academicJournals

Vol. 15(52), pp. 2874-2897, 28 December 2016
DOI: 10.5897/AJB2016.15464
Article Number: 528FF4562176
ISSN 1684-5315

http://www.academicjournals.org/AJB

African Journal of Biotechnology

Full Length Research Paper

Genetic diversity and population structure of common bean (*Phaseolus vulgaris* L) germplasm of Ethiopia as revealed by microsatellite markers

Zelalem Fisseha[1,2], Kassahun Tesfaye[1], Kifle Dagne[1], Matthew W. Blair[3], Jagger Harvey[4], Martina Kyallo[4], Paul Gepts[5*]

[1]Addis Ababa University, Department of Microbial, Cellular, and Molecular Biology, P.O.Box 1176, Addis Ababa, Ethiopia.
[2]Department of Dryland Crops, College of Dryland Agriculutre, Jigjiga University, Crop Research Directorate, Somali Region Pastoral and Agro-pastoral Research Institute (SoRPARI), P.O.Box 398, 1020, Jijiga, Ethiopia.
[3]Department of Agricultural and Environmental Sciences, Tennessee State University, Nashville, TN 37209, United States of America.
[4]Biosciences Eastern and Central Africa (BecA), ILRI, Nairobi, Kenya.
[5]Department of Plant Sciences / MS1, University of California, 1 Shields Avenue, Davis, CA 95616, USA.

Accepted 13 May, 2016; Accepted 9 December, 2016

The Ethiopian genetic center is considered to be one of the secondary centers of diversity for the common bean. This study was conducted to characterize the distribution of genetic diversity between and within ecological/geographical regions of Ethiopia. A germplasm sample of 116 landrace accessions was developed, which represented different common bean production ecologies and seed types common in the country. This sample was then analyzed with 24 simple sequence repeat (SSR) markers to assess the genetic diversity within and between common bean landraces, classifying them based on SSR clustering, and determining relationships between genetic and agroecological diversity. Representatives of both Andean and Mesoamerican gene pools were identified by STRUCTURE software analysis, as well as a high proportion of hybrid accessions as evidenced by a STRUCTURE K = 2 preset. At the optimum K = 5 preset value, mixed membership of Andean and Mesoamerican genotypes in some of the clusters was also seen, which supported previous findings. Cluster analyses, principal coordinate analysis, and analysis of molecular variance all indicated clustering of accessions from different collection sites, accompanied by high gene flow levels, highlighting the significant exchange of planting materials among farmers in different growing regions in the country. Values of allelic diversity were comparable to those reported in previous similar studies, showcasing the high genetic diversity in the landrace germplasm studied. Moreover, the distribution of genetic diversity across various bean-growing population groups in contrasting geographical/ecological population groups suggests elevated but underutilized potential of Ethiopian germplasm in common bean breeding. In summary, this study demonstrated the geographical, as well as gene pool diversity in common bean germplasm of Ethiopia. This substantial diversity, in turn, should be utilized in future common bean breeding and conservation endeavors in the nation.

Key words: Hybridity, simple sequence repeat, microsatellite, structure, seed exchange, gene flow.

INTRODUCTION

Common bean is the most widely consumed legume species of the genus *Phaseolus* (Freytag and Debouck, 2002). It is a pulse crop used since pre-Columbian times in the Americas and, since the 16th century, in other regions of the world (Gepts et al., 2008). It is a true diploid (2n = 2x = 22) with a small genome (580 Mbp; Broughton et al., 2003). Originating in the Neotropics, common bean was domesticated in Mesoamerica and the Andes (Gepts and Bliss, 1986; Gepts, 1988). The crop has high diversity that is broadly classified into six or seven domesticated races distributed into two gene pools (Singh et al., 1991a, b, c; Blair et al., 2007, 2010b; Pallottini et al., 2004; Kwak and Gepts, 2009; Kwak et al., 2012). The crop is a major legume in Eastern and Southern Africa, occupying more than 4 million ha annually and providing food for ≥100 million people (Wortmann et al., 1998; Fisseha, 2015). Of the total production in sub-Saharan Africa of over 3.5 million MT, 62% is in Eastern and Central African countries (Wortmann et al., 1998; Fisseha, 2015). Common bean became established with the African-European trade, even before the widespread era of colonization (Allen and Edje, 1990; Asfaw et al., 2009). Historical accounts show that common bean was introduced to Ethiopia in the 16th century by Portuguese traders and rapidly became an important component of the diet there (Assefa, 1985; Fisseha, 2015). Ever since the introduction of common bean into Ethiopia, farmers have developed farming practices adapted to local conditions by preservation and exploitation of useful alleles, which have resulted in a range of morphologically diverse landraces (Sperling, 2001). Moreover, recent efforts of the national bean-breeding program in Ethiopia have targeted improvement of on-farm common bean productivity and have benefited since the 1980's from continuous introduction of new germplasm from different parts of the world (Fisseha, 2015).

Today, Ethiopia is among the major bean producers in Sub-Saharan Africa (Wortmann et al., 1998). However, the national bean yield still lags behind the global average (Fisseha, 2015). This can be attributed partially to the low yielding capacity of cultivars under use (Assefa, 1990; Fisseha, 2015). To this end, it is essential to tap the potential of landrace genetic resources in order to introgress novel genes of adaptation, resistance to diseases and pests, and tolerance to abiotic stresses. According to Hornakova et al. (2003), landraces grown by small farmers are rich sources of valuable genes.

East Africa is a secondary center of diversity for common beans, due to the wide range of landraces there (Martin and Adams, 1987; Asfaw et al., 2009, Blair et al., 2010b). Understanding the patterns and levels of genetic diversity of bean landraces and cultivars can shed light on potential adaptation and direction and level of gene flow, and eventually help bean breeding and conservation in Ethiopia. Hence, this research project was undertaken with the principal goal of evaluating the genetic diversity within and between common bean landraces, to classify genotypes based on clustering and to understand the distribution of genetic diversity between and within ecological/geographical regions of Ethiopia.

MATERIALS AND METHODS

Plant materials

A total of 116 landrace accessions collected from a range of common bean production agro-ecologies in Ethiopia, four Ethiopian cultivars, three Kenyan cultivars, and two other cultivars, used as control genotypes for the Andean and Mesoamerican gene pools, respectively, were grown in August, 2012, in a greenhouse in the Biosciences Eastern and Central Africa (BecA-ILRI) hub in Nairobi, Kenya, for DNA extraction and analysis. The Ethiopian accessions were sampled from potential bean growing areas in the country (Supplementary Table 1 and Figure 1). The seeds of the control and commercial cultivars were acquired from the Ethiopian National Bean Research Project, based at Melkassa Agricultural Research Center, Adama, Ethiopia. The landrace accessions were provided by the Gene Bank of the Ethiopian Biodiversity Institute (EIB). A total of ten plants per each accession were planted in a single row in the screen house of BecA-ILRI hub, Nairobi, Kenya in August, 2012 for DNA extraction.

Genomic DNA extraction

For the molecular diversity assessment, total genomic DNA for each accession was isolated from a bulked leaf tissue sample of one week old plants from five randomly selected plants per accession using cetyltrimethylammonium bromide (CTAB) method (Doyle and Doyle, 1987) with some minor modifications, as described in supplementary part 1. However, 47 accessions did not produce enough genomic DNA, probably due to poor leaf sample qualities, which, in turn, imposed the need to repeat DNA extraction from the same, using the Zymoplant seed DNA extraction kit (descriptions on the protocol are presented in Supplementary Part 2).

Microsatellite amplification

Twenty-four (24) microsatellite markers from all the 11 linkage groups were selected based on their Polymorphic Information Content (PIC) values and dispersed map locations (Yu et al., 2000; Pedrosa-Harand et al., 2008; Kwak and Gepts, 2009). Out of the 24 SSR markers, 15 were genomic, and the remaining nine were non-genomic (genic) markers (Supplementary Table 2). Markers were PCR amplified with 6-FAM, NED, PET or VIC 5'-labeled forward primers and unlabeled reverse primers. The primers were run in multiplexes, based on their fluorescence dye and allele size using

*Corresponding author. E-mail: plgepts@ucdavis.edu . Tel: +1-530-752-7743.

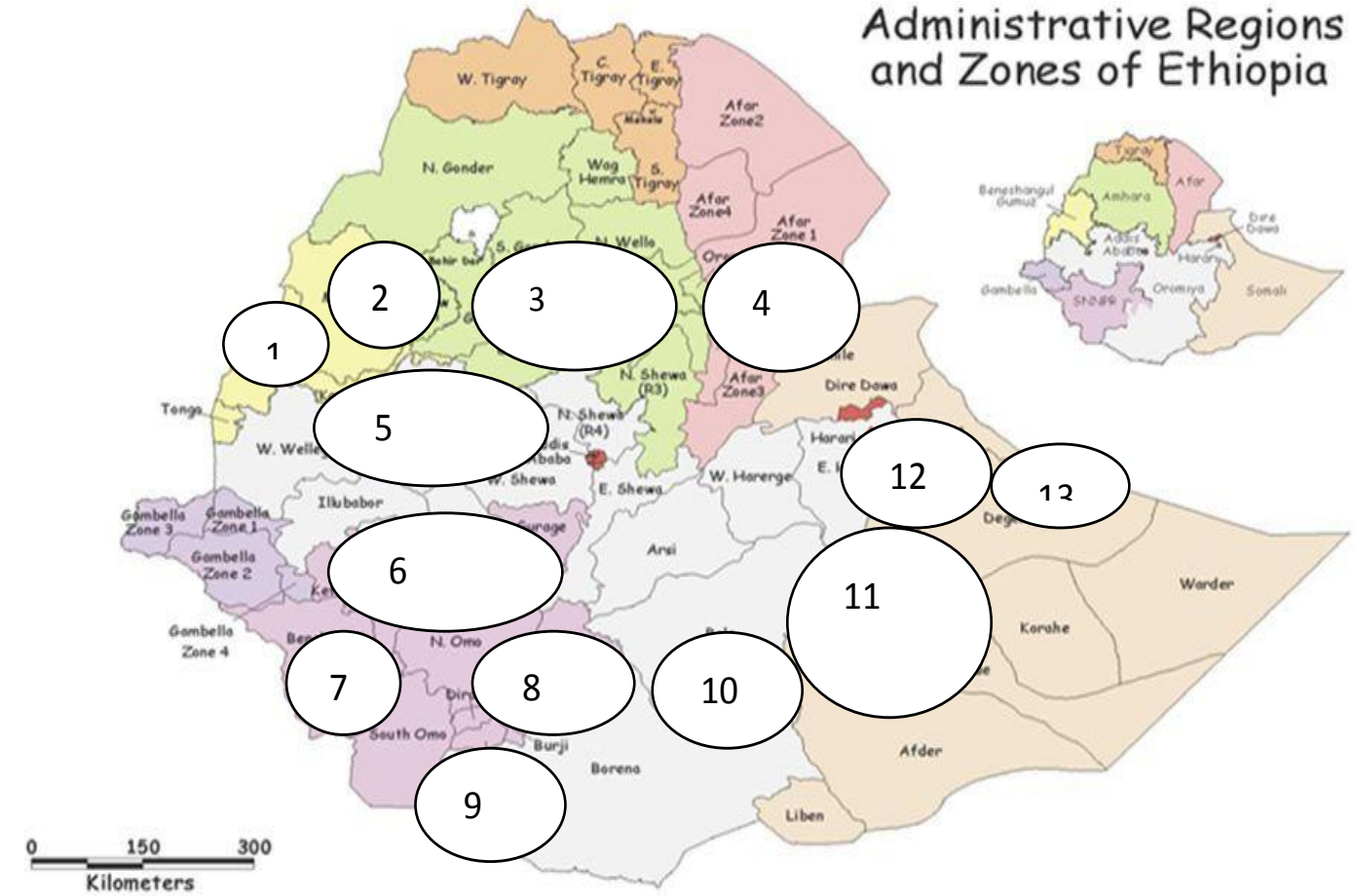


Figure 1. Map showing the collection sites. Key: 1 = Assosa; 2 = Metekel; 3 = Gojam; 4 = North Shewa & South Wello;; 5 = Wellega Gojam; 6 = Jimma and Illubabor; 7=Bench Maji; 8 = North Omo;; 9 = South Omo; 10 = Sidama and Others around; 11 = Bale & Arsi; 12 = East Hararghe;13 = West Hararghe) . The size of the bubbles does not correspond to number of genotypes sampled in each location.

BIONEER ACCUPOWER® Multiplex PCR Premix Kits (Supplementary part 3). Out of the 24 SSR markers, seven were dropped after preliminary evaluation, because they either produced no amplification (BM172 and BMd1) or were monomorphic (BM188, BM183, BMd16, PV-AG001, and PV-AT001). PCR products were run on an ABI PRISM 3730xl fragment analyzer (Applied Biosystems, Foster City, CA, USA) at the BecA-ILRI hub (Sequencing, genotyping, and Oligo unit, SegoLip), and allele sizes were determined by comparing with Genescan LIZ500 size standard using GeneMapper v. 3.7.3.7 software. The observed allele sizes were then adjusted for the discrete allele size using the AlleloBin software (http://test1.icrisat.org/gt-bt/download_allelobin.htm).

SSR genetic diversity analysis

Genalex 6.5b3 (Peakall and Smouse, 2012; http://biology.anu.edu.au/GenAlEx/) was used to calculate genetic diversity parameters, such as genetic distance, number of alleles (N_a); number of effective alleles (N_e); number of private alleles (N_{pa}); observed heterozygosity (H_o); expected heterozygosity (H_e); Shannon's information index (I); analysis of molecular variance (AMoVA); and principal coordinate analysis (PCoA). Genetic associations were determined using the neighbor-joining coefficient with Darwin V. 5.0 (http://darwin.cirad.fr/darwin). Genepop V.4 (Rousset, 2008) and Popgene32 (Yeh et al., 1999) programs were also used to determine genetic diversity, polymorphic loci, gene flow, levels of heterozygosity, fixation index, and F-values,. Finally, PowerMarker v. 3.25 (Liu and Muse, 2005) was used to estimate the number of alleles, polymorphic information content (PIC) values, genetic distance matrices, observed heterozygosity (H_o); and expected heterozygosity (H_e) for each marker across all genotypes and then across genotypes within and between gene pools.

