Research Square

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

Genetic diversity and population structure of Yam (Dioscorea species) from Western Ethiopia as revealed by simple sequence repeat markers

Fekadu Korsa (fikekh@yahoo.com)
Institute of Biotechnology, Addis Ababa University, Addis Ababa
Tileye Feyissa
Institute of Biotechnology, Addis Ababa University, Addis Ababa
Obssi Dessalegn
Ethiopian Agricultural Research Institute, National Agricultural Biotechnology research center
Tilahun Mekonnen
Institute of Biotechnology, Addis Ababa University, Addis Ababa

Research Article

Keywords: AMOVA, Dioscorea species, Genetic diversity, Population structure, SSR marker

Posted Date: August 26th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1905131/v1

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

Abstract

Background: Yam (*Dioscorea species*) is one of the most significant staples tuber crops in Ethiopia. However, few studies of genetic diversity have been conducted for Ethiopian yam germplasm accessions. Therefore, the present study was targeted to assess the genetic diversity and population structure of yams in Ethiopia using SSR markers.

Results: A total of 118 yam genotypes representing five populations collected from four zones in western Ethiopia were assessed using ten SSR markers. A total of 92 alleles with an overall mean of 9.2 per locus were detected. The average expected heterozygosity (He) and Shannon–Weaver diversity index (I) were 0.88 and 2.13, respectively. All the SSR markers were polymorphic and highly informative, with an overall mean polymorphic information content (PIC) of 0.87. The overall mean expected heterozygosity (He) and percentage of polymorphic loci within populations were 0.73 and 98, respectively. The highest within-population expected heterozygosity was recorded for yam populations of West and East Wollega. It was observed that landrace populations of Ethiopian *Dioscorea species* are more diverse than the released variety. A statistically highly significant moderate genetic differentiation (PhiPT=0.100, p <0.0001) was observed, as revealed by AMOVA. Most of the total genetic variation (91%) accounted for the within-population variation, leaving only 9% for the among-populations genetic variation. The neighbor-joining cluster analysis, PCoA, and population structure analysis failed to sharply separate the yam genotypes according to their geographical origins and breeding materials. STRUCTURE analysis based on the Bayesian model weakly inferred two subgroups (K=2), confirming the high potential of genetic admixtures and close relationships among the studied yam genotypes, likely due to the presence of high gene flow (NM= 2.39).

Conclusions: The present study generated baseline genetic diversity information that was useful for yam breeders to improve desired agronomic traits through breeding and for designing appropriate conservation strategies. Finally, rapid genetic erosion is occurring in the released variety 'Bulcha'. Thus, breeders are recommended to broaden the genetic base of Bulcha varieties by crossing them with yam landraces through artificial hybridization.

Background

Yam (*Dioscorea sp.*) is one of the staple tuber crops that contribute substantially to millions of people in the tropical and subtropical regions of Africa, Asia, the Pacific and Latin America (Tamiru *et al.*, 2017). West Africa, Southeast Asia, and tropical America are considered to be the centers of origin, diversity and domestication of yam (Andres *et al.*, 2017). The genus *Dioscorea* consists of approximately 650 species worldwide (Couto *et al.*, 2018), but only 10 of them are cultivated species of economic importance (Wilkin *et al.*, 2005). Hahn (1995) reported that the chromosome number of yam is variable, ranging from 2n = 20 to 2n = 140 in common food species.

Yam species reproduce using both sexual and asexual modes of reproduction. However, many landrace cultivars, including most popular varieties, reproduce exclusively by vegetative propagation using either small whole tubers or pieces of large tubers (Mondo *et al.*, 2020). Norman *et al.* (2020) reported that the sexual reproduction of yam involves hand or natural open pollination during flowering to generate genetically variable offspring for selection.

Globally, approximately 8.8 million hectares of land were cultivated for yam, with an annual production of 74.8 million tons in 2020 (FAO, 2021). The highest yam production (97%) comes from West and Central Africa, where approximately 60 million people depend on them. In 2020. Africa's yam production was approximately 73.2 million tons from a total cultivation area of 8.7 million hectares. Nigeria, Ghana, Ivory Coast, Benin, and Togo are the top five world yam producers. In Ethiopia, yam is widely distributed and grown by subsistence farmers in the southern, southwestern, and western parts of the country (Bekele and Bekele, 2020). The total production of yam in Ethiopia in 2020 was approximately 45,730 tons on 4874 hectares of land by 356,872 farmers (CSA, 2021). The planting seasons of yams in Ethiopia vary from region to region. It is planted in October in most parts of southern Ethiopia (Hildebrand, 2003) and from November to December in the southwestern and western parts of the country at the onset of the dry season, making use of soil moisture reserves from the preceding rains. In Ethiopia, yam is considered to be a gap-filling crop for farmers during the scarcity of food because it is mainly planted at the onset of the dry season in October and harvested from the early-maturing landraces during May and June. It is an important food and nutrition security crop in the country (Tamiru *et al.*, 2008).

Yams are essential sources of food consumed in boiled, baked, or fried forms. It brings food and nutritional security to approximately 300 million people in Africa, Asia, some parts of South America, the Caribbean, and the South Pacific Islands (Nanbol and Namo, 2019). According to Bekele and Bekele (2014), the fresh tubers of yam have a high nutritional content of protein, fiber and important minerals, including calcium and iron, but a relatively low fat content. Yam has a high starch content and is used as animal feed, a source of medicine, and an industrial raw material (Andres *et al.*, 2017). Irrespective of its importance, yam belongs to a group of crops labelled as "orphaned crops", which have not received research attention for an extended period, and hence, little improvement has been made to the crop (Otoo, 2009).

Genetic diversity analyses provide a better understanding of the extent of genetic variation available in germplasm collections (Tumwegamire *et al.*, 2011). Genetic diversity is a precondition for successful plant breeding (Ulukan, 2009). The extent of genetic diversity can be revealed using morphological, biochemical and DNA marker systems (Mohammadi and Prasanna, 2003). The low level of polymorphism, high environmental influence, and tissue specificity (in biochemical assays) made the former methods less preferred for genetic analysis (Rao, 2004). Currently, highly polymorphic, informative and less environmentally influenced DNA markers have become useful for genetic diversity analysis, mapping and genome-assisted breeding (Acquaah, 2009).

To date, several DNA marker systems have been used to analyse the genetic diversity of yam, including AFLPs (Rivera-Jiménez *et al.,* 2011), SSRs (Arnau *et al.,* 2017; Adewumi *et al*., 2020) and ISSRs (Kung *et al*., 2016; Castañeda-Cardona *et al*., 2020). In Ethiopia, the genetic diversity of yam has also been studied using AFLPs (Tamiru *et al*., 2007), ISSRs (Kedra Mohammed *et al*., 2019), and SSRs (Abebe *et al*., 2013; Bekele and Bekele, 2014; Mulualem *et al*., 2018), most of which have concentrated on the southern and southwestern yam collections. Simple sequence repeats (SSRs) are considered a marker of choice because they are highly polymorphic, reproducible, codominant, locus specific, analytically simple, readily transferable, abundant, and have good genome coverage (Kalia *et al*., 2011; Jiang, 2013, and Zeinalzadeh, 2014).

There is limited information on the genetic diversity of yam landraces in western Ethiopia, particularly in the Metekel zone of the Beneshangul Gumuze regional state and the East and West Wollega zones of Oromia. Moreover, previous genetic diversity analysis did not consider the two newly released yam varieties called Bulcha (*Dioscorea alata*) and Lalo (*Dioscorea rotundata*). Furthermore, yam is the most researched neglected crop, and more researchers are required for its improvement. Therefore, the present study was conducted to investigate the genetic diversity of yam landraces and released varieties grown in western Ethiopia using SSR markers to generate baseline information for yam breeding programs and for conservation purposes.

The general objective of our study was to analyse the genetic diversity and population structure of yam (*Dioscorea species*) landraces and varieties using simple sequence repeat markers. Its specific objectives were 1. To investigate the genetic diversity of yam (*Dioscorea species*) varieties and landrace of western Ethiopia using microsatellite markers. 2. To determine the population genetic structure of yam genotypes of western Ethiopia.

Results

Microsatellite locus polymorphism

All ten SSR markers used in the present study were 100% polymorphic and detected a total of 92 alleles across 118 genotypes of the *Dioscorea species* collected from four zones in the western parts of Ethiopia (Table 3). Of the detected alleles, 16 (17.4%) alleles had frequencies between 0.01 and 0.05 (rare), 37 (40.2%) had frequencies between 0.05 and 0.1, and 39 (42.4%) alleles had frequencies greater than 0.1 (Table 2). Moreover, the frequency of the most common allele for individual markers was less than 0.95 or 0.99.

The detected number of alleles ranged from seven to 12 with an overall mean of 9.2. The highest number (12) of alleles was detected by the Dab2CO5 and Dab2E07 markers, while the lowest (7) was observed for the MTI11B marker. The major allele frequency (MAF) value ranged from 0.13 (Dab2CO5 and Dab2E07) to 0.30 (MT12B), with an overall average of 0.18. The effective number of alleles varied from 5.83 (MT12B) to 10.89 (Dab2E07), with an overall mean of 8.09 (Table 3). The Shannon–Weaver diversity index (I) varied from 1.91 (MTI 11B) to 2.44 (Dab2E07), with an overall average of 2.13.

