Addis Ababa University

College of Natural and Computational Sciences,

Department of Microbial, Cellular and Molecular Biology

Genetic Diversity, Oil Content and Fatty Acid Composition of Vernonia [*Vernonia galamensis* (Cass.) Less] Germplasm in Ethiopia

By: Alemneh Mideksa

July/2020
Addis Ababa, Ethiopia

ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

Genetic Diversity, Oil Content and Fatty Acids Composition of Vernonia [Vernonia galamensis (Cass.) Less] Germplasm from Ethiopia

By

Alemneh Mideksa Egu

A Thesis Presented to the School of Graduate Studies of the Addis Ababa University in Partial Fulfillment of the Requirements for the PhD Degree in Biology (Applied Genetics)

Approved by Examining Board:

1. Dr. Kassahun Tesfaye (Advisor) _____
2. Prof. Kassahun Bantte (External Examiner) _____
3. Dr. Dawd Gudu (Internal Examiner) _____
4. Dr. Adey Feleke (Chair) _____

DECLARATION

I declare that the dissertation hereby submitted by me for the Degree of Doctor of Philosophy (PhD) in Biology (Applied Genetics) to the School of Graduate Studies of Addis Ababa University is my own work and has not previously been submitted by me or anybody else at another university. The materials obtained from other sources have been duly acknowledged in the dissertation.

PhD Candidate: Alemneh Mideksa _____ July, 2020

ACKNOWLEDGMENTS

First and foremost, I would like to extend my sincere thanks to Adama Science and Technology University, my home institution, for the study leave with my full salary during my study. I would like to extend my sincere thanks to Addis Ababa University, Department of Microbial, Cellular and Molecular Biology, for the admission, facility and monetary support during my study. I again extend my thanks to Ethiopian Biodiversity Institute (EBI) for providing the *Vernonia galamensis* accessions, and for permission to take the accessions to China for my study. Melka Werer and Wondo Genet Agricultural Research Centers also deserve special thanks for their kind support of providing land for field work and maintenance. Thanks to the Lab of Drug Discovery and Molecular Engineering, College of Plant Science and Technology, Huazhong (Central China) Agricultural University administration, staff members and technicians for the molecular lab facility and their good hospitality during my stay.

I would like to extend my thanks to my supervisors: Dr. Kassahun Tesfaye and Dr. Kifle Dagne for their excellent guiding, follow-up and support from the proposal development up to the end of my study. I received not only their valuable scientific comments and advices but also their hard-working habit, open mindedness, and their nice approach. I deeply saddened and socked by the death of Dr. Kifle Dagne, may his soul rest in peace. I want to also equally thankful to Prof. Xuebo Hu for accepting me to do part of my molecular research at Huazhong Agricultural University, China and for his remarkable supervision and guidance during molecular diversity study using SSR markers. I also extend my thanks to Mr. Seifu Juneidi for his valuable support and guidance at Huazhong Agricultural University.

Thanks to Dr. Adey Feleke (department head), Dr. Gurja Belay (former department head) and Dr. Dereje Beyene (stream coordinator) for their contentious support and encouragement during the study period. I extended my thanks to Dr. Geleta Dugassa for his support in reading materials and valuable scientific discussions we made. I am also grateful to Mr. Adane W/yohannes for his guidance in Biometrics and SAS software application, Dr.

Fikadu Gadissa for his technical and guidance of different software that I used for molecular analysis. I am also grateful to Oumer Abdi for his effort to produce map of my study area and, Agegnehu Wase and Hawi Nigusse for their moral, technical and material support.

Last but not least, my family deserves special thanks for their all sided support, my wife Yeshi Asrat and my children, Soliyana and Nanati thanks for their patience, love and generosity. My deepest gratitude goes to my Father Ato Mideksa Egu, and my mother W/o Shito Ferenji for the sacrifices they paid for my success. Especial thanks goes to my brothers and sisters, for their encouragement and moral supports. My families are my foundation and the source of my inspiration.

TABLE OF CONTENTS PAGE

LIST OF TABLES PAGE

LIST OF FIGURES PAGE

LIST OF APPENDICES PAGE

ACRONYMS AND ABBREVIATIONS

AFLP	Amplified Fragment Length Polymorphism
AMOVA	Analysis of molecular variance
ANOVA	Analysis of variance
CTAB	Cetyltriethyl ammonium bromide
EST-SSR	Expressed sequence tag-based simple sequence repeats
Fst	Genetic differentiation among populations
GC-FID	Gas chromatography with a flame ionization detector
GCV	Genotypic coefficient of variation
GLM	General linear model
Gst	Genetic differentiation among populations
He	Expected heterozygosity/Nei's gene diversity
Ho	Observed heterozygosity
I	Shannon information index
MSE	Mean square of error
MSG	Mean square of genotype
MWARC	Melka Werer Agricultural Research Center
MCMC	Markov Chain Mont Carol
Na	Number of allele
Ne	Number of effective allele
NJ	Neighbor joining
PCA	Principal component analysis
PCoA	Principal coordinate analysis
PCV	Phenotypic coefficient of variation
RAPD	Randomly amplified polymorphic DNA
SAS	Statistical Analysis System
SPSS	Statistical Package for Social Science
SSR	Simple sequence repeats
UPGMA	Unweighted Pair Group with Arithmetic Mean
WGARC	Wondo Genet Agricultural Research Center

ABSTRACT

Genetic diversity, oil content and fatty acid composition of Vernonia [*Vernonia galamensis* (Cass.) Less] germplasm in Ethiopia

Alemneh Mideksa

Addis Ababa University

Vernonia (*Vernonia galamensis*) is a potential novel industrial crop due to high sources of natural epoxidized oil, which can be used for the manufacturing of polyvinyl chloride, petrochemicals, cosmetic and pharmaceutical products. In spite of its high potential as an industrial plant, however, the plant is underutilized. No much attention has been given to research and conservation of this plant. The species is also under the threat of continued genetic erosion. This study was therefore initiated to assess the genetic diversity of *V. galamensis* accessions through morphological characterization and SSR molecular markers. In addition, oil content and fatty acid composition of the plant were investigated.

For agro-morphological trait study, 80 accessions of *V. galamensis* were grown in alpha-lattice design with two replications at Melka Werer and Wondo Genet Agricultural Research Centers. Data on a total of 13 quantitative traits were collected and analyzed using graph pad prism V-7, SAS V 9.0 and MINITAB V17.1 software. For molecular analysis, 20 SSR markers were used to analyses genetic diversity of a total of 10 populations using GenAlex software. Seed oil content was determined using Soxhlet method, and fatty acid composition was using Gas Chromatography Mass Spectroscopy. Data analysis for seed oil content and fatty acid composition were conducted using SAS V 9.0 and Minitab V17.1 software (Minitab, 2013).

Agro-morphological quantitative characters have highly significant variation ($p < 0.05$) for most of the quantitative traits among accessions except for days to emergence. The mean performance of seed yield per hectare ranged from 348.9 to 573.7 Kg ha^{-1}.The highest genotypic coefficient of variation (GCV) was observed for seed weight per plot (95.44), while the highest phenotypic coefficient of variation (PCV) was observed for branch length

(99.48). Seed yield per hectare showed significant and positive correlations with days to 50% heading (r = 0.50). High heritability coupled with high genetic advance as percentage of the mean was recorded for number of heads per plant and seed weight per plot. Number of heads per plant showed the highest positive direct effect (0.72) and genotypic correlation (r= 0.86) with seed yield per hectare. The result of principal component analysis indicated that those exhibited more than one eigenvalue were about 71.0% of variability.

The marker detected a total of 79 alleles/bands. The observed heterozygosity (Ho) values were quite low which ranged between 0.05 and 0.36, whereas the expected heterozygosity (He) ranged from 0.23 to 0.64. Two allelic patterns, East Showa and East Hararghe populations had the highest expected heterozygosity. The markers also showed the highest polymorphic information content that ranged between 0.50 (Vg-005) and 0.96 (Vg-002) with an average of 0.76. Molecular analysis of variance (AMOVA) showed 11% of variation among populations and 67% of the variation within populations. Dendrogram clusters, principal coordinate and population structure analyses showed that accessions collected from the same region of origin did not often grouped entirely together within a given major groups, suggesting that strong gene flow between populations.

The seed oil content and fatty acids composition analysis revealed significant variations, and the seed oil content ranged from 31.18% to 36.67% with an average of 33.24%. The fatty acids composition profiles also showed vernolic acid is the predominant fatty acid that ranged from 71.45% to 78.43% with an average of 73.72%.

Overall, the plant showed high genetic, phenotypic, and oil content variations, and will be used as the potential accessions (as potential parent sources) in *V. galamensis* improvement and genetic conservation endeavor in Ethiopia.

Key words: Vernonia, Epoxy oil, vernolic acid, dendrogram, heritability, heterozygosity, principal component

1. INTRODUCTION

1.1. Background and justification

Vernonia plant (*Vernonia galamensis* (Cass.) Less.; 2n = 18) belongs to the family Asteraceae (Compositae). It is a potential novel industrial oil crop due to high sources of natural epoxidized oil (Thompson *et al*., 1994a; Baye, 2002; Mebrahtu *et al.*, 2009; Shimelis and Hugo, 2011). The genus *Vernonia* contains more than 500 species that have two major centers of origin, South America and tropical Africa (Thompson *et al*., 1994b). *Vernonia galamensis* occurs naturally in Cape Verde, Senegal, Eritrea, Kenya, Ethiopia and Mozambique (Shimelis *et al*., 2013). The greatest diversity of *V.galamensis* is found in East Africa, mainly in Ethiopia and Kenya (Baye and Becker, 2005a).

Regarding its ecology, *V. galamensis* naturally grows as a wild weed in valleys or depressions in disturbed fields, river banks and on the road sides and along with acacia trees or in the forest (Shimelis *et al*., 2008; Shimelis and Gwata, 2013). In Eastern and South Eastern Ethiopia, the plant is found in farmlands, in the forests, and in the compounds of mosques and churches. The species is diversified in areas with well-drained soil and well defined seasons and with relatively warm climate. It occurs in areas with annual rainfall as low as 250 mm (Baye and Becker, 2005a). *Vernonia galamensis* subsp. *galamensis* variety *ethiopica* M. Gilbert was first identified by Perdue in 1964 in Eastern Ethiopia (Perdue *et al*., 1986). Later south and south-eastern Ethiopia was described as a natural habitat of this botanical variety. A number of studies demonstrated the presence of considerable variability in oil and vernolic acid (VOC) contents in this botanical variety (Thompson *et al*., 1994b).

Vernonia galamensis is herbaceous, usually annual but sometimes a short lived perennial (Perdue *et al*., 1986; Narina *et al*., 2012). Seeds are dicotyledonous and the embryo does not contain chlorophyll (Thompson *et al*., 1994). Fruit of Vernonia is non-fleshy and indehiscent. *V. galamensis* is propagated by seed that may show some dormancy for a few months after maturation and subsequent germination which on the average takes about 10 days (Mohammed *et al*., 1999).

Vernonia galamensis is a wild oil plant of economic importance and it is a rich source of novel epoxidized oil (vernolic acid) (Baye *et al*., 2005; Shimelis *et al*., 2011). In addition, the seed of this plant also contains other essential fatty acids such as linoleic acid, oleic acid, palmitic acid, stearic acid and a trace amount of arachidic acid (Thompson *et al*., 1994c; Mills, 1999). Oil rich in epoxidized fatty acid (Vernolic acid) is useful in the production of oleochemicals such as paints, plastic formulations and protective coatings. This oil is also useful in the manufacturing of polyvinyl chloride, adhesives, insecticides, cosmetic and pharmaceutical products (Baye *et al*.,2001; Shimelis *et al*., 2011). However, their production and management system is restricted, and their potential value for industries has been underutilized and underexploited.

Studying the genetic diversity of plants such as *V. galamensis* is important to enhance survival of plants in nature, and for effective utilization, conservation, management and improvement (Idrees and Irshad, 2014; Bhandari *et al*., 2017). The assessment of genetic diversity within and between plant populations is routinely performed on the basis of morphological, biochemical, and molecular markers (Singh *et al*., 2005). Agronomic and morphological traits-based systems are the most common and easily observable markers, usually used farmers, and breeders for classification and evaluation of yield or yield related traits. Furthermore, it provides good opportunity for effective selection and conservation measures (Govindaraj *et al*., 2015).

In addition, there is a need to characterize the diverse genetic resources using different statistical tools and utilize them in the breeding programmes (Burow *et al*., 2012). The process usually demands partitioning the overall variability of genotypes into its heritable and non-heritable components with the use of suitable genetic parameters such as genetic coefficient of variation, heritability, genetic advance and correlation, and multivariate analysis methods are also used especially for accessions/genotypes and agronomic traits that are several in numbers (Singh *et al*., 2005).

In this regards, limited studies with few accessions have been conducted to estimate the genetic diversity, oil content and fatty acid composition of *Vernonia galamensis* (Thompson *et al*., 1994c; Mohamed *et al*., 1999; Baye, 2002; Ramalema *et al*., 2010; Shimelis and Hugo, 2011; Shimelis and Gwata, 2013). Some of these studies were based on accessions from gene banks, while others were based on field collections. Its characterization involves growing the materials in a few locations. To date, only single molecular marker study with RAPD markers was reported by Ramalema *et al*. (2010). There were no SSR and other DNA markers based report(s) on *V. galamensis* genetic diversity in Ethiopia as well as elsewhere.

As a potential industrial plant, *V. galamensis*, limited information is available and it is considered only as a wild weed colonizing disturbed areas and bare agricultural lands (Baye 2002). The plant is also under threat of continued genetic erosion. This study is, therefore, important to include more accessions from all potential growing areas of the country, testing at multiple locations to clearly understand the extents of the existing genetic diversity for proper conservation and breeding programmes using morphological characterization, and SSR molecular markers. It also aimed at determining oil content and fatty acid composition of the plant.

1.2. Objectives

1.2.1. General objective

- To assess the genetic diversity of *Vernonia galamensis* plant in Ethiopia using morphological characteristics and molecular markers, and determine its oil content and fatty acid composition.

1.2.2. Specific objectives

- To assess the patterns of genetic diversity and population distributions of *Vernonia galamensis* germplasm using agro-morphological and molecular markers
- To estimate the magnitude of genetic variability, heritability of agro-morphological traits of *Vernonia galamensis*
- To determine the seed oil content and fatty acid composition of selected accessions of *Vernonia galamensis*

2. LITERATURE REVIEW

2.1. *Vernonia galamensis*

2.1.1. Origin and distribution

The genus *Vernonia* contains more than 500 species that have two major centers of origin, South America and tropical Africa (Perdue *et al.*, 1986). The species *Vernonia galamensis* is limited in its distribution primarily to eastern African countries including Ethiopia, Eritrea, Kenya and Tanzania. One variety has also been found in West Africa (Thompson, *et al.*, 1994a). Another species, *V. anthelmintica*, is originated in India (Gilbert, 1986), and a closely related to genus, also within the taxonomical tribe of *Vernonieae*, *Stokesia aster*, is native to the Southeastern U.S (Thompson, *et al.*, 1994b).

Vernonia galamensis subsp. *galamensis* variety *ethiopica* M. Gilbert was first identified by Perdue in 1964 in Eastern Ethiopia along the Harar-Jijiga road at 1700 m a.s.l (Perdue *et al*., 1986; Thompson, *et al.*, 1994 b). Later, it was recognized that South and Southeastern parts of Ethiopia are also natural habitat for the species (Baye *et al.*, 2001). Hence, Ethiopia is considered as the center of diversity for *V. galamensis* variety *ethiopica* (Perdue *et al*., 1986). *V. galamensis* is naturally distributed in marginal areas with less than 600 mm annual rainfall and at an elevation ranging from 700 to 2400 m above sea level in the eastern and southeastern parts of Ethiopia (Gilbert, 1986).

2.1.2. Taxonomy

Vernonia galamensis belongs to the family Asteraceae (Compositae), and the taxonomic works by Gilbert (1986) and Robinson (1999) have shown that *Vernonia* is better referred to the genus *Centrapalus* Cass., as *Centrapalus pauciflorus* (Willd.) H. Rob. (Syn. *Conyza pauciflora* Willd., and *Centrapalus galamensis* Cass.). *V. galamensis* is commonly known as ironweed in English, and locally it is known as Dunfare (Sidama), Fechatu (Afan oromo), Kefatheogie (Keficho), “Metaboko (Afan oromo) and Noya (Wolayita), different names in different localities (Baye *et al*., 2001).

According to Gilbert (1986) the taxonomic revision of *galamensis* is recognized to comprise the six subspecies such as *galamensis, mutomoensis, nairobensis, afromontana, gibbosa,* and *lushotoensis* (Baye *et al*., 2001). The sub-species *galamensis* is the most widely distributed; it is highly diverse and has four botanical varieties, namely var. *galamensis,* var*. petitiana,* var. *australis, and* var. *ethiopica* (Gilbert, 1986). In general, *V. galamensis* sub-species *afromontana* and sub-species *lushotoensis* often grow in the areas of higher elevation with high rainfall (Perdue *et al.*, 1986). The sub-species *nairobensis* and *gibbosa* occur in dry evergreen while the subspecies *galamensis* and *mutomoensis* are found in dry areas with low rainfall (Gilbert, 1986). The sub-species *galamensis* and *mutomoensis* are found in areas that receive 200 mm annual rainfall (Shimelis *et al*., 2008).

2.1.3. Ecology

Vernonia galamensis is endemic to East African countries, and naturally grows as a weed in the fields or in woodlands and adapted to the semi-arid tropics of dryland. It is a tropical, annual plant and requires a well-drained soil (Baye, 2002; Baye and Becker, 2005a). *V. galamensis* requires a rainy season that provides sufficient moisture to permit the main flower heads to develop (Baye *et al*., 2001). However, it can grow under low rainfall and marginal conditions that is suitable for dryland farming. But, a long rainy season permits developing secondary flower heads, resulting in non-uniformity of maturation and a risk of seed shattering (Hadebe *et al*., 2017). The plants tolerate substantial shading, high temperature and full sun as long as the soil moisture is sufficient (Shimelis *et al*., 2008). The plant grows best in a porous, well-drained and sandy soil, but does not live on heavy clay (Perdue *et al*., 1986).

2.1.4. Botany

Vernonia galamensis is herbaceous, usually annual but sometimes a short lived perennial plant (Mohammed *et al*., 1999; Narina *et al*., 2012). *V. galamensis* is propagated by seed that may show some dormancy for a few months after maturation and subsequent germination which on the average takes about 10 days (Thompson *et al*, 1994c). The shoot can grow up to 5 meters in height with single to multiple flower heads. The leaves are

alternate, sessile, membranous, hairy, and light to dark green in color (Gilbert, 1986). The apex is more or less acuminate and cuneate at the base. The margins irregularly dentate to sharply serrate to serrulate and upper surfaces range from asperous to sparsely long-pilose with the lower surfaces sparsely to quite densely long-pilose, sometimes the hairs dimorphic with small appressed hairs as well as the normal long erect hairs (Gilbert, 1986; Favi *et al*., 2008).

Figure 1. *Vernonia galamensis* grown at Wondo Genet Research Center

The color of florets is either pink to purple, or white (Figure 1). Flowering is indeterminate and can continue almost indefinitely so that a single plant can bear flower heads at every stage of development from young buds to mature seeds (Perdue *et al.*, 1986). The inflorescence consisting of a terminal flower head with lateral flower heads from the uppermost axis (Perdue *et al*., 1986; Favi *et al*., 2008). As a result, ripening of the heads of a plant is not uniform and shattering of mature fruiting heads occurs in most types (Thompson *et al*., 1994b). Dark brown to black seeds develop in seed heads. The seeds are crowned by pappi (Singular pappus) composed of numerous bristly hairs. Moreover, wide

variations were reported in the number of heads (capitula) per this species (Baye and Becker, 2005a).

2.1.5. Cytology

Chromosomes have been used to assign organisms to different taxa as members of the same species have similarity in their chromosome sets and related species have related chromosome sets (Gill and Singhal, 1998). In other words, the more closely related species are the more likely to have the same chromosome number, and the more distantly related, the more likely they are to have different chromosome number. Chromosome number may change in various ways and results in a new chromosome set which has effect on the general biology of the organism (Schubert, 2007). Polyploidy is the commonest of all changes in chromosome number, especially in plants (Stace, 2000). According to Dematteis (1998) polyploidy is common in the New World *Vernonia* species whereas the Old World species do not exhibit significant polyploidy.

The genus *Vernonia* is not well characterized cytologically. The chromosome number report shows that there is variation among *Vernonia* species (Hunter, 1964; Dematteis, 1998). Chromosome counts of n = 9, 10 and 20 were recorded from the Old World *Vernonia* species (Hunter, 1964). Ali (2008) reported 9 bivalents (2n = 18) meiotic chromosomes in *V. galamensis*. Kumbi (2009) also reported that chromosome numbers of 2n = 18 were observed by *V. galamensis* and *V. leopoldii*), whereas, 2n = 20 exhibited by *V. amygdalina, V. adoensis, V. schimpe*r and *V. myriantha*.

Another study on chromosomes of Ethiopian *Vernonia* species showed that *Vernonia* has evolved from closely related ancient species with x = 4 and x = 5 by mechanisms such as hybridization, polyploidization and aneuploidy decrease (Adegbite and Ayodele, 2004). *Vernonia* in the Old World has a dibasic chromosome number of n = 9 or 10 with polyploids of n = 18, 20 or 30, whereas in the New World it has basic chromosome number of n = 17 with polyploids of 34, 51, 58 or 68 (Hunter, 1964).

2.1.6. Mode of reproduction

Vernonia galamensis usually reproduces sexually by seed and it is a dicotyledonous, predominantly self-fertile plant. The natural outcrossing rate, ranges from 2.5 - 16.1% (Baye and Becker, 2004). Moreover, Thompson *et al*. (1994a) found that var. *petitiana* was self-sterile and allowing for ease of crossing with var. *ethiopica*. The authors also stated that the crosses generated the intra-species hybrids, and selections were performed for self-fertility, day-neutral photoperiodic response, good seed retention, non-dormant seed, and high yield (both for seed and oil) (Thompson *et al*., 1994a). Dispersal is commonly by wind, using hairy pappi (Singular pappus) (Perdue *et al*., 1986).

2.1.7. Seed oil content and fatty acid composition

2.1.7.1. Seed oil content of vernonia plant

Plant oils characterize as a resource of highly reduced carbon (Gunstone, 2001). There are wide-ranging of plants that contain oil. Soybean oil alone comprises nearly 60% of the world's vegetable oil market and is composed primarily of five fatty acids: palmitic acid, stearic acid, oleic acid, linoleic acid, and linolenic acids (Baye and Becker, 2005b). The commonly cultivated oil crops in Ethiopia include niger (noug), linseed, rapeseed/Ethiopian mustard, groundnut, sesame, sunflower, castor and safflower (Belayneh, 1986). The five oilseed plants most widely used in oleochemical industries are soybean, rapeseed, linseed, coconut, and castor. *V. galamensis* is one of the naturally occurring plants that contained epoxidized oils in its seeds known as *Vernonia* oil (Baye *et al*., 2005). Vernolic acid is the best quality fatty acid which is present as trivernolin (Figure 2).

2.1.7.2. Fatty acid composition

Seeds of *V. galamensis* are the major source of naturally occurring epoxidized fatty acid, vernolic acid, which comprises of 72-80% of the fatty acid in the oil, and other essential fatty acids such as linoleic acid (12-14%) and oleic acid (4-6%). The oil also contains saturated fatty acids namely palmitic acid (2-3%), stearic acid (2-3%) and trace amounts of

arachidic acid (< 2%) (Thompson *et al.*, 1994c; Baye *et al.*, 2005). The seed of *Vernonia* species produces useful natural epoxidized oil that is better than artificial epoxidized oils (Shimelis *et al.*, 2008). In addition, the leaves of *V. galamensis* also contain small amount of oil. The fatty acids compositions of the leaves include palmitic acid (12-22%), linolenic acid (41-59%), and a trace amount of vernolic acid (Baye *et al.*, 2005).

Many oilseed crop species have been investigated as potential sources of oils used in the oleochemical industry. Several of them contain a high proportion of industrially desirable fatty acids, such as linoleic acid (*Madia sativa* Molina), linolenic acid (*Lepidium sativum* L. and *Camelina sativa* (L.) Crtz.), and epoxy oils (*Euphorbia lagascae* Sprengel and *Vernonia galamensis* (Cass.) Less.) (Perdue, 1988). Plant oil in the form of triacylglycerol (TAG) is a natural resource to supplement or replace petroleum. Plant species have been found to contain high levels of unusual fatty acids (UFA) such as hydroxy, epoxy and acetylenic fatty acids (Perdue *et al.*, 1986)**.** For example, an epoxy fatty acid known as vernolic acid constitutes up to 30-90% of total fatty acids in the seeds of *Vernonia galamensis, Euphorbia lagascae, Stokesia laevis, Crepis palaestina* and *Bernardia pulchella* (Perdue, 1988; Pascual and Correal, 1992; Thompson *et al.*, 1994a).

The unique structure of vernolic acid enables a wide variety of characteristic of the ester group, the double bond, and the epoxy group to occur as shown in Figure 2.

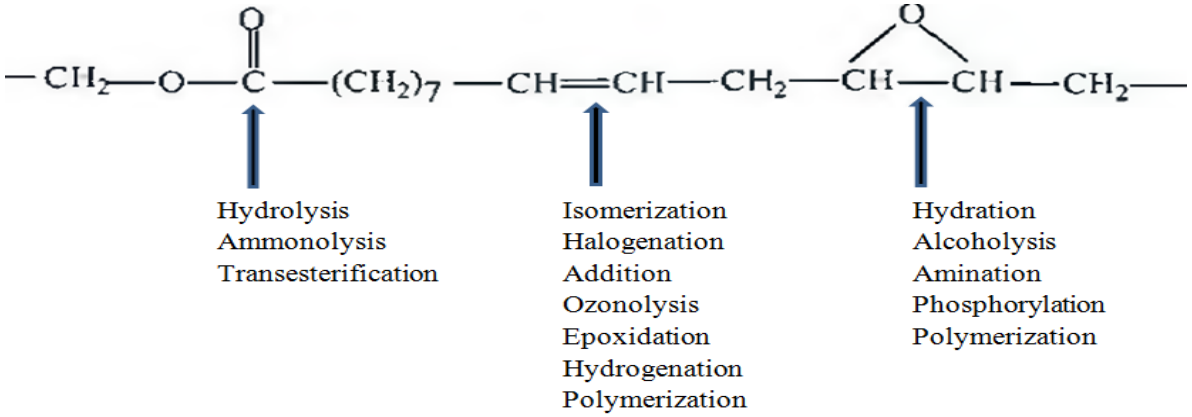

Figure 2. Structure of vernolic acid, arrows indicate the ester group, the double bond, and the epoxy group, respectively (Thompson *et al.*, 1994a)

The epoxy or oxirane group (the oxygen atom attached to two carbon atoms) of such triglyceride oil makes this material useful in plastics and coatings (Figure 3). It serves as highly reactive site where one triglyceride molecule can become attached to adjacent molecules, and these to others, to form interlocking polymer networks (Baye *et al*., 2001).

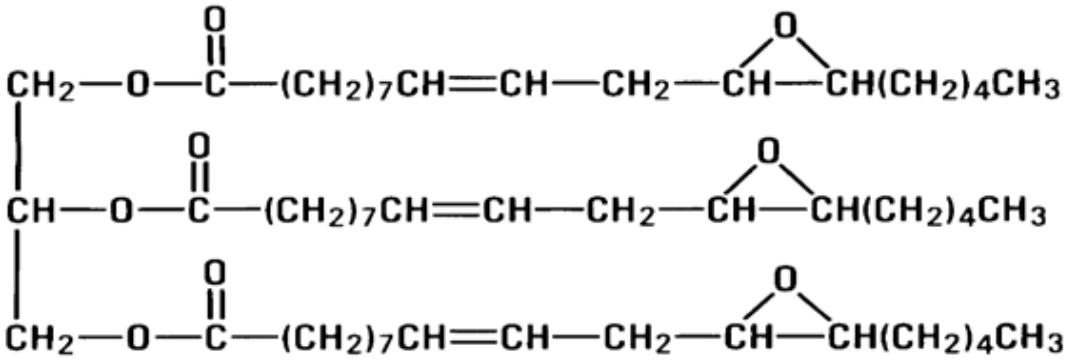

Figure 3. The structure of trivernolin (Thompson *et al.*, 1994a)

2.1.8. Economic importance

Vernonia galamensis is economically important because it is a new and novel oil sources (Baye, 2002). The seed of this species contain high amount of vernolic acid (Baye and Becker, 2005a; Shimelis *et al*., 2008), which has many applications in the chemical, pharmaceutical and agro-processing industries. This oil can be used in the paint industry as a component of low volatile organic solvent paints. It is also used in coatings, widely as a plasticizer of polyvinylchloride (PVC) and as a structural component of polymers (Thompson *et al*., 1994a).

The industrial use of *Vernonia* seed oil (vernolic acid) is important to reduce emissions of volatile organic compounds, and thus is environmentally friendly. Also it is less expensive (Bhardwaj *et al*., 2000). In contrast, chemically synthesized epoxidized soybean and linseed oil are highly viscous, expensive (need high epoxidation process) and contains volatile organic compounds (VOC) with high emission to the environment during processing and use. In addition, artificial epoxidized oil is also semi-solid at 10°C and can no longer be

poured at 0°C while vernolic acid can be poured even below the freezing point (Mebrahtu *et al*., 2009; Shimelis and Gwata, 2013). The low viscosity of vernonia oil makes it a good solvent in paint manufacture, while its highly reactive epoxy group makes it to chemically bind in the dried paint rather than evaporating to the atmosphere (Shimelis and Hugo, 2011).

Now-a-days, paint industries use VOC which is the major pollutants of the environment (Baye and Becker, 2005b). Therefore, the naturally occurring epoxidized oil from *V. galamensis* would be a better substitute for the volatile solvents used in paint industries because of their low VOC and environment friendly products. In addition, other products that are being developed from vernolic acids are degradable lubricants and lubricant additives, epoxy resins, adhesives, insecticides and insect repellants, crop-oil concentrates, and the formulation of carriers for slow-release pesticides (Bhardwaj *et al*., 2000; Narina *et al*., 2012). Moreover, vernonia oil provides excellent adhesion, flexibility and chipping resistance, resistance to acid, alkaline and non-polar solvents when incorporated into heat-baked films and coatings (Mebrahtu *et al*., 2009). The potential use of vernonia seed oil as a petroleum substitute is important since the demand for petroleum each year for the production of plastics and industrial petrochemicals is very high (Mebrahtu *et al*., 2009).

Many oils including vernonia seed oil are currently converted, through extensive oleochemical processing, to value added products, such as lubricants, polymer additives or surfactants (Shimelis and Hugo, 2011; Shimelis and Gwata, 2013). In most cases, these products are based on fatty acids, their methyl esters or fatty alcohols as intermediates, because the reactive head group of these compounds is a candidate for chemical derivatization (Baye and Becker, 2005b; Shimelis *et al*., 2008). Furthermore, press cake obtained from vernonia seed oil can be used as an animal food, and the seed oil also has unique medical applications when applied to human or animal skin to treat skin diseases and wound, in folk medicine to treat chest pain in Tanzania, and stomach pain in Kenya, are among other uses (Baye, 2002).Commercial production of *V. galamensis* tried in Ethiopia by Ver-Tech^{TM}. However, at a time large-scale commercial production was needed and no data on production is available (Perdue, 1988).

2.2. Genetic diversity of *Vernonia galamensis*

Genetic diversity can be expressed as any variation in the nucleotides, genes, chromosomes, or genomes of a species at the level of individual, population, species and sub-species for a given time (Fu, 2015). Genetic diversity reflects the presence of different alleles in the gene pool, and hence, different genotypes within populations. Genetic diversity can be measured by counting the number of different genes in a gene pool (Bhandari *et al*., 2017). Populations with greater genetic diversity are experienced with high fertility, low mortality and able to evolve in response to a changing environment than those with low genetic diversity even in good environmental conditions (Bhandari *et al*., 2017). The assessment of genetic diversity within and between plant populations is routinely performed on the basis of morphological, biochemical, and molecular markers (Shimelis *et al*., 2008; Govindaraj *et al*., 2015).

