ADDIS ABABA UNIVERSITY

INSTITUTE OF BIOTECHNOLOGY

GENETIC DIVERSITY AND POPULATION STRUCTURE ANALYSIS OF NOUG (*GUIZOTIA ABYSSINICA)* ACCESSIONS COLLETED FROM ETHIOPIA USING MICROSATELLITE MARKERS

MSc. Thesis

Submitted to the Institute of Biotechnology, Addis Ababa University for the Partial Fulfillment of the Requirements of Masters of Science in Biotechnology

By

Motbaynor Terefe

Addis Ababa, Ethiopia

June, 2021

ADDIS ABABA UNIVERSITY

INSTITUTE OF BIOTECHNOLOGY

Genetic Diversity and Population Structure Analysis of Noug (*Guizotia abyssinica*) Accessions Collected from Ethiopia Using Microsatellite Markers

BY

Motbaynor Terefe

Submitted to the Institute of Biotechnology, Addis Ababa University for the Partial Fulfillment of the Requirements of Masters of Science in Biotechnology

Addis Ababa, Ethiopia

June, 2021

INSTITUTE OF BIOTECHNOLOGY

MSc. THESIS APPROVAL SHEET

We certify that **Mr. Motbaynor Terefe's** MSc. Thesis entitled as **"Genetic Diversity and Population Structure Analysis of Noug (*Guizotia abyssinica*) Accessions Collected from Ethiopia Using Microsatellite Markers"** has been carried out under our supervision. Therefore, we kindly request the Institute of Biotechnology of Addis Ababa University to final approval and acceptance of the thesis.

Major advisor	Signature	Date
Genet Birmeta (PhD, Asst. Prof.), IoB, AAU	_____	_____
Co-advisors		
Mulatu Geleta (PhD, Assoc. Prof.), Alnrap, Sweden	_____	_____
Kassahun Tesfaye (PhD, Assoc. Prof.), IoB, AAU	_____	_____
Dejene Girma (PhD), NABRC, EIAR	_____	_____

This is to certify that the thesis prepared by **Motbaynor Terefe** entitled as "**Genetic Diversity and Population Structure Analysis of Noug (*Guizotia abyssinica*) Accessions Collected from Ethiopia Using Microsatellite Markers**" and submitted as partial fulfillment of the requirements for the degree of masters of Science in Biotechnology complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the examining committee:

Teklehaimanot Hailesilassie (PhD, Associ. Prof.)	_____	_____
Name of Examiner	Signature	Date
Demsachew Guadie (PhD)	_____	_____
Name of Examiner	Signature	Date
Tesfaye Sisay (PhD, Associ. Prof.)	_____	_____
Director, Institute of Biotechnology	Signature	Date

Genetic diversity and Population Structure Analysis of Noug (*Guizotia abyssinica)* accessions Collected from Ethiopia using microsatellite markers

Motbaynor Terefe, Addis Ababa University, Institute of Biotechnology, June, 2021

Abstract

Noug (*Guizotia abyssinica*) is an Ethiopian indigenous oilseed crop cultivated primarily for its oil and various economic importances. Its production and productivity is very low due to different constraints. The crop has received a little attention in implementing improvement programs and until very recently remained as underutilized oilseed crop. Crop improvement requires the availability and understanding of the genetic resource in terms of genetic variation and population structure. However, estimation of genetic diversity of Noug with a wider representation of Ethiopia's Noug growing regions using reliable molecular marker such as, microsatellite markers are rare. Thus, this study was aimed to uncover the extent of genetic diversity and population structure of 161 Noug genotypes collected from Ethiopia using 13 microsatellite markers. The analysis revealed a total of 158 alleles generated with a mean of 12.15 per locus. All the 13 microsatellite markers were polymorphic and highly informative with a mean polymorphic information content (PIC) value of 0.82. The genetic diversity indices across the 14 collection regions indicated the presence of higher genetic variability among genotypes; including the Shannon's information index (I) which ranged from 0.89 to 1.26 with a mean value of 1.57. Observed heterozygosity (Ho), expected heterozygosity or gene diversity (He), unbiased expected heterozygosity (uHe) was 0.17, 0.74, and 0.79, respectively. Genotypes from Shewa, Wollo, Gojjam, Tigray and Benishangul-Gumuz (B/G), shared the highest in both expected hetrozygosity or gene diversity and unbiased expected heterozygosity. Individuals in the populations had a fixation index (F) ranging from 0.67 to 0.80, with a mean of 0.77. The percentage of polymorphic loci (%Pl) in all the populations showed the highest (100%) and lowest (94.33%) with a mean of 99.19%. The pairwise Nei's genetic distance showed the highest between Eritrea and Bale (0.892) followed by Eritrea and Illubabor (0.875). AMOVA revealed that 96.06% of the total genetic variation was attributed to within populations while only 3.94% was attributed to among populations, indicating that the highest variation was between genotypes rather than variation due to geographic distances. The NJ or UPGMA cluster and PCOA poorly grouped the genotypes based on their origin indicating high gene flow or seed exchange across geographic regions. Furthermore, the model-based population structure analysis weakly inferred the predefined populations into K=2 with considerable genetic admixture among tested populations. In conclusion, the microsatellites used in the present study are highly informative and could be used in future marker assisted breeding. Genotypes from Shewa, Wollo, Gojjam, Tigray and B/G are identified as hotspot areas for their high genetic diversity. They also showed the highest number of private alleles in which special focus should be given in these areas to implement breeding and future improvement programs.

Key words/Phrases; Cluster analysis, Gene flow, Heterozygosity, hotspots, SSR Markers

ACKNOWLEDGMENT

First and foremost I would like to thank the almighty God with his mother St. Mary who gave me blessings, courage, patience, and strength in my entire life.

My heartfelt thank goes to my advisors; Dr. Genet Birmeta for accepting me as a student and giving critical comments, continuous support, and advises starting from proposal development and critical comments on the draft write up. It is my pleasure to express my thanks to Dr. Mulatu Geleta for accepting me as a student and for sending primers and PCR reagents, without his help this work would not have been possible. My deepest thank goes to Dr. Dejene Girma for his immense advices constructive suggestions and support during my lab work and writing up. I am also thankful to Dr. Kassahun Tesfaye for his valuable suggestions and critical comments. I felt honorable for advice by you all!.

I would like to acknowledge the following institutions; Ethiopian Institute of Agricultural Research (EIAR) and National Agricultural Biotechnology Research Center (NABRC) for allowing me study leave and for the laboratory facilities. I am grateful to the Addis Ababa University, Institute of Biotechnology for giving me the chance to join the institute for the study and the financial supports. My acknowledgement also goes to the Ethiopian Biodiversity Institute (EBI) for acquisition of accessions and for their nice hospitality.

I extend my deepest thankful to all my instructors at Institute of Biotechnology (IoB, AAU), specially Dr. Tileye Feyissa for his positive thought and for unreserved perusing of knowledge and skills.

My heartfelt gratitude goes to my colleagues Obssi Dessalegn, Kalkidane Tesfu, and Yitaseb Nigussie for their helps during my lab work.

My appreciation would be to my family and friends for the love and encouragements, also for their indirect contribution to my success. My final appreciation, acknowledgment and thanks goes to individuals whose names are not listed for their direct or indirect support during my study and research work.

Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

ABBERIVIATIONS

AFLP	Amplified fragment length polymorphism
AMOVA	Analysis of molecular variance
Bp	Base pair
DaRT	Diversity array technology DNA extraction protocol
Dntp	deoxy Nucleotide Tri-Phosphate
EBI	Ethiopian biodiversity institute
NABRC	National Agricultural Biotechnology Rsearch Center
EDTA	ethylene Diamine Tetra Acetic acid
ISSR	Interspecific sequence repeat
Masl	Meter above sea level
PCoA	Principal Component analysis
PIC	Polymorphic information content
RAPD	Random amplified polymorphic DNA
RFLP	Restriction fragment length polymorphism
SSR	Simple sequence repeat
TE	Tris-EDTA buffer
UPGMA	Unweight Pair Group Method with Arithmetic Mean
NJ	neighbor joining

1. INTRODUCTION

Noug (*Guizotia abyssinica*) is an edible oilseed crop that belongs to the *Compositaae* family and genus *Guizotia* (Baagoe, 1974). The genus has six species, five of which are originated in Ethiopia. All the species including Noug (*Guizotia abyssinica)* are diploids and comprise of 2n=2x=30 chromosome number (Murthy *et al*., 1993; Kifle Dagne, 1994). *G. abyssinica* is the only cultivated species in the genus (Kifle Dagne, 2001).

Noug is an annual crop cultivated mainly in Ethiopia, India, and Eritrea and in small scale in different parts of the world including Sudan, Uganda, Zaire, Tanzania, Malawi, Zimbabwe, Nepal, USA, Bangladesh, Myanmar and Bhutan (Weiss, 1983; Getinet Alemaw and Sharma, 1996). In Ethiopia it accounts for about 2.03% (257,950.4 hectares) national gross grain productions (CSA, 2019). In Ethiopia, it's mainly grown from 1600 masl to 2500 masl, where the average temperature ranges from 15°C to 23 °C and annual rainfall ranges from 500 mm to 1000mm (Getinet Alemaw and Sherma, 1996). Major Noug cultivation and production regions of the country are Gojjam, Gondor, Shoa, Wellega, Wello, Harerge, Arsi and Bale (Getinet Alemaw and Nigussie Alemayehu, 1997). The crop is grown usually as inter-cropped or rotated with tef and maize to suppress weed and sometimes as a sole crop (Hiruy Belayneh, 1987). It is also cultivated as a border for tef and sorghum crops (Mulatu Geleta *et al*., 2002).

Noug has several economic importances. The Noug seed is mainly used as a human food. It constitutes about 50 - 60% of Ethiopian (MoA, 2016) and 3% of Indian oilseed production (Getinet Alemaw and Sharma, 1996). The oil content of the Ethiopian Noug-seed is between 27- 56% and nutritionally essential fatty acid of

linoleic and oleic acid composition ranges from 54-85% and5-13%, respectively (Kilfe Dagne and Jonsson, 1997; Mulatu Geleta *et al*.,2011). The oil is used to protect against cardiovascular disorder, cancer and to treat burns (Abarsh *et al*., 2014), and also for cultural and medicinal value in Ethiopia (Mulatu Geleta *et al.*, 2002). The seed is warmed in a kettle over an open fire, crushed with a pestle in a mortar, and then mixed with crushed pulse seeds to prepare 'stew' and 'Chibto' (Tig/Amh) and 'litlit' (Amh). In Ethiopia, are also prepared from crushed Noug seed mixed with roasted cereals and is the preferred food for young boys (Getinet Alemaw and Sharma, 1996). The pressed cake left after oil extraction, used as the main livestock feed source containing 17-20% protein, 34-40% carbohydrate sources, and 13.5% fiber (Kandel and Porter, 2002).

Despite the crop's immense economic importance and cultivation in many locations, national production and productivity are fairly low in Ethiopia, contributing approximately 11.49 qt/ha (CSA, 2019). For this, different production constraints are reported. These can be categorized as (i) indeterminate growth habit, shattering, lodging, and self-incompatibility (Mulatu Geleta *et al*., 2007), (ii) important weed [(Dodder (*Cuscuta campestris*)], pests [Niger blight (*Alternaria* sp.), leaf spot)], and insects [Niger fly (*Dioxyna sororcula*, black pollen beetles (*Meligethes* spp.)] (Getinet Alemaw and Sharma, 1996), (iii) Modern breeding and enhancement for desired traits are lucking. So far, only five improved varieties under production namely (Ginchi-1, Kuyu, Fogera, Esete-1, and Shambu-1) are released in different times by Holeta Agricultural Research Center (MoA, 2016). This implies that the crop is given little attention and remained underutilized (Dempewolf *et al*., 2015). Further genetic improvement of the crop is also hindered by the semi-domestic nature (Dempewolf *et*

al., 2015) and the limited availability of genetic information. Hence, the quantification of genetic diversity will allow precise parental line selection for Noug modern breeding and improvement program to control the constraints the crop faced and for its proper conservation.

The extent of genetic diversity and pattern of distribution of Noug at the morphological, biochemical, and molecular level have been reported by different scholars. The variation in oil and fatty acid composition among and between taxa was determined by Kifle Dagne, and Jonsson, (1997) and obtained significant differences in oil content and almost similar in their fatty acid (linoleic acid, oleic acid, palmitic acid, and stearic acid) composition. Mulatu Geleta *et al*. (2011) reported higher oil content and oleic acid variation among 153 Ethiopian Noug accessions and positively correlated with an altitudinal variation. Determination of Noug genetic diversity using morphological marker has been conducted and high genetic variability was observed among tested Ethiopian and Indian genotypes (Adefris Teklewold and Adugna Wakjira, 2004; Paramesh Warappa *et al.*, 2009; Yadav *et al.*, 2012; Pulate *et al.*, 2013; Adarsh *et al.*, 2014; Goyal and Bisen, 2017). Mulatu Geleta *et al*. (2007) obtained higher genetic variation within population than among population of 70 Ethiopian Noug accessions using RAPD marker. Also relatively high genetic variations were observed among 18 Indian Noug genotypes using RAPD marker (Nagella *et al.,* 2008). Mulatu Geleta *et al*., (2008) analyzed the genetic diversity of seventeen Ethiopian Noug populations using AFLP marker, and observed higher genetic variation within populations. Studies were also reported on the extent and genetic diversity of Noug using ISSR marker in Ethiopia and India (Yohanes Petrose *et al*., 2008; Hussain *et al*., 2015). Few studies measured the genetic diversity of Noug

using SSR markers. Noug and its wild relatives were revealed using SSR and morphological markers (Dempewolf *et al.,* 2015). Another study of genetic diversity using SSR marker between 65 Noug accessions showed significance variation (Abebaw Misganaw and Solomon Abera, 2017), and also Birhanu Mengistu *et al*. (2019; 2020) reported considerable genetic variation in 100 Noug accessions using morphological and SSR markers. However, their samplings in these studies were not complete and exhaustive.

Ethiopia is considered as center of origin for Noug, it's also believed to be center of genetic diversity. The Ethiopian Biodiversity Institute (EBI) conserved more than one thousand four hundred (1400) Noug accessions collected from different growing regions (Personnel communication). Genetic diversity information on all the conserved accessions needed to be studied. The development, utilization and conservation of genetic resources of Noug depend on the understanding of its genetic diversity. Thus, genetic diversity study of Ethiopian Noug populations representing Noug growing areas using reliable and co-dominant molecular markers such as microsatellite marker is limited. Genetic diversity study using this molecular marker and determining its population structure helps to generate genetic information for future breeding program and conservation (Glaszmann *et al*., 2010). Hence, this study aims to assess the genetic diversity and to determine the population structure of Ethiopian Noug accessions collected from different growing regions using microsatellite markers.

General Objective

To study the extent of genetic diversity, population structure and pattern of population distribution of Ethiopian Noug accessions using microsatellite markers in order to generate basic information for future improvement and conservation.

Specific Objectives

- To evaluate the genetic diversity of Ethiopian Noug accessions using microsatellite markers.
- To determine the patterns of population structure in Ethiopian Noug accessions.

2. LITRATURE REVIEW

2.1. Origin and distribution of Noug (*Guizotia abyssinica*)

Noug is native to tropical Africa, specifically to the highlands of Ethiopia (Baagoe, 1974). Ethiopia is thought to be the center of genetic variation due to evidence that five of the six species in the genus Guizotia, with the exception of G. reptans, are found therein, as well as the presence of wild related populations (Kifle Dagne, 1994). Out of the six spp, *G. abyssinica* is the only cultivated one.

The exact time of its domestication and cultivation is not known. *G. abyssinica* is believed to evolve from the progenitor *G. scabra ssp. Schimperi* a common weed in Ethiopia through selection and cultivation (Baagøe, 1974; Murthy; Abebe *et al*, 1992; Hiremath and Salimath, 1993; Kifle Dagne, 1994) and it is assumed that its domestication started around 3,000 BC in Ethiopia (Hiremath and Murthy, 1988). Dempewolf *et al*. (2015) reported that Noug is still a semi-domesticated crop due to two probable reasons; the first is the earlier farmers may have selected for resilience to episodic drought or unintended environment, or maybe it's an outcrossing mating system. It's considered the slowest domesticated oil crop than its close relative, sunflower.

Noug is believed to be taken to India by Ethiopian immigrants thousands of years ago and gradually distributed to the rest of the world (Getinet Alemaw and Sharma., 1996). It's primarily grown in Ethiopia, India, and Eritrea (Dempewolf, *et al*., 2015). In Ethiopia, a diverse range of areas are important for Noug growth and cultivation. Its favorite temperature is range from 15-25 ^{0}C and grows at an altitude of 1600-2200 msl with annual rainfall of 500-1000 mm (Getinet Alemaw and Sharma, 1996; Getinet Alemaw and Nigussie Alemayehu, 1997).

2.2. Biology

Noug is a diploid (2n=2x = 30) edible oil crop extensively cultivated in Ethiopia (Kifle Dagne, 1995). The Noug plant is a stout, erect annual herb that grows up to a height of 2 m (Bulcha Woyessa, 2007). The root system is well developed with a tape root; especially in the upper 5 cm it has many lateral roots. The stems are soft, hairy, hallow, and branched, with up to 2 m diameters. Their color is pale green, often stained or dotted with purple, and become yellow with age. The leaves are opposite often with alternate at the apices of the stem. The leaf blade is lanceolate to obviate with 3-23 cm x 1-6 cm variable in shape, with a margin that is entire or toothed, ciliate, softly hairy on both surfaces. The leaves are normally dark green, but they have a distinct yellow color in the lower ones. The inflorescences are arranged up to 3 cm long in apical or axillary cymes. The flowers are capitula, varying in diameter from 15-50 mm, bright yellow, becoming golden yellow when maturing. Each flower produces about 50 seeds (Bulcha Woyessa, 2007). The seeds are small up to 3-6mm long x 1.5-4mm broad, glossy black color (Fig.1). The hermaphroditic disk florets are often up to 40-60 per capitulum arranged at the edge opening first, followed by progressively by the next in line to the center of the head like the sunflower. Flower anthesis begins early in the morning at around 5:30 am and pollen dehiscence begins about 2 hours later and continuous up to 10:00 am under optimum growing conditions (Adefris Teklewold and Nigussie Alemayehu, 2002).

Figure 1. The seed morphology of Noug (*Guizotia abyssinica*) collected from different areas in Ethiopia, where, 1= West Shewa, 2= Jimma, 3=West Gojjam, 4= Bale, 5= West Harerge, and 6= Metekel.

Noug is a completely outcrossing species with a property of self-incompatibility (Murthy *et al*., 1993). The self-incompatibility of *G. abyssinica* was further confirmed by the study carried out by Sileshi Nemomissa *et al*. (1999). They identified that the incompatibility is a sporophytic type, resulting in pollen germination inhibition and/or pollen tube twisting over the surface of papillae. Within in the Ethiopian gene pool, self-compatible Noug genotypes have been identified in very low frequencies, which can be only up to 5% in few populations (Getinet Alemaw and Sharma, 1996; Sileshi Nemomissa *et al*., 1999). Insects, particularly bees, are the major pollinators of Noug (Adefris Teklewold and Nigussie Alemayehu, 2002).

2.3. Taxonomy

G. abyssinica; named '*Noug'* in Amharic and '*Nigerseed'* in English is found in the family *compositae* (*Asteraceae*), tribe *Heliantheae*, sub-tribe *Coreopsidinae*. The taxonomic classification was given using morphological descriptors by Baggoe

(1974). According to this classification; the genus *Guizotia* was sub-divided into six species namely; *G. abyssinica (*L.f*)* Cass. *G. arborescenes, G. scarbia (vis) schimpero.spp. G. villosa, G.zavattarii* and *G. jacksonii.* And other additional taxa called 'Chelelu' and 'Ketcha' within the genus *G. abyssinica* included in the species (Kifle Dagne, 1994). The taxonomical relationship is further confirmed by self-compatibility test (Kifle Dagne, 1994); cytological evidence (Murthy, 1996; Kifle Dagne, 2001), and based on DNA sequence derived phylogenetic studies (Mulatu Geleta *et al*., 2010). The karyotype and chromosomal arrangement studies indicated a relationship within and among taxa (Hiremath and Murthy, 1992; Kifle Dagne, 1994). All the species are a diploid containing 2n=2x=30 chromosome number. Scientific classification of *Guizotia* is described as shown in the table below;-

Table 1. Taxonomic classification of Noug

Kingdom	*Plantea*
Order	*Asterales*
Class	*Magnoliopsida*
Family	*Compositea/Asteracea*/
Tribe	*Heliantheae*
Sub tribe	*Coreopsidinae*
Genus	*Guizotia*
Species	*G. abyssinica (L.f*) Cass.)

