



# Molecular analysis for genetic diversity and population structure of Ethiopian faba bean (*Vicia Faba L*) accessions

**Asnakech Tekalign<sup>1,2\*</sup>; John Derera<sup>1</sup>; Julia Sibiya<sup>1</sup>; Rita H Mumm<sup>3</sup>**

<sup>1</sup>African Centre for Crop Improvement, School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, South Africa

<sup>2</sup>Ethiopian Institute of Agricultural Research (EIAR), Holetta Agricultural Research Centre (HARC), Ethiopia

<sup>3</sup>University of Illinois at Urbana-Champaign, USA

## \*Corresponding Author(s): Asnakech Tekalign

African Centre for Crop Improvement, School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, South Africa

Tel: +251-911984122, Fax: +251-011-2370377

Email: asnakechtek@yahoo.com

## Abstract

Genetic diversity in germplasm is crucial for strategic breeding, however very little is known about the genetic potential of faba bean (*Vicia fabea*) for faba bean improvement and genomics research in Ethiopia. Therefore, forty landraces including ten improved faba bean varieties were characterized using thirty SSR markers to quantify the extent of genetic diversity and determine population structure. There were 220 alleles, per locus averaging 7.86; the (Polymorphic information content) PIC ranged between 0.0 and 0.87 with average of 0.62 using Power Marker and PAST software analysis. The PIC and gene diversity value averaged 0.62 and 0.63, respectively, indicating high genetic diversity among the faba bean collections. Analysis of molecular variance showed that 68.5% of total variation was found within population and population structure analysis showed three major cluster, clusters I and II are landrace and III genotypes from ICARDA. Therefore, these could potentially be used to improve the faba bean gene pool. Further, the generated knowledge about the level of diversity and population structure of faba bean germplasm is an important contribution to this crop breeding and conservation. Besides, the SSR markers used in the present study were effective, highly polymorphic and are recommended in future diversity studies of faba bean.

Received: Aug 31, 2018

Accepted: Jan 17, 2019

Published Online: Jan 21, 2019

Journal: Journal of Plant Biology and Crop Research

Publisher: MedDocs Publishers LLC

Online edition: <http://meddocsonline.org/>

Copyright: © Tekalign A (2018). This Article is distributed under the terms of Creative Commons Attribution 4.0 international License

**Keywords:** Cluster; Faba bean landraces; Genetic diversity; Principal component analysis; Population structure; Molecular markers.



**Cite this article:** Tekalign A, Derera J, Sibiya J, Mumm RH. Molecular analysis for genetic diversity and population structure of Ethiopian faba bean (*Vicia faba L.*) accessions. Plant Biol Crop Res. 2019; 1: 1010.

## Background

Faba bean (*Vicia faba* L.) is thought originated in the Near East and is one of the earliest domesticated legumes after chickpea and pea with the Mediterranean basin as the most important centre of diversity; China, Afghanistan and Ethiopia have also been reported as secondary centres of diversity for the crop [1,2]. It is fourth most important grain legume in tropical and subtropical regions of the world [3]. Faba bean belongs to the family Fabaceae (Leguminosae) with an estimated genome size of ~13000 Mb [4,5]. It is a diploid ( $2n=2x=12$ ) and predominantly a cross-pollinating species (35-55%) [6,7].

In Ethiopia faba bean is grown in the mid-altitude to highland areas with high rainfall and in various types of soils. Faba bean has been cultivated as excellent source of protein for resource poor farmers (27-34%) [8], as cash crop and a major break for cereal mono-cropping system to increase the soil fertility by fixing atmospheric nitrogen [9]. Although less productive, the faba bean landraces have shown good adaptation to local, often stressful, conditions and are preferred by the farmers and consumers for their good taste.

The landraces harbour great genetic potential with alleles to improve agronomic performance, biotic and abiotic stress tolerance and quality characteristics. Therefore, there is potential to develop improved varieties with traits sourced from landrace populations [10].

The potential for improvement of a plant is determined by the level of genetic diversity; thus their use in breeding results in improved food production [11]. For effective breeding and management of genetic diversity, germplasm collections need to be well-characterized. Diversity study in faba bean can be performed using phenotypic, molecular and biochemical markers [12]. However, the phenotypic characters are influenced by environmental factors [13]. Moreover, the phenotypic differences possibly determined by a small number of genes and may not represent genetic divergence in the entire genome [14].

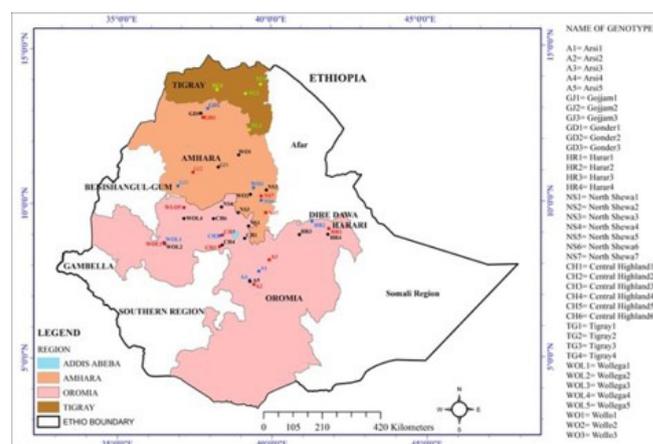
Diversity studies using molecular techniques have been conducted on local collections of faba bean in the world [15,1,16-19]. Terzopoulos and Bebeli [20] confirmed the existence of different germplasm pools in Mediterranean faba bean using ISSR markers. In contrast, narrow genetic diversity was observed in faba bean from China using EST-SSR markers [21]. In Ethiopia the existence of potential genetic diversity of faba bean was reported based on phenotypic characterisation [22]. However, information on genetic diversity is a major challenge for systematic use of faba bean in breeding programmes in Ethiopia. Consequently, the faba bean improvement programme depends on genotypes from other sources mainly the International Centre for Agricultural Research in the Dry Areas (ICARDA) [22]. Therefore, the objectives of this study were to estimate the genetic diversity, population structure and gene flow of faba bean landraces from major faba bean growing areas of Ethiopian highlands.