2.2.1. Morphological markers based genetic diversity

Several studies have been carried out to estimate the genetic diversity of *Vernonia galamensis* using agro-morphological traits to assess the phenotypic variability among its accessions (Thompson *et al*., 1994c; Mohamed *et al*., 1999; Baye, 2002; Ramalema *et al*., 2010; Shimelis and Hugo, 2011; Shimelis and Gwata, 2013), and oil content and fatty acid composition (Baye and Becker, 2005b; Mebrahtu *et al*., 2009). Some of these studies were based on accessions from gene banks, while others were based on field collections. The agro-morphological studies and fatty acids analysis indicated that a high genetic diversity existed in *V. galamensis* accessions (Baye *et al*., 2001).

Genetic variability analyses of quantitative characters involve the collection of germplasm from individual genotypes from the field/gene bank (Burton, 1952; Johnson*et al*., 1955). Its characterization involves growing the materials in a replicated experiment under uniform environmental conditions and measuring a range of vegetative and reproductive parameters (Kumar *et al*., 1985). Phenotypic variability is determined by the cumulative action of many genes (polygenes) and their interaction with the environment that can vary among

individuals over a given range to produce a continuous distribution of phenotypes (Kisua *et al*., 2015). Moreover, genetic variability study helps to understand evolutionary forces, the level of genetic relationship and the like that influence their survival (Govindaraj *et al*., 2015). Estimates of genetic parameters serve as a base for selection, and thus the degree of variability for a given character is a basic prerequisite for its improvement (Shimelis *et al*., 2011).

Quantitative trait diversity studies were conducted in *V. galamensis* by different authors who exhibited the existence of genetic variability among accessions (Perdue *et al*., 1986; Thompson *et al*., 1994c; Shimelis and Gwata, 2013). Baye and Becker (2005b) reported that highly significant mean squares differences were present for all the characters such as days to heading, days to flowering, days to maturity, plant height, number of heads, seed yield and oil content, except for days to emergence (Baye, 2002).

Perdue *et al*. (1986) conducted field experiments on *Vernonia* species collected from the countries of India, Pakistan, Ethiopia and Kenya, and reported the existence of a considerable genetic diversity among the studied accessions. They also indicated that African species had good natural seed retention and they are promising new crop for semiarid tropical areas. Thompson *et al*. (1994a) characterized the agro-morphology of *V. galamensis* that where grown in USA, Arizona and indicated that considerable variation exists among the accessions. Mohammed *et al*. (1999) evaluated *Vernonia galamensis* accessions for agronomic and chemical parameters and reported that significant differences existed for seed yield, oil content and vernolic acid. Baye *et al*. (2001) conducted field experiments at three sites (Haramaya, Harar and Babile) using *Vernonia galamensis* accessions that were collected from its grown area in Ethiopia, and reported that great variations were observed among the studied materials. Similarly, Shimelis *et al*. (2008) conducted field work in the Limpopo province of South Africa and reported that significant differences were found among *V. galamensis* var. *ethiopica* lines for agronomical traits as well as for seed oil content and fatty acids composition. This genetic variability could give the possibilities for genetic improvement of the germplasm.

With respect to oil content and fatty acid composition, Bhardwaji *et al*. (2007) conducted field experiments at the Randolph Farm of the Virginia State University and reported that significant variation was found among fatty acid concentrations of *V. galamensis*. Mohammed *et al*. (1999), Baye and Becker (2005a) and Shimelis *et al*. (2008) evaluated the seed oil content and fatty acid composition of *V. galamensis* germplasm and showed the existence of wide genetic variability for seed yield, oil content and vernolic fatty acid.

2.2.1.1. Genetic variability of quantitative traits

There are different statistical techniques used to evaluate the genetic variability of crop plants. Genetic variability in phenotypic characters is important for the improvement of the characters through selection in breeding program (Falconer, 1986). The phenotypic variation consists of genetic and environmental components. The genetic components include additive genetic variance, and variance due to non-additive gene action, such as dominance and epistasis. The non-genetic components include environmental and genotype-environment interactions (Johnson *et al*., 1955).

The parameters, phenotypic coefficient of variation (PCV) and genotypic coefficient of variation (GCV) are useful for comparing the relative amount of phenotypic and genotypic variations in different traits and they are very useful to estimate the scope for improvement by selection (Allard, 1991). Baye and Becker (2005a) reported that in *V. galamensis* phenotypic coefficients of variation (PCV) were the lowest for days to emergence, vernolic acid content, and seed yield per plant while seed yield per hectare had the highest PCV. The reliability of a parameter to be selected for breeding program, among other factors, is dependent on the magnitude of its coefficient of variations especially the GCV (Burton, 1955; Falconer, 1989). When the PCV is higher than the GCV for a character, it indicates that the environment has an effect on the expression of the traits (Johnson *et al*., 1955). However, if the GCV is higher than environmental coefficient of variations (ECV) for the traits, indicatives that genes influence the inheritance of the traits more than the environment does (Panse, 1957; Kumar *et al*., 1985).

2.2.1.2. Heritability and genetic advance

Heritability provides information on the extent to which a particular morphogenetic character can be transmitted to successive generations, an index of transmissibility (Fehr, 1987). Heritability is the percentage of phenotypic variance that is attributed to genetic variation. High heritability indicates that the environmental influence is minimal on characters (Johnson *et al*., 1955). In another word, if the percentage of heritability is high, the character is heritable but if it is small, the environment is correspondingly prominent in the character expression. Allard (1991) indicated that the heritability values for quantitative traits are low mainly due to their sensitivity to environmental factors. Moreover, heritability should be used along with genetic advance in predicting the efficiency of selection (Falconer, 1989).

Broad sense heritability is the ratio of the total genetic variance to the phenotypic variance, and it gives useful information about the relative contribution of genes to the phenotypic variation. Moreover, narrow sense heritability is the ratio of additive genetic variance to the total phenotypic variance (Allard, 1991). A high value of narrow sense heritability indicates that the character investigated is simply inherited, while a low estimate shows that the character is less heritable (Falconer, 1986).

High genetic advance (GA) with high heritability estimates offers the most effective condition for selection (Panse, 1957). Genetic advance as percentage of the mean gives more precise result in comparison to only genetic advance. In general, genetic variability, heritability and genetic advance are prerequisites for a breeding program and provide to the breeders the opportunities to select high yielding genotypes or to combine or transfer genes having desirable traits (Fehr, 1987). Baye (2002) reported that in *V. galamensis* plant height and seed yield per hectare showed high heritability with moderate genetic advance (GA), stem diameter, branch height, 1000-seed weight and oil content showed high heritability and low GA (Shimelis and Hugo, 2011). In addition, high genetic advance as percentage of mean coupled with high heritability was recorded for the characters such as number of seeds per head, number of branches per plant and number of heads per plant, indicating that

selection for these characters could be effective due to additive gene action (Shimelis *et al*., 2008).

2.2.1.3. Associations of major quantitative traits

Correlation coefficient is used to find out the degree and direction of the relationship between two or more variables (Miller *et al*., 1958). Correlation studies enable the breeder to know the strength of relationships between various characters as well as the magnitude and direction of changes expected during selection (Falconer, 1986). Genotypic and phenotypic correlation coefficient studies would provide reliable information about the nature, extent and the direction of selection especially when the breeder needs to combine high yield potential with desirable traits. The correlation coefficient analysis is useful in the selection of several traits that influence yield simultaneously (Johnson *et al*., 1955).

Association studies of quantitative traits in *V. galamensis* were conducted by several authors. Baye and Becker (2005a) reported that 1,000-seed weight showed strong negative associations with the palmitic, stearic, linoleic and oleic fatty acids, but a strong positive association with vernolic acid and oil content. Generally, the positive association among vernolic acid content, oil yield and seed yield per plant indicate the possibility of improving these important traits simultaneously. Bhardwaj *et al*. (2007) reported that significant association was existed between the concentrations of oil and various fatty acids, and vernolic acid was positively correlated with oil content but negatively correlated with palmitic, stearic, linoleic and oleic fatty acids. Shimelis and Hugo (2011) conducted a field experiment at Limpopo, South Africa, and reported that days to 50% flowering was positively and significantly correlated with plant height, but negatively associated with capsule per plant, seeds per head and seed yield per hectare.

However, association of traits determined by correlation coefficient analysis alone may not provide an exact picture of the relative importance of direct and indirect influence of each of yield components on seed yield (Miller *et al*., 1958). So, path coefficient analysis was used

to measure the direct influence of one variable upon the other, and permits separation of correlation coefficients into components of direct and indirect effects (Shimelis and Hugo, 2011). It is used to determine the interrelationship between grain yield and other yield components (Dewey and Lu, 1959). In addition, multivariate analysis (clustering and principal component analysis) is very useful for characterization, evaluation and classification of plant genetic material (Johnson and Wichern, 1988).

In general, morphological traits are crucial in determining adaptation, evolutionary forces, genetic diversity, and they are inexpensive, rapid and simple to score, and the traits cannot need sophisticated biochemical and molecular techniques (Kisua *et al*., 2015). However, morphological traits have a number of limitations, including low polymorphism, low heritability, late expression, and vulnerable to strong environmental influences. These limitations in turn may affect the estimation of genetic relationships. Therefore, to be useful, morphological measurements should be accomplished in replicated trials. This may be labor intensive and time consuming. However, if the trials are highly heritable, morphological markers are one of the choices for diversity studies because the inheritance of the marker can be monitored visually (Govindaraj *et al*., 2015).

2.2.2. Molecular markers based genetic diversity

Molecular markers are the most important and offer numerous advantages over conventional morphological and biochemical markers (Anumalla *et al*., 2015). DNA markers are not affected by environmental factors, easy to detect, applied to any part of the genome and abundant throughout the genome. Many DNA based marker techniques such as Restriction Fragment Length Polymorphisms (RFLP), Random Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphism (AFLP), Inter-Simple Sequence Repeat (ISSR), Simple Sequence Repeats (SSR), Single Nucleotide Polymorphism (SNP) and others have been used in ecological, evolutionary, taxonomical, phylogenetic and genetic studies of plant sciences (Mondini *et al*., 2009). However, each marker has its own advantages and disadvantages.

To date only a few studies are available that has dealt with the genetic diversity of *V. galamensis* accessions. RAPD markers study was done on *V. galamensis* in South Africa to determine the extent and distribution of genetic variations and reported that a moderate genetic diversity was observed (Ramalema *et al*., 2010). Narina *et al*. (2012) identified 63 microsatellite (SSR) from *V. galamensis* accessions collected from Eritrea. There were no other DNA marker based report(s) on *V. galamensis* diversity in Ethiopia as well as elsewhere.

2.2.2.1. SSR markers and its advantages over the other molecular markers

Simple sequence repeats (SSR) also known as microsatellites or short tandem repeats (STR), are tandem repeats motifs of 1-6 nucleotides found at high frequency in the nuclear genomes of most taxa. Microsatellites are classified according to the composition of the repeated sequences, as follows: perfect repeats, without any interruption, such as $(GT)_{12}$; imperfect repeats e.g.$(AT)_{12}GC\ (AT)_6$, interrupted, do not match the motif, example $(GT)_5\ A\ (GT)_8T$; and composite repeats, at which two or more microsatellite repeats (classes) are adjacent to one another, such as $(GT)_8\ (CA)_6$ (Abdel-Mawgood, 2012) .

Microsatellite (SSR) markers can be amplified for identification by the polymerase chain reaction (PCR), using two unique sequences which are complementary to the flanking regions as primers (Govindaraj *et al*., 2015). This is followed by agarose or polyacrylamide gel electrophoresis, and then the polymorphism is visualized by staining with silver or fluorescent (polyacrylamide) or with ethidium bromide solution (under ultraviolet light, agarose) (Mondini *et al*., 2009). The extent of polymorphism detected depends upon the amplification conditions, efficiency of gel assays, and clarity of gel profile (Saxena *et al*., 2011).

SSR markers are advantages over other DNA markers due to their random genome distribution, co-dominant inheritance (distinguishes homozygotes from that of heterozygotes), high level of polymorphism due to high mutation rate affecting the number

of repeat units, good genome coverage, high clarity and reproducibility, low operational cost, hyper-variability, amenability to automation, ease of multiplexing and use with low quality DNA (Govindaraj *et al*., 2015). SSR markers have been widely used in determination of genetic diversity, germplasm fingerprinting, heterosis analysis, tracing germplasm migrations, gene flow, genetic linkage mapping and phylogenetic studies (Varshney *et al*., 2005; Saxena *et al*., 2011).

According to Hosseinzadeh-Colagar *et al*. (2016) the excessive rate of mutation, high number of alleles and frequency in the genomic DNA, have made SSRs very effective molecular markers in population genetics, genome mapping, taxonomic study, linkage analysis, genetic fingerprinting and diversity (Varshney *et al*., 2005). The variations or polymorphism of SSRs is a result of polymerase slippage during DNA replication or unequal crossing over (Saxena *et al*., 2011; Govindaraj *et al*., 2015). SSR markers, however, have limitations such as genomic sequencing is needed to design specific primers; it is also not very cost effective and requires much discovery and optimization for each species before use (Abdel-Mawgood, 2012). To date, no information on the use of SSR markers to study the genetic diversity of *V. galamensis*.

3. MATERIALS AND METHODS

This study has three main parts that include agro-morphological diversity analysis of data collected from two experimental sites, diversity analysis using SSRs, and seed oil content and fatty acid composition of *V. galamensis* populations from Ethiopia.

3.1. Genetic diversity study using agro-morphological traits

3.1.1. Description of the study areas

For agro-morphological study, eighty accessions of *V. galamensis* were grown at two study sites, Melka Werer Agricultural Research Center (MWARC) and Wondo Genet Agricultural Research Center (WGARC) located in eastern and southeastern parts of Ethiopia, respectively (Table 1). Melka Werer Agricultural Research Center is located at elevation of 750 m above sea level and latitude of 9°16' N and a longitude of 40°9' E. It has an annual temperature ranging from 26.7 ^{o}C - 40.8 °C and receives an average annual rainfall of 590 mm, which is erratic and uneven in distribution. The site is dominated by vertisols and flui-soil textures (Altaye *et al*., 2016). The Wondo Genet Agricultural Research Center is at an altitude of 1765 m above sea level, 7°19’ N, 38°7’ E and has an annual average rainfall of 1376 mm. Its annual temperature ranges from 11.5 ^{o}C - 26.2 ^{o}C. The soil is fine sandy loam in texture with a pH of 6.40 (Hundesa and Mekonnen, 2017).

3.1.2. Plant material collection

A total of 150 *V. galamensis* accessions, representing 10 populations according to their geographic regions, were randomly collected from diverse agro-ecologies in three regional states of Ethiopia (Oromia, Amhara, and Southern Nations, Nationalities and Peoples Regional State (SNNPs). Most of the accessions were collected from the field (n = 90), and others were collected from the Ethiopia Biodiversity Institute (n = 45) and Wondo Genet Agricultural Research Center (n=15) (Figure 4) between November, 2016 to January, 2017.

At each collection areas, seed samples were taken from individual heads or seeds from plants with all matured flowers and kept in separate bags. The distances between consecutive sampling sites were about 5-10 Km. Then, accessions were cleaned and documented.

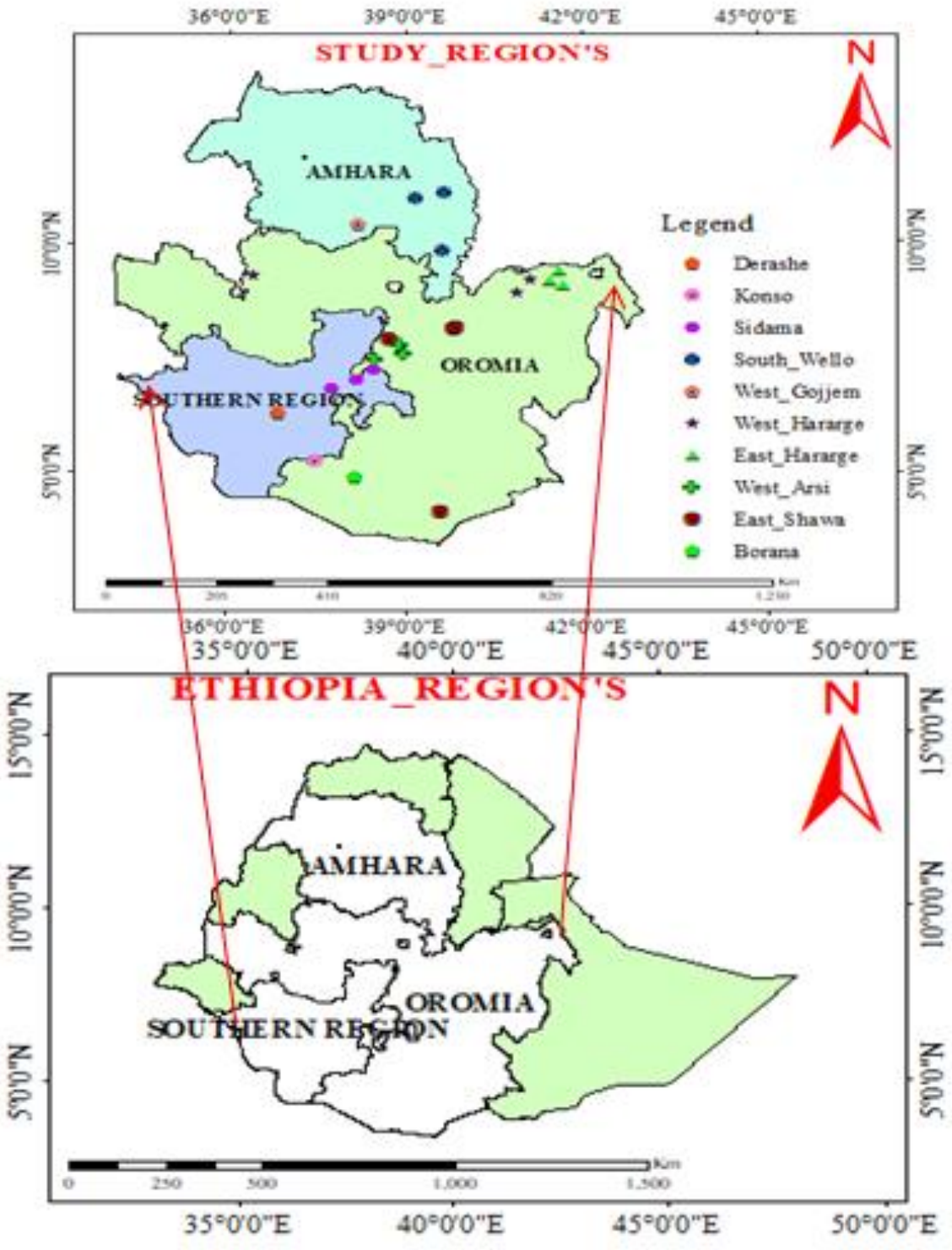

Figure 4. Sample collection sites in three regional states.

3.1.3. Experimental design and trial management

Experimental design of each site was layout in alpha-lattice design (8 x 10) with two replications. Each plot consisted of two rows of 2.5 meter size with spacing of 0.6 meter between plants and 0.6 meter between rows with 1.5 meter inter-block spacing, giving a total area of 9 m^2 (Shimelis and Hugo, 2011). To study various quantitative parameters, five individual plants in the middle of the plot were tagged to avoid bordering effect (Table 2). Other cultural practices and weeding were done whenever needed. Harvesting was started when 90% of the flower head was changed to dark/brown color (Baye *et al*., 2005).

Table 1. *Vernonia galamensis* accessions grown for evaluation of agro-morphological traits at Melka Werer and Wondo Genet agricultural research centers

No	Accesssion	Region	Woreda	Altitude	No	Accession	Region	Locality	altitude
1	VgEt-001B	Oromia	Yabelo	1092m	41	VgEt-076	Oromia	Meta	2174m
2	VgEt-002B	Oromia	Yabelo	1772m	42	VgEt-078	Oromia	Badessa	1626m
3	VgEt-003B	Oromia	Yabelo	1092m	43	VgEt-079	Oromia	Badessa	1822m
4	VgEt-004B	Oromia	Yabelo	1092m	44	VgEt-080	Oromia	Chiro Zuria	1889m
5	Vget-008B	Oromia	Yabelo	1774m	45	VgEt-082	Oromia	Chiro Zuria	1889m
6	VgEt-010B	Oromia	Yabelo	1770m	46	VgEt-085	Oromia	Mieso	1393m
7	VgEt-012B	Oromia	Yabelo	1778m	47	VgEt-088	Oromia	Mieso	1524m
8	VgEt-016W	SNNP	Aleta Wondo	1780m	48	VgEt-090	Oromia	Mieso	1393m
9	VgEt-017W	SNNP	Aleta Wondo	1780m	49	VgEt-094B	Amhara	Enemay	2560m
10	VgEt-019W	SNNP	Aleta Wondo	1775m	50	VgEt-095B	Amhara	Enemay	1700m
11	VgEt-020W	SNNP	Aleta Wondo	1778m	51	VgEt-096B	Amhara	Enemay	1876m
12	VgEt-022W	SNNP	Aleta Wondo	1782m	52	VgEt-098B	Amhara	Enemay	2560m
13	VgEt-024W	SNNP	Aleta Wondo	1790m	53	VgEt-100B	Amhara	Enemay	2630m
14	VgEt-025W	SNNP	Awassa Zuria	1708m	54	VgEt-102B	Amhara	Enemay	2400m
15	VgEt-027W	SNNP	Awassa Zuria	1800m	55	VgEt-103B	Amhara	Enemay	1890m
16	VgEt-030W	SNNP	Awassa Zuria	1795m	56	VgEt-105B	Amhara	Debra Sina	2600m
17	VgEt-032	Oromia	A/Tullu	1643m	57	VgEt-107B	Amhara	Debra Sina	2630m
18	VgEt-033	Oromia	A/Tullu	1643m	58	VgEt-110B	Amhara	Debra Sina	1570m
19	VgEt-035	Oromia	Dugda	1636m	59	VgEt-111B	Amhara	Legambo	1866m
20	VgEt-039	Oromia	Bora	1753m	60	VgEt-113B	Amhara	Legambo	1890m
21	VgEt-040	Oromia	Bora	1670m	61	VgEt-115B	Amhara	Legambo	1976m
22	VgEt-042	Oromia	Bora	1630m	62	VgEt-117B	Amhara	Dessie Zuria	2015m
23	VgEt-043	Oromia	Bora	1630m	63	VgEt-118B	Amhara	Dessie Zuria	2550m
24	VgEt-045	Oromia	Bora	1650m	64	VgEt-120B	Amhara	Dessie Zuria	1980m

25	VgEt-048	Oromia	Sheshemene	2044m	65	VgEt-122	SNNP	Konso	1650m
26	VgEt-049	Oromia	Sheshemene	1904m	66	VgEt-124	SNNP	Konso	1650m
27	VgEt-050	Oromia	Sheshemene	2004m	67	VgEt-126	SNNP	Konso	1750m
28	VgEt-053	Oromia	Arsi Negele	2143m	68	VgEt-130	SNNP	Konso	1562m
29	VgEt-055	Oromia	Arsi Negele	2020m	69	VgEt-131	SNNP	Konso	1830m
30	VgEt-056	Oromia	Arsi Negele	2015m	70	VgEt-132	SNNP	Konso	1780m
31	VgEt-058	Oromia	Shala	2000m	71	VgEt-134	SNNP	Konso	1890m
32	VgEt-060	Oromia	Shala	2040m	72	VgEt-135	SNNP	Konso	1650m
33	VgEt-061	Oromia	Bedeno	2044m	73	VgEt-138	SNNP	Derashie	1395m
34	VgEt-062	Oromia	Bedeno	2044m	74	VgEt-140	SNNP	Derashie	1390m
35	VgEt-063	Oromia	Bedeno	2050m	75	VgEt-141	SNNP	Derashie	1540m
36	VgEt-064	Oromia	Bedeno	1987m	76	VgEt-143	SNNP	Derashie	1395m
37	VgEt-069	Oromia	Malkabalo	2750m	77	VgEt-145	SNNP	Derashie	1450m
38	VgEt-071	Oromia	Meta	1574m	78	VgEt-147	SNNP	Derashie	1395m
39	VgEt-073	Oromia	Meta	1574m	79	VgEt-148	SNNP	Derashie	1835m
40	VgEt-075	Oromia	Meta	1673m	80	VgEt-150	SNNP	Derashie	1550m

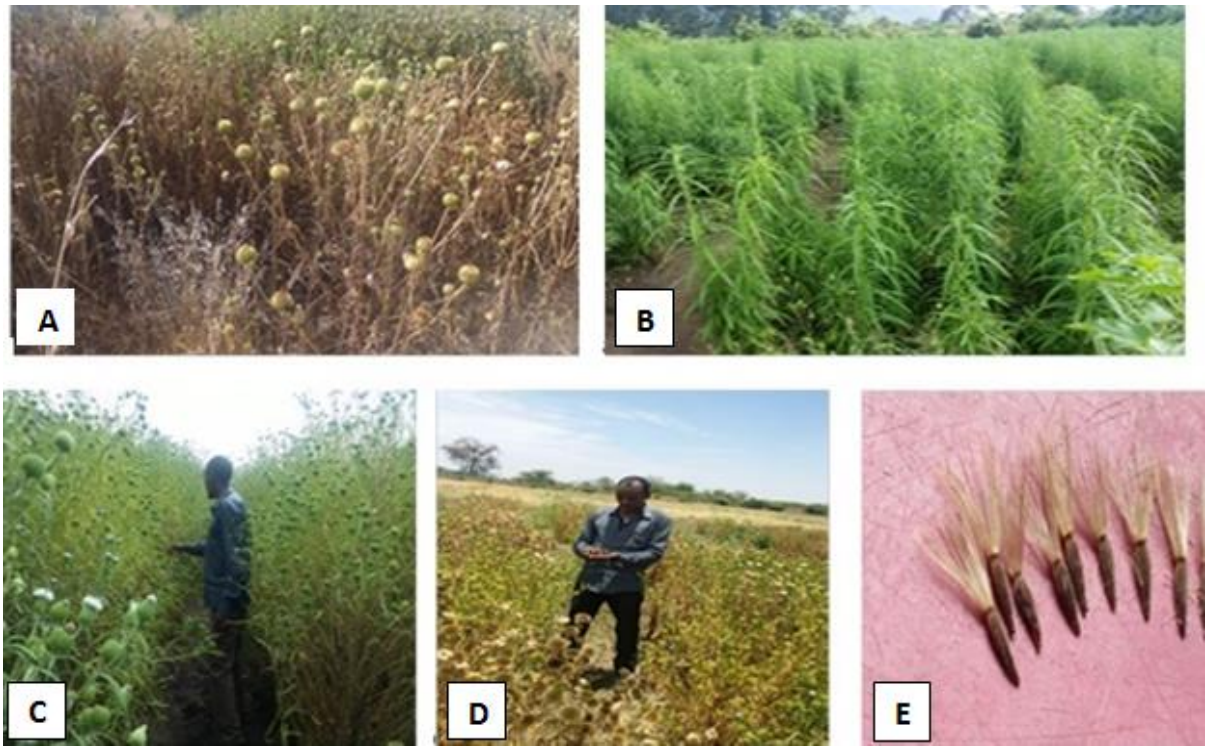

Figure 5. *Vernonia galamensis* growth stages: A. *Vernonia galamensis* plant growing Wild, B. Vegetative stage at Wondo Genet Agricultural Research Center, C. 50% flowering stage at Wondo Genet Agricultural Research Center, D. 90 % maturity stage at Melka Werer Agricultural Research center, E. *Vernonia galamensis* seed with hairy pappi at the tip.

3.1.4. Morphological traits scored

The descriptors used by Baye *et al*. (2001) and Shimelis *et al*., (2008) for *V. galamensis* were used in the present study. Phenological, agro-morphological and seed and seed related component measurements were done on five plants basis/plot and on plot basis as shown in Table 2.

Table 2. Morphological traits scored and method of scoring

Traits	**Description**	**Scores**
Days to emergence	Days from sowing to emergence (germination) of the seedlings above the ground level	Plot based
Days to 50% heading	Number of days from emergence to 50% of the plants showing heads	Plot based
Days to 50% flowering	Number of days from emergence to 50% of the plants showing flowers	Plot based
Days to 90% maturity	Number of days from date of emergence to the date when the plants in each plot attained 90% physiological maturity	Plot based
Plant height (cm)	The distance of the height of main stem from the ground to the tip of the flowering	5 plants/plot/rep/location
Branch length (cm)	The length of the branches from the base to the tip of the branch	5 plants/plot/rep/location
Branch number	Average number of branches formed per plant	5 plants/plot/rep/location
Number of heads per plant	Total number of heads per plant at the time of harvesting	5 plants/plot/rep/location
Number of seeds per head	Average number of seeds per head harvested from 5 plants	5 plants/plot/rep/location
Seed yield per plant (g)	Average seed weight of 5 plants	5 plants/plot/rep/location
Seed weight per plot (g)	The total weight of seeds per plot	Plot based
Seed yield per hectare (Kg/h)	Seed yield was determined after harvesting the two middle rows of each plot and later converted to kg/ha	Plot based
1000-seed weight (g)	Weight (g) of randomly selected 1000-seeds after harvesting	Plot based

3.2. Genetic diversity study using SSR molecular markers

3.2.1. Leaf sample collection and DNA extraction

Fresh young leaves of 150 *V. galamensis* accessions were collected from individual plants (3 individual plants per accession) of 80 accessions grown at experimental sites to evaluate the agro-morphological traits. The remaining 70 leaf samples were collected from accessions that were grown at Melka Werer Agricultural Research Center for the purpose of molecular analysis. The 150 accessions were grouped into 10 populations based on administrative zones. A collected individual plant was put in a sealed bag envelope and dried with silica gel, then kept at room temperature until used for DNA extraction following Gilbert *et al*. (1999). The dried leaf samples were transported to Huazhong Agricultural University, China, for genotyping. The genomic DNA extraction was made according to the modified CTAB protocol of Doye and Doye (1990). The extracted DNA was visualized on a 1% (w/v) agarose gel and quantified spectrophotometrically using a Nanodrop® 2000 (Thermo Scientific, USA), and stored at -20 °C for further use.

3.2.2. Primer screening and optimization

Simple sequence repeats (SSR) markers developed from *V. galamensis* by Narina *et al*. (2012) are available in the database (gene bank). Among these, 20 SSR markers were selected, and used for this study based on their high polymorphism (Table 3). Optimization was done by a sequential investigation of each reaction variable, testing different cycling conditions by varying the amount of DNA template, the concentration of primers, and the concentration of Taq PCR master mix at different combination.