Source: (Baggoe, 1974; Kifle Dagne, 1994)

2.4. Uses

Noug is the most economically important oil crop for human consumption. It contributes 50-60% of domestic edible oil requirement of Ethiopia (MoA, 2016) and 3% of the national oil requirement of India (Getinet and Sharma, 1996). The seed

serves as a good source of calories, protein, vitamins, carbohydrates, and other micronutrients providing bio-energy and industrial oil (Vollmann and Rjcan 2009).

Warming, grinding, and mixing with hot water are used to extract the seed oil of Noug in Ethiopian tradition, accompanied by centrifugation in an 'ensera' (a clay made pot). The pale yellow oil settles over the meal after an hour of centrifugation by hand on a smooth soft surface, and is mixed with crushed cereals to make 'stew' (Getinet Alemaw and Sharma., 1996). Noug is also crushed in big oil mills and small cottage expellers. By combining crushed Noug seed with roasted cereals, "Chibito" and "Litlit" are made.

Besides its value in human nutrition, Noug seed oil is used to make soaps, paints, lubricants and illuminants, as well as to clean machinery (Kandel and Porter, 2002) and traditional and medicinal purposes in Ethiopia (Mulatu Geleta *et al*., 2002). It is used to protect gainst cardiovascular disorders, cancer and to treat burns (Abarsh *et al*., 2014). Noug sprouts mixed with garlic and 'tej' are used to treat coughs (Abarsh *et al*., 2014).

The press-cake left after oil extraction serves for poultry and livestock main feed, as it contains 33-37% protein and is high in organic constituents and crude fiber (Kandel and Porter, 2002). In the United States and Europe, Nougseed is used as food for birds especially finches. The fresh plant is consumed by sheep. Noug is used to generate additional income for the farmers and generates foreign currency. Agar medium prepared from Noug can be used to isolate a fungus; *Cryptococcus neoformans* (Sant) which causes a serious brain ailment (Seegeler, 1983; Kandel and Porter, 2002). The whole plant used as foodstuff and green dung (Weiss, 1983). In Ethiopia, the Nougseed is one of the main exporting oilseed crops that bring foreign currency

(Yared Sertse *et al*., 2011). Over all, the crop has high economic importance for human welfare.

2.5. Production and productivity of Noug

2.5.1. Global production and productivity

India, Ethiopia, and Eritrea are the main producers of Noug seed and small scale production is in many countries including, Myanmar, Singapore, Sudan, Uganda, Zaire, Tanzania, Malawi, Zimbabwe, USA, Nepal, Bangladesh (Getinet Alemaw and Sharma, 1996; Hosalli, 2005). India is the leading growing country accounting for more than 50% of global area and production, which is cultivated on 231.1 million ha area and yields 76.17 million tons (Shaikh *et al*., 2019). Ethiopia is the second producer in which from 257,950.4 hectares obtained about 2,963,227.47 quintals which contributes about 2.03% of the national grain total production (CSA, 2019) average productivity of 11.49 q/ha. In the USA, after adaptation trial made to evaluate its agronomic potential in 2000 in Indian State by Jefferson Institute and obtained successful flowering and seed set (Quinn and Myers, 2002). Currently, Noug is grown in different States including, Canada, Ontario, Maryland, California, Virginia, and most east nearest States (Benelli, 2015). It's mainly grown as a bird feed seed (Quinn and Myers, 2002).

This low production in these countries is due to the farming methods, scarce resources, and Noug being grown under rainfed situations on infertile soils. Higher yields can be expected with improved production technology and modern breeding efforts.

2.5.2. Production and productivity in Ethiopia

Noug is the most important oilseed crop and contributes up to 50% of the Ethiopian oilseed crop (MoA, 2016) and 3% of Indian national production (Getinet Alemaw and Sharma, 1996). Major growing areas in Ethiopia are, Gonder, Gojjam, Shewa (Central Ethiopia), Wollega, Wollo, Arsi, Jimma and Tigray (Getinet Alemaw and Sharma, 1996). Farmers in Central highland grow the 'abat' Noug type which is mainly grown in mid-May to early June and harvest in December whereas, the lowland farmers grow the 'Bungne' type which is planted in July and harvested in October. It is the second oilseed crop next to sesame (*S. indicum* L.) in terms of production where from 257,950.4 hectares about 2,963,227.47 quintals were obtained which contributes about 2.03% of the national grain total production (CSA, 2019). It is cultivated, consumed, and sold by several smallholders with high socio-economic importance (Mulatu Geleta *et al*., 2002). The national productivity is very low accounting for about 11.49 q/ha (Fig. 2). The national production and productivity of Noug in Ethiopia are lower than other oilseed crops, including sesame (*Sesamum indicum*). Various factors contribute to its low productivity including shattering, lodging, strict self-incompatibility, indeterminate growth habit and some pests and diseases.

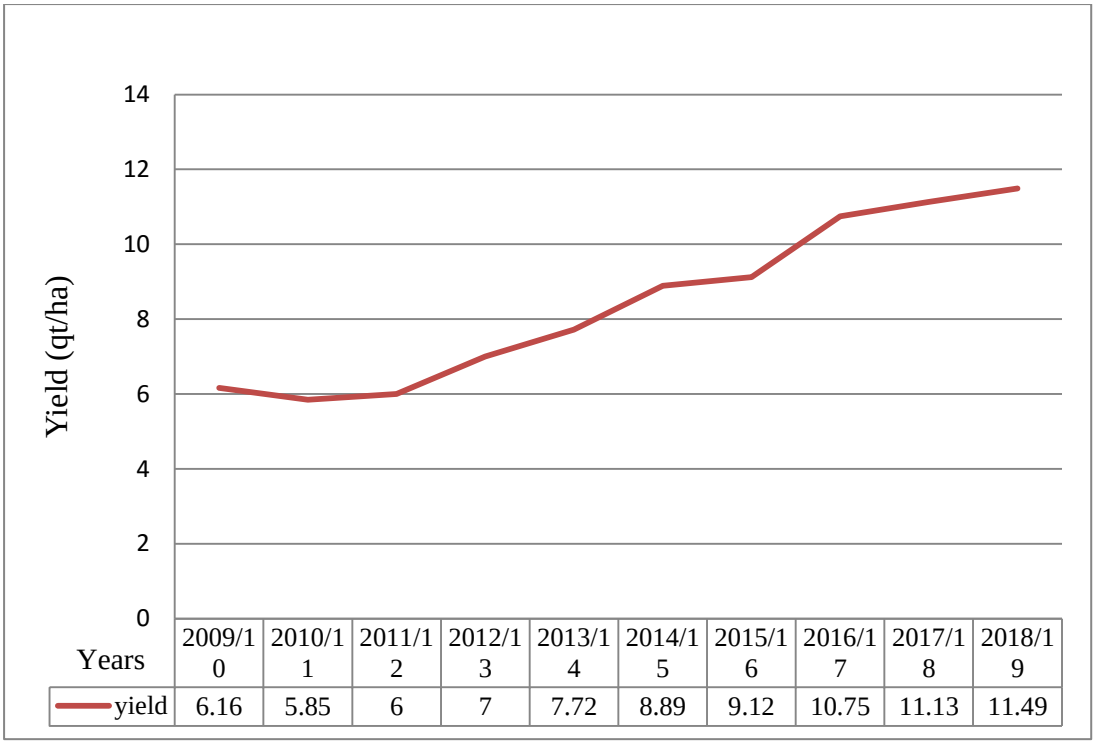


Figure 2. The average productivity of Noug for the last ten years in Ethiopia (qt/ha)

Source; CSA, (2019).

The agronomic practice for Noug is mostly similar to small grains such as sorghum, wheat, and millet (Hiruy Belayneh, 1987). The most used method of sowing is broadcasting or sometimes could be sown in a row. Noug is mainly sown as a sole crop in rotation with tef and maize, but usually sown as an intercrop with cereal crops (Getinet Alemaw Almaw and Sharma, 1996). It is a small-seeded crop and the seed rate varies between 5-10 kg/ha in Ethiopia. Whereas, the fertilizer requirement is 23 kg/ha nitrogen and 23 kg/ha phosphorus for initial growth of the crop as recommended by Hiruy Belayneh and Nigussie Alemayehu, (1987). However, Noug is less responsive for nitrogen and phosphorus fertilizers and has less for weed infestation. Noug usually takes 120-180 days after emergence for maturation and depends on the type and season of sowing. In Ethiopia, the recommended harvesting stage is when the buds turn from yellow to brown or when the moisture content is

between 50-60% (Hiruy Belayneh, 1987). Harvesting at the appropriate stage is crucial for minimizing yield loss due to shattering (Hiruy Belayneh, 1987).

2.5.3. Noug breeding goals

Breeding Noug (research) in Ethiopia started in 1961 at Debrezeit experimental station, Debreziet, and continued at Holeta Research Station, Holeta (Getinet Alemaw and Alemayehu, 1992). The overall breeding goal was to develop high yielding and disease-resistant varieties that are adapted to vertisols, with synchronized flowering and maturity, semi-dwarf stature and thin hull (Getinet Alemaw and Alemayehu Nigussie, 1992) and to increase oil content. So far, only five improved varieties are under production (namely; Ginchi-1, Shambu-1, Kuyu, Fogera, and Este-1) after their release in different years (MoA, 2016). This shows very few varieties were released compared to other newly released oilseed crops in Ethiopia, for example, twenty-six (26) improved varieties of sesame (MoA, 2016).

These Noug varieties were developed using conventional breeding approaches mostly selection. With the availability of molecular techniques, however, the efforts of Noug breeders can be complemented to enhance the efficiency of varietal development and increase genetic gain. However, the genetic elasticity should be primarily determined using highly resolving and efficient molecular markers.

2.6. Production and breeding constraints

Noug is an economically important oil crop and cultivated worldwide. However, its production and productivity is very low. Several constraints in Ethiopia hinder the production and productivity. These include; extremely low harvesting index, lodging, shattering, indeterminate growth habit, self-incompatibility, and a low response to input such as fertilizer. The indeterminate nature of the crop exposes to shattering.

Lodging and shattering are the two most important causes of low seed yield (Getinet Alemaw and Sharma, 1996). Noug has fewer diseases and insect pests as compared to other oilseed crops such as sunflower, sesame, and linseed (Getinet Alemaw and Adefris Teklewold, 1995). However, bacterial leave spots (*Xamthomona* sp.), blights (*Alternaria* sp.) and the Noug fly (*Dioynasorocula* and *Eutretosoma* spp.) which feed flower heads are some major diseases and pests. Also, Black pollen beetles (*Meligethes* spp.) are serious insect pests. In the Noug growing areas of Ethiopia, parasitic weed Dodder (*Cuscusta campastris*) has been reported to be a very serious problem. *Orobanche minor* was also a problem in Ethiopia but serious damage was not reported (Getinet Alemaw and Sharma, 1996).

The strict self-incompatibility nature of Noug (Mulatu Geleta and Ortiz, 2013) makes breeding work difficult and causes a serious problem for inbred line development and maintenance. This is concerning limited efficiency, number, and type of pollinating agents, as well as a limited number of self-incompatibility. The indeterminate growth habit of the crop makes it liable for shattering. Noug should be given the focus in modern breeding programs in Ethiopia to improve its production and productivity.

2.7. Genetic diversity study and roles for crop improvement and conservation

The total heritable variation present within and between populations of organisms is defined as genetic diversity, and it plays an important role in evolution by allowing a species to adapt to a new environment (Kremer, 1998). The genetic variation might arise from evolutionary forces like selection, mutation, migration and genetic drift act continuously and results in continuous changes in allelic frequency in a population

(Slatkin, 1987). Artificial selection favors few alleles at the cost of others resulting in increased frequency of selected alleles (Fu, 2015).

Consequently, domestication reduces genetic diversity as compared to the diversity in wild (Zhang *et al*., 2017). Natural selection also affects genetic diversity considerably (Fu, 2015).

Natural selection chooses the best fit between and within a population, in which there could not be adaptive evolution without genetic variation. Thus, genetic diversity is an essential raw material for evolution, which enables populations of the crop species to survive, adapt to new circumstances, and evolve to produce new genetic variants, where some of them may become the fit variants that meet long-term changes in the environment (Hedrick, 2011).

Estimating the genetic diversity within and among accessions is a prerequisite for further utilization and improvement of the crops or to implement conservation strategies. It is particularly useful in characterizing populations, plant varieties, and species, in detecting duplications of genetic materials in germplasm collections, and for studying the evolutionary ecology of populations. Similarly, genetic diversity is important to meet the primary goals of plant breeding including, breeding for increasing yield, wider adaptation and desirable quality. The greater the genetic diversity within a species, the greater is the species chance for a long-term survival and flourishing (Frankel *et al.,* 1995). Thus, understanding the genetic diversity of plant genetic resources is essential for efficient utilization and conservation of germplasms (Frankel *et al*., 1995). Thereafter, determining the genetic variation and population structure of underutilized crops like, Noug in Ethiopia helps to improve

desirable traits to combat production and breeding constraints and to implement suitable conservation strategies.

Genetic diversity could be revealed by genetic markers that are used to study population, species or individuals and to characterize crops accessions. Traits that serve as genetic markers are by definition polymorphic and the more polymorphic of the trait the greater its potential value to germplasm management (Seetharam *et al.,* 2009). Markers are entities that are heritable as simple Mendelian traits and are easy to score (Schulman *et al.,* 2004). The three types of marker systems are recognized as; morphological, biochemical, and molecular markers systems. Each of these has got different attributes that make it more or less desirable to use in certain applications. Moreover, molecular markers become the first choices for genetic diversity study, quantitative trait loci mapping and marker-assisted selection.

2.8. Measures of genetic diversity

To determine the amount of genetic variability among individuals within as well as between populations, various diversity parameters are used (Chakraborti and Rao, 1991). Gene diversity (H) usually called heterozygosity is one of the measures used to estimate population genetic diversity. It is characterized as the probability of receiving two different alleles at a locus if haploid groups were randomly selected from a population. It was popularized in genetic literature by Nei and Lewontin in the early 70's. Mathematically, written as follow:

$H= 1-\sum Xi^2$,

Where H is the population's genetic variation and xi is the frequency of alleles at the specific locus. The minimum value of gene diversity is zero. The value of gene diversity ranges from 0 to 0.5 and H=0.5 when the frequencies of two alleles are same in a certain loci. The maximum gene diversity will increase with increasing the number of alleles per locus in co-dominant markers, like microsatellites.

The magnitude of gene differentiation among sub-populations measured by G_{ST} equivalent to Wright's F_{ST} as a coefficient of gene differentiation is a method for dividing a population's gene diversity into distinct parts (Nei's, 1973). The genetic differentiation relative to the total population is given by:

$$F_{ST}= (H_T-H_S)/H_T$$

Where, F_{ST} is the proportion of overall genetic variation, H_T is the overall genetic variation of the populations. If genetic diversity is high within a populations but low among others, the FST value will be lower among sub-populations and F_{ST} will be higher when vice versa (Nei, 1973).

The other inter-population genetic diversity measuring parameters are Wright's F-statistics (F_{ST}, F_{IT}, and F_{IS}), Wright's fixation index determination of the dissimilarity coefficient or genetic distance (Wright, 1951). Analysis of molecular variance (AMOVA), cluster analysis, and population structure are also the parameters used to estimate population differentiation and genetic relationship.

Different inter-population and intra-population genetic diversity estimators are important to consider during crop genetic diversity study; including allele number

(Na), major allele frequency (MAF), gene flow (Nm), and polymorphic information content (PIC, estimated as,

PIC= 1- $\sum_{i=1}^{n}$ pi2 $-\sum_{i=1}^{n-1}$ $\sum_{i=1}^{n}$ 2pi2pj2)

Where, n is the number of allele, pi and pj is the allele frequency in the population *i and j*), the richness of allelic variants (A), the effective number of alleles (Ne), observed heterozygosity (Ho), the average expected heterozygosity or Nei's genetic diversity (He), Shannon's information index, equivalent to Shannon-weaver index (I) given by;

$I=\sum pi^2 \ln (pi))$

Where, pi is the frequencies of i^{th} allele for the population, sum pi2 is the sum of squared population allele frequencies), rarified allelic richness (Ar), private allele richness (Apr), and estimate of deviation from Hardy-Weinberg Equilibrium (HWE).

2.9. Applications of molecular markers for diversity study in crops

Molecular markers are based on naturally occurring polymorphisms in DNA sequences due to base-pair deletions, insertions, and substitutions (Gupta *et al*., 1999) or a gene whose phenotypic expression is often effectively recognized and used to identify an individual or as a probe to mark a chromosome, nucleus, or locus and the markers are transmitted by the law of inheritance from one generation to the next (King and Stansfield, 1990; Schulmann, 2004). Various molecular techniques exist which are different in their principles and methodologies (Semagn *et al.*, 2006; Kumar

et al., 2009). They differ from morphological and biochemical markers in that; they are easily available, the assay is rapid, and easy, reproducible and highly polymorphic, co-dominant inheritance and recurrent occurrence in the genome and selectively neutral to environmental conditions. Molecular markers could be either homozygous (dominant marker) or heterozygous (co-dominant markers) based on their ability of detection (Hartl, 1988).

2.9.1. Restriction fragment length polymorphism (RFLP)

Restriction fragment length polymorphism (RFLP) markers are the first molecular markers developed in the first '80s (Botstein *et al*., 1980). It is based on the differential hybridization of cloned DNA to DNA fragments in a sample of restriction enzyme digested DNA individuals of the same species exhibit polymorphism as a result of insertion/deletions, point mutations, translocations, duplications, and inversions. RFLP markers are suitable for linkage mapping analysis, genetic mapping, and genetic fingerprinting.

The RFLP marker system has got certain advantages including the need for no prior sequence information, abundant distribution in the genome of most species, high reproducibility and co-dominance nature (Karp *et al.,* 1997). In contrast, the requirement of a large amount of DNA for restriction enzyme and blotting, characterization of probes and the requirement of radioactive isotopes are some of its drawbacks. These make the analysis expensive and hazardous, time-consuming, labor-intensive assay. A few out of several markers may be polymorphic, which is highly inconvenient especially for crosses between closely related species (Joshi, 1999).

2.9.2. Random amplified polymorphic DNA (RAPD)

Random Amplified Polymorphism (RAPD) markers are the first PCR-based markers and are produced by PCR machines using genomic DNA and random (arbitrary) primers which act as forward and reverse primers in the creation of multiple copies of DNA strands (Jacobson and Hedre`n, 2007). Because of its simplicity, low-cost technique, and no need of prior sequence information, it has found a wide range of applications in many areas of biology. Some applications are; genetic mapping, population, and evolutionary genetics, developing markers linked to the desirable trait, etc. Even though, it has those advantages, the RAPD marker has its limitation which includes; specificity of the marker in genome scanning (Hadrys *et al*., 1992), unclear and non-reproducibility of amplification fragments (Williams *et al*., 1990).

In Noug*;* Mulatu Geleta *et al*., (2007) used the RAPD marker to evaluate the genetic diversity of seventeen Ethiopian Noug populations. Based on this study, they detected the genetic variation within the population 65.58% and among the populations 35.42%. Mulatu Geleta *et al*., (2007) also did a comparative analysis on both RFLP and RAPD markers to examine the genetic diversity of wild and weedy *Guizotia* species and revealed a variation among taxa ranges from 0.22 to 0.28 in RAPD and 0.32 to 0.37 in RFLP markers. The genetic diversity of 18 Indian Noug cultivars was estimated using 17 RAPD primers in which, about 41.2 % of polymorphism was observed which confirmed the existence of high genetic polymorphism in cultivars (Nagella *et al.,* 2008).

