation mechanism that enhances the stability of its seed oil across various climates. Wiltshire *et al*., 2021 examined the effect of seasonality on the properties of *M. oleifera* seed oil and found that dry-season oils had lower acidity and moisture content than the rainy season oils and that the dry season oils would be more suitable for application as feedstocks in the cosmetics, food, medicinal and pharmaceutical industries.

The health benefits of these oils are significantly influenced by genetic and environmental factors, as evidenced by regional variability (Anwar *et al*., 2005; Olson, 2002). Thus, mapping biochemical variation against environmental conditions is necessary to separate genetic effects from environmentally driven changes and to identify climate - linked genetic variants that are useful for breeding.

Tsaknis *et al*. (1999) demonstrated that various species and geographical regions of Moringa exhibit distinct compositions of fatty acids, phytosterols, and tocopherols in their seeds. Tropically cultivated *M. oleifera* may exhibit distinct fatty acid profiles compared to

its subtropical or temperate counterparts (Makkar and Backer, 1997). Ayerza 2019, reports high levels of monounsaturated fatty acids (MUFA) such as palmitoleic, oleic and gadoleic acids up to 77.17. 79.27 and 77.83 % respectively in varied arid locations.

Sarin *et al*. (2010) report that *M. peregrina* seeds exhibit significantly higher oil content and oleic acid concentrations than *M. oleifera*. Research by Tsaknis *et al*. (1999) indicates that *M. oleifera* seeds contain a higher tocopherol content compared to *M. peregrina* seeds; however, the concentration of phytosterols in Moringa seeds is influenced by both species and geographical origin. The total amounts of unsaturated and saturated fatty acids can be influenced by the germination process and various treatments applied to Moringa seeds (Abdulkarim *et al*., 2005) for instance seeds collected from temperate regions of Pakistan exhibit a high oleic acid concentration of 78.59% and an oil content between 38.00% and 42.00%.

3.1.2 Fatty Acid Composition

Moringa oil has a distinctive fatty acid composition derived from *M. oleifera* and its related species (Somia *et al*., 2024). Seed-oil fatty acids set the nutritional profile and oxidative stability of Moringa oil and, indirectly, reflect lipid metabolism relevant to stress adaptation. *Moringa oleifera* is naturally high-oleic (70–79% C18:1) with moderate saturated fatty acids (palmitic(C16:0) 6%, stearic (C18:0)5%) and notable very-long-chain behenic (C22:0) 7% (Cervera-Chiner *et al*., 2024; Ogunsina *et al*.,2014), a composition associated with brilliant oxidative or thermal stability and long shelf life, which are desirable targets for quality breeding and end-use (edible/cosmetic) applications. Accordingly, oleic acid (C18:1) is the unique trait, while linoleic C18:2 and α-linolenic C18:3 are known for stability and nutrition; behenic (C22:0) valued for its smoothing, lubricating and solvent evaporating retarder nature. In comparison, Olive oil contains 55-83% oleic acid, 7.5-20% palmitic acid, approximately 0.5-5% stearic acid, and 3.5-21% linoleic acid (Visioli *et al*., 2019). While both olive oil and Moringa oil contain significant amounts of heart-healthy oleic acid, Moringa oil has considerably lower levels of saturated fatty acids such as palmitic, stearic, and behenic acids, potentially offering health advantages.

Lipids generally serve as the primary storage form of energy and carbon source in plants, stored in the form of triglycerides (TAG) in oil body, (Baud and Lepiniec, 2009). Biochemically, C18:0- ACP (Acyl Carrier Protein) is desaturated by SAD (Stearoyl-ACP desaturase) to 18:1-ACP; in the Endoplasmic Reticulum, FAD2 (Fatty acid desaturase) converts 18:1→18:2 and FAD3 converts 18:2→18:3, whereas FAE-KCS (Fatty acid elongases - β-ketoacyl-CoA synthase) elongates C18 acyl-CoAs to VLCFAs (Very Long Fatty Acids) such as C22:0 Behenic acid; storage TAGs assemble via the Kennedy pathway (GPAT→LPAAT→PAP→DGAT) (Cai,Yu and Hu, 2025). Hence, the usual way to

boost oleic acid in seed oils is to lower FAD2 activity, while the total amount of oil the seed stores is mostly determined by DGAT and PDAT enzymes (Cai, Yu and Hu, 2025). Altogether, natural and engineered variation at these steps offers tractable markers and alleles for breeding oil quality and climate-resilient germplasm.

Keywords: Enzymes involved in TAG synthesis / Kennedy pathway glycerol-3-phosphate acyltransferase (GPAT), lysophosphatidic acid acyltransferase (LPAAT), diacylglycerol acyltransferase (DGAT), phospholipid:diacylglycerol acyltransferase (PDAT), phosphatidic acid phosphatase (PAP).

3.1.3 Phytosterols in Plant Adaptation

Plant sterols are made step wisely through a biosynthetic pathway, where β-Sitosterol is the end product sterol in most plants and most abundant in Moringa seed oil (Yang *et al.*, 2019). Δ^7-stigmastenol and Δ^5-avenasterol are intermediate sterol in the pathway. Δ^7-stigmastenol is a precursor sterol that occurs earlier in the pathway, before conversion into Δ^5-avenasterol and then more stable end sterols like β-Sitosterol or stigmasterol (Yang *et al.*,2007. By measuring the relative amounts of these sterols, one can readily depict how efficiently the pathway is flowing towards β-Sitosterol and whether environmental or genetic factors are shifting the balance at intermediate steps. Phytosterols regulate membrane fluidity in plant as well as formation of lipid rafts or microdomains (Valitova, *et al.*, 2016) which are important for signal transduction and stress response. There is generally a trigger in sterol composition changes when a plant is under stress like salt stress. In Brassica for instance sitosterol level decreases and stigmasterol reduces in roots under saline conditions (Shazad *et al.*, 2022), sterol ratios also increase under ultra violet UV-B exposure as seen in grapevine to reduce UV damage to the plant (Valitova, *et al.*, 2024). Production of sterols in plants are indicators of stress adaptation. They are also connected to important metabolic pathways and form precursors for plant hormones, with their ratios revealing the degree of metabolic regulation.

Moringa oil contains a high concentration of phytosterols (150-200mg per 100g), including β-sitosterol, and stigmasterol. In 100 grams of olive oil, there are 100 –250 milligrams of phytosterols. Soybean oil contains 300 – 400mg per 100g of total phytosterols, while sunflower oil contains 250–450mg per 100g of total phytosterols. Both edible oils, Sunflower oil and Soybean oil have elevated total phytosterol levels. One of the most significant health benefits of phytosterol is their capacity to reduce cholesterol levels by inhibiting the intestinal absorption of dietary cholesterol.

This study focusses on these specific phytosterols, β-Sitosterol, Δ^5-avenasterol, and Δ^7-stigmastenol because they are particularly important for controlling how cell membranes behave, protecting cells from damage, and helping plants adapt to stress.

β-Sitosterol enhances antioxidant defenses, stabilizes membranes under drought and temperature stress, and helping plants tolerate adverse environments as seen in rice (SujithKumar *et al.*, 2015) while dynamic shifts in sterol composition, including Δ^5-avenasterol and Δ^7-stigmastenol reflect underlying metabolic regulation and adaptation strategies. Focusing on these compounds provides both insight and breeding-relevant markers in Moringa, guiding selection for seed oil quality and resilience across diverse environments. Identifying genetic variants that influence these sterol levels allows plant breeders to select lines with improved membrane stability, metabolic robustness, and stress tolerance, which are all valuable for seed oil quality, oxidative stability and resilience. High β-sitosterol and Δ^5-avenasterol are important for both shelf-life and nutritional value.

While direct functional studies on Δ^5-Avenasterol in stress adaptation are limited, it is a precursor within the branched sterol biosynthesis pathway. Changing how this pathway works can alter the amounts of important sterols like sitosterol and stigmasterol, which help plants cope with stress and keep their cell membranes stable. Their levels therefore act as useful markers to show when the pathway has shifted and how that affects the plant's sterol balance and overall stability.

Δ^7-stigmastenol is ordinarily found in trace amounts and appears during sterol metabolism shifts serving as an indicator of sterol pathway adjustments under environmental conditions.

The relative abundance and variation of these sterols across accessions therefore provide biochemical markers that integrate genetic background and environmental influence. In the context of GWAS, coupling sterol profiling with dense SNP datasets and geographic information enables the identification of alleles that control sterol biosynthesis and composition, and tests whether certain variants are enriched in particular climates. This approach not only uncovers the genetic basis of oil quality traits (e.g. high β-sitosterol or oleic acid content) but also provides a framework for selecting climate-resilient genotypes, aligning seed oil improvement with both nutritional and ecological goals.

3.1.4 Tocopherols in Moringa seed oil

Tocopherols, a category of fat-soluble antioxidants, are classified into alpha α-, Beta β-, Gamma γ-, and Delta δ-isomers based on the number and configuration of methyl groups present on the ring structure. The synthesis of tocopherol primarily takes place in plants and is typically found in plant oils, seeds, nuts, and green leafy vegetables. Tocopherols (vitamin E compounds) are determinants of both the nutritional and ecological value of Moringa seed oil. They act as lipid-phase antioxidants that protect polyunsaturated fatty acids from peroxidation, thereby enhancing seed longevity, oil stability, and germination

success (Sattler *et al*., 2004). In plant tissues, α-tocopherol plays a central role in mitigating oxidative stress under drought, salinity, and high-light conditions, making it a key factor in ecological adaptation (Shah *et al*., 2021; Faizan *et al*., 2023). Among the four major isomers (α, β, γ, δ), α-tocopherol is nutritionally most relevant due to its highest vitamin E activity, while γ- and δ-tocopherols serve as direct biochemical precursors to α- and β-tocopherol, respectively, via the action of γ-tocopherol methyltransferase (VTE4) (Falk *et al*., 2010). The pathway begins with homogentisate (HGA) and phytyl diphosphate (PDP) condensation by homogentisate phytyltransferase (HPT or VTE2), followed by methylation (MPBQ Methyltransferase VTE3) and cyclization (Tocopherol cyclase/VTE1) to yield δ- and γ-tocopherols, which are subsequently converted to β- and α-tocopherols by VTE4 (Fritsche *et al*., 2017). γ- and δ-tocopherol are direct precursors to α- and β-tocopherol, respectively, and VTE4 controls that final conversion.

In Moringa, both α- and γ-tocopherols are abundant in seed oil, and much more than in sesame seed oils (Matthaus and Özcan, 2018), making them promising targets for breeding strategies aimed at improving oil stability and nutritional value. Mapping natural variation in tocopherol composition across environments can thus reveal alleles underlying antioxidant capacity and identify genotypes adapted to oxidative stress while delivering superior oil quality.

The tocopherol concentration in Moringa oil varies, *Moringa oleifera* contains 34-515 mg/kg of total tocopherols and *Moringa ovalifolia* contains 44.56 mg per 100mg including 33.94mg of alpha tocopherol, 6.64mg of beta tocopherol and 3.98mg of delta tocopherol according to Cheikhyoussef *et al*., 2020; Cheikhyoussef *et al*., 2018 respectively; according to Hanaa *et al*., nd.. *Moringa peregrinna* shows tocopherol levels of 20.35 mg/100g, while Tsaknis,1998 reports 145mg/kg for alpha tocopherol, 58mg/kg for gamma tocopherol and 66mg /kg for delta tocopherol. Olive oil demonstrates a greater tocopherol concentration, spanning from 100 to 250 mg per 100g. Sunflower oil contains a significant amount of α-tocopherol, with concentrations between 400 and 700 mg per 100g, whilst Canola oil has levels ranging from 150 to 300 mg per 100g. Moringa oil possesses lower overall tocopherol levels compared to Olive and Sunflower oils; nonetheless, its diverse tocopherol profile, especially the γ-tocopherol content, provides significant antioxidant characteristics that augment its health advantages.

3.1.5 Genetic Variability and Genome-Wide Association Studies (GWAS)

Genome-wide association studies (GWAS) provide a robust statistical framework for identifying genetic variants linked to phenotypic traits (Alseekh *et al*., 2021), even in non-model species such as *Moringa spp*., where targeted mutants and reference genetic resources are limited. By scanning millions of single nucleotide polymorphisms (SNPs)

distributed across the genome, GWAS can localize genomic regions, sometimes down to specific genes whose variation alters gene activity and ultimately influences trait expression (Chen *et al.*, 2021). GWAS also helps in understanding how to capitalize on how natural variations contribute to phenotypic diversity and adaptation in different environmental conditions (TibbsCortes *et al.*, 2021).

3.1.6 Problem statement, Aim and Objectives of the Study

Problem statement

While the seed oil of *Moringa oleifera* has been extensively explored for its major fatty acid composition and associated health benefits, significant gaps persist in understanding how regional variation influences oil yield and quality. Although oleic, behenic, and erucic acids are well characterized, comprehensive profiling of minor yet bioactive constituents including sterols, tocopherols, and polyphenols remains limited, despite their potential roles in regulating health effects and contributing to biological activity (Tsaknis *et al.*, 1999; Sánchez-Machado *et al.*, 2010). Moreover, there is a dearth of research into the long-term stability and degradation kinetics of Moringa oil under varying storage and environmental conditions, which are essential for developing optimal processing, packaging, and shelf-life strategies (Joshi and Pandit, 2023). Although studies suggest anti-inflammatory and antioxidant properties, the underlying molecular mechanisms remain largely speculative, necessitating more rigorous in vitro and in vivo investigations to substantiate therapeutic claims and clarify safe and effective dosages (Vergara-Jimenez *et al.*, 2017). Available research has not consisdered how minor compounds might interact synergistically with major fatty acids, an area with important implications for understanding oil efficacy and guiding plant breeding programmes aiming to enhance nutritional and functional properties. Compounding these knowledge gaps is the absence of standardized, eco-friendly extraction methods that could minimize environmental impact while preserving oil quality and yield (Abdulkarim *et al.*, 2005). A significant limitation in available literature on Moringa seed oil researches is the narrow geographic scope of most studies, which often focus on isolated regions without representing the diverse climates, soils, and cultivation practices across Moringa's native and introduced ranges (Lalas & Tsaknis, 2002; Leone *et al.*, 2016). This restricts the ability to generalize findings and develop region-specific breeding strategies. Compounding this issue are inconsistent methodologies for extraction, analysis, and reporting, which hinder cross-study comparisons and meta-analyses. Moreover, many studies fail to clearly identify the Moringa cultivar examined, despite documented variations in oil composition among genotypes, leading to confound results (Leone *et al.*, 2016). Equally unexplored is the influence of geographical region climate such as rainfall, temperature, soil nutrients, and

altitude on oil yield and composition, even though these variables are known to affect plant secondary metabolism (Azad *et al*., 2015).

To bridge these knowledge gaps, there is need for holistic data integration from diverse geographic locations, examining the effects of geographic locations on oil traits, investigations of genotype-environment interactions, and latitudinal assessments across multiple growing seasons. This would inform optimal breeding strategies for different environments and enable the use of oil composition profiles as biomarkers for selecting superior cultivars, ultimately supporting the development of regionally adapted, high-quality Moringa varieties.

Aim

This study therefore addressed 3 questions based on the dearth of current studies on Moringa seed oil and Genotype-Environment influence on oil yield and oil profile components. These include, does environmental factors/variables like altitude/latitude, rainfall or temperature influence the fatty acid, tocopherol and phytosterol profile of Moringa seed oil across diverse geographical regions? what are the genotype - environment interactions affecting oil yield and composition among *M. oleifera and M. stenopetala* accessions across various ecological zones? and, what patterns emerge from the studies of variations in fatty acid, phytosterol, and tocopherol compositions of oil extracted from Moringa seeds *(M. oleifera and M. stenopetala)* sourced from diverse geographical locations? Understanding these influences would guide region-specific cultivation for desired oil traits, identifying the genotype - environment interactions would support breeding programmes targeted at oil quality and stability and understanding the patterns in the variations of the diverse Moringa seed oil would inform both scientific and commercial strategies.

Objectives

Here, I will explore the biochemical diversity of Moringa seed oil, focusing on how geographical origin influences its composition. By examining seeds from *Moringa oleifera* and *Moringa stenopetala* across diverse locations, this study will uncover variations in their fatty acid, phytosterol, and tocopherol profiles. Moringa oil is renowned for its high oleic acid content, robust antioxidant properties, and versatility in food, cosmetic, and industrial applications. However, the role of environmental and genetic factors in shaping its composition remains under-explored. This work seeks to bridge that gap by investigating these key objectives. First, I will analyse how fatty acid, phytosterol, and

tocopherol composition varies across geographically sourced seeds, emphasizing differences in oleic, palmitic, and behenic acids which are critical contributors to the oil's nutritional and functional properties. Through these analyses, I aim to determine whether certain locations consistently promote higher levels of health-beneficial unsaturated fatty acids. Thirdly, I will evaluate whether the geographical origin of Moringa seeds significantly impacts the concentrations of fatty acids, phytosterols and tocopherols. These compounds are important for the oil's health-promoting and antioxidative attributes. By assessing their levels in seeds from diverse climates and regions, this study will establish the potential of specific Moringa accessions as sources of nutritionally rich or industrially valuable oils. The correlations between geographical origin, genetics and oil composition will reveal adaptive responses that Moringa species exhibit to environmental pressures. This study will also bridge the gap of lack of comparative biochemical profiling of Moringa seed oil across diverse geographical regions by leveraging plant science, crop breeding, phytochemistry and genomics standards to characterise biochemical variability in 48 Moringa seed oils, to identify marker compounds with breeding potential, and to link biochemical traits with geography and genetics.

These insights will inform breeding strategies and identify the Moringa accessions with optimum oil metabolic profiles for specific applications. Through these objectives, I will demonstrate the intricate interplay between geography, genetics, and biochemistry in defining Moringa oil's quality. This work will not only provide a deeper understanding of Moringa's potential as a sustainable oil source but also establish a framework for selecting or breeding superior accessions tailored to dietary, pharmaceutical, and industrial needs. By combining detailed compositional analyses with a focus on geographic variability, this study aims to unlock Moringa's full potential as a globally significant crop. This study highlights the impact of geographical variations and genetics on Moringa oil composition whilst leveraging its high unsaturated fatty acid components, sterol composition, and tocopherols.

3.2 Materials and Method

3.2.1 Sampling strategy and plant material

Seed samples (n = 48) were assembled from 48 geographic locations as follows: 33 samples from northern and southern Nigeria, 8 from Kenya, 3 purchased via eBay (UK), 2 from Mali, and 1 each from the Philippines, India and Burkina Faso. Collections prioritized mature, verified seed lots from single mother trees where possible, and sampling records include precise GPS coordinates and collector notes.

The seed oil metabolic traits of 48 Moringa accessions were obtained from Sub-Saharan Africa (Nigeria, Mali, Burkina Faso), East Africa (Kenya), Southeast Asia (Philippines), South Asia (India), Southwest Asia (Oman), and the commercial platform eBay (UK). The 33 Nigerian accessions are from the Southwest (Oyo, Osun, Ekiti and Ogun); North central (Kwara and Plateau) and Northwest (Katsina). The accessions include three distinct species: *Moringa oleifera*, *Moringa stenopetala*, and *Moringa peregrinna*. Ensuring diverse environmental conditions.

The climate of the locations where the seed samples were obtained gives insight on how they influence Moringa seed development and seed oil yield. For instance, Kenya climate varies from equatorial in the west to semi-arid in the north and east, with temperature ranging from 80°F (27°C) in July to 90°F (32°C) in October and February. The country experiences 2 rainfall patterns, long rains from March to May and Short rains from October to December. The dry season runs from May to October.

Mali is a semi-arid region with tropical climate. Temperature here ranges from 104°F to 113°F (40-45°C). Its long dry season runs from October to May and has a short rainy season between June and September with erratic rainfall of 300mm-800mm per year.

Burkina Faso is a tropical savanna and semi-arid area which experiences hot year-round with a distinct dry season between November and May, and a wet season between June and October. Annual rainfall is approximately 600 - 1,200 mm/year, decreasing northward. The climate in Oman is arid, it is a hot desert, with temperature of 40 °C in the hot summers. Rainfall is low, about 100mm/year occurring mostly in January to March. Has high humidity near the coast and dry inland.

India has four main seasons with strong seasonal variation and a predominantly tropical monsoon climate. It experiences hot summers around March to June, wet monsoon between June and September and mild winters between December and February. Generally, annual rainfall is 800-2,000mm/year.

Philippines is generally hot and humid all year round. It has a tropical maritime climate, with distinct dry season from December to May and a wet season which runs from June to November. Experiences very high annual rainfall of 2,000 to 4,000mm.

Nigeria's climate varies by region, usually the rainy season is longer in the south than in the north, and rainfall decreases from the coast inland. The Western Nigeria has tropical rainforest (humid) to savanna (sub-humid) with long rainy season from March to October and short dry season from November to February. Annual rainfall is 1,200 to 2,000 mm/year. While the Northern Nigeria is semi-arid, experiencing long dry season with intense heat and high radiation and an annual rainfall of 500 - 1,000mm/year.

Within the Southwest region, there exists a variation in climate ranging from derived savannah (Oyo, Osun, Ekiti) to humid, forest climate in Ogun. Each having distinct climatic features. Oyo has an equatorial climate with dry season from November to March, and wet season from April to October with average daily temp ranging from 77℉ (25℃) to 95℉ (35℃). Osun has a tropical savanna climate with an annual temperature of 64℉ (17.8℃) and about 596 inches of rainfall. The state is dry for 59 days a year. Ekiti has a tropical climate with a rainy season from April to October and a dry season from November to March with temperature ranging from 21℃ to 28 ℃. Ogun has a tropical wet and dry climate with average annual temp being 84.81℉ (29.34℃) and receives annual rainfall of 5.57 inches with 224.18 rainy days in a year.

Seed handling and storage prior to biochemical analysis

Seeds collected (Figure 3.1 and Table 3.1) were clean, debris-free, air-dried. They were stored in paper-polythene bags at room temperature in the Biology Research Laboratory, University of York, UK, for approximately one year before the oil extraction process.

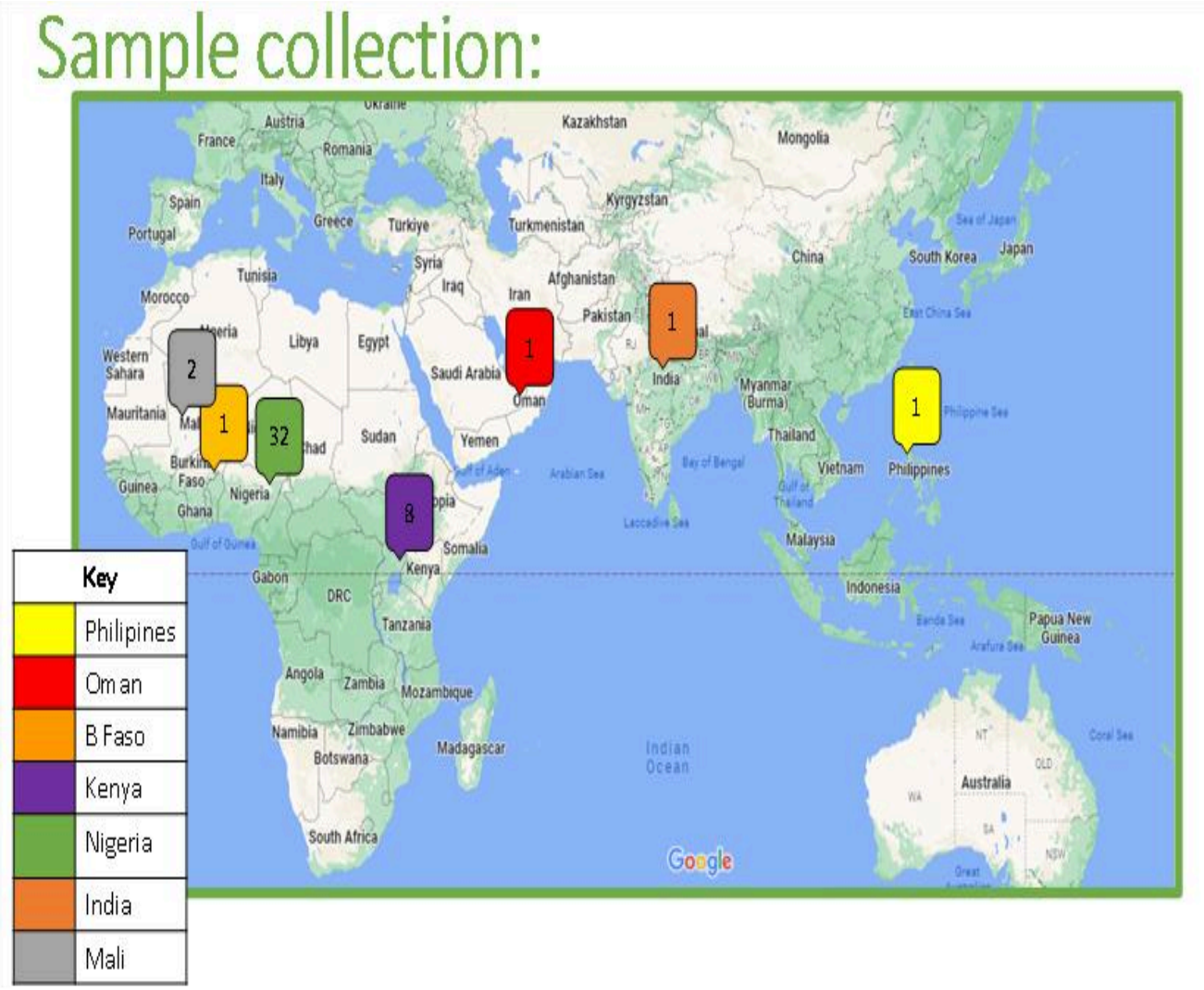


Figure 3.1 Map of Collection sites of Moringa seed samples

Collection note: Sources of seed collection include the commercial samples gotten from WOT Natural, Glasgow, UK (WOT, R1), Supreme products, London, UK (SAJINA, R3), BeeKind ReImagined, UK (BK, R2). The 8 Kenya, 2 Mali, 1 India, and 1 Philippines samples were gotten from ICRAF, Nairobi, Kenya, while the 33 Nigerian seed samples were gotten from Covenant University, Ota, Nigeria by Popoola and Obembe. Their collection methodology has been recorded in previous research (Popoola, *et al.*, 2014; Popoola, 2016; Popoola and Obembe, 2013).

Table 3.1 Collection Areas of the 48 Accessions of Moringa used for this study. All are field samples except R1, R2 and R3, which are commercially obtained from Ebay, UK (n=48). Coordinates obtained from Google Maps and rounded to match measurement accuracy.

S/N	Coordinates	Accession Number	Sample ID	Species	Latitude	Longitude	Collection Area
1	Point (3.530,8.207)	Mo1	Mo1	*M. oleifera*	8.207	3.530	Okaka,Oyo, NIG
2	Point (4.229,8.124)	Mo2	Mo2	*M. oleifera*	8.124	4.229	Ijeru Baptist, Ogbomoso, NIG
3	Point (3.383 ,8.666)	Mo3	Mo3	*M. oleifera*	8.666	3.383	Oke-Oro, Saki-West, NIG
4	Point (3.953,7.530)	Mo4	Mo4	*M. oleifera*	7.530	3.953	Aroro-Iroju road, Akinyele, Ibadan, NIG
5	Point (3.423,7.398)	Mo5	Mo5	*M. oleifera*	7.398	3.423	Abeokuta-Iseyin road, Eruwa, NIG
6	Point (4.187,7.372)	Mo10	Mo10	*M. oleifera*	7.372	4.187	Ikire,along Ife road, Osun NIG
7	Point (4.595,7.157)	Mo11	Mo11	*M. oleifera*	7.157	4.595	Abeere,Ede,Osun NIG
8	Point (4.659,7.934)	Mo12	Mo12	*M. oleifera*	7.934	4.659	Steel rolling Ikirun, Osun, NIG
9	Point (4.738, 8.185)	Mo16	Mo16	*M. oleifera*	8.185	4.738	Ijagbo,Offa, Kwara NIG
10	Point (4.816, 8.233)	Mo17	Mo17	*M. oleifera*	8.233	4.816	BHS,Ajase-Ipo, Kwara NIG
11	Point (4.609, 8.418)	Mo18	Mo18	*M. oleifera*	8.418	4.609	Ganmo,Kwara, NIG
12	Point (4.168, 8.247)	Mo22	Mo22	*M. oleifera*	8.247	4.168	Flora,Tanke, Iledu, Kwara, NIG
13	Point (4.672, 8.147)	Mo23	Mo23	*M. oleifera*	8.147	4.672	University of Ilorin Kwara NIG
14	Point (5.132, 8.0812)	Mo25	Mo25	*M. oleifera*	7.519	5.223	Kajola, Ikere- Ekiti, Ekiti NIG
15	Point (5.234, 7.613)	Mo26	Mo26	*M. oleifera*	7.613	5.234	Oye-Ekiti, Ekiti NIG
16	Point (5.068,7.912)	Mo27	Mo27	*M. oleifera*	7.912	5.068	Ijero-Ekiti, Ekiti NIG
17	Point (5.158,7.672)	Mo28	Mo28	*M. oleifera*	7.672	5.158	Iyin-Ekiti, Ekiti NIG
18	Point (5.340 ,7.833)	Mo29	Mo29	*M. oleifera*	7.833	5.340	Ado-Ekiti Ekiti NIG
19	Point (3.439, 7.181)	Mo31	Mo31	*M. oleifera*	7.181	3.439	FUNAAB, Ogun NIG
20	Point (3.892, 7.446)	MO51	MO51	*M. oleifera*	7.446	3.892	Kwara NIG
21	Point (4.536, 8.478)	MO52	MO52	*M. oleifera*	8.478	4.536	Kwara NIG
22	Point (4.476, 8.505)	MO53	MO53	*M. oleifera*	8.505	4.476	Kwara NIG
23	Point (4.716, 8.583)	MO55	MO55	*M. oleifera*	8.583	4.716	Kwara NIG
24	Point (8.898, 9.898)	MO56	MO56	*M. oleifera*	9.898	8.898	Plateau NIG
25	Point (8.801, 8.650)	MO57	MO57	*M. oleifera*	8.650	8.801	Plateau NIG
26	Point (11.423,11.361)	MO58	MO58	*M. oleifera*	11.36	11.423	Plateau NIG
27	Point (9.516, 8.872)	MO59	MO59	*M. oleifera*	8.872	9.516	Plateau NIG
28	Point (9.738,10.858)	MO60	MO60	*M. oleifera*	10.858	9.738	Plateau NIG
29	Point (7.501,12.590)	MO61	MO61	*M. oleifera*	12.590	7.501	Katsina NIG
30	Point (7.055,11.704)	MO62	MO62	*M. oleifera*	11.704	7.055	Katsina NIG
31	Point (7.346, 12.711)	MO64	MO64	*M. oleifera*	12.711	7.346	Katsina NIG
32	Point (7.496, 12.592)	MO65	MO65	*M. oleifera*	12.592	7.496	Katsina NIG

33	Point (-1.041, 13.068)	Mo70	Mo70	*M. oleifera*	13.068	-1.041	Burkina Faso
34	Point (37.618, 0.374)	ICRAF08214	KMs10	*M. stenopetala*	0.374	37.618	Isolo, Kenya
35	Point (56.295,20.614)	OmMp	OmMp	*M. peregrina*	20.614	56.295	Oman
36	Point (35.579, 0.346)	ICRAF08392	KMs8	*M. stenopetala*	0.346	35.579	Isolo, Kenya
37	Point (37.593, 0.351)	ICRAF08212	KMs5	*M. stenopetala*	0.351	37.593	Isolo, Kenya
38	Point (36.883, 1.205)	ICRAF07842	KMo2	*M. oleifera*	1.205	36.883	Nairobi, Kenya
39	Point (34.099, 0.461)	ICRAF05536	KMo4	*M. oleifera*	0.461	34.099	Busia, Kenya
40	Point (40.819, 2.181)	ICRAF08446	KMo7	*M. oleifera*	-2.181	40.819	Hindi- Lamu, Kenya
41	Point (36.822, 1.292)	ICRAF08403	KMo6	*M. oleifera*	-1.292	36.822	Ramogi, Kenya
42	Point (37.572, 0.394)	ICRAF08219	KMs11	*M. stenopetala*	0.394	37.572	Isolo, Kenya
43	Point (-6.246, 13.433)	ICRAF07633	MaMo3	*M. oleifera*	13.433	-6.246	Segou, Mali
44	Point (-7.961 ,12.589)	ICRAF07632	MaMo1	*M. oleifera*	12.589	-7.961	Niamakoro, Mali
45	Point (76.014, 31.488)	ICRAF07633	IMo9	*M. oleifera*	31.488	76.014	India
46	Point (123.440, 12.288)	ICRAF08029	PMo12	*M. oleifera*	12.288	123.440	Philippines
47	-	R1	R1	*M. oleifera*	-	-	Ebay ,UK -WOTNat
48	-	R2	R2	*M. oleifera*	-	-	Ebay, UK - BeeKind
49	Point (90.391,23.870)	R3	R3	*M. oleifera*	23.870	90.391	Ebay, UK -Sajina, Bangladesh, India.
50	Sample used for RNASeq (Drought) experiment	Mantig	Mantig	*M. oleifera*	-	-	Ebay, UK

3.2.2 Extraction of seed oil and Biochemical Profiling

Oil extraction (bulk oil for FAMEs, sterols and tocopherols)

Seed oil extraction was carried out in the Frederic Beaudoin Laboratory, Rothamsted Research, Hertfordshire, UK using the saponification/derivatisation with TMS and quantification by GC-FID method. (Fig.3.2). Subsamples were ground to fine powder in liquid nitrogen prior to extraction. Antioxidant butylated hydroxytoluene (BHT, 0.01% w/v) was included in solvents to minimize oxidative degradation during processing (Broughton & Beaudoin, 2021). The summary of the process is as follows: seed preparation, extraction, derivatization and saponification, instrument conditions and calibration /quantification.

Seed preparation

80mg of Moringa seed kernels were weighed and ground with ceramic beads in grinding tubes in Precellys 24 tissue homogenizer (Bertin Instruments, France) and followed subsequent processes as per Broughton & Beaudoin, 2021 protocol. The choice of seed

size of 80mg was based on previous pilot studies and considering the size of the tubes to be used for grinding. Grinding was performed at 5500rpm in three 25-second cycles, with a 30-second pause between each cycle. The reinforced tubes (CK28-R) allow grinding of hard samples like Moringa.

Extraction

1 ml of hexane solution containing the internal standards (50 µM cholesteryl heptadecanoate (17:0-cholesterol) and cholestanol; Avanti Polar Lipids, USA) was added to each sample. The hexane, n-hexane is of HPLC grade. Samples were vortexed and incubated at 30℃ in a rotary incubator for 1 hour. The samples were centrifuged in Centrifugal evaporator (Genevac, UK) at 20,000x for 1 minute to sediment seed material and the hexane supernatant was transferred to 3.5 ml screw top glass vials (Trident, SLS UK) using glass Pasteur pipettes. The extraction was repeated twice with 1 ml of pure hexane omitting the incubation step. The three hexane extracts were pooled in 3.5 ml glass vials and the lipid extracts were dried using a centrifugal evaporator keeping the temperature below 40℃, or alternatively under a stream of nitrogen. Lipids were re-suspended in 1 ml of hexane. Samples were stored at 20℃ for further analysis. All metabolites were analysed from the same total seed oil extracts adjusted to 50mg/ml in hexane. 60 µl was used for Fatty Acid Methyl Esters (FAMEs), 20 µl for tocopherols and 300 µl for phytosterols.

3.2.2.1 Fatty Acid Methyl Esters (FAMEs)

Gas chromatography (GC) (Winkler-Moser, 2011) was performed to analyse the fatty acid methyl esters (FAMEs) in the extracted oils to determine the fatty acid composition. Fatty acids were analysed by gas chromatography and flame ionization detector (GC-FID) after derivatisation into Fatty Acid Methyl Esters (FAMEs) which is the standard method (see protocol in Appendix, (Broughton *et al.,* 2018)). The process is as follows: 250 µl aliquot of the lipid extract was placed into a screw top glass tube and dried as described above (Fig.3.1). 1.5 ml of 1.5M potassium hydroxide in 95% (v/v) ethanol, was added and the tubes closed tightly with screw caps and incubated in a heating block or water bath at 80 for 1 hour. The samples were made to cool and 1 ml of a 1% (w/v) NaCl solution was added followed by 1 ml of hexane (n-hexane, HPLC grade), vortexed and centrifuged at 1000x for 1 minute to accelerate phase partitioning. The upper hexane phase was transferred to a glass vial. The extraction was repeated with 1 ml of hexane and both extracts were pooled. The samples were dried as described above.

3.2.2.2 Phytosterol

The methodology utilized saponification/derivatisation with TMS, succeeded by quantification through GC-FID (Fig.3.2 Broughton & Beaudoin, 2021). The overall sterol

content, which includes hydrolysed steryl esters (SE) and free sterols (FS), was measured as a whole, instead of evaluating the distinct compositions of SE and FS independently.

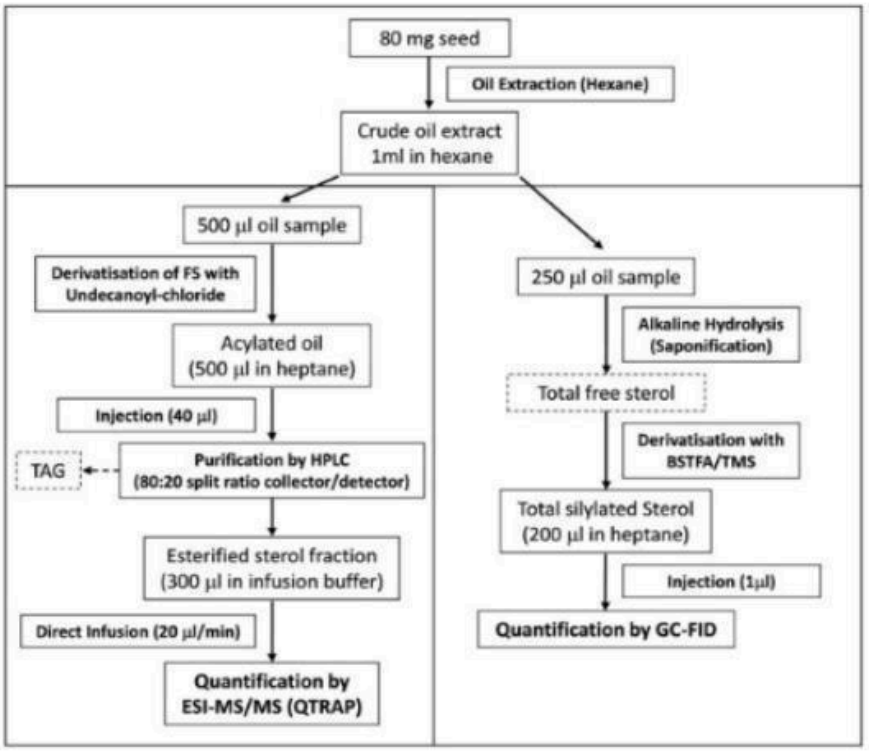


Figure 3.2 The flow chart for seed oil sterol quantification according to (Broughton & Beaudoin, 2021)

The left part shows the derivatization of the free sterols with undecanoyl chloride, purification of steryl esters by HPLC, and quantification of steryl esters by ESI-MS/MS (QTRAP). The right part shows the total sterol measurement after saponification and silylation with BSTFA BY GC-FID (Broughton & Beaudoin, 2021)

The method outlined herein entails the extraction of free and esterified phytosterols from seed oil, necessitating the division of samples for both GCFID/MS and ESI-MS/MS analyses, as isomers such as stigmasterol and Δ^5-avenasterol, or β-sitosterol and Δ^7-stigmastenol , cannot be differentiated by ESI-MS/MS (Broughton & Beaudoin, 2021). The calibration curve employed is a single Point calibration response curve as only one concentration of standard (16:0-phytol) is present in each sample.

Sterols were derivatized by adding 100 µl of BSTFA–TMCS (99:1, v/v; BSTFA (N,O-bis(trimethylsilyl)trifluoroacetamide) + 1% TMCS (trimethylchlorosilane): BSTFA/TMCS 99:1, v/v; Merck, UK) and incubated for 1 hour at 85℃. Once samples cooled down, 100 µl of heptane was added and 1 µl was injected for GC-FID and/or GC-MS analysis as described by Broughton *et al.,* 2018. Typically, GC column settings are the same for both GC-FID and GC-MS with splitless injection, helium used as the

carrier gas at 2 ml/min and the inlet temperature set to 325℃. The oven initial temperature was set to 200 ℃, a ramp of 6.5 ℃/min was applied until 325℃ and held for 4 minutes. The FID and/or the MSD transfer line temperature was set to 325℃, the source and quadrupole also set to 230 and 150 ℃, respectively, and the scan range between 42 and 520 m/z. GCFID (6890, Agilent) and/or GC-MS (6890 N/5975B, Agilent). Internal standard mixture containing 23.4 μM of each cholesteryl ester (16:0, 18:0, 18:1, 18:2, 22:0, 22:1) and cholesterol (Merck) in anhydrous dichloromethane. External standard mixture containing 25 μM free sterols (-β-sitosterol, stigmasterol, cholesterol) (Merck, UK) and cholestanol (Avanti Polar Lipids, USA) in anhydrous dichloromethane (Broughton *et al.,* 2018).

3.2.2.3 Tocopherols

HPLC-based methods (Lampi, 2024) was utilized to assess tocopherols and tocotrienols extracted from Moringa seed oil. Tocopherols were quantified directly from lipid extracts diluted in hexane with 0.01% BHT. Tocopherols was included as an internal standard. Samples were analysed by normal phase HPLC using a LiChrospher Si 60 column and a fluorescence detector (FLD: λ Excitation = 294 nm; λ Emission = 325 nm). The column sustained a temperature of 22℃, employing a mobile phase of 5% (v/v) 1, 4-dioxane in hexane for isocratic elution. Tocopherol standards from Lampi, AOCS, 2024 was used in the process. Identification and quantification of α-, γ-, and δ-isomers were based on authentic standards and external calibration curves.Tocopherols were separated on a diol column (250 × 4 mm, 5 μm) with hexane:isopropanol mobile phase and quantified by fluorescence detection at Ex 295/Em 330 nm. Quantification of all compounds was performed using calibration curves generated from authentic standards, normalized to the relevant internal standard, and reported as $\mu g \cdot g^{-1}$ seed or % total lipid.

3.2.2.4 Extraction Efficiency

The solvent volume utilized in the extraction process is denoted by the volume of hexane in a 50 mg/ml stock. The extraction process's efficacy was evaluated by comparing the quantity of oil recovered to the volume of hexane used. The concentration of 50 milligrams of oil per millilitre of the stock solution is represented by the hexane volume of a 50 mg/ml stock. The extract concentration was recorded and utilized for subsequent analyses or applications that necessitated a specific concentration of Moringa seed oil. An understanding of the efficacy of the extraction process was obtained by comparing the average oil weight to the volume of hexane used. The quantity of hexane used was effective for oil extraction. Adjustments to the extraction process may be required if extraction efficiency falls below anticipated levels. The extraction technique's reliability

was demonstrated by the consistent mean oil weight and hexane volume across all experiments.

3.2.3 Statistical Analyses of Seed and Seed oil traits

Correlation analysis was performed on the seed weight, oil weight and oil% yield. The Seed oil traits (FAMEs.txt, Tocopherols.txt, and Phytosterols.txt) were subjected to statistical analyses. Firstly, the means of 3 replicates of each of the 48 samples were taken (FAMEs means.txt, Tocopherols means.txt, and Phytosterols means.txt) and means_tables generated Supp.Table 3.1; Supp.Table 3.2; Supp.Table 3.3; Supp.Table 3.4. . Then a t-test to compare the means of the fatty acid compositions (FAMEs, Tocopherols, and Phytosterols) between the *Moringa oleifera* and *Moringa stenopetala* groups were performed. ANOVA was also conducted to compare the means across all samples; Correlation analysis was performed to examine relationships between each of the fatty acid composition variables and a regression analysis was conducted to model the relationships among the samples in relation to their latitudes (location). In addition, A one-way ANOVA was conducted to determine whether the mean concentrations of individual oil components (Fatty Acids, Tocopherols, and Phytosterols) significantly differed across geographical locations. Where ANOVA was significant ($p < 0.05$), Tukey's HSD post-hoc test was used to pinpoint which specific locations differed, that is, To reveal geographical influence on specific oil constituent (Moringa_oil_master_no_KMs). Furthermore, Pearson correlation and linear regression was done to determine if geographic latitude or genetic variation(PC1) was of significant influence on the Moringa oil variability (PCA_sample_table; PCA_PC_summary_table)

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) was employed to examine the correlations among the diverse phenotypic characteristics of Moringa seeds and their oil properties, alongside the genotypic differences among different accessions.

The SNPRelate package in R was employed to do a PCA analysis of the variations, as outlined by Zheng *et al.,*(2012) (CovariatePC_taxa), utilizing the function snpgdsPCA. The function snpgdsPCACorr was employed to ascertain the correlation between the SNPs and the PCA. The functions in SNPRelate for PCA include calculating the genetic covariance matrix from genotypes, computing the correlation coefficients between sample loadings and genotypes for each SNP, calculating SNP eigengenvectors (loadings) and estimating the sample loadings of a new dataset from specified SNP eigengenvectors.

3.2.4 Genome Wide Association Study (GWAS) - SNPs Data

In this study, GWAS was integrated with quality phenotypic data from oil metabolic profiling, dense SNP datasets derived from RNA sequencing, and location coordinates information. This approach enabled the detection of alleles associated with variation in fatty acid methyl esters (FAMEs), phytosterols, and tocopherols, and allows testing whether certain alleles are preferentially distributed in specific climatic or geographic contexts. Key breeding-relevant traits, such as elevated fatty acid, phytosterol or tocopherol content, can thus be linked to specific genomic loci. These associations advance understanding of the genetic regulation of seed oil composition and provide tangible targets for marker-assisted selection.

Importantly using GWAS in this study helped in combining phenotypic data (e.g., oleic acid, α-Tocopherol, β-sitosterol, Δ^5-avenasterol, Δ^7- stigmastenol content), their genetic differences, and details about their environment (such as local temperature, rainfall, and soil), hence, this helps to tell which traits come from their genes or heritable effects and which are shaped by the environment.

3.2.4.1 Seed sowing, growth and development

Seeds were pre-treated by being washed with a solution consisting of 1 litre of distilled water and 200 milliliters of 4.5% diluted bleach. The seeds of each accession were placed in 50 ml centrifuge tubes containing a mixture and allowed to settle for 5 minutes before disposing. The seeds were subsequently rinsed a minimum of four times with distilled water. The purified seeds were soaked for 24 hours to reduce seed dormancy. Sowing took place in the Rice block of the University of York Greenhouse, with lighting operating from 5:00 AM to 5:00 PM. Illumination is triggered when light levels decrease to below 120 watts/m². Consequently, the length of daylight is 12 hours.

Seventy pots were filled with 100 grams of mixed soil of loamy and sandy soil type (F2+S), into which pre-treated seeds were placed at a depth of 2 centimeters to promote the emergence of the cotyledons. Following this, 80 milliliters of water were added to each pot, which were subsequently covered with lids to facilitate germination. The majority of the seeds sown germinated within 5 to 7 days, except for a few (MO4, MO30, MO54, and MO63), which germinated around the 10th day. On the seventh day after sowing, 28 pots demonstrated seed germination. In the following week, an additional 18 seeds germinated, resulting in a total of 2 weeks. The germination success rate is 46/70. Those that did not germinate were discarded after three weeks.

The accession MO70 from Burkina Faso germinated on the eleventh day, whereas MO8 failed to germinate after fourteen days and was subsequently discarded. *Moringa peregrina* (Oman, OmMp) germinated solely following the removal of the seed coat /husk

and subsequent cultivation. The seeds were placed on MSO plates and stored in a 4℃ refrigerator for three days prior to transfer to the Plant Growth room, which maintained an air temperature of 20℃ (+/-2℃), a daylength of 14 hours (8:00-22:00 BST), a light intensity of 120 µmol (m-2s-1 PPFD), utilized LED Valoya L28 (NS12 Spectrum) lighting, and an ambient relative humidity of approximately 60% (+/-10%). Seed germination is defined by the onset of sprouting; in Moringa, this process is hypogeal, with the cotyledons remaining below the surface. The first sign of emergence was the epicotyl producing the initial true leaves, observed around the seventh day (Cirlini *et al.,* 2022).

3.2.4.2 RNA extraction, and sequencing protocols

Established procedures for RNA extraction, library preparation, and sequencing were followed for RNA sequencing (RNA-Seq) analysis, in accordance with the pertinent literature. RNA was isolated from leaf, stem, and root tissues using the Omega EZNA Plant RNA Kit Protocol as per the manufacturer's guidelines, (LabRulez LCMS, 2024; Fang *et al.,* 2015). The CTAB procedure was utilized for the challenging root samples of *Moringa peregrinna* (Wang and Stegeman, 2010). Sample quality check (QC) was done on all samples using 1% Agarose gel electrophoresis with noticeable clear bands, RNA amount and purity were done using Nanodrop spectrophotometer and Agilent 2100 Bioanalyzer system to assess the RNA Integrity Number (RIN). These were averagely >9 which is a good value as RIN values greater than 7 are often regarded as good. Likewise, pure RNA typically with A260/280 value of 1.9 – 2.1, and an A260/230 value of 2.0 –2.2 is considered good quality (Agilent Technical Overview ; NEB#T2110).

The RNA quality and concentration achieved a RIN threshold exceeding 7 (Schroeder *et al.,* 2006). Furthermore, RNA samples with an RNA Integrity Number (RIN) greater than 7 were subsequently sequenced at Novogene using an Illumina short-read genome library.

3.2.4.3 Read pre-processing

A quality assessment was conducted on 251.7G sequenced reads corresponding to 1,635,367,412 paired reads of 150bp sequence length. This assessment utilized a high-performance computing system named Viking and employed FastQC version 0.11.9 (Andrews, 2010; Cock *et al.,* 2010). Raw RNA sequencing reads were processed to remove low-quality reads and adapters using Trimmomatic v0.36 (Bolger *et al.,* 2014; Hansen *et al.,* 2010). RNA transcripts were aligned to the reference genome (AOCCv2) using the splice-aware alignment method known as STAR. BAM files were obtained for a variant calling study (Dobin, 2015, and Dobin *et al.*,2012; 2015; 2016).

3.2.4.4 Single Nucleotide Polymorphism (SNP) Data Analysis

Variant calling analysis was performed with GATK version 4.3.0.0 Picard toolkit (Koboldt, 2020). Upon designating the reference and ploidy, each job was executed independently, generating output results to a file. Variant hard filtering was performed utilizing GATK, with the following parameters: QUAL < 30.0, MQRankSum < -2.5, and QD < 2.0. All variants from the 48 samples were subsequently consolidated using VCFTools version 0.1.16 (Danecek *et al.,* 2011). The genotypes were filtered according to the following criteria: a minimum allele count of 5, at least 2 alleles, a maximum of 25% missing data across all individuals, and the exclusion of SNPs with more than 3 alleles and a quality score below 30. The VCF data were converted to geno format using the vcf2geno function from the LEA package in R, and were then used to assess population stratification with TASSEL v.5.0 (Bradbury *et al.,* 2007) and PSIKOv2 (Popescu and Huber, 2015).The package SNPRelate in R was used to infer a PCA analysis of the variants (function snpgdsPCA). Correlation between the SNPs and the PCA was also obtained using the function snpgdsPCACorr.

3.2.5 GWAS - Bioinformatics analysis

The bioinformatics analyses associated with the GWAS (Uffelmann *et al*., 2021) included the use of HPC cluster (Viking) for variant calling by GATK, for the extraction of the SNPs and EGGNOG DB/mapping done with the support of Sara Franco Ortega (Post The GWAS pipeline in this study involved three main steps: Data preparation including quality filtering and formatting of SNP genotypes and phenotypic measurements.

Association testing, applying a General Linear Model (GLM) to test each SNP for association with the trait of interest. Multiple testing correction for controlling false discovery rates due to the large number of comparisons, here using Bonferroni correction.

Significant SNP trait associations were then annotated against reference databases to identify proximity to known or predicted genes, followed by functional interpretation to assess potential biological relevance.

Doc, Harper Lab), DIAMOND alignment by Sandy MacDonalds and Fabiano Pais (Data Science, Technology facility). I performed the integration of the traits and SNPs data, tested the association models GLM (General Linear Model) and MLM (Mixed Linear Modelling) for the GWAS analysis and chose the GLM using TASSELv5.0 (Bradbury *et al*., 2007). Evaluated the association by plotting the QQ plot and the Manhanttan plots, exploratory data analysis and visualization plots made by myself using R and Python.

The GWAS pipeline in this study involved three main steps: Data preparation including quality filtering and formatting of SNP genotypes and phenotypic measurements.

Association testing, applying a General Linear Model (GLM) to test each SNP for association with the trait of interest. Multiple testing correction for controlling false discovery rates due to the large number of comparisons, here using Bonferroni correction. Significant SNP trait associations were then annotated against reference databases to identify proximity to known or predicted genes, followed by functional interpretation to assess potential biological relevance.

The SNP file, Trait file (including FAMEs means.txt, Phytosterols means.txt, Cholesterol means.txt, seed weight, oil weight, and oil %), and PCA file were concatenated into a single file and made accessible in data/TASSEL. Single nucleotide polymorphisms exhibiting minor allele frequencies under 5% were omitted from the analysis. The SNP file and Trait file were standardized; likewise, the PCA file. The data were filtered, a kinship matrix was generated, and the datasets were merged utilizing an intersection method. The newly selected file, the intersected joined file, and the Kinship matrix underwent Mixed Linear Modelling (MLM) analysis from the Association menu, resulting in three output files: residual files, allele effects, and a statistics file comprising p-values, R^2 values, F statistics, and variance estimates for each tested trait. After comparing the QQplots and manhattan plots for both the MLM and GLM testing, I employed the General Linear Model (GLM) to identify correlations between markers (SNPs) and mean phenotypic values, using population membership estimates as covariates to control for population structure. The GLM statistics file was chosen due to its reliablity, resulting in the generation of a QQ plot and a Manhattan plot.

In order to identify top SNPs in the trait-gene association, all chromosomes above the Bonferroni threshold was chosen, using the calculation: bon=0.05/n row (gwas_results). Where n is the number of sites (markers) = 43456

To annotate all the markers on the chromosome of interest, I employed annotatepvalue=0.0001. To find specific SNPs within/near the candidate genes associated with the related metabolic oil traits, I used the EggNOGv5.0 (Huerta-Cepas *et al.,* 2019), EggNOG-mapper v2 (Cantalapiedra *et al.,*2021) which made use of EGGNOG DB and DIAMOND (Buchfink *et al.,* 2015) for the alignment. This provided the functional annotation of my Moringa dataset (Moringa emapper_annot.xlsx and MoringaRefgtf.xlsx) from where the number of SNPs significantly associated with the key oil traits studied were obtained.