Table 3. Primer sequences, annealing temperature and product size that was used for genetic diversity studies of *Vernonia galamensis* accessions (Narina *et al*., 2012)

No	SSR primers	SSR motifs (5'-3')	Forward sequence (5'-3')	Reverse sequence (5'-3')	Annealing temp (o^C)	Product size
1	Vg-001	(AT)12	CTTGATTTTGTGGGGACCTAAGTG	TAGGAATGGAATAGAATGGATCGG	56	135
2	Vg-002	(TC)12	GGGTTGTGGGGAGAGATAGAGATA	AGCCAAGTTACGCATAGACATCTG	56	146
3	Vg-003	(TC)24	GTGAGCGGGGATCTTCACTTC	GAGAAAGCGAGCATCAACAGACTT	56	142
4	Vg-004	(CCA)12	ACCATACAGTCCCGCATGAATATC	GCTCCTGGAAATGGAGGATAGAAT	58	145
5	Vg-005	(AAG)12	AGCTTAAACAAGAACAAACCGCTG	TGCGAAGGCTTACCAGTTACAAAC	56	157
6	Vg-006	(ACA)12	ATCAGCGTTGCTTGAAAAGAGTG	AAACTCATCGCTCAAACTCAAACG	58	101
7	Vg-007	(TCT)18	ATGACGATGCAACTCACCGTT	CGGAGAGGTTTGGTAGGGTAAGAT	56	136
8	Vg-008	(GAT)24	ATGTCTTCCAAATCGAGGATGGTA	AATTTTTGCAGCTAAGCCAGTGAG	58	101
9	Vg-009	(GGA)15	ATTGAAGGATGAACGGACAGAGTC	GTATCACATCACGTCGTCCACATC	58	127
10	Vg-010	(ATG)15	CAAAGGGAAGATGCACCTAGAGAA	ATCAAACCTGCTGCTTTTCAAGTC	58	130
11	Vg-011	(TGA)12	GCACAATCAGACTTGAGACCAAGA	GCAGTGATCAGCCATAGTGCATAC	58	130
12	Vg-012	(CGC)12	GGGCTGAGCAAATACAGCAGAC	AGGATCTTCTTGTTGGTGTGGAAA	58	150
13	Vg-013	(CAG)15	GGGGCGTTTCCTTGATTTTG	CTCTTCACCTGCCATCTTTTCTGT	56	93
14	Vg-014	(CAA)12	GTAGCAGCAGCAGTTCACTACCAC	CAAAATCCTCACAAACTTCACACG	56	132
15	Vg-015	(GGC)18	GTGCTACAACGGTGGTACATCAAG	TCATTGATTCCATGCTGAAATAGC	56	159
16	Vg-016	(GGT)12	GTTAGAGATGGGTTTGAAGAGCGA	CCTTACCAACTCCAACACCACTTG	58	139
17	Vg-019	(CTTCAC)24	GGGTCTCCATCTATTCACCTTCAA	AAGGAGCGTGAGCTAGAAGAAGC	56	158
18	Vg-021	(GTC)15	TGAAGAAGAAGGTTTCCCAAATCA	GATGCATTGACATCACGTAGAAGC	56	155
19	Vg-024	(ATC)15	TTGGATGTCGAAAAAGATGAGGTT	TTCTCCCTCTGTTTCAACACCTTC	56	144
20	Vg-030	(CT)12	TCAAACACACTCCCCAATTTCTCT	GCTGCCGATTGATCAAATTACACT	56	100

3.2.3. Polymerase chain reaction (PCR) amplification

A total of 20 SSR markers were used for this study (Table 3). The amplification reaction was performed with T100^{TM} Thermal Cycler (USA) in a total volume of 10 µl reaction mixture, containing 5 µl 2 x Taq PCR master mix (Vazyme P213-01, China) and 1 µl of forward and reverse primers, 100ng/µl of template DNA and 3.0 µl of double distilled water. The PCR amplification was programmed at an initial denaturation step of 5 minutes at 94°C followed by 35 cycles of 30 s denaturation at 94°C, annealing at 56/58°C (depending on primers) for 30 s, initial extension at 72^{o}C for 1 minute and final extension at 72°C for 5 minutes. The amplified DNA samples were stored at 4°C until it was loaded on the agarose gel for electrophoresis, then the amplified PCR products were separated by electrophoresis using 3% agarose gel.

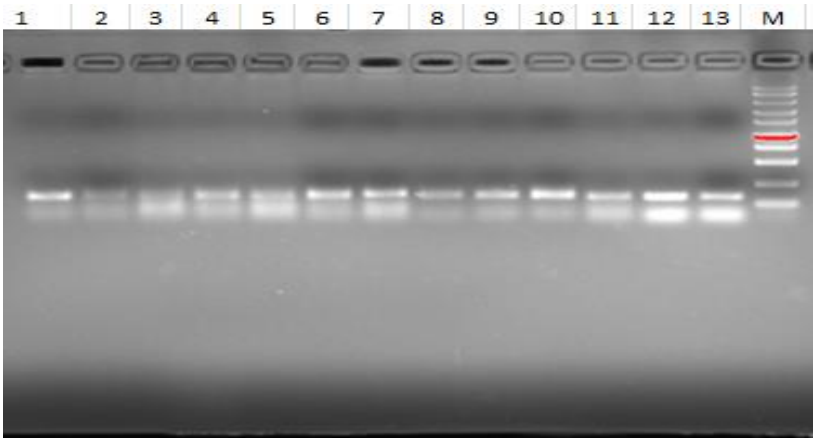

Figure 6. SSR assays for DNA samples isolated by CTAB methods. M = molecular marker

3.2.4. Band scoring and analysis

The gel band was photographed under UV illumination (Gel Doc^{TM} with Image Lab^{TM} software, BIO-RAD), in comparison of the amplified bands with the standard size of molecular marker (100 bp DNA ladder) (Figure 6). The amplified products were visually scored as molecular data matrix based on their migration from the gel photographs. Differently sized amplified bands were scored as different genotypes, and the bands were recorded as (11, 22, 33. . .) to represent homozygous genotypes or (12, 13, 23. . .), to indicate the heterozygous genotypes. For each marker, '?' was used for missing data.

3.3. Determination of oil content and fatty acid composition

3.3.1. Seed materials used for oil content and fatty acid composition

The seed samples used for oil content and fatty acids compositions analysis were selected from accessions grown at experiment sites, Melka Werer and Wondo Genet Agricultural Research Centers. In this study, 15 accessions were used from each experimental sites based on their 1000-seed weight performances. Accordingly, five seed samples were selected from each (low, medium and high) mean value of the accessions (Table 4).

Table 4. List of *Vernonia galamensis* accessions used for seed oil content and fatty acid composition analysis that grown at both Melka Werer and Wondo Genet agricultural research centers

No	Accessions	Collection woreda	Collection region
1	VgEt-017	Aleta Wondo	SNNP
2	VgEt-020	Aleta Wondo	SNNP
3	VgEt-027	Awassa Zuria	SNNP
4	VgEt-032	Adami Tullu	East Showa
5	VgEt-042	Bora	East Showa
6	VgEt-076	Meta	East Harerghe
7	VgEt-078	Badessa	West Harerghe
8	VgEt-080	Chiro Zuria	West Harerghe
9	VgEt-088	Mieso	West Harerghe
10	VgEt-090	Mieso	West Harerghe
11	VgEt-094	Enemay	West Gojjam
12	VgEt-111	Lagambo	South Wollo
13	VgEt-118	Dassie Zuria	South Wollo
14	VgEt-130	Konso	SNNP
15	VgEt-134	Konso	SNNP

3.3.2. Extraction of seed oil content

The seed samples were transported to Huazhong Agricultural University, China for seed oil content and fatty acids composition analysis. The oil was extracted according to the method of Folch *et al*. (1957) and Xin *et al.* (2013). The seeds were crushed in a crushing mill and reduced to 1.0 mm to 2.0 mm particle size. The crushed seeds and n-hexane solvent at a ratio of 1:5 (w/v) were fed into Soxhlet apparatus. The solvent and crushed mixtures were mixed for 5 minutes at 65^{o}C. After extraction, the mixture was allowed to settle and separated by a vacuum filtration. The solvent and oil in the filtrate were separated using a rotary evaporator at a temperature of 70^{o}C. The solvent was recovered using condenser. The oil was collected and stored to be filtered. Total extractable fat was determined gravimetrically and was expressed as percent fat (v/w) per 100g sample (Bhardwaj*iet al*., 2007).

3.3.3. Determination of fatty acid composition

Fatty acid composition of *V.galamensis* seeds was determined by Gas Chromatography with a Mass Spectrophotometry (GC-MS). Fatty acid methyl esters (FAME) were prepared using KOH /methanol reagent. The GC was fitted with an HP-5 column (30 m×0.25 mm×0.25 μm, Thermo Scientific, Bellefonte, PA, USA). The GC-MS conditions were modified according to Xin *et al.* (2013). The extracted oil was mixed with 2 ml hexane and converted into fatty acid methyl esters (FAMEs) by using 0.4M KOH in methanol. The oil samples (5 mg) were vortexed with 2 ml sulfuric acid/methanol (1:99 v/v) in 10 ml glass vials. The open vials were placed in a heating block at 90 ^{o}C until the sample volume was reduced to 0.5 ml. After cooling to room temperature, 1 ml hexane followed by 1 ml distilled water was added. The mixture was vortexed and the supernatant containing the FAME was taken and dried over anhydrous Na_2SO4 and transferred to a suitable vial for gas chromatographic analysis.

The GC column was operated under the conditions that the injector was maintained at 250 °C, with a transfer line temperature of 280 °C. The ion energy of electron impact ionization

was 70eV and the scanning range was 45-450 Da, with the ion source temperature set to 230℃. The flow rate of the helium (99.999%) carrier gas was 1 ml/min. Analytes absorbed on the fiber were desorbed for 3 min in the GC injector at 250 ℃ in split mode. The temperature isothermal was set at 40 ℃ for 3 min, then increased by 3 ℃ every minute until it reaches 73 ℃. The temperature was finally raised to 220 ℃ at the rate of 5 ℃/min, and held for 1 min. Peaks of fatty acids were identified using reference library. The area percent of each fatty acid was calculated by dividing its peak area by the total peak area of the fatty acids identified (Xu *et al.*, 2013).

3.4. Statistical analysis

The morphological quantitative traits and oil content among different accession, location and replication were computed using analysis of variance (one-way of ANOVA) after checking the normality of the data with the help of t-test. Before computing the combined analysis (ANOVA), error variance homogeneity test was performed using Hartley`s test (F-max test) (Gomez and Gomez, 1984).

3.4.1. Statistical Analysis for quantitative traits

The statistical analyses done include: descriptive statistics, the analysis of variance (ANOVA), computation of correlation coefficients, estimation of components of variances, principal component analysis, and clustering using Graph Pad Prism version 7.0 (Graph Pad Software, Inc. California, San Diego) and SAS version 9.0 (PROC GLM procedure) software.

The descriptive statistics used to compare accessions by their mean, range, phenotypic and genotypic variances and coefficient of variation. Variance components (genotypic [V_g], phenotypic [V_p] and environment [V_e] variances) were estimated using the formula of Prasad *et al*. (1981) as follows:

$V_g = [MSG – MSE / r]$

$V_p = [MSG / r]$

$V_e = [MSE / r]$, Where MSG, MSE and r are the mean square of genotype, the mean square of environment and number of replications, respectively.

Genotypic coefficient of variation (GCV), phenotypic coefficient of variation (PCV) and environment coefficient of variation (ECV) were calculated according to the methods of Burton (1952), Johnson *et al*. (1955) and Kumar *et al*. (1985) as:

$$GCV = \sqrt{(Vg / \overline{X})} x100$$

$$PCV = \sqrt{(Vp / \overline{X})} x100$$

ECV = $\sqrt{(Ve/\overline{X})}x100$, where Vg, Vp, Ve and $\overline{X}$ are the genotypic variance, phenotypic variance, environment variance and grand mean for each measured characters, respectively, for the character under consideration.

Broad sense heritability (h^2B) was estimated as the ratio of genotypic variance (Vg) to the phenotypic (Vp) variance as described by Allard (1991).

$$h^2B = \frac{Vg}{Vp}x100$$

Genetic advance (GA): Expected genetic advance, assuming selection of the superior 5% of the genotypes was estimated in accordance with the methods illustrated by Fehr (1987), using the formula,

$$GA = K\ (\sigma P)\ h^2B$$

where, K is the standardized selection differential at 5% (K = 2.063),
σP is the phenotypic standard deviation
h^2B is the broad sense heritability.

Genetic advance as % of the mean = $\frac{GA}{\overline{X}}x100$, where $\overline{X}$ refers to the grand mean for each measured traits.

The genotypic and phenotypic correlation of coefficients for all possible combinations were calculated using the formula suggested by Miller *et al*. (1958) and Johnson *et al*. (1955) as:

$$rp(XY) = \frac{P\,\mathrm{cov}(X,Y)}{\sqrt{\delta^2 px..\delta^2 py}}$$

Where r_p = phenotype correlation of coefficient, pcov (x, y) = phenotypic co-variance between variable x and y, σ^2px = phenotypic variance for variable x, σ^2py = phenotypic variance for variable y.

$$rg(XY) = \frac{G\,\mathrm{cov}(X,Y)}{\sqrt{\delta^2 gx..\delta^2 gy}}$$

Where r_g = genotype correlation of coefficient, Gcov (x.y) = genotypic co-variance between variable x and y, σ^2gx = genotypic variance for variable x, σ^2gy = genotypic variance for variable y.

Pearson's coefficient of correlation, regression coefficient analysis, cluster analysis and principal component analysis were performed using the correlation matrix between all pairs of quantitative characters and oil content using MINITAB® version 17.1 (Minitab, 2013) software.

3.4.2. Statistical analysis for molecular data

3.4.2.1. Population genetic diversity indices

The genetic diversity for each allele such as the number of different alleles (Na), the major allele frequency, the effective number of alleles (Ne), Shannon's diversity index (I), observed heterozygosity (Ho), expected heterozygosity (He), F-statistics values (Fis, Fit and Fst), polymorphic information content (PIC), Nei's genetic identities (Ji) and genetic distances (Ds), gene flow (Nm), were determined using GeneAlex version 6.503 software (Peakall and Smouse, 2012).

Mathematically:

- $Ne = \frac{1}{\sum pi^2}$, where *Ne* is number of effective alleles, Σpi^2 is the sum of squared population allele frequencies
- $He = 1 - \sum_{i=1}^{n} pi^2$, where *He* is the expected heterozygosity
- Polymorphic information content was calculated following (Anderson *et al.*, 1993) as

$$PIC = 1 - \sum_{i=1}^{n} pi$$, where *pi* is the frequency of the *i*th allele of each SSR marker

3.4.2.2. Analysis of molecular variance and population differentiation

The molecular variance analysis (AMOVA) of the populations, and population differentiation tests: Wrights, (1978) fixation indices (FIS, FST and FIT) and pairwise FST were computed using GenAIEx ver. 6.503 (Peakall and Smouse, 2012).

3.4.2.3. Population cluster analysis

Unweighted Pair Group Method with Arithmetic Mean (UPGMA) and Neighbor-Joining (NJ) tree were computed using DARwin version 6.0.19 software (Perrier and Jacquemoud-Collet, 2006). The resulting trees were displayed using Fig Tree version. 1.4.4 (Andrew,

2018). The patterns of variation among individual samples, a principal coordinated analysis (PCoA) was performed using GeneAlex version 6.503 software (Peakall and Smouse, 2012).

3.4.2.4. Analysis of population structure

The SSR markers data that subjected to a Bayesian model-based cluster analysis was performed using STRUCTURE version 2.3.4 software (Pritchard *et al.* 2000) to establish the number of hypothetical distinct populations, used to estimate allele frequencies in each of the *K* populations and the degree of admixture. To determine the most likely number of populations (K), a burn-in period of 50,000 was used in each run, and data were collected over 500,000 Markov Chain Monte Carlo (MCMC) replications for K = 1 to K = 10 using 20 iterations for each K. The optimum K value was determined according to Evanno *et al*. (2005) using the web-based (http://tyloro.biology.ucla.edu\structure Harvester\) STRUCTURE HARVESTER ver. 0.6.92 (Earl and vonHoldt, 2012). The results generated by this software were visualized in a graphical bar plot using Clumpak (Cluster Markov Packager Across K) (http://www.clumpak.tau.ac.il/) (Kopelman *et al*., 2015).

4. RESULTS

4.1. Quantitative traits-based genetic diversity analysis

4.1.1. Analysis of variation in vernonia plant

Mean squares of quantitative traits among different accession, location and replication were computed using one-way of ANOVA (Table 5).

The results of the analysis of variance for all characteristics revealed highly significant variation ($p < 0.05$) for most of the quantitative traits among accessions, except for days to emergence. The accession by location interaction was also highly significant for all characteristics except for days to emergence and thousand seed weight (Table 5).

Mean performance of days to 50% heading and plant height exhibited significant differences (Table 6). The earliest 50% heading was 63.50 days and the latest 50% heading was 73.0 days at Melka Werer Agricultural Research Center whereas the earliest 50% heading was 81.5 days and the latest 50% heading was 91.0 days at Wondo Genet Agricultural Research Center. The maximum plant height was 143.5 cm and 228.0 cm at Melka Werer and Wondo Genet, respectively.

The mean performance of seed yield ranged from 348.98 Kg ha^{-1} (VgEt-027) to 624.27 Kg ha^{-1} (VgEt-085), with an average of 474.39 Kg ha^{-1}.The seed weight per plot was in the range of 100.15 g/ plot to 187.43 g/ plot with an average of 141.63 g/ plot. Significant differences ($p < 0.05$) were observed among seed yield per hectare of the two study sites (Table 6). The highest seed yield per plant was observed in accession VgEt-140 (20.55 g) and the lowest was recorded for VgEt-063 (12.48 g), the overall average of 15.63 g. The highest 1000-seed weight was recorded for VgEt-094 (3.51 g) and the lowest 1000-seed weight was scored in VgEt-105 (2.77 g) with an average of 3.04 g (Appendix 2).

Table 5. Mean squares for 13 quantitative traits of 80 *Vernonia galamensis* accessions evaluated at Melka Werer and Wondo Genet agricultural research centers

Source of variation	Mean square and significant tests							
	DF	DE	DFH	PH	DFF	NB	BL	DNM
Accession	79	1.53	7.44	352.297**	8.76*	154.18	86.07*	152.49
Location	1	0.03	20704.61*	434992.26*	14137.90*	943.77	16.707*	20464.00*
Block (Rep)	7	1.25	10.86**	702.406**	2.16	214.02	20.22	90.189
Rep	1	1.25	42.05	6200.481*	21.53*	520.84	12.21**	20464.00*
Accession*Location	79	0.74	9.57**	391.120**	11.90*	52.16**	86.14**	33.93**
Error	152	0.83	5.84	221.309	1.08	86.25	38.30*	154.92
CV (%)		8.51	3.17	9.49	1.076	26.44	11.05	7.25

Table 5 continued

Source of variation	Mean square and significant tests						
	DF	NHP	SPH	SYP	TSW	SWP	SYH
Accession	79	536.14*	312.05	11.34*	0.075	1117.34*	12944.2*
Location	1	395606.26*	133501.89*	4130.73*	0.48*	24888.971*	2496910.68*
Block (Rep)	7	396.66*	194.96	10.8	0.06	2254.77	17208.1
Rep	1	854.78	599.68*	5.77	0.31*	8439.06*	67753.82*
Accession*Location	79	1310.478*	357.46*	13.90*	0.07	2269.08*	20889.64*
Error	152	146.5389	197.69	3.72	0.04	41.58	20.42
CV (%)		7.87	9.87	12.33	6.53	29.36	30.16

* Significant at $P < 0.05$. DE = days to emergence, DFH = days to 50% heading, PH = plant height, DFF = days to 50% flowering, NB = number of Branches, BL = branch length, DNM = days to 90% maturity, NHP = heads per plant, SPH = seeds per head, SYP = seed yield per plant, SWP = seed Weight per plot, SYH = seed yield per hectare

The mean performances of accessions for most quantitative traits including seed yield and yield grown at Wondo Genet Agricultural Research Center (WGARC) were higher than that of Melka Werer Agricultural Research Center (MWARC) (Table 6). The maximum length of time taken for 50% flowering was 93.50 days at Melka Werer and 111.50 days at Wondo Genet Agricultural Research Center (WGARC).

Furthermore, the maximum number of branches was 52.4 and 59.5 at Melka Werer and Wondo Genet, respectively. Similarly, the seed yield per hectare was the highest (1104.0 Kg ha^{-1}) at Wondo Genet and 822. 7 Kg ha^{-1} at Melka Werer with respective average of 562.2 Kg ha^{-1} and 385.5 Kg ha^{-1} (Table 6).

Table 6. Estimates of mean performance and ranges for 13 quantitative traits of *Vernonia galamensis* grown at Melka Werer (MWARC) and Wondo Genet (WGARC) agricultural research centers

Quantitative traits	Mean performance					
	Melka Werer			Wondo Genet		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Days to emergence	10.74±0.1	9.00	12.50	10.72±0.1	9.50	13.00
Days to 50% heading	68.13±0.3	63.50	73.00	84.22±0.3	81.50	91.00
Days to 50% flowering	89.82±1.3	85.00	93.50	103.11±1.7	96.00	111.50
Plant height (cm)	119.88±0.3	89.50	143.50	193.62±0.4	159.50	228.00
Number of branches	33.41±0.9	17.80	52.40	36.85±0.9	19.50	59.50
Branch length (cm)	56.26±0.8	46.00	71.90	72.81±0.8	43.00	76.00
Days to 90% maturity	163.65±0.9	151.00	177.00	179.64±0.8	165.50	198.50
Number of head/plant	118.73±2.9	84.90	188.40	189.06±2.2	132.40	239.20
Number of seed/head	121.99±1.8	93.10	163.50	162.84±1.2	141.20	189.00
Seed yield /plant (g)	12.04±0.2	8.70	16.80	19.22±0.3	14.90	25.70
Seed weight/plot (g)	113.74±39.5	58.60	246.80	169.51±44.3	116.80	331.20
1000-seed weight (g)	3.00.00±0.0	2.70	3.40	3.08±0.0	2.40	3.50
Seed yield (Kg/ha)	385.49±9.4	246.80	822.70	562.16±11.5	406.00	1104.00

Significant at $P < 0.05$

The mean performances of the combined data of accessions for most quantitative traits showed that the maximum days taken for 50% flowering were 87.92 while the minimum days 73.2 days (Table 7). Further, the minimum number of head per plant was 127.7 whereas the maximum was 188.7.

Table 7. Mean performance and ranges for 13 quantitative traits of *Vernonia galamensis* over two locations, Melka Werer and Wondo Genet agricultural research centers

Quantitative traits	Mean performance of data		
	Mean	Minimum	Maximum
Days to emergence	10.73	9.5	12.5
Days to 50% heading	76.18	72.75	80.5
Days to 50% flowering	81.22	73.2	87.92
Plant height (cm)	156.75	135.85	176.1
Number of branches	35.13	18.65	55.75
Branch length (cm)	56.04	46.03	69.08
Days to 90% maturity	171.65	158.5	185.5
Number of head/plant	153.89	127.7	188.65
Number of seed/head	142.42	122.94	164.81
Seed yield /plant (g)	15.63	12.48	20.55
Seed weight/plot (g)	141.63	100.15	187.43
1000-seed weight (g)	3.04	2.77	3.51
Seed yield (Kg/ha)	474.39	348.98	624.27

Significant at $P < 0.05$

4.1.2. Variance components

Phenotypic variance (Vp), genotypic variance (Vg), environmental variance (Ve), genotypic coefficient of variance (GCV), phenotypic coefficient of variance (PCV), environmental coefficient of variance (ECV), heritability (h^2B) and genetic advance as percentage of mean (GAM) were shown in Table 8. In the present study, the phenotypic variance (Vp) and phenotypic coefficient of variance (PCV) were higher than theirs corresponding genotypic variance (Vg) and genotypic coefficient of variance (GCV), respectively for all the characters studies.

The highest GCV was observed for seed weight per plot (95.44) followed by branch length (80.46). Likewise, the highest PCV was recorded for branch length (99.48) followed by seed weight per plot (97.81). Likewise, the lowest PCV and GCV were recorded for thousand seed weight (11.51, 8.14, respectively). Moreover, the highest environmental coefficient of variation (ECV) was recorded for seed weight per plot (75.18) whereas the lowest was recorded for days to 50% flowering (1.57). The estimated values of the PCV for different agro-morphological traits were higher than the corresponding GCV value (Table 8).

4.1.3. Heritability and genetic advance

Heritability estimates the relative contributions of differences in genetic and non-genetic factors to the total phenotypic variance in a population. In the present study, the magnitude of heritability for different traits was measured and it ranged from 11.69 to 95.22% (Table 8). The characters, seed weight per plot (95.22%), days to 50% flowering (94.78%) and number of heads per plant (87.72%) had high heritability, whereas days to 90% maturity exhibited the lowest heritability (11.69%). Values of genetic advance as a percentage of mean ranged from 2.25% for days to 90% maturity to 70.13% for seed weight per plot. Relatively high heritability coupled with high genetic advance as percentage of mean was recorded for number of heads per plant (50.74%) and seed weight per plot (70.13%) (Table 8).

Table 8. Estimates of mean, phenotypic, genetic and environmental variation, genotypic, phenotypic and environmental coefficients of variation (GCV, PCV & ECV), broad sense heritability (h^2B), genetic advance (GA) and genetic advance as a percentage of the mean (GAM) for 13 quantitative traits

Quantitative traits	Mean	Vp	Vg	Ve	PCV%	GCV%	ECV%	h^2B	GA	GAM
DE	10.73	0.84	0.42	0.42	27.96	19.83	3.92	50.3	1.05	9.83
DFH	76.18	5.68	2.76	2.92	27.31	19.03	3.83	48.59	8.58	11.26
DFF	96.47	10.54	9.99	0.55	33.05	32.18	1.57	94.78	14.25	14.77
PH	156.75	90.25	69.85	20.4	75.88	66.75	13.01	77.39	65.64	41.87
NB	35.13	12.32	8.08	4.24	59.22	47.96	12.07	65.58	15.09	42.94
BL	64.54	55.46	36.28	17.11	99.48	80.46	26.51	65.42	11.09	19.79
DNM	171.65	87.71	10.25	77.46	71.48	24.44	45.13	11.69	3.86	2.25
NHP	153.89	108.07	94.8	13.27	83.78	78.49	8.62	87.72	78.08	50.74
SPH	142.42	106.02	57.17	98.85	86.28	63.31	64.41	53.93	29.5	20.71
SWP	141.63	135.48	129.01	106.47	97.81	95.44	75.18	95.22	99.32	70.13
SYP	15.62	5.75	3.89	1.86	60.65	49.89	11.91	67.65	6.44	41.17
SYH	473.83	313.38	227.15	86.23	81.33	69.24	18.2	72.48	95.9	20.24
TSW	3.03	0.04	0.02	0.02	11.51	8.14	2.67	50.00	0.25	8.34

DE = days to emergence, DFH = days to 50 % heading, PH = plant height, DFF = days to 50 % flowering, NB = number of branches, BL = branch length, DNM = days to 90 % maturity, NHP = number of heads per plant, SPH = number of seeds per head, SYP = seed yield per plant, SWP = seed weight per plot, SYH = seed yield per hectare, TSW = 1000-seed weight, Vg = genotypic variation, Vp = phenotypic variation, Ve = environmental variation, GCV = genotypic coefficient of variation, PCV = phenotypic coefficient of variation, GA = genetic advance, GAM = genetic advance as percentage of mean at 5% selection intensity.

4.1.4. Correlation coefficient analysis in *Vernonia galamensis*

4.1.4.1. Correlation between traits

Estimates of Pearson correlation of coefficients for 13 quantitative traits for data from both locations are presented in Table 9. The result shows that days to 50% heading was positively correlated with days to 50% flowering (r = 0.66) but negatively correlated with number of head per plant (r =-0.34) at Wondo Genet agricultural research center while plant height had positive correlation with days to 50% flowering (r = 0.89) and branch length (r = 0.54) at Melka Werer. Further, days to 50% heading had high and positive correlation with plant height (r = 0.87) at Melka Werer, whereas days to 50% flowering had high negative correlation with number of head per plant (r = -0.45) at Wondo Genet. Moreover, seed yield per hectare had high significant and positive association with seed weight per plot (r = 0.88, r = 0.98) at Melka Werer and Wondo Genet, respectively (Table 9).

Table 9. Estimates of Pearson correlation of coefficients for 13 quantitative traits of *Vernonia galamensis* at Melka Werer agricultural research center (below diagonal) and Wondo Genet agricultural research center (above diagonal)

	DE	DFH	PH	DFF	NB	BL	DNM	NHP	SPH	SYP	SWP	TSW	SYH
DE		0.12*	0.07	0.06	0.08	-0.02	0.16*	-0.18	-0.13	-0.09	-0.05	-0.04	-0.04
DFH	0.27*		0.17*	0.66*	0.11	0.27*	0.49*	-0.34*	0.09	-0.03	0.09	0.14	0.06
PH	0.19*	0.87*		0.14	0.18*	-0.07	0.22*	-0.06	-0.05	-0.06	0.13	0.03	0.12
DFF	0.54*	0.04	0.89*		0.11	-0.10	0.23*	-0.45*	-0.06	-0.12	-0.05	0.07	-0.09
NB	0.02	0.03	-0.16	-0.10		0.06	0.27*	-0.03	-0.17	0.01	0.04	0.01	0.05
BL	0.08	-0.12	0.54*	-0.30*	-0.09		-0.18	0.09	0.15	-0.07	0.01	0.07	0.01
DNM	-0.06	-0.04	0.04	0.09	0.16	0.07		-0.15	0.12	0.01	0.05	0.22*	0.46*
NHP	0.19	-0.02	-0.08	0.14	0.44*	0.04	0.12		0.09	0.09	0.06	0.05	0.87*
SPH	-0.02	-0.02	-0.01	-0.13	0.00	0.02	0.05	0.03		0.08	0.11	0.06	0.11
SYP	0.21*	0.01	-0.01	0.20	0.04	0.15	0.07	0.06	0.08		-0.16	-0.03	-0.15
SWP	0.07	-0.06	0.08	-0.04	0.12	0.16	0.05	0.06	0.09	0.16		0.03	0.98*
TSW	0.09	-0.05	0.06	0.09	0.02	-0.01	0.04	-0.02	-0.15	-0.08	0.06		0.02
SYH	0.07	-0.04	0.09	-0.01	0.02	0.16	-0.03	0.72*	0.10	0.21*	0.88*	0.05	

*, significant at $P < 0.05$. DFH = days to 50% heading, PH = plant height, DFF = days to 50% flowering, NB = number of branches, BL = branch length, DNM = days to 90% maturity, NHP = heads per plant, SPH = seeds per head, SYP = seed yield per plant, SWP = seed weight per plot, SYH = seed yield per hectare, TSW = 1000-seed weight.

4.1.4.2. Correlation coefficients at genetic (G) and phenotypic (P) levels

In this experiment, estimates of genotypic (r_g) and phenotypic (r_p) correlation coefficients between each pair of the studied traits are presented in Tables 10. The results of correlation coefficient indicated, days to 50% heading showed high, significant ($p<0.05$) and positive correlations at both genotypic and phenotypic levels with plant height (r = 0.71G, 0.62P), days to 50% flowering (r = 0.72G, 0.59P), number of heads per plant (r = 0.84G, 0.73P), number of seeds per head (r = 0.94G, 0.70P), seed yield per plant (r = 0.88G, 0.72P) and seed yield per hectare (r = 0.99G, 0.86P) (Table 10).

Table 10. Estimates of genotypic (above diagonal) and phenotypic (below diagonal) correlation of coefficients for 13 quantitative traits of *Vernonia galamensis* over two locations, Melka Werer and Wondo Genet agricultural research centers

	DE	DFH	PH	DFF	NB	BL	DNM	NHP	SPH	SYP	SWP	SYH	TSW
DE		0.20	0.21	0.04	0.31	0.07	0.50	0.14	-0.25	0.09	0.03	0.05	0.10
DFH	0.06		0.71*	0.72*	0.24*	-0.16	0.99**	0.84*	0.94**	0.88**	0.86**	0.99**	0.31*
PH	0.04	0.62*		0.95**	0.34*	0.06	0.93**	0.84**	0.91**	0.83**	0.95**	0.83*	0.06
DFF	0.02	0.59*	0.91**		0.30	-0.09	0.89**	0.78*	0.85**	0.83**	0.91**	0.99**	0.31*
NB	0.08	0.38*	0.22*	0.20		0.02	0.84**	0.47*	0.16	0.27*	0.47*	0.45*	0.13
BL	0.05	-0.20	0.05	0.07	0.01		0.26*	0.02	0.21*	0.02	0.13	0.13	0.07
DNM	0.10	0.81**	0.66*	0.66*	0.42*	0.15		0.67*	0.84**	0.74*	0.91**	0.99**	0.59*
NHP	0.03	0.73*	0.79*	0.76*	0.31*	0.02	0.49*		0.83**	0.79*	0.92**	0.99**	0.25*
SPH	-0.07	0.70*	0.80**	0.77*	0.10	0.03	0.57*	0.74*		0.86**	0.96**	0.85**	0.17
SYP	0.04	0.72*	0.78*	0.78*	0.17	-0.01	0.52*	0.74*	0.74*		0.85**	0.97**	0.19
SWP	0.01	0.70*	0.70*	0.63*	0.22*	0.07	0.47*	0.62*	0.66*	0.55*		0.96**	0.40*
SYH	0.01	0.86**	0.71*	0.62*	0.19	0.07	0.45*	0.60*	0.67*	0.56*	0.66*		0.41*
TSW	0.07	0.37*	0.21*	0.22*	0.06	0.04	0.31*	0.17	0.11	0.13	0.20	0.18	

**, * Significant at $P < 0.05$, respectively. DFH = days to 50% heading, PH = plant height, DFF = days to 50% flowering, NB = number of branches, BL = branch length, DNM = days to 90% maturity, NHP = heads per plant, SPH = seeds per head, SYP = seed yield per plant, SWP = seed weight per plot, SYH = seed yield per hectare, TSW = 1000-seed weight.

