# Genetic diversity and population structure of Vernonia [Vernonia galamensis (Cass.) Less] populations from Ethiopia revealed by SSR markers

**Alemneh Mideksa Egu** ( alemnehmideksa@gmail.com )
Addis Ababa University https://orcid.org/0000-0002-0755-6525
**Kifle Dagne**
Addis Ababa University College of Natural Sciences
**Kassahun Tesfaye**
Addis Ababa University College of Natural Sciences
**Xuebo Hu**
Huazhong Agriculture University

**Research article**

**Keywords:** Allelic pattern, dendrogram, genetic diversity, polymorphism, simple sequence repeat, vernonia

**Posted Date:** August 15th, 2019

**DOI:** https://doi.org/10.21203/rs.2.12914/v1

# Abstract

Background Vernonia (Vernonia galamensis) is a potential novel industrial crop due to high demand to its natural epoxidised oil, which can be used for the manufacturing of polyvinyl chloride, adhesives, and petrochemicals, cosmetic and pharmaceutical products. This study is initiated for the systematic and intensive assessment of V. galamensis accessions genetic diversity through SSR molecular markers to minimize the existing research gaps, and provide a clue for germplasm conservation and further research. Results A total of 150 V. galamensis accessions were analyzed using 20 SSR markers. The markers detected a total of 79 with an average of 3.9 alleles per locus. The mean number of effective alleles was 3.06 and, the mean observed heterozygosity (Ho) was 0.15 across all the 20 markers evaluated. The marker also showed the highest percent of polymorphism that ranged from 0.50 to 0.96 with an average of 0.76. The analysis of molecular variance showed only 11% variation was among populations, 22% among individuals within populations and 67% within individuals. The largest number of migrants per generation was occurred between Derashie and Wollo (Nm=7.37) whereas the lowest values was between East Harerghe and West Harerghe (Nm =1.42). A factor analysis including dendrogram clusters and principal coordinates classified the 150 accessions into 4 groups. However, the Bayesian model based clustering (STRUCTURE) grouped into 3 (K = 3) major gene pools. These analyses showed accessions collected from the same region of origin did not often grouped entirely together within a given major groups. Conclusions Molecular genetic diversity analysis, using SSR markers was the first report in V. galamensis. All the markers used were polymorphic in the population studied. The markers detected the larger number of alleles, higher expected heterozygosity than observed heterozygosity. The markers applied to ten populations, in which East Showa and West Harerghe revealed higher genetic diversity, and can be considered as the hotspots for in-situ conservation of V. galamensis. In addition, the values of SSR markers such as heterozygosity, Shannon's index, polymorphic information content and population clusters are an important baseline information for future V. galamensis cultivation, breeding and genetic resource conservation endeavors in Ethiopia.


# Background

Vernonia (*Vernonia galamensis* (Cass.) Less.; 2n =18) belongs to the family Asteraceae (Compositeae), a potential novel industrial crop due to high demand for its natural epoxidised oil [1-5]. *Vernonia galamensis* subsp. *galamensis* variety *ethiopica* M. Gilbert was first identified by Perdue in 1964 in Eastern Ethiopia [6-10]. Seeds of this plant are the major source of naturally occurring epoxidised fatty acids and other essential fatty acids such as linoleic acid, oleic acid, palmitic acid, stearic acid and trace amounts of arachidic acid [11-13].

Study genetic diversity of crop plants is a valuable tool for wild populations such as *V. galamensis* to address about its conservation, levels of gene flow among populations and its improvement through breeding [14, 15]. The assessment of genetic diversity within and between plant populations is routinely performed on the basis of morphological, biochemical, and molecular markers [16, 17]. Further, there is a need to characterize the diverse genetic resources using different statistical tools and utilize them in the breeding programmes [18].

SSR markers are among the most commonly used molecular markers to evaluate the genetic diversity within species, to investigate phylogenetic relationships, to identify and test the paternity of cultivars, to study population structure and gene flow, and to develop a gene mapping [19, 20]. SSRs are highly versatile genetic markers because of their co-dominant inheritance (distinguishes homozygotes from that of heterozygotes), high abundance, highly polymorphic due to the high mutation rate affecting the number of repeat units, enormous extent of allelic diversity (good genome coverage), ease of assessing SSR size variation through PCR with pairs of flanking primers and high

reproducibility [15]. 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 [21]. To date, no information, no anyone used these SSR markers to study the genetic diversity of *V. galamensis*.

In Ethiopia, geo-ecological conditions are favorable for the cultivation of *V. galamensis*, and used as a source of raw material for agro-processing industries [2, 9, 10, 22]. However, the plant is neglected and considered only as a wild weed colonizing disturbed areas and bare agricultural lands [3]. As a result, the crop is not cultivated in any of the collection sites and/or elsewhere in the country. Moreover, lack of attention, negligence in research and conservation, priority has been given to other major crop plants while the potential industrial values of *V. Galamensis* is underestimated and underexploited. The plant is also under threat of continued genetic erosion. This study is therefore initiated for the systematic and intensive evaluation and characterization of *V. galamensis* accessions genetic diversity through molecular analysis using SSR markers to minimize the existing research gaps, and provide a clue for germplasm conservation and further research.

# Methods

### Plant material

A total of 150 *V. galamensis* accessions, representing 10 populations, were randomly collected from its diverse agro-ecologies of the three-potential growing regional states of Ethiopia (Figure 1; Table 1). The samples identity was confirmed from the description that available in the Flora of Ethiopia and Eritrea. Most of the study materials were collected from the field and, others were assembled from the Ethiopia Biodiversity Institute and Wondo Genet Agricultural Research Center.

At each collection areas, seed samples were collected from plants and kept in separate bags. To ensure that the distance between any two collecting site was about 5-10 Km. From Collection areas observations, *V. galamensis* naturally grows in hilly/depression, along the roadside, in valley, farm lands, in forest, in the compounds of mosques and church [23]. The collections were done by taking either seed samples of individual flower head or seeds from plants with all matured flowers, and then accessions were threshed, cleaned and documented. *V. galamensis* was not cultivated in any of the collecting sites.

### Leaf sample collection and DNA extraction

Fresh young leaves of 150 *V. galamensis* accessions were collected from individual plants that grown at experimental sites, representing 10 populations. A collected leaf samples was put in a sealed bag envelope and dried with silica gel (with 1:10 ratio of leaf samples to silica gel), then kept under room temperature until it used for later DNA extraction according to Gilbert et al. [24]. The dried leaf samples were transported to Huazhong Agricultural University, China for genetic analysis. The total genomic DNA extraction was made according to the modified CTAB protocol of Doye and Doye [25]. Extracted DNA was visualized on a 1% (w/v) agarose gel and quantified spectrophotometrically using a Nanodrop® 2000 (Thermo Scientific, USA). Finally, it was stored at -20 °C for further use.