Analysis of population structure

The software program STRUCTURE was run for K values ranging from 2 to 8. Each run was performed using the admixture model and 5,000 replicates for burn-in and 50,000 during the analysis (Pritchard et al., 2000). Evanno et al. (2005) test was performed after 10 simulations per K value. The repeated simulations were conducted for every subpopulation number from K = 2 to K=8 using 5,000 replicates for burn-in and 50,000 replicates according to previous suggestions (Rosenberg et al., 2002; Evanno et al., 2005; Ehrich, 2006). The Δ statistic showed that K = 5 was the optimal number of subpopulations in this analysis (Supplementary Figure 1). This ideal K value presented the highest peak for change in value from and to the previous and subsequent numbers of

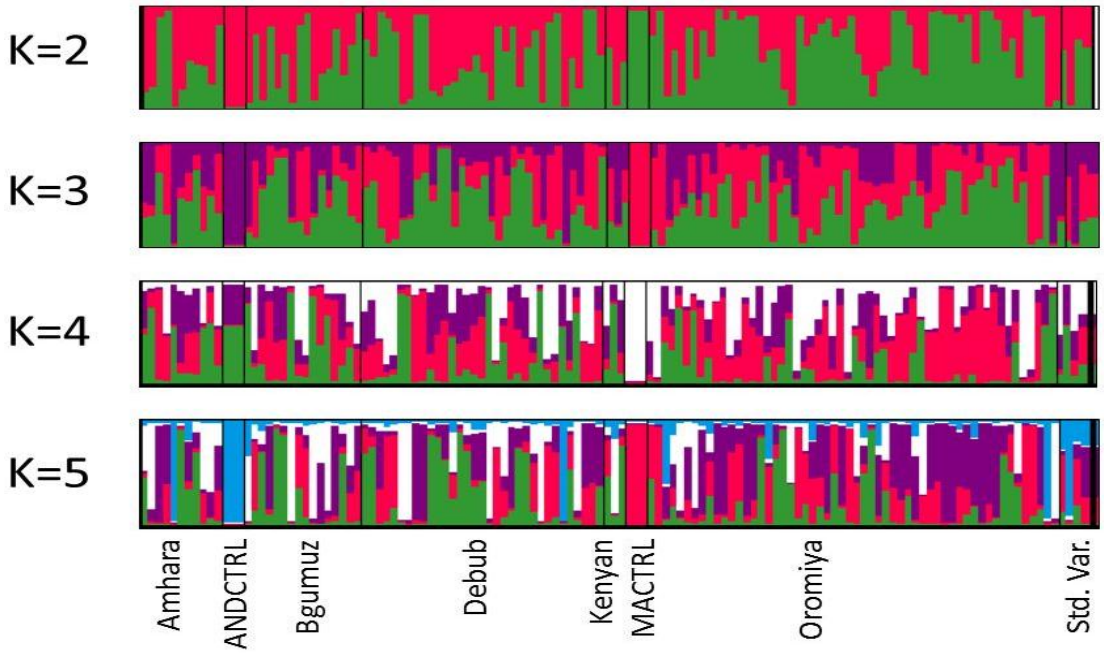


Figure 2. Population structure for 120 common bean accessions from different growing regions of Ethiopia and 3 Kenyan cultivars compared to Andean and Mesoamerican control genotypes at K = 2 to K = 5. Predetermined group names indicated below figure are: Amhara = Genotypes from Amhara Regional State; andectrl = Andean control genotypes; Bgumuz = Genotypes from Benishangul Regional State; Debub = Genotypes from Southern Nations and Nationalities Regional State; Kenyan = Kenyan accessions; MACTRL = Mesoamerican control genotypes; Oromiya = Genotypes from Oromiya Regional State; and Std. Var. = Standard Varieties.

subpopulations, respectively. This showed a gain in precision from subdividing the genotypes into five subpopulations versus any lower or higher numbers of subpopulations. The K=2 analysis was done with a particular interest of distinguishing between Andean and Mesoamerican accessions (Koenig and Gepts, 1989; Kwak and Gepts, 2009). To this end, five independent runs were performed with the admixture model and 5,000 replicates for burn-in and 50,000 replicates during analysis. The clustering in different runs was almost identical (similarity coefficient 0.9914). The run with the lowest likelihood value was selected among the five runs, and the accessions with more than 80% posterior assignment probability in the Mesoamerican cluster were assigned to the Mesoamerican gene pool (and vice versa for the Andean gene pool) (Supplementary Table 3). Lower posterior assignment probability values (that is, between 50 and 80%) may actually indicate hybrids or admixed accessions rather than ‘‘pure’’ accessions (Kwak and Gepts, 2009). Nonetheless, such accessions were included in the K=2 analysis, as they are important in future studies towards shedding light on the population structure of the common bean in Ethiopia, and as baseline information in breeding/improvement programs.

RESULTS

Population structure into the Andean and Mesoamerican gene pools in the common bean germplasm

The population subdivision (as determined by STRUCTURE) (Figure 2), the NJ tree (Figure 3), and the PCoA (Figure 4), showed significant Andean–Mesoamerican gene pool divergence as well as racial differentiation within gene pools. The accessions were assigned to the respective gene pools of origin, as per the methods explained in the “Materials and Methods” for K=2. Consequently, 78 accessions out of the total 125 fell into the Mesoamerican group, whereas the remaining 47 were classified into the Andean group. This classification was based on posterior assignment probabilities p>0.5. This split was generally maintained from K=2 to 3, but broke down for K = 4 and 5 (Figure 2; Supplementary Table 3). The analysis for K = 2 populations showed individual genotypes distributed between the two gene pools, which was congruent with the neighbor-joining and PCoA analyses, which clearly separated the Mesoamerican and Andean gene pools. At K=3, looking jointly into the bar-graphs produced and membership coefficient values, the Mesoamerican gene pool genotypes further separated into two sub-groups but no meaningful interpretation of population structure could be made, while the Andean gene pool genotypes did not show any separation. At K=4, the Mesoamerican accessions further subdivided into two groups with a mild level of admixture but no meaningful interpretation of population structure could be made. At K = 5, the Andean accessions further subdivided into three groups with

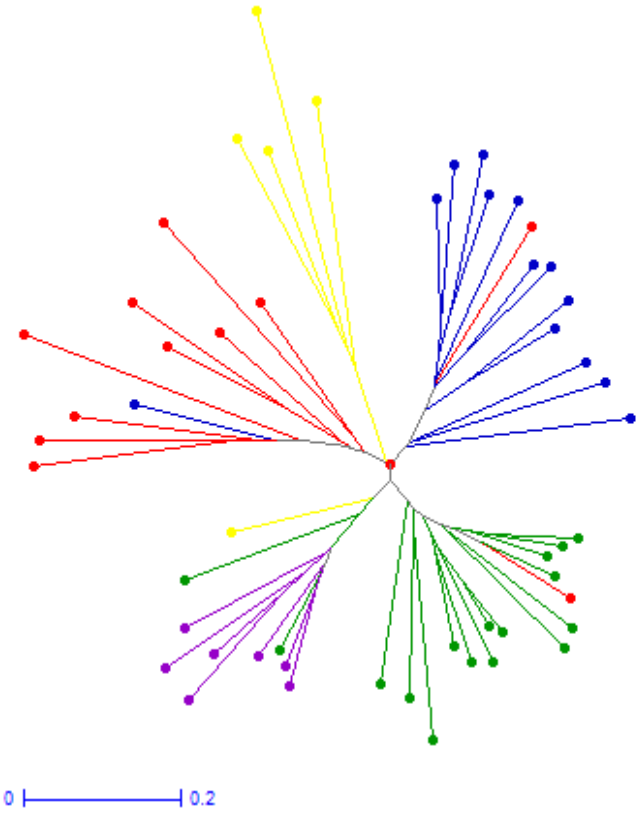


Figure 1. Neighbor-joining dendrogram depicting genetic relationship between common bean accessions from different bean growing populations in Ethiopia with respect to Andean and Mesoamerican control genotypes. Red: Andean Cluster1 (K4); Blue: Andean Cluster2 (K5); Yellow arrows: Andean Control (K1); Green: MA Cluster1 (K2); Purple arrows: MA Control (K3).

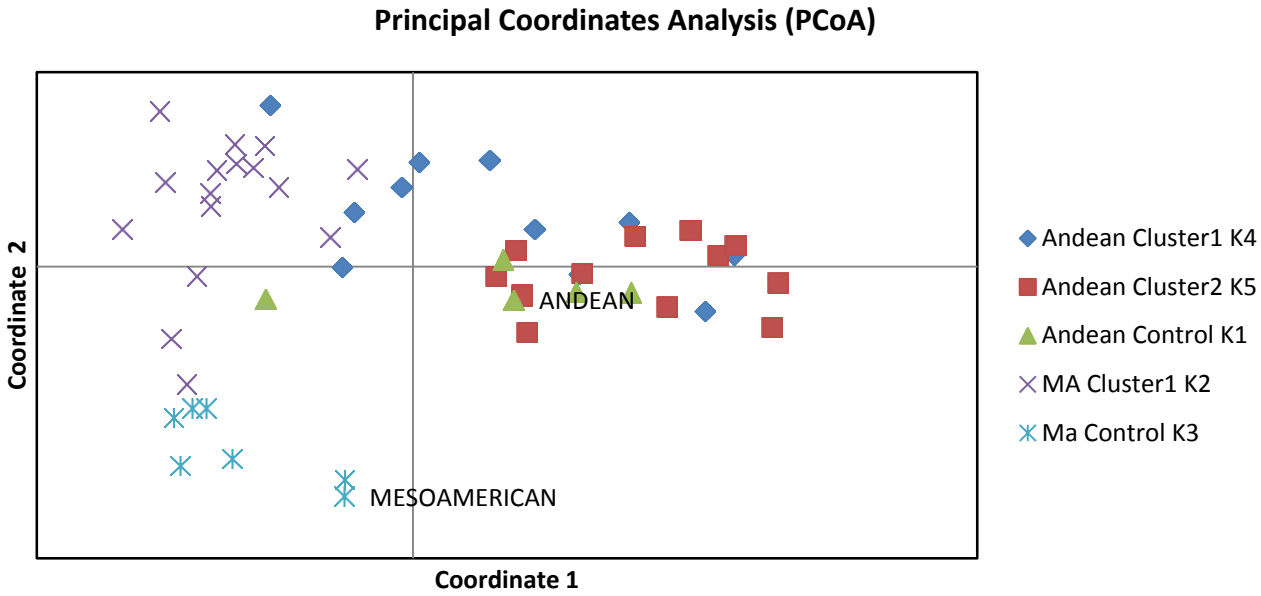


Figure 4. PCoA graph for the 53 accessions from different growing populations in Ethiopia.

Table 1. F_{st} values among five populations identified by STRUCTURE.

K	Andean Cluster 1 (K4)	Andean Cluster 2 (K5)	Andean Control (K1)	Mesoamerican Cluster 1 (K2)	Mesoamerican Control (K3)
5	0.239	0.356	0.547	0.135	0.264

Table 2. Proportion of non-hybrid accessions in K = 5 groups identified by STRUCTURE.

Groups	Total number of accessions	0.8 Cutoff	
		Number of accessions	% from total
Total	125	53	42.4
Mesoamerican	66	23	34.9
Mesoamerican Cluster1 (K2)	41	16	39.0
Mesoamerican control (K3)	25	7	28.0
Andean	59	30	50.9
Andean Cluster1 (K4)	27	11	40.7
Andean Cluster2 (K5)	26	14	53.9
Andean Control (K1)	6	5	83.3

some admixture level, whereas the Mesoamerican accessions did not subdivide further. In the following section, we describe in further details the five groups of K = 5.

Genetic diversity among accessions and cluster groups in STRUCTURE preset K=5

For K=5, the groups were identified as Andean Cluster 1 (K4); Andean Cluster 2 (K5); Andean control (K1); Mesoamerican Cluster 1 (K2) and Mesoamerican control (K3). On average, F_{st} values for Andean populations (K1, K4, and K5) were lower (0.213) compared to those of Mesoamerican populations (K2, and K3) (0.451) (Table 1). We also quantified population admixture for each accession (Figure 2; Supplementary Table 3). The Andean gene pool had a higher proportion of non-hybrid accessions than the Mesoamerican gene pool (51and 35% at the 0.8 cutoff, respectively; Table 2). The proportion of non-hybrid accessions in each K group ranged from 28% (Mesoamerican Controls K3) to 54% (Andean Cluster 2 K5) at the 0.8 cutoff values (Table 2).

The proportions of polymorphic loci were 100% in the Andean Cluster 1 (K4) genotypes; 94% in the Andean cluster 2 (K5), Andean control (K1), and the Mesoamerican cluster 1 (K2); 76% in the Mesoamerican control (K3) (Table 3). On average, the Andean groups had a higher number of alleles (N_a), number of effective alleles (N_e); Shannon Index (I), observed heterozygosity, expected heterozygosity, fixation index, percent of polymorphic loci; genetic distance; and number of private alleles. On the other hand, the Mesoamerican groups had higher hybridity rates than the Andean groups. The highest number of alleles, genetic distance (GD), observed heterozygosity (H_o), hybridity rate (t), and percent of polymorphic loci was recorded for the Andean cluster 1 (K5). The Andean control cluster had the highest Shannon index (I), fixation index (F), number of private alleles (N_{pa}); and number of effective alleles (N_e).