Likewise, the number of heads per plant exhibited highly significant and positive genotypic and phenotypic correlation with days to 50% heading (r =0.84G, 0.73P), days to 50% flowering (r = 0.78G, 0.76P), number of seeds per head (r = 0.83G, 0.74P), seed yield per plant (r= 0.79G, 0.74P) and seed yield per hectare (r = 0.99G, 0.60P). Moreover, seed yield per hectare revealed significant and positive genotypic and phenotypic correlations with days to 50% heading (r = 0.99G, 0.86P), days to 50% flowering (r = 0.99G, 0.62P), number of seeds per head (r = 0.85G, 0.67P), seed yield per plant (r = 0.97G, 0.56P) and seed weight per plot (r = 0.96G, 0.66P). In addition, 1000-seed weight showed significant and positive genotypic correlation with seed yield per hectare (r = 0.41) while, 1000-seed weight showed significant and positive phenotypic correlation with days to 50% heading (r = 0.37), days to 50% flowering (r = 0.22) and days to 90% maturity (r = 0.31) (Table 10).

4.1.5. Regression coefficient analysis in *Vernonia galamensis*

The results of regression coefficient showed, among the traits affecting on the seed yield per hectare (dependent variable), number of branches (0.42), the number of head per plant (0.72), the number of seed per head (0.35), seed yield per plant (0.33) and seed weight per plot (0.20) as the independent variables (Table 11). Moreover, the coefficient of determination and standardized coefficient of determination had high (81.0%), and residual effect was 0.19 (Table 11). The residual effect was 0.19 and the contribution of component characters on seed yield per hectare was 81.0%, by traits studied in path coefficient analysis, the rest 19.3% was the contribution of other factors, such as the traits not studied.

Table 11. Regression coefficient analysis with direct effects, residual, coefficient of determination and standardized coefficient of determination in *V. galamensis*

Dependent variable	Independent variable	Direct effect	Residual	Coefficient of determination	Standardized coefficient of determination
SYH	NB	0.42	0.19	0.81	0.81
	NHP	0.72			
	SPH	0.35			
	SYP	0.33			
	SWP	0.20			

NB = number of branches, NHP = number of heads per plant, SPH = number of seeds per head, SYP = seed yield per plant, SWP = seed weight per plot, SYH = seed yield per hectare

Regression coefficient analysis as shown in (Table 12), the number of heads per plant had the highest positive direct effect (72%), on the seed yield per hectare. Also, seed weight per plot had the lowest rate of the direct effect (20%), with the seed yield per hectare. However, the trait of number of seeds per head through seed weight per plot had the highest indirect and positive effect (64%).

Table 12. Regression coefficient analysis with indirect effects and total correlation with depend traits

Variable	Direct effect	Indirect effect by					Total correlation with seed yield per hectare
		NB	NHP	SPH	SYP	SWP	
NB	0.42	*	0.39	0.23	0.15	0.03	0.49
NHP	0.72	0.36	*	0.65	0.68	0.37	0.86
SPH	0.35	0.21	0.56	*	0.46	0.72	0.47
SYP	0.33	0.06	0.49	0.51	*	0.63	0.50
SWP	0.20	0.03	0.26	0.39	0.37	*	0.90

NB = number of branches, NHP = number of heads per plant, SPH = number of seeds per head, SYP = seed yield per plant, SWP = seed weight per plot, SYH = seed yield per hectare

4.1.6. Hierarchical clustering of phenotypic traits

The Unweighted Pair Group Method with Arithmetic Mean (UPGMA) dendrogram clustered all the 80 *V. galamensis* accessions, based on 13 quantitative traits, in eight major groups having 2 to 21 accessions per cluster (Table 13; Figure 7). The number of member accessions varied from cluster to cluster. Cluster-I had the largest number of accessions (21 accessions), characterized by relatively high seed yield per plant (16.26 g), seed weight per plot (173.99 g) and seed yield per hectare (607.75 Kg ha^{-1}), and lower in number of days to 50% heading (75.50%) and days to emergence (10.25). Accessions in cluster-II (n = 15) had the lowest value of number of heads per plant (143.73), and had high value of thousand seed weight (3.05 g). Cluster-III, had sixteen accessions, characterized by the highest number of seed per head (145.28), and the lowest value in seed weight per plot (123.29 g), plant height (152.48 cm) and number of branches per plant (29.12) (Table 13 and 14).

Mean comparison between clusters showed that cluster-IV (n = 6) is characterized by the highest number of branches (41.68) and higher value of days to 90% maturity (176.46), while it had the lowest value in the number of seed per head (133.09). Cluster-V (n = 3 accessions) was mainly distinguished by having the lowest value in thousand seed weight (2.89 g) (Table 13).

Table 13. Distribution of 80 *Vernonia galamensis* accessions into eight clusters using a mean of 13 agro-morphological quantitative traits

Cluster	Number of accessions	Accessions in each cluster
C-I	21	VgEt-001, VgEt-049, VgEt-005, VgEt-064, VgEt-076, VgEt-079, VgEt-080,VgEt-007, VgEt-053, VgEt-008, VgEt-040, VgEt-034, VgEt-036, VgEt-037, VgEt-070, VgEt-072, VgEt-012, VgEt-024, VgEt-039, VgEt-060, VgEt-074
C-II	15	VgEt-009, VgEt-013, VgEt-016, VgEt-047, VgEt-050, VgEt-073, VgEt-071, VgEt-025, VgEt-068, VgEt-054, VgEt-061, VgEt-051, VgEt-056, VgEt-065, VgEt-077
C-III	16	VgEt-002, VgEt-045, VgEt-066, VgEt-067, VgEt-003, VgEt-019, VgEt-020, VgEt-038, VgEt-057, VgEt-004, VgEt-014, VgEt-031, VgEt-018, VgEt-027, VgEt-021, VgEt-029
C-IV	6	VgEt-030, VgEt-055, VgEt-044, VgEt-063, VgEt-052, VgEt-061
C-V	3	VgEt-015, VgEt-059, VgEt-017
C-VI	5	VgEt-006, VgEt-041, VgEt-011, VgEt-033, VgEt-048
C-VII	12	VgEt-010, VgEt-035, VgEt-058, VgEt-032, VgEt-043, VgEt-062, VgEt-075, VgEt-022, VgEt-028, VgEt-0023, VgEt-026, VgEt-078
C-VIII	2	VgEt-042, VgEt-046

Cluster-VI had five accessions, and characterized by the highest number of heads per plant (176.08) and number of seeds per head (153.79). Cluster-VII comprised of twelve accessions and characterized with the highest number of days to emergence (11.17), days to 50% heading (77.92), plant height (161.52 cm), days to 50% flowering (83.53) and thousand seed weight (3.11 g). Cluster-VIII contained only two accessions and characterized by the highest branch length (57.94) (Table 13 and 14).

Table 14. Mean values of 13 quantitative traits used in the clusters of 80 *Vernonia galamensis* accessions

	Cluster							
	I	II	III	IV	V	V-I	V-II	V-III
DE	10.25	10.94	10.41	10.67	10.84	10.38	11.17	10.79
DFH	75.50	76.17	75.72	76.08	76.17	76.13	77.92	76.44
PH	158.47	157.78	152.48	159.79	157.55	158.00	161.52	156.68
DFF	81.41	81.63	79.53	82.18	81.52	81.50	83.53	81.30
NB	36.75	32.55	29.12	41.68	36.34	36.09	32.20	39.62
BL	54.19	56.65	55.52	55.23	54.70	56.66	54.27	57.94
DNM	172.00	173.08	169.52	176.46	170.23	165.13	174.25	172.15
NHP	154.67	143.73	149.21	173.00	153.26	176.08	143.83	160.37
SPH	139.59	141.48	145.28	133.09	140.82	153.79	138.23	145.13
SYP	16.26	15.15	15.76	15.72	15.49	16.01	15.03	15.99
SWP	173.99	160.45	123.29	124.49	137.72	158.36	107.69	148.77
SYH	607.75	533.10	411.89	396.14	462.10	560.32	358.99	499.70
TSW	3.03	3.05	3.03	3.01	2.89	2.98	3.11	3.04

Keys: DE = days to emergence, DFH = days to 50% heading, PH = plant height, DFF = days to 50% flowering, NB = number of branches, BL = branch length, DNM = days to 90% maturity, NHP = heads per plant, SPH = seeds per head, SYP = seed yield per plant, SWP = seed weight per plot, SYH = seed yield per hectare, TSW = 1000-seed weight.

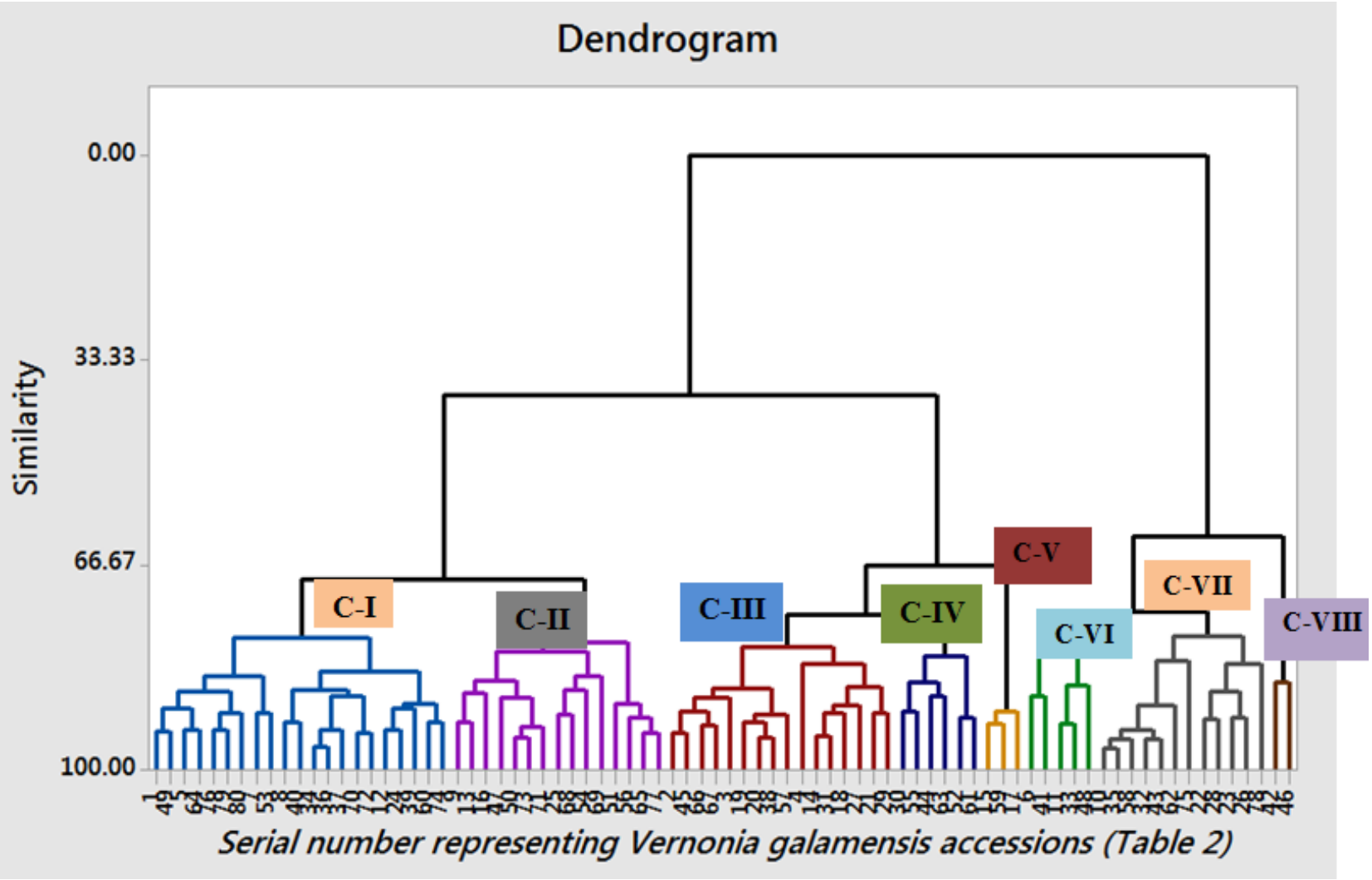

Figure 7. Genetic similarities among *Vernonia galamensis* accessions based on 13 major quantitative traits

4.1.7. Hierarchal clustering at regional level

According to regional based hierarchal clustering, all the accessions were grouped into five clusters, and are presented in (Table 15 and Figure 8). Accessions collected from West Arsi showed strong relations with those from Konso and Sidama, and grouped in the first cluster (C-I). This cluster was characterized by the highest mean number of heads per plant (169.26), number of seeds per head (151.35) and thousand seed weight (3.06 g), but the lowest in days to 50% heading (73.94), days to 50% flowering (78.04) and days to 90% maturity (169.41) (Table 15). The second cluster (C-II) comprised of accessions collected from East Showa and this group was distinguished by having the highest number of branches (40.51), but with the lowest number of seeds per head (137.51), seed yield per plant (15.35 g), seed yield per hectare (453.08 Kg/h) and thousand seed weight (3.00 g). The third cluster (C-III) comprised accessions obtained from West Gojjam and South Wollo, and this cluster was characterized with the lowest number of branches (29.58), number of heads per plant (143.09) and seed weight per plot (135.86) (Table 15).

Table 15. Mean of 13 quantitative traits per cluster at regional level

	Cluster				
	C-I	C-II	C-III	C-IV	C-V
DE	10.28	10.88	10.72	10.50	10.94
DFH	73.94	76.44	76.12	76.28	77.09
PH	149.89	154.83	158.09	157.42	159.08
DFF	78.04	80.71	81.64	81.40	82.37
NB	38.82	40.51	29.58	35.35	38.89
BL	55.98	54.49	56.20	53.39	58.61
DNM	169.41	173.02	171.35	169.53	173.05
NHP	169.26	164.71	143.09	162.41	152.74
SPH	151.35	137.51	143.10	138.50	143.45
SYP	15.92	15.35	15.64	16.18	15.47
SWP	142.77	139.21	135.86	151.95	149.82
SYH	475.78	453.08	459.10	525.06	497.43
TSW	3.06	3.00	3.03	3.01	3.03

Keys: DE = days to emergence, DFH = days to 50% heading, PH = plant height, DFF = days to 50% flowering, NB = number of branches, BL = branch length, DNM = days to 90% maturity, NHP = number of heads per plant, SPH = number of seeds per head, SYP = seed yield per plant, SWP = seed weight per plot, SYH = seed yield per hectare, TSW = 1000-seed weight.

Cluster-IV comprised accessions collected from the Borena zone alone and distinguished by having the highest seed yield per plant (16.18), seed weight per plot (151.95) and seed yield per hectare (525.06). Accessions collected from East Hararghe, West Hararghe and Derashie were grouped in the fifth cluster (C-V) and this cluster was characterized with the highest days to emergence (10.94), days to 50% heading (77.09), plant height (159.08), days to 50% flowering (82.37) and branch length (58.61) (Figure 8).

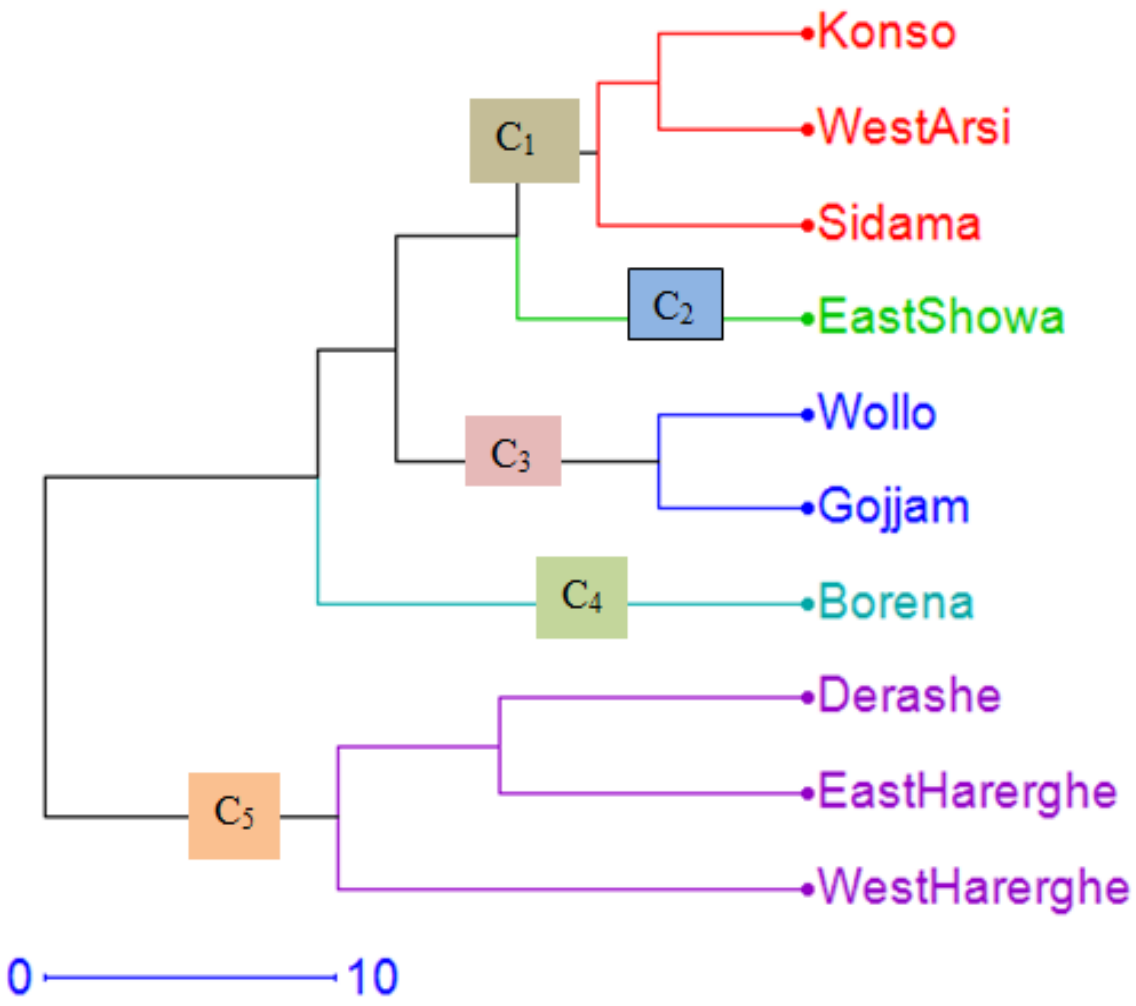

Figure 8. The genetic relations of 80 *Vernonia galamensis* accessions for 13 quantitative traits among regions

4.1.8. Genetic distance between clusters

The pairwise genetic distance (D^2) (Mahalanobis, 1936) was based on the means for the 13 quantitative traits, revealed that relatively wider genetic distance (variation) (Table 16). Inter-cluster D^2 values ranged from 9.29 to 73.47. The maximum pairwise generalized squared distances (D^2) were found between clusters C-II and C-IV (D^2 = 73.47), and the minimum pairwise occurs between cluster-III and cluster-VIII (D^2 = 9.29).

Table 16. Genetic distance for the 80 *Vernonia galamensis* accessions

Cluster	C-I	C-II	C-III	C-IV	C-V	C-VI	C-VII	C-VIII
C-I	0	28.23	35.40	53.07	31.42	16.64	25.77	21.10
C-II		0	26.13	73.47	47.94	44.94	42.03	39.96
C-III			0	71.06	43.07	55.54	57.07	9.29
C-VI				0	30.71	16.94	51.42	10.78
C-V					0	10.38	10.81	40.39
C-VI						0	15.56	21.69
C-VII							0	11.35
C-VIII								0

4.1.9. Principal component analysis of *Vernonia galamensis*

4.1.9.1. Accession based principal component analysis

Principal component analysis (PCA) indicated that the first vectors were more important than the second and other vectors. In the principal component analysis, only four principal components exhibited more than one eigenvalue and showed 71.0% of variation (Table 17). The Eigenvalues above 1 indicate the evaluated principal component (PC) weight values are reliable. The first PC accounted for 37.80% of the total variation, whereas the corresponding values for the second to the fourth PCs were 15.30%, 9.60% and 8.40%, respectively (Table 17). Those principal components having more than one eigenvalue were more varied and the best for selection of the diverse parents.

Agro-morphological quantitative traits such as days to 50% heading, plant height, days to 50% flowering, days to 90% maturity, number of heads per plant, number of seeds per head, seed yield per hectare and 1000-seed weight were the major contributors for the variations observed in the first principal component (Table 18). Hence, most of the variables in PC-1 were positive except for branch length (-0.02). PC-1 could be named components of productivity since it contains the highest factor loading for days to 50% flowering (0.89) and plant height (0.90) (Figure 9).

Table 17. Principal components for 13 quantitative traits of *Vernonia galamensis* accessions

Variables	PC1	PC2	PC3	PC4
DE	0.025	-0.114	0.599	0.33
DFH	0.93	-0.175	-0.097	0.037
PH	0.900	-0.073	-0.059	0.124
DFF	0.898	-0.181	-0.106	0.082
NB	0.240	-0.221	0.558	-0.328
BL	-0.016	0.445	0.174	0.730
DNM	0.612	-0.335	0.265	-0.153
NHP	0.817	-0.084	-0.053	0.106
SPH	0.801	-0.017	-0.225	0.122
SYP	0.794	-0.138	-0.16	0.159
SWP	0.704	0.613	0.143	-0.272
SYH	0.688	0.631	0.126	-0.273
TSW	0.207	-0.078	0.452	0.111
Eigenvalue	5.862	1.247	1.114	1.009
Proportion	0.378	0.153	0.096	0.084
Cumulative	0.378	0.530	0.626	0.710

DE = days to emergence, DFH = days to 50% heading, PH = plant height, DFF = days to 50% flowering, NB = number of branches, BL = branch length, DNM = days to 90% maturity, NHP = number of heads per plant, SPH = number of seeds per head, SYP = seed yield per plant, SWP = seed weight per plot, SYH = seed yield per hectare, TSW = 1000-seed weight.

The second principal component exhibited positive effects for branch length, seed weight per plot and seed yield per hectare, while days to 50% heading, plant height, days to 50% flowering and days to 90% maturity exhibited negative contribution. The PC-2 axis may consider as seed yield components because it showed the highest variability (0.63) for seed yield per hectare (Figure 9). Likewise, days to emergence, the number of branch, branch length, days to maturity and 1000-seed weight were among the major contributors to the variations in the third component. The fourth principal component is more related to days of emergence and branch length (Table 17).

A pattern of loading plots of the accessions was shown in Figure 9. The principal component analysis based on loading plots of the accessions using the first four principal components (PC1, PC2, PC3 and PC4) accounted for about 71.0% of the total variation. The patterns of distribution showed that plant height and days to 50% flowering distributed closer to each other at one plane of the PC, number of branches and branch length around the center of the plane while seed weight per plot and seed yield per hectare situated relatively far apart at other part in a plot of the PC (Figure 9).

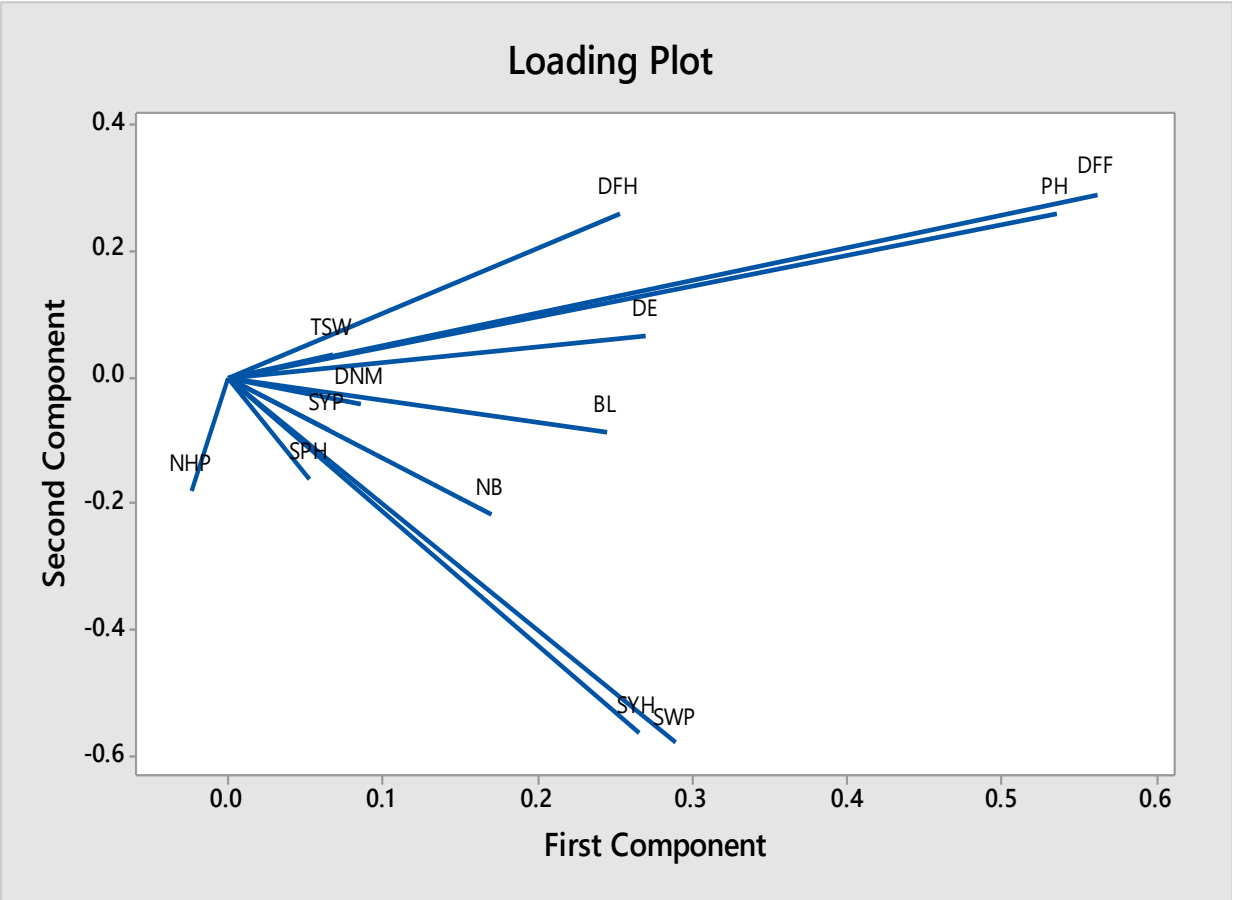

Figure 9. Principal component analysis loading plot used for 13 quantitative traits. DE = days to emergence, DFH = days to 50% heading, PH = plant height, DFF = days to 50% flowering, BN = number of branches, BL = branch length, DNM = days to 90% maturity, NHP = number of heads per plant, SPH = number of seeds per head, SYP = seed yield per plant, SWP = seed weight per plot, SYH = seed yield per hectare, TSW = 1000-seed weight.

The scree plot could be the percentage of variance associated with each principal component, and obtained by drawing a graph between eigenvalues and number of principal components (PCs) (Figure 10). According to scree plot diagram, the first four main principal component analyses are extracted from the complicated components, the total cumulative variance of these first four principal components (PC-1, PC-2, PC-3 and PC-4) accounted for 71.0% of the total variation. Hence, the first principal component (PC-1) showed the maximum variation, and then reduced gradually (Figure 10).

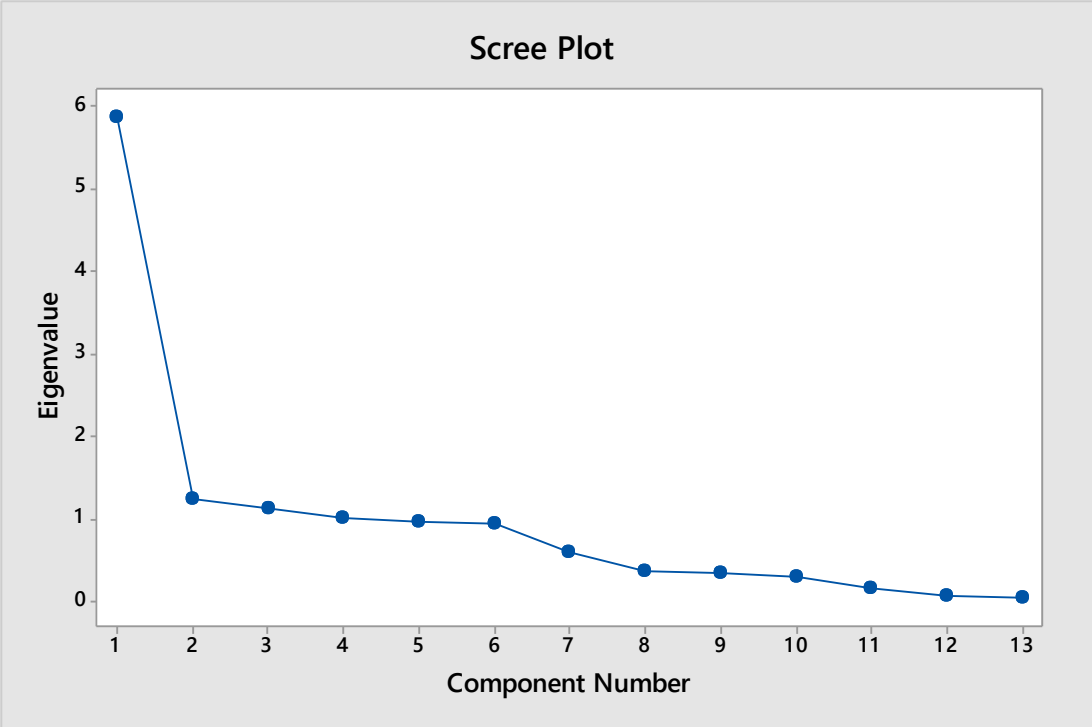

Figure 10. Scree plot of principal component analysis between Eigenvalue and number of PC

4.1.9.2. Region based principal component analysis

The presence of broad morphological differences among accessions was further confirmed by region based principal component analysis, which indicated that the overall diversity observed could be elucidated by a few eigenvectors. Accordingly, the first six principal components accounted for about 74.08% of total variation for all morphological traits and exhibited a high correlation among the characteristics analyzed (Table 18; Figure 11). From the eigenvector analysis, it was found that 2.41%, 1.99%, 1.60%, 1.29%, 1.26% and 1.07%

of the variation of morphological traits could be explained with respect to the first six principal components (Table 18; Figure 11).

Plant height (0.44) and days to 50% flowering (0.43) were the major contributors to the variation in the first principal component (PC-1). The second principal component (PC-2) analysis related to seed weight per plot (0.50) and seed yield per hectare (0.51). Number of branches (0.52) and number of heads per plant (0.56) were the main cause for variations observed in the third principal component (PC-3).