### Primer Screening and Optimization

About 63 simple sequence repeats (SSR) markers were developed by Narina et al. [26] and available in the database (gene bank). Among these, 30 SSR markers were selected, and finally 20 SSR markers were used for this study based on their high polymorphism and compatibility for multiplexing (Table 2). Optimization was carried out by a sequential investigation of each reaction variable, testing different cycling conditions and then by varying (1) the amount of DNA template, (2) the concentration of primer, and (3) the concentration of Taq PCR master mix.

### Polymerase Chain Reaction (PCR)

The amplification reaction was performed with a thermal cycler using 96-well plates (T100$^{TM}$ Thermal Cycler) in a total volume of 10 µl reaction mixture, containing 100 ng/ml of template DNA, 5 µl 2 x Taq PCR master mix (Vazyme P213-01, China), 1 µl of forward and reverse primers and 3.0 ml 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.

### Band Scoring and Analysis

The amplified products were visually scored based on its migration in comparison with the size standard (100 bp DNA ladder) from the gel photographed under UV illumination (Gel Doc$^{TM}$ with Image Lab$^{TM}$ software, BIO-RAD, in the lab of drug discovery and Technology).The genetic diversity for each alleles such as the number of different alleles (Na), the effective number of alleles (Ne), Shannon's diversity index (I), observed heterozygosity (Ho), expected heterozygosity (He), F-statistics values ($F_{is}$, $F_{it}$ and $F_{st}$), polymorphic information content (PIC), random segregation and distribution (Hardy-Weinberg equilibrium) of each genotype within populations for each locus, Nei's genetic identities (Ji), genetic distances (Ds) and gene flow (Nm) in *V. galamensis* populations were performed using GeneAlex version 6.503 software [27].

Simple matching dissimilarity coefficient-based Unweighted Pair Group Method with Arithmetic Mean (UPGMA) and Neighbor-Joining (NJ) tree was computed using DARwin version 6.0.19 software [28]. The resulting trees were displayed using Fig Tree version. 1.4.4 [29]. The principal coordinated analysis (PCoA) was performed using GeneAlex version 6.503 software [27]. The SSR markers data that subjected to a Bayesian model-based cluster analysis was performed using STRUCTURE version 2.3.4 software [30]. 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. [31] using the web-based (http://tyloro.biology.ucla.edu\structure Harvester\) STRUCTURE HARVESTER ver. 0.6.92 [32]. The results generated by this software were visualized in a graphical bar plot using Clumpak beta version (http://www.clumpak.tau.ac.il/) [33].

## Results

### Molecular based genetic diversity of *Vernonia galamensis* using SSR markers

Twenty SSR markers were used for the characterization and genetic diversity analysis of the 150 *V. galamensis* accessions, all of which were polymorphic (Table 3). A total of 79 alleles were identified, varied from 2 to 6 with an

average of 3.9 alleles per locus. The maximum number of effective allele (Ne) was 4.79 (Vg-03) and the least number of effective alleles was 1.99 (Vg-16). The highest major allele frequency (MAF) (0.85) was recorded by locus Vg-01 and the least MAF was (0.45) recorded by locus Vg-03. The observed heterozygosity (Ho) values were quite low that ranged between 0.05 (Vg-21) and 0.36 (Vg-03) with an average of 0.16 across all the 20 markers evaluated. The expected heterozygosity (*He*) mean was 0.50, ranged from 0.23 (Vg-11) to 0.65 (Vg-03) (Table 3).

Shannon-Weaver's information indices (I) ranged from 0.86 to 1.67, and averaged at 1.20. The index (F) compares *He* with *Ho*, estimating the degree of allelic fixation, and ranged from 0.25 (Vg-01) to 0.89 (Vg-21) with an average of 0.68. Finally, polymorphic information content (PIC) values ranged between 0.50 and 0.96 with an average of 0.76. Microsatellite markers such as Vg-02 and Vg-11 showed the highest polymorphism with 0.96, 0.93, respectively (Table 3). The results of diversity parameters showed that a high level of polymorphism among the 20 SSR markers, favoring the genetic variation within *V. galamensis* collection.

### Analysis of molecular variance (AMOVA) and genetic distances

The molecular analysis of variance (AMOVA) showed a 67% of the total variation was attributed to genetic variability among individuals from different populations, whereas 22% was due to variation among individuals within the same population. In contrast, a smaller portion (11%) was among populations variations (Table 4).

### Genetic distance among *Vernonia galamensis* populations

The maximum pairwise Nei's [34] genetic distance (GD) was observed between populations of Borena and East Harerghe (0.57), followed between populations of Sidama and West Harerghe (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 (I) was occurred between Konso and Derashie (0.80) population, while the least Nei's [34] genetic identity was observed between Borena and Konso (0.24) populations. The overall magnitude of pairwise population matrix of Nei genetic distance was relatively lower than that of Nei's genetic identity (Table 5).

### Cluster and principal co-ordinate analysis (PCoA)

Clustering analysis was performed based on the allelic frequency, grouped the 150 accessions into four (4) major clusters from the main node using neighbor-joining, with the DARrwin 6.0.19 software programs. Each of the four clusters comprises individual plants from different zones (geographic regions). The first and the third cluster further divided into sub-clusters, the samples grouped according to their geographic origin (Figure 2).

The first cluster constituted 41 accessions mainly from Borena (11) and West Harerghe (12), the second cluster contained 25 accessions, and mainly from Wollo (8), cluster three was characterized as the major group in clustering, composed of 59 accessions while the fourth cluster, $C_4$, comprised accessions mainly from West Arsi (7) (Figure 2). Generally, the cluster analysis revealed that accessions from different populations (collection sites) clustered together, and clusters did not follow a clear pattern of geographic origins.

The principal co-ordinate analysis (PcoA) showed that the majority of samples were placed at the center of a two-dimensional coordinate plane and roughly forms four groups (Figure 3). The first three axes of the PCoA accounted together 33.02% of the total variation.

### Population structure analysis

Analysis of population structure distinguished the 150 *V. galamensis* accessions using a model based Bayesian approach with the highest *ΔK* Value that ranging from K = 1 to K= 10 and 20 iterations for each K. According to Evanno et al. [31] and Gilbert et al. [35] STRUCTURE outputs were used for STRUCTURE Harvester and predicted K = 3 were most likely selected to describe the genetic structure of the 150 *V. galamensis* accessions (Figure 4). Based on this value, population structure (Clumpak result) revealed that accessions collected from the same region of origin did not often grouped entirely together within a given major groups. There was a wide admixture in structuring of *V. galamensis* populations, agreed with neighbor joining trees.