2.9.3. Amplified fragment length polymorphism (AFLP)

Amplified Fragment Length Polymorphism (AFLP) method was originally developed as a universal DNA fingerprinting analysis (Vos *et al*., 1995) and is robust and

relatively insensitive to PCR reaction conditions. PCR is done by combining AFLP markers and after digestion of DNA is performed (Lynch and Walsh, 1998). AFLP markers are cost-effective, dominant and there is no need for prior sequence information. In turn, AFLP requires more steps to produce the result, it requires template DNA free of inhibitor compounds that interfere with the restriction enzyme and higher cost of manipulation (Milbourne *et al*., 1997).

Seven AFLP primers combined were applied to 170 Ethiopian Noug individuals to determine genetic diversity (Mulatu Geleta *et al*., 2008). A total of 539 AFLP loci were identified, 90% of which were polymorphic, and all the individuals were genetically distinct (Mulatu Geleta *et al*., 2008). The majority of the heterogeneity was within the population and there was no substantial variation within populations' interms of cultivation and altitude.

2.9.4. Inter-simple sequence repeat (ISSR)

Inter simple sequence repeats (ISSR) are one of the molecular markers amplified by polymerase chain reaction (PCR) having a semi-arbitrary marker in the presence of one complementary primer to target microsatellite repeats. Target DNA segment amplification found at an amplifiable distance in between two identical microsatellite repeat regions oriented in opposite directions each band corresponds to a DNA sequence bordered by two inverted microsatellites (Tsumara *et al.,* 1996). ISSR uses comparatively longer primers than RAPD primers. The product sizes are usually between 200-2000 bp long and could be detected either in agarose or polyacrylamide gels, in which the use of radioactivity is not required. ISSR markers usually show high polymorphism. However, like RAPDs, dominant inheritance and homology of co-migrating amplification products are the main limitations of ISSRs (Semagn, 2006).

Different studies used ISSR markers to determine the genetic variation and population structure of Noug. Yohanes Petrose *et al*. (2008) studied the genetic diversity, and relationship of wild *Guizitia* species of 45 individuals each from five taxa collected from Ethiopia. The result revealed a total of 145 scorable bands generated from five ISSR markers out of these, the polymorphic percentage ranges from 68.2 to 88% between taxa. The genetic distance between taxa ranges from 0.0017 to 0.1754, and between populations ranges from 0.0125-0.0866. The variation within 37 Noug populations collected from major growing areas of Ethiopia was analyzed using inter-simple sequence repeats (ISSRs). The result showed more variability within the populations than among populations (Yohanes Petrose *et al*., 2007). Hussain *et al*. (2015) also reported the genetic diversity assessment of 35 Indian Noug accessions using ISSR markers for oil quality parameters. High genetic variability exhibited by the accessions. The variance component within populations accounted for 95 % while among population variance component accounted for only 5%. The genetic diversity within populations was higher as compared to among the populations. The study proved the existence of much more genetic diversity for oil quality parameters grown in different geographic locations.

2.9.5. Simple sequence repeats (SSR)

Microsatellites or simple sequence repeats (SSR), sometimes called variable number tandem repeats (VNTR) or short tandem repeats (STR) are repeated motifs of nucleotides from 1-6 found at a very high frequency in the genome of almost in all taxa (Beckmann and Weber, 1992). Like, repeats named mono containing ((A)11), di-((GT)12), tri-((ATT)9), tetra-((ATCG)8), penta-((TAATC)6), hexa-((TGTGCA)5 nucleotides. Microsatellite loci are mostly varying in length between 5 and 40 repeats, but also might be of long strings of repeats. The SSR marker's differences within the

microsatellite structure are based on their variable numbers of repeat units. It has got a wide range of applications including uses for detecting differences between and within species, DNA fingerprinting, marker-assisted selection, genetic linkage mapping, studying the structure of populations and determination of hybrids.

The advantages of SSR markers include Mendelian inheritance as a co-dominant marker, high reproducibility and polymorphism at each locus, high abundance and wide distribution throughout the genome, locus-specificity, revealing of high allelic diversity and allowing rapid generation of data from a relatively small amount of plant tissue (Mohan *et al*., 1997). SSR markers have the potential transfer to use in related species. This is possible because the flanking regions are conserved and the number of duplications is variable. Despite their advantage, SSR markers are expensive to generate and very laborious in certain species, and sequence information is required.

SSR markers could be developed from genomic regions or using express sequence tags (ESTs). EST-derived SSRs other than genome-derived SSR markers are advantageous in terms of rapid and cheap development using EST database mining (Koilkonda *et al*., 2012). In recent years, EST derived SSR markers have become the choice of marker for genetic diversity analysis, genotype identification, molecular tagging of genes and marker-assisted breeding (Wei *et al.*, 2011). There are many reports on which EST-derived SSR markers are applied in different crops for genetic diversity estimation, including Sesame (Wei *et al.*, 2011), Mango (Dillon *et al.*, 2014), and Sorghum (Ramu *et al.*, 2013).

Few studies have used SSR markers for molecular characterization of Noug accessions. Dempewolf *et al*. (2015) used 16 SSR primers previously identified by

Dempewolf *et al*. (2010) to analyze the genetic diversity of different Noug accessions and to determine the level of crop-wild gene flow with their pattern of domestication. Except for one population, microsatellite study failed to identify significant recent admixture between noug and its wild progenitor. The genetic variation estimation of 65 Noug accessions using 11 EST-derived SSR markers was investigated (Abebaw and Solomon, 2017). This analysis yielded a total of 66 alleles with a mean of 6 per locus. And 0.69 mean numbers of major allele frequency, 0.42 of gene diversity, and 0.50 of heterozygosity were observed.

Also, Birhanu Mengistu *et al*, (2019; 2020) investigated the genetic diversity of 100 Noug accessions using both morphological and SSR markers. The result showed a major allele frequency of 0.31 and polymorphic information content of 0.78. The genetic distance based on Nie's genetic distance ranges from 0.04-0.79. However, their sampling was not complete and exhaustive representation interms of both number and geographic coverage. Thus, studying the genetic variation of landraces collected from across the country is mandatory to forward comprehensive information that could help in the future breeding and conservation. Therefore, this study aims to infer the genetic diversity and population structure of Ethiopian Noug accessions collections from all over Noug growing areas including some from Eritrean origins using microsatellite markers to generate basic information.

3. MATERIAL AND METHODS

3.1. Plant material

The 161 Noug accessions used in this study were Ethiopian and some from Eritrean origins and were provided by the Ethiopian Biodiversity Institute (EBI). The accessions (Appendix 3) were systematically selected based on area coverage with varying altitudes and represent all growing regions including minor growing areas of the country (Fig. 3). The sampling areas (regions) and the sample size were; Arsi (9), Benishangul-Gumuz (B/G) (19), Bale (9), Tigray (18), Gonder (10), Gojjam (17), Harerge (5), Illubabor (9), Jimma (4), Shewa or Central Ethiopia (21), Wellega (8), Wollo (21), South Nation Nationalities and Peoples (SNNP) (7), and Eritrea (4).

The accessions was collected by the Ethiopian Biodiversity Institute (EBI) when Eritrea was part of Ethiopia (before it became an independent state), but at different years, Ethiopian new collections are added. The population names are assigned based on both the old and new administrative regions, only for the purpose of this study. Some of the new regions are; Benishangul-Gumuz (B/G) contains accessions from Assosa, Metekel, and Kamash zones. South Nation Nationalities and Peoples (SNNP) is a new regional state name, consists of accessions collected only from Gurage and Hadiya zones. Jimma is also one of the new population regions that contain accessions collected from different parts of the areas. The remaining populations are named after the previous administrative locations.

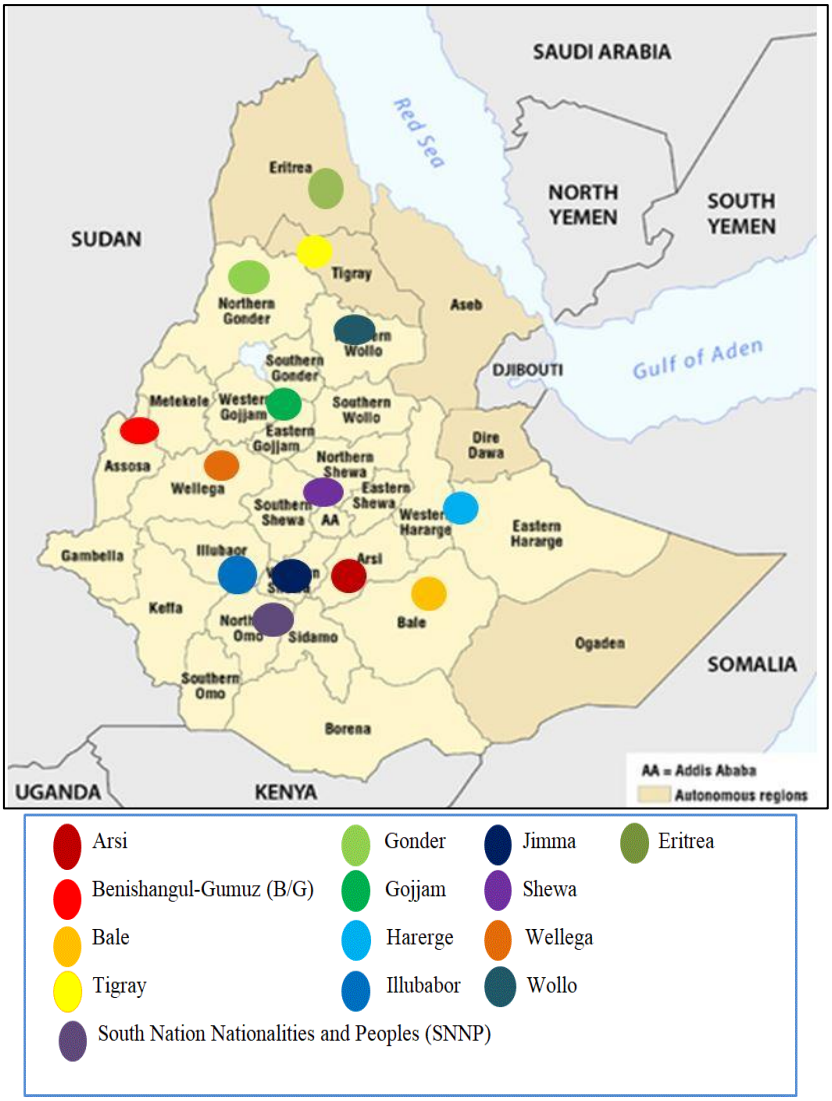


Figure 3. Map showing Noug sample collection regions (NB: all the sample areas, points, and boundaries are approximate and nothing to do with political boundaries).

3.2. Genomic DNA extraction and quality check

The seed of Noug accessions was planted at the National Agricultural Biotechnology Research Center (NABRC), Holeta on seedling pot in the greenhouse for germination. Genomic DNA was extracted from a two-week old seedling leaf sample using plant DNA extraction protocol for DArT (Diversity Array Technology) with some modification. From 15 days old seedling, about 100 mg of fresh leaf samples were harvested and placed in a 2 ml Eppendorf tube and were freeze dried (-80 ^{o}C) overnight. The leaves were further dried in liquid nitrogen and then grounded using

Geno Grinder (MM-200, Retsch) with 25 rpm/s for 3 minutes. Then genomic DNA was extracted according to the procedures in the Appendix 2. After DNA extraction, the concentration of genomic DNA was measured using Nano-drop spectrophotometer (ND-800, 8 sample spectrophotometer). The purity of the DNA was determined by the 260/280 absorbance ratio. The quality of the DNA was also further checked using 1% agarose-electrophoresis runned for 45 minutes at 100 voltages (was primarily mixed about 5 µl of DNA with 2 µl of loading dye). The DNA with good quality and high concentration was used for PCR analysis by diluting with a final concentration of 20 ng/µl.

3.3. Selection of primers and SSR survey

A total of 20 microsatellite primer-pairs (SSR) previously developed by Dempewolf *et al.* (2010) for Noug were used for optimization. From 20 primers tested thirteen (13) were selected and used for final genotyping on the basis of their successful amplification and polymorphism (Table 2).

Table 2. Description of microsatellite markers used for the study

SN	Marker Names	Repeated motifs	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')	Expected size (bp)	Remark
1	GA013	CTT	GGTAATGGTAATGGAGGTTCTGG	CCTCATCAGAGTTCTTCGGGTTAT	424-455	Polymorphic
2	GA018	AGC	GTTCCAGCCCATGAGTCATAAT	CTATCTCTATCTCGTGGGGTTTTG	353-358	Polymorphic
3	GA029	ATC&TC	CCATCATCAATGGCGTTACTC	GTCTCGTTCTAGAAGCTTCATCCT	270-276	Polymorphic
4	GA035	TGA	GATTTCTCAGGTGAAGGAGAAGAG	GCCCTCCCTACAACATACTTGATA	301-307	Polymorphic
5	GA037	TA&GAA	GGTGTTTTTGTGTAGTGGTCTGTC	GACTAGCCAGAAACCGAAGAATC	347-350	Polymorphic
6	GA054	TA	AACGGTTTAGGAGACCTTGG	TCACCTGGCTCAGACTTGTTT	247-265	Polymorphic
7	GA077	TC	TCAGCCAAACATTCCAAAGC	AAACAACGCGCTAAAAACGA	487-490	Polymorphic
8	GA081	TC	AATCTCGATTGGCTGAGTGG	AGGAAGTTGGGGCTTCGTAA	437-441	Polymorphic
9	GA117	CAC	CCCTTCATCCAATTCTAACGAC	AGGTCTAATCCCAGCCTCTCTAAT	336-339	Polymorphic
10	GA139	GAA	GTACATCCCAACTTTACCATCCAC	CTCTACAACCAACACCACTTTCC	223-241	Polymorphic
11	GA144	AGTT	GGTCCCACAAACCAATATGATG	CTAGGGCTTGTACCACACCTTAAA	331-347	Polymorphic
12	GA150	ACC	GTAATGACTTGTGAGGAACACGAC	GGGTTTGGAGGTACAGTGTAAGAT	279-298	Polymorphic
13	GA162	CAA	AGCCACTCTCTTGTTTGTTACC	CAAGTTCTGGTGGGTGGTATG	134-140	Polymorphic
14	GA003	GAT	CGCCCTAAAGCTACTTTCTTCC	CACACTCGCACTAGGA	399-402	Monomorphic
15	GA055	CT	CCTGAAACAAACCCCAACAA	CAGTACATCGCGGAGAGAGG	194-200	Monomorphic
16	GA082	TC	TGTCCGTATGAAACCCATTGA	CAATGATCATGGGGACTGCT	197-197	Monomorphic
17	GA127	CCT	CAATCTGCAACTACTGCCAATACC	CCAGTCAGAACCCTTGATCACTA	213-216	Failed to amplify
18	GA138	AAG	ATCAACTTCCCCATATACCTCTGG	CTTCCTCTGTCACTTCTTTTGGAC	360-378	Failed to amplify
19	GA143	TGA	GGATGGTGTACTTCTTTCTGACCT	TAGCGACGGTAACATACGAGTCT	296-312	failed to amplify
20	GA156	AAG	CCAGTTTGTGAGAATTCACCGTGT	GAGCTCCAGGTCTCTAGGGTTATC	158-173	Monomorphic

The PCR amplification was carried out in a 12.5 µl reaction volume containing 1.25 µl of 10x PCR buffer, 0.75 µl of 1.5 nM MgCl2, 0.25 µl of 200 µM dNTPs mix (each dATP, dTTP, dCTP, dGTP), 0.25 µL of Forward and Reverse primer, 0.2 µL DEMSO, 0.1 µl (2.5 U/µL) of Taq DNA polymerase, 7.45 µl of nuclease-free water and 2 µl of genomic DNA.

The PCR condition was programed to touchdown PCR using BIO-RAD T100 thermal cycler. The program is adjusted into two conditions based on their touchdown temperature of primers; for primers GA018, GA037, and GA144 the touchdown was programmed at an initial denaturation of 94 °C for 4 minutes, denaturation 94 °C for 45 seconds, annealing 60 °C to 50 °C for 1 min and extension 72 °C for 1 min and 30 sec running for 9 cycles (with 1^{o}c decrease in every cycle), and denaturation 94 °C for 45 sec, annealing 50 °C for 1min and extension 72 °C for 1min and 30sec running for 29 cycles, final extension 72 °C for 10 min in and stored at 4^{o}c. Whereas, for the rests of 10 primers the PCR was adjusted at an initial denaturation of 94 °C for 4 minutes, denaturation 94 °C for 45 seconds, 65 °C to 55 °C for 1 min and 72 °C for 1 min and 30 sec running for 9 cycles (with 1^{o}c decrease in every cycle), and 94 °C for 45 sec, 55 °C annealings for 1min and 72 °C for 1min 30 sec running for 29 cycles, 72 °C for 10 min in final extension and stored at 4 °C.

3.4. Gel-electrophoresis

After amplification, the PCR products were fractioned in 3% agarose gel (w/v) using 1xTAE buffer running 2:30 hr at 100 constant voltages. The samples were loaded on the gel with 5 µl of SSR amplification product and 2 µl of 6X loading dye (containing gel red). The molecular weight of each amplified product was estimated by comparing

the DNA bands with 100bp base-pair mixed DNA ladder (SIMOBIO, DM2100) loaded in the peripheral wells. An amplified product was visualized under UV light using BioDoc-IT^{TM} Imaging system (Cambrige, UK) to confirm successful amplification of the PCR products.

3.5. Data scoring and statistical analysis

The fragment sizes detected by each SSR region were estimated in reference to the size marker using PyElph 1.4 (Pavel *et al*., 2012) software package. Fragments with the same mobility were considered identical fragment size and treated as the same allele and bands of different molecular weight were considered as a different allele for a single locus. Different statistical software packages were employed to compute the genetic diversity indices and population structure of 161 Ethiopian Noug genotypes.

SSR marker-based diversity parameters including, major allele frequency (MAF), number of observed alleles (Na), and the polymorphic information content (PIC) were calculated using Power marker v3.25 software (Liu and Muse, 2005). Population diversity indices such as determination of the number of different alleles (Na), the effective number of allele (Ne), observed heterozygosity (Ho), expected heterozygosity (He), unbiased expected heterozygosity/gene diversity (uHe), Shannon's information index or Shannon's diversity index (I), fixation index (F), gene flow (Nm) and percentage of polymorphic loci (%Pl) were computed using GenAlex v 6.5 software (Peakall and Smouse, 2015). Rarified allelic richness (Ar) and private allele richness (Arp) were estimated using the rarefaction procedure implemented in the HP-Rare software package (Kalinowski, 2005).

To estimate population genetic differentiation among and within accessions based on their genetic origin, for this, analysis of molecular variance (AMOVA) was performed

using Arlquin 3.5.2.2 software (Excoffier and Lischer 2015). AMOVA compares the genetic makeup of populations using approximate F-statistics (genetic differentiation (F_{ST}), inbreeding coefficient (F_{IS}), and total fixation index (F_{IT}). The pairwise Nie's standard genetic distance and pairwise population differentiation (F_{ST}) was computed using GenAlex v 6.5 software (Peakall and Smouse, 2015).

Principal component analysis (PCOA) was carried out using GenAlex v 6.5 software (Peakall and Smouse, 2015) to examine and visualize the pattern of variation on two-dimensional plots among 161 Noug genotypes. Cluster analysis was performed by using the unweight neighbor-joining (NJ) algorithm in Darwin 6.0.2.1 (Perrier and Jacquemoud-Collet, 2006). A dendrogram was constructed based on the dissimilarity matrix to see the pattern of the cluster among and within tested genotypes. UPGMA clustering method of populations based on Nei's standard genetic distance was carried using POPTREE2 software (Takezaki and Tamura, 2010).