Table 18. Eigenvectors and eigenvalues of the first six region based principal components using agro-morphological traits in *Vernonia galamensis*

Variables	PC-1	PC-2	PC-3	PC-4	PC-5	PC-6
DE	0.12	0.03	0.02	0.11	0.04	-0.45
DFH	0.07	0.00	-0.11	0.50	-0.15	-0.16
PH	0.44	-0.03	0.01	-0.08	-0.01	0.05
DFF	0.43	-0.03	-0.01	0.02	-0.03	-0.01
NB	0.02	0.04	0.52	0.13	-0.03	0.00
BL	0.22	0.02	0.08	-0.24	0.26	0.27
DNM	-0.11	0.00	0.07	0.59	0.19	0.26
NHP	0.00	-0.07	0.56	-0.16	-0.03	0.02
SPH	-0.02	0.03	-0.14	0.01	0.53	0.07
SYP	-0.02	-0.10	0.08	0.05	0.58	-0.26
SWP	-0.02	0.50	0.01	0.02	-0.03	0.01
SYH	-0.03	0.51	-0.05	-0.02	-0.05	-0.07
TSW	0.07	-0.01	0.02	0.12	-0.05	0.69
Eigenvalue	2.41	1.99	1.60	1.29	1.26	1.07
Proportion	18.55	15.32	12.30	9.93	9.72	8.26
Cumulative	18.55	33.87	46.17	56.11	65.82	74.08

DE = days to emergence, DFH = days to 50% heading, PH = plant height, DFF = days to 50% flowering, BN = branch number, BL = branch length, DNM = days to 90% maturity, NHP = number of heads per plant, SPH = number of seeds per head, SYP = seed yield per plant, SWP = seed weight per plot, SYH = seed yield per hectare, TSW = 1000-seed weight.

Moreover, days to 50 % heading (0.50) and days to 90 % maturity (0.59) were the major traits contributing to the variation observed in the fourth principal component (PC-4), whereas number of seeds per head (0.53) and seed yield per plant (0.58) were the main contributing traits for the fifth principal component (PC-5). Lastly, branch length 0.27), days to 90 % maturity (0.26) and thousand seed weight (0.69) were the major observed factors for the sixth principal component (PC-6) (Table 18).

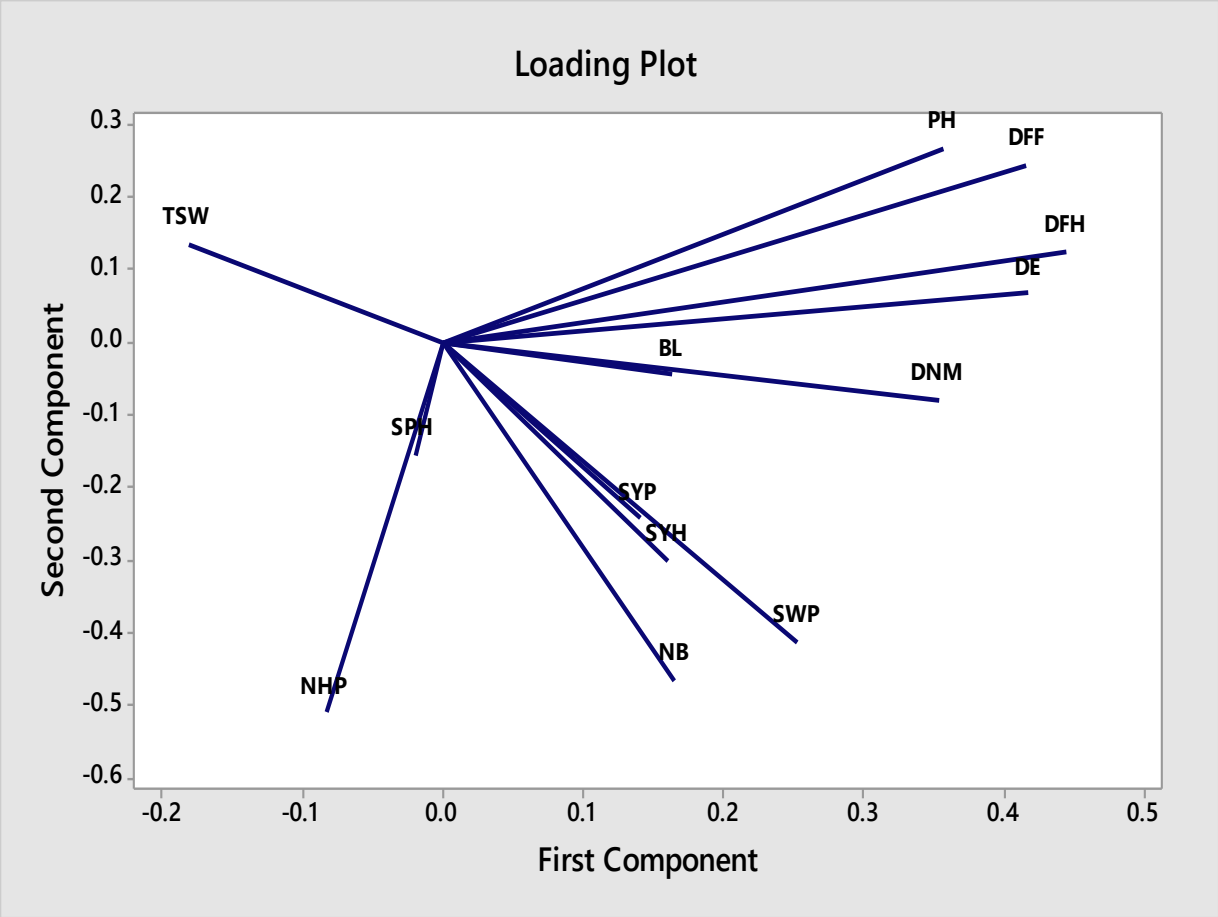

Figure 11. Region based PCA of 13 quantitative traits showing relationships among *Vernonia galamensis* accessions. DE = days to emergence, DFH = days to 50% heading, PH = plant height, DFF = days to 50% flowering, BN = branch number, BL = branch length, DNM = days to 90% maturity, NHP = number of heads per plant, SPH = number of seeds per head, SYP = seed yield per plant, SWP = seed weight per plot, SYH = seed yield per hectare, TSW = 1000-seed weight

4.2. Molecular based genetic diversity of *Vernonia galamensis* using SSR markers

4.2.1. Allelic frequency and patterns of diversity

A total of 20 SSR markers were used for the characterization and genetic diversity analysis of the 150 *V. galamensis* accessions (Table 19). The primers used contain di, tri and hexa-nucleotide repeats. A total of 79 alleles were identified, ranging from 2 to 6, with an average of 3.9 alleles per locus. The number of effective alleles (Ne) is an important parameter to measure genetic diversity, and varied from 1.99 (Vg-016) to 4.79 (Vg-003) with an average of 3.06. The highest major allele frequency (MAF) (0.85) was recorded for locus Vg-001 and the lowest MAF was (0.45) recorded for locus Vg-003.

Table 19. Summary of genetic parameters revealed by 20 SSR markers for *Vernonia galamensis* populations collected from different regions of Ethiopia

Locus	Na	Ne	MAF	Ho	He	uHe	F	PIC	I
Vg-001	3	2.8	0.85	0.25	0.48	0.49	0.25	0.80	1.06
Vg-002	5	3.26	0.77	0.13	0.50	0.51	0.73	0.96	1.36
Vg-003	6	4.79	0.45	0.36	0.65	0.64	0.77	0.62	1.67
Vg-004	3	2.28	0.52	0.07	0.59	0.61	0.88	0.64	0.90
Vg-005	3	2.01	0.64	0.13	0.60	0.63	0.78	0.50	0.89
Vg-006	5	3.69	0.54	0.13	0.44	0.45	0.70	0.86	1.44
Vg-007	4	3.46	0.61	0.20	0.58	0.60	0.66	0.57	1.31
Vg-008	5	3.26	0.59	0.20	0.61	0.63	0.67	0.89	1.36
Vg-009	3	2.6	0.61	0.23	0.45	0.47	0.49	0.87	1.01
Vg-010	4	3.81	0.60	0.07	0.46	0.48	0.86	0.86	1.36
Vg-011	4	2.03	0.56	0.13	0.23	0.24	0.42	0.93	0.95
Vg-012	4	2.53	0.57	0.07	0.46	0.48	0.86	0.84	1.01
Vg-013	4	3.57	0.49	0.13	0.48	0.50	0.72	0.73	1.33
Vg-014	3	2.27	0.60	0.27	0.48	0.50	0.44	0.77	0.95
Vg-015	4	2.68	0.76	0.08	0.50	0.52	0.85	0.88	1.16
Vg-016	2	1.99	0.64	0.07	0.36	0.37	0.81	0.78	0.86
Vg-019	4	3.46	0.58	0.33	0.64	0.67	0.48	0.70	1.31
Vg-021	5	4.41	0.51	0.05	0.60	0.62	0.89	0.86	1.55
Vg-024	5	3.81	0.67	0.27	0.62	0.65	0.57	0.53	1.46
Vg-030	3	2.53	0.61	0.07	0.28	0.29	0.76	0.55	1.01
Mean	3.9	3.06	0.61	0.16	0.50	0.52	0.68	0.76	1.20

Key: Na = number of different alleles, Ne = number of effective alleles, MAF = major allele frequency, Ho = observed heterozygosity, He = expected heterozygosity, uHe = unbiased expected heterozygosity, F = fixation index, PIC = polymorphic information content, I = Shannon's information index.

The observed heterozygosity (Ho) values were quite low and ranged between 0.05 (Vg-021) and 0.36 (Vg-003) with an average of 0.16 across all the 20 markers evaluated. The expected heterozygosity (*He*) mean was 0.50, ranging from 0.23 (Vg-011) to 0.65 (Vg-003) (Table 18). For most of the loci the *He* values were higher than *Ho* value (Table 19).

The Wright's fixation index (F) estimated and ranged from 0.25 (Vg-001) to 0.89 (Vg-021) with an average of 0.68. Likewise, the unbiased expected heterozygosity (uHe) means was 0.52 and ranged from 0.24 (Vg-011) to 0.67 (Vg-019).

Polymorphic information content (PIC) is generally used for characterization of marker polymorphism and the values ranged between 0.50 and 0.96 with an average of 0.76. Microsatellite markers such as Vg-002 and Vg-011 showed the highest polymorphism with 0.96 and 0.93, respectively. The results of diversity parameters showed a high level of polymorphism among the 20 SSR markers, indicating existence of genetic variation within *V. galamensis* collections (Table 20). Finally, Shannon-Weaver's information indices (I) ranged from 0.86 to 1.67 and averaged at 1.20 (Table 19).

On the other hand, the allelic patterns across the studied populations, the number of alleles (Na), number of effective alleles (Ne) and Shannon's diversity index (I) were presented in (Table 20). From the allelic pattern of variation among populations, East Showa and East Hararghe had the highest expected heterozygosity (0.51and 0.49, respectively).

As can be seen in Table 20, accessions from East Showa had the highest average number of alleles (5). Likewise, populations from East Showa and East Hararghe had high number of effective alleles (Ne), (4.74 and 3.68), respectively. This result showed that populations from East Showa and East Hararghe may be, having high number of alleles and number of effective alleles. As with the scenario observed regarding the number of effective alleles (Ne), East Showa and East Hararghe had again the highest genetic diversity measures, Shannon's index (0.82 and 0.79, respectively).

Table 20. Important allelic values recorded in the ten populations of *Vernonia galamensis*

Population	N	Na	Ne	I	Ho	He	Uhe	F
Borena	15	3.00	1.98	0.76	0.13	0.48	0.50	0.74
Sidama	15	4.00	2.03	0.74	0.12	0.47	0.51	0.72
East Showa	15	5.00	4.74	0.82	0.15	0.51	0.55	0.68
West Arsi	15	3.00	2.07	0.70	0.13	0.45	0.47	0.70
East Hararghe	16	4.00	3.68	0.79	0.12	0.49	0.53	0.73
West Hararghe	14	4.00	3.12	0.77	0.16	0.48	0.50	0.73
West Gojjam	14	3.00	2.33	0.70	0.13	0.47	0.49	0.64
South Wollo	16	4.00	2.74	0.72	0.14	0.48	0.52	0.68
Konso	15	3.00	2.21	0.71	0.13	0.45	0.48	0.72
Derashie	15	2.00	1.74	0.75	0.09	0.43	0.47	0.81
Average across population	15	3.50	2.70	0.75	0.13	0.47	0.50	0.72

Key: Na = number of different alleles, Ne = number of effective alleles, I = Shannon's diversity index, Ho = observed heterozygosity, He = expected heterozygosity, uHe = unbiased expected heterozygosity, F = fixation index, PIC = polymorphic information content

4.2.2. Analysis of molecular variance and genetic distances

4.2.2.1. Analysis of molecular variance (AMOVA)

The AMOVA of *V. galamensis* is presented in Table 21 and Figure 12, and it shows that 67% of the variation was attributed to genetic variability within populations, whereas 22% was due to variation among individuals within the same population. In contrast, a smaller portion (11%) of the variation was due variation among populations. F_{ST} values among all populations were significant ($F_{ST} = 0.15$, $p < 0.001$) (Table 21).

Table 21. Analysis of Molecular Variance (AMOVA) showing the distribution of genetic diversity within and among populations and among individuals of *Vernonia galamensis* collected from different regions of Ethiopia

Source	Df	SS	MS	Est. Var.	% Variation	F-Statistics	Value	P
Among populations	9	322.71	35.86	0.76	11%	Fst	0.15	0.001
Among individuals	140	1842.01	13.16	6.39	22%	Fis	0.946	0.001
Within individual	150	55.21	0.37	0.37	67%	Fit	0.951	0.001
Total	299	2219.93		7.52	100%			

Df = Degrees of Freedom; SS = Sum of Squares; MS = Mean Square; Est. Var. = Estimated Variability

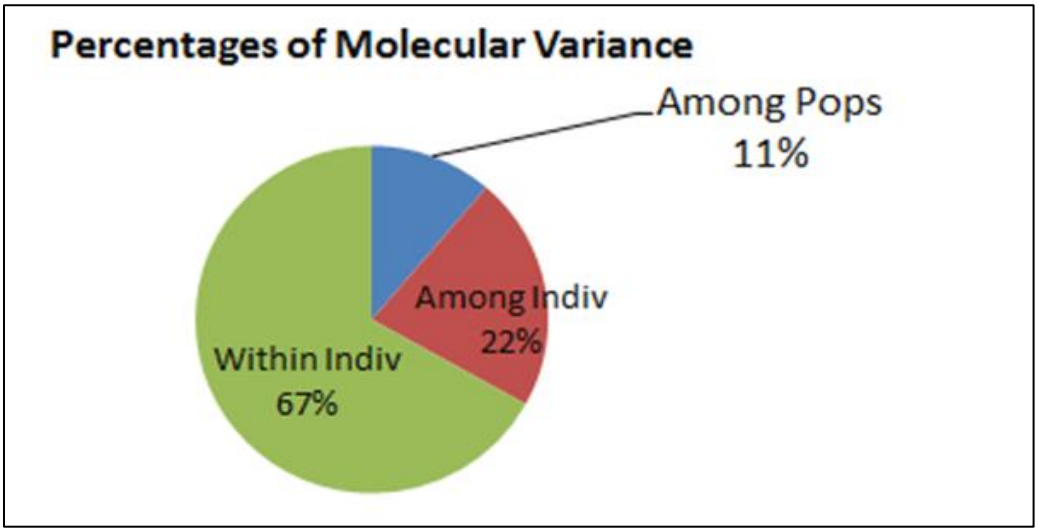

Figure 12. Analysis of molecular variation of pie chart for 150 *Vernonia galamensis* accessions from ten populations in Ethiopia

The values of the various genetic diversity indices of the 20 loci for *V. galamensis* populations are presented in Table 22. The total expected heterozygosity (Ht) ranged from 0.65 (Vg-001) to 0.83(Vg-008) with an average of 0.74. Dst (Nei's, 1978) ranged from -0.01 (Vg-012) to 0.54 (Vg-009) with overall diversity of 0.22. The overall *F*st value was 0.15 and ranged from 0.06 (Vg-012) to 0.33 (Vg-009), and the values of gene flow was ranged from 0.04 (Vg-001) to 13.32(Vg-006) with an average of 1.11 (Table 22). The average extent of differentiation was low (G_{st} = 0.09).

Table 22. Indices of genetic diversity of *Vernonia galamensis* population using 20 SSR loci

Locus	Hs	Ht	cHs	Dst	Fis	Gis	Fst	Gst	Nm	HWc
Vg-001	0.51	0.65	0.54	0.31	0.82	0.83	0.24	0.19	0.04	0.00**
Vg-002	0.58	0.66	0.61	0.15	0.33	0.36	0.12	0.08	0.64	0.29ns
Vg-003	0.71	0.81	0.76	0.24	0.99	0.99	0.12	0.06	0.28	0.00**
Vg-004	0.64	0.76	0.69	0.26	0.98	0.98	0.15	0.09	1.44	0.00**
Vg-005	0.70	0.81	0.75	0.25	1.00	1.00	0.13	0.07	0.15	0.00**
Vg-006	0.70	0.81	0.75	0.28	1.00	1.00	0.14	0.08	13.32	0.65ns
Vg-007	0.64	0.78	0.69	0.36	1.00	1.00	0.18	0.13	0.14	0.00**
Vg-008	0.68	0.83	0.73	0.42	1.00	1.00	0.18	0.12	0.44	0.00**
Vg-009	0.51	0.76	0.55	0.54	1.00	1.00	0.33	0.28	0.45	0.00**
Vg-010	0.63	0.72	0.68	0.16	1.00	1.00	0.12	0.07	0.22	0.00**
Vg-011	0.59	0.71	0.64	0.25	0.99	0.99	0.17	0.11	0.06	0.00**
Vg-012	0.64	0.68	0.68	-0.01	0.99	0.99	0.06	0.00	1.57	0.00**
Vg-013	0.58	0.71	0.63	0.27	1.00	1.00	0.18	0.13	0.04	0.00**
Vg-014	0.61	0.72	0.65	0.22	1.00	1.00	0.15	0.10	0.22	0.00**
Vg-015	0.65	0.76	0.70	0.22	1.00	1.00	0.14	0.08	0.19	0.00**
Vg-016	0.61	0.71	0.65	0.21	1.00	1.00	0.15	0.09	0.12	0.00**
Vg-019	0.67	0.74	0.72	0.10	1.00	1.00	0.10	0.04	0.54	0.00**
Vg-021	0.67	0.75	0.72	0.13	0.98	0.98	0.10	0.05	0.23	0.00**
Vg-024	0.64	0.69	0.69	0.04	1.00	1.00	0.08	0.02	1.97	0.70ns
Vg-030	0.68	0.74	0.73	0.08	1.00	1.00	0.09	0.03	0.15	0.89ns
Mean	0.63	0.74	0.68	0.22	0.95	0.96	0.15	0.09	1.11	

Hs = Mean Expected Heterozygosity, Ht = Total Expected Heterozygosity, cHs = Corrected Hs, adjusted for bias, Dst = Nei's (1978) unbiased average gene diversity between subpopulations, Fis = Inbreeding coefficient within individuals, Gis = Inbreeding coefficient within individuals, adjusted for bias, Fst = Inbreeding coefficient within subpopulations, relative to total genetic differentiation among populations, Gst = Analog of Fst, adjusted for bias, HWc = Overall Hardy-Weinberg test using Fit, Nm = Gene flow estimated from Fst (Nei, 1987); * $p<0.05$, ** $p<0.01$, ***$p<0.001$.

Pair-wise populations' numbers of migrants (Nm) values based on Fst values in *V. galamensis* populations are presented in Table 23. The pairwise Nm values among the ten populations, with the largest number of migrants per generation was between East Hararghe and West Hararghe (Nm=13.45), and the next highest value between Sidama and West Arsi (Nm= 10.91). The lowest number of migrants per generation (Nm) value was recorded between South Wollo and Derashie (Nm =1.60). This finding may be supported by the geographical proximity between the regions or collection sites for *V. galamensis* populations.

Table 23. Pair-wise populations' number of migrants (Nm) values (Nm ≥1) based on Fst values in *Vernonia galamensis* populations

	BOR	SID	ESH	WAS	EHG	WHG	GOJ	WOL	KON	DER
BOR	0.00									
SID	2.24	0.00								
ESH	8.55	4.80	0.00							
WAS	5.21	10.91	4.39	0.00						
EHG	4.13	2.42	5.28	5.69	0.00					
WHG	2.94	2.11	4.07	3.82	13.45	0.00				
GOJ	4.16	3.04	3.76	5.64	3.68	2.45	0.00			
WOL	4.94	2.05	5.73	3.50	4.24	5.07	9.07	0.00		
KON	6.72	2.37	4.81	3.83	7.30	9.82	3.10	4.84	0.00	
DER	3.69	2.94	6.66	5.00	3.91	3.65	2.03	1.60	4.62	0.00

Keys: BOR = Borena, SID = Sidama, ESH = East Showa, WAS = West Arsi, EHG = East Hararghe, WHG = West Hararghe, GOJ = West Gojjam, WOL = South Wollo, KON = Konso, DER = Derashie

4.2.2.2. Genetic distance among *Vernonia galamensis* germplasm

Pairwise population matrix of Nei genetic distance (Ds) and pairwise population matrix of Nei genetic identity are presented in Table 24. The result showed the maximum pairwise genetic distance between populations of Borena and East Hararghe (0.57), followed by populations of Sidama and West Hararghe (0.54), whereas the minimum genetic distance was observed between populations of Borena and Konso (0.24).

Further, the highest pairwise Nei's genetic identity was occurred between Konso and Derashie (0.80) populations, while the least Nei's genetic identity was observed between Borena and Sidama (0.51) populations. The overall magnitude of pairwise population matrix of Nei genetic distance was relatively lower than that of Nei's genetic identity (Table 24).

Table 24. Pairwise population matrix of Nei genetic distance (Ds) (above diagonal) and pairwise population matrix of Nei genetic identity (I) (below diagonal)

Population	BOR	SID	ESH	WAS	EHG	WHG	GOJ	WOL	KON	DER
BOR	**	0.47	0.37	0.43	0.57	0.40	0.46	0.36	0.24	0.41
SID	0.51	**	0.59	0.37	0.39	0.54	0.51	0.44	0.50	0.36
ESH	0.69	0.56	**	0.49	0.44	0.47	0.34	0.29	0.42	0.27
WAS	0.65	0.69	0.61	**	0.44	0.38	0.45	0.36	0.35	0.47
EHG	0.59	0.68	0.65	0.64	**	0.47	0.47	0.40	0.39	0.25
WHG	0.67	0.58	0.62	0.68	0.62	**	0.41	0.35	0.31	0.41
GOJ	0.63	0.60	0.71	0.64	0.76	0.66	**	0.36	0.37	0.38
WOL	0.70	0.65	0.75	0.70	0.67	0.70	0.70	**	0.37	0.42
KON	0.79	0.61	0.66	0.70	0.68	0.73	0.69	0.69	**	0.35
DER	0.67	0.69	0.76	0.63	0.78	0.66	0.79	0.70	0.80	**

Keys: BOR = Borena, SID = Sidama, ESH = East Showa, WAS = West Arsi, EHG = East Hararghe, WHG = West Hararghe, GOJ = West Gojjam, WOL = South Wollo, KON = Konso, DER = Derashie

4.2.3. Cluster and principal co-ordinate analyses of *Vernonia galamensis*

4.2.3.1. Cluster analysis

Clustering analysis performed based on the allelic frequency using neighbor-joining with the DARwin 6.0.19 software grouped the 150 accessions into four (4) major clusters from the main node. Each of the four clusters comprises accessions collected from different zones (geographic regions). The first and the fourth cluster further divided into sub-clusters, and the samples grouped according to their geographic origin (Figure 13).

Cluster 1 comprised of 39 accessions, and most of the accessions were come from Borena, Konso and East Hararghe. The second cluster was characterized as the second major group, contained 42 accessions, and most of the accessions were collected from South Wollo and West Hararghe. Cluster three was characterized as the smallest group in clustering comprising of 15 accessions that came from East Showa and Borena. The fourth cluster comprised of the major accessions (54 accessions), and most of the accessions are collected from West Arsi and Sidama (Figure 13). Generally, the cluster analysis revealed that accessions from different populations (collection sites) clustered together and no clear pattern of clustering on the basis of geographic origins, which may imply the presence of gene flow between populations of different collection sites.

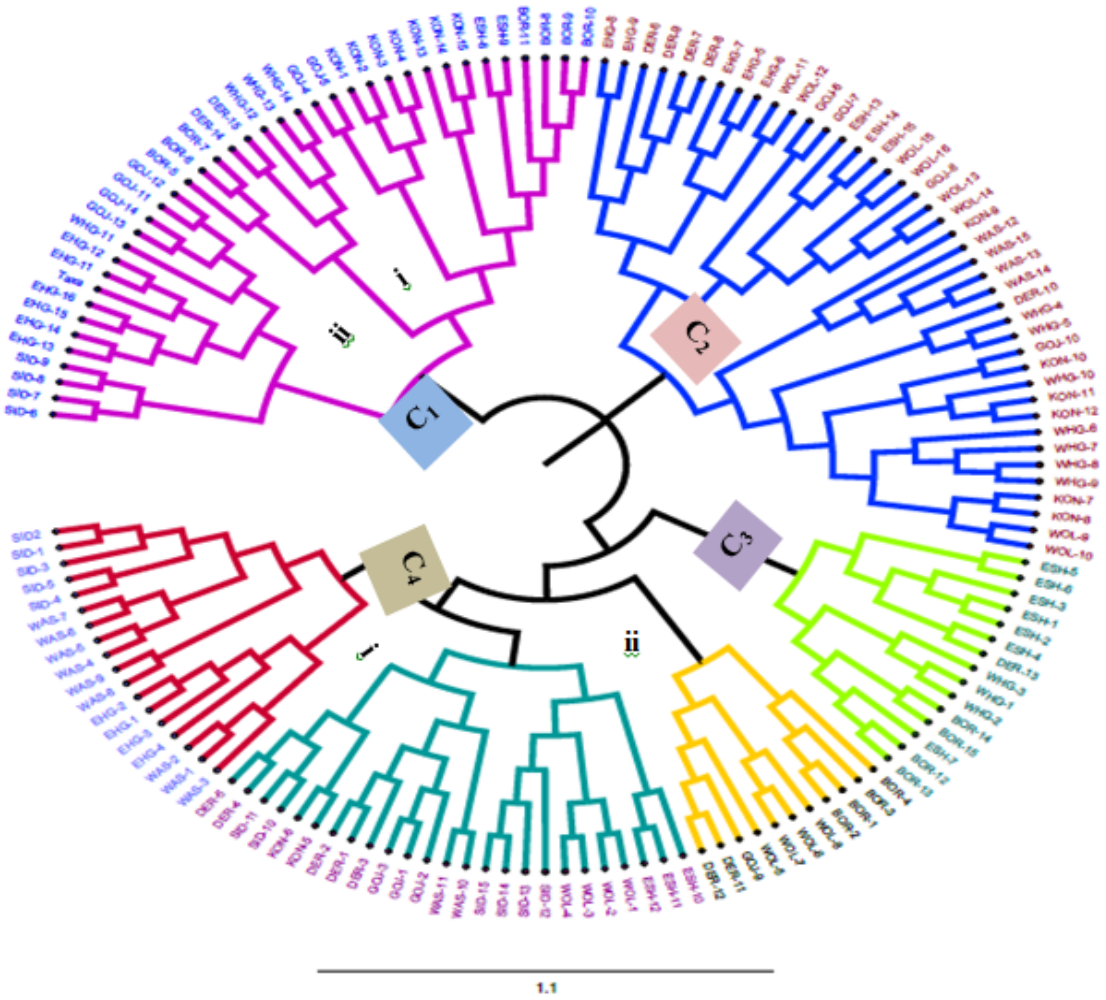

Figure 13. Neighbor-joining tree of the 150 *Vernonia galamensis* accessions constructed by Darwin software program. Keys: BOR = Borena, SID = Sidama, ESH = East Showa, WAS = West Arsi, EHG = East Hararghe, WHG = West Hararghe, GOJ = West Gojjam, WOL = South Wollo, KON = Konso, DER = Derashie

4.2.3.2. Principal co-ordinates analysis (PCoA)

The principal co-ordinate analysis (PcoA) conducted using 150 samples of *V. galamensis*, showed that the majority of samples were placed at the center of a two-dimensional coordinate plane and roughly forms four groups (Figure 14). The first three axes of the PCoA accounted together 33.02% of the total variation, with 12.83%, 10.87%, and 9.32% explained by PC axis 1, 2, and 3, respectively. Likewise, the result showed that accessions obtained from different collection areas often clustered together. This, again, agrees with the results of the NJ dendrogram in that there was no clear clustering among accessions from the same population/collection areas.

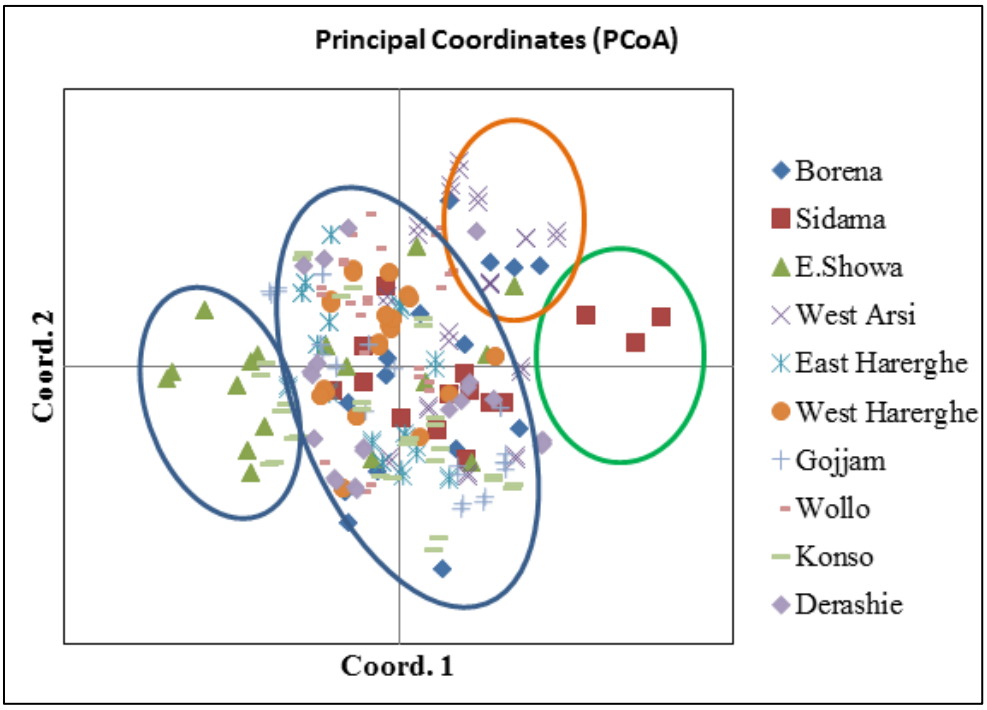

Figure 14. Two dimensional scaling principal coordinate analyses of 150 *Vernonia galamensis* accessions with ten populations

4.2.4. Population structure analysis

Bayesian model-based algorithm was applied to distinguish the 150 *V. galamensis* accessions to different populations and to determine the patterns of populations' structure. According to Evanno *et al*. (2005) and Gilbert *et al*. (2012), STRUCTURE outputs were predicted at K = 6, most likely selected (number of clusters) to describe the genetic structure of the *V. galamensis* accessions (Figure 15). Based on this value (Clumpak result) revealed that accessions collected from the same region of origin did not often grouped entirely together within a given major group. There was a wide admixture in structuring of *V. galamensis* populations, which agreed with neighbor joining trees.