3.3 Results

3.3.1 Trait data- Moringa seed oil Characteristics across Locations

From the oil extraction process, a total of forty-five distinct characteristics of seeds and seed oils was obtained. The attributes of seeds include their weight, the weight of oil, and the quantities of oil and hexane present in a 50 mg/ml stock solution. The trait data derived from Phytosterols comprises of Cholesterol, Cholestanol, Brassicasterol, Ergostadienol, Campesterol, Campestanol, Stigmasterol, Stigmastanol, Clerosterol, Obtusifoliol, β-Sitosterol, Δ^5-avenasterol, Δ^7-stigmastenol, Δ^7-Avenasterol, 24-Methyl-Cycloart, and Citrostadienol. The Tocopherols comprise α-Tocopherol (α-T), α-Tocotrienol (α-T3), β-Tocopherol (β-T), γ-Tocopherol (γ-T), γ-Tocotrienol (γ-T3), and γ-Tocopherol (δ-T). The derivatives of fatty acids classified as FAMEs include C14:0 (Myristic Acid), C16:0 (Palmitic Acid), C16:1Δ9 (Palmitoleic Acid), C17:0 (Margaric Acid), C18:0 (Stearic Acid), C18:1Δ9 (Oleic Acid), C18:1Δ11 (Vaccenic Acid), C18:2 (Linoleic Acid), C18:3n3 (α-Linolenic Acid, Omega-3FA), C20:0 (Arachidic Acid), C20:1n11 (Gadoleic Acid), C20:1n13 (Erucic Acid), C22:0 (Behenic Acid), C22:1n13 (Erucic Acid), and C24:0 (Lignoceric Acid).

The data on average seed weight, average oil weight, average oil content, and hexane volume for a 50 mg/ml stock pertains to the properties and extraction methodology of Moringa seed oil Supp.Table 3.1; Supp.Table 3.2; Supp.Table 3.3; Supp.Table 3.4. The oil yield, defined as the proportion of average oil weight to average seed weight, was calculated. This signifies the oil output for each individual seed unit. This metric is crucial for evaluating the efficacy of oil extraction processes. This denotes the proportion of oil, articulated as a percentage of the overall mass of the seeds.

Seed Variability

To standardize extraction processes and produce consistent results, deviations in the average seed weight could indicate variations in seed size within the samples. While *M. oleifera* seeds are smaller, somewhat triangular, and have a hard, dark brown outer shell, *M. stenopetala* seeds from Kenya (KMs5, KMs8, KMs10, KMs11) are triangular in shape, noticeably larger, and elongated, with a light brown to cream shell (Fig.3.3).

Figure 3.3 *Moringa oleifera* seeds (MO25, SW Nigeria;141.8mg ± 32.26) and *Moringa stenopetala* seeds (KMs11,Kenya; 626.0mg ± 32.19). **Photo**-Dorcas, Biology Research Lab, University of York, UK.

The seed weight varies across the samples with the *M. stenopetala* (KMs8, KMs11, KMs5, and KMs10) seeds with the highest average seed weight of 626mg, 516mg, 483.7mg and 405.3mg respectively) the smallest average seed weight of 141.8mg in *Moringa oleifera* MO25 (SW Nigeria). One would have expected that the highest oil yield would be obtained from the bigger seeds of the *M. stenopetala* accessions, but the highest oil yield (%) was obtained from the *M. oleifera* seeds from Mali (MaMo1, 45.5%) and SW Nigeria (MO25, 44.2%) with oil weight of 88.9mg and 62.1mg respectively. KMs8 and KMs11 has the highest oil weight, 203.6mg and 180.9mg respectively. From this data there is no correlation between seed weight and oil yield (Supp.Table 3.4) across regions and Supp. Figure3.2: Boxplot of oil content by region).

The largest seeds are those of Kenya *M. stenopetala* (KMs8) weighing 626 mg, has the highest oleic acid (C18:1D9) content of 875.90 Supp.Table 3.1, high β-sitosterol content of 1288.4mg Supp.Table 3.3 and highest α-tocopherol content of 63.78 Supp.Table 3.2 but with very low γ-tocopherol content of 1.69 Supp.Table 3.2. There is a huge difference in seed weight between accessions, and this does not correlate with oil content since the highest values are found in lines with smaller seeds (MaMo1-194.4mg, Mo25-141.8mg). Others like Mo55 are intermediate in both seed weight (size) (297.8mg) and oil % 41.1%. Two samples yielded very little oil (Mo70-B i.e 36.9mg and SJ-C i.e 36.4mg) despite having larger seeds than other repeats of the same accession.

The Kenya *M. oleifera* spp. showed high concentration Oleic acid (C18:1Δ9) 971.64mg/kg, and Vaccenic acid (C18:1Δ11) 111.92mg/kg. The concentration of *M. stenopetala* in Kenya is significantly low 7.5mg/kg in Vacenic acid compared to the *M.oleifera* samples 7.5mg/kg. Palmitoleic acid (C16:1Δ9) is present in *M. oleifera* (MO26) at a concentration of 34.49mg/kg highest in all the samples, while gadoleic acid (C20:1n11) is highest in MO16 at a concentration of 39.85mg/kg. The *M. oleifera* species

MO26 contains the highest linoleic acid (C18:2) concentration of 12.12 mg/100g among all the samples. α-linolenic acid (C18:3n3) is typically present in extremely low concentrations in *Moringa oleifera* samples (1.34mg/kg in MO17) and entirely absent from the Kenya *M. stenopetala* group (KMs8, KMs11). The average concentration of Myristic acid (C14.0) was 0.65mg/kg (0.09%), Erucic acid (C20:1n13) was 0.597mg/kg (0.09%), and Palmitic acid (C16.0) was 80.30mg/kg (6.26%). Only minimal quantities of deleterious fatty acids were detected.

The overall Mean values is as follows:

Average seed weight (mg) =259.70

Average Oil weight (mg) = 90.73

Average oil% = 35.25%

Correlation between Average Seed weight (mg) and Average Oil weight (mg): 0.97

Correlation between Average Seed weight (mg) and Average Oil %: -0.249

Correlation between Average Oil weight (mg) and Average Oil %: -0.026

There is a weak correlation between the average Seed weight and average Oil% (approx. -0.25) and between average Oil% and average Oil weight (-0.026), however there is strong positive correlation between average Seed weight and Oil weight (0.97) across all the samples. This indicates that an increase in seed weight is associated with a minor reduction in oil percentage; nevertheless, the correlation is weak.

In order to ascertain if the correlation among the seed weight, oil weight and oil percentage is statistically significant, a linear regression was made to predict oil% (Figure 3.4). There is a negative correlation between average seed weight and average oil percentage as seen with the regression line having a slight downward slope (Figure 3.4). As the seed weight increases, the oil percentage tends to decrease. The regression line can be used to estimate the corresponding oil percentage if the seed weight is known.

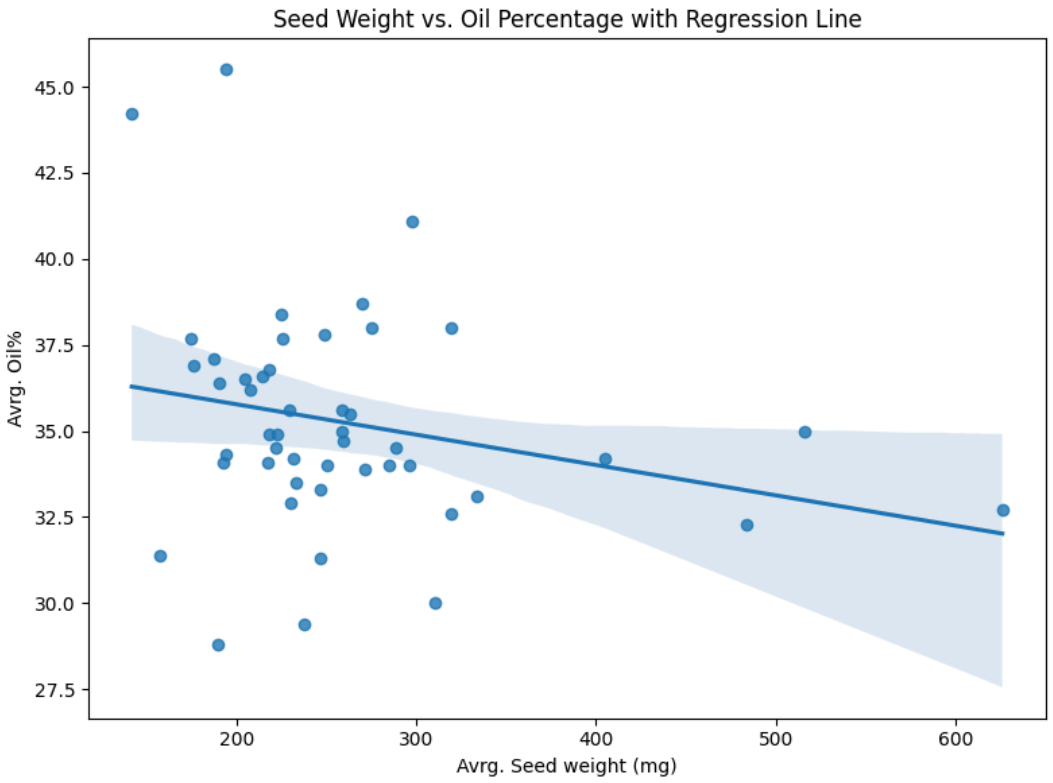


Figure 3.4 Regression seed weight vs oil percent.
Plot showing the relationship between average seed weight (mg) x-axis and average oil percentage (y-axis). Model, linear, coefficients R^2 equals 0.062 p equals 0.088 weak non-significant negative trend.

The R-squared value of 0.062 is low indicating that only about 6.2% of the variation in Oil percentage can be explained by the Seed weight. P>|t| for seed weight (mg) was 0.088 which is greater than 0.05, hence the relationship is not statistically significant at the 0.05 level. The coefficient for seed weight is -0.0088 which shows a very slight negative relationship, meaning that as seed weight increases, the oil percentage tends to slightly decrease, but this relationship is not statistically significant.

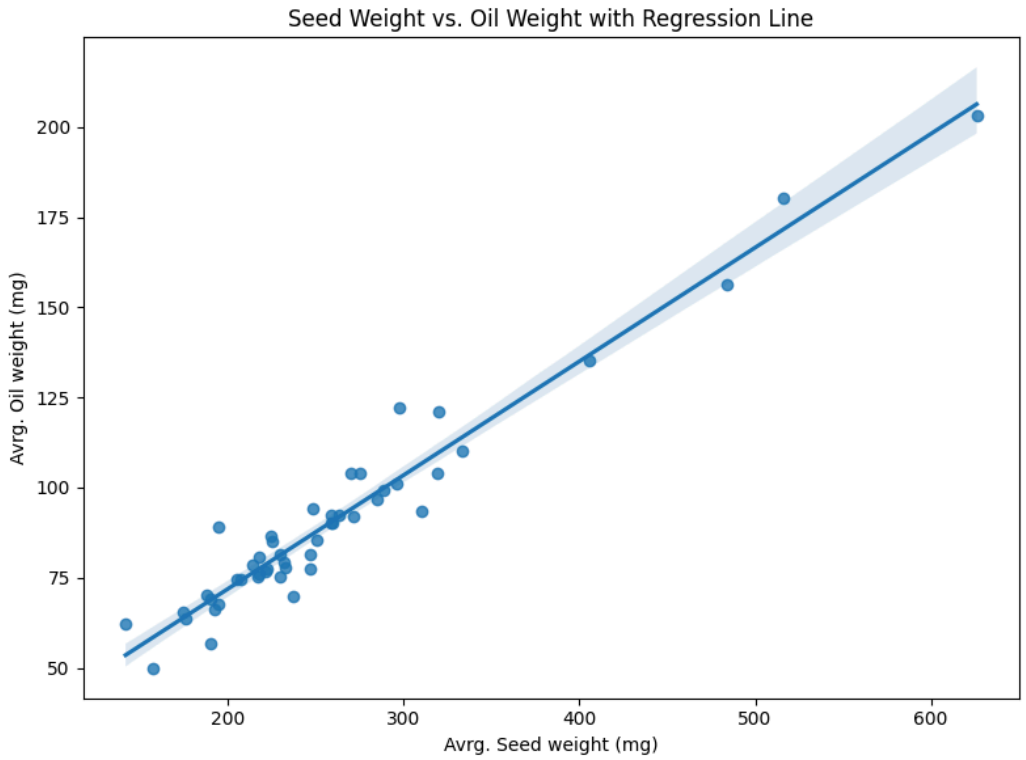


Figure 3.5 Regression seed weight vs oil weight
Plot showing the relationship between average seed weight mg (x-axis) and oil weight mg (y-axis). Linear model, coefficients 0.3156 R^2 equals 0.944 p less than 0.001 strong, significant positive linear relationship.

The R-squared value is 0.944, which is very high. This indicates that about 94.4% of the variation in Oil Weight can be explained by the Seed Weight. This is a strong relationship. P>|t| for seed weight (mg) is 0.000, which is highly statistically significant (much less than 0.05). This indicates a strong linear relationship between Seed Weight and Oil Weight. The coefficient is 0.3156. This indicates a strong positive relationship, meaning that for every one-unit increase in Seed weight (mg), the Oil Weight (mg) is estimated to increase by approximately 0.3156mg.

3.3.2 Fatty Acid Profile Comparison

This study produced 15 derivatives of fatty acids derived from Moringa seed oil and examined variations in Fatty Acid Methyl Esters (FAMEs) among Moringa accessions from 48 geographic locations. **Supp.Figure 3.2** shows that the *M. stenopetala* species yielded the least concentration of C18.3n3 α-Linolenic acid and C18.1D11 Vaccenic Acid (isomer of oleic acid) compared with the *M. oleifera* species. The Kenya *M. oleifera* group KMO2, KMO4, KMO6, KMO7 shows high yield of Vaccenic acid, an isomer of Oleic acid. Likewise C24.0 Lignoceric acid a unique FAME derivative produced in the Kenya *M. stenopetala* seed oil yielded considerably high concentration unlike other derivatives where they were seen to yield very low.

t -Test analysis was done to compare the means of the FAMEs between the *Moringa oleifera* and *Moringa stenopetala* groups to determine if there are significant differences in their means between the two groups. The t-test results indicate significant differences in the means of several fatty acids between the *Moringa oleifera* and *Moringa stenopetala* groups. Notably, the following fatty acids show statistically significant differences (p-value<0.05): **C16.1D9 Palmitoeic acid** (p-value 0.00), **C17.0 Margaric acid** (pvalue 0.02), **C18.1D11Vaccenic acid** (p-value 0.00), **C18.3n3 α-Linolenic acid** (p-value 0.00), **C20.1n13 Erucic acid** (p-value 0.00), **C22.0 Behenic acid** (p-value 0.00), C22.1n13 (p-value 0.01), **C24.0 Lignoceric acid** (p-value 0.00), Total (p-value 0.00). The t-test result is summarized in (**Supp.Table 3.2**). Furthermore, the ANOVA test conducted to compare the means across all samples presents the F-statistics values for each of the fatty acid composition, tested in the ANOVA (**Supp.Table 3.3**) where Palmitoleic acid (C16.1D9), 4E+02 corresponds to and shows significant differences across the groups considering the strong p-value (0E+00).

Based on the ANOVA and Tukey's HSD post-hoc tests the following fatty acids show statistically significant differences among the sample groups: C16_0 **Palmitic**, C17_0 **Margaric acid**, C18_1D11 **Vaccenic acid**,C18_3n3 **α- linolenic ac**id, C20_1n13 **Erucic acid**, C22_0 **Behenic acid** and C24_0 **Lignoceric acid** Supp.Table 3.1.

C16_0 Palmitic acid is significantly different in MO26 (135.44 ± 67.83^b), C24_0 Lignoceric acid in KMs5 (22.49 ± 1.69^c) and KMs8 (21.79 ± 1.12^d). C18_3n3 α-linolenic acid is significantly different in MO26 (3.97 ± 1.02^b) and KMs5 (0.28 ± 0.04^c) from other samples. C17_0 Margaric acid is significantly different in MO3(0.45 ± 0.77^b) from other samples with value.

C18_1D11 Vaccenic acid is significantly different in KMs5 (7.30 ± 0.39^b), KMs 8 (7.11 ± 1.23^c) and KMs10 (7.25 ± 0.79^d), C20_1n13 Erucic acid in KMO7 (2.08 ± 0.40^b) and KMs5 (0.00 ± 0.00^c) and C22_0 Behenic acid in KMs10 (47.94 ± 8.40^b). The unsaturated fatty acids among these significant FAMEs include α-linolenic acid, C20_1n13 Erucic acid, C18_1D11 Vaccenic acid and C24_0 Lignoceric acid while palmitic, behenic and margaric acids are saturated fatty acids.

Principal Component Analysis (PCA) of FAMEs variables

The evaluation of the Principal Component dimensions of the variables (FAMEs), **(Fig.3.6)**, and the individual Moringa accessions, gives insight to the key components affecting oil quality and distinguished the Moringa accessions based on their fatty acid.

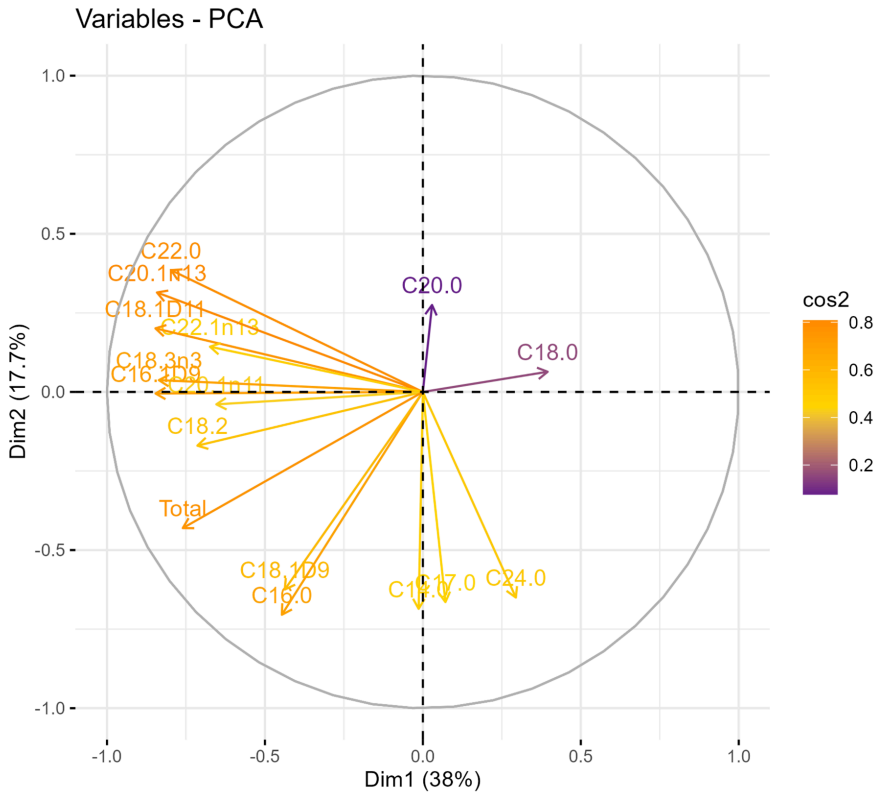


Figure 3.6 PCA of FAMEs variables. This is the PCA of FAMEs, x-axis/ Dim.1 38 percent and Dim.2/ y-axis 17.7 percent and 55.7 percent total. C18:0 Linolenic acid and C20:0 Arachidic acid show no correlation. They exhibit independent variation. Behenic acid (C22:0), Vaccenic acid (C18:1n11), and C22:1n13 in the left quadrat show a positive correlation with one another. They share comparable enzyme pathways, such as elongation and desaturation. Key shows the cos² values which measures the quality of the representation of the FAMEs such as C14:0 (Myristic Acid), C16:0 (Palmitic Acid), C16:1D9 (Palmitoleic Acid), C17:0 (Margaric Acid), C18:0 (Stearic Acid), C18:1D9 (Oleic Acid), C18:1D11 (Vaccenic Acid), C18:2 (Linoleic Acid), C18:3n3 (α-Linolenic Acid, Omega-3FA), C20:0 (Arachidic Acid), C20:1n11 (Gadoleic Acid), C20:1n13 (Erucic Acid), C22:0 (Behenic Acid), C22:1n13 (Erucic Acid), and C24:0 (Lignoceric Acid), on the PCA dimension with values ranging from 0 to 1, where higher values shown in orange implies strong representation on Dim1and Dim 2 and lower values, shown in purple, indicate weak representation. cos² is the squared cosine of the angle between the FAMEs variables and a principal component axis. Plot made using R.

FAMEs derivatives do not show any statistically significant difference with geographical locations (latitude) on evaluation with Tukey HSD at $p<0.05$ indicating stable expression of these compounds or possibly influenced by environmental or genetic factors.

3.3.3 Variability in Phytosterols

Based on the ANOVA and Tukey's HSD post-hoc tests, the following phytosterols show statistically significant differences among the sample groups: Ergostadienol (KMO4: 89.99 ± 40.97^b), Campesterol (KMO4: 1022.63 ± 679.91^b), Stigmasterol (KMO2: 1325.39 ± 360.29^b, KMO4: 1188.44 ± 863.44^c), Stigmastanol (KMO4: 74.33 ± 37.36^b), Clerosterol (KMO4:69.82 ± 36.17^b), Obtusifoliol (MO26: 72.53 ± 0.02^b;

KMs10: 57.30 ± 11.35^c), β_Sitosterol (KMO4: 3453.54 ± 2187.62^b), D5Avenasterol (KMO4: 833.35 ± 606.86^b), D7_Stigmastenol (MO26: 290.04 ± 182.85^b). Here KMo4 (Kenya *M. oleifera*) shows statistically significant variability in Ergostadienol, Campesterol, Stigmasterol, Stigmastanol, Clerosterol, β_Sitosterol and D5Avenasterol. Likewise, MO26 (Nigeria *Moringa oleifera*) shows statistically significant variability in Obtusifoliol and D7Stigmastenol.

Phytosterols such as Brassicasterol, Ergostadienol, Campesterol, Campestanol, Stigmastanol, Clerosterol, Obtusifoliol, β-Sitosterol, D5Avenasterol, D7Stigmastenol, D7-Avenasterol, 24-Methyl-Cycloart, and Citrostadienol were not found to be statistically significant with latitude, with Tukey HSD, ANOVA $p < 0.05$.

Principal Component Analysis (PCA) of Phytosterol variables

This PCA helped identify relationships and variations in Phytosterol concentrations among different Moringa accessions (Figure 3.7). In the Phytosterols plot, most compounds are clustered on the left side, indicating negative correlation with Dim1. Campesterol is strongly negatively correlated with PC1 and positively with PC2. Stigmasterol and β-sitosterol are positively correlated with PC1 and negatively with PC2. Delta5Avenasterol, Delta7-Stigmastenol, and Delta7-Avenasterol are clustered together, suggesting they have similar patterns. There appears to be a loose clustering along the Dim1 axis with some accessions spread towards the left (e.g., KMs5, KMs8) and others grouped towards the right (e.g., MO17, WOT).

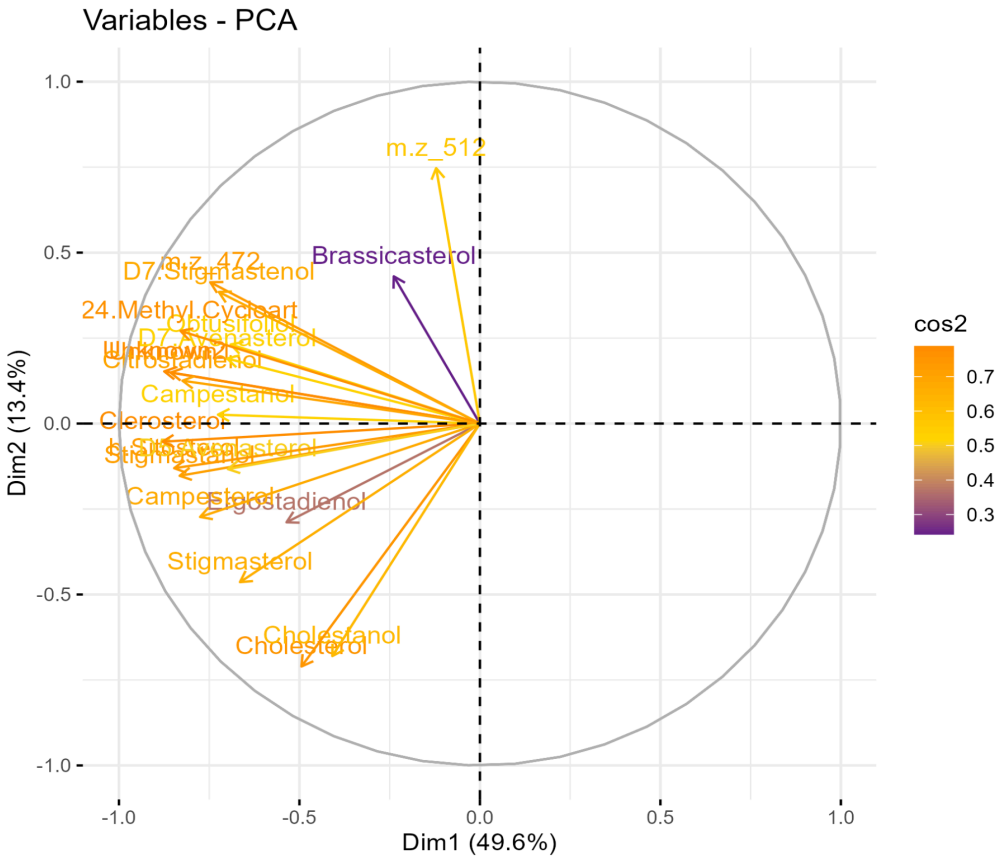


Figure 3.7 PCA of phytosterols individual variables.

This is the phytosterols PCA, with the x-axis, Dim1showing 49.6 percent variance and y-axis, Dim2 with 23.8 percent and total variance, 63.0 percent. Key shows the cos² values which measures the quality of the representation of the Phytosterols on the PCA dimension with values ranging from 0 to 1, where higher values shown in orange implies strong representation on Dim1and Dim 2 and lower values, shown in purple, indicate weak representation. cos² is the squared cosine of the angle between the Phytosterols variables and a principal component axis. The colour gradient from purple to yellow rereflects the true contribution of the concentration of Phytosterols. Plot made using R

3.3.4 Variability in Tocopherols

The analysis of the Moringa oil samples revealed notable concentrations of various tocopherols (Supp.Table 3.2**)** specifically: α-T at 63.78mg in KMs8, α-T3 at 49.80mg in Mo26, γ-T at 26.36mg in MaMo1, γ-T3 at 4.65mg in KMs11, and Delta at 2.99mg in KMo6. The Kenya *Moringa stenopetala* demonstrated the most elevated concentrations of α-T (63.78mg in KMs8) and γ-T3 (4.65mg in KMs11) in seed oil, among the 49 accessions examined. The *Moringa oleifera* exhibited the most elevated concentrations of Gamma Tocopherol (Mali accession), α-T3 (SW/Nigeria MO26), and Delta Tocopherol (Kenya).

The t-test results indicate significant differences in the means for some tocopherol measurements between the two groups. Specifically, the measurements for AlphaT3 and

GammaT3 show statistically significant p-values (0.0007 and <0.0001, respectively), suggesting that these tocopherols differ significantly between the KMS samples and the other samples.

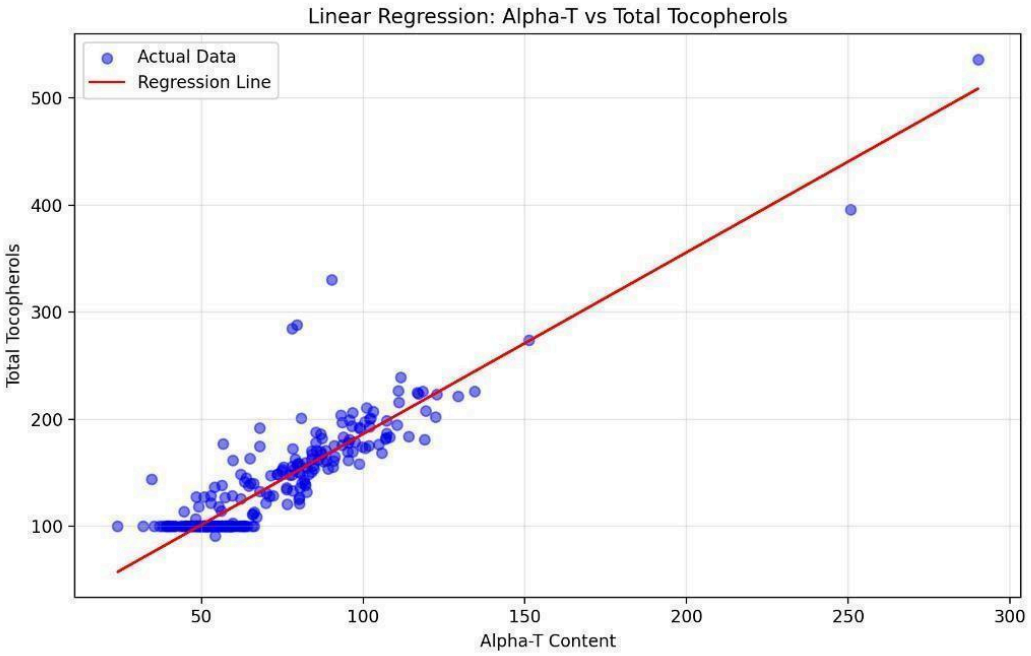


Figure 3.8 Regression of alpha(α)-T vs total tocopherols

Plot showing the relationship between alpha(α)-T(x-axis) and total tocopherols (y-axis). Linear model, coefficients R^2 equals 0.747 p 74.7percent variability in total tocopherols show strong, significant positive linear relationship.

Principal Component Analysis (PCA) of Tocopherol variables

The Principal Component Analysis (PCA) conducted on the tocopherol data of the various Moringa accessions (Fig.3.9) demonstrates an association between the tocopherols and the principal components. The vector representations of the variables' placements and lengths demonstrate their contributions to the PCs. The four distinct varieties of tocopherols are readily apparent in the tocopherols plot. Each point represents an individual, and their positions on the plot indicate their values for the respective principal components. Clusters or patterns in the distribution of individuals, suggest similarities or differences in their Tocopherol compositions as observed here. For example, MO17, WOT, and MO16 are positioned close together, indicating they have similar profiles. In contrast, individuals like KMS8, KMS5, KMS11 and KMS10 are scattered, suggesting they have distinct profiles. These two PCAs can be interpreted together to gain insights into the relationships between variables and individuals.

Variables that contribute significantly to a particular PC can be associated with the individuals that score highly on that PC.

This information helps to understand the factors responsible for the observed patterns and for identifying potential markers or targets for further investigation or breeding efforts about the genetic diversity, nutritional profiles, and other traits of interest in this Moringa research.

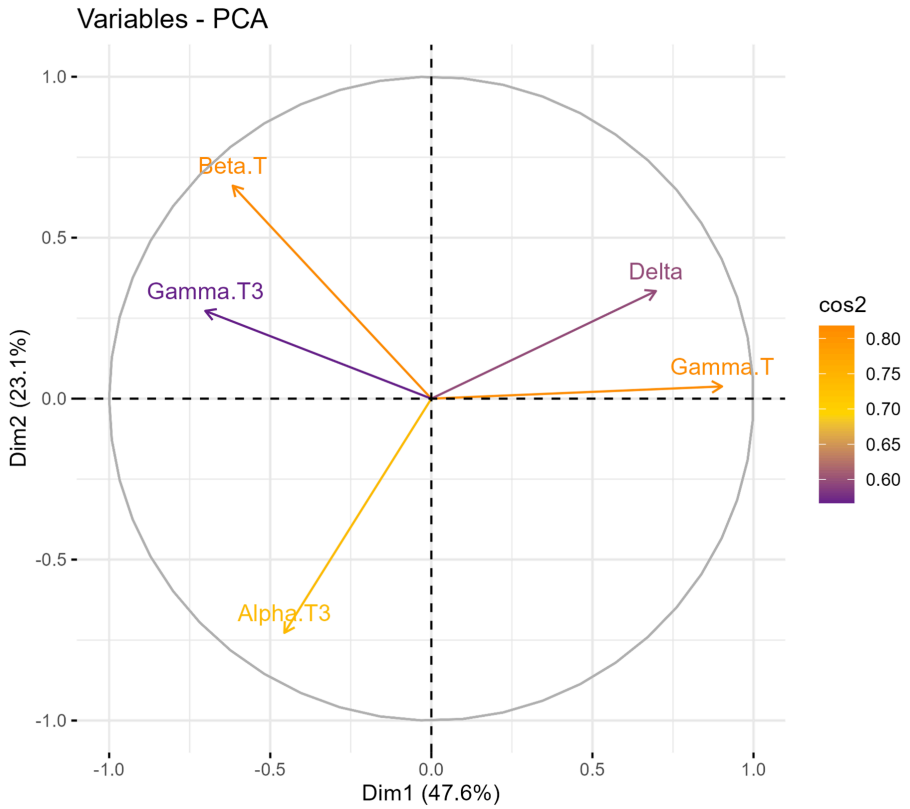


Figure 3.9 PCA of tocopherols individuals.
This is the tocopherols PCA, with the x-axis, Dim1 showing 47.6 percent variance and y-axis, Dim2 with 23.1 percent and total variance, 70.7 percent. The colour gradient from purple to yellow represents the concentration of tocopherols. The first principal component (Dim1) is primarily influenced by the factors Beta.T, Gamma.T3, Delta, and Gamma.T, which collectively account for 47.6% of the total variance. Alpha.T3 is also included in the second primary component (Dim2) (Fig.3.9). The correlation between Beta.T and Gamma.T3 is positive, while the correlation between Delta.T and Gamma.T is negative. Key shows the cos² values which measures the quality of the representation of the tocopherols on the PCA dimension with values ranging from 0 to 1, where higher values shown in orange implies strong representation on Dim1and Dim 2 and lower values, shown in purple, indicate weak representation. cos² is the squared cosine of the angle between the tocopherols variables and a principal component axis. Plot made using R.

A one-way ANOVA was conducted to determine whether the mean concentrations of individual oil components (Fatty Acids, Tocopherols, and Phytosterols) significantly differed across geographical locations. Where ANOVA was significant ($p < 0.05$), Tukey's HSD post-hoc test was used to pinpoint which specific locations differed Supp.Table 3.8.

3.4 Does Location influence oil constituents regardless of species?

Summary of ANOVA and Tukey HSD analysis of all samples as seen in (Supp.Table 3.7) reveals that Fatty acids such as: C22.1n13 (Erucic Acid), C20.0 (Arachidic Acid), C20.1n11 (Gadoleic Acid), C24.0 (Lignoceric Acid); Tocopherols such as Gamma T and Beta T; and Phytosterols such as Cholesterol, Cholestanol, Clerosterol, b.Sitosterol, Brassicasterol show statistically significant differences across locations.

On application of Tukey HSD, post-hoc pairwise comparisons were made to identify exact locations which differ significantly. For instance C20.0 (Arachidic Acid) varied significantly between Oyo NIG vs PlateauNIG; Cholesterol significantly varied between ['Burkina Faso vs Oyo NIG', 'Ekiti NIG vs Oyo NIG', 'India vs Oyo NIG', 'KatsinaNIG vs Osun NIG', 'KatsinaNIG vs Oyo NIG', 'Kenya vs Kwara NIG', 'Kenya vs Osun NIG', 'Kenya vs Oyo NIG', 'Kwara NIG vs Mali', 'Kwara NIG vs Oyo NIG', 'Mali vs Osun NIG', 'Mali vs Oyo NIG', 'Oyo NIG vs Philippines', 'Oyo NIG vs PlateauNIG'] and Gamma.T between ['Ekiti NIG vs Kenya', 'KatsinaNIG vs Kenya', 'Kenya vs Kwara NIG', 'Kenya vs Mali', 'Kenya vs Osun NIG', 'Kenya vs Oyo NIG'] and others as seen in (Supp.Table 3.7) . Results from Tukey's HSD post-hoc analysis highlighted specific location pairs with significant differences, which can guide breeding or regional selection decisions.

In terms of latitude regression, α-Tocopherol, C22.1n13 (Erucic Acid) showed a statistically significant relationship with latitude ($p = 0.011$), indicating a trend where southern locations (lower latitudes) produce more α-T and C22.1n13 (Erucic Acid) (Supp.Table 3.7). This finding would guide regional selection decisions for breeders and highlight specific oil profile as potential marker traits for identifying superior Moringa lines for oil quality.

The following fatty acids did not vary with location or latitude, indicating stable expression of these compounds, or possibly influenced by environmental or genetic factors. These are C14:0 (Myristic Acid), C16.0 (Palmitic Acid), C16:1Δ9 (Palmitoleic Acid), C17:0 (Margaric Acid), C18:0 (Stearic Acid), C18:1Δ9 (Oleic Acid), C18:1Δ11 (Vaccenic Acid), C18:2 (Linoleic Acid), C18:3n3 (α-Linolenic Acid, Omega-3FA), C20:1n11 (Gadoleic Acid), C20:1n13 (Erucic Acid), C22:0 (Behenic Acid), C24:0 (Lignoceric Acid). Particularly C16.1D9 (Palmitoleic)$F = 4.66$, $p = 0.00026$ and C22.0 (Behenic) $F = 3.11$, $p = 0.0048$.

Phytosterols such as Campesterol, Campestanol, Citrostadienol, likewise Tocopherols such as AlphaT, AlphaT3, Gamma T3 and Delta were not found to be statistically significant with ANOVA $p < 0.05$

3.5 Does location influence oil constituents within the *M oleifera* samples only?

Summary of ANOVA and Tukey HSD Results for all *M. oleifera* samples without the Kenya *Moringa stenopetala* group (KMs) Summary of the ANOVA, Regression and

Tukey HSD (Supp.Table 3.8) reveals that Beta Tocopherol and Phytosterols such as Cholesterol, Stigmasterol, and Cholestanol, show statistically significant differences across locations. On application of Tukey HSD, post-hoc pairwise comparisons were made to identify exact locations which differ significantly. The Tukey's HSD post-hoc analysis highlighted specific location pairs with significant differences, which can guide breeding or regional selection decisions.

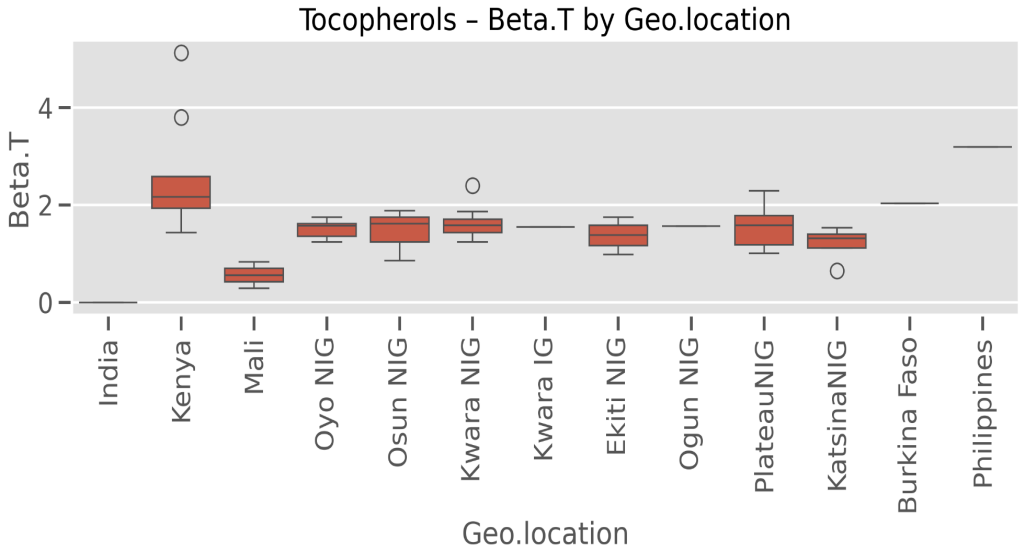


Figure 3.10 Beta (β)-Tocopherol by location boxplots.
The geographical regions are on the x-axis, and the Beta.Tocopherol concentrations (mg/100g). Sample size is 32, Model is OLS Regression, Python software. Beta.T (ANOVA p=0.0001***) shows there is high significance across the different Moringa accessions. There is a significant relationship between β-Tocopherol levels and location (Regression p=0.0012**) Tukey Significant Comparisons showed multiple pairwise differences, varying significantly between 'Burkina Faso vs India', 'Ekiti NIG vs Philippines', 'India vs Kenya', 'India vs Kwara NIG', 'India vs Oyo NIG', 'India vs Philippines', 'India vs PlateauNIG', 'KatsinaNIG vs Philippines', 'Kenya vs Mali', 'Kwara NIG vs Mali', 'Kwara NIG vs Philippines', 'Mali vs Philippines', 'Osun NIG vs Philippines', 'Oyo NIG vs Philippines', 'Philippines vs PlateauNIG'. There is significant contrasts between India and several African locations (Kenya, Kwara, Oyo, Plateau). Repeated differences involving the Philippines, suggesting it has a unique β-Tocopherol profile. Also notable are inter-African differences, e.g., Kwara vs Mali, and Kenya vs Mali.

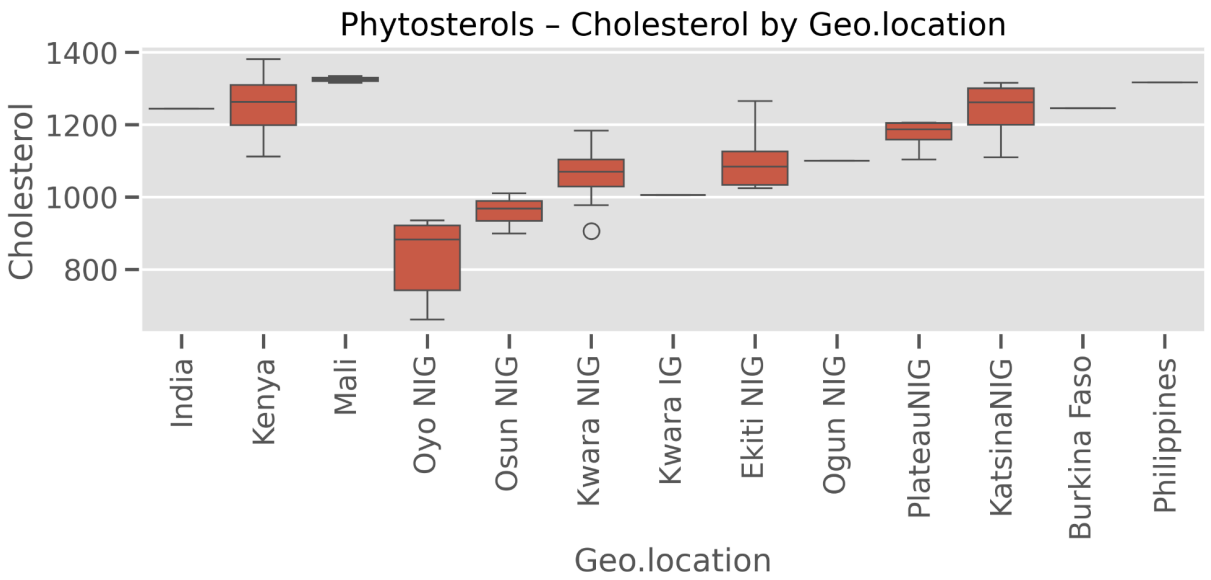


Figure 3.11 Cholesterol by location boxplots.
X-axis represent the geographical regions where seeds were collected and the Cholesterol concentrations (mg) represent the y-axis. Sample size is 32, Model is OLS Regression, Python software. There is a highly significant difference in Cholesterol levels between the different locations (countries or Nigerian states). ANOVA p-value of 0.0*** (typically means $p < 0.001$) confirms strong evidence that location affects cholesterol content in the Moringa oils. The Regression $p = 0.5654$ is well above the typical 0.05 threshold, no linear trend or predictive pattern was detected. Suggesting that the linear regression model used to assess the relationship between cholesterol and latitude or oil yield did not show a significant relationship. Tukey's significant pairwise comparisons show specific location pairs where cholesterol levels differ significantly. 'Burkina Faso vs Oyo NIG' and 'Kenya vs Oyo NIG' signifies that cholesterol content is significantly different between those countries. Oyo NIG appears frequently in the comparisons, suggesting that it is statistically distinct from many other regions.

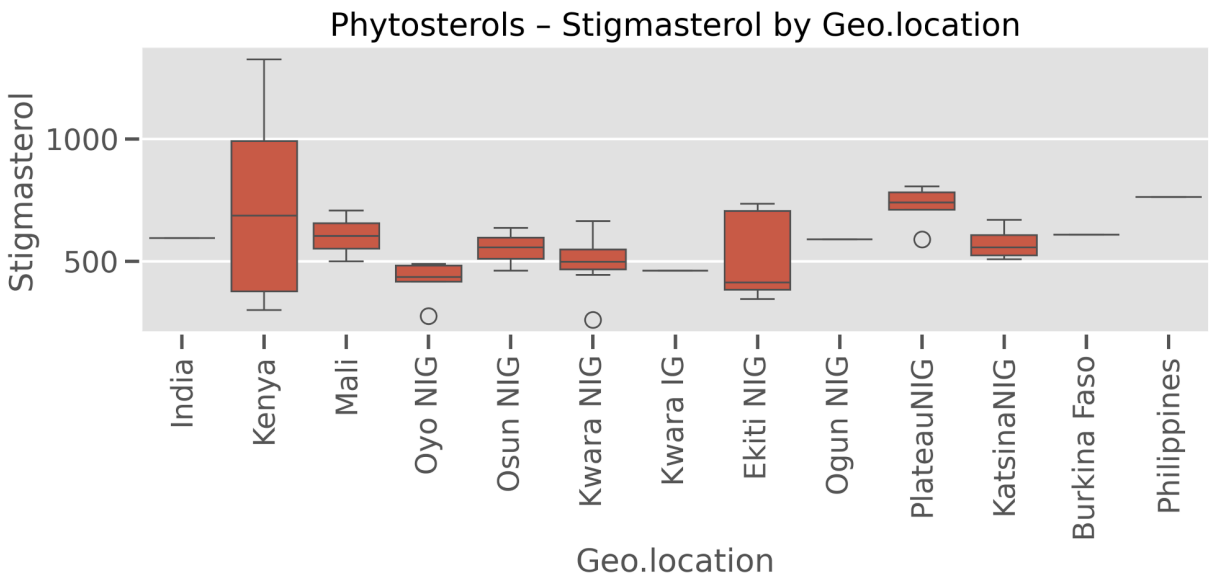


Figure 3.12 Stigmasterol by location Box plots.

X-axis represent the geographical regions where seeds were collected and the Stigmasterol concentrations (mg) represent the y-axis. Sample size is 32, Model is OLS Regression, Python software. ANOVA p-value of 0.0*** (typically means p < 0.001) confirms strong evidence that location affects Stigmasterol content in the oils. Regression p=0.0352* shows there is a significant linear relationship between stigmasterol content and geographical location of the seeds. There is a highly significant difference in Stigmasterol levels among the different countries or Nigerian states ('Ekiti NIG vs Kenya', 'KatsinaNIG vs Kenya', 'Kenya vs Kwara IG', 'Kenya vs Kwara NIG', 'Kenya vs Mali', 'Kenya vs Osun NIG', 'Kenya vs Oyo NIG', 'Kenya vs PlateauNIG', 'Oyo NIG vs PlateauNIG'.There are distinct patterns where Kenya appears repeatedly in the Tukey comparisons, suggesting that its stigmasterol profile is distinct from many other regions.

This infers a predictive trend that stigmasterol likely increases or decreases along a climate gradient such as temperature or dryness. The distinct patterns where Kenya appears repeatedly in the Tukey comparisons, suggests that its stigmasterol profile is distinctly higher from many other regions (**Figure 3.12**). Differences like 'Ekiti NIG vs Kenya' and 'Kenya vs Mali' suggest broad divergence across ecological zones (Kenya is semi-arid to arid; Ekiti and Mali are wetter/more humid).

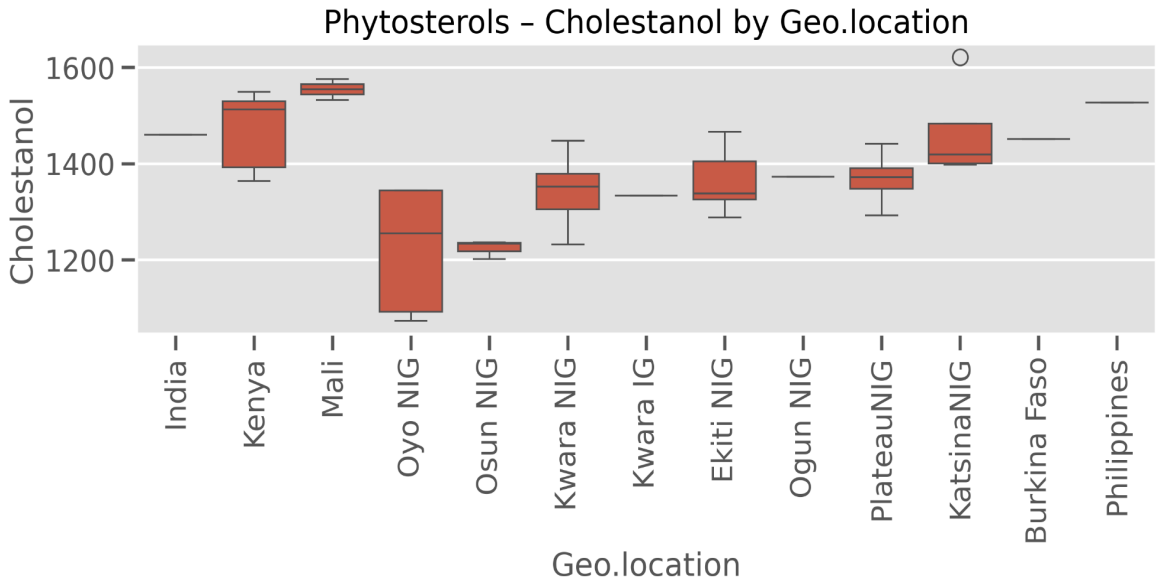


Figure 3.13 Cholestanol by location Boxplot

X-axis represent the geographical regions where seeds were collected and the Cholestanol concentrations (mg) represent the y-axis. Sample size is 32, Model is OLS Regression, Python software. ANOVA significant regression p equals 0.5192 non-significant. Tukey test identifies key differences between accessions, 'KatsinaNIG vs Osun NIG', 'KatsinaNIG vs Oyo NIG', 'Kenya vs Osun NIG', 'Kenya vs Oyo NIG', 'Mali vs Osun NIG', 'Mali vs Oyo NIG' particularly involving KatsinaNIG, Kenya, and Mali vs. Osun NIG or Oyo NIG indicating that these southern Nigerian accessions have lower cholestanol content

ANOVA confirms that location affects Cholestanol content, but unlike stigmasterol, this variation is not driven by a measurable linear environmental gradient (like temperature or rainfall). The lack of significance in regression ($p = 0.5192$) means that other factors perhaps genetic or post-transcriptional regulation drive cholestanol levels. (Figure 3.13).

Relationship between Geographic Latitude and Genetic Variation (PC1)

The association between latitude and the first principal component (PC1) of genetic variation was tested to examine whether geographic location influences genetic structure in the Moringa populations (Supp. Figure 3.11A and Supp. Figure3.11B). The pearson correlation analysis showed a weak, non-significant negative association between latitude and PC1 ($r = –0.23$, $p = 0.203$). Similarly, the simple linear regression indicated that latitude explained only a small proportion of the variation in genetic structure, with an R^2 value of 0.05. This suggests that geographic latitude accounts for approximately 5% of the observed variation in PC1 among the sampled Moringa populations. Altogether, the

findings from these analyses indicated that while there is a slight trend, genetic differentiation in these populations is not strongly structured by latitude.

Hierarchical clustering heatmap (Fig.3.14) was made to explore the patterns among Moringa accessions and across the Moringa species alongside the Moringa seed oil traits. Such patterns include cluster separation, trait co-clustering, branch length and population clustering. The colour intensity reflects the relative abundance of each trait across accessions, while the dendograms display clustering relationships. The populations that cluster closely together share more similar profiles for instance KMs10, KMs8, KMs11 and KMs5 show low abundance of C22, C18.1D11, C20.1n13, Δ^5-avenasterol and GammaT3. So also is C16.1D9, C18.3n3; Campesterol, Stigmasterol, campestanol and stigmasterol. All the FAMEs and tocopherols show low abundance in KMO4.

This analysis highlights distinct population groups and reveals trait clusters that co-vary across populations, suggesting shared regulatory or metabolic pathways. The following show high abundance of ergostadienol and D5Avenasterol (KM04), C16.0, C18.2, C161D9, C18.3n3, Beta T, almost all sterols including Δ^7-stigmastenol, X24.MethylCycloart (MO26), all sterols, especially Clerosterol, β-Sitosterol, Campesterol, Stigmasterol and Stigmastanol (KMO4).

KMo2, PMo12, KMO6, KMO7, KMO4, MO26, and SJ show high abundance of cholesterol, cholestanol, campesterol, stigmasterol, campestanol, stigmastanol.

MO3 shows extremely low abundance of C17.0, so also is MO17 for C18:1D9.