## Materials and methods

### Plant material

Forty Ethiopian faba bean landraces chosen randomly from collection representing the major faba bean growing areas as well as ten inbred lines were used for this study. The landraces were collected from different major faba bean growing areas of Ethiopian highlands (Figure 1); inbred lines were developed

by pulse crops breeding programme of Holetta Agricultural Research Centre (HARC) and International Centre for Agriculture Research in the Dry Area (ICARDA). The population designation was made by area where the materials collected and more detailed information for each genotype is given in (Table 1).



**Figure 1:** Map of Ethiopia showing faba bean collection sites of major faba bean growing areas of Ethiopian highlands.

### DNA extraction, quantification and quality analysis

Faba bean genotypes were planted in seedling trays at IN-COTEC laboratory in South Africa, the leaf samples were harvested from the two-week-old healthy leaves of one plant per genotypes, folded and stored in to a 14-15 mL test tube. The genomic DNA of the 50 faba bean genotypes was extracted based on the International Maize and Wheat Improvement Centre (CIMMYT) protocol [23]. The purified total DNA was checked for its concentration using spectrophotometer while the quality was examined using 0.7% Tris Borate EDTA (TBE) agarose gel electrophoresis. The final DNA concentration of all extracted DNA stock was adjusted to 10 ng/ $\mu$ l and the DNA samples were stored at 4°C.

### SSR markers and polymerase chain reaction (PCR) amplification

Thirty SSR markers used in the present study were obtained 6 from Gong et al. [21] study (P designation), 5 from Ma et al. [24] study (M designation) and 19 from Zeid et al. [25]. Study (VfG designation). These markers have been recommended by researchers [21,24,25] due to their level of polymorphism and close association with known gene functions. They were recommended for the basic studies, diversity in population or germplasm collection, genetic mapping and marker-assisted breeding of faba bean. Details of the 30 SSR loci used in this study are presented in (Table 2).

PCR amplification was performed using a Gene Amp PCR System 9700 (Applied Bio systems) thermal cycler. Reactions were executed using 12  $\mu$ l of a reaction mixture containing 1 $\times$ PCR reaction buffer, 2.5 mM Mg<sup>2+</sup>, 0.2  $\mu$ l each of dNTPs (Bioline), 1 unit Taq 42 polymerase (Bioline) and 10 ng of genomic DNA. Primers were labelled with a 104 fluorescent dye; two primers were provided for the amplification of each SSR locus: one tailed forward primer (0.05  $\mu$ mol), one normal reverse primer (0.25  $\mu$ mol). The initial denaturation step was performed at 94°C for 2 min, followed by 1 cycle at 94°C for 30 s, 63°C for 30 s and 72°C for 45 s. The annealing temperature was decreased by 1°C per cycle in subsequent cycles until it reached a temperature of 57°C. Products were subsequently amplified for 33 cycles at 94°C for 30 s, 57°C for 30 s, and annealing of primer

at primer specific temperature for 30 s, and 72°C for 45 s with a final extension for 20 min. Electrophoresis was done on the PCR products in 0.7% TBE- agarose gels. The gel then stained in ethidium bromide and observed under a UV transilluminator and the bands were scored.

### Data analysis

Basic statistics were calculated using the genetic analysis package as indicated below. Summary statistics that include the Number of Alleles (NA), Major Allele Frequency (MAF), the heterozygosity, Polymorphic Information Content (PIC) used to measure the informativeness of genetic marker for linkage analysis, and gene diversity [18] were calculated using Power Marker (ver.3.23) [26]. Variance components within and among populations were calculated using analysis of molecular variance (AMOVA) with the software ARLEQUIN V3.1 [27]. As a measure of genetic differentiation between populations a test for significance between population pairs was computed with the F-Statistics estimation (Fst), depict the statistically expected level of heterozygosity in a population, with AMOVA. Genetic relationships among individuals were assessed by multivariate Principal Component Analysis (PCA), which is used to detect patterns of variation in complex data sets, estimates of genetic similarity based on Jaccard's coefficient similarity were computed from the proportion of shared SSR alleles between every pair of faba bean genotypes [28] using PAST software V 3.0 [29]. The pattern of diversity among the genotypes was identified based on Eigen vectors. The genetic distance based clustering was performed with the un-weighted pair group method with arithmetic mean (UPGMA) and the dendrogram showing the relatedness among the 50 faba bean obtained by WADR clustering method using SAS software V. 9.3.[30].

### Population structure

The population structure was analysed using STRUCTUR 2.3.4 which implements a model-based clustering algorithm using the genotype data [31]. It assumes a model in which there are K populations (where K may be unknown), each of which is characterized by a set of allele frequencies at each locus. Individuals are assigned to populations according to their membership coefficients for each cluster. A series of K, from 1 to 20, was used to estimate the number of clusters under the admixture model with allele frequencies correlated. For each K, 20 independent runs of 10000 iterations were processed following a burn-in period of 50000 iterations. The optimum K value, which indicates the number of genetically distinct clusters in the data, was determined from 20 replicate runs for each of K [32]. The  $\Delta K$  was calculated based on the rate of change of the log-likelihood between successive K values. Software program Structure Harvester V 0.6.93 [33] was used for calculating parameters of Evanno et al. [32]. Following the method of Evanno et al. [32], the  $\Delta K$  were plotted against the K numbers of the groups. The software package CLUMPP was used to combine the STRUCTURE group-membership output data for each population from 20 replicates run from the molecular data for K=3 [34]. The optimal number genetic structure (the maximum value of  $\Delta K$  was graphically displayed using DISTRUCT [35].