Population structure was examined using Structure 2.3.1 (Pritchard *et al*., 2000) based on a Bayesian algorithm to infer the possible numbers of sub-groups. The analysis was computed based on 100,000 iterations and 100,000 Markov Chain Monte Carlo (MCMC) burn-in period in 10 independent runs from Delta k=1 to k=12. The Evano *et al*. (2005 approach was used to calculate the amount of optimal Delta K. The web-based software, STRUCTURE HARVESTOR ver 0.9.93 was used for this purpose (Earl and von Holdt, 2012) and StructureSelector (Li and Lui, 2018) was employed to extract the optimum number of sub-groups and the biplot of the optimum Delta K.

4. RESULTS

4.1. SSR markers polymorphism and levels of diversity

In this study, the analysis of 20 SSR markers revealed that 13 markers were polymorphic and effective in discriminating the 161 Noug genotypes, while 4 were monomorphic and 3 of them failed to amplify and excluded from further analysis. The 13 SSR markers detected a total of 158 alleles across the 161 Noug genotypes. The observed number of alleles (N_a) varied from 7 to 17 for locus GA054 and GA144 with a mean of 12.15 per locus (Table 3). The effective number of alleles (N_e) values ranged from 3.48 to 5.68 for locus GA054 and GA029, respectively, with a mean of 4.65 per locus. All the 13 SSR loci were highly polymorphic with the minimum and maximum PIC observed for locus GA054 (0.67) and locus GA144 (0.90) with a mean PIC of 0.82. All the 13 markers had a PIC value of > 0.50, the majority of them fell between 0.80 to 0.90.

The major allele frequency (MAF) was the lowest 0.15 for locus GA035 and the highest 0.46 for locus GA054 with a mean frequency of 0.28. The Shannon's information index (I) across the test genotypes in 13 SSR loci generated a mean value of 1.57 (ranging from 1.32 to 1.82). Similarly, the mean values of observed heterozygosity (Ho=0.17), expected heterozygosity (He=0.74), unbiased expected heterozygosity (uHe=0.79), fixation index (F=0.77) were recorded. Following Nei's (1987), an indirect estimate of the level of gene flow (Nm) based on the number of migrants per generation between the genotypes were calculated using the formula $Nm= 0.25(1-F_{ST}/F_{ST})$. The level of gene flow was higher for locus GA162 (2.69) and lower for locus GA013 (0.92) with a mean gene flow rate Nm=1.79.

Table 3. Genetic diversity indices for 161 genotypes across 13 microsatellite loci

SSR Loci	Diversity Indices									
	MAF	**Na**	**Ne**	**I**	**Ho**	**He**	**uHe**	**F**	**Nm**	**PIC**
GA013	0.26	14	3.82	1.32	0.00	0.67	0.71	0.96	0.92	0.85
GA018	0.36	15	4.26	1.56	0.13	0.74	0.79	0.87	2.08	0.86
GA029	0.24	16	5.68	1.82	0.71	0.79	0.84	0.28	2.36	0.88
GA035	0.15	13	5.38	1.69	0.10	0.78	0.82	0.86	1.85	0.87
GA037	0.19	14	5.52	1.72	0.19	0.77	0.81	0.69	1.77	0.88
GA054	0.46	7	3.48	1.34	0.00	0.68	0.72	0.92	1.46	0.67
GA077	0.24	13	5.08	1.69	0.17	0.77	0.82	0.77	1.71	0.87
GA081	0.31	10	4.49	1.54	0.02	0.74	0.79	0.96	1.80	0.78
GA117	0.28	11	4.31	1.52	0.19	0.74	0.78	0.76	1.84	0.82
GA139	0.33	9	4.26	1.51	0.00	0.75	0.79	0.89	1.73	0.73
GA144	0.17	17	5.66	1.75	0.22	0.78	0.83	0.74	1.76	0.90
GA150	0.27	11	4.39	1.48	0.00	0.71	0.75	0.91	1.34	0.84
GA162	0.42	8	4.17	1.50	0.53	0.74	0.79	0.43	2.69	0.70
Mean	**0.28**	**12.15**	**4.65**	**1.57**	**0.17**	**0.74**	**0.79**	**0.77**	**1.79**	**0.82**
SE	-	-	0.13	0.03	0.02	0.00	0.08	0.03	0.12	-

MAF, major allele frequency; Na, number of observed alleles; Ne, number of effective alleles; I, Shannon's information index or Shannon's diversity index; Ho, observed heterozygosity; He, expected heterozygosity or gene diversity; uHe, unbiased expected heterozygosity; F, fixation index; Nm, gene flow; PIC, polymorphic information content.

4.2. Magnitudes of genetic diversity

The gene diversity parameters for the fourteen studied Noug populations based on their geographic origin are summarized in Table 4. The percentages of polymorphic loci were lower (94.33%) for genotypes collected from Jimma and highest (100%) for those collected from most geographic origins. The highest number of different alleles (Na) was observed in Shewa (8.76) population and the lowest resulted from the Eritrean population (2.85) with an average of 6.05. Similarly, the number of effective alleles (Ne) was highest in the Shewa population and the lowest in the Eritrea population, the average record was 6.46. Allelic richness was recorded higher in Shewa (7.84) and lowest in Eritrea (2.85) with a mean of 5.77 in the overall populations. Additionally, the Shewa populations remarked the highest number of

private alleles 0.56 followed by Wollo, Gojjam, and Benisgangul Gumuz populations (0.36, 0.35, and 0.34, respectively) and there is no private alleles observed in Harerge, Illunabor, Jimma, SNNP, and Eritrea populations (Table 5). On other hand, the highest (7.69) number of different alleles with a frequency of ≥ 5% was found in Gonder and lowest (2.84) in the Eritrea population with a mean value of 5.54.

Table 4. Summary of genetic diversity parameters in14 populations using 13 microsatellite loci

Populations	**Gene diversity parameters**							
	Na	**Ne**	**I**	**Ho**	**He**	**uHe**	**F**	**%Pl**
Arsi	5.69	4.33	1.52	0.23	0.73	0.78	0.71	100.0
B/G	8.30	5.96	1.87	0.15	0.80	0.83	0.82	100.0
Bale	5.30	4.01	1.47	0.14	0.72	0.76	0.81	100.0
Tigray	7.53	5.65	1.79	0.15	0.80	0.82	0.82	100.0
Gonder	6.15	4.87	1.64	0.14	0.77	0.81	0.82	100.0
Gojjam	7.30	5.66	1.82	0.17	0.81	0.83	0.78	100.0
Harerge	4.00	3.38	1.26	0.18	0.68	0.76	0.74	98.21
Illubabor	6.07	4.61	1.63	0.17	0.76	0.81	0.77	100.0
Jimma	3.92	3.48	1.26	0.23	0.68	0.78	0.67	94.33
Shewa	8.76	6.45	1.97	0.16	0.83	0.85	0.80	100.0
Wellega	5.61	4.52	1.56	0.25	0.75	0.80	0.69	100.0
Wollo	8.53	5.76	1.90	0.15	0.81	0.83	0.80	100.0
SNNP	5.00	4.05	1.45	0.18	0.73	0.78	0.75	99.52
Eritrea	2.84	2.39	0.89	0.13	0.53	0.61	0.74	96.70
Mean	**6.07**	**4.65**	**1.57**	**0.17**	**0.74**	**0.79**	**0.77**	**99.19**
SE	0.17	0.133	0.03	0.02	0.00	0.08	0.02	0.00

B/G, Benishangul-Gumuz; SNNP, South Nation Nationalities and Peoples; Na, number of different alleles per population; Ne, effective number of alleles; I, Shannon's information index; Ho, observed heterozygosity; He, expected heterozygosity or gene diversity; uHe, unbiased expected heterozygosity; F, fixation index; %Pl, percentage of polymorphic loci.

The Shewa (Central Ethiopia) population exhibited the highest number of locally common alleles found in ≤25% of populations and the number of locally common alleles in ≤50%of populations.

Genotypes from Shewa (1.97) and Wollo (1.90) showed relatively the highest number of Shannon's information index with an average of 1.57. Likewise, the observed number of heterozygosity (Ho) was higher in Wellega (0.25) while it was least in Eritrea (0.13) population, overall mean (0.17). In Shewa (Central Ethiopia) population both expected heterozygosity (He)or gene diversity and unbiased heterozygosity(uHe) was the highest 0.83 and 0.85, respectively, while the lowest was in Eritrea (He=0.53 and uHe=0.61) population. The mean value of expected heterozygosity (He) and unbiased expected heterozygosity (uHe) in all populations were 0.74 and 0.79, respectively. In all populations, the observed heterozygosity was much lower than expected heterozygosity and unbiased expected heterozygosity. The data analysis revealed that the fixation index or inbreeding coefficient (F) was similar in Benishangul Gumuz, Tigray, and Gonder populations (0.82) which were the highest than others. Whereas, the lowest (F= 0.67) has resulted from the Jimma population.

Table 5. Allelic Patterns across populations

Populations	**Parameters**				
	Ar	**Arp**	**NAF (≥5%)**	**NLCA (≤25%)**	**NLCA (≤50%)**
Arsi	5.69	0.08	5.69	0.23	2.00
Benishangul-Gumuz (B/G)	7.42	0.34	6.15	0.07	2.92
Bale	5.31	0.08	5.30	0.07	2.15
Tigray	6.89	0.20	7.00	0.07	3.30
Gonder	6.15	0.15	7.69	0.07	3.76
Gojjam	6.92	0.35	7.07	0.23	2.76
Harerge	4.00	0.00	4.00	0.07	0.76
Illubabor	5.92	0.00	5.92	0.15	2.38
Jimma	3.85	0.00	3.84	0.00	1.23
Shewa	7.84	0.56	6.23	0.30	4.00
Wellega	5.62	0.26	5.61	0.07	1.76
Wollo	7.49	0.36	5.38	0.23	3.84
South Nation Nationalities and Peoples (SNNP)	4.92	0.00	4.92	0.00	1.53
Eritrea	2.84	0.00	2.84	0.00	0.61
Mean	**5.77**	**0.17**	**5.54**	**0.11**	**2.35**

Where, Ar=allelic richness; Arp=private allelic richness; NAF (≥5%) = number of different alleles with a frequency ≥5%; NLCA (≤25%) = number of locally common alleles (freq. ≥5%) found in 25% or fewer populations; NLCA (≤50%) = number of locally common alleles (freq.≥5%) found in 50% or fewer populations.

4.3. Genetic distance and genetic differentiation between populations

The pairwise Nei's standard genetic distance analysis between geographic origins of the 161 Noug genotypes is shown in the Table 6. The result was estimated based on Nei's 1978 genetic distance method. A very high genetic distance was revealed between different genotypes from different geographic regions ranging from 0.309 to 0.892. The largest genetic distance was observed between genotypes from Eritrea and

Bale (0.892), followed by Eritrea and Illubabor (0. 881). The lowest genetic distance (0.309) was recorded between genotypes from Shewa and Wellega populations. Noug genotypes from Eritrean origins appear to be consistently distant from the other Noug genotypes from Ethiopia and were followed by Noug genotypes from Jimma.

Table 6. Pairwise Nei's standard genetic distance among populations

Populations	Ars	B/G	Bal	Tig	Gon	Goj	Har	Ilu	Jim	She	Wel	Wol	SNP	Ert
Ars	0.000													
B/G	0.462	0.000												
Bal	0.438	0.530	0.000											
Tig	0.411	0.323	0.614	0.000										
Gon	0.719	0.315	0.591	0.473	0.000									
Goj	0.391	0.378	0.460	0.373	0.599	0.000								
Har	0.417	0.537	0.529	0.485	0.771	0.467	0.000							
Ilu	0.654	0.426	0.768	0.566	0.482	0.541	0.788	0.000						
Jim	0.820	0.770	0.732	0.685	0.853	0.619	0.678	0.875	0.000					
She	0.434	0.409	0.499	0.373	0.551	0.346	0.321	0.500	0.688	0.000				
Wel	0.481	0.439	0.498	0.378	0.555	0.318	0.391	0.601	0.459	**0.309**	0.000			
Wol	0.595	0.638	0.622	0.382	0.659	0.520	0.583	0.736	0.553	0.507	0.411	0.000		
SNP	0.523	0.526	0.496	0.435	0.660	0.547	0.516	0.627	0.684	0.416	0.332	0.492	0.000	
Ert	0.811	0.783	**0.892**	0.669	0.810	0.801	0.808	0.881	0. 790	0.801	0.78	0.71	0.765	0.000

Where, Ars=Arsi, B/G=Benishangul Gumuz, Bal=Bale, Tig= Tigray, Gon=Gonder, Goj=Gojjam, Har=Harerge, Ilu= Illubabor, Jim= Jimma, She= Shewa (Central Ethiopia), Wel= Wellega, Wol= Wollo, SNP= South Nation Nationalities and Peoples, Ert= Eritrea

The pairwise population genetic differentiation (F_{ST}) among genotypes ranged from 0.028 between Wellega and Shewa to 0.175 between Eritrea and Bale (Table 7). Low F_{ST} value implies that there is a high frequency of identical alleles and high gene flow among Wellega and Shewa populations.

Table 7. Pairwise genetic differentiation (F_{ST}) between Populations (most pairs were significant, $p<0.05$)

Populations	Ars	B/G	Bal	Tig	Gon	Goj	Har	Ilu	Jim	She	Wel	Wol	SNP	Ert
Ars	0.000													
B/G	0.053	0.000												
Bal	0.062	0.062	0.000											
Tig	0.049	0.033	0.070	0.000										
Gon	0.077	0.035	0.071	0.049	0.000									
Goj	0.047	0.036	0.055	0.036	0.056	0.000								
Har	0.065	0.068	0.078	0.065	0.098	0.062	0.000							
Ilu	0.072	0.044	0.084	0.056	0.053	0.053	0.093	0.000						
Jim	0.103	0.087	0.098	0.083	0.098	0.076	0.099	0.105	0.000					
She	0.048	0.036	0.056	0.034	0.050	0.031	0.046	0.047	0.078	0.000				
Wel	0.054	0.040	0.057	0.036	0.052	0.030	0.054	0.056	0.062	**0.028**	0.000			
Wol	0.073	0.064	0.078	0.046	0.071	0.055	0.081	0.077	0.077	0.052	0.046	0.000		
SNP	0.070	0.059	0.068	0.053	0.075	0.060	0.076	0.073	0.093	0.048	0.042	0.064	0.000	
Ert	0.164	0.130	**0.175**	0.126	0.155	0.141	0.163	0.172	0.170	0.139	0.131	0.140	0.147	0.000

Where, Ars=Arsi, B/G= Benishangul Gumuz, Bal= Bale, Tig= Tigray, Gon= Gonder, Goj= Gojjam, Har= Harerge, Ilu= Illubabor, Jim= Jimma, She= Shewa (Central Ethiopia), Wel= Wellega, Wol= Wollo, SNP= South Nation Nationalities and Peoples, Ert= Eritrea

4.4. Analysis of molecular variance (AMOVA)

AMOVA was carried out to determine the extent of the variation within and among populations. The result partitioned the total molecular variance within and among the sets of tested genotypes based on their genetic origins. AMOVA revealed that 3.94% and 96.06% of the total variation resulted from among population and among individuals within the population (Table 8). This indicates that out of the total variation the highest was found among individuals and lowest among populations. The genetic differentiation as measured by pairwise fixation index (F_{ST}) was low (0.039) among populations, implying there is a very high gene flow among populations. On the other hand, a higher differentiation within individuals (F_{IS}=0.96) was quantified.

Table 8. AMOVA showing the genetic differentiation within and among populations from different origins as revealed by 13 microsatellite markers

Source of variation	**Df**	**SS**	**Estimated variance**	**% of variation**	**F-statistics**	**P-value ***
Among populations (AP)	13	205.945	0.22227 va	3.94	F_{ST}=0.039	0.001
Among individuals within population (AI)	147	1593.409	5.41976 vb	96.06	F_{IS}=0.96	0.001
Total	160	1799.354	5.64203	100		

Df= degree of freedom, SS= sum square, * significant test at 1023 permutations, Va= variance components due to among populations, Vb= variance components due to among individuals.

4.5. Cluster analysis of genotypes

The unweight neighbor-joining cluster analysis grouped the 161 Noug genotypes into three major clusters (C-1, C-2, and C-3), and each major cluster is further grouped into six sub-clusters (Fig. 4). The three clusters are composed C-1 (46.6%), C-2 (30.5%) and C-3 (22.9%) of the total populations, respectively, forming different hierarchical sub-clusters. The first cluster was the major cluster consisting of 75 genotypes from all populations except populations from Arsi and Harerge, whereas the second cluster contains 49 genotypes excluding Jimma, SNNP and Eritrea populations. And the third cluster contains 37 genotypes consisting of genotypes from all populations except Gonder and Illubabor. However, genotypes assignment in each major cluster was considerably different (Fig. 4). Cluster one (C-1) consists more than 83% were genotypes from B/G (16%), Tigray (11%), Gonder (10%), Gojjam (12%), Shewa (11), Wollo (16%) and Wellega (7%). In C-2, most of the genotypes were mainly from Shewa (20%), Tigray (14%), Arsi (13%) and Illubabor (11%). The C-3 cluster constitutes genotypes from Bale (14%), Harerge (10%), SNNP (11%), Shewa (9%), and Jimma (8%). Genotypes from Shewa are mainly found in all the three major clusters.

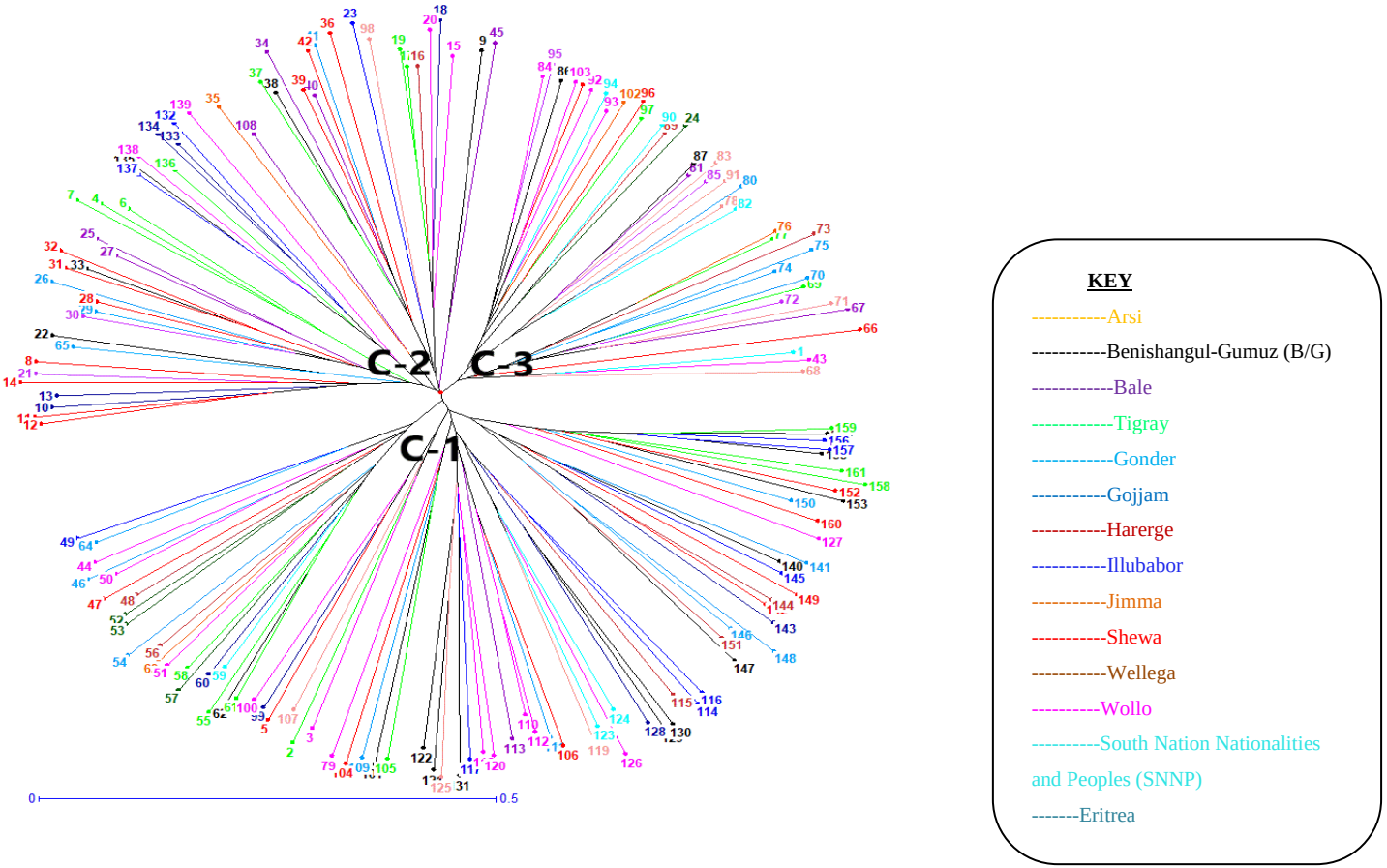


Figure 4. Neighbor –Joining tree generated based on dissimilarity matrix for 161 Noug genotypes using 13 microsatellite markers (the scale bar shows the coefficient of evolutionary distance)

Population grouping was carried out based on unweighted pair-group method with the arithmetic mean (UPGMA) method to determine the genetic relationship among the 14 populations (Fig. 5). The analysis divided the populations into three major-clusters (I, II, and II) consisting of Eritrea, Jimma and other populations in the main cluster. The third cluster further formed sub-cluster consisting Harerge as main sub-cluster and the rest as one sub-cluster. Populations from B/G, Tigray, Gojjam, Shewa, and Wellega originated from the same main sub-cluster. The clustering pattern indicated populations from geographically adjoining regions (Arsi vs Bale) and (Shewa vs Wellega) are sub-clustered together. On the other hand, populations from geographic distant regions (B/G vs Tigray) were found in the same sub-cluster.