The allelic richness ranged from 4.00 (MTI 11B to 7.20 (Dab2CO5) with a mean of 5.50. Likewise, the gene diversity ranged from 0.84 (MT12B) to 0.92 (Dab2E07), with an overall mean of 0.88. Only three markers (Dab2CO5, Da1A01, and MT12B) detected observed heterozygosity (Ho) with a value ranging from 0.14 (Da1A01) to 0.25 (Dab2CO5) with an overall average of 0.19. The lowest and highest genetic differentiation (Gst) were detected by the Dpr3D06A (0) and Dpr3D06A (0.12) markers, respectively, with an overall mean of 0.03. The highest (3.48) and lowest gene flow (Nm= 0.93) was detected for the loci Da1A01 and MTI 2B, respectively, with an overall mean of 1.68. All the microsatellite loci used had a high PIC (>0.5), with values ranging from PIC = 0.83 (MT12B) to 0.91 (Dab2E07). Furthermore, all the SSR markers used highly significantly deviated from Hardy-Weinberg equilibrium (Table 3).

Genetic Diversity of Population

Various genetic diversity parameters were computed to estimate the within- and among-population genetic variation for the five yam populations (three landraces and two released varieties). Among the study populations, the within-population, Ne, I, NPA, HO, and AR were the highest for the yam landrace populations of West Wollega and East Wollega (Table 4). In contrast, these diversity estimate parameters were the lowest for the cultivated variety Bulcha, followed by *cv* Lalo (Table 4). Moreover, the number of private alleles was absent in the *cv* Bulcha and Metekel landrace populations. Except for the Bulcha population, which showed 90%, all the other populations showed 100% polymorphisms across all the microsatellite loci, with an overall mean of 98%. The analysis revealed that the landrace *Dioscorea* populations showed relatively higher within-population genetic variation than the released yam populations (Table 4).

Genetic diversity among populations

The pairwise Nei's standard genetic distance and gene flow (Nm) for landraces and released yam genotypes are presented in Table 5. The pairwise Nei's genetic distance ranged from 0.37 - 1.21. The highest (1.21) pairwise Nei's genetic distance was observed between the landrace population of Metekel (MZ) and Bulcha with the lowest gene flow (N= 1.14). However, relatively lower gene flow (Nm = 1.25) was found between the Bulcha and East Wollega populations, with a high (1.20) coefficient of genetic distance. On the other hand, the lowest genetic distance (0.37) and highest (Nm = 6.08) gene flow were found between the landrace populations of East Wollega and West Wollega, respectively. Moreover, the lowest genetic differentiation (PhiPT) of the yam populations was scored between East and West Wollega (PhiPT = 0.039). However, statistically, the highest (PhiPT = 0.180, p = 0.001) genetic differentiation was recorded between the landrace population of the Metekel zone (MZ) and Bulcha varieties (Table 6).

Analysis of molecular variance (AMOVA)

AMOVA was computed for 118 genotypes of yam species based on their areas of collection and breeding materials as a criterion to identify the proportion of the genetic variation among and within populations using 999 permutations (Table 7). The AMOVA based on the sample collection areas showed 9% genetic variation among populations, leaving 91% of genetic variation within the populations. Furthermore, the analysis confirmed the presence of higher genetic variations within the studied populations with moderate genetic differentiation (PhiPT=0.10, P= 0.001) through high gene flow (2.39). On the other hand, the AMOVA based on landraces and released varieties revealed 6% and 94% of the genetic variation among and within the population, respectively. The overall genetic differentiation of landrace and released varieties of yam genotypes was moderate (PhiPT=0.06; P= 0.001) with higher gene flow (Nm: 3.86).

PCoA, cluster analysis, and population structure

Principal coordinate analysis (PCoA) was conducted for two categories of yam genotypes based on their sites of collection and breeding materials (Fig. 2a and b). For both categories of yam genotypes, the PCoA results showed that 6.26%, 4.57%, and 4.13% of the genetic variation was explained by the first, second, and third coordinate axes, respectively. Overall, the three coordinate axes explained approximately 14.97% of the total variation among the studied yam populations. The scatter plot obtained from PCoA indicated an almost uniform distribution of the yam landraces of all populations around the center of the two-dimensional coordinate plane with poor population clustering. Clustering did not clearly group the populations into distinct clusters according to their collection site and breeding materials (Fig. 2a, 2b).

The cluster analysis using the neighbor-joining method grouped the 118 yam genotypes into three major clusters (CI, CII, and CIII) (Fig. 3). Each major cluster was divided into two subclusters (i and ii). All Bulcha, most Lalo and Metekel, few West Wollega and East Wollega genotypes were assigned to cluster I. Cluster II consisted of most of the yam genotypes collected from both West Wollega and East Wollega, with a limited number of samples from Lalo varieties. Six yam genotypes from West Wollega, three from East Wollega and a single from Metekel zones were grouped in cluster III. The clustering analysis indicated that landraces and released yam varieties were clustered together (Fig. 3). In general, 45 (38.1%), 63 (53.4%), and 10 (8.5%) of the yam genotypes were grouped into major clusters I, II, and III, respectively. None of the yam genotypes from the four zones were assigned to specific clusters, indicating the presence of considerable admixture of the yam species. Likewise, the dendrogram based on the unweighted pair group method with arithmetic mean (UPGMA) grouped the five populations into three major clusters (I, II, and III) without corresponding to their collection sites (Figure 4).

The population structure of 118 yam genotypes was computed using Structure v 2.3.4 software based on the Bayesian model. The delta K value produced using an online structure harvester showed its maximum value at K=2 (Fig. 5B), indicating that the yam populations have two hypothetical subpopulations. Furthermore, the Clumpak result showed a greater degree of genetic intermixing (admixture) between the two yam subpopulations (Fig. 5A).

Discussion

Genetic diversity

All ten microsatellite loci used to profile the genetic diversity of 118 genotypes of *Dioscorea species* were found to be highly polymorphic with PIC > 0.5 (Botstein *et al*., 1980), indicating that they are highly informative and useful tools for genetic diversity analysis of yam. According to Botstein *et al*. (1980), the PIC value has four levels. These are as follows: 1) PIC near 1 is most desirable, 2) PIC > 0.5 is highly informative, 3) PIC: 0.5- 0.25 are reasonably informative, 4) PIC < 0.25 are slightly informative. The polymorphic information content (PIC) observed in the present study (0.83 to 0.91, mean 0.87) is in agreement with the average PIC value reported for the population from Ghana as reported by Adewumi *et al*. (2020) (PIC = 0.82) but in agreement with the report from Brazil (Siqueira *et al*., 2014) (0.91) and Ghana (0.65) (Otoo *et al*., 2015). However, the current result was higher than the yam genotypes of West Africa (0.65) (Obidiegwu *et al*., 2009), Brazil (0.79) (Nascimento *et al*., 2013), Ethiopia (0.30), (Mulualem *et al*., 2018), Seri Lanka (0.68) (Munaweera *et al*., 2020) and *Dioscorea species* using SSR markers. Therefore, the current results showed that all the markers used in this study had high discriminatory power and were useful for detecting the genetic diversity among the studied yam genotypes.

The markers produced a total of 92 alleles ranging from seven to 12, with an average of 9.2. The total number of alleles and their average per loci observed in the present study were considerably higher than those in previous studies by Abebe *et al*. (2013) (60, 8.6), Mulualem *et al*. (2018) (30, 3), Silva *et al*. (2016) in Brazil (33, 3.3), and Osuagwu *et al*. (2020) in West Africa (74,7.4). On the other hand, a higher number of alleles and their average per locus (256, 10.7) were reported for yam populations of France by Arnau *et al*. (2017) using 367 accessions with 24 SSR markers. These higher differences could be due to the high number of genotypes and the microsatellite loci level of polymorphism used in the study.

The average expected heterozygosity (He= 0.88) per locus observed in the present study was higher than previous studies by Arnau *et al*. (2017) in France (He= 0.66), Mulualem *et al*. (2018) in Ethiopia (He = 0.53), Adewumi *et al*. (2020) in Ghana (He = 0.81) for *Dioscorea species* using SSR markers. However, it was lower than the level of average gene diversity (He = 0.92) reported by Otoo *et al*. (2015). Therefore, the high genetic diversity observed in our current study might be the presence of a high level of gene flow among the studied yam genotypes and the collection of samples from a wide range of geographical areas with various levels of selection pressure. Moreover, the mean observed heterozygosity was considerably lower than the level reported by Tostain *et al*. (2006) (0.58) and Abebe *et al*. (2013) (0.48), likely due to high inbreeding and null alleles.

The average Shannon index information (I= 2.23) in the current study was higher than that reported by Nascimento *et al*. (2013) in Brazil (I= 0.40), Bekele and Bekele (2014) in Ethiopia (I= 0.49), and Adewumi *et al*. (2020) in Ghana (I= 1.94). The observed high Shannon index information in the current study

revealed the presence of high genetic diversity in the studied yam genotypes. The study also revealed the existence of significant deviation from Hardy-Weinberg equilibrium, likely due to the presence of high gene flow, nonrandom mating, and artificial selection by farmers for some target traits.

Population Genetic Diversity

The genetic diversity parameters used in the current study confirmed the presence of high genetic diversity within the Ethiopian *Dioscorea species*. The average within-population allelic richness (AR=5.50), number of alleles (Na=5.62), number of effective alleles (Ne=4.45), expected heterozygosity (He=0.73), number of private alleles (NPA= 1.42) and percent of polymorphism (%P=98) for Ethiopian *Dioscorea species* were higher than the previous studies reported by Olu-Olusegun *et al.* (2018), who observed Na= 4.8 and He= 0.68 for yam populations of Japan, and by Abebe *et al.* (2013) (Ne=6.09, Ho= 0.48, AR= 4.47, He=0.59 and %P 90.47) for yam populations of Ethiopian. However, the average within populations Na=1.93, Ne=1.63, I= 0.45 and He=0.33, and PPL = 58.6 reported by Mulualem *et al.* (2018) was substantially lower than the corresponding values in our study. The variation could be due to genotype differences and differences in the levels of marker polymorphism.