## Results

### Marker characterisation

#### Number of alleles gene diversity and polymorphic information content

Twenty eight of the 30 SSRs were polymorphic among the 50 entries (Table 3). The fragment size of the alleles ranged from 120 bp (VfG31 and VfG55) to 326 bp (VfG11). A total of 222 alleles were detected with 30 SSR markers. The number of alleles detected by a single SSR locus varied from 1 (for VfG87 and M10) to 24 (for VfG67) with an average of 7.4 per marker (Supplementary Table 1). These, markers VfG87 and M10 were no longer useful to this analysis and were dropped. Heterozygosity (He) across all loci was very high ranging from ranged from 0.39 (VfG19) to 1.00 (M17, VfG3, VfG 10, VfG27 and VfG 81), with a mean of 0.92. The major allele frequency ranged from 0.14 (VfG28) to 0.83 (VfG19) with a mean of 0.44 (Table 3). Gene diversity, or expected heterozygosity per locus in a population, is used to quantify the genetic variation, evaluate genetic divergence and population relationship and detect inbreeding. The gene diversity scores of the 28 polymorphic SSR loci ranged from 0.28VfG19) to 0.91 (VfG28) with mean of 0.68 (Table 3). The PIC value, for 28 SSR loci varied from 0.32 (VfG19) to 0.91 (VfG28). All except two loci (P41, and VfG19) showed high PIC values (>0.5) with an average of 0.67 (Table 3).

### Analysis of molecular variance (AMOVA)

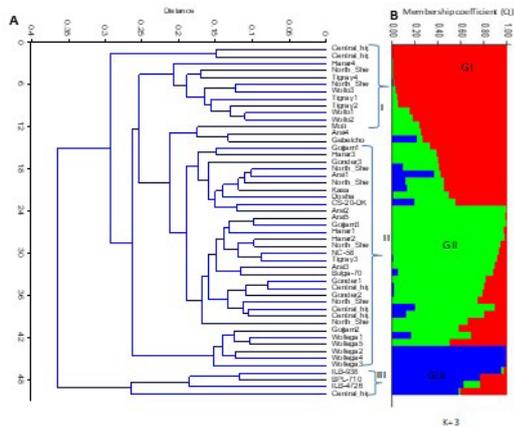
There was higher and significant ( $P<0.001$ ) distribution of genetic variation within population grouping based on geographical location (Table 4). There was 68.54% of the molecular variation within populations and 31.09% among populations. The Fixation index (Fst), a measure of population differentiation due to genetic structure, value of 0.31460 was observed showing the extent of differentiation of populations (Supplementary Table 2). Thus, population differed from one another and lines within population also exhibited diversity.

### Genetic differentiation analysis

The pair wise FST value revealed the lowest genetic differentiation (0.087) between Tigrai and Harer populations (Table 5, below diagonal). Alternatively, it was highest between the faba bean genotypes from ICARDA and other populations (Wollega 0.77, Wollo 0.72, North Shewa 0.69, Gojjam 0.69, Tigrai 0.68, Central highland 0.58 and HARC 0.55). Significant variation ( $P<0.001$ ) of population differentiation was observed between population of Arsi and Central highland, Arsi and Wollo, Arsi and North Shewa, Arsi and ICARDA materials, Central highland and North Shewa, improved materials from HARC and Central highland, North Shewa and Harar, Wollega and Central highland, Wollo and Central highland population (Table 5, above diagonal).

### Population structure and cluster analysis

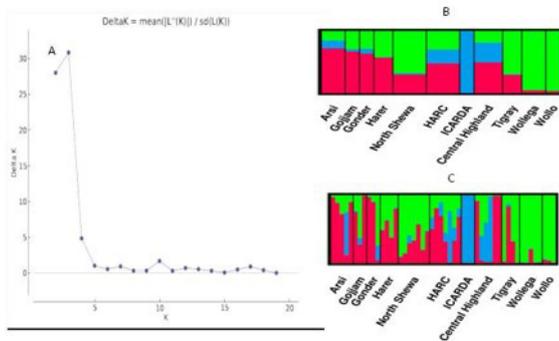
The dendrogram classified the germplasm into three major clusters (Figure 2). Cluster I and cluster II were further subdivided into sub-clusters. Cluster II comprised the landraces from different locations, and four old small seeded improved varieties (CS-20-DK, NC58, Bulga-70 and Kasa) and two recently released varieties, Moti and Dosha (Moti and Dosha). All faba bean landrace collections from different parts of Wollega (Wollwga1, Wollwga2, Wollwga3, Wollwga4 and Wollwga5) were also grouped in cluster II. Cluster III comprised the exotic genotypes from ICARDA (ILB-938, ILB-4726, and BPL-710). The Genotype Central highland4, a faba bean collection from the Central highland of West Shewa Chalya, Arsi4, collection from Arsi zone Dawa Bursa and Gebelcho an improved variety from HACR were grouped in cluster III.



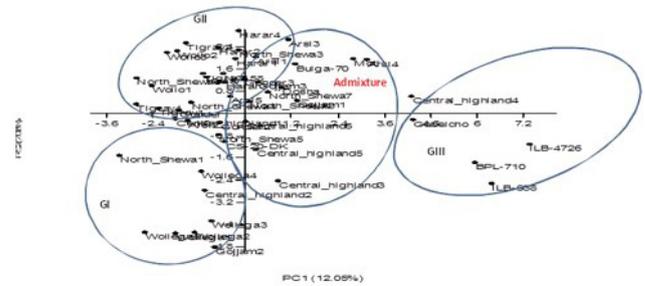
**Figure 2 (A):** MUPGMA dendrogram for fifty faba bean genotypes based on the Jaccard coefficient as revealed using SSR markers (the name indicated genotype name and collection place listed in (Table 1) and (B) bar plot of K=3, estimates of membership coefficient, each vertical bar represents the membership coefficient for an individual genotype grouped into three GI, GII and GIII.

