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
Kassahun Tesfaye
Addis Ababa University College of Natural Sciences
Kifle Dagne
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

DOI: https://doi.org/10.21203/rs.2.18928/v3

Abstract

Background: Vernonia (*Vernonia galamensis*) is a potential novel industrial crop due to high demand for its natural epoxidized oil, which can be used for the manufacturing of oleochemicals such as paints, plastic formulations (polyvinyl chloride), and pharmaceutical products. Assessment of the extent of genetic diversity in *V. galamensis* was the first and the most step in cultivation, breeding and genetic resource conservation. Hence, this study was aimed to assess the genetic diversity and population structure of this plant from different agro-ecologies in Ethiopia using SSR markers.

Results: Twenty SSR markers were used for genetic diversity analyses of 150 individual *V. galamensis* accessions representing 10 populations, from which a total of 79 bands were identified for the 20 loci. All the loci used showed high polymorphism that ranged from 0.50 to 0.96, while the mean observed heterozygosity (Ho) was 0.15 across all the 20 markers evaluated. The molecular variance analysis (AMOVA) showed significant variations among populations which accounted for 11% of the variations. Populations clustering showed that the dendrogram and principal coordinate's analysis roughly classified the accessions into four groups. However, the Bayesian model-based clustering (STRUCTURE) grouped into 6 (K = 6) major gene pools. Since, the cluster and the STRUCTURE analyses did not group the populations into sharply distinct clusters, due to presence of gene flow and mode of reproduction of the plant.

Conclusions: The SSR molecular markers used in this study are highly polymorphic. Among the ten populations, East Showa and East Hararghe revealed higher genetic diversity, signaled that these areas are 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 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 the high demand for its natural epoxidized oil [1-5]. *Vernonia galamensis* subsp. *galamensis* variety *ethiopica* M. Gilbert was first identified by Perdue in 1964 in Eastern Ethiopia [6, 7]. Seeds of *Vernonia* species are the major sources of naturally occurring epoxidized fatty acids and other essential fatty acids such as linoleic acid, oleic acid, palmitic acid, stearic acid and trace amounts of arachidic acid [9, 10, 11-14].

Studying genetic diversity of crop plants including *V. galamensis* is a valuable tool for effective utilization, conservation, management, and improvement during breeding and selection [15, 16]. Molecular markers are valuable tools for discrimination between accessions, easy and cost-effective to measure and not influenced by environmental factors. Microsatellites (SSR) markers have efficient advantages over other DNA markers that they are highly polymorphic, co-dominant inheritance with high mutation rates, relatively abundant and good genome coverage, which used to investigate taxonomical, genetic diversity and evolutionary relationship [17, 18]. However, SSR markers 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 [19]. To date, this study was the first report that used SSR markers to study the genetic diversity of *V. galamensis*.

In Ethiopia, geo-ecological conditions are favorable for the cultivation of *V. galamensis* and, use as a source of raw material for agro-processing industries. However, the plant is neglected and considered only as a wild weed colonizing disturbed 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 genetic diversity evaluation and characterization of *V. galamensis* accessions by molecular analysis using SSR markers, which serve as the basis for cultivation, breeding and sustainable conservation of the plant.

Methods

Plant material

A total of 150 *V. galamensis* accessions, representing 10 populations, were randomly collected from their diverse agro-ecologies of the three-potential growing regional states of Ethiopia (Figure 1; Table 1). The samples identified were confirmed with the descriptions

available in the Flora of Ethiopia and Eritrea. Most of the study materials were collected from the field and, others were assembled from the Ethiopian Biodiversity Institute and Wondo Genet Agricultural Research Center.

At each collection area, seed samples were collected from plants and kept in separate bags to ensure that the distance between any two collection sites was about 5-10 Km. From collection areas observations, *V. galamensis* naturally grows in hilly/depression, along the roadside, in the valley, in farmlands, in the forest, in the compounds of mosques and churches. The collections were done by taking either seed samples of the individual flower heads 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 collection sites.

Plant material collection and DNA extraction

Fresh young leaves of 150 *V. galamensis* accessions were collected from individual plants that were grown at experimental sites, representing 10 populations. Collected leaf samples were 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 used for later DNA extraction according to Gilbert et al. [20]. 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 [21]. 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.

