MORPHO-AGRONOMIC, BIOCHEMICAL AND MOLECULAR CHARACTERIZATION OF BLACK CUMIN (*Nigella sativa* L.) IN ETHIOPIA

PhD DISSERTATION

BASAZINEW DEGU GEBREMEDIN

HAWASSA UNIVERSITY, HAWASSA, ETHIOPIA

APRIL, 2025

MORPHO-AGRONOMIC, BIOCHEMICAL AND MOLECULAR CHARACTERIZATION OF BLACK CUMIN (*Nigella sativa* L.) IN ETHIOPIA

BASAZINEW DEGU GEBREMEDIN

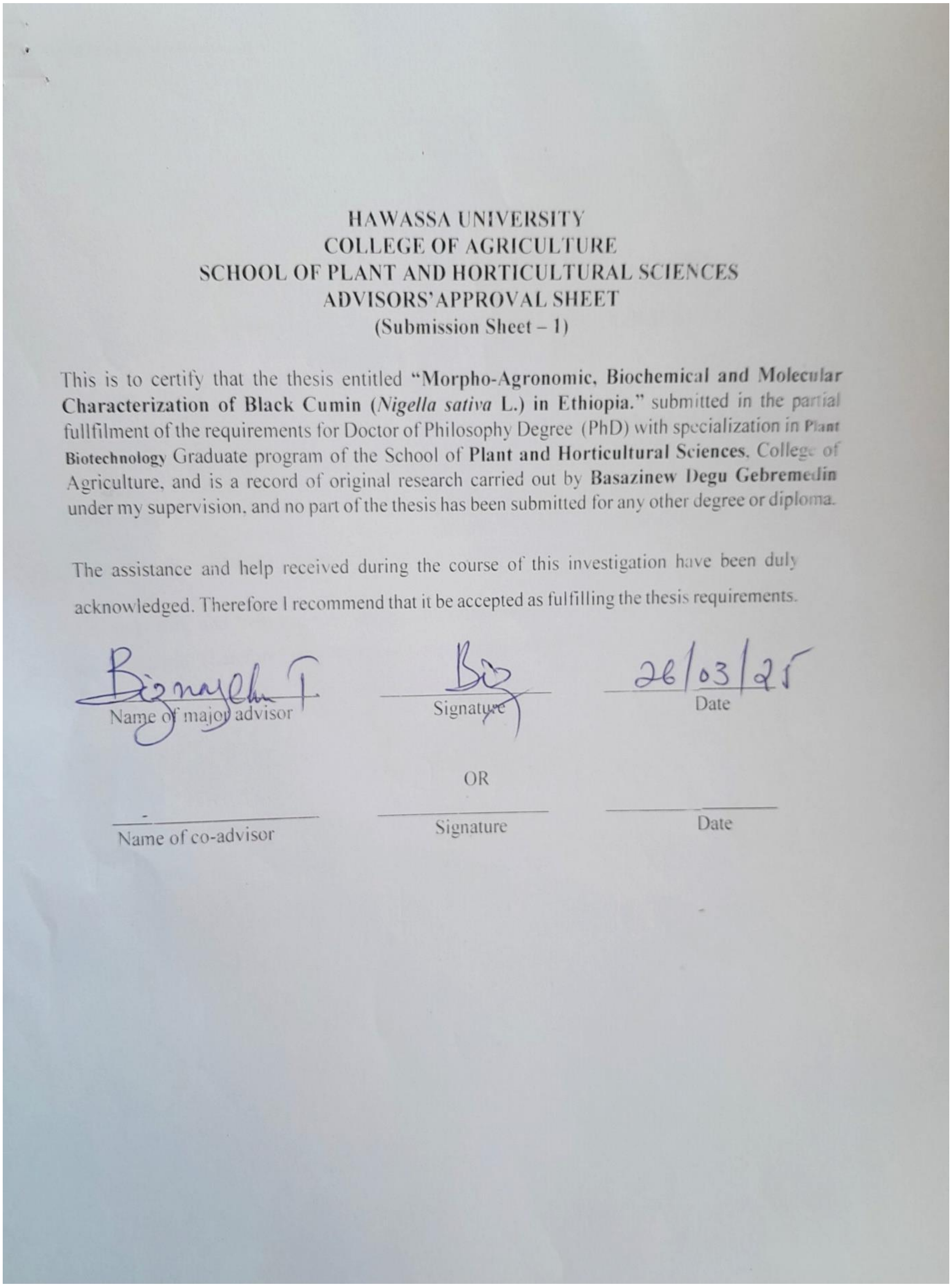
A THESIS SUBMITTED TO THE
SCHOOL OF PLANT AND HORTICULTURAL SCIENCES,
COLLEGE OF AGRICULTURE,
SCHOOL OF GRADUATE STUDIES, HAWASSA UNIVERSITY,
HAWASSA, ETHIOPIA

IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE
DEGREE OF

DOCTOR OF PHILOSOPHY IN PLANT SCIENCE
(SPECIALIZATION: PLANT BIOTECHNOLOGY)

MAJOR SUPERVISOR: BIZUAYEHU TESFAYE (PhD, Assoc. Prof)
CO- SUPERVISORS: WENDAWEK ABEBE (PhD, Assoc. Prof)
KEBEBEW ASSEFA (PhD, Lead Researcher)

APRIL, 2025



HAWASSA UNIVERSITY
COLLEGE OF AGRICULTURE
SCHOOL OF PLANT AND HORTICULTURAL SCIENCES
ADVISORS'APPROVAL SHEET
(Submission Sheet – 1)

This is to certify that the thesis entitled **"Morpho-Agronomic, Biochemical and Molecular Characterization of Black Cumin (*Nigella sativa* L.) in Ethiopia."** submitted in the partial fullfilment of the requirements for Doctor of Philosophy Degree (PhD) with specialization in **Plant Biotechnology** Graduate program of the School of **Plant and Horticultural Sciences**, College of Agriculture, and is a record of original research carried out by **Basazinew Degu Gebremedin** under my supervision, and no part of the thesis has been submitted for any other degree or diploma.

The assistance and help received during the course of this investigation have been duly acknowledged. Therefore I recommend that it be accepted as fulfilling the thesis requirements.

Bizuayehu T.		26/03/25
Name of major advisor	Signature	Date
	OR	
Name of co-advisor	Signature	Date

EXAMINERS' APPROVAL SHEET
SCHOOL OF GRADUATE STUDIES
HAWASSA UNIVERSITY EXAMINERS' APPROVAL SHEET
(Submission Sheet-2)

We, the undersigned, members of the Board of Examiners of the final open defense **Basazinew Degu Gebremedin** have read and evaluated his/her thesis entitled **Morpho-Agronomic, Biochemical and Molecular Characterization of Black Cumin (*Nigella sativa* L.) in Ethiopia** and examined the candidate. This is, therefore, to certify that the thesis has been accepted in partial fulfillment of the requirements for the degree

Bizuayehu T.		26/03/25
Name of Major Advisor	Signature	Date
Hewan Demisse		18/03/2025
Name of Internal Examiner I	Signature	Date
Zerihun Dennew		18/03/2025
Name of Internal Examiner II	Signature	Date
Tileye Feyissa		18/03/2025
Name of External Examiner	Signature	Date
ሔዋን ደምሴ (ዶ/ር) Hewan Demissie(PhD) የምርምርና ቴክኖሎጂ ሽግግር Research and Technology Transfer Associate Dean		01-07-2025
SGS Approval	Signature	Date

Final approval and acceptance of the thesis is contingent upon the submission of the final copy of the thesis to the school of Graduate Studies (SGS) through the School Graduate Committee (SGC) of the candidate's School.

Stamp of SGS   Date ____________________

Remark:

- ○ Use this form to submit the thesis with **<u>minor or without correction</u>** suggested by the examining board
- ○ 3 copies
- ○

DEDICATION

To my mother, Mrs. Aberash Dadore, and the memory of my father, the late Mr. Degu Gebremedin.

DECLARATION

I hereby declare that this PhD dissertation is my original work and has not been presented for a degree in any other university, and all sources of material used for this dissertation have been duly acknowledged.

Name: Basazinew Degu Gebremedin

Signature: ______________________

Date: __________________________

BIOGRAPHICAL SKETCH

Basazinew Degu Gebremedin was born at Metahara Haro-Adi kebele 01, Fentale District, East Shoa Zone of Oromia Regional State, Ethiopia, in 1986 G.C. He attended his primary education at Haro-Adi primary and medium secondary school in Metahara. Then he followed his secondary and Preparatory education at Merti Comprehensive Secondary School in Metahara and then after, he joined the higher education at Haramaya University College of Agriculture in 2005 where he got a B.Sc. Degree in Plant Science in July 2007. In 2008 he was employed as an Irrigation and drainage development expert at Merhabete district, North Shoa Zone, Amhara Regional State, and served for four months. In 2009 he was employed as unit farm manager of the cotton plantation at Middle Awash Agricultural Development Enterprise (Sidihafage section), Amibara district, Zone 03, Afar Regional State, and served for 2 years and 5 months. In 2011 he was employed as a researcher in the Ethiopian Institute of Agricultural Research, Wondo Genet Agricultural Research Center and he has been working here. In 2015, he got his M.Sc. degree in Horticulture from Hawassa University, College of Agriculture, School of Plant and Horticultural Science. Then, he joined Hawassa University, College of Agriculture, School of Graduate Studies to pursue his PhD Degree in Plant Biotechnology in 2019.

ACKNOWLEDGEMENTS

First of all, I thank the Almighty God who has given me additional age to live and health and passed me through all the difficult situations that are said to be insurmountable.

I want to express my special thanks to my major supervisor, Dr. Bizuayehu Tesfaye, who is an associate professor at Hawassa University College of Agriculture for his special and appreciated close supervision, technical guidance, support, and moral excellence as a father, brother, and friend from the inception of the research proposal writing until the completion of my study period. I also express my heartfelt appreciation to my co-supervisors, Dr. Wendawek Abebe, who is an associate professor at Addis Ababa University, and Dr. Kebebew Assefa, who is the lead researcher at the Ethiopian Institute of Agricultural Research (EIAR), for their critical comments, guidance and support starting from the proposal writing until the completion of my PhD study. I am also thankful to my instructors, who gave me their valuable knowledge of what the profession needs.

I want to express my profound gratitude to the EIAR for giving me a study leave and financial support. I appreciate Hawassa University for enrolling me as a PhD student in the School of Plant and Horticultural Sciences. I am indebted to the Norwegian Agency for Development Cooperation (NORAD) Project, Hawassa University, for their financial support. The financial support given by the Food Systems Resilience Program (FSRP) is duly appreciated. I am also grateful to the Wondo Genet Agricultural Research Center (WGARC) for providing the necessary facilities and support during the study period. I am highly thankful to the WGARC staff, especially the finance officers and researchers, for their positive moral support and timely responses to my questions related to my study.

My heartfelt appreciation also goes to the Ethiopian Biodiversity Institute (EBI), Debre Zeit, and Sinana Agricultural Research Centers for providing seeds of black cumin genotypes, which gave life to my planned experiment. Without seeds, it would have been impossible to think about my experiment. I appreciate Debre Zeit and Kulumsa Agricultural Research Centers for providing me with the research field and technically appropriate persons. I am very grateful to my sureties Ms. Workitu Tura, and Mr. Henok Tesfaye who believe in me, and Dr. Aynalem Gebre for her meek and invaluable assistance in transporting the black cumin seeds to a place where high-quality laboratory facilities of SEQART AFRICA located at Biosciences Eastern and Central Africa (BecA-ILRI) Hub in Nairobi, Kenya and facilitating the sequencing process. Without her support, it would not be possible to think about the completion of my PhD study by this time.

I am indebted to Mr. Assebe Getachew and Mr. Awol Hamza for their unreserved support during fieldwork and data collection, Mr. Demis Fikre, Mr. Desta Fekadu Mijena, Mr. Tewodros Lulseged, and Mr. Getachew Asefa for their positive assistance while doing this experiment. I am also thankful to Mr. Dinka Mulugeta from EIAR Food Science and Nutrition Laboratory for handling the Oleoresin analysis, WGARC natural product laboratory staff Mrs. Birtukan Yanera, Mrs. Agegnehush Dabi and Mrs. Selamawit Derese for handling the extraction of essential oils with me, Mr. Beriso Mieso and Mr. Abdela Befa for handling the Gas chromatography-mass spectrometry (GC-MS) analysis and assisting me how to summarize the results, WGARC Technology Multiplication and Seed Research staff Mr. Zerihun Beshir for providing me a woven polypropylene bags for data collection, Mr. Dejene Tadesse for providing me paper bags for storing seeds and center director Mr. Muluken Philipos for his positive and valuable support while doing this experiment.

Finally, I would like to express my sincere gratitude to my dear and beautiful mother, Mrs. Abrash Dadore, who raised me in difficulties and laid the foundation for who I am today, and to my family who have been by my side from the beginning to the end in thoughts and morals during this challenging period. I offer my heartfelt thanks to my beloved spouse, Mrs. Endalamaw Shewaye, and her entire family, who have been by my side from the bottom of my heart. In addition, in the name of God, I would like to thank my dear brothers Damtew Abewoy, Elias Meskelu, and others; my sisters, Selamawit Degu, Tsigie Worku, and others, and all my colleagues whose names have not mentioned your names. God bless you.

PAPER/MANUSCRIPT TITLE

1. Genetic Diversity of Ethiopian Black cumin (*Nigella sativa* L.) Based on Morpho-Agronomic Characteristics

Published as: Gebremedin, B.D., Asfaw, B.T., Mengesha, W.A., Mengesha, K.A. Genetic diversity of Ethiopian black cumin (*Nigella sativa* L.) based on morpho-agronomic characteristics. *Euphytica*, 220, 51 (2024). https://doi.org/10.1007/s10681-024-03315-4.

2. Biochemical Characterization of Ethiopian Black Cumin (*Nigella sativa* L.)

Published as: Basazinew Degu Gebremedin, Bizuayehu Tesfaye Asfaw, Wendawek Abebe Mengesha, Kebebew Assefa Abebe. Biochemical Characterization of Ethiopian Black Cumin (*Nigella sativa* L.). *International Journal of Food Science*, Volume 2024, Article ID 2746560, 12 pages. https://doi.org/10.1155/2024/2746560.

3. Diversity of Ethiopian Black Cumin (*Nigella sativa* L.) Based on Compositions of Essential Oil

Submitted to Biochemistry Research International as:

Basazinew Degu Gebremedin, Bizuayehu Tesfaye Asfaw, and Wendawek Abebe Mengesha. Diversity of Ethiopian Black Cumin (*Nigella sativa* L.) Based on Compositions of Essential Oil.

4. Genetic Diversity and Population Structure of Ethiopian Black Cumin (*Nigella sativa* L.) as Revealed by DArTseq SNP Markers

Submitted to Molecular Breeding as:

Basazinew Degu Gebremedin, Bizuayehu Tesfaye Asfaw, Wendawek Abebe Mengesha, and Kebebew Assefa Abebe. Genetic Diversity and Population Structure of Ethiopian Black Cumin (*Nigella sativa* L.) as Revealed by DArTseq SNP Markers

ABBREVIATIONS AND ACRONYMS

ANOVA	Analysis of Variance
AMOVA	Analysis of Molecular Variance
AOAC	Association of Official Analytical Chemists
BecA	Biosciences east and central Africa
DAPC	Discriminate Analysis of Principal Component
DArT	Diversity Array Technology
EIAR	Ethiopian Institute of Agricultural Research
EBI	Ethiopian Biodiversity Institute
GBS	Genotype by Sequencing
GC-MS	Gas Chromatography-Mass Spectrometry
GCV	Genotypic Coefficient of Variance
GLM	General Linear Model
GWAS	Genome-Wide Association Study
ILRI	International Livestock Research Institute
MoA	Ministry of Agriculture
PCA	Principal Component Analysis
PCoA	Principal Coordinate Analysis
PCV	Phenotypic Coefficient of Variance
SAS	Statistical Analysis System
SNP	Single Nucleotide Polymorphism
UPGMA	Unweighted Pair Group Method with Arithmetic mean
WGARC	Wondo Genet Agricultural Research Center

TABLE OF CONTENTS

Contents

LIST OF TABLES

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

LIST OF FIGURES

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

APPENDIX

LIST OF TABLES IN THE APPENDIX

LIST OF FIGURES IN THE APPENDIX

Abstract

MORPHO-AGRONOMIC, BIOCHEMICAL AND MOLECULAR CHARACTERIZATION OF BLACK CUMIN (*Nigella sativa* L.) IN ETHIOPIA

Black cumin (Nigella sativa L.) is a diploid annual flowering plant native to Southern Europe, North Africa, and Southwest Asia. It is cultivated worldwide for its medicinal and aromatic values. Ethiopia is an important center of black cumin genetic diversity. This study aims to characterize Ethiopian black cumin genotypes using morpho-agronomic, biochemical, and molecular markers (SNPs). A total of 64 genotypes, including 8 improved varieties and 56 genotypes from five Ethiopian regions, were characterized for morpho-agronomic and biochemical traits at Debre Zeit and Kulumsa Agricultural Research Centers. Essential oils and fixed oils were extracted, and their compositions analyzed using GC-MS. Molecular characterization was conducted using DArTseq SNP markers to determine genetic variation and population structure of 94 genotypes. Significant differences were found among genotypes for most morpho-agronomic traits, with substantial variability in essential oil compositions. Biochemical analysis revealed significant differences in fixed oil yield, essential oil content, and yield among genotypes. Molecular analysis indicated high levels of genetic diversity within regions and among genotypes, clustering into two distinct groups. The findings reveal significant diversity and variation among Ethiopian black cumin genotypes, offering valuable insights for conservation and breeding programs. The study emphasizes the need for equal attention to all growing areas for effective crop improvement.

Keywords: Biochemical, Cluster analysis, Diversity, DArTseq SNP, Essential oil, Essential oil composition, Fixed oil, Morpho-agronomic traits, Principal component analysis

CHAPTER 1

1. Introduction

1.1. Background and Justification

1.1.1. Black Cumin: General Perspectives

Black cumin (*Nigella sativa* L.) is an annual flowering plant from the Ranunculaceae family. The genus *Nigella* comprises about 20 species occurring predominantly in Europe, the Mediterranean, and Central Asia (Demel, 2000). It is native to Southern Europe, North Africa, and Southwest Asia. Black cumin is cultivated in India, Pakistan, Syria, Turkey, Saudi Arabia, South Europe (Khare, 2004), and Ethiopia (Demel, 2000). It is diploid (2n = 2x = 12) and an erect, usually profusely branched herb that grows up to 70 cm tall (Ermias, 2009). The crop is a minor cultivated crop in the regions extending from Morocco to Northern India; in sub-Saharan Africa in particular, it is grown in Niger, and eastern Africa, especially Ethiopia (Iqbal *et al.*, 2010). It grows in Southwest Asia, the Mediterranean Region (Toncer and Kizil, 2004), and Ethiopia.

The seeds of black cumin contain 26.7% protein, 28.5% fat, 24.9% carbohydrates, 8.4% crude fiber, and 4.8% total ash, vitamins, and minerals like Cu, P, Zn, Fe, etc (Cheikh-Rouhou et al., 2008). It contains all essential amino acids and is a rich source of vitamins and minerals (Abu-Jadayil *et al.*, 1999; Tierram, 2005). It is an important spice, commonly used to add flavor to bread. Black cumin seeds contain alkaloids (nigellicines and nigelledine), saponin (α- hederin), and fixed and essential oils (Ozel *et al.*, 2009). The fixed and essential oil of black cumin contains various bioactive molecules such as thymoquinone, thymol, tocopherol, trans-retinol, and selenium (Sultan *et al.*, 2009). Javed *et al*. (2010) have also described black cumin as a miraculous plant and is considered by earliest herbal specialists as “The herb from heaven”. Medicinally, it is used for many complaints such as asthma, emphysema, and bronchitis (Atta, 2003). Thymoquinone and nigellimine are among the main components (Ermias, 2009). Several findings also reported that black cumin has different pharmacological activities such as antioxidant (Ozdemir *et al*., 2018), antidiabetic (Daryabeygi-Khotbehsara *et al*., 2017), antihypertensive (Vasant *et al*., 2012), antibacterial (Abdallah, 2017), antifungal (Taha *et al*., 2010; Marongiu *et al*., 2013; Mahmoudvand *et al*., 2014; Aljabre *et al*., 2015; Shokri, 2016), antiviral (Barakat *et al*., 2013; Onifade *et al*., 2013; Forouzanfar *et al*., 2014; Onifade *et al*., 2015), antiparasitic (Bafghi *et al*., 2009; Okeola *et al*., 2011; Abd El-Hack *et al*., 2016; Bachek *et al*., 2016) and anticancer (Schneider-Stock *et al*., 2014).

The cultivation and use of spices and other essential oil-bearing plants are not new to Ethiopia. It is as old as the crops themselves, and its history can be traced back to the reign of Queen Sheba (ca.992 BC). Ethiopia has many of these plant species (Endashaw, 2007). So far in Ethiopia, there are eight released varieties of black cumin, namely Derbera, ADEN, DERSHYE, Gammachis, Sooressaa, Silingo, Kenna, and Qeneni (MOA, 2019).

Black cumin is cultivated in Amhara, Oromia, Tigray, Benshangul-Gumuz, and South West Ethiopia Peoples Regions at an altitude ranging between 1500 and 2500 m.a.s.l., often intercropped with cereals. Black cumin is one of the spices export commodities of Ethiopia. In 2009/10, the export of black and white cumin was 801 MT (metric tons) valued at 1.55 million USD. Out of these, black cumin had the greater share of 97.59% and 99.02% volume in MT and value in USD, respectively. The countries of major destinations with their share of the total value of cumin export are Indonesia, Saudi Arabia, and Sudan with a share of 33.8%, 28.7%, and 22.7%, respectively (Masresha, 2010).

In Ethiopia, black cumin serves as a flavor in bread and sauces. Most Ethiopian people use it in their household spice preparation. More recently, a great deal of attention has been given to the seed and oil yields of black cumin. Consequently, consumption has increased, and among the spices, black cumin is the second cash crop exported next to ginger in Ethiopia. Ethiopia's annual production of black cumin seed was 18,000 metric tons in 2014/15 (Ethiopian Investment Agency, 2015), and the national average of Black cumin productivity is 0.79 tons per hectare (Habtewold *et al*., 2017).

Due to the increased demand for black cumin seed for local consumption, such as oil and oleoresin for medicinal purposes, its export market, its potential in crop diversification, and income generation, and its importance in reducing the risk of crop failure, black cumin as a best alternative crop under the Ethiopian smaller land holdings (Dessalegn and Wubshet, 2018).

Natural or human-directed crop plant evolution is primarily based on existing genetic diversity in the population. The degree of differentiation between or within species is referred to as diversity. The foundation of all agricultural improvement initiatives is the existence of intra- and inter-specific differences (Bhandari *et al*., 2017). The basis of both crop improvement and plant survival in the natural world is genetic diversity. Given that genetic diversity may be the source of numerous novel traits that confer tolerance to various biotic and abiotic stresses, it becomes increasingly significant in the age of climatic change and the unanticipated events accompanying it. Plant breeders can create new and improved cultivars with desirable traits (high yield potential, large seed, etc.) and breeder-prefer traits (pest and disease resistance and photosensitivity, etc.) (Bhandari *et al.*, 2017). A healthy population depends on genetic diversity to preserve the variety of genes that may confer resistance against diseases, pests, or other stressful situations. Additionally, it helps individuals adapt to a range of biotic and abiotic stresses (Salgotra and Chauhan, 2023). Numerous significant agricultural phenomena, including heterosis and transgressive segregation, have a genetic basis. Both the development of new varieties and the correction of defects in commercial varieties require diverse lines. Therefore, the main objectives of any crop improvement program are to identify diverse lines (if available), create diversity (if not, or if it is limited), and then use that diversity.

Variations in morphology, anatomy, physiological behavior, or biochemical features can manifest as heritable character variation. Morphological, cytological, biochemical, and molecular characterization methods can all be used for diversity analysis (Bhandari *et al.,* 2017). Morphological markers were the first and are still used for diversity analysis. These were variations of a specific plant species that existed in the wild. Eventually, variations in a species' cytology and biochemistry that result from its genotypes were employed to measure genetic diversity. Molecular markers have emerged as the preferred approach for assessing genetic diversity (Bhandari *et al.,* 2017).

1.2. Statement of the Problem

Black cumin is a crucial crop for income generation and domestic use in Ethiopia. However, research on its characterization and improvement has been intermittent. Comprehensive studies on its use potential, genetic diversity, and chemical and nutritional composition are scarce in Ethiopia, limiting our understanding of this valuable crop.

Some efforts have been made so far on nutritional compositions and showing black cumin as a good source of carbohydrates (29.18%), protein (18.09%), and lipids (32.74%) (Mamun and Absar, 2018). Besides, several studies were made involving: analysis of secondary metabolites showed the presence of alkaloids (10.11%), flavonoids (3.78%), saponins (7.58%) and tannins (2.21%) in Bangladesh (Mamun and Absar, 2018); analysis of oleoresin yield (24.79 - 28.40%) in Ethiopia (Ermias *et al*., 2015); analysis of main essential oil components such as thymoquinone (67.7%), carvacrol (8.4%), junipene (4.8%), p-cymene (2.3%), 4-Terpineol (1.9%), longipinene (0.6%), bornylacetate (0.5%) and the main unsaturated fatty acid linoleic acid (39.20 - 43.74%) followed by oleic acid (33.41 - 37.75%) in Turkey (Nimet *et al*., 2015; Gulçin and Zehra, 2018); and evaluation of agronomic characters including seed yield per hectare in Ethiopia (Ermias *et al*., 2015; Gezahegn and Sintayehu, 2016), Turkey (Gulçin and Zehra, 2018; Nimet *et al*., 2015) and Jordan (Nasri *et al*., 2007).

Genetic diversity of some black cumin genotypes has previously been studied using a variety of molecular markers, such as random amplification of polymorphic DNA (RAPD) (Aydın, 2024; Iqbal *et al*., 2011; Khan *et al*., 2017; KorehKhosravi *et al*., 2018; Sudhir, 2016), inter simple sequence repeat (ISSR) (Birhanu *et al*., 2015; KorehKhosravi *et al*., 2018; Mehri *et al*., 2022), start codon targeted (SCoT) (El-Mahrouk *et al*., 2020; Golkar and Nourbakhsh, 2019; Mirzaei and Mirzaghaderi, 2015) and sequence-related amplified polymorphism (SRAP) (Golkar and Nourbakhsh, 2019).

Existing research efforts have been fragmented, focusing on specific aspects such as nutritional compositions and secondary metabolites in different regions. While some studies have used molecular markers to analyze genetic diversity, there has been no comprehensive study using a combination of morpho-agronomic, biochemical, and molecular markers.

The lack of detailed information on black cumin genotypes hampers efforts to collect, conserve, and improve the crop. Traditional identification methods using local names complicate the identification and improvement of genotypes. To address these issues, it is essential to assess the genetic diversity, population structure, and chemical composition of Ethiopian black cumin.

This study aims to provide valuable baseline information for breeders and policymakers, enabling the development of effective crop improvement and conservation strategies. By applying morpho-agronomic, biochemical, and molecular techniques, the research will enhance our understanding of black cumin's genetic diversity and potential, ultimately benefiting growers and consumers.

The findings will serve as a resource to address major production constraints and pave the way for the improvement of black cumin cultivation in Ethiopia.

1.3. Research questions and hypotheses

1.3.1. Research Questions

This study addressed the following research questions:

1) What is the extent of genetic diversity among black cumin genotypes growing in Ethiopia in terms of morpho-agronomic traits?
2) What biochemical traits are found in black cumin genotypes growing in Ethiopia?
3) What essential oil compositions are found in black cumin genotypes growing in Ethiopia?
4) What is the extent and pattern of genetic diversity and population structure of black cumin genotypes using GBS?

1.3.2. Hypotheses

The following null hypotheses were tested:

1) There is no significant genetic diversity among black cumin genotypes growing in Ethiopia in terms of morpho-agronomic traits.
2) There is no significant biochemical trait difference among black cumin genotypes growing in Ethiopia.
3) There is no significant difference in essential oil composition among black cumin genotypes growing in Ethiopia.
4) The DNA profile of the black cumin genotypes growing in Ethiopia are similar.

1.4. Research Objectives

General Objective

The study's overall objective was to contribute towards expanding the knowledge base of black cumin to enhance the conservation, improvement, and utilization of black cumin genetic resources in Ethiopia

Specific objectives

The specific objectives are:

- To characterize black cumin genotypes using morpho-agronomic traits and determine their variability, heritability, and genetic advance,
- To characterize black cumin genotypes using biochemical traits and determine their corresponding variability in Ethiopia,
- To characterize black cumin genotypes using essential oils composition of and determine their corresponding variability in Ethiopia, and
- To estimate the extent and pattern of genetic diversity and population structure of black cumin genotypes using GBS.

Outline of the thesis

The various chapters of this thesis each address one of the aforementioned specific goals. Since the chapters were written separately and interrelated, overlaps are most likely to occur in terms of content and references among the different chapters. The chapters are organized as follows:

Chapters	Title
1	General Introduction
2	Literature Review
3	Genetic Diversity of Ethiopian Black cumin (*Nigella sativa* L.) Based on Morpho-Agronomic Characteristics
4	Biochemical Characterization of Ethiopian Black Cumin (*Nigella sativa* L.)
5	Diversity of Ethiopian Black Cumin (*Nigella sativa* L.) Based on Compositions of Essential Oil
6	Genetic Diversity and Population Structure of Ethiopian Black Cumin (*Nigella sativa* L.) as Revealed by DArTseq SNP Markers
7	Summary, Conclusion, and Recommendation

CHAPTER 2

2. Literature Review

2.1. Distribution of black cumin (*Nigella sativa* L.*)*

Black cumin is native to Southern Europe, North Africa, and Southwest Asia, and it is cultivated in many countries in the world like South Europe, India, Pakistan, Syria, Turkey, Saudi Arabia (Khare, 2004), and Ethiopia (Demel, 2000).

In Ethiopia, black cumin is found in an altitudinal range between 1500-2500 m.a.s.l (Demel, 2000). Black cumin is widely cultivated in the Amhara Region, Northern Gondar Zone, Dembia district and Oromia region, Bale Zone (Sinana, Goro, Ebisa, Lega hide, Fedis, and Ginnir districts). It also grows in Western Arsi (Kofele and Dodola districts) and Arsi Zone (Shirka, Tena, and Silitana districts) (Edeget, 2016). It is highly cultivated in Kaffa and Keficho Zones (Ermias *et al*., 2015).

2.2. Taxonomic classification and common names of black cumin

N. sativa belongs to the family Ranunculaceae and genus *Nigella* (Sultana *et al*., 2015). In Ethiopia black cumin is known by different common names in different ethnic groups such as a afo in keffa; abasuda guraacha, abaesuda guraatii, habasuuda, nugaa guraacha, nugii guraacha in Oromia; awesda in Gurage and Tigray; habasuda in Adere; karesa sawo in Gamo; and kareta lawu'a in Welayta; and tqur azmud in Amharic (Demel, 2000). However, in the rest of the world it is known by *Rabi crop* in Bangladeshi (Mamun and Absar, 2018); *fennel flower, nutmeg flower, Roman coriander, blackseed or black caraway, black sesame* in English; *Assamese - kaljeera or kolajeera, Bengali - kalojeeray, Kannada – Krishna Jeerige, Tamil -*

karumjeerakam, Hindi/Urdu - kalaunji/mangrail in India; *Chernushka* in Russian; *Ketzakh* in Hebrew; *Cörek out* in Turkish; *habbat al-barakah* in Arabic; *siyâhdâne* in Persian; *jintanhitam* in Indonesisan; *þurekot* in Bosnian (Zohray and Hopf, 2000); *nigelle de Crète, touteépice* in French; *Schwarzkümmel* in Germany; *svartkummin* in Swedish; *cominho-negro* in Portuguese; and *ajenuz, arañuel* in Spanish (http://www.ars-grin.gov/cgi-bin/npgs/html/taxon.pl?25337).

2.2. Botanical description of black cumin

In Ethiopia, one species is widely cultivated (i.e., *Nigella sativa*), and the other (i.e., *N. damascena*) is ornamental (Demel, 2000). *N. damascena* is easily distinguished from *N. sativa* by its leaf-like involucre; the British call it 'love-in-the-mist' (Jansen, 1981). *N. sativa* has flowers without an involucre and a capsule, usually with 5 locules (Figure 2.1). Whereas, *N. damascena* has flowers closely surrounded by a much-divided involucre; capsule usually with 10 locules.

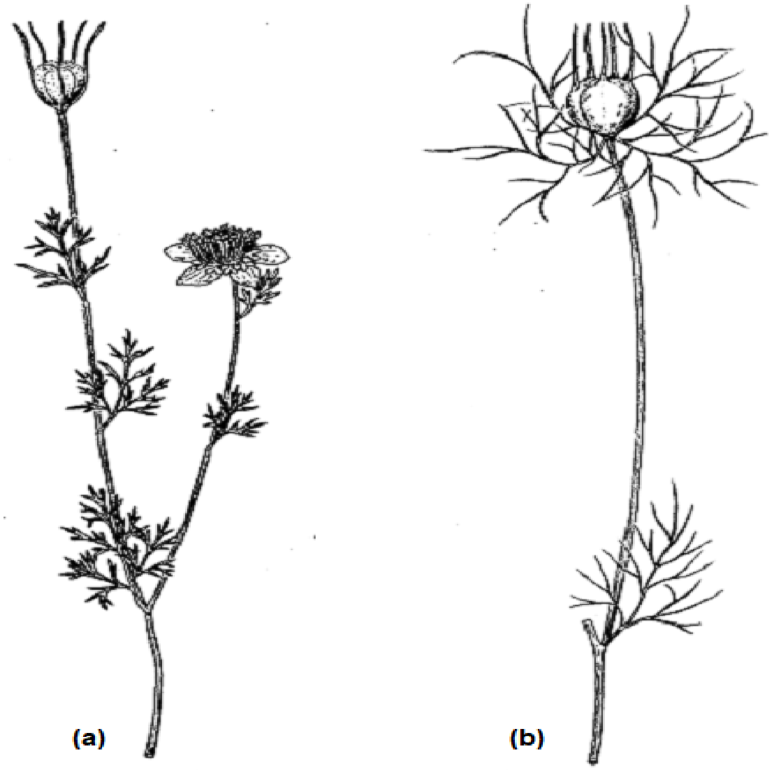


Figure 2.1. Image showing two Nigella species: a) *Nigella sativa*: - flowering and fruiting stem, from Sebsebe Demissew and Nigatu 1208; b) *N. damascena*: - the fruiting stem, from Mesfin Tadesse. 6110. Drawn by Damtew Teferra. (Source: Demel, 2000).

Black cumin is an erect annual herb attaining 30 cm to 67.6 cm (mean 52.18 cm ± 4.42) in plant height at maturity. Number of primary branches per plant ranges from 4 to 10 (mean 7.0 ± 0.71); leaf arrangement alternate, leaf phyllotaxy 1-2, pinnae of leaves broad, number of pinna per rachis 5-6; total branches per plant 22.5 ± 4.1 (6 - 48); flower hermaphrodite with determinate flowing patterns, main axis terminate with a solitary flower, delicate; flower size 2.74 cm × 2.78 cm; color - French blue (43/3 – Horticultural Color Chart) (Datta *et al*., 2012), white, yellow, pink, pale blue or pale purple (Sultana *et al*., 2015); flowers without any involucre of bracts, pedunculate; peduncle long, erect; petaloid sepals broad, ovate in a single whorl, 4-6 mostly 5 and characterized by the presence of nectarines; flower fertility 89.89%; stamens in 3 to 4 whorls, numerous (32 to 66; mean 49.6 ± 2.7) and shed their pollen as the filament bent outward during male phase; gynoecium 5, completely united follicles, each with a long indehiscent style and composed of variable number of multi ovule carpel, developing into a follicle after pollination; fruit single partially connected to form a capsule like structure (capsule: 5 to 45; mean 20.0 ± 3.37); capsule fertility: 94.5% dehiscence through junction; fruits (length: 0.4 to 1.7 cm, mean 1.03 cm ± 0.13); seta per capsule: 4 to 8, mean 5.10 ± 0.10 with numerous seeds (59.29 ± 3.2); average seed production/plant: 935 ± 177.9; seed yield/plant: 1.91 g; seed viability: 80% to 90%; seeds ovate, tetra angular, angles sharp, acute, more tapering at the end, color black; seed size: 2.33 mm ± 0.1 × 1.14 mm ± 0.02 (Datta *et al*., 2012).

Black cumin is a self-pollinating crop that releases pollen grains when the anthers reach a horizontal position. At the beginning of the male stage, stamen stand upright, curved outwards one by one, roughly in whorls, and strictly reflecting the order of initiation. The male phase of black cumin is initiated a few days before the stigmas become receptive and lasts for five days. Anther receptivity occurred between 8:00 am and 1:00 pm for a single day, and the male and female stages synchronized on the last day of flowering (Abu-Hammour and Wittmann, 2008). One of the most significant characteristics of stigmas is their capacity to promote pollen germination, a crucial step in successful fertilization that varies greatly between plant species. This capacity is known as stigmatic receptivity (Heslop-Harrison, 2000). Empty anthers curled upwards; the pollinator honey bee, one bee per flower, visits in the morning at around 7:00 am; the pollinated stigma is upright and forms an angle of 180º with the ovary; the style and anther length are almost equal to 1.73 cm; high temperatures affect the success of fertilization by influencing stigma receptivity and hastening ovule degeneration (Postweiler *et al*., 1985).

2.3. Importance and use values of black cumin

2.3.1. Medicinal uses

The seeds of black cumin are used in the treatment of various diseases like bronchitis, diarrhea, rheumatism, asthma, and skin disorders. It acts as a liver tonic, appetite stimulant, and emmenagogue. It is used in digestive disorders, to increase milk production in nursing mothers, to fight parasitic infections, and to enhance the immune system (Al-Ali *et al*., 2008). Seeds are also used in food as flavoring additives in bread and pickles because it has very low levels of toxicity (Yarnell and Abascal, 2011). Seeds are useful in the treatment of worms and skin eruptions. The oil is used as an antiseptic and local anesthetic externally. Roasted black cumin seeds are given internally to stop vomiting (Morsi, 2000).

2.3.2. Food applications

Black cumin is used as a food preservative, mainly by oxidation processes or by microorganisms, during processing, storage, and marketing is an important issue in the food industry. The food industry has used synthetic additives, which reduce microbial growth and delay the oxidation of oxidizable materials, such as lipids. Nevertheless, owing to the economic impact of spoiled foods and consumers' growing concerns over the safety of foods containing synthetic antioxidants, much attention has been paid to natural bioactive compounds (Alzoreky and Nakahara, 2003; Viuda-Marto *et al*., 2011). Black cumin is predominantly consumed in Africa and the Far East (Herms, 2015).

Black cumin seeds are added as a spice to a variety of foods such as yogurt, pickles, sauces, and salads (Hajhashemi *et al*., 2004; Venkatachallam *et al*., 2010). Seeds are used in the preparation of a traditional sweet dish and eaten with honey and syrup as well as for sprinkling on bread

(Cheikh-Rouhou *et al*., 2007; Hamrouni-sellami *et al*., 2008). Seeds are of importance as a carminative; often, they are used as a condiment in bread and other dishes (Burits and Bucar, 2000; Ramadan, 2007). Volatile and fixed black cumin seed oil is used worldwide for functional foods and nutraceuticals (Hassanien *et al*., 2015).

2.4. Biochemical composition and nutritional profile of black cumin

Black cumin is one of the essential oil-bearing seed spices. Essential oils are characterized by a strong odor and are formed by medicinal or aromatic plants as secondary metabolites. Recovery of essential oils could be obtained by expression, fermentation, or extraction. Essential oils are usually obtained by hydro-distillation, steam distillation, or dry distillation of a plant or some parts of it (Bakkali *et al*., 2008).

The most important active constituents are thymoquinone (30 - 48%), thymohydroquinone, dithymoquinone, *p*-cymene (7 - 15%), carvacrol (6 - 12%), 4-terpineol (2 - 7%), t-anethol (1 - 4%), sesquiterpene longifolene (1 - 8%), α-pinene and thymol, etc. Seeds contain two different types of alkaloids; i.e. isoquinoline alkaloids e.g. nigellicimine and nigellicimine- N-oxide, and pyrazol alkaloids or indazole ring-bearing alkaloids which include nigellidine and nigellicine. Moreover, *N. sativa* seeds also contain α-hederin, a water-soluble pentacyclic triterpene, and saponin, a potential anticancer agent (Al-Jassir, 1992).

The seeds of black cumin contain fatty oil rich in unsaturated fatty acids, mainly 50 - 60% linoleic acid, 20% oleic acid, 3% eicodadienoic acid, and 10% dihomolinoleic acid; about 30% or less saturated fatty acids (palmitic; stearic acid). Alpha-sitosterol is a major sterol, which accounts for 44% and 54% of the total sterols in Tunisian stigmasterol (6.57 - 20.92% of total sterols) (Bourgou *et al*., 2008).

2.5. Assessment of genetic diversity

Diversity is the essence of the biological world. No two living things (even maternal twins) are exactly similar to each other. The difference in one or a few traits of the organism is referred to as variability. In common phrasing, genetic variability and genetic diversity are considered synonyms to each other, which is mistaken. Genetic variability is the variation in alleles of genes or the variation in DNA/RNA sequences in the gene pool of a species or population. This expresses itself in terms of alternate forms in phenotype. Conversely, genetic diversity is a broad term encompassing all the variability occurring among different genotypes concerning the total genetic makeup of genotypes related to a single species or between species. Genetic diversity is the base for the survival of plants in nature and for crop improvement (Bhandari *et al*., 2017). Diversity in plant genetic resources provides an opportunity for plant breeders to develop new and improved cultivars with desirable characteristics, which include both farmer-preferred traits (high yield potential, large seed, etc.) and breeder-preferred traits (pest and disease resistance and photosensitivity, etc.).

Genetic resources of plants encompass a diversity of genetic material limited to traditional varieties, modern cultivars crop wild relatives, and other wild species. Genetic diversity provides options to develop new and more productive crops that are resistant to pests and diseases and adaptive to changing environments (Rao, 2004). Zhang *et al*. (1993) report, that measurement and characterization of genetic diversity has always been a primary concern in population genetic studies. Understanding the genetic variation within and among populations is crucial for the effective conservation, management, and efficient utilization of plant genetic resources. Adequate knowledge of existing genetic diversity in plant populations is of fundamental interest to basic science and applied aspects. The ability to identify genetic diversity is crucial to the effective management and utilization of genetic resources (Mondini *et al*., 2009). Thus, genetic resource conservation activities require characterization of the germplasm (Karp *et al*., 1997), which is performed using genetic markers (Mondini *et al*., 2009). According to Collard *et al*. (2005), genetic markers are broadly categorized into three major types: morphological markers, biochemical markers, and molecular markers.

2.5.1. Morphological markers

The traditional approach to variety identification is composed of the observation and the recording of morphological characters or descriptors (Nováková *et al*., 2010). Morphological markers are often multigenic, continuously expressed, and influenced by environmental interactions, making it difficult to assess them quickly and objectively; and requiring replication of observation Mba and Thome (2005) quoting as cited by Nováková *et al*. (2010).

Variations in qualitative morphological traits such as texture, color, and growth habit, or quantitative traits like yield potential, height, size, weight, etc. are used for the assessment of genetic diversity in crops. According to Rao (2004), morphological traits have their basis in genetic alterations that lead to visible differences in the phenotype. The morphological characterization does not require expensive technology. Using morphological traits alone is often undesirable but cannot be replaced by any biochemical or molecular techniques (Smith and Smith, 1989). According to Karp *et al* (1997), the results of molecular studies should be considered complementary to morphological characterization.

2.5.2. Molecular markers

Molecular markers can be classified into different groups based on the mode of transmission (biparental nuclear inheritance, maternal nuclear inheritance, maternal organelle inheritance, or paternal organelle inheritance), the mode of gene action (dominant or co-dominant markers), and the method of analysis (hybridization-based or PCR-based markers). These molecular markers include hybridization-based markers such as restriction fragment length polymorphism (RFLP); PCR-based markers like randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), inter-simple sequence repeats (ISSRs), and microsatellite or simple sequence repeats (SSRs), and sequence-based markers like single nucleotide polymorphism (SNP) (Gupta and Varshney, 2000).

Molecular markers and high-throughput genome sequencing initiatives have made it possible to characterize genetic diversity in the germplasm pool for almost any crop species with more accuracy and insight (Moose and Mumm, 2008). Since molecular markers are unaffected by the environment and the stage of plant development, and typically found in non-coding regions of DNA, they are selectively neutral. All phases of plant growth exhibit their presence (Winter and Kahl, 1995). In plant genetics, molecular markers are used for many purposes, including determining cultivar identity and the degree of genetic diversity in germplasm. By conducting selection on molecular markers associated with the desired trait rather than the trait itself, they present a significant opportunity to increase the effectiveness of traditional plant breeding (Mohan *et al*., 1997). They are effective diagnostic instruments that identify DNA polymorphism at the locus and entire genome levels (Laurentin, 2009). According to Agarwal *et al.* (2008), the most crucial factors in choosing a molecular marker type should be its informativeness and ease of genotyping for the particular crop system.

All types of molecular marker assays have different properties. An ideal molecular marker should possess the following features: (1) evenly distributed throughout the genome; (2) provide adequate resolution of genetic differences; (3) generate multiple, independent, and reliable markers; (4) be simple, quick and inexpensive; (5) need small amounts of tissue and DNA samples; (6) have linkage to distinct phenotypes and (7) require no prior information about the genome of an organism. Nevertheless, no molecular marker presents all the listed advantages. Thus, the most important criteria to determine the type of molecular marker should be informativeness and ease of genotyping for the specific crop system (Agarwal *et al*., 2008; Powell *et al*., 1996).