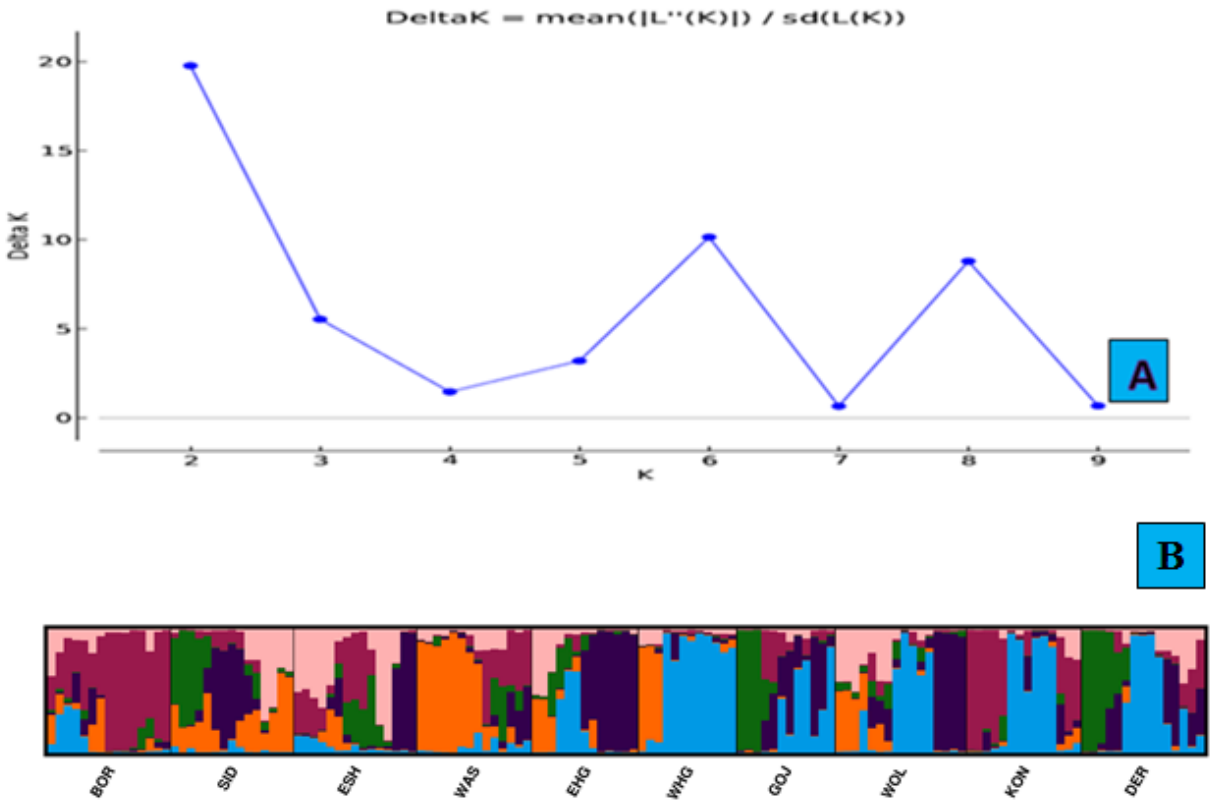

Figure 15. Delta K values estimated according to Evanno *et al*. (2005) Method (A). Bayesian model-based estimation of population structure (K = 6) (B) for 150 *Vernonia galamensis* accessions from different growing regions in Ethiopia. Keys: BOR = Borena, SID = Sidama, ESH = East Showa, WAS = West Arsi, EHG = East Hararghe, WHG = West Hararghe, GOJ = West Gojjam, WOL = South Wollo, KON = Konso, DER = Derashie

4.3. Seed oil content and fatty acid composition

The result demonstrated that there was a high significant variation ($p < 0.05$) for traits of seed oil content and fatty acids composition, except arachidic acid and lenolenic acid (Table 25). The seed oil content ranged from 28.70% to 35.32%, with an average of 32.17%.VgEt-094 was the dominant accession in oil content (35.32%) that grown at Wondo Genet Agricultural Research Center. In addition, vernolic acid was the predominant fatty acid, ranged from 68.78% (VgEt-111) to 77.31% (VgEt-094) with an average of 72.90%. Likewise, there was considerable variation in the proportion of other fatty acids. For instance, the average linoleic acid was 14.08% that ranged from 12.0% to 15.91%, whereas palmitic acid averaged at 3.10% and ranged from 2.80% to 3.38% (Table 25).

Table 25. The seed oil content and fatty acids composition of *Vernonia galamensis* that grown at Wondo Genet agricultural research center

S.No	Accession	OC %	VA (%)	LA (%)	OA (%)	PA (%)	SA (%)	AA (%)	NA (%)
1	VgEt-017	33.25	73.00	12.85	4.23	3.31	2.42	0.50	0.41
2	VgEt-020	28.70	73.90	15.91	4.78	3.15	2.01	0.50	0.38
3	VgEt-027	34.20	74.45	15.02	5.02	3.21	2.49	0.56	0.40
4	VgEt-032	29.92	70.85	14.78	4.21	3.16	2.42	0.53	0.38
5	VgEt-042	33.42	73.80	13.21	4.78	3.02	2.50	0.58	0.40
6	VgEt-076	32.67	75.01	15.05	5.12	3.38	3.02	0.52	0.45
7	VgEt-078	34.75	69.00	14.02	4.00	2.80	2.80	0.55	0.43
8	VgEt-080	33.50	73.61	13.42	4.21	2.90	3.15	0.51	0.40
9	VgEt-088	31.85	72.24	14.00	4.91	3.20	3.00	0.53	0.37
10	VgEt-090	29.20	75.34	15.20	4.15	3.11	2.95	0.58	0.38
11	VgEt-094	35.32	77.31	14.42	4.05	3.15	2.80	0.57	0.36
12	VgEt-111	29.67	68.78	13.08	4.53	3.01	2.56	0.55	0.39
13	VgEt-118	34.11	74.01	14.10	5.20	3.10	2.85	0.53	0.41
14	VgEt-130	30.82	72.20	12.00	4.23	3.02	2.80	0.51	0.42
15	VgEt-134	31.12	70.05	14.20	4.50	3.03	2.70	0.52	0.43
Mean		32.17	72.90	14.08	4.53	3.10	2.70	0.54	0.40
LSD		0.90	0.95	0.65	0.50	0.82	0.60	0.50	0.40

Significant at $p < 0.05$. Keys: OC= seed oil content; VA = vernolic acid; LA =linoleic acid; OA= oleic acid; PA = palmitic acid; SA = stearic acid; AA = arachidic acid; NA = lenolenic acid

Further, there was also a significant variation existed between accessions for seed oil contents and fatty acid composition that grown at Melka Werer agricultural research center. The seed oil content was ranged from 30.23% (VgEt-111) to 38.01% (VgEt-094) with an average of 34.31% (Table 26). The fatty acid composition, vernolic acid (VA) showed significant variation that ranged from 70.41% (VgEt-078) to 79.55% (VgEt-094) with an average of 74.53%. Likewise, there was considerable variation in the other fatty acids such as linoleic acid was ranged from 13.02 to 17.18% with an average of 14.89%, whereas oleic acid ranged from 4.00 to 6.00% with an average of 4.81%. Palmitic acid was averaged at 3.09%, and ranged from 2.80% to 3.42%. In the same way, stearic acid averaged at 2.81% and ranged from 2.49 to 3.18% (Table 26).

Table 26. The seed oil content and fatty acids composition of *Vernonia galamensis* that grown at Melka Werer agricultural research center

S.No	Accession	OC %	VA (%)	LA (%)	OA (%)	PA (%)	SA (%)	AA (%)	NA (%)
1	VgEt-017	36.47	73.91	14.47	5.27	3.33	2.89	0.48	0.82
2	VgEt-020	37.08	74.86	13.37	5.26	3.01	3.18	0.98	0.38
3	VgEt-027	36.12	76.00	16.23	5.25	3.30	2.49	0.56	0.42
4	VgEt-032	32.44	74.01	17.18	5.00	3.42	2.55	0.50	0.42
5	VgEt-042	36.00	76.04	13.37	4.02	3.40	2.75	0.51	0.42
6	VgEt-076	31.45	75.69	15.56	5.35	3.08	2.98	0.53	0.40
7	VgEt-078	35.33	70.41	15.58	4.54	3.02	2.83	0.54	0.44
8	VgEt-080	36.04	74.02	13.48	4.85	2.92	2.91	0.52	0.38
9	VgEt-088	33.77	72.60	14.39	4.01	3.00	3.09	0.54	0.39
10	VgEt-090	31.02	78.38	15.58	4.85	2.90	2.85	0.60	0.39
11	VgEt-094	38.01	79.55	16.15	4.75	3.01	2.88	0.58	0.33
12	VgEt-111	30.23	70.69	13.02	4.00	2.80	2.57	0.53	0.42
13	VgEt-118	33.81	74.74	16.00	6.00	3.18	2.84	0.52	0.42
14	VgEt-130	32.40	74.15	14.94	4.90	2.81	2.70	0.53	0.40
15	VgEt-134	34.54	72.85	14.02	4.03	3.16	2.69	0.53	0.44
Mean		34.31	74.53	14.89	4.81	3.09	2.81	0.56	0.43
LSD		0.96	0.97	0.65	0.55	0.85	0.56	0.36	0.40

Significant at $p < 0.05$. Keys: OC= seed oil content; VA = vernolic acid; LA =linoleic acid; OA = oleic acid; PA = palmitic acid; SA = stearic acid; AA = arachidic acid; NA = lenolenic acid

The result of seed oil content, important industrial fatty acid and 1000-seed weight are present in Table 27. The oil content was ranged from 29.95 to 36.67% with an average of 33.24%. Likewise, vernolic fatty acid ranged from 69.71 to 78.43% with an average of 73.72%. The other fatty acids such as linoleic and linolenic fatty acids had an average of 14.49% and 0.42%, respectively. VgEt-094 that had the highest 1000-seed weight (3.51 g) also performed the highest percentage of oil content 36.67% and vernolic acid (78.43%) (Table 27).

Table 27. The seed oil content and fatty acids composition and 1000-seed weight of *Vernonia galamensis*

Accession	OC %	VA (%)	LA (%)	NA (%)	TSW(g)
VgEt-017	34.86	73.46	13.66	0.62	3.17
VgEt-020	32.89	74.38	14.64	0.38	3.01
VgEt-027	35.16	75.23	15.63	0.41	3.01
VgEt-032	31.18	72.43	15.98	0.4	3.25
VgEt-042	34.71	74.92	13.29	0.41	3.09
VgEt-076	32.06	75.35	15.31	0.43	2.8
VgEt-078	35.04	69.71	14.8	0.44	3.21
VgEt-080	34.77	73.82	13.45	0.39	2.89
VgEt-088	32.81	72.42	14.2	0.38	3.09
VgEt-090	30.11	76.86	15.39	0.39	3.26
VgEt-094	36.67	78.43	15.29	0.35	3.51
VgEt-111	29.95	69.74	13.05	0.41	3.06
VgEt-118	33.96	74.38	15.05	0.42	2.8
VgEt-130	31.61	73.18	13.47	0.41	3.02
VgEt-134	32.83	71.45	14.11	0.44	3.22
Mean	33.24	73.72	14.49	0.42	3.09
LSD	0.95	0.97	0.82	0.09	0.28

Significant at $p < 0.05$. Keys : OC= seed oil content; VA = vernolic acid; LA =linoleic acid; NA = lenolenic acid; TSW = thousand seed weight

The minimum seed oil content was 28.70% at Wondo Genet while 30.23% at Melka Werer. Among fatty acid, vernolic acid was the predominant fatty acid, the maximum vernolic acid was 79.55% at Melka Werer whereas 77.31% at Wondo Genet (Table 28).

Table 28. Minimum, Maximum, Means (%) and standard error of the mean (SEM) for seed oil content and fatty acids composition in *Vernonia galamensi*s

Fatty acids	Melka Werer			Wondo Genet		
	Minimum	Maximum	mean ±SE	Minimum	maximum	mean±SE
OC	30.23	37.08	34.31±0.45	28.70	35.32	32.17±0.50
VA	70.69	79.55	74.53±0.21	68.78	77.31	72.90±0.32
LA	13.02	17.18	14.89±0.39	12.00	15.91	14.08±0.65
OA	4.00	6.00	4.81±0.26	4.00	5.20	4.53±0.13
PA	2.81	3.42	3.09±0.06	2.80	3.38	3.1±0.23
SA	2.49	3.18	2.81±0.18	2.01	3.15	2.7±0.12
AA	0.48	0.98	0.56±0.26	0.50	0.58	0.54±0.42
NA	0.33	0.82	0.43±0.04	0.36	0.45	0.40±0.07

Keys : OC= seed oil content; VA = Vernolic acid; LA = linoleic acid; OA = oleic acid; PA = palmitic acid; SA = stearic acid; AA = arachidic acid; NA =Linolenic acid

4.3.1. Relationship between seed oil content and level of fatty acids

The Pearson correlation coefficient is presented in Table 29. The result showed that seed oil content significantly high and positively correlated with vernolic acid (r = 0.51), linoleic acid (r =0.24), linolenic acid (r = 0.54) and 1000-seed weight.

Table 29. Pearson correlation coefficient of seed oil content and fatty acid composition of *Vernonia galamensis*

Fatty acid	OC	VA	LA	NA	TSW
OC		0.51*	0.24*	0.54*	0.24*
VA			0.01	0.42*	0.51*
LA				0.1	0.24*
NA					-0.07
TSW					

Keys: OC= seed oil content; VA = Vernolic acid; LA =linoleic acid; NA=linolenic acid

Likewise, vernolic acid positively correlated with linolenic acid (r = 0.42). Thousand seed weight was positively associated with oil content (r = 0.24) and vernolic acid (r = 0.51) (Table 29).

4.3.2. Component analysis for seed oil content and fatty acid composition

Component analysis reflects the importance of the largest contributor to the total variation at each axis of differentiation. Three principal components showed more than one Eigenvalue and showed about 72.0% of variability (Table 30, Figure 16). PC-I showed 28.0%, PC-II 25.0% and PC-III exhibited 11.0% variability with eigenvalue of 2.20, 1.99, and 1.57, respectively (Table 30).

Table 30. Principal component analysis of oil contents and fatty acids composition of *Vernonia galamensis*

Variable	PC-I	PC-II	PC-III
OC	0.56	0.03	-0.32
VA	0.52	-0.18	-0.09
LA	0.09	0.59	0.00
OA	0.11	0.60	0.15
PA	-0.21	0.18	-0.39
SA	0.47	-0.27	-0.09
AA	0.35	0.37	0.23
NA	-0.08	0.15	-0.68
Eigenvalue	2.20	1.99	1.57
Proportion	0.28	0.25	0.11
Cumulative	0.28	0.52	0.72

Keys : OC= seed oil content, VA = Vernolic acid, LA =linoleic acid, OA = oleic acid, PA = palmitic acid, SA = stearic acid, AA = arachidic acid, NA = linolenic acid

From the PCA, the traits such as seed oil content (0.56), vernolic acid (0.52), stearic acid (0.47) and arachidic acid (0.35) were the major contributors for the variations observed in the first principal component (PC-1). Most of the variables in PC-1 were positive except palmitic acid (-0.21) and linoleic acid (-0.08). The second principal component related to linoleic acid (0.59), oleic acid (0.60) and arachidic acid (0.37) (Figure 16). In addition, PC-3 composed of arachidic acid (0.23) and oleic acid (0.15), but oil content (-0.32), palmitic acid (-0.39) and linolenic acid (-0.68) were negative contribution for PC-3 (Table 30, Figure 16).

The loading plot distribution patterns showed in (Figure 17) that seed oil content, vernolic acid and stearic acid distributed closer to each other at the same plane on the PC whereas linoleic acid and oleic acid were distributed closer to each other at the other side plane of the PC.

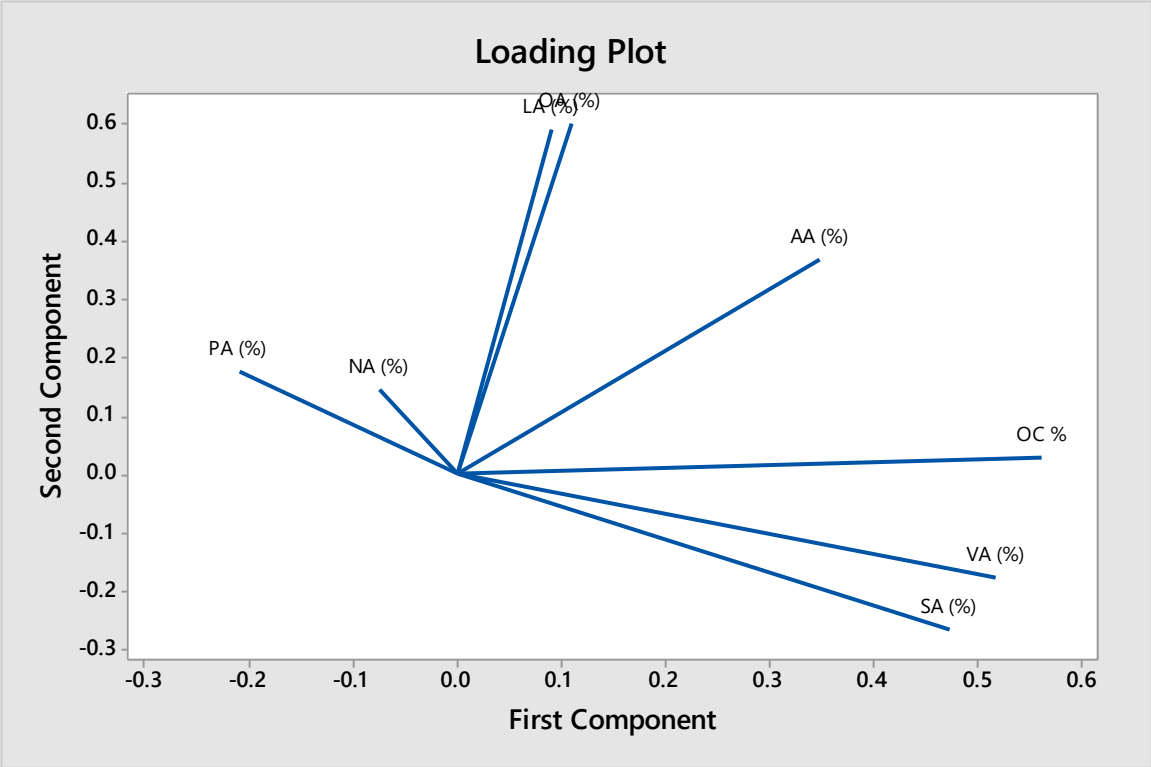

Figure 16. The two dimensional principal components of seed oil and fatty acid composition in *Vernonia galamensis* populations. Keys : VA = Vernolic acid; LA =linoleic acid; OA = oleic acid; PA = palmitic acid; SA = stearic acid; AA = arachidic acid; NA = linolenic acid; OC= seed oil content

The scree plot diagram showed that the main principal component analyses are extracted from the complicated component, and the first three PC (PC-1, PC-2 and PC-3) accounted for 72.0% of the total variations. Hence, the first principal component (PC-1) showed the maximum variation (28%), and then reduced gradually (Figure 17).

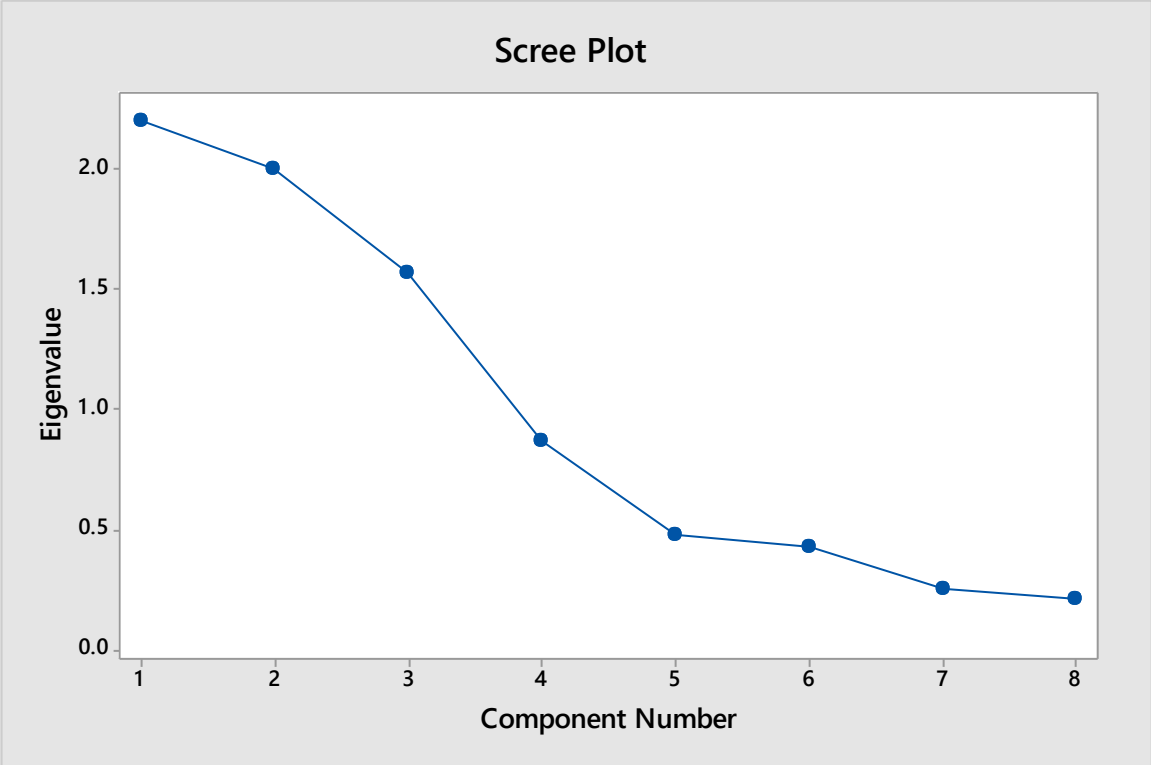

Figure 17. Scree plot of principal component analysis between eigenvalue and number of PC for seed oil content and fatty acids composition in *Vernonia galamensis*

5. DISCUSSION

5.1. Patterns of quantitative traits variation and their implications

Vernonia (*Vernonia galamensis*) is a plant species that contains naturally occurring epoxidized oils in its seeds. *V. galamensis* grows naturally in marginal areas at an elevation ranging from 700 to 2400 m above sea level in the eastern, southern and south-eastern parts of Ethiopia (Perdue *et al*., 1986, Baye, 2002). *V. galamensis* has a major potential as an industrial source of plant due to a high sources of natural epoxidized oil. Crops like soybean (*Glycine max* L.) and linseed (*Linum usitatissimum* L.) are used for epoxy oil production, but the oil still requires undergoing chemical epoxidation which is expensive (Thompson *et al*., 1994a).

Genetic diversity analysis of populations using agro-morphological traits is an initial step for crop improvement. In line with this, 80 accessions of *V. galamensis* were used for the agro-morphological traits, the mean square of the one-way ANOVA showed highly significant variation for most of the quantitative traits (such as, days to 50% heading, days to 50% flowering, plant height, days to 90% maturity, number of head per plant, number of seed per head and seed yield). These variations among accessions would provide a good opportunity for improvement through selection if the plant will be taken into cultivation. This finding is consistent with the previous report by Baye (2002) and Shimelis *et al*. (2008) in *V.galamensis*. Further, Baye and Becker (2005b) reported that genotype by location interaction mean squares were highly significant for all the characters. This indicates that field experiments at several environmental conditions are required for evaluating the diversity of plants.

The mean performances of all the quantitative characteristics exhibited significant variations, however, no statistically significant difference was observed for days to emergence at both Melka Werer Agriculture Research Center (MWARC) and Wondo Genet Agricultural Research Center (WGARC) sites. This might indicate that seeds have got similar germination capacity under the climatic conditions of both study sites. For example, phenotypic traits such as plant height, days to heading, days to flowering and days to

showed significant variation, indicating of higher genetic diversity occurring among the materials that can serve as the genetic resource for its improvement. Baye *et al*. (2001) reported that days to heading, days to flowering and days to maturity showed significant variation among *V. galamensis* accessions planted at three experimental sites (Alemaya, Babile and Harar). Bhardwaj *et al.* (2000) reported that morphological traits such as plant height, seed per head, days to maturity and seed yield per hectare manifested significant variation among the *V. galamensis* accessions.

Significant differences were also observed among seed yield and yield components. For example, five quantitative traits including number of heads per plant, number of seeds per head, seed yield per plant, seed weight per plot and seed yield per hectare scored more than three-folds of the minimum mean performance range. Such significant and wide range of genetic variation indicates the existence of large genetic variability among the tested accessions, and extrapolation to the crop as a whole could be used as a baseline information for further work in cultivation, conservation and breeding of the plant species. Baye *et al*. (2001), Shimelis *et al*. (2008) and Mebrahtu *et al.* (2009) reported a similar result but, with slight differences in the magnitude which could be attributed to differences in the number of samples studied and experimental sites used. The average seed yield per hectare (473.82 ±20.42 Kg/ha) was comparable with that reported by Mebrahtu *et al.*(2009) and slightly lower than that reported by Baye (2002) and Shimelis *et al*. (2008) in *V. galamensis*. Thompson *et al*. (1994b) reported that the number of head (capitula) per plant and the number of seed per head varied greatly among accession in *V.galamensis*.

Furthermore, the time from sowing to days of maturation were earlier, and the seed yield per hectare was remained by far lower at Melka Werer may be due to the hot and dry conditions at this location enhanced maturity to take place in a short period of time. This indicates that morphological traits are easily affected by environmental conditions.

However, the reverse is true for accessions grown at Wondo Genet Agricultural Research Center, prolonged period of time from the days to sowing to maturity. This may show that the moist and cool environment at Wondo Genet favors extended period of vegetative and

reproductive growth. Similarly, Baye *et al*. (2001) reported that the length of days to heading, days to flowering and days to maturity were shorter in dry environmental conditions than in the moist and cool environmental conditions, which indicates that the hot and dry conditions enhanced heading, flowering and maturity in a short period of time. Thompson *et al*. (1994c) reported that the mean number of days to flowering in the total populations was 78, which is earlier than the result of the present study. The climatic conditions of Wondo Genet are favorable for *V. galamensis* and thus resulted in a higher number of branches, heads per plant and seeds per head than that of populations grown at Melka Werer. Baye and Becker (2005a) grown *V.galamensis* at Babile and Haramaya and reported that the length of growing season was shorter at Babile, because the relatively hot conditions at Babile enhanced heading, flowering and maturity, while the cooler climate at Haramaya favored vegetative growth.

5.2. Genetic Variability and associations of traits: implication to improve the productivity of *Vernonia galamensis* in Ethiopia

The estimates of genetic variance components and coefficient of variation showed that the phenotypic variance (VP) and phenotypic coefficient of variance (PCV) were higher than their corresponding genotypic variance (VG) and genotypic coefficient of variance (GCV), respectively, for all the characters, suggesting the presence of environmental influence on the traits studied. The PCV and GCV values greater than 20% are regarded as high, values between 10-20% are regarded as medium and less than 10% are classified as low (Deshmukh *et al.,* 1986). Therefore, according to the classification high PCV and GCV values were recorded for number of branches, days to 90% maturity, number of heads per plant, number of seeds per head and seed yield per hectare, which suggests the possibility of improving this trait through selection.

On the other hand, the estimates of GCV and PCV were low in the present study for thousand seed weight (11.51%, 8.14%). Low values indicate the need for creation of variability either by hybridization or mutation followed by selection. Baye (2002) reported in *V.galamensis* that the phenotypic and genotypic variation were closer to each other for

most of the characters; however, the magnitude of the environment variation was relatively lower than the genotypic variation for most characters except days to 90% maturity and number of seed per head. Baye and Becker (2005a) reported that the estimated values of the PCV for different agro-morphological traits were higher than the corresponding GCV value which indicates all the traits were highly influenced by environmental factors.

Correlation coefficient is used to find out the degree and direction of the relationship between two or more variables. In our study, 50% heading showed high significant ($p < 0.05$) and positive correlations at both genotypic and phenotypic levels with the seed yield per hectare and other yield component traits. Moreover, number of head per plant revealed significant and positive genotypic and phenotypic correlations with plant height, seeds per head and seed yield per hectare. This association indicates that number of head per plant serves as the first selection criteria for the improvement of seed yield in *V. galamensis*. Baye and Becker (2005a) reported that strong and positive association existed among phenological traits such as days to heading, flowering and maturity, revealed the growth habit of *V. galamensis* to estimate date of maturity. In addition, the magnitudes of genotypic correlation of coefficient were higher than their corresponding phenotypic correlation of coefficient for all traits, except for days to emergence and branch length, which clearly indicate the presence of inherent association among most of the considered traits.

Contrary to the present finding, previous study on *Vernonia* by Shimelis *et al*. (2008) reported that negative correlations occur between the number of heads per plant with plant height and seed yield per hectare. The presence of significant positive correlation could be either due to the strong coupling linkage between the genes or was the result of pleiotropic genes that controlled these characters in the same direction.

5.3. Heritability of the traits and genetic advance for yield improvement

Heritability used to measure the phenotypic variance that is attributed to genetic causes, and it provides information on the extent to which a particular morphogenetic character can be transmitted to the offspring. Heritability was classified as low (below 40%), medium (40-59%), moderately high (60-79%) and very high (above 80%) as suggested by Singh (2001). Accordingly, very high heritability values were recorded for days to 50% flowering, number of head per plants and seed weight per plot (Mideksa *et al*., 2018). High heritability in broad sense values indicate that the characters under study are less influenced by environment in their expression. Days to 50% heading, number of seeds per head and thousand seed weight had medium heritability which indicates that improvement can be made through selection. In contrast, days to 90% maturity had low heritability which indicates greater role of environment on the expression of the traits. Thus, direct selection for these traits will be ineffective. Dudley and Moll (1969) reported that the characters having very high heritability indicate relative small contribution of the environment factors to the phenotype and selection for such characters could be fairly easy due to high genetic variation. Baye and Becker (2005a) reported that broad sense heritability of *V.galamensis* accessions were low for days to emergence and high for days to maturity. High heritability does not always indicate a high genetic gain, and thus heritability coupled with genetic advance and genotypic coefficient of variation should be considered in predicting the ultimate effect of selecting superior genotypes (Allard, 1991).

Genetic advance (GA) shows the degree of gain obtained in a character under a particular selection pressure. Johnson *et al.* (1955) categorized genetic advance as percent of mean (GAM) as; low (< 10%), moderate (10-20%), and high (>20%). Accordingly, values of genetic advance as a percentage of mean ranged from 2.25% for (days to 90% maturity) to 70.13% for (seed weight per plot). The traits for instance, plant height, number of heads per plant, number of seeds per head, seed weight per plot and seed yield per hectare had high GAM while moderate genetic advance as percent of mean was recorded for days to 50% heading, days to 50% flowering and branch length. Therefore, selection based on high and moderate genetic advance as percent of mean, result in the improvement of the genotypes

for the traits. Low estimates of genetic advance as percent mean was also noticed for days to 90% maturity, seed yield per plant and thousand seed weight. This indicates the characters governed by non-additive gene action and heterosis breeding will be useful.

Relatively high heritability coupled with high genetic advance as percentage of the mean was recorded for number of heads per plant and seed weight per plot, indicating that selection for these characters could be more effective. Furthermore, high heritability with high genetic advance could be an indication of additive gene action and expected genetic gain for these characters is high. Baye and Becker (2005a) reported about genetic advance as percent of mean varied from 1.3% for days to emergence to 39.5% for seed yield per plant. Jonhson *et al*. (1955) reported that high heritability estimates coupled with high GAM is usually more helpful in predicting gain under selection than heritability alone. Rosielle and Hamblin (1981) reported that heritability and genetic advance values may be concealed due to a greater genotype by environment under unfavorable condition. Ceccarelli (1994) on the contrary reported that there is no interrelationship between the type of the environment and the magnitude of heritability and genetic advance values. The magnitude of heritability and genetic advance values, according to the latter, is rather affected by the nature of the genetic material under consideration than the test environment.

5.4. Patterns of clustering within and among populations of *Vernonia galamensis*

The agro-morphological traits that exhibited considerable mean divergence are very important for deciding the purpose of further parents' selection. Falconer (1989) reported for selecting genotypes from a particular cluster, inter-cluster distance, cluster mean and percent performance needs to be taken into consideration. Accordingly, clustering was done based on the mean values of the agro-morphological traits, and 65% of the accessions were grouped under the first three clusters, indicates the existence of considerable differences in the mean values of different traits. In addition, accessions falling under clusters 1, 2, 3 and 7 had a maximum mean performance in traits such as seed yield per plant, seed yield per hectare, number of seeds per head, and number of heads per plant. Moreover, the mean

based clustering pattern indicates a wide diversity of accessions between different groups and their close relationship within a cluster. It is, therefore, logical to expect a higher genetic diversity from accessions of different clusters than within a cluster, which also important in breeding and parental selection.

Falconer (1981) reported that variation in origin (geographical separation), ancestral relationship, gene frequency and morphology are the probable sources of genetic diversity. Thus, plants differing in either one or more of these factors would differ by significant number of genes and hence, categorizing genotypes (accessions) on the bases of such variations into morphologically similar, more particularly genetically similar groups, is useful in selecting parents for crossing.