The populations of West and East Wollega showed Na (8.70, 7.60), AR (8.20, 7.50), Ne (6.44, 5.77), He (0.83, 0.82), I (1.95, 1.85), and PPL (100,100) values, respectively, which were higher than their respective grand means, indicating that these areas are genetic hot spots for *Dioscorea species*. Hence, they are appropriate sites for *in situ* conservation, further yam genetic analysis and sources of useful genes for yam improvement through breeding. Furthermore, the private alleles observed in the population of West Wollega (NPA = 0.3) demonstrate that the population might have evolved independently to generate and maintain its unique alleles at the population level. On the other hand, the population of Bulcha (released variety) showed the least genetic diversity, likely due to inbreeding, high gene flow and a high rate of artificial selection pressure.

Genetic differentiation of the population

A statistically highly significant moderate genetic differentiation (PhiPT=0.100, p <0.0001) was observed, as revealed by AMOVA. Most of the total genetic variation (91%) accounted for the within-population variation, leaving only 9% for the among-populations genetic variation. Similar lower population genetic variation was reported for *Dioscorea species* from different parts of the world, including Kenya (12%) (Mutamia *et al.*, 2013), Ethiopia (17%) (Mulualem *et al.*, 2018), Ghana (7%) (Adewumi *et al.*, 2020), and West Africa (14%) (Osuagwu *et al.*, 2020). Globally, the lowest (2%) and highest (40.9%) among population genetic variation were reported for the yam populations of Japan (Olu-Olusegun *et al.*, 2018) and France (Arnau *et al.*, 2017), respectively. The current study showed the existence of significant moderate genetic differentiation among the population (PhiPT=0.10, p <0.001) with higher gene flow (Nm = 2.39). Similar findings were reported by Mulualem *et al.* (2018) with a population differentiation rate of 0.060, Adewumi *et al.* (2020) with Fst=0.066, and Osuagwu *et al.* (2020) with Fst = 0.141. Lower (< 0.05) population differentiation rates were reported by Siqueira *et al.* (2014) and Loko *et al.* (2016). The higher gene flow observed in the present study might be due to the closeness of the sample collection zones, which might facilitate the exchange of planting materials between farmers through the common market.

The highest pairwise population differentiation was observed between the landrace population of the Metekel zone (MZ) and Bulcha varieties (0.18). This finding was further confirmed by the highest pairwise genetic distance recorded between them. This could be due to the existence of limited gene flow between these populations. On the other hand, the lowest pairwise genetic population differentiation and pairwise genetic distance were observed between the populations of West Wollega and East Wollega. This could be explained by the presence of higher yam seed exchange among the nearby districts of the two zones through traders and farmers that may contribute to the gene flow (Mulualem *et al.*, 2018).

Population Structure

Neighbour-joining cluster analysis grouped the studied yam populations into three major clusters based on their geographical origins. None of the major clusters sharply grouped the samples into distinct clusters, suggesting the occurrence of high rates of gene exchange among the studied populations collected from different zones. This could be likely due to high gene flow or might have a similar gene pool. UPGMA clustering based on Nei's genetic distance clustered the populations into three groups, supporting the results of neighbor-joining clustering irrespective of their geographical origins. The landrace clustering pattern is not strong enough to support the principle of "isolation by distance".

Moreover, UPGMA showed sharp separation between landraces and varieties. Furthermore, PCOA failed to sharply group the yam genotypes according to their geographical origins and breeding materials. Similar findings were reported in previous studies in Kenya (Mutamia *et al.*, 2013) and Brazil (Silva *et al.*, 2016). In contrast, strong population clustering was reported for yams (*Dioscorea alata* L.) populations of West African countries (Obidiegwu *et al.*, 2009). This could be due to the presence of high gene flow between the studied yam populations.

The results of population structure analysis supported the PCOA and neighbor-joining clustering, confirming the high potential of genetic admixtures and close relationships among the studied *Dioscorea species* genotypes collected from the different zones of western Ethiopia, likely due to high gene flow. The structure analysis weakly grouped the populations into two subpopulations (K=2), and each subgroup contained samples collected from the different zones. A similar weak population structure in yam was reported by Siqueira *et al.* (2014) and Adewumi *et al.* (2020). Conversely, Arnau *et al.* (2017) reported the existence of a strong population structure in yam following geographical areas of sampling. Surprisingly, no genetic structure was reported for *D. bulbifera* accessions in Brazil (Silva *et al.*, 2016), likely due to the occurrence of excess gene flow among farmers that made the genetic diversity similar among the populations.

Conclusion

Generally, knowledge of the genetic diversity of crops is a prerequisite and the foremost important step in plant breeding programs. In the present study, the genetic diversity and population structure of yam collected from western Ethiopia were uncovered using highly polymorphic microsatellite markers. The study confirmed the presence of high genetic diversity in Ethiopian yam populations, indicating the possibility of improvement through breeding and the importance of maintaining diversity by deploying appropriate conservation strategies. The study also indicated that most of the total genetic variation (91%) accounted for within populations, likely due to genetic recombination and mutation. The populations exhibited moderate genetic differentiation, possibly due to the presence of high gene flow. The highest genetic diversity indices were recorded for the yam populations of West Wollega followed by East Wollega, suggesting that these areas are hot spots for yam genetic diversity studies, sources of important alleles for breeding purposes, and *in situ* conservation strategies at the regional and/or national levels. Overall, the present study successfully generated baseline information on the genetic diversity and population structure of yam collected from western Ethiopia using highly informative simple sequence repeat markers. This information could be useful for yam improvement through selection breeding and proper germplasm management.

Materials And Methods

Plant materials

A total of 118 genotypes consisting of 95 landraces (38 from East Wollega, 46 from West Wollega, and 11 from the Metekel zone) were collected from farmers' fields, and 23 varieties were obtained from the Bako Agricultural Research Center (BARC) and were the subject of this study (Sup.1).

In the downstream diversity analysis, genotypes collected from the same administrative zone were considered as one population with the assumption that they were more likely shared within zones than among zones through markets and farmers' seed exchange. The two released varieties, namely, Bulcha and Lalo, collected from BARC were considered two different populations.

Yam tubers were collected following the main roads and accessible routes in a randomly selected district, and stops were made at 5 - 10 km intervals based on vehicle odometers where yam fields were observed. Collected yam tubers were planted in pots filled with sterilized topsoil and grown in the greenhouse of the National Agricultural Biotechnology Research Center (NABRC), Holeta, Ethiopia, located 29 km west of Addis Ababa. For 11 samples from the Metekel zone, young leaves were collected using silica gel and transported to the Molecular Biology Laboratory of NABRC for DNA extraction. Sample passport data, including administrative region, geographical location and elevations, were recorded from which collection map was produced (Fig. 1).

Leaves Sample collection and preparation

For 107 yam genotypes, two to three fresh and healthy leaf samples per genotype were collected into 2 ml Eppendorf tubes. The Eppendorf tubes with the sample were kept at -20°C for a minimum of 24 hours and then lyophilized. This made the samples ready for DNA extraction.

Genomic DNA extraction

Total genomic DNA was extracted using the cetyl trimethyl ammonium bromide (CTAB) protocol with minor modification (Turaki *et al.*, 2017) at NABRC, Molecular Biology Laboratory, Holeta, Ethiopia. One hundred milligrams of lyophilized yam leaves in a 2 ml Eppendorf tube was ground in the presence of two metal beads using a Geno Grinder (MM-200, Retsch).

Then, 1000 ml of pre warmed extraction buffer of (2% (w/v) CTAB, 100 mM Tris-HCl (pH 8.0), 20 mM EDTA, 1.4 M NaCl) containing freshly added 2% (w/v) polyvinylpyrrolidone-40 (PVP-40) and 1% (w/v) sodium sulfite (Na_2SO_3) was added to each ground sample. The mixture was vortexed and incubated in a water bath (60°C, 30 min) for cell lysis.

Then, 700 ml phenol: chloroform: isoamyl alcohol solution at a 25:24:1 (v/v/v) ratio was added to the incubated buffer, vortexed and then centrifuged at 15000 × g for 10 min. Then, 700 µl of the aqueous layer was transferred carefully into a 1.5 ml new Eppendorf tube, and 700 µl chloroform:isoamyl alcohol 24:1 (v/v) was added again, vortexed, and centrifuged at 15000 × g for 10 min.

After that, 700 µl of the aqueous layer was transferred carefully into a new 1.5 ml fresh Eppendorf tube. Then, 75 µl of 5 M NaCl and 450 µl of cold isopropanol were added to the mixture. The mixture was inverted 4 to 5 times and incubated at -20°C for 1 h. The mixture was centrifuged at 15000 × g for 10 min to pellet the nucleic acids. The supernatant was decanted carefully.

Then, the pellet was washed twice by adding 500 µl of 70% (v/v) ethanol and centrifuged at 15000 × g for 5 min. Finally, the pellet was dried in vacuum for 30 min, resuspended in 100 µl 1x TE buffer (10 mM Tris-HCl pH 8.0, 0.1 mM EDTA) and stored at -20°C.