CHAPTER 3

3. Genetic Diversity of Ethiopian Black cumin (*Nigella sativa* L.) Based on Morpho-Agronomic Characteristics

Abstract

This study was conducted to investigate the variability of Ethiopian black cumin genotypes by using morpho-agronomic traits. Sixty-four genotypes were tested at Debre Zeit and Kulumsa Agricultural Research Center in 2021 using an 8×8 simple lattice design with two replications. Analysis of variance revealed significant differences among the genotypes for all traits studied except the number of primary branches per plant. The effect of location was significant for all traits except the number of primary branches per plant. It is expected to improve all phenological traits as well as seed yield and yield-related qualitative traits by 4 to 41% over improved varieties by selecting the top 5% of black cumin accessions. Thus, through selection, it would also be possible to shorten the flowering and maturity periods of the genotypes. High broad-sense heritability values coupled with high to moderate genetic advance as a percentage of mean values were shown by the number of capsules per plant and plant height, indicating possibilities for improvement of these traits through selection. Plant height, number of primary branches per plant, number of seeds per capsule, and thousand-seed weight had a positive and significant phenotypic and genotypic association with seed yield per hectare. Plant height had a positive direct effect on seed yield per hectare, both phenotypically and genotypically. This would be a direct selection criterion for further improvement of the genotypes. The principal component analysis of 12 quantitative traits exhibited 81.5% of the total variance captured by the first four principal components (PCs). Days to 50% flowering, days to full blooming, days to maturity, number of seeds per plant, and seed yield per plant were the main contributor traits for the variation in the first and second PCs. The genotypes were grouped into three different clusters (C-I = 35.93%, C-II = 9.38%, and C-III = 54.69%) based on 12 quantitative traits with significant inter-cluster distances. This clearly showed that there was sufficient diversity among the genotypes, which can be exploited for the future black cumin improvement program.

Keywords: Cluster analysis, Diversity, Heritability, *Nigella sativa* L., Principal component analysis

3.1. Introduction

Genetic diversity can be described as the degree of differentiation between or within species (Bhandari *et al*., 2017). Plant genetic resources allow plant breeders to develop new and improved cultivars with desirable characteristics, including farmer and breeder-preferred traits (Govindaraj *et al*., 2015). In different countries, genetic diversity studies of black cumin have been conducted using morphological traits such as plant height, number of pods per plant, number of branches per plant, seed yield per hectare, etc. (Amdie and Teshome, 2021; Demis *et al*., 2023; Ermias and Addis, 2015; Gezahegn and Sintayehu, 2016; Nimet *et al*., 2015; Tewodros *et al*., 2018).

In this study, *N. sativa* seeds of the genotypes were planted in two different agroecological areas of the country to measure their diversity in terms of traits using univariate, bivariate, and multivariate methods.

Despite its invaluable spice and medicinal uses, there have been very few genetic diversity studies on the Ethiopian black cumin genotypes, particularly using morpho-agronomic characters. Therefore, this study aims to investigate the variability of Ethiopian black cumin genotypes using morpho-agronomic traits.

3.2. Materials and Methods

3.2.1. Planting materials

Seeds of 64 black cumin genotypes (44 from Debre Zeit Agricultural Research Center (DZARC) and 20 from Sinana Agricultural Research Center (SARC)) were used for this experiment

(Appendix Table 1). These genotypes were the improved varieties and those collected from potential black cumin growing areas of Ethiopia covering different parts of the country such as Oromia, Amhara, Tigray, Benshangul-Gumuz, and South West Ethiopia Peoples regions. The individual genotypes acquired were initially collected in collaboration with Ethiopian Biodiversity Institute (EBI) staff from sites at least 5 km apart from each other unless they were clearly distinguished morphologically by their appearance.

3.2.2. Description of experimental sites

The experiment was conducted at Debre Zeit and Kulumsa Agricultural Research Centers` experimental sites under field conditions in Ethiopia during the 2021 cropping season.

Debre Zeit Agricultural Research Center is located about 47 km East of Addis Ababa, in East Shewa Zone, Ada district, 08^{o}46'04" N latitude and 39^{o}00'08" E longitude at an altitude of 1865 m.a.s.l (meter above sea level). The area receives an average annual rainfall of 851 mm while having minimum and maximum temperatures of 19.03 °C and 26.91 ^{o}C, respectively. The soil type of the experimental area is black soil or Vertisol, which has a high water-holding capacity (Tewodros *et al*., 2018).

Kulumsa Agricultural Research Center (KARC) is located about 170 km South of Addis Ababa, in Oromia National Regional State, Arsi Administrative Zone, Tiyo district, 8^{o}01'04' N latitude and 39^{o}09'24" E longitude at an altitude of 2177 m.a.s.l. The area receives an average annual rainfall of 820 mm while having minimum and maximum temperatures of 10.5 °C and 22.8 ^{o}C, respectively (Zerihun *et al*., 2018). The soil type of the experimental area is Vertic Luvisol (Sahlemedhin *et al.*, 2003). The meteorological data and site maps of both testing locations are illustrated in Figures 3.1 and 3.2, respectively. The site maps were made by using ArcGIS Desktop Advanced version 10.8 (ESRI, 2011).

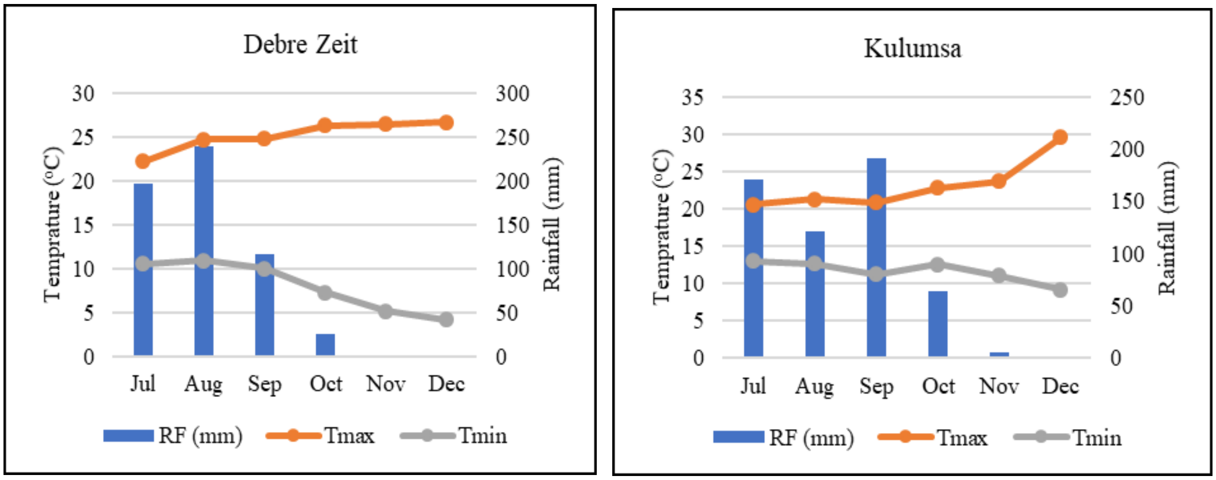


Figure 3.1. Monthly total rainfall and average temperature (maximum and minimum) at Debre Zeit and Kulumsa during 2021.

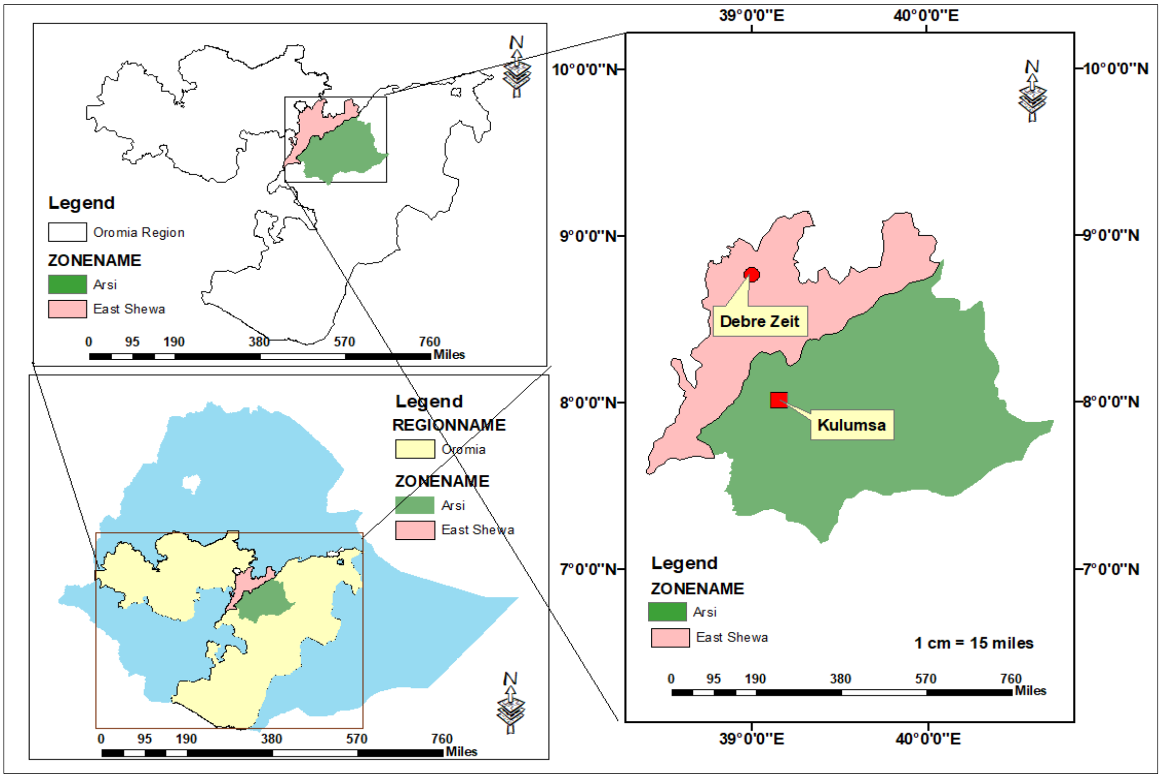


Figure 2.2. Map of Ethiopia, Oromia Region, and East Shewa and Arsi Zones showing where the testing locations are.

3.2.3. Treatments and experimental procedures

For this experiment, an 8 × 8 simple lattice design was used. The layout and randomization were done as per the standard procedure set by Cochran and Cox (1957). A distance of 30 cm was kept between rows in the same plot, 1 m between plots and blocks in the same replication, and 2 m between replications. The plots had a size of 2 m × 1.5 m (length × width). The seeds of 64 black cumin genotypes were used for this experiment. Genotypes were comprised of 18 from Oromia, 24 from Amhara, 6 from Tigray, 4 from Benshangul-Gumuz, 4 from Southern Nations, Nationalities and Peoples Regional (SNNPR) State, and 8 improved varieties (Appendix Table 1). The seeds of each genotype were sown directly on the field at a soil depth of 3 cm after the field was prepared well with a seed rate of 15 kg/ha. Sowing took place at the commencement of the main rainy season. Appropriate crop management practices, like weeding and hoeing, were applied as required. Thinning was done when the plants reached a height of 15 cm or at the true leaf stage. The R package “agricolae” (De Mendiburu, 2014) was used for the randomization of treatments using R-software version 4.2.2 (R Core Team, 2022).

3.2.4. Data collection

Morpho-agronomic data

Five randomly selected plants were used for recording quantitative character data for every genotype, except days to flowering, days to 50% flowering, days to full blooming, and days to maturity, which were recorded on a plot basis. However, the most frequent character state was recorded for the qualitative characters (Appendix Table 3). During this experiment, the following parameters were measured:

Days to flowering: The number of days from sowing to flowering of at least one plant in the row bear flower was recorded carefully.

Days to 50% flowering: The number of days from sowing to 50% of the plants in the row bear at least one open flower was recorded.

Days to full blooming: The number of days from sowing till 100% of the plants in the row bear flower.

Days to maturity: The number of days from sowing to all plants in a plot matured fully was recorded.

Plant height (cm): When the plants reached their maximum growth stage, the height of the main stem from ground level to the tip of each sampled plant was measured using a tape meter, and the average height was determined.

Number of primary branches per plant: This was determined by counting the number of primary productive branches when the plants reached the maximum growth stage and determining the average branch number.

Number of capsules per plant: When they matured, the capsules of each sampled plant from the central rows of each plot were counted carefully, and the average capsule number per plant was determined.

Number of seeds per capsule: The number of seeds from the five randomly selected matured capsules of the sampled plant was counted carefully, and the average number of seeds per capsule was determined.

Number of seeds per plant: The number of seeds from the capsules of each sampled plant was counted carefully, and the average number of seeds per plant was determined.

Seed yield per plant (g): The seeds of five randomly selected plants from each genotype were collected from the capsules when dried and weighed carefully using a sensitive analytical electronic balance. Then, the average seed yield per plant was determined.

1000-seed weight (g): This was estimated by taking 1000 seeds randomly from the composite sample of each genotype five times and weighing them using a sensitive analytical electronic balance, and then the average value was determined.

Seed yield per plot (kg): This was determined by harvesting the matured capsules from the net plot area by excluding the borders, threshed, cleaned, and weighted critically using a sensitive analytical electronic balance**.**

Seed yield per hectare (t): This was determined by the following formula after seed yield per plot was estimated carefully:

$$\text{Seed yield per hectare (t)} = \frac{\text{Seed yield per plot (kg)}\ x\ 10\ m^2}{\text{Harvested}\ plot\ area\ (m^2)} \tag{3.1}$$

3.2.5. Data Analysis

3.2.5.1. Univariate analysis

Bartlett’s test was used to assess the homogeneity of error variances for each of the measured traits to determine the validity of the individual experiment before performing the analysis. Accordingly, most of the parameters across the two locations were homogeneous. Then, the combined analysis of variance of the 12 quantitative traits data across the two locations was performed based on the model for simple lattice design following Gomez and Gomez (1984) by using SAS version 9.4 (SAS Institute Inc. 2019). Analysis of variance (ANOVA) was performed by PROC GLM procedure to manage the imbalance of the treatments in the combined analysis. The Least Significant Difference (LSD) test procedure at a 5% level of significance was used to identify significant differences among means of genotypes. During the analysis genotypes were treated as fixed variables, while the replications, blocks, and locations were considered random variables. The following model was used to make ANOVA for a simple lattice design:

$$Y_{ijkl} = \mu + \alpha_i + \beta_j(i) + \gamma_k(ij) + \delta_l + (\delta\alpha)_{li} + \varepsilon_{ijkl} \quad (3.2)$$

Where, Y_{ijkl} = the response trait Y measured at i^{th} location under j^{th} replication within location i under k^{th} incomplete block within location i and replication j in l^{th} genotype, μ = overall mean, α_i = the effect of location i, $\beta_j(i)$ = the effect of replication j within location i, $\gamma_k(ij)$ = the effect of block k within location i and replication j, δ_l = the effect of genotype l, $(\delta\alpha)_{li}$ = the interaction effect between genotype and location, and ε_{ijkl}= the effect of intra-block error.

Estimations of variance components

Variance components partitioning into genotypic variance (σ^2_G), phenotypic variance (σ^2_P), environmental variance (σ^2_E), a genotype by environment interaction (σ^2_{GxL}). As suggested by Singh and Chaudhary (2005) and Johnson *et al*. (1955), this was performed from the analyses of variance by assuming various observed mean squares equal to their expected mean squares. The following formulas were used to estimate these variance components.

$$\sigma^2_G = \frac{MS_G - MS_{GxL}}{rl} \quad (3.3)$$

$$\sigma^2_{GxL} = MS_{GxL} - MSe \quad (3.4)$$

$$\sigma^2_E = \frac{MSe}{r} \quad (3.5)$$

$$\sigma^2_P = \sigma^2_G + \frac{\sigma^2_{GxL}}{l} + \frac{\sigma^2_E}{rl} \quad (3.6)$$

Where: σ^2_G= genotypic variance, σ^2_P= phenotypic variance, σ^2_E= environmental variance, MS_G = mean square of genotypes, MS_e = error mean square, MS_{GxL} = mean square of genotypes by environmental interaction, σ^2_{GxL} = variance of genotypes by environmental interaction, r = the number of replications and l = number of locations.

Based on these, the phenotypic (PCV) and genotypic (GCV) coefficients of variation were computed as follows:

$$\text{Genotypic coefficient of variation (GCV)} = \frac{\sqrt{\sigma^2{}_{\mathrm{G}}}}{\bar{x}} \times 100 \qquad (3.7)$$

$$\text{Phenotypic coefficient of variation (PCV)} = \frac{\sqrt{\sigma^2{}_{\mathrm{P}}}}{\bar{x}} \times 100 \qquad (3.8)$$

Where: $\sigma^2{}_{G}$= genotypic variance, $\sigma^2{}_{\mathrm{P}}$= phenotypic variance, and $\bar{x}$ = grand mean of a trait.

Estimation of broad sense heritability

Broad sense heritability ($H_b{}^2$) of all the traits was determined by the following formula as described by Johnson *et al*. (1955) and Hanson *et al*. (1956).

$$\mathrm{H}_b{}^2(\%) = \frac{\sigma^2{}_{\mathrm{G}}}{\sigma^2{}_{\mathrm{P}}} \times 100 \qquad (3.9)$$

Where: $H_b{}^2$= broad sense heritability, $\sigma^2{}_{G}$= genotypic variance, and $\sigma^2{}_{\mathrm{P}}$= phenotypic variance.

Estimation of genetic advance

Genetic advance (GA) was determined by the following formula as described by Johnson *et al*. (1955).

$$\mathrm{GA} = \mathrm{K}\,(\sigma_{\mathrm{P}})\,\mathrm{H}_b{}^2 \qquad (3.10)$$

Where: K = the selection differential (K = 2.06 at 5% selection intensity), σ_{p}= the phenotypic standard deviation of the trait, and $H_b{}^2$= broad sense heritability.

The genetic advance as a percentage of the mean (GAM) was calculated by the formula as described by Johnson *et al*. (1955) as follows:

$$GAM\ (\%) = \frac{GA}{\bar{x}} \times 100 \qquad (3.11)$$

Where: GAM = genetic advance as a percentage of the mean, GA= genetic advance, and $\bar{x}$= grand mean of a trait.

Descriptive statistics was used to estimate the variations among populations carried out using the SAS version 9.4 software statistical package (SAS Institute Inc., 2019).

3.2.5.2. Bivariate analysis

The standard procedures suggested by Johnson et al. (1955), and Singh and Chaudhary (1985) were used to compute the genotypic and phenotypic correlation coefficients using the following formula:

$$r_{gxy} = Cov_{gxy}/\sqrt{(\sigma_{gx}^2 x \sigma_{gy}^2)} \qquad (3.12)$$

$$r_{pxy} = Cov_{pxy}/\sqrt{(\sigma_{px}^2 x \sigma_{py}^2)} \qquad (3.13)$$

Where, r_{gxy} = genotypic correlation coefficient between characters x and y, Cov_{gxy} = genotypic covariance between characters x and y, $\boldsymbol{\sigma}^2_{gx}$= genotypic variance for character x and $\boldsymbol{\sigma}^2_{gy}$ = genotypic variance for character y, r_{pxy} = phenotypic correlation coefficient between characters x and y, Cov_{pxy} = phenotypic covariance between characters x and y, $\boldsymbol{\sigma}^2_{px}$= phenotypic variance for character x and $\boldsymbol{\sigma}^2_{px}$ = phenotypic variance for character y. The correlation coefficients were tested using the ‘r’ tabulated value at n - 2 degrees of freedom, at 5 and 1% probability levels, where n is the number of treatments (genotypes) as described by Robertson (1959).

3.2.5.3. Multivariate analysis

Path-coefficient analysis was performed by considering seed yield per hectare as a response trait, and the remaining traits were considered independent. The direct and indirect effects of the independent traits on seed yield per hectare were estimated by a simultaneous equation using the formula as applied by Dewey and Lu (1959).

$$r_{ij} = p_{ij} + \sum r_{ik}\, p_{kj} \qquad (3.14)$$

Where, r_{ij} = mutual association between the independent trait (i) and dependent traits (j) as measured by correlation coefficients; p_{ij} = components of direct effects of the independent traits (i) on the dependent traits (j); $\sum r_{ik}\, p_{kj}$ = summation of components of the indirect effect of a given independent trait (i) on the dependent traits (j) via all other k independent traits.

The mean values of 12 quantitative traits were subjected to multivariate analysis methods, including principal component analysis (PCA) and cluster analysis (CA) using R-software version 4.2.2 (R Core Team, 2022). The dataset of recorded quantitative traits was standardized to means of zero and variances of unity before performing consequent analysis to avoid bias due to differences in measurement scales (Manly and Alberto, 2016).

Hierarchical cluster analysis was performed by using R-software program packages "phylogram" (Wilkinson and Davy, 2018), "factoextra" (Kassambara and Mundt, 2020), "cluster" (Maechler *et al*., 2018), and "igraph" (Csardi and Nepusz, 2006). Clustering was used to examine the aggregation patterns of the 64 black cumin genotypes based on their similarity concerning the corresponding means of all the quantitative traits that were collected. The best number of clusters for the data set was determined by using an R-software package "Nbclust" (Charrad *et al*., 2014). The genetic distance was measured using Euclidean distance and the distance matrix was used to construct the dendrograms using Ward`s minimum variance method (Ward, 1963). Intra- and inter-cluster distances were calculated based on the standardized Mahalanobis's D2 statistics (Mahalanobis, 1936) using R-software packages "biotools" (Da Silva, 2021) and "MASS" (Venables and Ripley, 2002).

For the groups of collections, hierarchical cluster analysis was run using the regional means for the 12 quantitative traits. The measure of dissimilarity was Euclidean distance and the clustering method Ward`s minimum variance method (Ward, 1963).

The PCA was employed by using R-software packages “factoextra”, “ggplot2” (Wickham, 2016), “corrplot” (Wei and Sinko, 2021), and “ggsignif” (Ahlmann-Eltze, 2022) to identify the traits contributing to a large part of the total variation among the populations and groups.

3.3. Results and Discussion

3.3.1. Morpho-agronomic traits performance of the genotypes

Combined analysis of variance over the two locations revealed a significant ($p < 0.001$) effect for location and genotypes in all quantitative traits measured, except the number of primary branches per plant. A similar result was reported by Arega and Solomon (2021) on the number of capsules per plant and seed yield per hectare of black cumin and Demis *et al*. (2023) on black cumin seed yield per hectare. Location × genotype interaction effects were also significant ($p < 0.001$) in all the studied quantitative traits except the number of primary branches per plant (Appendix Table 2). Supporting results were reported by Demis *et al*. (2023) on black cumin seed yield per hectare. In the case of the 15 qualitative traits measured, there was no difference among 64 genotypes of Ethiopian black cumin (Figure 3.3, Appendix Table 3).

Separate analyses of variance showed significant differences among the genotypes for all the measured traits, except for the number of primary branches per plant in the case of both sites and the thousand seed weight in the case of Kulumsa only, which exhibited non-significant differences. This result is in line with the findings of Arega and Solomon (2021) on the number of capsules per plant and seed yield per hectare of black cumin and Demis *et al*. (2023) on black cumin seed yield per hectare. Similarly, combined analyses of variance exhibited significant differences in all parameters studied except the number of primary branches per plant.

Significant variability was also detected from the wider ranges observed between the minimum and the maximum values of the traits measured (Table 3.1).

The variability of the genotypes for days to maturity creates a great opportunity for developing early-maturing varieties by improving the traits that are associated with days to maturity. This result would help the breeders to develop improved and suitable varieties for different agro-ecologies of the country.

Table 3.1. Descriptive statistics, F-test, and coefficient of variation of 12 quantitative traits of the 64 black cumin genotypes of Ethiopia at each testing site and pooled during the 2021 cropping season

Trait	Debre Zeit				Kulumsa				Combined			
	Mean ± SE	Range	F-Test	CV (%)	Mean ± SE	Range	F-Test	CV (%)	Mean ± SE	Range	F-Test	CV (%)
DF	58.4 ± 0.20	56-80	*	3.34	86.02 ± 0.15	77-87	***	1.03	72.2 ± 0.87	56-87	***	2.08
DFPF	74.5 ± 0.02	70-78	***	0.47	98.02 ± 0.16	92-99	***	0.22	86.3 ± 0.75	70-99	***	0.34
DFB	80.1 ± 0.14	76-82	***	0.31	107.3 ± 0.26	93-109	***	1.32	93.7 ± 0.86	76-109	***	1.08
DM	143.2 ± 0.29	136-148	***	0.44	159.4 ± 0.22	154-161	***	0.33	151.3 ± 0.54	136-161	***	0.39
PH	53.9 ± 0.56	32-68	**	9.87	59.8 ± 0.46	47-72	***	4.03	56.8 ± 0.41	32-72	***	7.43
NBP	4.6 ± 0.07	4-7	ns	16.52	5.4 ± 0.09	4-9	ns	17.51	5.1 ± 0.06	4-9	ns	16.72
NCP	9.6 ± 0.18	4-16	***	15.06	14.6 ± 0.33	7-26	*	22.04	12.1 ± 0.24	4-26	***	20.12
NSC	90.6 ± 0.69	67-108	***	6.23	105.9 ± 0.72	87-127	*	6.75	98.2 ± 0.69	67-127	***	6.48
NSP	869.3 ± 18.43	387-1725	***	16.92	1545 ± 37.07	590-2788	**	22.14	1207.3 ± 29.6	387-2788	***	21.11
TSW	2.31 ± 0.01	2.1-2.6	***	3.54	2.10 ± 0.01	1.9-2.3	ns	3.7	2.21 ± 0.01	1.9-2.6	**	3.3
SYP	2.02 ± 0.04	1.08-4	***	13.05	3.24 ± 0.08	1.18-5.86	*	23.26	2.63 ± 0.06	1.18-5.08	***	20.11
SYH	1.18 ± 0.03	0.24-1.93	***	14.83	0.58 ± 0.02	0.19-1	***	10.01	0.88 ± 0.02	0.19-1.93	***	15.25

Where SE= Standard error, DF= Days to flowering, DFPF= Days to 50% flowering, DFB= Days to full blooming, DM= Days to maturity, PH=Plant height (cm), NBP= Number of primary branches per plant, NCP= Number of capsules per plant, NSC= Number of seeds per capsule, NSP= Number of seeds per plant, TSW= Thousand seed weight (g), SYP= Seed yield per plant (g), SYH= Seed yield per hectare (t). ns = non-significant at $p \geq 0.05$, *, ** and *** = Significant at $p < 0.05$, at $p < 0.01$ and at $p < 0.001$, respectively.

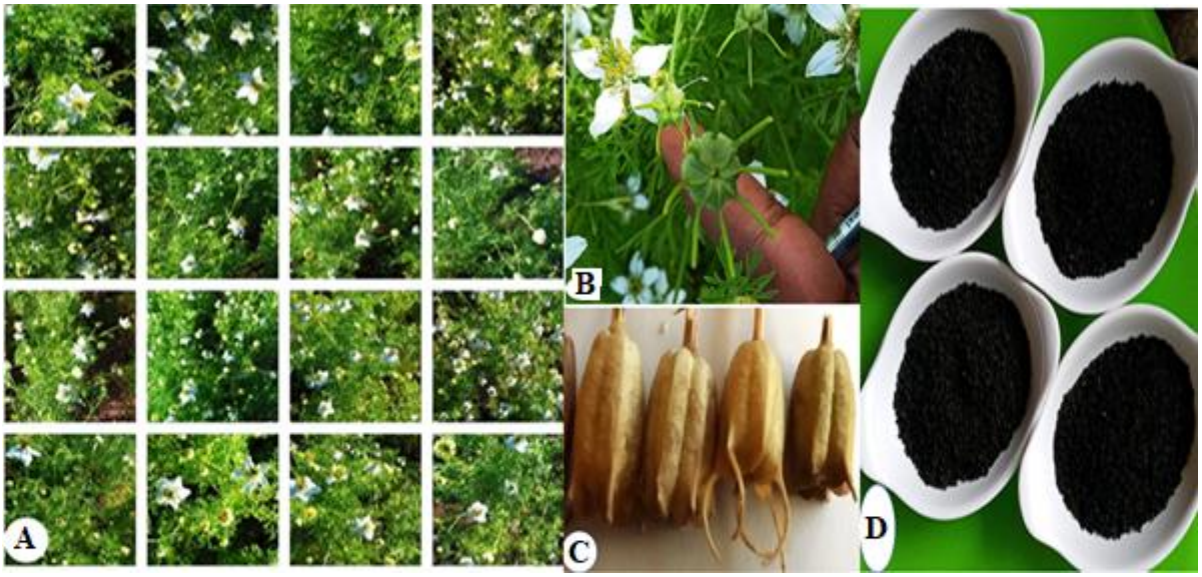


Figure 3.3. Images showing: A) the morphological appearance, B) the leaf, flower, and immature capsule, C) matured capsules, and D) seeds of some Ethiopian black cumin genotypes after threshing.

Phenological attributes

Based on the combined result presented in Table 3.1, there was a highly significant ($p < 0.001$) difference among genotypes for days to flowering, days to 50% flowering, days to full blooming, and days to maturity with wider ranges between the minimum and maximum values. Tewodros *et al*. (2018) reported supporting data for days to flowering, days to 50% flowering, and days to maturity of black cumin, as did Kedir *et al.* (2025) for black cumin days to 50% flowering.

The days to flowering performance ranged from 56 days to 87 days, with a mean value of 72.2 days. The highest mean days to flowering were obtained from genotype 219970 (from Tigray), whereas the lowest mean value was obtained equally from five genotypes, namely 90504 and

237989 (from Oromia), 230040 (from Tigray), 242225 and 002_ATH (from Amhara) (Appendix Table 4).

The days to 50% flowering performance ranged from 70 days to 99 days, with a mean value of 86.3 days. The highest mean days to 50% flowering was obtained from genotype 242841 (from Oromia) followed by 9069, 215319, 003_ATH, 010_ATH, and 014_ATH (from Amhara); whereas, the least mean value was obtained from genotype 237989 followed by 90504 (from Oromia), 230040 (from Tigray), 242225 and 002_ATH (from Amhara) (Appendix Table 4).

The days to full blooming performance ranged from 76 days to 109 days, with a mean value of 93.7 days. The highest mean days to full blooming were obtained from genotype 242841 (from Oromia) followed by 219970 (from Tigray); whereas, the least mean value was obtained equally from five genotypes, namely 90504 and 237989 (from Oromia), 230040 (from Tigray), 242225 and 002_ATH (from Amhara) (Appendix Table 4).

The days to maturity performance ranged from 136 days to 161 days, with a mean value of 151.3 days. The highest mean days to maturity were obtained from genotype 242841 (from Oromia), whereas the lowest mean value was obtained equally from three genotypes, namely 90504 and 237989 (from Oromia) and 242225 (from Amhara), followed by 230040 (from Tigray) and 002_ATH (from Amhara). This result indicates the highest and least number of days, which implies the lateness and earliness of the genotype to reach maturity, respectively. Similar results were reported by Arega and Solomon (2021) and Demis *et al*. (2023) on black cumin. Accordingly, genotype 242841 (from Oromia) matured late, whereas genotypes 90504 and 237989 (from Oromia) and 242225 (from Amhara), followed by genotype 230040 (from Tigray) and 002_ATH (from Amhara), matured earlier (Appendix Table 4). The summary statistics indicated a large variation among the tested genotypes in the phenological attributes measured.

Growth, yield, and yield-related attributes

There was a significant difference among genotypes for plant height, number of capsules per plant, number of seeds per capsule, number of seeds per plant, thousand seed weight, seed yield per plant, and seed yield per hectare, with wider ranges between the minimum and the maximum values (Table 3.1). Tewodros *et al.* (2018) found similar results for black cumin plant height, thousand seed weight, and seed yield per plant, as did Kedir *et al.* (2025) for black cumin plant height, number of capsules per plant, number of seeds per plant, seed yield per plant, and per hectare.

The seed yield performance ranged from 0.19 t/ha to 1.93 t/ha, with a mean of 0.88 t/ha. The highest mean seed yield per hectare was obtained from genotype 90504 (from Oromia), followed by genotype 219970 (from Tigray); whereas, the lowest mean value was obtained from genotype 237989 (from Oromia) (Appendix Table 4). Complementary findings were documented by Arega and Solomon (2021) as well as by Demis *et al*. (2023) regarding black cumin. This reflects the presence of ample amounts of variation among the tested genotypes. The variation in seed yield could be due to environmental, genetic, or interaction among the two factors.

The presence of variability among the black cumin genotypes was witnessed by the wide range of most of the traits studied. This showed that there is a possibility of improving these traits through selection. Therefore, selecting the top 5% of genotypes is predicted to improve quantitative traits by 4% to 41% through selection (Table 3.2). With this regard, genotype 90504 (from Oromia), 219970 (from Tigray), and 019_ATH (from Amhara) were the top 5% best-performed genotypes over improved varieties selected for seed yield per hectare (Table 3.2).

Table 3.2. Comparison of mean performances of 5% of the best-performed genotypes selected for quantitative traits over the mean performance of improved varieties

Traits	Mean values		Comparative advantage over mean values of improved Varieties (%)
	Top 5% genotypes	Improved varieties	
Days to flowering	69	71.9	- 4.03
Days to 50% flowering	81.5	86.1	- 5.34
Days to full blooming	88.5	93.5	- 5.35
Days to maturity	145	151.3	- 4.16
Plant height (cm)	63.3	59.3	6.75
Number of capsules per plant	15	12	25
Number of seeds per capsule	107	98	9.18
Number of seeds per plant	1588	1199	32.44
Thousand seed weight (g)	2.31	2.18	5.96
Seed yield per plant (g)	3.40	2.57	32.3
Seed yield per hectare (t)	1.275	0.907	40.57

3.3.2. Components of variance

The presence of variability for a given trait is mandatory for any crop improvement program. Accordingly, Table 3.3 presents the estimates of the genotypic coefficient of variation (GCV), phenotypic coefficient of variation (PCV), heritability in a broad sense (H_b^2), genetic advance (GA), and genetic advance as a percentage of the mean (GAM) for 12 quantitative traits.

The values of GCV ranged from 0.66-16.12% for days to maturity and number of seeds per plant, respectively. Likewise, the values of PCV ranged from 1.37-21.32% for days to maturity and seed yield per hectare, respectively (Table 3.3). According to Sivasubramanian and

Madhavamenon (1973), GCV and PCV values were categorized as low (0-10%), moderate (10-20%), and high (> 20%). Hence, moderate GCV and high PCV were obtained for the number of seeds per plant, and seed yield per hectare, respectively. The same result was reported by Tewodros *et al.* (2018) on black cumin. Moderate to high values of GCV and PCV indicated the presence of moderate to high variability among genotypes, respectively. Low GCV and moderate PCV were obtained for the number of capsules per plant. This could be due to low variability of genotypes in the number of capsules per plant genotypically, but moderately varied phenotypically by the influence of environment. Low GCV and PCV were obtained for days to flowering, days to 50% flowering, days to full blooming, days to maturity, plant height, number of seeds per capsule, and thousand seed weight. This is in line with the finding of Tewodros *et al.* (2018) on black cumin days to 50% flowering and days to maturity. This study showed the PCV values were greater than the GCV values for all the traits considered, which reflects the existence of environmental influence on the expression of genotypes performance. Therefore, this ensured for the breeders that there is a possibility for further improvement of the genotypes through selection. This result is supported by the finding of Kumhar *et al.* (2020) on cumin.

Broad sense heritability for the studied quantitative traits ranged from the least 8.16% for number of seeds per capsule to the highest 89.47% for thousand seed weight. As Robertson (1959) suggested, heritability was categorized as low (0-30%), moderate (30-60%) and high (> 60%). Accordingly, high heritability values were obtained on thousand seed weight (89.47%), plant height (78.3%), days to flowering (76.65%), days to full blooming (74.71%), number of capsules per plant (66.86%), and days to 50% flowering (62.56%). Supporting results were reported by Tewodros *et al.* (2018) on black cumin days to flowering and days to 50% flowering. Moderate heritability in a broad sense was recorded on the number of seeds per plant, seed yield per hectare, number of primary branches per plant, and seed yield per plant. A similar result was reported by Tewodros *et al.* (2018) on black cumin seed yield per plant. However, low broad-sense heritability values were shown by days to maturity and the number of seeds per capsule (Table 3.3). Tewodros *et al.* (2018) reported the same result on black cumin days to maturity. Heritability by itself provides no hint of the amount of genetic advancement that could be possible to make by selecting the best individuals (Johnson *et al*., 1955). GCV, together with a heritability estimate, would seem to give the best picture of the amount of advance to be expected from selection (Burton, 1952, as cited by Umamaheswar *et al*., 2024). Therefore, through selection, it would be possible to improve the number of capsules per plant, the number of seeds per plant, the seed yield per plant, and the seed yield per hectare.

Table 3.3. Estimated values of the variability of 12 quantitative traits in 64 black cumin genotypes of Ethiopia at Debre Zeit and Kulumsa in the 2021 cropping season

Traits	Mean	GCV (%)	PCV (%)	H_b^2 (%)	GA	GAM (%)
DF	72.23	1.88	2.15	76.65	2.45	3.39
DFPF	86.27	1.80	2.27	62.56	2.52	2.93
DFB	93.67	2.13	2.47	74.71	3.56	3.80
DM	151.3	0.66	1.37	23.20	0.99	0.66
PH	56.84	7.45	8.42	78.30	7.72	13.58
NBP	5.02	4.88	6.71	52.81	0.37	7.30
NCP	12.05	15.65	19.14	66.86	3.18	26.37
NSC	98.23	1.00	3.51	8.16	0.58	0.59
NSP	1207.3	16.12	20.89	59.61	309.62	25.65
TSW	2.21	4.53	4.79	89.47	0.19	8.84
SYP	2.63	13.73	19.05	51.95	0.54	20.39
SYH	0.88	16.01	21.32	56.39	0.22	24.77

Where, GCV = Genotypic coefficients of variation, PCV = phenotypic coefficients of variation, H_b^2 = Broad sense heritability, GA = Genetic advance, GAM = Genetic advance as a percentage of the mean; DF = Days to flowering, DFPF = Days to 50% flowering, DFB = Days to full blooming, DM = Days to maturity, PH = Plant height (cm), NBP = Number of primary branches per plant, NCP = Number of capsules per plant, NSC = Number of seeds per capsule, NSP = Number of seeds per plant, TSW = Thousand seed weight (g), SYP = Seed yield per plant (g), SYH = Seed yield per hectare (t).

The estimate of GA values ranged from 0.19 for thousand seed weight to 309.62 for the number of seeds per plant. High GA was recorded on the number of seeds per plant, whereas low values were shown by the rest of the quantitative traits studied. Moreover, GAM values ranged from 0.59% for the number of seeds per capsule to 26.3% for the number of capsules per plant. The value of GAM is categorized as low (0-10%), moderate (10-20%), and high (above 20%) (Johnson *et al.*, 1955). Based on this, high values of GAM were recorded on the number of capsules per plant (26.37%), number of seeds per plant (25.65%), seed yield per hectare (24.77%), and seed yield per plant (20.39%). Similar results were reported by Kumhar *et al.* (2020) on seed yield per hectare of cumin. A moderate value of GAM was recorded for plant height (13.58%). However, the rest of the quantitative traits showed low values of GAM (Table 3.3). According to Panes and Sukhatme (1995), traits that show high heritability estimates along with high GAM could be used as a powerful tool for selection. Therefore, selection based on the number of capsules per plant and seed yield per plant would be more effective than others.

3.3.3. Correlations among traits

Correlation coefficient analysis is commonly used to measure the degree and direction of relationships between 12 quantitative traits, including seed yield per hectare. Significant genotypic and phenotypic correlation coefficients were found among seed yield and other morpho-agronomic traits studied (Table 3.4).

Phenological attributes

Days to flowering had significant ($p < 0.01$) positive association with days to 50% flowering ($r_g = 0.97$, $r_p = 0.71$), days to full blooming ($r_g = 0.98$, $r_p = 0.80$), and days to maturity ($r_g = 0.94$, r_p

= 0.71). Genotypically, a similar result was reported by Zigyalew *et al.* (2020). Genotypically, days to flowering had significant ($p < 0.01$ or $p < 0.05$) positive association with plant height ($r_g = 0.25$*), number of primary branches per plant ($r_g = 0.36$**), number of seeds per capsule ($r = 0.39$**), thousand seed weight ($r_g = 0.27$*), and seed yield per plant ($r_g = 0.25$*). On the contrary, days to flowering were not significantly associated with the number of capsules per plant, number of seeds per plant, and seed yield per hectare both phenotypically and genotypically. Days to 50% flowering had a significant ($p < 0.01$) positive association with days to full blooming ($r_g = 0.93$**, $r_p = 0.85$**), and days to maturity ($r_g = 0.96$**, $r_p = 0.93$**). On the contrary, days to 50% flowering were not significantly correlated with the number of primary branches per plant, number of capsules per plant, number of seeds per plant, and seed yield per hectare both phenotypically and genotypically.

Growth, yield, and yield-related attributes

Plant height had significant ($p < 0.01$ or $p < 0.05$) positive association with the number of primary branches per plant ($r_g = 0.63$**, $r_p = 0.20$*), number of capsules per plant ($r_g = 0.54$**, $r_p = 0.24$**), number of seeds per capsule ($r_g = 0.64$**, $r_p = 0.40$**), and number of seeds per plant ($r_g = 0.51$**, $r_p = 0.33$**). This suggests that the number of major branches per plant, the number of capsules per plant, the number of seeds per capsule, and the number of seeds per plant were all directly correlated with plant height. A similar result was reported by Zigyalew *et al.* (2020) on the number of black cumin capsules per plant. On the contrary, plant height was not significantly associated with thousand seed weights.

Seed yield per plant had significant ($p < 0.01$ or $p < 0.05$) positive association with plant height ($r_g = 0.47^{**}$, $r_p = 0.33^{**}$), number of primary branches per plant ($r_g = 0.99^{**}$, $r_p = 0.20^{*}$), number of capsules per plant ($r_g = 0.99^{**}$, $r_p = 0.89^{**}$), number of seeds per capsule ($r_g = 0.93^{**}$, $r_p = 0.36^{**}$) and number of seeds per plant ($r_g = 0.99^{**}$, $r_p = 0.97^{**}$). This suggests that the plant height, the number of primary branches per plant, the number of capsules per plant, the number of seeds per capsule, and the number of seeds per plant were all directly correlated with seed yield per plant.

Seed yield per hectare had a significant ($p < 0.01$ or $p < 0.05$) positive correlation with plant height ($r_g = 0.65^{**}$, $r_p = 0.53^{**}$), number of primary branches per plant ($r_g = 0.51^{**}$, $r_p = 0.20^{*}$), number of seeds per capsule ($r_g = 0.31^{*}$, $r_p = 0.31^{**}$), and thousand seed weight ($r_g = 0.26^{*}$, $r_p = 0.23^{**}$). This indicated that these traits are the major contributing traits for seed yield improvement in black cumin genotypes. Generally, the genotypic correlation is greater than the phenotypic correlation of most quantitative traits studied. This result aligns with the finding of Tewodros *et al.* (2018) on black cumin.

Table 3.4. Genotypic (above diagonal) and phenotypic (below diagonal) correlation coefficient of the 12 quantitative traits

Traits	DF	DFPF	DFB	DM	PH	NBP
DF		0.97**	0.98**	0.94**	0.25*	0.36**
DFPF	0.71**		0.93**	0.96**	0.15	0.02
DFB	0.80**	0.85**		0.89**	0.14	0.14
DM	0.71**	0.93**	0.84**		0.12	-0.03
PH	0.14	0.09	0.08	0.06		0.63**
NBP	0.09	-0.02	0.04	0.01	0.20*	
NCP	0.06	-0.02	0.09	-0.01	0.24**	0.19*
NSC	0.13	0.14	0.07	0.12	0.40**	0.06
NSP	0.10	0.03	0.11	0.04	0.33**	0.21*
TSW	0.11	0.06	0.04	0.09	0.06	0.08
SYP	0.10	0.04	0.10	0.04	0.33**	0.20*
SYH	0.12	0.04	0.04	0.08	0.53**	0.20*

Where, DF= Days to flowering, DFPF= Days to 50% flowering, DFB= Days to full blooming, DM= Days to maturity, PH=Plant height (cm), NBP= Number of primary branches per plant.
*= Significant at p < 0.05, **= Significant at p < 0.01.

Table 3.4. Continued

Traits	NCP	NSC	NSP	TSW	SYP	SYH
DF	0.05	0.39**	0.15	0.27*	0.25*	0.22
DFPF	-0.06	0.23	0.04	0.15	0.06	0.05
DFB	0.22	0.12	0.19	0.08	0.22	0.05
DM	0.01	0.18	0.09	0.11	0.11	0.1
PH	0.54**	0.64**	0.51**	-0.08	0.47**	0.65**
NBP	0.99**	0.51**	0.99**	-0.69**	0.99**	0.51**
NCP		0.88**	0.99**	-0.55**	0.99**	0.13
NSC	0.01		0.79**	0.52**	0.93**	0.31*
NSP	0.93**	0.32**		-0.13	0.99 **	0.14
TSW	-0.02	0.26**	0.02		0.17	0.26*
SYP	0.89**	0.36**	0.97**	0.17		0.22
SYH	0.07	0.31**	0.16	0.23**	0.20*	

Where, NCP= Number of capsules per plant, NSC= Number of seeds per capsule, NSP= Number of seeds per plant, TSW= Thousand seed weight (g), SYP= Seed yield per plant (g), and SYH= Seed yield per hectare (t).
*= Significant at p < 0.05, **= Significant at p < 0.01.

3.3.4. Path coefficient analysis

The path coefficient is simply a standardized partial regression coefficient that fundamentally measures the direct impact of one trait upon another and permits the separation of the correlation coefficient into direct and indirect effects of the components other than the response trait (Dewey and Lu, 1959). The use of this method requires a cause-and-effect condition among the traits studied. Accordingly, the analysis was performed to assess the magnitude of the contributions of various traits to the response trait of seed yield per hectare.