**Estimation of the optimal number of cluster in structure**

There were three different genetic groups (GI, GII and GIII) (Figure 2A and B). The maximum  $\Delta K$  occurred at K=3 and at K=3, the faba bean genotypes divided into three clusters (Figure 2B & Figure 3A). All ICARDA faba bean genotypes were fully grouped together in the DISTRIBUTE plot (Figure 3B) the rest of the population showed admixture (Figure 4). As shown in Figure 3C some of the individual genotypes from populations of Gojjam, Central highland, Tigray, Wollega, and Wollo were found not admixtures. All the improved faba bean materials from HARC were admixtures, shared genetic components across groups.



**Figure 3:** Inferred population structure of faba bean genotypes: Plot of (A) the relationship between  $\Delta K$  and K showing the highest peak at K=3; (B) DISTRIBUTE plot for 50 faba bean genotypes based on the STRUCTURE analysis, each color represent a different cluster, and black segments separate the population, population names are below the Figure and (C) individual genotypes from each population (collection site and source) is represented by a single vertical line partitioned in to K=3. Each color represents one cluster, and the length of the colored segment shows the genotype's estimated proportion of membership in that cluster as calculated by STRUCTURE. The color code for the inferred three cluster is 1=Red, 2=Green, 3= Blue.



**Figure 4:** Scatter plot of PC1 and PC2 based on the similarity of 30 SSR markers for 50 faba bean germplasm. Different genotypes grouped into three main groups and one admixture group.

**Individual genotype identification**

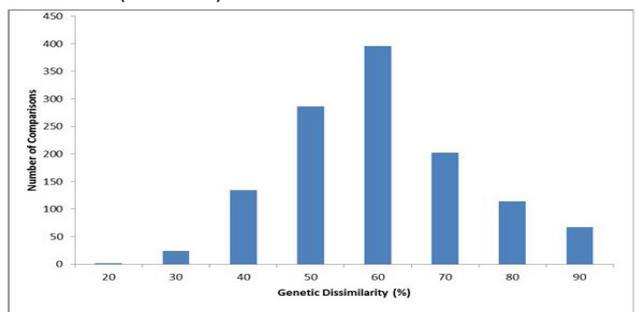
The individuals were assigned to clusters (Supplementary Table 2). Genotypes with membership coefficient > 0.800 were assigned to the respective cluster completely and individuals with <0.800 indicated that they were admixed and were assigned to two or more population clusters. All genetic groups comprised individuals with a high estimated membership coefficient for the respective cluster.

**Principal component analysis**

The first four principal components explained 30% (12.05% for PC1, 7.08% for PC2, 5.16% for PC3 and 4.67% for PC4) of the total diversity. The test genotypes were divided into three distinct groups with the ICARDA materials fully separated from the others (Figure 4). In the PCA plot, the faba bean germplasm were divided into three main groups and one admixture with 12, 16, 5 and 17 genotypes in Cluster I, II, III and admixture respectively. Group III consisted of faba bean exotic varieties from ICARDA (FBV26-ILB-938, FBV28-ILB-4726, and FBV31-BPL 710). This group also comprised of improved faba bean variety Gebelcho-FBV29 and landrace collection (FBColl-36). Thus, the grouping obtained by unweighted pair-group method with arithmetic mean (UPGMA) dendrogram and plot from STRUCTUR analysis was confirmed by Principal Component Analysis (PCA).

**Genetic dissimilarity among genotypes**

Genetic dissimilarity >50% was observed from about 86.94% of the pair wise comparison among the faba bean genotypes (Figure 5). The highest genetic dissimilarity coefficient (93%) was observed between faba bean collection from West Shewa (FBColl-036) and faba bean collection from Arsi Zone (FBColl-003). The lowest value of genetic dissimilarity coefficient (29%) was observed between faba bean varieties ILB-4726 (FBV-028) and ILB-938 (FBV-026).



**Figure 5:** Histogram showing the distribution of 1245 pair wise genetic dissimilarity coefficients among 50 faba bean germplasm.

## Discussion

In this study on average 7.4 alleles were detected per locus whereas Gong et al. [36] detected 2.3 alleles per locus among lines tested. Such considerable differences in the number of alleles detected may arise from differences in the diversity of the test genotypes used, the number of genotypes examined and the genotyping method used. The SSR markers exhibit relative abundance and co-dominant inheritance and are useful for estimating genetic relationship and diversity [37]. In this study, the SSR markers were able to detect considerable level of genetic diversity present among the tested faba bean genotypes.

High average heterozygosity of 0.85 was detected in this study, which could reflect the partial cross-pollination nature of faba bean [10]. This heterozygosity is important in creating genetic variability in faba bean [38, 39]. The results have implications for breeding new varieties. There is potential for selection within the populations, which is consistent with previous studies. Link et al. [40] reported the possibility of considerable potential for selection within populations for specific traits from highly heterogeneous and heterozygous plants in faba bean.

Polymorphic information content (PIC) value provides an estimate of discriminating power of a marker based on the number of alleles at a locus and relative frequencies of these alleles. In the present study nearly 86% of these markers had high PIC values (>0.5) with an average of 0.62. This indicates that the markers were highly polymorphic, informative and were useful in discriminating the faba bean genotypes. This is considered to be high based on the previous classification by Botstein et al. [41], who indicated that PIC values >0.5 represents highly informative markers, 0.5 > PIC >0.25 is informative marker, and PIC < 0.25 is a slightly informative marker. In the present study the high polymorphic rate for most (86%) of the markers and PIC value, together with more than 50% genetic dissimilarity for about 87% pair wise comparison in the test faba bean genotypes, suggests a high level of heterogeneity. Furthermore, the SSR markers used in the present study were effective and highly polymorphic. Therefore, the set of markers used are recommended in other future evaluations of faba bean germplasm.

