Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Genetic Diversity and Population Structure Analysis of Grass pea (*Lathyrus Sativus* L.) Accessions Collected from North-Western Ethiopia Using SSR Markers

Desalegn Abebe (desalegn.abebe16@arari.gov.et)
Amhara Regional Agricultural Research Institute https://orcid.org/0000-0001-7929-5638
Adane Abraham
Botswana International University of Science and Technology
Happiness Oselebe
Ebonyi State University Faculty of Agricultural Sciences
Celestine Afiukwa
Ebonyi State University Faculty of Biological Sciences
Oluyinka Ilesanmi
International Institute of Tropical Agriculture
Tiegist Dejene Abebe
Bahir Dar University College of Agriculture and Environmental Sciences

Research Article

Keywords:

Posted Date: April 7th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-397389/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

Grass pea (Lathyrus sativus L.) is a legume crop known to be an excellent source of protein, tolerant to drought, waterlogging, and salinity. The crop is used as an alternative source of protein to reduce malnutrition for resource-poor people and farmers leaving in marginal areas. However, due to the presence of a neurotoxin that causes lathyrism in the crop, it has been neglected and underutilized. As a necessary first step towards, therefore, this investigation was undertaken to assess the genetic diversity and population structure that existed within grass pea accessions collected from the North-Western part of Ethiopia using simple sequence repeat markers. Twenty-five grass pea accessions collected from the Ethiopian Biodiversity Institute were planted at the College of Agricultural Science, Ebonyi State University, Nigeria. The genomic DNA was extracted using Quick-DNA™ ZR Plant/Seed Miniprep Kit and amplified by ABI Veriti PCR with 10 pairs of SSR markers in IITA, Ibadan, Nigeria. Out of 10 SSR primers, only eight primers were polymorphic. A total of 41 alleles were detected with an average of 5.13. The Polymorphic information content and gene diversity values ranged from 0.074 to 0.944 with a mean of 0.474 and 0.536 respectively. The largest pairwise genetic distance (0.365) was detected between North Gondar and East Gojam populations. The ΦPT (analogs of FST test) estimated through AMOVA were 0.24, 0.20, and 0.05 for within accessions, among regions, and population within regions respectively. The highest genetic differentiation value (76%) resides within accessions followed by 20% among regions. Both population structure and cluster analysis grouped the 25 grass pea accessions into two distinctive subgroups. This grouping pattern indicates the presence of gene flow among geographic regions. In general, the findings of this study indicate that despite few numbers of SSR markers it was possible to detect genetic diversity among grass pea accessions indicating the power of the SSR marker in picking up the existing genetic diversity within the studied accessions.

Introduction

Grass pea (*Lathyrus sativus* L.) is the most important legume crop in Ethiopia which is rich in protein (28%-32%); and micronutrients, and used for the human diet and animal feed (Urga *et al.,* 2005). Worldwide, 1.2 M t of grass pea is produced from ~ 1.5 M ha of land. Ethiopia is the third-largest producer of the crop next to Bangladesh and India (Gupta *et al.,* 2018). The genus *Lathyrus* has as large as 187 species and can be grown in temperate and tropical areas of the world (Soren *et al.,* 2020; Wang *et al.,* 2015). The exact center of origin is not known but it has many centers of diversity as indicated by different authors, South-west and Central Asia, Mediterranean, Iran-Turanian regions, Bangladesh and East Africa mainly in Ethiopia (Dixit *et al.,* 2016; Parihar *et al.*, 2015, 2013; Wang *et al.,* 2015; Talukdar, 2009). Ethiopia is also considered one of the centers of diversity for grass pea (Girma and Korbu, 2012). Recently, there are 578 grass pea accessions in the collection preserved at the Ethiopian Biodiversity Institute (EBI), in Addis Ababa, Ethiopia (https://www.ebi.gov.et/biodiversity/conservation/database-ms/). It is the fifth major grain legume following fava bean, field pea, haricot bean, and chickpea regarding production, and area coverage (CSA, 2020). It has many significant traits like drought-tolerance, resistance to insect-pests, adaptability to numerous kinds of soil, and climatic conditions. Its resilient and

penetrating root system allows the crop to grow in an extensive range of soil types including very poor soils and heavy clays. It has been realized that grass pea can serve as a survival foodstuff in difficult situations (Parihar *et al.,* 2013).

Although the crop is marginalized and neglected by governmental research institutes in Ethiopia, it is widely cultivated by farmers as a bonus crop in fallow lands because of its nitrogen-fixing ability and also used as a break crop between rice due to the drought-tolerant nature of the crop (Gupta *et al.,* 2018). Its seed is used as food in Ethiopia, prepared in various forms sauce (Shiro), boiled (nifro), and roasted (kolo) (Urga *et al.,* 2005). Currently, the price of this crop is comparable with other cereal crops in the local market and it is being used as the alternative protein source and food security crop by resource-poor farmers who live in marginal lands in Ethiopia.

Despite the possession of economically important traits, grass pea has the drawback of having a major anti-nutritional compound known as β-N-oxalyl-L-α, β di-amino propionic acid (β-ODAP) also called β-N-oxalyl amino-L-alanine (BOAA), a neurotoxin that causes leg paralysis in animals including humans (Hillocks and Maruthi, 2012). To date, very limited research attention has been given to it for improving this very essential legume plant. The fundamental reason for this less privileged research effort is the presence of β-ODAP. As a result, the crop is marginalized by both governmental and non-governmental organization donor groups (Dixit *et al.,* 2016). Thus, grass pea is among many African orphan crops' with little scientific information available on it, particularly at the molecular level (Tadesse and Bekele, 2003a, b).

Crop characterization based on morphology is generally not reliable due to phenotypic plasticity. Hence, for significant progress in the breeding program of grass pea, the use of molecular marker and genomic knowledge is a pre-requisite (Akter *et al.,* 2015). Molecular markers that have so far been used successfully for genomic studies in *L.sativus* includes random amplification of polymorphic DNA (RAPD) (Barik *et al.,* 2007; Croft *et al.,* 1999), restriction fragment length polymorphism (RFLP) (Chtourou-Ghorbel *et al.,* 2001), amplified fragment length polymorphism (AFLP) (Tavoletti and Iommarini, 2007), inter-simple sequence repeat (ISSR) (Belaid *et al.,* 2006), and simple sequence repeats (SSR) (Wang *et al.,* 2015; Shiferaw *et al.,* 2012; Lioi *et al.,* 2011). In the present investigation, we applied 10 polymorphic SSR markers selected from Wang *et al.,* (2015) to assess and characterize the genetic diversity of 25 North-Western Ethiopian grass pea populations.

Materials And Methods

Plant materials

The study was carried out using 25 grass pea accessions collected from the Ethiopian Biodiversity Institute (EBI), Ethiopia. These accessions mainly originated from four of the major grass pea growing administrative zones of the Amhara region, East Gojjam (seven accessions) South Gondar (six accessions), North Gondar (six accessions), and West Gojjam (six accessions), and the detailed passport data is presented below (Table 1).

DNA extraction

Fresh leaves of five individual plants were collected in each of the 25 grass pea accessions grown for three weeks in plastic bags at Ebonyi State University, College of Agricultural Science (EBSU, CAS) campus, Abakaliki, Nigeria, and dried with silica gel before DNA extraction. About, 100 mg of dry leaves were ground in buffer with mortar and pestle. Genomic DNA was extracted using Quick-DNA™ ZR Plant/Seed Miniprep Kit (Zymo Research, California, USA). DNA quality and quantity were assessed by the nano-drop apparatus followed by 1% agarose gel electrophoresis using Ethidium Bromide staining.

SSR markers screening

Ten SSR markers, based on polymorphic information content (PIC) values ranging from 0.4107 to 0.7292, were selected from published SSR markers for *L.sativus* and related species which were developed by Yang *et al.*, (2014) and used by Wang *et al.,* (2015). The marker screening and genetic diversity analyses were done at the Bio-Science Center Laboratory of the International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria.