Determination of DNA concentration and quality

The extracted genomic DNA concentration was determined using a Nanodrop Spectrophotometer (Nano Drop® ND-8000). The DNA quality was checked using gel electrophoresis by loading a mix of 5 µl sample DNA on a 1% agarose gel at 100 V for one hour. The gel was stained with gel red, and the gel was visualized using a gel documentation system (Bio Doc-ItTM imaging system) under UV light.

Genotyping

For polymerase chain reaction (PCR), the genomic DNA was adjusted to 20 ng/µl working solution using nuclease-free water. Finally, the working solution was stored at -20°C until used for PCR. For genotyping, a pair of ten polymorphic SSR markers were selected from published literature (Table 1). The annealing temperatures of the primers were optimized using the touch-down PCR program of the Bio-Rad T100^{TM} thermal cycler (Korbie *et al*., 2008). The final volume for PCR amplification reactions was 10 µl consisting of 3.70 µl of nuclease-free water, 0.25 µl of each forward and reverse primer (10 mM), 5 µl of 2X Go Taq green master mix, and 0.8 µl of a template DNA sample. The PCR conditions were adjusted at initial denaturation at 94°C for 5 min, followed by 15 cycles of annealing at 60°C for 25 s (temperature decrease of 1°C per cycle), a second annealing temperature of 40°C for 45 s, and extension at 72°C for 1 min, followed by a final extension at 72°C for 10 minutes. Finally, the holding temperature was adjusted to 4°C with an infinites time.

Gel electrophoresis

The amplified PCR products were detected using 2% agarose gel electrophoresis stained with gel red. The gel was run for three hours at 80 V and 100 A. A 100 bp ladder was used to estimate the length of alleles. A negative control was used by adding nuclease-free water to the master mix instead of DNA samples. The gel was visualized using UV light, and the image was captured using gel documentation (Bio Doc-ItTM imaging system).

Data scoring and analysis

The clear and visible amplified bands of 118 genotypes of yam species using 10 SSR markers were scored using the PyElph 1.4 software package (Pavel and Vasile, 2012). The scored base pairs were analysed using different software. POPGENE version 1.32 software (Yeh *et al*., 1999) was used to compute the genetic diversity parameters such as the total number of alleles per locus (Na), the number of effective alleles (Ne) (Kimura *et al*., 1964), the Shannon information index (I) (Lewontin, 1972), the expected heterozygosity (He), the observed heterozygosity (Ho), and the percentage of polymorphic loci (PPL).

Nei's genetic identity (NGI), genetic distance (GD) (Nei, 1972) and gene flow (Nm) were calculated as 0.25(1-Gst)/Gst, allele frequency, number of private alleles, analysis of molecular variance (AMOVA), and coefficient of genetic differentiation, PHWE (P value for deviation from Hardy Weinberg equilibrium), and principal coordinate analysis (PCoA) was calculated using GeneAlex version 6.5 software packages (Peakal and Smouse, 2012). Furthermore, H-Rare software was used to compute the allelic richness (A) (Kalinowski, 2005).

A genetic dissimilarity matrix was computed based on the continuous Euclidian dissimilarity index, and neighbor-joining (NJ) and Nei's standard genetic distance (DST, corrected) (Nei, 1972)-based unweighted pair group method with arithmetic mean (UPGMA) trees were generated using Powermaker software (Liu and Muse, 2005). The same software was used to calculate the major allele frequency (MAF), gene diversity (GD), and polymorphic information content (PIC). The phylogenetic tree was drawn using Darwin 6 software (Perrier and Jacquemoud-Collet, 2006).

STRUCTURE software version 2.3.4 (Pritchard *et al*., 2000) was used to determine the population structure and admixture patterns of the yam genotypes based on the Bayesian model clustering algorithm. The true number of population clusters (K) was estimated using a burn-in period of 10,000 for each run and 100,000 Markov chain Monte Carlo (MCMC) to collect data with replications of K between 1 and 10 using 20 iterations for each K. The simulation method of Evanno *et al.* (2005) was applied to predict the optimum value of K using the online STRUCTURE HARVESTER version. 0.6.94 (Earl and Von Holdt, 2012). A bar plot for the optimum K was determined using the online CLUMPAK beta version (Kopelman *et al*., 2015).

Declarations

Acknowledgements

Not applicable.

Funding

Not applicable.

Abbreviations

BARC: Bako Agricultural Research Center, NABRC: National Agricultural Biotechnology Research Center, CSA: Central Statistical Agency

Availability of supporting data and materials

The data set of the studied materials in our study are deposited in the national herbarium of Ethiopia under accession numbers WWGS01-MZG23.

Supplementary information

Sup 1. 118 Specimen voucher information from National Herbarium of Ethiopia.

Sup 2 DNA bands of studied genotypes.

Sup 3 Photo take during sample collection.

Declarations

Ethics approval and consent to participate

Study protocol

The study protocol was complied with the institutional, national, and international guidelines and legislation.

Permission of plant materials collection

Permission was obtained from the institute of Biotechnology, Addis Ababa University and National Agricultural Biotechnology Research Center of Ethiopia for the collection of studied plant materials from farmer's field.

Informed consent

Verbal agreement was made between the researcher and farmers for the collection of the studied plant materials from their own farm lands.

Informed Consent for publication

Not applicable.

Identification of the plant materials

The studied plant material were formally identified by the taxonomists of the National Herbarium of Ethiopia.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

F.K: Preparations of original thesis proposal, methodology, sample collections, Laboratory works, Data analysis, and write up of the final thesis report. T.F: Supervision, writing-review and editing. O.D: providing all necessary laboratory facilities, chemicals and ten SSR markers. TM: supervision, writing-review and editing. Providing samples collected from Metekel zones. The author(s) read and approved the final manuscript.

References

1. Abebe W., Demissew S., Fay, M.F., Smith, R.J., Nordal, I. and Wilkin, P. (2013). Genetic diversity and population structure of Guinea yams and their wild relatives in South and South West Ethiopia as revealed by microsatellite markers. *Genetic Resources and Crop Evolution*. **60**(2):529-541.
2. Acquaah, G. (2009). *Principles of plant genetics and breeding*. John Wiley and Sons. pp. 454-458.
3. Adewumi, A.S., Taah, K.J., Asare, P.A., Adu, M.O. and Agre, P.A. (2020). Assessment of genetic diversity of *Dioscorea praehensilis* (Berth.) collected from Central Region, Ghana using simple sequence repeat (SSR) markers. *ES Journal of Agriculture and Current Research.* **1**(1): 1005-1012.
4. Andres, C., AdeOluwa, O.O. and Bhullar, G.S. (2017). Yam (*Dioscorea spp.*)-A rich staple crop neglected by research. In *Encyclopedia of Applied Plant Sciences*. Academic Press. **2**: 435-441.
5. Bekele, A. and Bekele, E. (2014). Studies on Molecular Genetic Diversity and Useful Genomic Traits of Yam (Dioscorea spp) Germplasm Collections from Ethiopia. Doctoral dissertation, Addis Ababa University.
6. Bekele, A. and Bekele, E. (2020). Identification of Ethiopian Yam (Dioscorea spp.) Collections and Their Phenotypic Diversity Study. Agricultural Sciences. 11(11):1116-1132.
7. Botstein, D., White, R.L., Skolnick, M. and Davis, R.W., 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American journal of human genetics, **32**(3): 314-331.
8. Castañeda-Cardona, C.C., Morillo-Coronado, Y. and Morillo, A.C. (2020). Assessing the genetic diversity of *Dioscorea alata* and related species from Colombia through intersimple sequence repeat (ISSR) markers. *Chilean journal of agricultural research*. **80**(4): 608-616.
9. Couto, R.S., Martins, A.C., Bolson, M., Lopes, R.C., Smidt, E.C. and Braga, J.M.A. (2018). Time calibrated tree of Dioscorea (Dioscoreaceae) indicates four origins of yams in the Neotropics since the Eocene. *Botanical Journal of the Linnean Society*. **188**(2):144-160.
10. CSA (2021). Reports on area and production of crops (Private Peasant Holdings, Meher Season). Addis Ababa, Ethiopia.