# Discussion

### Determination of SSR-markers based genetic diversity with genetic parameters

*Vernonia galamensis* is a potential novel industrial crops, contains naturally occurring epoxidized oil. However, its potential values are neglected, underestimated and underexploited. In addition, it also exposed to genetic erosion. Therefore, assessment of genetic diversity with SSR markers generally in plants and particularly in *V. galamensis* is important for in-situ and ex-situ conservation and efficient management, for selection and improvement of the available genetic resource [16]. The SSR study showed considerable genetic diversity, the average number of alleles (3.9) detected in this study was higher than that reported by Keneni et al. [36] , an average of 3.36 alleles per locus using 155 chickpea accessions with 33 SSR markers. But lower than that reported by Olango et al. [37] and Gadissa et al. [38], who reported an average number of alleles 5.94 and 6.40 using 70 enset and 174 *Plectranthus edulis* accessions with 34 and 20 SSR markers, respectively. Number of effective alleles (Ne) is an important parameter to measure genetic diversity in a finite population, averaged 3.06 and Varied from 1.99 to 3.05.

Polymorphic information content (PIC) is generally used for characterization of marker polymorphism. In this study, the PIC values ranged between 0.50 (Vg-05) and 0.96 (Vg-02) with an average of 0.76, higher than that reported by Adugna et al [39] and Olango et al. [37], reported average PIC of 0.62 and 0.54, using 160 cultivated sorghum bicolor and 70 enset (*Ensete ventricosum* (Welw.) accessions with 12 and 34 SSR markers, respectively. The diversity parameters showed that a high level of polymorphism among the 20 SSR markers, favoring the genetic variation within *V. galamensis* collection. For most of the loci, expected heterozygosity ($H_e$) values were higher than that of observed heterozygosity ($H_o$), revealing a high homozygosity at the given loci among the accessions [38, 39].

### Genetic Differentiation and Gene Flow

The (AMOVA) demonstrated that *V. galamensis* had low variation among population (11%). On the other hand, 67% of the total variation was attributed to genetic variability among individuals from different populations and 22% was due to variation among individuals within the same population. The result is similar to the previously reported in chickpea [36], cultivated Sorghum bicolor [38] and Ethiopia potato [39]. In addition, *F*st has important in discriminating genetic differentiation among the studied populations, according to IPGRI and Cornell University [40], *F*st values ranging from 0.0 to 0.05 was small in genetic differentiation, from 0.05 to 0.15 correspond to moderate, and from 0.15 to 0.25 imply large, and those greater than 0.25 was very large genetic differentiation among populations in terms of allele frequencies. In line with these, the extent of genetic differentiation among the ten populations in terms of allele frequencies measured was moderate ($F_{st}$= 0.101), which implied individuals within similar populations was significant. The same trends were reported by Adugna et al. [39], Olango et al. [37] and Gadissa et al. [38].

Genetic distance is the measure of the allelic substitutions per locus that have occurred during the separate evolution of two populations, and in this study the largest genetic distance was observed between Borena and East Harerghe (0.57) populations, while the minimum genetic distance was observed between Borena and Konso (0.24). The overall magnitude of pairwise population matrix of Nei genetic distance was relatively lower than that of Nei's genetic identity. The genetic identity of two populations could be due to interspecific hybridization that has occurred throughout their evolution, which favors allele sharing [36].

### Clustering and principal co-ordinates among *Vernonia galamensis* accessions

In the present study, a phylogenetic tree was constructed based on the 150 accessions of *V. galamensis* collected from different geographic and agro-ecological regions. *V.galamensis* accessions were clustered into four (4) major clusters based on the allelic frequency. Cluster 1 was characterized as the second major group, comprised of 41 accessions, the second cluster contained 25 accessions, the third cluster composed of 59 accessions, and the fourth groups contained 25 accessions that collected from different regions of origin. Generally, the cluster analysis revealed a poor clustering pattern, accessions from different collection sites was clustered together; clusters did not follow a clear pattern of geographic origins, which may imply the presence of gene flow between and within populations/regions/collection sites. Similarily, Adugna (2014) reported that 160 cultivated sorghum bicolor grouped into 3 major clusters, and pattern of the population structure was weak intra-regional similarity. Gadissa *et al*. (2018) also reported similarly, 12 populations of Ethiopian potatos clustered into four major clusters, and mixed clustering was observed among accessions from different geographic regions (low levels of intra-regional similarities). In contrary, Keneni *et al*. (2012) reported 155 chickpea grouped into 5 clusters, and the clustering pattern showed the existence of definite pattern of relationships between geographical origins and genetic diversity (high levels of intra-regional similarities).

Principal components (PC) analysis explores complex data sets and transforms a number of associated variables into a smaller number of PCs. In the present investigation, the principal component analysis revealed that the majorities of samples were placed at the center of a two-dimensional coordinate plane and roughly forms three groups with the total variation of 30.04%. This, in turn, agrees with the results of the NJ dendrogram in that there was no unique clustering among accessions from the same population/collection areas. The presence of gene flow between and within populations/collection areas, accompanied by the prevalence of inter-gene pool introgressions/hybrids between the gene pools of origin may be the most probable explanations behind the mixed clustering of accessions from different populations/collection areas together. The result in the PCoA further supported by the previous results of Adugna (2014), mixed grouping of populations was observed among accessions from different collection areas.

### Populations genetic structure in *Vernonia galamensis*

The structure analyses of 150 *V. galamensis* accessions using a model based Bayesian approach based on highest $\Delta K$ Value, according to Gilbert *et al*. (2012) and Evanno *et al*. (2005) method. Three sub-populations were detected when K = 3 according to STRUCTURE results (STRUCTURE Harvester). The patterns of population structure was certainly supported by the UPGMA and PCoA analyses, however, accessions collected from the same region of origin did not often grouped entirely together within a given major groups. There was a wide admixture in structuring of *V. galamensis* populations, which again agreed with neighbor joining trees (Mondini et al., 2009; El-Esawi *et al*., 2018; Gadissa *et al*., 2018). Similarly, Adugna (2014) used STRUCTURE analysis of SSR markers in a study of 160 sorghum bicolor accessions and identified two (k =2) sub-groups. Most of the magnitudes identified for the SSR markers were important information for *V. galamensis* cultivation, breeding and genetic resource conservation.