Analysis of Molecular Variance (AMoVA) among accessions and cluster groups in STRUCTURE preset K=5

The AMOVA results showed that 50% of allelic diversity was attributed to individuals within gene pool ($P<0.001$), 31% among individuals in the total population, and the remaining19% was attributed to the diversity among populations (Figure 5). A highly significant genetic differentiation among subpopulations (0.186, $P<0.01$) was observed. Some lower level of gene flow between different cluster of accessions was also reported (that is, 1.1), with higher values among accessions from different Andean gene pool clusters (that is, 1.6) values observed among different Mesoamerican clusters (i.e. 0.3) (Table 4). The average Nei's unbiased genetic distance was higher within each gene pool (0.8), but slightly lower between the Andean and Mesoamerican gene pools (0.7). Within gene pool, the Mesoamerican representatives presented lower genetic distances (0.7) than the Andean gene pool representatives (0.8) (Table 4).

Genetic associations among accessions

Genetic associations among accessions from different

Table 3. Mean SSR diversity for 17 microsatellite loci in five clusters of Ethiopian common bean genotypes.

Parameter	N	Na	Ne	I	H_e	H_o	GD	F	P (%)	N_{pa}	t
Andean Cluster 1 (K4)	11	4.118	2.598	1.032	0.545	0.325	0.304	0.380	100.00	0.154	0.449
Andean Cluster 2 (K5)	14	4.000	2.562	0.990	0.526	0.495	0.286	0.034	94.12	0.211	0.934
Andean Control (K1)	5	3.765	3.007	1.103	0.597	0.382	0.372	0.383	94.12	0.277	0.446
Mean Andean group	-	**3.961**	**2.722**	**1.042**	**0.556**	**0.401**	**0.321**	**0.266**	**96.1**	**0.214**	**0.610**
Mesoamerican Cluster 1 (K2)	16	3.647	2.077	0.819	0.445	0.363	0.229	0.209	94.12	0.174	0.654
Mesoamerican Control (K3)	7	2.412	1.606	0.524	0.298	0.272	0.356	0.067	76.47	0.106	0.874
Mean Mesoamerican group	-	**3.030**	**1.842**	**0.672**	**0.372**	**0.318**	**0.293**	**0.138**	**85.3**	**0.140**	**0.764**
General Mean	**-**	**3.588**	**2.370**	**0.894**	**0.464**	**0.367**		**0.223**	**91.76**	**0.185**	**0.687**

N number of genotypes, N_A number of different alleles, N_E effective number of alleles, N_{PA} number of private alleles, GD gene diversity according to Nei (1978), He expected heterozygosity, H_o observed heterozygosity, I Shannon's information index, F fixation index, t = (1-F)/(1 + F) out-crossing rate, P (%) percent polymorphic loci.

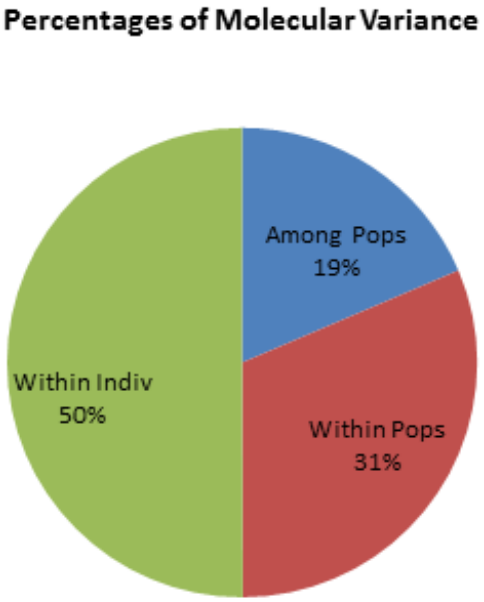


Figure 5. AMOVA pie-chart for the percentage of variation explained among individuals in a population; among populations; and within individuals in all the populations Pops=Populations; Indiv=Individuals.

populations in Ethiopia with respect to Andean and Mesoamerican control genotypes were identified using variation for fluorescent microsatellite markers (Figures 3 and 4). Both the PCoA and Neighbor-Joining graphs indicated the clustering of the bean genotypes into either of the Andean or Mesoamerican control genotypes. In the context of the geographical sample collection sites (Supplementary Table 1), genotypes from the same collection site were often in different clusters and likewise accessions from different collection sites often clustered together (Figure 6), indicating the possibility of gene flow by seeds between sites and regions within Ethiopia.

A principal coordinate analysis (PCoA) was conducted using five populations identified by STRUCTURE. The overall variation explained by the PCoA was 64% with dimensions 1, 2 and 3 explaining 26, 21and 19%, respectively. PCoA separated the bean genotypes into their corresponding centers of domestication (Andean/ Mesoamerican) along the first axis (Figure 4). Exceptions were Andean Cluster 4 genotypes in the second quadrant (four in number) and one genotype of the Andean Control cluster (quadrant III), which showed mixed cluster membership with the Mesoamerican Cluster. The mixed membership of Andean Cluster 1 (K4) was consistent between the STRUCTURE and neighbor-joining analysis results. However, the mixed clustering of Andean Control Cluster (K1) with the Mesoamerican groups was exhibited only in the PCoA and neighbor-joining tree.

Microsatellite diversity of Ethiopian common bean landrace accessions with respect to collections sites

Allelic patterns/diversity

A total of 149 alleles were identified, giving an average of 8.8 alleles per locus for the 17 microsatellites evaluated, of which 12 were genomic markers and 5 were genic (gene-based) (Supplementary Table 2). The range in allele number was 4 to 15, with the marker BM143 having the highest number of alleles, followed by GATS91, GATS54 and BM140, with 14, 13, and 13 alleles, respectively. All these markers were

Table 1. Pairwise population matrix of Nei unbiased genetic distance (below diagonal); Pair-wise N_m values (above diagonal); and F-values of the five cluster groups identified at Structure preset K=5.

Parameter	Andean Cluster1 K4	Andean Cluster2 K5	Andean Control K1	MA Cluster1 K2	MA Control K3	F-values
Andean Cluster1 K4	**0.000**	**1.8**	**1.5**	**1.5**	**0.8**	F_{st}=0.186
Andean Cluster2 K5	0.327	**0.000**	**1.4**	**1.1**	**0.7**	F_{is}=0.385
Andean Control K1	0.525	0.435	**0.000**	0.99	0.7	F_{it}=0.500
MA Cluster1 K2	0.182	0.259	0.384	**0.000**	0.75	
MA Control K3	0.488	0.451	0.517	0.321	**0.000**	

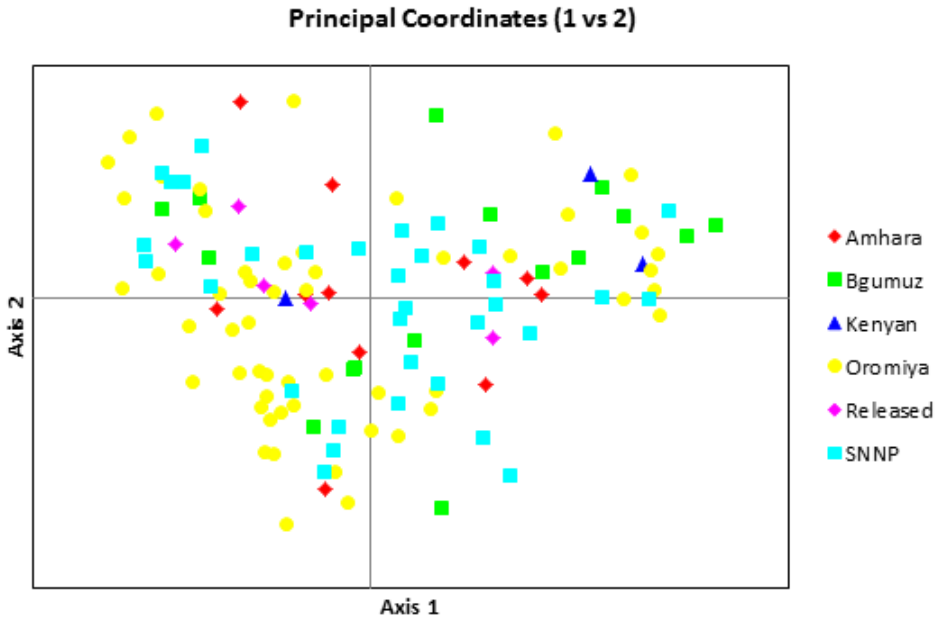


Figure 6. PCoA graph of the 125 common bean accessions from 6 populations.

genomic. The highest number of alleles found for a gene-based microsatellite was for BMd53 with 9 alleles, followed by BMd36 and BMd42 having 6 alleles each. The mean number of alleles for genomic microsatellites was 1.5 times more than that of genic microsatellites. The observed heterozygosity on average was 0.51 across all the 17 markers evaluated. The markers with the highest levels of observed heterozygosity were GATS91 (0.68) and BM143 (0.67), whereas the genic marker PV-CCTT001 had the lowest value, 0.01. With respect to the values recorded for expected heterozygosity (H_e), the SSR markers had an average of 0.564, with the highest being the genomic SSR, GATS91 (0.817), and the lowest for the genic SSR marker, PV-CCTT001 (0.011).

On the other hand, the allelic patterns across the studied populations are presented in Supplementary Figure 2. The figure also depicts the number of alleles, number of effective alleles, Shannon's diversity index, number of private alleles, and number of less common alleles in bars of different colors. The line above the bars indicates pattern of variation in expected heterozygosity among the different groups of accessions. The 'Amhara' and Southern Nations, Nationalities, and People (SNNP) had the highest expected heterozygosity. Nonetheless, the overall variation observed in accessions from different populations (collection sites) vis-à-vis the expected heterozygosity values was moderate. The calculated values for each of the aforementioned allelic measures are given in Table 5. The table corroborates the patterns depicted by Supplementary Figure 2. According to Table 5, accessions from Oromiya and SNNP had the highest average number of alleles (6.9and 6.4, respectively) (Table 5). On the other hand, accessions from 'Amhara' and the released varieties' group had the highest number of alleles with frequencies ≥5% (measurement taken to alleviate the sampling error associated with the sampling of race or distinct alleles, that is, with frequencies ≤ 5%), (N_a Freq. ≥5%), whereas accessions from 'Amhara' and

Table 5. Observed/effective number of alleles, genetic diversity, PIC, total number of alleles, average and expected heterozygosity and Shannon index of the 17 SSR markers used in the study.

Locus	Sample Size	n_a*	n_e*	I*	Average heterozygosity	F_{st}	Genetic Diversity	PIC	H_e
BM205	258	8	2.69	1.4	0.5428	0.25	0.618	0.591	0.590
AG-1	224	4	1.66	0.76	0.3716	0.33	0.409	0.376	0.438
GATS91	234	14	6.57	2.19	0.684	0.28	0.845	0.831	0.817
GATS54	254	13	3.06	1.52	0.5277	0.27	0.665	0.627	0.633
BMd42	242	6	2.89	1.38	0.567	0.28	0.638	0.608	0.607
PV-CCTT001	250	4	1.07	0.08	0.0126	0.11	0.025	0.025	0.011
BMd53	258	9	3.03	1.40	0.6134	0.13	0.664	0.605	0.659
BM156	250	11	2.98	1.40	0.56	0.284	0.655	0.595	0.602
BM187	216	11	2.86	1.35	0.651	0.421	0.648	0.584	0.574
BMd18	216	5	1.69	0.8	0.35	0.423	0.420	0.384	0.381
BMd36	220	6	3.06	1.30	0.53	0.34	0.669	0.621	0.610
BM151	218	8	3.44	1.43	0.56	0.336	0.705	0.656	0.596
BM140	232	13	4.1	1.72	0.64	0.31	0.752	0.717	0.677
BM141	242	7	2.56	1.18	0.48	0.31	0.603	0.536	0.589
BM143	242	15	4.17	1.91	0.67	0.223	0.757	0.736	0.765
BM165	226	6	3.51	1.39	0.54	0.366	0.717	0.671	0.559
BM139	244	9	1.88	1.08	0.41	0.251	0.457	0.437	0.479
Mean	**237**	**8.8**	**3.01**	**1.31**	**0.51**	**0.289**	**0.603**	**0.565**	**0.564**
St. Dev		**3.53**	**1.245**	**0.471**	**0.16**				**0.042**

SNNP had the highest number of effective alleles (N_e) (Table 5). From the perspective of this study, the 'Amhara' and SNNP regions may be the most important population of accessions owing to the higher number of alleles with frequencies ≥5% (excluding rare alleles) and number of effective alleles. Similar to the observations regarding the number of effective alleles (N_e), 'Amhara' and SNNP had the highest genetic diversity measures (Shannon's index=I) (Table 5). This may further strengthen the argument made above regarding the two populations, namely that the 'Amhara' and SNNP regions contain the highest level of bean diversity. Furthermore, accessions from 'Oromiya' and SNNP had the highest numbers of private alleles (0.82 and 0.59, respectively), and fewer common alleles with frequencies less than 50% (1.77and 1.60, respectively). This may imply that, upon further determination of what functional characters, if any, these private/less common alleles encode for or which genome region they mark, it may be possible to harness the potential of accessions in the population in future common bean breeding/improvement and genetic conservation endeavors in Ethiopia. Finally, the highest values for both expected and unbiased expected heterozygosity were recorded for accessions from 'Amhara', SNNP, and the released varieties' group.

Analysis of Molecular Variance (AMOVA) in the ecological/geographic population groups

Results of AMOVA are presented in Table 6 and Figure 7. Figure 7 shows that 58% of the total variation was attributed to genetic diversity prevalent within individuals from different populations, whereas 40% was due to variation among individuals within the same population. In contrast, a smaller portion (2%) of the total variation differentiated populations. In comparison, when the cluster groups identified at STRUCTURE preset K=5 (discussed below) were considered, AMOVA showed that 50% of allelic diversity was attributed to individuals within each of the groups ($P<0.001$); 31% among individuals in the total population; and the rest 19% was attributed to the diversity among populations. Moreover, highly-significant genetic differentiation among subpopulations (0.186, $P<0.01$) was observed.