11. Earl, Dent A. and vonHoldt, Bridgett M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*. **4** (2):359-361.
12. FAO (2021). Food and Agriculture Organization of the United Nations Statistics database, FAOSTAT. Retrieved from https://www.fao.org/faostat/en/#data/QCL assessed at April 15, 2022
13. Hahn, S.K., 1995. Yams: *Dioscorea spp.(Dioscoreaceae*) In J. Smaartt and Simmonds NW (Eds). Evolution of crop plants.
14. Hildebrand, E.A. (2003). Motives and opportunities for domestication: an ethnoarchaeological study in southwest Ethiopia. *Journal of Anthropological Archaeology*. **22**(4):358-375.
15. Hochu, I., Santoni, S. and Bousalem, M. (2006). Isolation, characterization and cross-species amplification of microsatellite DNA loci in the tropical American yam Dioscorea trifida. *Molecular Ecology Notes*. **6**(1):137-140.
16. Jiang, G.L. (2013). Molecular markers and marker-assisted breeding in plants. *Plant Breeding from Laboratories to Fields*.pp.45-83.
17. Kalinowski, ST. (2005). HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes.* **5**:187-189.
18. Kalia, R.K., Rai, M.K., Kalia, S., Singh, R. and Dhawan, A.K. (2011). Microsatellite markers: an overview of the recent progress in plants. *Euphytica*. **177**(3):309-334.
19. Kimura, M. and Crow, J.F. (1964). The number of alleles that can be maintained in a finite population. *Genetics*. **49**(4):725-738.
20. Kung, T.L., Lin, K.H. and Lin, S.F. (2016). Genetic diversity of *Dioscorea japonica* germplasm in Taiwan revealed by intersimple sequence repeat DNA markers. *Plant Genetic Resources*. **14**(3):211-218.
21. Lewontin, R.C. (1972). Testing the theory of natural selection. *Nature*. **236**(5343):181-182.
22. Liu, K. and Muse,SV. (2005). Power Marker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*. **21**(9):2128-2129.
23. Mohammed K., Tesfaye T., and Hailesilassie T. (2019).Genetic diversity assessment of yams(*Dioscorea spp.*) from Ethiopia using inter simple sequence repeat (ISSR) markers. *African Journal of Biotechnology*. **18**(30):970-977.
24. Mohammadi, S.A. and Prasanna, B.M. (2003). Analysis of genetic diversity in crop plants—salient statistical tools and considerations. *Crop science*. **43**(4):1235-1248.
25. Mulualem, T., and WeldeMichael, G., (2013). Agronomical evaluation of aerial yam (*Dioscorea bulbifera*) accessions collected from South and Southwest Ethiopia. *Greener Journal of Agricultural Sciences. 3(9):693-704.*
26. Mulualem, T. (2016). Genetic diversity, path coefficient analysis, classification and evaluation of yams (Dioscorea spp.) in Southwest Ethiopia. Doctoral dissertation, Haramaya University.
27. Mulualem, T., Mekbib, F., Shimelis, H., Gebre E.,and Amelework, A. (2018). Genetic diversity of yam (Dioscorea spp.) landrace collections from Ethiopia using simple sequence repeat markers. *Australian Journal of Crop Science*. 12(8):1222.
28. Mulualem, T., Mekbib, F., Shimelis, H., and Gebre E., (2021). Agro morphological diversity of Yams (*Dioscorea spp.*) Landraces from Southwest Ethiopia assessed through quantitative and qualitative traits. *Research Journal of Pharmacognosy and Phytochemistry*. 13(2):63-72.
29. Munaweera, K., Samarasinghe, G., Shreemathie, E. and Dasanayake, N. (2020). Genetic Variation of Yam *Dioscorea species* (Family Dioscoreaceae) Accessions of Sri Lanka. **5**(1): 6-12
30. Muthamia, Z.K., Morag, F.E., Nyende, A.B., Mamati, E.G. and Wanjala, B.W. (2013). Estimation of genetic diversity of the Kenyan yam (*Dioscorea spp.*) using microsatellite markers. *African Journal of Biotechnology*. ***12***(40):5846-5851
31. Mondo, J.M., Agre, P.A., Edemodu, A., Adebola, P., Asiedu, R., Akoroda, M.O. and Asfaw, A.(2020). Floral Biology and Pollination Efficiency in Yam (*Dioscorea species.*). *Agriculture*. **10**(11):560.
32. Muluneh Tamiru, Becker, H. C., & Maass, B. L. (2007). Genetic diversity in yam germplasm from Ethiopia and their relatedness to the main cultivated *Dioscorea species* assessed by AFLP markers. *Crop Science*. **47**(4):1744-1753.
33. Nascimento WF, Rodrigues JF, Koehler S, Gepts P, *et al*. (2013). Spatially structured genetic diversity of the endangered Amerindian yam (*Dioscorea trifida* L.) assessed by SSR and ISSR markers in southern Brazil. *Genetic Resource Crop Evolution*. **60**: 2405-2420.
34. Nanbol, K. and Namo, O. (2019). The Contribution of Root and Tuber Crops to Food Security: A Review. *Journal of Agricultural Science and Technology B*. **9**(4):221-233.
35. Norman, P.E., Paterne, A.A., Danquah, A., Tongoona, P.B., Danquah, E.Y., De Koeyer, D., Ikeogu, U.N., Asiedu, R. and Asfaw, A. (2020). Paternity Assignment in White Guinea Yam (*Dioscorea Rotundata*) Half-Sib Progenies from Polycross Mating Design Using SNP Markers. *Plants*. **9**(4):527-542
36. Olu-Olusegun, F., Babil, P., Kikuno, H., Irie, K., Toyohara, H. and Shiwachi, H. (2018). Intra–specific genetic diversity in water yam (*Dioscorea alata* L.) accessions collected from Myanmar, Papua New Guinea and Japan. *J ISSAAS*. **24**(1):107-117.
37. Otoo, E., Akromah, R., Kololesnikova-Allen, M. and Asiedu, R. (2009). Ethno-botany and morphological characterization of the yam pona complex in Ghana. In *Africa Crop Science Conference Proceedings*.**9**:407-414.
38. Osuagwu, A.N. and Edem, U.L. (2020). Evaluation of genetic diversity in Aerial Yam (*Dioscorea bulbifera* L) Using Simple Sequence Repeats (SSR) Markers. *Agro technology*.**9**:202.
39. Pavel, A. B., and Vasile, C. I. (2012). PyElph - a software tool for gel images analysis and phylogenetics. *BioMed Central Bioinformatics*.**13**:9
40. Perrier, X. and Jacquemoud-Collet, J.P. (2006). DARwin.

41. Pritchard, J.K., Stephens, M. and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*. **155**(2):945-959.
42. Rao, N.K. (2004). Plant genetic resources: Advancing conservation and use through biotechnology. *African Journal of Biotechnology*. **3**(2):136-145.
43. Rivera-Jiménez, H.J., Álvarez-Soto, A., Palacio-Mejía, J.D., Barrios-Leal, D.Y. and López-Álvarez, D. (2011). Diversidad genética intra e interespecífica de ñame (*Dioscorea spp*.) de la región Caribe de Colombia mediante marcadores AFLP. *Acta Agronómica*. **60**(4):328-338.
44. Scarcelli, N., Cubry, P., Akakpo, R., Thuillet, A.C., Obidiegwu, J., Baco, M.N., Otoo, E., Sonké, B., Dansi, A., Djedatin, G. and Mariac, C. (2019). Yam genomics supports West Africa as a major cradle of crop domestication. *Science Advances*. **5**(5):1947.
45. Siqueira, M.V., Dequigiovanni, G., Corazon-Guivin, M.A., Feltran, J.C. and Veasey, E.A. (2012). DNA fingerprinting of water yam (*Dioscorea alata*) cultivars in Brazil based on microsatellite markers. *Horticultura Brasileira*. **30**(4):653-659.
46. Siqueira MVBM, Bonatelli ML, Günther T, Gawenda I. (2014). Water yam (*Dioscorea alata* L.) diversity pattern in Brazil: an analysis with SSR and morphological markers. *Genetics Resources Crop Evolution*. **61**: 611-624.
47. Tamiru, M., Becker, H. C., & Maass, B. L. (2007). Genetic diversity in yam germplasm from Ethiopia and their relatedness to the main cultivated *Dioscorea species* assessed by AFLP markers. *Crop Science*. 47(4):1744-1753.
48. Tamiru, M., Becker, H.C. and Maass, B.L. (2008). Diversity, distribution and management of yam landraces (*Dioscorea species*) in Southern Ethiopia. Genetic Resources and Crop Evolution. 55(1):115-131.
49. Tamiru, M., Yamanaka, S., Mitsuoka, C., Babil, P., Takagi, H., Lopez-Montes, A., Sartie, A., Asiedu, R. and Terauchi, R. (2015). Development of genomic simple sequence repeat markers for yam. *Crop Science*. 55(5):2191-2200.
50. Tamiru, M., Natsume, S., Takagi, H., White, B., Yaegashi, H., Shimizu, M., Yoshida, K., Uemura, A., Oikawa, K., Abe, A. and Urasaki, N. (2017). Genome sequencing of the staple food crop white Guinea yam enables the development of a molecular marker for sex determination. *BioMed Central Biology*. 15(1):1-20.
51. Tostain, S., Scarcelli, N., Brottier, P., Marchand, J.L., Pham, J.L. and Noyer, J.L. (2006). Development of DNA microsatellite markers in tropical yam (*Dioscorea spp*.). *Molecular Ecology Notes*. **6**(1):173-175.
52. Tumwegamire, S., Kapinga, R., Rubaihayo, P.R., LaBonte, D.R., Grüneberg, W.J., Burgos, G., Zum Felde, T., Carpio, R., Pawelzik, E. and Mwanga, R.O. (2011). Evaluation of dry matter, protein, starch, sucrose, β-carotene, iron, zinc, calcium, and magnesium in East African sweetpotato [*Ipomoea batatas* (L.) Lam] germplasm. *HortScience*. **46**(3):348-357.
53. Turaki, A.A., Ahmad, B., Magaji, U.F., Abdulrazak, U.K., Yusuf, B.A. and Hamza, A.B. (2017). Optimized cetyltrimethylammonium bromide (CTAB) DNA extraction method of plant leaf with high polysaccharide and polyphenolic compounds for downstream reliable molecular analyses. *African Journal of Biotechnology*. **16**(24):1354-1365.
54. Ulukan, H. (2009). The evolution of cultivated plant species: classical plant breeding versus genetic engineering. *Plant Systematics and Evolution*. **280**(3):133-142.
55. Wilkin, P., Schols, P., Chase, M.W., Chayamarit, K., Furness, C.A., Huysmans, S., Rakotonasolo, F., Smets, E. and Thapyai, C. (2005). A plastid gene phylogeny of the yam genus, Dioscorea: roots, fruits and Madagascar. *Systematic Botany*. **30**(4):736-749.
56. Yeh, F.C., Yang, R.C. and Boyle, T. (1999). Microsoft windows-based freeware for population genetic analysis, quick user guide. *Edmonton: Center for International Forestry Research, University of Alberta*.
57. Zeinalzadeh., H. T. (2014). Genetic diversity of sunflower genotypes using SSR and TRAP markers. Ph.D. Thesis. Graduate School of Natural and Applied Sciences. Ataturk University, Erzurum, Turk.