Likewise, genetic diversity is associated with geographical diversity, and they are not necessarily directly related. There was moderate tendency of association between geographical proximity and genetic diversity for accessions, woreda and population cluster analysis among the tested accessions. For instance, the distribution of accessions over entire regional cluster revealed that clusters I, III and V (West Gojjam and South Wollo, East and West Hararghe) were clustered according to their geographical proximity, indicating that the genetic diversity of these accessions is associated with their geographical regions. The similarity of accessions between neighboring regions may be due to the gene flow, migration of seed and pollen dispersal by wind.

However, accessions from Derashie population fell into cluster-V with East Hararghe and West Hararghe populations; this may be due to intentional mixing during seed collection. The pattern of distribution of the accessions from different regions of origin over different clusters was apparently random, showing that there was no clear association between geographic sources of origin and genetic diversity, and could be accessions from different sources of origin may share similar genetic background. Some accessions from the same places of origin fell into different clusters and vice versa. The clustering pattern with the local collections apparently revealed no definite relationship with origin as accessions from

the same origin fell into different clusters and no significant inter-cluster distance. It is believed that the genetic architecture of a population is the result of breeding system, gene flow within and between populations, isolation mechanisms and prolonged selection by various natural and artificial forces (Chandel and Joshi, 1983).

Moreover, geographic level clustering also shows some separate populations, for instance, accessions from East Showa (C-II) and others from Borena (C-IV) were grouped separately, may be due to unique in phenotypic characters which separate them from the other groups. Hence, accessions at regional level falling under clusters 1, 2 and 4, which had maximum deviation from the total mean, showed higher performance in traits such as, number of head per plant, number of branches, number of seed per plant, days to 50% flowering and seed yield per hectare. These traits are believed to be directly influencing yield and therefore helpful for selection of parents. Furthermore, regions of accessions under these clusters are recommended as superior groups and can be used for further improvement programs since superior lines in terms of genetic diversity and agronomical properties are important for selection. According to Singh and Chaudhary (1985) members in clusters with non-significant distance were assumed to have more close relationships with each other than they are with those in significantly distant clusters. On the other hand, clustering analysis is helpful for genetic recombination purpose and hence, genotypes separate by cluster with the largest statistical distance show the maximum divergence and are important sources for hybridization. Populations from areas far separated geographically with mountains and valleys, having complex environments and varied ecological conditions are normally expected to accumulate enormous genetic diversity as geographical separation with physical barriers and genetic barriers to inter-crossing are believed to give rise to genetic diversity among genetic materials (Chandel and Joshi, 1983).

The principal component analysis (PCA) reduces relatively a large series of data into a smaller number of components by looking for groups that have very strong inter-correlation in a set of variables. The first principal component is the largest contributor to the total variation in the population followed by subsequent components. Hence, traits with relatively

larger absolute values of eigenvector weights in PC1 had the largest contribution to the differentiation of the accessions into clusters, and assumed that traits with larger absolute values closer to unity within the first principal component influence the clustering more than those with lower absolute values closer to zero (Johnson and Wichern, 1988). In our study, four PCs exhibited more than one eigenvalue that showed 71.0% of the variation. PC-1 and PC-2 were accounted for 53.1% of the total variation. The PCs showed more than one eigenvalue, used as one of the criterion for selection of the diverse parents.

The variations in PC-1 and PC-2 were more influenced by traits such as days to 50% heading, days to 90% maturity, number of heads per plant, number of seeds per head, seed weight per plot and seed yield per hectare. Shimelis *et al*. (2013) reported that two major PCA which explaining 79.26% of the total variation. This study showed that after the principal component four (PC-4), it did not describe much variation. The scree graph revealed the maximum variation present in the first principal component (PC-1), so that selection of accessions from PC-1 would be useful. Shimelis *et al*. (2013) reported that PC-1 correlated with the number of head per plant and seed yield, and PC-2 showed positive association with head per plant but negative correlation with thousand seed weight in *V.galamensis*.

5.5. SSR-marker based genetic diversity of *Vernonia galamensis*

5.5.1. SSR markers polymorphism

Vernonia galamensis is characterized by its wide range of distribution (in the eastern, southern and south-eastern parts of Ethiopia), wind-pollination and its mating system allowing the species to develop a higher genetic diversity. The current study evaluated the genetic relatedness among a collection of 150 *V.galamensis* accessions using 20 polymorphic SSR markers. Gene diversity or expected heterozygosity (*He*) is the most widely used parameter to estimate genetic variability within populations (Nei, 1973). Surprisingly, the overall gene diversity or expected heterozygosity detected in this study was 0.48. Hence, heterozygosity used to determine the proportion of heterozygous individuals at a locus in populations, and the result showed differences between observed heterozygosity (*Ho*) and expected heterozygosity (*He)*. The *He* value was higher than *Ho* value (*Ho* = 0.15), the heterozygosity is expected in outcrossing species that maintain their heterozygosity through natural and selection pressure. Aikpokpodion *et al*. (2018) reported that the average gene diversity of genus *Vernonia* (*Vernonia amygdalina*) was 0.72 (range from 0.30 to 0.80), which was higher than the present study, implies significantly high genetic diversity. The SSR analysis showed considerable genetic diversity; a total of 79 alleles detected, varying from 2 to 6.

Moreover, the number of effective alleles (*Ne*) is an important parameter to measure genetic diversity in a finite population, with an average of 3.06 whereas an average number of alleles (Na) were 3.90. The variations on the number of effective alleles and number of alleles per locus were possibly due to differences in *V.galamensis* accessions, variability of markers used, and methods to detect and score PCR products. The number of alleles indicates the richness of the population. Since SSR are short tandem repeats, generally allele numbers of 2 to 6 per locus are considered good as seen in this study.

The high value of Wright's fixation index (ranged from 0.25 to 0.89) and Shannon's information index (ranged from 0.86 to 1.67) in the present study were other indications of

the presence of high genetic diversity in the *V.galamensis* accessions under consideration. Similarly, Nwakanma *et al*. (2018) reported higher Shannon-Weaver diversity index (0.76), implies that there is high polymorphism among genus *Vernonia*.

Polymorphic information content (PIC) values reflect the relative allelic polymorphism of a particular marker and their potential to differentiate the genotypes based on their genetic relationships. The PIC values are also a good indication of the usefulness of markers for linkage analysis when determining the inheritance between offspring and parental genotypes (Guo and Thompson, 1992). The informativeness level of markers based on PIC is usually defined as low (PIC < 0.25), medium (0.5 > PIC > 0.25) or high (PIC > 0.5) (Powell *et al*., 1996). Accordingly, the average percent polymorphism per population revealed in the present study across the 20 SSR loci (76%) is by far greater than that reported by Ramalema *et al*. (2010) in *V.galamensis* accessions (40%). This indicates that the majority of markers were able to distinguish differences among the studied *V. galamensis* accessions. Microsatellite markers such as Vg-002 and Vg-011 showed the highest polymorphism of 0.96, 0.93, respectively. Slightly lower PIC ranging from 0.27 to 0.78 in genus *Vernonia* was reported by Aikpokpodion *et al*. (2018). The diversity parameters showed that a high level of polymorphism among the 20 SSR markers, and the genetic variation within *V. galamensis* collection.

5.5.2. Genetic differentiation and gene flow

The analysis of molecular variance (AMOVA) partitioned the variance between groups and within groups, and, 67% of the variation was due to genetic variability within populations and 11% of the variation was due to among populations. On the other hand, 22% was due to variation among individuals within the same population. This indicates the existence of high genetic differentiation within the populations. The result is in-line with earlier reports (Nwakanma *et al.*, 2018). High genetic differentiation is very important within the population for creating a desirable heterotic group. Thus genetic diversity characterization is very important as it provides the basis for conservation, utilization, breeding and improvement of vernonia plant.

In addition, *F*st is important in discriminating genetic differentiation among the studied populations due to genetic structure. According to IPGRI and Cornell University (2003), *F*st value ranging from 0 to 0.05 is small in genetic differentiation, from 0.05 to 0.15 is moderate, and from 0.15 to 0.25 is large, and greater than 0.25 is very large genetic differentiation among populations in terms of allele frequencies. In-line with this, the extent of genetic differentiation among the ten populations in terms of allele frequencies measured was low (F_{st} = 0.15), which implies the presence of gene flow among populations in different regions. The pairwise populations of gene flow (N_m) values among the ten populations were estimated. N_m > 1 reveals little differentiation among populations, and migration is more important than genetic drift (McDermott and McDonald, 1993). In the present study, the highest N_m value was 13.32, indicates the presence of high genetic exchange or gene flow (in line with F_{st} value) and led to a low genetic differentiation between sub-populations. The largest number of migrants per generation was found between East Hararghe and West Hararghe whereas the lowest number of migrants per generation (N_m) value was recorded between South Wollo and Derashie, this assumption may be strengthened by the geographical proximity between the regions and or/collection sites for *V. galamensis* populations from Ethiopia. Wright (1965) reported that N_m value less than one indicate limited gene exchange among subpopulations.

5.5.3. Clustering and principal co-ordinates among *Vernonia galamensis*

All the 150 accessions were clustered into four (4) major clusters based on the allelic frequency. Each of the four clusters comprised individual plants from different zones (geographic regions). Cluster 1 comprised of 39 accessions, the second group contained 42 accessions, the third cluster composed of 15 accessions, and the fourth groups contained the largest and diverse group that contained 54 accessions. The clusters analysis revealed a poor clustering pattern as accessions from different populations (collection sites) clustered together. In other words, the clustering did not follow a clear pattern of geographic origins which may imply the presence of gene flow between populations/regions/ collection sites. *V. galamensis* was not observed growing under cultivation at any of the collection sites.

Principal coordinates analysis (PCoA) is a method to explore and visualize similarities or dissimilarities of data. In this study, the principal co-ordinate analysis was similar to the pattern of clustering. The majorities of samples were placed at the center of a two-dimensional coordinate plane and roughly forms four groups with the total variation of 30.04%. The study showed that accessions from different collection areas often clustered together. This, in turn, agrees with the results of the NJ dendrogram in that there was no unique clustering among accessions from the same population or collection areas. The presence of gene flow between and within populations or collection sites, accompanied by the prevalence of inter-gene pool introgressions or hybrids between the gene pools of origin and the seed dispersal by wind may be the most probable explanations behind the mixed clustering of accessions from different populations/collection areas together. Although UPGMA and PCoA analyses showed a certain level of populations clustering according to their geographical regions, the clustering pattern is weak, not clustered by their regions of origin.

5.5.4. Populations genetic structure and pattern of diversity in *Vernonia galamensis*

A plant population's genetic structure is determined by the interaction of processes such as gene flow, mutation, selection and mating strategy (Dudley, 1993; Schaal *et al*., 1998). Bayesian-model based approach to infer population structure, assign individuals to populations, as well as to identify migrants and admixture in individuals according to Gilbert *et al*. (2012) and Evanno *et al*. (2005) method. Accordingly, the 150 accessions were inferred into six (K = 6) with the help of STRUCTURE outputs. The patterns of population structure were certainly supported by the UPGMA and PcoA analyses. The structure analysis revealed that all individual plants analyzed have alleles originated from the six clusters, which supports the presence of a strong gene flow that led to poor population differentiation. Hence, accessions collected from the same region of origin did not group entirely together within a given major groups. There was a wide admixture in populations, which is the first report in *V.galamensis*.

The genetic variations across the ten populations were also investigated. Accordingly, the East Showa population had the highest average number of alleles. Likewise, populations from East Showa and East Hararghe again had the highest number of effective alleles, heterozygosity and Shannon's diversity index. This result indicates that East Showa and East Hararghe may be the most important populations with high genetic diversity, and were the potentials accessions in future *V. galamensis* improvement and genetic conservation endeavor in Ethiopia. So the researchers, breeders and other concerned bodies should take this information into consideration for cultivation of this species. Hence, these populations may serve as potential sources of new genetic variation of important traits (hotspots for in-situ and ex-situ conservation), and used in breeding programs as potential parental sources. Since, in plant natural populations, spatial distribution of genetic variation is primarily determined by seed and pollen dispersal, habitat distribution, micro-environmental selection and genetic drift (Guo and Thompson, 1992).

Generally, the present study showed that there was ample allelic diversity among the *V. galamensis* accessions studied, based on the values recorded for all the SSR markers used in the study (like heterozygosity, Shannon's diversity index, polymorphic information content, etc). Most of the values identified for the SSR markers are important baseline information for future *V. galamensis* cultivation, breeding/genetic resource conservation endeavors in Ethiopia. In line with these, cluster analyses, PCoA and population structure together showed the presence of gene flow among the accessions of regional proximity which should be carefully considered in future genotype sampling and collection regarding biodiversity evaluation and conservation in the *V. galamensis*.

5.6. Seed oil content and fatty acid composition, and their implications in breeding

5.6.1. Seed oil content and fatty acid variation

Many oilseed crop species are potential sources of oils that contain desirable fatty acids such as linoleic acid (*Madia sativa* Molina), linolenic acid (*Lepidium sativum* L. and *Camelina sativa* (L.) Crtz.), and epoxy oils (*Euphorbia lagascae* Sprengel and *Vernonia galamensis* (Cass.) Less.). For example, an epoxy fatty acid known as vernolic acid constitutes up to 30-90% of total fatty acids in the seeds of *Vernonia galamensis, Euphorbia lagascae, Stokesia laevis, Crepis palaestina* and *Bernardia pulchella* (Perdue, 1989; Pascual and Correal, 1992; Thompson *et al*., 1994b).

In the present study, oil content and different fatty acid composition showed a high significant variation ($p < 0.05$) among the 15 selected accessions. The highest seed oil content was recorded for VgEt-094 (36.7%) with an average of (33.2%) while the lowest was recorded for VgEt-032 (31.2%). The difference in seed oil content was significant between VgEt-032 and VgEt-094 and this is probably due to genotypic differences between the accessions. The observed variation in seed oil content was in agreement with the observations from other similar studies (Baye and Becker, 2004; Ramalema *et al*., 2010; Shimelis *et al*., 2011). Baye *et al*. (2001) in his studies of *V. galamensis* accessions sown at Haramaya, Harar and Babile, reported that seed oil content averaged 38.7, 39.5 and 34.7% for each respective location. Mebrahtu *et al.* (2009) in other studies reported that seed oil content ranged from 20.6 to 29.7% with an average of 25.0. Ramalema *et al.* (2010) also reported seed oil content averaged 27.2% and varied from 22.4 to 33.1%. Similarly, Bhardwaj *et al.* (2000) reported that seed oil content ranges from 32.0-39.0%, almost similar to that of the present study. The present study varies from previous research reports may be due to differs in number of accessions, growth locations, and the accessions used in the study.

The fatty acid composition profile showed significant variations, for instance, vernolic acid (VA) content ranged from 71.5% to 78.4% with an average of 73.7% in the present study. Baye and Becker (2005b) reported that vernolic acid content showed significant variation, which ranged from 34.0-87.0% with an average of 74.3%, and higher than the result of present study. Vernolic acid variation existed in seeds of *Euphorbia lagascae, Stokesia laevis*, *Crepis palaestina* and *Bernardia pulchella* (Perdue, 1989; Pascual and Correal, 1992; Thompson *et al*., 1994a). Epoxy oil also occurred in crops like soybean (*Glycine max* L.) and linseed (*Linum usitatissimum* L.) but the oil still requires undergoing chemical epoxidation (not as trivernolin) which is expensive (Thompson *et al*., 1994a).

Likewise, there was considerable variation in the proportion of other fatty acids such as linoleic, oleic, palmitic, linolenic and arachidic fatty acids in the present study. Baye and Becker (2005a) and Ramalema *et al.* (2010) had reported similar result in *V. galamensis*. For instance, the following fatty acids components showed significant variation, the average linoleic, palmitic and stearic fatty acids are 14.31%, 3.18% and 2.72%, respectively. Trace amounts (less than 2%) of arachidic and linolenic acids in descending order, were also observed. Emphasis will be given to the accessions that gave the highest vernolic acid content. The fatty acids like linoleic and oleic fatty acids were existed in the seed of sun flower (Osorio *et al*., 1995), *Brassica carinata* (Barro *et al*., 2001) and sesame (Savant and Konthekar, 2011).

The relationship between 1000-seed weight (g) and seed oil content was estimated. For example, accession VgEt-094 used in the present study scored the highest 1000-seed weight (3.5g), which also produced the highest oil content (36.7%). This correlation shows that the accessions with relatively heavier seed weight had high amount of oil content. Therefore, selection for increased thousand seed weight is the recommended morphological selection criteria for increased oil content in *V. galamensis*. Contrary to this study, Mebrahtu *et al*. (2009) reported that the accessions with bigger seed size tended to have lower oil content, and the genotypes with smaller seed size had higher oil content.

5.6.2. Association and variance components of seed oil content and fatty acid composition

Significant positive and negative correlations were existed between the seed oil content and various fatty acids, and between fatty acids. The seed oil content is positively correlated with vernolic acid. Furthermore, vernolic acid is positively correlated with linolenic acid, but negatively correlated with oleic acid. A significant, strong positive association was found between linoleic acid and oleic acid. In addition, arachidic acid again positively associated with linoleic acid. Similar report was found in *V. galamensis* (Shimelis and Hugo, 2011). This indicates that improvement for seed oil content can be best achieved through direct selection for high vernolic fatty acid. Bhardwaj *et al.* (2007) reported that the concentration of vernolic fatty acid was positively correlated with oil content but negatively correlated with linoleic, oleic, palmitic, linolenic and arachidic fatty acids. This indicating that there was poor associations between oil content and other fatty acids.

According to Baye and Becker (2005a), the fatty acid biosynthesis pathway, oleic acid is the precursor for linoleic acid, and stearic acid. In the present study, an increase in linoleic acid leads to the corresponding increase in oleic acid. But the increase/decrease in all the other fatty acids leads a corresponding decrease/increase in vernolic acid content, except that of linolenic fatty acid. Most other oilseed plants such as sunflower (Seiler and Brothers, 1999), *Brassica napus* (Mandal *et al*., 2002), soybean, peanut, sesame (Weiss, 1983), have significant negative correlations among stearic, oleic and linoleic acids. These differences perhaps indicate differences in the different species in the activities of the different steps in the fatty acid biosynthesis pathways (Fern´andez-Mart´ınez *et al*., 1993; Barro *et al*., 2001).

In addition, the Pearson correlation coefficient showed a strong and positive association between vernolic and stearic fatty acids. Thousand seed weight was positively associated with vernolic and arachidic fatty acids, and negatively correlated with linoleic fatty acid. Thus, increased oil content in vernonia can be achieved through simultaneous selection of lines that display high seed weight. Similarly, Baye and Becker (2005a) reported that thousand seed weight showed strong positive associations with vernolic fatty acid and oil

yield, but strong and negative associations with palmitic, stearic, linolenic and oleic fatty acids. Shimelis *et al*. (2008) reported that a significant positive correlation was found between thousand seed weight and seed oil content. Shimelis and Hugo (2011) reported that a significant positive association was found between seed oil content and thousand seed weight. These indicate that accessions with higher thousand seed weight had high performance in oil content.

6. CONCLUSIONS AND RECOMMENDATIONS

6.1. Conclusions

The study revealed that there was significant genetic variability for all the quantitative traits ($p < 0.05$), except that of days to emergence among *V. galamensis* accessions. Accession VgEt-094 recorded the highest seed yield (1104 Kg ha^{-1}) followed by accessions VgEt-078 (1064 Kg ha^{-1}) and VgEt-141 (1058.7 Kg ha^{-1}) at Wondo Genet Agricultural Research Center, due to a good climatic condition at Wondo Genet promotes vegetative and reproductive growth of *V.galamensis*, resulting in high yield performance. Seed yield was positively and significantly associated with number of heads per plant, number of seeds per head and seed yield per plant, indicating the importance of these traits for seed yield improvement.

The magnitude of phenotypic coefficient of variance (PCV) was higher than their corresponding genotypic coefficient of variance (GCV), for all the characters, suggesting the presence of environmental influence on the traits studied. Relatively high heritability coupled with high genetic advance as percentage of the mean was recorded for number of heads per plant and seed weight per plot, indicating that selection for these characters could be more effective. Furthermore, high heritability with high genetic advance for seed weights per plot, indicating that the presence of additive gene action and expected genetic gain for these characters was high. The result of regression coefficient analysis revealed that the number of heads per plant showed the highest positive direct effects (0.72) and high positive correlation ($r = 0.86$) with seed yield per hectare.

In addition, four PCs exhibited more than one eigenvalue that showed 71.0% of the variation. PC-1 and PC-2 were accounted for 53.1% of the total variation. The PCs showed more than one eigenvalue, used as one of the criterion for selection of the diverse parents.

The SSR analysis showed considerable genetic diversity; a total of 79 alleles detected, varying from 2 to 6. The polymorphic information content of the present study revealed an

average of 76% across 20 SSR loci, a good indication of the usefulness of markers for genetic diversity analysis. The clusters analysis revealed a total of four clustering, and accessions from different populations (collection sites) clustered together, indicating that poor clustering pattern was observed. Further, there was a wide admixture in populations, which is the first report in *V.galamensis*. This may imply the presence of gene flow between germplasm/ regions/ collection sites. In-line with these, cluster analyses, PCoA and population structure together showed the presence of gene flow among the accessions of regional proximity which should be carefully considered in future accessions sampling and collection regarding biodiversity evaluation and conservation in *V. galamensis*.

The oil content and different fatty acid composition showed a high significant variation. The highest seed oil content was recorded for VgEt-094 (36.7%) while the lowest was recorded for VgEt-032 (31.2%), due to the genotypic differences between accessions. VgEt-094 accession also recorded the highest seed yield per hectare at Wondo Genet Agricultural Research Center, Vernolic acid (VA) was the predominant fatty acid, and its values ranging from 71.5% to 78.4%. Variation in seed oil content and fatty acid composition could be used as the baseline for appropriate selection, conservation and improvement of *V. galamensis*.

Generally, the present study showed that there was ample genetic, phenotypic and oil content variability among the *V. galamensis* accessions studied. Most of the values of variability detected are important baseline information for future *V. galamensis* cultivation, breeding and conservation in Ethiopia.

6.2. Recommendations

The accessions taken under consideration showed wide range of variation for quantitative characters. Since, it is a potential industrial plant, it will be recommended to be included in the national breeding system of the country.

The present SSR markers study (the first report) showed that accessions collected from East Showa and East Hararghe had maximum genetic diversity, may serve as potential parent sources for *V. galamensis* improvement and genetic conservation endeavor in Ethiopia. So the researchers, breeders and other concerned bodies should take this information into consideration for cultivation, improvement and conservation of this species.

Concerning the seed oil content and fatty acids composition analyses report, accession VgEt-094 scored the highest seed oil content and vernolic acid, can be used as the promising candidate to check the oil yield and fatty acid composition. However, only few accessions were used in this study, further detailed work and investigations may be needed using more germplasm from multi-locations of the country.

7. REFERENCES

Abdel-Mawgood, A. L. (2012). DNA based techniques for studying genetic diversity, genetic diversity in microorganisms, DOI: 10.5772/33509. Available from: https://www.intechopen.com/books/genetic-diversity-in-microorganisms/dna-based techniques-for-studying-genetic-diversity.

Adegbite, A. E. and Ayodele, M. S. (2004). Cytogenetic and phylogenetic studies in the genus *Vernonia* SCHREB. *Feddes Reportorium*, **115** (7): 513-518.

Aikpokpodion, P. O., Abebe, J. and IGWE, D. (2018). Genetic diversity in *Vernonia amygdalina* Delile accessions revealed by random amplified polymorphic DNAs (RAPDs). *Journal of Biotechnology*, **99** (2): 147–152.

Allard, R. W. (1991). *Principles of plant breeding*, 2nd ed. New York: John Wiley and Sons.

Altaye, T., Mekibib, F. and Woyessa, B. (2016). Genetic gain in grain yield potential and associated traits of sesame (*Sesamum indicum* L.). *Journal of Biology, Agriculture and Healthcare*, **6** (19):104-119.

Andrew, R. (2018). FigTree: *Tree figure drawing tool*, Version 1.4.4. Institute of Evolutionary Biology, University of Edinburgh.

Anderson, J., Churchill, G., Autrique, J., Tanksley, S., Sorrells, M. (1993). Optimizing parental selection for genetic linkage maps. *Genome,* **36**: 181–186.

Anumalla, M., Roychowdhury, R., Geda, C. K., Mazid, M. and Rathoure, A. K. (2015). Utilization of plant genetic resources and diversity analysis tools for sustainable crop improvement with special emphasis on rice. *International Journal of Advanced Research*, **3** (3): 1155-1175.

Barro, F., Fernandez-Escobar, J., Vega, M. D. and Martian, A. (2001). Doubled haploid lines of *Brassica carinata* with Modified erucic acid content through mutagenesis by EMS treatment of isolated microspores. *Plant Breeding*, **120**: 262-264.

Baye, T. (2002). Genotypic and phenotypic variability in *Vernonia galamensis* germplasm collected from eastern Ethiopia. *The Journal of Agricultural Science*, **139**.

Baye, T. and Becker, H. C. (2004). Natural outcrossing rate in *Vernonia galamensis*. *Plant Breeding*, **123**: 398-399.

Baye, T. and Becker, H.C (2005a). Genetic variability and interrelationship of traits in the industrial oil crop *Vernonia galamensis*. *Euphytica,* **142**: 119–129.

Baye, T. and Becker, H.C. (2005b). Exploration of *Vernonia galamensis* in Ethiopia, and Variation in Fatty Acid Composition of Seed Oil. *Genetic Resources and Crop Evolution*, **52**: 805–811.

Baye, T. Becker, H.C. and Witzke-Ehbrecht, S. V. (2005). *Vernonia galamensis*, a natural source of epoxy oil: variation in fatty acid composition of seed and leaf lipids. *Industrial Crops and Products* **21**: 257–261.

Baye, T. Kebede, H. and Belete, K. (2001). Agronomic evaluation of *Vernonia galamensis* germplasm collected from eastern Ethiopia. *Industrial Crops and Products*. **14** (3):179-190.

Belayneh, H. (1986). An overview of oil crops research in Ethiopia. In: J. M. M. Engles (cd.), The conservation and utilization of Ethiopian germplasm. Proceedings of an international symposium held in Addis Ababa. pp 212-221.

Bhardwaj, H. L., Hamama, A. A. and Dierig, D. A. (2007). Fatty acids in vernonia produced in the mid-Atlantic region of the United States. *JAOCS, Journal of the American Oil Chemists' Society*, **84**:393–397.

Bhardwaj, H. L., Hamama, A. A., Rangappa, M. and Dierig, D. A. (2000). Vernonia oilseed production in the mid-Atlantic region of the United States. *Industrial Crops and Products*, **12** (2): 119-124.

Bhandari, H. R,. Bhanu, A. N,. Srivastava, K,. Singh, M. N. and Shreya, A.H. (2017). Assessment of genetic diversity in crop plants - An overview. *Adv. Plants. Agric. Res*. **7**(3): 00255. DOI: 10.15406/apar.2017.07.00255.

Birhanu Kumbi (2009). Chromosome study of six *Vernonia* species (asteraceae) from central Ethiopia. MSc.Thesis. Addis Ababa University, Addis Ababa, 59 pp.

Burow, G., Franks, C. D., Xin, Z. and Burke, J.J. (2012). Genetic diversity in a collection of Chinese sorghum landraces assessed by microsatellites. *American Journal of Plant Sciences*, DOI:10.4236/ajps.2012.312210.

Burton, G. W. (1952). Quantitative inheritance in grasses. Proceedings of the 6th International Grassland Congress, **1**: 277–283.

Ceccarelli, S. (1994). Specific adaptation and breeding for marginal conditions. *Euphytica* **77**:205-219.

Chandel, K. P. S. and Joshi, B. S. (1983). Multivariate analysis in green seeded pea. *Indian Journal Agricultural Sciences*, **53**:198-200.

Dansi, A., Vodouhè, R., Azokpota, P., Yedomonhan, H., Assogba, P., Adjatin, A., Loko, L., Dossou-Aminon, I. and Akpagana, K. (2012). Diversity of the neglected and underutilized crop species of importance in Benin. *The Scientific World Journal*, 1-19; doi:10.1100/*2012*/932947.

Dematteis, M. (1998). Chromosome study of some *Vernonia* species (Asteraceae*). Genet. Mol. Biol.* **21**: 381-385.

Deshmukh, S. N., Basu, M. S., Reddy, P. S. (1986). Genetic variability, character association and path coefficient analysis of quantitative traits in *Virginia bunch* varieties of groundnut. *Indian J. Agr. Sci.* **56**: 816-821.

Dewey, D. R. and Lu, K. H. (1959). A Correlation and path-coefficient analysis of components of crested wheatgrass seed production. *Agronomy Journal*. doi: 10.2134/agronj1959.00021962005100090002x.

Doyle, J. J. and Doyle, J. L. (1990). Isolation of plant DNA from fresh tissue. *Focus*, **12**:13-15.

Dudley, J. W. (1993). Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. *Crop Sci*. **33**: 660–668.

Dudley, J. W. and. Moll, R. H (1969). Interpretation and use of estimates of heritability and genetic variances in plant breeding. *Crop Sci*. **9** (3): 257-262.

Earl, D. A. and Von Holdt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Cons. Genet. Reso*. **4**: 359-361.

Evanno, G, Regnaut, S, Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol*. **14**: 2611-2620.

Falconer, D. S. (1981). *Introduction to quantitative genetics*, Ed. 2. Longmans Green, London/New York.

Falconer, D. S. (1986). *Introduction to quantitative genetics*, Longman Press, London.

Falconer, D.S. (1989). *Introduction to quantitative genetics.* 3rd edn., Longman, London.

Favi, F., Cantrell, C. L., Mebrahtu, T. and Kraemer, M. E. (2008). Leaf peltate glandular trichomes of *Vernonia galamensis* ssp. *galamensis* var. *ethiopica* Gilbert: development, ultrastructure, and chemical composition. *Int. J. Plant Sci*.https://doi.org/10.1086/533598.

Fern´andez-Mart´ınez, J., Rio, M. and Haro, A. (1993). Survey of safflower (*Carthamus tinctorius* L.) germplasm for variants in fatty acid composition and other seed characters. *Euphytica*, **69**:115–122.

Fehr, W. R. (1987). *Principles of cultivar development*, Vol.1. New York: Macmillan.

Folch, J., Lees, M. and Stanley, H. S. (1957). Simple method for the isolation and purification of total lipids from animal tissues. *J. Biol. Chem*. **226**:497-509.

Fu, B.Y. (2015). Understanding crop genetic diversity under modern plant breeding. *Theor. Appl. Genet*. **128**:2131–2142.

Getachew Ali (2008). Genetic diversity of agro-morphological traits, self-compatibility and chromosomal cytology in Ethiopian populations of *Vernonia galamensis* (L). MSc. Thesis. Addis Ababa University, Addis Ababa, 61 pp.

Gilbert, M. G. (1986). East African *Vernonieae* (Compositae), a revision of the *Vernonia galamensis* complex. Kew Bulletin, **41** (1): 19-35.

Gilbert, J. E., Lewis, R. V., Wilkinson, M. J and Caligari, P.D.S. (1999). Developing an appropriate strategy to assess genetic variability in plant germplasm collections. *Theor. Appl. Genet*. **98**:1125–1131.

Gilbert, K.J., Andrew, R.L., Bock, D.G., Franklin, M.T., Kane, N.C. *et al*. (2012). Recommendations for utilizing and reporting population genetic analyses: the reproducibility of genetic clustering using the program STRUCTURE. *Mol. Ecol*. **21**: 4925–30.

Gill, B. S. and Singhal, V. K. (1998). Chromosomes, chromosomal techniques and chromosomal evolution. **In**: *Forest Genetics and Tree Breeding*, pp. 168-193, (Mandal, A. and Gibson, G., eds). CBB Publishers and Distributors.

Gomez, K. A. and Gomez, A. (1984). Statistical Procedures for Agricultural Research, 2nd edn. pp 680. John Wiley & Sons. New York

Govindaraj, M., Vetriventhan, M. and Srinivasan, M. (2015). Importance of genetic diversity assessment in crop plants and its recent advances. *Gene. R. Inter.* DOI: 10.1155/*2015*/431487.