The AMOVA results indicated high genetic variation which is consistent with previous studies on faba bean. The findings from this study indicate that faba bean populations from the Ethiopian highlands are highly variable. This is in line with the study by Kwon et al. [16] who reported a large amount of variation within faba bean landraces. Recently, Oujii et al. [19] reported 74.3% genetic diversity within nine populations of Tunisian faba bean than that of among population which is also congruent with findings from the current study. Similarly, Terzopoulos and Bebeli [20] reported a high level of genetic diversity within the populations of Mediterranean-type faba bean. The study, therefore, adds crucial evidence that variation within populations is large in faba bean.

Higher genetic diversity within populations is expected [42] than that of among populations since faba bean is largely an out-crossing crop [43]. The small differentiation, low percentage of variation (31.46%) partitioned among populations of faba bean from different locations observed in this study could be attributed to the exchange of faba bean seed among farmers. Ethiopian faba bean growers get their seed mostly by informal seed exchange system. In this line there has been a lot of genetic exchange among farmers' landrace for seed [44]. However, in this study large  $F_{ST}$  value (0.17-0.23) between the population of

Arsi, Central highland, and North Shewa suggests that although they are located geographically near each other, the collections from these locations have high genetic differentiation.

The clustering of genotypes into three major groups reflects the origin of the genotypes and known pedigree relationships. For example, the improved variety Gebelcho grouped with the ICARDA materials because it is derived from a cross of ILB-4726×Tesfa and released from HARC [45]. The results showed well-defined distribution patterns of the materials, according to the genetic distance and the membership coefficient relationships among them [35]. In the PCA plot, the faba bean germplasm were divided into three main groups and admixture. These groups reflect the breeding history and the relation between improved, exotic and landrace collection. Admixture situation in the genotypes shows good agreement with pedigree information. Both the PCA and STRUCTURE analysis also suggested the existence of three major groups. The admixture from small membership coefficient value in the PCA is explained as the presence of gene flow of the genotype which had infiltrated into other population clusters [46]. This observed admixture situation is due to the largely out-crossing nature of faba bean [47] and artificial crossing and hybridization.

The high percentage genetic dissimilarity coefficient for the majority of pair wise comparison among the faba bean genotypes suggests lesser amount of genetic relatedness and elucidates low genetic similarity. The present result is different from Kwon et al. [16] who reported that high genetic similarity exhibited for the majority of the pair wise comparison among worldwide collected faba bean entries. It was observed substantial genetic diversity and clear population structure using unexploited set of landraces. Since the landraces included in this study possess desirable attributes like resistance to abiotic factors, they qualify as suitable parental choice for varietal development in faba bean.

## Conclusion

The present study suggests that the faba bean collections used were genetically variable and shows evidence of considerable gene flow among the geographical locations where the materials were collected. The choice of parents within faba bean population would be recommended to explore the inter-clusters variability, followed by crosses among different populations to explore the inter-population variability for traits of interest of the breeding program. This information is helpful for developing appropriate science based strategies for faba bean breeding. However, it is suggested that more molecular work such as sequencing of the faba bean genomes is vital to explore the available diversity and to have better understanding of the presence of genetic variability in faba bean and consequently utilization of existing variability for improvement of faba bean for the challenges of faba bean production in Ethiopia.

## Acknowledgement

The authors are grateful to the African Centre for Crop Improvement (ACCI) and the Alliance for a Green Revolution in African (AGRA) for financial support to the first author. The Ethiopian Institute of Agricultural Research (EIAR) and Holetta Agricultural Research Centre (HARC) are gratefully acknowledged for hosting the study and providing research facilities.

## Tables

**Table 1:** Description of faba bean genotypes used for the study

Genotype name	Population designation	Sources of the genotyp	Population type	Altitude (masl)
Arsi1	Pop1	Arsi/Bekoji	Landrace	2784
Arsi2	Pop1	Arsi/ArsiRobe	Landrace	2410
Arsi3	Pop1	Arsi/LimuJara	Landrace	2853
Arsi4	Pop1	Arsi/DawaBursa	Landrace	2908
Arsi5	Pop1	Arsi/Chole	Landrace	3050
Gojjam1	Pop2	West Gojjam/Yilmanadensa	Landrace	2240
Gojjam2	Pop2	West Gojjam/ Awi	Landrace	2610
Gojjam3	Pop2	East Gojjam/ Hulet-Eju Enebse	Landrace	2412
Gonder1	Pop3	North Gonder Zone/Wogera	Landrace	2961
Gonder2	Pop3	North Gonder Zone/Debark	Landrace	2739
Gonder3	Pop3	North Gonder Zone/Wogera	Landrace	2951
Harar1	Pop4	Harer/Borda	Landrace	2240
Harar2	Pop4	Harer/Kulubi	Landrace	2380
Harar3	Pop4	Harer/Arberkoti	Landrace	2266
Harar4	Pop4	Harer/Gerawa	Landrace	2288
North Shewa1	Pop5	North Shewa/MoretnaJiru	Landrace	2663
North Shewa2	Pop5	North Shewa/Basonaworana	Landrace	3012
North Shewa3	Pop5	North Shewa/Mehalmeda	Landrace	2667
North Shewa4	Pop5	North Shewa/ Ankober	Landrace	3200
North Shewa5	Pop5	North Shewa/ Molale	Landrace	3100
North Shewa6	Pop5	North Shewa/ Tarmaber	Landrace	3058
North Shewa7	Pop5	North Shewa/Hageremariyam	Landrace	2670
CS-20-DK	Pop6	HARC	Inbred line	
NC-58	Pop6	HARC	Inbred line	
Bulga-70	Pop6	HARC	Inbred line	
Kasa	Pop6	HARC	Inbred line	
Gebelcho	Pop6	HARC	Inbred line	
Moti	Pop6	HARC	Inbred line	
Dosha	Pop6	HARC	Inbred line	
ILB-938	Pop7	ICARDA	Inbred line	
ILB-4726	Pop7	ICARDA	Inbred line	
BPL-710	Pop7	ICARDA	Inbred line	
Central highland1	Pop8	East Shewa/Chefe Donsa	Landrace	2263
Central highland2	Pop8	South West Shewa/TuluBolo	Landrace	2192
Central highland3	Pop8	South West Shewa/ Tole	Landrace	2127
Central highland4	Pop8	West Shewa/ Chalya	Landrace	2839
Central highland5	Pop8	West Shewa/ Dendi	Landrace	2270
Central highland6	Pop8	South West Shewa/ Elu	Landrace	2100
Tigray1	Pop9	Central Tigray/MayTimeket	Landrace	1855
Tigray2	Pop9	Eastern Tigray/Saese Tsaedamba	Landrace	2547
Tigray3	Pop9	Eastern Tigray/Atsiwemberta	Landrace	2840
Tigray4	Pop9	East-southern Tigray/Degua Tembien	Landrace	2770
Wollega1	Pop10	Horo Guduru Wollega	Landrace	2451
Wollega2	Pop10	Horo-guduru wollega	Landrace	2296
Wollega3	Pop10	East Wollega/Jima Arjo	Landrace	2424