Polymerase Chain Reaction (PCR)

PCR amplification was done in a total reaction's volume of 25 µL comprising 2.5 µL of 10 x buffer, 1µL (2 mM) of 50 mM $MgCl_2$, 2 µL of 2.5 mM dNTP mix (Bioline, USA), 200 nM each of forward and reverse primers, 20 ng μL^{-1} of genomic DNA and one unit of Taq DNA polymerase (Bioline). The amplification was carried out in a thermal cycler machine (Applied Biosystem, ABI Veriti 96 well Thermal Cycler, Singapore) in IITA, Bioscience Lab, Ibadan, Nigeria; under the following conditions: initial denaturation step at 95 °C for 4 minutes followed by 35 cycles of 95 °C for 20 seconds, annealing at respective annealing temperature (Ta) for each primer for 30 seconds, elongation at 72 °C for 30 seconds and final elongation at 72 °C for 5 minutes.

Allele scoring and data analysis

Amplified products were resolved by the gel electrophoresis using 2% SFR agarose gel and the size of the amplicon determined by the 50bp DNA ladder (Bioline), as a reference. Henceforth, the amplicon image was converted to an inverted image using the gel capture software for good visualization of bands then polymorphic bands in each SSR marker were scored as binary data (zero for absence and one for the presence of specific bands). The resulting binary matrix data of allelic information were compiled on Microsoft Excel software for further genetic analysis.

Estimation of genetic diversity parameters

The allelic information above was utilized to estimate the values of polymorphic information content (PIC), gene diversity, and a major allele frequency using the power marker software v. 3.25 (Liu and Muse, 2005), while other diversity parameters including Shannon information indices, expected heterozygosity,

and the number of effective alleles was computed using GenAlEx 6.5.1b2 software (Peakall and Smouse, 2012).

Population differentiation analysis

To partition, the grass pea accessions based on the genetic variation among and within the examined population, analysis of molecular variance (AMOVA), and Phi-statistic were computed using GenAlEx v.6.51b2 software (Peakall and Smouse, 2012). Genetic differentiation (ΦPT analogous of Fst) was rated as proposed by Wright, (1978), as low (<0.05), medium (0.05 to 0.15), moderate (0.15 to 0.25) and high >0.25. Genetic differentiation among the population was computed via the following AMOVA procedure applied by many authors (Michalakis and Excoffier, 1996; Peakall *et al.,* 1995; Huff and Bara, 1993; Excoffier *et al.,* 1992).

Principal coordinate analysis

The principal coordinate analysis investigation was computed using the genetic distance standardized method. A tri genetic distance matrix as input data was applied for computing two-dimensional principal coordinate analysis using GenAlEx v.6.51b2 software (Peakall and Smouse, 2012).

Genetic structure analysis

To figure out the genetic structure of the grass pea populations, STRUCTURE software V. 2.3.4 (Falush *et al.,* 2003; Pritchard *et al.,* 2000) was used. The investigation was computed by utilizing the ancestors model of 100,000 iterations and 300,000 Markov Chain Monte Carlo /MCMC/ burn-in, using the admixture model and allele frequency correlated model for 10 independent runs from k=1 to k=10. The Evano *et al.,* (2005) technique was applied to determine the number of k which have the remarkable delta K value. For this reason, a web-based program, STRUCTURE HARVESTER ver 0.6.94 was applied to separate the ideal number of a subgroup (Earl, 2012). Afterward, the web-based CLUMPAK program-Clustering Markov Packager (Kopelman *et al.,* 2015) was accustomed to recognizing clustering modes and packaging population structure inferences across K. This program was used for the summation and graphical plotting of population inferences and illustration of the results obtained by STRUCTURE. For this function, formatted Q-matrices data and population data labels were used to combine individual membership coefficients generated from 10 independent runs. To distinguish accessions that were admixed, each accession was allotted to its particular group on membership coefficient (Q). The limit for the membership coefficient was 90%, in which accessions membership more than that were allotted to particular subgroups, whereas, those that were lower than the limit were considered as admixed.

Cluster analysis

Cluster analysis was computed by NTSYSpc Software v 2.02e (Rohlf, 1997) using the Jaccard similarity matrix with the option of Sequential Agglomerative Hierarchic and Non-overlapping (SAHN) technique using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) method was applied for cluster dendrogram construction.

Spatial genetic structure

The spatial autocorrelation investigation was conducted using a genetic similarity matrix and a geographic distance matrix produced from coordinates of accessions. The investigation was done utilizing the “spatial” option applied in GenAlEx ver.6.51b2 software (Peakall and Smouse, 2012). The significance of the spatial autocorrelation and heterogeneity was determined by generating a two-tailed 95 % confidence interval close to the null hypothesis of no spatial structure, which was r = 0. The estimation was executed with a preference of an even distance class of 20 km, and permutations of 9999 and a bootstrap of 1000 were used to estimate the confidence interval around the null hypothesis.

Results

Genetic diversity estimated with SSR markers

Among the total of 10 SSR markers tested eight markers amplified scorable bands with a fragment size ranging from 150 to 750bp and PIC values from 0.074 to 0.941. These eight markers detected a total of 41 alleles in the 25 grass pea accessions ranging from two alleles in marker (G9) to 21 in (G17) with an average of 5.13 alleles per locus. Gene diversity values varied from 0.077 in marker (G19207) to 0.944 in (G17) with an average value of 0.536 and the heterozygosity values ranged from 0.121 to 0.395 with an average of 0.326 per locus. The major allelic frequency ranged from 0.12 to 0.96 with an average of 0.55 per locus (Table 2 and Figure 1).

Diversity among grass pea populations from different geographic zones of North-West Ethiopia

To measure diversity among the population from the regions, data of accessions collected from different geographic zones were pooled together (Figure 2). The number of different alleles and effective alleles per locus ranged from 1.53 to 1.90 and 1.48 to 1.70 with an overall mean of 1.78 and 1.58, respectively. The smallest Shannon’s diversity index was observed in East Gojjam (0.370) on the contrary, the highest value was observed for the West Gojjam (0.553) with an overall mean of 0.474. East and West Gojjam again followed the same pattern with the lowest (0.259) and highest (0.384) expected heterozygosity, respectively with a mean value of 0.326. The largest percentage of polymorphism was observed in accessions collected from North Gondar and West Gojjam zones (89.47 %). While the lowest percentage polymorphism was observed for accessions collected from East Gojjam (57.89 %).

Genetic variation analysis

The genetic variation among populations and subpopulations was determined using analysis of molecular variance (AMOVA). In this analysis, Gondar and Gojjam were considered as regions, whereas, zones were assigned as populations within regions. The result of AMOVA showed a highly significant difference ($P< 0.001$) among regions. The percentage of genetic variation among regions was 20% of the total genetic variance. Whereas, contributions of zones within the region to the total variance were not significant (4 %), indicating that zones within the region play a slight role in the total variability in the

studied grass pea. Moreover, the largest genetic variance (76 %) was explained among individuals within populations, demonstrating that most of the genetic variation was introduced due to the accessions rather than the geographic locations, as determined through the standardized permutation analyses (Table 3).

Phi-statistics analysis

Phi-statistics (Φpt analogous to Fst) among regions (0.20) and within populations among accessions (0.24) showed a moderate level of genetic differentiation. Whereas, Φpt values among regions within populations i.e. between adjacent zones (South and North Gondar) and between (East and West Gojam) showed a low level of genetic differentiation (0.053) as indicated in table 4.

Genetic differentiation between populations based on pairwise genetic distance

The pairwise genetic differentiation (PhiPT) values among four zones were varied from 0.033 to 0.365. There was highly significant genetic differentiation ($P< 0.01$) between zones (East Gojjam and North Gondar), (East Gojjam and South Gondar), and (North Gondar and West Gojjam) population. In this study, the mean genetic differentiation (Φpt) values among Gondar and Gojjam regions was 0.24 which was categorized as moderate level by Wright, (1978). The average genetic differentiation within South Gondar, North Gondar, West Gojjam, and East Gojjam was 0.175 and rated as moderate. The largest pairwise genetic differentiation values were observed between North Gondar and East Gojjam populations (0.365). While a low genetic differentiation value was observed between South Gondar and North Gondar populations (0.053) (Table 5).