Phenotypically, plant height had the highest positive direct effect (0.5058) on seed yield per hectare, followed by seed yield per plant (0.4084), days to maturity (0.3774), and thousand seed weight (0.1095) (Appendix Table 5). Plant height and seed yield per plant showed a significant positive correlation with seed yield per hectare. This indicated that plant height and seed yield per plant are the most influential traits on seed yield per hectare compared to the others. All the measured traits exerted indirect positive effects on seed yield per hectare via plant height and seed yield plant. This demonstrated that the relationship between these traits and seed yield per hectare was predominantly due to indirect effects. The number of seeds per capsule had a negative and low direct effect (-0.0213) on seed yield per hectare but had a significant positive correlation with seed yield per hectare (Appendix Table 5). Thousand-seed weight and the number of primary branches per plant had a positive and low direct effect on seed yield per hectare but had a significant positive correlation with seed yield per hectare. This implied that these traits might probably contribute to seed yield per hectare through an indirect effect requiring indirect selection. Based on this result, plant height and seed yield per plant are the most influential traits on seed yield per hectare.

Genotypically, days to flowering had the highest positive direct effect (2.6195) on seed yield per hectare followed by days to maturity (0.8893), plant height (0.7482), the number of primary branches per plant (0.3344), the number of seeds per plant (0.1353) and seed yield per plant (0.2814) (Appendix Table 6). Plant height and the number of primary branches per plant showed a significant positive correlation with seed yield per hectare. This indicated that plant height and the number of primary branches per plant are the most influential traits on seed yield per hectare compared to the others. Days to flowering, days to 50% flowering, days to full blooming, number of capsules per plant, number of seeds per capsule, number of seeds per capsule, number of seeds per plant and seed yield per plant had a positive indirect effect on seed yield per hectare via plant height and number of primary branches per plant. This demonstrated that the relationship between these traits and seed yield per hectare was predominantly due to indirect effects. Thousand-seed weights also had a low positive direct effect (0.0170) on seed yield per hectare but had a significant positive correlation with seed yield per hectare. The number of seeds per capsule had a low negative direct effect (-0.5329) on seed yield per hectare but had a significant positive correlation with seed yield per hectare (Appendix Table 6). This implied that these traits might probably contribute to seed yield per hectare through an indirect effect requiring indirect selection. Based on this result, plant height and the number of primary branches per plant are the most influential traits on seed yield per hectare.

In general, plant height is the most influential trait on seed yield per hectare, both phenotypically and genotypically. Therefore, this would be an important selection criterion for further improvement programs of the genotypes.

3.3.5. Diversity based on quantitative traits

3.3.5.1. Principal component analysis

Principal Components Analysis (PCA) is a multivariate statistical technique used for dimension reduction and to further an understanding of the underlying groups of variables in the data (Jolliffe, 1986). The PCA grouped the 12 morpho-agronomic traits into 12 principal components, which explained the entire 100% of the variability among the studied genotypes. The first four principal components (PCs) explained 81.5% of the variation that existed among the studied genotypes (Table 3.5).

There are different methods recommended to decide how many components to preserve. Out of these methods, an eigenvalue > 1 indicates that PCs account for more variance than accounted by one of the original variables in standardized data commonly used as a cutoff point for which PCs are retained (Kaiser, 1960). The number of components is determined at the point, beyond which the remaining eigenvalues are relatively small and of comparable size (Jolliffe, 2002; Peres-Neto *et al*., 2005) Therefore, based on the Scree plot (Figure 3.4) and eigenvalues four PCs having eigenvalues between 1.06 and 3.98, extracted a cumulative variance of about 81.5% of the total phenotypic diversity maintained (Table 3.5).

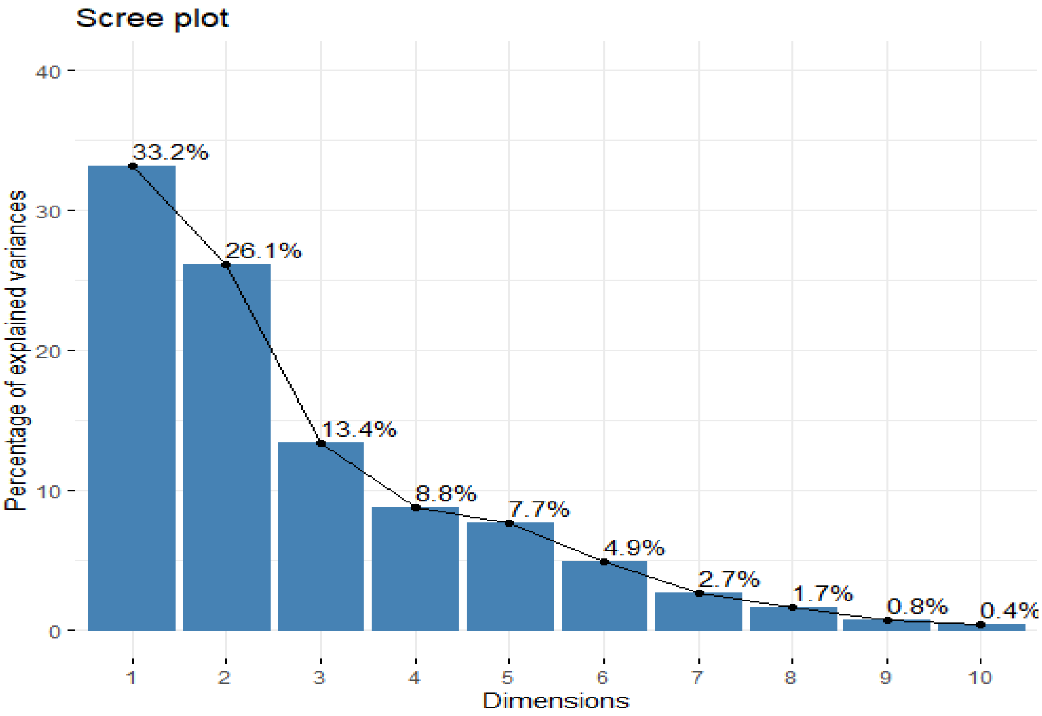


Figure 3.4. Scree plot of principal Components (PCs) for 12 quantitative characters in 64 black cumin genotypes.

Table 3.5. The eigenvalues and eigenvectors of the first four principal components (PCs) for 12 quantitative traits of 64 black cumin genotypes of Ethiopia

Variable	Principal components			
	PC1	PC2	PC3	PC4
Eigenvalue	3.98	3.13	1.61	1.06
Proportion of variance (%)	33.2	26.1	13.4	8.8
Cumulative variance (%)	33.2	59.3	72.7	81.5
		Eigenvectors		
Days to flowering	**0.39**	-0.28	-0.004	-0.07
Days to 50 % flowering	**0.36**	**-0.36**	0.03	0.03
Days to full blooming	**0.39**	**-0.31**	0.14	-0.02
Days to maturity	**0.37**	**-0.34**	0.07	0.06
Plant height (cm)	0.23	0.21	**-0.38**	**0.47**
Number of primary branches per plant	0.15	0.20	-0.05	**-0.34**
Number of capsules per plant	0.23	**0.37**	**0.42**	0.03
Number of seeds per capsule	0.24	0.18	**-0.36**	-0.02
Number of seeds per plant	**0.31**	**0.40**	0.24	0.03
Thousand seed weight (g)	0.09	0.04	**-0.37**	**-0.76**
Seed yield per plant (g)	**0.32**	**0.39**	0.16	-0.11
Seed yield per hectare (t)	0.17	0.12	**-0.55**	0.24

Chahal and Gosal (2002) suggested that the influential PCs for clustering were the characters with the largest absolute values closer to unity than those with lower absolute values closer to zero. Accordingly, due to the main contribution of the variations in days to flowering, days to full blooming, days to maturity, days to 50% flowering, seed yield per plant, and number of seeds per plant the first principal component explained up to 33.2% of the total variance (Table 3.6). The number of seeds per plant, seed yield per plant, number of capsules per plant, days to 50% flowering, days to maturity, and days to full blooming were the main contributor traits for the variation in the second PC, which contributed 26.1% of the total variations. The main contributor traits for the third PC were variations in the seed yield per plant, number of capsules per plant, plant height, thousand seed weight, and number of seeds per capsule. Moreover, thousand seed weight, plant height, and number of primary branches per plant were the main contributor traits for the variations in the fourth PC (Table 3.6). The proportion of the total phenotypic variance of the genotypes accounted for by the third and fourth PCs was about 13.4% and 8.8%, respectively (Table 3.6). Hence, the PCA confirmed that the collected Ethiopian black cumin genotypes have high diversity and most of the traits considered seemed to have high contributions toward the total phenotypic variability.

As Kassambara (2017) described, the variable correlation plot shows the relationship between all variables; the closer a variable is to the circle, the more important it is to interpret these components, and variables that are close to the center of the plot are less important for the first components. Thus, a thousand seed weight and seed yield per hectare contribute less for the first components (Figure 3.5). Traits such as days to flowering, days to full blooming, days to maturity, days to 50% flowering, seed yield per plant, and number of seeds per plant were the main contributors to the variation in the first PC; whereas, the number of seeds per plant, seed yield per plant, number of capsules per plant, days to 50% flowering, days to maturity and days to full blooming were the main contributor to the variation in the second PC (Figure 3.5).

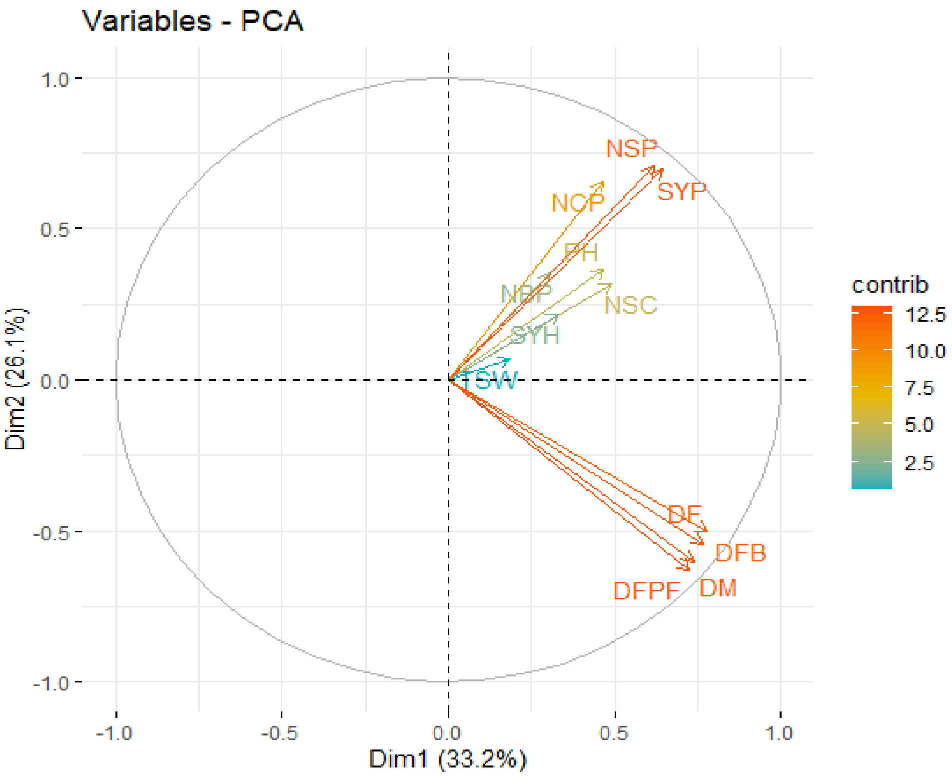


Figure 3.5. Variable correlation plot showing the relationships between 12 quantitative traits in 64 black cumin genotypes of Ethiopia.

3.3.5.2. Cluster analysis

Cluster analysis was done based on 12 quantitative traits of the 64 black cumin genotypes of Ethiopia to examine the aggregation patterns of the 64 black cumin genotypes based on their similarity. According to Charrad *et al.* (2014), the best number of clusters for the data set was determined to be three. The ward`s minimum variance clustering method classified the 64 black cumin genotypes of Ethiopia into three different clusters (Figure 3.6). The number and name of genotypes in each cluster with their collection regions were presented in Appendix Table 1.

As shown in Figure 3.6 and Table 3.6, the third cluster (C-III) was the largest group with 35 (54.69%) genotypes from all collection regions and half of the improved varieties. This group was characterized by late flowering, late maturing, and low-yielding genotypes, with a mean value of 73 days, 152 days, and 0.78 t/ha, respectively (Table 3.7). Furthermore, this group's number of seeds per plant and seed yield per plant were the lowest. However, the remaining genotypes belonged to the first (n = 23 (35.93%)) and the second (n = 6 (9.38%)) clusters. The first (C-I) and second cluster (C-II) consisted of the genotypes from the five (SNNP, Oromia, Amhara, B/Gumuz, and Tigray) and three regions (Amhara, Tigray, and Oromia), respectively and some of the improved variety/ies. The first group was characterized by the highest mean value of plant height (60 cm), number of capsules per plant (13), number of seeds per capsule (103), number of seeds per plant (1353), seed yield per plant (2.99 g), and seed yield per hectare (0.93 t). Moreover, the second cluster was characterized by early flowering and maturing genotypes, with the mean value of 69 and 145 days, respectively; and having intermediate performance on other morpho-agronomic traits studied (Table 3.7). However, the number of primary branches per plant and thousand seed weight remains constant for the three clusters.

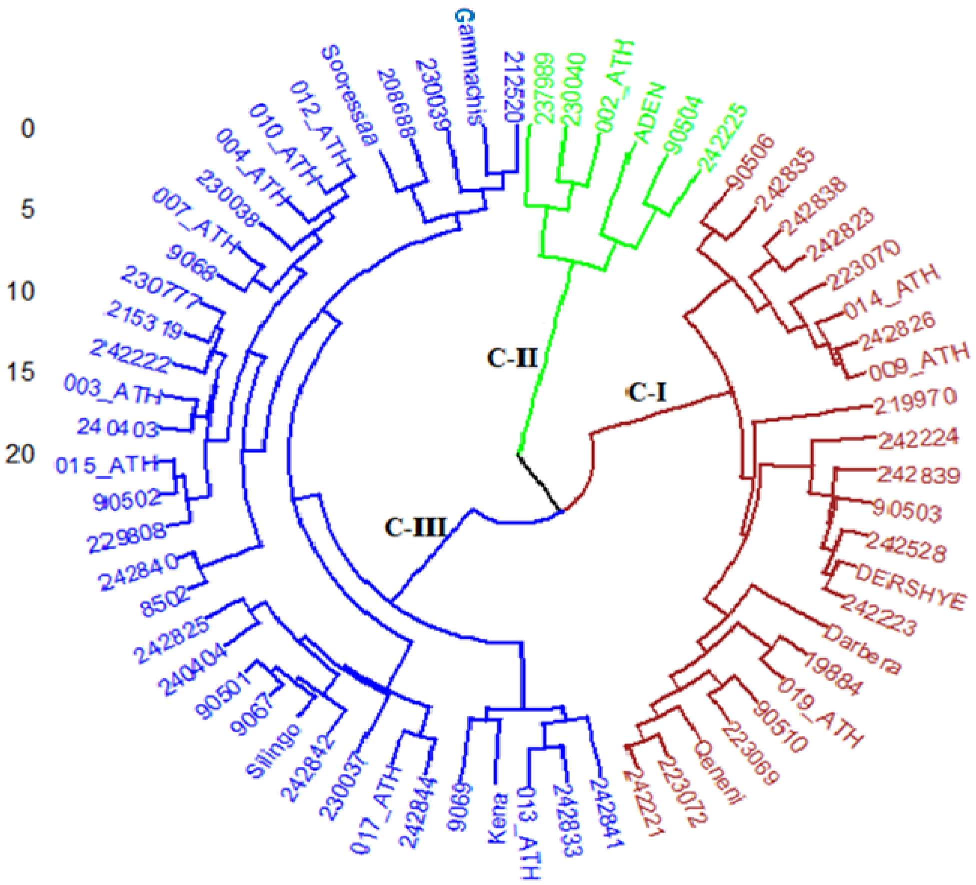


Figure 3.6. Phylogram showing the relationships among the 64 black cumin genotypes of Ethiopia evaluated for 12 quantitative traits. (Brown, green, and blue colors represent clusters I, II, and III, respectively.)

Table 3.6. Clustering of 64 black cumin genotypes of Ethiopia into three clusters using the mean of 12 quantitative traits.

Cluster	No. of genotypes	Genotypes included	Collection region
I	23 (35.93%)	90506, 242835, 242838, 242823, 223070, 014_ATH, 242826, 009_ATH, 219970, 242224, 242839, 90503, 242528, DERSHYE, 242223, Darbera, 19884, 019_ATH, 90510, 223069, Qeneni, 223072, 242221	Improved, SNNP, Oromia, Amhara, B/Gumuz, Tigray
II	6 (9.38%)	237989, 230040, 002_ATH, ADEN, 90504, 242225	Improved, Amhara, Tigray, Oromia
III	35 (54.69%)	212520, Gammachis, 230039, 208688, Sooressaa, 012_ATH, 010_ATH, 004_ATH, 230038, 007_ATH, 9068, 230777, 215319, 242222, 003_ATH, 240403, 015_ATH, 90502, 229808, 242840, 8502, 242825, 240404, 90501, 9067, Silingo, 242842, 230037, 017_ATH, 242844, 9069, Kena, 013_ATH, 242833, 242841	Improved, Oromia, Amhara, SNNP, Tigray, B/Gumuz

Where, B/Gumuz = Benshangul-Gumuz, SNNP = Southern Nations, Nationalities, and Peoples.

Table 3.7. The mean value of 12 quantitative traits of 64 black cumin genotypes of Ethiopia in each cluster

Trait	Cluster		
	I	II	III
Days to flowering	73	69	73
Days to 50% flowering	87	82	87
Days to full blooming	95	89	95
Days to maturity	152	145	152
Plant height (cm)	60	56.5	56
Number of primary branches per plant	5	5	5
Number of capsules per plant	13	11.5	11
Number of seeds per capsule	103	93.5	96
Number of seeds per plant	1353	1137.5	1107
Thousand seed weight (g)	2.2	2.2	2.2
Seed yield per plant (g)	2.99	2.43	2.39
Seed yield per hectare (t)	0.93	0.80	0.78

Intra- and inter-cluster distance analysis

The pairwise generalized squared distance (D^2) among the two clusters based on Mahalanobis's D^2 statistics are shown in Table 3.8. The significant ($p < 0.05$ or $p < 0.001$) intra-cluster genetic distances were observed within cluster I ($D^2 = 35.73$) and cluster II ($D^2 = 23.9$) indicating the existence of wide genetic variability within the genotypes of these two clusters. But intra-cluster genetic distance was not observed in cluster III ($D^2 = 19.03$) indicating there is no existence of wide genetic variability within the genotypes that belong to this cluster. The significant ($p < 0.001$) inter-cluster genetic distances were observed between clusters I and II ($D^2 = 64.78$), clusters I and III ($D^2 = 947.45$), and clusters II and III ($D^2 = 1300.44$). The result revealed the presence of high genetic divergence among genotypes within and between the clusters.

Table 3.8. Average intra-cluster (bolded diagonal) and inter-cluster (off-diagonal) distances among clusters in black cumin

Cluster	I	II	III
I	**35.73*** **		
II	64.78***	**23.9***	
III	947.45***	1300.44***	**19.03^{ns}**

$\chi 2$ = 19.68, 24.73, and 31.26 at 5, 1, and 0.1% probability levels, respectively. * and *** = Significant at $p < 0.05$ and $p < 0.001$, respectively.

Phenotypic diversity in groups of genotypes

The genotypes from all regions except those from SNNP and B/Gumuz, including improved varieties, were spread into all three clusters but in different proportions (Table 3.9). This indicates no definite relationship between places of origin and genotype clustering patterns. This might be related to wide-ranging seed exchange between farmers across the regions.

Table 3.9. Clustering of 64 black cumin genotypes based on collection groups

Cluster	Group						Total
	Amhara	Oromia	Tigray	SNNP	B/Gumuz	Improved	
I	7	6	2	2	3	3	23
II	2	2	1	-	-	1	6
III	15	10	3	2	1	4	35
Total	24	18	6	4	4	8	64

Cluster analysis, based on means for groups of origin for 12 quantitative traits, was used to obtain a dendrogram of the regions of origin (Figure 3.7). The genotypes were clustered into three. The first cluster comprised the genotypes from Amhara and improved cultivars, whereas the genotypes from Oromia, Benshangul-Gumuz, and SNNP were grouped into the second cluster. The genotypes from Tigray were grouped into the third cluster. The dendrogram clearly showed the close relationship between genotypes in each cluster. The first cluster was characterized by the lowest mean value of the number of capsules per plant (12), number of seeds per plant (1169), and seed yield per plant (2.54 g); and the highest mean value of seed yield per hectare (0.93 t) (Table 3.10). This indicates that the genotypes from this group are higher seed yielders than the others. The second cluster was characterized by late maturity, the highest number of seeds per capsule, number of seeds per plant, and seed yield per plant, with mean values of 152 days, 100, 1316, and 2.8 g, respectively. Whereas, the third group was characterized by early maturing, and the shortest plant height, with a mean value of 150 days and 56 cm, respectively, and intermediate in the remaining morpho-agronomic traits studied (Table 3.10). However, the number of primary branches per plant remains the same for the three clusters.

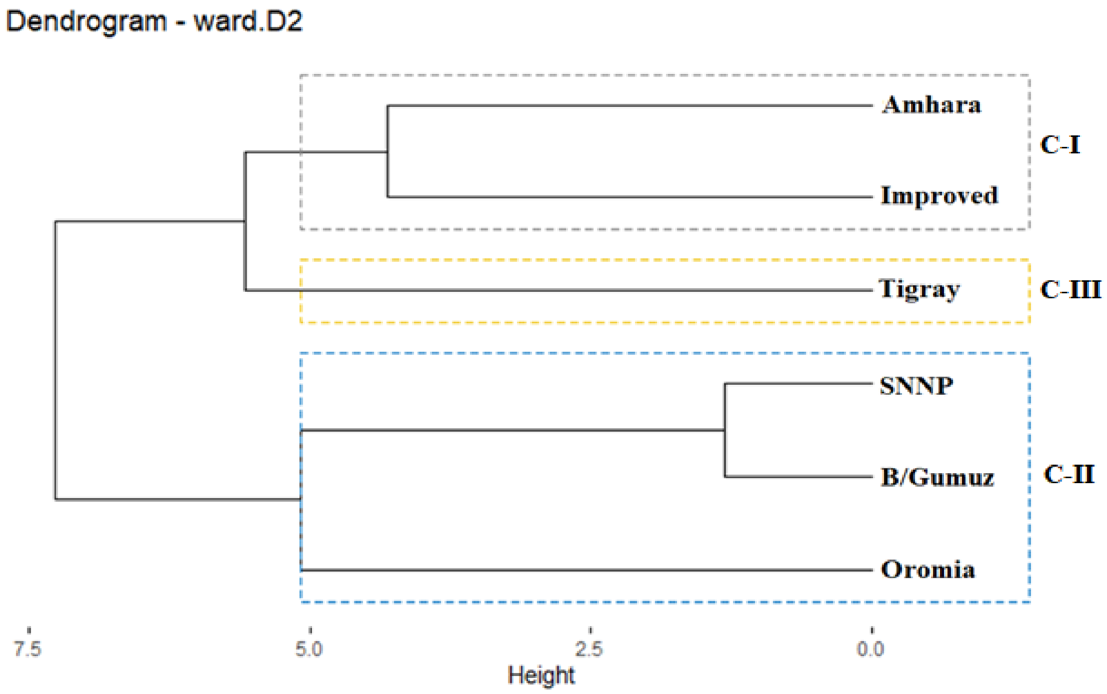


Figure 3.7. Dendrogram showing the relationships among the six groups of black cumin genotypes evaluated for 12 quantitative traits.

Table 3.10. The mean value of 12 quantitative traits of the groups of black cumin genotypes evaluated during the 2021 cropping season

Traits	Cluster		
	I	II	III
Days to flowering	72.5	73	73
Days to 50% flowering	86.5	87	86
Days to full blooming	94	94	93
Days to maturity	151.5	152	150
Plant height (cm)	58	58	56
Number of primary branches per plant	5	5	5
Number of capsules per plant	12	13	13
Number of seeds per capsule	97	100	97
Number of seeds per plant	1169	1316	1258
Thousand seed weight (g)	2.2	2.2	2.2
Seed yield per plant (g)	2.54	2.87	2.76
Seed yield per hectare (t)	0.93	0.84	0.87

3.4. Conclusions

There is a significant genetic variation among black cumin genotypes in Ethiopia for the studied morpho-agronomic traits. This suggests there is a high potential of the studied genotypes for the genetic improvement of those traits. The number of capsules per plant and plant height exhibited high heritability in a broad sense, coupled with high to moderate genetic advance as a percentage of the mean, which indicates the possibility of enhancing these traits through selection. Plant height and number of primary branches per plant are the most influential traits on seed yield per hectare; this would be an important selection criterion for further seed yield improvement programs of those genotypes. About 81.5% of the variation that occurred among the studied genotypes was explained by the first four principal components. The PCA confirmed the presence of high phenotypic variability among the black cumin genotypes of Ethiopia. The studied genotypes were grouped into three distinct clusters. This implies that there is a presence of significant quantitative traits of diversity that can be exploited for further black cumin improvement programs. In general, this study confirmed there is a reality about the presence of enough variation in most of the morpho-agronomic traits studied, which can create an enabling environment for breeders to design active genotype collection, conservation, and use strategies. Thus, it is recommended that further characterization and evaluation studies be conducted at a wider agroecological condition of the country to exploit the existence of diversity and enhance the improvement strategies.

CHAPTER 4

4. Biochemical Characterization of Ethiopian Black Cumin (*Nigella sativa* L.)

Abstract

Black cumin (Nigella sativa L.) seed oil has been used for its medicinal and aromatic values. Some studies revealed the presence of variability among black cumin genotypes in seed oil content and yield. In Ethiopia, very few studies were conducted to investigate the variability of black cumin genotypes by using biochemical traits. Thus, the objective of this study was to investigate the variability of Ethiopian black cumin genotypes based on biochemical traits. The study was conducted at Debre Zeit and Kulumsa Agricultural Research Centers' experimental sites under field conditions during the 2021 cropping season. Sixty-four genotypes were used and arranged in an 8x8 simple lattice design with two replications. Essential oils and fixed oils were extracted by the respective methods of hydrodistillation and solvent extraction. The univariate, bivariate, and multivariate analyses of the collected data were performed. Combined analysis of variance revealed significant differences among genotypes in fixed oil yield per hectare (FOY), essential oil content (EOC), and essential oil yield per hectare (EOY). EOY had a significant positive correlation with FOY and EOC. It is expected to improve all biochemical traits by 17.39% to 94.62% over the improved varieties by selecting the top 5% of black cumin accessions. Therefore, genotypes 90504, 219970, and 013_ATH were the top 5% best-performed accessions by FOY and EOY over the improved varieties. So, through selection, it would also be possible to improve the studied biochemical traits of the genotypes. The principal component analysis of four biochemical traits showed 85.86% of the total variance captured by the first two principal components (PCs). EOY and FOY were the main contributor traits to the variation in the first PC, whereas FOC and EOC were the main contributor traits to the variation in the second PC. The genotypes were grouped into three different clusters based on four biochemical traits with significant inter-cluster distance. This showed that there was sufficient diversity among the genotypes, which can be exploited for the future black cumin improvement program in Ethiopia.

Keywords: Association, Cluster, Essential oil, Fixed oil, Principal component

4.1. Introduction

Black cumin (*Nigella sativa* L.) is cultivated for its seed and biochemical traits, such as fixed oil and essential oils. Fixed oils are extracted from ground black cumin seeds with an organic solvent that contains volatile and nonvolatile constituents. However, Essential oils (EOs) are natural, secondary metabolites and volatile complex compounds characterized by the aroma of their corresponding aromatic plants (Bakkali *et al*., 2008; Dhifi *et al*., 2016; Mironescu and Georgescu, 2021).

In Ethiopia, black cumin has been cultivated and used for its spice and medicinal value. Black cumin seeds fixed oil, locally known as “Ye Tikur Azmud Zeyit,” is well adapted and sold in many pharmacies to treat the common cold. It is now a source of income for small-scale households by selling whole seeds and fixed oil. The black cumin oil market in the United States of America was valued at more than 15 million USD in 2018. It is anticipated that this amount will rise to about 25 million USD annually by 2025. Due to an increase in black cumin plantations, the market for its oil in the Asia Pacific, which is dominated by Australia, South Korea, Japan, China, and India, is expected to grow to a value of over 10 million USD by 2025 (Haque *et al*., 2021).

Evaluation and documentation of the existing genetic diversity are very important for the maintenance and exploration of the variability of the black cumin for breeding programs. Different scholars have reported the genetic diversity of black cumin from the potentially producing countries of the world (Edris, 2010; Gharby *et al*., 2015; Beheshti *et al*., 2018; Hosseini *et al*., 2018; Palabıyık and Aytaç, 2018; Bayati *et al*., 2020; Bozdemir *et al*., 2022). In

Ethiopia, the existence of genetic diversity of black cumin seed oil was reported by Abdela (2020), Asefa and Beriso (2021), Demis *et al*. (2023), and Fekadu and Gizaw (2023) on 3, 14, 11, and 3 germplasms, respectively. Those were limited in number.

Biochemical characterization is another alternative criterion for the evaluation and documentation of genotypes. The fixed oil content, fixed oil yield, essential oil content, and essential oil yield were analyzed separately for univariate and combined for bivariate and multivariate like cluster and principal components. Based on this result, there is a high level of genetic variability in black cumin genotypes of Ethiopia that can be exploited for the identification and selection of the spice and medicinal plants' breeding program. This provides a piece of baseline information for academia, researchers, industries, policymakers, and societies that rely on this crop.

As mentioned, some authors reported the genetic diversity of a few Ethiopian black cumin collections by seed oil content and yield. This would hinder the potential of black cumin to be exploited for future breeding programs. This should be addressed by using many collections of black cumin cultivated in the country to exploit the existing full potential. Therefore, this study was designed to evaluate the Ethiopian black cumin genotypes by their seed oil content and yield.

4.2. Materials and Methods

4.2.1. Planting materials and experimental procedures

For this experiment, all the planting materials (Appendix Table 1) and the experimental procedures used were the same as in Chapter 3. ArcGIS Desktop Advanced version 10.8 software created the site map of the collection areas from where the planting materials originated (ESRI, 2011) (Figure 4.1).

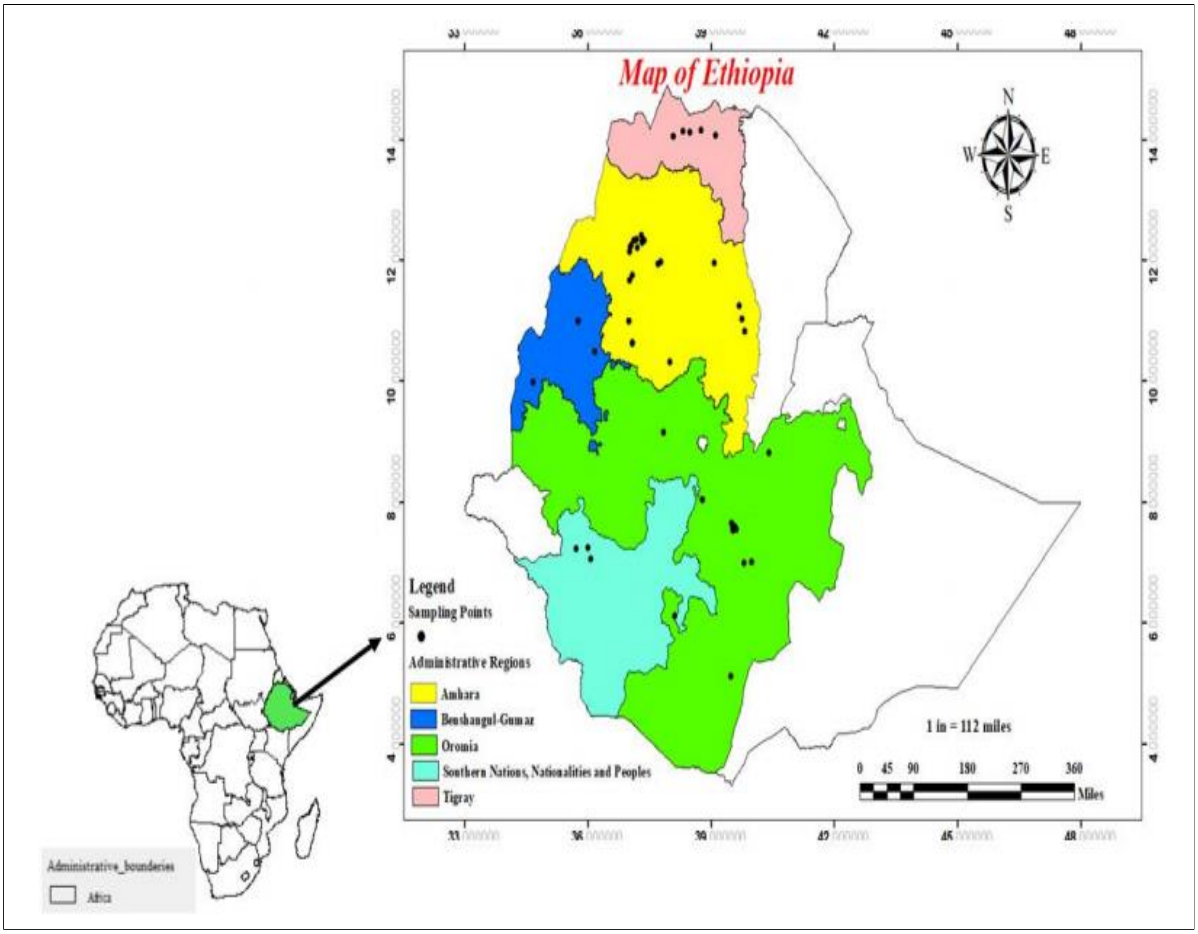


Figure 4.1. The diagram shows the collection sites of each *N. sativa* L. genotype from different regions.

4.2.2. Data collection

Biochemical data such as fixed oil content (%), fixed oil yield per hectare (kg), essential oil content (%), and essential oil yield per hectare (kg) were critically recorded from the collected dried seeds of *N. sativa* by using the following respective procedures.

Fixed oil (crude fat) content (%)

The fixed oil content of the samples was determined using the Soxtec Extraction system (Foss Soxtec^{TM} 8000 Extraction unit, Sweden (Appendix Figure 3) according to the AOAC (2005) method. Three grams of composite dried and crushed black cumin seed samples from each germplasm were weighed into thimbles lined with cotton at the bottom. The mass of the cooled cups was then measured. The thimbles containing the samples were placed into the Soxtec^{TM} 8000 Extraction system, and 50 ml of n-hexane was added as a solvent to each cup using a dispenser. The extraction process consisted of 20 minutes of boiling, 30 minutes of rinsing, and 10 minutes of recovery. The cups with their residue were then removed from the Soxtec system and placed in a drying oven at 105 °C for 30 minutes. The cups were then cooled in desiccators for an hour. The mass of each cooled cup together with its fixed oil contents was weighed, and then the mass of the cup was subtracted from each sample. Finally, the fixed oil content (FOC) was calculated by the following formula (Daniel *et al*., 2009):

$$\text{FOC}\left(\frac{w}{w},\%\right) = \frac{\text{Weight of fixed oil (g)}}{\text{Weight of sample (g)}} \text{ x } 100 \qquad (4.1)$$

Fixed oil yield per hectare (kg)

Fixed oil yield per hectare (FOY) is the fixed oil yield obtained from the ground black cumin seeds that were harvested from the central rows of each plot and converted into yield per hectare based on FOC and seed yield per hectare. It was determined by the following formula:

$$\mathrm{FOY\ (kg)} = \frac{\mathrm{FOC}\left(\frac{\mathrm{w}}{\mathrm{w}}, \%\right) \mathrm{x\ Seed\ yield\ per\ hectare\ (kg)}}{100} \tag{4.2}$$

Essential oil content (%)

To determine essential oil content (w/w%), 100 g of composite ground black cumin seeds samples were taken and put in 0.5 l water from each genotype separately and then subjected to hydrodistillation for 3 hours using Clevenger apparatus (Appendix Figure 4) at Wondo Genet Agricultural Research Center, Natural Products Laboratory (Figure 4.2) according to the standard procedure described by Baj *et al.* (2015). The amount of extracted essential oil was measured by using a measuring pipette. Essential oil content (EOC) was determined by the following formula (Daniel *et al.*, 2009).

$$\mathrm{EOC}\left(\frac{\mathrm{w}}{\mathrm{w}}, \%\right) = \frac{\mathrm{Weight\ of\ essential\ oil\ (g)}}{\mathrm{Weight\ of\ sample\ (g)}} \mathrm{\ x\ 100} \tag{4.3}$$

Figure 4.2. Essential and fixed oil extraction process: A) Black cumin seeds, B) Seed grinder, C and D) Ground the seed, put it in the flask containing water, and assemble the Clevenger apparatus, respectively, and E) The extracted essential oil in the flask.

Essential oil yield per hectare (kg)

Essential oil yield per hectare (EOY) is the essential oil yield obtained from the ground black cumin seeds that were harvested from the central rows of each plot and converted into yield per hectare based on EOC and seed yield per hectare. It was determined by the following formula:

$$\text{EOY (kg)} = \frac{\text{EOC } \left(\% \frac{\text{w}}{\text{w}}\right) \text{ x Seed yield per hectare (kg)}}{100} \qquad (4.4)$$

4.2.3. Data Analysis

Before conducting the analysis, the validity of each experiment was evaluated by using Bartlett's test to determine the homogeneity of error variances for each of the measured traits. As a result, it was discovered that most of the parameters were similar between the two sites. Next, the combined analysis of variance of FOC, FOY, EOC, and EOY data from the two locations was carried out based on the model of simple lattice design proposed by Gomez and Gomez (1984) using SAS version 9.4 computer software programs (SAS Institute Inc., 2019). The analysis of variance (ANOVA) was performed using the PROC GLM procedure. To find significant differences between genotype means, the Least Significant Difference (LSD) test procedure was employed at a 5% significance level. Genotypes were regarded as fixed variables in the analysis, whereas replications, blocks, and locations were regarded as random variables.

Principal component analysis (PCA) and cluster analysis were applied to the mean of the genotype data that were collected using R-software version 4.2.2 (R Core Team, 2022). The R software packages "factoextra" (Kassambara, 2017), "cluster" (Maechler *et al.*, 2021), "class" (Venables and Ripley, 2002), and "clv" (Lukasz, 2020) were used to carry out the hierarchical cluster analysis. The aggregation patterns of the 64 *N. sativa* genotypes and populations that were selected based on their similarity to the corresponding means of all the collected traits were examined by using clustering. Using the R software package "Nbclust," the ideal number of clusters for the data set was ascertained (Charrad *et al.*, 2014). Euclidean distance was used to calculate the genetic distance, and the ward D2 linkage method was used to create the dendrogram using the distance matrix. The inter-cluster distance was calculated based on the standardized Mahalanobis's D2 statistics (Mahalanobis, 1936) as:

$$D^2{}_{ij} = (xi - xj)' cov^{-1} (xi - xj) \tag{4.5}$$

Where D^2ij = the distance between cases i and j; xi and xj = vectors of the values of the variables for cases i and j; and cov^{-1}= the pooled within-groups variance-covariance matrix. The significance of D^2 values between any two clusters was tested both at 1% and 5% probability levels against the tabulated chi-square ($\chi 2$) values at p-1 degrees of freedom, where p refers to the number of quantitative characters considered (Singh and Chaudhary, 1985).

The genotypes with higher levels of desirable traits were selected for further production and improvement programs, and clusters containing these genotypes were identified. The R-software packages "factoextra," "ggplot2" (Wickham, 2016), "corrplot" (Wei and Sinko, 2021), and "ggsignif" (Ahlmann-Eltze, 2022) were utilized in conjunction with PCA to determine the traits that account for a significant portion of the overall variation between the populations and groups.

4.3. Results and Discussion

Combined analysis of variance over the two locations revealed a significant ($p < 0.001$) effect for location and genotypes in FOY, EOC, and EOY. Supporting results were reported by Demis *et al.* (2023) on black cumin FOC. However, FOC was non-significant ($p \geq 0.05$) in the case of location. Location x genotype interaction effects were also significant ($p < 0.001$) in FOC, FOY, EOC, and EOY (Table 4.1).

Separate analyses of variance showed significant differences among the genotypes by FOC, FOY, EOC, and EOY. Similarly, combined analyses of variance exhibited significant differences in those measured biochemical traits. Significant variability was also detected from the wider ranges observed between the minimum and the maximum values of the traits measured (Table 4.2).

The variability of the genotypes for EOY and FOY creates a great opportunity for developing high-yielding varieties by improving the traits that are associated with them. This result would help the breeders to develop improved and suitable varieties for different agroecology of the country.

Table 4.1. Combined analysis of variance of 4 biochemical traits recorded on 64 N. sativa genotypes of Ethiopia at Debre Zeit and Kulumsa during the 2021 cropping season

Traits	Mean Square			CV (%)
	Location (DF = 1)	Genotype (DF = 63)	Location*Genotype (DF = 63)	
FOC (%)	1.80ns	31.90***	3.80***	0.81
FOY (kg)	3400241***	19813.23***	11402***	17.68
EOC (%)	0.38***	0.03***	0.03***	11.04
EOY (kg)	512.12***	4.89***	2.75***	18.75

Where FOC = Fixed oil content, FOY = Fixed oil yield per hectare, EOC = Essential oil content, EOY = Essential oil yield per hectare. DF = Degree of freedom, CV = Coefficient of variation, ns = non-significant at $p \geq 0.05$, ***= Significant at $p < 0.001$.

4.3.1. Mean performance of the genotypes on biochemical traits

Based on the combined result, there was a significant ($p < 0.001$) difference among the evaluated genotypes for all biochemical traits studied (Table 4.1).

The FOC ranged from 30.1% to 46.47%, with a mean value of 37.89%. The highest mean FOC was obtained from genotype 19884 (from SNNP) followed by genotype 9068 (from Amhara) and the improved variety DERSHYE; whereas, the least mean value was obtained from genotype 242840 (from Oromia) (Appendix Table 7). Studies from different countries reported *N. sativa* FOC was below this range (Abdela, 2020; Ali and Hassan, 2014; Atta, 2003; Bayati *et al*., 2020; Hamed *et al*., 2023; Rezaei-Chiyaneh *et al*., 2021), and within this range (Aksu *et al*., 2021; Bozdemir *et al*., 2022; Demis *et al*., 2023; Gharby *et al*., 2015; Hosseini *et al*., 2018; Kara *et al*., 2021; Kizil *et al*., 2008; Kökdil and Yilmaz, 2005). This shows, that there is a possibility of developing improved black cumin varieties with high fixed oil content out of the Ethiopian collections.

FOY ranged from 76.04 kg to 788.18 kg, with a mean value of 334.51 kg (Table 4.2). The highest mean FOY was obtained from genotype 90504 (from Oromia) followed by genotype 013_ATH (from Amhara) and 219970 (from Tigray); whereas, the lowest value was obtained from genotype 237989 (from Oromia) (Appendix Table 7). Abdou *et al.* (2023), Bayati *et al.* (2020), Hamed *et al.* (2023), Hosseini *et al.* (2018), Rezaei-Chiyaneh *et al.* (2021) and Salaheldin *et al.* (2020) reported that *N. sativa* seeds FOY were found within this range. Seed yield per hectare is the main determinant of fixed oil yield per hectare. This indicates, that there is a possibility of developing high fixed oil-yielding black cumin varieties out of the Ethiopian collections.

Table 4.2. Descriptive statistics, F-test, and coefficient of variation of four biochemical traits of the 64 black cumin genotypes of Ethiopia at each testing site and pooled during the 2021 cropping season

Location	**Statistics**	**FOC (%)**	**FOY (kg)**	**EOC (%)**	**EOY (kg)**
Debre Zeit	Mean ± SE	37.89 ± 0.31	449.76 ± 11.52	0.39 ± 0.01	4.61 ± 0.18
	Range	30.10-46.47	108.55-788.18	0.09-0.65	0.64-9.75
	F-Test	***	***	***	***
	CV (%)	0.89	17.22	11.53	16.77
Kulumsa	Mean ± SE	37.72 ± 0.24	219.27 ± 6.21	0.31 ± 0.01	1.78 ± 0.08
	Range	32.88-43.69	76.04-410.23	0.09-0.83	0.41-5.88
	F-Test	***	***	***	***
	CV (%)	0.68	11.56	8.81	11.31
Combined	Mean ± SE	37.81 ± 0.20	334.51 ± 9.73	0.35 ± 0.01	3.20 ± 0.13
	Range	30.10-46.47	76.04-788.18	0.09-0.83	0.41-9.75
	F-Test	***	***	***	***
	CV (%)	0.81	17.68	11.04	18.75

Where SE = Standard error, FOC = Fixed oil content, FOY = Fixed oil yield per hectare, EOC = Essential oil content, EOY = Essential oil yield per hectare.
*** = Significant at $p < 0.001$.

EOC ranged from 0.09% to 0.83%, with a mean value of 0.35% (Table 4.2). The highest mean EOC was obtained equally from genotype 90504 (from Oromia) and 90501 (from Amhara) followed by genotype 215319 (from Amhara) and 229808 (from Benshangul-Gumuz); whereas, the least mean value was obtained equally from genotype 240404 (from SNNP) and the improved variety Silingo (Appendix Table 7). This might be a result of the variations in the genetic makeup of the genotypes. Results from different countries reported *N. sativa* EOC was within this range (Abdela, 2020; Bayati *et al*., 2020; Kara *et al*., 2021; Kizil *et al*., 2008; Rezaei-Chiyaneh *et al*., 2021; Sileshi and Biruk, 2020). This shows, that there is a possibility of developing black cumin varieties with high essential oil content from the Ethiopian collections.

EOY ranged from 0.41 kg to 9.75 kg, with a mean value of 3.2 kg (Table 4.2). The highest mean EOY was obtained from genotype 90504 (from Oromia), followed by genotype 219970 (from Tigray) and 013_ATH (from Amhara), whereas the lowest mean value was obtained from genotype 240404 (from SNNP) (Table 3.3). Hosseini *et al.* (2018) and Bayati *et al.* (2020) from Iran and Abdou *et al.* (2023) from Egypt reported that *N. sativa* EOY was within this range. This indicates that there is a possibility of developing black cumin varieties with high essential oil yield from the Ethiopian collections. Like FOY, seed yield per hectare is the main determinant of EOY. They have a direct relationship.