# Conclusions

Molecular genetic diversity analysis, using SSR markers was the first report in *V. galamensis*. All the markers used were polymorphic in the population studied. The markers detected the larger number of alleles, higher expected heterozygosity than observed heterozygosity. The markers applied to ten populations, in which East Showa and West Harerghe revealed higher genetic diversity, and can be considered as the hotspots for in-situ conservation of *V. galamensis*. Generally, the results of the present study showed that there was ample allelic diversity among the *V. galamensis* accessions studied, such as heterozygosity, Shannon's index, polymorphic information content and population clusters. Most of the values identified for the SSR markers were important baseline information for future *V. galamensis* cultivation, breeding/genetic resource conservation endeavors in Ethiopia.

# Abbreviations

AMOVA: Analysis of molecular variance; CTAB: Cetyltriethyl ammonium bromide; He: Expected heterozygosity; Ho: observed heterozygosity; NJ: Neighbor joining; PCoA: Principal coordinate analysis; PIC: Polymorphic information content; SSR: simple sequence repeat; UPGMA: Unweighted pair group with arithmetic mean

# Declarations


### Acknowledgements

This is the part of author's PhD thesis work. The authors would like to thank Addis Ababa University and Adama Science and Technology University for material and technical supports for this research. We would also like to thank Melka Werer and Wondo Genet Agricultural Research Center for they created good working condition during field work, and the Ethiopian Biodiversity Institute (EBI) for providing accessions and for allowing the transfer of the study material to China, and the Huazhong Agricultural University for providing laboratory facilities to conduct this research.

### Funding

This work is financially supported by Addis Ababa University's Research Project and Huazhong Agricultural University, China through the link of Swedish International Development Cooperation Agency (SIDA).The role of the funding is limited to the field work and laboratory facilities.


### Availability of data and materials

Pass port data of 150 *Vernonia galamensis* samples representing the 10 populations used in the current study are provided in Additional file 1. Unweighted neighbor joining based clustering of 150 *Vernonia galamensis* accessions for 20 polymorphic SSR markers are provided in Additional file 2. AMOVA variation pie chart for 150 *Vernonia galamensis* accessions from ten populations in Ethiopia are provided in Additional file 3.

### Authors' contributions

AM, KD and KT designed the study. AM and XH coordinate and carried out the laboratory work. AM performed statistical data and wrote the manuscript. All the authors read and approved the final manuscript.

### Ethics approval and consent to participate

Not applicable

### Consent for publication

Not applicable

### Competing interests

The authors declare that they have no competing interests.


### Author details

[1]Department of Microbial, Cellular and Molecular Biology, Addis Ababa University, Box 1176, Addis Ababa, Ethiopia. [2]Department of Applied Biology, Adama Science and Technology University, Box 1888, Adama, Ethiopia. [3]Ethiopian Biotechnology Institute, Ministry of Science and Technology, Box 32853, Addis Ababa, Ethiopia. [4]Drug Discovery and Molecular Engineering, College of Plant Science and Technology, Huazhong Agricultural University, China


## References


1. Thompson AE, Dierig DA, Johnson ER, Dahlquist GH, Kleiman R. Germplasm development of *Vernonia galamensis* as a new industrial oilseed crop. Ind. Crops Prod. 1994a; https://doi.org/10.1016/0926-6690(94)90066-3.
2. Baye T, Kebede H, Belete K. Agronomic evaluation of *Vernonia galamensis* germplasm collected from eastern Ethiopia. Industrial Crops and Products. 2001; doi: 10.1016/S0926-6690(01)00082-6.
3. Baye T. Genotypic and phenotypic variability in *Vernonia galamensis* germplasm collected from eastern Ethiopia. The Journal of Agricultural Science. 2002; 139(02). doi: 10.1017/S0021859602002459.
4. Mebrahtu T, Gebremariam T, Kidane A, Araria W. Performance of Vernonia galamensis as a potential and viable industrial oil plant in Eritrea: yield and oil content. African Journal of Biotechnology. 2009;8(4): 635-640.
s
5. Shimelis H, Mashela P, Hugo A. Genotype by environment interaction of seed and oil yield in vernonia (Vernonia galamensis variety ethiopica). Industrial Crops and Products. 2011; 33(3):756–760. doi: 10.1016/j.indcrop.2011.01.018.
6. Gilbert MG. East African *Vernonieae* (Compositae), a revision of the *Vernonia galamensis* complex. Kew Bulletin. 1986; 41(1): 19-35.

7. Perdue RE, Carlson KD, Gilbert MG. *Vernonia galamensis*, Potential new crop source of epoxy acid. Economic Botany. 1986; doi: 10.1007/BF02858947.
8. Thompson AE, Dierig DA, Kleiman R. Characterization of *Vernonia galamensis* germplasm for seed oil content, fatty acid composition, seed weight, and chromosome number. Industrial Crops and Products. 1994b; doi: 10.1016/0926-6690(94)90121-X.
9. Baye T, Becker HC. Analyzing seed weight, fatty acid composition, oil, and protein contents in Vernonia galamensis germplasm by near-infrared reflectance spectroscopy. JAOCS, Journal of the American Oil Chemists' Society. (2004; doi: 10.1007/s11746-004-955-y.
10. Baye T, Becker HC. Exploration of *Vernonia galamensis* in Ethiopia, and Variation in Fatty Acid Composition of Seed Oil.*Genetic* Resources and Crop Evolution. 2005; 52(7): 805–811. doi: 10.1007/s10722-003-6086-5.
11. Thompson AE, Dierig DA, Kleiman R. Variation in *Vernonia galamensis* flowering characteristics, seed oil and vernolic acid contents. Industrial Crops and Products. 1994c; doi: 10.1016/0926-6690(94)90065-5.
12. Baye T, Becker HC, Witzke-Ehbrecht, SV. *Vernonia galamensis*, a natural source of epoxy oil: variation in fatty acid composition of seed and leaf lipids. Industrial Crops and Products. 2005; 21(2): 257–261. doi: 10.1016/j.indcrop.2004.04.003.
13. Bhardwaj HL, Hamama AA, Dierig DA. Fatty acids in vernonia produced in the Mid-Atlantic region of the united states. Journal of the American Oil Chemists Society. 2007; 84(4): 393–397. doi: 10.1007/s11746-007-1043-7.
14. Idrees M, Irshad M. Molecular markers in plants for analysis of genetic diversity*:* European Academic Research. 2014; 1(1).
15. Govindaraj M, Vetriventhan M, Srinivasan M. Importance of genetic diversity assessment in crop plants and its recent advances. Genet Res Int. *2015;* 431487. doi: 10.1155/*2015*/431487*.*
16. Ramalema SP, Shimelis H, Ncube I, Kunert KK, Mashela PW. Genetic analysis among selected *vernonia* lines through seed oil content, fatty acids and RAPD DNA markers. African Journal of Biotechnology. 2010; 9 (2):117-122.
17. Bhandari HR, Bhanu AN, Srivastava K, Singh MN, Shreya AH. Assessment of genetic diversity in crop plants - An overview. Adv. Plants. Agric. Res. 2017; **7**(3): 00255. DOI: 10.15406/apar.2017.07.00255.
18. Anumalla M, Roychowdhury R, Geda CK, Mazid M, Rathoure AK. Utilization of plant genetic resources and diversity analysis tools for sustainable crop improvement with special emphasis on rice. International Journal of Advanced Research. 2015; 3 (3): 1155-1175.
19. Varshney RK, Graner A, Orrells ME. Genic microsatellite markers in plants: features and applications. TRENDS in Biotechnology. 2005; 23(1): 48-55.
20. Saxena B, Kaur R, Bhardwaj SV. Assessment of genetic diversity in cabbage cultivars using RAPD and SSR markers. J. Crop Sci. Biotech. 2011; 14 (3): 191-196.
21. Abdel-Mawgood AL. DNA based techniques for studying genetic diversity, genetic diversity in microorganisms. 2012; ISBN: 978-953-51-0064-5.
22. Shimelis H, Mashela PW, Hugo A. Performance of *Vernonia* as an alternative industrial oil crop in Limpopo Province of South Africa. Crop Science. 2008; 48(1): 236. doi: 10.2135/cropsci2007.06.0331.
23. Mideksa A, Tesfaye K, Dagne K. *Centrapalus pauciflorus* (Willd.) H. Rob. Neglected potential oil crop of Ethiopia, agro-morphological characterization. Genet. Resour. Crop. Evol. 2018; https://doi.org/10.1007/s10722-018-0719-1.