Classification and principal coordinate analysis among all studied accessions

The genetic relationship among individual accessions according to geographical origin was inspected based on principal coordinate analysis using the genetic distance matrix as the input file. The two-dimensional biplot grouped all the accessions from East Gojjam together with only three accessions from West Gojjam. to some extent grouped according to their geographic origin and slightly differ from other accessions. On the contrary, all accessions from North Gondar, 83.33% of accessions from South Gondar, and half of the accessions from West Gojjam were grouped in the same eclipse despite the wide distance between them. The contribution of the first and the second principal coordinates to the total genetic variation was 45.97% and 12.12, respectively (Figure 3). The overall genetic variation explained by the three-dimensions was 66.31%.

Population structure analysis.

The population structure analysis indicated that $K = 2$ is the ideal number of subgroups (Figure 4). The membership coefficient ($\geq$ 90 %) was used to assign individuals to their respective subgroups. Based on this value, 52 % of the accessions were assigned to the first subgroup. While 24% of the accessions were assigned to subgroup two. The remaining accessions (24 %) were considered admixed given the ordinarily shared ancestors with individuals allocated to the detected subgroups. The arrangement of

every individual was represented by two different colors. The blue color indicates subgroup one, and the orange color indicates subgroup two (Figure 5 and 6). Figure 6 clarifies the genetic assignment of each grass pea accession represented by a single line into two subgroups (blue and orange color) as inferred from structure analysis at k=2 using SSR markers data. This result has shown that both subgroups were signified by each individual, which has distinctive membership coefficients, collected from all zones showing the absence of distinct geographic origin-based genetic structuring.

Cluster analysis

Cluster analysis classified the accessions into two distinct groups using the Jaccard genetic similarity matrix. Cluster I and II consisted of 18 accessions and seven accessions, respectively (Figure 7). The threshold value of the similarity coefficient determining the number of clusters was 70 % similarity. The clustering pattern revealed that accessions from various geographic regions were grouped in the same cluster and vice-versa. Accessions from North Gondar and South Gondar showed more genetic diversity within zones. Whereas accessions from East and West Gojjam showed low genetic diversity within zones.

Spatial autocorrelation analysis

The spatial autocorrelation result exhibited a significant correlation between a geographic distance and genetic similarity matrices for grass pea accessions collected in the range of 30 km. The grass pea accessions which were collected at a distance within 30 km were shown to have spatial autocorrelation. It was observed that as the geographic distance starts to increases beyond 30 km the genetic similarity begins to diminish, and the heterogeneity of the accessions starts to increases. Even if it was not significant the correlation between the geographic distance and genetic similarity matrix was positive until it reached the intercept point around 102.11 Km (Figure 8). When the sampling distance begins to increase, the Mantel correlation coefficient began to decrease resulting in a non-significant correlation. This result exhibited that when the geographic distance increase, genetic similarity started to decline. The maximum geographic distance covered between two grass pea accessions was 320 km.

Discussion

Grass pea is the most widely cultivated legume by the resource-poor farmers living in marginal lands in Ethiopia despite the presence of neurotoxin β-ODAP content in its seeds and foliage that is known to cause leg paralysis in animals including man. The cultivation of the crop is favored in the country due to poor performance and failure of other crops in times of drought since grass pea is known to withstand and yield well under harsh environmental conditions because of its stress-tolerant nature. The rich protein content of the crop (28%-32%) and minerals are other reasons the crop is widely cultivated in Ethiopia where it is used as alternative protein sources to reduce malnutrition and food insecurity of the overgrowing population of the country (Soren *et al.*, 2015; Urga *et al.*, 2005).

Nowadays when the impact of climate change is gradually overwhelming crop production making it increasingly difficult to feed the fastest growing population, the importance of grass pea that has the

natural ability to maintain high yield under harsh environmental conditions is increasing. However, there is still a need to improve the crop for better yield and increased tolerance to biotic and abiotic stresses in the face of increasing climate change. To achieve this, understanding the genetic variations and relationships within and among grass pea populations in Ethiopia is a pre-requisite. In this study, previously developed SSR markers of *L. sativus* and its relatives were used to assess the diversity of the North-Western part of Ethiopian grass pea populations (Yang et al., 2014, Wang et al., 2015).

Henceforth, the present investigation exhibited the existence of a moderate level of genetic diversity among the North-Western Ethiopian grass pea population revealed by AMOVA and genetic differentiation analysis. Thus, the highest genetic differentiation was explained by within population. This indicates an equivalent level of genetic variation harbored among zones. the present study was relatively in a close population that would be important in germplasm management and breeding program. Piloting a close study on a particular population could be desirable when conducting genetic improvement in grass pea. Taddese and Bekele, (2003a) reported based on phenotypic characterization of Ethiopian grass pea showed the genetic variation was distributed across regions and Shiferaw *et al.*, (2012) who worked on genetic diversity of Ethiopian grass pea using EST-SSR markers showed an equivalent level of genetic variation harbored among their studied regions but there study area covered distantly regions than the present study. The present finding and the above two reports showed geneflow among regions and zones. This due to germplasm exchange among farmers in Ethiopian regions. However, gene flow from where to where was unknown. To identify geneflow origin needs further studies. The other reason for the occurrence of gene flow was regardless of the floral biology of grass pea as a predominantly self-pollinating crop, in contrary to this, there were around 27 % outcrossing by bees (Gutierrez-Marcos *et al.*, 2006).

The average allele number among the studied accessions was 5.13 overall loci. This result was lower than the result obtained by Wang *et al.*, (2015) where 30 polymorphic SSR markers were applied for the genetic diversity study of *L.sativus* and relative species detecting an average allele number of 8.6 overall loci. This difference might be due to the small number of markers and grass pea accessions used in this study.

The polymorphism testing of studied accessions through eight SSR loci has been evaluated by a major allele frequency was in the range of 0.12 to 0.96 with a mean value of 0.55. A related result has been pointed out by (Shiferaw et al., 2012) who studied exploring the genetic diversity of the Ethiopian grass pea using 24 EST-SSR markers among which 11 EST-SSR markers detected a major allele frequency ranging from 0.29 to 0.88 with a mean of 0.62. Moreover, close results have been noted by (Soren et al., 2015) who have studied EST-SSR analysis providing insights about genetic relatedness, population structure, and the gene flow in the grass pea. This has indicated a major allele frequency of 19 EST-SSR loci for 176 Indian grass pea accessions which were from 0.39 to 0.97 with an average of 0.63.

In the present study, gene diversity, and PIC values were ranged from 0.074 to 0.944 with a mean of 0.536 and 0.474 respectively. A similar result has been described by (Wang et al., 2015) who have investigated

on genetic diversity of the grass pea and its relative species evaluated by SSR markers among 30 polymorphic SSR primers for 266 accessions comprise 17 relative species, and the other was originated from Europe, Asia, and African *L.sativus* the gene diversity and PIC were indicated in the range of 0.0688 to 0.8505 with a mean of 0.534 and 0.4817 respectively. Furthermore, our study agreed with the work of Soren *et al.*, (2015); (Shiferaw et al., 2012, Soren et al., 2015) in 30 EST-SSR analysis gene diversity was ranged from 0.04 to 0.73 with a mean of 0.45 and 11 EST-SSR markers gene diversity was varied from 0.205 to 0.805 with a mean of 0.477, respectively.

Expected heterozygosity performed among four populations has ranged from 0.259 to 0.384 with a mean of 0.326. The mean was a fall in the result obtained by (Soren et al., 2015) using 19 polymorphic EST-SSR primers that ranged from 0.23 to 0.50 and a mean of 0.42. However, the discrete result was obtained by (Shiferaw et al., 2012) who have investigated 11 polymorphic EST-SSR primers for seven regions of Ethiopian 20 grass pea accessions of expected heterozygosity was varied from 0.354 to 0.478 with a mean of 0.430. This finding differs from the present investigation the present work has used close geographical regions grass pea accessions and may be as results of different genotype, primer type and number used.