GraphPad Software, Inc (2018) Accessed 29 June 2018. http:// www.graphpad.com/faq/viewfaq.cfm?faq=8000.

Gunstone, F. D. (2001). *Basic oleochemical products and new industrial oils*. In: F. D. Gunstone and R. J. Hamilton (eds.), Oleochemical manufacture and applications. Sheffield Acadcmic Press, pp 1-22.

Guo, S. and Thompson, E. A. (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics,* **48**: 361–372.

Hadebe, T. S., Modi, A. T. and Shimelis, H. A. (2017). Determination of optimum ethyl methane sulfonate conditions for chemical mutagenesis of selected *Vernonia* (*Centrapalus pauciflorus*) accessions. *South African Journal of Plant and Soil*, 1-7, DOI: 10.1080/02571862.2017.1317851.

Hosseinzadeh-Colagar, A., Haghighatnia, M.J., Zahra Amiri, Z., Mohadjerani, M. and Tafrih, M. (2016).Microsatellite (SSR) amplification by PCR usually led to polymorphic bands: Evidence which shows replication slippage occurs in extend or nascent DNA strands. *Molecular Biology Research Communications,* **5**(3):159-166.

Hunter,G. E. (1964). Chromosome number in *Vernonia*, section *Lepidaploa*: Subsection *Paniculatae* verae. *The Southwestern Naturalist*, **9** (4): 238-244.

Hundesa, N. and Mekonnen, M. (2017). Evaluation of fungicides for management of powdery mildew (*Erysiphecichoracearum)* on Vernonia (*Vernonia Galamensis*). *International Journal of Research Studies in Agricultural Sciences,* **3**(3): 19-23.

Idrees, M. and Irshad, M. (2014).Molecular markers in plants for analysis of genetic diversity*: European Academic Research*, **1** (1).

IPGRI and Cornell University. (2003). Genetic diversity analysis with molecular data: Learning module, pp.1-71.

Johnson, H. W., Robinson, H. F. and Comstock, R. E. (1955). Estimates of genetic and environmental variability in soybeans. *Agronomy Journal*, **47**, 314-318.

Johnson, R.A. and. Wichern, D.W. (1988). *Applied multivariate statistical analysis*. Prentice-Hall.

Kisua, J., Mwikamba, K., Makobe, M. and Muigai, A. (2015). Genetic diversity of sweet and grain sorghum populations using phenotypic markers. *Int. J. Biosci*.**6** (9): 34-46.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A. and Mayrose, I. (2015). CLUMPAK: A program for identifying clustering modes and packaging population structure inferences across k. *Mole. Ecol. Res*. **15** (5): 1179-1191.

Kumar, A., Misra, S. C., Singh, Y. P. and Chauhan, B. P. (1985). Variability and correlation studies in triticale. *Journal of the Maharashtra Agri. University*, **10**: 273-275.

Mahalanobis, P. C. (1936). Generalized distance in statistics, *Proceedings National Institute of Science, India*, **12**: 49–55.

Mabhaudhi, T., Vimbayi, G. P., Chibarabada, C. T. and Albert T. Modi, A.T. (2017). Developing a roadmap for improving neglected and underutilized crops: A Case study of South Africa. *Frontiers in Plant Science*, doi: 10.3389/fpls.2017.02143.

Mandal, S., Yadav, S., Singh, R., Begum, G., Suneja, P. and Singh, M. (2002). Correlation studies on oil content and fatty acid profile of some cruciferous species. *Genet. Resour. Crop. Evol.* **49**: 551-556.

Mayes, S., Massawe, F.J., Alderson, P.G., Roberts, J.A., Azam-Ali, S.N., Hermann, M. (2012). The potential for underutilized crops to improve security of food production. *Journal of Experimental Botany*, **63** (3): 1075–1079. doi: 10.1093/jxb/err396.

Mbugua, M. N., Yusuf, A. O., Gitu, P. M. and Bhatt, B. M. (2007). Conversion of *Vernonia galamensis* oil to pyridinyl vernolamides and their antimicrobial activities. *Bull. Chem. Soc. Ethiop*.**21** (1): 103-110.

McDermott, J. M., and McDonald, B. A. (1993). Gene flow in plant pathosystems. *Annu. Rev. Phytopathol*. **31**: 53–373. doi: 10.1146/annurev.py.31.090193.002033.

Mebrahtu, T., Gebremariam, T., Kidane, A. and Araia, W. (2009). Performance of *Vernonia galamensis* as a potential and viable industrial oil plant in Eritrea: yield and oil content. *African Journal of Biotechnology*, **8** (4):635-640.

Mideksa, A., Tesfaye, K. and Dagne, K. (2018). *Centrapalus pauciflorus* (Willd.) H. Rob. neglected potential oil crop of Ethiopia, agro-morphological characterization. *Genet. Resour. Crop. Evol*. **66** (1); https://doi.org/10.1007/s10722-018-0719-1.

Miller, P. A., Williams, J.C., Robinson, H. F. and Comstock, R. E. (1958). Estimates of genotypic and environmental variances and covariances in upland cotton and their implications in selection1. *Agronomy Journal*, **50**: 126.

Mills, D. (1999). Response of *Vernonia galamensis* grown in the greenhouse to liquid fertilization. *Journal of Plant Nutrition*, **22** (11): 1665-1678.

Minitab, (2013). Minitab Statistical Package for Windows. Minitab Inc., Sales (USA).

Mohamed, A. I., Mebrahtu, T. and Andebrhan, T. (1999). Variability in oil and vernolic acid contents in the new *Vernonia galamensis* collection from East Africa. p. 272-274 *J. Janick* (ed.), ASHS Press, Alexandria, VA.

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

Narina, S. S., Andebrhana, T., Mohamed, A. I., Bhardwaj, H. L. (2012). Functional classification of ESTs from vernonia (*Vernonia galamensis* L.) cDNA library. *Industrial Crops and Products*, **36**: 370-375.

Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genet*. **89**: 583–90.

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proce. Nat. Acad. Sci. USA*, **70**: 3321-3323.

Nwakanma, N. M., Oboh, B. O., Adekoya, K. O. and Ogunkanmi, L. A. (2018). Genetic diversity of *Vernonia* as revealed by random amplified polymorphic DNA (RAPD) markers. *Egypt. Acad. J. Biolog. Sci*. **9** (1): 27-38. ISSN 2090-3812.

Osorio, J., Fernandez-Martinez, J., Mancha, M. and Garces, R. (1995). Mutant sunflowers with high concentration of saturated fatty acids in the oil. *Crop Science*, **35**: 739-742.

Pascual, M. J. and Correal, E. (1992). Mutation studies of an oilseed spurge rich in vernolic acid. *Crop Science*, **32**: 95-98.

Panse, V. G. (1957). Genetics of quantitative characters in relation to plant breeding. *Ind. J. genet*. **17**: 318-328.

Prasad, S. R., Prakash, R. Sharma, C. M. and Haque, M. F. (1981). Genotypic and phenotypic variability in quantitative characters in oat. *Indian J. Agric. Sci*. **54**: 480-482.

Peakall, P. R. and Smouse, P. P. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research– an update. *Bioin.* **28**: 2537–2539.

Perdue, R. E., Carlson, K. D. and Gilbert, M. G. (1986). *Vernonia galamensis*, Potential new crop source of epoxy acid. *Economic Botany*, **40** (1): 54-68.

Perdue, R. E.(1988). Systematic Botany in the Development of *Vernonia galamensis* as a New Industrial Oilseed Crop for the Semi-arid Tropics. Symbiotic Botany Upsal. **28**: 125-135.

Perdue, R. E. (1989). Vernonia - Bursting with Potential. *Agricultural Engeneering*, **70**: 1113.

Perrier, X. and Jacquemoud-Collet, J. P. (2006). DARwin software. http://darwin.cirad.fr/ darwin.

Pritchard, J. K., Stephens, M. and Donnelly, P. (2000). Inference of population structure using multi locus genotype data. *Genetics*, **155**: 945–959.

Powell, W., Machray, G. C. and Provan, J. (1996). Polymorphism revealed by simple sequence repeats. *Trends in plant science*, **1**: 215–222.

Ramalema, S. P., Shimelis, H., Ncube, I., Kunert, K. K. and Mashela, P.W. (2010. Genetic analysis among selected vernonia lines through seed oil content, fatty acids and RAPD DNA markers. *African Journal of Biotechnology*, **8** (4): 635-640.

Robinson, H. (1999). Revisions in paleotropical *Vernonieae* (*Asteraceae*). *Proc. Biol. Soc. Wash.***112**:220–246.

Rosielle, A. A. and Hamblin, J. (1981). Theoretical aspects of selection for yield in stress and non-stress environments. *Crop Science* **21**:943-946.

SAS Institute Inc. (2008) SAS, Version 9.2, SAS Institute Inc., Cary, NC

Savant, K. D. and Kothekar, V. S. (2011). Induction of Variability in Fatty Acid Profile in Sesame (*Sesamum indicum* L.). *Journal of Phytology*, **3**: 01-03.

Saxena, B., Kaur, R. and Bhardwaj, S.V. (2011). Assessment of genetic diversity in cabbage cultivars using RAPD and SSR markers.*J. Crop Sci. Biotech*. **14** (3) : 191-196.

Seiler, G. J. and Brothers, M. E. (1999). Oil concentration and fatty acid composition of achenes of *Helianthus* species (Asteraceae) from Canada. *Econ. Bot.* **53**: 273-280.

Schaal, B., Hayworth, D., Olsen, K. M., Rauscher, J. and Smith, W. (1998). Phylogeographic studies in plants: problems and prospects. *Molecular Ecology*, **7**: 465-474.

Schubert, I. (2007). Chromosome evolution. *Current Opinion in Plant Biology*, **10**: 109-115.

Shimelis, H. and Hugo, A. (2011). Determination of selection criteria for seed yield and seed oil content in Vernonia (*Vernonia galamensis* variety ethiopica). *Industrial Crops and Products*, **33** (2):436-439.

Shimelis, H., Mashela, P. and Hugo, A. (2011). Genotype by environment interaction of seed and oil yield in vernonia (*Vernonia galamensis* variety *ethiopica*). *Industrial Crops and Products*, **33**: 756–760.

Shimelis, H., Mashela, P.W. and Hugo, A. (2008). Performance of *Vernonia* as an alternative industrial oil crop in Limpopo Province of South Africa. *Crop Science*, **48**: 236.

Shimelis, H. A., Mashela, P. W. and Hugo, A. (2013). Principal agronomic and seed oil traits in the industrial oil crop vernonia (*Centrapalus pauciflorus* var. *ethiopica*). *South African Journal of Plant and Soil*, **30**: 131–137.

Shimelis, H., and Gwata, E. T. (2013). Evaluation of the Agronomic performance of Vernonia (*Vernonia galamensis*) germplasm. DOI: 10.5772/55016.

Stace, C. A. (2000). Cytology and cytogenetics as a fundamental taxonomic resource for the 20th and 2st centuries. *Taxon* **49**: 451-477.

Singh, B. D. (2001). *Plant Breeding: Principles and methods*. Kalyani Publishers, New Delhi, India, Pages: 896.

Singh, R. K. and Chaudhary, B. D. (1985). *Biometrical methods in quantitative genetic analysis*. Kalyani Publishers, New Delhi, pp.13-55.

Thompson, A. E., Dierig, D. A., Johnson, E. R., Dahlquist, G. H. and Kleiman, R. (1994a). Germplasm development of *Vernonia galamensis* as a new industrial oilseed crop. *Industrial Crops and Products*, **3**: 185-200.

Thompson, A. E., Dierig, D. A. and Kleiman, R. (1994b). Characterization of *Vernonia galamensis* germplasm for seed oil content, fatty acid composition, seed weight, and chromosome number. *Industrial Crops and Products*, **2** (4): 299-305.

Thompson, A. E., Dierig, D. A. and Kleiman, R. (1994c). Variation in *Vernonia galamensis* flowering characteristics, seed oil and vernolic acid contents. *Industrial Crops and Products*, 2:175–183.

Ullah, H., Khalđl, I.H., Lđghtfoot, D., Ullah, I., and Rahman, H. (2010). Estimation of genetic divergence among elite mungbean (*Vigna radiata*) genotypes by RAPD analysis. *Biol. Diver. Cons.* **3** (1): 75-82.

Varshney, R. K., Graner, A. and Sorrells, M. E. (2005). Genic microsatellite markers in plants: features and applications. *TRENDS in Biotechnology*, **23** (1): 48-55.

Weiss, E. A. (1983). *Oilseed crops*. Tropical series, pp. 603, Longman Inc. Leonard Hill Books, New York.

White, T. L., Adams, W.T. and Neale, D. B. (2007). *Forest genetics*. CABI, USA.

Wright S. (1978). Evolution and the Genetics of Populations, Vol.4. Variability within and among natural populations. Univ. Chicago Press, Chicago, IL.

Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*, **19**: 395–420. doi: 10.1111/j.1558-5646.1965.tb01731.x.

Xin, H. P., Wu, B. H., Zhang, H. H., *et al*., (2013). Characterization of volatile compounds in flowers from four groups of sweet osmanthus (*Osmanthus fragrans*) cultivars. *Can. J. Plant Sci.* **93** (5):923-931.

8. APPENDICES

Appendix 1, Table 1. *Vernonia galamensis* sample collection sites

Regional state	Zone	Woreda	Co-ordinates		Altitudes	Accession codes	Accession number	Total/ zone
			Latitude	Longitude				
Oromia	Borena	Yabelo	04°53'N	38°08'E	1092m	VgEt001-VgEt015B	15	15
	East Showa	Adami Tullu & Jido Kombolcha	7°56'N	38°43'E	1643m	VgEt031-VgEt033	3	15
		Dugda	8°9'N	39°49'E	1636m	VgEt034-VgEt038	5	
		Bora	8°10'N	38°50'E	1630m	VgEt039-VgEt045	7	
	West Arsi	Sheshemene Zuria	07°28'N	38°30'E	2044m	VgEt046-VgEt051	6	
		Arsi Negele	07°36'N	38°60'E	2143m	VgEt052-VgEt056	5	15
		Shala	07°47'N	38°55'E	2000m	VgEt057-VgEt060	4	
	East Hararghe	Bedeno	09°06'N	41°38'E	2044m	VgEt061-VgEt065	5	16
		Malkabalo	09°12'N	41°25'E	2750m	VgEt066-VgEt070	5	
		Meta	09°25'N	41°34' E	1574m	VgEt071-VgEt076	6	
	West Hararghe	Badessa	08°56'N	40°52'E	1626m	VgEt077-VgEt079	3	14
		Chiro Zuria	09°19'N	42°27'E	1889m	VgEt080-VgEt084	5	
		Mieso	09°13'N	41°06'E	1393m	VgEt085-VgEt090	6	
Amhara	West Gojjam	Enemay	10°27'N	38°12'E	2560m	VgEt091-VgEt104B	14	14
	South Wollo	Debra Sina	9°51'N	39°36'E	2630m	VgEt105-VgEt110B	6	16
		Legambo	11°0'N	39°10'E	1866m	VgEt111-VgEt116B	6	
		Dessie Zuria	11°8′N	39°38′E	2550m	VgEt117-VgEt120B	4	
Southern Nations, Nationalities and Peoples Regional state	Sidama	Aleta Wondo	06°51'N	37°45'E	1780m	VgEt016-VgEt024W	9	15
		Awassa Zuria	7°15'N	38°27'E	1708m	VgEt025-VgEt030W	6	
	Konso	Konso	5°15′N	37°29′E	1650m	VgEt121-VgEt135	15	15
	Derashie	Derashie	6°18′N	36°53′E	1395m	VgEt136-VgEt150	15	15
Total							150	

Key: B = accessions obtained from Ethiopian Biodiversity Institute; W = accessions obtained from Wondo Genet Agricultural Research Center, the remaining accessions collected from the field

Appendix 2, Table 2. The mean performance of yield and yield components of *Vernonia galamensis* grown at Melka Werer Agricultural Research Center (MWARC) and Wondo Genet Agricultural Research Center (WGARC).

Accession	DE	DFH	PH	DFF	NB	BL	DNM	NHP	SPH	SYP	SWP	SYH	TSW
VgEt-001	9.75	73.25	148.60	77.20	39.50	46.53	170.25	165.15	159.11	17.92	150.35	500.84	2.85
VgEt-002	10.00	72.75	151.90	78.22	42.70	59.85	167.75	159.15	153.55	18.36	129.28	430.95	2.94
VgEt-003	10.00	73.75	135.85	73.20	32.45	46.30	171.75	163.55	135.40	13.79	128.64	428.13	3.16
VgEt-004	10.00	76.00	135.90	73.97	25.85	62.03	174.00	164.10	159.43	15.44	119.15	397.18	3.02
VgEt-008	10.75	73.50	148.00	77.42	40.90	58.40	173.25	173.95	151.20	13.80	148.03	493.42	3.30
VgEt-010	10.25	73.75	154.50	79.50	39.70	58.48	164.50	176.05	161.08	15.86	164.09	546.88	2.99
VgEt-012	10.50	74.00	151.65	78.72	45.05	55.60	164.75	186.80	141.98	14.47	153.65	512.18	3.20
VgEt-016	11.00	74.50	172.75	86.08	44.38	60.68	169.00	165.30	149.03	17.75	149.00	496.67	3.03
VgEt-017	10.50	74.50	161.50	82.17	33.50	66.08	163.00	159.10	160.46	16.92	141.53	471.67	3.17
VgEt-019	11.00	74.50	158.55	81.35	34.48	55.73	167.00	156.75	129.09	14.92	158.30	527.59	3.15
VgEt-020	10.50	74.75	168.55	84.60	34.43	58.05	175.75	151.40	129.34	13.59	165.90	553.18	3.01
VgEt-022	10.50	76.25	154.25	80.33	37.23	56.18	163.00	136.45	131.86	13.18	153.73	512.42	2.98
VgEt-024	10.50	77.25	168.45	85.40	34.03	52.63	165.00	147.55	160.63	13.07	138.83	462.76	2.90
VgEt-025	9.75	75.50	167.65	84.30	27.90	56.10	161.50	138.45	147.32	13.08	119.30	397.67	2.97
VgEt-027	11.25	79.00	165.30	85.18	35.95	56.25	179.00	148.35	142.63	15.22	104.68	348.98	3.01
VgEt-030	10.75	76.75	149.75	79.08	27.65	52.90	172.00	148.50	150.21	18.41	119.53	480.18	3.16
VgEt-032	10.75	76.50	155.00	80.75	23.85	49.85	177.50	136.05	137.10	16.31	109.15	363.83	3.25
VgEt-033	10.25	76.75	157.75	81.58	22.73	47.78	168.00	143.05	141.65	17.62	100.15	392.84	3.19
VgEt-035	10.00	76.50	145.75	77.42	22.73	46.03	172.25	153.50	139.56	18.26	122.70	409.01	2.97
VgEt-039	9.75	77.25	151.80	79.60	29.75	53.03	173.50	143.45	138.24	13.44	123.93	413.34	3.15
VgEt-040	11.50	77.00	165.25	84.58	32.50	57.20	164.75	132.10	154.50	14.12	123.95	413.17	3.24
VgEt-042	11.75	77.50	161.80	83.68	18.65	49.35	176.75	130.40	136.05	14.35	166.05	553.52	3.09
VgEt-043	11.00	75.75	150.90	79.22	22.20	50.63	165.75	129.65	144.38	13.68	139.45	536.99	3.12

VgEt-045	11.25	76.25	156.65	81.38	24.88	54.85	172.50	134.40	140.45	15.19	154.40	514.67	2.87
VgEt-048	10.50	76.50	149.15	78.72	28.68	54.70	173.50	127.95	142.84	15.07	135.35	451.17	2.85
VgEt-049	10.50	76.00	141.80	76.10	26.33	52.90	173.00	127.70	141.25	14.33	159.98	533.25	2.83
VgEt-050	10.75	74.50	143.40	76.22	23.65	58.73	176.00	137.95	145.44	16.75	120.10	400.33	3.02
VgEt-053	11.25	76.00	165.10	84.12	20.55	56.33	165.75	140.95	151.32	18.36	164.60	548.83	3.08
VgEt-055	10.50	76.00	173.00	86.50	34.48	50.55	182.00	142.00	151.03	18.13	119.68	399.92	3.04
VgEt-056	10.00	75.50	166.70	84.07	31.30	64.28	182.00	153.80	125.30	16.26	121.55	389.83	3.23
VgEt-058	9.75	75.50	162.70	82.65	24.08	60.73	164.50	145.90	136.74	15.91	119.30	397.68	2.89
VgEt-060	12.00	75.50	162.35	83.28	36.68	58.00	168.25	143.65	140.92	15.49	158.40	528.04	3.37
VgEt-061	10.75	74.50	152.25	79.17	41.85	67.03	178.75	151.65	131.29	15.99	165.33	551.00	3.08
VgEt-062	9.50	75.25	158.25	81.00	30.68	53.53	158.50	144.25	141.91	14.01	145.20	484.01	3.04
VgEt-063	10.50	75.25	155.55	80.43	38.23	56.45	166.00	154.05	134.82	12.48	156.80	522.67	2.88
VgEt-064	11.00	77.00	161.10	83.03	35.10	55.50	163.50	144.80	135.69	14.83	146.38	487.92	3.47
VgEt-069	10.00	76.50	151.25	79.25	33.55	55.85	174.00	148.20	141.56	15.90	146.20	487.33	3.22
VgEt-071	11.25	76.25	162.75	83.42	30.63	53.68	177.75	150.65	138.06	14.75	122.83	409.43	3.12
VgEt-073	11.50	76.75	153.75	80.67	33.00	63.00	166.25	149.00	143.58	14.46	133.58	513.75	2.79
VgEt-075	10.50	76.75	167.40	84.88	28.53	68.03	160.75	161.70	144.07	13.75	147.65	492.17	3.13
VgEt-076	10.50	78.50	161.50	83.50	32.48	54.85	165.75	176.10	146.51	16.17	152.63	573.75	2.80
VgEt-078	10.50	76.50	176.10	87.70	39.88	56.38	168.25	156.05	133.37	17.16	160.03	616.75	3.21
VgEt-79	11.00	76.50	161.10	82.87	34.65	60.30	177.00	150.90	137.85	15.77	156.40	521.33	3.14
VgEt-080	11.00	77.50	159.00	82.50	35.58	49.90	162.75	178.05	128.83	17.10	113.98	379.93	2.89
VgEt-082	10.50	76.50	145.60	77.53	33.55	52.85	164.00	158.85	143.75	17.00	128.78	429.25	3.08
VgEt-085	10.75	75.00	152.75	79.50	41.50	48.63	169.00	154.10	148.35	17.16	187.27	528.23	2.92
VgEt-087	10.25	74.75	156.75	80.58	36.33	46.63	170.75	171.40	132.32	14.65	141.89	472.97	3.09
VgEt-094	9.50	75.00	146.55	77.02	28.88	57.58	178.75	153.85	137.06	14.47	174.67	624.27	3.51
VgEt-095	9.75	75.00	154.80	79.85	32.60	50.85	175.25	164.30	149.32	15.19	152.77	554.09	3.26

VgEt-096	11.00	74.50	168.70	84.73	40.30	52.83	168.25	157.95	140.50	15.90	141.57	471.91	3.26
VgEt-097	12.50	76.50	147.50	78.83	45.23	51.45	166.25	183.40	134.33	18.10	144.69	449.15	3.23
VgEt-098	11.00	78.00	166.90	85.30	47.53	54.25	173.25	188.65	132.43	15.60	125.15	417.16	2.92
VgEt-100	10.75	78.00	146.55	78.43	49.98	51.38	173.00	181.80	125.13	15.22	149.02	496.71	3.04
VgEt-102	10.00	76.75	138.75	75.17	46.56	51.03	182.00	139.15	134.95	14.94	134.07	446.91	2.96
VgEt-103	10.75	75.75	154.70	80.40	50.78	50.43	180.00	150.70	122.94	14.80	125.60	388.00	3.16
VgEt-105	11.75	76.50	150.00	79.42	30.88	52.18	165.50	159.50	136.48	15.61	133.32	444.39	2.77
VgEt-107	10.50	77.25	151.80	79.85	32.50	55.78	165.75	162.40	144.93	14.46	136.83	411.50	3.17
VgEt-110	10.50	76.75	157.10	81.45	31.68	50.10	173.75	157.35	136.51	15.96	158.27	527.50	2.93
VgEt-111	11.50	78.25	164.25	84.67	36.80	56.70	166.25	147.10	134.96	13.58	109.25	364.17	3.06
VgEt-113	10.50	77.25	149.65	79.13	38.38	54.45	169.50	146.95	126.98	15.04	148.98	496.60	3.06
VgEt-115	10.25	74.50	152.00	78.92	38.93	57.55	177.75	184.65	146.71	13.59	125.55	418.49	3.05
VgEt-117	10.25	74.75	162.55	82.52	39.05	64.70	178.50	161.25	154.58	16.07	157.85	526.17	3.04
VgEt-118	11.00	75.25	159.45	81.90	45.98	54.95	183.00	182.15	142.31	16.96	135.10	383.45	2.80
VgEt-120	12.00	78.00	152.50	80.83	41.00	63.20	170.25	168.00	137.09	14.67	149.39	497.96	2.97
VgEt-122	11.00	73.25	145.05	76.43	39.60	56.50	176.25	163.50	134.80	15.75	136.82	456.07	2.66
VgEt-124	11.75	73.00	141.75	75.50	27.98	69.08	164.25	146.10	148.80	17.69	126.88	422.91	2.87
VgEt-126	10.25	77.00	146.75	78.00	22.45	58.63	164.50	146.20	146.04	13.46	131.13	436.99	2.97
VgEt-130	10.75	75.25	152.65	79.55	28.40	59.48	178.75	141.30	147.67	14.69	141.13	470.42	3.02
VgEt-131	10.75	77.50	175.50	87.92	34.95	61.73	169.00	142.95	135.67	15.75	132.93	442.99	2.94
VgEt-132	11.50	78.25	158.85	82.87	30.23	57.45	173.50	143.40	155.17	16.44	149.81	499.37	2.93
VgEt-134	10.50	77.50	171.20	86.40	40.53	59.00	173.25	144.60	128.54	14.96	141.51	471.69	3.22
VgEt-135	10.25	76.50	164.00	83.58	34.35	57.70	181.50	141.90	164.07	20.07	146.97	489.88	3.02
VgEt-138	12.00	75.50	174.50	87.33	40.33	55.88	165.00	149.55	138.44	15.09	138.85	462.84	3.00
VgEt-140	11.50	73.25	151.05	78.60	44.25	61.25	164.75	153.25	137.15	20.55	152.94	509.84	2.98
VgEt-141	11.00	78.00	173.75	87.58	29.53	64.93	182.00	140.75	159.15	16.60	187.43	524.78	3.17

VgEt-143	11.50	79.50	160.25	83.75	41.48	60.85	182.25	172.85	157.29	14.13	152.16	507.20	3.02
VgEt-145	11.25	80.50	153.05	81.60	43.75	48.63	176.75	171.55	133.43	14.87	136.36	454.52	2.92
VgEt-147	11.25	80.50	142.25	78.00	50.43	50.95	185.50	132.40	164.81	16.41	164.32	547.73	3.06
VgEt148	9.75	79.50	161.05	83.43	55.75	62.05	181.75	165.25	152.73	14.24	149.64	498.82	3.02
VgEt150	11.75	78.50	167.00	85.75	45.15	60.20	185.25	154.00	141.25	19.40	151.45	504.83	3.09
Mean	10.73	76.18	156.75	81.22	35.13	56.26	171.65	153.89	142.41	15.63	141.63	473.82	3.03

Appendix 3, Table 3. Estimates of mean, ranges, mean square and critical variation (CV) for 13 quantitative traits of *Vernonia galamensis*

Quantitative traits	Locations	Mean ± SE	Range		Mean square	CV (%)
			Maximum	Minimum		
Days to emergence	MWARC	10.74±0.09	12.50	9.00	1.24	9.72
	WGARC	10.72±0.086	13.00	9.50	0.99	7.37
Days to 50% heading	MWARC	68.13±0.26	73.00	63.50	9.39**	2.01
	WGARC	84.22±0.26	91.00	81.50	6.20	2.87
Days to 50% flowering	MWARC	89.82±1.31	93.50	85.00	4.11**	0.86
	WGARC	103.11±1.73	111.50	96.00	14.61**	1.14
Plant height (cm)	MWARC	119.88±0.25	143.50	89.50	229.94**	5.99
	WGARC	193.62±0.38	228.00	159.50	440.8	10.78
Number of branches	MWARC	33.41±0.97	52.35	17.80	106.28**	16.59
	WGARC	36.85±0.96	59.50	19.50	88.13	33.18
Branch length (cm)	MWARC	56.26±0.76	71.90	46.00	74.41**	6.21
	WGARC	72.81±0.78	76.00	43.00	88.60*	12.61
Days to 90% maturity	MWARC	163.65±0.86	177.00	151.00	120.87**	1.71
	WGARC	179.64±0.75	198.50	165.50	60.09	4.49
Number of head/plant	MWARC	118.73±2.94	188.40	84.90	1025.94**	11.69
	WGARC	189.06±2.23	239.20	132.40	195.48*	5.76
Number of seed/head	MWARC	121.99±1.80	163.50	93.08	432.29*	13.50
	WGARC	162.84±1.19	189.00	141.20	201.11	6.87
Seed yield /plant (g)	MWARC	12.04±0.22	16.84	8.70	7.15*	17.26
	WGARC	19.22±0.33	25.65	14.85	16.74	10.07
Seed weight/plot (g)	MWARC	113.74±39.5	246.80	58.60	1019.71**	34.76
	WGARC	169.51±44.3	331.20	116.80	1635.78*	26.12
1000-seed weight (g)	MWARC	3.00.00±0.02	3.35	2.65	0.058**	8.54
	WGARC	3.08±0.02	3.45	2.42	0.095*	3.93
Seed yield (Kg ha-1)	MWARC	385.49±9.42	822.67	246.80	10723.23**	32.15
	WGARC	562.16±11.54	1104.00	406.00	16588.91**	28.38

**, * Significant at P<0.01 and P<0.05, respectively. Max = maximum, Min = minimum, MWARC = Melka Werer Agricultural Research Center, WGARC = Wondo Genet Agricultural Research Center, CV= critical variation.

Appendix 4, Table 4. Nei's original measures of unbiased genetic distance (Ds) (above diagonal) and unbiased genetic identity (I) (below diagonal)

Population	BOR	SID	ESH	WAS	EHG	WHG	Gojjam	WOL	KON	DER
BOR		0.61	0.31	0.36	0.48	0.34	0.40	0.30	0.18	0.34
SID	0.55		0.52	0.31	0.33	0.48	0.45	0.37	0.44	0.30
ESH	0.74	0.59		0.43	0.38	0.41	0.28	0.22	0.35	0.28
WAS	0.70	0.74	0.65		0.39	0.32	0.39	0.30	0.29	0.40
EHG	0.62	0.72	0.69	0.68		0.42	0.29	0.34	0.33	0.19
WHG	0.71	0.62	0.66	0.73	0.66		0.35	0.29	0.25	0.35
GOJ	0.67	0.64	0.76	0.68	0.81	0.71		0.29	0.31	0.37
WOL	0.74	0.69	0.80	0.74	0.71	0.75	0.75		0.31	0.45
KON	0.83	0.65	0.70	0.75	0.72	0.78	0.73	0.73		0.29
DER	0.71	0.74	0.81	0.67	0.83	0.71	0.84	0.75	0.86	

Keys: BOR = Borena, SID = Sidama, ESH = East Showa, WAS = West Arsi, EHG = East Hararghe, WHG = West Hararghe, GOJ = West Gojjam, WOL = South Wollo, KON = Konso, DER = Derashie

Appendix 5, Figure 1. Unweighted neighbor joining based clustering of 150 *Vernonia galamensis* accessions for 20 polymorphic SSR markers. Key: BOR = Borena, SID = Sidama, ESH = East Showa, WAS = West Arsi, EHG = East Hararghe, WHG = West Hararghe, GOJ = West Gojjam, WOL = South Wollo, KON = Konso, DER = Derashie