In view of the F-statistics values (Table 7), the extent of genetic differentiation among the six populations in terms of allele frequencies measured was small (F_{st}=0.015*). Furthermore, the pair-wise Nm values among the six populations studied indicate that the highest values for putative gene flow were recorded for the following pairs of populations: BenishangulGumuz and SNNP (N_m=63); BenishangulGumuz and Kenya (N_m=55); and Oromiya and SNNP (N_m=31) (Supplementary Table 4).

Cluster analysis and Principal Coordinate Analysis (PCoA)

Cluster analysis with respect to populations (collection sites) was performed on the allelic frequency data using

Table 6. Important allelic values recorded in the landrace and control genotypes in six population groups.

Parameters	Populations					
	Amhara	**Bgumuz**	**Kenyan**	**Oromiya**	**Released**	**SNNP**
Na	5.059	4.647	2.529	**6.882**	3.824	**6.353**
Na Freq. ≥ 5%	**3.824**	3.765	2.529	3.294	**3.824**	3.647
Ne	**3.262**	2.791	2.195	2.612	2.815	**2.970**
I	**1.239**	1.114	0.757	1.179	1.065	**1.236**
No. Private Alleles	0.353	0.176	0.000	**0.824**	0.412	**0.588**
No. Less common Alleles (≤50%)	1.294	1.412	0.588	**1.765**	0.765	**1.588**
He	**0.619**	0.571	0.457	0.565	**0.574**	**0.597**
uHe	**0.652**	0.593	0.578	0.571	**0.627**	**0.607**

N_a (number of alleles), N_a Freq >= 5% (number of alleles with frequencies greater than or equal to 5%, N_e (number of effective alleles), I (Shannon's index), number of private alleles, number of less common alleles (with frequencies less than or equal to 25% and 50%, and He (expected heterozygosity). Populations refer to geographical administrative regions from which accessions had been collected.

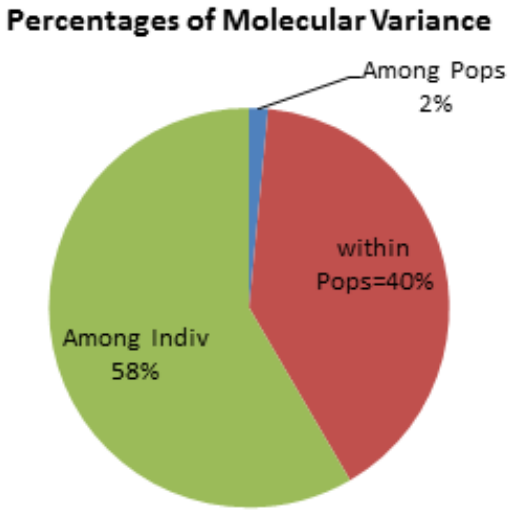


Figure 7. AMoVA variation pie chart for 125 common bean accessions from six populations in Ethiopia.

Table 7. Values of sum of squares; mean squares; and F-values among populations; among individuals in a population; and among individuals in all the populations.

Source	df	SS	MS	Est. Var.	%	F-Statistics	Value	P (random ≥ data)
Among Pops	5	53.742	10.75	0.085	2	F_{st}	0.015	0.020
Within pops	119	924.562	7.77	2.247	40	Fi_s	0.407	0.010
Among Indiv	125	409.500	3.28	3.276	58	F_{it}	0.416	0.010
Total	249	1387.804		5.608	100	N_m	16.282	

the Neighbor-joining method as implemented in the Darwin 5 and PowerMarker V3.25 software programs. Figure 8 shows the dendrogram clustering pattern for individual accessions in different populations (collection sites). As can be seen from the dendrogram, five different groups were identified. Furthermore, accessions from

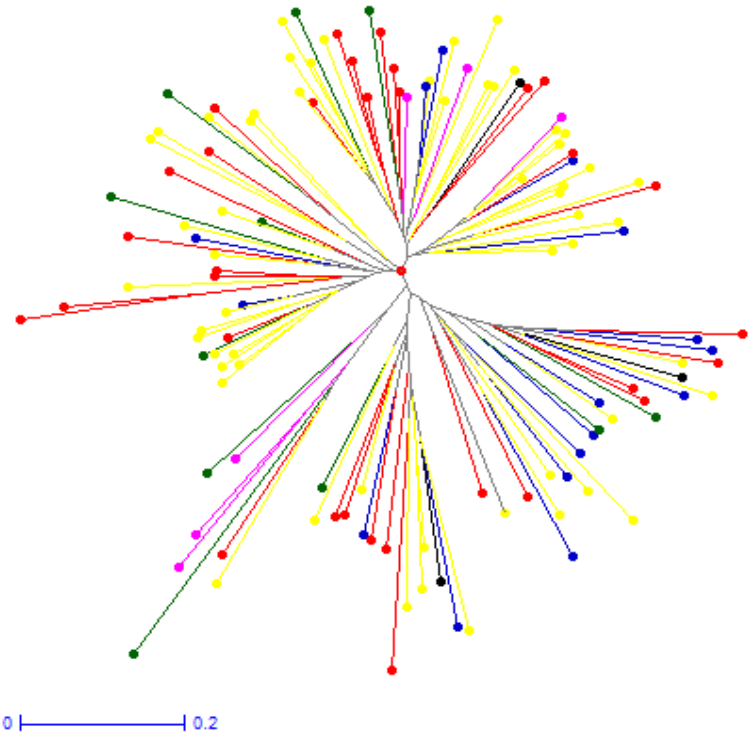


Figure 2. Neighbor-joining dendrogram of the 125 common bean accessions constructed by Darwin V5 software program. Green: Amhara; Blue: Benishangul Gumuz; Yellow: Oromiya; Red: Southern region; Purple: Released varieties; Orange: Kenya.

different populations (collection sites) clustered together. On the other hand, Supplementary Figure 3 shows the results of the cluster analysis done based on Nei's average unbiased genetic distance (Nei, 1983) among the accessions studied. Based on these results, four groups of populations were identified among the common bean landrace accessions from six different populations. Group 1 belonged to accessions from the Amhara region/population; the second group comprised accessions from the southern Ethiopia and Oromiya regions/populations. Another neighbor-joining dendrogram was constructed based on the shared-allele frequency genetic distances measured (Supplementary Figure 4). In comparison, this dendrogram identified five groups (compared to the four groups identified in the Nei's genetic distance NJ dendrogram), with Oromiya and Southern regions in the farthest end and Benishangul-Gumuz and Amhara, being group 3 and 4. Finally, yet importantly, the shared-allele frequency NJ dendrogram, similarly with the Nei's NJ dendrogram, clustered accessions from Kenya and the released varieties' group.

On the other hand, the first three axes of the PCoA accounted together for 65 % of the total variation, with 27, 21 and 17% explained by PC axis 1, 2, and 3, respectively. Results of the PCoA are displayed in Figure 6. It can be seen from this figure that accessions from different collection sites often clustered together.

DISCUSSION

The hierarchical classification scheme into subpopulations comprised of Andean and Mesoamerican genotypes obtained here was in agreement with that reported for common bean germplasm in various studies (Singh et al., 1991a; Gepts, 1998; Diaz and Blair, 2006; Blair et al., 2007, 2010a, 2011; Okii et al., 2014b). Moreover, the moderate to mostly large differentiation among subpopulations (F_{st} values) were higher than was reported in other studies (Asfaw et al., 2009; Okii et al., 2014b). On the other hand, the higher differentiation recorded in the present study among Mesoamerican subpopulations compared to their Andean counterparts was in contrast with the results of Asfaw et al. (2009) and Okii et al. (2014b). The separation of bean accessions into the two gene pools was also evidenced in the NJ and PCoA analyses. Five cluster groups were identified, which supports the findings reported in previous studies (Kwak and Gepts, 2009; Burle et al., 2011). Furthermore, the presence of moderate admixture level agrees with previous reports (Asfaw et al., 2009; Blair et al., 2010b; Okii et al., 2014b). Similarly, the concurrence of

STRUCTURE results with that of PCoA and NJ analyses was in agreement with that reported by Asfaw et al. (2009) and Okii et al. (2014b).

Five subpopulations with moderate admixture level and some switching of membership were observed in the present study. In line with this, Okii et al. (2014b) noted that the high level presence of admixture is indicative of the considerable mixing of common bean germplasm in planting and consumption and in hybridization in breeding. The considerable presence of admixture and switching of membership in some instances was supported by the PCoA and NJ analyses. These agree with some previous reports (Asfaw et al., 2009; Blair et al., 2010b; Burle et al., 2011; Okii et al., 2014b). In addition, the PCoA and NJ analyses in terms of geographic/ecological sampling of accessions indicated that accessions from different collection sites clustered together, which implied there was significant exchange of planting materials among farmers in different growing regions in the country. Moreover, the analysis of hybrid/ non-hybrid accessions indicated the Mesoamerican genotypes had higher instances of hybridity than the Andean counterparts. This observation supported previous results (Asfaw et al., 2009; Kwak and Gepts, 2009). On the other hand, from a population differentiation viewpoint, Andean genotypes were more differentiated than those from the Mesoamerican gene pool, which concurs with the findings reported previously for East African common bean germplasm (Asfaw et al., 2009; Okii et al., 2014b).

Accessions with Andean origin had higher allelic parameter values (Na, Ne, PIC, etc.) than Mesoamerican accessions. This contrasted with results from other related studies (Asfaw et al., 2009; Burle et al., 2011; Okii et al., 2014 a,b). Such differences may be attributed to differences in genetic samples and respective sampling methods employed. On the other hand, the hybridity values recorded in our study were much higher than those reported previously (Blair et al., 2010b; Okii et al., 2014a). This might be explained by the fact that most of the accessions (>90%) were acquired from the National Gene Bank, which, in turn, had collected these accessions from subsistence farmers with a culture of keeping mixed seeds for consumption and subsequent planting seasons. Moreover, the higher allelic values of genomic than genic markers were comparable to those reported in some previous studies (Asfaw et al., 2009; Blair et al., 2010b; Okii et al., 2014b). The high genetic diversity of the Ethiopian common bean landraces was also evident when considering their ecological or geographical distribution. Such presence of high diversity in terms of both gene pools and the existence of ecologically- or geographically differentiation populations can have potential applications prospective common bean breeding programs in Ethiopia.

The presence of higher levels of gene flow within each gene pool than that found between gene pools observed in our study agrees with the result of Asfaw et al. (2009). This may be explained, in part, due to the lack of flowering synchronization, which could reduce inter-gene pool gene flow. A larger proportion of the accessions (i.e., 58%) were introgressions, which contradicts the report of Asfaw et al. (2009) about the lower level of introgression with Ethiopian and Kenyan bean landraces/cultivars. This, in turn, negates the assumption of the aforementioned authors implying that the genetic divergence in Ethiopian bean germplasm could be mainly due to the original differences in introduced germplasm from the primary centers of origin. Rather, the presence of a higher number of introgressions may be partially explained by the fact that the accessions were gene bank collections from farmers' fields often characterized by a higher level of mixtures. The common practice of subsistence farmers in the country who cultivate for consumption and save segregant genotypes, resulting from any natural hybridization, as planting materials for subsequent generations, could result in such type of introgressions (Blair et al., 2010b; Worthington et al. 2012). A final noteworthy remark may be the fact that inter-gene pool introgressions are often endowed with useful combination of traits, including enhanced adaptation to environmental stresses, higher resistance to diseases and pests, and higher nutritional quality; hence, the introgressions identified in this study are of considerable importance in future bean breeding and conservation endeavors in Ethiopia. These merits of hybrids were evidenced in Islam et al. (2005) and Blair et al. (2010b), who reported that introgressions had higher mineral compositions than their respective non-hybrid parents. Consequently, it may be essential to tap into the useful genetic diversity found in such types of inter-gene pool introgressions, to be harnessed in further common bean breeding, improvement, and genetic conservation programs of beans in Ethiopia.

Conclusion

This study formulates new insights about the pattern and extent of genetic diversity and population structure of common bean landrace germplasm in Ethiopia. The results in the context of both the two gene pools of origin and ecological/geographic populations shed light on the presence of adequate genetic diversity organized into the Andean and Mesoamerican gene pools, and distributed across various ecological/geographic populations. This in turn should be strengthened by identifying the cluster groups identified by STRUCTURE via integrating molecular marker evaluations with phenotypic data.

Conflicts of interest

The authors have not declared any conflict of interest.

ACKNOWLEDGMENTS

We are grateful to the financial and material support provided by the Rural Capacity Building Project (RCBP)-Ministry of Agriculture, Ethiopia; the African Biosciences Challenge Fund (ABCF) at the BecA-ILRI hub, Nairobi, Kenya; and Addis Ababa University. Moreover, we are thankful to the Ethiopian Biodiversity Institute, Ethiopia, and the National Common Bean Research Project based at Melkassa Agricultural Research Center, Ethiopia for availing the plant materials for our study.

REFERENCES

Allen DJ, Edje OT (1990). Common bean in Africa Farming System. In: 1990 Progress in Improvement of Common Bean in Eastern and Southern Africa; Smithson JB. Ed. Bean Research, CIAT Africa Workshop Series. 5(20):32.

Asfaw A, Blair MW, Almekinders C (2009). Genetic diversity and population structure of common bean (*Phaseolus vulgaris* L.) landraces from the East African highlands. Theor. Appl. Genet. 120:1-12.

Blair MW, Díaz LM, Buendía HF, Duque MC (2010a). Genetic diversity, seed size associations and population structure of a core collection of common beans (*Phaseolus vulgaris* L.). Theor. Appl. Genet. 119:955-73

Blair MW, Díaz LM, Gill-Langarica HR, Rosales-Serna R, Mayek-Perez N, Acosta-Gallegos JA (2011). Genetic relatedness of Mexican common bean cultivars as revealed by microsatellite markers. Crop Sci. 51: 1–13.

Blair MW, Giraldo MC, Buendia HF, Tovar E, Duque MC, Beebe SE (2007). Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.). Theor. Appl. Genet. 113:100-109.

Blair MW, Gonzalez LF, Kimani PM, Butare L (2010b). Genetic diversity, inter-gene pool introgression and nutritional quality of common beans (*Phaseolus vulgaris* L.) from Central Africa. Theor. Appl. Genet. 121:237-248.