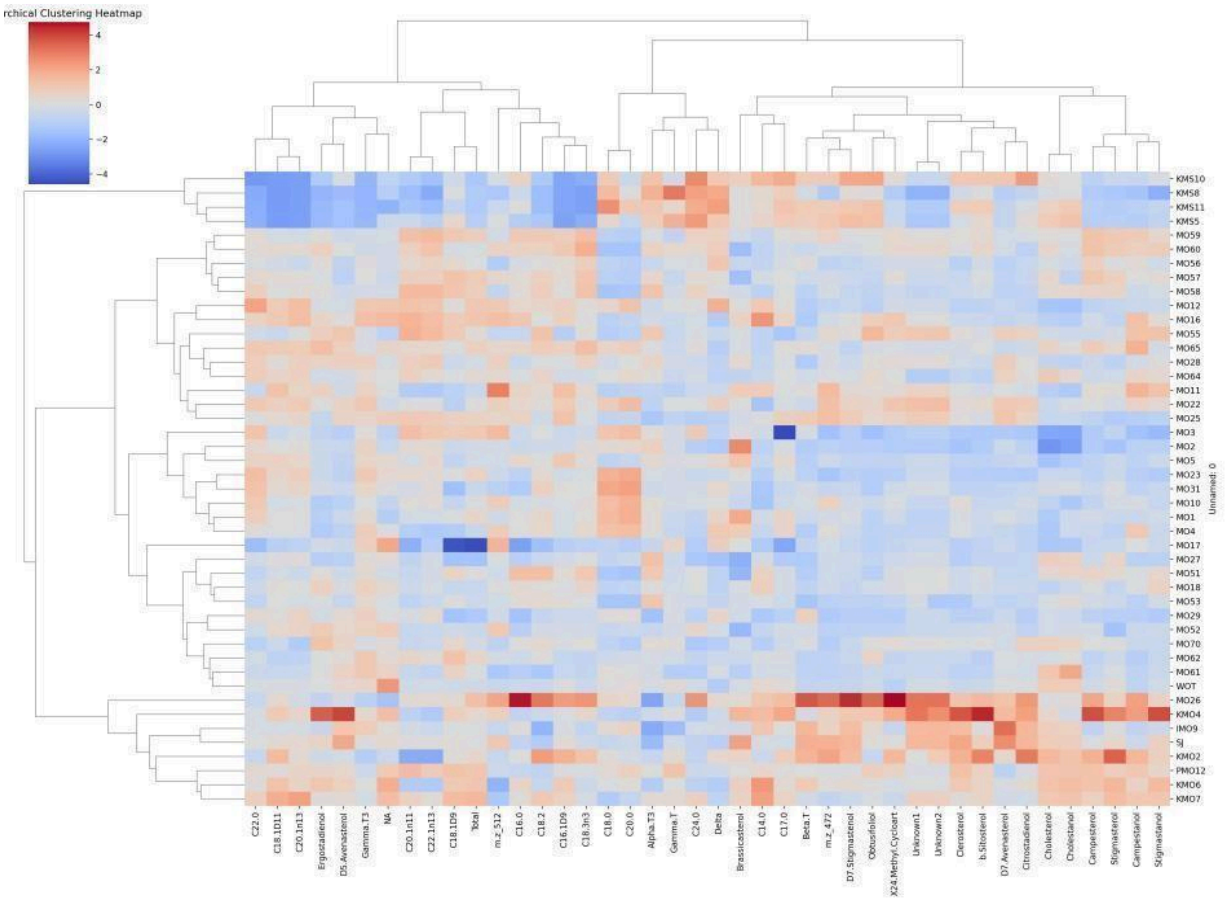


Figure 3.14 Hierarchical Clustering Heatmap Moringa oil traits and accessions. This 2-dimensional visualization combines a heatmap (coloured intensity matrix of data values) with dendograms (tree-like structures) that cluster both rows and columns. Heatmap displays variation in the seed oil traits across samples with colour intensity indicating low-to-high values. The dendograms represent similarity relationships, Moringa samples (rows) and Seed oil traits (columns) are grouped based on their pairwise distances, with closer branches indicating higher similarity.

3.6 Trait-Gene Associations with GWAS

A Genome-Wide Association Study was performed to discover possible SNPs linked to variations in Moringa seed oil components across diverse genotypes from different regions. The trait data, PCA, and SNP data were integrated using the General Linear Model (GLM), with the relationships depicted via QQ plots (Supp. Figure 3.9) and a Manhattan plots (Supp. Figure 3.10). The analysis included 48 accessions, with 149,740 SNPs distributed across 90,751 loci and 43,456 sites. The nominal significance was $p<0.05$ and Bonferroni-corrected significance threshold used was 5.47×10^{-7}. All SNPs which exceed the Bonferroni threshold included candidates linked to important oil biosynthesis genes.

3.6.1 Association mapping of Phytosterols - Δ5 Avenasterol

The quantile-quantile (QQ) plot for Δ^5-avenasterol (Figure 3.15) demonstrated a modest deviation from the expected distribution under the null hypothesis, indicating the presence of genuine associations beyond what would be expected by chance. This deviation from the diagonal line, infers potential SNPs influencing the variation in Δ^5-avenasterol concentration.

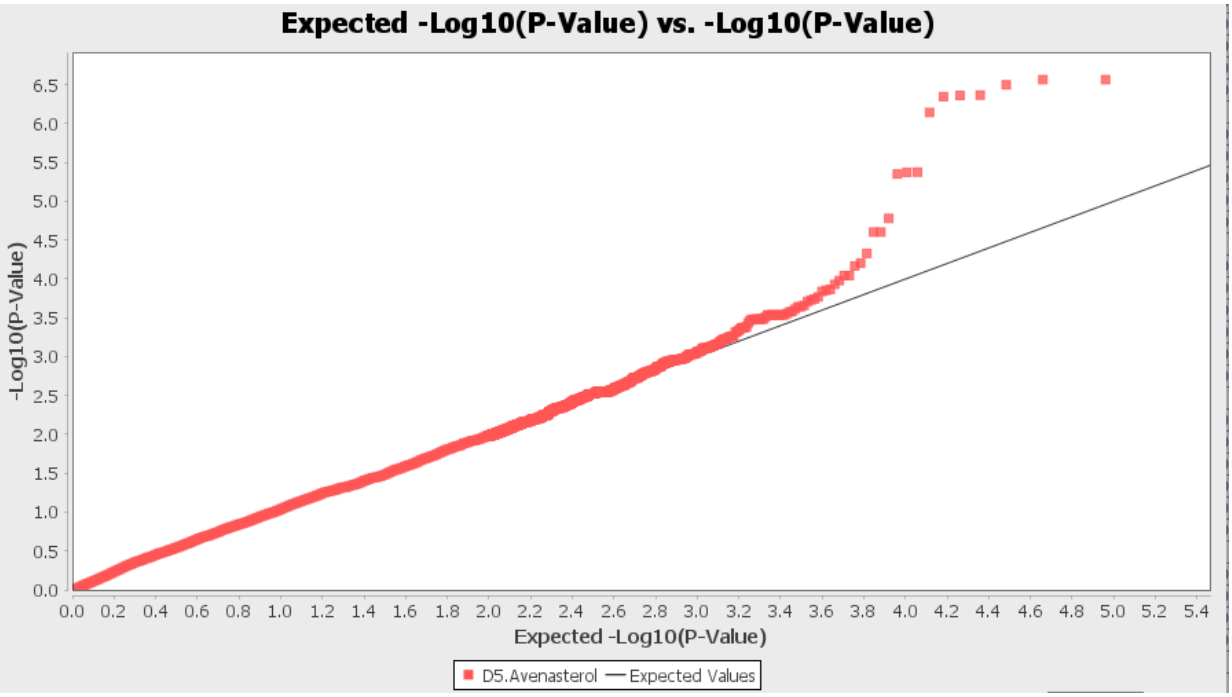


Figure 3.15 Genome-wide association analysis for Δ^5-avenasterol concentration. Quantile-quantile (QQ) plot of the expected vs observed $-\log_{10}(p)$ values from the GWAS model showing deviation indicative of true associations. The x-axis represents the expected distribution and the y-axis shows the observed distribution.

The Manhattan plot (Figure 3.16) revealed multiple peaks of significant single nucleotide polymorphisms (SNPs) across the genome. Notably, a strong association signal was observed on Chromosome 13, exceeding the genome-wide significance threshold. To further investigate this region, a focused Manhattan plot of Chromosome 13 (Figure 3.17) was generated, which includes gene annotations. On examining SMorol13g04900_461 (GO: 0140359), an ATP-binding cassette that is a member of the ABC transporter G family linked to the Δ5 Avenasterol trait implies that variations in or near this gene could affect a plant's ability to adapt to environmental stresses.

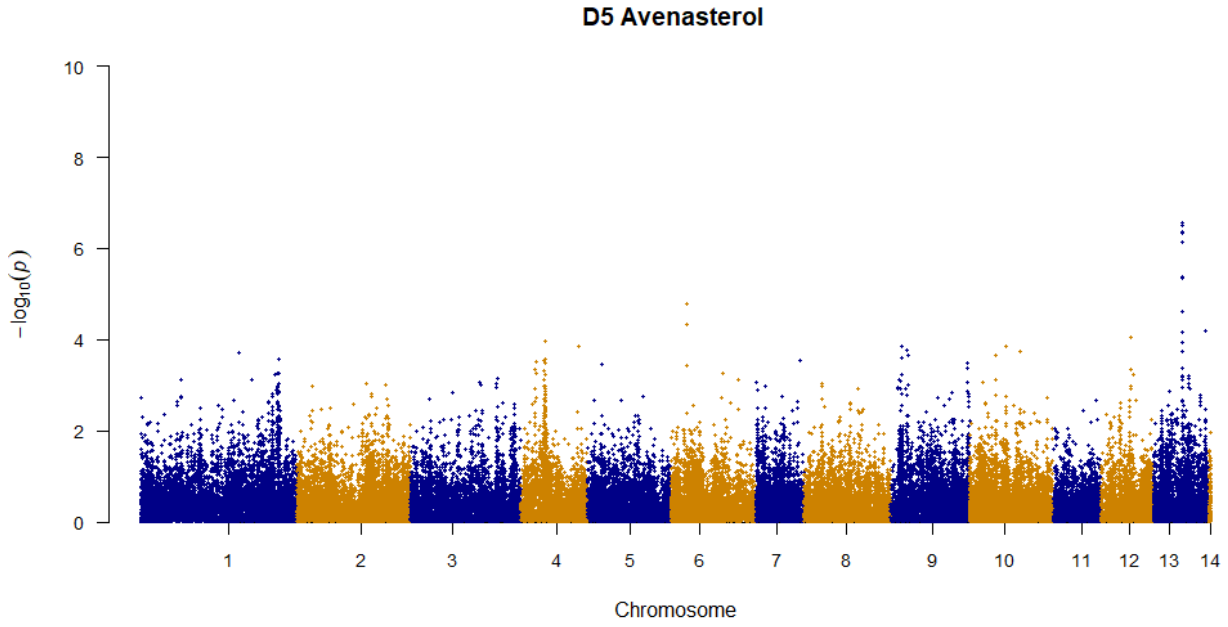


Figure 3.16 Manhattan plots of Δ^5-avenasterol

This plot shows SNP-trait associations identified using a general linear model (GLM) across the transcriptome-derived SNP dataset. Analysis included 48 accessions with 149,740 SNPs distributed across 90,751 loci and 43,456 sites. The x-axis represents the chromosome co-ordinates in bases (SNP order), while the y axis shows the–$\log_{10}$ of the association p-values. Significant peak observed on Chromosome 13 demonstrates the significance of each genetic variant across the genome, typically showing the negative logarithm of the p-value in relation to the SNP order of the variants.

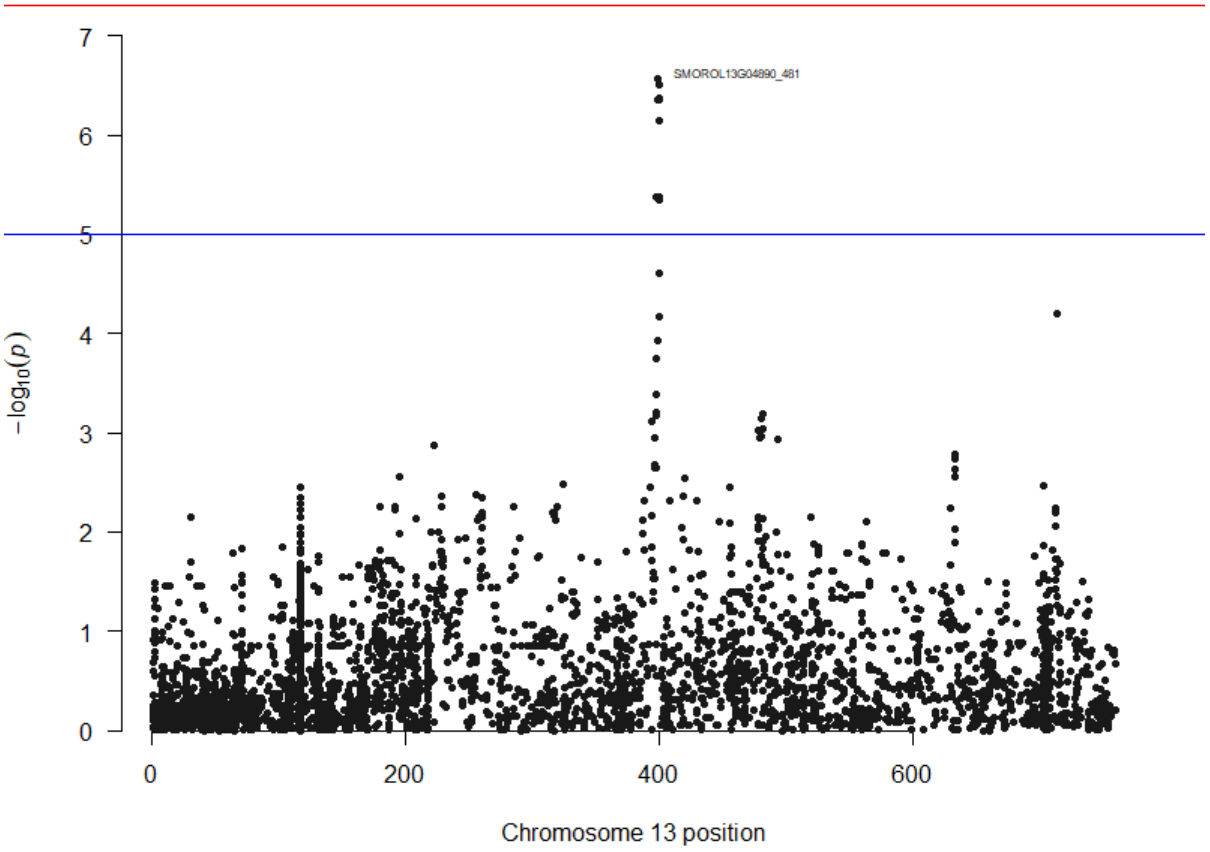


Figure 3.17 Detailed view of Chromosome 13 associated with Δ5 Avenasterol.

Regional Manhattan plot of Chromosome 13 associated with Δ5 Avenasterol shows the x-axis representing the chromosome co-ordinates in bases (SNP order), while the y-axis shows the $-\log_{10}$ of the association p-values. The horizontal red dashed line marks the Bonferroni- corrected significance threshold (5.47 x 10-7), while the blue line indicates the normal significance cut off at p <0.05. SNPs above this blue line are statistically significant at the uncorrected level and include candidates linked to key oil biosynthesis genes such as Morol13g04900_461

Δ7 Stigmastenol

The QQ plot (Figure 3.18) for Δ7 Stigmastenol, indicated a deviation from the expected line under the null, though with more inflation than for Δ5 Avenasterol. This QQ pattern implies the presence of loci contributing to Δ7 Stigmastenol variations.

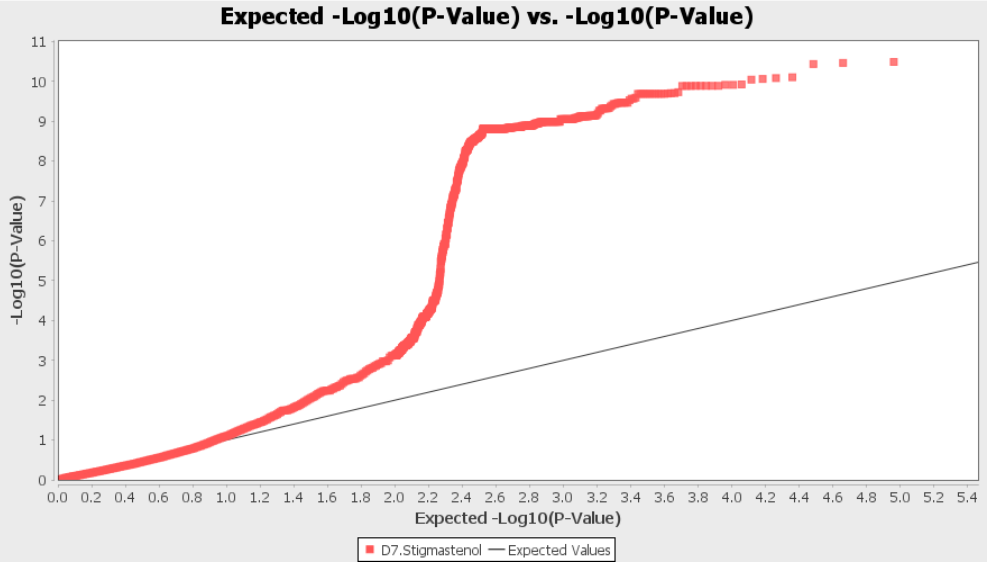


Figure 3.18 Genome-wide association analysis for Δ7Stigmastenol concentration. Quantile-quantile (QQ) plot of the expected vs observed–$\log_{10}$(p) values from the GWAS model showing deviation indicative of true associations. The x-axis represents the expected distribution, and the y-axis shows the observed distribution.

The Manhattan plot (Figure 3.19) depicts the –$\log_{10}$(p-value) of SNP-trait associations along Chromosome 5 for the oil compound Δ7 Stigmastenol. The leading SNPs surpass the blue line, which represents a suggestive significance threshold (p < 1e-5), and some SNPs exceed the red line, which likely represents a genome-wide significance threshold (Bonferroni-corrected p < 5e-8). This indicates statistically robust association of SNPs in this region with Δ7Stigmastenol levels (Figure 3.20). The background noise and other loci depicts the rest of the chromosome, a relatively uniform scatter of SNPs with low significance, exhibiting minimal background association which suggests that this trait, Δ7 Stigmastenol is likely influenced by a limited number of loci, supporting oligogenic control on this chromosome (that is the trait being influenced by a small number of genes/ a few genetic loci with moderate to large effects on the phenotype). SMorol05g04660_75 (GO:0009249) was identified.

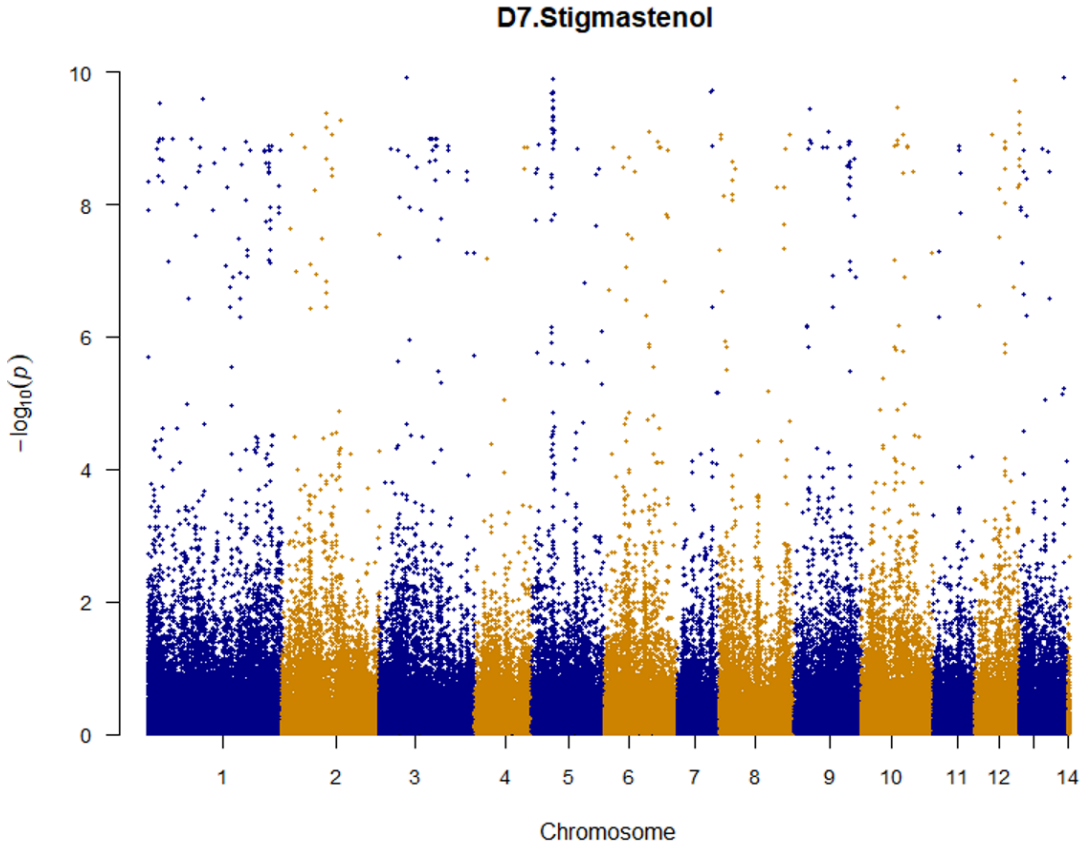


Figure 3.19 Manhattan plots of Δ7 Stigmastenol.
This plot shows SNP-trait associations identified using a general linear model (GLM) across the transcriptome-derived SNP dataset. Analysis included 48 accessions with 149,740 SNPs distributed across 90,751 loci and 43,456 sites. The x-axis represents the chromosome co-ordinates in bases (SNP) order, while the y axis shows the $-\log_{10}$ of the association p-values. Significant peak observed on Chromosome 5 demonstrates the significance of each genetic variant across the genome, typically showing the negative logarithm of the p-value in relation to the SNP position of the variants.

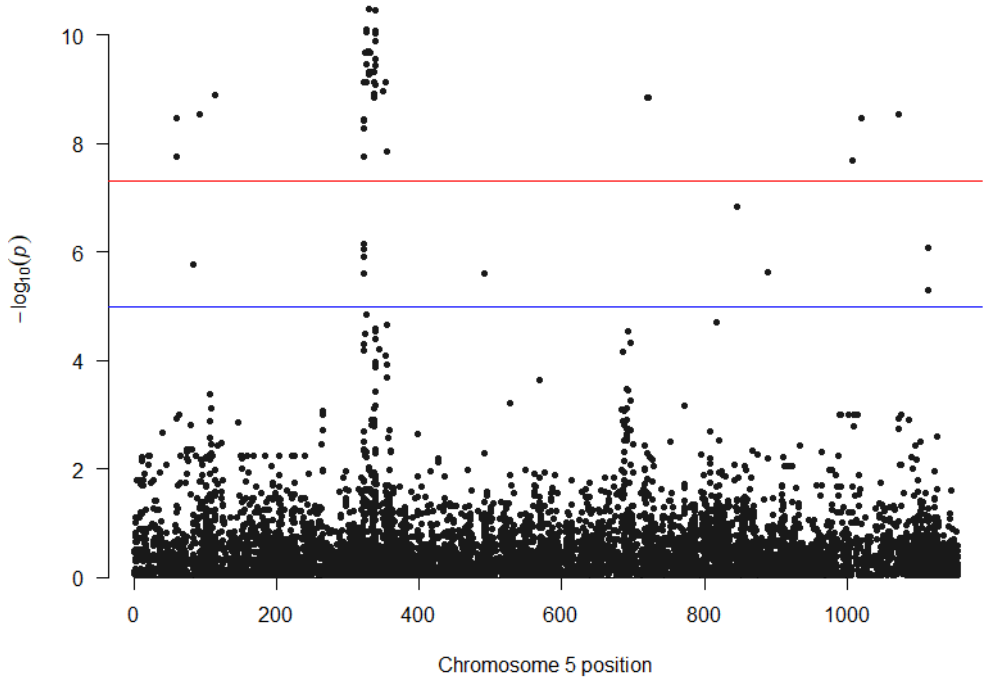


Figure 3.20 Detailed view of Chromosome 5 associated with Δ7 Stigmastenol

Regional Manhattan plot highlighting annotated genes near the lead SNPs, suggesting involvement of sterol biosynthesis-related enzymes. SNP-trait associations are plotted as $-\log_{10}$ (p-values) y-axis against the chromosome co-ordinates in bases on Chromosome 5 (x-axis). The horizontal red dashed line marks the Bonferroni- corrected significance threshold (5.47 x 10-7), while the blue line indicates the normal significance cut off at p <0.05. A cluster of SNPs surpass the genome-wide corrected significance threshold (red line), suggesting a candidate locus for variation in **Δ7 Stigmastenol** levels. SNPs exceeding the Bonferroni threshold include candidates linked to key oil biosynthesis genes such as Morol05g04660_75

Obtusifoliol

The quantile-quantile (QQ) plot for Obtusifoliol (**Figure 3.21**) demonstrated a modest deviation from the expected distribution under the null hypothesis, indicating the presence of genuine associations beyond what would be expected by chance. This deviation from the diagonal line, particularly in the tail of the distribution, suggests potential SNP position influencing the variation in Obtusifoliol concentration.

The Manhattan plot (**Figure 3.22**) revealed multiple peaks of significant single nucleotide polymorphisms (SNPs), notably a strong association signal was observed on Chromosome 12, exceeding the genome-wide significance threshold. To further

investigate this region, a focused Manhattan plot of Chromosome 12 (**Figure 3.23**) was generated.

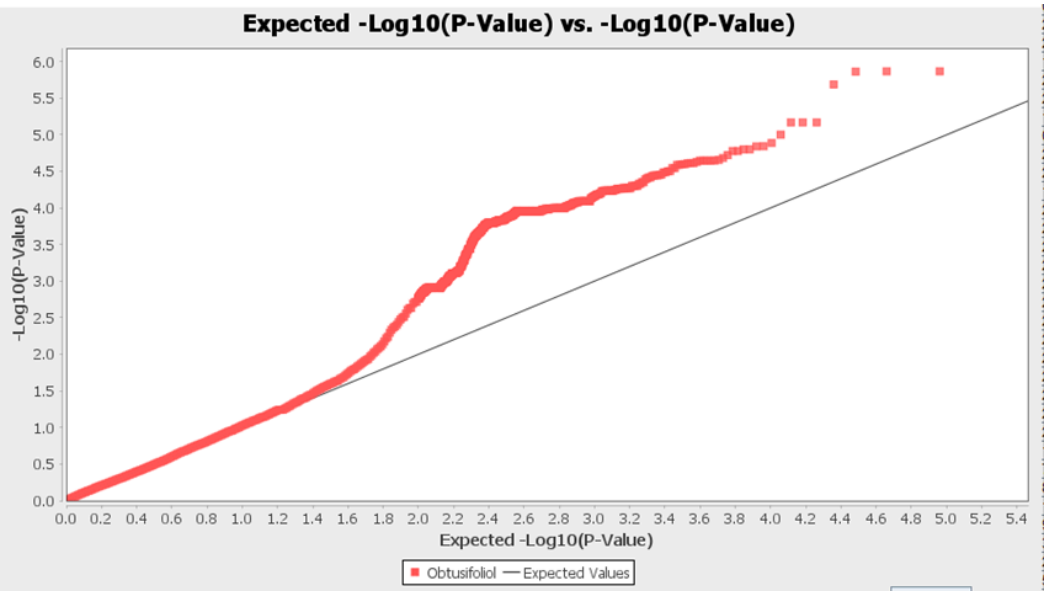


Figure 3.21 Genome-wide association analysis for Obtusifoliol concentration. Quantile-quantile (QQ) plot comparing expected and observed -log$_{10}$(p-values), showing deviation indicative of true associations.

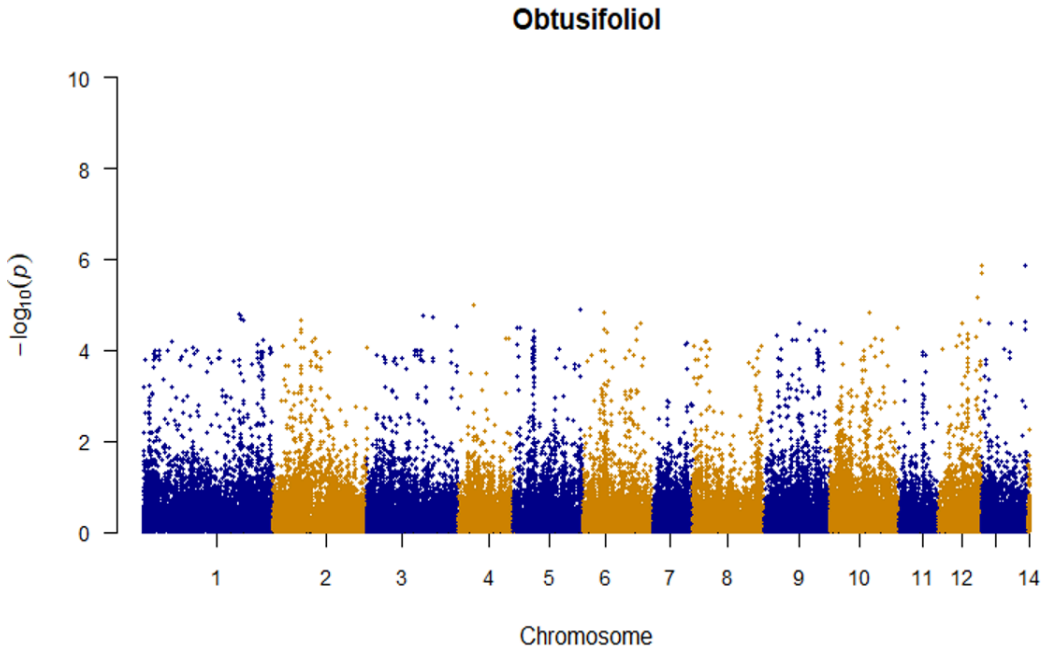


Figure 3.22 Manhattan plots of Obtusifoliol.

This plot shows SNP-trait associations identified using a general linear model (GLM) across the transcriptome-derived SNP dataset. Analysis included 48 accessions with 149,740 SNPs distributed across 90,751 loci and 43,456 sites. The x-axis represents the

chromosome co-ordinates in bases (SNP order), while the y axis shows the - $\log_{10}$ of the association p-values. Significant peak observed on Chromosome 12 demonstrates the significance of each genetic variant across the genome, typically showing the negative logarithm of the p-value in relation to the SNP position of the variants.

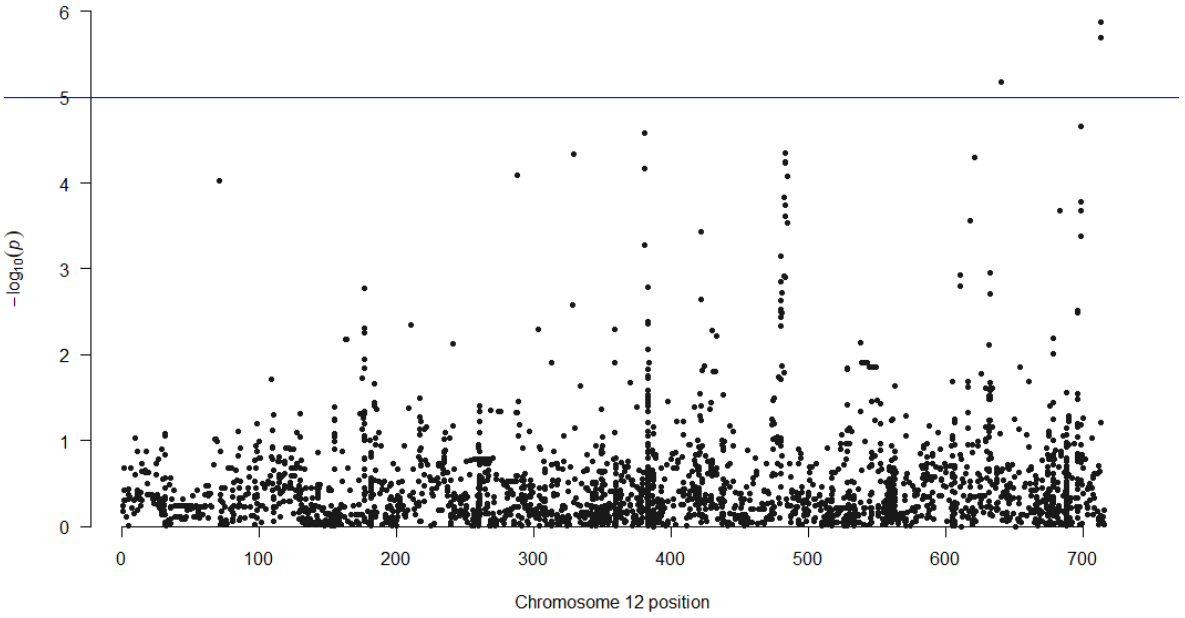


Figure 3.23 Detailed view of Chromosome 12 associated with Obtusifoliol. Regional Manhattan plot highlighting annotated genes near the lead SNPs, suggesting involvement of sterol biosynthesis-related enzymes shows the x-axis representing the chromosome co-ordinates in bases (SNP order), while the y- axis shows the - $\log_{10}$ of the association p-values. The horizontal red dashed line marks the Bonferroni- corrected significance threshold (5.47 x 10-7), while the blue line indicates the normal significance cut off at p <0.05. SNP-trait associations are plotted as $-\log_{10}$ (p-values) against physical position on Chromosome 12. A cluster of SNPs surpass the genome-wide significance threshold blue line, suggesting a candidate locus for variation in Obtusifoliol levels. The blue line represents a significance threshold. . SNPs above this blue line are statistically significant at the uncorrected level and include candidates linked to key oil biosynthesis genes such as Morol12g11600

Database search using EggNOG-mapper v2 was done and the number of SNPs located near the candidate genes associated with Phytosterols was assessed, Morol12g11600 (Cysteine-type Endopeptidase) found between 13040628 and 13042332 (1704 bp) on Chromosome12 associated with Obtusifoliol synthesis. Morol05g04660 (P450 CYP82D47) found between 2859077 and 2860702 (1625 bp) on Chromosome 5 associated with Δ7Stigmastenol. While Morol13g04900 (ABC Transporter G family) between 3008419 and 3010671bp (2,252bp) on Chromosome 13 associated with Δ^5-avenasterol.

3.6.2 FAMEs: C18.1D9 - Oleic acid

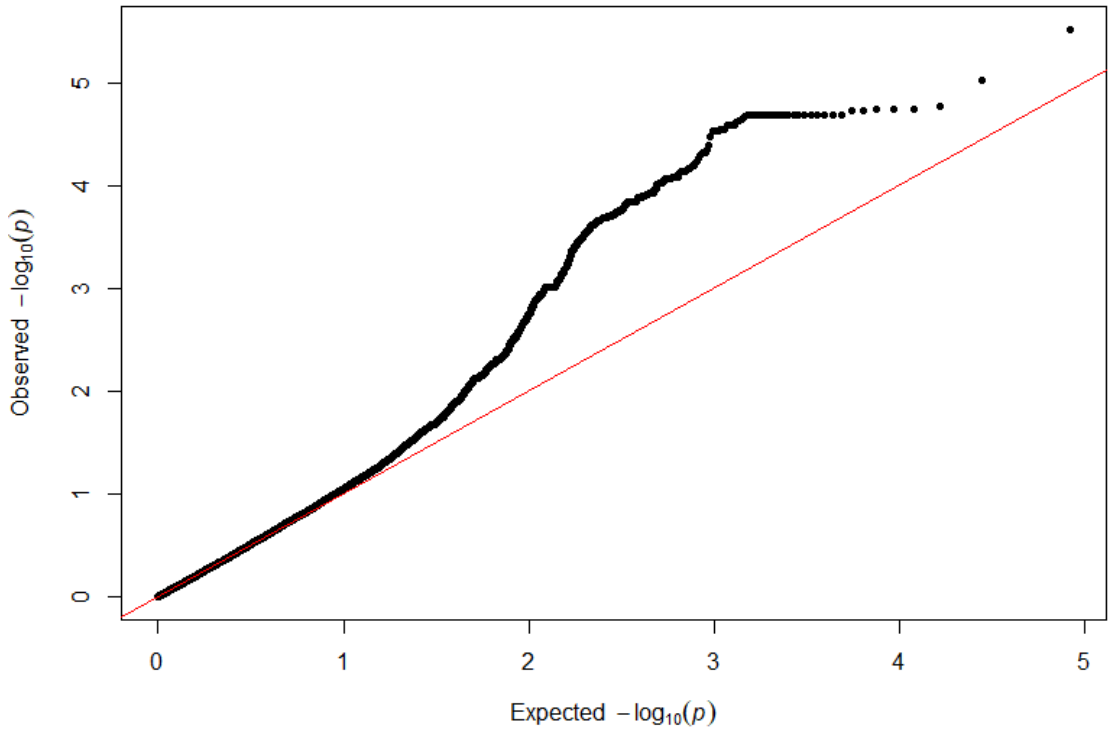


Figure 3.24 Genome-wide association analysis for C18.1D9-Oleic acid concentration. Quantile-quantile (QQ) plot comparing expected and observed -$\log_{10}$ (p-values), showing deviation from the diagonal especially at the tail. The deviation from the expected distribution indicates the presence of true associations beyond what would be expected. This plot suggests potential SNP order influencing the variation in C18.1D9 - Oleic acid concentration.

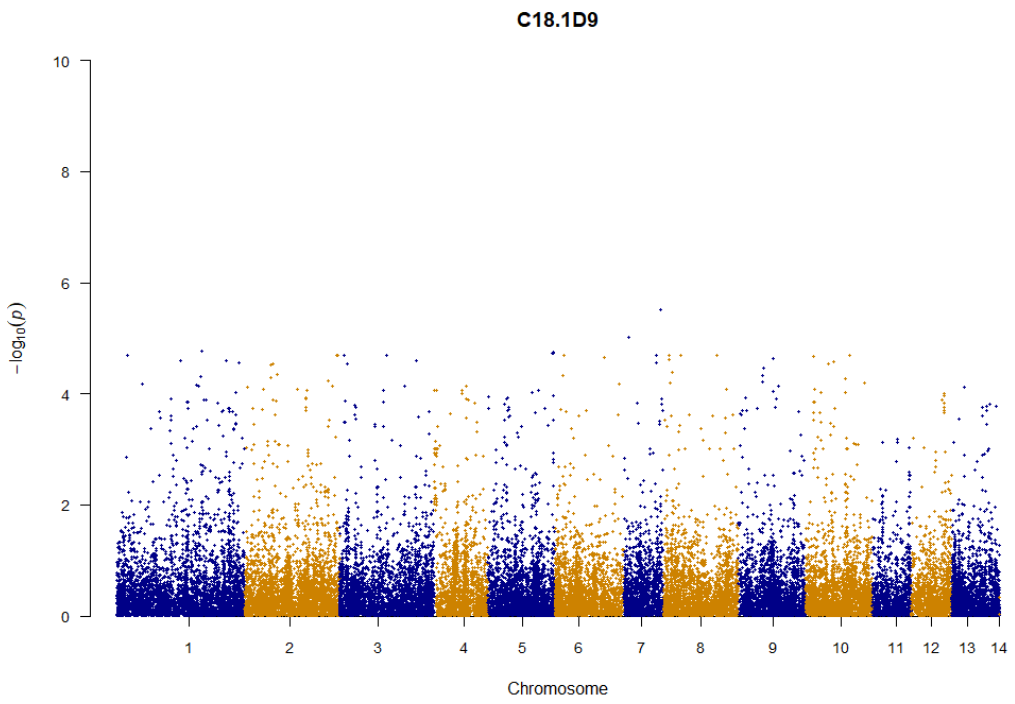


Figure 3.25 Manhattan plots of C18.1D9 -Oleic acid.
This plot shows SNP-trait associations identified using a general linear model (GLM) across the transcriptome-derived SNP dataset. Analysis included 48 accessions with 149,740 SNPs distributed across 90,751 loci and 43,456 sites. The x-axis represents the chromosome co-ordinates in bases (SNP order), while the y axis shows the - $\log_{10}$ of the association p-values. Significant peak observed on Chromosome 7 demonstrates the significance of each genetic variant across the genome, typically showing the negative logarithm of the p-value in relation to the SNP order of the variant.

The Manhattan plot (Figure 3.25) revealed multiple peaks of significant single nucleotide polymorphisms (SNPs) across the genome. Notably, a strong association signal was observed on Chromosome 7, exceeding the genome-wide significance threshold. To further investigate this region, a focused Manhattan plot of Chromosome 7 (Figure 3.27) was generated, which includes gene annotations.

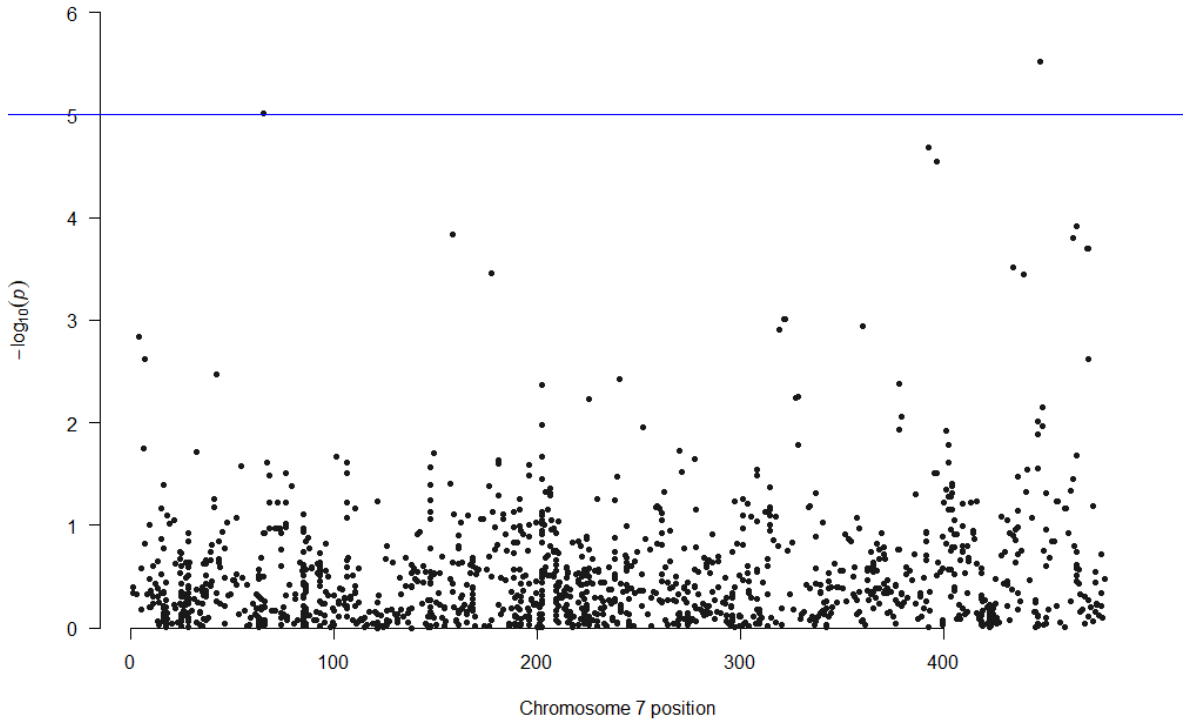


Figure 3.26 Detailed view of Chromosome 7 associated with C18.1D9 - Oleic acid. Regional Manhattan plot highlighting annotated genes near the lead SNPs, suggesting involvement of sterol biosynthesis-related enzymes. SNP-trait associations are plotted as $-\log_{10}$(p-values) against the chromosome co-ordinates in bases (SNP order) on Chromosome 7. A cluster of SNPs surpass the genome-wide significance threshold blue line, the nominal threshold ($p < 0.05$) without correction. SNPs above this blue line are statistically significant at the uncorrected level and include candidate locus SMorol07g09570_49 for variation in C18.1D9 - Oleic acid levels. The blue line represents a significance threshold.

C18.2 Linoleic acid

The quantile-quantile (QQ) plot for C18.2 Linoleic acid (Figure 3.27) demonstrated a modest deviation from the expected distribution under the null hypothesis, indicating the presence of genuine associations beyond what would be expected by chance. This deviation from the diagonal line, particularly in the tail of the distribution, suggests potential SNP order in C18.2 Linoleic acid concentration. The Manhattan plot (Figure 3.28) revealed multiple peaks of significant single nucleotide polymorphisms (SNPs) across the genome. Notably, a strong association signal was observed on Chromosome 5, exceeding the genome-wide significance threshold. To further investigate this region, a focused Manhattan plot of Chromosome 5 (Figure 3.29) was generated, which includes gene annotations.

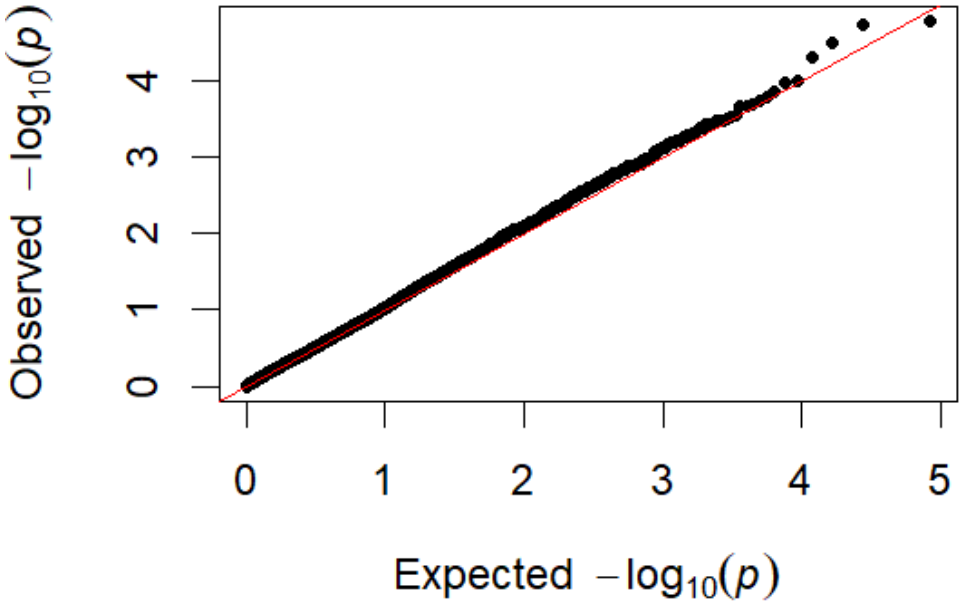


Figure 3.27 Genome-wide association analysis for C18.2 Linoleic acid concentration. Quantile-quantile (QQ) plot comparing expected and observed -$\log_{10}$ (p-values), showing deviation is indicative of true associations.

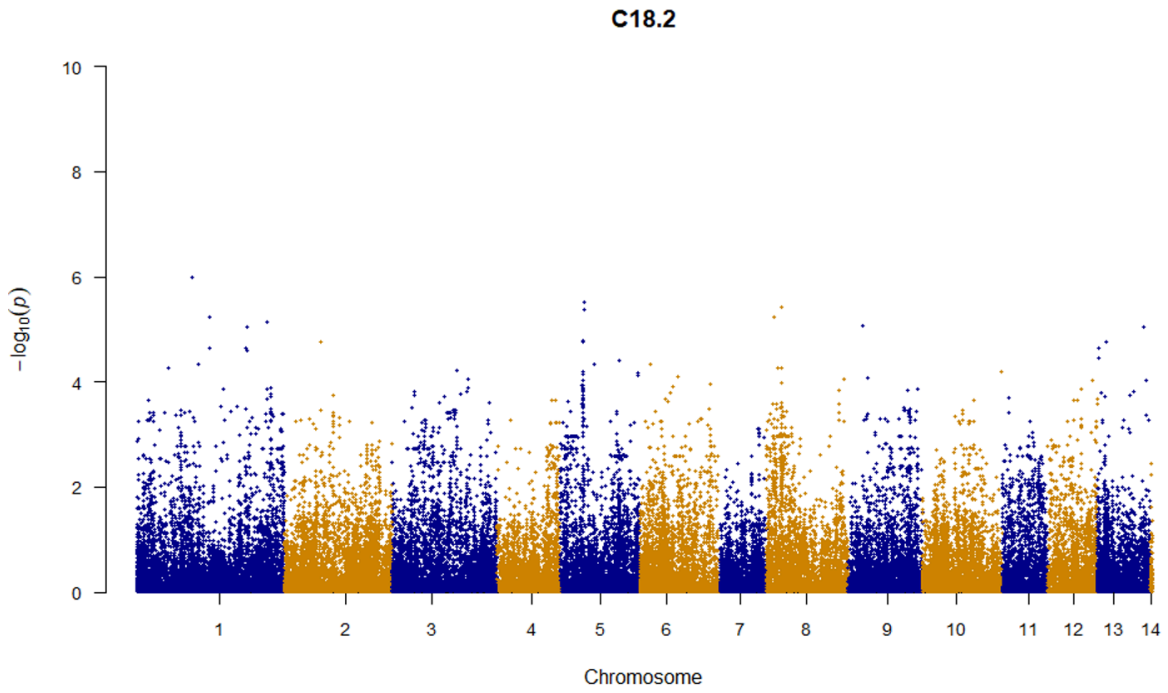


Figure 3.28 Manhattan plots of C18.2 - Linoleic acid.
This plot shows SNP-trait associations identified using a general linear model (GLM) across the transcriptome-derived SNP dataset. Analysis included 48 accessions with 149,740 SNPs distributed across 90,751 loci and 43,456 sites. The x-axis represents the chromosome co-ordinates in bases (SNP order), while the y axis shows the - $\log_{10}$ of the association p-values. Significant peak observed on Chromosome 5 demonstrates the significance of each genetic variant across the SNPs, typically showing the negative logarithm of the p-value in relation to the SNP order of the variant.

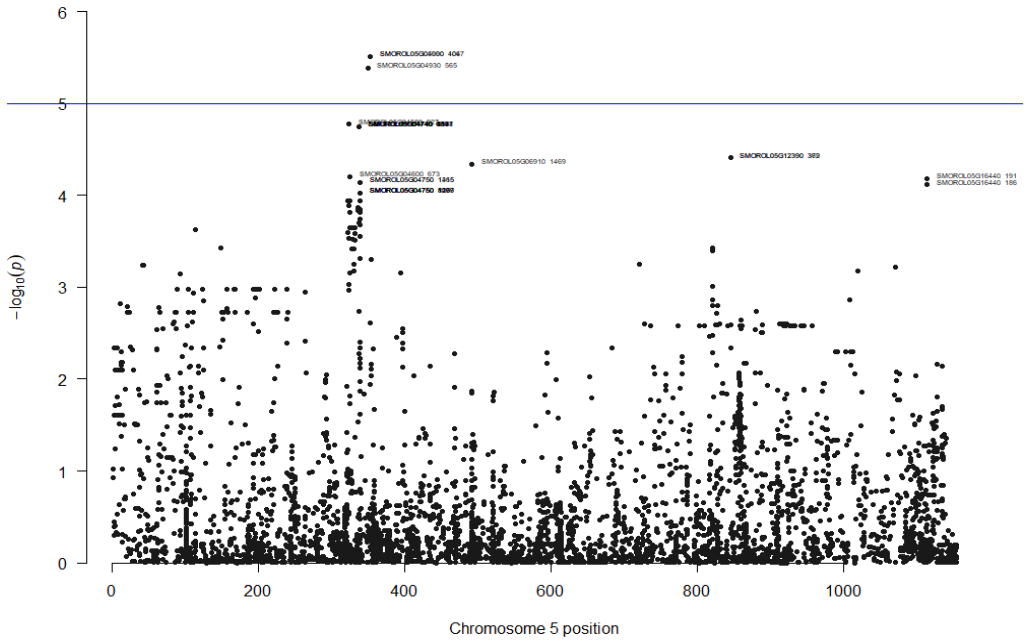


Figure 3.29 Detailed view of Chromosome 5 associated with C18.2 - Linoleic acid. Regional Manhattan plot highlighting annotated genes near the lead SNPs, suggesting involvement of sterol biosynthesis-related enzymes shows the x-axis representing the chromosome co-ordinates in bases (SNP order), while the y- axis shows the - $\log_{10}$ of the association p-values. The horizontal red dashed line marks the Bonferroni- corrected significance threshold (5.47 x 10-7), while the blue line indicates the normal significance cut off at p <0.05. SNP-trait associations are plotted as $-\log_{10}$(p-values) against physical position on Chromosome 5. The several labelled points mark SNPs with strong association signals linked to the trait, A cluster of SNPs surpass the genome-wide significance threshold blue line, suggesting a candidate locus for variation in **C18.2 - Linoleic acid** levels. The blue line represents a significance threshold. SNPs exceeding the blue line are statistically significant at the uncorrected level and include candidates linked to key oil biosynthesis genes such as Morol05G04930

Database search using EggNOG-mapper v2 was done and the number of SNPs located near the candidate genes associated with FAMEs was assessed and Morol07g09570 identified as Translocase of Inner Mitochondrial Membrane (TIM22) and responsible for protein transmembrane transporter activity, significant for Oleic acid was found on Chromosome 7 between site 14574844 and 14576678 with the length of interval being 1,834bp. The ENT-kaurene oxidase thought to be encoded by the gene MOROL05G04930 found significant for Linoleic acid was identified on Chromosome5 between site 3057992 and 3061232 with 3,240 bp.

3.6.3 Tocopherols- Beta Tocopherol

The quantile-quantile (QQ) plot for Beta-Tocopherol (Figure 3.30) demonstrated a modest deviation from the expected distribution under the null hypothesis, indicating the

presence of genuine associations beyond what would be expected by chance. This deviation from the diagonal line, particularly in the tail of the distribution, suggests potential SNP order influencing the variation in Beta-Tocopherol concentration. The Manhattan plot (Figure 3.31) revealed multiple peaks of significant single nucleotide polymorphisms (SNPs) across the genome. Notably, a strong association signal was observed on Chromosome 2 & 8, exceeding the genome-wide significance threshold. To further investigate this region, a focused Manhattan plot of Chromosome 2 & 8 (Figure 3.32) was generated, which includes gene annotations.

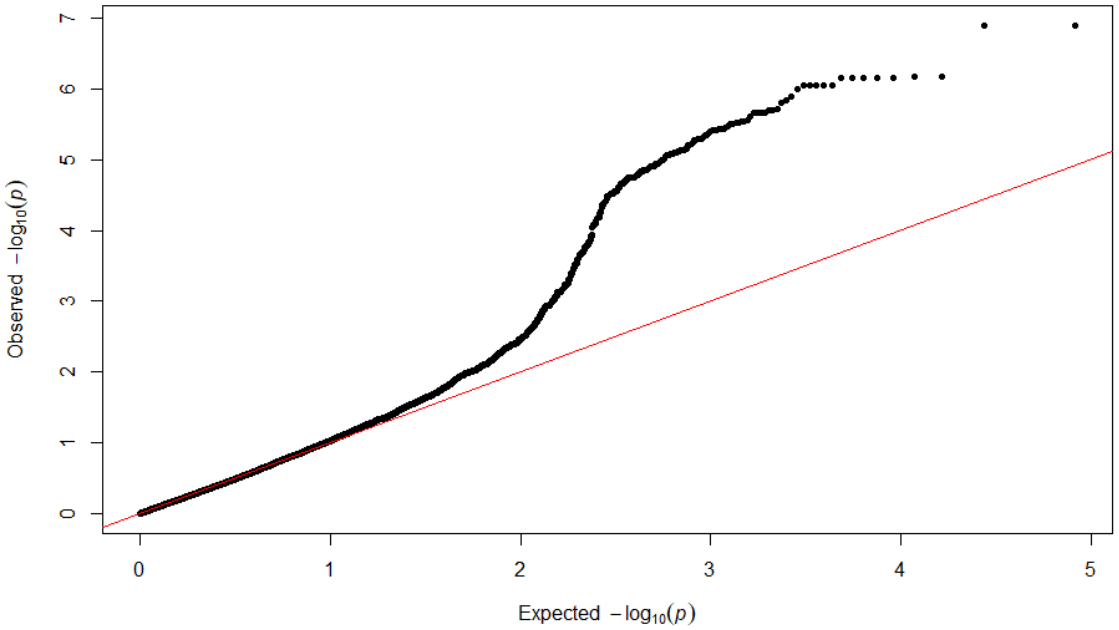


Figure 3.30 Genome-wide association analysis for Beta-Tocopherol concentration. Quantile-quantile (QQ) plot comparing expected and observed -$\log_{10}$(p-values), showing deviation indicative of true associations.