Tables

Table 1

Sequence of primers, annealing temperature, and fragment size range detected using 10 SSR markers

Primer Code	Sequence 5′-3′	Repeat motif	AT (C^o)	Size (bp)	References
Dab2C05	F: CCCATGCTTGTAGTTGT R: TGCTCACCTCTTTACTTG	(GA)19	50.4	177-211	Tostain *et al.* (2006)
Dab2E07	F: TTGAACCTTGACTTTGGT R: GAGTTCCTGTCCTTGGT	(CT)23	52.5	126-183	Tostain *et al.* (2006)
DalA01	F: TATAATCGGCCAGAGG R: TGTTGGAAGCATAGAGAA	(GT)8	48.2	196-240	Tostain *et al.* (2006)
Dpr3D06A	F: ATAGGAAGGCAATCAGG R: ACCCATCGTCTTACCC	(GA)15	53.1	128-220	Tostain *et al.* (2006)
Dpr3F04A	F: AGACTCTTGCTCATGT R: GCCTTGTTACTTTATTC	(AG)15	48.2	107-133	Tostain *et al.* (2006)
MTI11B	F: CTCTTITGCTICTCATTTCA R: ATGTAGCCAATCCAAAATAG	$(CAA)_7$	51.6	131-153	Hochu *et al.* (2006)
MTI12B	F: CTGCCAGCGTTCCGATTC R: CGTAGGACCTCTCGCATCAG	(CTT)8	60.7	100-133	Hochu *et al.* (2006)
MTI2B	F:TCATCAAGAGCATCAAAAAAC R: GCCTCGTCTTTGAAGTTGGT	(CAA)6	54.9	118-146	Hochu *et al.* (2006)
Yam 09C	F: AGGAACATTCCC ACTCAGTTATG R: ATTGGGCAAGTGTGGTGTG	(CTT)12	57.9	132-221	Muluneh Tamiru *et al.* (2015)
Yam 12C	F: TGAGCATTCTTGTTTTGCC G R: CTTTCAGGGCGTGCATGG	(ATC)5.(AAC)8	57.6	195-235	Muluneh Tamiru *et al.* (2015)

Table 2

Summary of allele numbers and their frequencies

Marker	Rare alleles (0.01-0.05)	Common alleles (0.05-0.1)	Most frequent allele (>0.1)	Total
Dab2CO5	5	4	3	12
Dab2E07	3	5	3	12
Da1A01	1	4	3	8
Dpr3D06A	2	5	2	9
DPr3F04A	0	5	4	9
MTI11B	0	3	4	7
MTI12B	1	3	4	8
MTI2B	0	1	7	8
Yam 09C	2	5	3	10
Yam 12C	2	2	5	9
Total	16	37	39	92
Percentage	17.4%	40.2%	42.4%	100%

Table 3

Summary of genetic diversity parameters of ten microsatellite markers across 118 Dioscorea species genotypes

Markers	N	MAF	Na	Ne	I	AR	He	HO	Gst	Nm	PIC	PhiPT	P value	PHWE	%P
Dab2CO5	118	0.13	12.00	10.76	2.43	7.20	0.91	0.25	0.02	1.64	0.90	0.09	0.001	0.000***	100
Dab2E07	111	0.13	12.00	10.89	2.44	6.40	0.92	0.00	0.03	1.71	0.91	0.05	0.002	0.001***	100
Da1A01	116	0.19	8.00	7.39	2.04	6.40	0.87	0.14	0.01	3.48	0.85	0.05	0.003	0.000***	100
Dpr3D06A	106	0.16	9.00	7.08	2.03	3.60	0.88	0.00	0.12	0.38	0.86	0.24	0.001	0.000***	100
DPr3F04A	117	0.18	9.00	8.06	2.14	6.20	0.88	0.00	0.00	2.35	0.87	0.04	0.007	0.000***	100
MTI 11B	114	0.18	7.00	6.58	1.91	4.00	0.86	0.00	0.04	1.41	0.84	0.09	0.001	0.000***	100
MTI 12B	111	0.30	8.00	5.83	1.93	6.40	0.84	0.18	0.01	2.14	0.83	0.05	0.002	0.004**	100
MTI 2B	114	0.17	8.00	7.28	2.03	4.80	0.87	0.00	0.03	0.93	0.86	0.13	0.001	0.000***	100
Yam 09C	117	0.16	10.00	9.06	2.25	4.80	0.89	0.00	0.01	0.96	0.88	0.14	0.001	0.001***	100
Yam 12C	114	0.17	9.00	7.92	2.13	5.20	0.88	0.00	0.08	1.76	0.87	0.07	0.001	0.000***	100
Mean	114	0.18	9.20	8.09	2.13	5.50	0.88	0.19	0.03	1.68	0.87	0.10	0.001		100

AR= allelic richness, GD = gene diversity; Gst and PhiPT= genetic differentiation statistics by locus; I = Shannon's information index; MAF = major allele frequency; N = number of observed samples; Na = number of observed alleles; Ne = effective number of alleles; Nm = estimate of the number of migrants (gene flow) from GST at Nm = 0.25(1 - Gst)/Gst; p = differentiation statistics probabilities; and PIC = polymorphic information content. PHWE= p value for deviation from Hardy Weinberg equilibrium, ns = not significant, * = $p < 0.05$, ** = $p < 0.01$ and ***= $p<0.001$

Table 4

Summary of genetic diversity parameters over the five populations of Dioscorea species genotypes collected from the western parts of Ethiopia.

Population	Size	Na	NaFreq. >= 5%	Ne	I	NPA	NLCA <=50%	HO	He	AR	PPL	%P
WW	46	8.70	6.50	6.44	1.95	0.30	3.10	0.08	0.83	8.20		100
EW	38	7.60	6.90	5.77	1.85	0.10	2.40	0.06	0.82	7.50		100
LV	13	4.40	4.40	3.77	1.35	0.10	0.40	0.01	0.71	4.40		100
BV	10	3.60	3.60	3.08	1.11	0.00	0.70	0.02	0.60	3.60		90
MZ	11	3.80	3.80	3.21	1.21	0.00	0.50	0.00	0.67	3.80		100
Grand mean	23.6	5.62	5.04	4.45	1.49	0.10	1.42	0.03	0.73	5.50		98
Landrace mean	31.67	6.70	5.73	5.14	1.67	0.13	2.00	0.05	0.77	6.50		100
ReleasedV mean	11.5	4	4	3.43	1.23	0.05	0.55	0.015	0.66	4		95

AR= allelic richness, BV= Bulcha variety, EW= East Wollega, He = expected heterozygosity, Ho = observed heterozygosity, LV= Lalo variety, MZ= Metekel zone, Na = No. of different alleles, Na (Freq >= 5%) = No. of different alleles with a frequency >= 5%, Ne = No. of effective alleles, I = Shannon's information index, NLCA= No. of locally common alleles (Freq. >= 5%), NPA= number of private alleles, %p= percent of polymorphic loci

Table 5

Pairwise population matrix of Nei’s standard genetic distance (below diagonal) and gene flow (Nm) (above diagonal) values for landraces and released yam genotypes.

Pop	WW	EW	LV	BV	MZ
WW	****	6.08	2.35	1.45	2.09
EW	0.37	****	2.14	1.25	1.94
LV	0.82	0.90	****	1.74	2.09
BV	0.98	1.21	0.65	****	1.14
MZ	1.00	1.04	0.79	1.20	****

WW= West Wollega; EW= East Wollega; LV=Lalo Variety; MZ=Metekel Zone

Table 6

Pairwise population genetic differentiation (PhiPT) (below the diagonal) between the five populations of Dioscorea species with p values above the diagonal.

Pop	WW	EW	LV	BV	MZ
WW	***	0.001	0.001	0.001	0.001
EW	0.04	***	0.001	0.001	0.001
LV	0.10	0.11	***	0.001	0.001
BV	0.15	0.17	0.13	***	0.001
MZ	0.11	0.11	0.11	0.18	***

WW= West Wollega; EW= East Wollega; LV=Lalo Variety; MZ=Metekel Zone

Table 7

Analysis of molecular variance (AMOVA) of 118 Dioscorea species genotypes based on areas of collection and breeding materials and genetic differentiation

Criteria	Source	Df	SS	MS	Est. Var.	Variation (%)	Genetic differentiation	P Value
Collection areas	Among Pops	4	205.65	51.41	1.68	9%	PhiPT=0.10	0.001
	Within pops	113	1807.93	16.00	16.00	91%	Fis=0.94	0.001
	Total	117	2013.59		17.68	100%		
Gene flow	2.39							
Landrace and released variety	Among Pops	1	57.32	57.32	1.09	6%	PhiPT=0.06	0.001
	Within pops	116	1956.26	16.86	16.86	94%	Fis 0.94	0.001
	Total	117	2013.59		17.96	100%		
Gene flow	3.86							

df: degree of freedom, Fis= coefficient of inbreeding, ms: mean of square, PhiPT= the coefficient of genetic differentiation, calculated using formula PhiPT = AP/(WP + AP), where AP indicates = the variance among populations and WP = the variance within populations, Nm =gene flow, calculated using formula = [(1/PhiPT) - 1]/4, ss: sum of squares.