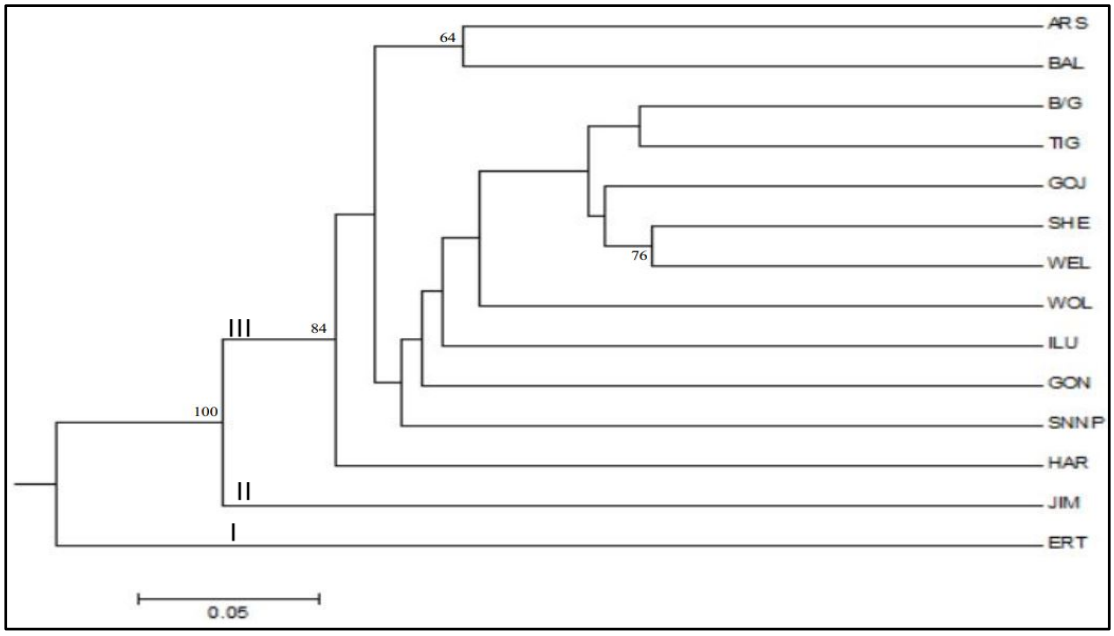


Figure 5. UPGMA dendrogram showing the genetic dissimilarity among the fourteen populations generated by 13 microsatellite markers (Observations are; ARS= Arsi, B/G= Benishangul Gumuz, BAL= Bale, TIG= Tigray, GON= Gonder, GOJ= Gojjam, HAR= Harerge, ILU= Illubabor, JIM= Jimma, SHE= Shewa or Central Ethiopia, WEL= Wellega, WOL= Wollo, SNP= South Nation Nationalities and Peoples, ERT= Eritrea). The scale bar shows the coefficient of evolutionary distance and the numbers above branches represents bootsrap values with 1000 replications, and bootstrap values less than 60% are not displayed.

4.6. Principal coordinate analysis (PCoA)

The genetic relatedness of the 161 Noug genotypes was further investigated using principal coordinate analysis (PCoA). The results indicated that about 21.2% of the overall variation was clarified by the first three most informative co-ordinates. From the total variation, the first, the second, and the third co-ordinates described 7.8%, 6.7%, and 6.6%, respectively.

Patterns of genotype distribution on a two-dimensional plot revealed a scattered plot with a uniform distribution of genotypes in different coordinates without separate clustering of populations (Fig. 6). The two-dimensional (2D) plot revealed poor patterns of grouping based on their geographic origins and the individuals were intermixed.

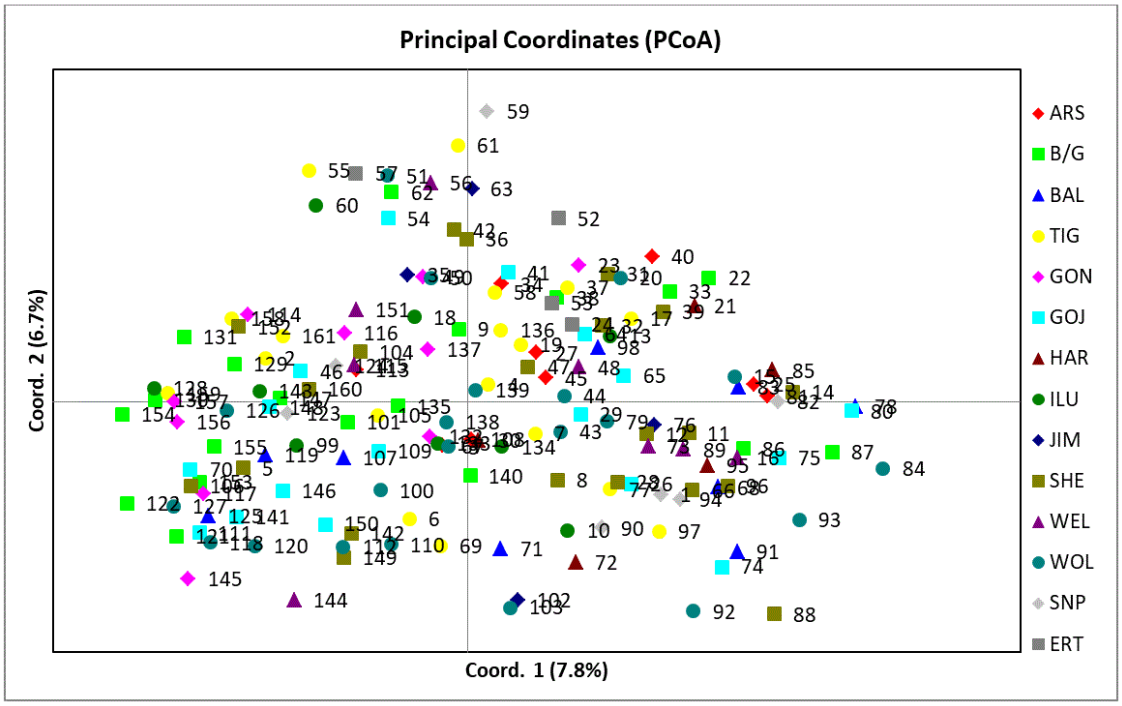
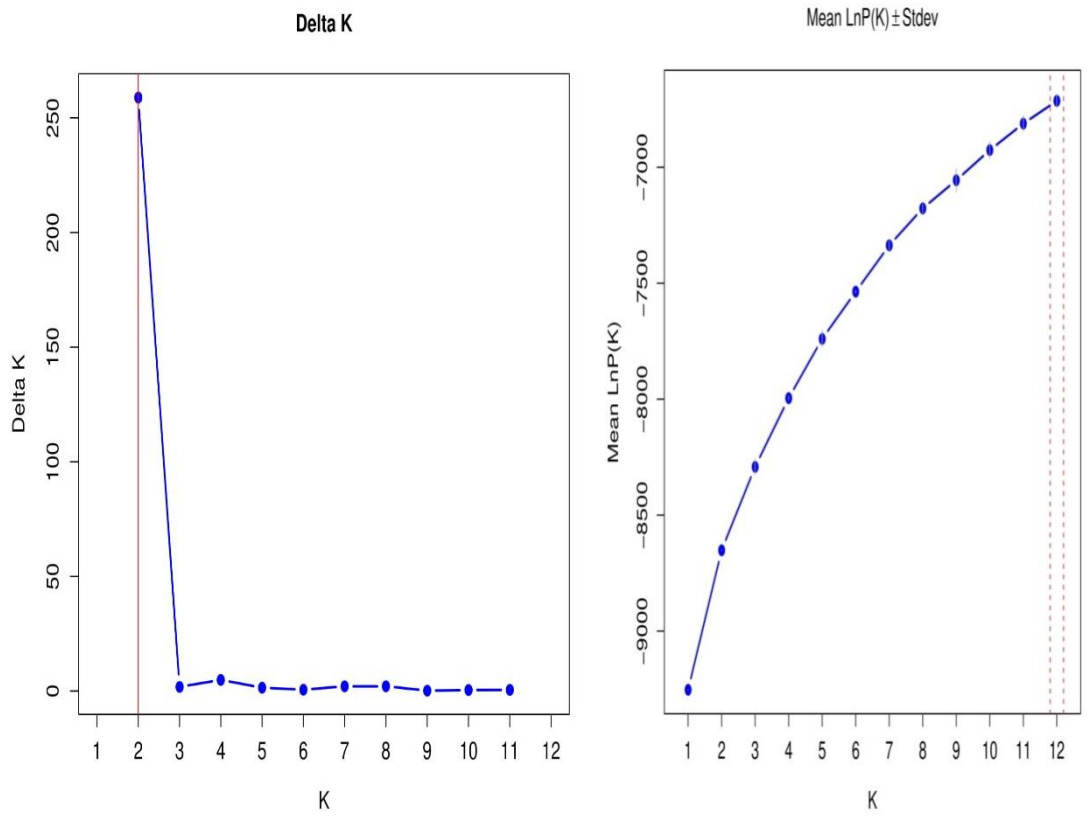


Figure 6. Principal coordinates analysis (PCoA) of the 161 Noug genotypes as revealed by 13 microsatellite markers (Observations are; ARS=Arsi, B/G= Benishangul Gumuz, BAL=Bale, TIG=Tigray, GON=Gonder, GOJ=Gojjam, HAR=Harerge, ILU= Illubabor, JIM= Jimma, SHE=Shewa 0r Central Ethiopia, WEL= Wellega, WOL=Wollo, SNP= South Nation Nationalities and Peoples, ERT= Eritrea).

4.7. Population structure analysis

A model-based approach was carried out in STRUCTURE software package (Pritchard *et al*., 200) to see the genetic admixture of genotypes. The delta K values were used to estimate the number of clusters of genotypes based on Evanno's method. The analysis assigned the individuals into sub-populations originated from different geographic region into K= 2 (Fig. 7; Table 9). The bar plot result showed a moderate genetic admixture and failed to show some structure based on clustering of the pre-defined populations (Fig. 8).



A **B**

Figure 7. The inferred population structure of 161 Noug genotypes (A) A biplot detected the optimum number of clusters at **Δ**K=2 based on Evano *et al*. (2005) method of estimation (B) the mean L(K) ± SD over 12 runs for each K value.

Table 9. The Evanno table output

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	10	-9253.210000	0.422821	—	—	—
2	10	-8651.440000	0.934761	601.770000	242.010000	258.900455
3	10	-8291.680000	34.949958	359.760000	63.700000	1.822606
4	10	-7995.620000	8.298166	296.060000	40.220000	4.846854
5	10	-7739.780000	35.923895	255.840000	52.830000	1.470609
6	10	-7536.770000	5.414600	203.010000	3.130000	0.578067
7	10	-7336.890000	19.655898	199.880000	40.350000	2.052819
8	10	-7177.360000	18.331042	159.530000	38.230000	2.085533
9	10	-7056.060000	49.907096	121.300000	8.420000	0.168713
10	10	-6926.340000	36.295552	129.720000	15.260000	0.420437
11	10	-6811.880000	34.493020	114.460000	16.600000	0.481257
12	10	-6714.020000	29.274745	97.860000	—	—

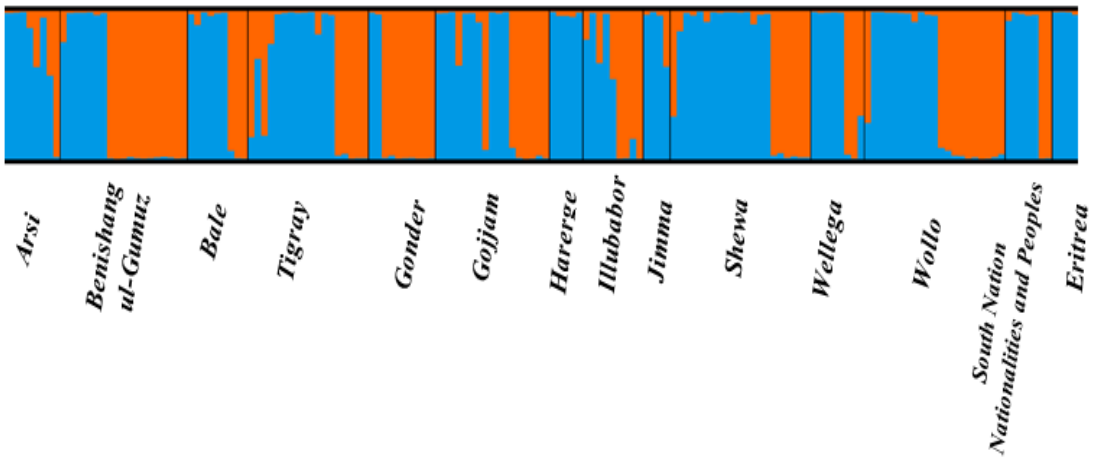


Figure 8. The population structure of Noug genotypes collected from Ethiopia and Eritrea. The structure bar plot shows the estimated membership of 161 individuals into K=2 clusters (where the color represents each clusters, the length of color segment is the estimated share of membership of the genotypes in that cluster).

5. DISCUSSION

5.1. Microsatellite markers reveal a high level of polymorphism in Noug

In this study, express sequence tags derived from simple sequence markers (EST-SSR) were used to analyze the genetic variation and population differentiation of Ethiopian and Eritrean Noug genotypes. EST-SSR markers are more informative to anchor genetic variability and evolutionary study in crops as reported by different scholars (Fu *et al*., 2014; Feng *et al*., 2016).

All the 13 microsatellite markers used in this study were highly polymorphic with a mean PIC value of 0.82. All the loci used had a PIC value >0.5, implying the high discriminating ability of the markers. This is highly supported by Botstein *et al*. (1980) where they asserted that markers with PIC value <0.25 are considered as less informative, PIC value between 0.25 and 0.5 are informative, whereas markers with PIC value >0.5 are highly informative. The PIC values reported in our study are also higher than those reported earlier (Abebaw Misganaw and Solomon Abera, 2017; Birhanu Mengistu *et al*., 2020).

In this study, 158 alleles were detected in 161 genotypes using 13 markers, with an average of 12.15 per locus. This is higher than previously reported by Birhanu Mengistu *et al*. (2020), in which a total of 135 alleles from 100 accessions were discovered across 14 loci, with a mean of 9.6 per locus. Abebaw Misganaw and Solomon Abera, (2017) found 66 alleles in 65 accessions using 11 loci, with a mean of 9 alleles per locus. The number of alleles in the current study is also higher than estimated by Dempolf *et al*. (2010) who obtained 137 alleles using SSR markers. The probable reason for obtaining a higher number of alleles than earlier reports may be

due to the size of genotypes, the genetic diversity of selected genotypes, and the markers difference used for this investigation.

5.2. Gene diversity analysis along with populations

Elucidating the genetic divergence and relationship of genetic resources is a prerequisite in making a parental selection for breeding and implementing genetic conservation strategies (Muhinyuza *et al*., 2015). The evolutionary forces for population differentiation are mostly selection, genetic drift, and gene flow that act continuously and results in continuous changes in allelic frequency in a population (Slatkin, 1987). Understanding the frequency and distribution of alleles is important to identify genetically distinct parents (Iskakova *et al*., 2014). The gene diversity indices were measured among and within the pre-determined populations based on Noug growing regions.

Heterozygosity can be recognized as one of the measure of the amount of genetic diversity within genotypes (Nei and Roychoudhury, 1974). This parameter shows how much the population difference happens and how the change speared through the alleles of the examined markers. The mean heterozygosity values obtained in our study, except observed heterozygosity were highest, suggests that the highest intra-populations genetic diversity and SSR markers are powerfulness to resolve heterozygosity and homogeneity (Filippi *et al*., 2015). The expected heterozygosity or equivalent to gene diversity is higher than that obtained by Dempewolf *et al*. (2015) and Birhanu Mengistu *et al*. (2020) (He= 0.49 and He= 0.63, respectively) using SSR markers. Thereby, indicating that the highest level of genetic diversity in the genotypes used in our study than their counterparts. The average expected heterozygosity of the current study is also higher than estimated based on the RAPD

marker (Mulatu Geleta *et al*., 2008) and AFLP marker (Mulatu Geleta *et al*., 2007) where the values were 0.18 and 0.21, respectively. This is most likely because RAPD and AFLP are both dominant markers that disparage the marker's locus polymorphism.

The observed heterozygosity was much lower than expected heterozygosity under Hardy-Weinberg Equilibrium (HWE). Noug is a strictly self-incompatible crop (Mulatu Geleta and Bryngelsson, 2010), and hence the observed heterozygosity (Ho) is predicted to exceed from the expected heterozygosity (He) when all the HWE assumptions are met (Sewalem Tsehay *et al*., 2020). Our result is in agreement with the one reported by Birhanu Mengistu *et al*. (2020) who obtained lower observed heterozygosity (Ho=0.20) than expected heterozygosity (He=0.63) using SSR markers. Using similar markers, Dempewolf *et al*. (2015) reported Ho and He values of 0.49 and 0.54, respectively in 29 Noug and its progenitor populations. The slight difference occurred in each study including our study are because of the sample size difference and the efficiency of the marker of resolving heterozygosity. Furthermore, the current finding concurs with that of Filippi *et al*. (2015) and Zeinalzadeh-Tabrizi *et al*. (2018), who observed lower observed heterozygosity than with that of expected heterozygosity in sunflower, an out-crossing species and the closest oilseed crop for Noug.

The other measure of gene diversity is Shannon's information index or Shannon-weaver diversity index (I). If Shannon's information index value is close to one or above, it indicates that there is variation in the tested populations and that the markers are suitable for studying diversity (Nassiry *et al*., 2009). The values obtained in our study is ranged from I= 0.89 to I=1.97 with a mean of I= 1.57. Except for Eritrea, all the thirteen populations had Shannon's information index values of greater than

one(1). This implies that the Ethiopian noug genotypes have a more diversified genetic makeup than those genotypes from Eritrea.

In most of gene diversity parameters computed, the Noug populations from Shewa, Wollo, Benishangul-Gumuz (B/G), Gojjam, and Tigray showed higher values, which signifies that these locations could be hotspots for Noug. This finding is somewhat consistent with the findings of Getinet Almaw and Ketema Belete, (2000), who identified Shewa, Gonder, and Gojjam as a potential hotspot area for Noug diversity, based on morphological descriptors. These areas are also higher in allelic richness and private alleles. All populations had private alleles (Arp) except Noug populations collected from Harerge, Illubabor, Jimma, SNNP, and Eritrea. The occurrence of private alleles pointed out the existence of genetic uniqueness among the genotypes in the respective regions in certain loci. Private alleles might illustrate the availability of important traits for future breeding and improvement programs. This may also increase the allelic richness in the gene bank and could be exploited in the future.