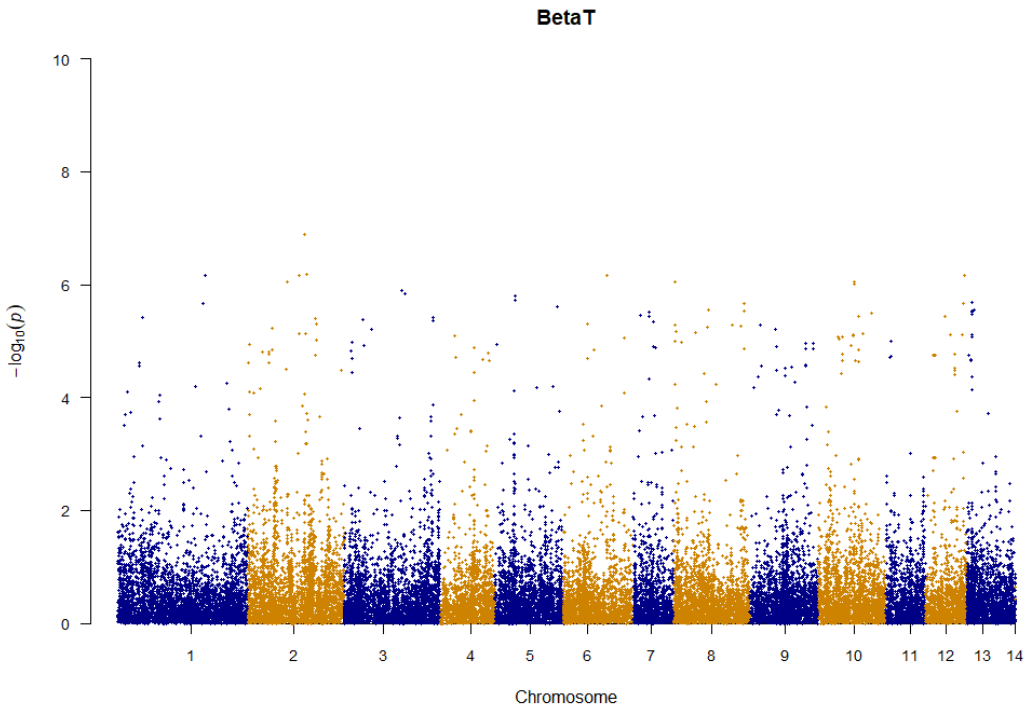


Figure 3.31 Manhattan plots of Beta Tocopherol.

This plot shows SNP-trait associations identified using a general linear model (GLM) across the transcriptome-derived SNP dataset. Analysis included 48 accessions with 149,740 SNPs distributed across 90,751 loci and 43,456 sites. The x-axis represents the chromosome co-ordinates in bases (SNP order), while the y axis shows the - $\log_{10}$ of the association p-values. Multiple peaks of significant SNPs were observed across the transcriptome / genome demonstrates the significance of each genetic variant across the genome, typically showing the negative logarithm of the p-value in relation to the SNP order of the variant.

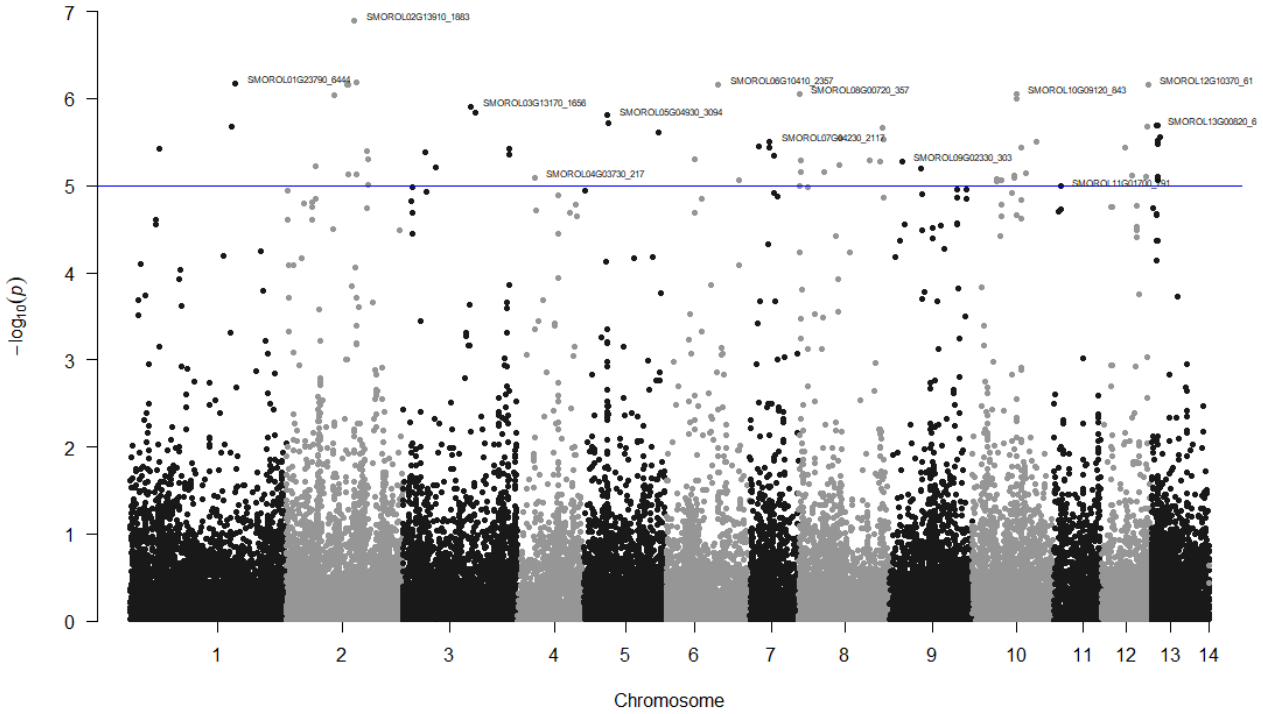


Figure 3.32 Annotated view of Manhattan plots of Beta Tocopherol. Regional Manhattan plot with multiple chromosome view, shows the x-axis representing the the chromosome co-ordinates in bases (SNP order), ordered by loci, while the y- axis shows the - $\log_{10}$ of the association p-values. The blue line indicates the normal significance cut off at p <0.05. SNPs which exceed threshold blue line, the several labelled points marks SNPs with strong association signals linked to the trait, especially the strongest peaks at chromosomes 2 and 8 indicate candidate genes such as Morol02g15870, Morol08g09340 and Morol10g10410 are statistically significant at the uncorrected level and linked to variation in **Beta Tocopherol** levels.

Database search using EggNOG-mapper v2 was done and the number of SNPs located near the candidate genes associated with Tocopherols was assessed and significant for β-tocopherols was found on Chromosome 2: Morol02g15870 identified for the regulation of photomorphogenesis (LCA) was found between sites 10669439 and 10672454 (3015); Morol08g09340 (SPB) also found on Chromosome 8 was found between 9763345 and 9753391(46 bp) and likewise Morol10g10410_2423 (TFIIE) between 10279819 and 10282271(2452 bp) on Chromosome10.

3.7 Reproducibility of Results

The data trends were reproducible across replicates and experiments as the oil yield obtained were consistent across experiments for the same accession. Code and scripts in Github repository (https://github.com/andreaharper/HarperLabScripts/DorcasOjo_thesis scripts/R). Analysis was run with varying data set at least three times to validate outcomes.

3.8 Discussion

Seed Trait Discrepancies

Unusually, one of the *Moringa oleifera* seeds had brown endosperm instead of white but did not show a different oil content than other repeats. This observation can be attributed to several factors which include genetic mutations, environmental conditions and potential fungal contamination. Genetic variations sometimes cause differences in seed colour and composition (Anwar *et al.,*2005). The brown endosperm could be a result of mutations affecting the biosynthesis pathways of pigments or storage compounds as evidenced in *Brassica napus* (Kaidi Yu *et al*.,2023). On the other hand, fungal pathogens may be present as the seeds had been stored at room temperature for about a year before the oil extraction was done, which could alter the seed's endosperm colour but not necessarily alter the seed's biochemical composition (Wawrzyniak *et al.,* 2018). This might suggest that the fungus is not significantly affecting the oil-producing tissues. The oil consistency despite the brown colour of the endosperm could be that the oil content does not necessarily correlate with endosperm colour as plants may have compensatory mechanisms that adjust oil production despite variations in seed morphology or colour (Song, *et al.,* 2020; Vanhercke, *et al.,*2018).

Low Oil Yield Despite Large Seeds

The connection between how big seeds are and how much oil they produce is complicated. It is not just about the genes of the plant; other things matter too. For example, a healthy plant finds it simpler to translate nutrients into oil (Cai, *et al.,*2023). But if the plant is stressed by bad weather or pests, it might focus on surviving instead of making oil, which can lower the oil content, no matter the seed size. When to harvest the seeds is also important. If seeds are picked too early, they will not have enough oil (Kale*, et al.,*2020). If they are harvested too late, they might spoil or lose quality, making it harder to get oil from them (Zhang, *et al.,* 2021; 2020). Farming methods, like how the soil is managed, how plants are watered, and how fertilizers are used can greatly affect both seed growth and oil production. For instance, having the right nutrients can boost oil output, while a lack of them can hinder growth and reduce oil levels *(*Zhang*, et al.,2020)*. Additionally, how crops are rotated or grown together can impact soil health and nutrient supply, which also plays a role. In short, while bigger seeds might seem better, they may not always mean more oil. A complete approach that looks at genetics, the environment, and farming practices is necessary to get the most oil from seed crops (Subedi, *et al.,*2020).

Seed Variability and Oil Yield

Low oil weight (156.3, 203.3, 135.3, 180.3mg/100g) and low oil% (32.3,32.7,34.2,35.0) was observed in specific samples (KMs5, KMs8, KMs10 and KMs11) despite their larger seed weight (483.7, 626.0, 405.3, 516.0mg). There was a lack of correlation between seed weight and oil yield, as noted with smaller seeds (e.g. MaMo1, 194.4mg; MO25,141.8mg) producing higher oil weight (88.9mg/100g, 62.1mg/100g) and low oil% (45.5%, 44.2%) compared to larger seeds (e.g. KMs5, KMs8, KMs10 and KMs11).

Oil Extraction Efficiency

The mean Oil Extraction Efficiency (OEE) calculated as average oil weight (mg) divided by average Oil% multiplied by the average seed weight (mg) multiplied by 100.

Mean OEE = 0.9996 or 99.96% (2 dp)

df['OilExtractionEfficiency'] = (df['Avrg.\n Oil weight\n (mg)'] / (df['Avrg. Oil%'] * df['Avrg.\n Seed weight\n (mg)'])) * 100 is 0.9996135

Hexane Extraction Efficiency

The Average Oil weight and Hexane volume showed consistency across multiple experiments, especially concerning samples like KMs5, KMs8, KMs10 and KMs11 which showed deviations from the expected efficiency. This indicates the reproducibility of the extraction method.

Outlier Sample Analysis

The Oil consistency despite the brown colour of the endosperm in MO64-A sample (*Moringa oleifera*, Katsina, Nigeria) could be due to the storage condition which might have allowed fungal growth, genetic mutation or environmental factors. Whatever might have caused it, the oil content does not necessarily correlate with endosperm colour as plants may have compensatory mechanisms that adjust oil production despite variations in seed morphology or colour (Song, *et al.,* 2020; Vanhercke, *et al.,*2018).

Influence of Geographic Latitude on Genetic Variation in *Moringa spp..*

Results from the PCA when integrated with correlation and regression analyses, indicate that geographic latitude exerts only a marginal influence on genetic variation among the Moringa populations that were sampled. The observed weak negative correlation ($r = -0.23$, $p=0.203$) and the minimal explanatory power derived from a regression analysis ($R^2 = 0.05$) suggest that latitude accounts for merely 5% of the variation observed in Principal Component 1 (PC1). I can infer from this result that genetic variation is not dependent of

latitudinal gradients. Several factors may contribute to this observed pattern. Firstly, these Moringa seeds could have undergone extensive gene flow among populations through mechanisms of natural dispersal or anthropogenic seed translocation, which can serve to homogenize genetic variation and mask geographic structuring. Secondly, alternative environmental or ecological determinants such as soil composition, precipitation patterns, or local adaptive strategies could exert a more significant influence on shaping genetic variation than latitude in isolation. Lastly, the genetic structure represented by PC1 may be indicative of historical population dynamics or founder events rather than contemporary spatial arrangements. Altogether, these findings imply that although latitude may exert some degree of influence, it is not a principal determinant of genetic differentiation within Moringa. Future research endeavors could incorporate additional environmental variables, landscape characteristics, or higher-resolution genomic markers to enhance the understanding of the factors driving population structure.

Variability in Fatty acid profile of Moringa seed oil

This research produced 15 derivatives of fatty acids derived from Moringa seed oil. Oleic acid (C18:1D9) is the predominant fatty acid in MO57, exhibiting a concentration of 956.58mg, which accounts for 68.51% of the total derivative composition documented in the study. This result is consistent with the research conducted by Leone *et al*., 2016 and Basuny (2016), which reported similar oleic acid concentrations of over 70% and 65% in *Moringa oleifera*, respectively. Rahman *et al*., (2014), reported a concentration exceeding 80% oleic acid. All the Moringa samples yielded considerably high concentration of Oleic acid C18.1D9 (Figure 3.6) which is supported by previous research on *Moringa stenopetala* (Pluháčková *et al.,* 2023) and *Moringa oleifera* seed oil (Nadeem & Imran, 2016). These insights will give informed information for crop breeders where these traits are desirable.

Behenic acid (C22:0) is a saturated fatty acid that's generally considered less desirable in edible oils. However, in Moringa oil, it makes up only 6.3% of the total fatty acid composition, which is a relatively small amount. At this low level, it's unlikely to pose any health concerns and may actually be beneficial, particularly for industrial applications such as cosmetics, lubricants, detergents, coatings, and pharmaceuticals. In addition, Moringa oil is rich in unsaturated fatty acids, which further offsets the presence of behenic acid and makes the oil suitable for cooking as well.

The statistical evaluation (ANOVA, Tukey HSD, regression) revealed geographical influence on specific oil constituents, notably within fatty acids (FAMEs), tocopherols, and phytosterols. While not all compounds showed significant variation, a select group

exhibited clear geographical trends, environmental associations (e.g., with latitude), or inter-location differences.

Does location influence Fatty Acids (FAMEs) in Moringa overall regardless of species?

Variation and Environmental Effects

C22.1n13 (Erucic acid), C20.0 (Arachidic acid), C20.1n11 (Gadoleic acid), C24.0 (Lignoceric acid) showed statistically significant variation (ANOVA $p < 0.05$ indicating that geography may influence their biosynthesis.

C22.1n13 (Erucic acid) showed significant association with latitude, suggesting that environmental factors such as altitude or temperature linked to latitude might affect its accumulation.

Post-hoc Tukey HSD tests further confirmed that some regional means differ significantly, especially between high-yielding and low-yielding zones.

These fatty acids, C22.1n13 (Erucic acid), C20.0 (Arachidic acid), C20.1n11 (Gadoleic acid), C24.0 (Lignoceric acid) can be considered key markers for selecting superior accessions suitable for dietary and industrial purposes.

Geographical Influence on Specific Fatty Acids

The high concentration of vaccenic acid an isomer of oleic acid in Kenya *M. oleifera* and its absence in Kenya *M. stenopetala* can be attributed to a combination of genetic, metabolic, environmental, and ecological factors. *M. oleifera* and *M. stenopetala* belong to different species within the Moringaceae family. Their distinct genetic make-ups can lead to variations in metabolic pathways, influencing the types and concentrations of fatty acids produced. Certain regulatory mechanisms may control the expression of genes involved in fatty acid synthesis, resulting in *M. oleifera* producing higher levels of vaccenic acid. It may also have a greater capacity to store and accumulate vaccenic acid in its tissues compared to *M. stenopetala*. Likewise, the nutrient profile of the soil in which these species grow, variations in temperature, humidity, and sunlight can affect plant metabolism can influence their biochemical pathways and the synthesis of secondary metabolites. *M. oleifera* may have adapted to specific ecological niches in Kenya that favour the production of vaccenic acid. Differences in agricultural practices

between the two species such as irrigation, fertilization, and pest management may also impact the chemical composition of the plants.

Unique Fatty Acid Profiles

The presence of lignoceric acid (C24:0) uniquely in Kenya *M. stenopetala* accessions and its absence in other samples may be due to variations in soil types, nutrient availability, climate, and agricultural practices. *M. stenopetala* may have a more efficient or specialized pathway for elongating fatty acids, allowing it to produce lignoceric acid as it is produced through the elongation of its precursors palmitic acid /stearic acid which are shorter fatty acids. In addition, *M. stenopetala* may have developed unique adaptations to its environment that include producing specific fatty acids like lignoceric acid, which may serve protective roles (for instance in stress responses or defence against herbivores).

With regards to whether location influence FAMEs in *Moringa oleifera* samples only, all the fatty acids did not vary with location or latitude. This reveals the stable expression of FAMEs as they were not influenced by environmental or genetic factors.

β-Tocopherol in *Moringa spp.*: Environmental Influence, Genetic Variation, and Nutritional Potential

Although β-tocopherol (β-T), a key antioxidant form of vitamin E, has been detected in *Moringa* seed oil, its environmental drivers remain largely unexplored. Unlike α-tocopherol, which has received more attention, there are no studies directly linking β-tocopherol levels in *Moringa oleifera* to environmental variables such as latitude, temperature, or rainfall. One recent investigation by Kesler *et al*., (2025) examined α-tocopherol responses in leaf and seed extracts under varying environmental conditions but did not focus on β-T. Previous reports often noted β-T as being below detectable levels in *Moringa* oil (Tsaknis *et al*., 1999, 2002; Anwar & Bhanger, 2003; Özcan *et al*., 2019). However, Pluháčková *et al*., 2023 provided updated evidence, detecting β-Tocopherol at 3.6 mg/100g in *M. oleifera* and 6.09 mg/100g in *M. stenopetala*, where it was the second most abundant tocopherol isomer after α-Tocopherol (115.60 mg/100g), these are similar to my findings of β-Tocopherol in KMs8 (5.13 ± 2.50) and PMO12 (3.20 ± 0.87)mg/kg, but lower α-Tocopherol of 60.77 ± 1.82mg/kg in MO27 (Supp.Table 3.2). Similarly, Julia *et al*., (2015) previously reported β-Tocopherol concentrations of 9.04 mg/100g, though without examining environmental influences. Despite various reviews on the phytochemical

diversity of *Moringa* by Panova *et al.*,2025 and Pareek *et al.*, 2023, the environmental influence of β-Tocopherol in seed oil has not been fully addressed.

Environmental and Genetic Influences on β-Tocopherol

Tocopherol biosynthesis in plants is known to respond to environmental stress. β-Tocopherol helps protect cellular membranes under heat, drought, and oxidative stress, functioning as part of the plant's defense system. Similar responses have been observed in cereals and *Arabidopsis*, where β-T levels varied across latitude and climatic zones (Szymańska *et al.*, 2014; Savignac *et al.*, 2022), suggesting that β-Tocopherol in *Moringa* may also be under environmental control. Notably, differences in β-Tocopherol content between Indian and African *Moringa* accessions may be attributed to genetic variation in tocopherol biosynthetic genes such as VTE1 and VTE4, or localized environmental pressures like humidity, temperature, or solar intensity. For instance, the Philippine's shows a unique tocopherol profile, likely due to climatic influences and distinct gene regulation.

Tocopherol Variation and Breeding Potential

While β-Tocopherol has lower vitamin E activity compared to α-T, it remains important for nutritional quality, oil stability, and industrial applications in nutraceuticals and cosmetics. Certain accessions, such as Kenyan *M. stenopetala* (KMs8), exhibit high α-tocopherol (63.78 ± 0.57 mg/100g) with very low γ-Tocopherol (1.64 ±0.58 mg/100g), suggesting an active VTE4 gene, which converts γ-T to α-T (Zhang *et al.*, 2022; Stahl *et al.*, 2019; Bergmüller *et al.*, 2003). This supports the hypothesis that tocopherol composition is shaped by both genetic and environmental factors Supp.Table 3.2.

Comparative analysis also reveals that *Moringa* oils can outperform olive oil in total tocopherol content, offering a potent antioxidant profile (Abdulkarim *et al.*, 2005). The high γ-Tocopherol content in Malian *M. oleifera* (MaMo1) further reflects species-specific metabolic differences, influenced by both genetics and agroecological conditions.

The considerable variation in β-tocopherol content across *Moringa* species and regions point out the importance of integrating genetics with environmental analysis. Understanding the connection between genotype (e.g., VTE gene variants) and local climate could enhance breeding strategies aimed at producing *Moringa* cultivars with optimized antioxidant profiles, tailored for both nutritional enhancement and climate resilience.

Does location influence Tocopherol yield in all samples regardless of species?

α-Tocopherol (AlphaT) exhibited a statistically significant negative regression with latitude, implying higher concentrations in southern (lower-latitude) locations. This is important because α-Tocopherol is the most bioactive vitamin E component, and its variation suggests that Moringa from certain environments may be more nutritionally potent. Breeders could select low-latitude accessions to enhance nutritional antioxidant profiles in future cultivars.

Phytosterol Diversity and Environmental influence in Moringa Seed Oil

The phytosterol composition of Moringa oilseed: *Moringa oleifera* and *Moringa stenopetala* demonstrate significant variability across accessions, particularly for β-sitosterol, ergostadienol, campesterol, stigmasterol, stigmastanol, clerosterol, obtusifoliol, Δ^5-avenasterol, and Δ^7- stigmastenol. This variation is shaped by both genetics and environmental influences.

For instance, Δ^5-avenasterol levels in *Kenya Moringa oleifera* (e.g., Mo4) showed high intra-accession variability, suggesting either heterogeneous genetic backgrounds or fluctuating environmental conditions. In contrast, accessions like KMs5 and KMs8 (Kenya *M. stenopetala*) exhibited consistently low levels, possibly reflecting genetic uniformity or stable environmental conditions. Accessions such as IMo9 (India) and WOT (a commercial *M. oleifera* line) presented outlier profiles in Δ^5-avenasterol content, pointing to unique genetic traits or cultivation environments that may be harnessed in targeted breeding or nutraceutical development. Similarly, Δ^7-stigmastenol levels varied notably among accessions, low in MO2, Mo3, Mo29, and WOT, yet more variable in Mo26 indicating that both genotype and environment modulate its biosynthesis (Moreau *et al*., 2002; Piironen *et al*., 2000).

Despite numerical differences in other minor sterols like campesterol, campestanol, and citrostadienol, statistical analysis (ANOVA and regression) found no significant geographical associations, suggesting that local environmental variation may mask broader regional effects.

Interestingly, cholesterol and cholestanol which are typically minor in plants were found in significant levels in some Moringa accessions. Cholesterol, although less abundant than β-sitosterol or stigmasterol, contributes to membrane stability under abiotic stress conditions such as heat, dehydration, and oxidative environments (Rogowska & Szakiel, 2020; Hartmann, 1998). Elevated cholesterol levels in samples from arid regions may reflect metabolic adaptation to ecological stress. Cholestanol (dihydrocholesterol), biosynthesised from both campestanol and cholesterol, was detected in high amount across all samples including the *M. oleifera* and *M. stenopetala* species. As cholestanol's saturated ring makes it more stable and less prone to oxidative damage, it may contribute to oil shelf life and stress resilience.

Stigmasterol, an important functional sterol, was detected consistently but at varying levels. As much as it is known for lowering cholesterol and fighting against inflammation, stigmasterol also helps maintain membrane fluidity under stress and enhances the oil's nutraceutical value (Aboobucker & Suza, 2019; Stermer *et al.*, 1994; Bach, 1986). The relative balance of stigmasterol and β-sitosterol often shifts in response to abiotic factors such as salinity or drought, serving as a proxy for environmental adaptation.

Comparative analyses across regions for instance Nigeria, Kenya, and Burkina Faso reveals the influence of genotype–environment (G-E) interactions on phytosterol profiles. These compounds, particularly stigmasterol, cholestanol, and cholesterol could be regarded as chemotaxonomic markers and indicators of environment - dependent trait expression.

From a nutritional and industrial standpoint, these phytosterols are of high interest. β-sitosterol, stigmasterol, and campesterol are recognized for their cardiovascular benefits and antioxidant properties (Moreau *et al.*, 2002; Piironen *et al.*, 2000). Meanwhile, cholestanol and Δ^7- stigmastenol contribute to oxidative stability, enhancing oil shelf life and suitability for pharmaceutical, cosmetic, and functional food applications. The sterol fingerprint of Moringa oil offers promise for traceability and authentication in premium markets. Variation in sterol composition, if well-characterized, could help standardize Moringa oil products by geographic origin or functional profile.

The GWAS findings for the phytosterols components, especially such as Δ5 Avenasterol, Δ7 Stigmastenol, and Obtusifoliol identified SNP regions associated with these compounds particularly the strong signals on chromosome 13 of Δ5 Avenasterol, chromosome 5 of Δ7 Stigmastenol, chromosome 12 of Obtusifoliol.

The ABC transporter family, which includes the protein encoded by this gene SMorol13g04900_130, facilitates the movement of various molecules including proteins and phytohormones across cellular membranes, and is also involved in functions such as cholesterol efflux. This mechanism may be especially beneficial for developing drought-resistant plants. Key regions on chromosome 13, house genes involved in transport, storage, membrane fusion, redox regulation, and cellular repair. Specifically, the ABC transporter G family member 6-like gene play a vital role in cholesterol transport and the broader movement of molecules across membranes, as highlighted by Gräfe & Schmitt (2021).

The Manhattan plot for Chromosome 5 related to Δ7 Stigmastenol showed multiple significant peaks, indicating the presence of genetic loci or regulatory elements that may play roles in its biosynthesis, transport, or accumulation. Among the key genes, Morol05g04660 gene (annotated with GO:0009249 and GO:0033819) which encodes Cyt P450 CYP82D47 enzyme is likely involved in synthesis of steroids, detoxification, neutralizing environmental pollutants, defense against dehydration (Chakraborty, *et al*., 2023), pathogens (Wang *et al*., 2024) and biosynthetic pathways.

The synthesis of obtusifoliol is known to occur through a series of enzymatic reactions, starting with the transformation of farnesyl pyrophosphate into squalene, then progressing through 2,3-oxidosqualene, lanosterol, and cycloartenol. Specific enzymes drive oxidation, demethylation, and structural rearrangements that modify the sterol backbone (Harshad *et al*.,2025).

The significant sterol variation observed in β-sitosterol, ergostadienol, campesterol, stigmasterol, stigmastanol, clerosterol, obtusifoliol, Δ^5-avenasterol, and Δ^7- stigmastenol across the various samples can be biologically explained in several ways. Firstly, in the context of plant sterol biosynthetic pathway, many of these sterols are precursors, intermediates, branch point products and end products (**Supp. Figure 3.12).** Generally, seeds sterols arise from the mevalonate pathway, producing squalenne and subsequently cycloartenol, a universal precursor. Through enzymatic actions at varying steps, cycloartenol branches into the diverse phytosterols identified in this study. Obtusifoliol which is a direct precursor to several downstream phytosterols through the activity of sterol C-14 reductase and sterol demethylases. Campesterol vs β-sitosterol balance shows sterol methyltransferase (SMT2) activity, while β-sitosterol - stigmasterol interconversion is controlled by CYP710A (C-22 desaturase). The geographical or regional differences may signify allelic variation in these key enzymes. Secondly, the sterol findings

give adaptive signals in the sense that their increase or decrease could be linked to drought or disease tolerance (Valitova *et al*., 2024), For instance high concentrations of stigmasterol biosynthesised from β-sitosterol could help maintain the membrane fluidity, stability and drought tolerance. The multi-trait associations of these sterols implies that a shift in one of the sterols for example campesterol will be correlated with others like β-sitosterol, stigmasterol and others.

The differences observed across the regions in the Moringa sterol profiles may result from either pleiotropy, where a single enzyme such as SMT (sterol methyltransferase) or CYP710A (C–22 desaturase) directly controls the levels of multiple sterols or from linkage disequilibrium, where several sterol biosynthetic genes are physically close together in the geme and inherited as a block causing their effects to appear correlated even if each acts on different steps.

FAMEs - C18.1D9 - Oleic acid

The GWAS findings for the FAMEs, include the Morol07g09570, located on chromosome 7, has been linked to oleic acid levels in *Moringa* seed oil. Morol07g09570 appears to play roles in mitochondrial protein transport, embryo development ending in seed dormancy, and response to drought stress. All 48 *Moringa* samples studied showed high oleic acid content, and notably, all originated from tropical regions with long dry seasons, supporting this association.

The Translocase of the Inner Membrane 22 (TIM22) family known for transporting proteins into mitochondria, an essential process for fatty acid metabolism and energy production suggests that genetic variation in this gene is linked to differences in oleic acid content among the *Moringa* accessions analysed. Moreover, because TIM22-4 has also been associated with oleic acid levels in other crops like olives, identifying favorable alleles could help breed oilseed varieties with improved oleic acid content, enhancing both nutritional value and market potential.

C18.2- Linoleic acid

Morol05g04930 linked to chromosome 5 of Linoleic acid is important for metabolic processes, development control, ENT-Kaurene oxidation, heme binding and gibberellin biosynthesis. These plant hormones called gibberellins (GAs) aid in stem elongation, plant height, seed germination, fruit formation and blooming time. Since this gene is involved in gibberellin production, it is possible that plant development and growth are affected by linoleic acid concentrations. The ubiquitination process, which prepares proteins for proteasomal degradation, is linked to MOROL05G04600. Plant metabolism, and fatty acid metabolism in particular, are impacted by the control of protein degradation processes. It is possible that this gene regulates the amounts of linoleic acid and other fatty acids in plants by affecting the stability of proteins involved in lipid production and stress response.

Tocopherols - Beta Tocopherol

Genes associated with β-tocopherol production were identified at key loci on chromosomes 2, 8, 10, and 12, including bHLH63, SPB, GTFIIE, and LCA. These genes appear to coordinate the actions of enzymes involved in tocopherol biosynthesis, and interact with stress-response genes, potentially enhancing the stress resilience of *Moringa* plants. On chromosome 2, several genes were linked to β-tocopherol accumulation and are involved in various biological processes such as metabolic regulation, stress response, gene expression, and cell structure maintenance. These include: Morol02g12450 (GO:0003824), Morol02g13910 (GO:0003723), Morol02g13920 (GO:0031011), Morol02g14460 (GO:0005886), Morol02g15970 (GO:0005634), Morol02g16120 (GO:0003779), Morol02g16130 (GO:0005975). Among them, Morol02g15970 encodes a sterol regulatory element-binding protein that functions as a DNA-binding transcription factor. This gene, part of the bHLH protein family, likely regulates tocopherol biosynthesis genes by acting as a transcriptional activator or repressor, potentially influenced by stress or metabolic signals. On chromosome 8, Morol08g09340 encodes a Squamosa Promoter Binding Protein-Like 7 (SBP), which may regulate gene expression in response to environmental and developmental stimuli. Although its exact role in β-tocopherol production remains unconfirmed, SBPs are known to regulate genes involved in tocopherol biosynthesis (Wang *et al.*, 2018; Zhang *et al.*, 2016). On chromosome 10, Morol10g10410 (GO:0005673) encodes the General Transcription Factor IIE Subunit 2, likely involved in controlling the transcription of genes related to tocopherol synthesis, particularly under changing environmental or developmental conditions. Lastly, Morol12g09630, located on chromosome 12 and associated with GO:0005737, encodes a Lecithin Cholesterol Acyltransferase-Like 4

protein. Its role in lipid metabolism suggests it may influence the availability of lipid precursors required for tocopherol production, thereby impacting overall tocopherol levels in the plant.

These results offer promising avenues for metabolic engineering or marker-assisted selection aimed at improving oil composition, particularly in species where oil quality is a key agronomic trait. Furthermore, the resilience of plants is improved by the coordination of stress-related genes with those involved in tocopherol synthesis.

3.9 Limitations and future work.

This study entailed biochemical profiling on Moringa seed oil, GWAS, bioinformatic analysis, RNA sequence analysis, multiple testing correction and statistical testing of the various findings. Since GWAS cannot identify all genetic determinants of the complex traits identified, there is need for validation of the results, which is yet to be done. The validation of the identified markers by confirming the SNP-trait associations in the populations.

It may be important to conduct additional research that employs exhaustive analytical techniques and larger sample sizes than 48 from more geographical regions to clarify the interactions between environmental factors and seed composition in more Moringa species.

3.10 Conclusion and Recommendations

Conclusively, Moringa oil is a multifunctional "superfood oil" due to its distinctive composition and properties, which are responsible for its diverse characteristics. This study outlines the composition of Moringa oil, emphasizing its value as a result of a high ratio of unsaturated fatty acids (mono- or poly-) to saturated fatty acids, as well as the presence of β-tocopherol and sterols. The GWAS findings for all the oil components especially phytosterols such as Δ5 Avenasterol, Δ7 Stigmastenol, Obtusifoliol; Fatty Acid methyl esters such as C18.1D9 - Oleic acid, C18.2 Linoleic acid; and Tocopherol such as β-Tocopherol highlight the coordinated action of multiple enzyme families and transcription factors which influences the synthesis, modification and degradation of various types of lipid. The genes identified at significant loci of the varying chromosomes highlighted in the phytosterols, fatty acids derivatives, and β-tocopherol showed responses to stress, transport as well as regulation of lipid synthesis and metabolism.

Traits to monitor in future agroforestry or breeding trials

The GWAS study highlights key genetic loci and candidate genes involved in the biosynthesis and regulation of important phytochemicals in *Moringa*, particularly oleic acid, linolenic acid, obtusifoliol, phytosterols such as Δ5 avenasterol and Δ7 stigmastenol, and β-tocopherols. Through GWAS analysis, genes, enzymes and transcription factors such as Morol07g09570 (TIM22-4; protein transmembrane transporter), Morol05g04930 (ENT-Kaurene; heme binding; gibberellin biosynthesis), Morol12g11600 (Cysteine-type Endopeptidase), Morol13g04900_130 (ABC transporter G family member 6-like gene), Morol05g04660 (P450 CYP82D47) and Morol08g09340 (SPB/ Biological regulation), Morol08g01360 (lipid A biosynthetic process, lipidA disaccharide synthase, lipid X metabolic process), Morol08g17400 (triglyceride biosynthesis; diacylglycerol O-acyltransferase activity), Morol08g17050 (MAPK cascade; MAP3Kinase activity) and Morol10g10410_2423 (TFIIE complex) were identified as being closely associated with metabolic pathways that contribute to plant stress tolerance, energy metabolism, and lipid composition.

The association of TIM22-4 with oleic acid levels emphasize the link between mitochondrial protein transport and fatty acid metabolism, particularly in drought-prone tropical environments. Similarly, genes implicated in sterol biosynthesis, including P450 CYP82D47, suggest an intricate regulatory network involving enzymatic modification, lipid transport, and adaptation to abiotic stress.

Notably, genes on chromosomes 2, 8, and 10 related to β-tocopherol biosynthesis such as bHLH63, SPB-like proteins, GTFIIE, and LCAT-like enzymes demonstrate the complex interplay between transcriptional regulation, lipid metabolism, and plant defense mechanisms. The integration of stress-responsive elements with core biosynthetic genes suggests potential targets for genetic improvement of *Moringa* as a nutritionally valuable, climate-resilient crop. Overall, these findings provide a deeper understanding of the genetic basis of key phytochemicals and open avenues for biofortification, stress-adaptive breeding, and functional genomics in *Moringa oleifera* and related species.

Breeding regions with desirable profiles

The following regions exhibit desirable FAMEs profiles. The Kenyan *M.oleifera* high in Vaccenic acid C18.1n11) an isomer of oleic acid, Nigerian *M.oleifera* MO57 high in Oleic acid C18.1n9, Kenya *M.stenopetala* high in Lignoceric acid C24.0. Desirable Phytosterols were found among the Nigerian *M.oleifera* MO26 (D7Stigmastenol, Obtusifoliol), Kenyan *M.oleifera* KMO4 (Ergostadienol, Campesterol, Stigmasterol, β-sitosterol, Δ^5-avenasterol, Stigmastanol), Kenyan *M.stenopetala* KMs10 (Obtusifoliol).

Novelty of the Study

There has been no specific GWAS on *Moringa spp.* in existing genetic studies, such studies only provide a foundation for such studies by identifying trait-specific markers, understanding of genetic variation and potential for breeding programmes in Moringa species. This study provides the first integrative analysis of phytosterol variability in *Moringa* seed oil across diverse geographic accessions, revealing novel patterns of β-sitosterol, Δ^5-avenasterol, and Δ^7-stigmastenol distribution linked to genotype–environment interactions. It is also among the first to report detectable levels of cholesterol and cholestanol in *Moringa oleifera* seed oil, suggesting a previously under-explored role in stress adaptation. The identification of extreme sterol outliers and region-specific biochemical signatures offers new chemotaxonomic markers and opens breeding opportunities for oil stability and nutraceutical enhancement.

Recommendations for Oil improvement programmes

Based on the results of this research on Moringa oils its versatility would be of great importance for dual breeding purpose, hence oil improvement programmes may focus on the following: enhancing the oleic acid content through targeted modification of FAD2 gene whilst reducing polyunsaturated fatty acids (linoleic acid, α-linolenic acid) , improving the oil stability, oxidative resistance and its suitability for edible and cosmetic purposes. In other words FAD2 gene which converts oleic acid to linolenic can be silenced or knocked out, to obtain oil which is richer in monounsaturated oleic acid, with higher resistance to oxidation, higher shelf-life and excellent cooking and frying capability.

One can aim to reduce the very long chain fatty acids (VLCFAs) for edible purposes by selecting against high activity of (beta-ketoacyl-CoAsynthases / Fatty acid elongation complex) KCS/FAE1 elongases which can decrease behenic acid (C22.0) and other similar VLCFAs thereby improving the taste and edibile quality for humans, while retaining its industrial use properties in separate breeding lines.

Also, harnessing the sterol diversity for adaptation and nutrition would assist nutraceutical industries, as B-sitosterol - stigmasterol balance and enrichment of Δ^5-avenasterol were identified as good markers for stress adaptation and nutraceuticals hence they could be exploited for breeding health-promoting oils.

Another way of improving the nutritional value of Moringa oil is to optimise the tocopherol profile by increasing α-Tocopherol levels through varying VTE4 and maintaining the γ- and δ- tocopherols for oxidative stability would be beneficial for industrial use.

Findings from the GWAS could help identify climate - resilient alleles and encourage growing accessions which is suitable and adaptable for specific regions.

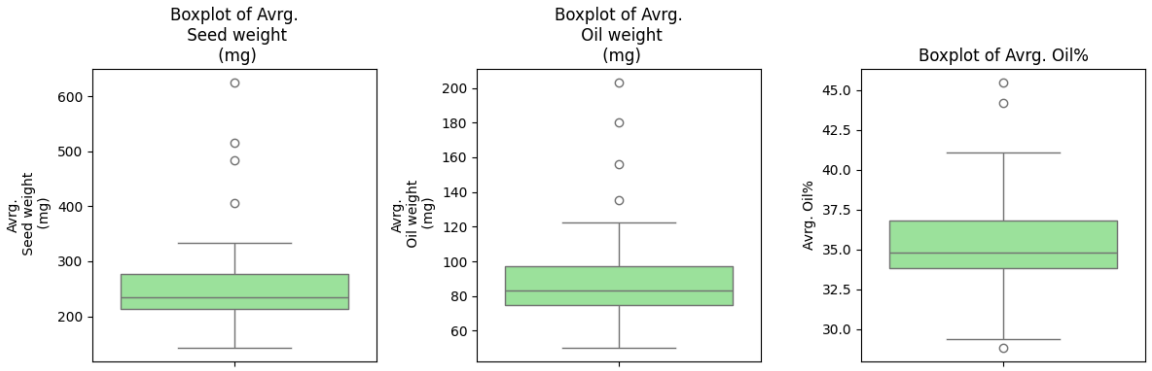
This research can also guide in adopting dual-purpose breeding pipelines, where separate improvement tracks would be pursued. For instance, an edible oil line with high oleic, low behenic, enriched alpha-tocopherol and β-sitosterol may be an option or an industrial oil line with maintained VLCFAs such as behenic acid and balanced tocopherol profiles excellent for use as lubricants, biodiesel and bio-based polymers.

Recommendations for Nutritional or industrial cultivar development

The diversity of the nutritional profile of Moringa oil indicates both nutritional and industrial relevance which could be harnessed for potential applications in the food, medicinal, and nutraceutical industries. The presence of nutritionally valuable FAMEs including margaric, vaccenic and α-linolenic acid C18.3n3 show strong nutritional selling point as food oil, while behenic acid, erucic acid C22.1n9, and lignoceric acid C24.0 makes Moringa oil attractive for cosmetic, lubricants and biodiesel industries. The statistically significant variation across the accessions point out that genotype and environment interactions influence oil quality which may guide breeding and selection either for edible or industrial purposes.

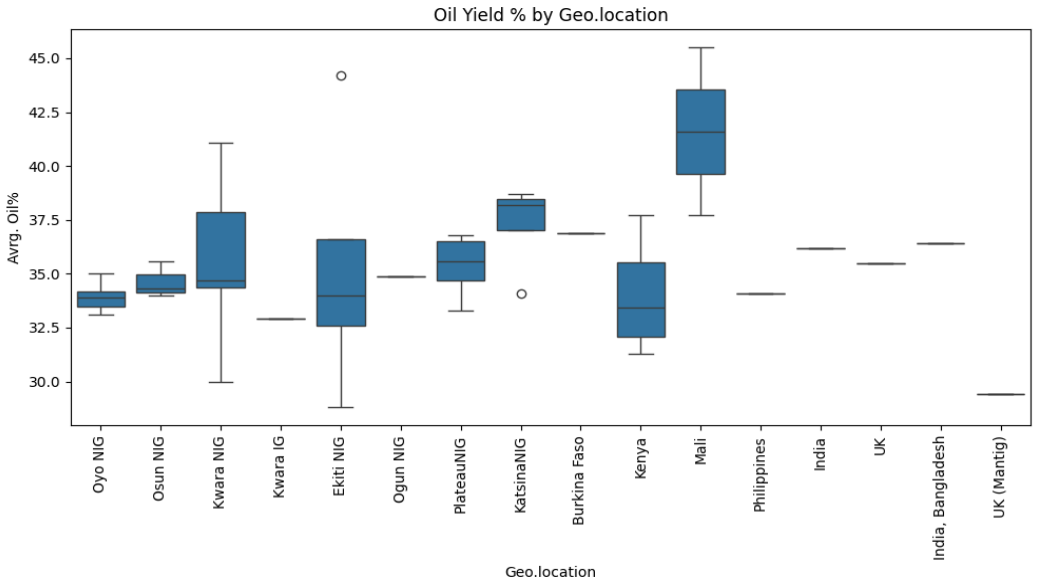
Appendix

Supplementary Figures

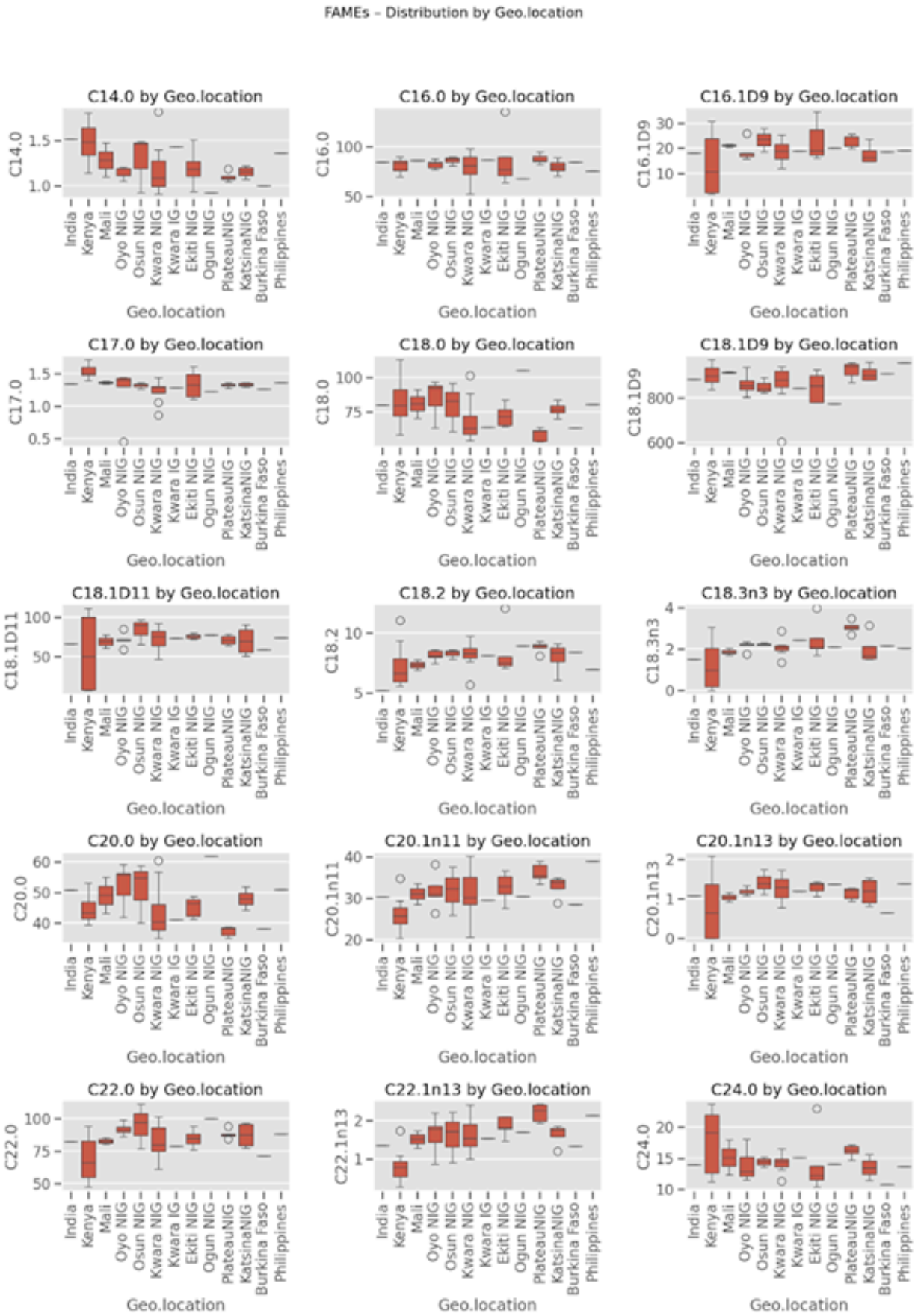


Supp.Figure 3.1: Boxplots of Seed and Oil Attributes

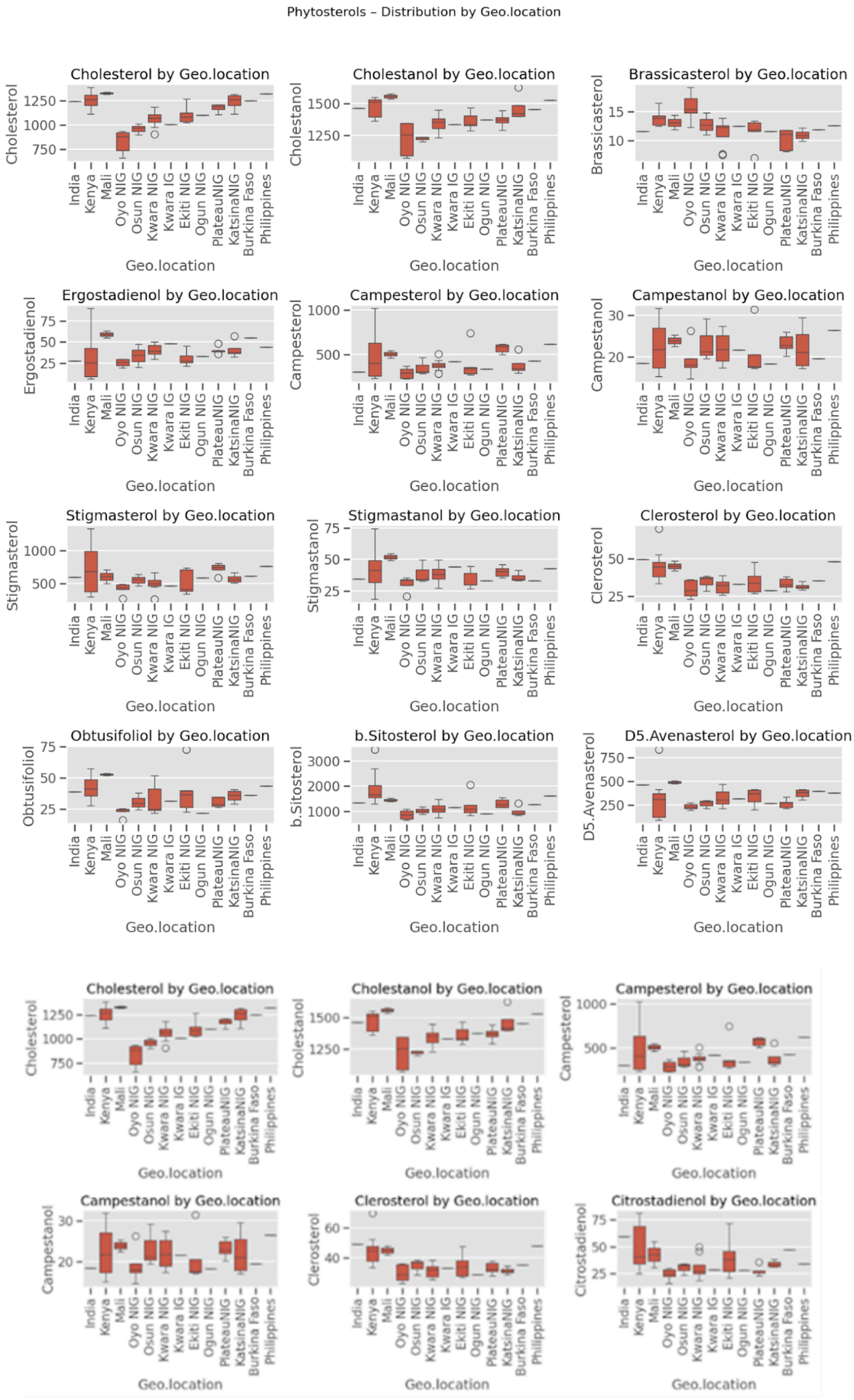
The boxplots for 'Avrg. Seed weight (mg)', 'Avrg. Oil weight (mg)', and 'Avrg. Oil%' are the means of three replicates of each sample. Plots show the distribution of data and identifying potential outliers, which are shown as individual points that fall significantly outside the whiskers/ the interquartile range.



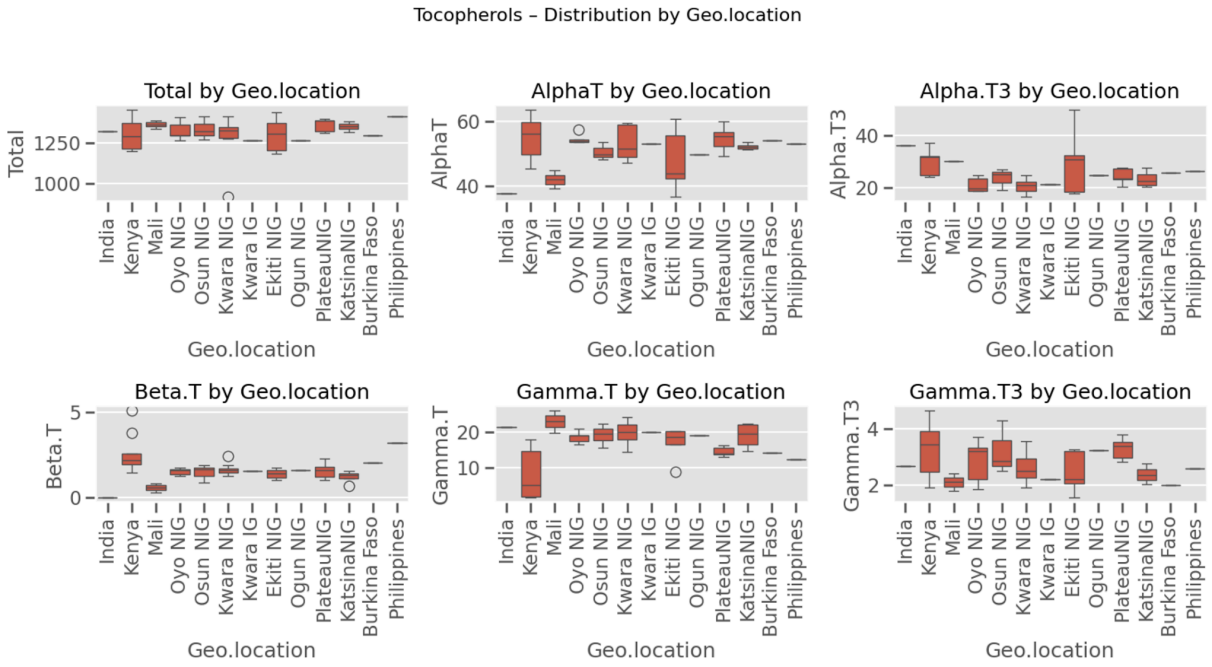
Supp.Figure 3.2 Histogram of Oil yield (%) by geographical locations.
The mean oil yield (%) is being compared across locations.



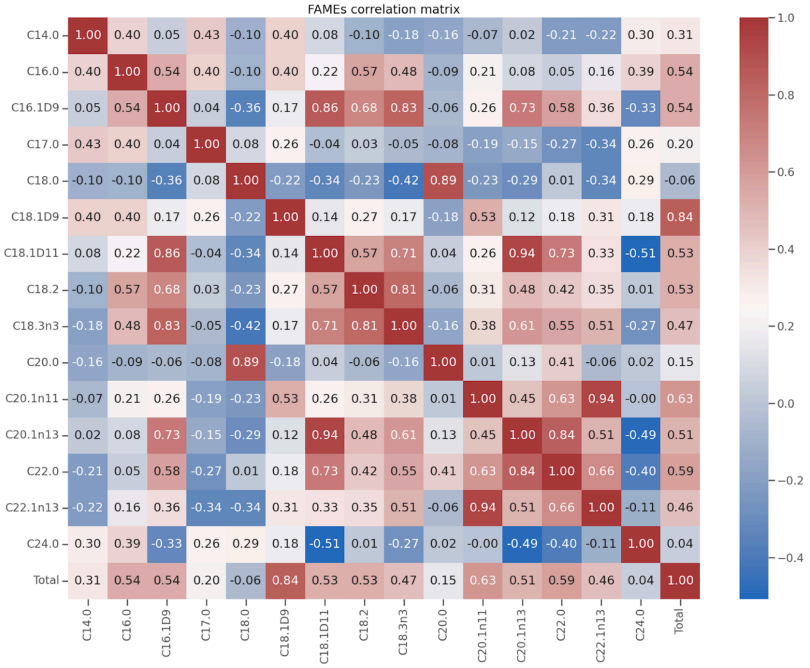
Supp.Figure 3.3 Box plots show 15 Fatty Acid Methyl Esters (FAMEs) derivatives by geographical location across the 48 samples.