Figures

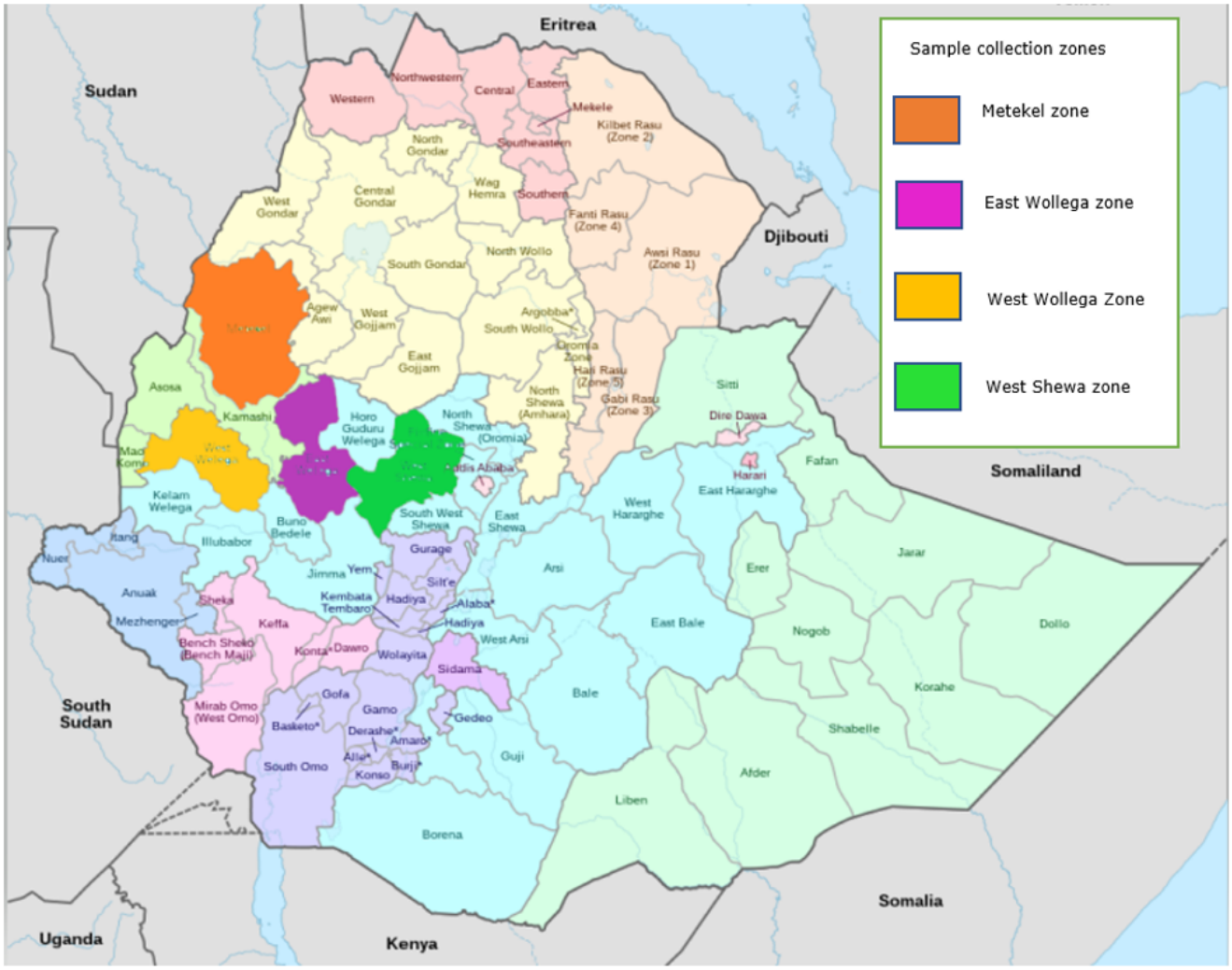
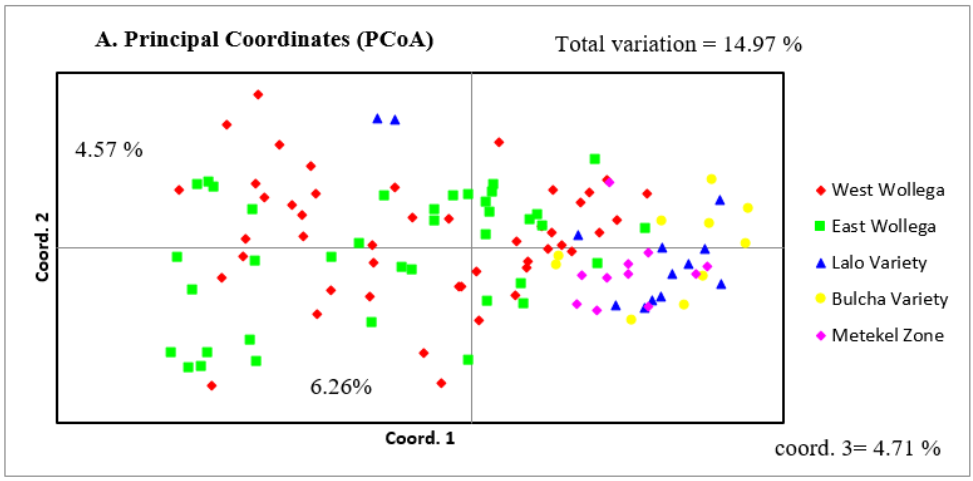


Figure 1

Map of Ethiopia indicating sample collection areas from four zones. The colors indicate the zones of sample collection (West Shewa, West Wollega, East Wollega and Metekel). The map was constructed using the global positioning system (GPS) of the geographic coordinates and elevation of the sample collection site.



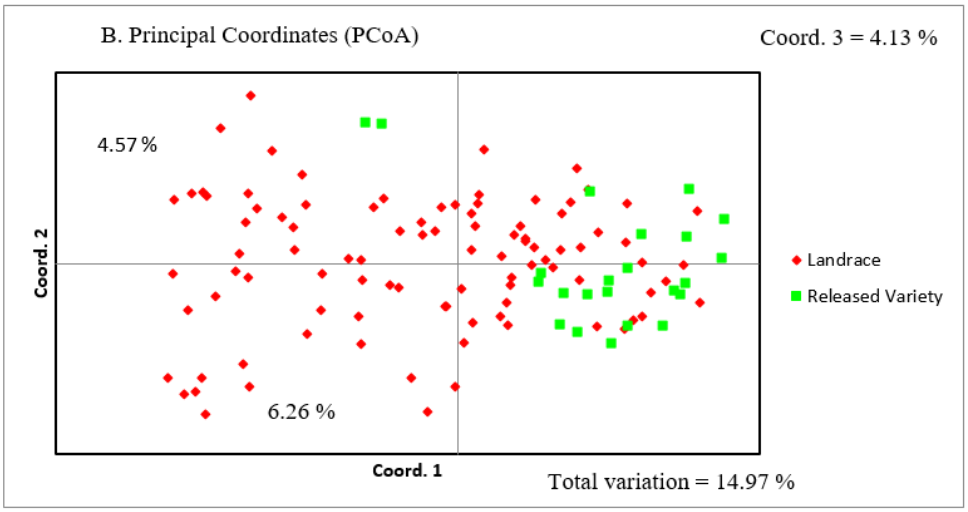


Figure 2

Principal coordinates analysis (PCoA) plot showing the clustering pattern of 118 genotypes of yam using 10 SSR markers. A) Based on collection sites B) Based on breeding materials (Landrace and Released varieties). Samples coded with the same color belong to the same population.