5.3. Genetic distance and patterns of population differentiation

Based on Nei's (1987) distance matrix, the pairwise Nei's standard genetic distance between populations was estimated. Accordingly, Noug genotypes from Eritrea and Bale were the most distant showing only10.8% similarity. This might be due to the genotypes are from geographically distant regions. Comparatively, the lowest genetic distance was observed between genotypes from Wellega and Shewa (0.309) showing 69% similarity. The pairwise genetic distance was significant across pairs which are in harmony with the idea that a genetic distance higher than 0.25 is regarded to be a higher genetic distance.

According to Wright's (1951), the level of population differentiation is determined by fixation indices or F-statistics (F_{ST}, F_{IS} and F_{IT}). Theoretically, F_{ST} value can be categorized as low (0 - 0.05), moderate (0.05 to 0.15), high (0.15 to 0.25), and very high (> 0.25) population differentiation. In this study, the pairwise population differentiation was found in between 0.028 to 0.175 or in the range of low to higher. However, most of the pairwise population differentiation was moderate. The global population differentiation (F_{ST}) based on AMOVA was 0.039, which is low among the fourteen populations. This is probably due to the occurrence of high gene flow through seed exchange irrespective of their geographic distance. Similar results of low population differentiation were reported by Mulatu Geleta *et al*. (2007; 2008) and Birhanu Mengistu *et al*. (2020) using RAPD, AFLP, and SSR markers, respectively.

The AMOVA confirmed the weak differentiation of genotypes in this study, which revealed that a large proportion of the variation was found among the individuals within the regions of the collection (AI) than among populations (AP). Similar findings reported in Ethiopian Noug accessions by Mulatu Geleta *et al*. (2007) using RAPD markers who obtained the highest (65%) variation were within the individuals and lowest (35%) among populations. Mulatu Geleta *et al*. (2008) using AFLP markers obtained 77% variance components within populations and 23% among populations. In addition, about 95% of the variation was attributed within populations and 5% among populations in Noug genotypes using ISSR and SSR markers (Hussain *et al*., 2015; Birhanu Mengistu *et al*., 2020). Even though a relatively higher proportion of within-population variation was revealed by different reports, most of the results demonstrated that the total variation is found within populations rather than among populations. Most certainly, most of the genetic variability in cross-pollinating plants distributed within individuals, and a small proportion of its variation is

attributed to among populations (Hamrick *et al*., 1990). Indeed, Noug is a strictly out-crossing crop and all the AMOVA results are agreed with this stated presumption.

5.4. Cluster analysis and population structure

It has been suggested that cluster analysis helps to harvest optimum information explaining genetic distance and their evolutionary relationship among genotypes. In this study, we clustered the 161 Noug genotypes into three major groups and six sub-groups based on the neighbor-joining (NJ) method. All the three clusters contain genotypes from geographic distance and did not clearly divide the populations into geographic origins, demonstrating the occurrence of substantial gene flow along with human movement. This study is congruent with previously reported by Mulatu Geleta *et al*. (2007) who obtained three major clusters from 70 populations using RAPD marker. Similarly, the 17 Noug populations were cluster into three major clusters based on their geographic proximity to a considerable degree as revealed by AFLP markers (Mulatu Geleta *et al*. (2008). Also, Birhanu Mengistu *et al*. (2020) reported about 100 Noug accessions from different geographic regions have been clustered into three as assessed by SSR markers. Likewise, the current study is also in agreement with different reports in which the clustering pattern is weak enough to support isolation by distance in other cross-pollinating crops (Barati and Arzani, 2012; Zia *et al*., 2014; Mwangi *et al*., 2019).

However, the majority of the populations were significantly grouped based on geographic closeness using the UPGMA method constructed based on the mean dissimilarity values of each population. Populations from nearby areas (Shewa and Wellega) and (Arsi and Bale), were found in the same sub-cluster. Geographic closeness plays a prominent factor in genetic similarity due to the high possibility of

seed exchange (Mulatu Geleta *et al*., 2007). Sometimes, populations from geographically distant regions are found in the same sub-cluster, for example, Benishangul-Gumuz (B/G) vs Tigray. This might be explained by the inclusion of genotypes from the same genetic backgrounds. This further supports previously reported (Yohanes Petrose *et al*., 2007), where clustering may be according to the type of noug growing (*Bugne or Abat*). The case of Jimma further supports this explanation, on which that, populations from Jimma showed clearly unique cluster. This is because that Jimma is found between Shewa and Wellega, but the Noug growing areas of Jimma are not as similar as the Noug growing areas of Shewa and Wellega. This is because of two great arid valleys separated it from these regions. On the Northeast, the Gibe valley separates Jimma and Shewa and on the Northwest, the Diddessa valley separates Jimma from Wellega. Hence, this isolates and determines the type of Noug growing in the area (Yohanes Petrose *et al*., 2007).

The Eritrean population clustering trend appears to be at odds with the above assumptions. This is most likely because Eritrean genotypes were gathered when the country was still part of Ethiopia or before it became an independent country, but Ethiopian genotypes are updated and new collections are added at different periods. As a result of sexual recombination, continuous cultivation may result in allelic difference, and selection for desirable traits favors few alleles at the cost of others resulting in increased frequency of selected alleles (Fu, 2015).

The principal coordinate analysis (PCoA), a 2D display clustering showed no clear pattern of population clustering based on their region of origin. There is intermixing of genotypes in different coordinates. This suggests a significant degree of genetic mixing in each coordinate and a lack of a distinct pattern of grouping. This result substantially validates our NJ dendrogram clustering. The PCoA result confirmed by

prior findings, in which a weak clustering was detected from distinct geographic sources (Mulatu Geleta *et al.,* 2008; Hussian *et al*., 2015; Birhanu Mengistu *et al*., 2020).

Similarly, the structures of populations based on their geographic locality have been poorly inferred. This is similar as the PCoA of our result, where the geographic distance was not strongly correlated with the genetic background. The Noug populations have been categorized into K=2 sub-population (genetic origin) meaning that each individual shares genetic backgrounds inherited from both clusters, with a considerable level of genetic admixture. But, some of the predefined populations appeared genetic homogeneity in which the genotypes acquired dominantly from either of the clusters. Accordingly, a more structured populations were reported with K=2 genetically distinct of Noug accessions and its putative progenitor with a little genetic admixture (Dempewolf *et al*., 2015. In contrast, the structure analysis weakly inferred the 24 Noug accessions collected from different regions into K=3 genetic background with a very high genetic admixture (Sewalem Tsehay *et al*. (2020). This confirmed the presence of potential genetic admixture within the Noug accessions even if collected from different regions.

6. CONCLUSION AND RECOMMENDATION

6.1. Conclusion

Noug is an oilseed crop mainly grown in Ethiopia, Eritrea, and India for its oil and different socio-economic importances. It has received very little attention in research and development and remained as a neglected or underutilized oilseed crop. The depiction of the genetic diversity and population structure of this neglected crop aids breeders to focus on the available genetic resource and for implementing improvement programs. The microsatellite markers used were more informative for diversity and population structure analysis of Noug, and could be applied in future association mapping studies. The result of this study revealed that there is a significant level of genetic variation, both within and among Noug genotypes

Noug has a wide range of genetic diversity in Ethiopia and widely distributed and grows in different parts of the country. However, populations from Shewa, Gojjam, Gonder, Tigray, and Benishangul-Gumuz showed the highest variation in most of the gene diversity parameters. These areas are also showed the highest number of private alleles and are rich in alleles and can be considered as hotspots areas for Noug. This indicates the presence of high genetic resources in the gene bank and could be exploited and utilized in the future improvements.

The AMOVA revealed that individual differentiation was much higher than among population differentiation (geographic regions). Thus, Individual diversity should be prioritized above just focusing on their places of origin. Cluster analysis using NJ or UPGMA, PCOA, and population structure showed that most Noug genotypes were not correlated with their corresponding geographic origin. These confirm the existence of potential gene flow among populations.

6.2. Recommendation

The following recommendations are forwarded based on the current study;

- Microsatellite markers are highly informative and should be used for assessing the gentic diversity of Noug accessions maintained at the gene bank at the Ethiopian Biodiversity Institute (EBI);
- Single Nucleotide Polymorphism (SNP) markers should be applied for genome-wide assessment of genetic variation in Ethiopian Noug;
- Noug breeding and insitu-conservation program should focus on populations from Shewa, Wollo, Gonder, Gojjam, Tigray, and Benishangul-Gumuz (B/G), identified as hotspot areas for Noug;
- Desirable traits identification and its association should be worked out for Noug populations in Shewa, Wollo, Gojjam, and Benishangul-Gumuz (B/G) populations which showed higher numbers of private alleles;
- New germplasm collection should be further carried out in Jimma, Harerge, and SNNP regions which showed scarce representations in the gene bank.

7. REFERENCES

Abebaw Misganaw and Solomon Abera (2017). Genetic diversity assessment of *Guizotia abyssinica* using EST-derived simple sequence repeat markers. *Afr.J.Plant Sci*.**11**: 79-85.

Abebe Demissie, Dawit Tadesse, Getahun Mulatu and Debritu Beyene (1992). Ethiopia's oilseed genetic resources. In first national oilseeds workshop, Addis Ababa, Ethiopia, 3-5 Dec 1991, IAR.

Adarsh, M.N., Poonam, K. and Shilpa, D. (2014). A review of *Guizotia abyssinica* : A multi-purpose plant with an economic prospective. *J.Ind.Pollut. Control.* **30**:277-280.

Adefris Teklewold and Adugna Wakjira.(2004). Seedfilling and oil-accumulation in Noug. SINET:*Ethiop.J.Sci*. **27**:25-32.

Adefris Teklewold and Nigussie Alemayehu (2002). Studies on the floral characterstics of nigerseed (*Guizotia abyssinica*). *J.App.Bot.* **76**: 163-167.

Allison, P.D. (2010). Survival analysis using SAS: a practical guide. Sas Institute.

Baagøe, J. (1974). The genus Guizotia (Compositae) a taxonomic revision. *Botanisk Tidsskrift*. **69**: 1-39.

Barati, M. and Arzani, A. (2012). Genetic diversity reveald by EST-SSR markers in cultivated and wild safflower. *Biochem. Syst. Ecol*. **44**: 117-123.

Benelli, V. G. (2015). Comparison of seed yield, oil and phenotypic traits among selected parents and acrosses of niger. PhD Thesis; University of Tennesse.

Birhanu Aboye, Wosene Gebrselassie and Tesfaye Disasa (2020). Estimating the genetic diversity of Ethiopian Noug (*Guizotia abyssinica*) genotypes using SSR markers. *Adv Crop Sci Tech*.**7**:2

Birhanu Mengistu, Wosene Gebrselassie and Tesfaye Disasa (2020). Diversity analysis in *Guizotia abyssinica* (Lf) Cass. Germplasms collected from Ethiopia. *Chemical and Biomolecular Engineering* **5**:8-14.

Botstein, D., White, R.L., Skolnick, M. and Davis, R.W. (1980). Construction of a genetic linkage map in man using RFLPs. *Am. J.Hum. Genet.* **32:** 314-331.

Bulcha Woyessa (2007). *Guizotia abyssinica* (Lf) Cass. *Record from PROTA4U.van der Vossen, HAM, and Mkamilo, GS (Editors). PROTA (Plant Resources of Tropical Africa/Ressources végétales de l'Afrique tropicale), Wageningen, Netherlands*

Chakraborty, R. andRao, C.R. (1991). Measurment of genetic variation for evolutionary studies. *Handbook of statistics* **8**: 271-316.

CSA, (2019). ' Report on area and production for major crops (private peasant holdings, Meher season)' Statistical Bulletin No. 589,I.

Dempewolf, H., Kane, N.C., Ostevik, K.L., Mulatu Geleta, Barker, M.S., Lia, Z and Rieseberg, L.H. (2010). Establishing genomic tools and resources for *Guizotia*

abyssinica (Lf) Cass---the development of library of expressed sequence tags, microsattelite loci, and the sequencing of its chloroplast genome. *Mol. Ecol. Resour.* **10**: 1048-1058.

Dempewolf, H., Mistru Tesfaye, Abel Teshome, Borkman, A.D., Andrew, R.L., Scascitelli, M., Black, S., Endashaw Bekele, Engels, J.M., Cronk, Q.C and Rieseberg, L.H. (2015). Patterns of domestication in the Ethiopian oilseed crop Noug (*Guizotia abyssinica*). *Evol.appl.* **8**: 464-475

Dillon, N.L., Innes, D.J., Bally, I.S., Wright, C.L., Devitt, L.C. and Dietzgen, R.G. (2014). Expressed sequence tag-simple sequence repeat (EST-SSR) marker resources for diversity analysis of mango (*Mangifera indica* L.). *Diversity* ***6***:72-87.

Earl, D.A. and Von Holdt, B.M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evano method. *Conservation Genetics Resources* **4**: 359-361.

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

Excoffier, L., Laval,G. and Schneider,S. (2005). ARLQUIEN ver.3.0: An integrated software package for population genetics data analysis. *Evol. Bioinform.online*.**1**: 47-50.

Feng, S., He, R., Lu, J., Jiang, M., Shen, X., Jiang, Y., Wang, Z.A. and Wang, H. (2016). Development of SSR markers and assessment of genetic diversity in medicinal Chrysanthemum morifolium cultivars. *Front. Genet.* **7**:113.

Filippi, C.V., Aguirre, N., Rivas, J.G., Zubrzycki, J., Puebla, A., Cordes, D., Moreno, M.V., Fusari, C.M., Alvarez, D., Heinz, R.A. and Hopp, H.E. (2015). Population Structure and Genetic Diversity of Sunflower and association mapping of Populations using SSR and SNPs markers. *BMC Plant Biol*.**15**:52

Frankel, O. H., Brown, A. H. and Burdon, J. J. (1995). *The conservation of plant biodiversity*. Cambridge University Press.

Fu, N., Wang, P.Y., Liu, X.D. and Shen, H.L. (2014). Use of EST-SSR markers for evaluating genetic diversity and fingerprinting celery (*Apium graveolens* L.) cultivars. *Molecules* **19**:1939-1955.

Fu, Y.B. (2015). Understanding crop genetic diversity under modern plant breeding. *Theor. Appl. Genet.* ***128***:2131-2142.

Genet Tsige and Ketema Belete (2000). " Phenotypic diversity in Ethiopian noug Gerplasms". *Afr. Crop Sc. J*.**8**:137-143.

Getinet Alemaw and Adefris Teklewold (1995). An agronomic and seed quality evaluation of noug (*Guizotia abyssinica*) germplasms in Ethiopia. *Plant breed.* **114**: 375-376.

Getinet Alemaw and Alemayehu Nigussie (1992). Production and Oilseeds Research in Getinet Alemaw and Sharma, S. M. (1996). *Niger, Guizotia abyssinica (*Lf) Cass (Vol. **5**). Bioversity International.

Getinet Alemaw and Alemayehu Nigussie (1997). *Highland oil crops: A three decade research experience in Ethiopia* (Research Report No. 30, pp. 22-27).

Getinet Alemaw and Sharma, S.M. (1996). Niger (*Guizotia abyssinica* (L.F.)) Cass. In Promoting the Conservation and Use of Underutilized and Neglected Crops; International Plant Genetic Resources Institute (IPGRI): Rome, Italy.

Glaszmann, J. C., Kilian, B., Upadhyaya, H. D. and Varshney, R. K. (2010). Accessing genetic diversity for crop improvement. *Curr.Opin. Plant Biol.* **13**: 167-173.

Goyal, V.K. and Bisen, R. (2017). Assessment of genetic divergence in Niger germplasm. *Int. J. Chem. Stud*. **5**:1482-1485.

Gupta, P. K., Varshney, R. K., Sharma, P. C. and Ramash.B. (1990). Molecular markers and their application in wheat breeding. *Plant Breed.* **118**: 369-390.

Hadrys, H., Balick, M. and Schierwater, B. (1992). Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. *Mol. Ecol*, ***1***: 55-63.

Hamrick JL. and Godt MJ. (1990). Allozyme diversity in plant species. In: Genetics, Breeding, and Genetic Resources. pp. 43-63. Sinauer Associates Inc. Sunderland, Massachusetts, USA,

Hartl, D. L. (1988). *A primer of population genetics*. Sinauer Associates, Inc..

Hedrick, P. (2011). *Genetics of populations*. Jones and Bartlett Learning.

Hirmath, S.C. and Murthy, H.N. (1992). Cytological studies in *Guizotia* (*Asteraceae*). *Caryologia* **45**: 69-82.

Hirmath, S.C.and Murthy, H.N. (1988). Domestication of niger (*Guizotia abyssinica*). *Euphytica* **37**: 225-228.

Hiruy Belayneh and Nigussie Alemayehu. (1987). Progress in Niger Research in Ethiopia. In: Proceeding of the Third Oil crops Networks, 6-10 October 1986, pp. 6-10, IAR, Addis Ababa, Ethiopia.

Hosalli, S. (2005). Studies on Self Fertility and Autogamy in Niger *(Guizotia abyssinica Cass.)* (Doctoral dissertation, UAS, Dharwad.

Hussain, Z., Yadav, S., Kumar, S., Suneja, P., Nizar, M. A., Yadav, S. K. and Dutta, M. (2015). Molecular diversity studies of Niger-seed (*Guizotia abyssinica*) germplasmsfor oil parameters. *Indian J.Biotechnol*. **14**: 344-350.

Iskakova, A.N., Romanova, A.A., Voronina, E.N., Sikhayeva, N.S., Belozerceva, A.B., Filipenko, M.L. and Ramanculov, E., (2014). Allele frequency and genotype distribution of 9 SNPs in the Kazakh population. *J Pharmacogenomics and Pharmacoproteomics* **5**:2153-0645.

Jacobson, A. and Hedrén, M. (2007). Phylogenetic relationships in Alisma (Alismataceae) based on RAPDs, and sequence data from ITS and trnL. *Plant Syst. Evol***. 265**: 27-44.

Joshi, S. P., Ranjekar, P. K. and Gupta, V. S. (1999). Molecular markers in plant genome analysis. *Curr. Sci*. **77**: 230-240.

Kalinowski, S.T. (2005). HP-RARE 1.0: A Computer program for performing rarefaction on measures of allelic richness. *Mol. Ecol. Notes*. **5**: 187-189.

Kandel H. and Porter p. (2002). Niger (*Guizotia abyssinica*).Production in Northwest Minnesota. University of Minnesota extension service.

Karp, A., Seberg, O. L. E. and Buiatti, M. (1996). Molecular techniques in the assessment of botanical diversity. *Ann. Bot.* **78**: 143-149.

Kifle Dagne (1994). Meosis in interspecific hybrids and genomic interrelationships in *Guizotia* (*Compositae*). *Hereditas* **121**: 119-129.

Kifle Dagne (1995). Karyotypes, C-Banding and nucleolar numbers in *Guizotia* (*Compositae*). *Plant Syst. Evol*. **195**: 121-135.

Kifle Dagne (2001). Cytogenetics of new *Guizotia* (*Compositae*) interspecific hybrids pertaining to genomic and phylogenetic affinities. *Plant Syst. Evol.* **230**:1-11.

Kifle Dagne and Johnsson, A. (1997). Oil content and fatty acid composition of seeds of *Guizotia* (*Compositae*). *J.Sci. Food Agric*. **73**: 274-278.

King, R. C. and Stansfield, W. D. (1990). *Encyclopedic dictionary of genetics*.VCH Verlagsgesellschaft mbH.

Koilkonda, P., Sato, S., Tabata, S., Shirasawa, K., Hirakawa, H., Sakai, H., Sasamoto, S., Watanabe, A., Wada, T., Kishida, Y. and Tsuruoka, H. (2012). Large scale development of EST-derived simple sequence repeat markers and diversity study in *Arachis* spp. *Mol Breed*. **30**: 125-138.