The presence of variability among the black cumin genotypes was observed by the wide range of biochemical traits studied. This indicates there is a possibility of improving these traits through selection. Selection of the top 5% of genotypes is predicted to improve biochemical traits by 17.39% to 94.62% through selection (Table 4.3). Genotype 90504 (from Oromia), 219970 (from Tigray), and 013_ATH (from Amhara) were the top 5% best-performed accessions over improved varieties selected for FOY and EOY (Appendix Table 7). Basazinew *et al.* (2024) reported that genotypes 90504 and 219970 were the top best-performed accessions over the improved varieties in seed yield per hectare.

Table 4.3. Comparison of mean performances of 5% of the best-performed accessions selected for 4 biochemical traits over the mean performance of improved varieties

Traits	Mean values		A comparative advantage over mean values of improved varieties (%)
	Top 5% accessions	Improved varieties	
FOC (%)	44.02	37.5	17.39
FOY (kg)	530.42	329.02	61.21
EOC (%)	0.54	0.33	63.64
EOY (kg)	6.15	3.16	94.62

Where FOC = Fixed oil content, FOY = Fixed oil yield per hectare, EOC = Essential oil content, and EOY = Essential oil yield per hectare.

4.3.2. Relationship among biochemical traits

Information on the nature and extent of association between any two characters is provided by correlation studies (Russinga, 2020). To measure the degree and direction of the relationship between biochemical traits, correlation analysis was made (Table 4.4). EOY had a significant positive correlation with fixed oil yield per hectare (r = 0.83***) and EOC (r = 0.69***); the association between EOC and EOY with FOC was non-significant. Hosseini *et al*. (2018) reported similar relationships between FOY and EOC on black cumin.

Table 4.4. The correlation coefficient of the four biochemical traits of 64 black cumin genotypes

Variables	FOC	FOY	EOC	EOY
FOC (%)	1.00			
FOY (kg)	0.25***	1.00		
EOC (%)	0.07ns	0.26***	1.00	
EOY (kg)	0.10ns	0.83***	0.69***	1.00

Where, FOC = Fixed oil content, FOY = Fixed oil yield per hectare, EOC = Essential oil content, and EOY = Essential oil yield per hectare.
ns = non-significant at $p \geq 0.05$, and ***= Significant at $p < 0.001$.

4.3.3. Cluster analysis

Cluster analysis was done based on four biochemical traits of the 64 black cumin germplasms of Ethiopia to examine the aggregation patterns of the 64 black cumin germplasms based on their similarity. The best number of clusters for the data set was determined, which was three (Charrad *et al*., 2014).

Cluster I: This cluster was the largest group, having 32 (50%) genotypes and consisting of accessions from all collection regions and most of the improved varieties (Figure 4.5 and Table 4.5). This group was characterized by the highest FOC with a mean value of 39.09% (Table 4.6). Three of the 32 genotypes recorded above 43% of FOC: 19884 (44.97%), 9068 (44.21%), and DERSHYE (43.23%) (Appendix Table 7).

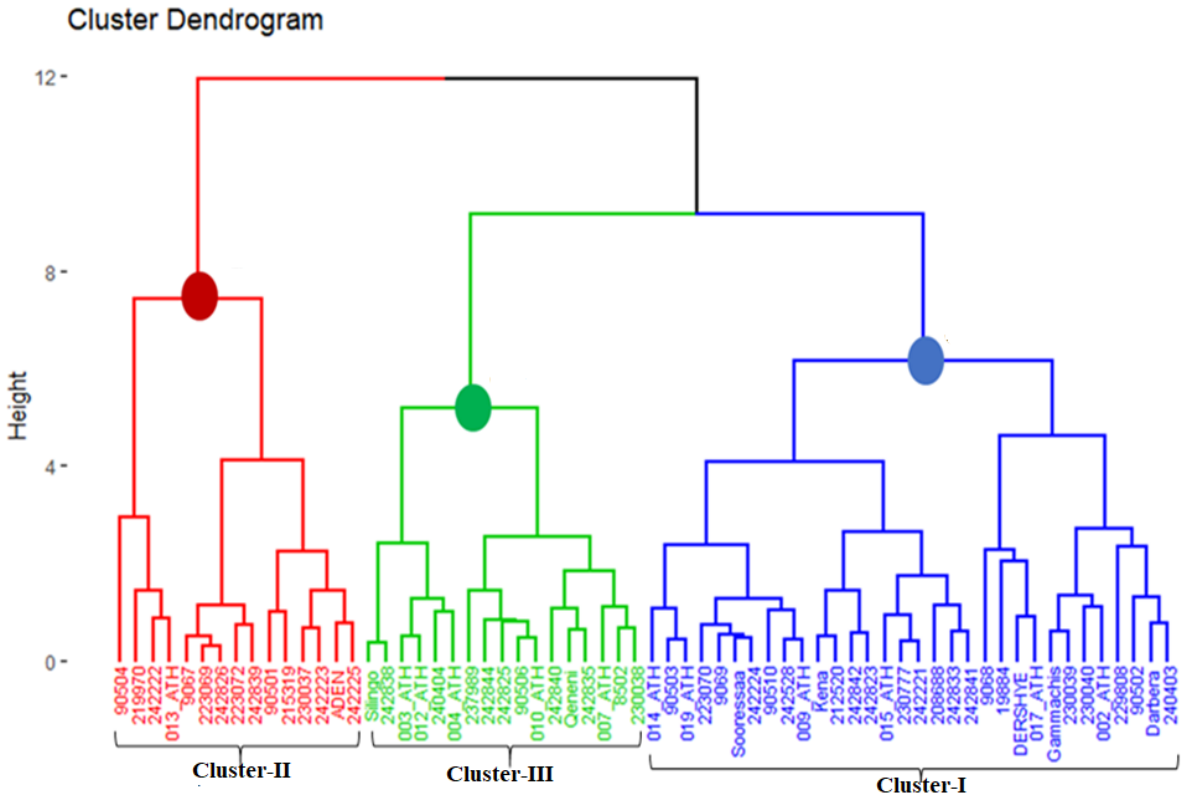


Figure 4.3. Dendrogram showing the relationships among the 64 black cumin genotypes of Ethiopia evaluated for four biochemical traits.

Cluster II: This group contained 15 (23.44%) genotypes and consisted of accessions from the Oromia, Amhara, Tigray, and Benshangul-Gumuz regions and only one of the improved varieties (Figure 4.5 and Table 4.5). This group was characterized by the highest FOY, EOC, and EOY with the mean value of 367.86 kg, 0.45%, and 4.69 kg, respectively (Table 4.6). Three of the 15 genotypes recorded above 495 kg of EOY: 90504 (582.03 kg ha^{-1}), 013_ATH (509.28 kg ha^{-1}), and 219970 (499.95 kg ha^{-1}) (Appendix Table 7).

Cluster III: The remaining genotypes 17 (26.56%) from the five regions (SNNP, Oromia, Amhara, Benshangul-Gumuz, and Tigray) and two of the improved varieties (Silingo and Qeneni) belonged to this group and are characterized by the least mean values of all the studied biochemical traits (Figure 4.5 and Table 4.5).

Table 4.5. Clustering of 64 black cumin genotypes of Ethiopia into three clusters using the mean of 4 biochemical traits

Cluster	Number of genotypes	Genotypes included	Collection region
I	32 (50%)	Kena, Gammachis, 230039, 208688, 90503, 242841, 230040, 242823, 242833, 242842, 009_ATH, 019_ATH, DERSHYE, 240403, Darbera, 242224, 19884, Sooressaa, 017_ATH, 014_ATH, 230777, 015_ATH, 242528, 9068, 223070, 229808, 212520, 90510, 9069, 002_ATH, 90502, 242221	Improved, SNNP, Oromia, Amhara, B/Gumuz, Tigray
II	15 (23.44%)	90504, 219970, 242222, 013_ATH, 9067, 223069, 242826, 223072, 242839, 90501, 215319, 230037, 242223, ADEN, 242225	Improved, Oromia, Amhara, Tigray, B/Gumuz
III	17 (26.56%)	Silingo, 242838, 003_ATH, 012_ATH, 240404, 004_ATH, 237989, 242844, 242825, 90506, 010_ATH, 242840, Qeneni, 242835, 007_ATH, 8502, 230038	Improved, Amhara, SNNP, Oromia, Tigray

Table 4.6. The mean value of 4 biochemical traits of 64 black cumin genotypes of Ethiopia in each cluster

Traits	Cluster		
	I	II	III
FOC (%)	39.09	36.31	35.51
FOY (kg)	348.56	367.86	257.31
EOC (%)	0.33	0.45	0.30
EOY (kg)	3.08	4.69	1.95

Where FOC = Fixed oil content, FOY = Fixed oil yield per hectare, EOC = Essential oil content, and EOY = Essential oil yield per hectare.

The genotypes from the Amhara, Oromia, and Tigray regions and the improved varieties were spread into all three clusters but in different proportions (Table 4.7). However, the genotypes from the SNNP and Benshangul-Gumuz regions were spread only into the two clusters with different proportions. Most of the genotypes from Amhara, Oromia, SNNP, Benshangul-Gumuz regions, and improved varieties were grouped under the first cluster (Table 4.7).

Table 4.7. Clustering of 64 black cumin genotypes based on collection groups

Groups	Cluster			Total
	I	II	III	
Amhara	11	7	6	24
Oromia	8	3	7	18
Tigray	2	3	1	6
SNNP	3	-	1	4
Benshangul-Gumuz	3	1	-	4
Improved	5	1	2	8
Total	32	15	17	64

Intra- and inter-cluster distance analysis

The significant ($p < 0.001$) intra-cluster genetic distances were observed within cluster I (D^2 = 103.65), cluster II (D^2 = 102.55), and cluster III (D^2 = 153.87). This indicates the existence of wide genetic variability within the genotypes of the three clusters (Table 4.8).

Significant ($p < 0.001$) inter-cluster genetic distances were observed between clusters I and II ($D^2 = 634.78$), clusters I and III ($D^2 = 555.05$), and clusters II and III ($D^2 = 2054.45$) were also significant ($p < 0.001$). The result shows the presence of high genetic divergence among genotypes within and between the clusters (Table 4.8).

Table 4.8. Average intra-cluster (bolded diagonal) and inter-cluster (off-diagonal) distances among clusters in black cumin

Cluster	I	II	III
I	**103.65*****		
II	634.78***	**102.55*****	
III	555.05***	2054.45***	**153.87*****

$\chi 2$ = 7.82, 11.35, and 16.27 at 5, 1, and 0.1% probability levels, respectively.
*** = Significant at $p < 0.001$.

A dendrogram of the regions of origin was created using cluster analysis, which was based on means for groups of origin for four biochemical traits (Figure 4.6). Three groups of genotypes were identified. Genotypes from the Benshangul-Gumuz, SNNP, and Amhara regions, and improved varieties made up the first cluster; genotypes from Tigray and Oromia were grouped into the second and third clusters, respectively. The close relationships between the genotypes in each cluster were amply displayed by the dendrogram. The highest mean value of FOY (343.74 kg) and the intermediate mean values of FOC (38.69%), EOC (0.33%), and EOY (3.12 kg) were the characteristics of the first cluster (Table 4.9). This suggests that the genotypes in this group have a higher FOY than the others. Out of all the biochemical traits measured, the second cluster had the lowest mean value. In contrast, the third group had the highest FOC (39.03%), EOY (3.65 kg), EOC (0.40%) and intermediate in FOY with a mean value of 339.38 kg (Table 4.9).

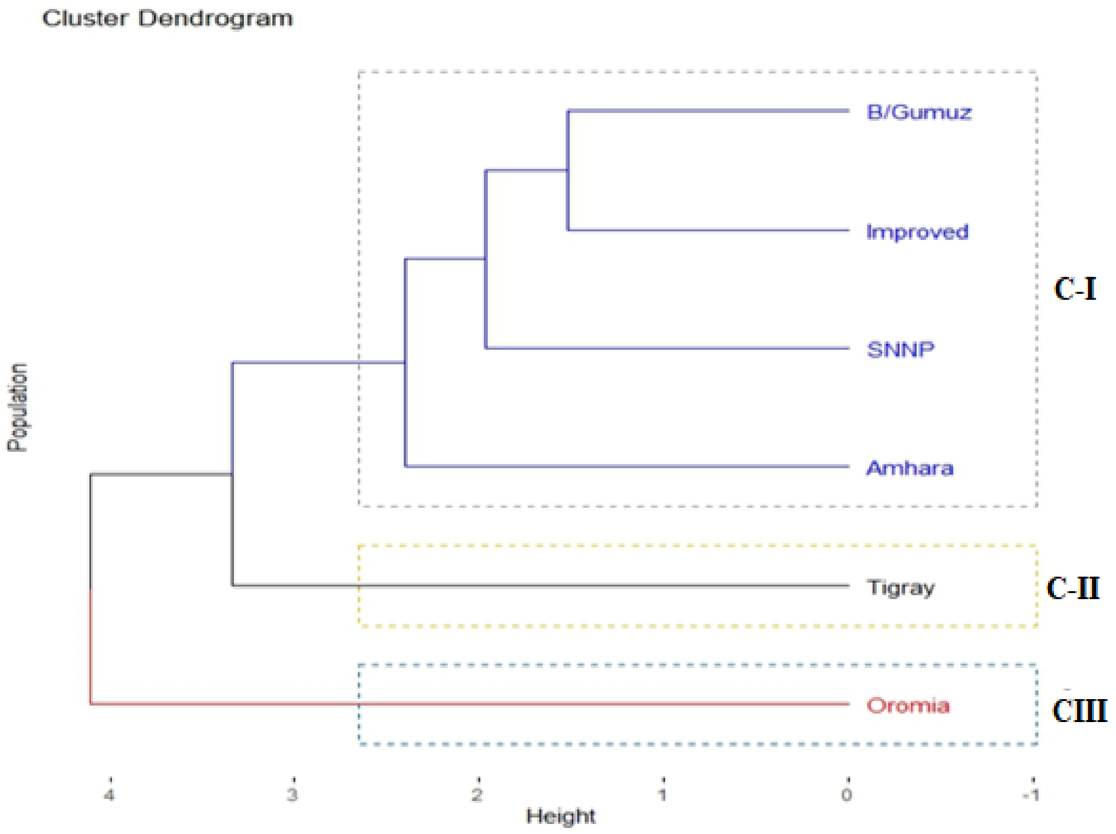


Figure 4.4. Dendrogram showing the relationships among the six groups of black cumin genotypes evaluated for four biochemical traits. (Blue, black, and red colors represent clusters I, II, and III, respectively.)

Table 4.9. The mean value of four biochemical traits of the groups of black cumin genotypes evaluated during the 2021 cropping season

Traits	Cluster		
	I	II	III
FOC (%)	38.69	35.02	39.03
FOY (kg)	343.74	293.07	339.38
EOC (%)	0.33	0.32	0.4
EOY (kg)	3.12	2.63	3.65

Where FOC = Fixed oil content, FOY = Fixed oil yield per hectare, EOC = Essential oil content, and EOY = Essential oil yield per hectare.

4.3.4. Principal component analysis

The principal components analysis grouped the four biochemical traits into four principal components presented in Table 4.10, which explained the entire 100% of the variability among the studied genotypes. The first two principal components (PCs) explained 85.86% of the variation that existed among the studied genotypes.

An eigenvalue greater than one indicates that PCs account for more variance than accounted by one of the original variables in standardized data commonly used as a cutoff point for which PCs are retained (Kaiser, 1960). The number of components is determined at the point beyond which the remaining eigenvalues are relatively small and of comparable size (Jolliffe, 2002; Peres-Neto *et al*., 2005). Therefore, based on the eigenvalues, two PCs having eigenvalues between 1.17 and 2.26 extracted a cumulative variance of about 85.86% of the total phenotypic diversity maintained.

The influential PCs for clustering were the characters with the largest absolute values closer to unity than those with lower absolute values closer to zero (Chahal and Gosal, 2002). So, due to the main contribution of the variations in FOY and EOY, the first principal component explained up to 56.61% of the total variance. However, the second principal component explained about 29.24% of the total variance by the main contribution of FOC and EOC. The principal component analysis confirmed that the collected Ethiopian *N. sativa* genotypes have high diversity, and most of the traits considered seemed to have high contributions toward the total phenotypic variability.

Table 4.10. The eigenvalues and eigenvectors of the first two principal components (PCs) for 4 biochemical traits of 64 black cumin genotypes of Ethiopia

Variable	Principal components	
	PC1	PC2
Eigenvalue	2.26	1.17
Proportion of variance (%)	56.61	29.24
Cumulative variance (%)	56.61	85.86
	Eigenvectors	
Fixed oil content (%)	-0.28	**0.74**
Fixed oil yield per hectare (kg)	**-0.55**	0.35
Essential oil content (%)	-0.48	**-0.51**
Essential oil yield per hectare (kg)	**-0.63**	-0.25

4.4. Conclusions

The result revealed the existence of significant variation for the studied biochemical traits among Ethiopian black cumin genotypes. The mean performance of the genotypes discovered the wider ranges between the least and greatest values of all biochemical traits, this showed the presence of significant variation among black cumin genotypes included in this study. This might be the result of the difference in the genetic makeup of the genotypes. EOY had a significant positive correlation with FOY and EOC, but it did not correlate with FOC. FOY had a strong positive and significant correlation with FOC. About 85.86% of the variation that occurred among the studied genotypes was explained by the first two principal components. EOY and FOY were the main contributor traits for most of the variation occurring among the genotypes in the first PC, whereas FOC and EOC were the main contributor traits for the second PC. The PCA confirmed the existence of high variability among black cumin genotypes of Ethiopia in all the biochemical traits studied. Genotypes of black cumin were partitioned into three distinct clusters with significant variation in the distance among genotypes within and between them. Genotypes 90504, 219970, and 013_ATH were the top 5% best-performed accessions over improved varieties selected for FOY and EOY. In general, this study confirmed the presence of enough variation in most of the biochemical traits studied in Ethiopian black cumin genotypes, which can create an enabling environment for the breeders to design active genotype collection, conservation, and use strategies. Moreover, this finding has a significant advantage for academia, industries, researchers, policymakers, and societies at large. Finally, it is recommended to conduct further evaluation studies on wider agroecological conditions of the country to enhance the improvement strategies by exploiting the existing diversity.

CHAPTER 5

5. Diversity of Ethiopian Black Cumin (*Nigella sativa* L.) Based on Compositions of Essential Oil

Abstract

No comprehensive investigation has been conducted on the variability of black cumin (Nigella sativa L.) seeds in essential oil compositions (EOCs). This would hinder the potential of the plant to exploit its medicinal and aromatic values. This study aimed to investigate the variability of Ethiopian black cumin genotypes by their EOCs. Seeds of 64 black cumin genotypes were used for this experiment. Composite samples of 100 g seeds were collected and ground roughly from each genotype. Extraction was made by hydro-distillation using a Clevenger-type apparatus for 3 h, and the essential oil was collected by measuring the amount using a measuring pipette. The essential oil samples were stored in a refrigerator at 4 °C until Gas Chromatography-Mass Spectrometry (GC-MS) analysis. Descriptive statistics was used to estimate the variations among populations' combined mean values of EOCs using the SAS version 9.4 software package. The study revealed the presence of variability among black cumin genotypes in EOCs. A total of 27 EOCs were detected from the essential oil of 64 genotypes using Gas Chromatography-Mass Spectrometry (GC-MS), out of which major constituents covered 92 to 100% of the total EOCs. The essential oils were dominated by ρ-cymene, thymoquinone, α-thujene, carvacrol, trans-4-methoxy thujane, longifolene, terpinen-4-ol, 2-ethylhexyl isohexyl phthalate, β-pinene, α-pinene, phthalic acid, 2-fluorobenzyl heptyl ester, and d-limonene. Among these, ρ-cymene, thymoquinone, α-thujene, trans-4-methoxy thujane, and carvacrol were the most abundant constituents investigated in all genotypes, while the rest varied among the genotypes. It is expected to improve all major EOCs by 13.2 to 152.1% over improved varieties by selecting the top 5% of black cumin accessions. The abundant EOC thymoquinone had a significant and positive correlation with carvacrol and a strong and significant negative correlation with α-thujene, α-pinene, β-pinene, ρ-cymene, and d-limonene. The principal component analysis of 12 major EOCs showed that 72.03% of the total variance was captured by the first two principal components (PCs). Based on the major EOCs, cluster analysis grouped the 64 genotypes into two different chemotypes. Chemotype A is characterized by a high content of thymoquinone, carvacrol, and longifolene, while Chemotype B is distinguished by a high content of ρ-cymene and α-thujene. This study disclosed the existence of a significant diversity of EOCs among the Ethiopian black cumin genotypes, which can be exploited for future breeding and improvement programs.

Keywords: Cluster, Chemotype, Essential oil compositions, GC-MS, Principal component

5.1. Introduction

The essential oil of black cumin (*Nigella sativa* L.) contains various bioactive molecules such as thymoquinone, thymol, tocopherol, trans-retinol, and selenium (Sultan *et al*., 2009) and is used worldwide for functional foods and nutraceuticals (Javed *et al*., 2012). Javed *et al* (2012) have also described black cumin as a miraculous plant and considered by earliest herbal specialists as “The herb from heaven”.

Herbal medicine is in high demand worldwide and continues to expand. Various technologies have been implemented to enhance the presence of bioactive compounds in medicinal plants. Nowadays, bioactive compounds that are extracted from plants are utilized in food additives, agrochemicals, pharmaceuticals, flavor and fragrance ingredients, and pesticides. One of the essential plant components, secondary metabolites, have significant economic value as pesticides, medications, food additives, fragrances, and pigments (Rao and Ravishankar, 2002). Secondary metabolites are known to play a significant role in helping plants adapt to their surroundings and are a major source of pharmaceuticals (Siahsar *et al*., 2011).

The variability of the black cumin genotype based on essential oil volatile compounds was reported by different authors from different countries (Edris, 2010; Palabıyık and Aytaç, 2018; Sileshi and Biruk, 2020). Thymoquinone was reported as the first main component of black cumin seeds essential oils (Edris, 2010; Palabıyık and Aytaç, 2018; Albakry *et al*., 2022). In addition, ρ-cymene was reported as the first main component of the three Ethiopian black cumin varieties of seeds essential oil (Sileshi and Biruk, 2020). Thymoquinone has been used worldwide for its biological activities among the main components of black cumin seeds essential oil with the highest share.

For the past 3000 years, black cumin has been cultivated in Ethiopia. Currently, it has been widely distributed in the Oromia region (Bale and Arsi zone), Amhara region (East Gojjam, West Gojjam, North Gondar, South Gondar, North Wollo and South Wollo zones), Tigray region (Central, Western and Northwestern zones), Benshangul-Gumuz region (Metekel and Asosa zones) and Southern, Nations and Nationalities Peoples Region (Keficho-Shekicho zone) of Ethiopia. It has been used as a flavor in bread and sauces, as well as an ingredient in the "berbere" spice mix.

In Ethiopia, whole seeds were used for a long period and fixed oil (oleoresin) of black cumin has been used recently for its spicy and medicinal values. The variability of some of the existing genotypes was made by their compositions of essential oil (Sileshi and Biruk, 2020), however, the genotypes involved in characterization is limited. Therefore, this would hinder the potential of the genotypes to be exploited for future breeding and improvement programs. The finding from this study was believed to show a clear variability of black cumin genotypes in the case of EOCs thereby, providing a piece of baseline information for further breeding and improvement programs. Thus, this study aimed to investigate the variability of Ethiopian black cumin genotypes by their essential oil compositions.

5.2. Materials and Methods

5.2.1. Genotypes collection

The matured dried seeds of 64 black cumin genotypes (8 improved varieties and 56 accessions) obtained from the first experiment were used for this experiment (Appendix Table 1).

5.2.2. Sample preparation and extraction of essential oil by hydrodistillation

Composite samples of 100 g black cumin seeds were collected and ground roughly from each genotype. Extraction was made in 2022 at Wondo Genet Agricultural Research Center, Natural Products Laboratory, by hydrodistillation using a Clevenger-type apparatus (Coffey *et al*., 1989) for 3 h, and the essential oil was collected in brown glass bottles by measuring the amount using a measuring pipette. The essential oil samples were kept in a refrigerator at 4^{o}C until Gas Chromatography-Mass Spectrometry (GC-MS) analysis.

5.2.3. GC-MS analysis of essential oil

The collected essential oil was dried with anhydrous Na_2SO_4 and 0.2% (w/v) of the essential oil solution was prepared by diluting in n-hexane and the instrument was conditioned with a split less injector mode. To identify the composition of essential oil, GC-MS analyses were performed on an Agilent GC model 7820A series equipped with an autosampler (Agilent model G4513A), coupled to a mass spectrometer detector (Agilent model 5975 series) quadrupole analyzer operating in the EI mode at 70ev. Helium was used as carrier gas (1.5 ml/minute) with a capillary column of HP-5ms (30 m × 0.25 mm id; 0.25 µm film thickness). The GC–MS transfer line temperature was set at 200°C. The column temperature program was the same as that used for the GC analyses. The inlet and ionization source temperatures were set at 250°C and 160°C, respectively. Electron-impact mass spectra were recorded over the range of 40–450 amu at 0.5 scan. The injected volume was 1 µl of the oil. The GC-MS run time for each germplasm was 40 minutes with a solvent delay time of 2 minutes. The quantitative composition of the essential oils was ascertained using GC-MS. Quantitatively, peak area approximated peak height approximation was used to express the relative amounts of the constituents of the essential oils as a percentage (Wang *et al*., 1996). The GC-MS analysis was done in 2022 at Wondo Genet Agricultural Research Center, Natural Products Laboratory (Appendix Figure 5).

5.2.4. Data Analysis

Descriptive statistics (mean and range) were used to estimate the variations among populations combined mean values of essential oil compositions using the SAS version 9.4 software package (SAS Institute Inc., 2019). The combined mean values of major volatile compounds were subjected to bivariate analysis (i.e. correlation analysis) and multivariate analysis (i.e. cluster and principal component analysis) methods using R-software version 4.2.2 (R Core Team, 2022).

Hierarchical cluster analysis was performed by using R-software packages “factoextra” (Kassambara and Mundt, 2020), “cluster” (Maechler *et al*., 2018), “class” (Venables and Ripley, 2002), and “clv” (Lukasz, 2020). Clustering was used to examine the aggregation patterns of the 64 black cumin genotype populations and the six groups of genotypes collected based on their similarity concerning the corresponding means of all the volatile compounds that were collected. The best number of clusters for the data set was determined by using an R-software package “Nbclust” (Charad *et al.,* 2014). The genetic distance was measured using Euclidean distance and the distance matrix was used to construct the dendrograms using Ward`s minimum variance method (Ward, 1963).

The PCA was employed by using R-software packages "factoextra", "ggplot2" (Wickham, 2016), "corrplot" (Wei and Sinko., 2021), and "ggsignif" (Ahlmann-Eltze, 2022) to identify the traits contributing to a large part of the total variation among the populations and groups.

5.3. Results and Discussion

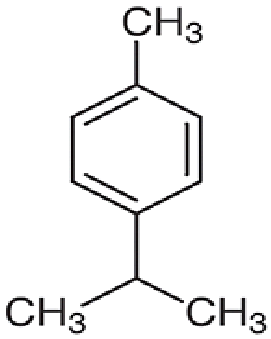
5.3.1. Chemical composition of black cumin essential oils

Essential oils extracted from 64 black cumin genotypes were characterized by GC-MS, allowing for the identification of 27 volatile compounds, of which 12 are major essential oil compounds covering 92–100% of the total essential oil composition (Table 5.1 and Appendix Table 8). So far 113, 85, 48, 40, 22, 20, and 19 volatile compounds have been detected in Algerian (Benkaci–Ali, 2007), Ethiopian (Sileshi and Biruk, 2020), Polish (Wajs *et al*., 2008), Iranian (Rezaei-Chiyaneh *et al*., 2021), Italian (D'Antuono *et al*., 2002), Tunisian (Bourgou *et al*., 2010) and Egyptian (Edris, 2010) black cumin, respectively. Among the major volatile compounds, ρ-cymene (23.93 – 45.61%), thymoquinone (14.81 – 42.44%), α-thujene (0.91 – 12.97%), carvacrol (3.01 – 11.48%), trans-4-methoxy thujane (4.82 – 11.29%), and longifolene (2.1 – 6.97%) were the most abundant volatile constituents found in most of the genotypes (Figure 5.1 and Table 5.1). Out of them, thymoquinone is the most important bioactive compound for phytomedicinal values. The highest thymoquinone content (67.7 and 63.3%) was reported by Edris (2010) and Palabıyık and Aytaç (2018) from Egyptian and Turkish black cumin, respectively. Black cumin essential oil is used for functional foods and nutraceuticals/pharmaceuticals (Mukhtar *et al*., 2019). Thus, Ethiopia is the third most important country next to Turkey and Egypt because of its rich potential of black cumin with the highest thymoquinone content to be exploited for functional foods and pharmaceuticals.

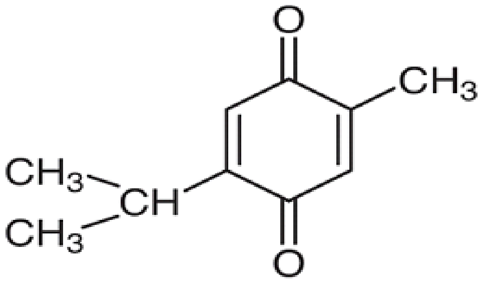
Table 5.1. Retention times, mean, and range of chemical compound (%) were identified across 64 black cumin genotypes of Ethiopia at each testing site and pooled during the 2021 cropping season

Chemical compounds	RT (min)	Debre Zeit		Kulumsa		Combined	
		Mean ± SE	Range	Mean ± SE	Range	Mean ± SE	Range
CC1	2.71	5.69 ± 0.45	0 - 14.25	5.82 ± 0.53	0.62 - 20.96	5.76 ± 0.33	0.91 - 12.97
CC2	2.80	1.30 ± 0.11	0 - 3.17	1.18 ± 0.12	0 - 4.47	1.24 ± 0.07	0 - 2.75
CC3	3.37	0.73 ± 0.06	0 - 1.67	0.50 ± 0.07	0 - 1.85	0.61 ± 0.05	0 - 1.35
CC4	3.42	1.68 ± 0.11	0 - 3.49	1.56 ± 0.09	0 - 3.36	1.62 ± 0.07	0 - 2.82
CC5	4.11	-	-	0.14 ± 0.04	0 - 1.2	0.07 ± 0.02	0 - 0.6
CC6	4.28	31.76 ± 0.79	13.54 - 46.13	38.56 ± 0.75	27.72 - 54.77	35.16 ± 0.56	23.93 - 45.61
CC7	4.34	1.22 ± 0.06	0 - 2.22	1.01 ± 0.07	0 - 2.49	1.11 ± 0.05	0.37 - 1.86
CC8	4.39	-	-	0.01 ± 0.01	0 - 0.94	0.01 ± 0.01	0 - 0.47
CC9	4.97	0.02 ± 0.02	0 - 0.97	0.72 ± 0.19	0 - 5.86	0.37 ± 0.09	0 - 2.93
CC10	5.82	1.33 ± 0.09	0 - 6.02	1.14 ± 0.07	0 - 2.02	1.24 ± 0.05	0 - 3.01
CC11	6.03	-	-	0.08 ± 0.05	0 - 2.26	0.04 ± 0.02	0 - 1.13
CC12	6.37	8.51 ± 0.18	5.39 - 12.15	9.05 ± 0.24	3.51 - 13.4	8.78 ± 0.15	4.82 - 11.29
CC13	6.99	-	-	0.06 ± 0.04	0 - 1.5	0.03 ± 0.02	0 - 0.75
CC14	7.63	0.01 ± 0.01	0 - 0.66	-	-	0.01 ± 0.01	0 - 0.33
CC15	7.89	1.68 ± 0.10	0 - 4.03	1.08 ± 0.12	0 - 3.02	1.38 ± 0.08	0 - 3.34
CC16	8.66	1.40 ± 0.06	0 - 2.63	1.06 ± 0.09	0 - 2.4	1.23 ± 0.06	0 - 2.46
CC17	10.05	37.15 ± 1.24	14.66 - 64.81	25.01 ± 1.05	0 - 37.14	31.08 ± 0.80	14.81 - 42.44
CC18	10.77	0.04 ± 0.03	0 - 1.3	0.04 ± 0.03	0 - 1.73	0.04 ± 0.02	0 - 0.87
CC19	11.34	-	-	0.17 ± 0.09	0 - 4.17	0.08 ± 0.04	0 - 2.09
CC20	11.59	3.44 ± 0.17	1.67 - 7	8.37 ± 0.44	0 - 16.6	5.91 ± 0.24	3.01 - 11.48
CC21	12.9	0.07 ± 0.04	0 - 1.58	0.02 ± 0.02	0 - 1.21	0.05 ± 0.02	0 - 1.22
CC22	14.48	3.68 ± 0.18	1.73 - 8.01	3.65 ± 0.16	1.64 - 8.24	3.67 ± 0.13	2.1 - 6.97
CC23	14.95	0.01 ± 0.01	0 - 0.85	0.06 ± 0.04	0 - 2.1	0.04 ± 0.02	0 - 1.05
CC24	15.95	-	-	0.03 ± 0.03	0 - 1.68	0.01 ± 0.01	0 - 0.84
CC25	17.34	0.05 ± 0.04	0 - 1.92	0.06 ± 0.04	0 - 1.93	0.05 ± 0.03	0 - 0.97
CC26	19.12	0.23 ± 0.07	0 - 2.02	0.45 ± 0.10	0 - 3.1	0.34 ± 0.07	0 - 2.05
CC27	19.91	-	-	0.16 ± 0.10	0 - 5.44	0.08 ± 0.05	0 - 2.72

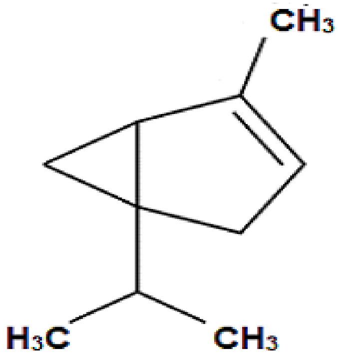
Where, RT = Retention time, SE = Standard error, CC = Chemical compound: CC1 = *α*-thujene, CC2 = *α*-pinene, CC3 = *β*-terpinene, CC4 = *β*-pinene, CC5 = 4-carene, CC6 = *ρ*-cymene CC7 = ***d***-limonene, CC8 = eucalyptol, CC9 = *γ*-terpinene, CC8 = terpinen-4-ol, CC9 = phthalic acid, CC10 = 2-ethylhexyl isohexyl phthalate, CC11 = thujone, CC12 = trans-4-methoxy thujane, CC13 = camphor, CC14 = bis(2-ethylhexyl) phthalate, CC15 = terpinen-4-ol, CC16 = phtalic acid, 2-fluorobenzyl heptyl ester, CC17 = thymoquinone, CC18 = 2-(octyloxycarbonyl) benzoic acid, CC19 = thymol, CC20 = carvacrol, CC21 = α-longipinene, CC22 = longifolene, CC23 = 1,2-benzenedicarboxylic acid, mono(2-ethylhexyl) ester, CC24 = humulene, CC25 = 1-(2-ethyl-[1,3]dithian-2-yl)-3-methyl-butan-1-ol, CC26 = 1,4-dimethoxy-2,3-dimethylbenzene, and CC27 = naphthalene, decahydro-4a-methyl-1-methylene-7-(1-methylethylidene)-, (4aR-trans)-



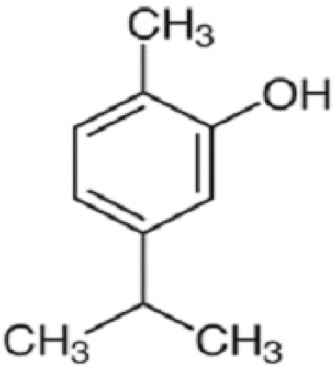
ρ-cymene



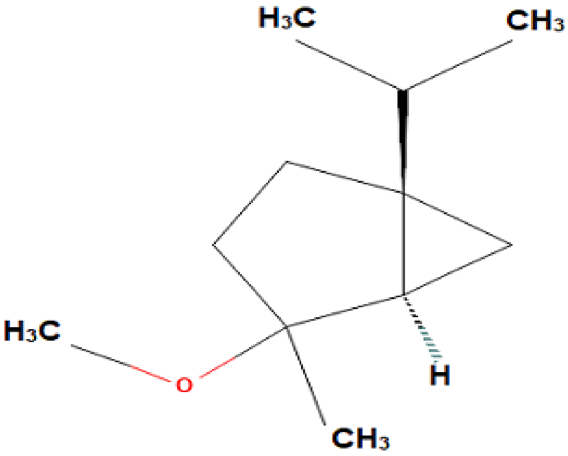
thymoquinon



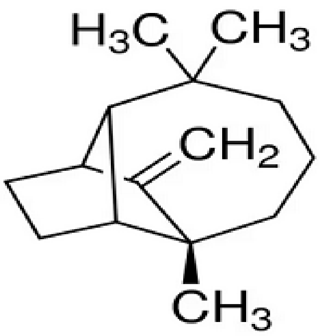
α-thujene



carvacrol



trans-4-methoxy thujane



longifolene

Figure 5.1. Chemical structures of the six most abundant volatile compounds of black cumin (*N. sativa* L.) essential oils from Ethiopia.

Pico-cymene was the first dominant constituent ranging from 23.93% (9068) to 45.61% (242221) across genotypes with an average of 35.16%. Six genotypes, 242221 (45.61%), 8502 (45.3%), 9069 (44.44%), 242825 (42.63%), 242839 (42.52%), and 90510 (40.71%), recorded more than 40% *ρ*-cymene content (Appendix Table 8). The result is in line with the report of Nickavar *et al.* (2003) and Rezaei-Chiyaneh *et al.* (2021), who reported that one of the major volatile constituents of Iranian Black cumin seeds was *ρ*-cymene (14.80% and 15.59%, respectively). Similarly, D'Antuono *et al.* (2002), Benkaci–Ali (2007), Wajs *et al.* (2008), Palabıyık and Aytaç (2018), Ramadan (2015), Kabir *et al*. (2020), Sileshi and Biruk (2020), and Farhan *et al.* (2021) reported *ρ*-cymene as one of the major volatile constituents of black cumin seeds. Pico-cymene is an important volatile constituent used in pharmaceutical industries for the production of anti-inflammatory (De Souza Siqueira Quintans *et al*., 2013), anticancer (Zhou *et al*., 2013), antioxidant (Bibi Sadeer *et al*., 2019; Luisi *et al*., 2019), and antimicrobial (Mouwakeh *et al*., 2019) drugs.

Thymoquinone (TQ) was the second most abundant constituent, ranging from 14.81% (8502) to 42.44% (242835) across genotypes with an average of 31.08%. Four genotypes, 242835 (42.44%), 9068 (41.14%), 242841 (40.35%), and 014_ATH (40.01%) recorded more than 40% TQ content (Appendix Table 8). Authors from different countries reported TQ among the main constituents in the essential oil of black cumin seeds (D'Antuono *et al*., 2002; Benkaci–Ali, 2007; Bourgou *et al*., 2010; Edris, 2010; Ali and Hassan, 2014; Ramadan, 2015; Palabıyık and Aytaç, 2018; Kabir *et al*., 2020, Sileshi and Biruk, 2020; Farhan *et al*., 2021; Albakry *et al*., 2022). Thymoquinone has been reported to be used as a natural drug traditionally (Khader and Eckl, 2014). Wang *et al.* (2021) reported TQ as an effective natural antimicrobial preservative that has potential applications in regulating food contamination and foodborne diseases caused by *Bacillus cereus*. Thymoquinone is an important volatile constituent used in pharmaceutical industries for the production of antibacterial, antioxidant (Bourgou *et al*., 2010), anti-inflammatory (Salem, 2005), anticancer (Shafiq *et al*., 2014), hepatoprotective (Ghadlinge *et al*., 2014), antidiabetic (Heshmati and Namazi, 2015), fertility-enhancing (Tüfek *et al*., 2015; Akour *et al*., 2016), analgesic (Huseini *et al*., 2015), and antimicrobial (Mouwakeh *et al*., 2019) drugs.

Alpha-thujene was the third dominant constituent ranging from 0.91% (9068) to 12.97% (242221) across genotypes with an average of 5.76%. Six genotypes, 242221 (12.97%), 90501 (11.99%), 012_ATH (11.13%), 90510 (10.62%), 229808 (10.42%), and 90504 (10.41%), recorded more than 10% α-thujene content (Appendix Table 8). This result is in line with the findings of Bourgou *et al.* (2010), Ramadan (2015), Kabir *et al*. (2020), and Sileshi and Biruk (2020), who reported α-thujene as one of the major volatile constituents from Turkish, Egyptian, and Ethiopian black cumin seeds, respectively. Alpha-thujene is an important volatile constituent used in pharmaceutical industries to produce antimicrobial drugs (Gerige *et al*., 2009).

Carvacrol was the fourth abundant constituent ranging from 3.01% (90504) to 11.48% (242838) across genotypes with an average of 5.91%. Four genotypes, 242838 (11.48%), Silingo (10.08%), 007_ATH (9.75%), and 9068 (9.23%), recorded more than 9% carvacrol content (Appendix Table 8). Supporting results were reported by Benkaci–Ali (2007), Palabıyık and Aytaç (2018), Ramadan (2015), and Rezaei-Chiyaneh *et al.* (2021) from Algeria, Turkey, Egypt, and Iran with respective values of 12.90%, 8.40%, 2.12%, and 4.65%. Similarly, Kabir *et al*. (2020) reported carvacrol as the fourth abundant volatile compound of Bangladesh and Indian black cumin, with respective values of 2.25% and 3.65%. Rattanachaikunsopon and Phumkhachorn (2010) reported that carvacrol has a good inhibitory effect against *Vibrio cholerae*. Carvacrol is an important volatile constituent used in pharmaceutical industries to produce antioxidant (Huseini *et al*., 2015) and antimicrobial (Mouwakeh *et al*., 2019) drugs.

Trans-4-methoxy thujane was the fifth dominant constituent ranging from 4.82% (242844) to 11.29% (242825) across genotypes with an average of 8.78%. Ten genotypes: 242825 (11.29%), 9068 (11.22%), 007_ATH (10.98%), 003_ATH (10.93%), 242838 (10.70%), 015_ATH (10.53%), 019_ATH (10.34%), 004_ATH (10.31%), 009_ATH (10.16%) and 010_ATH (10.07%) recorded more than 10% trans-4-methoxy thujane content (Appendix Table 8). This result agrees with the finding of Sileshi and Biruk (2020), who reported that one of the major volatile constituents of Ethiopian Black cumin seeds was trans-4-methoxy thujane (8.86%).

Longifolene was the sixth dominant constituent ranging from 2.1% (90501) to 6.97% (004_ATH) across genotypes with an average of 3.67%. Seven genotypes, 004_ATH (6.97%), 242835 (6.27%), 019_ATH (5.99%), 003_ATH (5.86%), 9068 (5.75%), 242838 (5.52%), and 007_ATH (5.42%), recorded more than 5% longifolene content (Appendix Table 8). Supporting results were reported by Ramadan (2015) and Sileshi and Biruk (2020) from Egypt and Ethiopia, respectively. Longifolene is an important volatile constituent used in pharmaceutical industries to produce antibacterial drugs (Bourgou *et al*., 2010).

The essential oil of black cumin seeds has different biological activities due to the presence of various kinds of volatile compounds. Among the major volatile compounds, thymoquinone is used for antioxidant, antiparasitic, hepatoprotective, antidiabetic, analgesic, anticancer, antimicrobial, anti-inflammatory, antibacterial, and fertility-enhancing activities. Carvacrol is another major volatile compound that has antioxidant and antimicrobial activity. Furthermore, *p*-cymene and longifolene have antimicrobial and antibacterial activity, respectively.

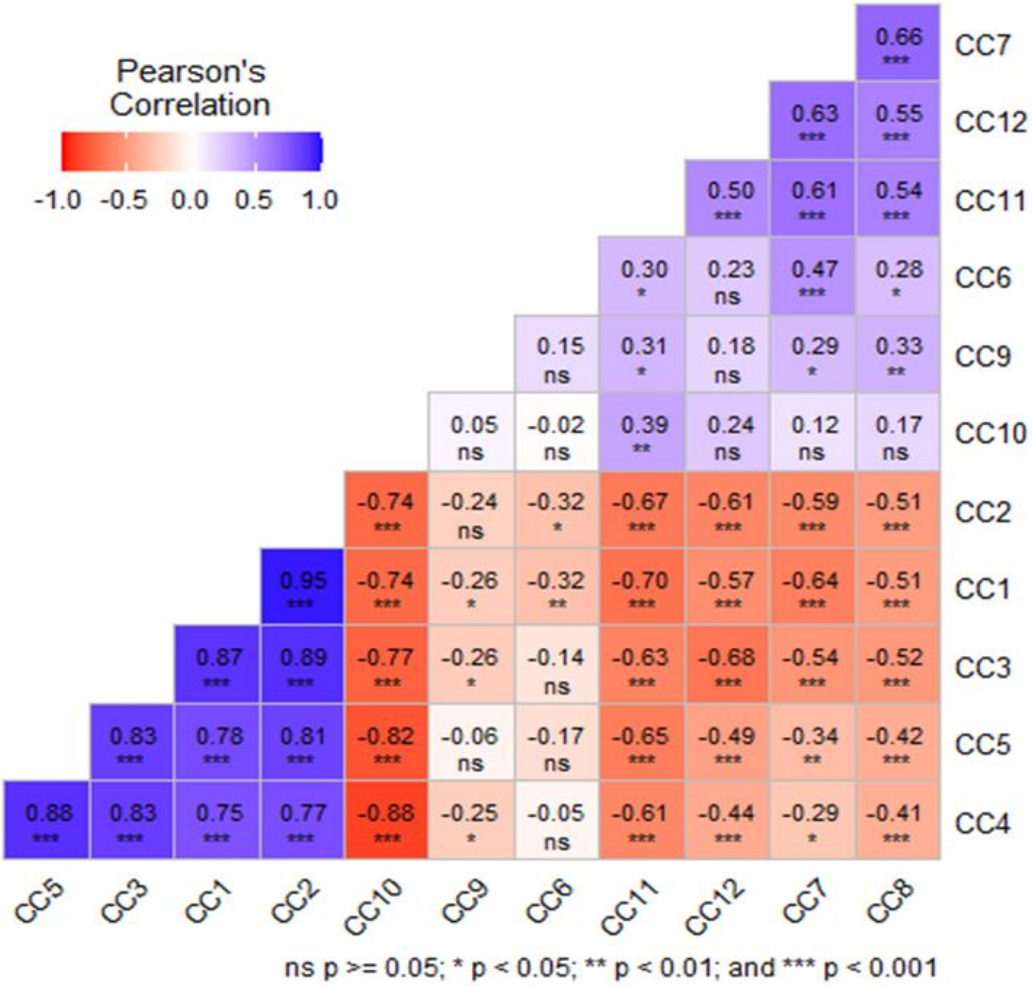
The wide range of major volatile compounds studied witnessed the presence of variability among the black cumin genotypes (Tables 5.1 and 5.2). This shows that there is a possibility of improving these traits through selection. Hence, it is expected to improve the biochemical traits of the top 5% of the genotypes by 13.2% to 152.1% through selection (Table 5.2). Based on this, genotypes 242835 (from Oromia), 9068, and 014_ATH (from Amhara) were the top 5% best-performed accessions over improved varieties selected for the production of major volatile compounds (Appendix Table 8). Therefore, they can be directly promoted for the production of those major volatile compounds.