Wollega4	Pop10	Horo Guduru Wollega/Horo	Landrace	2435
Wollega5	Pop10	East Wollega/Jima Arjo	Landrace	2460
Wollo1	Pop11	South Wollo/ Ambasel	Landrace	2972
Wollo2	Pop11	North Wollo/Wadla	Landrace	2930
Wollo3	Pop11	South Wollo/Wereilu	Landrace	2640

**Table 2:** Description of the 30 SSR markers used in the study.

SSR Markers	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')	Size range (bp)
P27	< HEX > CGGGTTTATTCCATCATTT	CGTTATTGTTGTCGCTATTT	210-235
P28	< HEX > CCATCTCCACCACCAG	CACAACGGCTTCAAAT	195-215
P41	< HEX > AACTGTTCACCGGTAT	GGAGTGATTAGAAGGTAG	152-166
P99	< FAM > GATGGGAGCTGAGAATGAT	TTAAACCAAACACCAGAGT	192-212
P139	< FAM > AACCCATCTGGAAGAAACA	TTGAGAATCCGAAGAAACC	195-222
M10	CCATGGTTACTGCAGTCGAA	GCCGTCGATTGATTCGTATT	202-216
M17	CTCCAACGAAGGCAGAGAAG	CATGATTCCTATAGCCTTGC	220-260
M41	CAACGCGGCAGTTAAGAAT	CAGGTATGGCTGACACCTCA	180-255
VfG1	TTTCAGCAAACCTAGAACCAATC	GGCATTGAGTTTTTACCTTGTA	218-272
VfG3	TTCTTTGGTCTCTCTATC	GCACTGTTGTTGCTGATACAA	164-184
VfG4	AAGGGGAGGGCATAACAGAA	AATCCGCAAGGGTCTTCTTT	225-250
VfG10	ACCAAAACGCGCACTTATCA	AAGAGAGAGAAGAGAGCTTC	205-240
VfG13	GGTTGGGATCTTTAGGTTGAA	TGGCCTTATATCCGTCGAAT	195-217
VfG87	AGGGCCAGCGTGATCCAATA	TGGGTTGGGATCTTTGGTTG	232-245
VfG19	AGCGATGGTGCTCATGCTTA	TCTCTCACGGAATCACATCTTT	180-210
VfG28	AGAGTCCCAAAGAGTGGGTT	CCAAAGGCAAAAATGAGGGCTT	215-257
VfG31	ATAAGAGAGAACGAGGGAGAA	TTATGGTGGGACGCTTACAT	120-150
VfG34	GCACTCGAAGGAATTAATTTT	GAACAGTTGTTTCGTGTCGTA	202-215
VfG53	GGTTCATGAAAAGAGGTTAG	CATTTCCGTTCTCTCTCTA	234-266
VfG55	ATCATCCAGGAGGGAGAAAA	ATGGGCAGAGAGGATAAAAA	120-180
VfG11	GCAAAAGGAGAGCAAGGGAA	CGAAAGAGGGGACATTTTGT	310-326
VfG67	GTTTCATCAAGCACAATCTAAAC	TCAATTTGGTTTATCTCTCTCTCT	122-190
VfG27	CCCAAAAAGAGACGAAGTGTAT	AGGGTTCATACGTTTGGCTT	196-236
P11	< FAM > CGTGGTTGATTTGCTTC	ACCTCCATCTTCGCCTC	166-192
VfG9	GGTTTTGAATAGAAATGCAA	AAGATGTGTCAATATTGTTTT	148-196
VfG41	CAAGCTTGTGAGAGCCAAA	GAACGAGGCTCACGAAAATA	142-170
M49	GCGTTATTAGCCGCTGTAA	AAAACCGTGGCTCGAATATTTA	240-310
VG47	CGATTGTTGCAGAGGAGATA	ACAGAGAGGGACAGAGAGAA	290-306
M57	TGCAGAGAAGCTAAGCACCA	TCGATGGTACAGTAGCAAAA	225-272
VfG81	GTCCTGAAAAAAGAAAGAGA	AAAGAACTCTCTCTCCAT	141-207

**Table 3:** Summary statistics for the 28 polymorphic SSR markers used in this study.

SSR Markers	Alleles number	Heterozygosity	Major allelic frequency	Gene Diversity	PIC value
P27	3	0.92	0.46	0.62	0.59
P28	8	0.98	0.24	0.82	0.81
P41	2	0.92	0.54	0.5	0.38
P99	3	0.98	0.4	0.69	0.68
P139	8	0.94	0.44	0.69	0.67
M17	3	1	0.5	0.55	0.46
M41	7	0.94	0.47	0.65	0.66