Shannon's information index detected in the present study ranged from 0.370 to 0.553 with a mean value of 0.474. The mean of Shannon's information index was found in the range of the work of (Soren et al., 2015) who have reported on 19 polymorphic EST-SSR primers for 176 Indian grass pea accessions of Shannon's information index was ranged from 0.42 to 0.69 with a mean of 0.61. A contrasting result has been found in the result of (Shiferaw et al., 2012) using 11 EST-SSR primers in seven regions of 20 accessions of Ethiopian grass pea genotype Shannon's information index per population was from 0.595 to 0.855 with a mean of 0.760. This indicates there was a difference in the studied genotype, wider geographical population, and the type of primer used than the present work indeed.

The comparable result has been computed by (Shiferaw et al., 2012) who have evaluated the genetic differentiation based on 11 EST-SSR markers data of 20 grass pea accessions from seven regions of Ethiopia: Shewa, Wollo, Gojjam, Gondar, Tigray, Arsi, and Hararge. The AMOVA showed high significance among regions, among populations, and within populations. Likewise, the related result has been reported by (Wang et al., 2015) who have estimated AMOVA using 30 SSR markers for 266 grass pea accessions, and 17 relative species from Europe, Asia, and Africa based on geographical origin. The genetic differentiation between species, among populations, and geographical origin showed high significance.

In this study, genetic differentiation (Φpt analogous to Fst) values within individual accessions (0.24) and among regions (0.20) based on Wright, (1978) were rated as a moderate level of genetic differentiation. A related result has been explained by (Shiferaw et al., 2012) genetic differentiation (Φpt) values of 0.15 and was categorized as a moderate level. In contrast, the genetic differentiation (Φpt) values among regions within zones or genetic differentiation within adjacent zones (0.05) as Wright, (1978)

categorization was low-level genetic differentiation. Therefore, when the geographic distance between zones decreased, the genetic differentiation becomes lower.

According to (Wright, 1978) the genetic differentiation level of pairwise population, genetic distance (PhiPT) between North Gondar and East Gojjam population was 0.365. This result showed a very large level of gene differentiation. Likewise, pairwise genetic differentiation (PhiPT) values of 0.23 and 0.21 between populations of (East Gojjam and South Gondar) and (North Gondar and West Gojjam) respectively, indicates there was a moderate level of gene differentiation between those populations. While between populations of South Gondar and West Gojjam pairwise genetic distance (PhiPT) value of 0.14 indicates there was a medium level of gene differentiation between those populations. Thus, the AMOVA analysis also revealed that large genetic differentiation resides among studied accessions, this indicates the existences of the gene flow across geographical regions. Henceforth, genetic differentiation between some accessions among zones was low, in contrast, some accessions showed high genetic differentiation. Correspondingly, when the geographic distant increased the level of genetic differentiation was also increased.

In the present study, principal coordinate analysis through a genetic distance matrix was discriminating 25 grass pea accessions based on their geographical origin. Genetic variation explained by the first principal coordinate was 45.95 %, and the second principal coordinate was 12.12 %. A related result has been noted by (Wang et al., 2015) who used 30 polymorphic SSR primers for 266 accessions of Europe, Asia, and Africa were discriminated in two-dimensional principal coordinates. They reported genetic variation explained by the first principal coordinate was 43.42 % and the second principal coordinate explained was 29.17% out of the total variation.

In this study, population genetic structure analysis showed two Subgroups. This result revealed that both subgroups were denoted by each accession, which has distinguishing membership coefficients, collected from all zones indicates the absence of zone-based genetic structuring. The clustering pattern was revealed that accessions from various geographic regions were grouped in the same cluster and vice-versa. Thus, population genetic structure analysis and clustering pattern in the present study indicated the presence of gene flow through seed exchange across all geographical locations. The possible reason for accessions from different geographical origins for being grouped into the same cluster and subgroups could be gene flow between adjacent zones or exchange of seed among farmers between zones or both. This indicated that different zones harbor an equivalent level of variations. Understanding the distribution of genetic variation among geographic regions would be important for the grass pea improvement program and planning germplasm collection and conservation in this crop interest. The germplasm exchange among farmers is a basis for increasing the diversity of the local population which might affect the spreading of alleles among various populations regardless of their geographical distance (Louette *et al.*, 1997). Grass pea's reproductive biology could also contribute to allelic distribution among populations. However, its floral biology is predominantly self-pollination (Shiferaw *et al.*, 2012; Yadav and Bejiga, 2006; Campbell, 1997). To some degree, several reports manifest extensive

out-crossing up to 27 % which is upon bees, environment, and genetic factors (Soren *et al.*, 2015; Shiferaw *et al.*, 2012; Gutierrez-Marcos *et al.*, 2006; Chowdhury and Slinkard 1997;).

The related results reported by (Shiferaw et al., 2012) who have studied the genetic diversity through 11 polymorphic EST-SSR markers for 20 grass pea accessions originated in Ethiopia from seven regions of the study areas. It was grouped into three clusters so that, Cluster I consists of accessions from Tigray, Gojjam Gondar, and Wollo. Cluster II contained accessions from all regions while cluster III was mainly composed of Shewa and Gojjam and some accessions categorized in this cluster from Gondar and Wollo. This showed that a gene flow took place across all regions.

(Tadesse and Bekele, 2003b, Tadesse and Bekele, 2003a) who have reported clustering 50 grass pea accessions in population-based phenotypic and morphological traits showed some of Gondar and Gojjam accessions grouped in the same cluster, Tigray and Wolega region which are distantly but grouped in the same cluster. This indicates there was a gene flow in the Ethiopian grass pea populations.

Furthermore, the associated result has been pointed out by (Wang et al., 2015) using 30 polymorphic SSR markers for 266 grass pea accessions, and 17 relative species originated from Europe, Asia, and Africa. The cluster analysis had been discriminating grass pea accessions, and relative species accessions, the African and European accessions were grouped in cluster II and relative species were grouped in cluster III. This indicated there was a high gene flow between Africa and Europe grass pea accessions. Accordingly, the cultivation of the grass pea was obscured as a result, the center of the origin was not known.

The spatial autocorrelation analysis figured to distinguish how far the grass pea accessions are separated upon the geographic distance showed a substantial correlation for accessions collected in a range of 30 km. This outcome designated that as the range of the collection between two accessions arises distant from each other, the genetic resemblance among the accessions started to drops in an account of the geographic distance.

Conclusion

In conclusion, this finding was confirmed that the microsatellite/SSR markers previously developed by (Yang et al., 2014) and used by (Wang et al., 2015) for *L.sativus* and relative species were a reproducible and vital tool for the genetic diversity study. The result showed the existence of a moderate level of genetic diversity among the North-Western part of Ethiopian grass pea populations which largely reside within each accession. This indicated that different zones harbor an equivalent level of variations. Thus, the present study was relatively in a close population that would be important in the breeding program. Piloting a close study on a particular population could be desirable when conducting genetic improvement in grass pea. Even if, the present investigation was based on a limited number of SSR primers, this investigation should be considered in planning future conservation, *in situ* maintenance, and breeding programs for grass pea. The recent grass pea collection in Ethiopian Biodiversity Institute comprises the majority of the accessions was from the Shewa region which was about 45 % (Shiferaw et

al., 2012). Thus, in the present studied regions of Gojjam and Gondar, a smaller number of accessions were maintained in Ethiopia Biodiversity Institute/EBI which was 19 % and 8 % respectively. Henceforth, it could be vital for increasing representative samples mainly in the Gondar region and other regions in general to get the large diversity.

Declarations

Acknowledgments

Special thanks to Ebonyi State University (EBSU), Faculty of Science and Department of Biotechnology, Abakaliki, Nigeria, and GENEs Intra-Africa Mobility Project EACEA/2017/2552 with Ref number EBSU/2018/GENES/MSc/ SCH-M/03 for hosting and sponsoring this research work. I appreciate Ethiopian Biodiversity Institute (EBI) for providing grass pea accessions.

My superior acknowledgments goes to Dr. Melaku Gedil allowing IITA Bioscience Lab, for his technical assistance, and providing consumable chemicals used for the molecular part of the study throughout the Lab work.

References

Akter, L., Mahbub, M., Habib, M. A., and Alam, S. S. (2015). Characterization of three varieties of Lathyrus sativus L. by fluorescent karyotype and RAPD analysis. Cytologia **80**, 457-465.