Polymerase Chain Reaction (PCR) Amplification

The twenty SSR markers used for this study were selected based on their high polymorphism from the work of Narina et al. [22] (Table 2). 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 3% agarose gel electrophoresis for visualization.

Band Scoring and Analysis

The amplified products were visually scored based on their migration in comparison with the standard size DNA (100 bp DNA ladder) photographed under UV gel illumination (Gel Doc^{TM} with Image Lab^{TM} software, BIO-RAD). The genetic diversity for each allele 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 the 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 [23].

Similarity test, Unweighted Pair Group Method with Arithmetic Mean (UPGMA) and Neighbor-Joining (NJ) tree were computed using DARwin version 6.0.19 software [24], and the significance was tested based on 1000 bootstraps. The resulting trees were displayed using Fig Tree version 1.4.4 [25]. A Bayesian model-based cluster analysis was performed using STRUCTURE version 2.3.4 software [26]. Each of the groups of populations composed of accessions collected from different administrative zones/geographic regions of origin. 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. [27] using the web-based (http://tyloro.biology.ucla.edu\structure Harvester\) STRUCTURE HARVESTER ver. 0.6.92 [28]. The results generated by this software were visualized in a graphical bar plot using Clumpak beta version (http://www.clumpak.tau.ac.il/) [29].

Results

SSR Polymorphism

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). The results of diversity parameters showed a high level of polymorphism among the 20 SSR markers. Polymorphic information content (PIC) 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 (Table 3). A total of 79 alleles were identified, varied from 2 to 6 with an average of 3.95 alleles per locus. The maximum number of effective alleles (Ne) was 4.79 (Vg-003) and the least number of effective alleles was 1.99 (Vg-016). The highest major allele frequency (MAF) (0.85) was recorded by locus Vg-001 and the least MAF was (0.45) recorded by locus Vg-003. The observed heterozygosity (Ho) values were quite low ranging 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 (Table 3). Fixation index (F) ranged from 0.25 (Vg-001) to 0.89 (Vg-021) with an average of 0.68. Finally, Shannon-Weaver's information indices (I) ranged from 0.86 to 1.67 (Table 3). On the other hand, East Showa and East Hararghe had the highest in effective alleles, heterozygosity, genetic diversity (Shannon diversity index) and fixation index (Table 4).

Analysis of molecular variance (AMOVA)

The analysis of molecular variance (AMOVA) was conducted using accessions grouped according to their administrative regions, and revealed that 67% of the variation was attributed to genetic variability within populations, 11% was variation among populations and 22% was due to variation among individuals within the same population (Table 5).

Genetic distance between populations

The Nei's [30] standard genetic distance (GD) between populations ranged from 0.24 to 0.59. The highest pairwise genetic distance was observed between populations of Sidama and East Showa (0.59), followed by populations of Borena and East Hararghe (0.57). The minimum pairwise genetic distance was observed between populations of Borena and Konso (0.24) (Table 6).

Cluster, principal co-ordinate analysis (PCoA) and population structure

UPGMA cluster analysis was performed using neighbor-joining, with the DARwin 6.0.19 software programs, genetic distance-based analysis, and grouped the 150 accessions into four (4) major clusters. Each of the four clusters comprises mixed of accessions collected from different administrative zones (geographic regions). The first and the fourth cluster further divided into sub-clusters according to their geographic origin (Figure 2).

The first cluster constituted 39 accessions, the second cluster contained 42 accessions, while cluster three was characterized as the smallest group in clustering comprising of 15 accessions and the fourth cluster was the major group of clustering and comprised of 54 accessions (Figure 2). The cluster analysis revealed that accessions from different populations (collection sites) clustered together with no clear pattern of geographic origins. The principal coordinate 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 together accounted for 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 ranged from K = 1 to K = 10 and 20 iterations for each K. According to Evanno et al. [27] and Gilbert et al. [31], STRUCTURE outputs were predicted K = 6, 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 group. There was a wide admixture in structuring of *V. galamensis* populations.