SuppFigure 3.4 Box plots show 18 phytosterols by geographical location across the 48 samples.

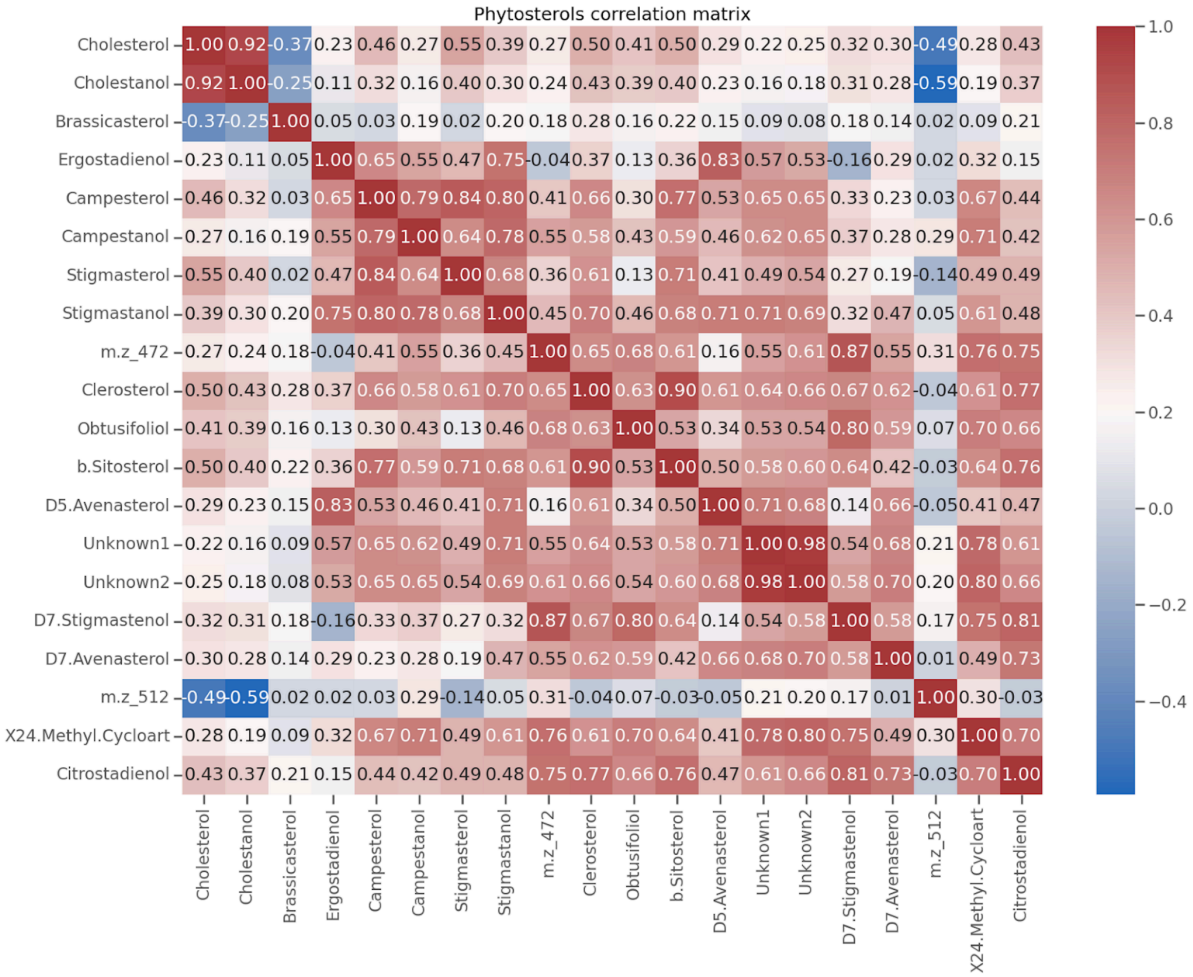


Supp. Figure 3.5 Box plots show 4 tocopherol isomers by geographical location across the 48 samples.



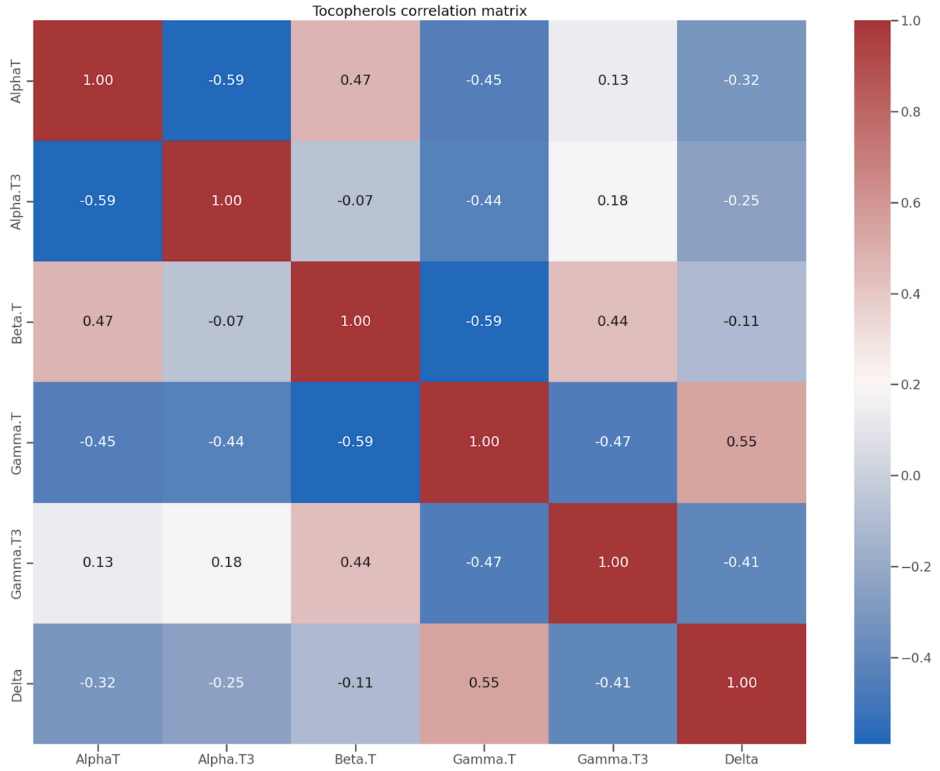
SuppFigure 3.6 Correlation matrix of fatty acid methyl esters visualized using a heatmap.

The heatmap infers inverse relationships between FAMEs such as seen when C18.1D11, C18.2 and C18.3n3 increases, C17.0 and C18.0 decreases and so on. Whereas a number of FAMEs compounds show similar colour intensities depicting strong positive or negative correlation inferring co-occurrence patterns, that is these might be functionally related. There are a few compounds with unique signatures C18.0, which may act as biomarkers for environmental conditions as they are seen to be high only in specific locations. These observed trends may give hints of shared biosynthetic pathways.

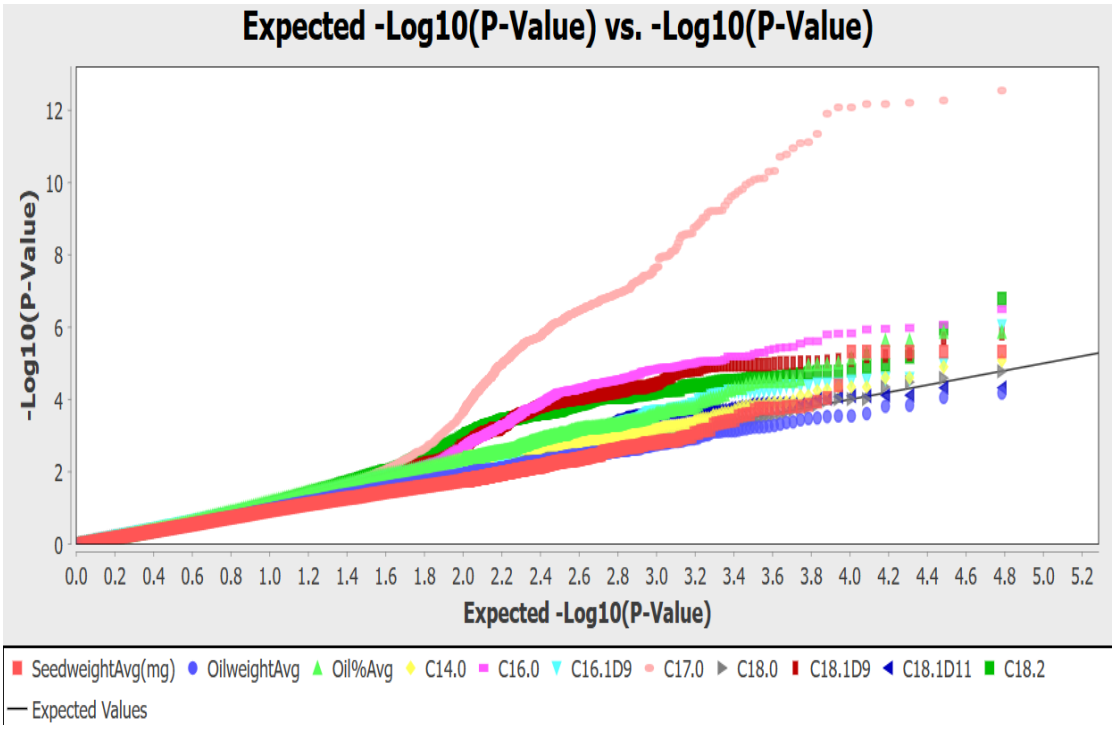


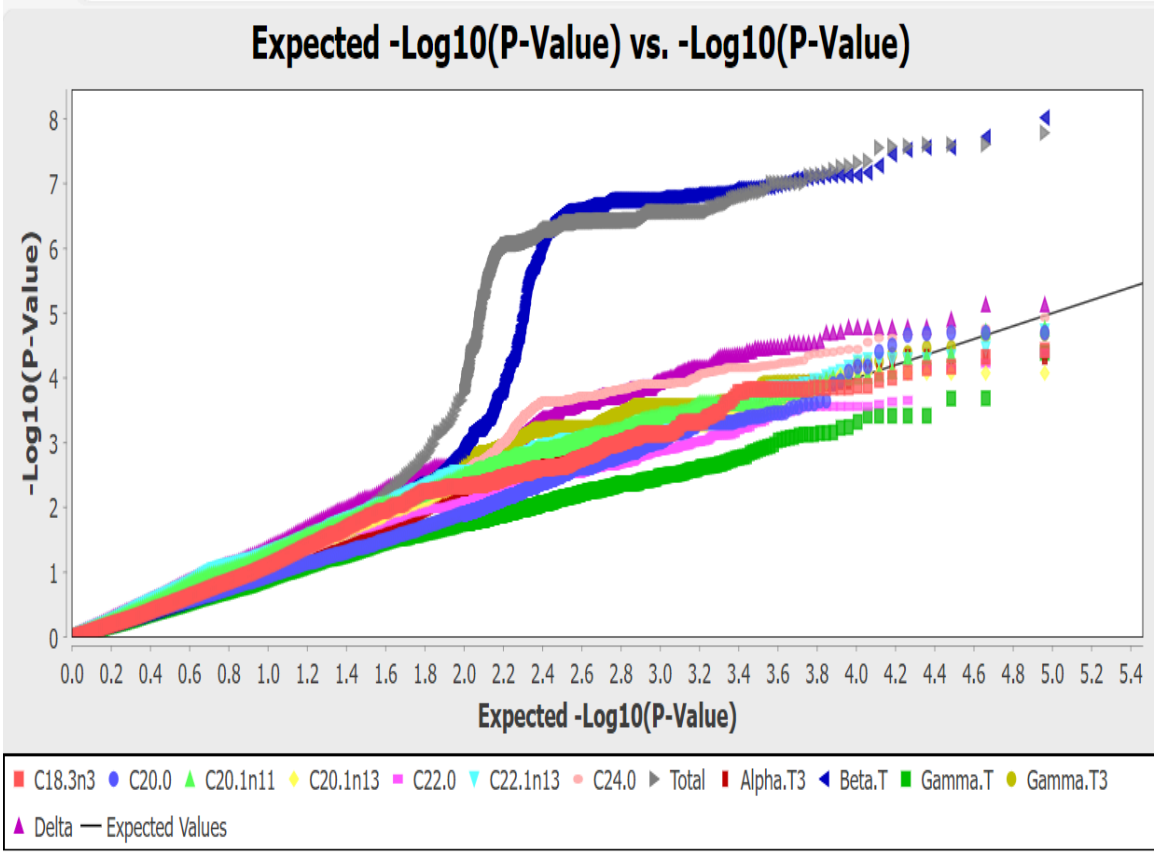
Supp. Figure 3.7 Correlation matrix of phytosterols visualized using a heatmap

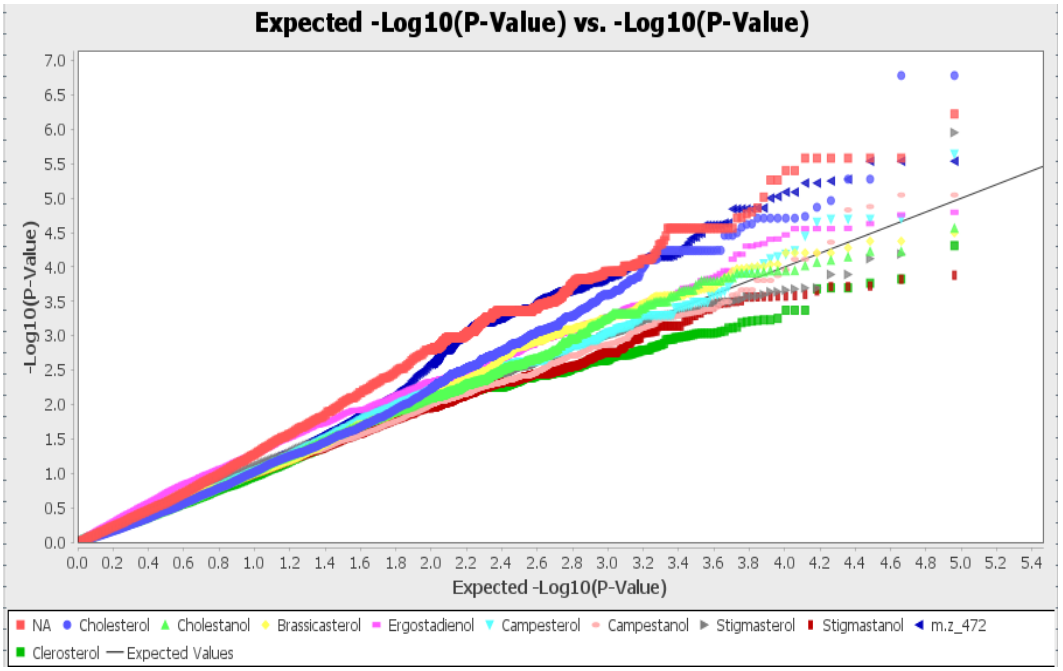
The heatmap infers inverse relationships between sterols such as seen when Cholesterol and Cholestanol increases, Brassicasterol decreases; likewise as D7Stigmastenol decreases, Citrostadienol and Obtusifoliol increases and so on. Whereas a number of sterol compounds show similar colour intensities depicting strong positive or negative correlation inferring co-occurence patterns, that is these might be functionally related. There are a few compounds with unique signatures (Ergostadienol), which may act as biomarkers for environmental conditions as they are seen to be high only in specific locations. These observed trends may give hints of shared biosynthetic pathways.

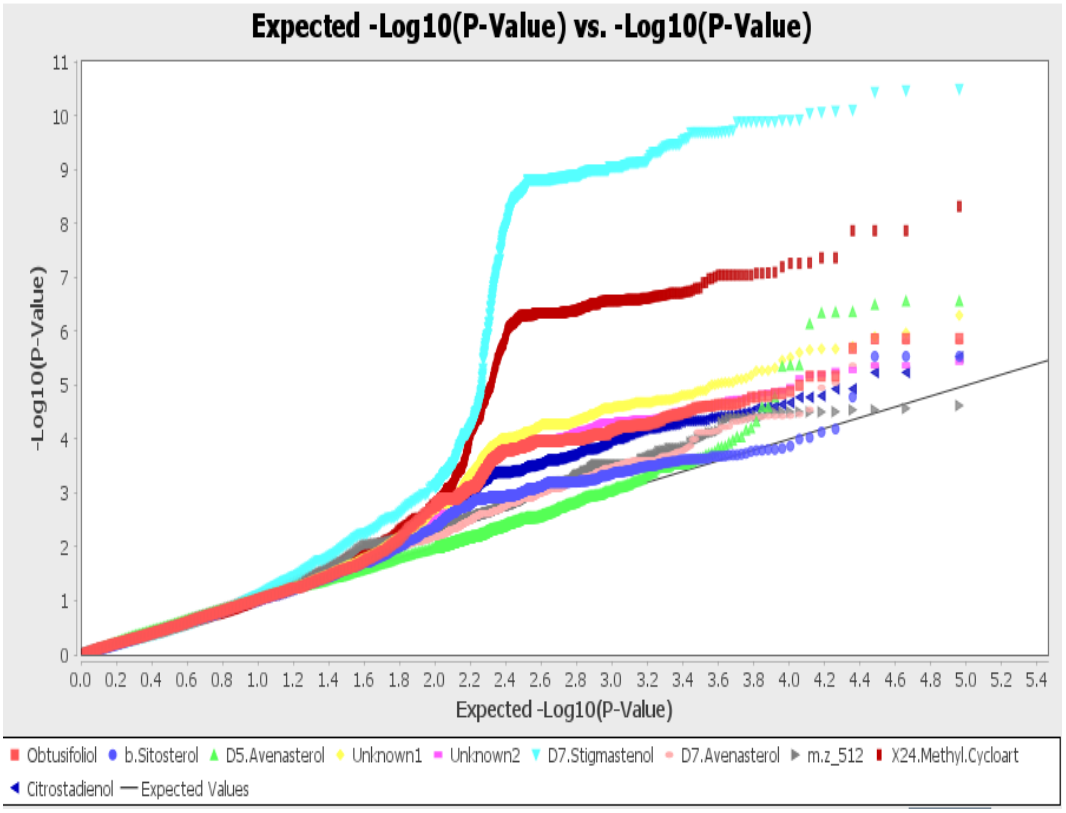


Supp. Figure 3.8 Correlation matrix of tocopherols visualized using a heatmap indicating the strength and direction of relationships. A gradient from blue (negative correlation) to red (positive correlation) signifies which tocopherols are positively or negatively correlated with each other. As the level of Alpha.T increases, Gamma - T3 also increases.


Expected -Log10(P-Value) vs. -Log10(P-Value)
-Log10(P-Value)
Expected -Log10(P-Value)
SeedweightAvg(mg)
OilweightAvg
Oil%Avg
C14.0
C16.0
C16.1D9
C17.0
C18.0
C18.1D9
C18.1D11
C18.2
Expected Values

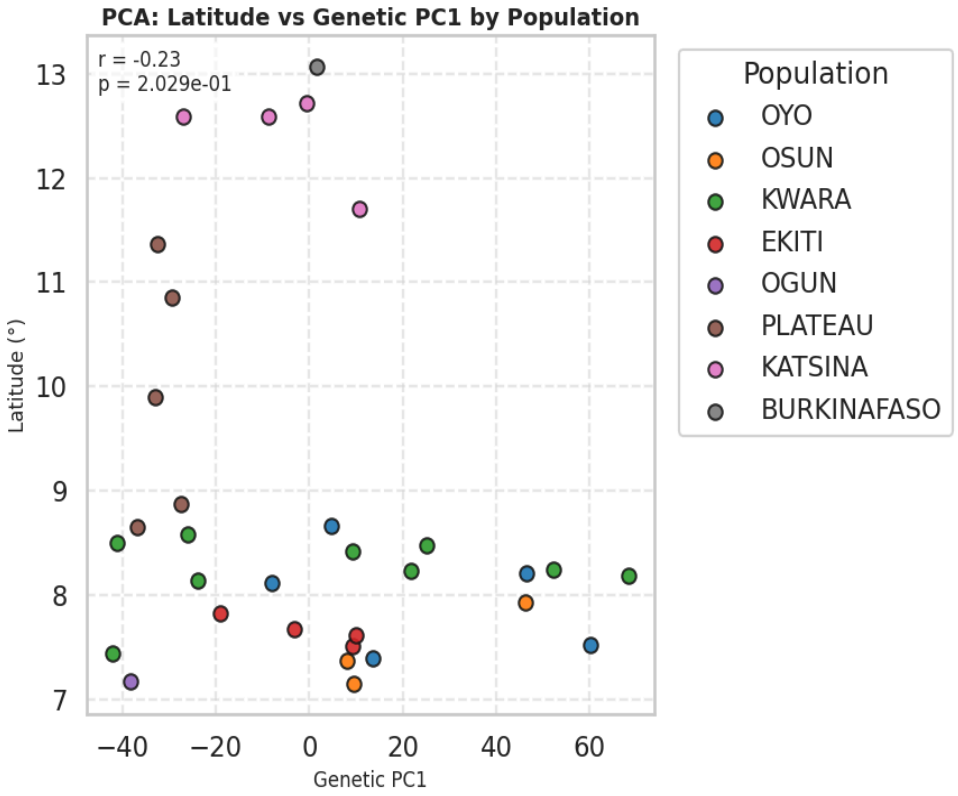

Expected -Log10(P-Value) vs. -Log10(P-Value)
-Log10(P-Value)
Expected -Log10(P-Value)
C18.3n3
C20.0
C20.1n11
C20.1n13
C22.0
C22.1n13
C24.0
Total
Alpha.T3
Beta.T
Gamma.T
Gamma.T3
Delta
Expected Values



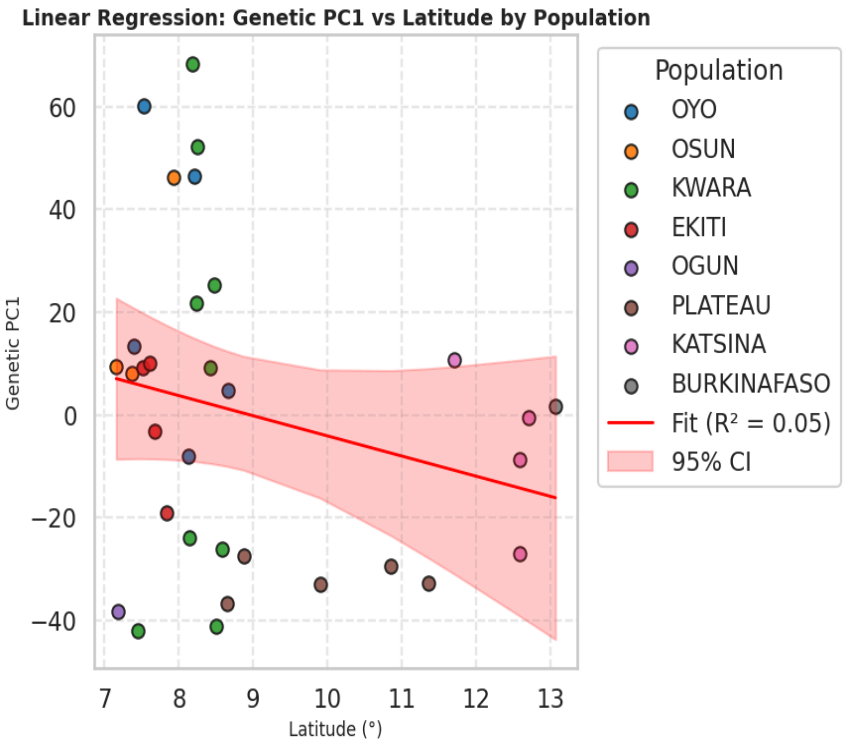


Supp. Figure 3.9 QQ-plots for all FAMEs, Tocopherols and Phytosterol

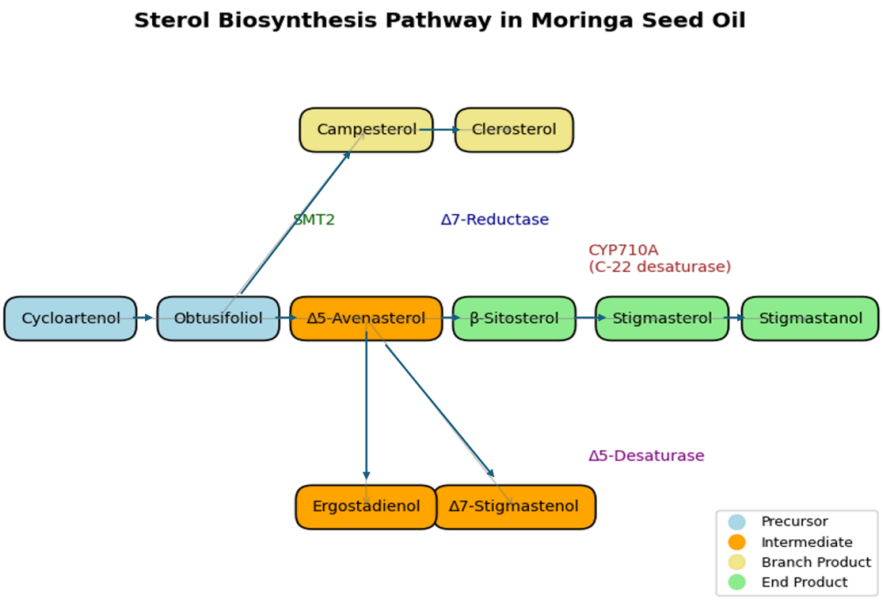
Supp. Figure 3.10 Manhanttan plots



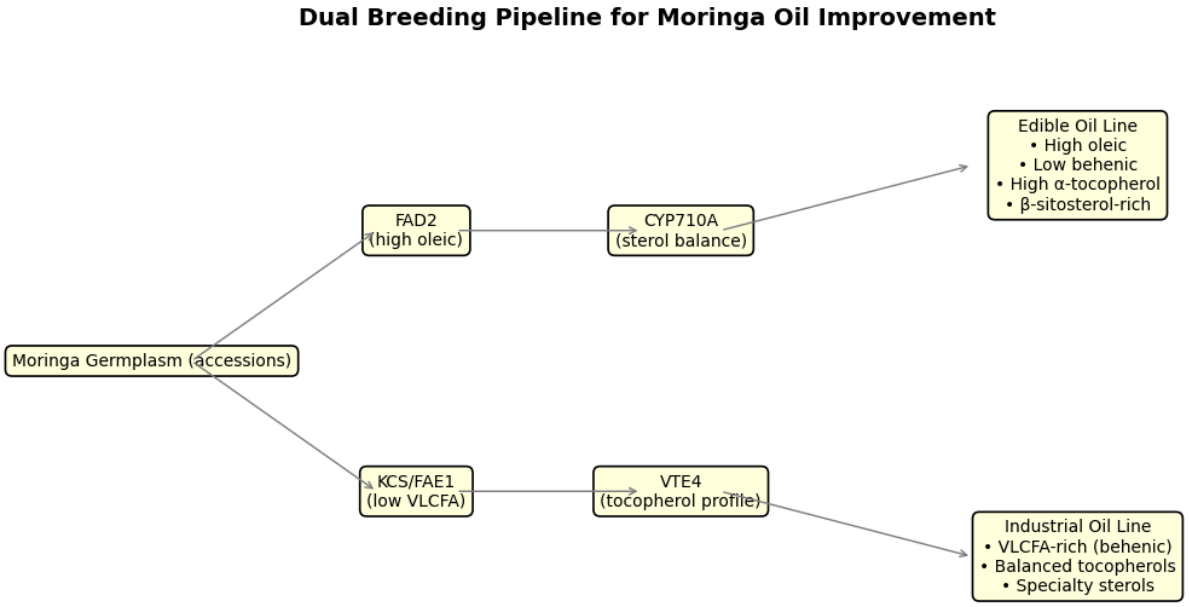
Supp. Figure 3.11A Principal Component Analysis (PCA) showing PC1from SNP/marker dataset plotted against Latitude of the location of seeds collection. The downward slope which is a negative correlation indicates Latitude increases (North-southward) as PC1 scores gradually increases. Regression line showing genetic-geographic correlation.



Supp. Figure 3.11B Principal Component Analysis (PCA) revealed a strong negative association between PC1 and latitude (Supp.Fig.3.11). Seed accessions from higher latitudes clustered at lower PC1 values, whereas those from lower latitudes clustered at higher PC1 values, indicating a latitudinal gradient in genetic variation across moringa populations.



Supp. Figure 3.12 Significant regional differences observed for Moringa sterols spanning precursors (obtusifoliol), intermediates (Δ^5-avenasterol, and Δ^7- stigmastenol), and end products (β-sitosterol, stigmasterol). These observed patterns imply an interconnectedness of genetic regulation (SMT, CYP710A, sterol desaturases) and environmental influences which has direct implications for the stability of oil, resilience to stress and overall breeding potential.



Supp.Figure.3.13 Dual breeding pipeline for Moringa Oil Improvement. Aligning oil quality traits with both nutritional and industrial priorities positions Moringa as a versatile multipurpose crop resilient to climate variability.

Supplementary Tables

Supp.Table 3.1fames.xlsx

Supp.Table 3.2TOCOPHEROLs_means_table .xlsx

Supp.Table 3.3PHYTOSTEROL_means_table.xlsx

Supp.Table 3.4 SeedData_means_table.xlsx

Supp.Table 3.5 t-test and F statistics of the comparisons of the FAMEs means between the *Moringa oleifera* and *Moringa stenopetala* groups.

	FAMEs	**t-statistics**	**p-value**	**F-statistics**	**p-value**
C14.0	Myristic acid	0.30	0.78	9E-02	**8E-01**
C16.0	Palmitic acid	-0.49	0.65	2E-01	6E-01
C16.1D9	Palmitoleic acid	-18.87	0.00	**4E+02**	**0E+00**
C17.0	Margaric acid	3.82	0.02	1E+01	1E-02
C18.0	Stearic acid	1.02	0.35	1E+00	4E-01
C18.1D9	Oleic acid	-1.60	0.16	3E+00	2E-01
C18.1D11	Vaccenic acid	-26.04	0.00	7E+02	0E+00
C18.2	Linoleic acid	-1.73	0.17	3E+00	1.E-01
C18.3n3	α-Linolenic Acid	-10.90	0.00	1E+02	0E+00
C20.0	Arachidic Acid	-2.46	0.05	6E+00	5E-02
C20.1n11	Gadoleic acid	-2.42	0.06	6E+00	5E-02
C20.1n13	Erucic Acid	-18.20	0.00	3E+02	0E+00
C22.0	Behenic Acid	-11.40	0.00	1E+02	0E+00
C22.1n13	Erucic Acid	-3.44	0.01	1E+01	1E-02
C24.0	Lignoceric Acid	6.00	0.00	4E+01	0E+00
Total		-7.61	0.00	6E+01	0E+00

Supp.Table 3.6 Comparing means between the KMs group and the *Moringa oleifera* group with t test.

Tocopherols	**t-stat**	**p-value**
Alpha T3	5.91	0.00
BetaT	2.34	0.07
GammaT	1.86	0.16
GammaT3	-26.82	3.71
Delta	5.33	0.01

Chapter 4

Integrative Transcriptomic Approach for *Moringa oleifera*, *Moringa stenopetala* and *Moringa peregrinna*.

Abstract

This study sought to utilize an efficient transcriptome assembly pipeline that would enable the transcriptomic study of three Moringa species: *M. oleifera*, *M. stenopetala*, and *M. peregrina* collected from diverse geographical regions in order to provide a resource dataset for Moringa research. A total of 49 cDNA libraries were constructed from a pool of three different tissues (leaf, root, and stem) of *M. oleifera*, *M. stenopetala* and *M. peregrinna* species. These libraries were sequenced in duplicates using the Illumina HiSeq 2500 platform and generated around 1635367412 million reads with an average length of 150bp. High-quality clean reads were chosen for the assembly after removing ambiguous reads identified by FASTQC. Trinity was used to perform a *de novo* assembly for the combined tissues of the 49 Moringa samples. Furthermore, a reference - guided *de novo* approach whereby the *M. oleifera* AOCCv2 genome was used as a reference for all the assembled transcriptomes. The reads from *M. stenopetala* and *M. peregrinna* species yielded an alignment rate of 91%. The assemblies were tested for completeness using BUSCO against Eudicots database. The Kenya *M. oleifera* (KMo6) had the least completeness (86.5%) against the Eudicot database. All others regardless of their species had BUSCO 97.6% completeness. The 49 novel transcriptome assemblies would serve as resources for species comparison and molecular marker development for the selection of traits such as drought stress and oil composition.

4.1 Introduction

The genetic characteristics of non-model organisms, such as the under-researched Moringa species, could be understood through transcriptome assembly. The *de novo* assembly is an important bioinformatics process that enables the re-construction of novel RNA sequences (transcriptome) from short reads, especially for orphan crops which have no existing reference. Each *de novo* assembler possesses distinctive attributes that make it particularly well-suited for specific categories of data. Using a *de novo* assembly tool like TRINITY (Grabherr *et al*., 2011) for the reconstruction of transcriptiomes from RNA-seq data have been widely used but the process is challenging due to their short length.

Due to sparse transcriptome studies and genetic data resources on a number of Moringa species, employing an integrative analysis that integrates *de novo* and reference-guided *de novo* methods could make comprehensive genetic databases of Moringa species (Moreton *et al*., 2015). In order to overcome this, an integrative transcriptomic approach of Reference-guided *de novo* Assembly, which suits this diploid plant, Moringa (*M. oleifera, M. stenopetala,* and *M. peregrinna*) was applied. By utilizing these advanced bioinformatics approaches high-quality transcriptomes that account for the unique challenges posed by non-model organisms like Moringa have been developed. This foundational database of Moringa transcriptomes will offer a comprehensive understanding of the diverse applications of genetic and genomic breeding enhancements necessary for Moringa and similar plants. It will help in the identification of traits and associated genes that have not yet been recognised in related, well-studied organisms and are not detectable in genomic analyses of model organisms. These findings will support practical applications in evolutionary studies, regulatory gene networks (Cassan *et al.,* 2021), conservation of genetic diversity, and tracing orthologous gene families among species within the genus (Wen *et al.,* n.d). This database composed of comprehensive 49 Moringa transcriptomes across three species *M. oleifera*, *M. stenopetala*, *and M. peregrina w*ill be a very useful document for researches in drought-tolerant plants.

Genome-guided *de novo* approaches exhibit distinct advantages, thereby establishing them as a viable alternative to the conventional pure *de novo* assembly method. It is worthy of note that genome-guided denovo assembly has its limitations, particularly because the most recent nearly-complete AOCCv2 Moringa reference genome is

specific to the *M. oleifera* species as this may introduce bias against *M. stenopetala* and *M. peregrinna* which may lead to an under-representation of transcripts from genes that are less conserved, as they are not adequately represented in the reference (Jackson et al., 2024). Also, the estimation of transcripts or gene expression may be biased due to the use of a congeneric reference genome, improvements in the quality of the reference may progressively mitigate the various limitations and drawbacks associated with alignment with a reference genome. The accuracy and success of assemblies are significantly influenced by the quality of sequencing data (Miller *et al.,* 2010), as well as the computational tools, methodologies employed, affordability and accessibility of various sequencing technologies (Salzberg *et al.,* 2012) and the choice of sequencing technology (Davis-Turak *et al.,* 2017).

Despite a number of molecular studies on *Moringa oleifera* (Panes and Discar, 2021; Yang *et al*., 2021; Shafi *et al*., 2020; Pasha *et al*., 2020; Ojeda-López *et al*., 2020; Panes and Baoas, 2018; Deng *et al*., 2016), *Moringa concanensis* (Shafi *et al*., 2022;), *Moringa peregrinna* (AbdAlla *et al*.,2023) there has been less attention on other members of the genus. Other species have remained uncharacterized at the transcriptomic level therefore, the aim of this study is to provide transcriptome databases for understudied Moringa species including *Moringa stenopetala* and *Moringa peregrinna* as well individual accessions of *Moringa oleifera* from diverse geographical locations which would be an excellent resource for upstream analyses such as differential gene expression, comparative transcriptomics, cross-species comparison of gene families, lineage-specific adaptations and pathway conservation. It also aims to provide a searchable resource for identifying genes linked to oil synthesis, stress responses, nutrition, and many others.

4.2 Materials and Methods

Initially, the effectiveness of several assembly tools to meet these questions were evaluated and assessed to identify the most efficient tools. Subsequently, the disparities in transcriptome assemblies among the three Moringa species was examined, evaluating the quality, prospective uses, and biological insights obtained from the assembled transcriptomes. The prospective applications of these tools in future research and the efficacy of various assembly methods was evaluated. Likewise, the computational demands of each tool, and the reproducibility of results across several datasets was assessed.

4.2.1 Sample collection

The seed samples of 49 Moringa accessions were collected from Sub-Saharan Africa (Nigeria, Mali, Burkina Faso), East Africa (Kenya), Southeast Asia (Philippines), South Asia (India), Southwest Asia (Oman), and a commercial source from eBay (UK) (Fig.3.1 and Table 3.1).

4.2.2 Tissue extraction and RNA sequencing protocols

RNA sequencing (RNA-Seq) analysis followed established protocols for RNA extraction, library preparation, and sequencing, as detailed in the relevant literature. RNA was extracted from leaf, stem, and root tissues using the Omega EZNA Plant RNA Kit Protocol, following the manufacturer's instructions (Fang *et al.,* 2016).

Leaf, stem, and root tissues were collected from 6-week old Moringa plants. To avoid RNA degradation by RNases, I made sure the integrity of my RNA samples were preserved prior to sequencing by promptly collecting tissues while wearing protective gloves, utilizing sterilized equipment such as scissors thoroughly cleaned with RNAseZAP, rapidly freeze-drying the tissues in liquid nitrogen shortly after cutting, and conducting the extraction process as soon as possible post-collection using a reliable RNA extraction kit and protocol that eliminates buffer residues after washing (Omega EZNA RNA kit). The workbench was subjected to comprehensive decontamination using RNAseZAP at the beginning of each extraction session and was consistently treated thereafter to maintain a sterile environment.

RNA samples were promptly stored in a -70℃ freezer following the extraction of 5µL aliquots to minimize freeze-thaw cycles. The aliquot samples underwent 1% agarose gel electrophoresis for visualization of rRNA bands and were subsequently evaluated for quality using the Nanodrop RNA 6000 Nano assay. Additionally, RNA concentration and RIN quality measurements were conducted through capillary electrophoresis on the Agilent Bioanalyser 2100 v1.3. The RNA quality and concentration must exceed a RIN threshold of 7 (Schroeder *et al.,* 2006). RNA samples from Moringa tissues in this study having RNA Integrity Number (RIN) greater than 8.5 were sequenced using an Illumina short-read library, following the manufacturer's protocols, resulting in 251.7G of sequence data from complementary DNA (cDNA). The A260/A280 ratio which is the ratio of ultraviolet (UV) absorbance at 260nm and 280nm was also considered to identify the level of protein contamination in the RNA sample. The Moringa samples generally yielded a ratio of 1.8 - 2.0 which is generally accepted as pure for RNA. While the rRNA 25S :18S ratio presented as 2.4 averagely in the Moringa samples, which shows the RNA being intact and not degraded.

RNA samples from the Oman tissues yielded low quality RNA but was repeated using a more extensive protocol (CTAB) *see Appendix.*

Library preparation and sequencing were conducted in accordance with the RNA Sequencing protocol established by Novogene. This process involves segmenting the sequences into 200-250 base pair lengths and subsequently attaching specific adapter sequences to the ends of the chosen fragments. Messenger RNA was isolated from total RNA utilizing poly-T oligo-attached magnetic beads. The library was constructed via end repair, A-tailing, adaptor ligation, size selection, amplification, and purification. The library was assessed via Qubit, quantified using real-time PCR, and its size distribution analysed with a bioanalyser. The quantified libraries were pooled and sequenced on the Illumina paired-end Novaseq 6000 platform, according to effective library concentration and data volume (Novogene, 2022).

4.2.3 Read pre-processing

A quality assessment of the sequenced reads was conducted using the high-performance computing system Viking, employing FastQC version 0.11.9 (Andrews, 2019; Cock *et al.,* 2010). The raw RNA sequencing reads were treated to exclude low-quality reads and adaptor sequences. The paired-end reads were trimmed with the Illumina adaptor utilizing TrimGalore, version 0.6.4 (Martin, 2011). The adapter sequence employed was 'AGATCGGAAGAGC' (Illumina TruSeq, Sanger iPCR), and trimming was performed using Cutadapt version 2.3 and Python version 3.7.0. A Phred quality score cutoff of 28 was established, and both Read1 and Read2 sequences were subjected to trimming.The output file was compressed in the .gzip format.

The transcriptome assembly process can be summarised in Fig.4.1 below. The sequenced reads passed through both *de novo* assembly and Reference - Guided *de novo* assembly using TRINITY aligner. This was done because the Moringa species : *M. stenopetala* and *M. peregrinna* used in this research which are not of the same species with the standard Moringa genome reference AOCCv2 used in this study. Utilizing a *de novo* approach for reconstructing the transcripts based on actual read sequences will help to capture sequence variations in *M. stenopetala* and *M. peregrinna* and resolve the paralogous transcripts whilst the Reference-guided *De novo* Assembly approach help assess the assemblies by comparing the outputs for quality and completeness.

To compile transcript sequences from Illumina RNA-Seq data composed of short reads of 150bp paired-end reads, the *de novo* approach using TRINITY assembly was applied through three processes: Inchworm, Chrysalis, and Butterfly, resulting in the production of full-length transcripts (Grabher *et al.,* 2011, Miller *et al.,* 2010).

4.2.4 *De novo* transcriptome assembly

TRINITY v2.11.0 (Grabherr *et al.,* 2011) was utilized for the *de novo* transcriptome assembly of this non-model plant (Moringa) due to its robustness and ease of usage. I chose the TRINITY assembler to clearly sort out where the reads came from (origin) and to rebuild the original sequences (Raghavan *et al*., 2022). It is highly effective for reconstructing transcripts and alternatively spliced isoforms (Haas *et al*., 2013). Also, TRINITY is the most prominent De Bruijn graph-based assembler which has been cited several times.

The TRINITY software used key parameter settings which involve read quality filtering, read normalization, selection of k-mer size and stranded library, yielding the identification of about 100,000 "unitigs," representing the assembled elements of the transcriptome. I then ran Trinity as using the code below only after filtering, trimming, FastQC, Kmer selection and other default parameters were set.

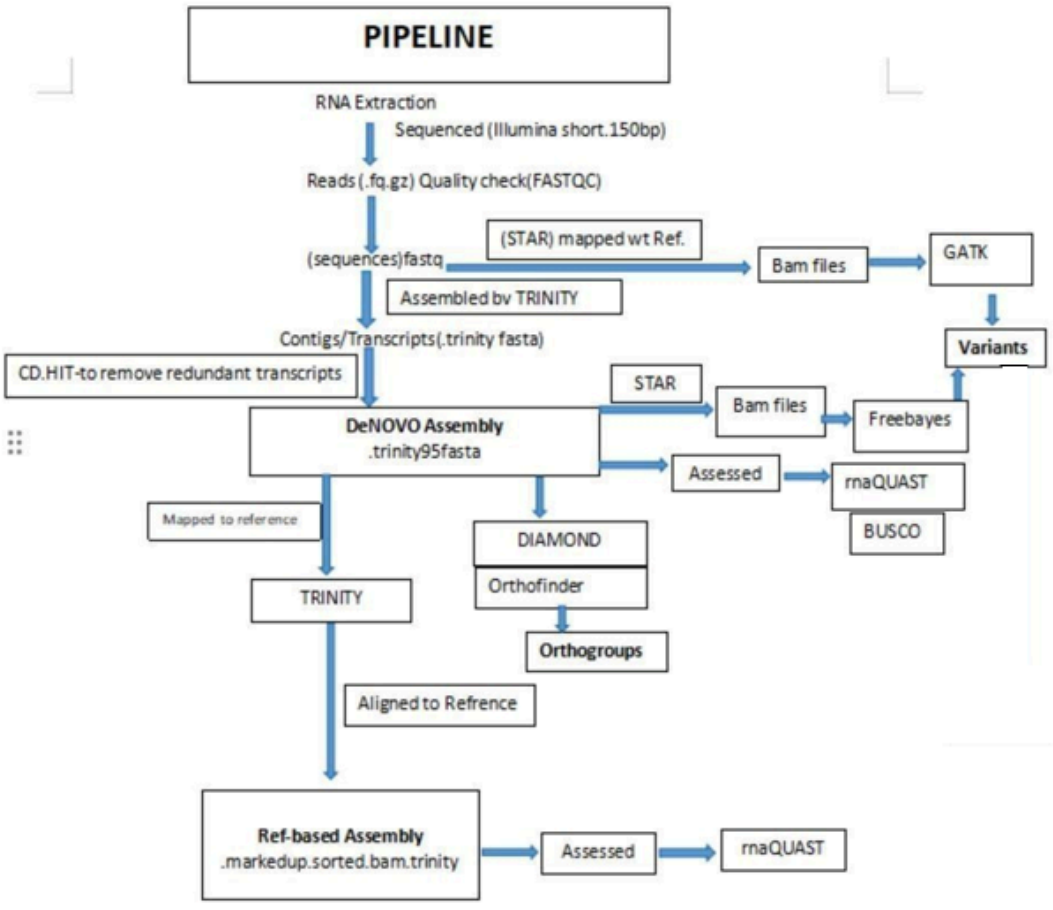


Figure. 4.1 Pipeline of the Reference-guided *De novo* Assembly

Read quality filtering

For trimming my paired-end reads, I used TrimGalore v0.6, and run parameters are as follows: Cutadapt version: 2.3; Python version: 3.7.0; Number of cores used for trimming: 4; Quality Phred score cutoff: 28; Quality encoding type selected: ASCII+33.

Adapter sequence is 'AGATCGGAAGAGC' (Illumina TruSeq, Sanger iPCR; auto-detected). Maximum trimming error rate: 0.1 (default); Minimum required adapter overlap (stringency): 1bp. Minimum required sequence length for both reads before a sequence pair gets removed: 20 bp. All Read 1 sequences was trimmed by 9 bp from their 5' end to avoid poor qualities or biases and all Read 2 sequences was trimmed by 9 bp from their 5' end to avoid poor qualities or biases (e.g. M-bias for BS-Seq applications).

FastQC was run on the data once trimming was completed and output file(s) were GZIP compressed. Cutadapt was up-to-date (version 2.3). Setting -j 4. Final adapter and quality trimmed output was written to MO1_1_trimmed.fq.gz and subsequently for all other 47 samples. Then performed quality (cut off '-q 28') and adapter trimming in a single pass for the adapter sequence: 'AGATCGGAAGAGC' from file MO1/MO1_1.fq.gz with 20000000 sequences processed using cutadapt 2.3 with Python 3.7.0. Command line parameters: -j 4 -e 0.1 -q 28 -O 1 -a AGATCGGAAGAGC MO1/MO1_1.fq.gz

26668260 sequences were processed in total. The length threshold of paired-end sequences gets evaluated later on (in the validation step). The same process ran for the second read (Reverse read) and the total number of sequences analysed being 26668260 for sample MO1.

72819 (0.27%) sequence pairs were removed because at least one read was shorter than the length cutoff (20 bp). FastQC run on the validated data MO1_1_val_1.fq.gz and MO1_2_val_2.fq.gz

Read normalization

Trinity phase 1 which is clustering of RNA-Seq reads involved In silico Read Normalization where excess reads beyond 200 coverage were removed.

Clustering

Multiple contigs which could have been generated from the TRINITY assembly representing different portions of the same transcript were clustered. This was done applying CD-HIT v4.8.1 (Ono *et al.,* 2015; Li & Godzik, 2006) to minimize redundancy in the assembled contigs through the clustering of nucleotide sequences from separate contigs that represent various regions of the same transcript resulting in a unitig

sequence count ranging from 85,169 to 120,422 transcripts. This mechanism enables the incorporation of alternative splicing variants or isoforms from a single gene into a cohesive transcript.

4.2.5 Reference-guided *De novo* Assembly

The RNASeq NF-core v3.15.0 pipeline was used to analyse 49 samples' raw sequence data (fastq.gz files). This pipeline included quality control (QC), trimming (Trim Galore), alignment of the denovo output (.trinity.fasta files) using STAR, partitioning of the aligned reads based on their genomic loci, and then *de novo* assembly of each partitioned read cluster using TRINITY.

Coordinate-sorted.bam.files were mapped to a reference genome MoringaV2.genome (AOCCv2) (Chang *et al.,* 2022). The Moringa reference genomes AOCCv2 which have been extensively studied and documented in previous studies, was obtained from GenBank. Based on the nf-core set of processes (Ewels *et al.,* 2020), the reads were processed using nf-core/rnaseq v3.14.0. Bioconda and Biocontainers provided repeatable software environments (da Veiga Leprevost *et al.,* 2017). Nextflow v24.04.2 (Di Tommaso *et al.,* 2017) was used to execute the pipeline, with the optimization stages and parameters detailed in the Appendix.

Here, the AOCCv2 *M. oleifera* genome, which is a chromosome-scale version was chosen as reference for this research. This updated version has 14 pseudo-chromosomes, which are n=14 haploid chromosomes, and reflects a better genome assembly that was produced from long-read sequencing. The alignment of the paired short reads (150 bp) acquired from Illumina sequencing of Moringa to the reference genome was performed using the splice-aware STAR Tool (Chang *et al.,* 2022; Dobin, et al., 2012; Dobin *et al.,* 2015; 2016). According to Musich *et al*. (2021), this tool demonstrated a higher level of tolerance for soft-clipped sequences and mismatches than HISAT2. The Samtools were then used to sort the .bam files. The assemblies were further mapped back to the transcriptomes using TRINITY, these are the reference-guided *de novo* assemblies. These were assessed using rnaQUASTv.2.3.0 (Bushmanova *et al.,* 2016). TRINITY groups transcripts into clusters based on the shared sequence content. The transcript cluster is referred to as a gene. The TRINITY output encodes the TRINITY “gene” and “isoform” information.

4.2.5.1 Mapping rates

The mapping rate represents the proportion of uniquely mapped reads compared to the total number of input reads (Benjamin *et al.,* 2014). A range of quality metrics for read mapping rates was assessed, including depth of coverage (50 to 100x), read lengths (generally 150 base pairs in Paired-end mode), the percentage of reads aligning to the consensus sequence, RNA quality, assembly score, N50 read lengths, and N50 contig lengths. A mapping rate of 95% of reads aligning with the reference transcriptome typically signifies a successful assembly. A score of 1 in assembly, which signifies the retrieval of a single contig, was regarded as a successful result, while a score of 0 indicates total assembly failure (Dida and Yi, 2021).

4.2.5.2 *Assembly quality assessment*

The rnaQuality Assessment Tool for Genome Assemblies (rnaQUAST v.2.3.0), Bushmanova *et al.,* 2016) was utilized to assess the quality of the assembled RNA-Seq Moringa transcripts and evaluate how well the assembled transcripts match known genes. The Moringa assembled transcripts were aligned to the reference genome (AOCCv2) in FASTA format and its gene database in gff/gtf format using GMAP (gmap.avx2). The alignments were analysed to calculate simple metrics which were then matched against the isoforms from the gene database. Output gave statistics such as database coverage, percentage of correctly assembled isoforms and transcript completeness and correctness of the assembly. Supp.Table 4.3 display the several metrics that were evaluated. The presented metrics include contig length, the number of contigs aligned to the reference genome, hidden contigs, accurate contigs, misassembled contigs, as well as the corresponding legend and N50 statistic. Wang & Wang (2023) asserts that a substantial overall assembly size, an extensive number of contiguous, and an elevated N50 value are markers of a high-quality assembly. A high L50 value signifies a minimal amount of gaps in the assembly, indicating that most of the material is consolidated into a few big contigs. An assembly with a low L50 value generally consists of multiple smaller contigs (Castro and Fei, 2017).

4.2.5.3 Completeness assessment (BUSCO analysis)

The evaluation of transcriptome quality and completeness was conducted utilizing BUSCO version 5.4.3, following the methods established by Simão *et al.,* (2015). This evaluation employed the BUSCO hmmsearch tool (version 3.3) in conjunction with the lineage dataset eudicots_odb10, as specified in the eukaryotic classification revised on

September 10, 2020. The results were examined to evaluate the overall quality of the transcriptome.

4.2.5.4 GC %

The GC % was calculated as the total number of G and C nucleotides in the assembly, divided by the total length of the assembly.

4.3 Results

4.3.1 Assembly metrics

A total of 1,635,367,412 raw reads was assembled and the average assembly's N50 value was 2.8Kbp. The mean transcript count before the implementation of CDHIT is approximately 138,801, with a range spanning from 103,512 to 191,372. The average count, following the removal of extraneous transcripts, is approximately 103,488, with a range of 78,842 to 142,700. The average Total Assembly Length is approximately 178,367,680 bp, with a minimum of 117,305,226 (about 117 million) and a maximum of 272,851,364 (close to 273 million) base pairs. The majority of samples exhibit a minimum length of 201 base pairs; however, a few exceptions lower the average to approximately 199 base pairs. The maximum transcript length varies between 8,426 and 20,269 base pairs. The average length is approximately 1,714 base pairs, with a range from 1,273 to 2,129 base pairs.

The length of the shortest contig at 50% of the total assembly length (N50), the number of contigs that spans 50% of the total assembly length (L50) and the total length of the combination of all contigs in the assembly (Assembly length) are all pointers to the quality of the Reference-guided *De novo* Assembly of these 49 Moringa samples (Supp. Table 4.3)

Although the Moringa genome is relatively small compared to other flowering plants, it is compact yet complex owing to the integration of its plastid DNA into the nuclear genome and gene family expansions especially those related to chloroplast and photosynthetic functions (Ojeda-López *et al*., 2020). Generally, for complex plant genomes as Moringa having longer contigs above 20Kbp and N50 values of at least 2Kbp ensures better contiguity and completeness which leads to a higher - quality genome assembly. Nevertheless, the low N50 values obtained in all the samples is expected as the assembly was made from Illumina short reads. On the other hand, 2 of the *Moringa oleifera* accessions - R3 and Mo5 have contig sizes (Largest contig) of 20.3Kbp and 19.1Kbp, while others have lower values (Supp.Table 4.3).

Although the L50 values (ranging from 19.9K to 30.8K) across the samples implies that a smaller number of contigs cover 50% of the assembly suggesting a robust assembly, the L50 value needs to be assessed alongside the other metrics.