VfG1	13	0.93	0.5	0.72	0.76
VfG3	5	1	0.45	0.72	0.76
VfG4	5	0.98	0.48	0.65	0.62
VfG10	14	1	0.4	0.78	0.77
VfG13	9	0.92	0.45	0.71	0.73
VfG19	3	0.39	0.83	0.28	0.32
VfG28	18	0.97	0.14	0.91	0.91
VfG31	9	0.96	0.44	0.7	0.71
VfG34	4	0.98	0.51	0.59	0.55
VfG53	5	0.87	0.54	0.61	0.58
VfG55	12	0.98	0.33	0.82	0.81
VfG11	7	0.95	0.36	0.73	0.74
VfG67	24	0.96	0.33	0.85	0.84
VfG27	10	1	0.35	0.8	0.87
P11	5	0.9	0.43	0.65	0.66
VfG9	12	0.96	0.22	0.87	0.87
VfG41	2	0.96	0.52	0.5	0.38
M49	4	0.67	0.64	0.51	0.59
VG47	4	0.67	0.52	0.61	0.59
M57	17	0.92	0.45	0.75	0.74
VfG81	4	1	0.45	0.67	0.62
Total	220	25.65	12.39	18.95	18.7
Mean	7.86	0.92	0.44	0.68	0.67

Where: PIC: Polymorphic Information Content

**Table 4:** Analysis of molecular variance of 50 faba bean genotypes grouped in to population based on their geographical location.

Source of variation	d.f	Sum of squares	Mean squares	Variance component	Percentage of variation	Probability
Among populations	10	385.59	38.559	3.44019 Va	31.09	
Within populations	89	678.61	7.6248	7.62483 Vc	68.54	0.0001
Total	99	1064.2	10.75	11.065		
Fixation Index	<b>FST: 0.31460</b>					

**Table 5:** Population pair wise FST values, (below diagonal) and their P values (above diagonal).

	Arsi	Gojjam	Gonder	Harer	North shewa	HARC	ICARDA	Central highland	Tigray	Wollega	Wollo
Arsi	0	0.189	0.739	0.315	0.018	0.108	0.009	0.027	0.315	0.018	0
Gojjam	0.202	0	0.45	0.468	0.018	0.081	0.054	0.09	0.333	0.054	0.153
Gonder	0.107	0.263	0	0.901	0.009	0.225	0.072	0.072	0.622	0.009	0.081
Harer	0.143	0.165	0.108	0	0.009	0.054	0.045	0.036	0.892	0.036	0.153
North shewa	0.169	0.269	0.254	0.222	0	0.009	0	0	0.577	0.009	0.018
HARC	0.144	0.175	0.165	0.181	0.178	0	0.018	0	0.162	0.009	0.027
ICARDA	0.65	0.687	0.668	0.675	0.687	0.551	0	0.018	0.009	0.018	0.09
Central highland	0.225	0.242	0.247	0.243	0.216	0.217	0.575	0	0.045	0.009	0.018
Tigray	0.147	0.177	0.156	0.087	0.09	0.148	0.677	0.207	0	0.685	0.955
Wollega	0.283	0.338	0.31	0.21	0.191	0.272	0.769	0.302	0.092	0	0.324
Wollo	0.307	0.345	0.305	0.252	0.247	0.243	0.716	0.295	0.098	0.166	0