Table 5.2. Comparison of mean performances of 5% of the best-performed accessions selected for 12 major volatile compounds over the mean performance of improved varieties

No.	Traits	Mean values		A comparative advantage over the mean values of improved varieties (%)
		Top 5% accessions	Improved varieties	
1.	α-thujene	12.03	5.33	125.70
2.	*α*-pinene	2.55	1.18	116.10
3.	*β*-pinene	2.70	1.66	62.65
4.	*ρ*-cymene	45.12	33.86	33.25
5.	*d*-limonene	1.77	0.95	86.32
6.	2-ethylhexyl isohexyl phthalate	1.72	1.52	13.16
7.	trans-4-methoxy thujane	11.17	8.84	26.36
8.	terpinen-4-ol	2.95	1.17	152.14
9.	phthalic acid, 2-fluorobenzyl heptyl ester	2.05	1.26	62.70
10.	thymoquinone	41.51	33.12	25.33
11.	carvacrol	10.15	6.65	52.63
12.	longifolene	6.41	3.17	102.21

5.3.2. Relationship among volatile compounds

A correlation analysis was made among the 12 major volatile compounds to determine the degree and direction of the relationship among them (Figure 5.2). Alpha-thujene had a strong positive and significant correlation with *α*-pinene (r = 0.95***), *β*-pinene (r = 0.87***), *d*-limonene (r = 0.78**) and *ρ*-cymene (r = 0.75***); are negatively correlated with thymoquinone (r = - 0.74***), carvacrol (r = - 0.70***), trans-4-methoxy thujane (r = - 0.64***), longifolene (r = - 0.57***), and terpinen-4-ol (r = - 0.51***). The significant positive correlation of *α*-thujene with *α*-pinene, *β*-pinene, *d*-Limonene, and *ρ*-cymene indicated that the higher the genotype with *α*-pinene, *β*-pinene, *d*-limonene, and ρ-cymene, the higher would be the *α*-thujene content. They have a direct relationship. Therefore, those compounds can be improved simultaneously. The major volatile chemical constituent *ρ*-cymene correlated positively with *d*-limonene (r = 0.88***), and negatively with thymoquinone (r = - 0.88***), carvacrol (r = - 0.61***), longifolene (r = - 0.44***), and terpinen-4-ol (r = - 0.41***). Thymoquinone had a positive correlation with longifolene (r = 0.24), terpinen-4-ol (r = 0.17) and trans-4-methoxy thujane (r = 0.12); a significant positive correlation with carvacrol (r = 0.39**); and a significant negative correlation with *ρ*-cymene (r = - 0.88***), *d*-limonene (r = - 0.88***), *β*-pinene (r = - 0.77***), α-thujene (r = - 0.74***) and *α*-pinene (r = - 0.74***). The significant negative correlation of thymoquinone with *ρ*-cymene, *d*-limonene, *β*-pinene, *α*-thujene, and *α*-pinene indicated that the higher the genotype with *ρ*-cymene, *d*-limonene, *β*-pinene, *α*-thujene, and *α*-pinene content, the lower the content of thymoquinone. This implied that those traits cannot be improved simultaneously. However, it is possible in the case of thymoquinone and carvacrol.



	CC5	CC3	CC1	CC2	CC10	CC9	CC6	CC11	CC12	CC7	CC8
CC7											0.66 ***
CC12										0.63 ***	0.55 ***
CC11									0.50 ***	0.61 ***	0.54 ***
CC6								0.30 *	0.23 ns	0.47 ***	0.28 *
CC9							0.15 ns	0.31 *	0.18 ns	0.29 *	0.33 **
CC10						0.05 ns	-0.02 ns	0.39 **	0.24 ns	0.12 ns	0.17 ns
CC2					-0.74 ***	-0.24 ns	-0.32 *	-0.67 ***	-0.61 ***	-0.59 ***	-0.51 ***
CC1				0.95 ***	-0.74 ***	-0.26 *	-0.32 **	-0.70 ***	-0.57 ***	-0.64 ***	-0.51 ***
CC3			0.87 ***	0.89 ***	-0.77 ***	-0.26 *	-0.14 ns	-0.63 ***	-0.68 ***	-0.54 ***	-0.52 ***
CC5		0.83 ***	0.78 ***	0.81 ***	-0.82 ***	-0.06 ns	-0.17 ns	-0.65 ***	-0.49 ***	-0.34 **	-0.42 ***
CC4	0.88 ***	0.83 ***	0.75 ***	0.77 ***	-0.88 ***	-0.25 *	-0.05 ns	-0.61 ***	-0.44 ***	-0.29 *	-0.41 ***

Figure 5.2. The correlation coefficient of the 12 major volatile compounds of 64 Black cumin genotypes of Ethiopia. (CC1 = α-thujene, CC2 = α-pinene, CC3 = β-pinene, CC4 = ρ-cymene, CC5 = *d*-limonene, CC6 = 2-ethylhexyl isohexyl phthalate, CC7 = trans-4-methoxy thujane, CC8 = terpinen-4-ol, CC9 = phthalic acid, 2-fluorobenzyl heptyl ester, CC10 = thymoquinone, CC11 = carvacrol, CC12 = longifolene).

5.3.3. Cluster analysis

Cluster analysis was done based on 12 major volatile compounds of the 64 black cumin genotypes of Ethiopia to examine the extent and pattern of the genotypes based on their similarity. The relevant number of clusters for the data set was determined according to Charrad *et al.* (2014). Based on this, the ward`s minimum variance clustering method was used to classify the 64 black cumin genotypes of Ethiopia into two different clusters (Figure 5.3). The number and name of genotypes in each cluster with their collection regions were presented in Appendix Table 1.

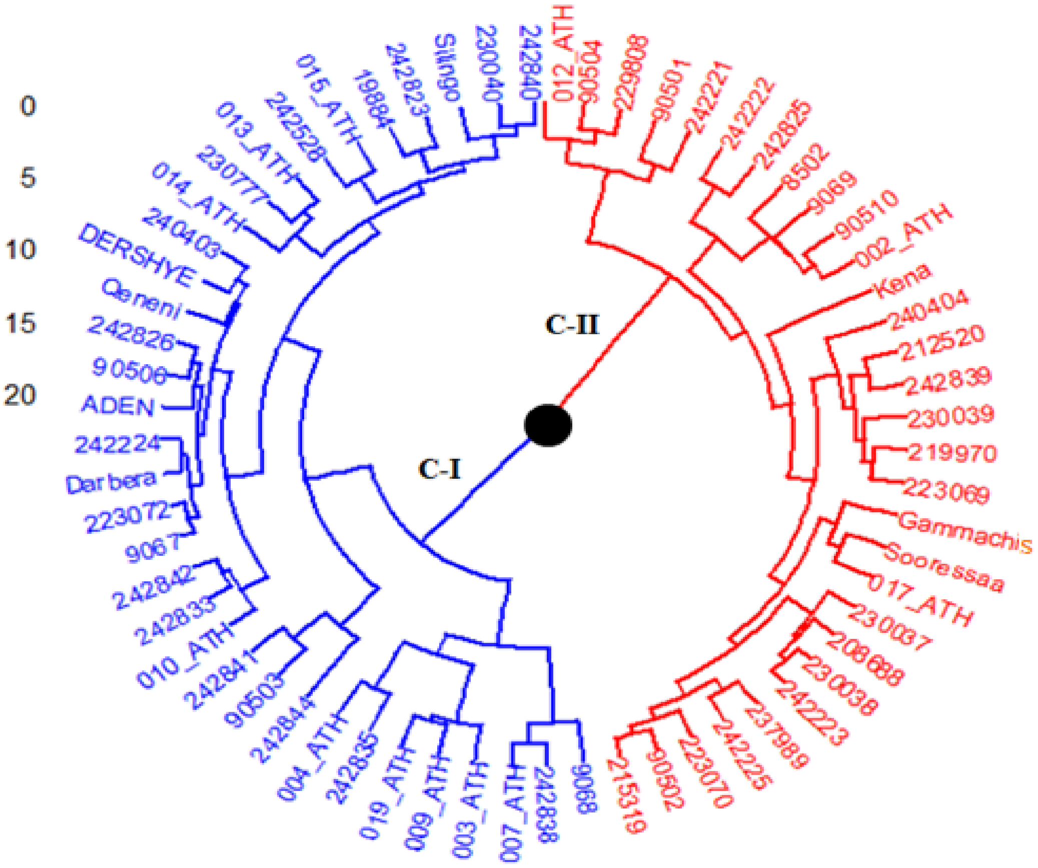


Figure 5.3. Phylogram showing the relationship between the 64 Black cumin genotypes of Ethiopia based on the 12 major volatile compounds using the Euclidean similarity index. (Blue and red colors represent clusters I and II, respectively.)

Cluster I: This was the largest group, having 34 (53.12%) genotypes comprised from all collection regions and most of the improved varieties (Figure 5.3 and Table 5.3). The group was characterized by a significant content of thymoquinone/carvacrol/longifolene, which represents chemotype A (Table 5.4). Genotype 242835 (from Oromia) contained a considerably high content of thymoquinone (42.44 %), followed by 9068 (41.14%) and 014_ATH (40.35%). Considerably high content of carvacrol was present in genotype 242838 (11.48%), followed by Silingo (10.08%) and 9068 (9.23%). Longifolene content was considerably high in genotype

014_ATH (6.97%) followed by 242835 (6.27%) and 019_ATH (5.99%) (Appendix Table 8). Genotypes better fitted for the production of pharmaceutical products may be identified through selection in this group by considering its chemotype. This group is highly suggested for the production of pharmaceutical products.

Cluster II: About 30 (46.88%) genotypes from all collection regions and some of the improved varieties were included in this group (Figure 5.3 and Table 5.3) and is characterized by the higher mean value of *ρ*-cymene/*α*-thujene, which represents chemotype B (Table 5.4). Genotype 242221 (from Amhara) contained a considerably high content of *ρ*-cymene (45.61%), followed by 8502 (45.30%) and 9069 (10.98%). Considerably high content of *α*-thujene (12.97%) was present in genotype 242221 (from Amhara), followed by 90501 (11.99%) and 012_ATH (11.13%) (Appendix Table 8). Chromatograms of the compounds representing the different clusters are shown in Figure 5.4.

Table 5.3. Clustering of 64 black cumin genotypes of Ethiopia into two clusters using the mean of 12 major volatile compounds

Cluster	Number of genotypes	Genotypes included	Collection region
I	34 (53.12 %)	242840, 230040, Silingo, 242823, 19884, 015_ATH, 242528, 013_ATH, 230777, 014_ATH, 240403, DERSHYE, Qeneni, 242826, 90506, ADEN, 242224, Darbera, 223072, 9067, 242842, 242833, 010_ATH, 242841, 90503, 242844, 004_ATH, 242835, 019_ATH, 009_ATH, 003_ATH, 007_ATH, 242838, 9068	Improved, SNNP, Oromia, Amhara, B/Gumuz, Tigray
II	30 (46.88 %)	012_ATH, 90504, 229808, 90501, 242221, 242222, 242825, 8502, 9069, 90510, 002_ATH, Kena, 240404, 212520, 242839, 230039, 219970, 223069, Gammachis, Sooressaa, 017_ATH, 230037, 208688, 230038, 242223, 237989, 242225, 223070, 90502, 215319	Improved, SNNP, Oromia, Amhara, B/Gumuz, Tigray

Where, B/Gumuz = Benshangul-Gumuz

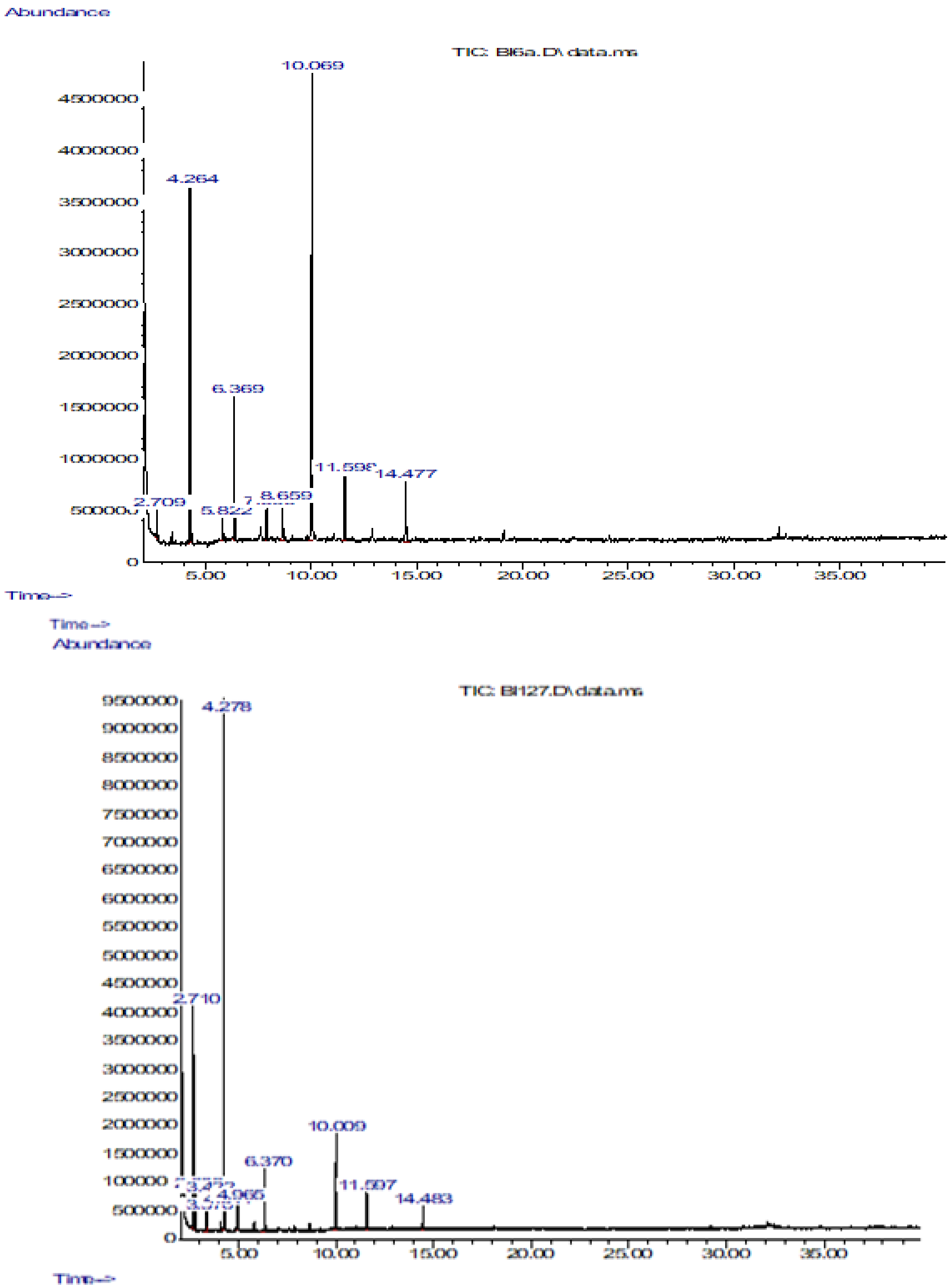


Figure 5.4. GC-MS chromatogram of Ethiopian black cumin essential oils for each cluster.

Table 5.4. The mean value of 12 major volatile compounds (%) of the groups of Ethiopian black cumin genotypes

No.	Compounds	Cluster	
		I	**II**
1.	*α*-thujene	3.82	**7.30**
2.	*α*-pinene	0.87	1.62
3.	*β*-pinene	1.36	2.05
4.	*ρ*-cymene	31.97	**38.82**
5.	*d*-limonene	0.95	1.41
6.	2-ethylhexyl isohexyl phthalate	1.34	1.19
7.	trans-4-methoxy thujane	9.18	8.14
8.	terpinen-4-ol	1.65	0.97
9.	phthalic acid, 2-fluorobenzyl heptyl ester	1.51	1.22
10.	thymoquinone	**34.72**	26.41
11.	carvacrol	**6.57**	4.46
12.	longifolene	**3.74**	3.26

Volatile compounds diversity in groups of genotypes

The genotypes from Amhara, Oromia, Tigray, SNNP regions, and improved varieties were spread into the two clusters but in different proportions (Table 5.5). However, the genotypes from the Benshangul-Gumuz region were spread into the two clusters with the same proportion. More than 50% of the genotypes from Amhara, Oromia, and SNNP regions, and improved varieties grouped under the first cluster.

Table 5.5. Clustering of 64 black cumin genotypes based on collection groups

Groups	Cluster		Total
	I	II	
Amhara	13	11	24
Oromia	10	8	18
Tigray	1	5	6
SNNP	3	1	4
Benshangul-Gumuz	2	2	4
Improved	5	3	8
Total	34	30	64

SNNP = Southern Nations, Nationalities and Peoples.

5.3.4. Principal component analysis

The principal component analysis grouped the 12 major volatile compounds into 12 principal components (PCs), which explained the entire 100% of the variability among the studied genotypes. Kaiser (1960) suggested that an eigenvalue > 1 indicates that PCs account for more variance than accounted by one of the original variables in standardized data commonly used as a cutoff point for which PCs are retained. Based on this, the first two PCs having eigenvalues between 1.85 and 6.80 explained 72.03% of the variation that existed among the studied genotypes in volatile compounds (Table 5.6).

The influential PCs for clustering were the traits with the largest absolute values closer to unity than those with lower absolute values closer to zero (Chahal and Gosal, 2002). Accordingly, the first principal component (PC) explained about 56.65% of the total variance due to the main contribution of the variation in *α*-pinene, *α*-thujene, *β*-pinene, *ρ*-cymene, *d*-limonene, and carvacrol (Table 5.6). Thymoquinone, trans-4-methoxy thujane, 2-ethylhexyl isohexyl phthalate, terpinen-4-ol, and *ρ*-cymene were the main contributor traits for the variation in the second PC, which contributed 15.38% from the total variations. The principal component analysis showed that there was significant variability among the 64 black cumin genotypes in most of the 12 major volatile compounds considered. Therefore, these traits should be given special emphasis during the selection for the improvement program of Ethiopian black cumin.

Table 5.6. The eigenvalues and eigenvectors of the first two principal components (PCs) for 12 major volatile compounds of 64 black cumin genotypes of Ethiopia

Variable	Principal components	
	PC1	PC2
Eigenvalue	6.80	1.85
Proportion of variance (%)	56.65	15.38
Cumulative variance (%)	56.65	72.03
	Eigenvectors	
α-thujene	**0.36**	-0.03
α-pinene	**0.36**	-0.05
β-pinene	**0.36**	-0.12
ρ-cymene	**0.33**	**-0.29**
d-limonene	**0.33**	-0.26
2-ethylhexyl isohexyl phthalate	-0.12	**-0.41**
trans-4-methoxy thujane	-0.26	**-0.45**
terpinen-4-ol	-0.25	**-0.34**
phthalic acid, 2-fluorobenzyl heptyl ester	-0.12	-0.27
thymoquinone	-0.28	**-0.46**
carvacrol	**-0.30**	-0.14
longifolene	-0.27	-0.22

5.4. Conclusions

Using GC-MS analysis of sixty-four genotypes of black cumin essential oil, twenty-seven compounds were identified. Based on the results, the essential oils that made up 92% to 100% of the total composition were *ρ*-cymene, *α*-thujene, carvacrol, trans-4-methoxy thujane, longifolene, terpinen-4-ol, 2-ethylhexyl isohexyl phthalate, *β*-pinene, *α*-pinene, phthalic acid, 2-fluorobenzyl heptyl ester, and *d*-limonene. The most prevalent components across all genotypes were five compounds: trans-4-methoxy thujane, *α*-thujene, *ρ*-cymene, thymoquinone, and carvacrol. The remaining constituents differed depending on the genotype. Choosing the top 5% of black cumin accessions was predicted to improve all major volatile compounds by 13.2% to 15.1% compared to improved varieties. Thymoquinone exhibited a strong and significant negative correlation with *α*-thujene, *α*-pinene, *β*-pinene, *ρ*-cymene, and *d*-limonene and a significant positive correlation with carvacrol. The first two principal components (PCs) accounted for 72.03% of the total variance, according to the principal component analysis of 12 major volatile constituents. This demonstrated that the Ethiopian black cumin genotypes in EOCs exhibit significant variability. Cluster analysis grouped the 64 genotypes into two chemotypes. Chemotype A has a high thymoquinone content, while Chemotype B has a high *ρ*-cymene concentration. The discovered high-volatile constituent variabilities demonstrated considerable diversity across the 64 Ethiopian black cumin genotypes, which can be used in future breeding and improvement programs. The most abundant volatile constituents, thymoquinone and ρ-cymene, are highly applicable to pharmaceutical and nutraceutical industries. Therefore, genotypes 242835, 9068, and 014_ATH are promising for pharmaceutical and food industries due to the presence of the most important volatile constituents. This finding provided baseline information for the beneficiaries to conserve, utilize, and improve the genotypes.

CHAPTER 6

6. Genetic Diversity and Population Structure of Ethiopian Black Cumin (*Nigella sativa* L.) as Revealed by DArTseq SNP Markers

Abstract

Black cumin (Nigella sativa L.) is one of the most important medicinal and aromatic plants in the Ranunculaceae family. The crop has a long history of cultivation and use in Ethiopia. Ethiopia is also considered a center of diversity. There is less knowledge about DNA fingerprinting in black cumin utilizing third-generation marker systems such as SNPs. In this study, we characterized the genetic diversity and population structure of 94 black cumin genotypes collected from various growing regions in Ethiopia using 1,391 high-quality SNPs derived from DArTseq. Genetically, the genotypes were grouped into two. The analysis of molecular variance (AMOVA) revealed the existence of higher levels of genetic diversity among genotypes within populations and within genotypes (74.3% and 14.3%, respectively) than among populations (11.4%). STRUCTURE analysis, discriminant analysis of principal components, and principal coordinate analysis clustered the black cumin genotypes into two well-differentiated groups. The unweighted pair group method with arithmetic means (UPGMA) was used to do clustering. The clustering pattern was independent of the geographic origin of the genotypes. These findings revealed substantial genetic variation in black cumin and a pattern of diversity consistent with its long history of cultivation, a decentralized and diversified selection of the crop in Ethiopia.

Keywords: DArTseq, Diversity, Molecular variance, *Nigella sativa* L., Population structure, SNP

6.1. Introduction

Black cumin is the world's most important and widely used aromatic and medicinal plant. Its whole seeds and oils (fixed and volatile) are used worldwide for functional foods and nutraceuticals (Javed *et al*., 2012) and to treat respiratory conditions like asthma, emphysema, and bronchitis (Atta, 2003).

Plant identification may be crucial for production and processing, intellectual property rights enforcement, and plant breeding. In genetic analyses related to evolution and conservation as well as germplasm selection in plant breeding, the identification of genetic relationships is necessary (Henry, 2012). Molecular markers are effective tools frequently used to characterize genetic resources and provide fundamental knowledge about population genetics, evolution, and ecology (Arzani and Ashraf, 2016).

The smallest possible genetic variation unit is a single nucleotide. The analysis of SNP (single nucleotide polymorphism) markers has increased with the availability of more sequence data (Henry and Edwards, 2009). Measurements of insertions and deletions (indels) are commonly made using SNP analysis (Henry, 2012). An SNP marker is a co-dominant DNA marker essential for plant breeding because it makes germplasm analysis and feature mapping quick and inexpensive. These differences deepen our understanding of genetics, which could change how we develop new cultivars. The fact that the desired trait is genetically controlled allows for the faster pursuit of phenotypic experiments. In addition to carrying out early trait selection, the breeder can also spread the desired allele to numerous populations (Morgil *et al*., 2020). There is a lack of knowledge regarding DNA fingerprinting in black cumin, which calls for more research. Thus far, single nucleotide polymorphism (SNP) molecular markers have not been used to evaluate the genetic diversity of the black cumin genotypes.

Knowledge regarding the extent of genetic variation and genetic relationships between genotypes provides valuable information for understanding the relationships between genotypes. It facilitates their characterization and classification, determination of population structure, and so on thereby providing information for designing sound breeding strategies for crop improvement. Genetic diversity analysis of plants is a critical component of plant genetics, breeding, conservation, and evolution (Peterson *et al*., 2014). Therefore, the present study aimed to reveal Ethiopian black cumin's genetic diversity and population structure using DArTseq SNP markers.

6.2. Materials and Methods

6.2.1. Experimental materials

The seeds of 94 black cumin collections were used in the study. Out of this the seeds of 64 black cumin accessions were obtained from research institutes (Ethiopian Institute of Agricultural Research and Oromia Agricultural Research Institute) and the remaining 30 black cumin accessions seeds were obtained from the Ethiopian Biodiversity Institute (EBI). These accessions were comprised of 8 improved varieties, 36 from Oromia, 25 from Amhara, 11 from SNNP, 10 from Tigray, and 4 from Benshangul-Gumuz regions (Appendix Table 9).

The black cumin test materials used in this study, hereafter called genotypes, are considered under different populations, and each grouping, based on geographic regions and genetic groups, is designated as a sub-population. The original collection areas of the black cumin collections in Ethiopia are given in Figure 6.1 which was constructed by using ArcGIS Desktop Advanced version 10.8 (ESRI, 2011).

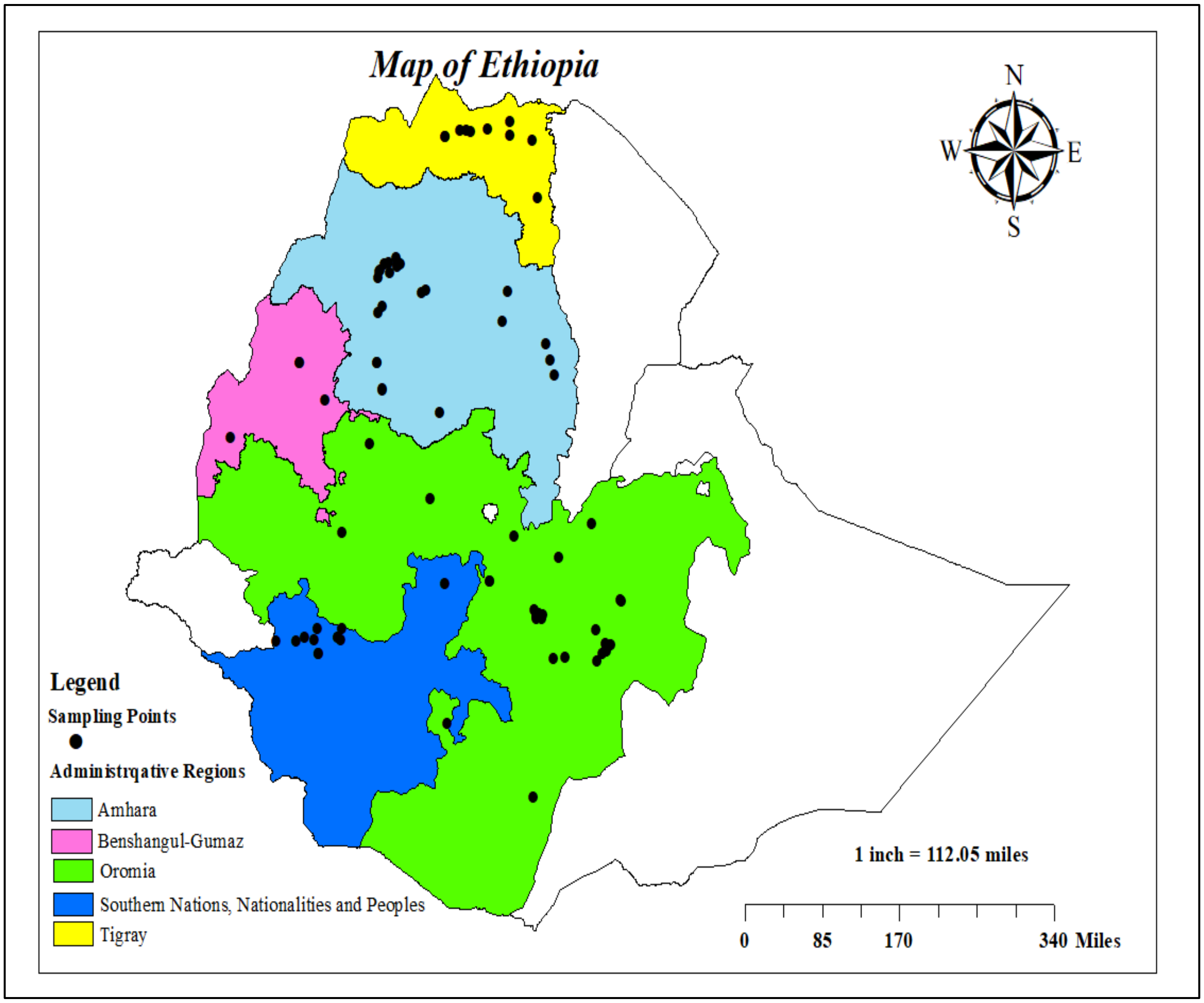


Figure 6.1. The diagram shows the collection sites for the 94 black cumin genotypes from different eco-geographical regions.

6.2.2. Genomic DNA extraction and sequencing

Genomic DNA (gDNA) of each genotype was extracted by the high-quality laboratory facilities of SEQART AFRICA located at Biosciences Eastern and Central Africa (BecA-ILRI) Hub in Nairobi, Kenya. Black cumin seeds were ground using a Geno/Grinder, and a NucleoMag plant

DNA extraction kit was used to extract the DNA. The quantity of extracted DNA was checked using a Thermo Scientific NanoDrop Spectrophotometer 2000c. The quality of the DNA was confirmed on 0.8% agarose gel run in 1% TAE buffer at 70V for 45 minutes.

As described by Elshire *et al.* (2011), 40µl of a 50ng/µl gDNA of each sample of 94 black cumin genotypes was sent for whole genome scanning using genotyping-by-sequencing (GBS) technology using DArTseq^{TM} technology of the Integrated Genotype Service and Support platform (https://www.diversityarrays.com/) after the quality was checked. Genotyping-by-sequencing was performed following protocols described in Wells and Dale (2018) by using a combination of DArT complexity reduction methods and next-generation sequencing (HiSeq 2500). The complexity reduction method involved digestion with the methylation-sensitive restriction enzymes. In conjunction with digestion using this relatively rarely cutting restriction enzyme (Gruenbaum *et al*., 1981), the enzymes PstI and MseI that have frequent cutting capabilities were used. Polymerase chain reaction (PCR) adapters were ligated to the PstI and MseI fragment ends, and the PCR amplification was performed using primers complementary to the PstI and MseI adapters. Only those fragments with adapters at both ends were amplified.

6.2.3. SNP calling, imputation and filtering

Diversity array technology (DArT) provides two different dataset formats: SilicoDArTs are scored in a binary fashion, representing genetically "dominant" markers, with "1" for presence and "0" for absence of a restriction fragment with the marker sequence in the genomic representation of the sample; "-" represents calls with non-zero counts, but too low counts to score confidently as "1" (often representing heterozygotes). The second format was SNPs in one-row format ("0" for reference allele homozygote, "1" for SNP allele homozygote, "2" for heterozygote, and "-" for double null/null allele homozygote (absence of fragment with SNP in genomic representation)) and two-row format (Each allele scored in a binary fashion ("1" for presence and "0" for absence)).

However, for this study, a one-row SNP dataset format was used to determine the genetic diversity and population structure. The alignment was made with the reference genome of *Nigella damascena,* which is present in the National Center for Biotechnology Information (NCBI) database with the SRR13275798 code (https://trace.ncbi.nlm.nih.gov/Traces/?-view=run_brow-serandacc=SRR13275798anddisplay=download). Data imputation was employed for imputing missing values first by a method of k-nearest neighbors (KNN) using R software version 4.4.1. (R Core Team, 2024). SNPs were filtered based on bi-allelic variants following the DArTseq^{TM} technology based on the following criteria; (i) markers and individuals with reproducibility less than 99%; (ii) call rate less than or equal to 95%, and (iii) minor allele frequency (MAF) less than 5% were discarded from further analysis using “BiocManager” (Morgan and Ramos, 2024) and “dartR” (Mijangos *et al.*, 2022) packages in R software version 4.4.1. (R Core Team, 2024). The most informative SNPs were selected based on the above criteria. The major allele frequencies and polymorphic information content (PIC) were performed using Power Marker Software v.3.25 (Liu and Muse, 2005).

6.2.4. Population structure analysis

To infer the population structure of Ethiopian black cumin genotypes, three complementary methods, such as Bayesian model-based clustering algorism (STRUCTURE v.2.3.4 software) (Pritchard, 2000), discriminant analysis of principal components (DAPC), and principal coordinate analysis (PCoA), were used.

DAPC was determined from the genetic data stored in “genind” object using “usethis” (Bryan, 2024), “devtools” (Wickham *et al*., 2022), and “adegenet” (Jombart and Ahmed, 2011) packages; whereas, PCoA was determined using GenAIEx v.6.51b2 (Peakall and Smouse, 2012).

Ten independent runs were conducted for each value of K using a burning period of 50,000 with 100,000 Markov Chain Monte Carlo (MCMC) iterations, assuming an admixture model and uncorrelated allele frequencies. The most probable value of K for each test was detected by ΔK (Evanno *et al*., 2005) using the web-based program StructureSelector (Li and Liu, 2018). Following Jakobsson and Rosenberg (2007), CLUMPP v.1.1.2 was used to align cluster assignments from independent runs using the in-files generated by the StructureSelector. Bar plots were generated with average results of runs for the most probable K value, using DISTRUCT v.1.1 (Rosenberg, 2004).

6.2.5. Genetic diversity analysis

Analysis of molecular variance (AMOVA) and genetic diversity was performed using GenAIEx v.6.51b2 (Peakall and Smouse, 2012) after grouping the genotypes based on the source of collection (i.e., geographical origins) and the genetic groups that resulted from structure analysis. The genotyping data were used to assess locus-based diversity indices including the number of alleles (Na), the effective number of alleles (Ne), Shannon’s information index (I), observed heterozygosity (Ho), expected heterozygosity (He), Wright fixation index (F_{ST}, F_{IS}, F_{IT}) (Wright, 1978) and percentage of polymorphic loci (PPL). Allelic richness (Ar) was performed using R software version 4.4.1 (R Core Team, 2024).

6.2.6. Phylogenetic analysis

The phylogenetic relationships of the sub-populations were generated based on the tri-matrix genetic distances of 94 black cumin genotypes using the "cluster" (Maechler *et al*., 2023) and "factoextra" (Kassambara and Mundt, 2020) packages in R software version 4.4.1 (R Core Team, 2024). The distance matrix between individuals was calculated according to Euclidean distance as implemented in the R environment. The resulting dissimilarity matrix was subjected to tree construction using the unweighted pair group method with arithmetic means (UPGMA), employing the same software with the "ggplot2" (Wickham, 2016) packages. The phylogenetic tree was constructed in R, implementing the hclust algorithm, with the UPGMA-relevant agglomeration method.

6.3. Results and Discussion

6.3.1. SNPs calling and data filtering

The GBS dataset showing the spread of scores across SNP markers is displayed in Figure 6.2. The alignment of the raw reads from the 94 black cumin genotypes to the reference genome resulted in the discovery of 41,442 SNPs and 9,564 silico markers. The SNP markers were then filtered using DArT-specific criteria (reproducibility > 99%, call rate > 95%, and minor allele frequency (MAF) ≥ 5%). First of all, the SNP dataset with the individual metadata was converted into the genlight object, and the quality of the SNP was provided. During the filtering process, the above filtering procedures were kept consecutively. The successive filtering criteria depended on the prior filtered genlight object.

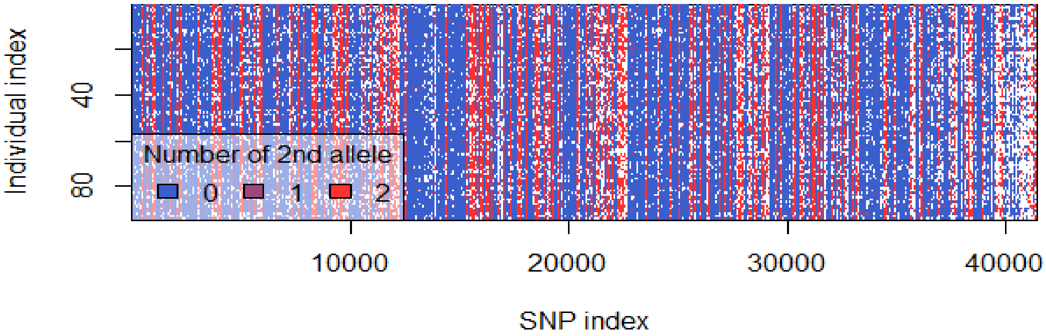


Figure 6.2. A smear plot showing the spread of scores across markers for each black cumin genotype. ("0" = Reference allele homozygote, "1" for SNP allele homozygote, "2" for heterozygote, and "-" for double null/null allele homozygote (absence of fragment with SNP in genomic representation))

Of the 41,442 SNP markers, only 1391 (3.36%) informative were deployed for downstream analysis; Table 6.1 displayed the diversity of markers kept after filtering. With a mean value of 1.00, the markers demonstrated excellent call rate and repeatability. The major allele frequency ranged from 0.92 to 0.98 with an average value of 0.96, while the minor allele frequency ranged from 0.005 to 0.50 with a mean of 0.014.

Table 6.1. After filtering, the characterization and distribution of SNP marker quality parameters of the black cumin genetic dataset

Markers quality	**Minimum**	**Maximum**	**Average**
Reproducibility	1.00	1.00	1.00
Call rate	1.00	1.00	1.00
Polymorphic information content (PIC)	0.03	0.15	0.07
Minor allele frequency (MAF)	0.005	0.50	0.014
Major allele frequency	0.92	0.98	0.96

Botstein *et al*. (1980) categorized polymorphic information content (PIC) as slightly informative (< 0.25), moderately informative (between 0.25 and 0.5), and highly informative (> 0.5). The polymorphic information content (PIC) ranged from 0.03 to 0.15, with a mean value of 0.07.

These SNPs retained after the refinement had a mean PIC value of 0.07. A supporting result was reported by Mirzaei and Mirzaghaderi (2015) in 39 Iranian black cumin collections using the SCoT marker.

6.3.2. SNP Markers distribution

The SNP marker distribution of the refined data across the black cumin genome is presented in Table 6.2. The marker alleles observed were A/G (15.82%), T/C (15.46%), G/A (20.92%), C/T (21.06%), C/A (4.03%), A/T (4.03%), T/A (4.24%), G/T (4.17%), A/C (4.17%), T/G (2.80%), C/G (1.65%), and G/C (1.65%). The transition-type SNPs (73.26%) are more abundant than the transversion-type SNPs (26.74%) in the black cumin genomes. Across the entire genome, the transitions to transversions (Ts/Tv) ratio is typically around 2.74:1. Compared to A/G and T/C, more G/A and C/T transition-type SNPs were found. High frequencies of transition mutations (C to T and G to A) after methylation may be indicated by the observation of a transition/transversion (Ts/Tv) ratio.

Table 6.2. Allelic information and proportion of transition and transversion-type SNP markers detected

Allelic information	**Transitions**				**Transversions**							
	A/G	T/C	G/A	C/T	C/A	A/T	T/A	G/T	A/C	T/G	C/G	G/C
Number of allelic sites	220	215	291	293	56	56	59	58	58	39	23	23
Allelic site (%)	15.82	15.46	20.92	21.06	4.03	4.03	4.24	4.17	4.17	2.80	1.65	1.65
Sum	1019				372							
Percentage	73.26				26.74							

6.3.3. Population structure

An effective statistical technique for determining the genetic makeup and relative ancestry of a population is population structure analysis. In this study, the population structure of 94 black cumin genotypes was assessed using various methodologies. The structure analysis revealed the existence of two ancestral populations in the collected black cumin genotype as indicated by the Delta k values (Figure 6.3). The black cumin populations were divided into two probable subpopulations based on population structure analysis, independent of geographic origin. A similar result was reported by Birhanu *et al.* (2015) in a molecular diversity study on 84 Ethiopian black cumin accessions using ISSR markers. Additional successful groupings were reported by Mirzaei and Mirzaghaderi (2015) and Mehri *et al.* (2022) on Iranian black cumin landraces by using SCoT and ISSR markers, respectively. In these studies, most of the landraces collected from different parts of Iran were grouped under the same subpopulation.

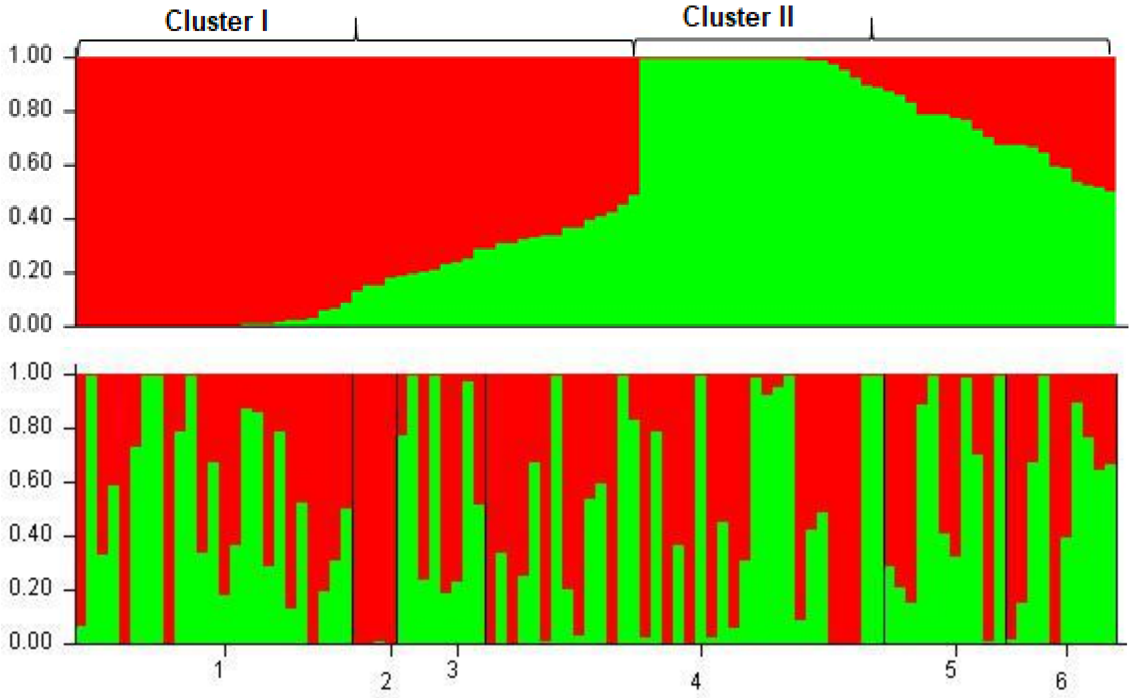


Figure 6.3. Structure results: the genetic structure of all black cumin populations in major modes at K = 2 using 1391 SNP markers. Numbers represented populations (1: Amhara, 2: B/Gumuz, 3: Improved, 4: Oromia, 5: SNNP, and 6: Tigray), respectively.

The scatter plots of DAPC (Figure 6.4) and principal coordinate analysis (Figure 6.5) also confirmed the grouping of the test genotypes into two groups. This result is consistent with the structure output. Therefore, the most probable number of subpopulations in the collected black cumin genotypes was two (K = 2).

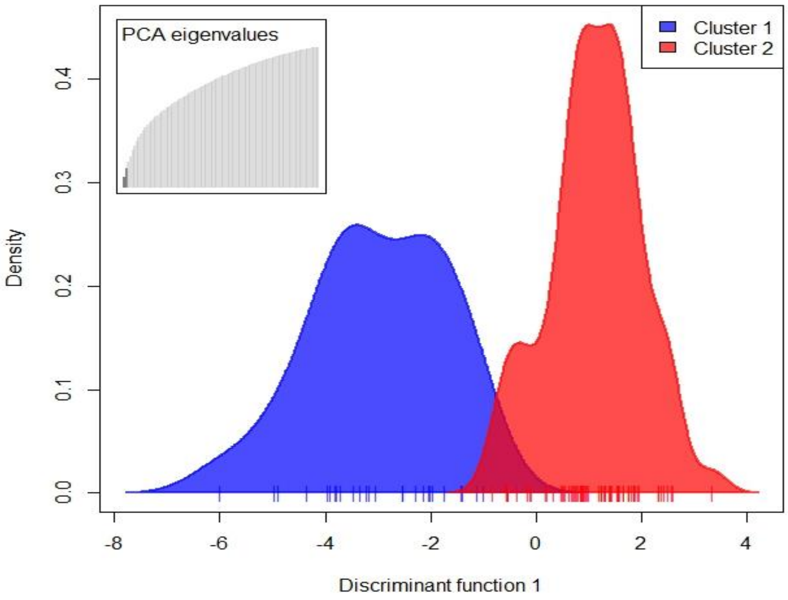


Figure 6.4. The scattered plot of DAPC at K=2. Each color represents one cluster.