Barik, D. P., Acharya, L., Mukherjee, A. K., and Chand, P. K. (2007). Analysis of genetic diversity among selected grass pea (Lathyrus sativus L.) genotypes using RAPD markers. Zeitschrift für Naturforschung C **62**, 869-874.

Belaid, Y., Chtourou-Ghorbel, N., Marrakchi, M., and Trifi-Farah, N. (2006). Genetic diversity within and between populations of the Lathyrus genus (Fabaceae) revealed by ISSR markers. Genetic Resources and Crop Evolution **53**, 1413-1418.

Campbell, C. G. (1997). *Grass pea, Lathyrus sativus L* (Vol. 18). Bioversity International.

Chowdhury, M. A., & Sllnkard, A. E. (1997). Natural outcrossing in grass pea. *Journal of Heredity*, **88**, 154-156.

Chtourou-Ghorbel, N., Lauga, B., Combes, D., and Marrakchi, M. (2001). Comparative genetic diversity studies in the genus Lathyrus using RFLP and RAPD markers. Lathyrus Lathyrism Newsletter **2**, 62-68.

Croft, A., Pang, E., and Taylor, P. (1999). Molecular analysis of Lathyrus sativus L.(grass pea) and related Lathyrus species. Euphytica **107**, 167-176.

CSA. 2020. The Federal Democratic Republic of Ethiopia, Central Statistical Agency, Agricultural

Sample Survey 2019/2020 (2012 E.C.). Volume I. Report on Area and Production of Major Crops (Private Peasant Holdings, Meher Season). Statistical Bulletin, April 2020, Addis Ababa, Ethiopia.

Dixit, G. P., Parihar, A. K., Bohra, A., and Singh, N. P. (2016). Achievements and prospects of grass pea (Lathyrus sativus L.) improvement for sustainable food production. The Crop Journal **4**, 407-416.

Earl, D. A. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation genetics resources*, **4**, 359-361.

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

Excoffier, L., Smouse, P. E., and Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics **131**, 479-491.

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics, **164,** 1567-1587.

Girma, D., and Korbu, L. (2012). Genetic improvement of grass pea (Lathyrus sativus) in Ethiopia: an unfulfilled promise. Plant breeding **131**, 231-236.

Gupta, P., Udupa, S. M., Gupta, D. S., Kumar, J., and Kumar, S. (2018). Population structure analysis and determination of neurotoxin content in a set of grass pea (Lathyrus sativus L.) accessions of Bangladesh origin. The Crop Journal **6**, 435-442.

Gutiérrez-Marcos, J. F., Vaquero, F., de Miera, L. E. S., & Vences, F. J. (2006). High genetic diversity in a worldwide collection of Lathyrus sativus L. revealed by enzymatic analysis. *Plant Genetic Resources*, **4,** 159-171.

Hillocks, R., and Maruthi, M. (2012). Grass pea (Lathyrus sativus): Is there a case for further crop improvement? Euphytica **186**, 647-654.

Huff, D. R., and Bara, J. M. (1993). Determining genetic origins of aberrant progeny from facultative apomictic Kentucky bluegrass using a combination of flow cytometry and silver-stained RAPD markers. Theoretical and Applied Genetics **87**, 201-208.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak:a program for identifying clustering modes and packaging population structure inferences across K. Molecular ecology resources, **15**, 1179-1191.

Lioi, L., Sparvoli, F., Sonnante, G., Laghetti, G., Lupo, F., and Zaccardelli, M. (2011). Characterization of Italian grass pea (Lathyrus sativus L.) germplasm using agronomic traits, biochemical and molecular markers. Genetic resources and crop evolution **58**, 425-437.

Liu, K., and Muse, S. V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics **21**, 2128-2129.

Louette, D., Charrier, A., and Berthaud, J. (1997). In situ conservation of maize in Mexico: genetic diversity and maize seed management in a traditional community. Economic Botany **51**, 20-38.

Michalakis, Y., and Excoffier, L. (1996). A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. Genetics **142**, 1061-1064.

Parihar, A., Dixit, G., and Singh, D. (2013). Multivariate analysis of various agronomic traits in grass pea (Lathyrus spp) germplasm. Indian Journal of Agricultural Sciences **83**, 570-575.

Parihar, A., Dixit, G., and Singh, D. (2015). Genetic variability analysis for quantitative traits in a germplasm set of grass pea (Lathyrus spp.). Legume Research-An International Journal **38**, 461-464.

Peakall, R., and Smouse, P. (2012). GenAlEx Tutorials-Part 2: genetic distance and analysis of molecular variance (AMOVA). Bioinformatics **28**, 2537-2539.

Peakall, R., Smouse, P. E., and Huff, D. R. (1995). Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss Buchloe dactyloides. Molecular Ecology **4**, 135-148.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics, **155**, 945-959.

Rohlf, F. J. (1997). NTSYS-pc. Numerical taxonomy and multi variance analysis system version2.02 e. Exeter Software, New York, USA.

Shiferaw, E., Pè, M., Porceddu, E., and Ponnaiah, M. (2012). Exploring the genetic diversity of Ethiopian grass pea (Lathyrus sativus L.) using EST-SSR markers. Molecular Breeding **30**, 789-797.

Soren, K. R., Konda, A. K., Gangwar, P., Tiwari, V. A., P. S. Shanmugavadivel, P. S., Parihar, A. K., Dixit, G. P and Singh, N. P. 2020. Development of SSR markers and association studies of markers with phenology and yield-related traits in grass pea (Lathyrus sativus). Crop & Pasture Science **71**, 768–775

Soren, K. R., Yadav, A., Pandey, G., Gangwar, P., Parihar, A. K., Bohra, A., Dixit, G. P., Datta, S., and Singh, N. P. (2015). EST-SSR analysis provides insights into genetic relatedness, population structure, and gene flow in grass pea (Lathyrus sativus). Plant Breeding **134**, 338-344.

Tadesse, W., and Bekele, E. (2003a). Phenotypic diversity of Ethiopian grass pea (Lathyrus sativus L.) with geographical regions and altitudinal range. Genetic Resources and Crop Evolution **50**, 497-505.

Tadesse, W., and Bekele, E. (2003b). Variation and association of morphological and biochemical characters in grass pea (Lathyrus sativus L.). Euphytica **130**, 315-324.

Talukdar, D. (2009). Recent progress on genetic analysis of novel mutants and aneuploid research in grass pea (Lathyrus sativus L.). African Journal of Agricultural Research **4**, 1549-1559.

Tavoletti, S., and Iommarini, L. (2007). Molecular marker analysis of genetic variation characterizing a grass pea (Lathyrus sativus) collection from central Italy. Plant breeding **126**, 607-611.

Urga, K., Fufa, H., Biratu, E., and Husain, A. (2005). Evaluation of Lathyrus sativus cultivated in Ethiopia for proximate composition, minerals, β-ODAP, and anti-nutritional components. African Journal of Food, Agriculture, Nutrition, and Development **5,** 1-16.

Wang, F., Yang, T., Burlyaeva, M., Li, L., Jiang, J., Fang, L., Redden, R., and Zong, X. (2015). Genetic diversity of grass pea and its relative species revealed by SSR markers. PloS one **10**, e0118542.

Wright, S. (1978). "Evolution and the genetics of populations: a treatise in four volumes: Vol. 4: variability within and among natural populations," University of Chicago Press.

Yadav, S. S., Bejiga, G., Brink, M., & Belay, G. (2006). Lathyrus sativus L. record from protabase. *PROTA (Plant Resources of Tropical Africa).* Wageningen, Netherlands *(http://database. prota. org/search. htm)*.

Yang, T., Jiang, J., Burlyaeva, M., Hu, J., Coyne, C. J., Kumar, S., Redden, R., Sun, X., Wang, F., and Chang, J. (2014). Large-scale microsatellite development in grass pea (*Lathyrus sativus* L.), an orphan legume of the arid areas. *BMC plant biology* **14**, 1-12.

Tables

Table 1. Passport data of grass pea accessions used for the study.