Kremer, A., R.J. Petik and O. Pons. (1998). Measures of polymorphism within and among populations. p. 301–311. In A. Karp et al. (ed.) Molecular tools for screening biodiversity. Chapmon and Hall, London.

Kumar B.N. (2008). Studies on maximizing seed yield and quality in niger (*Guizotia abyssinica* Cass.) M.S. thesis, Dharwad University of Agricultural Sciences, Karnataka, India.

Kumar, P., Gupta, V.K., Misra,A.K., Modi, D.R. and Pandy, B.K. (2009). The potential of molecular markers in plant biotechnology. *Plant omics* **2**:141.

Li, Y.L. and Liu, J.X. (2018). Structerselector: a web based software to select and visualize the optimal number of clusters using multiple methods. *Mol. Ecol. Resou*. **18**: 176-177.

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

Lynch, M. and Walsh, B. (1998). Genetics and analysis of quantitative traits. *Sunderland, MA: Sinauer.* **1**: 535-557.

Milbourne, D., Meyer, R., Bradshaw, J.E., Baired,E., Bonar, N., Provan, J., Powell, W. and Waugh, R. (1997). Comparison of PCR based marker systems for the analysis of genetic relationships in cultivated potato. *Mol breeding*. **3**: 127-136.

MoA, (2016). MINSTRY OF AGRICULTURE AND NATURAL RESOURCES AND SEED QUALITY CONTROL DIRECTORATE, (19), ADDIS ABABA, ETHIOPIA.

Mohan, M., Nair, S., Bhagwat, A., Krishna, T. G., Yano, M., Bhatia, C. R. and Sasaki, T. (1997). Genome mapping, molecular markers and marker-assisted selection in crop plants. *Mol breeding* **3**: 87-103.

Muhinyuza, J.B., Shimelis, H., Melis, R., Sibiya, J., Gahakwa, D. and Nzaramba, M.N. (2015). Assessment of genetic relationship among potato promising genotypes in Rwanda using SSR markers. *Aust. J. Crop Sci*. **9**:696-700.

Mulatu Geleta and Bryngelsson, T. (2010). Population genetics of self-incompatibility and developing self-compatible genotypes in niger (*Guizotia abyssinica*). *Euphytica*, ***176***:417-430.

Mulatu Geleta and Ortiz, R. (2013). The importance of *Guizotia abyssinica* (niger-seed) for sustainable food secuirity in Ethiopia. *Genet. Resour. Crop Evol*. **60**: 1763-1770.

Mulatu Geleta, Bryngelsson, T., Endashaw Bekele and Kifle Dagne (2008). Assesment of genetic diversity of *Guizotia abyssinica* (*Asteraceae*) from Ethiopia using AFLP markers. *Plant Genet. Resour*. **6**:41-51.

Mulatu Geleta, Bryngelsson, T., Endashaw Bekele and Kifle Dagne (2007). Genetic diversity of *Guizotia abyssinica* (*Asteraceae*) from Ethiopia as revealed by RAPD markers.*Genet Resour. Crop Evol.* **54**: 601-614.

Mulatu Geleta, Endashaw Bekele, Kifle Dagne and Bryngelsson, T. (2010). Phylogenetics and taxonomic delimitation of the genus *Guizotia abyssinica* based on sequences derived from various chloroplast DNA regions. *Plant Syst. Evol.***289**: 77-89.

Mulatu Geleta, Stymne, S. and Bryngelsson, T. (2011).Variation and inheritance of oil content and fatty acid composition in Niger. *J. Food Compos. Anal*.**24**:995-1003.

Mulatu Geleta, Zemede Asfaw, Endashaw Bekele and Abel Teshome (2002). Edible oilcrops and their integration with the major cereals in North Shewa and South Welo, Central Highlands of Ethiopia: An ethno-botanical prospective. *Hereditas* **137**: 29-40.

Murthy, H. N. (1996). Cytogenetics of *Guizotia* Species and Origin of Niger. *Proc.Indian.Acad.Sci*.**62**: 221-226.

Murthy, H.N., Hiremath, S.C. and Salimath, S.S. (1993). Origin, evolution and genetic differentiation of *Guizotia abyssinica* and its wild relative species. *Theor. Appl.Genet*. **87**: 587-592.

Mwangi, E.W., Lee, M.C., Sung, J.S., Marzougui, S. and Bwalya, E.C. (2019). Genetic Diversity Analysis of Maintaining Lines for Kenyan Sunflower (*Helianthus annus L*.) Using Allele Specific SSR Markers. In *Proceedings of the Plant Resources Society of Korea Conference* (pp. 61-61). The Plant Resources

Society of Korea.
https://www.koreascience.or.kr/article/CFKO201921956453484.page

Nagella, P., Hosakatte, N. M., Ravishankar, K. V., Hahn, E. J. and Paek, K. Y. (2008). Analysis of genetic diversity among Indian niger (*Guizotia abyssinica*) cultivars based on RAPD markers. *Electron. J. Biotechnol*. **11**: 140-144.

Nassiry, M.R., Javanmard, A. and Reza, T. (2009). Application of statistical procedures for analysis of genetic diversity in domestic animal populations. *Am. J Anim Vet Sci*. 4:136-141.

Nei, M. (1778). Estimation the average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583-590.

Nei, M. and Roychoudhury, A.K., (1974). Sampling variances of heterozygosity and genetic distance. *Genetics* **76**:379-390.

Parameshwarappa, S. G., Salimath, P. M. and Palakshappa, M. G. (2009). Assessment of genetic diversity in niger (*Guizotia abyssinica* (L) Cass). *Karnataka J. Agric. Sci.* **22**: 879-880.

Pavel , A. B. and Vasile, C. (2012). PyElph: A software tool for gel-image analysis and phylogenetics. *BMC Bioinform*.**13**:9.

Peakal, R. and Smouse, P. (2015). GENALEX: Genetic analysis in excel. Population genetic software for teaching and research. *Mol. Ecol. Notes*. **6**: 288-295.

Perrier, X. and Jacquemoud-Collet, J. P. (2015). DARwin software. 2006.

Prichard, J., Stephens, M. and Donnelly, P. (2000). Inference of population structure using multil-locus genotypic data. *Genetics* **155**: 945-959.

Pulate, S., Patil, H. and Patil, M. (2013). Multivariate analysis of genetic divergence among Niger genotypes in relation to seed oil quality traits. *Bioscan*. **8**: 829-833.

Quinn, J. and Myers, L. (2002). Niger seed: specialty grain opportunity for Midwestern 7 US. In: Janick, J., Whipkey, A. (Eds.), Trends in New Crops and New Uses. ASHS 8 Press, Alexandria, VA, pp. 174–182.

Ramu,P., Billot, C., Rami, J., Senthilvel, S., Upadhyaya, H., Reddy, L. and Hash, C. (2013). Assesment of genetic diversity in the sorghum reference set using EST-SSR markers.*Theor.Appl.Genet*.**126**:2051-2064.

Riley K W, and Hiruy Belayneh (1989). Niger. In: Oil Crops of the World, eds Robbelen G, DowneyRK and Ashri A. McGraw Hill, New York, USA, pp 394-403.

Saeed, A., Hovsepyan, H., Darvishzadeh, R., Imtiaz, M., Panguluri, S. K. and Nazaryan, R. (2011). Genetic diversity of Iranian accessions, improved lines of chickpea (*Cicer arietinum* L.) and their wild relatives by using SSR. *Plant Mol.Biol.Rep*. **29**: 848-858.

Schulman A.H., Flavell A.J. and Ellis T.H.N. (2004). The Application of LTR Retrotransposons as Molecular Markers in Plants. In: Miller W.J., Capy P. (eds)

Mobile Genetic Elements. *Methods Mol. Biol*. vol 260. *Humana Press*. https://doi.org/10.1385/1-59259-755-6:145.

Seetharam, K., Thirumeni, S. and Paramasivam, K. (2009). Estimation of the genetic diversity of rice genotypes using simple sequence repeats and morphological characters. *Afr. J. Biotechnol*. **8**:10.

Semagn, K., Bjornstad, A. and Ndjiondjop, M.N. (2006). An overview of molecular marker methods for plants. *Afr. J. Biotechnol.* **5**:25.

Sewalem Tsehay, Ortiz, R., Johansson, E., Endashaw Bekele., Kassahun Tesfaye, Hammenhag, C. and Mulatu Geleta (2020). New Transcriptom-based SNP markers for noug (*Guizotia abyssinica*) and their conversion to KASP markers for population genetic analysis. *Genes* **11**:1373.

Shaikh, M.F., Mirza, I.A.B. and Shaikh, S.N. (2019). Performance of Niger (*Guizotia abyssinica* Lf. Cass) Varieties under Varied Weather Condition of Marathwada Region. *Ind. J. Pure App. Biosci.***7:** 473-477.

Sileshi Nemomissa, Endashaw Bekele and Kifle Dagne (1999). Self-incompatibility system in the Ethiopian population of Guizotia abyssinica (niger-seed). *SINET: Ethiop. J.Sci*. **22**: 67-88.

Slatkin, M. (1987). Gene flow and the geographic structure of natural populations. *Science* **236**: 787-792.

Takezaki, N., Nei, M. and Tamura, K. (2010). POPTREE2: Software for constructing population tree from allele frequency data and computing statistics with window interface. *Mol Biol Evol*. **27**: 747-752.

Tsumura, Y., Ohba, K. and Strauss, S.H. (1996). Diversity and inheritance of inter-simple sequence repeat polymorphisms in Douglas-fir (*Pseudotsuga menziesii*) and sugi (*Cryptomeria japonica*). *Theor. Appl. Genet*. **92**:40–45.

Vollmann, J. and Rajcan, I. (2009). Oil-crop breeding and genetics. In *Oil Crops* (pp. 1-30). Springer, New York, NY.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Lee, T., Hornes, M. and Zabeau, M. (1995). AFLP: A new technique for DNA fingerprinting. *Nucleicc Acids Res.***23**: 4407-4414.

Wei, W., Qi, X., Wang, L., Zhang, Y., Hua, W., Li, D., Lv, H. and Zhang, X., (2011). Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. *BMC genomics* **12**:451.

Weiss, E. A. (1983). *Oilseed crops*. Longman Group, London

Weiss, E. A. (2000). *Oilseed crops*. Blackwell Science Ltd, London.

Williams, J.G., Kubelik, A.R., Livak, K.J., Rafalski, J.A. and Tingey, S.V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleicc Acids Res.***18**: 6531-6535.

Wright, S. (1951). The genetic structure of populations. *Annls of Eugenics* **15**:323-354.

Yadav, S., Hussain, Z., Suneja, P., Nizar, M. A., Yadav, S. K. and Dutta, M. (2012). Genetic divergence studies in Niger (*Guizotia abyssinica*) germplasm. *Biomass and Bioenergy* **44**:64-69.

Yared Serste, De Ruyter de Wildt, M., Dijkxhoorn, Y. and Danse, M. (2011). Small-Scale Edible Oil Milling Operations: Alternative Business Models for Ethiopia. The Hague: LEI Memorandum 11-005.

Yohanes Petros, Merker, A. and Habtamu Zeleke (2007). Analysis of genetic diversity of *Guizotia abyssinica* from Ethiopia using ISSR markers. *Hereditas* **144**: 18-24.

Yohanes Petros, Merker, A. and Habtamu Zeleke (2008). 'Analysis of genetic diversity and relationships of wild Guizotia species from Ethiopia using ISSR markers'. *Genet. Resour. Crop Evol.* **55**: 451–458.

Zeinalzadeh-Tabrizi, H., Haliloglu, K., Ghaffari, M. and Hosseinpour, A. (2018). Assessment of genetic diversity among sunflower genotypes using microsatellite markers. *Mol. Biol. Res. Commun.***7**:143.

Zhang, J., Wang, X., Yao, J., Li, Q., Liu, F., Yotsukura, N., Krupnova, T.N. and Duan, D., 2017. Effect of domestication on the genetic diversity and structure of *Saccharina japonica* populations in China. *Scientific reports* **7**:1-11.

Zia, Z.U., Sadaqat, H.A., Tahir, M.H.N., Sadia, B., Bushman, B.S., Hole, D., Michaels, L. and Malik, W. (2014). Estimation the genetic diversity of sunflower using simple sequence markers. *Russ.J.Genet*. **50**: 498-507.

8. APPENDICES

Appendix 1.Seedling establishment and microsatellites polymorphism

Fig 1.1. A two-week old seedlings ready for DNA extraction

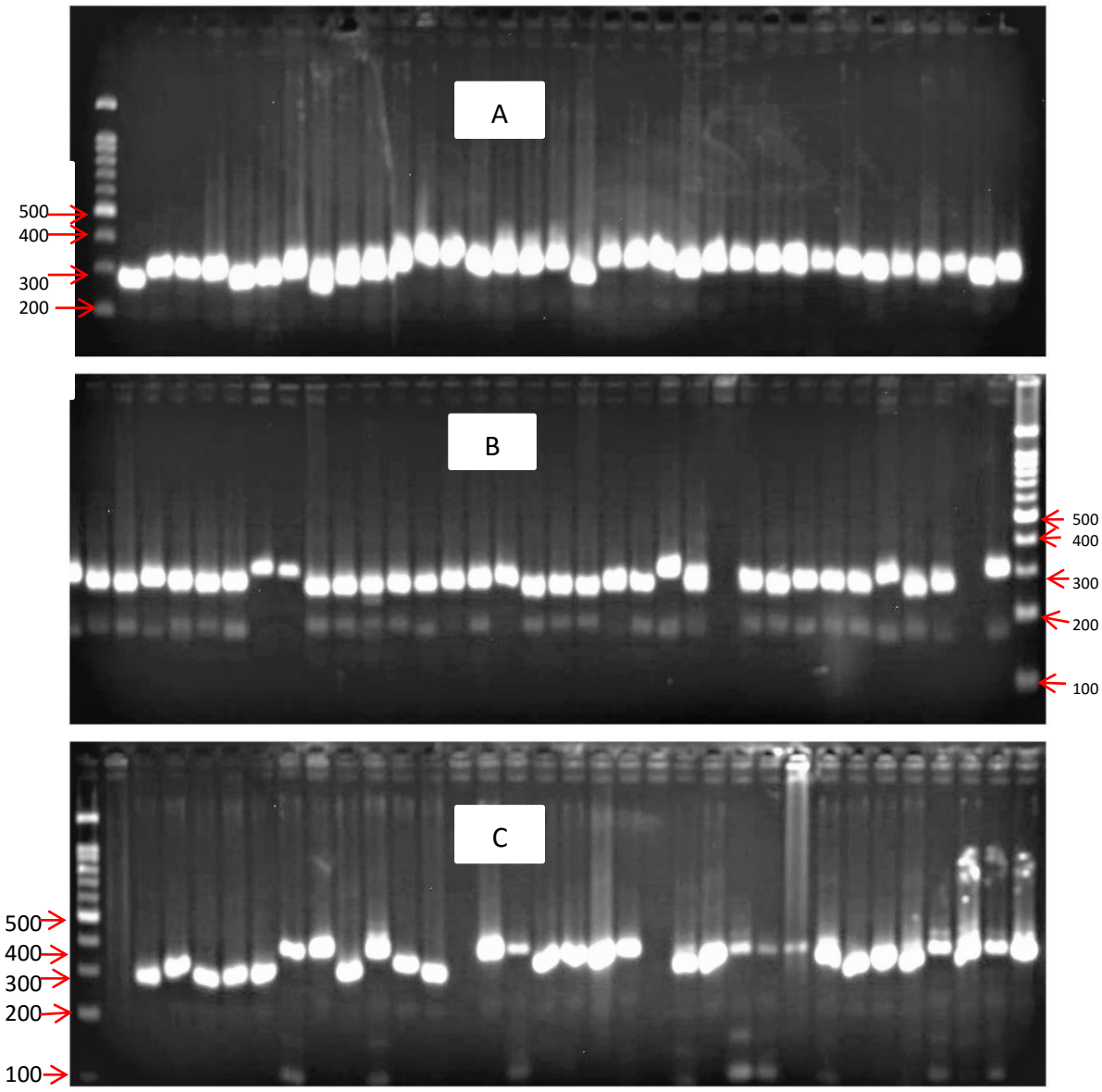


Fig 1.2. Gel-image of PCR products showing polymorphic markers (A, B and C) in some samples of noug with 100 bp DNA ladder runs at 3% agarose gel.

Appendix 2. Genomic DNA extraction procedures based on DArT protocol

Extraction buffer stock to make in 500ml

- sorbitol 0.35 M 31.9g sorbitol, 0.1 M TrisHCl pH 8.0 50 ml, 1M Tris-HCl 5 pH 8.0, 0.5 M EDTA pH 8.0, fill up to 500 ml dH2O

Lysis buffer stock to make 500 ml

- 100 ml 0.2 M Tris HCl pH 8.0, 50 ml 1M Tri HCl pH 8.0, 0.05 M EDTA pH 8.0, 0.5 M EDTA pH 8.0, 2M NaCl, 200 ml 5 M NaCl, 10 g of 2% CTAB, fill to 500 ml with dH20

SDS 5% (w/v)

Fresh buffer working solution

For example to make 120 ml:

To make 120 ml of working solution; add 50 ml of extraction buffer stock, add 0.6 g of sodiumdisulfite (sodium metabisulfite) and 2.4 g of PVP-40(K29-32) and dissolve with 50 ml of lysis buffer stock and 20 ml SDS stock.

For example to make 30 ml:

To make 30 ml of working solution; add 12.5 ml of extraction buffer stock, add 0.15 g sodiumdisulfite (sodium metabisulfite) and 0.6 g PVP-40 (K29-32) and dissolve with 12.5 ml lysis buffer stock and 5 ml SDS stock.

*On standing, this buffer can separate into two layers. Until connecting to the extraction tube, heat to 65 ^{o}C and shake instantly.

Procedures

For 2 ml Eppendorf tubes:

- Using a Geno Grinder, ground the required volume (consistent across all samples) of plant material to a fine powder.
- Suspend the fine powder in 1 ml ‘'fresh buffer" solution stored at 65 ^{o}C (make sure there are no clumps and vortex if necessary),

- Incubate at 65 ^{o}C for 1 hour (extend for an additional 30 minutes if necessary), invert tubes with gentle shaking every 20 minutes.
- Allow to cool for 5 minutes before adding 1ml of chloroform: isoamyl alcohol (24:1) combination and thoroughly blend for 30 minutes.
- Spin for 20 minutes at 10000 x g, RT,
- Remove the supernatant and wash the pellet with 2 ml EtOH (70%)
- Remove the EtOH, dry the pellet, and dissolve it in 250 µl of 1XTE (10Nm Tris HCL PH 8.0, 1nM EDTA PH 8.0).
- Check the DNA quality and quantity on a 0.8% agarose gel.