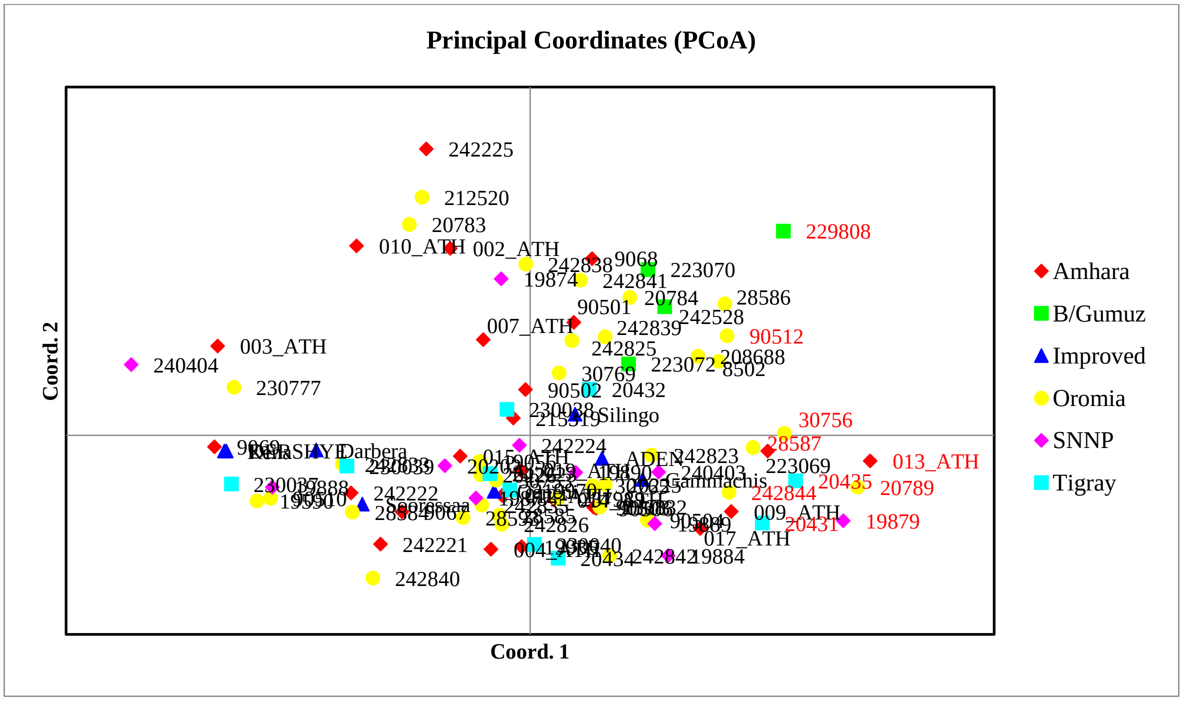


Figure 6.5. PCoA shows clustering patterns of 94 black cumin test genotypes based on genetic distances (GD) using 1391 SNP markers. (Genotypes written in black and red colors represented clusters I and II, respectively.)

6.3.4. Genetic diversity

The subpopulations of Ethiopian black cumin were categorized based on the geographical origins (pre-defined populations, n = 6) and the genetic group (n = 2). Table 6.3 displays the estimated genetic parameters for the pre-defined subpopulations and the genetic groups.

In pre-defined subpopulations

The average number of alleles varied from 1.487 to 1.983 (mean = 0.774 ± 005) (Table 6.3). The highest average number of alleles (Na = 1.983) was observed for genotypes originating from the Oromia region, followed by the Amhara region (Na = 1.923). In contrast, the population from the Benshangul-Gumuz region had the lowest value (Na = 1.487). The number of effective alleles varied from 1.312 to 1.34 (mean = 1.323 ± 003). The highest number of effective alleles (Ne = 1.34) was observed in genotypes originating from the Tigray region, followed by the Oromia region (Ne = 1.327), whereas genotypes originating from the SNNP region showed the lowest value (Ne = 1.312). This indicates the occurrence of an average number of alleles is more than the effective number of alleles. The variation in the number of alleles indicated the presence of genetic variability in black cumin genotypes.

Allelic richness ranged from 1.46 to 1.607 (mean = 1.564 ± 027) (Table 6.3). The genotypes originating from the Oromia region observed the highest allelic richness (Ar = 1.607), while the lowest value (Ar = 1.46) was shown by the Benshangul-Gumuz region. This indicates the variability of the genotypes originating from different regions of Ethiopia. The measure of allelic richness indicates the genetic diversity of a population and reflects its potential for adaptability and long-term persistence (Greenbaum *et al*., 2014).

The Shannon information index varied from 0.277 to 0.364 (mean = 0.331 ± 002) (Table 6.3). Genotypes originating from the Oromia region (I = 0.364) and the Amhara region (I = 0.352) were the richest subpopulations. In contrast, genotypes originating from the Benshangul-Gumuz region showed the lowest value. This confirmed the presence of genetic diversity between the black cumin genotypes.

The observed heterozygosity varied from 0.033 to 0.043 (mean = 0.037 ± 001) (Table 6.3). The highest observed heterozygosity (Ho = 0.043) was shown by the genotypes originating from the SNNP region, followed by the Tigray region (Ho = 0.04). In contrast, the lowest value (Ho = 0.033) was observed in the genotypes from the Amhara region. The expected heterozygosity varied from 0.187 to 0.224 (mean = 0.22 ± 002). The highest expected heterozygosity was found in genotypes originating from the Oromia and Tigray regions (He = 0.224 and He = 0.219, respectively). In contrast, the lowest value (He = 0.187) was observed in the genotypes from the Benshangul-Gumuz region. This indicates the observed heterozygosity was lower than the expected heterozygosity. This can be attributed to inbreeding. If the observed heterozygosity is lower than predicted, we could expect some degree of inbreeding (Kanaka *et al*., 2023).

The inbreeding coefficient varied from 0.369 to 0.783 (mean = 0.766) (Table 6.3). The highest inbreeding coefficient (F = 0.783) was observed in genotypes originating from the Oromia and Amhara regions (F = 0.783 and F = 0.734, respectively); the lowest value (F = 0.369) was observed in the genotypes from the Benshangul-Gumuz region. The higher inbreeding coefficient values indicate that the parents are more closely related than the lower value. This indicates the presence of genetic diversity is higher in the improved varieties and genotypes originating from Benshangul-Gumuz, SNNP, and Tigray regions as compared to Amhara and Oromia regions.

The percentage of polymorphic loci varied from 48.74 to 98.27% (mean = 77.37%) (Table 6.3). The genotypes originating from the Oromia and Amhara regions had the highest percentage of polymorphic loci (PPL = 98.27% and PPL = 92.31%, respectively). However, the lowest percentage of polymorphic loci was observed in the genotypes from the Benshangul-Gumuz region (PPL = 48.74%). This might be due to a decline in the proportion of polymorphic loci to the total number of populations. Similar results were reported by Birhanu *et al.* (2015) on Ethiopian black cumin collections.

Table 6.3. Genetic diversity in the populations of 94 Ethiopian black cumin genotypes

Population	Genetic parameters							
	Na	Ne	Ar	I	Ho	He	F	PPL
Based on geographical origin (Population = 6)								
Amhara	1.923	1.324	1.587	0.352	0.033	0.216	0.734	92.31
B/Gumuz	1.487	1.317	1.46	0.277	0.035	0.187	0.369	48.74
Improved	1.675	1.316	1.533	0.314	0.039	0.214	0.501	67.51
Oromia	1.983	1.327	1.607	0.364	0.035	0.224	0.783	98.27
SNNP	1.792	1.312	1.571	0.334	0.043	0.208	0.588	79.15
Tigray	1.782	1.34	1.585	0.346	0.04	0.219	0.58	78.22
Total								
Mean	1.774	1.323	1.564	0.331	0.037	0.22	0.766	77.37
SE (±)	0.005	0.003	0.027	0.002	0.001	0.002	0.005	7.26
Based on genetic groups (Population = 2)								
Group I	2	1.322	1.845	0.366	0.035	0.22	0.806	100
Group II	1.686	1.371	1.665	0.341	0.047	0.224	0.516	68.58
Total								
Mean	1.843	1.347	1.755	0.353	0.041	0.222	0.784	84.29
SE (±)	0.007	0.006	0.090	0.004	0.002	0.003	0.007	15.71

Where, B/Gumuz = Benshangul-Gumuz, SNNP = Southern Nations Nationalities and Peoples, Na = average number of alleles; Ne = effective number of alleles; Ar = Allelic richness; I = Shannon’s information index; Ho = observed heterozygosity; He = expected heterozygosity; F = Fixation Index/Inbreeding coefficient; PPL = percentage of polymorphic loci; SE = Standard error.

In the genetic groups

The average number of alleles varied from 2 to 1.686 (mean = 1.843 ± 0.007) (Table 6.3). The higher average number of alleles (Na = 2.0) was observed in genotypes from group I, whereas the lower value (Na = 1.686) was observed in genotypes from group II. The number of effective alleles varied from 1.322 to 1.371 (mean = 1.347 ± 0.006). The higher number of effective alleles (Ne = 1.371) was observed in genotypes from group II, whereas the lower value (Ne = 1.322) was observed in genotypes from group I.

Allelic richness ranged from 1.686 to 1.852 (mean = 1.769 ± 0.083) (Table 6.3). The genotypes from group I observed the higher allelic richness (Ar = 1.852), while the lower value (As = 1.686) was observed by genotypes from group II. The Shannon information index varied from 0.366 to 0.341 (mean = 0.353 ± 0.004) (Table 6.3). Genotypes originating from group I (I = 0.366) were the richer subpopulation than the genotypes from group II. The observed heterozygosity varied from 0.035 to 0.047 (mean = 0.041 ± 0.002) (Table 6.3). The higher observed heterozygosity (Ho = 0.047) was observed in genotypes from group II. In contrast, the lower value (Ho = 0.035) was observed in genotypes from group I. The expected heterozygosity varied from 0.22 to 0.224 (mean = 0.222 ± 0.003). The higher expected heterozygosity was found in genotypes from group II (He = 0.224). In contrast, the lower value (He = 0.22) was observed in the genotypes from group I.

The inbreeding coefficient varied from 0.516 to 0.806 (mean = 0.784 ± 007) (Table 6.3). The higher inbreeding coefficient (F = 0.806) was observed in genotypes from group I. In contrast, the lower inbreeding coefficient value (F = 0.516) was observed in genotypes from group II. The percentage of polymorphic loci varied from 68.58% to 100% (mean = 84.29% ± 15.71)

(Table 6.3). The genotypes from group I had the higher percentage of polymorphic loci (PPL = 100%), whereas the lower percentage of polymorphic loci was observed in the genotypes from group II (PPL = 68.58%).

6.3.5. Analysis of molecular variance

The AMOVA for the pre-defined subpopulations and genetic groups of 94 black cumin genotypes is displayed in Table 6.4.

In pre-defined subpopulations

The results revealed that the highest genetic variations (83%) were observed among genotypes within populations, followed by within genotypes (16%) based on geographical origins. However, the lowest value (1%) was observed among populations. This might result from comparatively higher seed exchange rates between geographical origins at its regional and trans-regional markets, which could cause genotypes to mix. The high probability of new genetic variation by mutation within the population might be the other reason. Supporting results were reported on black cumin collections by Birhanu *et al.* (2015) from Ethiopia using ISSR markers and Golkar and Nourbakhsh (2019) from Iran using sequence-related amplified polymorphisms (SRAP) and start codon targeted (SCoT) markers. The average fixation index is crucial for measuring genetic diversity among and within populations. The genetic differentiation was lowest among populations (Fst = 0.014), nevertheless, it was higher among genotypes within populations (Fis = 0.841) and within genotypes (Fit = 0.843). This is all corroborated by the estimated variance values (Table 6.4). The value of the overall gene flow was 17.35 (Table 6.4). This low variation among populations may be caused by larger-scale agroecological adaptation of the crop, the limited introduction of new varieties to the farming system, and regional exchange of genotypes.

In the genetic groups

The highest genetic variation (74.3%) was observed among genotypes within populations, followed by within genotypes (14.3%). However, the lowest genetic variation (11.4%) was observed among populations. The genetic differentiation was lowest among populations (Fst = 0.114), nevertheless, it was higher among genotypes within populations (Fis = 0.839) and within genotypes (Fit = 0.857). This is all corroborated by the estimated variance values (Table 6.4). The overall gene flow was 1.937 (Table 6.4).

According to Hartl and Clark (2007), there were many migrants in every generation in the case of the pre-defined subpopulations. This would promote an increase in variability within individuals of Ethiopian black cumin. This also confirms the presence of significant genetic differentiation among individuals. Gene flow facilitates gene migration, increases genetic variability, and may speed up the process of evolution (Mallory-Smith and Zapiola, 2008).

Table 6.4. Results of the analysis of molecular variance (AMOVA) among and within test black cumin subpopulations

Source	Df	SS	MS	EV	% EV	F-statistics	p-value
Based on geographical origin (Population = 6)							
Among populations	5	1782.42	356.483	2.28709	1	Fst = 0.014	0.001
Among genotypes	88	25706	292.114	133.429	83	Fis = 0.841	0.001
Within genotypes	94	2374	25.2553	25.2553	16	Fit = 0.843	0.001
Total	187	29862.5		160.972	100	Nm = 17.35	
Based on genetic groups (Population = 2)							
Among populations	1	1009.94	1009.94	20.203	11.40	Fst = 0.114	0.001
Among genotypes	92	26478.5	287.81	131.277	74.30	Fis = 0.839	0.001
Within genotypes	94	2374	25.255	25.255	14.30	Fit = 0.857	0.001
Total	187	29862.5		176.735	100.00	Nm = 1.937	

Where, Df = Degrees of freedom; SS = Sum of squares; MS = Mean square; EV = Estimated variance, % EV = Percent of estimated variance; Fst = Genetic differentiation, Fis = Fixation index or inbreeding coefficient; Fit = Overall fixation index, Nm = Gene flow, and p-value = Probability value.

This is also confirmed by the pairwise Nei's genetic distance results presented in Table 6.5. The genetic distance ranged from 0.009 to 0.091. Furthermore, the highest genetic distance was observed between the black cumin genotypes from the Improved varieties and Benshangul-Gumuz regions, followed by the Tigray and Benshangul-Gumuz regions. However, the lowest genetic distance was observed between genotypes from the Oromia and Amhara regions. This indicates the genetic differentiation between black cumin genotypes from the improved varieties and Benshangul-Gumuz regions is high, followed by the Tigray and Benshangul-Gumuz regions, whereas the genetic differentiation was little between genotypes from the Oromia and Amhara regions.

The pairwise Nei's genetic distances of the population matrix were calculated based on allele frequencies, and the results are shown in Table 6.6. The pairwise Nei's genetic distance of the two groups were 0.054 and 0.948, respectively (Table 6.6). This indicates the occurrence of moderate genetic diversity among the groups of black cumin.

Table 6.5. Pairwise Nei's genetic distance between 94 black cumin genotypes based on 1391 SNP markers

Population	Amhara	Benshangul-Gumuz	Improved	Oromia	SNNP
Amhara					
Benshangul-Gumuz	0.057				
Improved	0.027	0.091			
Oromia	0.009	0.056	0.027		
SNNP	0.020	0.076	0.038	0.016	
Tigray	0.023	0.081	0.035	0.017	0.026

Where SNNP = Southern Nations, Nationalities, and Peoples

Table 6.6. Pairwise Nei's genetic distance (below diagonal) and genetic identity (above diagonal) of the structure-inferred genetic groups of 94 black cumin genotypes based on 1391 SNP markers

	Group-I	Group-II
Group-I		0.948
Group-II	0.054	

6.3.6. Phylogenetic relationships

The 94 black cumin test genotypes were divided into two separate clusters using a UPGMA dissimilarity dendrogram created using the distance matrix produced (Figure 6.6).

Cluster I: This group comprised 84 (89.36%) genotypes of which eight were improved varieties and 76 accessions (Table 6.7). There was a clear tendency for all the improved varieties, and most of the genotypes from the five regions to group in the first cluster. Based on the prior morpho-agronomic and biochemical studies on the 64 black cumin genotypes, most genotypes (ca. 61) grouped under this cluster were characterized by higher seed yield, fixed oil, and essential oil yields per hectare. Furthermore, this group comprised genotypes from the two chemotypes (thymoquinone and ρ-cymene types).

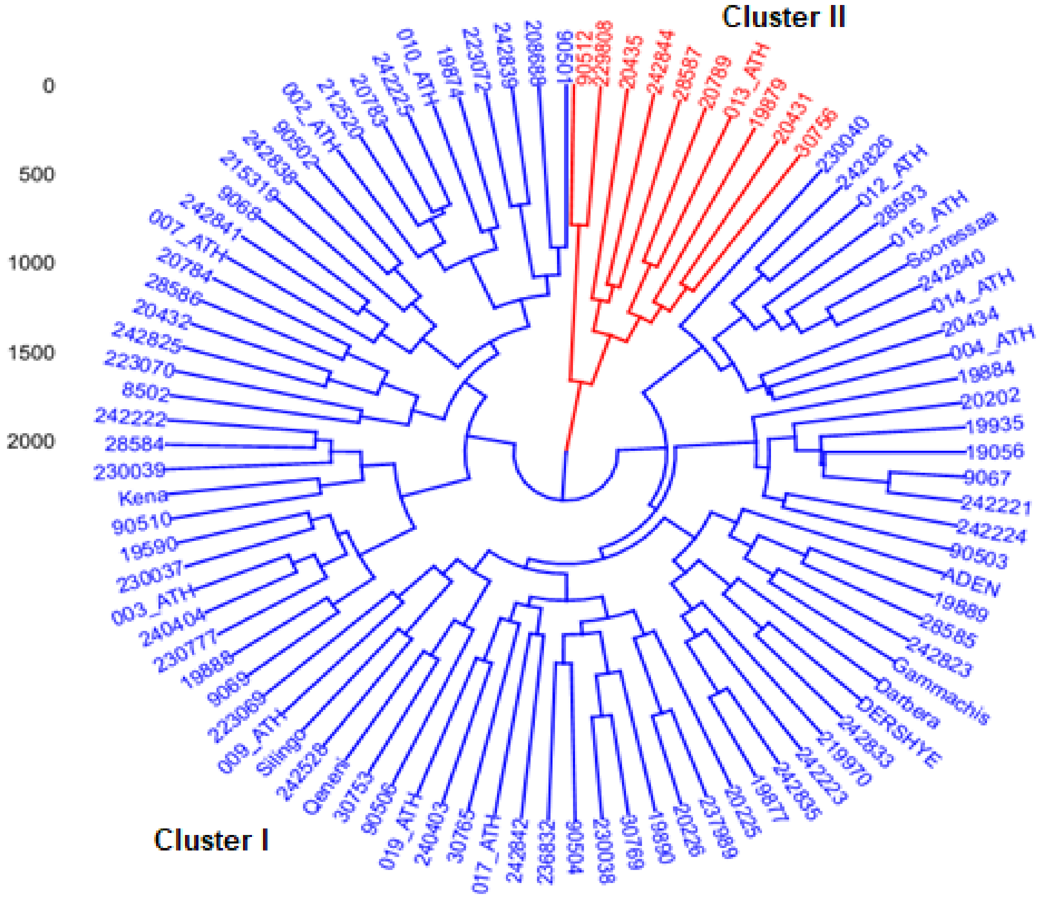


Figure 6.6. UPGMA dendrogram showing the genetic relationships among Ethiopia's 94 black cumin genotypes using 1391 SNP markers. (Blue and red colors represent clusters I and II, respectively.)

Cluster II: The remaining 10 (10.64%) genotypes were assigned to this cluster (Table 6.7). Of this, only three genotypes were included from the previous experiments and seven from EBI. The categories were not based on the geographical areas from which the black cumin genotypes originated. This could be due to the movement of planting material between geographical locations, which increases gene flow and allele distribution across various populations. Based on the previous result, the group was characterized by lower seed yield, fixed, and essential oil yields per hectare. Chemotypically, they were grouped under the thymoquinone/carvacrol/ longifolene type.

Table 6.7. Clustering of 94 black cumin genotypes of Ethiopia into two distinct clusters based on genetic distance

Cluster	Number of genotypes	Genotypes included	Collection region
I	84 (89.36%)	223072, 20432, 90503, 30756, Gammachis, 19056, 20435, 19935, Darbera, 28587, 230038, 003_ATH, 004_ATH, Qeneni, 215319, 242842, 28586, 30769, 90501, 20431, 90506, 015_ATH, 20434, 90512, 014_ATH, 007_ATH, 236832, 90504, 242841, 28584, 90502, 223069, 212520, 19884, 30753, 230039, 30765, 28593, 20783, 19879, DERSHYE, 242224, 28585, 19877, 230037, 242844, 230777, 208688, 20226, 19890, 013_ATH, 19874, Silingo, 010_ATH, 20225, 230040, 242223, 240403, 8502, Kena, 242823, 009_ATH, 19590, 19889, 219970, 012_ATH, 017_ATH, 019_ATH, 19888, 20784, 240404, 242840, 242839, 242838, 242835, 242833, 242826, 90510, 237989, 20202, 223070, 002_ATH, Sooressaa, 242528, 242225, 242222, 9068, 242221, 9069, 242825, 9067, 20789, ADEN	Oromia, Tigray, Improved, Amhara, SNNP, B/Gumuz
II	10 (10.64%)	90512, 242844, 28587, 20789, 30756, 20431, 013_ATH, 20435, 229808, 19879	Oromia, Tigray, Amhara, B/Gumuz, SNNP

B/Gumuz = Benshangul-Gumuz; SNNP = Southern Nations Nationalities and Peoples.

6.4. Conclusion

Advanced molecular techniques were utilized to generate the SNP data in this study, which provides valuable insights for upcoming genetic and breeding ventures involving black cumin. The study revealed high genetic variation within Ethiopian black cumin germplasm. It has also shown that much of this genetic variation is present among genotypes within populations and within genotypes than among populations. In addition, multiple modelling approaches partitioned the Ethiopian black cumin germplasm into two sub-populations. These findings were parallel to the findings obtained through morphological and biochemical studies of black cumin in Ethiopia and can serve as valuable input in the conservation and improvement of the crop.

CHAPTER 7

7. Summary and Recommendations

Black cumin whole seeds, fixed and essential oils, are economically important for flavoring foods and treating various ailments due to their many bioactive compounds. It has grown for a long period by small-scale farmers in different agroecologies of Ethiopia. It has different local names from the dwellers of the growing areas, including farmers. Small-holder farmers have undertaken the production, processing, and marketing. Currently, the whole seeds as a spice and fixed oil ("*Ye Tikur Azmud Zeyit*") have been sold for their aroma and medicinal value by different private investors in the country. Despite their greater adaptability, availability, and economic importance, there is a shortage of information on the diversity of genetic resources. Thus, the present studies were conducted to investigate the variability of Ethiopian black cumin based on morpho-agronomic and biochemical traits and using DArTseq SNP markers.

Four different experiments were conducted to achieve the above objectives. In the first two studies, a total of 64 black cumin genotypes composed of 8 improved varieties and 56 genotypes collected from five different former regions of Ethiopia were characterized for morpho-agronomic and biochemical traits at Debre Zeit and Kulumsa Agricultural Research Centers using an 8 × 8 simple lattice design with two replications. Essential and fixed oils were extracted by the methods of hydrodistillation and solvent extraction, respectively. The third experiment was conducted to investigate the variability of the 64 black cumin genotypes by essential oil compositions using Gas chromatography-mass spectrometry (GC-MS) analysis at Wondo Genet Agricultural Research Center Natural Products Laboratory. In the fourth experiment, the genetic diversity of 94 black cumin genotypes composed of 64 genotypes obtained from the first experiment and 30 additional materials obtained from the Ethiopian Biodiversity Institute were determined using DArTseq SNP markers. The univariate, bivariate, and multivariate analyses of the collected data were performed.

There is a significant genetic variation among black cumin genotypes in Ethiopia for the studied morpho-agronomic traits. This suggests there is a high potential of the studied genotypes for the genetic improvement of those traits. The number of capsules per plant and plant height exhibited high heritability in a broad sense, coupled with high to moderate genetic advance as a percentage of the mean, which indicates the possibility of these traits through selection. Plant height and number of primary branches per plant were the most influential traits on seed yield per hectare; this would be an important selection criterion for further seed yield improvement programs of those genotypes. About 81.5% of the variation that occurred among the studied genotypes was explained by the first four principal components. The PCA confirmed the presence of high phenotypic variability among the black cumin genotypes of Ethiopia. The studied genotypes were grouped into three distinct clusters. This implies that there is a presence of significant quantitative traits of diversity that can be exploited for further black cumin improvement programs. In general, this study confirmed that there is a reality about the presence of enough variation in most of the morpho-agronomic traits studied, which can create an enabling environment for breeders to design active genotype collection, conservation, and use strategies.

The mean performance of the genotypes discovered the wider ranges between the least and greatest values of all biochemical traits, this showed the presence of significant variation among black cumin genotypes included in this study. This might be the result of the difference in the genetic makeup of the genotypes. Essential oil yield per hectare had a significant positive correlation with FOY and EOC, but it did not correlate with FOC. The fixed oil yield per hectare had a strong positive and significant correlation with FOC. About 85.86% of the variation that occurred among the studied genotypes was explained by the first two principal components. Essential and fixed oil yield per hectare were the main contributor traits for most of the variation occurring among the genotypes in the first PC, whereas FOC and EOC were the main contributor traits for the second PC. The PCA confirmed the existence of high variability among black cumin genotypes of Ethiopia in all the biochemical traits studied. Genotypes of black cumin were partitioned into three distinct clusters with significant variation in the distance among genotypes within and among them. Genotypes 90504, 219970, and 013_ATH were the top 5% best-performed collections over improved varieties selected for FOY and EOY. In general, this study confirmed the presence of high variation in most of the biochemical traits studied in Ethiopian black cumin genotypes. Moreover, this finding has a significant advantage for academia, industries, researchers, policymakers, and societies at large.

Twenty-seven compounds were identified using GC-MS analysis of 64 black cumin genotypes essential oils. Based on the results, the essential oils that made up 92% to 100% of the total composition were *ρ*-cymene, *α*-thujene, carvacrol, trans-4-methoxy thujane, longifolene, terpinen-4-ol, 2-ethylhexyl isohexyl phthalate, *β*-pinene, *α*-pinene, phthalic acid, 2-fluorobenzyl heptyl ester, and *d*-limonene. The most prevalent components across all genotypes were five compounds: trans-4-methoxy thujane, α-thujene, *ρ*-Cymene, thymoquinone, and carvacrol. The remaining constituents differed depending on the genotype. Choosing the top 5% of black cumin accessions is predicted to improve all major volatile compounds by 13.2 to 15.1% compared to improved varieties. Thymoquinone exhibited a strong and significant negative correlation with

α-thujene, α-pinene, β-pinene, ρ-cymene, and *d*-limonene and a significant positive correlation with carvacrol. The first two principal components (PCs) accounted for 72.03% of the total variance, according to the principal component analysis of 12 major volatile constituents. This demonstrated that the Ethiopian black cumin genotypes in EOCs exhibit significant variability. The genotypes under study were divided into two chemotypes by cluster analysis. Chemotype A is characterized by high thymoquinone, carvacrol, and longifolene, while a high content of ρ-cymene and α-thujene distinguishes Chemotype B.

Advanced molecular techniques were utilized to generate the SNP data in this study, which provides valuable insights for upcoming genetic and breeding ventures involving black cumin. The study revealed high genetic variation within Ethiopian black cumin genotypes. It has also shown that much of this genetic variation is present among genotypes within populations and within genotypes rather than among regions. In addition, multiple modeling approaches partitioned the Ethiopian black cumin genotypes into two sub-populations. These findings were in line with the findings obtained through morphological and biochemical studies of black cumin in Ethiopia. They can serve as valuable input in the conservation and improvement of the crop.

Based on information generated from diversity analysis studies, the following recommendations and associated future works are suggested.

- Further characterization and evaluation studies at a wider agroecological condition of the country are required to understand the G x E interactions among morpho-agronomic and biochemical traits of black cumin diversity and enhance the improvement strategies.

- The genotypes need to be nutritionally characterized, and the full data needs to be documented.
- A well-planned breeding program needs to be established by giving priority to desirable traits such as seed yield, oil yield, and major compounds (i.e., Thymoquinone).
- Conventional plant breeding should be supplemented with targeted and trait-based association studies to come up with varieties with specific performance capacities that have a proven track record of success when it comes to phenotypic selection alone.
- A high-resolution study of the morpho-agronomic and biochemical features of the promising accessions is recommended to improve black cumin productivity and standardized cosmetic and pharmaceutical product development.
- GWAS – the next level of crop improvement – based on gene/SNPs – trait association studies are required.

References

Abd El-Hack, M.E., M. Alagawany, M.R. Farag, R. Tiwari, K. Karthik and K. Dhama. 2016. Nutritional, healthical and therapeutic efficacy of black cumin (*Nigella sativa*) in animals, poultry and humans. International Journal of Pharmacology 12(3): 232–248.

Abdallah, E.M. 2017. Black Seed (*Nigella sativa*) as antimicrobial drug: a mini-review. Novel Approaches in Drug Designing and Development 3(2): 1–5.

Abdela Befa Kinki. 2020. Physico-Chemical Characteristics of Released and Improved Black Cumin (*Nigella sativa* L.) Varieties. World Scientific Research 7(1): 1–4. https://doi.org/10.20448/journal.510.2020.71.1.4.

Abdou, N.M., M.H.H. Roby, A.A. AL-Huqail, A. Elkelish, A.A.S. Sayed, B.M. Alharbi, H.A.A. Mahdy and A.I.B. Abou-Sreea. 2023. Compost Improving Morphophysiological and Biochemical Traits, Seed Yield, and Oil Quality of *Nigella sativa* under Drought Stress. Agronomy 13(4): 1–19. https://doi.org/10.3390/agronomy13041147.

Abu-Hammour, K. 2008. Pollination of Medicinal Plants (*Nigella sativa* and *Coriandrum sativum*) and *Cucurbita pepo* in Jordan. Rheinischen Friedrich- Wilhelms-Universitat, Bonn.

Abu-Hammour, K.A. and D. Wittmann. 2011. Pollination of *Nigella sativa* L. (Ranunculaceae) in Jordan Valley to improve seed set. Advances in Horticultural Science 25(4): 212-222.

Abu-Jadayil, S., S.K.H. Tukan. and H.R. Takruri. 1999. Bioavailability of iron from four different local food plants in Jordan. (Electronic Version) Plant Foods for Human Nutrition 54: 285–294. https://www.ncbi.nlm.nih.gov/pubmed/10798339.

Agarwal, M., N. Shrivastava and H. Padh. 2008. Advances in molecular marker techniques and their applications in plant sciences. Plant Cell Reports 27(4): 617–631. https://doi.org/10.1007/s00299-008-0507-z

Ahlmann-Eltze, C. 2022. Package ‘ggsignif.’ 4. https://const-ae.github.io/ggsignif/.

Ahlmann-Eltze, C. and I. Patil. 2021. ggsignif: R Package for Displaying Significance Brackets for 'ggplot2. PsyArxiv. doi:10.31234/osf.io/7awm6.

Akour, A., V. Kasabri, F.U. Afifi and N. Bulatova. 2016. The use of medicinal herbs in gynecological and pregnancy-related disorders by Jordanian women: a review of folkloric practice vs. evidence-based pharmacology. Pharmaceutical Biology 54(9): 1901–1918. https://doi.org/10.3109/13880209.2015.1113994.

Aksu, M., G. Ozkan, S.S. Kiralan, M. Kiralan and M.F. Ramadan. 2021. Composition and Functionality of *Nigella sativa* Essential Oil. In: Fawzy Ramadan, M. (eds) Black cumin (Nigella sativa) seeds: Chemistry, Technology, Functionality, and Applications. Food Bioactive Ingredients. Springer, Cham. https://doi.org/10.1007/978-3-030-48798-0_26.

Al-Ali, A., A.A. Alkhawajah, M.A. Randhawa and N.A. Shaikh. 2008. Oral and intraperitoneal LD50 of thymoquinone, an active principle of *Nigella sativa*, in mice and rats. Journal of Ayub Medical College Abbottabad 20(2): 25-27.

Albakry, Z., E. Karrar, I.A.M. Ahmed, E. Oz, C. Proestos, A.F. El Sheikha, F. Oz, G. Wu and X. Wang. 2022. Nutritional Composition and Volatile Compounds of Black Cumin (*Nigella sativa* L.) Seed, Fatty Acid Composition and Tocopherols, Polyphenols, and Antioxidant Activity of Its Essential Oil. Horticulturae 8(575): 1-10. https://doi.org/10.3390/horticulturae8070575.

Alboukadel, K. and M. Fabian. 2020. Factoextra: extract and visualize the results of multivariate data analyses. R package version 1.0.7.

Ali E. and F. Hassan. 2014. Bio-production of *Nigella sativa* L. seeds and oil in Taif area. International Journal of Current Microbiology Applied Sciences 3(1): 315-328.

Aljabre, S.H.M., O.M. Alakloby and M.A. Randhawa. 2015. Dermatological effects of *Nigella sativa*. Journal of Dermatology and Dermatologic Surgery 19(2): 92–96.

Al-Jassir, M.S. 1992. Chemical composition and microflora of black cumin (*Nigella sativa* L.) seeds growing in Saudi Arabia. Food Chemistry 45: 239-242.

Alzoreky, N.S. and K. Nakahara. 2003. Antimicrobial activity of extracts from some edible plants commonly consumed in Asia. Int J Food Microbiol 80: 223–230.

Amdie, A. and S. Teshome. 2021. The adaptability of black cumin (*Nigella sativa* L.) varieties in the midland areas of Guji Zone, Southern Ethiopia. Journal of Science and Development 9(1): 46–51.

AOAC. 2005. American official methods of analysis, 10th edition. AOAC, International, Association of Official Analytical Chemists. In AOAC International, Gaithersburg, MD, USA, Official Method. Washington, DC, USA.

Arega Amdie and Solomon Teshome. 2021. The adaptability of black cumin (*Nigella sativa* L.) varieties in the midland areas of Guji Zone, Southern Ethiopia. Journal of Science and Development 9(1): 46-51.

Arzani, A., and M. Ashraf. 2016. Smart Engineering of Genetic Resources for Enhanced Salinity Tolerance in Crop Plants. Critical Reviews in Plant Sciences 35(3): 146–189. https://doi.org/10.1080/07352689.2016.1245056.

Atta, M.B. 2003. Some characteristics of nigella (*Nigella sativa* L.) seed cultivated in Egypt and its lipid profile. Food Chemistry 83(1): 63–68. https://doi.org/10.1016/S03088146(03)00038-4.

Aydın, A. 2024. Determining the genetic diversity of some black cumin genotypes collected in different regions of Türkiye using RAPD markers. International Journal of Agriculture Environment and Food Sciences 8(2): 294–300. https://doi.org/10.31015/jaefs.2024.2.6.

Bachek, N.F., M.A. Assi and M.H.M. Noor. 2016. The various effects of *Nigella sativa* on multiple body systems in human and animals. Pertanika Journal of Scholarly Research Reviews 2(3): 1–19.

Bafghi, A.F., A.R. Vahidi, M.H. Anvari, K. Barzegar and M. Ghafourzadeh. 2009. The in vivo antileishmanial activity of alcoholic extract from *Nigella sativa* seeds. African Journal of Microbiology Research 13(3): 001-007.

Baj, T., E. Sieniawska, R. Kowalski, M. Wesołowski and B. Ulewicz-Magulska. 2015. Effectiveness of the Deryng and Clevenger-type apparatus. Acta Poloniae Pharmaceutica - Drug Research 72(3): 507–515.

Bakkali, F., S. Averbeck, D. Averbeck and M. Idaomar. 2008. Biological effects of essential oils – A review. Food and Chemical Toxicology 46(2): 446–475. https://doi.org/10.1016/J.FCT.2007.09.106.

Barakat, E.M.F., L.M. El Wakeel and R.S. Hagag. 2013. Effects of *Nigella sativa* on outcome of hepatitis C in Egypt. World Journal of Gastroenterology 19(16): 2529–2536. https://doi.org/10.3748/wjg.v19.i16.2529.

Basazinew Degu Gebremedin, Bizuayehu Tesfaye Asfaw, Wendawek Abebe Mengesha and Kebebew Assefa Abebe. 2024. Genetic diversity of Ethiopian black cumin (Nigella sativa L.) based on morpho-agronomic characteristics. Euphytica 220(4): 1–22. https://doi.org/10.1007/s10681-024-03315-4.

Bayati, P., H. Karimmojeni and J. Razmjoo. 2020. Changes in essential oil yield and fatty acid contents in black cumin (*Nigella sativa* L.) genotypes in response to drought stress. Industrial Crops and Products 155: 1-7. https://doi.org/10.1016/j.indcrop.2020.112764.

Beheshti, F., F. Norouzi, A. Abareshi, A. Anaeigoudari and M. Hosseini. 2018. Acute Administration of *Nigella sativa* Showed Anxiolytic and Anti-Depression Effects in Rats. Current Nutrition and Food Science 14(5): 422-431. https://doi.org/10.2174/1573401313666170607155858.

Benkaci–Ali, F., A. Baaliouamer, B.Y. Meklati and F. Chemat. 2007. Chemical composition of seed essential oils from Algerian *Nigella sativa* extracted by microwave and hydrodistillation. Flavour and Fragrance Journal 22(2): 148-153. https://doi.org/10.1002/ffj.1773

Bhandari, H.R., A.N. Bhanu, K. Srivastava, M.N. Singh, Shreya and A. Hemantaranjan. 2017. Assessment of Genetic Diversity in Crop Plants - An Overview. Advances in Plants and Agriculture Research 7(3): 279–286. https://doi.org/10.15406/apar.2017.07.00255.

Birhanu Kapital, Tileye Feyissa, Yohannes Petros and Said Mohammed. 2015. Molecular diversity study of black cumin (*Nigella sativa* L.) from Ethiopia as revealed by inter simple sequence repeat (ISSR) markers. African Journal of Biotechnology 14(18): 1543–1551. https://doi.org/10.5897/ajb2015.14567.

Botstein, D., R.L. White, M. Skolnick and R.W. Davis. 1980. Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms. American Society of Human Genetics.32:314-331. https://pubmed.ncbi.nlm.nih.gov/6247908/.

Bourgou, S., A. Pichette, B. Marzouk and J. Legault. 2010. Bioactivities of black cumin essential oil and its main terpenes from Tunisia. South African Journal of Botany 76(2): 210–216. https://doi.org/10.1016/j.sajb.2009.10.009.

Bozdemir, Ç., R.B. Bağdat, İ. Subaşı, N. Akci and N. Çinkaya. 2022. Determination of Yield and Quality Characteristics of Various Genotypes of Black Cumin (*Nigella Sativa* L.) Cultivated Through Without Fertilizers. International Journal of Life Sciences and Biotechnology 5(3): 386–406. https://doi.org/10.38001/ijlsb.1111198.

Bryan, M.J. 2024. Package ‘usethis .’ https://usethis.r-lib.org.

Burits, M. and F. Bucar. 2000. Antioxidant Activity of *Nigella sativa* Essential Oil. Phytotherapy Research 14(5): 323–328.

Chahal, G.S. and S.S. Gosal. 2002. Principles and procedures of plant breeding: Biotechnological and conventional approaches. Alpha Science International Ltd, Oxford, United Kingdom.

Charrad, M., N. Ghazzali, V. Boiteau, and A. Niknafs. 2014. Nbclust: An R package for determining the relevant number of clusters in a data set. Journal of Statistical Software 61(6): 1–36. https://doi.org/10.18637/jss.v061.i06.

Cheikh-Rouhou, S., S. Besbes, B. Hentati, C. Blecker, C. Deroanne and H. Attia. 2007. *Nigella sativa* L.: chemical composition and physicochemical characteristics of lipid fraction. Food Chem 101:673–681.

Cheikh-Rouhou, S., S. Besbes, G. Lognay, C. Blecker, C. Deroanne and H. Attia. 2008. Sterol composition of black cumin (*Nigella sativa* L.) and Aleppo pine (*Pinus halpensis* Mill.) seed oils. Journal of Food Composition and Analysis 21(2): 162-168.

Clevenger, L.A., K.R. Coffey, K. Barmak, D.A. Rudman and C.V. Thompson. 1989. Experimental evidence for nucleation during thin-film reactions. Applied physics letters 55(9): 852-854.

Cochran, W.G. and G.M. Cox. 1957. Experimental Designs, 2nd ed. Wiley, New York.

Coffey, K.R., L.A. Clevenger, K. Barmak, D.A. Rudman, and C.V. Thompson. 1989. Experimental evidence for nucleation during thin-film reactions. In Applied Physics Letters 55(9): 852–854. https://doi.org/10.1063/1.102447.

Collard, B.C.Y., M.Z.Z. Jahufer, J.B. Brouwer and E.C.K. Pang. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142:169–196.

Csardi, G. and T. Nepusz. 2006. The igraph software package for complex network research.

D'Antuono, L.F., A. Moretti, and A.F.S. Lovato. 2002. Seed yield, yield components, oil content and essential oil content and composition of *Nigella sativa* L. and *Nigella damascena* L. Industrial Crops and Products 15(1): 59–69. https://doi.org/10.1016/S0926-6690(01)00096-6.

Da Silva, A.R. 2021. Tools for Biometry and Applied Statistics in Agricultural Science. R Journal.

Daniel Bisrat, Solomon Abate and Wossen Kebede. 2009. Laboratory manual for plant products analysis. Technical Manual 1(23): 3–18.

Daryabeygi-Khotbehsara, R., M. Golzarand, M. P. Ghafari and K. Djafarian. 2017. *Nigella sativa* improves glucose homeostasis and serum lipids in type 2 diabetes: A systematic review and meta-analysis. Complementary Therapies in Medicine 35: 6–13. https://doi.org/10.1016/j.ctim.2017.08.016.

Daryabeygi-Khotbehsara, R., M. Golzarand, M.P. Ghaffari and K. Djafarian. 2017. *Nigella sativa* improves glucose homeostasis and serum lipids in type 2 diabetes: A systematic review and meta-analysis. Complementary Therapies in Medicine 35: 6–13. https://doi.org/10.1016/j.ctim.2017.08.016.

Datta, A.K., A. Saha, A. Bhattacharya, A. Mandal, R. Paul and S. Sengupta. 2012. Black cumin (*Nigella sativa* L.) – A Review. Journal of Plant Development Sciences 4(1): 1-43.

De Mendiburu, F. 2014. Agricolae: Statistical procedures for agricultural research. R package version 1.2-0. https://cran.r-project.org/package=agricolae.

De Souza Siqueira Quintans, J., P.P. Menezes, M.R.V. Santos, et al. 2013. Improvement of p-cymene antinociceptive and anti-inflammatory effects by inclusion in β-cyclodextrin. Phytomedicine 20(5): 436–440. https://doi.org/10.1016/j.phymed.2012.12.009.

Demel Teketay. 2000. Ranunculaceae. In: Edwards S, Mesfn T, Sebsebe D, Hedberg I (eds) Flora of Ethiopia and Eritrea. Volume 2, Part 1. A Joint publication of the National Herbarium, Biology Department, AAU, Ethiopia and the Department of Systematic Botany, Uppsala University, Sweden, pp 31–32.

Demis Fikre, Fekadu Gebretensay Mengistu, Dasta Tsagaye, Awoke Ali, Nimona Fufa and Gizaw Wegayehu. 2023. Evaluation of Black Cumin (*Nigella sativa* L.) Genotypes for Yield and Yield Related Parameters in Potential Growing Areas of Ethiopia. International Journal of Bio-Resource and Stress Management 14(7): 1037-1045. https://doi.org/10.23910/1.2023.3518.

Dessalegn Anshiso and Wubeshet Teshome. 2018. Economic Value of Black Cumin (*Nigella sativa* L.) Conservation at Bale Zone of Oromia Region, Ethiopia. American Journal of Business, Economics and Management 6(4):104-109.

Dewey, D.R. and K.H. Lu. 1959. A Correlation and Path-Coefficient Analysis of Components of Crested Wheatgrass Seed Production. Agronomy Journal 51(9): 515–518. https://doi.org/10.2134/agronj1959.00021962005100090002x.

Dhifi, W., S. Bellili, S. Jazi, N. Bahloul and W. Mnif. 2016. Essential Oils' Chemical Characterization and Investigation of Some Biological Activities: A Critical Review. Medicines 3(25): 1-16. https://doi.org/10.3390/medicines3040025.

Edeget Merawi. 2016. Identification of *Nigella sativa* for Access and Benefit sharing purpose. Addis Ababa, Ethiopia. pp. 3

Edris, A.E. 2010. Evaluation of the Volatile Oils from Different Local Cultivars of *Nigella sativa* L. Grown in Egypt with Emphasis on the Effect of Extraction Method on Thymoquinone. Journal of Essential Oil-Bearing Plants 13(2): 154-164.

EL-Mahrouk M.E., M.K. Maamoun, O.F.A. El-leel, Y.H. Dewir, A.N. El-Banna, Y. Naidoo and S.K. Datta. 2020. Morpho-agronomical and biochemical traits screening and genetic variability in selected black cumin (*Nigella sativa*) mutant lines. Sains Malaysiana 49(3): 503–515. https://doi.org/10.17576/jsm-2020-4903-05.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6(5): 1–10. https://doi.org/10.1371/journal.pone.0019379.

Endashaw Bekele. 2007. Study on Actual Situation of Medicinal Plants in Ethiopia. Japan Association for International Collaboration of Agriculture and Forestry 1–5. http://www.endashaw.com.

Ermias Assefa, Addis Alemayehu and Teshom Mamo. 2015. Adaptability Study of Black Cumin (*Nigella sativa* L.) Varieties in the Mid and High Land Areas of Kaffa Zone, South West Ethiopia. Agriculture, Forestry and Fisheries 4(1): 14-17. https://doi.org/10.11648/j.aff.20150401.13.

Ermias Dagne. 2009. Natural Database for Africa. Plant Catalogue Software, Addis Ababa, Ethiopia.