Broughton WJ, Hernandez G, Blair MW, Beebe SE, Gepts P, Vanderleyden J (2003). Beans (*Phaseolus spp*.) model food legumes. Plant Soil 252:55-128.

Burle L, Fonseca JR, Peloso M, Melo L, Temple SR, Gepts P (2011). Integrating phenotypic evaluations with a molecular diversity assessment of a Brazilian collection of common bean landraces. Crop Sci. 51:2668-2680.

Diaz LM, Blair MW (2006). Race structure within the Mesoamerican gene pool of common bean (*Phaseolus vulgaris* L.) as determined by microsatellite markers. Theor. App. Genet. 114:143-154.

Doyle JJ, Doyle JL (1987). A rapid DNA isolation procedure from small quantities of fresh leaf tissue. Phytochem. Bull. 19:11-15.

Ehrich D (2006). AFLPDAT: a collection of R functions for convenient handling of AFLP data. Mol. Ecol. Notes 6:603-604.

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

Freytag GF, Debouck DG (2002). Taxonomy, Distribution, and Ecology of the Genus *Phaseolus* (Leguminosae-Papilionoideae) in North America, Mexico and Central America. Bot. Res. Inst. Of Texas, Brit. Press, Ft. Worth, TX.

Gepts P (1988). A Middle American and an Andean Common Bean Gene Pool. In. Genetic resources of *Phaseolus* beans; Gepts, P. Ed. Kluwer, Dordrecht, The Netherlands. Pp. 375-390.

Gepts P (1998). Origin and evolution of common bean: Past events and recent trends. HortScience 33:1124-1130.

Gepts P, Aragão FJL, Barros E, Blair MW, Broughton RBW, Galasso I, Hernández G, Kami J, Lariguet P, McClean P, Melotto M, Miklas PN, Pauls P, Pedrosa-Harand A, Porch T, Sánchez F, Sparvoli F, Yu K (2008). Genomics of *Phaseolus* Beans, a Major Source of Dietary Protein and Micronutrients in the Tropics. In. Genomics of Tropical Crop Plants; Moore PH, Ming R. Eds. Springer Science+Business Media, New York, USA. 1: 132-162.

Gepts P, Bliss FA (1986). Phaseolin variability among wild and cultivated common beans (*Phaseolus vulgaris*) from Colombia. Econ. Bot. 40:469-478.

Habtu A (1990). Research Highlights on Haricot Bean in Ethiopia: An Assessment of the Progress, Priorities, and Strategies. In. Proceedings of National Workshop; 1-3 October, 1990, Addis Ababa, Ethiopia. pp. 50-53.

Hornakova O, Zavodna M, Zakova M, Kraic J, Debre F (2003). Diversity of common bean landraces collected in the Western and Eastern Carpatien. Czech. J. Genet. Plant Breed. 39(3):73-83.

Imru A (1985). Bean production in Ethiopia. In. Regional Workshop on Potential for Field Beans in West Asia and North Africa. Allepo, Syria, 10-13 October 1983, pp. 15-38.

Islam FM, Basford KE, Jara C, Redden RJ, Beebe S (2005). Seed compositional and disease resistance differences among gene pools in cultivated common bean. Genet. Res. Crop. Evol. 49:285-293.

Koenig R, Gepts P (1989). Allozyme diversity in wild *Phaseolus vulgaris*: further evidence for two major centers of diversity. Theor. Appl. Genet. 78:809-817.

Kwak M, Gepts P (2009). Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). Theor. Appl. Genet. 118:979-992.

Kwak M, Toro O, Debouck D, Gepts P (2012). Multiple origins of the determinate growth habit in domesticated common bean (*Phaseolus vulgaris* L.). Ann. Bot. 110:1573-1580.

Liu KJ, Muse SV (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128-2129.

Martin GB, Adams MW (1987). Landraces of *Phaseolus vulgaris* (Fabaceae) in Northern Malawi: Generation and maintenance of variability. Econ. Bot. 41:204-215.

Okii D, Tukamuhabwa P, Kami J, Namayanja A, Paparu P, Ugen M, Gepts P (2014b). The genetic diversity and population structure of common bean (*Phaseolus vulgaris* L.) germplasm in Uganda. Afr. J. Biotechnol. 29:2935-2949.

Okii D, Tukamuhabwa P, Namayanja A, Mukabaranga J, Paparu P, Gepts P (2014a). Morphological diversity of tropical common bean germplasm. Afr. Crop Sci. J. 22:59-67.

Pallottini L, Garcia E, Kami J, Barcaccia G, Gepts P (2004). The genetic anatomy of a patented yellow bean. Crop Sci. 44:968-977.

Peakall R, Smouse PE (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics 28:2537-2539.

Pedrosa-Harand A, Porch T, Gepts P (2008). Standard nomenclature for common bean chromosomes and linkage groups. Bean Improvement Cooperative, East Lansing, MI, USA. https://naldc.nal.usda.gov/download/IND44058845/PDF.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002). Genetic structure of human populations. Science 298:2381-2385.

Rousset F (2008). Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Mol. Ecol. Resour. 8:103-106.

Singh SP, Gepts P, Debouck DG (1991a). Races of common bean (*Phaseolus vulgaris* L., Fabaceae). Econ. Bot. 45:379-396.

Singh SP, Gutierrez JA, Molina A, Urea C Gepts P (1991b). Genetic diversity in cultivated common bean: II. Marker-based analysis of morphological and agronomic traits. Crop Sci. 31:23-29.

Singh SP, Nodari R, Gepts P (1991c). Genetic diversity in cultivated common bean. I. Allozymes. Crop Sci. 31: 19–23.

Sperling L (2001). The effect of the civil war on Rwanda's bean seed systems and unusual bean diversity. Biodivers. Conserv. 10:989-1009.

Worthington M, Soleri D, Aragón-Cuevas F, Gepts P (2012). Genetic composition and spatial distribution of farmer-managed bean plantings: an example from a village in Oaxaca, Mexico. Crop Sci. 52:1721-1735.

Wortmann CS, Kirkby RA, Eledu CA, Allen DJ (1998). Atlas of common

bean (*Phaseolus vulgaris* L.) production in Africa. CIAT Pan-African Bean Research Alliance, Vol. 133.
Yeh FC, Yang R, Boyle T (1999). POPGENE VERSION 1.31-Microsoft Window-based Freeware for Population Genetic Analysis. University of Alberta Edmonton, AB Canada T6G 2H1.
Yu K, Park S, Poysa V, Gepts P (2000). Integration of simple sequence repeat (SSR) markers into a molecular linkage map of common bean (*Phaseolus vulgaris* L.). J. Hered. 91:429-434.
Fisseha Z (2015). Genetic diversity and population structure of common bean (*Phaseolus vulgaris*, L.) germplasm from Ethiopia. PhD Thesis. Addis Ababa University, Addis Ababa, Ethiopia.

Supplementary Text 1: Genomic DNA Extraction.

For the molecular diversity assessment, total genomic DNA for each accession was isolated from a bulked leaf tissue sample of five randomly selected, one-week-old plants per accession using cetyltriethylammonium bromide (CTAB) method (Doyle and Doyle, 1990) with some minor modifications, as described in the following sections.

About 200 mg of fresh leaf tissue samples/leaf were placed in a 2 ml autoclaved and labeled Eppendorf tubes, covered by paraffin paper with a small slot at one side for air circulation, and freeze-dried for two days at-80^{0}C. Subsequently, a drop of polyvinyl polypyrrolidone (PVPP) was added to the Eppendorf tubes. Then, 500 µl of 1× CTAB was added to each tube to break open cells and soluble cellular contents. Next, the contents in each tube were mixed using a Vortex, and kept in a gently-shaking water bath for 1 hour at 65°C. After the samples were taken out of the water bath, they were centrifuged at 14,000 rpm for 30 min, using an Eppendorf centrifuge (5417R). Afterwards, the supernatant suspension was transferred into new Eppendorf tubes, and 250 µl of potassium acetate was added. A total of 400 µl of ice-cold isopropanol was added to the supernatant solution harvested, after centrifuging the samples at 14,000 rpm for 30 min. At this point, the samples were left at -20°C overnight. The following day, the samples were removed from the -20°C freezer; centrifuged at 14,000 rpm for 30 min at -4°C. The supernatant was then poured off and the pellet dried. In order to remove the remaining isopropanol drops, the tubes were placed upside down on a paper towel. The pellets were air-dried for 30 min at room temperature.

Subsequently, 200 µl of TE and 3 µl of RNAse were added to each tube, which were then left in a water bath at 37°C. Following this, chlorophyll and some denatured proteins were removed by dissolving in a 200 µl mixture of phenol, chloroform, and isoamyl alcohol at a ratio of 25:24:1, which was mixed with manual inversions from 5 to 10 times. Next, the samples were incubated at room temperature for 10 min. Subsequently, a fixed volume of supernatant (180 µl) was harvested from each tube into new sets of 1.5 ml Eppendorf tubes. Three hundred µl of ice-cold 100% ethanol plus 15 µl of sodium acetate (at pH 5.2) was added to each. Following incubation at -80°C for 5 min, centrifugation was performed at 14,000 rpm for 30 min; the supernatant was poured off and the inside of each tube was washed with 200 µl 70% ethanol, and another centrifuging was applied at 14,000 rpm for 30 min at -4°C. Following this, DNA pellets were air-dried for an hour, and re-suspended with 30 µl of low salt buffer. DNA quality and quantity were measured by gel electrophoresis (using 1% agarose gel for 1 hour using λ-DNA as a size marker) (Figure 11).

Supplementary Text 2.Pictorial display of the Zymoplant seed DNA extraction kit.

BIONEER
innovation . Value . Discovery

AccuPower® Gold Multplex PCR PreMix

Bioneer Corporation
8-11, Munpyeong-ro
Daedeok-gu, Daejeon 306-220
South Korea
Tel. (Korea) 1588-9788
(International) + 82-42-930-8777
Fax. + 82-42-930-8600
E-mail: sales@bioneer.com

Bioneer, Inc.
1000 Atlantic Avenue
Alameda, CA 94501 USA
Toll free: + 1-877-264-4300
Fax: + 1-510-865-0350
E-mail: orderusa@bioneerusa.com

Bioneer Trade(Shanghai) Co., Ltd
403 Room, Building 88, no 887
Zuchongzhi Road, Zhangjiang High Tech Park
PuDong New District, Shanghai 201203 China
Tel. + 86-21-5080-0969
Fax. + 86-21-5080-1620
E-mail: salescn@bioneer.com

Seoul Office
2nd fl. Sansu B/D, 496, Gangnamdaero
30-gil, Seocho-gu, Seoul 137-130
South Korea
Tel. + 82-2-598-1064
Fax. + 82-2-598-1096

Order
Domestic only: 1588-9788
E-mail: order@bioneer.co.kr
URL: www.bioneer.com

6. Maintain the reaction mixture at 4°C after amplification. The sample can be stored at -20°C until use.
7. Load 5 μl of the reaction mixture directly on agarose gel without adding a loading dye to analyze the PCR products.

X. Reaction Example

1. Reaction mixture

Component	Volume	Concentration
Template	1 μl	100 ng/μl
Primers	2 μl	1 pmole/μl each
D.W	17 μl	
Total	20 μl	

2. PCR cycling condition

Step	Temperature	Time	No. of Cycles
Pre-denaturation	95°C	5 min	1
Denaturation	95°C	30 sec	30
Annealing	57°C (65°C*)	30 sec	
Extension	72°C	1 min	
Final-Extension	72°C	5 min	1

Figure 1, Figure 2. (a)
Figure 2. (b)

XI. Experimental Data

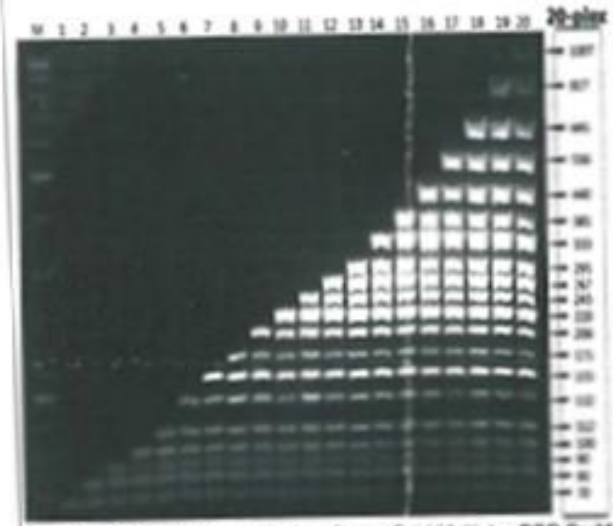
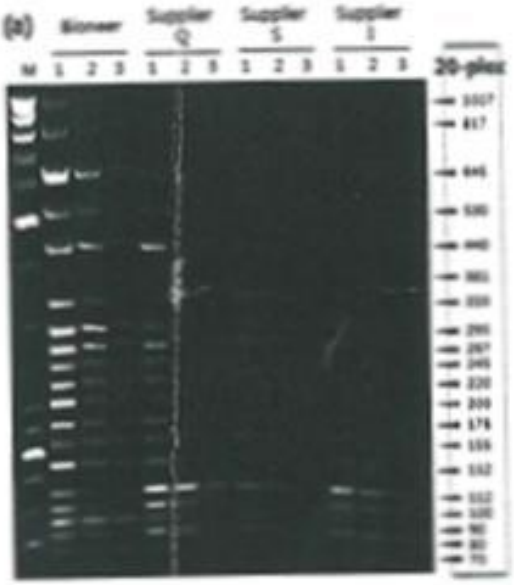


Figure 1. High specificity of *AccuPower* Gold Multiplex PCR PreMix. Each line from left to right represents the progressive number of primer sets up to 20 included in *AccuPower* Gold Multiplex PCR PreMix.