24. Gilbert JE, Lewis RV, Wilkinson MJ, Caligari PDS. Developing an appropriate strategy to assess genetic variability in plant germplasm collections. Theor. Appl. Genet. 1999; 98:1125–1131.
25. Doyle JJ, Doyle JL. Isolation of plant DNA from fresh tissue. Focus. 1990; 12:13–15.
26. Narina SS, Andebrhana T, Mohamed AI, Bhardwaj HL. Functional classification of ESTs from vernonia (*Vernonia galamensis* L.) cDNA library. Industrial Crops and Products. 2012; 36: 370-375.
27. Peakall R, Smouse PE. GenAlEx 6.503: genetic analysis in excel. Population genetic software for teaching and research-an update. Bioinformatics. 2012; 28: 2537-2539.
28. Perrier X, Jacquemoud-Collet JP. DARwin software. 2006; http://darwin.cirad.fr/ darwin.
29. Andrew R. FigTree: *Tree figure drawing tool*, Version 1.4.4. Institute of Evolutionary Biology, University of Edinburgh. 2018.
30. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multi locus genotype data. Genetics. 2000; 155:945–959.
31. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 2005; 14: 2611-2620.
32. Earl DA, Von Holdt BM. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Cons. Genet. Reso. 2012; 4: 359-361.
33. Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. CLUMPAK: A program for identifying clustering modes and packaging population structure inferences across k. Mole. Ecol. Res. 2015; 15(5): 1179-1191.
34. Nei M. Analysis of gene diversity in subdivided populations. Proce. Nat. Acad. Sci. USA. 1972; 70: 3321-3323.
35. Gilbert KJ, Andrew RL, Bock DG, Franklin MT, Kane N C. et al. Recommendations for utilizing and reporting population genetic analyses: the reproducibility of genetic clustering using the program STRUCTURE. Mol. Ecol. 2012; 1–6.
36. Keneni G, Imtiaz M, Dagne K, Getu E, Assefa F. Genetic piversity and population structure of Ethiopian chickpea (Cicer arietinum L.) germplasm accessions from different geographical origins as revealed by microsatellite markers. Plant Molecular Biology Reporter. 2012; 30(3):654–665. doi: 10.1007/s11105-011-0374-6.
37. Olango TM, Tesfaye B, Pagnotta MA, Pe ME, Catellani M. Development of SSR markers and genetic diversity analysis in enset (Ensete ventricosum (Welw.) Cheesman), an orphan food security crop from Southern Ethiopia. BMC Genetics. 2015; 16(1): 98. doi: 10.1186/s12863-015-0250-8.
38. Gadissa F, Tesfaye K, Dagne K, Geleta M. Genetic diversity and population structure analyses of Plectranthus edulis (Vatke) Agnew collections from diverse agro-ecologies in Ethiopia using newly developed EST-SSRs marker system. BMC Genetics. 2018; 19(1): 92. doi: 10.1186/s12863-018-0682-z.
39. Adugna A. Analysis of in situ diversity and population structure in Ethiopian cultivated *Sorghum bicolor* (L.) landraces using phenotypic traits and SSR markers. Springer Plus. 2014; 3(1): 212. doi: 10.1186/2193-1801-3-212.
40. IPGRI and Cornell University. Genetic Diversity Analysis with Molecular Data: Learning module. 2003; 1-71.
41. El-Esawi M. et al. Analysis of the genetic diversity and population structure of Austrian and Belgian wheat germplasm within a regional context based on DArT markers. Genes. 2018; 9(1): 47. doi: 10.3390/genes9010047.

42. Mondini L, Noorani A, Pagnotta MA. Assessing plant genetic diversity by molecular tools. Diversity. 2009; 1: 19-35; doi: 10.3390/d1010019.

# Tables

**Table 1**. *Vernonia galamensis* samples used in the current study, indicating regional states, sample size, altitude ranges and co-ordinates

| Population | Sample size | Altitude ranges (m) | Co-ordinates | |
|---|---|---|---|---|
| | | | Latitude ranges | Longitude ranges |
| Borena | 15 | 1090-1200 | 4°88'- 4°90'N | 39°35'- 39°40'E |
| East Showa | 15 | 1630-1643 | 7°56'- 8°90' N | 38°43'- 39°70'E |
| West Arsi | 15 | 2000-2143 | 7°15'- 8°90' N | 38°38'- 38°42'E |
| East Harerghe | 16 | 1574-2750 | 9°06'- 9°25' N | 41°25'- 41°38'E |
| West Harerghe | 14 | 1393-1889 | 8°56'- 9°13' N | 40°52'- 42°27'E |
| West Gojjam | 14 | 1205-2560 | 10°27'- 10°30' N | 38°12'- 39°09'E |
| South Wollo | 16 | 1866-2630 | 9°51'- 11°08'N | 39°10'- 39°38'E |
| Sidama | 15 | 1708-1780 | 6°51'- 7°15' N | 37°45'- 38°27'E |
| Konso | 15 | 1500-1650 | 5°15'- 5°20' N | 37°27'- 37°40'E |
| Derashie | 15 | 1395-1450 | 6°18'- 6°25' N | 36°53'- 37°09'E |