ESRI. 2011. ArcGIS Desktop: Release 10.

Ethiopian Investment Agency. 2015. Investment opportunity Profile for Spice Processing in Ethiopia. 7: 8–14.

Evanno, G., S. Regnaut and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Molecular Ecology 14(8): 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x.

Farhan, N., N. Salih and J. Salimon. 2021. Physiochemical properties of Saudi *Nigella sativa* L. ('Black cumin’) seed oil. OCL - Oilseeds and Fats, Crops and Lipids 28: 1–9. https://doi.org/10.1051/ocl/2020075.

Fekadu Gebretensay Mengistu and Gizaw Wegayehu Tilahun. 2023. Effects of Seed Ageing on Germination and Oil Content in Ethiopian Black Cumin Varieties. Preprints (www.preprints.org). DOI:10.20944/preprints202308.0917.v1.

Forouzanfar, F., B.S.F. Bazzaz and H. Hosseinzadeh. 2014. Black cumin (*Nigella sativa*) and its constituent (thymoquinone): A review on antimicrobial effects. Iranian Journal of Basic Medical Sciences 17(12): 929–938.

Gerige, S.J., M.K.Y. Gerige, M. Rao and Ramanjaneyulu. 2009. GC-MS analysis of *Nigella sativa* seeds and antimicrobial activity of its volatile oil. Brazilian Archives of Biology and Technology 52(5): 1189–1192. https://doi.org/10.1590/S1516-89132009000500016

Getachew Asefa and Beriso Mieso. 2021. Evaluation of black cumin genotypes for yield and yield related parameters in bale mid altitude, southeastern Ethiopia. International Journal of Agricultural Research, Innovation and Technology 10(2): 35–37. https://doi.org/10.3329/ijarit.v10i2.51574.

Gezahegn Assefa and Sintayehu Girma. 2016. Evaluation and Selection of Black Cumin (*Nigella sativa* L.) Varieties at Mid Highland of West Hararghe Zone, East Ethiopia. Journal of Biology, Agriculture and Healthcare 6(23): 1-5.

Ghadlinge, M.S., J.B. Jaju, R.D. Chandane, et al. 2014. A study of effect of *Nigella sativa* oil in paracetamol induced hepatotoxicity in albino rats. International Journal of Basic and Clinical Pharmacology 3(3): 539–546.

Gharby, S., H. Harhar, D. Guillaume, A. Roudani, S. Boulbaroud, M. Ibrahimi, M. Ahmad, S. Sultana, T.B. Hadda, I. Chafchaouni-Moussaoui and Z. Charrouf. 2015. Chemical investigation of *Nigella sativa* L. seed oil produced in Morocco. Journal of the Saudi Society of Agricultural Sciences 14(2): 172–177. https://doi.org/10.1016/j.jssas.2013.12.001.

Golkar, P. and V. Nourbakhsh. 2019. Analysis of genetic diversity and population structure in *Nigella sativa* L. using agronomic traits and molecular markers (SRAP and SCoT). Industrial Crops and Products 130: 170–178.

Gomez, K.A. and A.A. Gomez. 1984. Statistical procedures for agricultural research (2nd ed.). John Wiley, and Sons.

Govindaraj, M., M. Vetriventhan and M. Srinivasan. 2015. Importance of genetic diversity assessment in crop plants and its recent advances: An overview of its analytical perspectives. Genetics Research International 1-14. https://doi.org/10.1155/2015/431487.

Gruenbaum, Y., R. Stein, H. Cedar and A. Razin. 1981. Methylation of CpG sequences in eukaryotic DNA. FEBS Letters 124(1): 67–71. https://doi.org/10.1016/0014-5793(81)80055-5.

Gulçin, A.P. and A. Zehra. 2018. Chemical Composition of the Fixed and Essential Oils of *Nigella sativa* L. from Turkey. Current Perspectives on Medicinal and Aromatic Plants 1: 19-27

Gupta, P.K. and R.K. Varshney. 2000. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113(3): 163–185. https://doi.org/10.1023/A:1003910819967.

Habtewold Kifelew, Demes Fikere, Tewodros Luleseged, Dejene Bekele, Haimanot Mitiku and Wakjira Getachew. 2017. Seed Spices Production Guideline: Ethiopian Institute of Agricultural Research. Available: http://www.ublication.eiar.gov.et.

Hajhashemi, V., A. Ghannadi and H. Jafarabadi. 2004. Black cumin seed essential oil, as a potent analgesic and anti-inflammatory drug. Phytotherapy Research 18: 195–199.

Hamed, E., W. Toaima and W. Abd El-Aleem. 2023. Impact of Different Planting Locations on *Nigella Sativa* L. Yield in Egypt. Egyptian Journal of Desert Research 73(1): 23–38. https://doi.org/10.21608/ejdr.2023.198316.1130.

Hamrouni-Sellami, I., M.E. Kchouk and B. Marzoul. 2008. Lipid and aroma composition of black cumin (*Nigella sativa* L.) Seeds from Tunisia. Journal of Food Biochemistry 32: 335–352.

Hanson, C.H., H.F. Robinson, and R.E. Comstock. 1956. Biometrical Studies of Yield in Segregating Populations of Korean Lespedeza 1. Agronomy Journal 48(6): 268–272. https://doi.org/10.2134/agronj1956.00021962004800060008x.

Haque, M., Sapna, R. Singh, A. Nadeem, S. Rasool, et al. 2021. *Nigella sativa*: A promise for industrial and agricultural economic growth. Pharmacological and Therapeutic Applications. pp 439–460. https://doi.org/10.1016/B978-0-12-824462-3.00010-X.

Hartl, D.L. and A.G. Clark. 2007. Principles of population genetics (4th ed.). Sunderland, Massachusetts: Sinauer Associates, Inc.

Hassanien, M.F.R., A.M.A. Assiri, A.M. Alzohairy and H.F. Oraby. 2015. Health-promoting value and food applications of black cumin essential oil: an overview. Journal of Food Science and Technology 52(10): 6136–6142.

Heidari, E.F., M. Rahimmalek, S. Mohammadia and M.H. Ehtemam. 2016. Genetic structure and diversity of ajowan (*Trachyspermum ammi*) populations based on molecular, morphological markers, and volatile oil content. Industrial Crops and Products 92: 186–196. https://doi.org/10.1016/j.indcrop.2016.08.014

Henry, R. and K. Edwards. 2009. New tools for single nucleotide polymorphism (SNP) discovery and analysis accelerating plant biotechnology: Editorial. Plant Biotechnology Journal 7(4): 311. https://doi.org/10.1111/j.1467-7652.2009.00417.x.

Henry, R.J. 2012. Molecular Markers in Plants. In Molecular Markers in Plants. https://doi.org/10.1002/9781118473023.

Herms, S. 2015. Business Opportunities Report Spices #6 in the series written for the "Ethiopian Netherlands business event, Rijswijk, The Netherlands”

Heshmati, J. and N. Namazi. 2015. Effects of black seed (*Nigella sativa*) on metabolic parameters in diabetes mellitus: A systematic review. Complementary Therapies in Medicine 23(2): 275–282. https://doi.org/10.1016/j.ctim.2015.01.013.

Heslop-Harrison, Y. 2000. Control Gates and Micro-ecology: The Pollen-Stigma Interaction in Perspective. Annals of Botany 85: 5-13.

Hosseini, S.S., F. Nadjafi, M.H. Asareh and H. Rezadoost. 2018. Morphological and yield related traits, essential oil and oil production of different landraces of black cumin (*Nigella sativa*) in Iran. Scientia Horticulturae 233: 1–8.

Huseini, H.F., S. Kianbakht, M.H. Mirshamsi and A.B. Zarch. 2015. Effectiveness of Topical *Nigella sativa* Seed Oil in the Treatment of Cyclic Mastalgia: A Randomized, Triple-Blind, Active, and Placebo-Controlled Clinical Trial. In Planta Medica 82(4): 285–288. https://doi.org/10.1055/s-0035-1558208.

Iqbal, M.S., A. Ghafoor and A.S. Qureshi. 2010. Evaluation of *Nigella sativa* L. for genetic variation and ex-situ conservation. Pakistan Journal of Botany 42(4): 2489-2495.

Iqbal, M.S., S. Nadeem, S. Mehboob, A. Ghafoor, M.I. Rajoka, A.S. Qureshi and B.Niaz. 2011. Exploration of genotype specific fingerprinting of Nigella sativa L. using RAPD markers. Turkish Journal of Agriculture and Forestry 35(6): 569–578. https://doi.org/10.3906/tar-

1001-622.

Jakobsson, M. and N.A. Rosenberg. 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23(14): 1801–1806. https://doi.org/10.1093/bioinformatics/btm233.

Jansen, P.C.M., 1981. Spices, condiments and medicinal plants in Ethiopia, their taxonomy and agricultural significance. Centre for Agricultural Publishing and Documentation, Wageningen, 81 pp.

Javed, S., A.A. Shahid and M.S. Haider. 2010. Nutritional, phytochemical potential and pharmacological evaluation of *Nigella Sativa* (Kalonji) and *Trachyspermum Ammi* (Ajwain). Journal of Medicinal Plants Research 6(5): 768–775.

Javed, S., A.A. Shahid, M.S. Haider, A. Umeera, R. Ahmad and S. Mushtaq. 2012. Nutritional, phytochemical potential and pharmacological evaluation of *Nigella sativa* (Kalonji) and *Trachyspermum Ammi* (Ajwain). Journal of Medicinal Plants Research 6(5): 768–775. DOI: 10.5897/JMPR11.1341.

Johnson, H.W., H.F. Robinson and R.E. Comstock. 1955. Estimates of genetic and environmental variability in Soybeans. Experimental Agriculture 47(7): 314–318. https://doi.org/10.1017/S0014479700009510.

Jolliffe, I.T. 1986. Principal Component Analysis. 1st ed. Springer, New York. https://doi.org/10.1007/978-0-387-98135-2.

Jolliffe, I.T. 2002. Principal Component Analysis, Second Edition. In Principal Component Analysis. Springer, New York. http://link.springer.com/10.1007/b98835.

Jombart, T. and I. Ahmed. 2011. adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. Bioinformatics 27(21): 3070–3071. https://doi.org/10.1093/bioinformatics/btr521.

Kabir, Y., Y. Akasaka-Hashimoto, K. Kubota and M. Komai. 2020. Volatile compounds of black cumin (*Nigella sativa* L.) seeds cultivated in Bangladesh and India. 6(10): e05343. doi: 10.1016/j.heliyon.2020.e05343.

Kaiser, H.F. 1960. The Application of Electronic Computers to Factor Analysis. Educational and Psychological Measurement 20(1): 141–151. https://doi.org/10.1177/001316446002000116.

Kanaka, K.K., N. Sukhija, R.C. Goli, S. Singh, I. Ganguly, S.P. Dixit, A. Dash and A.A. Malik. 2023. On the concepts and measures of diversity in the genomics era. Current Plant Biology 33(December 2022): 100278. https://doi.org/10.1016/j.cpb.2023.100278.

Kara, N., G. Gürbüzer, M. Biyikli and H. Baydar. 2021. Effect on Yield and some Quality Characteristics of Seed Harvest at Different Stages of Maturity in *Nigella sativa* L. Tarım Bilimleri Dergisi 27(3): 341–346. https://doi.org/10.15832/ankutbd.657045

Karp, A., S. Kresovich, K.V. Bhat, W.G. Ayad and T. Hodgkin. 1997. Molecular tools in plant genetic resources conservation: a guide to the technologies. International Plant Genetic Resources Institute, Rome Italy.

Kassambara, A. 2017. Practical Guide to Principal Component Methods in R. STDHA. http://www.sthda.com.

Kassambara, A. and F. Mundt. 2020. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. Package Version 1.0.7. R Package Version.

Kedir Jaleto Bento, Gizaw Wegayehu and Demis Fikre. 2025. Evaluation of Ethiopian Black Cumin (*Nigella sativa* L.) Accessions for Yield and Yield Related Characters at Kulumsa, South-Eastern Ethiopia. Global Journal of Research in Agriculture & Life Sciences 5(2): 32-38. https://gjrpublication.com/gjrals/

Khader, M. and P.M. Eckl. 2014. Thymoquinone: An emerging natural drug with a wide range of medical applications. Iranian Journal of Basic Medical Sciences 17(12): 950–957.

Khan, N., F.H. Wattoo, I. Ahmad, I. Muhammad, S.G. Afridi, M.H.S. Wattoo and M.K.S. Safi. 2017. Assessment of Genetic Diversity and Phytochemical Analysis of *Nigella sativa* Genotypes from Pakistan. Asian Journal of Biological Sciences 10(2): 56–63. https://doi.org/10.3923/ajbs.2017.56.63.

Khare, C.P. 2004. Encyclopedia of Indian medicinal plants. Springer, Berlin.

Kizil, S., S. Kirici, Ö. Çakmak, and K.M. Khawar. 2008. Effects of sowing periods and P application rates on yield and oil composition of black cumin (*Nigella sativa* L.). Journal of Food, Agriculture and Environment 6(2): 242–246.

Kökdil, G., and H. Yilmaz. 2005. Analysis of the fixed oils of the genus Nigella L. (Ranunculaceae) in Turkey. Biochemical Systematics and Ecology 33(12): 1203–1209. https://doi.org/10.1016/j.bse.2005.07.013.

KorehKhosravi, S.H., A. Masoumiasl, and M. Dehdari. 2018. A comparative analysis of RAPD and ISSR markers for assessing genetic diversity in Iranian populations of *Nigella sativa* L. Cellular and Molecular Biology 64(1): 52–59. https://doi.org/10.14715/cmb/2018.64.1.10.

Kumhar, S.R., B.R. Choudhary, R. Ramesh and M.L. Mehriya. 2020. Genetic diversity studies of cumin (*Cuminun cyminum* L.) genotypes in western plains of Rajasthan. Journal of Spices and Aromatic Crops 29(1): 67–71. https://doi.org/10.25081/josac.2020.v29.i1.6335.

Laurentin, H. 2009. Data analysis for molecular characterization of plant genetic resources. Genetic Resource and Crop Evolution 56(2): 277–292. https://doi.org/10.1007/s10722-008-9397-8

Li, Y.-L. and J.-X. Liu. 2018. STRUCTURESELECTOR: A web-based software to select and visualize the optimal number of clusters using multiple methods. Molecular Ecology Resources 18(1): 176–177. https://doi.org/10.1111/1755-0998.12719.

Liu, K. and S.V. Muse. 2005. PowerMaker: An integrated analysis environment for genetic maker analysis. Bioinformatics 21(9): 2128–2129.

Luisi, G., A. Stefanucci, G. Zengin, M.P. Dimmito and A. Mollica. 2019. Anti-oxidant and tyrosinase inhibitory in vitro activity of amino acids and small peptides: New hints for the multifaceted treatment of neurologic and metabolic disfunctions. In Antioxidants 8(1): 1-14. https://doi.org/10.3390/antiox8010007.

Lukasz, N. 2020. “clv: Cluster Validation Techniques,” R package version 0.3-2.2. https://CRAN.R-project.org/package=clv

Maechler, M., P. Rousseeuw, A. Struyf, M. Hubert and K. Hornik. 2021. “cluster: Cluster Analysis Basics and Extensions. R package version 2.1.2.

Maechler, M., P. Rousseeuw, A. Struyf, M. Hubert and K. Hornik. 2023. Cluster package. https://cran.r-project.org/package=cluster.

Maechler, M., P. Rousseeuw, A. Struyf, M. Hubert, K. Hornik, M. Studer, P. Roudier, J. Gonzalez and K. Kozlowski. 2018. cluster: Cluster Analysis Basics and Extensions. R Package Version.

Mahalanobis, P.C. 1936. The generalized distance in statistics. Proceedings of the National Institute of Science of India 2: 49-55.

Mahmoudvand, H., A. Sepahvand, S. Jahanbakhsh, B. Ezatpour, and S.A.A. Mousavi, 2014. Evaluation of antifungal activities of the essential oil and various extracts of *Nigella sativa* and its main component, thymoquinone against pathogenic dermatophyte strains. Journal of Medical Mycology 24(4): 155–161. https://doi.org/10.1016/j.mycmed.2014. 06.048

Mallory-Smith, C. and M. Zapiola. 2008. Gene flow from glyphosate-resistant crops: Review. Pest Management Science 63(11): 1100–1106. https://doi.org/10.1002/ps.1517.

Mamun, M.A. and N. Absar. 2018. Major nutritional compositions of black cumin seeds – cultivated in Bangladesh and the physicochemical characteristics of its oil. International Food Research Journal 25(6): 2634-2639.

Manly, B.F.J. and J.A.N. Alberto. 2016. Multivariate Statistical Methods A Primer, Fourth Edition. In Multivariate Statistical Methods: A Primer, Fourth Edition.

Marongiu, B., A. Piras and A. Rosa. 2013. Chemical composition and in vitro bioactivity of the volatile and fixed oils of *Nigella sativa* L. extracted by supercritical carbon dioxide. Industrial Crops and Products 46: 317–323.

Masresha Yimer. 2010. Market Profile on Spices: Ethiopia. To UNCTAD ICT. Addis Ababa, Ethiopia.

Mehri, N., M. Mohebodini, M. Behnamian and K. Farmanpour-Kalalagh. 2022. Phylogenetic, Genetic Diversity, and Population Structure Analysis of Iranian Black Cumin (*Nigella sativa* L.) Genotypes Using ISSR Molecular Markers. International Journal of Horticultural Science and Technology 9(2): 151–163.

Mijangos, J.L., B. Gruber, O. Berry, C. Pacioni, and A. Georges. 2022. dartR v2: An accessible genetic analysis platform for conservation, ecology and agriculture. Methods in Ecology and Evolution 13(10): 2150–2158. https://doi.org/10.1111/2041-210X.13918.

Miller P.A., J.C. Williams Jr., H.F. Robinson and R.E. Comstock. 1958. Estimates of Genotypic and Environmental Variances and Covariances in Upland Cotton and Their Implications in Selection. Agronomy journal 50(3): 126-131.

Ministry of Agriculture (MOA). 2019. Plant Variety Release, Protection and Seed Quality Control Directorate Crop Variety Register. ISSUE No. 22, Addis Ababa, 292-294 p.

Mironescu, M. and C. Georgescu. 2021. Comparative analysis and antimicrobial action of some essential oils from plants. BIO Web of Conferences 30, 01011. https://doi.org/10.1051/bioconf/20213001011.

Mirzaei, K. and G.. Mirzaghaderi. 2015. Genetic diversity analysis of Iranian *Nigella sativa* L. landraces using SCoT markers and evaluation of adjusted polymorphism information content.pdf. Plant Genetic Resources: Characterization and Utilization 1–8. https://doi.org/10.1017/S1479262115000386.

Mohan, M., S. Nair, A. Bhagwat, T.G. Krisnhna, M. Yano, C.R. Bhatia and T. Sasaski. 1997. Genome mapping, molecular markers and marker-assisted selection in crop plants. Molecular Breeding 3(2): 87–107. https://doi.org/10.1023/A:1009651919792

Mondini, L., A. Noorani and M.A. Paguotta. 2009. Assessing plant genetic diversity by molecular tools. Diversity 1: 19–35.

Moose, S.P. and R.H. Mumm. 2008. Molecular Plant Breeding as the Foundation for 21st Century Crop Improvement. Plant Physiology 147: 969–977.

Morgil, H., Y.C. Gercek and I. Tulum. 2020. Single Nucleotide Polymorphisms (SNPs) in Plant Genetics and Breeding. The Recent Topics in Genetic Polymorphisms 1–12. https://doi.org/10.5772/intechopen.91886.

Morsi, N.M. 2000. Antimicrobial effect of crude extracts of *Nigella sativa* on multiple antibiotics-resistant bacteria. Acta Microbiologica Polonica 49(1): 63-74.

Mouwakeh, A., A. Kincses, M. Nové, T. Mosolygó, C. Mohácsi-Farkas, G. Kiskó and G. Spengler. 2019. *Nigella sativa* essential oil and its bioactive compounds as resistance modifiers against *Staphylococcus aureus*. Phytotherapy Research 33(4): 1010–1018. https://doi.org/10.1002/ptr.6294.

Mukhtar, H., A.S. Qureshi, F. Anwar, M.W. Mumtaz and M. Marcu. 2019. *Nigella sativa* L. seed and seed oil: potential sources of high-value components for development of functional foods and nutraceuticals/pharmaceuticals. Journal of Essential Oil Research 31(3): 171–183. https://doi.org/10.1080/10412905.2018.1562388.

Nasri, I.H., A.T. Khaled, I.H. Butros and K. Kamal. 2007. Effect of Some Agricultural Practices on the Productivity of Black Cumin (*Nigella sativa* L.) Grown under Rainfed Semi-Arid Conditions. Jordan Journal of Agricultural Sciences 3(4): 385-397.

Nickavar, B., F. Mojab, K. Javidnia and M.A.R. Amoli. 2003. Chemical Composition of the Fixed and Volatile Oils of *Nigella sativa* L. from Iran. Zeitschrift Fur Naturforschung - Section C Journal of Biosciences 58(9–10): 629–631. https://doi.org/10.1515/znc-2003-9-1004.

Nimet, K., K. Duran and B. Hasan. 2015. Yield and quality of black cumin (*Nigella sativa* L.) populations: The effect of ecological conditions. Turkish Journal of Field Crops 20(1): 9-14.

Nováková, A., K. Šimáčková, J. Bárta and V. Čurn. 2010. Utilization of DNA markers based on microsatellite polymorphism for identification of potato varieties cultivated in the Czech Republic. Journal of Central European Agriculture 11(4): 415–422.

Okeola, V.O., O.A. Adaramoye, C.M. Nneji, C.O. Falade, EO. Farombi and O.G. Ademowo. 2011. Antimalarial and antioxidant activities of methanolic extract of *Nigella sativa* seeds (black cumin) in mice infected with *Plasmodium yoelli* nigeriensis. Parasitology Research 108(6): 1507–1512. https://doi.org/10.1007/s00436-010-2204-4.

Oksanen J, G. Simpson, F. Blanchet, R. Kindt, P. Legendre, et al. 2024. vegan: Community Ecology. Package_. R package version 2.6-8. https://cran.r-project.org/package=vegan.

Onifade, A., A. Jewell and W. Adedeji. 2013. *Nigella Sativa* Concoction induced sustained seroreversion in HIV patient, African Journal of Traditional, Complementary and Alternative Medicines 10(5): 332–335.

Onifade, A.A., A.P. Jewell and A.B. Okesina. 2015. Seronegative conversion of an HIV-positive subject treated with *nigella sativa* and honey. African Journal of Infectious Diseases 9(2): 47–50.

Ozdemir, N., M.N. Kantekin-Erdogan, T. Tat and A. Tekin. 2018. Effect of black cumin oil on the oxidative stability and sensory characteristics of mayonnaise. Journal of Food Science and Technology 55(4): 1562–1568.

Ozel, A., U. Demirel, I. Guler and K. Erden. 2009. Effect of different row spacing and seeding rate on black cumin (*Nigella sativa* L.) yields and some agricultural characters. Harran University Journal of Agriculture Faculty 13(1): 17-25.

Palabıyık, G.A. and Z. Aytaç. 2018. Chemical Composition of the Fixed and Essential Oils of Nigella sativa L. from Turkey. Current Perspectives on Medicinal and Aromatic Plants 1(1): 19–27. https://dergipark.org.tr/en/pub/cupmap/issue/38636/427413.

Panes V.G. and P.V. Sukhatme. 1995. Statistical methods for agricultural workers. 3rd ed. ICAR, New Delhi. 58 p.

Peakall, R. and P.E. Smouse. 2012. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—An update. Bioinformatics 28:2537–2539.

doi:10.1093/bioinformatics/bts460.

Peres-Neto, P.R., D.A. Jackson and K.M. Somers. 2005. How many principal components? stopping rules for determining the number of non-trivial axes revisited. Computational Statistics and Data Analysis 49(4): 974–997.

Peterson, G.W., Y. Dong, C. Horbach and Y-B Fu. 2014. Genotyping-by-sequencing for plant genetic diversity analysis: a lab guide for SNP genotyping. Diversity 6: 665-680.

Peterson, G.W., Y. Dong, C. Horbach, and Y.B. Fu. 2014. Genotyping-by-sequencing for plant genetic diversity analysis: A lab guide for SNP genotyping. Diversity 6(4): 665–680. https://doi.org/10.3390/d6040665.

Piras, A., A. Rosa, B. Marongiu, S. Porcedda, D. Falconieri, M.A. Dessì, B. Ozcelik and U. Koca. 2013. Chemical composition and in vitro bioactivity of the volatile and fixed oils of *Nigella sativa* L. extracted by supercritical carbon dioxide. Industrial Crops and Products 46: 317–323. https://doi.org/10.1016/j.indcrop.2013.02.013.

Postweiler, K., R. Stosser and S.F. Anvari. 1985. The effect of different temperatures on the viability of ovules in cherries. Scientia Horticulturae 25: 235-239.

Powell, W., M. Morgate, C. Chadre, M. Hanafey, J. Vogel, S. Tingey and A. Rafasalki. 1996. The comparison of RFLP, AFLP, RAPD and SSR markers for germplasm analysis. Molecular Breeding 2: 225–238.

Pritchard, J.K., M. Stephens and P. Donnelly. 2000. Inference of Population Structure Using Multilocus Genotype Data. Genetics Society of America 155: 945–959.

R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

R Core Team. 2024. R: A language and environment for statistical computing (4.2.3). R Foundation for Statistical Computing. https://www.r-project.org.

Ramadan, M.F. 2007. Nutritional value, functional properties and nutraceutical applications of black cumin (*Nigella sativa* L.) oil seeds: an overview. International Journal of Food Science and Technology 42: 1208–1218.

Ramadan, M.F. 2015. Black Cumin (*Nigella sativa*) Oils. In Essential Oils in Food Preservation, Flavor and Safety (Issue Figure 2). Elsevier Inc.

Ramos, M. and M. Morgan. 2024. BiocManager: Access the Bioconductor Project Package Repository. R package version 1.30.23. https://cran.r-project.org/package=BiocManager.

Rao, K.N. 2004. Plant genetic resources: Advancing conservation and use through biotechnology. African Journal of Biotechnology 3: 136–145.

Rao, S.R. and G.A. Ravishankar. 2002. Plant cell cultures: Chemical factories of secondary metabolites. Biotechnology Advances 20(2): 101–153. https://doi.org/10.1016/S0734-9750(02)00007-1.

Rattanachaikunsopon P. and P. Phumkhachorn. 2010. Assessment of factors influencing antimicrobial activity of carvacrol and cymene against Vibrio cholerae in food. Journal of Bioscience and Bioengineering 110(5): 614–619.

Rezaei-Chiyaneh, E., M.L. Battaglia, A. Sadeghpour, F. Shokrani, A.D.M. Nasab, M.A. Raza, and M.V. Cossel. 2021. Optimizing Intercropping Systems of Black Cumin (*Nigella sativa* L.) and Fenugreek (*Trigonella foenum-graecum* L.) through Inoculation with Bacteria and Mycorrhizal Fungi. Advanced Sustainable Systems 5(9): 1–14. https://doi.org/ 10.1002/adsu.202000269.

Robertson, A. 1959. The Sampling Variance of the Genetic Correlation Coefficient. International Biometric Society 15(3): 469–485. http://www.jstor.org/stable/2527750.

Rosenberg, N.A. 2004. DISTRUCT: A program for the graphical display of population structure. Molecular Ecology Notes 4(1): 137–138. https://doi.org/10.1046/j.1471-8286.2003.00566.x.

Russinga, A.M. 2020. Correlation Studies on Yield and Yield Contributing Traits in Rice (*Oryza sativa* L.). Indian Journal of Pure and Applied Biosciences 8(5): 531–538. https://doi.org/10.18782/2582-2845.8334.

Sadeer, N.B., E.J. Llorent-Martínez, K. Bene, *et al*. 2019. Chemical profiling, antioxidant, enzyme inhibitory and molecular modelling studies on the leaves and stem bark extracts of three African medicinal plants. Journal of Pharmaceutical and Biomedical Analysis 174: 19–33. https://doi.org/10.1016/j.jpba.2019.05.041.

Sahlemedhin Sertsu, Abayneh Esayas and Demeke Tafesse. 2003. Soils of Kulumsa Agricultural Research Center. Federal Democratic Republic of Ethiopia. Ethiopian Institute of Agricultural Research. Technical Paper No. 76, p 7.

Salaheldin, S., S.F. Hendawy, M.S. Hussein and W.S. Soliman. 2020. Assessment the Yield and Quality of *Nigella sativa* Under Different Environmental Conditions. International

Journal of Pharmacy and Pharmaceutical Sciences 12(10): 29–33. https://doi.org/10.22159/ijpps.2020v12i10.38995.

Salem, M.L. 2005. Immunomodulatory and therapeutic properties of the *Nigella sativa* L. seed. International Immunopharmacology 5(13–14): 1749–1770.

Salgotra, R.K. and B.S. Chauhan. 2023. Genetic Diversity, Conservation, and Utilization of Plant Genetic Resources. Genes 14(1): 174. https://doi.org/10.3390/genes14010174

SAS Institute Inc. 2019. System Requirements for SAS®9.4 Foundation for Microsoft Windows for x64, Cary, NC: SAS Institute Inc.

Schneider-Stock, R., I.H. Fakhoury, A.M. Zaki, C.O. El-Baba and H.U. Gali-Muhtasib. 2014. Thymoquinone: Fifty years of success in the battle against cancer models. Drug Discovery Today 19(1): 18–30. https://doi.org/10.1016/j.drudis.2013.08.021.

Shafiq, H., A. Ahmad, T. Masud and M. Kaleem. 2014. Cardio-protective and anti-cancer therapeutic potential of *Nigella sativa*. Iranian Journal of Basic Medical Sciences 17(12): 967–979.

Shokri, H. 2016. A review on the inhibitory potential of *Nigella sativa* against pathogenic and toxigenic fungi. Avicenna Journal of Phytomedicine 6(1): 21–33.

Siahsar, B., M. Rahimi, A. Tavassoli, and A. Raissi. 2011. Application of Biotechnology in Production of Medicinal Plants. American-Eurasian Journal of Agriculture and Environmental Science 11(3): 439–444.

Sileshi Abera and Biruk Hirko. 2020. Chemical Composition of Essential Oils of Released Black Cumin Varieties Grown in Ethiopia. Chemistry and Materials Research 12(2): 9-14. https://doi.org/10.7176/cmr/12-2-02.

Singh R.K. and B.D. Chaudhary. 2005. Biometrical methods in quantitative genetics analysis, 2nd ed. Kalyani Publishers, New Delhi, India.

Singh, R.K. and B.D. Chaudhary. 1985. Biometrical methods in quantitative genetic analysis. Kalayani Publishers, New Delhi-Ludhiana, India.

Sivasubramanian S. and P. Madhavamenon. 1973. Genotypic and phenotypic variability in rice. Madras Agricultural Journal 60: 1093-1096.

Smith, J.S. and O.S. Smith. 1989. The description and assessment of distances between inbred lines of maize: The use of morphological traits as descriptors. Maydica 34: 141–150.

Sudhir, S.P., A. Kumarappan, J. Malakar and H.N. Verma. 2016. Genetic Diversity of *Nigella* *sativa* from Different Geographies Using RAPD Markers. American Journal of Life Sciences 4(6): 175-180. https://doi.org/10.11648/j.ajls.20160406.15.

Sultan, M.T., M.S. Butt, F.M. Anjum, A. Jamil, S. Akhtar and M. Nasir. 2009. Nutritional profile of indigenous cultivar of black cumin seeds and antioxidant potential of its fixed and essential oil. Pakistan Journal of Botany 41(3): 1321-1330.

Sultana, S., H.M. Asif, N. Akhtar, A. Iqbal, H. Nazar and R.U. Rehman. 2015. *Nigella sativa*: Monograph. Journal of Pharmacognosy and Phytochemistry 4(4): 103-106.

Taha, M., A. Azeiz and W. Saudi. 2010. Antifungal effect of thymol, thymoquinone and thymohydroquinone against yeasts, dermatophytes and non-dermatophyte molds isolated from skin and nails fungal infections. The Egyptian Journal of Biochemistry and Molecular Biology 28(2): 109- 126. https://doi.org/10.4314/ejbmb.v28i2.60802.

Takruri, H.R.H. and M.A.F. Dameh, 1998. Study of the nutritional value of black cumin seeds (*Nigella sativa* L). Journal of the Science of Food and Agriculture 76(3): 404–410.

Tewodros Lulseged, Firew Mekbib and Kebebew Assefa. 2018. Correlation and Path Analysis for Yield and Yield component in Black Cumin (*Nigella sativa* L.). International Journal of Current Research and Academic Review 6(11): 56–63.

Tierram, M. 2005. *Nigella sativa*, commonly known as "Love in the Mist": A beautiful Middle Eastern herb with many uses. https://planetherbs.com/research-center/specific-herbs-articles/nigella-sativa//

Toncer, O. and S. Kizil. 2004. Effect of seed rate on agronomic and technologic characters of *Nigella sativa* L. International Journal of Agriculture and Biology 3: 529-532.

Tufek, N.H., M.E. Altunkaynak, B.Z. Altunkaynak and S. Kaplan. 2015. Effects of thymoquinone on testicular structure and sperm production in male obese rats. Systems Biology in Reproductive Medicine 61(4): 194–204.

Umamaheswar, N., S.K. Roy, A. Kundu, L. Hijam, M. Chakraborty, S. Sen, B. Das, R. Barman and S. Vishnupriya. 2024. Genetic Variability and Character Association Studies in Diverse Rice (*Oryza sativa* L.) Genotypes for Agro-Morphological Traits in Terai Region of West Bengal. Journal of Advances in Biology and Biotechnology 27(5): 805–820. https://doi.org/10.9734/jabb/2024/v27i5843.

Vasant, O.K., B.G. Vijay, S.R. Virbhadrappa, N.T. Dilip, M.V. Ramahari and B.S. Laxamanrao, 2012. Antihypertensive and Diuretic Effects of the Aqueous Extract of *Colocasia* *esculenta* Linn. Leaves in Experimental Paradigms. Iranian Journal of Pharmaceutical Research 11(2): 621–634.

Venables, W.N. and B.D. Ripley. 2002. Modern Applied Statistics with S (4th ed.). Springer, New York. https://doi.org/10.1214/aoms/1177697510.

Venkatachallam, S.K.T., Pattekhan, H., Divakar, S. and Kadimi, U.S., 2010. Chemical composition of *Nigella sativa* L. seed extracts obtained by supercritical carbon dioxide. J Food Sci Technol. 47: 598–605.

Viuda-Marto, M., M.A. Mohamady, J. Fernández-López, K.A. Abd ElRazik, E.A. Omer, J.A. Pérez-Alvarez and E. Sendra. 2011. In vitro antioxidant and antibacterial activities of essentials oils obtained from Egyptian aromatic plants. Food Control 22: 1715–1722.

Wajs, A., R. Bonikowski and D. Kalemba. 2008. Composition of essential oil from seeds of *Nigella sativa* L. cultivated in Poland. Flavour and Fragrance Journal 23(2): 123–132. https://doi.org/10.1002/ffj.1866.

Wang, R., D. Sweeney, S.E. Gandy and S.S. Sisodia. 1996. The profile of soluble amyloid β protein in cultured cell media. Detection and quantification of amyloid β protein and variants by immunoprecipitation-mass spectrometry. Journal of Biological Chemistry 271(50): 31894–31902. https://doi.org/10.1074/jbc.271.50.31894.

Wang, S., H. Deng, Y. Wang, W. Rui, P. Zhao, Q. Yong, D. Guo, J. Liu, X. Guo, Y. Wang and C. Shi. 2021. Antimicrobial activity and action mechanism of thymoquinone against bacillus cereus and its spores. Foods 10(12): 1-18.

Ward, J.H. 1963. Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association 58(301): 236–244.

Wei, T. and V. Sinko. 2021. R package 'corrplot': Visualization of a Correlation Matrix (Version 0.92). https://github.com/taiyun/corrplot.

Wei, T., V. Simko, M. Levy, Y. Xie, Y.J. Jin and J. Zemla. 2017. Visualization of a Correlation Matrix. R package “corrplot”. Statistician.

Wei., T. and V. Sinko. 2021. R package 'corrplot': Visualization of a Correlation Matrix (Version 0.92). https://github.com/taiyun/corrplot.

Wells, S.J. and J. Dale. 2018. Contrasting gene flow at different spatial scales revealed by genotyping-by-sequencing in *Isocladus armatus*, a massively colour polymorphic New Zealand marine isopod. PeerJ 6:e5462; DOI 10.7717/peerj.5462.

Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. In Media. Springer-Verlag New York.

Wickham, H., J. Bryan, J. Hester and W. Chang. 2022. Package 'devtools' R topics documented. https://devtools.r-lib.org/.

Wilkinson, S.P. and S.K. Davy. 2018. phylogram: an R package for phylogenetic analysis with nested lists. Journal of Open Source Software 3(26): 790.

Winter, P. and G. Kahl. 1995. Molecular marker technologies for plant improvement. World Journal of Microbiology and Biotechnology 11: 438–448.

Wright, S. 1978. Evolution and the genetics of populations. Vol. 4. Variability within and among natural populations. University of Chicago Press.

Yarnell, E. and K. Abascal. 2011. *Nigella sativa:* holy herb of the Middle East. Alternative and Complementary Therapies 17(2): 99-105.

Zerihun Tadesse, Habtemariam Zegeye, Dawit Asnake, Tafesse Solomon, Yewubdar Shewaye and Tolessa Debele. 2018. Identification of Stable Bread Wheat (*Triticum aestivum* L) Genotypes using AMMI Analysis in Ethiopia. International Journal of Research in Agriculture and Forestry 5(6): 6–14.

Zhang, Q., M.A. Saghai and A. Kleinhofs, 1993. Comparative diversity analysis of RFLPs and isozymes within and among populations of *Hordeum vulgare* ssp. spontaneum. Genetics 134: 909–916.

Zhou, Q., P. Li, R. Lu, Q. Qian, X. Lei, *et al*. 2013. Synthesis, X-ray diffraction study, and cytotoxicity of a cationic p-cymene ruthenium chloro complex containing a chelating semicarbazone ligand. Zeitschrift Fur Anorganische Und Allgemeine Chemie 639(6): 943–946. https://doi.org/10.1002/zaac.201300142.

Zigyalew Gashaw, Wosene Gebreselassie and Girma Hailemichael. 2020. Correlation and Path Coefficient Analysis in Yield and Yield-Related Components of Black Cumin (*Nigella sativa* L.) Accessions, at Jimma, Southwest Ethiopia. International Journal of Agronomy 1–9. https://doi.org/10.1155/2020/8837794.

Zohary, D. and M. Hopf. 2000. Domestication of Plants in the Old World: The Origin and Spread of Cultivated Plants in West Asia, Europe and the Nile Valley. Clarendon Press, Oxford.

APPENDICES

Appendix Table 1. Lists of black cumin (*N. sativa* L.) genotypes with their area of collection in Ethiopia

No.	Genotype	Region	Zone	Latitude (ºN)	Longitude (ºE)	Source	Status
1.	Darbera	Oromia	Bale	-	-	SARC	Improved
2.	DERSHYE	Oromia	Bale	-	-	SARC	Improved
3.	ADEN	Oromia	East Shewa	-	-	DZARC	Improved
4.	Sooressaa	Oromia	Bale	-	-	SARC	Improved
5.	Gammachis	Oromia	Bale	-	-	SARC	Improved
6.	Silingo	Oromia	East Shewa	-	-	DZARC	Improved
7.	Kena	Oromia	Bale	-	-	SARC	Improved
8.	Qeneni	Oromia	Bale	-	-	SARC	Improved
9.	8502	Oromia	Bale	7.0000	39.8000	DZARC	Accession
10	9067	Amhara	West Gojjam	11.6856	37.0200	DZARC	Accession
11	9068	Amhara	West Gojjam	11.7611	37.0844	DZARC	Accession
12	9069	Amhara	West Gojjam	10.6467	37.0858	DZARC	Accession
13	19884	SNNP	Keficho Shekicho	7.0644	36.0667	DZARC	Accession
14	90501	Amhara	West Gojjam	10.6392	37.0869	DZARC	Accession
15	90502	Amhara	South Gondar	11.9500	37.7000	DZARC	Accession
16	90503	Amhara	South Gondar	11.9833	37.7667	DZARC	Accession
17	90504	Oromia	Arsi	8.0500	38.7833	DZARC	Accession
18	90506	Amhara	East Gojjam	10.3333	38.0000	DZARC	Accession
19	90510	Oromia	West Shewa	9.1667	37.8333	SARC	Accession
20	208688	Oromia	West Hararghe	8.8167	40.4167	SARC	Accession
21	212520	Oromia	Bale	7.0167	39.9833	SARC	Accession
22	215319	Amhara	East Gojjam	11.0022	37.0031	SARC	Accession
23	219970	Tigray	Western	14.1367	38.3094	DZARC	Accession
24	223069	Amhara	East Gojjam	11.0022	37.0031	SARC	Accession
25	223070	B/Gumuz	Metekel	11.0000	35.7625	SARC	Accession
26	223072	B/Gumuz	Metekel	11.0000	35.7625	DZARC	Accession
27	229808	B/Gumuz	Metekel	10.5000	36.1667	DZARC	Accession
28	230037	Tigray	Central	14.0667	38.0833	SARC	Accession
29	230038	Tigray	Central	14.1667	38.7500	SARC	Accession
30	230039	Tigray	Northwestern	14.0667	38.0833	DZARC	Accession
31	230040	Tigray	Central	14.0833	39.1000	SARC	Accession
32	230777	Oromia	Borena	5.1167	39.4833	SARC	Accession
33	237989	Oromia	Bale	8.0500	38.7833	SARC	Accession
34	240403	SNNP	Keficho Shekicho	7.2342	35.7092	SARC	Accession
35	240404	SNNP	Keficho Shekicho	7.2500	36.0000	SARC	Accession
36	242221	Amhara	South Wollo	10.8411	39.8167	DZARC	Accession
37	242222	Amhara	North Wollo	11.9681	39.0722	DZARC	Accession
38	242223	Tigray	Western	14.1208	38.4747	DZARC	Accession
39	242224	SNNP	Arbaminch	6.1186	38.1186	SARC	Accession
40	242225	Amhara	South Wollo	11.0333	39.7542	DZARC	Accession
41	242528	B/Gumuz	Asosa	9.9858	34.6675	DZARC	Accession
42	242825	Oromia	Arsi	7.5606	39.6081	DZARC	Accession

Appendix Table 1. Continued

No.	Genotype	Region	Zone	Latitude (ºN)	Longitude (ºE)	Source	Status
43.	242826	Oromia	Arsi	7.5744	39.5969	DZARC	Accession
44.	242833	Oromia	Arsi	7.6508	39.4961	DZARC	Accession
45.	242835	Oromia	Arsi	7.6036	39.5636	DZARC	Accession
46.	242838	Oromia	Arsi	7.6031	39.5414	DZARC	Accession
47.	242839	Oromia	Arsi	7.5719	39.5283	DZARC	Accession
48.	242840	Oromia	Arsi	7.5544	39.5333	DZARC	Accession
49.	242841	Oromia	Arsi	7.5511	39.5408	DZARC	Accession
50.	242842	Oromia	Arsi	7.5356	39.5364	DZARC	Accession
51.	242844	Oromia	Arsi	7.6061	39.5233	DZARC	Accession
52.	242823	Oromia	Arsi	7.6036	39.5619	DZARC	Accession
53.	002_ATH	Amhara	North Gondar	12.3129	37.3220	DZARC	Accession
54.	003_ATH	Amhara	North Gondar	12.3401	37.3688	DZARC	Accession
55.	004_ATH	Amhara	North Gondar	12.3522	37.3389	DZARC	Accession
56.	007_ATH	Amhara	North Gondar	12.4253	37.2989	DZARC	Accession
57.	009_ATH	Amhara	North Gondar	12.2181	37.1950	DZARC	Accession
58.	010_ATH	Amhara	North Gondar	12.3575	37.1872	DZARC	Accession
59.	012_ATH	Amhara	North Gondar	12.3417	37.1092	DZARC	Accession
60.	013_ATH	Amhara	North Gondar	12.3072	37.3158	DZARC	Accession
61.	014_ATH	Amhara	North Gondar	12.2475	37.0406	DZARC	Accession
62.	015_ATH	Amhara	North Gondar	12.2300	37.0325	DZARC	Accession
63.	017_ATH	Amhara	North Gondar	12.1522	37.0150	DZARC	Accession
64.	019_ATH	Amhara	South Wollo	11.2636	39.6803	DZARC	Accession

Where, B/Gumuz = Benshangul-Gumuz, SNNP = Southern Nations, Nationalities, and Peoples, DZARC = Debre Zeit Agricultural Research Center, SARC = Sinana Agricultural Research Center.

Appendix Table 2. Combined analysis of variance of 12 morpho-agronomic traits recorded on 64 black cumin genotypes of Ethiopia at Debre Zeit and Kulumsa during the 2021 cropping season

Traits	Mean Square			CV (%)
	Location (DF = 1)	Genotype (DF = 63)	Location*Genotype (DF = 63)	
Days to flowering	48675.39***	6.33***	4.49***	2.1
Days to 50% flowering	35297***	8.23***	5.83***	0.3
Days to full blooming	47198***	10.42***	6.43***	1.1
Days to maturity	16949***	14.59***	13.59***	0.4
Plant height (cm)	2232.56***	55.67***	37.73***	7.4
Number of primary branches per plant	39.06ns	0.86ns	0.92ns	16.7
Number of capsules per plant	1600***	9.38*	12.94***	20.1
Number of seeds per capsule	14976***	83.19***	84.16***	6.5
Number of seeds per plant	29247816***	129767***	167665.65***	21.1
Thousand seed weight (g)	3.04*	0.02***	0.01**	3.3
Seed yield per plant (g)	95.48***	0.63***	0.76***	20.1
Seed yield per hectare (t)	23.28***	0.10***	0.08***	15.2

Where**, *=** statistically significant at $p < 0.05$, ***= Significant at $p < 0.01$, ***= Significant at $p < 0.001$, and ns= Non-significant at $p \geq 0.05$.