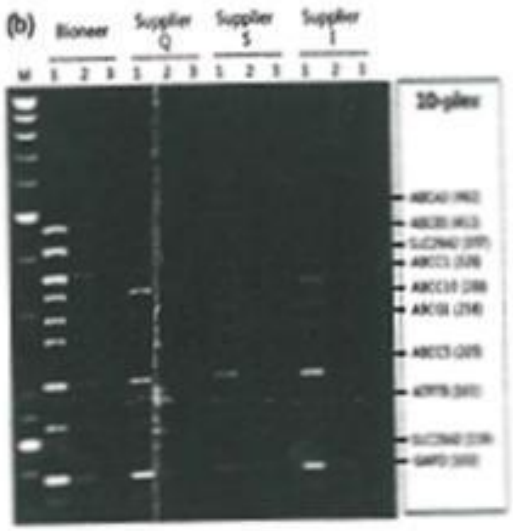


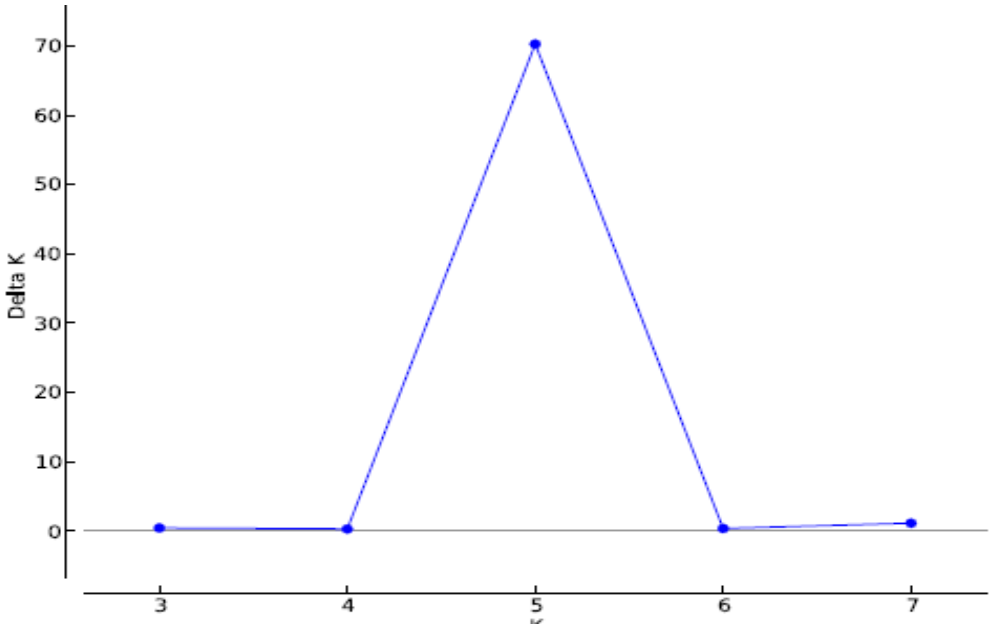
Figure 2. Comparison of amplification quality between *AccuPower* Gold Multiplex PCR PreMix and other supplier's Multiplex PCR kits.

Supplementary Table 1. ID number and names of collection site for the germplasm used in the study.

No.	Accession ID	Region/ Collection Area	No.	Accession ID	Region/ Collection Area	No.	Accession ID	Region/ Collection Area
1	211315	E. Hararghe	43	235692	Bench Maji	85	208703	Wellega
2	211317	E. Hararghe	44	235697	Bale &Arsi	86	211266	Gojam
3	211318	E. Hararghe	45	201066	Jimma	87	211267	Gojam
4	211349	Metekel	46	201293	W. Hararghe	88	211277	South Omo
5	241736	Sidama	47	201294	W. Hararghe	89	211279	South Omo
6	241756	Bench Maji	48	207933	Assosa	90	211290	Bench Maji
7	241757	Bench Maji	49	MWITEMA	Kenyan	91	211291	Bench Maji
8	244805	Sidama	50	E7	Kenyan	92	211299	W. Hararghe
9	211286	South Omo	51	WANJIRU	Kenyan	93	211300	W. Hararghe
10	211294	North Omo	52	211298	W. Hararghe	94	211305	W. Hararghe
11	211293	North Omo	53	211294	North Omo	95	211319	E. Hararghe
12	211301	W. Hararghe	54	211304	W. Hararghe	96	211320	E. Hararghe
13	211331	Somali	55	240190	Jimma	97	211322	E. Haraghe
14	211340	Wellega	56	211347	Metekel	98	211323	E. Haraghe
15	211341	Wellega	57	211349	Metekel	100	211327	W. Haraghe
16	211345	Metekel	58	211361	Metekel	101	211329	W. Haraghe
17	208647	Somali	59	211362	Metekel	102	211332	E. Haraghe
18	208705	Wellega	60	211379	Bale & Arsi	103	211295	E. Haraghe
19	211269	Gojam	61	211382	Shewa & Wello	104	211337	Wellega
20	211271	Wellega	62	211483	Bench Maji	105	211338	Wellega
21	211389	Shewa & Wello	63	219234	E. Hararghe	106	211339	Wellega
22	211394	South Omo	64	219235	E. Hararghe	107	211342	Wellega
23	211551	Shewa & Wello	65	AWASH 1	MA Control	108	211344	Metekel
24	211552	North Omo	66	211386	Shewa & Wello	109	211350	Metekel
25	MELKADIMA	Andean Control	67	211481	Bench Maji	110	211377	Bale &Arsi
26	CHERCHER	Standard Variety	68	241807	Gojam	111	211388	Wellega
27	GOBERASHA	Standard Variety	69	241737	Sidama	112	211546	North Omo
28	NASER	Standard Variety	70	241738	North Omo	113	211349	Metekel
29	237993	North Omo	71	241739	North Omo	114	212860	South Omo
30	240173	Jimma and Illubabor	72	241748	Sidama	115	216819	E. Hararghe
31	240512	Metekel	73	216819	E. Haraghe	116	216820	E. Hararghe
32	241730	Bale &Arsi	74	211278	South Omo	117	211337	Wellega
33	207934	Assosa	75	211292	Bench Maji	118	240522	Metekel
34	207938	Assosa	76	237078	Bale & Arsi	119	241733	Sidama
35	207949	Jimma and Illubabor	77	211348	Metekel	120	241750	North Omo
36	208638	W. Hararghe	78	211356	Metekel	121	241752	Bench Maji
37	212861	Bale &Arsi	79	211378	Bale & Arsi	122	241753	Bench Maji
38	212978	North Omo	80	211387	Shewa & Wello	123	241755	Bench Maji
39	213046	Bench Maji	81	208646	Somali	124	241814	Gojam
40	215719	Shewa & Wello	82	208695	Wellega	125	MEXICAN-142	Standard Variety
41	219233	West Hararghe	83	208698	Wellega			
42	230779	Bale &Arsi	84	208702	Wellega			

Supplementary Table 2. List of microsatellite (SSR) markers with forward/reverse nucleotide sequence, dye color, repeat motif, chromosomal location, and annealing temperature.

No.	SSR marker	Nucleotide Sequence	Dye Color	Repeat motif	Chromo-somal location	Annealing temperature
1	BM139-F BM139-R	TTAGCAATACCGCCATGAGAG ACTGTAGCTCAAACAGGGCAC	NED	(CT)25	2	55 °C
2	BM140-F BM140-R	TGCACAACACACATTTAGTGAC CCTACCAAGATTGATTTATGGG	PET	(GA)30	4	55°C
3	BM-141-F BM-141-R	TGAGGAGGAACAATGGTGGC CTCACAAACCACAACGCACC	VIC	(GA)29	9	55-58°C
4	BM143-F BM143-R	GGGAAATGAACAGAGGAAA ATGTTGGGAACTTTTAGTGTG	6- FAM	(GA)35	2	55-58°C
5	BM151-F BM151-R	CACAACAAGAAAAGACCTCCT TTATGTATTAGACCACATTACTTCC	NED	(TC)14	8	55°C
6	BM156-F BM156-R	CTTGTTCCACCTCCCATCATAGC TGCTTGCATCTCAGCCAGAATC	NED	(CT)32	10	55-58°C
7	BM165-F BM165-R	TCAAATCCCACACATGATCG TTCTTTCATTCATATTATTCCGTTCA	VIC	(TA)3(CA)9	8	52°C
8	BM172-F BM172-R	CTGTAGCTCAAACAGGGCACT GCAATACCGCCATGAGAGAT	6- FAM	(GA)23	2	50°C
9	BM183-F BM183-R	CTCAAATCTATTCACTGGTCAGC TCTTACAGCCTTGCAGACATC	NED	(TC)14	7	52°C
10	BM187-F BM187-R	TTTCTCCAACTCACTCCTTTCC TGTGTTTGTGTTCCGAATTATGA	PET	(CT)10 (CT)14	6	50-52°C
11	BM188-F BM188-R	TCGCCTTGAAACTTCTTGTATC CCCTTCCAGTTAAATCAGTCG	VIC	(CA)18 (TA)7	9	55°C
12	BM205-F BM205-R	CTAGACCAGGCAAAGCAAGC TGAGCTGGGATTTCATTTCTG	6-FAM	(GT)11	7	50°C
13	AG1-F AG1-R	CATGCAGAGGAAGCAGAGTG GAGCGTCGTCGTTTCGAT	NED	GA)8GGTA (GA)5GGGG	3	50°C
14	GATS54-F GATS54-R	GAACCTGCAAAGCAAAGAGC TCACTCTCCAACCAGATCGAA	PET	ACG (GA)5AACAGAGTC	10	56°C
15	GATS91-F GATS91-R	GAGTGCGGAAGCGAGTAGAG TCCGTGTTCCTCTGTCTGTG	VIC	(GA)8(AG)4 (GA)17	2	58°C
16	BMd53-F BMd53-R	TGCTGACCAAGGAAATTCAG GGAGGAGGCTTAAGCACAAA	6-FAM	(GTA)5	5	50°C
17	BMd36-F BMd36-R	CATAACATCGAAGCCTCACAGT ACGTGCGTACGAATACTCAGTC	NED	(TA)8	3	50°C
18	BMd42-F BMd42-R	TCATAGAAGATTTGTGGAAGCA TGAGACACGTACGAGGCTGTAT	PET	(AT)5	10	55°C
19	BMd1-F BMd1-R	CAAATCGCAACACCTCACACAA GTCGGAGCCATCATCTGTTT	VIC	(AT)9	3	54°C
20	BMd16-F BMd16-R	ATGACACCACTGGCCATACA GCACTGCGACATGAGAGAAA	6-FAM	(CATG)4	4	55°C
21	BMd18-F BMd18-R	AAAGTTGGACGCACTGTGATT TCGTGAGGTAGGAGTTTGGTG	NED	(TGAA)3	2	50-53°C
22	PV- AG001-F PV- AG001-R	CAATCCTCTCTCTCTCATTTCCAATC GACCTTGAAGTCGGTGTCGTTT	PET	(GA)1	11	50°C
23	PV- AT001-F PV- AT001-R	GGGAGGGTAGGGAAGCAGTG GCGAACCACGTTCATGAATGA	VIC	(TA)22	11	53°C
24	PV- CTOO1-F PV- CT001-R	CCAACCACATTCTTCCCTACGTC CGCAGGCAGTTATCTTTAGGAGTG	6- FAM	(CTT)3	4	56°C



Supplementary Figure 1. c Results of the Evano et al. (2005) test for ΔK between different sub-groupings of 123 common bean accessions/cultivars and two control genotypes based on analysis of allelic diversity at 17 microsatellite loci.

Supplementary Table 3. Membership coefficients and posterior probability values for K values from 1-5.

No.	Accession	Member Coefficient				Posterior Probability Values			
1	211269	0.09	0.132	0.082	0.696	(0.000,0.301)	(0.000,0.389)	(0.000,0.263)	(0.393,0.976)
2	241807	0.06	0.032	0.043	0.866	(0.000,0.206)	(0.000,0.106)	(0.000,0.145)	(0.666,0.997)
3	211266	0.853	0.017	0.066	0.064	(0.648,0.996)	(0.000,0.052)	(0.000,0.216)	(0.000,0.210)
4	211267	0.316	0.014	0.628	0.042	(0.069,0.553)	(0.000,0.046)	(0.406,0.838)	(0.000,0.145)
5	241814	0.012	0.964	0.011	0.013	(0.902,1.000)	(0.902,1.000)	(0.000,0.034)	(0.000,0.039)
6	211389	0.242	0.425	0.114	0.218	(0.000,0.606)	(0.004,0.833)	(0.000,0.331)	(0.000,0.605)
7	211551	0.565	0.301	0.107	0.027	(0.270,0.859)	(0.033,0.552)	(0.000,0.325)	(0.000,0.086)
8	215719	0.765	0.022	0.049	0.164	(0.545,0.968)	(0.000,0.068)	(0.000,0.166)	(0.001,0.341)
9	211382	0.017	0.016	0.4	0.568	(0.000,0.052)	(0.000,0.050)	(0.191,0.601)	(0.370,0.774)
10	211386	0.19	0.02	0.024	0.766	(0.000,0.422)	(0.000,0.061)	(0.000,0.080)	(0.528,0.985)
11	211387	0.252	0.016	0.498	0.234	(0.000,0.562)	(0.000,0.051)	(0.237,0.742)	(0.040,0.452)
12	ANDEAN Ctrl	0.012	0.964	0.008	0.015	(0.000,0.038)	(0.901,1.000)	(0.000,0.027)	(0.000,0.050)
13	207934	0.015	0.341	0.027	0.617	(0.000,0.047)	(0.174,0.519)	(0.000,0.089)	(0.430,0.796)
14	207938	0.035	0.053	0.745	0.167	(0.000,0.116)	(0.000,0.177)	(0.540,0.935)	(0.002,0.361)
15	207933	0.544	0.291	0.018	0.146	(0.237,0.814)	(0.001,0.603)	(0.000,0.057)	(0.000,0.387)
16	211349	0.876	0.025	0.07	0.029	(0.655,0.999)	(0.000,0.075)	(0.000,0.238)	(0.000,0.088)
17	211345	0.881	0.022	0.022	0.075	(0.673,0.999)	(0.000,0.066)	(0.000,0.070)	(0.000,0.249)
18	240512	0.879	0.057	0.035	0.029	(0.717,0.996)	(0.000,0.167)	(0.000,0.117)	(0.000,0.095)
19	211347	0.017	0.015	0.019	0.949	(0.000,0.056)	(0.000,0.046)	(0.000,0.061)	0.860,1.000)
20	211349	0.148	0.044	0.668	0.14	(0.000,0.384)	(0.000,0.150)	(0.444,0.894)	(0.000,0.374)
21	211361	0.063	0.038	0.826	0.073	(0.000,0.217)	(0.000,0.124)	(0.608,0.994)	0.000,0.245)
22	211362	0.044	0.032	0.014	0.91	(0.000,0.141)	(0.000,0.100)	(0.000,0.046)	(0.781,0.998)
23	211348	0.442	0.016	0.018	0.524	(0.057,0.736)	(0.000,0.050)	(0.000,0.055)	(0.237,0.881)
24	211356	0.063	0.012	0.046	0.879	(0.000,0.208)	(0.000,0.039)	(0.000,0.156)	(0.689,0.998)
25	211344	0.829	0.021	0.116	0.035	(0.624,0.991)	(0.000,0.067)	(0.000,0.288)	0.000,0.111)

Table 3. Contd.