Discussions

Determination of SSR-markers based genetic diversity with genetic parameters

Vernonia (*Vernonia galamensis*) is a plants species that contains naturally occurring epoxidized oils in its seeds. However, its potential values are neglected, underestimated and underexploited. In addition, it is also exposed to genetic erosion. Therefore, assessment of genetic diversity with SSR markers generally in plants and particularly in *V. galamensis* is important for cultivation, breeding and genetic resource management. A total of 79 alleles with an average of 3.9 alleles were detected in this study, which was higher than the one reported by Ramalema et al. [13] (3.14) in *V. galamensis*. The author also stated that less genetic diversity was observed among different *Vernonia* lines. According to Aikpokpodion et al. [32], a total of 29 bands were detected in genus *Vernonia*. Hence, the

large number of alleles detected in the present study indicated the suitability of microsatellites for genetic linkage and QTL mapping of desirable traits applied to marker-assisted selection (MAS) in breeding programmes.

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 [33]. In this study, the PIC values ranged between 0.50 (Vg-005) and 0.96 (Vg-002) with an average of 0.76. Ramalema et al. [13] reported the mean PIC value 0.40, which was lower than our result. Microsatellite markers such as Vg-002 and Vg-011 showed the highest polymorphism with 0.96 and 0.93, respectively. Aikpokpodion et al. [32] also reported the PIC value ranged from 0.27 to 0.78 in *V. galamensis* accessions, again little lower than the present study. This indicates that the majority of markers were able to distinguish differences among the studied *V. galamensis* accessions. For most of the loci, expected heterozygosity 0.50 values were higher than that of observed heterozygosity (*0.16*), revealing a high homozygosity at the given loci among the accessions. The diversity parameters showed high level of polymorphism among the 20 SSR markers, favoring the genetic variation within *V. galamensis* collection.

Genetic Differentiation and Gene Flow

AMOVA demonstrated that *V. galamensis* had low variation among the populations (11%). On the other hand, 67% of the total variation was attributed to genetic variability within the populations and 22% was due to variation among individuals within the same population. Similarly, Ramalema [13] reported 40% of genetic variation due to among different *vernonia* lines. Aikpokpodion [32] reported 30% of variation due to among the studied samples and 70% of variation due to within the studied samples. Nwakanma [34] reported that 36% of variation due to among species and 64% were within species variations. According to IPGRI and Cornell University [35], *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 moderate (F_{st} = 0.101), the result of STRUCTURE and F_{st} showed a strong gene flow among populations that collected from different regions. This may be self-pollination in nature, should be carefully considered in future accessions sampling and collection regarding biodiversity evaluation and conservation in *V. galamensis.*

Genetic distance is the measure of the allelic substitutions per locus that have occurred during the separate evolution of two populations. In this study, the largest genetic distance was observed between Sidama and East Showa (0.59) populations, while the minimum genetic distance was observed between Borena and Konso (0.24). The overall magnitude of the pairwise population matrix of Nei genetic distance was relatively low [30].

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

In the present study, a dendrogram 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 comprised of 39 accessions, the second cluster characterized as the second major clustering, contained 42 accessions, the third cluster composed of 15 accessions, and the fourth groups consisted of the major clustering which comprised 54 accessions that were collected from different regions of origin. The cluster analysis revealed a week clustering pattern. Hence, accessions collected from different geographic regions/zones of origins were clustered together, since clusters did not follow a clear pattern of geographic origins, which may due to the presence of strong gene flow, or epistatic gene interaction which may require more investigation. Similarly, Nwakanma et al. [34] reported that 49 *Vernonia* lines fingerprinted were grouped into four major clusters with no clear-cut separation among accessions related to their origin. Aikpokpodion et al. [32] also reported that genus *Vernonia* grouped into four major clusters and showed the existence pattern of relationships between geographical origins and genetic diversity.