**Table 2.** Primer sequences, annealing temperature and product size that used for genetic diversity studies in *Vernonia galamensis*

| SSR primers | SSR motifs (5'-3') | Forward Sequence (5'-3') | Reverse Sequence (5'-3') | Annealing tem | Product Size |
|---|---|---|---|---|---|
| Vg-001 | (AT)12 | CTTGATTTTGTGGGGACCTAAGTG | TAGGAATGGAATAGAATGGATCGG | 56 | 135 |
| Vg-002 | (TC)12 | GGGTTGTGGGGAGAGATAGAGATA | AGCCAAGTTACGCATAGACATCTG | 56 | 146 |
| Vg-003 | (TC)24 | GTGAGCGGGGATCTTCACTTC | GAGAAAGCGAGCATCAACAGACTT | 56 | 142 |
| Vg-004 | (CCA)12 | ACCATACAGTCCCGCATGAATATC | GCTCCTGGAAATGGAGGATAGAAT | 58 | 145 |
| Vg-005 | (AAG)12 | AGCTTAAACAAGAACAAACCGCTG | TGCGAAGGCTTACCAGTTACAAAC | 56 | 157 |
| Vg-006 | (ACA)12 | ATCAGCGTTGCTTGAAAAGAGTG | AAACTCATCGCTCAAACTCAAACG | 58 | 101 |
| Vg-007 | (TCT)18 | ATGACGATGCAACTCACCGTT | CGGAGAGGTTTGGTAGGGTAAGAT | 56 | 136 |
| Vg-008 | (GAT)24 | ATGTCTTCCAAATCGAGGATGGTA | AATTTTTGCAGCTAAGCCAGTGAG | 58 | 101 |
| Vg-009 | (GGA)15 | ATTGAAGGATGAACGGACAGAGTC | GTATCACATCACGTCGTCCACATC | 58 | 127 |
| Vg-010 | (ATG)15 | CAAAGGGAAGATGCACCTAGAGAA | ATCAAACCTGCTGCTTTTCAAGTC | 58 | 130 |
| Vg-011 | (TGA)12 | GCACAATCAGACTTGAGACCAAGA | GCAGTGATCAGCCATAGTGCATAC | 58 | 130 |
| Vg-012 | (CGC)12 | GGGCTGAGCAAATACAGCAGAC | AGGATCTTCTTGTTGGTGTGGAAA | 58 | 150 |
| Vg-013 | (CAG)15 | GGGGCGTTTCCTTGATTTTG | CTCTTCACCTGCCATCTTTTCTGT | 56 | 93 |
| Vg-014 | (CAA)12 | GTAGCAGCAGCAGTTCACTACCAC | CAAAATCCTCACAAACTTCACACG | 56 | 132 |
| Vg-015 | (GGC)18 | GTGCTACAACGGTGGTACATCAAG | TCATTGATTCCATGCTGAAATAGC | 56 | 159 |
| Vg-016 | (GGT)12 | GTTAGAGATGGGTTTGAAGAGCGA | CCTTACCAACTCCAACACCACTTG | 58 | 139 |
| Vg-019 | (CTTCAC)24 | GGGTCTCCATCTATTCACCTTCAA | AAGGAGCGTGAGCTAGAAGAAGC | 56 | 158 |
| Vg-021 | (GTC)15 | TGAAGAAGAAGGTTTCCCAAATCA | GATGCATTGACATCACGTAGAAGC | 56 | 155 |
| Vg-024 | (ATC)15 | TTGGATGTCGAAAAAGATGAGGTT | TTCTCCCTCTGTTTCAACACCTTC | 56 | 144 |
| Vg-030 | (CT)12 | TCAAACACACTCCCCAATTTCTCT | GCTGCCGATTGATCAAATTACACT | 56 | 100 |

**Table 3**: Summary of genetic parameters revealed by using 20 SSR markers for *Vernonia galamensis* populations collected from different regions of Ethiopia

| Locus | MAF | Na | Ne | I | Ho | He | uHe | F | PIC |
|---|---|---|---|---|---|---|---|---|---|
| Vg-01 | 0.85 | 3 | 2.80 | 1.06 | 0.25 | 0.48 | 0.49 | 0.25 | 0.80 |
| Vg-02 | 0.77 | 5 | 3.26 | 1.36 | 0.13 | 0.50 | 0.51 | 0.73 | 0.96 |
| Vg-03 | 0.45 | 6 | 4.79 | 1.67 | 0.36 | 0.65 | 0.64 | 0.77 | 0.62 |
| Vg-04 | 0.52 | 3 | 2.28 | 0.90 | 0.07 | 0.59 | 0.61 | 0.88 | 0.64 |
| Vg-05 | 0.64 | 3 | 2.01 | 0.89 | 0.13 | 0.60 | 0.63 | 0.78 | 0.50 |
| Vg-06 | 0.54 | 5 | 3.69 | 1.44 | 0.13 | 0.44 | 0.45 | 0.70 | 0.86 |
| Vg-07 | 0.61 | 4 | 3.46 | 1.31 | 0.20 | 0.58 | 0.60 | 0.66 | 0.57 |
| Vg-08 | 0.59 | 5 | 3.26 | 1.36 | 0.20 | 0.61 | 0.63 | 0.67 | 0.89 |
| Vg-09 | 0.61 | 3 | 2.60 | 1.01 | 0.23 | 0.45 | 0.47 | 0.49 | 0.87 |
| Vg-10 | 0.60 | 4 | 3.81 | 1.36 | 0.07 | 0.46 | 0.48 | 0.86 | 0.86 |
| Vg-11 | 0.56 | 4 | 2.03 | 0.95 | 0.13 | 0.23 | 0.24 | 0.42 | 0.93 |
| Vg-12 | 0.57 | 3 | 2.53 | 1.01 | 0.07 | 0.46 | 0.48 | 0.86 | 0.84 |
| Vg-13 | 0.49 | 4 | 3.57 | 1.33 | 0.13 | 0.48 | 0.50 | 0.72 | 0.73 |
| Vg-14 | 0.60 | 3 | 2.27 | 0.95 | 0.27 | 0.48 | 0.50 | 0.44 | 0.77 |
| Vg-15 | 0.76 | 4 | 2.68 | 1.16 | 0.08 | 0.50 | 0.52 | 0.85 | 0.88 |
| Vg-16 | 0.64 | 2 | 1.99 | 0.86 | 0.07 | 0.36 | 0.37 | 0.81 | 0.78 |
| Vg-19 | 0.58 | 4 | 3.46 | 1.31 | 0.33 | 0.64 | 0.67 | 0.48 | 0.70 |
| Vg-21 | 0.51 | 5 | 4.41 | 1.55 | 0.05 | 0.60 | 0.62 | 0.89 | 0.86 |
| Vg-24 | 0.67 | 5 | 3.81 | 1.46 | 0.27 | 0.62 | 0.65 | 0.57 | 0.53 |
| Vg-30 | 0.61 | 3 | 2.53 | 1.01 | 0.07 | 0.28 | 0.29 | 0.76 | 0.55 |
| Mean | 0.61 | 3.9 | 3.06 | 1.20 | 0.16 | 0.50 | 0.52 | 0.68 | 0.76 |