Appendix Table 3. Morphological traits of the 64 black cumin genotypes of Ethiopia

No.	**Genotype**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
1	Darbera	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
2	DERSHYE	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
3	ADEN	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
4	Sooressaa	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
5	Gammachis	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
6	Silingo	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
7	Kena	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
8	Qeneni	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
9	8502	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
10	9067	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
11	9068	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
12	9069	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
13	19884	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
14	90501	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
15	90502	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
16	90503	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
17	90504	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
18	90506	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
19	90510	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
20	208688	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
21	212520	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
22	215319	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det

Qualitative traits: **1**= Seed color, **2**= Seed shape, **3**= Leaf color, **4**= Leaf shape, **5**= Leaf type, **6**= Leaf margin, **7**= Leaf phyllotaxy, **8**= Petal color, **9**= Sepal color, **10**= Flower symmetry, **11**= Flower type, **12**= Color of immature capsules, **13**= Color of matured capsules, **14**= Capsule shape, **15**= Growth habit. Where, B=Black, Trian= Triangular, DD=Deeply-divided, Cpd= Compound, Zygo= Zygomorphic, Com= Complete, Det= Determinate, PD= Pinnately dissected, and Cyl= Cylinderical.

Appendix Table 3. Continued

No.	**Genotype**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
23	219970	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
24	223069	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
25	223070	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
26	223072	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
27	229808	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
28	230037	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
29	230038	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
30	230039	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
31	230040	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
32	230777	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
33	237989	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
34	240403	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
35	240404	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
36	242221	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
37	242222	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
38	242223	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
39	242224	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
40	242225	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
41	242528	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
42	242825	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
43	242826	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
44	242833	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
45	242835	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
46	242838	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det

Qualitative traits: **1**= Seed color, **2**= Seed shape, **3**= Leaf color, **4**= Leaf shape, **5**= Leaf type, **6**= Leaf margin, **7**= Leaf phyllotaxy, **8**= Petal color, **9**= Sepal color, **10**= Flower symmetry, **11**= Flower type, **12**= Color of immature capsules, **13**= Color of matured capsules, **14**= Capsule shape, **15**= Growth habit. Where, DD=Deeply-divided, B=Black, Trian= Triangular, Cpd= Compound, Zygo= Zygomorphic, Com= Complete, Det= Determinate, PD= Pinnately dissected, and Cyl= Cylinderical.

Appendix Table 3. Continued

No.	**Genotype**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
47	242839	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
48	242840	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
49	242841	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
50	242842	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
51	242844	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
52	242823	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
53	002_ATH	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
54	003_ATH	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
55	004_ATH	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
56	007_ATH	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
57	009_ATH	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
58	010_ATH	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
59	012_ATH	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
60	013_ATH	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
61	014_ATH	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
62	015_ATH	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
63	017_ATH	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det
64	019_ATH	Black	Trian	Green	DD	Cpd	PD	Alternate	White	Green	Zygo	Com	Green	Yellow	Cyl	Det

Qualitative traits: **1**= Seed color, **2**= Seed shape, **3**= Leaf color, **4**= Leaf shape, **5**= Leaf type, **6**= Leaf margin, **7**= Leaf phyllotaxy, **8**= Petal color, **9**= Sepal color, **10**= Flower symmetry, **11**= Flower type, **12**= Color of immature capsules, **13**= Color of matured capsules, **14**= Capsule shape, **15**= Growth habit. Where, DD=Deeply-divided, B=Black, Trian= Triangular, Cpd= Compound, Zygo= Zygomorphic, Com= Complete, Det= Determinate, PD= Pinnately dissected, and Cyl= Cylinderical.

Appendix Table 4. Mean performance of 12 morpho-agronomic traits in 64 black cumin genotypes of Ethiopia tested at Debre Zeit and Kulumsa during the 2021 cropping season

No.	**Genotype**	Morpho-agronomic traits											
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	Darbera	70	86.5	91	150	60	12	103	1291	2.15	2.70	0.26	0.850
2	DERSHYE	72.5	86.5	95	152	61	15	100	1486	2.20	3.17	0.24	0.810
3	ADEN	70	83.5	91	148	58	9	94	858	2.20	1.90	0.36	1.185
4	Sooressaa	72.5	86.5	95	152	56	11	97	1072	2.18	2.30	0.31	1.018
5	Gammachis	72.5	86.5	95	152	58	14	91	1286	2.15	2.69	0.23	0.758
6	Silingo	72.5	86.5	95	152	54	12	99	1157	2.25	2.61	0.23	0.778
7	Kena	72.5	86.5	95	152	67	11	107	1115	2.23	2.44	0.30	1.005
8	Qeneni	72.5	86.5	95	153	62	13	98	1327	2.08	2.72	0.25	0.850
9	8502	72.5	86.5	95	152	50	11	92	980	2.15	2.10	0.19	0.615
10	9067	72.5	86.5	95	152	55	12	98	1243	2.28	2.81	0.29	0.950
11	9068	72.5	86.5	95	153	51	12	95	1150	2.10	2.39	0.20	0.670
12	9069	72.5	88	94	152	60	8	102	821	2.23	1.82	0.27	0.905
13	19884	72.5	86.5	95	153	63	14	95	1359	2.20	2.90	0.30	1.005
14	90501	72.5	86.5	95	153	54	12	96	1173	2.28	2.63	0.30	0.998
15	90502	72.5	86.5	95	151	57	11	97	1017	2.18	2.21	0.24	0.803
16	90503	72.5	86.5	95	153	59	14	99	1469	2.18	3.18	0.35	1.165
17	90504	69	81.75	89	145	62	13	104	1357	2.25	3.02	0.40	1.343
18	90506	72.5	87	94	153	55	12	106	1279	2.20	2.80	0.21	0.695
19	90510	72.5	86.5	95	152	58	12	104	1255	2.15	2.66	0.30	0.993
20	208688	72	87	94	153	59	12	95	1181	2.13	2.49	0.36	1.185
21	212520	72.5	86.5	95	152	60	14	95	1354	2.13	2.79	0.28	0.935
22	215319	72.5	88	94	152	57	11	100	1106	2.18	2.39	0.27	0.893
23	219970	78	86.5	95	152	64	13	107	1330	2.28	3.00	0.38	1.255
24	223069	72.5	86.5	95	152	61	13	101	1363	2.08	2.79	0.28	0.930
25	223070	72.5	87.5	94	153	60	13	108	1353	2.28	3.11	0.27	0.900
26	223072	72.5	87	94	152	59	12	103	1302	2.23	2.83	0.25	0.843
27	229808	72.5	86.5	95	152	56	11	99	1100	2.18	2.31	0.20	0.660
28	230037	71.5	84.5	94	149	54	13	92	1168	2.25	2.63	0.31	1.015
29	230038	72.5	86.5	95	151	55	11	89	954	2.18	2.06	0.21	0.688
30	230039	72.5	86.5	95	152	48	13	95	1377	2.13	2.89	0.21	0.705
31	230040	69	81.5	89	145	46	11	92	968.5	2.23	2.15	0.17	0.570
32	230777	72.5	86.5	95	152	57	12	98	1157	2.20	2.54	0.27	0.905
33	237989	69	81.3	89	145	53	14	93	1330	2.18	2.84	0.16	0.530
34	240403	72.5	86.5	95	152	54	12	99	1293	2.15	2.68	0.22	0.723
35	240404	72.5	87	94	152	48	11	96	1105	2.25	2.47	0.21	0.705
36	242221	72.5	86.5	95	151	61	13	100	1339	2.15	2.87	0.28	0.923
37	242222	72.5	86.5	95	152	54	11	98	1108	2.18	2.38	0.34	1.123
38	242223	72.5	86.5	95	152	61	15	98	1540	2.15	3.28	0.28	0.933

Quantitative traits: 1 = Days to flowering, **2** = Days to 50% flowering, **3** = Days to full blooming, **4** = Days to maturity, **5** = Plant height (cm), **6** = Number of capsules per plant, **7** = Number of seeds per capsule, **8** = Number of seeds per plant, **9** = 1000-seed weight (g), **10** = Seed yield (g/plant), **11** = Seed yield (kg/plot), **12** = Seed yield (t/ha).

Appendix Table 4. Continued

No.	**Genotype**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
39	242224	72.5	86.5	95	151	63	15	104	1645	2.25	3.60	0.28	0.923
40	242225	69	81.5	89	145	61	11	102	1174	2.18	2.55	0.30	1.008
41	242528	72.5	86.5	95	152	62	15	103	1579	2.20	3.37	0.30	0.983
42	242825	72.5	86.5	95	151	55	10	101	1086	2.25	2.34	0.21	0.683
43	242826	72.5	86.5d	95	153	57	12	105	1249	2.33	2.87	0.29	0.958
44	242833	72.5	87	94	153	58	10	99	958.5	2.23	2.10	0.31	1.040
45	242835	72.5	86.5	95	151	53	13	107	1363	2.18	3.05	0.23	0.763
46	242838	72.5	86.5	95	153	56	14	99	1378	2.30	3.11	0.21	0.700
47	242839	72.5	86.5	95	151	62	14	103	1452	2.33	3.33	0.29	0.958
48	242840	72.5	86.5	95	152	53	11	95	1009	2.20	2.19	0.20	0.665
49	242841	73.5	88.5	96	154	57	10	100	967	2.35	2.28	0.31	1.023
50	242842	72.5	86.5	95	151	59	12	94	1145	2.28	2.57	0.28	0.943
51	242844	72.5	86.5	95	151	57	12	97	1180	2.25	2.55	0.23	0.775
52	242823	72.5	86.5	95	153	54	13	102	1350	2.25	2.99	0.25	0.845
53	002_ATH	69	82.5	89	145	55	12	93	1100	2.15	2.31	0.18	0.593
54	003_ATH	72.5	88	94	153	54	11	98	1110	2.20	2.47	0.22	0.733
55	004_ATH	72.5	86.5	95	151	54	12	93	1150	2.20	2.54	0.22	0.738
56	007_ATH	72.5	86.5	95	152	57	12	97	1110	2.10	2.28	0.23	0.763
57	009_ATH	72.5	87	94	153	56	13	101	1305	2.28	2.89	0.30	1.010
58	010_ATH	72.5	88	94	153	55	12	92	1077	2.20	2.37	0.21	0.710
59	012_ATH	72.5	87	94	153	56	12	93	1097	2.23	2.42	0.21	0.693
60	013_ATH	72.5	87	94	153	58	10	94	931	2.25	2.05	0.36	1.210
61	014_ATH	72.5	88	94	153	56	13	106	1345	2.30	3.01	0.33	1.093
62	015_ATH	72.5	86.5	95	152	58	11	95	1018	2.20	2.20	0.23	0.775
63	017_ATH	72.5	86.5	95	152	57	11	94	1073	2.20	2.32	0.28	0.920
64	019_ATH	72.5	86.5	95	152	59	14	94	1300	2.20	2.84	0.37	1.228
Mean		72.2	86.3	93.7	151.3	56.9	12	98	1207	2.21	2.63	0.27	0.884
LSD (5%)		2.99	3.41	3.6	5.21	8.7	5.08	13	578.6	0.1	1.23	0.12	0.39
CV (%)		2.1	0.3	1.1	0.4	7.4	20.1	6.5	21.1	3.3	20.1	15.4	15.2

Quantitative traits: **1** = Days to flowering, **2** = Days to 50% flowering, **3** = Days to full blooming, **4** = Days to maturity, **5** = Plant height (cm), **6** = Number of capsules per plant, **7** = Number of seeds per capsule, **8** = Number of seeds per plant, **9** = 1000-seed weight (g), **10** = Seed yield (g/plant), **11** = Seed yield (kg/plot), **12** = Seed yield (t/ha).

Appendix Table 5. Phenotypic path-coefficient analysis: Direct effect (diagonal) and indirect effect (off-diagonal)

	DF	**DFPF**	**DFB**	**DM**	**PH**	**NBP**	**NCP**	**NSC**	**NSP**	**TSW**	**SYP**	**RP**
DF	**0.0691**	-0.2684	-0.0447	0.268	0.0708	0.0073	-0.0178	-0.0028	-0.0143	0.012	0.0408	**0.12**
DFPF	0.0491	**-0.3781**	-0.0475	0.351	0.0455	-0.0016	0.0059	-0.003	-0.0043	0.0066	0.0163	**0.04**
DFB	0.0553	-0.3214	**-0.0558**	0.317	0.0405	0.0032	-0.0268	-0.0015	-0.0158	0.0044	0.0408	**0.04**
DM	0.0491	-0.3516	-0.0469	**0.3774**	0.0304	0.0008	0.003	-0.0026	-0.0057	0.0099	0.0163	**0.08**
PH	0.0097	-0.034	-0.0045	0.0226	**0.5058**	0.0162	-0.0714	-0.0085	-0.0473	0.0066	0.1348	**0.53****
NBP	0.0062	0.0076	-0.0022	0.0038	0.1012	**0.081**	-0.0565	-0.0013	-0.0301	0.0088	0.0817	**0.20***
NCP	0.0041	0.0076	-0.005	-0.0038	0.1214	0.0154	**-0.2974**	-0.0002	-0.1334	-0.0022	0.3635	**0.07**
NSC	0.009	-0.0529	-0.0039	0.0453	0.2023	0.0049	-0.003	**-0.0213**	-0.0459	0.0285	0.147	**0.31****
NSP	0.0069	-0.0113	-0.0061	0.0151	0.1669	0.017	-0.2766	-0.0068	**-0.1434**	0.0022	0.3962	**0.16**
TSW	0.0076	-0.0227	-0.0022	0.034	0.0304	0.0065	0.0059	-0.0055	-0.0029	**0.1095**	0.0694	**0.23***
SYP	0.0069	-0.0151	-0.0056	0.0151	0.1669	0.0162	-0.2647	-0.0077	-0.1391	0.0186	**0.4084**	**0.20***

Appendix Table 6. Genotypic path-coefficient analysis: Direct effect (diagonal) and indirect effect (off-diagonal)

	DF	**DFPF**	**DFB**	**DM**	**PH**	**NBP**	**NCP**	**NSC**	**NSP**	**TSW**	**SYP**	**RG**
DF	**2.6195**	-2.3616	-1.03	0.8359	0.187	0.1204	-0.0386	-0.2078	0.0203	0.0046	0.0704	**0.22**
DFPF	2.5409	**-2.4347**	-0.9775	0.8537	0.1122	0.0067	0.0463	-0.1226	0.0054	0.0025	0.0169	**0.05**
DFB	2.5671	-2.2642	**-1.051**	0.7915	0.1047	0.0468	-0.1699	-0.064	0.0257	0.0014	0.0619	**0.05**
DM	2.4623	-2.3373	-0.9354	**0.8893**	0.0898	-0.01	-0.0077	-0.0959	0.0122	0.0019	0.031	**0.1**
PH	0.6549	-0.3652	-0.1471	0.1067	**0.7482**	0.2107	-0.417	-0.3411	0.069	-0.0014	0.1323	**0.65****
NBP	0.943	-0.0487	-0.1471	-0.0267	0.4714	**0.3344**	-1.166	-0.2718	0.1462	-0.0117	0.2871	**0.51****
NCP	0.131	0.1461	-0.2312	0.0089	0.404	0.505	**-0.7722**	-0.469	0.1353	-0.0093	0.2814	**0.13**
NSC	1.0216	-0.56	-0.1261	0.1601	0.4788	0.1706	-0.6795	**-0.5329**	0.1069	0.0088	0.2617	**0.31***
NSP	0.3929	-0.0974	-0.1997	0.08	0.3816	0.3612	-0.7722	-0.421	**0.1353**	-0.0022	0.2814	**0.14**
TSW	0.7073	-0.3652	-0.0841	0.0978	-0.0599	-0.2308	0.4247	-0.2771	-0.0176	**0.017**	0.0478	**0.26***
SYP	0.6549	-0.1461	-0.2312	0.0978	0.3516	0.3411	-0.7722	-0.4956	0.1353	0.0029	**0.2814**	**0.22**

DF= Days to flowering, DFPF= Days to 50 % flowering, DFB= Days to full blooming, DM= Days to maturity, PH=Plant height (cm), NBP= Number of primary branches per plant, NCP= Number of capsules per plant, NSC= Number of seeds per capsule, NSP= Number of seeds per plant, TSW= Thousand seed weight (g), SYP= Seed yield per plant (g), RP= Phenotype correlation, RG = Genotype correlation..

*= Significant at $p < 0.05$, **= Significant at $p < 0.01$

Appendix Table 7. Mean performance of 4 biochemical traits in 64 black cumin genotypes of Ethiopia tested at Debre Zeit and Kulumsa during the 2021 cropping season

Genotype	Biochemical traits			
	ORC (%)	ORY (kg)	EOC (%)	EOY (kg)
Darbera	40.46	345.60	0.40	3.80
DERSHYE	43.23	349.87	0.33	2.68
ADEN	36.22	417.83	0.43	5.15
Sooressaa	39.10	362.79	0.31	3.20
Gammachis	38.82	292.12	0.38	3.02
Silingo	34.63	258.51	0.18	1.61
Kena	33.81	335.32	0.30	3.12
Qeneni	33.74	270.14	0.35	2.71
8502	34.25	211.37	0.40	2.65
9067	33.70	319.77	0.38	3.49
9068	44.21	298.56	0.23	1.49
9069	38.16	344.37	0.33	3.51
19884	44.97	456.67	0.25	2.70
90501	35.45	346.35	0.55	5.16
90502	42.63	345.36	0.47	3.68
90503	39.07	457.43	0.30	4.05
90504	42.89	582.03	0.55	7.18
90506	35.51	247.00	0.33	2.27
90510	37.86	381.49	0.35	4.35
208688	35.16	419.16	0.24	2.86
212520	34.97	324.56	0.29	2.83
215319	38.14	336.49	0.54	4.69
219970	40.18	499.95	0.49	5.91
223069	33.52	311.42	0.41	4.01
223070	38.51	347.25	0.38	3.44
223072	34.28	289.70	0.45	3.91
229808	40.18	263.10	0.53	2.17
230037	38.45	397.28	0.43	4.77
230038	35.64	248.34	0.40	2.74
230039	39.08	275.88	0.35	2.50
230040	41.11	233.07	0.40	2.63
230777	37.81	344.49	0.25	2.38
237989	35.13	185.97	0.35	1.81
240403	40.39	291.72	0.40	3.38
240404	39.98	272.60	0.18	1.13
242221	38.11	351.85	0.27	2.76
242222	40.37	455.81	0.39	5.18
242223	39.97	371.52	0.45	4.63
242224	39.32	363.72	0.35	3.37
242225	36.31	367.86	0.45	4.69
242528	39.57	388.68	0.38	3.75
242825	37.47	257.31	0.28	1.95
242826	33.98	324.05	0.40	3.81
242833	35.97	376.01	0.30	3.11
242835	34.47	263.35	0.30	2.50
242838	34.49	241.54	0.20	1.41
242839	32.96	315.43	0.45	4.48
242840	32.55	215.21	0.30	2.29
242841	37.39	384.81	0.30	2.71

Appendix Table 7. Continued

Genotype	**Biochemical traits**			
	ORC (%)	**ORY (kg)**	**EOC (%)**	**EOY (kg)**
242842	35.99	340.29	0.35	3.04
242844	35.70	274.40	0.30	1.42
242823	36.59	309.52	0.38	3.22
002_ATH	40.86	243.50	0.31	2.13
003_ATH	37.90	276.19	0.23	1.47
004_ATH	40.16	299.27	0.25	1.74
007_ATH	33.12	252.02	0.40	3.50
009_ATH	40.13	404.61	0.35	3.64
010_ATH	36.29	257.06	0.30	2.13
012_ATH	37.13	258.44	0.20	1.38
013_ATH	41.85	509.28	0.40	5.35
014_ATH	40.88	448.80	0.28	3.12
015_ATH	39.22	384.54	0.22	2.24
017_ATH	40.93	377.05	0.33	3.02
019_ATH	38.94	433.25	0.31	3.72
Mean	37.81	334.51	0.35	3.20
LSD (5%)	2.75	150.88	0.24	2.34
CV (%)	0.81	17.68	11.04	18.75

FOC = Fixed oil content, FOY = Fixed oil yield per hectare, EOC = Essential oil content, EOY = Essential oil yield per hectare.

Appendix Table 8. The 12 major volatile chemical compounds (%) of 64 black cumin genotypes of Ethiopia across two locations during 2021 cropping seasons

Genotype	**Chemical compound (%)**												**Total**
	CC1	**CC2**	**CC3**	**CC4**	**CC5**	**CC6**	**CC7**	**CC8**	**CC9**	**CC10**	**CC11**	**CC12**	
Darbera	5.1	1.08	1.43	34.74	0.95	1.32	9.31	1.62	1.62	32.58	6.6	3.28	**100**
DERSHYE	6.91	1.57	1.77	32.44	1.32	1.26	8.55	1.58	1.36	30.58	6.53	3.71	**98**
ADEN	4.42	1.2	1.5	31.38	1.08	1.57	9.27	1.59	1.37	33.99	8.59	3.49	**99**
Sooressaa	4.4	1.08	1.57	34.18	1.21	1.46	9.41	–	1.24	35.07	5.42	3.47	**98**
Gammachis	5.7	1.44	1.86	35.92	0.74	0.72	9.4	0.7	0.7	31.54	6.9	2.88	**98**
Silingo	3.65	0.67	1.36	30.53	0.62	1.61	9.74	1.66	1.68	32.75	10.08	3.17	**98**
Kena	7.06	1.14	2.07	37.11	0.68	3.01	7.32	0.64	0.61	32.03	4.68	2.91	**99**
Qeneni	5.39	1.25	1.72	34.58	0.98	1.2	7.77	1.56	1.52	36.4	4.38	2.47	**99**
8502	9.45	2.03	2.67	45.3	1.76	1.52	9.27	1.32	1.47	14.81	5.65	3.4	**99**
9067	4.26	1.16	1.53	34.12	1.04	1.32	9.11	1.51	1.26	34.5	5.88	3.86	**100**
9068	0.91	–	–	23.93	0.37	1.52	11.22	2.62	2.46	41.14	9.23	5.75	**99**
9069	8.04	1.64	2	44.44	1.86	0.64	8.16	0.87	1.63	20.71	4.11	3.27	**97**
19884	3.95	0.98	1.41	35.01	1.02	1.25	9.2	0.94	1.7	30.58	8.7	3.78	**99**
90501	11.99	2.75	2.61	38.53	1.56	1.06	6.09	0.66	1.2	21.68	3.79	2.1	**94**
90502	7.02	1.53	2.05	39.87	1.47	1.07	7.85	0.76	0.71	28.12	3.7	2.98	**97**
90503	4.57	0.81	1.45	34.46	1.12	1.05	7.52	–	0.55	40.94	4.27	2.82	**100**
90504	10.41	2.11	2.23	36.42	1.37	0.54	6.98	0.56	1.32	29.44	3.01	2.13	**97**
90506	5.34	1.25	1.68	33.04	1.01	1.49	8.76	1.53	1.71	34.07	5.84	3.86	**100**
90510	10.62	2.21	2.38	40.71	1.53	1.25	7.93	0.98	1.51	21.37	3.67	2.62	**97**
208688	7.12	1.14	1.84	35.8	1.25	1.06	7.29	0.86	1.48	30.13	5.08	2.42	**95**
212520	8.19	1.47	2.25	38.89	1.5	1.34	8.64	1.37	0.44	24.59	7.03	3.58	**99**
215319	7.3	1.5	2.04	39.9	1.48	1.43	7.96	0.96	1.01	26.24	4.28	3.4	**97**
219970	7.54	1.59	2	37.16	1.37	1.32	9.51	1.74	1.23	27.15	4.25	3.23	**98**
223069	7.12	1.57	2.14	39.52	1.41	0.58	8.9	1.48	1.1	26.03	4.54	3.46	**98**
223070	8.28	1.86	2.23	39.1	1.39	1.39	8.38	0.72	0.45	28.64	4.42	2.42	**99**
223072	4.4	1.05	1.56	35.33	1.01	1.43	9.16	1.57	1.29	34.37	5.09	3.31	**100**
229808	10.42	1.99	2.25	36.21	1.43	0.7	7.25	1.01	1.48	25.88	3.61	3.25	**95**
230037	8.07	1.75	2.25	36.16	1.38	1.12	8.12	0.94	1.49	28.51	4.5	3.33	**98**
230038	6.66	1.69	1.91	34.35	1.23	1.16	8.03	0.85	1.23	32.32	5.22	2.87	**98**
230039	6.9	1.53	1.79	35.76	1.3	0.7	8.2	0.9	0.97	25.96	6.03	4.49	**95**
230040	3.8	0.73	0.97	29.8	0.69	1.37	9.61	1.81	1.34	38	7.71	3.75	**100**
230777	3.62	0.93	1.23	32.08	0.68	1.22	9.32	1.76	0.65	36.93	7.69	3.53	**100**
237989	4.6	1.26	1.62	35.49	1.27	1.12	7.74	1.02	0.74	34.08	4.08	2.62	**96**
240403	5.64	1.57	1.7	33.26	1.23	1.31	9.35	1.9	1.27	34.3	5.35	2.66	**100**
240404	8.26	1.58	1.68	37.8	1.24	0.65	9.2	2.02	–	28.29	3.5	4.16	**98**
242221	12.97	2.66	2.82	45.61	1.68	0.66	6.82	0.74	0.52	16.61	3.36	3.09	**98**
242222	7.29	1.78	1.44	39.66	1.43	1.58	9.65	1.69	1.76	20.82	7.51	3.59	**98**
242223	6.5	1.68	1.89	34.99	1.36	1.22	7.37	1.14	1.29	33.11	4.64	2.46	**98**
242224	4.74	1.1	1.72	35.61	1.16	1.32	8.88	1.47	1.59	32.93	5.58	3.01	**99**
242225	6	1.37	1.8	38.89	1.36	1.26	8.82	1.41	0.67	29.98	4.15	3.12	**99**
242528	4.53	1.23	1.54	35.8	0.84	1.64	9.11	1.55	0.6	29.94	8.45	4.31	**100**
242825	5.58	1.39	2.06	42.63	1.54	1.82	11.29	1.73	1.5	18.38	6.66	4.56	**99**
242825	5.58	1.39	2.06	42.63	1.54	1.82	11.29	1.73	1.5	18.38	6.66	4.56	**99**
242826	5.24	1.26	1.71	32.32	0.99	1.26	8.82	1.83	1.6	33.29	7.58	3.57	**99**
242833	3.44	0.46	1.11	31.39	0.95	1.46	8.9	1.69	1.49	39.62	5.52	3.69	**100**

Appendix Table 8. Continued

Genotype	**Chemical compound (%)**												**Total**
	CC1	**CC2**	**CC3**	**CC4**	**CC5**	**CC6**	**CC7**	**CC8**	**CC9**	**CC10**	**CC11**	**CC12**	
242835	1.96	0.28	0.52	29.34	0.43	1.36	8.41	0.99	0.72	42.44	7.27	6.27	**100**
242838	1.22	—	0.5	25.86	0.39	1.47	10.7	3.34	1.59	37.29	11.48	5.52	**99**
242839	7.18	1.82	2.35	42.52	1.51	1.39	8.9	1.76	1.03	20.56	5.83	3.78	**99**
242840	3.1	0.47	1.19	31.71	0.89	1.19	9.88	2.01	1.58	34.94	8.86	4.18	**100**
242841	5.17	0.97	1.03	31.86	0.7	0.65	8.18	—	0.6	40.01	5.46	4.19	**99**
242842	4.44	0.92	1.43	30.76	0.95	1.11	9.04	1.63	1.39	38.42	4.99	3.39	**98**
242844	6.49	1.25	1.2	29.33	0.77	—	4.82	0.97	1.59	39.7	5.2	3.72	**95**
242823	3.63	1.01	1.32	30.6	0.96	0.85	8.81	0.58	1.83	38.75	8.12	3.55	**100**
002_ATH	8.09	1.74	2.21	39.45	1.61	1.58	7.46	1.41	1.85	23.96	4.27	3.68	**97**
003_ATH	3.37	0.81	0.5	34.4	1.08	1.49	10.93	2.13	1.72	31.91	5.8	5.86	**100**
004_ATH	2.55	0.56	0.54	28.69	0.47	1.47	10.31	2.24	0.79	37.59	7.04	6.97	**99**
007_ATH	1.32	—	0.55	28.33	0.45	1.55	10.98	2.88	1.74	36.38	9.75	5.42	**99**
009_ATH	3.84	0.81	1.39	37.9	1.22	1.46	10.16	2.07	1.53	29.29	4.59	4.54	**99**
010_ATH	3.42	0.69	0.88	33.88	1.04	1.37	10.07	1.66	1.58	34.47	5.67	3.99	**99**
012_ATH	11.13	2.24	1.97	38.74	1.4	0.65	7.63	1.05	0.65	24.71	3.2	3.37	**97**
013_ATH	3.75	0.73	1.35	31.09	0.7	1.25	8.06	1.63	0.62	35.21	5.42	2.61	**92**
014_ATH	2.81	0.77	1.18	30.73	0.47	0.62	8.4	2.29	0.91	40.35	7.92	3.55	**100**
015_ATH	3.6	0.66	1.36	31.35	0.9	1.57	10.53	1.76	0.94	32.61	8.41	4.94	**99**
017_ATH	4.34	1.02	1.86	39.84	1.44	1.39	9.11	0.88	1.3	26.58	6.55	3.29	**98**
019_ATH	3.73	0.65	1.46	33.52	1.04	1.7	10.34	1.83	1.72	25.74	5.28	5.99	**93**
Mean	**5.76**	**1.24**	**1.62**	**35.16**	**1.11**	**1.24**	**8.78**	**1.38**	**1.23**	**31.08**	**5.91**	**3.67**	
Minimum	**0.91**	0	0	**23.93**	0.37	0	**4.82**	0	0	**14.81**	**3.01**	**2.1**	
Maximum	**12.97**	2.75	2.82	**45.61**	1.86	3.01	**11.29**	3.34	2.46	**42.44**	**11.48**	**6.97**	

CC = Chemical composition: CC1 = α-Thujene, CC2 = α-Pinene, CC3 = β-Pinene, CC4 = ρ-Cymene, CC5 = D-Limonene, CC6 = 2-ethylhexyl isohexyl phthalate, CC7 = trans-4-methoxy thujane, CC8 = Terpinen-4-ol, CC9 = Phthalic acid, 2-fluorobenzyl heptyl ester, CC10 = Thymoquinone, CC11 = Carvacrol, CC12 = Longifolene.

Appendix Table 9. Lists of 94 black cumin (*N. sativa* L.) genotypes with their area of collection in Ethiopia

No.	Genotype	Population	Latitude (ºN)	Longitude (ºE)	Collection site	Elevation (m.a.s.l)
1	Darbera	Oromia	8.771	39.0021	East Shewa	1865
2	19889	SNNP	7.2481	36.4303	Kefisho Shekicho	2603
3	20789	Oromia	7.5475	39.6106	East Arsi	2309
4	30765	Oromia	7.1636	40.6478	Bale	2008
5	8502	Oromia	7	39.8	Bale	1953
6	90504	Oromia	8.05	38.7833	Arsi	.
7	223070	B/Gumuz	11	35.7625	Metekel	.
8	237989	Oromia	8.05	38.7833	Bale	.
9	242528	B/Gumuz	9.9858	34.6675	Asosa	1480
10	242841	Oromia	7.5511	39.5408	Arsi	2140
11	009 ATH	Amhara	12.2181	37.195	North Gondar	1861
12	20431	Tigray	13.2358	39.5397	Southern	2088
13	DERSHYE	Oromia	7.12	40.17	Bale	2400
14	19890	SNNP	7.0644	36.0667	Keficho Shekicho	1612
15	28584	Oromia	7.7989	40.8611	Bale	1488
16	30769	Oromia	7.0756	40.5725	Bale	1653
17	9067	Amhara	11.6856	37.02	West Gojam	1840
18	90506	Amhara	10.3333	38	East Gojam	.
19	223072	B/Gumuz	11	35.7625	Metekel	.
20	240403	SNNP	7.2342	35.7092	Keficho Shekicho	1540
21	242825	Oromia	7.5606	39.6081	Arsi	2300
22	242842	Oromia	7.5356	39.5364	Arsi	2155
23	010 ATH	Amhara	12.3575	37.1872	North Gondar	1834
24	20432	Tigray	14.2561	39.1061	Central	2014
25	ADEN	Oromia	8.771	39.0021	East Shewa	1865
26	20202	SNNP	7.2336	35.3869	Sheka	1350
27	28585	Oromia	7.7828	40.8756	Bale	1508
28	90512	Oromia	8.3667	39.8833	Arsi	2520
29	9068	Amhara	11.7611	37.0844	West Gojam	1854
30	90510	Oromia	9.1667	37.8333	West Shewa	.
31	229808	B/Gumuz	10.5	36.1667	Metekel	1700
32	240404	SNNP	7.25	36	Keficho Shekicho	.
33	242826	Oromia	7.5744	39.5969	Arsi	2310
34	242844	Oromia	7.6061	39.5233	Arsi	2360
35	012 ATH	Amhara	12.3417	37.1092	North Gondar	1886
36	20434	Tigray	14.1469	38.4117	North West	1962
37	Sooressaa	Oromia	7.12	40.17	Bale	2400
38	19056	Oromia	7.3806	40.4842	Bale	2051
39	28586	Oromia	7.2014	40.6297	Bale	2058
40	236832	Oromia	8.0167	38.0833	West Shewa	2320
41	9069	Amhara	10.6467	37.0858	West Gojam	2002
42	208688	Oromia	8.8167	40.4167	West Harerge	.
43	230037	Tigray	14.0667	38.0833	Egnaw	2400
44	242221	Amhara	10.8411	39.8167	South Wollo	1530
45	242833	Oromia	7.6508	39.4961	Arsi	2290
46	242823	Oromia	7.6036	39.5619	Arsi	2320
47	013 ATH	Amhara	12.3072	37.3158	North Gondar	1867
48	20435	Tigray	14.0161	39.4556	Eastern	2214
49	Gammachis	Oromia	7.12	40.17	Bale	2400

Appendix Table 9. Continued

No.	Genotype	Population	Latitude (ºN)	Longitude (ºE)	Collection site	Elevation (m.a.s.l)
50	20225	Oromia	9.9072	36.8781	East Wellega	1978
51	28587	Oromia	7.1644	40.6475	Bale	2007
52	19935	Amhara	11.5625	38.985	North Wollo	2431
53	19884	SNNP	7.2914	36.3669	Keficho Shekicho	2471
54	212520	Oromia	7.0167	39.9833	Bale	.
55	230038	Tigray	14.1667	38.75	Egnaw	1940
56	242222	Amhara	11.9681	39.0722	North Wollo	2130
57	242835	Oromia	7.6036	39.5636	Arsi	2340
58	002 ATH	Amhara	12.3129	37.322	North Gondar	1791
59	014 ATH	Amhara	12.2475	37.0406	North Gondar	1871
60	19874	SNNP	7.0644	36.0667	Keficho Shekicho	1612
61	Silingo	Oromia	7.12	40.17	Bale	2400
62	20226	Oromia	8.7086	36.4447	East Wellega	1499
63	28593	Oromia	6.9683	40.4864	Bale	1786
64	19590	Oromia	8.6631	39.1828	North Shewa	1993
65	90501	Amhara	10.6392	37.0869	West Gojam	.
66	215319	Amhara	11.0022	37.0031	East Gojam	.
67	230039	Tigray	14.0667	38.0833	Egnaw	.
68	242223	Tigray	14.1208	38.4747	Western	2080
69	242838	Oromia	7.6031	39.5414	Arsi	2355
70	003 ATH	Amhara	12.3401	37.3688	North Gondar	1798
71	015 ATH	Amhara	12.23	37.0325	North Gondar	1864
72	19877	SNNP	7.4083	36.4447	Keficho Shekicho	1385
73	Kena	Oromia	7.12	40.17	Bale	2400
74	20783	Oromia	7.5964	39.5575	East Arsi	2333
75	30753	Oromia	7.1081	40.6406	Bale	1957
76	90502	Amhara	11.95	37.7	South Gondar	1840
77	219970	Tigray	14.1367	38.3094	Western	.
78	230040	Tigray	14.0833	39.1	Egnaw	1950
79	242224	SNNP	6.1186	38.1186	Arbaminch	.
80	242839	Oromia	7.5719	39.5283	Arsi	2290
81	004 ATH	Amhara	12.3522	37.3389	North Gondar	1800
82	017 ATH	Amhara	12.1522	37.015	North Gondar	1851
83	19879	SNNP	7.2836	35.8517	Keficho Shekicho	1898
84	Qeneni	Oromia	7.12	40.17	Bale	2400
85	20784	Oromia	7.5967	39.6389	East Arsi	2335
86	30756	Oromia	7.1806	40.7172	Bale	1898
87	90503	Amhara	11.9833	37.7667	South Gondar	1900
88	223069	Amhara	11.0022	37.0031	East Gojam	.
89	230777	Oromia	5.1167	39.4833	Borena	1220
90	242225	Amhara	11.0333	39.7542	South Wollo	1800
91	242840	Oromia	7.5544	39.5333	Arsi	2160
92	007 ATH	Amhara	12.4253	37.2989	North Gondar	1821
93	019 ATH	Amhara	11.2636	39.6803	South Wollo	2142
94	19888	SNNP	7.3983	36.0519	Keficho Shekicho	1567

Where, B/Gumuz = Benshangul-Gumuz, SNNP = Southern Nations, Nationalities, and Peoples. ***Improved varieties**: Darbera, DERSHYE, ADEN, Sooressaa, Gammachis, Silingo, Kena, and Qeneni.

Appendix Figure 1. The photo gallery illustrated some of the phenomena at the testing site Kulumsa

Land preparation, layout and seeding; field performance of the genotypes at field condition

Field supervision with advisor, threshing and seed cleaning

Appendix Figure 2. The photo gallery illustrated some of the phenomena at the testing site Debre Zeit

Field performance of the genotypes at field condition

Field supervision with advisors, threshing and seed cleaning

Appendix Figure 3. FOSS SoxtecTM 8000 fixed oil extracting instrument

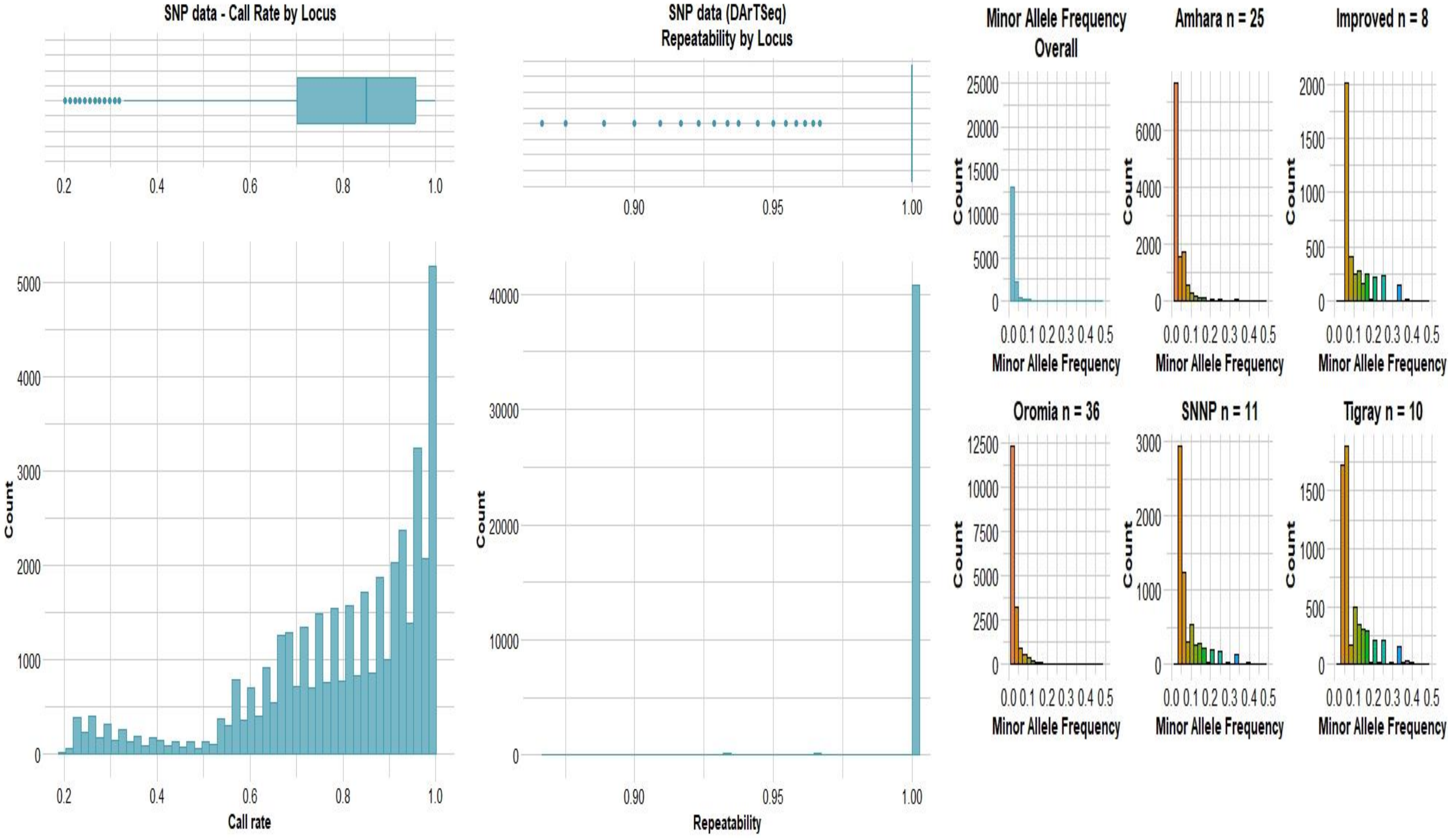
Appendix Figure 4. The assembled Clevenger apparatus setups

Appendix Figure 5. The photo gallery illustrated some of the phenomena during essential oil composition analysis using GC-MS

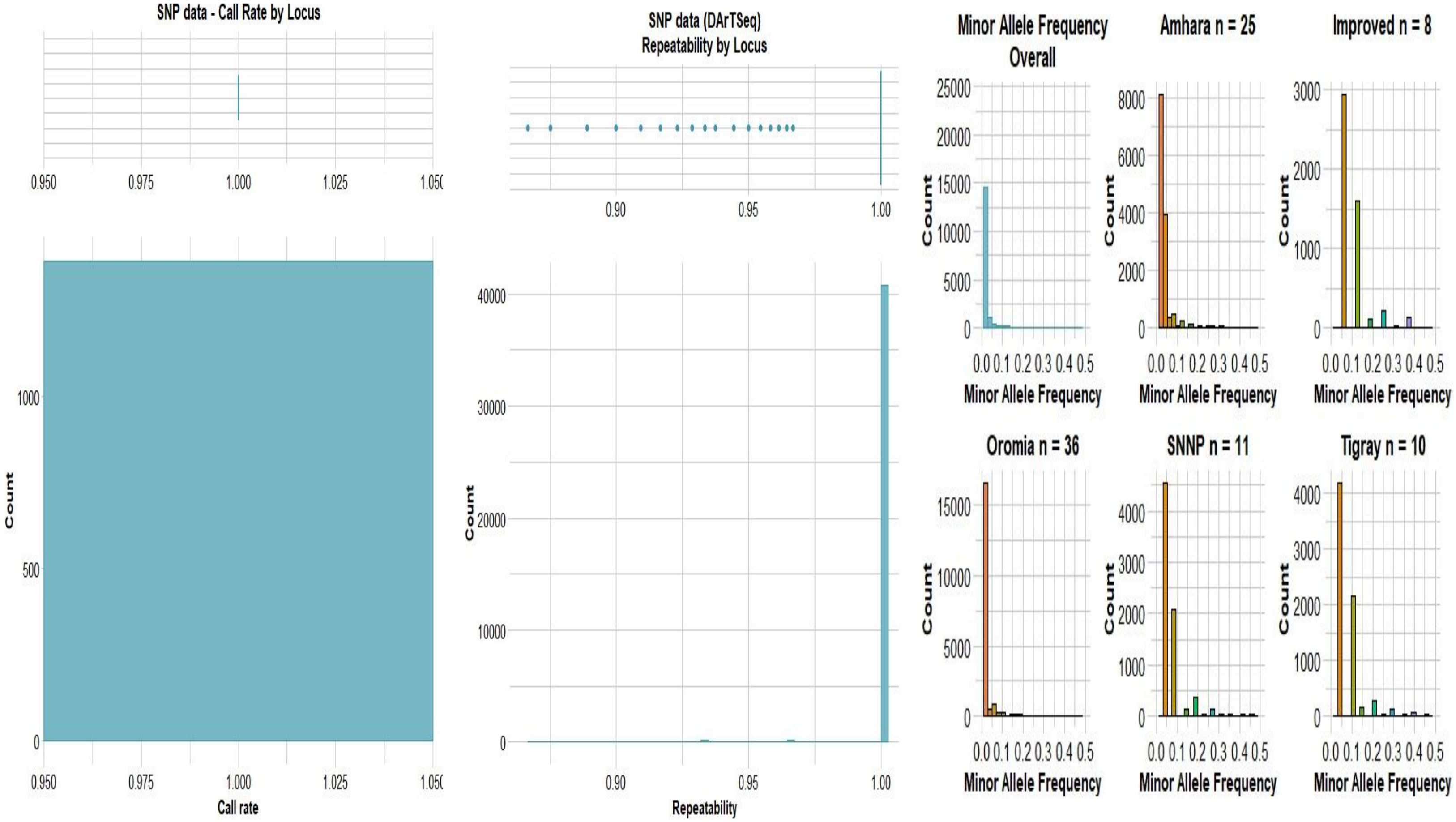
Conditioning of the GC-MS and sample preparation for GC-MS

Analyzing the samples with GC-MS

Appendix Figure 6. Diagrams show the report of DArTseq SNPs data quality report before imputation based on the filtered criteria



Appendix Figure 7. Diagrams show the report of DArTseq SNPs data quality report after imputation based on the filtered criteria



Appendix Figure 8. Diagrams show the report of DArTseq SNPs data quality report after imputation during filtering