No	Accessions	Zone	District	Latitude	Longitude	Altitude
1	238935	South_Gondar	Fogera	11°-55'-24"-N	37°-54'-21"-E	2130.00
2	236702	South_Gondar	Dera	-	-	1800.00
3	212741	South_Gondar	Kemekem	12°-06'-00"-N	37°-42'-00"-E	2000.00
4	226014	South_Gondar	Este	11°-27'-00"-N	37°-59"-00'-E	2645.00
5	238931	South_Gondar	Fogera	11°-57'-45"-N	37°-43'-35"-E	1920.00
6	226011	South_Gondar	Este	11°-03'-00"-N	38°-09'-00"-E	2520.00
7	226017	North_Gondar	Dembia	12°-30'-00"-N	37°-24'-00"-E	1905.00
8	226018	North_Gondar	Gondar zuria	12°-31'-00"-N	37°-20'-00"-E	1990.00
9	236708	North_Gondar	Dabat	-	-	2730.00
10	238928	North_Gondar	Gondar zuria	12°-24'-01"-N	37°-33'-09"-E	1990.00
11	242216	North_Gondar	Gondar zuria	12°-30'-00"-N	37°-32'-00"-E	1975.00
12	46107	North_Gondar	Gondar zuria	12°-37'-00"-N	37°-10'-00"-E	1950.00
13	236697	West_Gojjam	Adet	-	-	2260.00
14	236712	West_Gojjam	Jabi Tehnan	-	-	1820.00
15	238947	West_Gojjam	Dega Damot	10°-38'-55"-N	37°-23'-58"-E	1900.00
16	238920	West_Gojjam	Merawi	11°-25'-09"-N	37°-09'-54"-E	2050.00
17	238942	West_Gojjam	Achefer	11°-44'-33"-N	36°-59'-06"-E	2050.00
18	238945	West_Gojjam	Achefer	11°-48'-02"-N	36°-59'-59"-E	2030.00
19	24812	East_Gojjam	Hulet Ej Enese	11°-05'-45"-N	37°-52'-28"-E	2452.00
20	30357	East_Gojjam	Enarj Enawaga	10°-40'-57"-N	38°-11'-13"-E	2517.00
21	238908	East_Gojjam	Dejen	09°-57'-28"-N	38°-18'-26"-E	2450.00
22	26626	East_Gojjam	Debay Telatgen	10°-25'-26"-N	38°-07'-33"-E	2573.00
23	26627	East_Gojjam	Baso Liben	10°-04'-44"-N	37°-44'-54"-E	2350.00
24	238910	East_Gojjam	Enemay	10°-21'-29"-N	38°-09'-36"-E	2440.00
25	26633	East_Gojjam	Awabel	10°-14'-14"-N	38°-03'-24"-E	2439.00

Table 2. Details of the SSR markers applied for 25 grass pea accessions with the number of alleles, major allele frequency, annealing temperature (Ta), and product size.

Table 3. Summary of AMOVA partitioning of populations based on regions.

Note: (df= degree of freedom, SS= sum of squares, MS = mean squares, Est. var. = estimate of variance, %= percentage of total variation based on 999 permutations.

Table 4. Phi statistics values of grass pea accessions from different regions.

Primer	Repeat Motif	Forward primer sequence (5'-3')	Reverse primer sequence (3'-5')	Ta/0 C	Product size (bp)	No of Allele	Major Allele Frequency
G9	(AAC)6	CAACCAGAGCAACCACAAGA	GGTTGCAAGAGGTTGCAGAT	55	170-300	2	0.68
G17	(AAT)5	CAGGTCCGGCTTATCTCTCA	TTGGTTTCAACCCACTCCTC	56	150-750	21	0.12
G68	(AC)9	GCACACAAGGGCACACTG	TGCGTCGTGTGTATGTGTTG	52	250	2	0.52
G157	(CAA)6	ACATCCAATCCCCACCATAA	AATGCATGGTTGTTGCTTGA	52	170-200	2	0.72
G245	(TG)6	CGTTGGTTGTTAGTCGGTCA	GAACGAAACAACGACGACAA	52	260-400	5	0.32
G15771	(TCG)5	AGTGCCTGATGGGAGTCAGT	CCGACGACGACGACTACTAA	56	200-300	2	0.64
G17922	(CCA)5	CACCACCATAACCACCTCCT	ATGCGATTGAAGGGATGAAC	55	140-370	5	0.44
G19207	(AAG)5	ATCGTAAACCGTGAGGGTCA	AAGCTTGTGGTGGCTACTGC	56	200	2	0.96

Source Variation	df	SS	MS	Est. Var.	%
Among Regions	1	16.38	16.38	0.93	20
Among Population within Region	2	9.48	4.74	0.20	4
Within Population	21	73.74	3.51	3.51	76
Total	24	99.60		4.64	100

Phi Statistic	value	P rand ≥ data
PhiRT	0.20	0.001
PhiPR	0.05	0.099
PhiPT	0.24	0.001

Table 5: Pairwise PhiPT values (below diagonal) and numerical values based on 999 permutations (above diagonal) between geographic origins of grass pea populations.

	South Gondar	North Gondar	West Gojjam	East Gojjam
South Gondar	0.000	0.291	0.043	0.003
North Gondar	0.033	0.000	0.004	0.001
West Gojam	0.137*	0.210**	0.000	0.072
East Gojam	0.231**	0.365**	0.072	0.000

Note: ** indicates high significance at (P<0.01) and * indicates significance at (P<0.05) P-value based on 999 permutations.

Figures

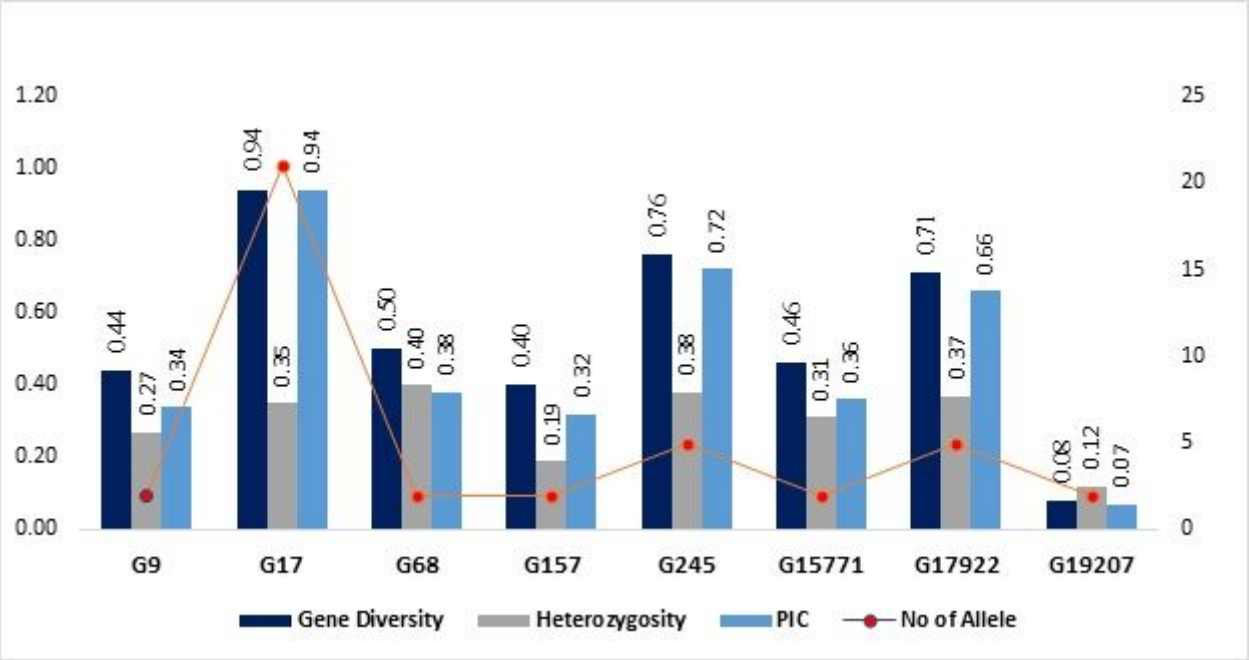

Figure 1

This figure illustrates gene diversity, heterozygosity, PIC, and a number of the allele of each SSR marker among 25 grass pea accessions.