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 majority of samples were placed at the center of a two-dimensional coordinate plane and roughly forms four groups with a total variation of 30.04%. This, in turn, agrees with the results of the NJ dendrogram in which there was no unique clustering among accessions from the same population/collection areas. The presence of gene flow among populations/collection areas, accompanied by wind dispersal to a long distance due to the hairy pappus at the tip of seed, may be the most probable explanation behind the mixed clustering of accessions. Although UPGMA and PCoA analyses also showed a certain level of population clustering according to their geographical regions, the clustering pattern

is weak, not clustered by their regions of origin. Aikpokpodion et al [32] reported geographical differentiation observed among *Vernonia* accessions, indicates the plant isolation by distance, and may be due founder effect as a result of genetic drift as well as a local adaptation. Founder effect is the loss of genetic variation due to the establishment of a new population by a very small number of individuals from the original population.

Population's genetic structure in *Vernonia galamensis*

The structure analyses of 150 *V. galamensis* accessions using a Bayesian-model based approach with the highest *ΔK* Value, according to Gilbert et al. [31] and Evanno et al. [27] method. The samples from 10 collection zones were inferred into six (K = 6) structure outputs. The structure analysis revealed a close relationship (weak sub-division) of the population from 10 collection zones, and in general, six inferred groups (K = 6) with potential admixtures have been observed. It is interesting to indicate that all individual plants analyzed have alleles originated from the six clusters, which supports the presence of a strong gene flow that led to low population differentiation. Hence, accessions collected from the same region of origin did not often group entirely together within a given major groups. There was a high level of population admixture in the structuring of *V. galamensis* populations, and population genetic structure study using SSR markers, which is the first report in *V.galamensis*. Most of the magnitudes identified for the SSR markers were important information for cultivation, breeding and genetic resource conservation. Nwakanma et al. [34] reported clustering in *Vernonia* species may have arisen from the same putative parents or may have diversed from each other in the course of their evolutionary adaptation in different agro-ecology. In addition, huge intra-species relationships were existed within the species due to ecotypes variation. 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 [33].

The genetic variations across the ten populations showed that populations from East Showa and East Hararghe had the highest number of effective alleles, heterozygosity and Shannon's diversity index. This 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. 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 [36].

Conclusions

The present genetic diversity analysis by using SSR markers was the first report in *V. galamensis*. All the twenty markers used were polymorphic in the population studied. The SSR markers 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 populations/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 this plant. Most of the values identified for the SSR markers were important baseline information for future *V. galamensis* cultivation and breeding/genetic resource conservation endeavors in Ethiopia.

Additional File Information

Additional file 1: Table S1. Pass port data of 150 *Vernonia galamensis* samples used in the current study.

Additional file 2: Figure S1. Unweighted neighbor joining based clustering of 150 *Vernonia galamensis* accessions for 20 polymorphic SSR markers.

Additional file 3: Figure S2. AMOVA variation pie chart for 150 *Vernonia galamensis* accessions from ten populations 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

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

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.

Competing interests

The authors declare that they have no competing interests.

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.

Authors' contributions

AM, KT and KD 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.

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

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. 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.
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.
5. 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.
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. 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.
10. 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.
11. 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.
12. 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.
13. 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.
14. 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.
15. Idrees M, Irshad M. Molecular markers in plants for analysis of genetic diversity*:* European Academic Research. 2014; 1 (1).
16. 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.
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. Varshney RK, Graner A, Orrells ME. Genic microsatellite markers in plants: features and applications. TRENDS in Biotechnology. 2005; 23 (1): 48-55.
19. 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.
20. 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.
21. Doyle JJ, Doyle JL. Isolation of plant DNA from fresh tissue. Focus. 1990; 12:13–15.
22. 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.
23. 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.
24. Perrier X, Jacquemoud-Collet JP. DARwin software. 2006; http://darwin.cirad.fr/ darwin.
25. Andrew R. FigTree: *Tree figure drawing tool*, Version 1.4.4. Institute of Evolutionary Biology, University of Edinburgh. 2018.

26. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multi locus genotype data. Genetics. 2000; 155:945–959.
27. 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.
28. 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.
29. 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.
30. Nei M. Analysis of gene diversity in subdivided populations. Proce. Nat. Acad. Sci. USA. 1972; 70: 3321-3323.
31. 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.
32. Aikpokpodion PO, Abebe J, IGWE D. Genetic diversity in *Vernonia amygdalina* Delile accessions revealed by random amplified polymorphic DNAs (RAPDs). Journal of Biotechnology. 2018; 99(2): 147–152.
33. Guo S, Thompson EA. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics. 1992; 48: 361–372.
34. Nwakanma NM, Oboh BO, Adekoya KO, Ogunkanmi LA. Genetic Diversity of *Vernonia* as Revealed By Random Amplified Polymorphic DNA (RAPD) Markers. Egypt. Acad. J. Biolog. Sci. 2018; 9(1): 27-38. ISSN 2090-3812
35. IPGRI and Cornell University. Genetic Diversity Analysis with Molecular Data: Learning module. 2003; 1-71.
36. Chandel KPS, Joshi BS. Multivariate analysis in green seeded pea. Indian Journal Agricultural Sciences. 1983; 53:198-200.

Tables

Table 1. *Vernonia galamensis* samples used in the current study

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 amplicon sizes which were used for genetic diversity studies in *Vernonia galamensis*

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

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

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	4	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	3.95	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. 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
E. Harerghe	16	4.00	3.68	0.79	0.12	0.49	0.53	0.73
West Harerghe	14	4.00	3.12	0.77	0.16	0.48	0.50	0.73
Gojjam	14	3.00	2.33	0.70	0.13	0.47	0.49	0.64
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

Table 5. 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%			

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

Table 6. Pairwise population matrix of Nei standard genetic distance (GD)

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		**	0.59	0.37	0.39	0.54	0.51	0.44	0.50	0.36
ESH			**	0.49	0.44	0.47	0.34	0.29	0.42	0.27
WAS				**	0.44	0.38	0.45	0.36	0.35	0.47
EHG					**	0.47	0.47	0.40	0.39	0.25
WHG						**	0.41	0.35	0.31	0.41
GOJ							**	0.36	0.37	0.38
WOL								**	0.37	0.42
KON									**	0.35
DER										**

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

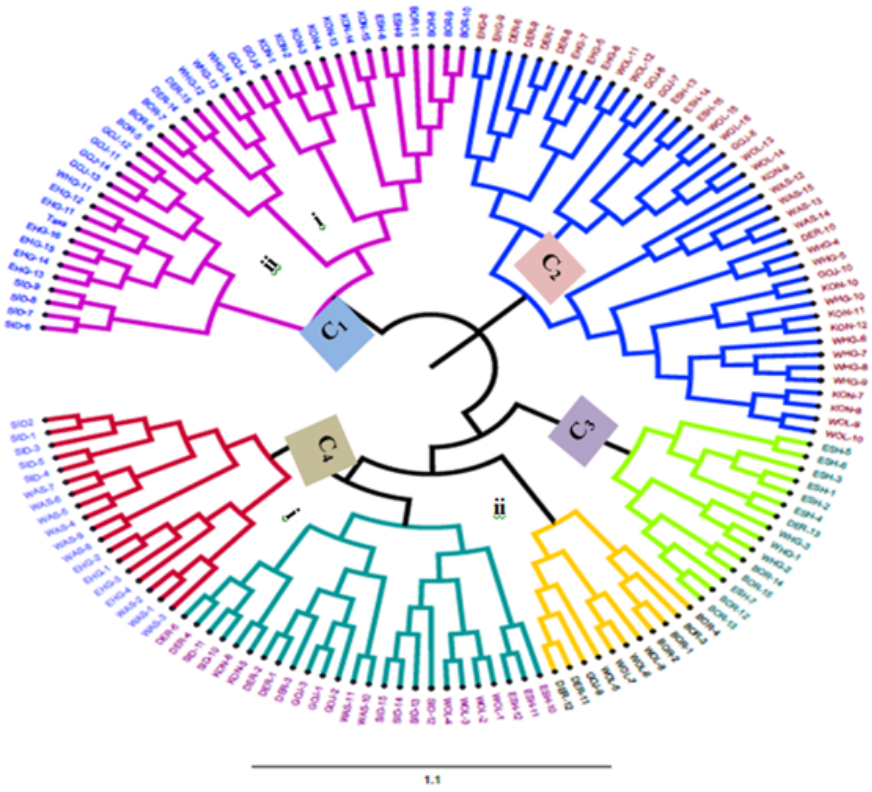

Figure 2

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