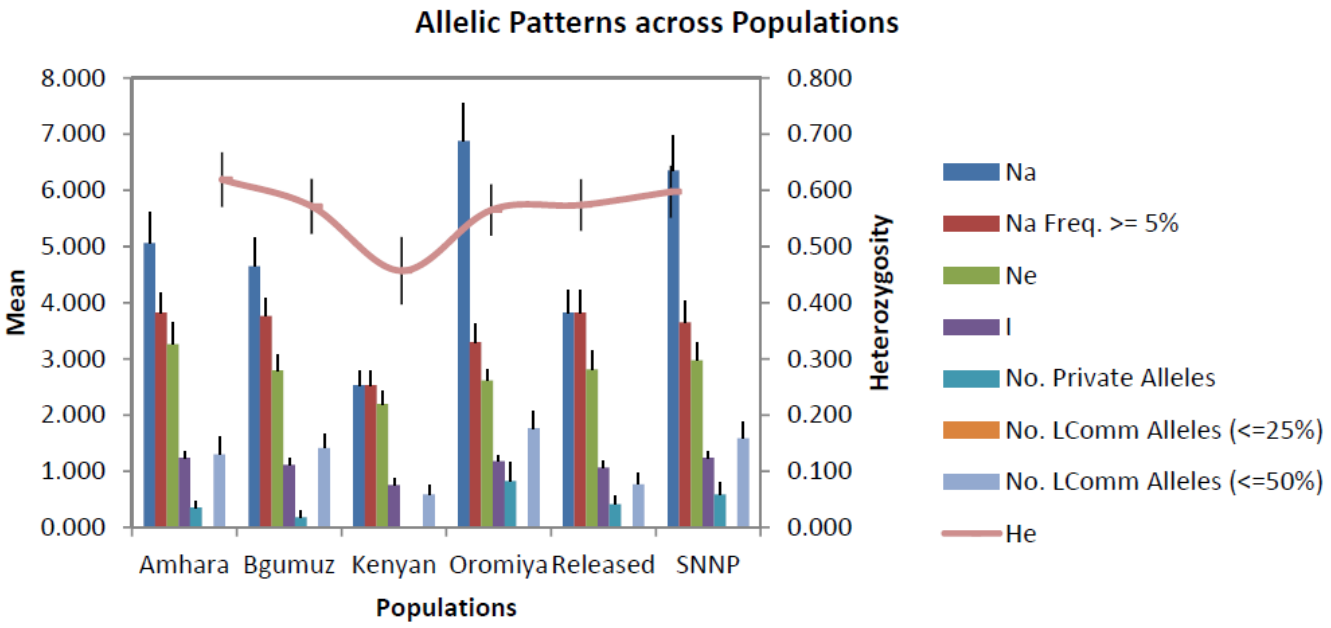
26	211350	0.448	0.074	0.135	0.343	(0.009,0.855)	(0.000,0.264)	(0.000,0.319)	(0.000,0.794)
27	211349	0.13	0.016	0.421	0.434	(0.000,0.406)	(0.000,0.050)	(0.164,0.646)	(0.234,0.634)
28	240522	0.214	0.028	0.736	0.022	(0.000,0.467)	(0.000,0.092)	(0.486,0.972)	(0.000,0.068)
29	241756	0.69	0.026	0.256	0.028	(0.406,0.967)	(0.000,0.084)	(0.001,0.524)	(0.000,0.089)
30	241757	0.697	0.041	0.194	0.069	(0.324,0.988)	(0.000,0.137)	(0.000,0.509)	(0.000,0.232)
31	213046	0.098	0.017	0.805	0.079	(0.000,0.333)	(0.000,0.054)	(0.531,0.996)	(0.000,0.266)
32	235692	0.027	0.069	0.871	0.033	(0.000,0.086)	(0.000,0.178)	(0.733,0.986)	(0.000,0.109)
33	211483	0.021	0.072	0.809	0.099	(0.000,0.067)	(0.000,0.216)	(0.636,0.965)	(0.000,0.273)
34	211481	0.023	0.018	0.012	0.947	(0.000,0.075)	(0.000,0.058)	(0.000,0.038)	(0.853,1.000)
35	211292	0.028	0.022	0.015	0.934	(0.000,0.093)	(0.000,0.073)	(0.000,0.047)	(0.822,1.000)
36	211290	0.814	0.014	0.148	0.024	(0.603,0.990)	(0.000,0.044)	(0.000,0.348	(0.000,0.074)
37	211291	0.876	0.013	0.061	0.05	(0.684,0.998)	(0.000,0.042)	(0.000,0.206)	(0.000,0.166)
38	241752	0.33	0.304	0.13	0.237	(0.000,0.802)	(0.000,0.739)	(0.000,0.372)	(0.000,0.751)
39	241753	0.923	0.032	0.012	0.032	(0.796,0.999)	(0.000,0.109)	(0.000,0.038)	(0.000,0.105)
40	241755	0.955	0.014	0.011	0.02	(0.876,1.000)	(0.000,0.046)	(0.000,0.033)	(0.000,0.063)
41	211294	0.044	0.189	0.346	0.422	(0.000,0.144)	(0.000,0.446)	(0.108,0.577	(0.130,0.719)
42	211293	0.698	0.098	0.047	0.157	(0.413,0.966)	(0.000,0.336)	(0.000,0.160)	(0.000,0.410)
43	211552	0.498	0.095	0.154	0.253	(0.223,0.751)	(0.000,0.302)	(0.000,0.392)	(0.000,0.550)
44	237993	0.655	0.148	0.129	0.068	(0.313,0.942)	(0.000,0.356)	(0.000,0.399)	(0.000,0.222)
45	212978	: 0.797	0.139	0.028	0.036	(0.514,0.995)	(0.000,0.409)	(0.000,0.093)	(0.000,0.119)
46	211294	0.027	0.018	0.026	0.929	0.000,0.090)	(0.000,0.057)	(0.000,0.086)	(0.810,0.999)
47	241738	0.013	0.011	0.742	0.234	(0.000,0.040)	(0.000,0.034)	(0.582,0.885)	(0.095,0.391)
48	241739	0.182	0.014	0.292	0.513	(0.000,0.458)	(0.000,0.044)	(0.002,0.554)	(0.299,0.733)
49	211546	0.885	0.011	0.043	0.061	(0.720,0.996)	(0.000,0.033)	(0.000,0.148)	(0.000,0.175)
50	241750	: 0.760	0.022	0.019	0.2	(0.506,0.988)	(0.000,0.069)	(0.000,0.061)	(0.000,0.449)
51	241736	0.815	0.021	0.136	0.028	(0.574,0.994)	(0.000,0.065)	(0.000,0.366)	(0.000,0.091)
52	244805	0.082	0.03	0.609	0.279	(0.000,0.283)	(0.000,0.099)	(0.196,0.957)	(0.000,0.617)
53	241737	0.016	0.026	0.014	0.943	(0.000,0.052)	(0.000,0.087)	(0.000,0.044)	(0.847,1.000)
54	241748	0.036	0.049	0.762	0.153	(0.000,0.119)	(0.000,0.131)	(0.601,0.909)	(0.012,0.308)
55	241733	0.106	0.045	0.835	0.014	(0.000,0.306)	(0.000,0.146)	(0.641,0.988)	(0.000,0.043)
56	211286	0.014	0.959	0.013	0.014	(0.000,0.044)	(0.890,1.000)	(0.000,0.042)	(0.000,0.044)
57	211394	0.165	0.208	0.43	0.197	(0.000,0.414)	(0.000,0.523)	(0.225,0.636)	(0.000,0.490)
58	211278	0.084	0.012	0.032	0.872	(0.000,0.281)	(0.000,0.038)	(0.000,0.106)	(0.657,0.999)
59	211277	0.124	0.011	0.842	0.023	(0.000,0.386)	(0.000,0.035)	(0.580,0.998)	(0.000,0.074)
60	211279	0.065	0.011	0.908	0.016	(0.000,0.223)	(0.000,0.035)	(0.734,0.999)	(0.000,0.050)
61	212860	0.923	0.032	0.025	0.02	(0.802,0.999)	(0.000,0.103)	(0.000,0.082)	(0.000,0.063)
62	MWITEMA	0.107	0.159	0.565	0.169	(0.000,0.349)	(0.000,0.416)	(0.296,0.811)	(0.000,0.439)
63	E7	0.026	0.253	0.014	0.707	(0.000,0.083)	(0.003,0.501)	(0.000,0.046)	(0.456,0.962)
64	WANJIRU	0.362	0.194	0.269	0.176	(0.009,0.701)	(0.000,0.500)	(0.000,0.598)	(0.000,0.508)
65	MA Ctrl	0.01	0.011	0.968	0.01	(0.000,0.032)	(0.000,0.036)	(0.914,1.000)	(0.000,0.031)
66	241730	0.104	0.149	0.691	0.055	(0.000,0.310)	(0.000,0.352)	(0.487,0.887)	(0.000,0.192)
67	212861	0.014	0.021	0.956	0.01	(0.000,0.042)	(0.000,0.068)	(0.882,1.000)	(0.000,0.031)
68	230779	0.271	0.677	0.021	0.031	(0.026,0.489)	(0.468,0.895)	(0.000,0.068)	(0.000,0.101)
69	235697	0.204	0.189	0.262	0.346	(0.000,0.521)	(0.001,0.399)	(0.000,0.547)	(0.006,0.692)
70	211379	0.053	0.014	0.018	0.916	(0.000,0.180)	(0.000,0.042)	(0.000,0.056)	(0.770,0.999)
71	237078	0.596	0.014	0.121	0.269	(0.337,0.828)	(0.000,0.043)	(0.000,0.368)	(0.093,0.468)
72	211378	0.067	0.016	0.064	0.854	(0.000,0.227)	(0.000,0.050)	(0.000,0.210)	(0.646,0.997)
73	211377	0.581	0.061	0.286	0.073	(0.259,0.880)	(0.000,0.191)	(0.029,0.558)	(0.000,0.234)
74	216819	0.211	0.183	0.396	0.211	(0.000,0.483)	(0.000,0.416)	(0.194,0.599)	(0.024,0.420)
75	211319	0.79	0.009	0.181	0.019	(0.574,0.984)	(0.000,0.027)	(0.000,0.397)	(0.000,0.063)
76	211320	0.877	0.059	0.049	0.015	(0.706,0.996)	(0.000,0.173)	(0.000,0.168)	(0.000,0.047)
77	211322	0.15	0.018	0.808	0.025	(0.000,0.340)	(0.000,0.059)	(0.624,0.969)	(0.000,0.082)

Table 3. Contd.