The assembly lengths are fairly good (Supp. Table 4.3). The top samples by number of transcripts and average length include KMs8 (265.7Mbp), KMs10 (261Mbp), KMs11 (260Mbp), KMs5 (243.3Mbp) (Kenya *Moringa stenopetala*), IMO9 (236.7Mbp) (India *Moringa oleifera*) and MO29 (226.9Mbp) (Nigeria *Moringa oleifera).*

4.3.2 BUSCO (Benchmarking Universal Single-Copy Orthologs)

Evaluating assembly quality is of paramount importance. The findings highlight the importance of utilizing diverse assessment metrics, including N50, L50, and BUSCO scores, to assess the continuity and accuracy of genomic assemblies, especially among closely related but distinct species. The compiled transcripts were evaluated for completeness and quantified utilizing BUSCO v5.4.3 with the core eudicots gene database, as illustrated in (Supp.Table 4.1). Among the 2326 predicted genes for the 49 Moringa accessions, an average of 94.65% were categorized as complete BUSCO genes; 36.71% were single-copy, whereas 57.74% were identified as duplicated. Among the remaining 5.54% of BUSCO genes, 1.59% were categorized as fragmented, while 3.95% were classified as absent. The recorded N50 values ranged from 2.2 Kbp to 3.1 Kbp, indicating a predominantly high-quality assembly. The assemblies exhibited total lengths between 111.3 Mbp and 265.7 Mbp. The average BUSCO completeness of 94.65%, indicated that the majority of expected genes were preserved, thus establishing a strong foundation for subsequent genetic research. 1.59% of the genes classified as fragmented genes must have lost some of their parts and so the full length were not adequately captured. Fragmented genes could hinder functional analyses and understanding gene function, but 1.59% fragmented genes in this study may be insignificant.

There was an average of 3.95% Missing genes which could also lead to incomplete annotations and may impact downstream analyses such as expression studies, evolutionary analyses and functional genomics.

4.3.3 Assessment of *De novo* Assembly using rnaQUAST

In terms of the completeness and correctness levels of the assemblies, a variation appears in total transcripts, number of transcripts longer than 500 base pairs (Transcripts>500bp), and average number of exons per isoform generated. The total

transcripts varied greatly among the samples, with 62,678 being the lowest (MO53) and the highest being 85,784 (R1), with an average approximately of 5.493 exons per isoform. This variation may reflect differing levels of gene expression between the samples or differences in assembly quality. There was also variability in the counts of transcripts > 500 bp, with some gene assemblies revealing from 47,688 (MO53) to 66,550 (IMo9). This measure is very important since, longer transcripts are more likely to be biologically relevant. Likewise, the numbers of transcripts > 1000 bp range from 36,528 (KMs5) to 53,202 (IMo9) Supp.Table 4.3. This means that a proportion of the transcripts is normally long, which is normally a property of functional genes.

The N50 and NG50 also represent assembly quality. N50 is the length of the shortest contig longer than any other contig comprising 50% of the total length of the genome. NG50 likewise is defined in the similar manner but considers the expected genome size. Assemblies showed different qualities whereby certain samples obtained high values of N50 indicating good continuity. An example is IMO9 assembly with N50 2480 bp, thus assuring quite a good quality of assembly. The genome coverage can perfectly be figured by the N50 and NG50 described previously. From the genome fraction percentages, the genome coverage by some assemblies was very substantial Fig. 4.2 and Fig.4.3.

The GC content in the 48 assembled transcriptomes was consistently around 71% Supp.Table 4.3. This possibly reflects some form of assembly bias and inherent transcriptome differences. Misassemblies and Local Misassemblies reflect the number of structural errors in the assembly; fewer misassemblies imply more accurate assembly. The number of misassemblies varies to a great risk, with samples like IMo9 maintaining 1317 misassemblies highlighting the potential for weaknesses in the assembly process that need to be rectified. Duplication Ratio, the ratio of duplicated sequences in the assembly are preferable when they are lower values, as they indicate fewer redundant sequences. The duplication ratios and error rates (mismatches and indels) are relatively high in some samples suggesting the need for further refinement and error correction in the assemblies. Lower values of mismatches and insertions/deletions per 100,000 base pairs. indicate higher sequence accuracy.

Correlation plots

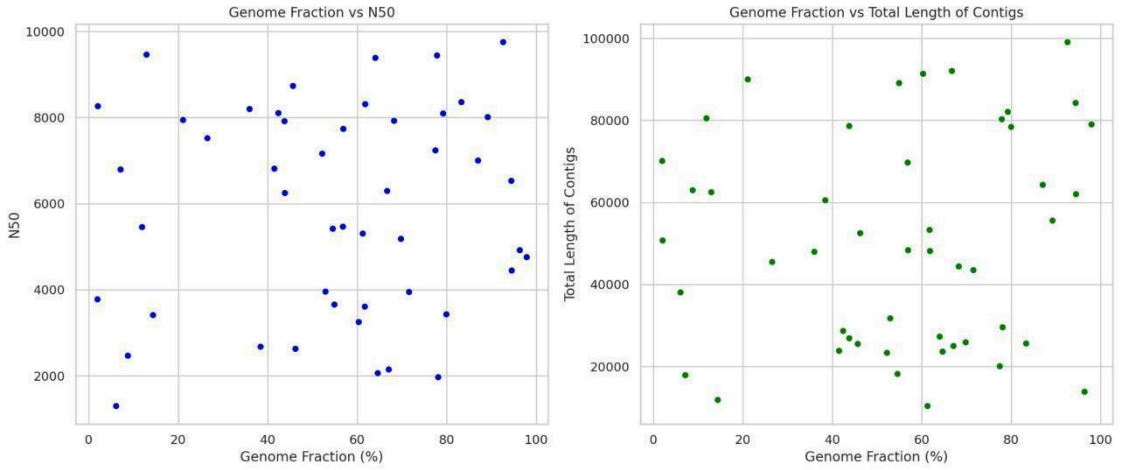


Figure 4.2: Scatter plots illustrating the correlation between Genome fraction and both N50 and total length of contigs of 48 Moringa Transcript Assemblies

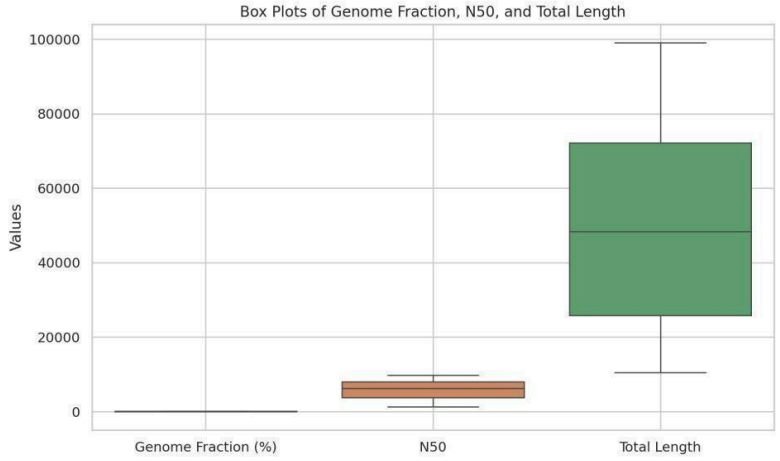


Figure 4.3: Box plots comparing N50 and Total length of transcripts

4.3.4 BUSCO 2 (Benchmarking Universal Single-Copy Orthologs)

BUSCO checked the completeness of the 48 assemblies by looking for a set of universal single-copy genes that most eukaryotes should have. The statistics show an average of 90% completeness which is excellent, however the assemblies generally have high duplication rate of 67% on the average (Supp.Table 4.2). This is typical for *de novo* transcriptomes because of the multiple isoforms TRINITY keeps for each gene. Also, alternative splicing and allelic variants appear as duplicates. In order to manage this situation for effective downstream analysis, the duplicate transcripts were collapsed and only the longest isoform per Trinity gene (longest ORF per TRINITY gene) were kept for Orthofinder analysis.

4.3.5 Annotation of un-characterized transcripts and isoforms

To identify un-characterised transcripts, that is RNA sequences that have not been previously identified or annotated in the existing reference genome databases, assembled transcripts were compared against existing databases using Pfam (InterPro). These un-annotated transcripts may arise from various biological processes and provides insight into gene regulation, alternative splicing and non-coding RNA functions (Alhabsi *et al.,* 2025). Reads mapped to intronic and intergenic regions may represent novel genes or transcripts, though these regions are often overlooked or discarded providing our understanding of the transcriptome ((Lloyd *et al.,* 2019)). The potential functions of these novel transcripts were explored using InterPro.

Two genes, Morol01g1284 (TRINITY_DN6300_c0_g1_i3) was identified to decrease in the Moringa stem under drought stress; and Morol08g10460 (LFC-3.9786; pvalue of -5.11E-34; padj.value of -5.18E-31; Cyan module) with transcript ID, TRINITY_DN6221_c0_g1_i1 was identified as one of the top 10 differentially expressed genes (DEGs) in Moringa drought stress but unannotated in the existing reference genome and could be regarded as novel. Their transcript id was searched from DIAMOND to obtain their amino acid sequences which served as input for the InterPro search where the protein family membership, domain relationship, TMHMM and functional annotation were identified.

Morol01g1284

>TRINITY_DN6300_c0_g1_i3 translated
ISSRTKMACWSAENATKAYLKTLKMGQKAKEPNTAEFISALAAGSNAQLMVVA
CAGAATS
SALALVAAAYQTGGRVVCILNSLEELTLSKDLLGVDACHIQFVLGDAQSLLSTH
YKEADFVLIDCNLENHEGIFKAVEQAGRKQNGAMVVGYNAFSKGSWRTSTESRTQLLPI
G
EGLLVTRMPASAKNDNGGHRFGKRSHWVVKVDKCTGEEHVFRVRLPPGKEIHA
Protein of unknown function DUF1442,This family consists of several hypothetical *Arabidopsis thaliana* proteins of around 225 residues in length. The function of this family is unknown.

Morol08g10460 with transcript id:

>TRINITY_DN6221_c0_g1_i1 translated

EMARSLSCAKSLAASVAGGLSLSISRGYAAASQGALPGSLARAGPRTAMMGKL

QKKAANKDDTEASSAWAPDPVTGYYRPGNCTQEIDPAALREMLLGHKVKPH

Annotation was done using InterPro (IPR004926), classified with PANTHER as Late

Embryogenis Abundant Protein 2-Related - Late Embryogenis Abundant Protein 2Related(PTHR33509) and Pfam as LEA_3a - Late embryogenesis abundant protein

(LEA_3a subfamily-PF03242) whose expression is induced by salt, drought and heat stress. Plant LEA proteins have been found to accumulate to high levels during periods of water deficit in vegetative organs.

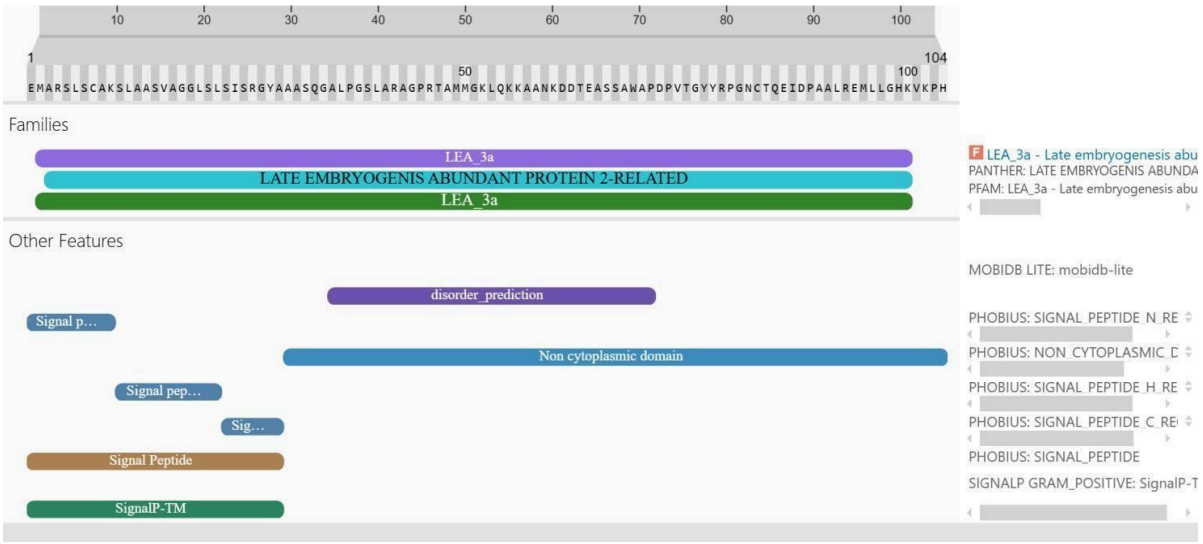


Fig.4.4 Identification of Late Embryogenis Abundant Protein 2Related(PTHR33509) from transcript id: >TRINITY_DN6221_c0_g1_i1 (Morol08g10460). PANTHER GO Terms, biological process: response to stress (GO : 0006950); Cellular component: mitochondrion (GO : 0005739)

4.3.6 Species-Specific Insights

Several variations were observed in the assembly metrics across the three species *M. oleifera, M. stenopetala,* and *M. peregrina.* The N50 value recorded for the *M. oleifera samples* were around 2.2 and 2.8, 2.5 in *M. peregrina* and 3.0 -3.1 in the *M. stenopetala* samples. The *M. stenopetala* samples showed the highest assembly length with KMs8 having 265.7Mbp, KMs10-261.0Mbp, KMs11-260.9Mbp and KMs5-243.3Mbp. The Oman *M. peregrina* sample has assembly length of 176.3Mbp, while the Kenya *M. oleifera* sample has the least assembly length of 113.3Mbp and the India *M. oleifera* sample IMo9 showed high value of 236.7Mbp like the *M. stenopetala* samples.

About 44 samples showed over 91 % complete genes after the BUSCO assessment except KMo2 - 87.8%, KMo6 - 88.8%, MO70 - 87.9% and Mo65 - 88.2% whereas some Nigeria *M. oleifera samples* Mo70, Mo65 showed 9.2%, Mo57 - 6.9% and Mo60 6.4% genes missing. The *M. stenopetala* samples showed lower value of 1.3 - 2.5% missing genes.

4.4 Discussion

This chapter presents a detailed analysis of transcriptome assembly metrics and quality assessments for various Moringa plant accessions, illustrating the efficacy of reference-guided *de novo* assembly technique. The comprehensive analysis of

sequencing quality and transcriptome assembly across the 49 Moringa accessions revealed several key findings that contribute to the understanding of the genomic landscape and the potential biological implications of these results in breeding enhancement.

De novo Assembly Metrics and BUSCO Assessment

The data on assembly metrics, including N50 values, Contig size, assembly length and related quality indicators, revealed significant variability in assembly quality among the samples. Variations in assembly metrics among the different Moringa species were notable. The N50 values and assembly lengths demonstrated that *M. stenopetala* samples exhibited the highest assembly lengths, indicating a more complete assembly compared to the *M. oleifera* samples.

The N50, contig size, and assembly length values obtained across all samples were generally low. This is expected as Illumina short reads were used for this *de novo* assembly. This can be improved by incorporating long read sequencing technologies such as Oxford Nanopore as this will provide longer reads which span repetitive regions and gaps. Notwithstanding, the *de novo* approach gave a hint of the assembly metrics of *Moringa stenopetala* as there has not been a published reference genome specifically for this *Moringa stenopetala* species. Most of the genomic research and resources available focus on other species within the genus such as *Moringa oleifera* (Shyamli *et al.,* 2021; Yang *et al.,* 2015).

The BUSCO assessment revealed that while 44 samples showed over 91% completeness, several Nigeria *M. oleifera* samples exhibited concerning levels of missing genes (e.g., Mo70 and Mo60 with only 6.4% complete genes). This discrepancy emphasises the need for careful consideration of geographical and biological factors that may influence genomic assembly and completeness.

GC Content Analysis

The average GC content across the Moringa accessions was found to be averagely 71% across all assembled transcriptomes, deviating from the typical range for eukaryotic organisms (40-60%). This high GC content could influence transcript stability which can also be linked to Moringa's adaptability to dry climates. This can be seen as an ecological advantage like complex gene regulation. Although, this might impact downstream analyses if not well managed. To manage this situation for effective downstream analysis,

transdecoder analysis was to identify the longest isoform per Trinity gene (longest ORF per TRINITY gene) kept for Orthofinder analysis.

Uncharacterised transcript/isoforms/genes

This transcriptome analysis identified uncharacterised transcripts/isoforms/genes linked to drought tolerance such as LEA_3a responsive to drought condition was discovered. The identification of this uncharacterised transcript will aid in the development of molecular markers for marker-assisted selection, allowing breeders to select more effectively for traits that enhance drought resistance. This knowledge is crucial for breeding programmes focused on improving drought resilience in Moringa species and expedite the development of Moringa varieties suited for arid conditions, significantly contributing to food security and agricultural sustainability.

Conclusion

Overall, the insights gained from this analysis reflect the importance of thorough quality assessment in RNA-Seq studies. The findings reveal both the potential and the limitations of the current dataset, emphasizing the need for further investigation into the biological implications of GC content, mapping quality, and transcriptome completeness. Addressing the challenges associated with low mapping rates by augmenting the short reads with long reads using nanopore sequencing. Likewise, exploring the underlying factors contributing to assembly variations will be crucial for better utilization of the Moringa transcriptome and its functional significance.

Appendix

Supp.Table 4.1 **BUSCO 1**
Supp.Table 4.2 **BUSCO 2**
Supp.Table 4.3.**rnaQUAST *de novo* Assembly**
Supp.Table 4.4. SoftwareTools and Versions:

Software versions	**Tool Versions**

Quality Control : FastQC	0.011.9
Trimming: Trimmomatic Trim Galore Cutadapt	0.36 0.6.4 2.3
ReadAlignment: STAR	2.7.10b
alling: GATK Freebayes	4.3.0.0 1.3.6
Bed tools Bcftools_merge	2.4.41 1.19
VCFTools VCF	0.1.16 4.2
Population Structure Analysis: TASSEL	 5.0
Bioconductor SNPRelate	1.40.0
PSIKO	2
Bioconductor GWASTools	3.20
Seqkit	2.3.1
Orthofinder	2.5.4
bamtools	2.5.1
BUSCO	5.4.3
QUAST	5.2.0
SAMtools	1.1.0
Salmon	1.10.0
TRINITY	2.15.1
CutAdapt	4.2
PLINK	2.00
CD-HIT	4.8.1

Supp.Table 4.5 nf-core/rnaseq Software Versions collected at run time from the software output.

Process Name	Software	Version
BEDTOOLS_GENOMECOV	bedtools	2.30.0
CUSTOM_DUMPSOFTWAREVERSIONS	python	3.11.7

	yaml	5.4.1
CUSTOM_GETCHROMSIZES	getchromsizes	1.16.1
DESEQ2_QC_STAR_SALMON	bioconductor-deseq2	1.28.0
	r-base	4.0.3
DUPRADAR	bioconductorduprada r	1.28.0
	r-base	4.2.1
FASTQC	fastqc	0.12.1
FQ_SUBSAMPLE	fq	0.9.1 (2022-02-22)
GFFREAD	gffread	0.12.1
GTF2BED	perl	5.26.2
GTF_FILTER	python	3.9.5
MAKE_TRANSCRIPTS_FASTA	rsem	1.3.1
	star	2.7.10a
PICARD_MARKDUPLICATES	picard	3.0.0
QUALIMAP_RNASEQ	qualimap	2.3
RSEQC_BAMSTAT	rseqc	5.0.2
RSEQC_INFEREXPERIMENT	rseqc	5.0.2
RSEQC_INNERDISTANCE	rseqc	5.0.2
RSEQC_JUNCTIONANNOTATION	rseqc	5.0.2
RSEQC_JUNCTIONSATURATION	rseqc	5.0.2
RSEQC_READDISTRIBUTION	rseqc	5.0.2
RSEQC_READDUPLICATION	rseqc	5.0.2
SALMON_INDEX	salmon	1.10.1
SALMON_QUANT	salmon	1.10.1
SAMTOOLS_FLAGSTAT	samtools	1.17
SAMTOOLS_IDXSTATS	samtools	1.17
SAMTOOLS_INDEX	samtools	1.17
SAMTOOLS_SORT	samtools	1.17
SAMTOOLS_STATS	samtools	1.17
SE_GENE	Bioconductor summarized experiment	1.24.0
	r-base	4.1.1
STAR_ALIGN	gawk	5.1.0
	samtools	1.16.1

	star	2.7.9a
STAR_GENOMEGENERATE	gawk	5.1.0
	samtools	1.16.1
	star	2.7.9a
STRINGTIE_STRINGTIE	stringtie	2.2.1
TRIMGALORE	cutadapt	3.4
	trimgalore	0.6.7
TX2GENE	python	3.9.5
TXIMPORT	bioconductor-tximeta	1.12.0
	r-base	4.1.1
UCSC_BEDCLIP	ucsc	377
UCSC_BEDGRAPHTOBIGWIG	ucsc	445
Workflow	Nextflow	24.04.2
	nf-core/rnaseq	3.14.0

Supp.Table 4. 6 nf-core/rnaseq Workflow Summary
This information was collected when the pipeline started.
Core Nextflow options
Revision 3.14.0
runName dorcas_rnaseq
containerEngine apptainer
launchDir/mnt/scratch/projects/biol-tf-2018/nextflow/run
workDir/mnt/scratch/projects/biol-tf 2018/nextflow/work
projectDir/mnt/scratch/projects/biol-tf-2018/nextflow/work/.nextflow/pipelines/490b65b5/nf-core/rnaseq
userName tld529
Profile york_viking
configFiles N/A
Input/output options
Input/mnt/scratch/projects/biol-tf-2018/projects/P2024-SJDCRJI/metadata/rnaseq_spreadsheet_samples.csv
Outdir /mnt/scratch/projects/biol-tf-2018/projects/P2024-SJD-CRJI/metadata/rnaseq_outdir

Reference genome options
Fasta/mnt/scratch/projects/biol-tf-2018/projects/P2024-SJDCRJI/reference_MoringaV2/MoringaV2.genome.fa
Gff/mnt/scratch/projects/biol-tf-2018/projects/P2024-SJDCRJI/reference_MoringaV2/MoringaV2.gff

Alignment
optionsmin_mapped_reads 5

Institutional config options
config_profile_description The University of York Viking profile
config_profile_contact Matthew Care
config_profile_url https://vikingdocs.york.ac.uk/

Max job request options
max_cpus 40 **max_memory** 500.GB **max_time** 48.h

Generic options
publish_dir_mode symlink

Chapter 5

Discussion & Conclusion

5.1 Discussion

5.1.1 Drought - Response in Moringa

It was important to do a research on the widely acclaimed drought-tolerant Moringa because though there are a number of researches acknowledging that *Moringa spp.* possess the ability to survive drought conditions (Bekka *et al.*, 2022), some have been able to report that its root systems, its leaf phenology, ability to control its stomata, and

osmotic adjustment (Hajaji, *et al*., 2024) have supported its adaptability to arid conditions, others have been able to reveal that drought could come by as a result of heat wave hence reported the role of heat shock transcription factors(HSF), (Shyamli *et al*., 2021), while only a few have been able to explain and provide genetic data on the mechanisms and network that control the drought-tolerance response. The limited information around the mechanisms responsible for its drought-tolerance which would clarify how Moringa responds to drought stress, the genes that interact to co-ordinate this adjustment and how this understanding can be used as a basis for plant breeding in areas with low rainfall or water shortage as well as improving biomass (leaves, pods and seeds) yield for food and industrial use.

As a result of this, I carried out an RNASeq experiment to explore the genetic basis of drought tolerance in a 5-day droughted *Moringa oleifera* plant, identified variations in the genes response to drought through differential expression analysis, studied the network of those genes (module) that increased as a result of drought that is co-expression network and identified the hub genes through WGCNA.

Findings from this study provided insights into both scientific and practical applications of drought-tolerance mechanisms of Moringa, which would be a valuable resource for plant breeding and selecting superior markers in the application of genetic engineering, thus helping mitigate the effect of climate change and enhancing food security.

Overview and Interpretation of findings

Investigating the response of Moringa species to drought has helped clarify the transcriptional networks enabling their survival in water-limited conditions. The key findings are compiled here as follows:

Differentially Expressed Genes (DEGs): The study exposed a complex transcriptional response to drought stress involving 9,249 differentially expressed genes (DEGs) over 49 co-expression modules. This huge dataset emphasises the complicated genetic network turned on in Moringa plants under dry conditions.

Notable Genes and Expression Patterns: While overexpression was seen for genes like Morol12g05670 (SAG12) and Morol01g26440, notable downregulation was seen for genes like Morol02g21300 (Peroxidase N1-like) and Morol02g21290 (Cationic Peroxidase 2-like). These trends, which differ based on gene and tissue type, suggest that different tissues may respond differently to drought stress.

Tissue-Specific Responses: The expression of genes varied across different tissues. For example, Morol03g14060 (Gawky-like protein) increased in stem tissue, whereas Morol07g03510 (Late embryogenesis abundant protein-related) decreased. This highlights the necessity for tailored responses in various plant structures during drought.

Identification of uncharacterised Genes: Finding uncharacterized genes is another intriguing result. One such a gene with significant differential expression was Morol08g10460. This result suggests that other genes yet unidentified that react to drought could exist. This creates chances for more research into their functions. Locating uncharacterised transcripts and isoforms (Chapter 4).

Co-expression Modules and Functional Insights: There was a substantial correlation between drought responses and modules like cyan, turquoise, blue, and green. The Cyan Module is rich in genes that shield against oxidative stress and genes that regulate gene expression. One such gene that might regulate the expression of genes reacting to drought is Morol10g08670 (CBFA/NFYB). Crucially for maintaining cellular integrity during a drought, the Turquoise Module is associated with stomatal regulation and cell wall remodelling including MYB-like transcription factors. The blue module contains phytohormone and osmo-protectant transporters, including ABC transporters; negative regulators of ABA signalling are located in the green module, which coordinates the plant's response to drought; and autophagy is associated with the red module, indicating increased cellular recycling to conserve resources during drought.

5.1.2 Moringa Seed Oil

In spite of the versatility of Moringa seed oil and its potential for great nutritional, medicinal, and industrial value, there is a need to fully understand the metabolic profile and genetics of its quality characteristics and their variance across geographical locations to put them into specific use. Although there have been studies on identifying specific genes like Heat shock transcription factors (Shyamli *et al.*, 2021), genetic diversity (Ondieki *et al.*,2023), comparative genomics of Moringa with implications for evolution of Brassicales (AbdAlla *et al.*, 2023) nutritional and pharmacological properties (Pareek *et al.*, 2023; Azlan *et al.*, 2023), oxidative stability, majorly on the fatty acid composition (Leone *et al.*, 2016; Nadeem & Imran, 2016), with less attention on the phytosterols and tocopherols, and evolutionary areas. There is almost no GWAS study integrating genomic data, with the oil metabolic profile and diverse geographical locations (latitudinal variation). This gap limits targeted breeding of Moringa for quality oil profiles and adaptaion to varying geographical locations.

In a bid to bridge this gap, I hypothesised that specific SNP variants are significantly associated with variation in Moringa fatty acids, sterols and tocopherol profiles; the metabolic oil profile traits show geographically structured variation which reflects genotype-environment interactions; and combining GWAS, biochemical profiling, geographical location to discover stable, breeding-relevant alleles.

This study therefore engaged a GWAS to identify specific genetic markers (SNPs) linked to 45 oil traits across 49 Moringa populations by integrating their genetic profiles and detailed oil analyses focusing on fatty acids, sterols and tocopherols to identify these associations and generate markers that would guide breeding for both edible and industrial uses. It also focused on providing leads to generating marker-assisted tools for breeding Moringa seeds with quality oil for dual purposes. Establishing a dual breeding pipelines including edible oil lines with enriched oleic acid, α-tocopherol, β-sitosterol, reduced behenic acid and industrial oil lines having VLCFA-rich, specialised sterols and balanced tocopherols. And identifying Moringa as a versatile, drought- resistant crop which can contribute to food security, nutrition, biofuel energy, cosmetics industries and agriculture.

This chapter examined seed trait data and seed oil profiles of 49 Moringa accessions consisting 2 species, *Moringa oleifera and M stenopetala*, for their phytosterol, tocopherol and fatty acid composition, seed germination, and oil extraction. The study included varied geographic locations. The results of Moringa seed oil and seeds drew attention to several crucial aspects influenced by species and environment. The seed oil properties of Moringa vary greatly depending on environmental factors and genetic background; moreover, the seeds showed hypogeal germination.

This chapter examined seed trait data and seed oil profiles of 49 Moringa accessions from varied geographic locations consisting 2 species, *M. oleifera* and *M. stenopetala*, for their phytosterol, tocopherol and fatty acid composition, seed germination, and oil extraction. The study included GWAS analysis, statistical analyses of the seed, seed oil traits, SNP Data analysis, bioinformatics analysis and GLM association mapping of the oil profiles. The results of Moringa seed oil and seeds inferred influence by species and environment, as the seed oil properties of Moringa vary depending on environmental factors and genetic background. Recommendations for improving Moringa oil, nutritional and industrial cultivar development highlight the importance of integrating genomic, biochemical and ecological information to guide Moringa breeding.

Overview and Interpretation of the findings

The mean Oil extraction efficiency with hexane of 99.96% in this research suggests that there is a correlation between seed weight, oil production, and extraction efficiency which exceeds the reports from Mani *et al.*, 2004 - hexane extraction; Chen *et al.*,

2021-supercritical CO2 extraction with 98.43%; Dinesha *et al*., 2018- Soxhlet extraction with 76.29%.

The differences in oil yield and composition between *M. oleifera* and *M. stenopetala* are consistent with earlier studies. According to Pluháčková e*t al.,* 2023, *M. oleifera* yielded higher oil content and oleic acid content than *M. stenopetala* but lower palmitic acid and total tocopherols in the Ethiopian populations studied. This data in line with my research where, *M. oleifera* seeds generated the most oil even though *M. stenopetala* seeds were bigger. This offers more proof that oil quality is defined in great part by species-specific traits including genetic variables influencing oil composition and yield. This study shows that Moringa oil has high yield levels of useful unsaturated fatty acid derivatives especially C16.1D9 Palmitoleic acid (34.49) in MO26, C18.1D9 Oleic acid (971.64) in KMO7, C18.2 Linolenic acid (12.12) in MO26 , C18.3n3 AlphaLinoleic acid (3.97) in MO26 , C20.1n11 Gadoleic acid (39.85) in MO16, C20.1n13 Erucic acid (2.08) in KMO7, and C24.0 Lignoceric acid (23.58) in KMs10. These give Moringa oil beneficial properties for nutritional and pharmacological applications. These findings are in conformity with Aly, Maraei and Ali, 2016, who reported up to 76.29% oleic acid, Zhao *et al*. 2019 reported total unsaturated fatty acids amounted up to 76.0%, Leone *et al*., 2016 among others.

Geographical factors like latitude, altitude, seasonality and climate which describe location, can influence Moringa seeds' oil content and fatty acid profiles. Locations have been reported to cause variations in Moringa products such as leaves and seeds which in the long run affect oil quality, composition and quantity as reported by Ayerza 2019; Haile *et al*., 2019; Ukwuezu *et al*.,2019. While seeds from warmer climates are more likely to yield increased saturated fatty acids, seeds from colder climates are more likely to yield increased polyunsaturated fatty acids. Locations from where seeds were obtained for this research are generally warm to hot with temperature ranging from 26℃ in the Philipines, SW Nigeria to 40℃ in Oman, except the Plateau, Nigeria which is generally cool at temperature of 18-22℃. There was statistically significant geographical influence (ANOVA $p<0.05$) on specific oil constituents across all accessions and species in this study as was seen in some FAMEs, including Erucic C22.1n13, Arachidic C20.0, Gadoleic C20.1n11 and Lignoceric C24.0 acid. These could be regarded as key markers for selecting superior accessions suitable for edible and industrial use. These findings emphasise the importance of considering local growing conditions while assessing oil quality. On the other hand, phytosterols including campesterol, campestanol, citrostadienol were not statistically significant, suggesting minimal geographical influence across all sample locations despite their numerical increases in the regions. While

phytosterols may have other biological relevance, FAMEs and alpha-tocopherol data may provide good breeding decisions.

Among all *Moringa oleifera* samples, excluding *M. stenopetala* accessions, specific phytosterols such as cholesterol, stigmasterol, cholestanol along with Beta-Tocopherol revealed significant differences between location pairs from Nigeria, Oyo, Burkina Faso and Kenya that could guide breeding or regional selection decisions.

Single nucleotide polymorphisms (SNPs), which are genetic features connected to the fatty acid composition of Moringa seeds from different geographical locations, are associated with variations in phytosterol, tocopherol, and fatty acid levels. These SNPs serve as markers that help explain the important traits, including increased oleic acid levels. Some genes at well-known loci, such TIM22-4 on chromosome 7, which is linked to oleic acid (C18.1D9) concentrations, help to transport proteins, oxidise fatty acids, and react to drought stress.

Furthermore, genes linked to phytosterols include those associated with the Δ^7-stigmastenol level and P450CYP82 on Chromosome 5, Δ^5-avenasterol), as well as genes on chromosomes 2, 8, and 10 respectively that are linked to β-tocopherol production (bHLH63, SPB, GTFIIE, and LecithinCholesterol Acytransferase), along with additional genes on chromosome 5 and 4 also related to phytosterols. Δ^5-avenasterol enhances plant tolerance and lipid metabolism by interactions with the ABC transporter G family, found on chromosome 13. Environmental elements have several effects on the oil content of Moringa seeds.

Phytosterol and Tocopherol Content: Two healthy compounds, phytosterols and tocopherols, were detected in different amounts among several Moringa accessions. The study indicates that some accessions from specific sites contain more levels of these compounds, which are recognised to have health advantages. This variation affects the therapeutic qualities and nutritional worth of Moringa oil. The Kenya *M. stenopetala* sample (KMs8) is also likely to exhibit a VTE mutation. VTE4 codes this enzyme, which converts γ-tocopherol into α-tocopherol. Should the VTE4 gene be active, α-tocopherol levels will be high. Variations in the VTE4 gene could influence the following features of the Kenya *M. stenopetala* (KMs8) seed. The function of the VTE4 gene is connected with the very low γ-tocopherol concentration (1.69%) and highest α-tocopherol level (63.59%) in KMs8. An effective mechanism for converting γ-tocopherol to α-tocopherol results in low amounts of γ-tocopherol should VTE4 be present. Although VTE4 mutations are not directly linked to sterol production, their impact on the metabolic network as a whole may be significant given tocopherol composition is the most notable

effect. This is consistent with the signs of an effective sterol biosynthesis pathway, the high amounts of oleic acid (73.92%) and β-sitosterol (61.1%).

Studies in this chapter helped to clarify how both genetic (species-specific) and environmental (geographical) factors affect the quality of Moringa oil and would help one to maximise developing techniques and oil extraction technologies for commercial and health-related purposes.

5.1.3 Transcriptomics

Transcriptome studies on *Moringa oleifera, Moringa stenopetala* and *Moringa peregrinna* provide transcriptomic database resource for series of downstream analysis including those required for improving Moringa varieties that can survive under harsh conditions, thereby enabling them to remain sustainable crops despite the ongoing challenges of climate change affecting agriculture as well as providing evolutionary insights, oil metabolic profiling, and applied uses in industry and agriculture. The rationale for constructing *de novo* assemblies of the 49 Moringa accessions was to bridge the gap of limited molecular resources for other *Moringa spp.,* and to provide species-specific databases that would enable higher-level analyses.

Reference - guided transcriptomes were compiled from 49 Moringa plant accessions collected from seven geographical locations of the world. I applied methods for building transcriptome libraries from scratch such as TRINITY, applied a reference-guided approach, assessed the quality using rnaQUAST, assessed the completeness and quality of the annotations based on evolutionarily informed expectations of gene content using BUSCO and compared the assembly metrics of the reference-guided and *de novo* transcriptome assembly to make decisions of the quality of the assemblies. This study found varying success rates for reference-guided and *de novo* assemblies depending on the selected approach. Laboratory validation is essential to confirm the assemblies. The final output will be a great tool for researchers investigating species of the Moringa genus. Comparatively, I can identify significant variations and evaluate the assemblies' genetic features.

This study is significant in many ways, firstly this study provides the first transcriptomic resources for genomic orphans such as *Moringa stenopetala* and *Moringa peregrinna*. These assembled transcriptomes provide foundational resources where there were none existing. With the findings of this study, comparative analyses of the 3 species are made possible. This would allow for better understanding of species-specific conserved genes across the Moringa genus, adaptation to varied environments and expanding the Moringa genomic landscape. Identification of candidate genes in stress response network and

biosynthetic pathways of seed oil metabolism is made possible due to this study. This study will also bring these species which are unpopular, understudied and locally important to East Africa and Middle East into global sustainability development goals discussions on food security and climate resilience. However, there were some limitations and potential biases associated with the reference-guided assembly process.

Using a reference genome similar to *M. oleifera* but distinct from the closely related *M. stenopetala* and *M. peregrina* caused biases to the outcomes as previously reported by Smith-Unna, *et al.,* 2016. This could be the result of the areas either missing the reference or being more divergent. Reconstructing assemblies or variants unique to particular species can be difficult, and the target assembly could have less diversity than the real genome. To lessen the impact of this problem, improvement to the assemblies could make them more accurate and thorough hence I have followed the reference-guided and *de novo* assembly approaches suggested by Sim *et al*., 2023; Lischer and Shimizu, 2017. In line with Behera *et al*., (2022), I have also assessed the assemblies using other criteria outside contiguity, such as BUSCO completeness, read mapping rates, transcript metrics such as transcript length, alignment metrics such as aligned, uniquely aligned, assembly completeness (sensitivity) such as database coverage, percentage of assembled isoforms, assembly (specificity) such as 50%-matched, 95%- matched rates therefore ensuring a comprehensive assessment of assembly quality.

5.2 Integration of Results: Implications and Applications

This work provides a whole picture of the process from genetic, evolutionary, and metabolic aspects, therefore revealing how Moringa responds to drought stress. Components of the drought response system in Moringa (Kim *et al.,* 2024) included molecular chaperones, Abscisic acid (ABA) detection and signalling, and activation of genes producing Late embryogenesis abundant (LEA) proteins, Morol07g03510 and Morol08g10460. The results showed declining LEA expression inside the stem tissues. The ABCG gene Morol01g26020 (blue module) most likely controls phytohormone (ABA) trafficking in response to a drought. The green module also comprises the AT1G67300 gene (Morol06g06410) and the PP2CG1 gene (Morol05g00780), both of which interact with other drought-responsive pathways to improve plant coordination and function as negative regulators of ABA signalling. Using RNA-Seq and differential expression, research has shown genes sensitive to drought including Morol12g05670 - SAG12, Morol01g09210 - Vignain, Morol02g21300 - KNAT2, Morol12g01460 - SBH1, and Morol03g14060 - Gawky. The same is true for enzymes such as

Xyloglucanglycosyltransferase6 (Morol08g14580) and Glutathione reductase (Morol10g09360) as well as transcription factors CBFA/NFYB (Morol10g08670), bHLH85 (Morol05g16040) and MYB-like TF (Morol02g05650). Co-expressed genes included the modules, Yellow (which included the top genes up-regulated in the stem and leaves; SAG12, Metallothionein-like, and Ferritin-3), Cyan (CBFA/NFYB), and Red. Turquoise (Morol05g00210), Galactan β-1,4-galactosyltransferase (GALS1), Green (Morol06g06410, AT1G67300), and Hub genes known to be the main controllers of drought response in Moringa were the genes identified as down-regulated ones.

Oil production was shown to be correlated with expression of transcription factors, genes, and enzymes linked with drought stress. PRX Morol04g12350; Morol13g04900; and Morol05g06560, which is associated with Chrom5 of Δ^7- stigmastenol and Chrom12 of Obtusifol; PRX Morol04g123; and ABC Transporter Morol13g04900, which is related with Chrom13 of Δ^5-avenasterol; down-regulation of the following genes showed by the drought-stressed Moringa leaves: PRX Morol04g12350; and Morol13g04900. Just one instance of how plants adapt to their surroundings is survival over development and oil production. They use stress-responsive transcription factors to do this since they inhibit genes linked to oil production so conserving resources and energy (Kim *et al.,* 2024).

One of the high differentially expressed genes (DEGs) Morol08g10460 in drought stress in Moringa was identified as transcript with the ID TRINITY_DN6221_c0_g1_i1 (LFC3.9786; pvalue-5.11E-34; *padj*. value -5.18E-31); current Moringa reference genome (AOCCv2) does not contain information on Morol08g10604. Pfam has classified this protein as PTHR33509, Late Embryogenesis Abundant Protein 2-Related, LEA 3a member of the late embryogenesis abundant protein subfamily (PF03242).

Analysing the metabolic profile of Moringa seed oil from 49 various sources helped to better understand the biochemical pathways affected by drought stress and the genes engaged in the synthesis of important metabolites. For instance, Morol05g06560 (CytP450) helps sterol intermediates hydroxylate by means of formation of sterols including Δ^7- stigmastenol and Obtusifoliol. Morol04g12350 (PRX), most likely engaged with peroxisomal functions, is fundamental for fatty acid β-oxidation and lipid metabolism. Δ^5-avenasterol among several other types of lipids is translocated across membranes by the ABC transporter Morol13g04900.

In the context of genetic engineering, agriculture, conservation, and breeding programmes, the findings of this study could find application in several different directions. Breeding techniques can be directed by the identification of genes and processes sensitive to drought, therefore enabling Moringa cultivars resistant to drought. The results allow one to genetically edit Moringa plants to be more drought resistant by either raising

or lowering their expression of specific genes or by altering their expression of others. Understanding the metabolic changes that take place during a drought helps one to maximise agricultural activities to increase Moringa output under pressure. Ultimately, phylogenetic understanding can aid to better maintain populations of moringa that are both drought-adapted and genetically varied.

5.3 Impact Statement

This RNASeq experiment clarifies the molecular basis of drought tolerance in *Moringa oleifera,* as the regulatory modules and hub genes were revealed through differential expression and co-expression networks that makes Moringa's adaptive response to drought possible. This study offers functional genomic resources for *Moringa oleifera,* a plant known for its resilience but lacking comprehensive molecular characterisation. As much as these findings strengthen the genetic basis for using Moringa in climate-resilient agriculture, they also enable the transfer of the important tolerance mechanisms identified in Moringa into other crops. Therefore, I highlight the key contributions of this experiment to scientific knowledge.

1. This research goes beyond the general claim that Moringa is drought-tolerant but specifies genes which are activated or suppressed under drought stress, thereby providing gene–level resolution of drought response.
2. This research sheds light on the co-ordinated responses of a group of genes, unveiling a network of genes rather than focusing on isolated genes.
3. The identification of hub genes as master regulators provides marker-assisted selection for breeding in Moringa and other crops.
4. I am creating a valuable dataset that can be utilised by future researchers in fields such as comparative genomics, stress physiology or biotechnological applications.
5. This study links the molecular mechanisms of Moringa's drought resilience to agriculture.

My study on Moringa seed oil, which integrated genotype-phenotype-environment data, is novel in orphan crops, as it is among the first to combine GWAS with RNA-seq SNPs, biochemical oil profiling and geo-location data in *Moringa spp.,* which sets a methodological precedent. The following are key impacts of this study:

1. Identification of SNP markers linked to fatty acids, phytosterols and tocopherols which has provided genetic insights into the quality of Moringa oil and clarified how oil traits are regulated in Moringa, which was previously poorly understood.
2. The proposal of Δ^5-avenasterol, Δ^7-stigmastenol, β-Sitosterol as chemotaxonomic and functional markers is novel, as my research identified sterol and tocopherol diversity as breeding targets, linked them to stress adaptation and phytochemical value as most previous studies emphasized yield and fatty acid profiles.

3. Identification of potential climate-resilient alleles contributing to the understanding of genotype - environment interactions in perennial oil seed trees like Moringa.
4. The study proposes the use of climate-associated alleles to enhance oil stability and improve edible or nutritional quality.
5. Methodological contribution through established pipelines: i) Trinity —SNP discovery—----GWAS—---Statistical validation (PCA, Clustering, Regression) —---- biochemical integration ii) proposed a marker validation framework and dual breeding pipelines (edible vs industrial oil) which can be adapted to other underutilized crops.
6. Conceptualised a dual-purpose breeding pipeline: Edible oil lines where accessions with properties such as high oleic, low behenic, enriched α-tocopherol and β-Sitosterol- rich are the ultimate choice. An industrial oil line characterised by VLCFA-rich content, balanced tocopherols and specialized sterols is also a key feature.
7. The research provides the first evidence that the tocopherol and sterol profiles in *Moringa spp*. vary systematically by geography.
8. Genetics and location impact behenic acid accumulation, thereby providing a method to reduce it in edible oil.

These findings do not only expand scientific knowledge of oilseed biology in orphan crops such as Moringa but also have great benefits and implications for agriculture, food, cosmetics, pharmacological and renewable industries.

It enables farmers access to climate - resilient, higher-value Moringa varieties and improved seed quality which translates to higher prices and new market opportunities. This finding also opens up diversification benefits of Moringa as a cash and perennial crop, along with serving as a nutritious food source. Plant breeders are able to select for high oleic and α-tocopherol oils to obtain nutritional quality and oxidative stability.

Through the transcriptome study,

1. 49 novel transcriptome assemblies were generated that have not been previously characterised.
2. The dataset created provides a comparative framework within the Moringa genus for species comparison, gene families, species-specific adaptations and pathway conservation.
3. The assemblies provide resources for molecular marker development, such as SNPs and for selection of traits related to stress tolerance, oil composition or nutrition.
4. The multiple assembled transcriptomes also allow researchers to infer evolutionary relationships and understand how genera diversify in drought-prone climates, thus serving as a basis for phylogenomics.

5.4 Future Research Directions

Although some clear differences exist, the *de novo* assemblies generally align well with the AOCCv2 reference genome. These differences show the complexity of the Moringa genome and the possibilities for genetic heterogeneity among accessions. Resolving these discrepancies could help to clarify the functional biology of Moringa species by means of long-read sequencing technology or bettering the quality of reference genomes, therefore enabling increased dependability of genomic analysis. Further research on these variations and their biological relevance will help one to grasp the genetic terrain and evolution of some significant plants.

In addition, there is a need for laboratory validation and the use of transcriptome data as a resource for further genetic studies. Validation of the presence, expression and modifications of candidate genes encoding enzymes leading to differentially expressed genes, the synthesis of FAMEs, sterols and tocopherols, using quantitative RT-qPCR, a reliable and robust technique to validate the presence and expression of genes. Functional experiments in model systems, such overexpression or knockdown, help one to better appreciate the importance of genes in drought resistance.

5.5 Overall Conclusions

Numerous research and development opportunities exist to improve food security and agricultural practices using Moringa as a sustainable agricultural resource. With the global population estimated to reach 10 billion by 2050, food production must double to cater for this huge population, bearing in mind other challenging factors affecting food production such as climate change, greenhouse gas emissions and the over-exploitation of natural resources such as land, water and biodiversity. Climate change effects, such as flooding, droughts and disease outbreaks usually disrupt food production and food availability for a lot of vulnerable populations; therefore, alternative climate-resilient crops that are nutritious and can play environmental and food production roles are needed.

By producing drought-resistant crop varieties with guidance from markers identified in Moringa, stable yields even during droughts is ensured, irrigation is optimised as less water is required, and this supports sustainable farming practice. The Moringa trees benefits the environment by contributing to carbon sequestration as they help mitigate climate change by absorbing carbon dioxide from the atmosphere. They are also valuable for land restoration due to their ability to grow in degraded soils and improve soil fertility. Moringa's high nutritional value presents it as a leafy food source for combating malnutrition and improving food security in vulnerable regions. Incorporating Moringa into local diets diversifies food sources and enhances food security by providing alternative

food options during periods of crop failure. Moringa cultivation also presents economic opportunities where farmers can achieve better livelihoods and economic stability.

This work linked genomes, transcriptomics, proteomics, and metabolomics to offer a fresh approach for multi-omics research of plant responses to drought stress. It exposed regulatory networks and connections at various molecular levels and found important genes, proteins, and metabolites linked in oil metabolism and stress responses. This approach drives development in plant biotechnology since it opens the path for fresh research in fields such stress tolerance, nutritional enhancement, sustainable agriculture, and climate change mitigating technologies.

The findings of Moringa seed oil span many different fields. It is renewable and sustainable as another energy source for the creation of biofuel. Its minimal environmental effect, great oxidative stability, and remarkable cetane number about 67 improve its storage and transit qualities, therefore providing a consistent source of biofuel. It advances both sustainability and energy security. Its profusion of monounsaturated fats, sterols, and tocopherols makes it a great dietary supplement for supporting general health. Many cosmetics and personal care products have moringa oil as a common ingredient since of its cleansing and moisturising properties. Making use of Moringa seed oil will help one progress sustainability, health, and general well-being.

BIBLIOGRAPHY

Abdalla, M., Wanga, V.O, Mkala, E.M., Amenu, S.G., Amar, M.H., Chen, L. and Wang, Q. (2023). Comparative genomics analysis of endangered wild Egyptian *Moringa peregrina* (Forssk.) Fiori plastome, with implications for the evolution of Brassicales order. *Frontiers in Genetics*, 14. doi:https://doi.org/10.3389/fgene.2023.1131644.

Abdulkarim, S.M., Long, K., Lai, O.M., Muhammad, S.K.S. and Ghazali, H.M. (2005). Some physico-chemical properties of *MoringVa oleifera* seed oil extracted using solvent and aqueous enzymatic methods. *Food Chemistry*, 93(2), pp.253–263. doi:https://doi.org/10.1016/j.foodchem.2004.09.023.

Abdullah, M.A., Ab Rashid, M.F.F. and Ghazalli, Z. (2018). Optimization of Assembly Sequence Planning Using Soft Computing Approaches: A Review. *Archives of Computational Methods in Engineering*, 26(2), pp.461–474. doi:https://doi.org/10.1007/s11831-018-9250-y.

Aboobucker, S.I and Suza, W.P (2019). Why Do Plants Convert Sitosterol to Stigmasterol? Front. Plant Sci. 10:354. doi: 10.3389/fpls.2019.00354

Aboobucker, S. I., & Suza, W. P. (2019). Regulation of phytosterol biosynthesis under abiotic stress. *Plant Cell Reports, 38*(4), 487–497.

Adesina B.T, Oguntuga, O.A, Raimi, K. A. A, Ogunremi, B. J (2013). Guide to Large Scale Production of *Moringa oleifera* Lam for Sustainable Aquaculture Development in Nigeria Prospects and Challenges. *Agrosearch* 13:3 https://www.researchgate.net/publication/287812626

Adzigbe, J., Frimpong,F., Danquah,F., Eric Yirenkyi Danquah, Isaac Kojo Asante, Samuel Oppong Abebrese, Richard Dormatey, Charles Afriyie-Debrah, Priscilla Francisco Ribeiro, Eric Owusu Danquah, Kennedy Agyeman, Ralph Kwame Bam, Maxwell Darko Asante,(2025).The responses and adaptations of rice (Oryza sativa L.) to drought stress: A review, *Climate Smart Agriculture*, Volume 2, Issue 4,2025, 100080,ISSN 2950-4090, https://doi.org/10.1016/j.csag.2025.100080.https://www.sciencedirect.com/science/article/pii/S2950409025000413

Agrictechmea, (2023) Nigeria unveils plans to establish 15,000ha of moringa tree crop farms (Updated November 1, 2023) www.agritechmea.com/15000ha-of-moringa-tree-crop-farms-to-be-established-in-nigeria

Ahmad, P., Jamsheed, S., Hameed, A., Rasool, S., Sharma, I., Azooz, M. M., and Hasanuzzaman, M., (2014). Chapter11-Drought Stress Induced Oxidative Damage and

Antioxidants in Plants. Science Direct pp 345-367ISBN 9780127999630 http://dx.doi.org/10.1016/B978-0-12-799963-0.00011-3

Ahmad S, da Costa Gonzales L J, Bowler-Barnett E H, Rice D L, Kim M, Wijerathne S, Luciani A, Kandasaamy S, Luo J, Watkins X, Turner E, Martin M J, UniProt Consortium The UniProt website API: facilitating programmatic access to protein knowledge Nucleic Acids Research, gkaf394 (2025)

Ahmad, P., Umar, S., Amjad, H., Srivastava, D.K., Abdin, M.Z., Thakur, A.K., Farooqi, H. (2021). Role of Antioxidants in Plants. In: Eds. Ahmad, P and Umar,S Oxidative Stress: Role of Antioxidants in Plants. Stadium Press India Pvt. Ltd.

Ahmad, Z., Waraich, E.A., Akhtar, S., Anjum, S., Ahmad, T., Mahboob, W., Hafeez, O.B.A., Tapera, T., Labuschagne, M. and Rizwan, M. (2018). Physiological responses of wheat to drought stress and its mitigation approaches. *Acta Physiologiae Plantarum*, 40(4). doi:https://doi.org/10.1007/s11738-018-2651-6.

Ahn, H.K., Kang, Y. W., Lim, H. M., Hwang, I., & Pai, H.-S. (2015). Physiological Functions of the COPI Complex in Higher Plants. *Molecules and Cells.* Elsevier BV. https://doi.org/10.14348/molcells.2015.0115

Alavilli, H., Poli, Y., Verma, K.S., Kumar, V., Gupta, S., Chaudhary, V., Jyoti, A., Sahi, S.V.,Kothari, S.L. and Jain, A. (2022). Miracle Tree *Moringa oleifera*: Status of the Genetic Diversity, Breeding, In Vitro Propagation, and a Cogent Source of Commercial Functional Food and NonFood Products. *Plants*, 11(22), p.3132. doi:https://doi.org/10.3390/plants11223132.

Alhabsi, A., Ling, Y., Crespi, M., Reddy, A.S.N. and Mahfouz, M. (2025). Alternative Splicing Dynamics in Plant Adaptive Responses to Stress. *Annual review of plant biology*, [online] pp.10.1146/annurev-arplant-083123-090055. doi:https://doi.org/10.1146/annurev-arplant083123-090055.

Al-Khalifah, N.S. and Shanavaskhan, A.E. (2017). *Moringa oleifera* Lam., a promising crop species for arid conditions of Saudi Arabia and *Moringa peregrina* (Forssk.) Fiori, a native wild species for crop improvement. Acta Hortic. 1158, 159-170
DOI: 10.17660/ActaHortic.2017.1158.19
https://doi.org/10.17660/ActaHortic.2017.1158.19

Al Khateeb, W., Bahar, E., Lahham, J. et al.(2013) Regeneration and assessment of genetic fidelity of the endangered tree Moringa peregrina (Forsk.) Fiori using Inter Simple Sequence Repeat (ISSR). Physiol Mol Biol Plants 19, 157–164 (2013). https://doi.org/10.1007/s12298-012-0149-z

Alseekh, S., Kostova, D., Bulut, M. *et al.* Genome-wide association studies: assessing trait characteristics in model and crop plants. *Cell. Mol. Life Sci.* 78, 5743–5754 (2021). https://doi.org/10.1007/s00018-021-03868-w

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, [online] 215(3), pp.403–410. doi:https://doi.org/10.1016/S0022-2836(05)80360-2.

Alvarezbuylla, E, Benitez, M., Davila, E., Chaos A, Espinosasoto C and Padillalongoria P, (2006). Gene Regulatory Network Models for Plant Development. *Current Opinion in Plant Biology*, 10(1), pp.83–91. doi:https://doi.org/10.1016/j.pbi.2006.11.008.