## References

1. Zong X, Liu X, Guan J, et al. Molecular variation among Chinese and global winter faba bean Germplasm. *Theoretical and Applied genetics*. 2009; 118: 971-978.
2. Maxted N, Kell SP. Establishment of a global network for the in situ conservation of crop wild relatives: Status and needs. FAO commission on genetic resources for FAO, Rome. Italy. 2009.
3. FAOSTAT Data base. Available at: <http://faostat.fao.org/beta/en/#data/QC>. 2016.
4. Maxted N, Bennett S. Plant genetic resources of legumes in the Mediterranean. *Current plant science and biotechnology in agriculture*. Kluwer Academic Press. Netherlands. 2001; 1-386.
5. Torres AM, Rom B, Avila CM, et al. Faba bean breeding for resistance against biotic stresses: Towards application of marker technology. *Euphytica*. 2006; 147: 67-80.
6. Metz PLJ, Buiel AAM, Norel AV, Helsen JPF. Rate and inheritance of cross-fertilization in faba bean (*Vicia faba* L.). *Euphytica*. 1993; 66: 127-133.
7. Suso MJ, Pierre J, Moreno MT, et al. Variation in outcrossing levels in faba bean cultivars: role of ecological factors. *The Journal of Agricultural Science*. 2001; 136: 399-405.
8. Hove EL, King S, Hill GD. Composition protein quality and toxins of seeds of the grain legumes *Glycine max*, *Lupinus spp.* *Phaseolus spp.* *Pisum sativum*, and *Vicia faba*. *Journal of Agricultural Research*. 1978; 21: 457-462.
9. López-Bellido L, López-Bellido RJ, Redond R, Benítez J. Faba bean nitrogen fixation in a wheat-based rotation under rainfed Mediterranean conditions: Effect of tillage system. *Field Crops Research*. 2006; 98: 253-260.
10. Gnanasambandam A, Paull J, Torres A, et al. Impact of molecular technologies on faba bean (*Vicia faba* L.) breeding strategies. *Agronomy*. 2012; 2: 132-166.
11. Khodadadi M, Fotokian MH, Miransari M. Genetic diversity of wheat (*Triticum aestivum* L.) genotypes based on cluster and principal component analyses for breeding strategies. *Australian Journal of Crop Sciences*. 2011; 5:17-24.
12. Chahal GS, Gosal SS. Principles and procedures of plant breeding: biotechnological and conventional approaches Fla: CRC Press. Boca Raton. 2002.
13. Fufa H, Baenziger PS, Beecher BS, et al. Comparison of phenotypic and molecular marker-based classifications of hard red winter wheat cultivars. *Euphytica*. 2005; 145: 133-146.
14. Brown-Guedira GL, Thompson JA, Nelson RL, Warburton ML. Evaluation of Genetic Diversity of Soybean Introductions and North American Ancestors Using RAPD and SSR Markers. *Crop Science*. 2000; 40: 815-823.
15. Zeid M, Schon CC, Link W. Genetic diversity in recent elite faba bean lines using AFLP markers. *Theoretical and Applied genetics*. 2003; 107: 1304-1314.
16. Kwon SJ, Jinguo H, Clarice JC. Genetic diversity and relationship among faba bean (*Vicia faba* L.) germplasm entries as revealed by TRAP markers. *Plant Genetic Resources*. 2010; 8: 204-213.
17. Zong, X, Ren J, Guan J, et al. Molecular variation among Chinese and global germplasm in spring faba bean areas. *Plant Breeding*. 2010; 129: 508-513.
18. Alghamdi SS, Al-Faifi SA, Migdadi HM, et al. Molecular Diversity Assessment Using Sequence Related Amplified Polymorphism (SRAP) Markers in *Vicia faba* L. *International Journal of Molecular Sciences*. 2012; 13: 16457-16471.
19. Ouji A, El-Bok S, Syed NH, et al. Genetic diversity of faba bean (*Vicia faba* L.) populations revealed by Sequence Specific Amplified Polymorphism (SSAP) markers. *African Journal of Biotechnology*. 2012; 11: 2162-2168.
20. Terzopoulos PJ, Bebeli PJ. Genetic diversity analysis of Mediterranean faba bean (*Vicia faba* L.) with ISSR markers. *Field Crops Research*. 2008; 108: 39-44.
21. Gong YM, Xu SC, Mao WH, et al. Generation and characterization of 11 novel EST derived microsatellites from *Vicia faba* (Fabaceae). *American Journal of Botany*. 2010; 97: 69-71.
22. Gemechu K, Mussa J, Tezera W. Faba bean (*Vicia faba* L.) Genetics and Breeding Research in Ethiopia: A Review. In: *Food and Forage legumes of Ethiopia: Progress and Prospects. Proceedings of the Workshop on Food and Forage Legumes* (Eds Kemal, A. et al.): ICARDA: Aleppo. Syria. 2006; 43-66.
23. CIMMYT. Laboratory Protocols: CIMMYT Applied Molecular Genetics Laboratory. Third Edition. Mexico, DF: CIMMYT. 2005.
24. Ma Y, Yang T, Guan J, et al. Development and characterization of 21 EST-derived Microsatellite Markers in *Vicia faba* (Faba bean). *American Journal of Botany*. 2011; 98: e22-e24.
25. Zeid M, Mitchell S, Link W, et al. Simple Sequence Repeats (SSRs) in faba bean: new loci from Orobanche-resistant cultivar 'Giza 402'. *Plant Breeding*. 2009; 128: 149-155.
26. Liu KJ, Muse SV. Power Marker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*. 2005; 21: 2128-2129.
27. Excoffier L, Laval G, Schneider S. Arlequin ver 3.1: An Integrated Software Package for Population Genetics Data Analysis. *Evolutionary Bioinformatics*. 2006; 1: 47-50.
28. Jaccard P. Nouvelles recherches sur la distribution florale. *Bulletin Society Vaud Science National*. 1908; 44: 223-270.
29. Hammer O, David ATH, Paul DR. Paleontological statistics software package for education and data analysis. *Palaeontologia electronica*. 2001; 4: 1-9.
30. SAS. SAS proprietary software. Release 9.3 SAS Inst, Cary, NC, USA. 2012.
31. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000; 155: 945-959.
32. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*. 2005; 14: 2611-2620.
33. Earl DA, vonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 2012; 4: 359-361. <http://taylor0.biology.ucla.edu/structureHarvester/>.
34. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*. 2007; 23: 1801-1806.
35. Rosenberg NA. DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*. 2004; 4: 137-138.
36. Gong YM, Sheng-chun X, Wei-hua M, et al. Genetic Diversity Analysis of Faba Bean (*Vicia faba* L.) Based on EST-SSR Markers. *Agricultural Sciences in China*. 2011; 10: 838-844.

- 
37. Kumar P, Gupta VK, Misra AK, et al. Potential of Molecular Markers in Plant Biotechnology. *Plant Omics Journal*. 2009; 2: 141-162.
38. Gregorius HR. The relationship between the concepts of genetic diversity and differentiation. *Theoretical and Applied Genetics*. 1987; 74: 397-402.
39. Hegaya S, Geletab M, Bryngelssonb T, et al. Comparing genetic diversity and population structure of common beans grown in Kyrgyzstan using microsatellites. *Scientific Journal of Crop Science*. 2012; 1: 63-75.
40. Link W, Schill B, Kittlitz EV. Breeding for wide adaptation in faba bean. *Euphytica*. 1996; 92: 185-190.
41. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human genetics*. 1980; 32: 314-331.
42. Smýkal P. Development of an efficient retrotransposon- based fingerprinting method for rapid pea variety identification. *Genetic Resource and Crop Evolution*. 2006; 47: 221-230.
43. Suso MJ, Gilsanz S, Duc G, et al. Germplasm management of faba bean (*Vicia faba* L.): Monitoring intercrossing between accessions with inter-plot barriers. *Genetic Resource and Crop Evolution*. 2006; 0: 1-13.
44. Alemu D, Rashid S, Tripp R. Seed system potential in Ethiopia: Constraints and opportunities for enhancing the seed sector. International Food Policy Research Institute CGIAR. Addis Ababa. 2010.
45. Ministry of Agriculture and Rural Development, MOARD. Crop Development Department Crop variety register. 2006.
46. Zhang P, Li J, Li X, et al. Population structure and genetic diversity in a rice core collection (*Oryza sativa* L.) investigated with SSR markers. *PLoS ONE*. 2011; 6: e27565.
47. Suso MJ, Harder L, Moreno MT, Maalouf F. New strategies for increasing heterozygosity in crops: *Vicia faba* mating system as a study case. *Euphytica*. 2005; 143: 51-65.