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

**Table 4**. 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% | $F_{st}$ | 0.101 | 0.001 |
| Among Individual | 140 | 1842.01 | 13.16 | 6.39 | 67% | $F_{is}$ | 0.946 | 0.001 |
| Within individual | 150 | 55.21 | 0.37 | 0.37 | 22% | $F_{it}$ | 0.951 | 0.001 |
| Total | 299 | 2219.93 | | 7.52 | 100% | Nm | 2.24 | |

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

**Table 5**. Pairwise population matrix of Nei genetic distance (GD) (above diagonal) and pairwise population matrix of Nei genetic identity (GI) (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 Harerghe, WHG = West Harerghe, GOJ = Gojjam, WOL = Wollo, KON = Konso, DER = Derashie

# Figures

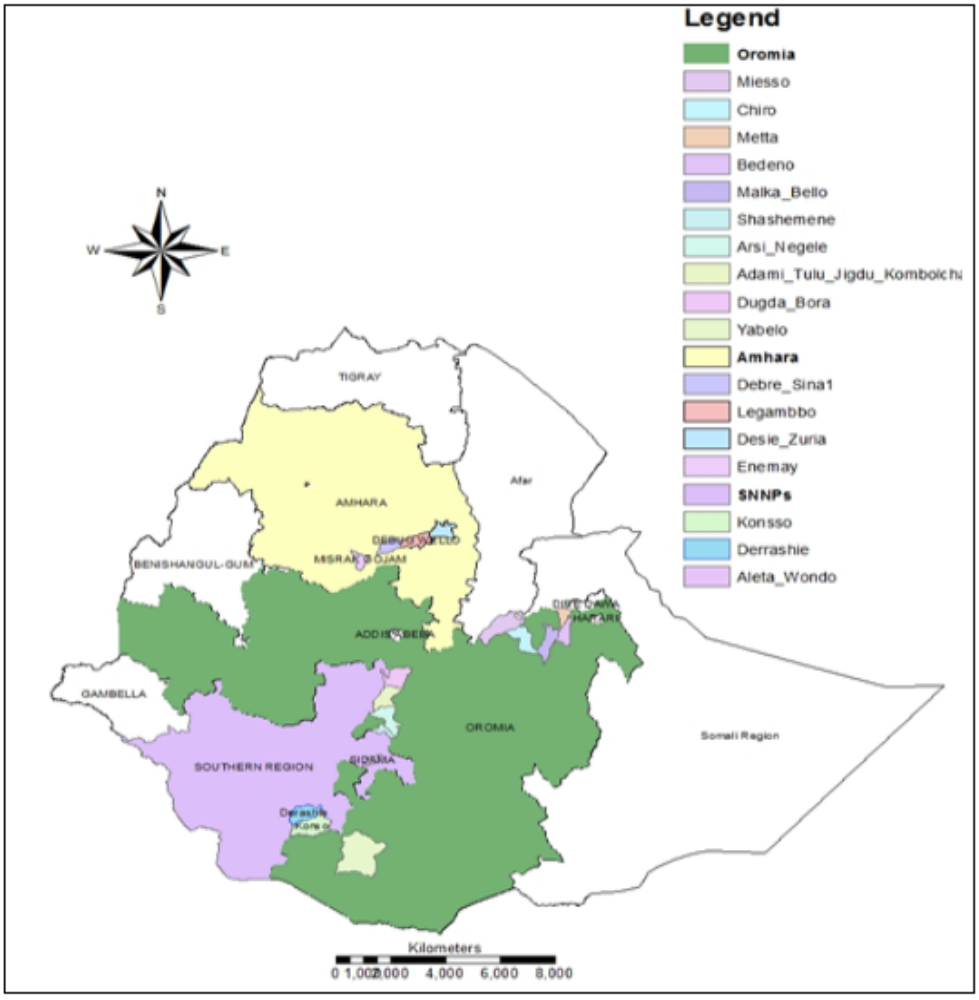



**Figure 1**

Map of Ethiopia showing Vernonia galamensis collection sites within three of the Federal Regions. The map was original and constructed using geographic coordinates and elevation data gathered from each collection sites using

global positioning system (GPS) that 4represent the 10 populations.