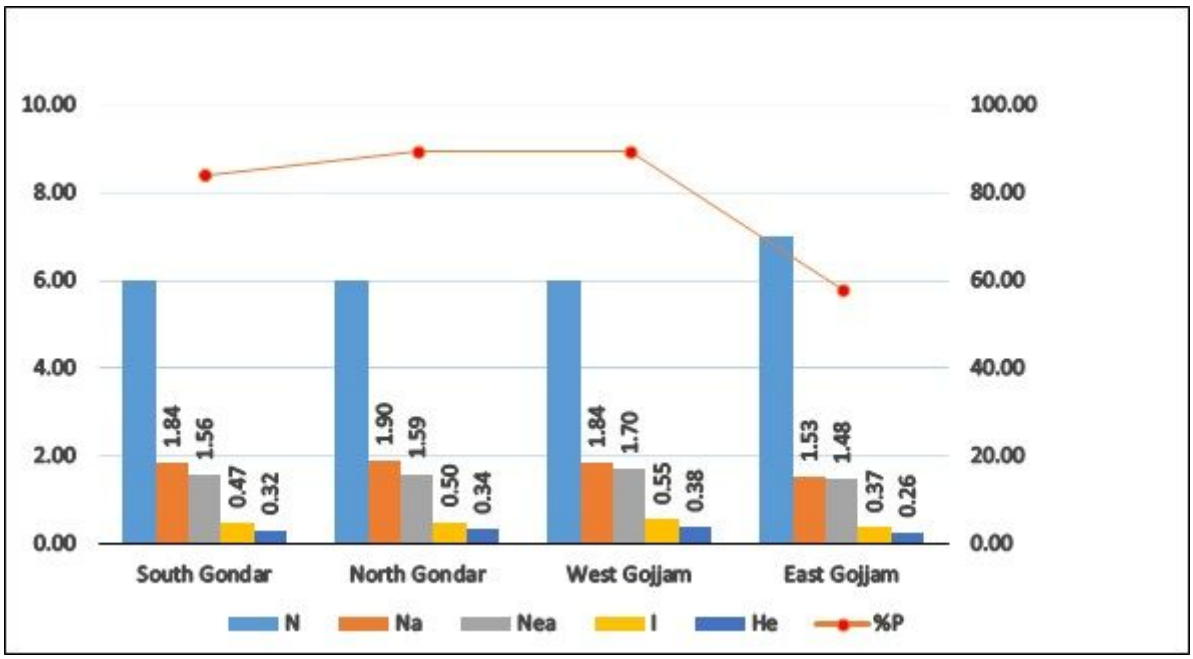

Figure 2

Diversity of grass pea populations obtained from the analysis of eight SSR loci based on geographic regions. Where: N=population size, Na = number of different alleles, Nea = number of effective alleles, I = Shannon's information index, He = expected heterozygosity and % P = percentage of polymorphic loci in each population.

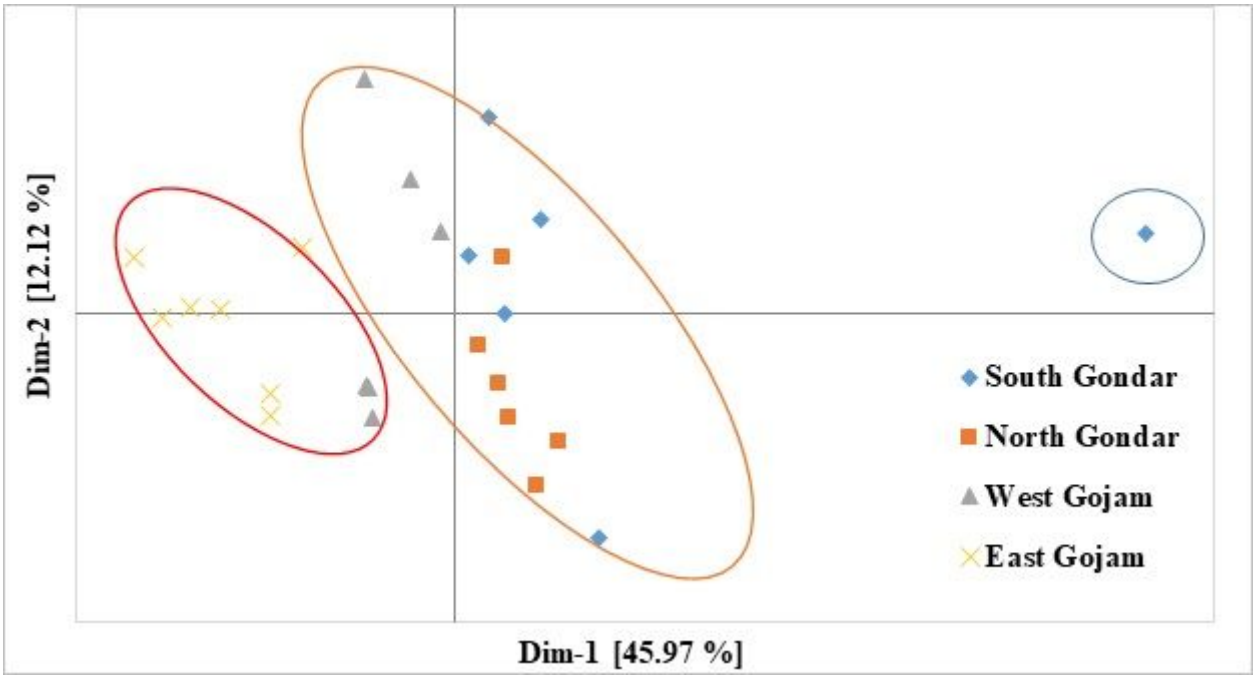

Figure 3

Two-dimensional principal coordinate analysis [PCoA] among 25 grass pea accessions based on geographical location origin.

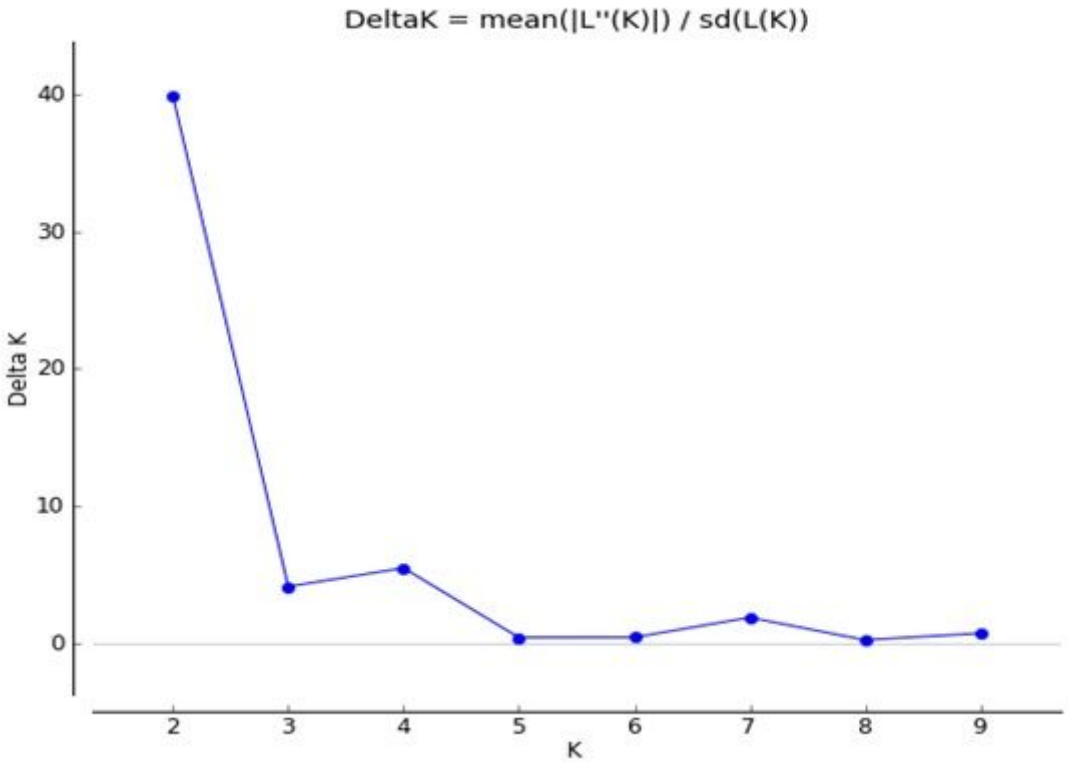

Figure 4

Population genetic structure estimation was designed for determining the number of subgroups for k ranging from 1 to 10 using SSR markers data.