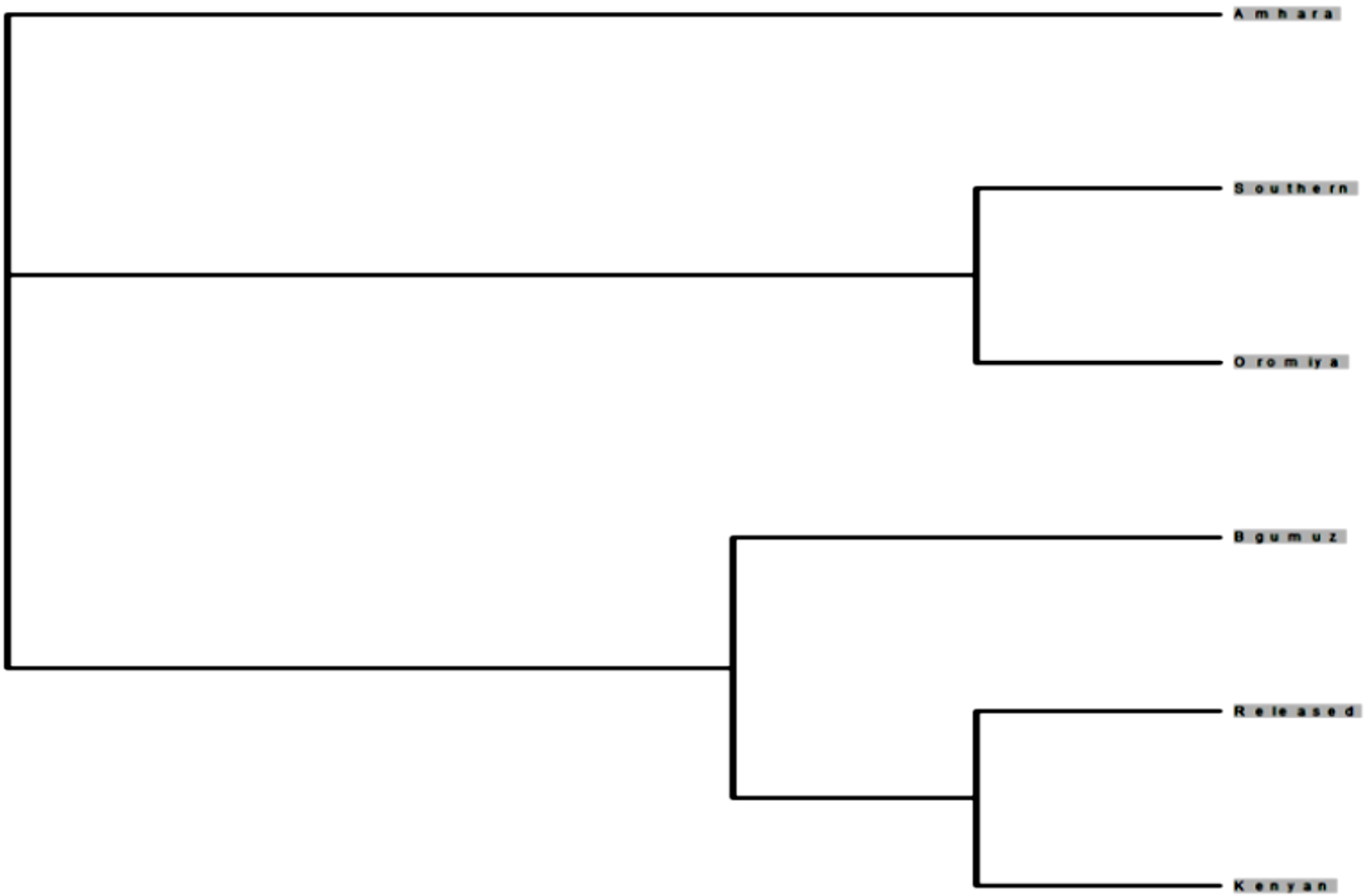
78	211323	0.342	0.054	0.584	0.02	(0.134,0.565)	(0.000,0.159)	(0.368,0.791)	(0.000,0.061)
79	211332	0.805	0.013	0.169	0.013	(0.594,0.980)	(0.000,0.042)	(0.002,0.372)	(0.000,0.041)
80	211295	0.49	0.048	0.44	0.021	(0.218,0.760)	(0.000,0.161)	(0.188,0.700)	(0.000,0.069)
81	211315	0.937	0.024	0.02	0.019	(0.828,1.000)	(0.000,0.079)	(0.000,0.061)	(0.000,0.058)
82	211317	0.032	0.354	0.568	0.046	(0.000,0.104)	(0.188,0.526)	(0.386,0.740)	(0.000,0.160)
83	211318	0.536	0.206	0.203	0.056	(0.206,0.880)	(0.000,0.447)	(0.000,0.441)	(0.000,0.196)
84	211331	0.373	0.145	0.041	0.441	(0.001,0.756)	(0.000,0.500)	(0.000,0.136)	(0.008,0.900)
85	208647	0.116	0.118	0.013	0.753	(0.000,0.333)	(0.000,0.346)	(0.000,0.039)	(0.528,0.976)
86	219234	0.016	0.016	0.909	0.059	(0.000,0.053)	(0.000,0.052)	(0.771,0.999)	(0.000,0.182)
87	219235	0.06	0.028	0.647	0.265	(0.000,0.204)	(0.000,0.086)	(0.323,0.947)	(0.000,0.584)
88	208646	0.102	0.23	0.644	0.024	(0.000,0.303)	(0.093,0.385)	(0.440,0.826)	(0.000,0.078)
89	216819	0.566	0.017	0.397	0.02	(0.340,0.787)	(0.000,0.055)	(0.182,0.618)	(0.000,0.062)
90	216820	0.703	0.028	0.245	0.023	(0.456,0.938)	(0.000,0.093)	(0.025,0.477)	(0.000,0.071)
91	240173	0.219	0.053	0.643	0.085	(0.000,0.486)	(0.000,0.188)	(0.399,0.922)	(0.000,0.273)
92	207949	0.923	0.021	0.035	0.021	(0.793,0.999)	(0.000,0.069)	(0.000,0.117)	(0.000,0.064)
93	201066	0.021	0.062	0.9	0.016	(0.000,0.069)	(0.000,0.164)	(0.778,0.995)	(0.000,0.053)
94	240190	0.057	0.017	0.423	0.504	(0.000,0.183)	0.000,0.052)	(0.228,0.620)	(0.311,0.699)
95	211340	0.62	0.231	0.064	0.085	(0.303,0.913)	(0.000,0.479)	(0.000,0.220)	(0.000,0.274)
96	211341	0.274	0.517	0.141	0.068	(0.000,0.806)	(0.029,0.900)	(0.000,0.407)	(0.000,0.215)
97	208705	0.373	0.229	0.122	0.277	(0.000,0.733)	(0.001,0.494)	(0.000,0.345)	(0.000,0.747)
98	211271	0.077	0.035	0.357	0.532	(0.000,0.250)	(0.000,0.116)	(0.176,0.543)	(0.333,0.727)
99	208695	0.282	0.012	0.684	0.022	0.000,0.591)	(0.000,0.036)	(0.380,0.977)	(0.000,0.073)
100	208698	0.607	0.01	0.324	0.06	(0.256,0.966)	(0.000,0.031)	(0.000,0.675)	(0.000,0.209)
101	208702	0.766	0.021	0.195	0.018	(0.545,0.973)	(0.000,0.069)	(0.000,0.415)	(0.000,0.056)
102	208703	0.496	0.185	0.054	0.265	(0.060,0.812)	(0.001,0.389)	(0.000,0.185)	(0.000,0.632)
105	211337	0.019	0.024	0.281	0.677	(0.000,0.061)	(0.000,0.076)	(0.067,0.492)	(0.462,0.892)
106	211338	0.378	0.026	0.567	0.03	(0.082,0.661)	(0.000,0.086)	(0.296,0.837)	(0.000,0.093)
107	211339	0.493	0.012	0.481	0.014	(0.279,0.711)	(0.000,0.037)	(0.265,0.694)	(0.000,0.044)
108	211342	0.93	0.013	0.036	0.021	(0.809,1.000)	(0.000,0.039)	(0.000,0.121)	(0.000,0.065)
109	211388	0.915	0.038	0.028	0.019	(0.788,0.998)	(0.000,0.118)	(0.000,0.093)	(0.000,0.059)
110	211337	0.383	0.086	0.482	0.05	(0.082,0.654)	(0.000,0.207)	(0.254,0.706)	(0.000,0.175)
111	211298	0.775	0.014	0.196	0.015	(0.548,0.975)	(0.000,0.044)	(0.003,0.418)	(0.000,0.047)
112	211299	0.787	0.171	0.021	0.02	(0.615,0.945)	(0.025,0.329)	(0.000,0.071)	(0.000,0.066)
113	211300	0.962	0.009	0.017	0.012	(0.895,1.000)	(0.000,0.028)	0.000,0.053)	(0.000,0.038)
114	211305	0.727	0.011	0.248	0.015	(0.506,0.943)	(0.000,0.033)	(0.033,0.465)	(0.000,0.047)
115	211325	0.931	0.02	0.016	0.032	(0.831,0.999)	(0.000,0.066)	(0.000,0.051)	(0.000,0.106)
116	211301	0.245	0.03	0.058	0.667	(0.000,0.688)	(0.000,0.099)	(0.000,0.202)	(0.235,0.985)
117	208638	0.018	0.012	0.96	0.01	(0.000,0.057)	(0.000,0.037)	(0.892,1.000)	(0.000,0.030)
118	219233	0.074	0.025	0.883	0.018	(0.000,0.213)	(0.000,0.085)	(0.732,0.995)	(0.000,0.057)
119	201293	0.875	0.019	0.083	0.023	(0.661,0.998)	(0.000,0.060)	(0.000,0.282)	(0.000,0.075)
120	201294	0.012	0.945	0.016	0.026	(0.000,0.037)	(0.854,1.000)	(0.000,0.052)	(0.000,0.087)
121	211304	0.027	0.018	0.026	0.929	(0.000,0.087)	(0.000,0.058)	(0.000,0.087)	(0.810,0.999)
122	CHERCHER	0.039	0.366	0.58	0.015	(0.000,0.128)	(0.188,0.547)	(0.388,0.766)	(0.000,0.046)
123	GOBERASHA	0.014	0.925	0.025	0.036	(0.000,0.045)	(0.805,0.999)	(0.000,0.083)	(0.000,0.125)
124	NASER	0.242	0.272	0.469	0.018	(0.008,0.473)	(0.099,0.459)	(0.272,0.658)	(0.000,0.057)
125	Mexico-142	0.418	0.237	0.311	0.035	(0.185,0.648)	(0.053,0.428)	(0.128,0.501)	(0.000,0.112)

Supplementary Table 4. Pair-wise number of migrants (Nm) based on Fst values.

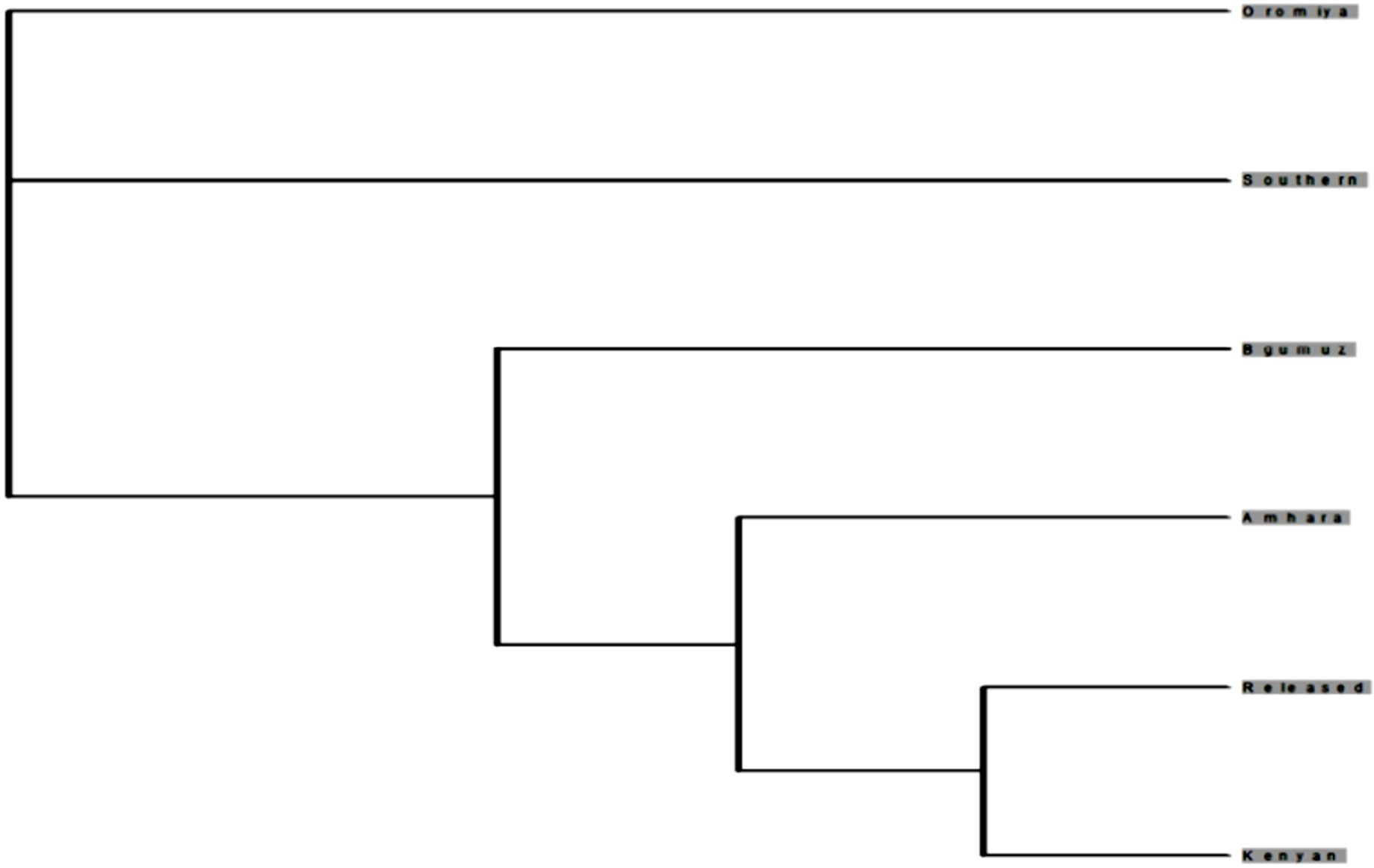
	Amhara	Bgumuz	Kenyan	Oromiya	Released	SNNP
Amhara	0.000					
Bgumuz	27.570	.000				
Kenyan	19.226	**54.601**	0.000			
Oromiya	14.973	14.168	5.207	0.000		
Released	10.537	3.838	2.738	6.193	0.000	
SNNP	0.000	**63.186**	9.461	**30.790**	6.480	0.000



Supplementary Figure 2. Patterns of allelic variation observed in the study populations along with important allelic values. Na (number of alleles), Na Freq>= 5% (number of alleles with frequencies greater than or equal to 5%, Ne (number of effective alleles), I (Shannon'sindex), number of private alleles, number of less common alleles (with frequencies less than or equal to 25% and 50%, and He (expected heterozygosity).



Supplementary Figure 3. Neighbor-joining dendrogram for the six (geographical) populations based on Nei's unbiased genetic distance (Nei,1983). Populations (from top to bottom): 'Amhara'; Southern; 'Oromiya'; 'Bensihangul-Gumuz'; 'standard' or 'released'; Kenyan.



Supplementary Figure 4. Neighbor-joining dendrogram for the six (geographical) populations based on shared-allele genetic distance values measured. Populations (from top to bottom): 'Oromiya'; Southern; 'Bensihangul-Gumuz'; 'Amhara'; 'standard' or 'released'; Kenyan.