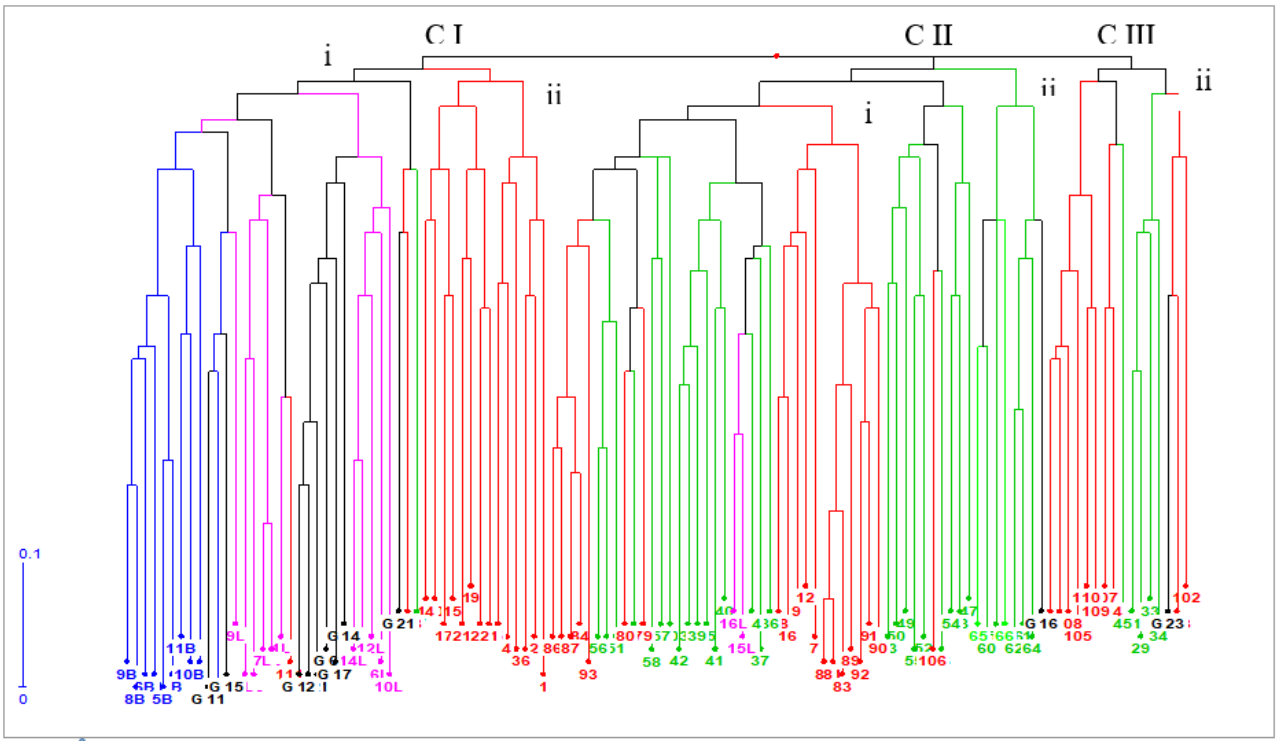


Figure 3

Dendrogram for 118 yam genotypes collected from the western parts of Ethiopia based on unweighted neighbor-joining. Clustering based on collection sites of yam genotypes. C I, C II, and C III indicate the major clusters. i, ii, and iii indicate the subclusters. The color of numbers shows the five populations: West Wollega (Red), East Wollega (Green), Lalo variety (Pink), Bulcha variety (Blue), and Metekel (black).

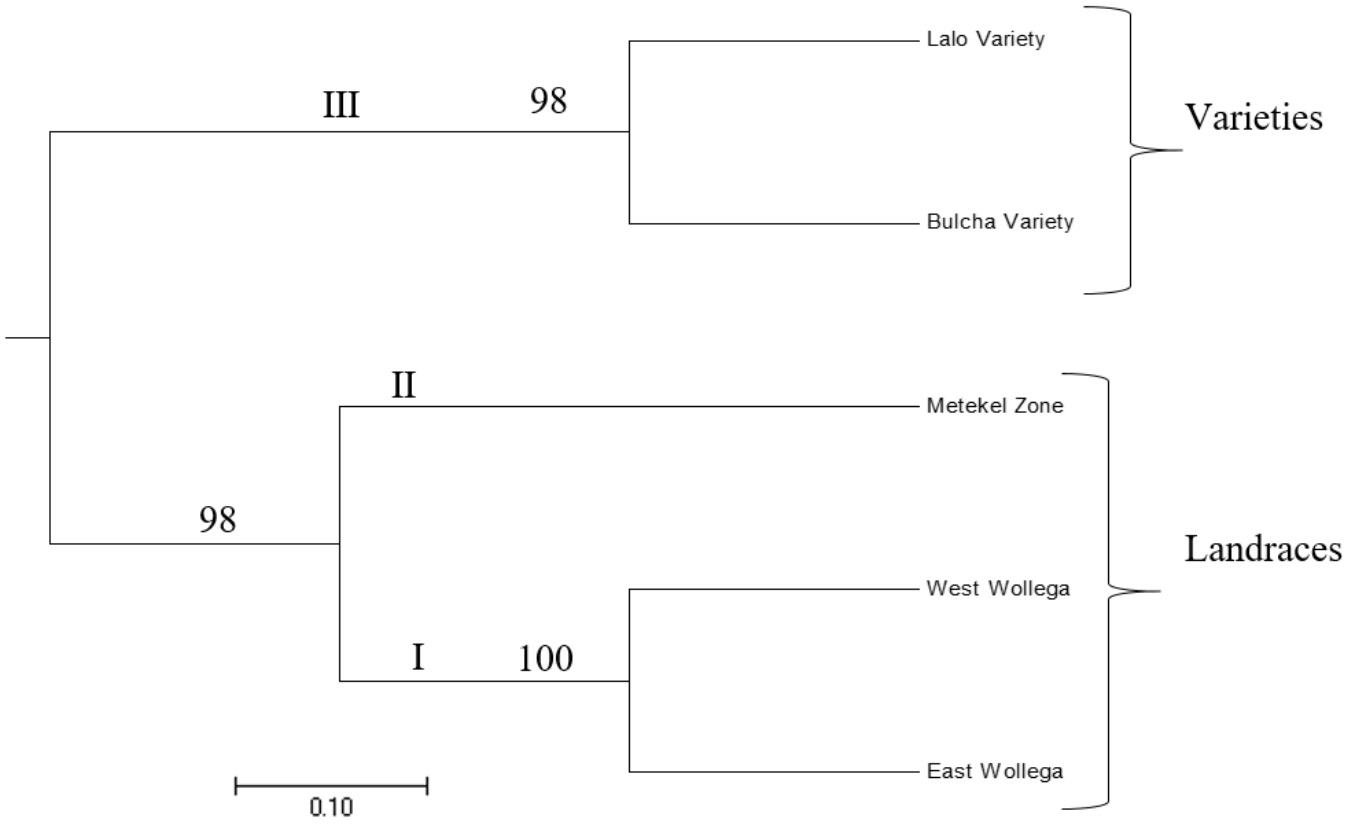
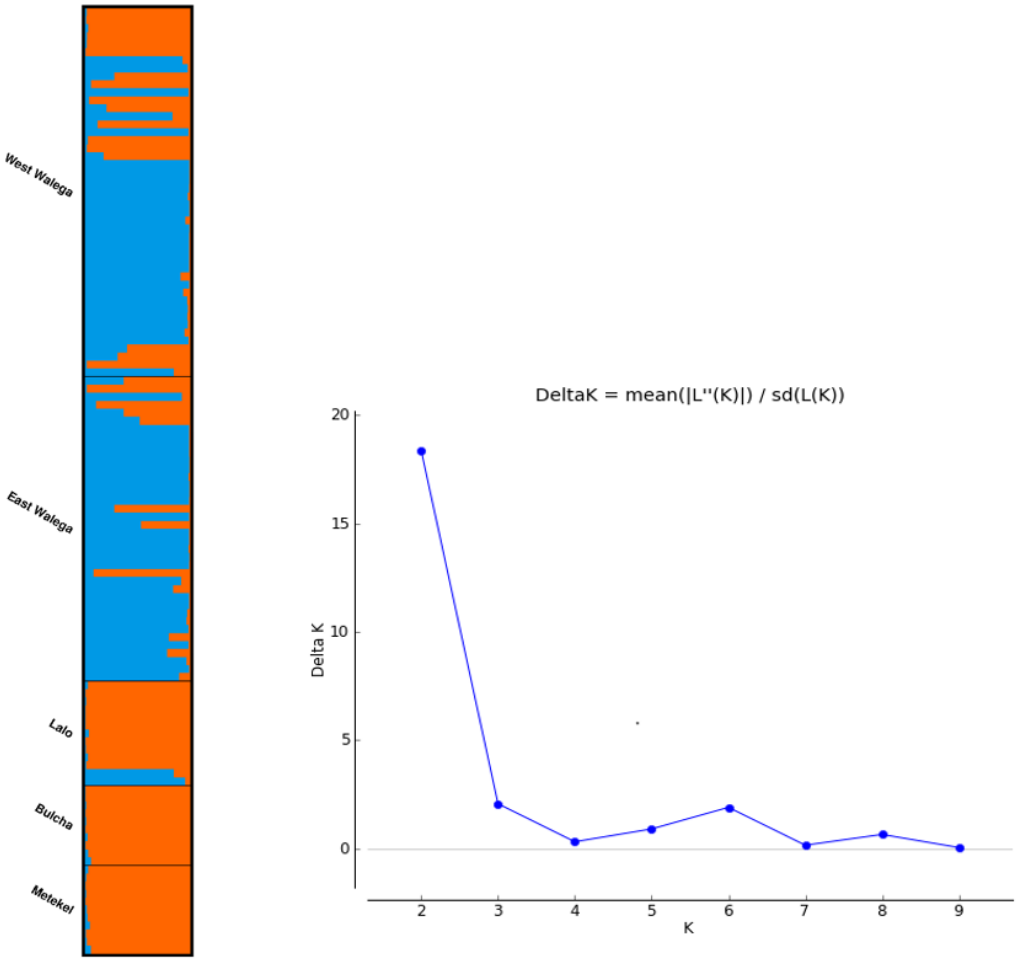


Figure 4

Unweighted pair-group method with arithmetic mean (UPGMA) dendrogram showing genetic relationships among the five populations of yam from western Ethiopia based on Nei's 1972 genetic distance over 1000 replicates. Numbers indicate the percentage of bootstrap replicates.



A). Population structure of yam with k =2 B). Plot of K against delta K

Figure 5

Population structure of the five yam populations obtained from the STRUCTURE analysis, A) population structure of yam with K = 2 (shown on the left) and B) the plot of K against delta K (shown on the right). The different (blue and orange in 5A) columns represent genetic groups or subpopulations designated by Structure Harvester: the x-axis represents the proportion of ancestry to each cluster, and the y-axis represents individual samples.

Supplementary Files

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

- Sup1.118SpecimenvoucherinformationfromNationalHerbariumofEthiopia.pdf
- Sup.2DNAbands.docx
- Sup.3Samplecollectionphoto.docx