Aly, A.A, Maraei, R.W., and Ali, H.G.M., (2016) Fatty Acids Profile and Chemical Composition of Egyptian *Moringa oleifera* Seed Oils Journal of the American Oil Chemists' Society 93:3

Amaranth - Wikiwand (nd) https://www.wikiwand.com/en/articles/Amaranth (Accessed: 31 December 2025).

Amin, N, Du, Y., Lu, L., Khalifa, M.A.S, Ahmad, N., Ahmad, S., Wang, P. (2024). *GmNAC3* acts as a key regulator in soybean against drought stress. *Science Direct* https://www.sciencedirect.com/science/article/pii/S2214662824000288?via%3Dihub

Amiteye, S. (2021). Basic concepts and methodologies of DNA marker systems in plant molecular breeding. *Heliyon*, 7(10), p.e08093. doi:https://doi.org/10.1016/j.heliyon.2021.e08093.

Ana Clara Sans Salomão Brunow Ventura, Thalita de Paula, Jenifer Pendiuk Gonçalves, Bruna da Silva Soley, Ananda Beatriz Munhoz Cretella, Michel Fleith Otuki, Daniela Almeida Cabrini (2021).The oil from *Moringa oleifera* seeds accelerates chronic skin wound healing, *Phytomedicine Plus*, Volume 1, Issue 3,2021,100099, ISSN 2667-0313,https://doi.org/10.1016/j.phyplu.2021.100099. (https://www.sciencedirect.com/science/article/pii/S2667031321000816)

Anbazhakan, S., Dhandapani, R., Anandhakumar, P. and Balu S. (2007). Traditional medicinal knowledge on *Moringa concanensis* Nimmo of Perambalur district, Tamilnadu. Ancient Science of Life 26: 42-45. 11

Andrews, S. (2010). *Babraham Bioinformatics - FastQC a quality control tool for high throughput sequence data*. [online] Babraham.ac.uk. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Anwar, F., Ashraf, M. and Bhanger, M.I. (2005). Interprovenance variation in the composition of *Moringa oleifera* oilseeds from Pakistan. *Journal of the American Oil Chemists' Society*, 82(1), pp.45–51. doi:https://doi.org/10.1007/s11746-005-1041-1.

Anwar, F. and Bhanger, M.I. (2003). Analytical Characterization of *Moringa oleifera* Seed Oil Grown in Temperate Regions of Pakistan. *Journal of Agricultural and Food Chemistry*, 51(22), pp.6558–6563. doi:https://doi.org/10.1021/jf0209894.

Anwar F, Bhanger MI, Yasmeen S. (2003). Antioxidant activity of some natural extracts in corn oil. In: Murata N, Yamada M, Nishida I, Okuyama H, Sekiya J, Hajime W, editors. Advanced research of plant lipids. Netherlands: Kluwer Publishers; 2003. p. 24.

Anwar, F., Latif, S., Ashraf, M. and Gilani, A.H. (2006). *Moringa oleifera*: a food plant with multiple medicinal uses. *Phytotherapy Research*, [online] 21(1), pp.17–25. doi:https://doi.org/10.1002/ptr.2023.

AOCS (2024). *Analysis of Tocopherols and Tocotrienols by HPLC – AOCS*. [online] AOCS. Available at: https://www.aocs.org/resource/analysis-of-tocopherols-and-tocotrienols-by-hplc/ [Accessed 15 Feb. 2025].

Atteya, A.K.G., Albalawi, A.N., El-Serafy, R.S., Albalawi, K.N., Bayomy, H.M, Genaidy, E.A.E.(2021) Response of *Moringa oleifera* Seeds and Fixed Oil Production to Vermicompost and NPK Fertilizers under Calcareous Soil Conditions. Plants (Basel). 2021

Sep 24;10(10):1998. doi: 10.3390/plants10101998. PMID: 34685807; PMCID: PMC8538915.

Azad, A. K., et al. (2015). *Moringa oleifera*: A review of the medicinal evidence for its nutritional, therapeutic, and prophylactic properties. Part 2: Leaves, seeds, and oils. *Phytotherapy Research*, 29(9), 1102–1114.

Azam S, Nouman W, Ubaid-ur Rehman U, Ahmed T, Gull M. Shaheen (2019). Adaptability of *Moringa oleifera Lam*. Under different water holding capacities. *South African Journal of Botany* 2019;11:123.

Azevedo, Vetriventhan, M., Ramachandran, S., G, R.V., Singh, P., Kumar, V., Babu, R., Naresh, D., Pawar, G., R, M.S. and D, U.H. (2019). Wide variability in the ICRISAT germplasm collections as a source for genetic enhancement of crop cultivars - OAR@ICRISAT. *Icrisat.org*. [online] doi:http://oar.icrisat.org/11422/1/PBCWR-First-2019-AbstractBook.pdf.

Azlan, U. K., Khairul Annuar, N. A., Mediani, A., Aizat, W. M., Damanhuri, H. A., Tong, X., Yanagisawa, D., Tooyama, I., Wan Ngah, W. Z., Jantan, I., & Hamezah, H. S. (2023). An insight into the neuroprotective and anti-neuroinflammatory effects and mechanisms of *Moringa oleifera*. Frontiers in pharmacology, 13, 1035220. https://doi.org/10.3389/fphar.2022.1035220

Azza, S.M. (2014). Morpho-anatomical variations of leaves and seeds among three Moringa species. Life Science Journal 11: 827-832.

Ayerza, R.(2019).Seed characteristics, oil content and fatty acid composition of moringa *(Moringa oleifera* Lam.) seeds from three arid land locations in Ecuador, *Industrial Crops and Products*. (140), 111575, ISSN 0926-6690, https://doi.org/10.1016/j.indcrop.2019.111575.https://www.sciencedirect.com/science/article/pii/S0926669019305862

Bach, T.J., (1986). Hydroxymethylglutaryl-CoA reductase, a key enzyme in phytosterol synthesis?. *Lipids*, *21*(1), pp.82-88.

Barratt, L.J., Reynolds, I.K., Franco Ortega, S. and Harper, A.L. (2023). Transcriptomic and coexpression network analyses on diverse wheat landraces identifies candidate master regulators of the response to early drought. *Frontiers in Plant Science*, 14. doi:https://doi.org/10.3389/fpls.2023.1212559.

Basuny, A.M. (2016). Biochemical Studies on *Moringa oleifera* Seed Oil. *MOJ Food Processing & Technology*, 2(2). doi:https://doi.org/10.15406/mojfpt.2016.02.00030.

Baud, S. and Lepiniec, L. (2009). Regulation of de novo fatty acid synthesis in maturing oilseeds of Arabidopsis, *Plant Physiology and Biochemistry*, Volume 47, Issue 6,2009,Pages 448-455, ISSN 0981-9428, https://doi.org/10.1016/j.plaphy.2008.12.006. (https://www.sciencedirect.com/science/article/pii/S0981942808002428)

Behera S, Voshall A, Moriyama EN. Plant Transcriptome Assembly: Review and Benchmarking. In: Helder I. N, editor. Bioinformatics [Internet]. Brisbane (AU): Exon Publications; 2021 Mar 20. Chapter 7. PMID: 33877767.

Beiki, H., Liu, H., Huang, J., Manchanda, N., Nonneman, D., Smith, T.P.L., Reecy, J.M. and Tuggle, C.K. (2019). Improved annotation of the domestic pig genome through

integration of Iso-Seq and RNA-seq data. *BMC Genomics*, 20(1). doi:https://doi.org/10.1186/s12864-0195709-y.

Benjamin, M.A., Nichols, M., Burke, T.W., Ginsburg, G.S., Lucas, J. (2014). Comparing reference-based RNA-Seq mapping methods for non-human primate data. BMC genomics. 15. 570. 10.1186/1471-2164-15-570.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300. http://www.jstor.org/stable/2346101

Bergmüller, E., Porfirova, S. and Dörmann, P. (2003). Characterization of an*Arabidopsis*mutant deficient in γ-tocopherol methyltransferase. *Plant Molecular Biology*, 52(6), pp.1181–1190. doi:https://doi.org/10.1023/b:plan.0000004307.62398.91.

Beyene, S.D. (2023). The impact of food insecurity on health outcomes: Empirical evidence from sub-Saharan African countries. *BMC Public Health*, [online] 23(1). doi:https://doi.org/10.1186/s12889-023-15244-3.

Bhattacharya, A. and Mandal, S. (2004) 'Pollination, pollen germination and stigma receptivity in *Moringa oleifera* Lamk', *Grana*, 43(1), pp. 48–56. doi: 10.1080/00173134.2004.11877463.

Bi H, Liu Z, Liu S, Qiao W, Zhang K, Zhao M, Wang D. Genome-wide analysis of wheat xyloglucan endotransglucosylase/hydrolase (XTH) gene family revealed TaXTH17 involved in abiotic stress responses. BMC Plant Biol. 2024 Jul 6;24(1):640. doi: 10.1186/s12870-024-05370-4. PMID: 38971763; PMCID: PMC11227136.

Blighe, K, S Rana, and M Lewis. (2018). "EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling." https://github.com/kevinblighe/EnhancedVolcano.

Bolger, A.M., Lohse, M. and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), pp.2114–2120. doi:https://doi.org/10.1093/bioinformatics/btu170.

Boopathi, N.M., Abubakar, B.Y. (2021). Botanical Descriptions of *Moringa spp.*., In: Boopathi, N.M., Raveendran, M., Kole, C. (eds) The Moringa Genome. Compendium of Plant Genomes. Springer, Cham. https://doi.org/10.1007/978-3-030-80956-0_2

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*, [online] 23(19), pp.2633–2635. doi:https://doi.org/10.1093/bioinformatics/btm308.

Brendel V, Kurtz S, Walbot V. (2002) Comparative genomics of Arabidopsis and maize: prospects and limitations. *Genome Biol*. 2002;3(3): REVIEWS1005. doi: 10.1186/gb-2002-3-3-reviews1005. Epub 2002 Feb 14. PMID: 11897028; PMCID: PMC139024.

Broughton R, and Beaudoin F. (2021). Analysis of Free and Esterified Sterol Content and Composition in Seeds Using GC and ESI-MS/MS. Methods Mol Biol. 2021;2295:179-201. doi: 10.1007/978-1-0716-1362-7_11. PMID: 34047978.

Broughton, R., Ruíz-Lopez, N., Hassall, K.L., Martínez-Force, E., Garcés, R., Salas, J.J. and Beaudoin, F. (2018). New insights in the composition of wax and sterol esters in common and mutant sunflower oils revealed by ESI-MS/MS. *Food Chemistry*, 269, pp.70–79. doi:https://doi.org/10.1016/j.foodchem.2018.06.135.

Broccoli - Wikiwand (nd). https://www.wikiwand.com/en/articles/Broccoli (Accessed: 31 December 2025).

Buchfink, B., Xie, C. and Huson, D.H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12(1), pp.59–60. doi:https://doi.org/10.1038/nmeth.3176.

Buet A, Costa ML, Martínez DE, Guiamet JJ. Chloroplast Protein Degradation in Senescing Leaves: Proteases and Lytic Compartments. *Front Plant Sci*. 2019 Jun 19;10:747. doi: 10.3389/fpls.2019.00747. PMID: 31275332; PMCID: PMC6593067.

Bushmanova, E., Antipov, D., Lapidus, A., Suvorov, V. and Prjibelski, A.D. (2016). rnaQUAST: a quality assessment tool for*de novo*transcriptome assemblies: Table 1. *Bioinformatics*, 32 (14), pp. 2210 –2212. doi:https://doi.org/10.1093/bioinformatics/btw218.

Cai, Z., Xian, P., Cheng, Y., Zhong, Y., Yang, Y., Zhou, Q., Lian, T., Ma, Q., Nian, H. and Ge, L. (2023). MOTHER-OF-FT-AND-TFL1 regulates the seed oil and protein content in soybean. *New Phytologist*, 239(3), pp.905–919. doi:https://doi.org/10.1111/nph.18792.

Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P. and Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, 38(12), pp.5825–5829. doi:https://doi.org/10.1093/molbev/msab293.

Cassan, O., Lèbre, S. and Martin, A. (2021). Inferring and analyzing gene regulatory networks from multi-factorial expression data: a complete and interactive suite. *BMC Genomics*, 22(1). doi:https://doi.org/10.1186/s12864-021-07659-2.

Castro, C.J. and Fei, T. (2017). U_{50}: A New Metric for Measuring Assembly Output Based on Non-Overlapping, Target-Specific Contigs. *Journal of Computational Biology*, 24(11), pp.1071– 1080. doi:https://doi.org/10.1089/cmb.2017.0013.

Cater N.B and Denke M.A. (2001). Behenic acid is a cholesterol-raising saturated fatty acid in humans. *Am J Clin Nutr.* 2001 Jan;73(1):41-4. doi: 10.1093/ajcn/73.1.41. PMID: 11124748.

Cattivelli, L., Rizza, F., Badeck, Franz-W., Mazzucotelli, E., Mastrangelo, A.M., Francia, E., Marè, C., Tondelli, A. and Stanca, A.M. (2008). Drought tolerance improvement in crop plants: An integrated view from breeding to genomics. *Field Crops Research*, 105(1-2), pp.1–14. doi:https://doi.org/10.1016/j.fcr.2007.07.004.

Cervera-Chiner, L., Pageo, S., Juan-Borrás, M., García-Mares, F.J., Castelló, M.L. and Ortolá,M.D. (2024). Fatty Acid Profile and Physicochemical Properties of *Moringa oleifera* Seed Oil Extracted at Different Temperatures. *Foods*, [online] 13(17), p.2733. doi:https://doi.org/10.3390/foods13172733.

Chakraborty P, Biswas A, Dey S, Bhattacharjee T, Chakrabarty S. Cytochrome P450 Gene Families: Role in Plant Secondary Metabolites Production and Plant Defense. J Xenobiot. 2023 Jul 25;13(3):402-423. doi: 10.3390/jox13030026. PMID: 37606423; PMCID: PMC10443375.

Chang, J., Marczuk-Rojas, J.P., Waterman, C., Garcia-Llanos, A., Chen, S., Ma, X., Hulse-Kemp, A., Van Deynze, A., Van de Peer, Y. and Carretero-Paulet, L. (2022). Chromosome-scale assembly of the *Moringa oleifera* Lam. genome uncovers polyploid history and evolution of secondary metabolism pathways through tandem duplication. *The Plant Genome*, 15(3). doi:https://doi.org/10.1002/tpg2.20238.

Chaudhary, K., and Amin, D. (2022). Polygalacturonase. In: Amin, D., Amaresan, N., Ray, S. (eds) Biotic Elicitors. Springer Protocols Handbooks. Humana, New York, NY. https://doi.org/10.1007/978-1-0716-2601-6_20

Chai, L.X., Dong, K., Liu, S.Y., Zhang, Z., Zhang, X.P., Tong, X., Zhu, F.F., Zou, J.Z. and Wang, X.B., 2020. A putative nuclear copper chaperone promotes plant immunity in Arabidopsis. Journal of Experimental Botany, 71(20), pp.6684-6696.

Chaves M.M, Maroco J.P, Pereira J.S.(2003) Understanding plant responses to drought - from genes to the whole plant. Funct Plant Biol. 2003 Mar;30(3):239-264. doi: 10.1071/FP02076. PMID: 32689007.

Chen, Z., Lancon-Verdier, V., Le Signor, C. *et al. (2021)* Genome-wide association study identified candidate genes for seed size and seed composition improvement in *M. truncatula*. *Sci Rep* 11, 4224 (2021). https://doi.org/10.1038/s41598-021-83581-7

Chitiyo, S.T. Ncube, B., Ndhala, R.A, Tsvuura, Z. (2021). Biochemical responses of *Moringa oleifera* Lam. Plants to graded moisture deficit. *South African Journal of Botany*. 2021 138:l41-49

Chin, C.H, Chen, S.H, Wu, H.H, Ho, C.W, Ko, M.T, Lin, C.Y(2014). cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol. 2014;8 Suppl 4(Suppl 4):S11. doi: 10.1186/1752-0509-8-S4-S11. Epub 2014 Dec 8. PMID: 25521941; PMCID: PMC4290687.https://pubmed.ncbi.nlm.nih.gov/25521941/

Chukwuebuka, E. (2015). *Moringa oleifera* 'The Mother's Best Friend'. *International Journal of Nutrition and Food Sciences*, [online] 4(6), p.624. doi:https://doi.org/10.11648/j.ijnfs.20150406.14.

Chung, M., Bruno, V.M., Rasko, D.A., Cuomo, C.A., Muñoz, J.F., Livny, J., Shetty, A.C., Mahurkar, A. and Dunning Hotopp, J.C. (2021). Best practices on the differential expression analysis of multi-species RNA-seq. *Genome Biology*, 22(1). doi:https://doi.org/10.1186/s13059021-02337-8.

Cirlini, M., Lorenzo, D.V., Leto,L., Russo, F., Dellafiora, L., Guarrasi, V., Chiancone, B. (2022). Sprouts of *Moringa oleifera* Lam.: Germination, Polyphenol Content and Antioxidant Activity. Molecules. 27. 8774. 10.3390/molecules27248774.

Cock, P.J.A., Fields, C.J., Goto, N., Heuer, M.L. and Rice, P.M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, [online] 38(6), pp.1767–1771. doi:https://doi.org/10.1093/nar/gkp1137.

Cui,M., Zhang,W., Zhang,Q., Xu,Z., Zhu,Z., Duan, F., Wu,R. (2011). Induced over-expression of the transcription factor OsDREB2A improves drought tolerance in rice, *Plant Physiology and Biochemistry*, Volume 49, Issue 12,Pages 1384-1391, ISSN 0981-9428, https://doi.org/10.1016/j.plaphy.2011.09.012.https://www.sciencedirect.com/science/article/pii/S0981942811002804

da Veiga Leprevost, F., Grüning, B.A., Alves Aflitos, S., Röst, H.L., Uszkoreit, J., Barsnes, H., Vaudel, M., Moreno, P., Gatto, L., Weber, J., Bai, M., Jimenez, R.C., Sachsenberg, T., Pfeuffer, J., Vera Alvarez, R., Griss, J., Nesvizhskii, A.I. and Perez-Riverol, Y. (2017). BioContainers: an open-source and community-driven framework for software standardization. *Bioinformatics*, 33(16), pp.2580–2582. doi:https://doi.org/10.1093/bioinformatics/btx192.

Dadamouny, M.A., Unterseher, M., Konig, P., and Schnittler, M., (2016). Population performance of *Moringa peregrina* (Forssk)Fiori (Moringaceae) at Sinai Pennisula, Egypt in the last decades: Consequences for its conservation. Journal for Nature Conservation 34 DOI:10.1016/j.jnc.2016.08.005

Dai X, and Shen L. Advances and Trends in Omics Technology Development. Front Med (Lausanne). 2022 Jul 1;9:911861. doi: 10.3389/fmed.2022.911861. PMID: 35860739; PMCID: PMC9289742.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R; (2001). 1000 Genomes Project Analysis Group. The variant call format and VCFtools. Bioinformatics. 2011 Aug 1;27(15):2156-8. doi: 10.1093/bioinformatics/btr330. Epub 2011 Jun 7. PMID: 21653522; PMCID: PMC3137218.

Daniel, B., Wallner,S., Steiner, B., Oberdorfer, G., Kumar, P, Eric van der Graff, Roitsch, T, Sensen, C.W, Gruber, K., Macheroux, P. (2016) Structure of a Berberine Bridge Enzyme-Like Enzyme with an Active site specific to the plant family Brassicaceae. *PLOSOne Journal*. https://doi.org/10.1371/journal.pone.0156892

Daniel, B., Pavkov-Keller, T., Steiner, B., Dordic, A., Gutmann, A., Nidetzzky, B., Sensen, C.W., Eric van der Graff, Wallner, S., Gruber, K, Macheroux, P.(2015). Oxidation of Mongolignols by members of the Berberine Bridge Enzyme family suggests a role in plant cell wall metabolism. Journal of Biological Chemistry 290(30) 24 July 2015, Pages 18770-18781

Davis-Turak, J., Courtney, S.M., Hazard, E.S., Glen, W.B., da Silveira, W.A., Wesselman, T., Harbin, L.P., Wolf, B.J., Chung, D. and Hardiman, G. (2017). Genomics pipelines and data integration: challenges and opportunities in the research setting. *Expert Review of Molecular Diagnostics*, 17(3), pp.225–237. doi:https://doi.org/10.1080/14737159.2017.1282822.

Delphin, S, Versanne, R., Urda, C., Galben, R.D., and Ona, A (2023). Haitian Moringa, the Plant that Nourishes, Healsand Enriches. Hop and Medicinal Plants, Year XXX, No. 1-2, 2022 ISSN 2360 – 0179 print, ISSN 2360 – 0187 electronic

Deng L-T, Wu Y-L, Li J-C, OuYang K-X, Ding M-M, Zhang J-J, et al. (2016) Screening Reliable Reference Genes for RT-qPCR Analysis of Gene Expression in *Moringa oleifera*. PLoS ONE 11(8): e0159458. https://doi.org/10.1371/journal.pone.0159458

Devkota, S., & Bhusal, K. K. (2020). *Moringa oleifera*: A miracle multipurpose tree for agroforestry and climate change mitigation from the Himalayas – A review. *Cogent Food & Agriculture*, 6, 1805951. https://doi.org/10.1080/23311932.2020.1805951

Dhakar, R.C., Maurya, S.D., Pooniya, B.K., Bairwa, N., Gupta, M. and Sanwarmal (2011). *Moringa: The Herbal Gold To Combat Malnutrition*. [online] papers.ssrn.com. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3774645.

Di Tommaso, P., Chatzou, M., Floden, E.W., Barja, P.P., Palumbo, E. and Notredame, C. (2017). Nextflow enables reproducible computational workflows. *Nature Biotechnology*, 35(4), pp.316– 319. doi:https://doi.org/10.1038/nbt.3820.

Dida, F. and Yi, G. (2021). Empirical evaluation of methods for *de novo* genome assembly. *PeerJ Computer Science*, 7, p.e636. doi:https://doi.org/10.7717/peerj-cs.636.

Dinesha, B L, Nidoni, U, Ramachandra, C.T., Naik, N. K. and Sankalpa. (2018). Effect of extraction methods on physicochemical, nutritional, antinutritional, antioxidant and antimicrobial activity of Moringa (*Moringa oleifera* Lam.) seed kernel oil. *Journal of Applied and Natural Science*. 10. 10.31018/jans.v10i1.1619.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2012). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), pp.15–21. doi:https://doi.org/10.1093/bioinformatics/bts635.

Dobin, A. & Gingeras, T. (2016). Optimizing RNA-seq mapping with STAR. Methods in molecular biology (Clifton, N.J.) 1415:245-262 DOI:10.1007/978-1-4939-3572-7_13 In book: Data Mining Techniques for the Life Sciences (pp.245-262)

Dobin, A. and Gingeras, T.R. (2015). Mapping RNA-seq Reads with STAR. *Current Protocols in Bioinformatics*, [online] pp.11.14.1–11.14.19. doi:https://doi.org/10.1002/0471250953.bi1114s51.

Du, Z., Zhou, X., Ling, Y., Zhang, Z. and Su, Z. (2010). agriGO: a GO Analysis Toolkit for the Agricultural Community. *Nucleic Acids Research*, [online] 38(suppl_2), pp. W64–W70. doi:https://doi.org/10.1093/nar/gkq310.

Ejigu, G.F. and Jung, J. (2020). Review on the Computational Genome Annotation of Sequences Obtained by Next-Generation Sequencing. *Biology*, 9(9),p.295. doi:https://doi.org/10.3390/biology9090295.

Emms, D.M. and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, 16(1). doi:https://doi.org/10.1186/s13059-015-0721-2.

Emms, D.M. and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, 20(1). doi:https://doi.org/10.1186/s13059-019-1832-y.

Emms, D.M. and Kelly, S.L. (2017). STRIDE: Species Tree Root Inference from Gene Duplication Events. *Molecular Biology and Evolution*, 34(12), pp.3267–3278. doi:https://doi.org/10.1093/molbev/msx259.

Emongor, V.E. (2011). MORINGA (*Moringa oleifera*Lam.): A REVIEW. *Acta*

Ewels, P.A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M.U., Di Tommaso, P. and Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology*, 38(3), pp.276–278. doi:https://doi.org/10.1038/s41587-020-0439-x.

Faizan, M., Alam, P., Rajput, V. D., Shareen, Kaur, K., Faraz, A., Minkina, T., Maqbool Ahmed, S., Rajpal, V. R., & Hayat, S. (2023). Potential role of tocopherol in protecting crop plants against abiotic stresses. *Physiology and molecular biology of plants : an*

international journal of functional plant biology, *29*(10), 1563–1575. https://doi.org/10.1007/s12298-023-01354-0

Falk, Jon & Munné-Bosch, Sergi. (2010). Tocochromanol functions in plants: Antioxidation and beyond. Journal of experimental botany. 61. 1549-66. 10.1093/jxb/erq030

Fang, H., Bergmann, E.A., Arora, K., Vacic, V., Zody, M.C., Iossifov, I., O'rawe, J.A., Wu, Y., Jimenez Barron, L.T., Rosenbaum, J. and Ronemus, M., (2016). Indel variant analysis of short-read sequencing data with Scalpel. Nature protocols, 11(12), pp.2529-2548.

Fao, I. (2023). *The State of Food Security and Nutrition in the World 2023*. [online] *openknowledge.fao.org*. FAO ; IFAD ; UNICEF ; WFP ; WHO ; Available at: https://openknowledge.fao.org/handle/20.500.14283/cc3017en.

Farber, C.R., Atila van Nas, Anatole Ghazalpour, Aten, J.E., Doss, S., Sos, B.C., Schadt, E.E., Ingram-Drake, L., Davis, R.A., Horvath, S., Smith, D.J., Drake, T.A. and Lusis, A.J. (2009). An Integrative Genetics Approach to Identify Candidate Genes Regulating BMD: Combining Linkage, Gene Expression, and Association. 24(1), pp.105–116. doi:https://doi.org/10.1359/jbmr.080908.

Feng, X., Jia, L., Cai, Y., Guan, H., Zheng, D., Zhang, W., Xiong, H., Zhou, H., Wen, Y., Hu, Y., Zhang, X., Wang, Q., Wu, F., Xu, J., & Lu, Y. (2022). ABA-inducible DEEPER ROOTING 1 improves adaptation of maize to water deficiency. *Plant biotechnology journal*, 20(11), 2077–2088. https://doi.org/10.1111/pbi.13889

Ferreira, P.M.P., Farias, D.F., Oliveira, J.T. de A. and Carvalho, A. de F.U. (2008). *Moringa oleifera*: bioactive compounds and nutritional potential. *Revista de Nutrição*, [online] 21(4), pp.431–437. doi: https://doi.org/10.1590/s1415-52732008000400007.

Ferfuia, C, Turi, M and Vannozzi, G.P (2015). "Variability of Seed Fatty Acid Composition to Growing Degree-Days in High Oleic Acid Sunflower Genotypes" *Helia*, vol. 38, no. 62, 2015, pp. 61-78. https://doi.org/10.1515/helia-2014-0022

Foidl, N. & Makkar, Harinder & Becker, K.. (2001). The potential of *Moringa oleifera* for agricultural and industrial uses *Moringa oleifera* Webpage.. The Potential of *Moringa oleifera* for Agricultural and Industrial Uses, The Miracle Tree: The Multi Uses of Moringa. 45-76.

Fritsche S, Wang X, Jung C. (2017). Recent Advances in our Understanding of Tocopherol Biosynthesis in Plants: An Overview of Key Genes, Functions, and Breeding of Vitamin E Improved Crops. Antioxidants (Basel). 2017 Dec 1;6(4):99. doi: 10.3390/antiox6040099. PMID: 29194404; PMCID: PMC5745509.

Fuller, T.F., Anatole Ghazalpour, Aten, J.E., Drake, T.A., Lusis, A.J. and Horvath, S. (2007). Weighted gene co-expression network analysis strategies applied to mouse weight. *Mammalian Genome*, 18(6-7), pp.463–472. doi:https://doi.org/10.1007/s00335-007-9043-3.

Gandji, K., Chadare, F.J., Idohou, R., Salako, V.K., Assogbadjo, A.E. and Kakaï, R.L.G. (2018). Status and utilisation of *Moringa oleifera* Lam: A review. *African Crop Science Journal*, 26(1), p.137. doi:https://doi.org/10.4314/acsj.v26i1.10.

Godino, M., Arias, C. and Izquierdo, M.I., (2015). *Moringa oleifera*: Potential areas of cultivation on the Iberian Peninsula. In I International Symposium on Moringa 1158 (pp. 405-412).

Gómez-Martínez,M., Garcia-Ortiz,J.D., Gómez-Martínez,S., Michel, M.R.,López-Badillo,C.M., Pintado,M., Rodríguez-Herrera,R (2025). Location and tissue effect on physicochemical, mineral and amino acid profile of *Moringa oleifera*.*Journal of Trace Elements and Minerals*, Volume 13, 2025, 100258,ISSN 2773-0506,https://doi.org/10.1016/j.jtemin.2025.100258.https://www.sciencedirect.com/science/article/pii/S2773050625000485

Gordon, L.J., Finlayson, C.M. and Falkenmark, M. (2010). Managing water in agriculture for food production and other ecosystem services. *Agricultural Water Management*, 97(4), pp.512– 519. doi:https://doi.org/10.1016/j.agwat.2009.03.017.

Goudarzi, A (2019). The recent insights into the function of ACAT1: Apossible anti-cancer therapeutic target. *Life Sciences*. Vol.232 , 116592 https://doi.org/10.1016/j.lfs.2019.116592

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K. and Friedman, N. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), pp.644–652. doi:https://doi.org/10.1038/nbt.1883.

Green, M. and Sambrook, J. (2012). *This is a free sample of content from Molecular Cloning: A Laboratory Manual, 4th edition. Click here for more information or to buy the book. Molecular Cloning*. [online] Available at: https://www.cshlpress.com/pdf/sample/2013/MC4/MC4FM.pdf.

Gui, S., Wei, W., Jiang, C., Luo, J., Chen, L., Wu, S., Li, W., Wang, Y., Li, S., Yang, N., Li, Q., Fernie, A.R. and Yan, J. (2022). A pan-Zea genome map for enhancing maize improvement. *Genome biology*, 23(1). doi:https://doi.org/10.1186/s13059-022-02742-7.

Guo Y et al. (2015) Gene expression profiling of drought tolerant and sensitive *Brassica rapa* accessions under control and osmotic stress conditions. NCBI Gene Expression Omnibus (GEO) accession number GSE73963 http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE73963

Guo Y.M, Samans B, Chen S, Kibret K.B et al.(2015) Drought-Tolerant Brassica rapa Shows Rapid Expression of Gene Networks for General Stress Responses and Programmed Cell Death Under Simulated Drought Stress. *Plant Mol Biol Report* 2017;35(4):416-430. PMID: 28751801

Gupta, P.K, Baylan, H.S, and Gahlaut, V (2017). QTL Analysis for drought Tolerannce in Wheat: Present status and future possibilities. *Agronomy* 2017, *7*(1),5; https://doi.org/10.3390/agronomy7010005

Gupta, O.P. , Pandey, V., Saini, R., Khandale, A.S., Malik, V.K., Narwal, S., Ram,S., Gyanendra, P.S., (2021). Comparative physiological, biochemical and transcriptomic analysis of hexaploidy wheat (*T. aestivum* L.) roots and shoots identifies potential pathways and their molecular regulatory network during Fe and Zn starvation. *Genomics* 113:5 pp3357-3372 Elsievier. https://doi.org/10.1016/j.ygeno.2021.07.029

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., … Regev, A. (2013). de novo transcript sequence reconstruction from RNA-seq using the

Trinity platform for reference generation and analysis. *Nature protocols*, *8*(8), 1494–1512. https://doi.org/10.1038/nprot.2013.084

Habtemariam, S. (2017). The African and Arabian Moringa Species. In: Chemistry, Bioactivity and Therapeutic Applications. *Elsevier Science*, Dordecht, Netherlands.

Hajaji, A.N., Heikal, Y.M., Hamouda, R.A.E.F., Abassi, M. and Ammari, Y. (2024). Multivariate investigation of *Moringa oleifera* morpho-physiological and biochemical traits under various water regimes. BMC Plant Biology, 24(1). doi:https://doi.org/10.1186/s12870-024-05040-5.

Haile, M, Duguma, H.T, Chameno, G. and Kuyu, C.G (2019). Effects of location and extraction solvent on physico chemical properties of Moringa stenopetala seed oil. *Heliyon.* 5. doi:10.1016/j.heliyon.2019.e02781.

Haris, M., Sajid, M., Hassan, M., Zahid, A., Abbas, M.T., Ammara, F., Zulfiqar,B., Ahmad,Z., Maqsood, S., Hasan, S.Z(2024). Evaluating the impact of *Moringa oleifera* Leaf Extract on drought stress mitigation to enhance growth and yield in Mung bean (*Vigna radiata*). 3(2)*J.Agric.Vet.Sci* DOI: https://doi.org/10.55627/agrivet.003.02.0805

Harshad A. Shirke, Ashwini M. Darshetkar, Vikas B. Naikawadi, P.B. Kavi Kishor, Tukaram D. Nikam, Vitthal T. Barvkar(2025). Genomics of sterols biosynthesis in plants: Current status and future prospects, *Plant Science*, Volume 353,2025,112426,ISSN 0168-9452, https://doi.org/10.1016/j.plantsci.2025.112426.
(https://www.sciencedirect.com/science/article/pii/S0168945225000433)

Hartmann, M. A. (1998). *Plant sterols and the membrane environment*. *Trends in Plant Science*, 3(5), 170–175.

Hassanein, A.M.A. and Al-Soqeer, A.A. (2018). Morphological and genetic diversity of *Moringa oleifera* and Moringa peregrina genotypes. *Horticulture, Environment, and Biotechnology*, 59(2), pp.251–261. doi:https://doi.org/10.1007/s13580-018-0024-0.

Hebert, P.D.N., Cywinska, A., Ball, S.L. and deWaard, J.R. (2003). Biological Identifications through DNA Barcodes. *Proceedings*. *Biological Sciences*, 270(1512), pp.313–21.

High Oleic Oils : development, properties and uses. (2021). S.L.: Amer Oil Chemist Society.

Hiroyuki, K. (2018). Status and Outlook of Global Food Security and the Role of Underutilized Food Resources: Sago Palm. *Springer eBooks*, pp.3–16. doi:https://doi.org/10.1007/978-981-10-5269-9_1.

Hodas, F., Zorzenon, M.R.T. and Milani, P.G. (2021). *Moringa oleifera* potential as a functional food and a natural food additive: a biochemical approach. *Anais da Academia Brasileira de Ciências*, [online] 93. doi:https://doi.org/10.1590/0001-3765202120210571.

Hölzer, M. and Marz, M. (2019). *de novo* transcriptome assembly: A comprehensive crossspecies comparison of short-read RNA-Seq assemblers. *GigaScience*, 8(5). doi:https://doi.org/10.1093/gigascience/giz039.

Horvath R, Hudson G, Ferrari G, Fütterer N, Ahola S, Lamantea E, Prokisch H, Lochmüller H, McFarland R, Ramesh V, Klopstock T, Freisinger P, Salvi F, Mayr JA, Santer R, Tesarova M, Zeman J, Udd B, Taylor RW, Turnbull D, Hanna M, Fialho D, Suomalainen A,

Zeviani M, Chinnery PF. Phenotypic spectrum associated with mutations of the mitochondrial polymerase gamma gene. Brain. 2006 Jul;129(Pt 7):1674-84. doi: 10.1093/brain/awl088. Epub 2006 Apr 18. PMID: 16621917.

Huang, Y., Hong, Y., & Liang, K. (2024). Impact of tree root structure on soil Water dynamics: drought adaptation mechanisms and ecological significance. Huang | Molecular Soil Biology. https://bioscipublisher.com/index.php/msb/article/html/3942

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L., von Mering, C. and Bork, P. (2018). eggNOG5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, 47(D1), pp. D309–D314. doi:https://doi.org/10.1093/nar/gky1085.

Hussain, Q., Asim, M., Zhang, R., Khan, R., Farooq, S., & Wu, J. (2021). Transcription Factors Interact with ABA through Gene Expression and Signalling Pathways to Mitigate Drought and Salinity Stress. Biomolecules, 11(8), 1159. https://doi.org/10.3390/biom11081159

Idris, A., Nour, A., Ishag, O., Mahmoud ALİ, Ibrahim ERWA and NOUR, A. (2020). Physicochemical Properties and Fatty Acids Composition of Sudanese *Moringa oleifera* Seed Oil. *Journal of the Turkish Chemical Society Section a Chemistry*, 7(3), pp.911–920. doi:https://doi.org/10.18596/jotcsa.771260.

Iqtree.org. (2017). *IQ-TREE version 1.5.3*. [online] Available at: http://www.iqtree.org/release/v1.5.3/ [Accessed 8 Mar. 2025].

Islam, S., Ali, M., Banerjee, A., Shah, M.H., Rahman, S.S., Faijuddin Ahammad and Hossain, A.(2023). *Moringa oleifera* under Stressful Conditions. *CRC Press eBooks*, pp.201–222. doi:https://doi.org/10.1201/9781003242963-11.

Jackson, D.J., Cerveau, N. & Posnien, N. (2024). *de novo* assembly of transcriptomes and differential gene expression analysis using short-read data from emerging model organisms – a brief guide. *Front Zool* 21, 17 (2024). https://doi.org/10.1186/s12983-024-00538-y

James, M., Poret, M., Masclaux-Daubresse, C., Marmagne,A., Coquet,L., Jouenne, T., Chan, P., Trouverie, J., Etienne, P. (2018). SAG12, a Major Cysteine Protease Involved in Nitrogen Allocation during Senescence for Seed Production in *Arabidopsis thaliana*, *Plant and Cell Physiology*, Volume 59, Issue 10, October 2018, Pages 2052–2063, https://doi.org/10.1093/pcp/pcy125

Jarzyniak KM, Jasiński M.(2014). Membrane transporters and drought resistance - a complex issue. Front Plant Sci. 2014 Dec 4;5:687. doi: 10.3389/fpls.2014.00687. PMID: 25538721; PMCID: PMC4255493.

Jattan, M., N, K., Kumar, R., A, K., B, R., D.S, P., S., K. and P, K. (2021). *Moringa (Moringa oleifera L.) : an Underutilized and Traditionally Valued Tree Holding Remarkable Potential*. https://jhs.iihr.res.in/index.php/jhs/article/view/1073/887.

Jones LB, Bean R, McLachlan GJ, Zhu JX. (2006) Mixture models for detecting differentially expressed genes in microarrays. Int J Neural Syst. 2006 Oct;16(5):353-62. doi: 10.1142/S0129065706000755. PMID: 17117496.

Jones, N., Ougham, H., Thomas, H. and Pasakinskiene, I. (2009). Markers and mapping revisited: finding your gene. *The New Phytologist*, [online] 183(4), pp.935–966. doi:https://doi.org/10.1111/j.1469-8137.2009.02933.x.

Joshi, M.R. and Pandit, S., 2023. *Moringa oleifera*: A boon to agriculture. Journal of the Institute of Agriculture and Animal Science, pp.113-120.

Julia, P., Coppin, H., et al. (2015) Chapter 17. Variations in Polyphenols and Lipid Soluble Vitamins in *Moringa oleifera*. In: Preedy, V., Ed., Processing and Impact on Active Components in Food, Elsevier, Amsterdam, 655-663.
https://doi.org/10.1016/B978-0-12-404699-3.00079-2

Jyothi, P.V., Atluri, J.B. & Reddi, C.S. (1990). Pollination ecology of *Moringa oleifera* (Moringaceae). *Proc. Indian Acad. Sci.* 100, 33–42.). New Delhi: Springer India. https://doi.org/10.1007/BF03053466

Kaidi Yu,H-L., Zhang,Z., Yu,Y., Wan, J., He,H., Fan, C.(2023). Targeted mutagenesis of flavonoid biosynthesis pathway genes reveals functional divergence in seed coat colour, oil content and fatty acid composition in *Brassica napus* L. Plant Biotechnology Journal Volume 22, Issue 2 pp. 445-459 https://doi.org/10.1111/pbi.14197

Kale - Wikiwand (n.d). Available at: https://www.wikiwand.com/en/articles/Kale (Accessed: 31 December 2025).

Kale, S., Bertrand Matthäus, Fahad Aljuhaimi, Ahmed, M., Mehmet Musa Özcan, Ghafoor, K., Babiker, E.E., Osman, M.A., Gassem, M.A. and Hesham A. S. Alqah (2020). A comparative study of the properties of 10 variety melon seeds and seed oils. *Journal of food processing and preservation*, 44(6). doi:https://doi.org/10.1111/jfpp.14463.

Katoh, K. and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, 30(4), pp.772–780. doi:https://doi.org/10.1093/molbev/mst010.

Kavitha, M.P., Dr Balakumbahan,R., Maheswari, U.M.,Sidhdharth, G, Nageswari, K, Sudhakar, G, Chidambaram, P., Mohan, M.M, Desigan, J. (2024). Strategic agronomic interventions for leaf yield and quality improvement in moringa (*Moringa oleifera* L.) under high density planting system. Plant Science Today. 11. 10.14719/pst.5487.

Kessler, J.C., Manrique,Y.A., Martins,I.M., Ribeiro,A., Rodrigues,A.E., Barreiro,M.F., Dias, M.M. (2025). Influence of environmental factors on α-tocopherol levels in *Moringa oleifera* extracts. *Industrial Crops and Products*,Volume 228,2025,120906,ISSN 0926-6690, https://doi.org/10.1016/j.indcrop.2025.120906.
(https://www.sciencedirect.com/science/article/pii/S0926669025004522)

Khan S, Ibrar D, Hasnain Z, Nawaz M, Rais A, Ullah S, Gul S, Siddiqui MH, Irshad S. Moringa Leaf Extract Mitigates the Adverse Impacts of Drought and Improves the Yield and Grain Quality of Rice through Enhanced Physiological, Biochemical, and Antioxidant Activities. Plants (Basel). 2023 Jun 30;12(13):2511. doi: 10.3390/plants12132511. PMID: 37447069; PMCID: PMC10347012.

Kim, Y., Chung, Y. S., Lee, E., Tripathi, P., Heo, S., & Kim, K. H. (2020). Root Response to Drought Stress in Rice (Oryza sativa L.). *International journal of molecular sciences*, 21(4), 1513.

Kim, J.-S., Satoshi Kidokoro, Yamaguchi-Shinozaki, K. and Shinozaki, K. (2024). Regulatory networks in plant responses to drought and cold stress. *Plant physiology*, 195(1). doi:https://doi.org/10.1093/plphys/kiae105.

Knez, M., Mattas, K., Gurinovic, M., Gkotzamani, A. and Koukounaras, A. (2024). Revealing the power of green leafy vegetables: Cultivating diversity for health, environmental benefits, and sustainability. *Global Food Security*, [online] 43, p.100816. doi:https://doi.org/10.1016/j.gfs.2024.100816.

Koboldt, D.C. (2020). Best practices for variant calling in clinical sequencing. *Genome Medicine*, 12(1). doi:https://doi.org/10.1186/s13073-020-00791-w.

Kumar, S., Bhattacharya, A., Tiwari, P. and Sahu, P. (2018). A review of the phytochemical and pharmacological characteristics of *Moringa oleifera*. *Journal of Pharmacy And Bioallied Sciences*, 10(4), p.181. doi:https://doi.org/10.4103/jpbs.jpbs_126_18.

Kumar S., Kumar H., Gupta V., Kumar A., Singh C. M., Kumar, M., Singh A. K., Panwar G. S., Kumar, S, Singh A. K., Kumar R. (2023) Capturing agro-morphological variability for tolerance to terminal heat and combined heat–drought stress in landraces and elite cultivar collection of wheat. *Frontiers in Plant Science.* Volume 14 - 2023. https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2023.1136455 DOI:10.3389/fpls.2023.1136455

Kumar, S., Trivedi, P.K. (2018). Glutathione S-Transferases: Role in Combating Abiotic Stresses Including Arsenic Detoxification in Plants. Front Plant Sci. 2018 Jun 7;9:751. doi: 10.3389/fpls.2018.00751. PMID: 29930563; PMCID: PMC5999759.

Kumar, Y., Thakur, T. K., Sahu, M. L., & Thakur, A. (2017). A multifunctional wonder tree: *Moringa oleifera* Lam open new dimensions in field of agroforestry in India. *International Journal of Current Microbiology and Applied Sciences*, 6, 229–235.

Kumssa, D.B., Joy, E.D.M., Young, S.D., Odee, D.W., Ander, E, L., Magare, C., Gitu, J. and Broadley, M.R (2017) Challenges and Opportunities for Moringa growers in Southern Ethiopia and Kenya. *PLOSOne* https://doi.org/10.1371/journal.pone.0187651

LabRulez LCMS. (2024). *RNA Quality Assessment with the Agilent Automated Electrophoresis Systems*. [online] Available at: https://lcms.labrulez.com/paper/27177 [Accessed 26 Feb. 2025].

Lakshmidevamma, T.N., Jayashree Ugalat, Apoorva, K.A., Suresh, G., Mrityunjay Doddamani, Kadam, S.S., R.S Nayana, R.C. Jagadeesha and B. Fakrudin (2021). Genetic Diversity of Moringa (*Moringa oleifera* Lam.).In: Boopathi,N.M., Raveendram,M., Kole, C.(eds)The Moringa Genome. *Compendium of Plant Genomes the Moringa Genome*. doi:https://doi.org/10.1007/978-3-030-80956-0_7.

Lalas, S. and Tsaknis, J. (2002). Characterization of *Moringa oleifera* Seed Oil Variety 'Periyakulam 1'. *Journal of Food Composition and Analysis*, 15(1), pp.65–77. doi:https://doi.org/10.1006/jfca.2001.1042.

Lampi, A-M (2019). Analysis of tocopherols and tocotrienols by HPLC. *AOCS* Department of Food and Environmental Sciences, PO Box 27, Latokartanonkaari 11, FI-00014

Langfelder, P. and Horvath, S. (2008). WGCNA: an R Package for Weighted Correlation Network Analysis. *BMC Bioinformatics*, [online] 9(1). doi:https://doi.org/10.1186/1471-2105-9559.

Leone, A., Spada, A., Battezzati, A., Schiraldi, A., Aristil, J. and Bertoli, S. (2016). *Moringa oleifera* Seeds and Oil: Characteristics and Uses for Human Health. *International Journal of Molecular Sciences*, 17(12), p.2141. doi:https://doi.org/10.3390/ijms17122141. PMID: 27999405; PMCID: PMC5187941

Letsara, R., Razafindrahaja, V. & Razanatsoa, J. (2019). *Moringa hildebrandtii*. The IUCN Red List of Threatened Species 2019: e.T65523618A66155121. https://dx.doi.org/10.2305/IUCN.UK.2019-3.RLTS.T65523618A66155121.en. Accessed on 11 September 2025.

Li J, Han G, Sun C, Sui N. (2019). Research advances of MYB transcription factors in plant stress resistance and breeding. Plant Signal Behav. 2019;14(8):1613131. doi: 10.1080/15592324.2019.1613131. Epub 2019 May 14. PMID: 31084451; PMCID: PMC6619938.

Li, J., Zhou, D., Qiu, W., Shi, Y., Yang, J.-J., Chen, S., Wang, Q. and Pan, H. (2018). Application of Weighted Gene Co-expression Network Analysis for Data from Paired Design. *Scientific Reports*, [online] 8(1). doi:https://doi.org/10.1038/s41598-017-18705-z.

Li, W. and Godzik, A. (2006). CD-Hit: a Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences. *Bioinformatics*, 22(13), pp.1658–1659. doi:https://doi.org/10.1093/bioinformatics/btl158.

Lin, W., Dai, S., Chen, Y., Zhou, Y. and Liu, X. (2019). The Complete Chloroplast Genome Sequence of *Moringa oleifera* Lam. (Moringaceae). *Mitochondrial DNA Part B*, 4(2), pp.4094– 4095. doi:https://doi.org/10.1080/23802359.2019.1627922.

Lischer, H.E.L. and Shimizu, K.K. (2017). Reference-guided *de novo* assembly approach improves genome reconstruction for related species. *BMC Bioinformatics*, 18(1). doi:https://doi.org/10.1186/s12859-017-1911-6.

Liu, H., Song, S., Liu, M., Mu, Y., Li, Y., Xuan, Y., Niu, L., Zhang, H., & Wang, W. (2023). Transcription Factor ZmNAC20 Improves Drought Resistance by Promoting Stomatal Closure and Activating Expression of Stress-Responsive Genes in Maize. International journal of molecular sciences, 24(5), 4712. https://doi.org/10.3390/ijms24054712

Liu, J., Cai, H.-H., Li, H.-Q., Liu, Z.-Y., Zheng, C., Shi, C. and Niu, Y.-F. (2019). The chloroplast genome of *Moringa oleifera* (Moringaceae). *Mitochondrial DNA Part B*, 4(1), pp.646–647. doi:https://doi.org/10.1080/23802359.2018.1545550.

Liu, S., Wang, X., Wang, H., Xin, H., Yang, X., Yan, J., Li, J., Tran, L. S., Shinozaki, K., Yamaguchi-Shinozaki, K., & Qin, F. (2013). Genome-wide analysis of ZmDREB genes and their association with natural variation in drought tolerance at seedling stage of Zea mays L. PLoS genetics, 9(9), e1003790. https://doi.org/10.1371/journal.pgen.1003790

Liwanag AJ, Ebert B, Verhertbruggen Y, Rennie EA, Rautengarten C, Oikawa A, Andersen MC, Clausen MH, Scheller HV. (2012). Pectin biosynthesis: GALS1 in Arabidopsis thaliana is a β-1,4-galactan β-1,4-galactosyltransferase. *Plant Cell*. 2012 Dec;24(12):5024-36. doi: 10.1105/tpc.112.106625. Epub 2012 Dec 14. PMID: 23243126; PMCID: PMC3556973.

Lloyd, J.P., Bowman, M.J., Azodi, C.B., Sowers, R.P., Moghe, G.D., Childs, K.L. and Shiu, S.H. (2019). Evolutionary characteristics of intergenic transcribed regions indicate rare novel genes and widespread noisy transcription in the Poaceae. *Scientific Reports*, [online] 9(1). doi:https://doi.org/10.1038/s41598-019-47797-y.

Love, M.I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), p.550.

Long, Q., Qiu, S., Man, J., Ren, D., Xu, N., & Luo, R. (2023). OsAAI1 Increases Rice Yield and Drought Tolerance Dependent on ABA-Mediated Regulatory and ROS Scavenging Pathway. Rice (New York, N.Y.), 16(1), 35. https://doi.org/10.1186/s12284-023-00650-3

Lyons, G., Gondwe, C., Banuelos, G., Mendoza, C., Haug, A., Christophersen, O. and Ebert, A.W. (2017). Drumstick tree (*Moringa oleifera*) leaves as a source of dietary selenium, sulphur and pro-vitamin A. Acta Hortic. 1158, 287-292 DOI: 10.17660/ActaHortic.2017.1158.32 https://doi.org/10.17660/ActaHortic.2017.1158.32

Maechler, M. and Rousseeuw, P. (2025). *R-packages - Revision 8496: /trunk/cluster*. [online] Rproject.org. Available at: https://svn.r-project.org/R-packages/trunk/cluster/ [Accessed 20 Feb. 2025].

Mahmood, T., Khalid, S., Abdullah, M., Ahmed, Z., Shah, M. K. N., Ghafoor, A., & Du, X. (2019). Insights into Drought Stress Signalling in Plants and the Molecular Genetic Basis of Cotton Drought Tolerance. *Cells*, *9*(1), 105. https://doi.org/10.3390/cells9010105

Mallenakuppe, R., Homabalegowda, H., Gouri, M.D., Basavaraju, P.S. and Chandrashekharaiah, U.B. (2019). History, Taxonomy and Propagation of *Moringa oleifera*- A Review. *SSR Institute of International Journal of Life Sciences*, 5(3), pp.2322–2327. doi:https://doi.org/10.21276/ssriijls.2019.5.3.7.

Manavalan, N, Boopathi, N.M and Raveendran, M (2021). Medicinal and Therapeutic Properties of Moringa. *Compendium of plant genomes*, pp.31–39. doi:https://doi.org/10.1007/978-3-03080956-0_4.

Mani, S., Jaya, S., and Narayanan, L. (2004). Solvent Extraction of Oil from Moringa (*Moringa oleifera*). 10.13031/2013.16940. DOI:10.13031/2013.16940 Conference: 2004, Ottawa, Canada August 1 - 4, 2004 Published by the American Society of Agricultural and Biological Engineers, St. Joseph, Michigan www.asabe.org

Martin, J.A. and Wang, Z. (2011). Next-generation Transcriptome Assembly. *Nature Reviews Genetics*, [online] 12(10), pp.671–682. doi: https://doi.org/10.1038/nrg3068.

Martin, M. (2011). Cutadapt Removes Adapter Sequences from high-throughput Sequencing Reads. *EMBnet.journal*, [online] 17(1), p.10. doi:https://doi.org/10.14806/ej.17.1.200.

Mashamaite, C.V., Pieterse, P.J., Mothapo, P.N. and Phiri, E.E. (2021). *Moringa oleifera* in South Africa: A review on its production, growing conditions and consumption as a food source. *South African Journal of Science*, 117(3/4). doi:https://doi.org/10.17159/sajs.2021/8689.

Matthäus, B., and Özcan, M. M. (2018). 'Fatty Acid Composition and Tocopherol Contents of some Sesame Seed Oils', *Iranian Journal of Chemistry and Chemical Engineering*, 37(5), pp. 151-155. doi: 10.30492/ijcce.2018.35427

Merewitz, E.B., Gianfagna, T. and Huang, B., (2011). Photosynthesis, water use, and root viability under water stress as affected by expression of SAG12-ipt controlling cytokinin synthesis in Agrostis stolonifera. Journal of experimental botany, 62(1), pp.383-395.

Milla, P.G., Peñalver, R. and Nieto, G. (2021). Health Benefits of Uses and Applications of *Moringa oleifera* in Bakery Products. *Plants*, 10(2), p.318. doi:https://doi.org/10.3390/plants10020318.

Miller, J.R., Koren, S. and Sutton, G. (2010). Assembly Algorithms for next-generation Sequencing Data. *Genomics*, 95(6), pp.315–327. doi:https://doi.org/10.1016/j.ygeno.2010.03.001.

Minerdi, D., Savoi, S., & Sabbatini, P. (2023). Role of Cytochrome P450 Enzyme in Plant Microorganisms' Communication: A Focus on Grapevine. *International Journal of Molecular Sciences*, *24*(5), 4695. https://doi.org/10.3390/ijms24054695

Mohammed, G. M., & Hawar, S. N. (2022). Green Biosynthesis of Silver Nanoparticles from *Moringa oleifera* Leaves and Its Antimicrobial and Cytotoxicity Activities. International Journal of Biomaterials, 2022, 1–10. https://doi.org/10.1155/2022/4136641

Mohan, V. R., MacDonald, M. T., & Abbey, L. (2025). Impact of Water Deficit Stress on *Brassica* Crops: Growth and Yield, Physiological and Biochemical Responses. *Plants*, *14*(13), 1942. https://doi.org/10.3390/plants14131942

Moore, G., Devos, K.M., Wang, Z. and Gale, M.D. (1995). Cereal Genome Evolution: Grasses, line up and form a circle. *Cereal Genome Evolution: Grasses, Line up and Form a Circle*, 5(7), pp.737–739. doi:https://doi.org/10.1016/s0960-9822(95)00148-5.