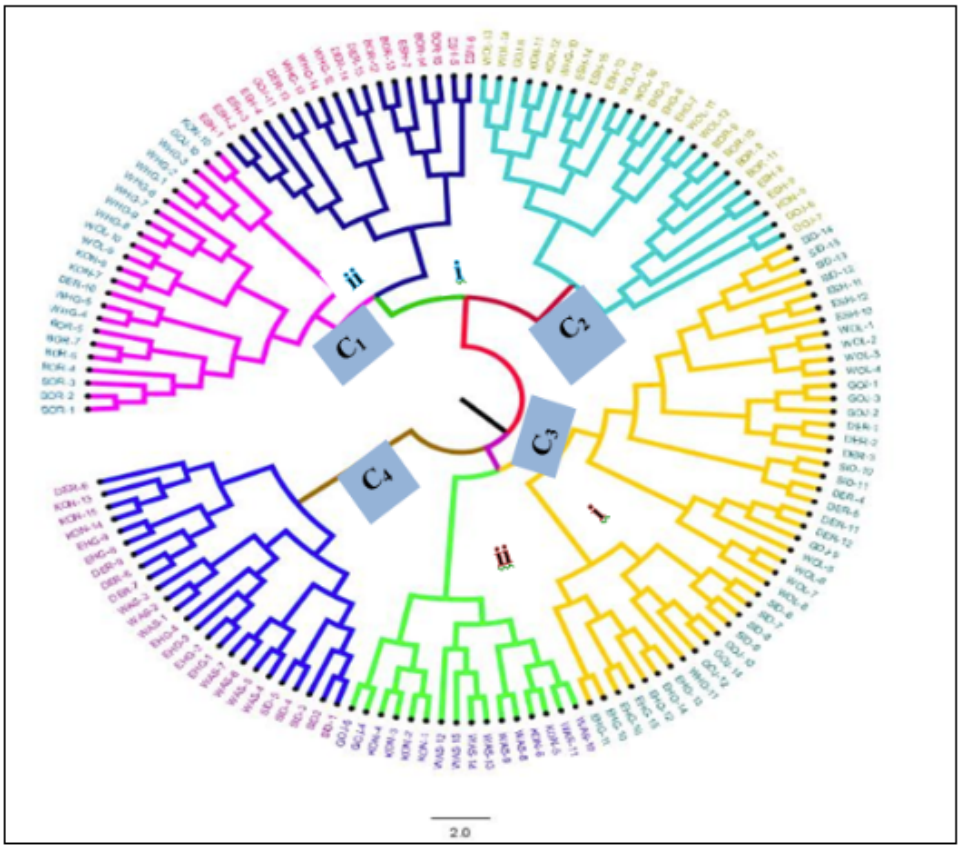
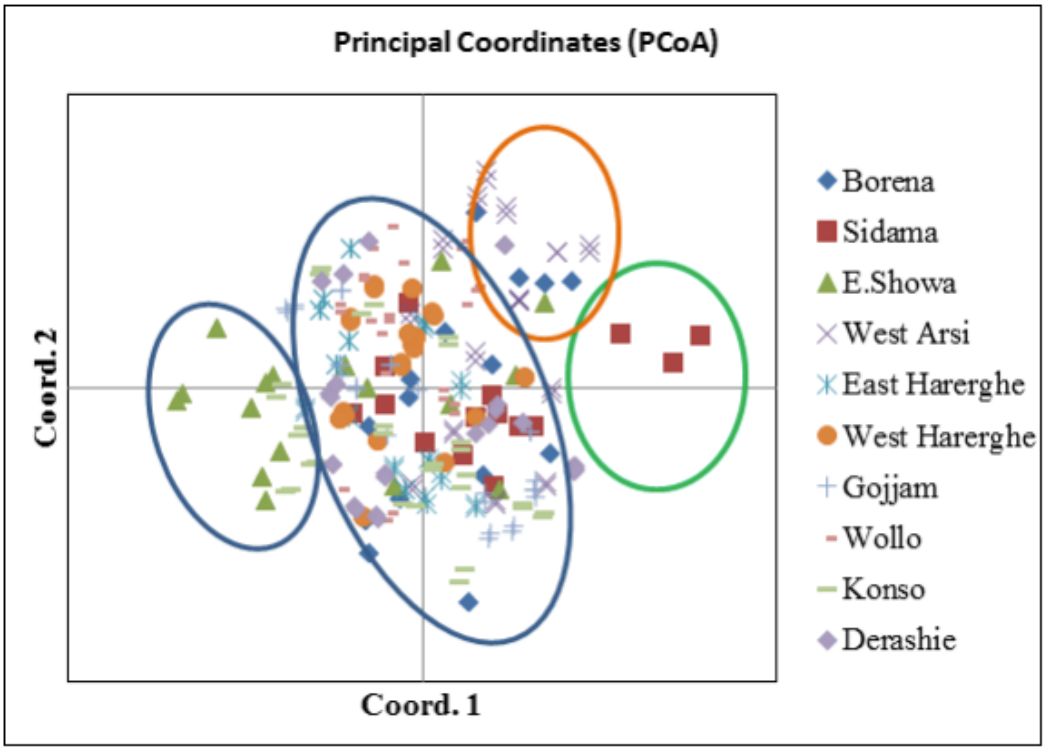


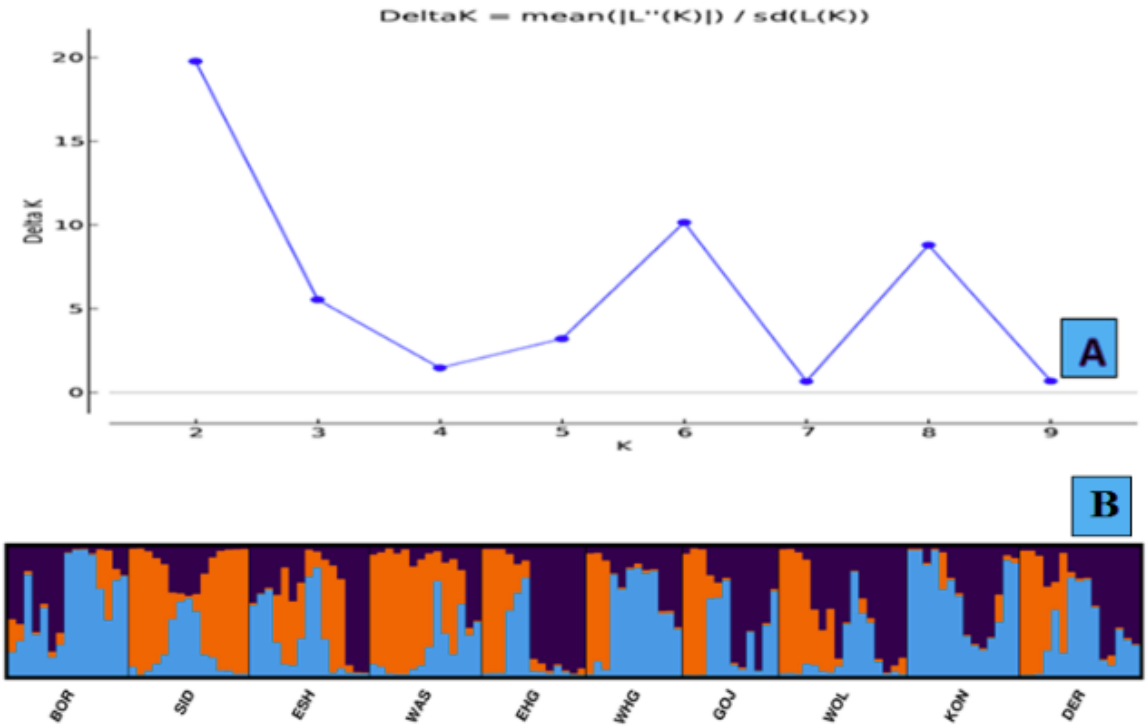

**Figure 2**

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 Harerghe, WHG = West Harerghe, GOJ = Gojjam, WOL = Wollo, KON = Konso, DER = Derashie




## Figure 3

Two dimensional scaling principal coordinate analysis of 150 Vernonia galamensis accessions of ten populations.




## Figure 4

Delta K values estimated according to Evanno et al. 2005 Method (A) and Bayesian model-based estimation of population structure (K = 6) (B) for 150 Vernonia galamensis accessions from different growing regions of Ethiopia.

# Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfiles.docx