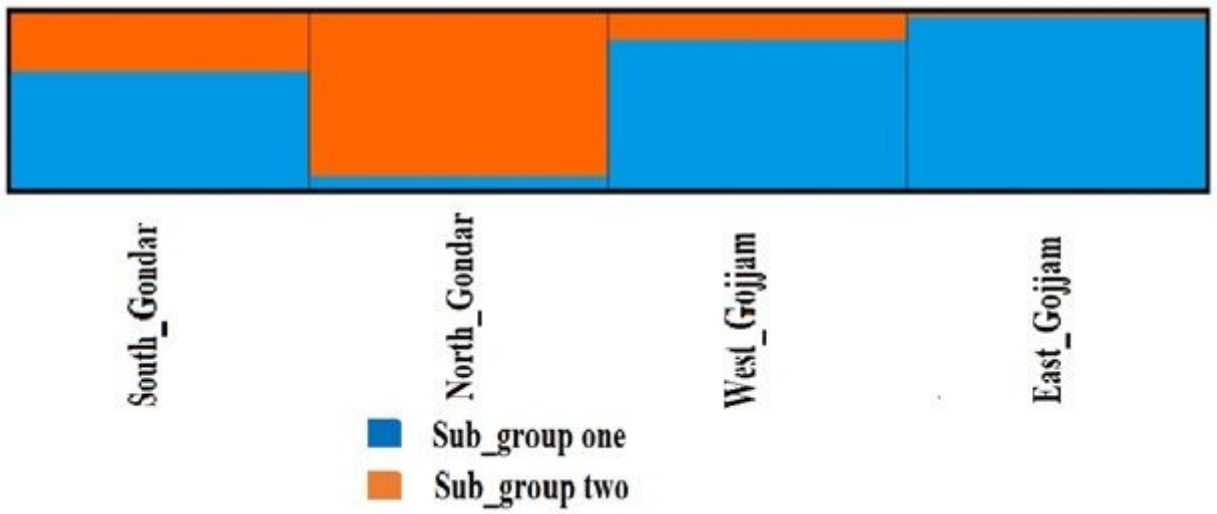

Figure 5

Genetic assignment of grass pea populations inferred from structure analysis at k =2 based on SSR markers data. The analysis assigned the populations into two subgroups (1st = blue color and 2 nd = orange color).

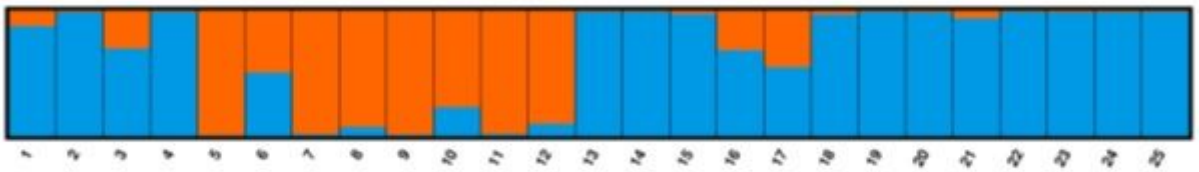

Figure 6

Genetic assignment of each grass pea accession represented by a single line into two subgroups (blue and orange color) as inferred from structure analysis at k=2 using SSR markers data.

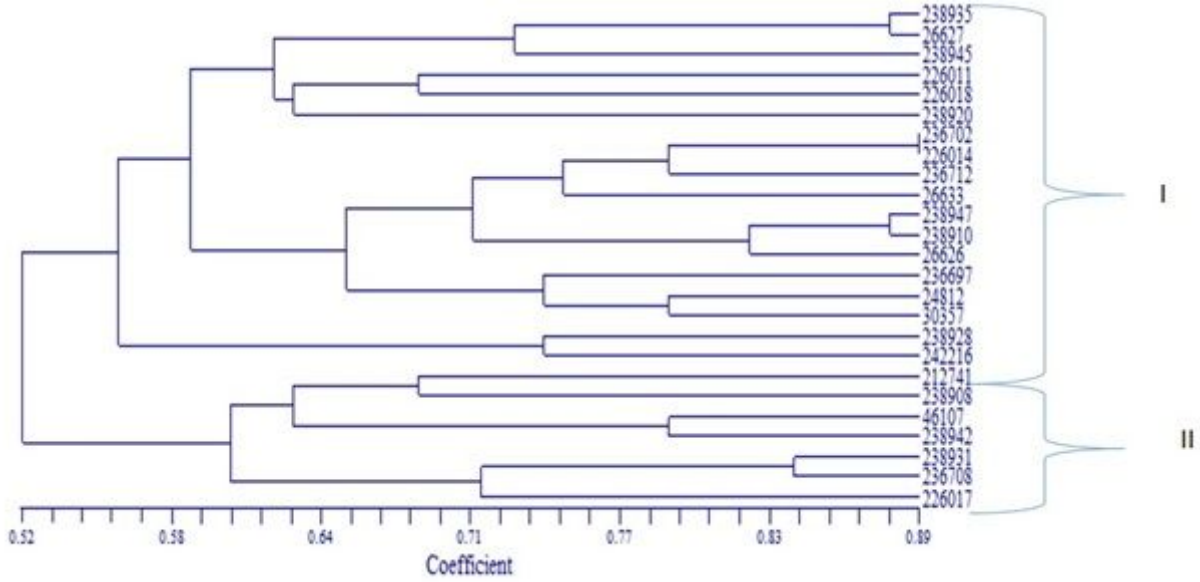

Figure 7

Dendrogram constructed by UPGMA method based on the Jaccard genetic similarity matrix data among 25 grass pea accessions.

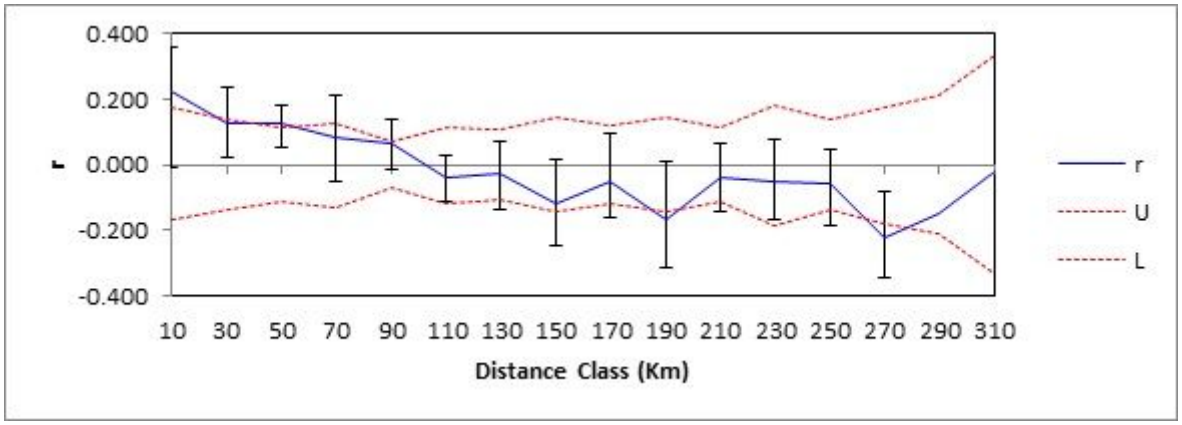

Figure 8

Spatial correlation between geographic distance and genetic similarity matrices for North-West Ethiopian grass pea accessions. The broken lines indicate the 95% confidence interval for the Null hypothesis of random distribution, the solid line in the center indicates the correlation coefficient (Mantel r), and the whiskers indicating the magnitude of error after bootstrapping 1000.