Moreton J, Izquierdo A, Emes RD. Assembly, Assessment, and Availability of *de novo* Generated Eukaryotic Transcriptomes. Front Genet. 2016 Jan 11;6:361. doi: 10.3389/fgene.2015.00361. PMID: 26793234; PMCID: PMC4707302.

Moreau, R.A., Nyström, L., Whitaker, B.D., Winkler-Moser, J.K., Baer, D.J., Gebauer, S.K. and Hicks, K.B. (2018). Phytosterols and their derivatives: Structural diversity, distribution, metabolism, analysis, and health-promoting uses. *Progress in Lipid Research*, 70, pp.35–61. doi:https://doi.org/10.1016/j.plipres.2018.04.001.

Moreau, R.A., Whitaker, B.D. and Hicks, K.B. (2002). Phytosterols, phytostanols, and Their Conjugates in foods: Structural diversity, Quantitative analysis, and health-promoting Uses. *Progress in Lipid Research*, 41(6), pp.457–500. doi:https://doi.org/10.1016/s01637827(02)00006-1.

Moringa oleifera - Wikiwand (nd) https://www.wikiwand.com/en/articles/Moringa_oleifera (Accessed: 31 December 2025).

Muniandi, S.K., Ariff, F.F.M., Pisar, M.M., Harun, S.T., Abdullah, M.Z., Abdullah, F., Hashim, S.N.A.M., Bahari, S.N.S. and Saffie, N. (2024). Crop Improvement of *Moringa oleifera* L. through Genotype Screening for the Development of Clonal Propagation Techniques of High Yielding Clones in Malaysia. *Biology*, 13(10), p.785. doi:https://doi.org/10.3390/biology13100785.

Musich, R., Cadle-Davidson, L. and Osier, M.V. (2021). Comparison of Short-Read Sequence Aligners Indicates Strengths and Weaknesses for Biologists to Consider. *Frontiers in Plant Science*, 12. doi:https://doi.org/10.3389/fpls.2021.657240.

Mwami, P. Manourova, A. Hendre, P.S, Muchugi,A., Verner, V., Kariuki, P., Sulaiman, N., Polesny,Z. (2024). Traditional Knowledge, Use, And Management of *Moringa oleifera* among the Mijikenda Community in Kilifi, Kenya. *Plants* 2024, *13*(24), 3547; https://doi.org/10.3390/plants13243547

Nadeem, M. and Imran, M. (2016). Promising features of *Moringa oleifera* oil: recent updates and perspectives. *Lipids in Health and Disease*, 15(1). doi:https://doi.org/10.1186/s12944-0160379-0.

Nakashima, K & SUENAGA, Kazuhiro. (2017). Toward the Genetic Improvement of Drought Tolerance in Crops. Japan Agricultural Research Quarterly: JARQ. 51. 1-10. 10.6090/jarq.51.1.

Nardino M, Perin EC, Aranha BC, Carpes ST, Fontoura BH, de Sousa DJP, et al. (2022) Understanding drought response mechanisms in wheat and multi-trait selection. PLoS ONE 17(4): e0266368. https://doi.org/10.1371/journal.pone.0266368

National Sunflower Association (2001). The sunflower, National Sunflower Association, Bismarck, N. Dakota. 2001; March/ April 2001. Page 8–9

NCBI Resource Coordinators , Database resources of the National Center for Biotechnology Information, *Nucleic Acids Research*, Volume 46, Issue D1, 4 January 2018, Pages D8–D13, https://doi.org/10.1093/nar/gkx1095

Nguyen, K,H., Mostofa, M.G, Watanabe,Y, Tran, C.D ., Md. Rahman, M, Tran, Lam-Son, P. (2019). Overexpression of *GmNAC085* enhances drought tolerance in *Arabidopsis* by regulating glutathione biosynthesis, redox balance and glutathione-dependent detoxification of reactive oxygen species and methylglyoxal. *Environmental and Experimental Botany*. Vol.161pp242-254 https://doi.org/10.1016/j.envexpbot.2018.12.021

Nnamchi, C, I, Onyetugo, C.A and Amarachi,I.N (2021) “Peroxidase, an Example of Enzymes with Numerous Applications”, Asian Journal of Chemical Sciences, 10(2), pp. 11–22. doi: 10.9734/ajocs/2021/v10i219087.

Nouman, W. (2013). Biomass Production and Nutritional Quality of *Moringa oleifera* as Field Crop. *Turkish Journal of Agriculture and Forestry*. doi:https://doi.org/10.3906/tar-1206-29.

Oderinde, F.O., Akano, O.I., Adesina, F.A. and Omotayo, A.O. (2022). Trends in climate, socioeconomic indices and food security in Nigeria: Current realities and challenges ahead. *Frontiers in Sustainable Food Systems*, 6. doi:https://doi.org/10.3389/fsufs.2022.940858.

Ogbe, A.O. and Affiku, J.P. (2011). Proximate Study, Mineral And Anti-Nutrient Composition of *Moringa oleifera* Leaves Harvested From Lafia, Nigeria: Potential Benefits in Poultry Nutrition and Health. *Journal of microbiology, biotechnology and food sciences*, [online] 1(3), pp.296–308. Available at: https://office2.jmbfs.org/index.php/JMBFS/article/view/4464.

Ogunsina, B.S., Indira, T.N., Bhatnagar, A.S., Radha, C., Debnath, S. and Gopala Krishna, A.G. (2011). Quality characteristics and stability of *Moringa oleifera* seed oil of Indian origin. *Journal of Food Science and Technology*, 51(3), pp.503–510. doi:https://doi.org/10.1007/s13197-011-0519-5.

Ojeda-López, J., Marczuk-Rojas, J.P., Polushkina, O.A., Purucker, D., Salinas, M. and Carretero-Paulet, L. (2020). Evolutionary analysis of the *Moringa oleifera* genome reveals a recent burst of plastid to nucleus gene duplications. *Scientific Reports*, [online] 10(1), p.17646. doi:https://doi.org/10.1038/s41598-020-73937-w.

Ojiako, F., Adikuru, N. and Emenyonu, C. (2011). Critical issues in Investment, Production and Marketing of *Moringa oleifera* as an Industrial Agricultural raw material in Nigeria. J. Agric. Res. Dev. 10.

Olson, M.E (2019). Introduction to the Moringa Family. *in Fuglie, L., Ed., the Miracle Tree the Multiple Attributes of Moringa, CTA and CWS, Dakar, 11-28. - References - Scientific Research Publishing*. [online] Scirp.org. Available at: https://www.scirp.org/reference/referencespapers?referenceid=2453538 [Accessed 20 Feb. 2025]

Olson, M.E. (2002). *Combining Data from DNA Sequences and Morphology for a Phylogeny of Moringaceae Verdcourt, B. "A Synopsis of the Moringaceae." Kew Bulletin, vol. 40, no. 1, 1985, pp. 1–ix. JSTOR, https://doi.org/10.2307/4108470. Accessed 1 June 2025.(Brassicales)*, *SystematicBotany*.27(1):55–73.doi:https://doi.org/10.1043/03636445-27.1.55.

Oluduro, A.O (2025). Evaluation of Antimicrobial Properties and Nutritional Potentials of *Moringa oleifera* Lam. Leaf in South-Western Nigeria. *Malaysian Journal of Microbiology*, [online] pp.59–67. Available at: https://pesquisa.bvsalud.org/portal/resource/pt/wpr-625633 [Accessed 20 Feb. 2025].

Ondieki, S.K, Hunja, C.W, Muluvi, G.M, Korir, C.J, Mutiso, F., Kitheka, J.U, Kioko, D., Kimatu, J., Ndufa, J.K, and Mutati, K. (2023). Genetic diversity of selected *Moringa oleifera* Lam. Provenances from the coastal region of Kenya. https://doi.org/10.30574/msarr.2023.8.2.0113

Ono, H., Ishii, K., Kozaki, T., Ogiwara, I., Kanekatsu, M. and Yamada, T. (2015). Removal of redundant contigs from *de novo* RNA-Seq assemblies via homology search improves accurate detection of differentially expressed genes. *BMC Genomics*, 16(1). doi:https://doi.org/10.1186/s12864-015-2247-0.

Ost, C., Hieu Xuan Cao, Thuy Linh Nguyen, Axel Himmelbach, Mascher, M., Stein, N. and Klaus Humbeck (2023). Drought-Stress-Related Reprogramming of Gene Expression in Barley Involves Differential Histone Modifications at ABA-Related Genes. *International journal of molecular sciences (Online)*, [online] 24(15), pp.12065–12065. doi:https://doi.org/10.3390/ijms241512065.

Palada, M.C. (2017). The moringa industry in the Philippines: status, challenges and opportunities. *Acta Hortic*. 1158, 447-454 DOI:10.17660/Acta Hortic.2017.1158.52 https://doi.org/10.17660/ActaHortic.2017.1158.52

Palafox-Carlos, H., Ayala-Zavala, J.F. and González-Aguilar, G.A. (2011). The Role of Dietary Fiber in the Bioaccessibility and Bioavailability of Fruit and Vegetable Antioxidants. *Journal of Food Science*, 76(1), pp.R6–R15. doi:https://doi.org/10.1111/j.1750-3841.2010.01957.x.

Palomares, M.A., Dalmasso, C., Bonnet, E., Derbois, C., Brohard-Julien, S., Ambroise, C., Battail, C., Deleuze, J.-F. and Olaso, R. (2019). Systematic analysis of TruSeq, SMARTer and SMARTer Ultra-Low RNA-seq kits for standard, low and ultra-low quantity samples. *Scientific Reports*, 9(1). doi:https://doi.org/10.1038/s41598-019-43983-0.

Pandey A, Pradheep K, Gupta R, Nayar ER, Bhandari DC (2011) 'Drumstick tree' (*Moringa oleifera* Lam.): a multipurpose potential species in India. Genetic Resour Crop Evol 58(3):453–460

Pandian BA, Sathishraj R, Djanaguiraman M, Prasad PVV, Jugulam M(2020). Role of Cytochrome P450 Enzymes in Plant Stress Response. Antioxidants (Basel). 2020 May 25;9(5):454. doi: 10.3390/antiox9050454. PMID: 32466087; PMCID: PMC7278705.

Panes, V.A. and Baoas, R.D. (2018). Analysis of the transcription factors expressed in the mature seed embryos of *Moringa oleifera* Lam. using RNA-sequencing and de novo transcriptome assembly. ActaHortic. 1205, 705-716 DOI: 10.17660/ActaHortic.2018.1205.87 https://doi.org/10.17660/ActaHortic.2018.1205.87

Panes, V.A. and Discar, M.D.K.A. (2021). Discovering genes involved in the synthesis of secondary metabolites from the seeds of *Moringa oleifera* through transcriptome analysis. Acta Hortic. 1306, 51-58 DOI: 10.17660/ActaHortic.2021.1306.7 https://doi.org/10.17660/ActaHortic.2021.1306.7

Panova N, Gerasimova A, Gentscheva G, Nikolova S, Makedonski L, Velikova M, Beraich A, Talhaoui A, Petkova N, Batovska D, Nikolova K. *Moringa oleifera* Lam.: A Nutritional Powerhouse with Multifaceted Pharmacological and Functional Applications. Life (Basel). 2025 May 29;15(6):881. doi: 10.3390/life15060881. PMID: 40566534; PMCID: PMC12194112.

Pardey, P.G., Beddow, J.M., Hurley, T.M., Beatty, T.K.M. and Eidman, V.R. (2014). A Bounds Analysis of World Food Futures: Global Agriculture Through to 2050. *Australian Journal of Agricultural and Resource Economics*, 58(4), pp.571–589. doi:https://doi.org/10.1111/14678489.12072.

Pareek, A., Pant, M., Gupta, M.M., Kashania, P., Ratan, Y., Jain, V., Pareek, A. and Chuturgoon, A.A. (2023). *Moringa oleifera:* An Updated Comprehensive Review of Its Pharmacological Activities, Ethnomedicinal, Phytopharmaceutical Formulation, Clinical, Phytochemical, and Toxicological Aspects. *International Journal of Molecular Sciences*, [online] 24(3), p.2098. doi:https://doi.org/10.3390/ijms24032098.

Pasha SN, Shafi KM, Joshi AG, Meenakshi I, Harini K, Mahita J, Sajeevan RS, Karpe SD, Ghosh P, Nitish S, Gandhimathi A, Mathew OK, Prasanna SH, Malini M, Mutt E, Naika M, Ravooru N, Rao RM, Shingate PN, Sukhwal A, Sunitha MS, Upadhyay AK, Vinekar RS, Sowdhamini R. (2020). The transcriptome enables the identification of candidate genes behind medicinal value of Drumstick tree (*Moringa oleifera*). Genomics. 2020 Jan;112(1):621-628. doi: 10.1016/j.ygeno.2019.04.014. Epub 2019 Apr 29. PMID: 31048014.

Pasternak, D., Keatinge, J. D. H., Mamoudou, Z., (2017). Moringa research and cultivation in Niger. *Acta Horticulturae*, No.1158171-177. http://www.actahort.org/books/1158/1158_20.htm

Paterson A.H, Freeling, M., Sazaki, T. (2005) Grains of knowledge: genomics of model cereals. *Genome Res* 15:1643–1650 https://www.researchgate.net/publication/225230262.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. and Kingsford, C. (2017). Salmon Provides Fast and bias-aware Quantification of Transcript Expression. *Nature Methods*, 14(4), pp.417–419. doi:https://doi.org/10.1038/nmeth.4197.

Pirrò, S., Zanella, L., Kenzo, M., Montesano, C., Minutolo, A., Potestà, M., Sobze, M.S., Canini, A., Cirilli, M., Muleo, R., Colizzi, V. and Galgani, A. (2016). MicroRNA from *Moringa oleifera*: Identification by High Throughput Sequencing and Their Potential

Contribution to Plant Medicinal Value. *PLOS ONE*, 11(3), p.e0149495. doi:https://doi.org/10.1371/journal.pone.0149495.

Piironen, V., Lindsay, D.G., Miettinen, T.A., Toivo, J., Lampi, A.M. (2000). Plant sterols: biosynthesis, biological function and their importance to human nutrition. Journal of the Science of Food and Agriculture. 80. 939-966. 10.1002/(SICI)1097-0010(20000515)80:7<939::AID-JSFA644>3.3.CO;2-3.

Pluháčková, H., Weichtová, L., Lojková, L., Němec, P., Svoboda, Z., Boško, R., Pernica, M. and Benešová, K. (2023). The Content of Bioactive Compounds in *Moringa oleifera* and *Moringa stenopetala* Grown in Ethiopia. *South African Journal of Botany*, [online] 155, pp.355–360. doi:https://doi.org/10.1016/j.sajb.2023.02.010.

Poli, A., Marangoni, F., Corsini, A., Manzato, E., Marrocco, W., Martini, D., Medea, G. and Visioli, F. (2021). Phytosterols, Cholesterol Control, and Cardiovascular Disease. *Nutrients*, [online] 13(8), p.2810. doi:https://doi.org/10.3390/nu13082810.

Popescu, A.-A. and Huber, K.T. (2015). PSIKO2: a fast and versatile tool to infer population stratification on various levels in GWAS. *Bioinformatics*, [online] 31(21), pp.3552–3554. doi:https://doi.org/10.1093/bioinformatics/btv396.

Popoola J.O, Oluyisola B.O and Obembe, O (2014). Genetic diversity in *Moringa oleifera* from Nigeria using fruit morpho-metric characters and Random Amplified Polymorphic DNA (RAPD) markers. *Covenant Journal of Physical and Life Sciences (CJPL) Vol. 1, No. 2. May,2014.*[online]1(2), pp.43–60. Available at: https://www.researchgate.net/publication/265508338

Popoola, J. (2016). Phenotypic Intraspecific Variability among Some Accessions of Drumstick (*Moringa oleifera* Lam.). *Phenotypic Intraspecific Variability among Some Accessions of Drumstick (Moringa oleifera Lam.)*, [online] 10(1), pp.3681–3693. https://www.researchgate.net/publication/289504933

Prochazkova, D., Sairam, R.K., Srivastava, G.C., Singh, D.V.. (2001). Oxidative stress and antioxidant activity as the basis of senescence in maize leaves. *Plant Science*. 161(4): 765-771. doi:10.1016/S0168-9452(01)00462-9

Qiao T, Zhang L, Yu Y, Pang Y, Tang X, Wang X, Li L, Li B, Sun Q (2022). Identification and expression analysis of xyloglucan endotransglucosylase/hydrolase (XTH) family in grapevine (*Vitis vinifera* L.). PeerJ. 2022 Jun 13;10:e13546. doi: 10.7717/peerj.13546. PMID: 35722264; PMCID: PMC9202548.

R Core Team (2019). *R: The R Project for Statistical Computing*. [online] R-project.org.Available at: http://www.r-project.org/.

Rahman F, Nadeem M, Azeem MW, Zahoor Y. (2014). Comparison of the chemical characteristics of high oleic fraction of *Moringa oleifera* oil with some vegetable oils. Pak J Anal Environ Chem. 2014;15:80–3.

Rashid, U., Anwar, F., Moser, B.R. and Knothe, G. (2008). *Moringa oleifera* oil: A possible source of biodiesel. *Bioresource Technology*, 99(17), pp.8175–8179. doi:https://doi.org/10.1016/j.biortech.2008.03.066.

Reed RC, Bradford KJ, Khanday I. (2022) Seed germination and vigor: ensuring crop sustainability in a changing climate. Heredity (Edinb). 2022 Jun;128(6):450-459. doi:

10.1038/s41437-022-00497-2. Epub 2022 Jan 10. PMID: 35013549; PMCID: PMC9177656.

Rees DC, Johnson E, Lewinson O. (2009) ABC transporters: the power to change. N*at Rev Mol Cell Biol*. 2009 Mar;10(3):218-27. doi: 10.1038/nrm2646. PMID: 19234479; PMCID: PMC2830722.

Rejwinkel, Ansmant, Gatfield and Izaurralde, (2005). A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. http://www.rnajournal.org/cgi/doi/10.1261/rna.2191905.

Rensink W.A, Buell C.R (2004) Arabidopsis to rice. Applying knowledge from a weed to enhance our understanding of a crop species. *Plant Physiol* 135: 622–629.

Rebeca Rivas, Marciel T. Oliveira, Mauro G. Santos, (2013). Three cycles of water deficit from seed to young plants of *Moringa oleifera* woody species improves stress tolerance, Plant Physiology and Biochemistry,Volume 63, 2013, Pages 200-208,ISSN 0981-9428, https://doi.org/10.1016/j.plaphy.2012.11.026. (https://www.sciencedirect.com/science/article/pii/S098194281200335X)

Robiansyah, I., Hajar, A.S., Al-Kordy, M.,and Ramadan, A. (2014). Current status of economically important plant Moringa peregrina (Forsk.) in Saudi Arabia: A review. *Int. J. Theor. App. Sci.*. 6. 79-86.

Rogowska, A. and Szakiel, A. (2020). The Role of Sterols in Plant Response to Abiotic Stress.*Phytochemistry Reviews*, 19(6), pp.1525–1538. doi:https://doi.org/10.1007/s11101-020-09708-2.

Roser, M., Ritchie, H. and Rosado, P. (2023). *Hunger and Undernourishment*. [online] Our World in Data. Available at: https://ourworldindata.org/hunger-and-undernourishment.

Rurek, M., & Smolibowski, M. (2024). Variability of plant transcriptomic responses under stress acclimation: a review from high throughput studies. *Acta biochimica Polonica*, *71*, 13585. https://doi.org/10.3389/abp.2024.13585

Ruttarattanamongkol, K. and Petrasch, A. (2015). Antimicrobial Activities of *Moringa oleifera* Seed and Seed Oil Residue and Oxidative Stability of Its Cold Pressed Oil Compared with Extra Virgin Olive Oil. *J. Sci. Technol*, 37(5), pp.587–594.

Sahoo, J.P., Mohapatra, U., Sahoo, S. and Samal, K.C. (2020). Insights into the Miracle Plant *Moringa oleifera*. *The Pharma Innovation*, 9(7), pp.473–479. doi:https://doi.org/10.22271/tpi.2020.v9.i7h.4978.

Saini, A., Manuja, S., Upadhyay, R.G. *et al.* Assessing the effect of soil cultivation methods and genotypes on crop yield components, yield and soil properties in wheat (*Triticum aestivum* L.) and Rice (*Oryza sativa* L.) cropping system. *BMC Plant Biol* 24, 349 (2024). https://doi.org/10.1186/s12870-024-05001-y

Sakuraba,Y, Kim,Y-S., Han, S-H., Lee,B-D., Paek,N-C (2015). The Arabidopsis Transcription Factor NAC016 Promotes Drought Stress Responses by Repressing *AREB1* Transcription through a Trifurcate Feed-Forward Regulatory Loop Involving NAP, *The Plant Cell*, Volume 27, Issue 6, June 2015, Pages 1771–1787, https://doi.org/10.1105/tpc.15.00222

Salgotra, R.K. and Chauhan, B.S. (2023). Genetic Diversity, Conservation, and Utilization of Plant Genetic Resources. *Genes*, [online] 14(1), p.174. doi:https://doi.org/10.3390/genes14010174.

Salzberg, S.L., Phillippy, A.M., Zimin, A., Puiu, D., Magoc, T., Koren, S., Treangen, T.J., Schatz, M.C., Delcher, A.L., Roberts, M., Marçais, G., Pop, M. and Yorke, J.A. (2012). GAGE: A critical evaluation of genome assemblies and assembly algorithms. *Genome Research*, [online] 22(3), pp.557–567. doi:https://doi.org/10.1101/gr.131383.111.

Sánchez-Machado, D. I., et al. (2010). Fatty acids, tocopherols, and phytosterols content of *Moringa oleifera* oil extracted with different solvents. *Journal of Food Composition and Analysis*, 23(5), 419–425.

Sarwar, M., (2017). Biodiversity and Convservation of Moringa species. *Journal of Biodiversity and Conservation. JBC-APRF*-1(1): 5-5, 2017

Sattler SE, Gilliland LU, Magallanes-Lundback M, Pollard M, DellaPenna D. Vitamin E is essential for seed longevity and for preventing lipid peroxidation during germination. Plant Cell. 2004 Jun;16(6):1419-32. doi: 10.1105/tpc.021360. Epub 2004 May 21. PMID: 15155886; PMCID: PMC490036.

Sato, H., Suzuki, T., Takahashi, F., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2019). NF-YB2 and NF-YB3 Have Functionally Diverged and Differentially Induce Drought and Heat Stress-Specific Genes. *Plant Physiology*, *180*(3), 1677–1690. https://doi.org/10.1104/PP.19.00391

Savignac, J-M., Atanasova, V., Chereau, S., Ortega,V., and Richard-Forget, R. (2022). Role of Tochochromanols in tolerance of cereals to biotic stresses: specific focus on pathogenic and toxigenic fungal species. *Int. J. Mol. Sci.* 2022, *23*(16), 9303; https://doi.org/10.3390/ijms23169303

Shafi, K.M., Sajeevan, R.S., Kouser, S. *et al.* Transcriptome profiling of two *Moringa* species and insights into their antihyperglycemic activity. *BMC Plant Biol* 22, 561 (2022). https://doi.org/10.1186/s12870-022-03938-6

Shah W, Ullah S, Ali S, Idrees M, Khan MN, Ali K, et al. (2021) Effect of exogenous alpha-tocopherol on physio-biochemical attributes and agronomic performance of lentil (*Lens culinaris* Medik.) under drought stress. PLoS ONE 16(8): e0248200. https://doi.org/10.1371/journal.pone.0248200

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003 Nov;13(11):2498-504. doi: 10.1101/gr.1239303. PMID: 14597658; PMCID: PMC403769.

Sheoran S., Kaur Y., Kumar S., Shukla S., Rakshit S., Kumar, R. (2022).Recent Advances for Drought Stress Tolerance in Maize (Zea mays L.): Present Status and Future Prospects. Frontiers in Plant Science.Volume 13 - 2022. https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2022.872566. DOI=10.3389/fpls.2022.872566 ISSN=1664-462X

Shih M.C, Chang C.M, Kang S.M, Tsai M.L. (2011). Effect of different parts (leaf, stem and stalk) and seasons (summer and winter) on the chemical compositions and antioxidant activity of *Moringa oleifera*. *Int J Mol Sci*. 2011;12(9):6077-88. doi: 10.3390/ijms12096077. *Epub* 2011 Sep 19. PMID: 22016645; PMCID: PMC3189769.

Shrestha, R., et al. (2021) Exploring Phenotypic Variation and Associations in Root Nodulation, Morphological, and Growth Character Traits among 50 Guar Genotypes. Industrial Crops and Products, 171, Article ID: 113831.

Shyamli, P. S., Pradhan, S., Panda, M., & Parida, A. (2021). de novo Whole-Genome Assembly of *Moringa oleifera* Helps Identify Genes Regulating Drought Stress Tolerance. *Frontiers in plant science*, 12, 766999. https://doi.org/10.3389/fpls.2021.766999

Schaarschmidt, S., Fischer, A., Zuther, E. and Hincha, D.K. (2020). Evaluation of Seven Different RNA-Seq Alignment Tools Based on Experimental Data from the Model Plant *Arabidopsis thaliana*. *International Journal of Molecular Sciences*, 21(5), p.1720. doi:https://doi.org/10.3390/ijms21051720.

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M. and Ragg, T. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology*, 7(1), p.3. doi:https://doi.org/10.1186/1471-2199-7-3.

Sefrji, F. O., Abulfaraj, A. A., Alshehrei, F. M., Al-Andal, A., Alnahari, A. A., Tashkandi, M., Baz, L., Barqawi, A. A., Almutrafy, A. M., Alshareef, S. A., Alkhatib, S. N., Abuauf, H. W., Jalal, R. S., & Aloufi, A. S. (2025). Comprehensive analysis of orthologous genes reveals functional dynamics and energy metabolism in the rhizospheric microbiome of *Moringa oleifera*. *Functional & integrative genomics*, *25*(1), 82. https://doi.org/10.1007/s10142-025-01580-7

Sekhar C, Venkatesan, N, Murugananthi, D, Vidhyavathi,A, (2018). Status of Value Addition and Export of Moringa Produce in Tamil Nadu – a Case Study. *International Journal of Horticulture*, 2018,8:3, 16-28 DOI:10.5376/ijh.2018.08.0003

Selvam, A.B.D. (2005). Distribution, Phenology and Utilization of *Moringa oleifera* Lamk. - An indigenous Medicinal Plant of India. Journal of Economic and Taxonomic Botany. 29. 102-108.

Senthilkumar, A., Karuvantevida, N., Rastrelli, L., Kurup, S.S. and Cheruth, A.J. (2018). Traditional Uses, Pharmacological Efficacy, and Phytochemistry of Moringa peregrina (Forssk.) Fiori. —A Review. *Frontiers in Pharmacology*, 9. doi:https://doi.org/10.3389/fphar.2018.00465.

Shahzad, U., Khan, M.A., Jaskani, M.J., Khan, I.A. and Korban, S.S. (2013). Genetic Diversity and Population Structure of *Moringa oleifera*. *Conservation Genetics*, 14(6), pp.1161–1172. doi:https://doi.org/10.1007/s10592-013-0503-x.

Shahzad, B., Rehman, A., Tanveer, M., Wang, L., Park, S. K., Ali, A. (2022). Salt stress in brassica: effects, tolerance mechanisms, and management. *J. Plant Growth Regul.* 41 (2), 781–795. doi: 10.1007/s00344-021-10338-xShannon, P. (2003). Cytoscape: a Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, 13(11), pp.2498–2504. doi:https://doi.org/10.1101/gr.1239303.

Sharma, D. (2019). *Moringa oleifera* Lam.: The Honey Bee Heaven Plant in Jammu and Kashmir. *Bee World*, *96*(4), 120–122. https://doi.org/10.1080/0005772X.2019.1638688

Sharma, P., Sumita Kachhwaha, Mahenderakar Damodhar Mahendrakar, Kothari, S.L. and Singh, R.B. (2024). Assessment of the Genetic Diversity and Population Structure in *Moringa oleifera* Accessions Using DNA Markers and Phenotypic Descriptors. doi:https://doi.org/10.2139/ssrn.4720653.

Shyamli, P.S., Pradhan, S., Panda, M. and Parida, A. (2021). *de novo* Whole-Genome Assembly of *Moringa oleifera* Helps Identify Genes Regulating Drought Stress Tolerance. *Frontiers in Plant Science*, 12(12). doi:https://doi.org/10.3389/fpls.2021.766999.

Sim, M., Lee, J., Kwon, D., Lee, D., Park, N., Wy., S., Ko, Y., and Kim, J.(2023). Reference-based read clustering improves the de novo genome assembly of microbial strains. Computational and Structural Biotechnology Journal, Volume 21, 444 - 451

Simão F.A., Waterhouse R.M., Ioannidis P., Kriventseva E.V., Zdobnov E.M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015 Oct 1;31(19):3210-2. doi: 10.1093/bioinformatics/btv351. Epub 2015 Jun 9. PMID: 26059717.

Singh, A., Pandey, H., Pandey, S., Lal, D., Chauhan, D., Aparna, Antre, S.H., B. Santhosh and Kumar, A. (2023). Drought stress in maize: stress perception to molecular response and strategies for its improvement. *Functional & Integrative Genomics*, 23(4). doi:https://doi.org/10.1007/s10142-023-01226-6.

Singh, A.K., Rana, H.K., Tshabalala, T., Kumar, R., Gupta, A., Ndhlala, A.R. and Pandey, A.K. (2020). Phytochemical, nutraceutical and pharmacological attributes of a functional crop *Moringa oleifera* Lam: An overview. *South African Journal of Botany*, 129, pp.209–220. doi:https://doi.org/10.1016/j.sajb.2019.06.017.

Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J.M. and Kelly, S., 2016. TransRate: reference-free quality assessment of de novo transcriptome assemblies. Genome research, 26(8), pp.1134-1144.

Somia Ajab, N, A., Nour, A., Abdulgadir, J. and Akhtar, M. (2024). Physicochemical Properties and Fatty Acids Composition of Moringa Seed Oil Compared with Olive Oil. *Egyptian Academic Journal of Biological Sciences. C, Physiology and Molecular Biology/Egyptian Academic Journal of Biological Sciences. C, Physiology and Molecular Biology*, 16(1), pp.243–251. doi:https://doi.org/10.21608/eajbsc.2024.344273.

Soneson, C., Love, M.I. and Robinson, M.D. (2016). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*, 4, p.1521. doi:https://doi.org/10.12688/f1000research.7563.2.

Song, G., Li, X., Munir, R., Khan, A.R., Azhar, W., Yasin, M.U., Jiang, Q., Bancroft, I. and Gan, Y. (2020). The WRKY6 transcription factor affects seed oil accumulation and alters fatty acid compositions in *Arabidopsis thaliana*. *Physiologia Plantarum*, 169(4), pp.612–624. doi:https://doi.org/10.1111/ppl.13082.

Song, X., Zhao, Y., Wang, J., Lu M.Z (2021). The transcription factor KNAT2/6b mediates changes in plant architecture in response to drought via down-regulating GA20ox1 in *Populus alba* and *P. glandulosa*. *J Exp Bot.* 2021 Jul 28;72(15):5625-5637. doi: 10.1093/jxb/erab201. PMID: 33987654.

Stahl, E., Hartmann, M., Scholten, N. and Zeier, J. (2019). A Role for Tocopherol Biosynthesis in Arabidopsis Basal Immunity to Bacterial Infection. *PLANT PHYSIOLOGY*, 181(3), pp.1008– 1028. doi:https://doi.org/10.1104/pp.19.00618.

Stark, A.H. and Madar, Z. (2002). Olive Oil as a Functional food: Epidemiology and Nutritional Approaches. *Nutrition Reviews*, [online] 60(6), pp.170–176. doi:https://doi.org/10.1301/002966402320243250.

Statista. (2022). *Malnutrition: share of people by region 2022 | Statista*. [online] Available at:https://www.statista.com/statistics/273291/number-of [Accessed 13 Feb. 2025].

Stephens, M. (2016). False discovery rates: a new deal. *Biostatistics*, 2(18), p.kxw041. doi:https://doi.org/10.1093/biostatistics/kxw041.

Stephenson, K.K. and Fahey, J.W., (2004). Development of tissue culture methods for the rescue and propagation of endangered *Moringa spp..* germplasm. Economic botany, 58(1), pp.S116-S124.

Stermer, B.A, Bianchini, G.M., Korth, K.L (1994). Regulation of HMG-CoA reductase activity in plants. *Journal of Lipid Research*, Volume 35, Issue 7, 1133 - 1140

Su P, Sui C, Niu Y, Li J, Wang S, Sun F, Yan J and Guo S (2023) Comparative transcriptomic analysis and functional characterization reveals that the class III peroxidase gene TaPRX-2A regulates drought stress tolerance in transgenic wheat. *Front. Plant Sci.* 14:1119162. doi: 10.3389/fpls.2023.1119162

Subedi, U., Jayawardhane, K.N., Pan, X., Ozga, J., Chen, G., Foroud, N.A. and Singer, S.D. (2020). The Potential of Genome Editing for Improving Seed Oil Content and Fatty Acid Composition in Oilseed Crops. *Lipids*, 55(5), pp.495–512. doi:https://doi.org/10.1002/lipd.12249.

Sujith Kumar, M.S., Ali, K., Dahuja, A., Tyagi,A.(2015). Role of phytosterols in drought stress tolerance in rice, *Plant Physiology and Biochemistry*, Volume 96, 2015, Pp 83-89, ISSN 0981-9428, https://doi.org/10.1016/j.plaphy.2015.07.014.(https://www.sciencedirect.com/science/article/pii/S0981942815300607)

Sun, L., Sufen D., Yinbing G., Jose P. F., Zachary R., Kirankumar M., Perdeep M. "DiVenn: An Interactive and Integrated Web-based Visualization Tool for Comparing Gene Lists." Frontiers in Genetics 10 (2019):421.

Sutarno, S., and Rosyida, R. (2020). The growth and yield of *Moringa oleifera* Lam. As affected by plant spacing and cutting interval. *IOP Conference Series Earth and Environmental Science 518:012044* DOI:10.1088/1755-1315/518/1/012044

Szrok-Jurga, S., Czumaj, A., Turyn, J., Hebanowska, A., Swierczynski, J., Sledzinski, T., & Stelmanska, E. (2023). The Physiological and Pathological Role of Acyl-CoA Oxidation. *International Journal of Molecular Sciences*, *24*. https://doi.org/10.3390/ijms241914857

Szymanska,R., Nowicka,B., Gabruk,M., Glinska,S., Michlewska, S., Dluzewska,J., Sawicka, A., Kruk,J., Laitinen, R.(2014). Physiological and antioxidant responses of two accessions of *Arabidopsis thaliana* in different light and temperature conditions. *Physiologia Plantarum* 154(2), pp.194-209. https://doi.org/10.1111/ppl.12278

Thonglim, A., Bortolami, G., Delzon, S., Larter, M., Offringa, R., Keurentjes, J. J. B., Smets, E., Balazadeh, S., & Lens, F. (2023). Drought response in Arabidopsis displays synergistic coordination between stems and leaves. *Journal of experimental botany*, 74(3), 1004–1021. https://doi.org/10.1093/jxb/erac446

Tian, Y., Zeng, Y., Zhang, J., Yang, C., Yan, L., Wang, X., Shi, C., Xie, J., Dai, T., Peng, L. and Zeng Huan, Y., (2015). High quality reference genome of drumstick tree (*Moringa oleifera* Lam.), a potential perennial crop. Science China Life Sciences, 58(7), pp.627-638.

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., Xu, W. and Su, Z. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Research*, [online] 45(W1), pp.W122–W129. doi:https://doi.org/10.1093/nar/gkx382.

Tibbs Cortes L, Zhang Z, Yu J. Status and prospects of genome-wide association studies in plants. Plant Genome. 2021 Mar;14(1):e20077. doi: 10.1002/tpg2.20077. Epub 2021 Jan 13. PMID: 33442955.

Todaka,D., Takahashi,F., Yamaguchi-Shinozaki,K., Shinozaki,K. (2019). ABA-responsive gene expression in response to drought stress: cellular regulation and long distance signalling. Advances in Botanical Research. Vol.92 pp83-113 https://doi.org/10.1016/bs.abr.2019.05.001

Tomczak, A., Mortensen, J.M., Winnenburg, R., Liu, C., Alessi, D.T., Swamy, V., Vallania, F., Lofgren, S., Haynes, W., Shah, N.H., Musen, M.A. and Khatri, P. (2018). Interpretation of biological experiments changes with evolution of the Gene Ontology and its annotations. *Scientific Reports*, [online] 8(1), pp.1–10. doi:https://doi.org/10.1038/s41598-018-23395-2.

Trigo, C., Castelló, M.L. and Ortolá, M.D. (2022). Potentiality of *Moringa oleifera* as a Nutritive Ingredient in Different Food Matrices. *Plant Foods for Human Nutrition*, 78. doi:https://doi.org/10.1007/s11130-022-01023-9.

Trigo, C., Castelló, M.L., Ortolá, M.D., García-Mares, F.J. and Desamparados Soriano, M. (2020). *Moringa oleifera*: An Unknown Crop in Developed Countries with Great Potential for Industry and Adapted to Climate Change. *Foods*, 10(1), p.31. doi:https://doi.org/10.3390/foods10010031.

Tsaknis, J. (1998). *Characterisation of Moringa peregrina arabia seed oil. Grasas Aceites*. [online] Academia.edu. Available at: https://www.academia.edu/72819513/Characterisation_of_Moringa_peregrina_arabia_seed_oil_ Grasas_Aceites?uc-sb-sw=110923549 [Accessed 26 Feb. 2025].

Tsaknis J, Lalas S, Gergis V, Dourtoglou V, Spiliotis V. (1998) A total characterization of *Moringa oleifera* Malawi seed oil. Riv Ital Sostanza Grasse. 1998;75:21–7

Tsaknis, J., Lalas, S., Gergis, V., Dourtoglou, V. and Spiliotis, V. (1999). Characterization of *Moringa oleifera* Variety Mbololo Seed Oil of Kenya. *Journal of Agricultural and Food Chemistry*, 47(11), pp.4495–4499. doi:https://doi.org/10.1021/jf9904214.

Tshabalala, T., Ncube, B., Madala, N.E., Nyakudya, T.T., Moyo, H.P., Sibanda, M. and Ndhlala, A.R. (2019). Scribbling the Cat: A Case of the 'Miracle' Plant, *Moringa oleifera*. *Plants*, 8(11),p.510. doi:https://doi.org/10.3390/plants8110510.

Uffelmann, E., Huang, Q.Q., Munung, N.S. *et al.* Genome-wide association studies. *Nat Rev Methods Primers* 1, 59 (2021). https://doi.org/10.1038/s43586-021-00056-9

Ukwueze, C.K, Okogwu, O.I, Ebem, C.E., Nwonumara, G.N., Nwodo, J.N.Okogwu, O.I. (2019). Evaluation of the Influence of Geographical Location on Phytochemical Composition of *Moringa oleifera* Seeds. World Applied Sciences Journal. 37. 196-201. 10.5829/idosi.wasj.2019.196.201.

Valitova J.N, Sulkarnayeva A.G, Minibayeva F.V. (2016). Plant Sterols: Diversity, Biosynthesis, and Physiological Functions. *Biochemistry* (Mosc). 2016 Aug;81(8):819-34. doi: 10.1134/S0006297916080046. PMID: 27677551.

Valitova, J., Renkova, A., Beckett, R., & Minibayeva, F. (2024). Stigmasterol: An Enigmatic Plant Stress Sterol with Versatile Functions. *International journal of molecular sciences*, *25*(15), 8122. https://doi.org/10.3390/ijms25158122

Vanhercke, T., Srinivas Belide, Taylor, M.C., Anna El Tahchy, Okada, S., Rolland, V., Liu, Q., Mitchell, M., Shrestha, P., Venables, I., Ma, L., Blundell, C., Mathew, A., Ziolkowski, L., Niesner, N., Hussain, D., Dong, B., Liu, G., Godwin, I.D. and Lee, J. (2018). Up-regulation of lipid biosynthesis increases the oil content in leaves of*Sorghum bicolor*. *Plant Biotechnology Journal*, 17(1), pp.220–232. doi:https://doi.org/10.1111/pbi.12959.

Vaser, R., Sović, I., Nagarajan, N. and Šikić, M. (2017). Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Research*, [online] 27(5), pp.737–746. doi:https://doi.org/10.1101/gr.214270.116.

Velasco, L., Fernandez-Cuesta, A, Fernandez-Martinez,J.M (2014) New sunflower seeds with high contents of phytosterols. Institute for Sustainable Agriculture (IAS-CSIC) *OCL*, 21(6)

Vélez-Gavilán, J. (2022). *Moringa oleifera* (horse radish tree). *CABI Compendium*, CABI Compendium. doi:https://doi.org/10.1079/cabicompendium.34868.

Verdcourt, B. (1985) “A Synopsis of the Moringaceae.” *Kew Bulletin*, 40:1, 1985, pp. 1–ix. *JSTOR*, https://doi.org/10.2307/4108470. Accessed 1 June 2025.

Vergara-Jimenez, M., Almatrafi, M. and Fernandez, M. (2017). Bioactive Components in *Moringa oleifera* Leaves Protect against Chronic Disease. *Antioxidants*, 6(4), p.91. doi:https://doi.org/10.3390/antiox6040091.

Vijayaragavan, V., Sumathi, T., Irene vethamoni, P., Ravichandran, V. and Raju, M. (2025). Mechanisms of drought tolerance in Moringa: Strategies for mitigation and adaptation – Review. *Plant Science Today*. [online] doi:https://doi.org/10.14719/pst.4279.

Visioli, F., Davalos, A., López de las Hazas, M.C., Crespo, M.C. and Tomé‐Carneiro, J., 2020. An overview of the pharmacology of olive oil and its active ingredients. British journal of pharmacology, 177(6), pp.1316-1330.

Vlahakis, C. and Hazebroek, J., (2000). Phytosterol accumulation in canola, sunflower, and soybean oils: Effect of genetics, planting location, and temperature. *J. Am. Oil Chem. Soc.*, 77, 49-53, 2000.

Wairich, A., Vitali, A., Adamski, J. M., Lopes, K. L., Duarte, G. L., Ponte, L. R., Costa, H. K., Menguer, P. K., Santos, R. P. D., Fett, J. P., Sperotto, R. A., & Ricachenevsky, F. K. (2023). Enhanced expression of OsNAC5 leads to up-regulation of OsNAC6 and changes rice (Oryza sativa L.) ionome. *Genetics and molecular biology*, 46(1 Suppl 1), e20220190. https://doi.org/10.1590/1678-4685-GMB-2022-0190

Wang, B., Liu, C., Zhang, D., He, C., Zhang, J. and Li, Z. (2019). Effects of maize organ specific drought stress response on yields from transcriptome analysis. *BMC Plant Biology*, 19(1). doi:https://doi.org/10.1186/s12870-019-1941-5.

Wang J, Ding H, Zhang A, Ma F, Cao J, Jiang M. A novel mitogen-activated protein kinase gene in maize (Zea mays), ZmMPK3, is involved in response to diverse environmental cues. J Integr Plant Biol. 2010 May;52(5):442-52. doi: 10.1111/j.1744-7909.2010.00906.x. PMID: 20537040.https://doi.org/10.3390/ijms21041513

Wang, J., Xu, J., Yang, X., Xu, S., Zhang, M. and Lu, F. (2021). Boosting the power of transcriptomics by developing an efficient gene expression profiling approach. *Plant Biotechnology Journal*, 20(1), pp.201–210. doi:https://doi.org/10.1111/pbi.13706.

Wai, L. and Martti Tapani Tammi (2023). *Practical Bioinformatics For Beginners: From Raw Sequence Analysis To Machine Learning Applications*. World Scientific.

Wang, P., Chen, D., Zheng, Y., Jin, S., Yang, J. and Ye, N. (2018). Identification and Expression Analyses of SBP-Box Genes Reveal Their Involvement in Abiotic Stress and Hormone Response in Tea Plant (*Camellia sinensis*). *International Journal of Molecular Sciences*, 19(11), pp.3404–3404. http://doi.org/10.3390/ijms19113404.

Wang, P. and Wang, F. (2023). A Proposed Metric Set for Evaluation of Genome Assembly Quality. *Trends in Genetics*, 39(3), pp.175–186. doi: https://doi.org/10.1016/j.tig.2022.10.005.

Wang, X., Niu, Y. and Zheng, Y. (2021). Multiple Functions of MYB Transcription Factors in Abiotic Stress Responses. *International journal of molecular sciences*, [online] 22(11). doi:https://doi.org/10.3390/ijms22116125.

Wang Y.G, Fu F.L, Yu H.Q, Hu T, Zhang Y.Y, Tao Y, Zhu J.K, Zhao Y, Li W.C.(2018) Interaction network of core ABA signaling components in maize. *Plant Mol Biol*. Feb; 96(3):245-263. doi: 10.1007/s11103-017-0692-7. Epub 2018 Jan 17. PMID: 29344831.

Wang, Z., Gerstein, M. and Snyder, M. (2009). RNA-Seq: a Revolutionary Tool for Transcriptomics. *Nature Reviews Genetics*, 10(1), pp.57–63. doi:https://doi.org/10.1038/nrg2484.

Wang, Z., Cheng, K., Wan, L. et al. (2015). Genome-wide analysis of the basic leucine zipper (bZIP) transcription factor gene family in six legume genomes. BMC Genomics 16, 1053 (2015). https://doi.org/10.1186/s12864-015-2258-x

Wawrzyniak, J., Gawrysiak-Witulska, M. and Ryniecki, A. (2018). Management Control Points Related to the Lag Phase of Fungal Growth in a Stored Rapeseed Ecosystem. *Journal of the American Oil Chemists'Society*, 95(9), pp.1223–1235. doi:https://doi.org/10.1002/aocs.12130.

Wellcome, 2024. How are diet, food and climate change connected?

Wiltshire F.M.S, de França Santos A, Silva L.K.B, Carvalho de Almeida ,L., Dos Santos Freitas.L, Lima A.S, Fricks A.T, Dariva C, Soares C.M.F. (2021) Influence of seasonality on the physicochemical properties of *Moringa oleifera* Lam. Seed oil and their oleochemical potential. *Food Chem* (Oxf). 2021 Dec 27;4:100068. doi: 10.1016/j.fochms.2021.100068. PMID: 35415675; PMCID: PMC8991620.

Wise, M.J. and Tunnacliffe, A., (2004). POPP the question: what do LEA proteins do?. Trends in plant science, 9(1), pp.13-17.

Wu, T.D. and Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*, [online] 26(7), pp.873–881. doi:https://doi.org/10.1093/bioinformatics/btq057.

Xiao, J., Cheng, H., Li, X., Xiao, J., Xu, C., & Wang, S. (2013). Rice WRKY13 regulates cross talk between abiotic and biotic stress signaling pathways by selective binding to

different cis-elements. Plant physiology, 163(4), 1868–1882. https://doi.org/10.1104/pp.113.226019

Xing,Y., Wang,X., Mustafa,A. (2025). Exploring the link between soil health and crop productivity. *Ecotoxicology and Environmental Safety*, Volume 289,2025,117703,ISSN 0147-6513, https://doi.org/10.1016/j.ecoenv.2025.117703.(https://www.sciencedirect.com/science/article/pii/S0147651325000399)

Xing,J., Yang,Y., Zhang,Q., Yin,T., Zhao,R., Hao,G., Liu, X., Chi, Y., Zhang,J., (2023). A cytochrome P450 gene, *LpCYP72A15*, confers drought tolerance in perennial ryegrass. *The Journal of the British Grassland Society*. https://doi.org/10.1111/gfs.12642

Xing S, Zhu H, Zhou Y, Xue L, Wei Z, Wang Y, He S, Zhang H, Gao S, Zhao N, Zhai H, Liu Q (2022). A cytochrome P450 superfamily gene, IbCYP82D47, increases carotenoid contents in transgenic sweet potato. Plant Sci. 2022 May; 318:111233. doi: 10.1016/j.plantsci.2022.111233. Epub 2022 Feb 23. PMID: 35351305.

Yadav, S., Kalwan, G., Gill, S.S and Jain, P.K (2025) The ABC transporters and their epigenetic reegulation under drought stress in chickpea. *Plant Physiology and Biochemistry* vol.223 https://doi.org/10.1016/j.plaphy.2025.109903

Yang, C., McDonald, J.G., Cohen, J.C. and Hobbs, H.H. (2007). Sterol intermediates from cholesterol biosynthetic pathway as liver X receptor ligands. *The FASEB Journal*, 21(6). doi:https://doi.org/10.1096/fasebj.21.6.a979-a.

Yang, C., Huang, Y., Lv, P., Antwi-Boasiako, A., Begum, N., Zhao, T., & Zhao, J. (2022). NAC Transcription Factor *GmNAC12* Improved Drought Stress Tolerance in Soybean. *International journal of molecular sciences*, *23*(19), 12029. https://doi.org/10.3390/ijms231912029

Yang, E., Zheng, M., Zou, X., Huang, X., Yang, H., Chen, X. and Zhang, J. (2021). Global Transcriptomic Analysis Reveals Differentially Expressed Genes Involved in Embryogenic Callus Induction in Drumstick (*Moringa oleifera* Lam.). *International Journal of Molecular Sciences*, 22(22), p.12130. doi:https://doi.org/10.3390/ijms222212130.

Yang, R., Xue, L., Zhang, L., Wang, X., Qi, X., Jiang, J., Yu, L., Wang, X., Zhang, W., Zhang, Q. and Li, P. (2019). Phytosterol Contents of Edible Oils and Their Contributions to Estimated Phytosterol Intake in the Chinese Diet. *Foods*, 8(8), p.334. doi:https://doi.org/10.3390/foods8080334.

Yang, R.-Y., Chang, L.-C., Hsu, J.-C., Weng, B., Palada, M., Chadha, M. and Levasseur, V. (2006). *Nutritional and Functional Properties of Moringa Leaves − from Germplasm, to Plant, to Food, to Health*. [online] Available at: https://formadenvironnement.org/Yang_ghana_2006.pdf [Accessed 20 Feb. 2025].

Yang, S., Xu, K., Chen, S. *et al.* A stress-responsive bZIP transcription factor *OsbZIP62* improves drought and oxidative tolerance in rice. *BMC Plant Biol* 19, 260 (2019). https://doi.org/10.1186/s12870-019-1872-1

Yang, T., Zeng, Y., Zhang, J., Yang, C., Yan, L., Wang, X., Shi, C.-Y., Xie, J., Dai, T., Lei, P., Huan, Y., Xu, A., Huang, Y., Zhang, J.-J., Ma, X., Dong, Y., Hao, S. and Sheng, J. (2015). High quality reference genome of drumstick tree (*Moringa oleifera* Lam.), a potential perennial crop. *Sci. China Life Sci*, 58(7), pp.627–638. doi:https://doi.org/10.1007/s11427-015-4872-x.

Yip, A.M. and Horvath, S. (2006). The Generalized Topological Overlap Matrix for Detecting Modules in Gene Networks. *Proceedings of the 2006 International Conference on Bioinformatics & Computational Biology, BIOCOMP'06, Las Vegas, Nevada, USA, June 26-29,2006*, [online] pp.451–457. Available at: https://www.researchgate.net/publication/221051399_The_Generalized_Topological_Overlap_ Matrix_for_Detecting_Modules_in_Gene_Networks.

Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics : a journal of integrative biology*, *16*(5), 284–287. https://doi.org/10.1089/omi.2011.0118

Zarina, Ab Waheed Wani, Rawat, M., Kaur, H., Das, S., Kaur, T., Akram, N., Faisal, Z., Syed Saad Jan, Nabila Nusrat Oyshe, Mahbubur Rahman Khan and Yasir Abbas Shah (2024). Medicinal utilization and nutritional properties of drumstick (*Moringa oleifera*)—A comprehensive review. *Food science & nutrition*, 7(12). doi:https://doi.org/10.1002/fsn3.4139.

Zhao, B., Li. H, Lan, T., Wu, D., Chen, Z. (2019) Characterization of the Chemical Composition of Chinese *Moringa oleifera* Seed Oil. *Journal of the American Oil Chemists' Society* 96:5

Zhang, B. and Horvath, S. (2005). A General Framework for Weighted Gene Co-Expression Network Analysis. *Statistical Applications in Genetics and Molecular Biology*, 4(1). doi:https://doi.org/10.2202/1544-6115.1128.

Zhang, H., Shi, Y., Sun, M., Hu, X., Hao, M., Shu, Y., Zhou, X.-R., Hu, Q., Li, C. and Mei, D. (2022). Functional Differentiation of BnVTE4 Gene Homologous Copies in α-Tocopherol Biosynthesis Revealed by CRISPR/Cas9 Editing. *Frontiers in Plant Science*, 13. doi:https://doi.org/10.3389/fpls.2022.850924.

Zhang, H.-X., Jin, J.-H., He, Y.-M., Lu, B.-Y., Li, D.-W., Chai, W.-G., Khan, A. and Gong, Z.-H. (2016). Genome-Wide Identification and Analysis of the SBP-Box Family Genes under *Phytophthora capsici* Stress in Pepper (Capsicum annuum L.). *Frontiers in Plant Science*, 7. doi:https://doi.org/10.3389/fpls.2016.00504.

Zhang, J., Li, J., Geng, G., Hu, W., Ren, T., Cong, R., Li, X. and Lu, J. (2020). Combined application of nitrogen and potassium reduces seed yield loss of oilseed rape caused by Sclerotinia stem rot disease. *Agronomy Journal*, 112(6), pp.5143–5157. doi:https://doi.org/10.1002/agj2.20410.

Zhao, Z., Wu, S., Gao, H., Tang, W., Wu, X., & Zhang, B. (2023). The BR signaling pathway regulates primary root development and drought stress response by suppressing the expression of PLT1 and PLT2 in Arabidopsis thaliana. Frontiers in plant science, 14, 1187605. https://doi.org/10.3389/fpls.2023.1187605

Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C. and Weir, B.S. (2012). A high performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28(24), pp.3326–3328. doi:https://doi.org/10.1093/bioinformatics/bts606.

Zhou, Y., Chen, L., Fan, X. and Bian, Y. (2014). *de novo* Assembly of *Auricularia polytricha* Transcriptome Using Illumina Sequencing for Gene Discovery and SSR Marker Identification. *PLoS ONE*, 9(3), p.e91740. doi:https://doi.org/10.1371/journal.pone.0091740.

Zion Market Research, (2023). Global Moringa Market Size Envisioned at USD 18.2 Billion by 2030, Increasing a CAGR of 8.5% | Healthcare Revolution https://www.zionmarketresearch.com/report/moringa-market