Appendix 3. List of Noug accessions used for the study and their geographic origins

SN	Accession code	Administrative region/Zone	Woreda/District	Locality/collection site	Latitude	Longitude	Altitude (masl)
1	15137	Gurage	Goro	Goro	08-32-00-N	37-59-00-E	2030
2	9970	West Tigray	Adi Daero	Myni 30 km Enda-Selasi to Adi Daero	14-10-78-N	38-13-90-E	1872
3	15041	South Wollo	Kalu	Kombolcha			1660
4	15148	East Tigray	Wukro	Agula	13-41-00-N	39-36-00-E	2000
5	212493	North Shewa	Lay betna tach bet	Weira Amba	09-58-00-N	38-59-00-E	2000
6	15149	Central Tigray	Laelay Maychew	Axum	14-07-00-N	38-48-0 -E	2050
7	15150	East Tigray	Hawzen	Hawzen	13-59-00-N	39-26-00-E	2300
8	15076	West Shewa	Dendi	Ginchi	09-00-00-N	38-09-00-E	2340
9	23584	Metekel	Wenbera	Senkera 1 from wenbera to senkara zuria 26 km	10-34-41-N	35-43-01-E	2541
10	15627	Illubabor	Bedele	About 20km. from bedele road to arjo	08-34-00-N	36-22-00-E	1920
11	15027	West Shewa	Ejerie (Addis Alem)	Welenkomi			2400
12	17469	West Shewa	Welisso	Senkele hale marami kebele haya village about 3 km from dile	08-39-52-N	38-04-12-E	2391
13	15007	Illubabor	Ale	Ale	08-27-00-N	36-23-00-E	1980
14	212497	North Shewa	Lay betna tach bet	Gonde Meskel	10-11-00-N	38-44-00-E	2530
15	15181	Wollo/Oromiya special	Bati	Bati	11-11-00-N	40-01-00-E	NA
16	15015	East Wellega	Diga leka	Diga			2088

17	219891	Central Tigray	Naeder Adet	Zahay Axum Market			NA
18	15078	Illubabor	Bedele	Bedele	08-27-00-N	36-23-00-E	1980
19	207597	Central Tigray	Laelay Maychew				NA
20	15770	Wollo/Oromiya special	Chefe Golana Dewera	Debegna kebele dawira kamido wirda 24km. from kemese to bora			1660
21	15110	West Harerge	Tulo				1900
22	23568	Metekel	Guba	Bakambel from Guba to Beshata 21 km			664
23	15096	North Gonder	Dembiya				1970
24	215604	Eritrea	Asmara				2325
25	200438	Arsi	Sherka				2600
26	15104	West Gojjam	Jabi Tehnan				1980
27	9302	Arsi	Jeju		8 -14-17-N	39-33-31-E	2533
28	212498	North Shewa	Weremo Wajetuna Mid	Kewetina Temamit 10km from Meragno on the way to Were-Ilu	10-15-00-N	39-15-00-E	2640
29	241916	West Gojjam	Achefer	18km North of Durbete town the right of Kunzla Tana town	11-38-00-N	37-10-00-E	1990
30	15112	West Harerge	Habro				1900
31	15027	West Shewa	Ejerie (Addiss Alem)	Welenkomi			2400
32	15116	North Shewa	Debrebirhan Zuria				2777
33	15529	Metekel	Bullen	About 60km.from bulen			2630
34	17477	Arsi	Sherka	Hela mekana kebele usmana village about 7 km from gobessa	07-38-59-N	39-32-42-E	2331

35	213876	Jimma	Limu Seka				1750
36	15023	East Sheaw	Shashemene	Shashemene	07-12-00-N	38-38-00-E	1980
37	15151	East Tigray	Hawzen	Hawzen	13-59-00-N	39-26-00-E	2300
38	29471	Asossa	Menge	Kashief found on the main road of Homesha to Menge town	10-19-43-N	34-42-41-E	1358
39	17471	West Shewa	Wenchi	Haro senkole kebele woldo village about 6 km from chitu town	08-39-06-N	37-53-54-E	2259
40	15557	Arsi	Sude	Dereba 1km. on the way to Kula	08-23-00-N	39-41-00-E	2600
41	18025	Agew Awi/Gojjam	Faggeta Lekoma	Abla ledeta 18 km from Kossober to Dangella	11-05-16-N	37-53-04-E	2442
42	212491	North Shewa	Siyadebrina Wayu Ens	Kecha 28kms away from Muke Ture on the way to Lemi	09-39-00-N	39-00-00-E	2630
43	15769	Wollo/oromiya special	Chefe Golana Dewera	5km. from Kemeseto Denkeye gelam (east of the road)	10-49-00-N	39-53-00-E	1600
44	15186	North Wollo	Guba Lafto	Hara	11-50-00-N	39-44-00-E	
45	200437	Arsi	Chole				2600
46	200427	Bahirdar special/Gojjam	Bahirdar	Woramit farmer's Association market NA			NA
47	15122	North Shewa	Geramidirna Keya Geb	Mehal Meda	10-14-00-N	39-36-00-E	2960
48	17744	East Wellega	Wayu Tuqa		08-57-51-N	36-41-50-E	1655
49	15171	North Gonder	Gonder Zuria	Azezo	12-33-00-N	37-27-00-E	2080
50	15164	North Wollo	Guba Lafto	Sirinka	11-43-00-N	39-35-00-E	NA
51	15165	South Wollo	Kutaber	Kutaber	11-25-00-N	39-34-00-E	2700
52	215602	Eritrea	Asmara	Suburb			2325
53	215603	Eritrea	Asmara	Suburb			2325

54	208391	Bahirdar Special/Gojjam	Bahirdar	Dekie 7Km away from Bahir Dar on the way to Addis Abeba	11-30-00-N	37-18-0 -E	1920
55	234124	Central Tigray	Adwa	Nebela 8km. from Adwa Asmara	14-17-00-N	38-49-00-E	1950
56	17747	East Wellega	Limu		09-50-48-N	36-31-47-E	2165
57	215605	Eritrea	Asmara	Suburb			2325
58	243782	West Tigray	Tselemti	15.1 Km from Boya to Tseada karni kebale			1200
59	17468	Gurage	Sodo	Negesa kebele kormachi village 5 km from bui town towards	08-16-53-N	38-31-43-E	1972
60	15064	Illubabor	Metu	Metu	09-18-00-N	35-37-00-E	1710
61	243779	Central Tigray	Kola Temben	6.3 Km from the junction of Abyi Ady road to Atakilty kebale	13-37-00-N	39-00-00-	1990
62	29476	Kamash	Balogiganfoy	Sense found on the main road Nekemte to Asossa	09-02-10-N	36-08-45-E	1195
63	215020	Jimma	Mana	Tessaba farmers Association			1750
64	243794	Agew Awi/Gojjam	Banja	21 Km west from Injibara to chagni	10-57-00-N	36-52-00-E	2200
65	18029	East Gojjam	Gozamen	Wonka 10 km Debre Markos to Bahir Dar the junction to cherte	10-20-45-N	37-39-55-E	2503
66	203198	West Shewa	Alem Gena	5Kms from Sebeta to Gibe			2060
67	200445	Arsi	Aseko				2600
68	15560	Bale	Sinana Dinsho	Gabo Jena 23Km from Robe to Sedik	07-41-00-N	39-45-00-E	2430
69	15039	Southern Tigray	Ofla	Korem			NA
70	18028	West Gojjam	Fenoteselam	Aba Nuguse 12 km Fnoteselam to Debre Markos	10-40-41-N	37-19-31-E	1830
71	15552	Bale	Gaserana Gulelcha	About 33km. from Dega on the way to Gasre	07-22-00-N	40-11-00-E	NA
72	230809	East Harerge	Girawa	Dogo Mudifara Near the Tawn	09-01-00-N	41-58-00-E	2050

73	244989	East Wellega	Amuru Jarte	8.5km N of Jardega town			2343
74	228428	Agew Awi/Gojjam	Guangua	Dangula			2541
75	243795	Agew Awi/Gojjam	Guangua	53.6 Km from Injibara to chagni	10-57-00-N	36-47-00-E	1750
76	207969	Jimma	Sorku	Gibe close to the Gibe bridge on the way to Jimma	07-50-00-N	37-10-00-E	1100
77	234123	Central Tigray	Adwa	Akasin	14-12-00-N	38-51-00-E	2110
78	237927	Bale	Ginir	Dolo Sebro Market			NA
79	202455	Wollo/Oromiya special	Chefe Golana Dewera	22km. From Chaeffa Robi to Artuma			1795
80	208392	West Gojjam	Merawi	Kolkotima 29Km away from Bahir Dar on the way to Addis Abeba	11-15-00-N	37-22-00-E	2100
81	200441	Arsi	Jeju				NA
82	17467	Gurage	Sodo	Eurshi kebele dereki village 5 km from bui town to words	08-20-02-N	38-31-57-E	1985
83	15060	Bale	Sinanana Dinsho	Gabo jena 23km. from robe on sedika road	07-41-00-N	39-45-00-E	2430
84	15535	South Wollo	Dessie	About 6.5km. from dessie wegeltena junction	11-28-00-N	39-17-00-E	1940
85	15113	West Harerge	Habro				NA
86	15579	Asossa	Asossa	Asosa market			NA
87	29478	Asossa	Bambasi	Arba Soge found around 12km of northern Bambasi	09-48-58-N	34-47-35-E	1420
88	15119	North Shewa	Efratana Gidim				NA
89	15035	West Wellrga	Gimbi	Gimbi	09-09-00-N	35-52-00-E	1960
90	17468	Gurage	Sodo	Negesa kebele kormachi village 5 km from bui town towards	08-16-53-N	38-31-43-E	1972

91	19048	Bale	Gasera	Darie Beneha Gurada about 8 km NE of Gasera town	07-24-32-N	40-15-02-E	2398
92	9296	Wollo/Oromiya Special	Kemise		10-48-14-N	39-49-36-E	1466
93	15166	North Wollo	Kobo	Kobo	12-10-00-N	39-38-00-E	
94	15133	Hadiya	Limo	Hosana	07-33-00-N	37-55-00-E	2350
95	15019	West Harerge	Chiro	Asbe Teferi	09-05-00-N	40-51-00-E	NA
96	15061	West Shewa	Ejerie(Addis Alem)	Chalia Goda Dalo			2400
97	237518	Central Tigray	Tahtay Maychew	Tach mychew	14-10-00-N	38-45-00-E	2200
98	15562	Bale	Gingir	About 2km. from Ginir on the way to jara	07-09-00-N	41-00-00-E	2010
99	15016	Illubabor	Ale	Ale	08-27-00-N	36-23-00-E	
100	9297	Wollo/Oromiya Special	Artuma Furdina Jile		10-34-28-N	39-55-3 -E	1485
101	23583	Metekel	Wenbera	Senkera 1 from wenbera to senkara zuria 25 km	10-34-41-N	35-43-01-E	2541
102	207968	Jimma	Limu Seka	Olmie Abo about 30Km away from Jima on the way to Bonga	07-34-00-N	36-28-00-E	1750
103	202452	Wollo/Oromiya Special	Chefe Golana Dewera	4km. From Cheffa Robi to Artuma			1760
104	229141	North Shewa	Weremo Wajetuna Mid	Budilib kebele market	14-12-00-N		NA
105	237515	Central Tigray	Enticho	12km. from Inticho to Adigrat		38-56-00-E	2010
106	15003	West Shewa	Ambo	Ambo			NA
107	237928	Bale	Gaserana Gololcha	Wote chamo 5km. from Gassera to Ali town			2330
108	15502	Arsi	Robe	11.6k. from robi ticho road Bokie	07-51-00-N	39-35-00-E	2500

109	9977	Agew Awi/Gojjam	Dangela	Tukur 3 km Dangela to D/Markos	11-14-02-N	36-51-39-E	2146
110	15784	South Wollo	Tehuledere	7km. from Hayik to bist			1930
111	15516	Agew Awi/Gojjam	Dangela	9km away from dangla on the way to chara			2065
112	15776	Wollo/Oromiya special	Bati	Kersa 39km. for kombolcha to bati			1590
113	15503	Arsi	Hitosa	Huruta market source from dodie about 10km.towards arbgebeya	08-09-00-N	39-21-00-E	2050
114	15094	North Gonder	Dabat				2500
115	236370	East Wellega	Abay Chomen	Achane 16Km. from Fincha to Shambo			2430
116	15715	South Gonder	Kemkem	25k. from Adiszen road to Gonder	12-28-00-N	37-28-00-E	1920
117	15547	North Gonder	Alefa	Bedrkoon	12-13-00-N	37-30-00-E	1910
118	213076	South Wollo	Jama	Kebela 018 farmers association			NA
119	29937	Bale	Gassera	Sendebo oda is located about 19km south of Gassera town	07-16-53-N	40-13-24-E	2352
120	202459	North Wollo	Habru	46km. From Hayk to Weldiya			1580
121	29474	Kamash	Balogiganfoy	Say dalecha found around 31km from Balo-Soge on the Arjo town on main road	09-12-40-N	36-12-27-E	1231
122	29473	Kamash	Balogiganfoy	Dedesa found 12km on the north of Soge town	09-22-34-N	36-05-59-E	1192
123	15501	Gurage	Meskana Mareko	83.9km. south of Addis on the road to butajira	08-29-00-N	38-38-00-E	2250
124	15193	Hadiya	Limo	Achamo	07-39-00-N	38-05-00-E	NA
125	212839	Bale	Goro	Beliya Burka			NA

126	202451	Wollo/Oromiya special	Chefe Golana Dewera	1km. From Cheffa Robi to Artuma			1520
127	202454	Wollo/Oromiya special	Artuma Fursina Jile	14km. from Cheffa Robi surrounding			1730
128	207967	Illubabor	Gechi	Gole about 10Km E of Gechi market place	08-20-00-N	36-26-00-E	1912
129	207966	Asossa	Asossa	Ashora Horal Gaseb 41Km away from Asosa on the way to Gizen	10-16-00-N	34-41-00-E	1340
130	23572	Metekel	Dangur	Bengus memder from Dangur to Bengus 42 km	11-11-53-N	35-50-31-E	774
131	229778	Metekel	Dibate	Addis Zemen farmer ass.			1600
132	35241	Central Gonder	Misrak Dembiya	about 8km south west of Kolla diba town	12-25-54-N	37-17-12-E	1869
133	245007	Illubabor	Bedele	3.1km NW of Kone town			1982
134	15006	Illubabor	Bedele	Bedele			NA
135	23580	Metekel	Wenbera	Tach Lope from wenbear to Lope 6 km	10-36-29-N	35-40-23-E	2507
136	236188	Southern Tigray	Ofla	Market			NA
137	35242	Central Gonder	Mirab Dembiya	about12km west of Kolla diba	12-21-50-N	37-15-07-E	1912
138	202456	Wollo/Oromiya special	Chefe Golana Dewera	Jarango, Kemissie Market			NA
139	15595	South Wollo	Dessie	About 6.5km. from dessie wegeltena junction	11-28-00-N	39-17-00-E	1940
140	29475	Kamash	Ballogiganfoy	Say dalecha found around 30km from Balo-Soge on the Arjo town	09-12-32-N	36-12-43-E	1227
141	202276	East Gojjam	Dejen	Kurar			1850
142	229140	North Shewa	Laybet ena Tach Bet	Kabi Derami 4km, N of Gundo meskal			2570
143	245009	Illubabor	Gechi	4.4km NE of Gachi town			1912

144	211509	West Wellega	Ayra Guliso	Mendo Galiso, 3km SN of Guliso	09-09-0 -N	35-26-00-E	1960
145	35235	South Gonder	Derra	about 6km wes of Andebesa town	11-45-36-N	37-36-14-E	1940
146	15109	Agew Awi/Gojjam	Dangela				2100
147	29472	Asossa	Kurmuk	Abachi founf around 5km of eastern	10-31-09-N	34-32-59-E	1346
148	15005	West Gojjam	Jebi Tehuan				NA
149	15117	North Shewa	Moretna Jiru				2700
150	228408	Agew Awi/Gojjam	Dangela	Gissa			NA
151	211508	West Wellega	Boji	Amuma Agello, 3km N.of Billa Tawn	09-23-00-N	35-35-00-E	1960
152	15124	West Shewa	Becho	Asgori	08-47-00-N	38-20-00-E	2050
153	229775	Metekel	Dibate	Addis Zemen farmer ass.			1600
154	23570	Metekel	Guba	Ketena 3 from Guba to Almehal 54 km	11-35-12-N	35-09-42-E	664
155	23569	Metekel	Guba	Albeda from Guba to Albeda 46 km	11-24-46-N	35-20-17-E	709
156	15585	South Gonder	Farta	Kemer Dengay market	11-48-00-N	38-28-00-E	3114
157	15001	North Gonder	Gonder Zuria	Azezo	12-33-00-N	37-27-00-E	2080
158	15086	West Tigray	Medebay Zana	Slehleka	14-06-00-N	38-17-00-E	1891
159	15085	Central Tigray	Laelay Maychew	Axum	14-07-00-N	38-48-00-E	2134
160	15118	North Shewa	Efratana Gidim				NA
161	15087	West Tigary	Tahtay Koraro	Indesilase	14-06-00-N	38-17-00-E	1891

Appendix 4. Summary of allelic patterns

Sample code	Accessions	Populations	No. Loci with Private Alleles	Loci with Private Alleles
27	9302	1	1	GA018
34	17477	1	1	GA018
40	15557	1	1	GA018
45	200437	1	1	GA018
22	23568	2	1	GA035
62	29476	2	1	GA013
86	15579	2	1	GA013
101	23583	2	1	GA029
129	207966	2	2	GA029 GA144
130	23572	2	1	GA144
131	229778	2	1	GA029
147	29472	2	1	GA035
155	23569	2	1	GA081
98	15562	3	1	GA029
6	15149	4	1	GA037
19	207597	4	1	GA037
158	15086	4	1	GA117
161	15087	4	1	GA117
23	15096	5	1	GA018
49	15171	5	2	GA018 GA035
46	200427	6	1	GA035
54	208391	6	2	GA018 GA117
64	243794	6	1	GA117
70	18028	6	1	GA150
75	243795	6	1	GA150
109	9977	6	1	GA150
148	15005	6	2	GA018 GA144
150	228408	6	1	GA144
5	212493	10	1	GA037

8	15076	10	1	GA029
11	15027	10	2	GA013 GA029
14	212497	10	1	GA013
32	15116	10	1	GA029
36	15023	10	3	GA029 GA037 GA162
42	212491	10	1	GA162
47	15122	10	1	GA029
66	203198	10	1	GA029
88	15119	10	1	GA162
96	15061	10	1	GA162
104	229141	11	1	GA029
106	15003	11	2	GA029 GA162
142	229140	11	1	GA018
149	15117	11	2	GA035 GA162
152	15124	11	1	GA162
160	15118	11	1	GA018
73	244989	11	1	GA117
15	15181	12	1	GA077
20	15770	12	1	GA077
44	15186	12	1	GA144
79	202455	12	2	GA077 GA139
100	9297	12	1	GA013
118	213076	12	1	GA144
126	202451	12	1	GA144
127	202454	12	1	GA144
138	202456	12	1	GA144
139	15595	12	2	GA013 GA077

Appendix 5. Noug accessions and population codes used for genetic diversity and population structure analysis

Sample code	Accessions code	Population Name	Population code
1	15137	SSNP	13
2	9970	Tigray	4
3	15041	Wollo	12
4	15148	Tigray	4
5	212493	Shewa	10
6	15149	Tigray	4
7	15150	Tigray	4
8	15076	Shewa	10
9	23584	B/G	2
10	15627	Illubabor	8
11	15027	Shewa	10
12	17469	Shewa	10
13	15007	Illubabor	8
14	212497	Shewa	10
15	15181	Wollo	12
16	15015	Wellega	11
17	219891	Tigray	4
18	15078	Illubabor	8
19	207597	Tigray	4

20	15770	Wollo	12
21	15110	Harerge	7
22	23568	B/G	2
23	15096	Gonder	5
24	215604	Eritrea	14
25	200438	Arsi	1
26	15104	Gojjam	6
27	9302	Arsi	1
28	212498	Shewa	10
29	241916	Gojjam	6
30	15112	Harerge	7
31	15027	Shewa	10
32	15116	Shewa	10
33	15529	B/G	2
34	17477	Arsi	1
35	213876	Jimma	9
36	15023	Shewa	10
37	15151	Tigray	4
38	29471	B/G	2
39	17471	Shewa	10
40	15557	Arsi	1
41	18025	Gojjam	6

42	212491	Shewa	10
43	15769	Wollo	12
44	15186	Wollo	12
45	200437	Arsi	1
46	200427	Gojjam	6
47	15122	Shewa	10
48	17744	Wellega	11
49	15171	Gonder	5
50	15164	Wollo	12
51	15165	Wollo	12
52	215602	Eritrea	14
53	215603	Eritrea	14
54	208391	Gojjam	6
55	234124	Tigray	4
56	17747	Wellega	11
57	215605	Eritrea	14
58	243782	Tigray	4
59	17468	SNNP	13
60	15064	Illubabor	8
61	243779	Tigray	4
62	29476	B/G	2
63	215020	Jimma	9

64	243794	Gojjam	6
65	18029	Gojjam	6
66	203198	Shewa	10
67	200445	Arsi	1
68	15560	Bale	3
69	15039	Tigray	4
70	18028	Gojjam	6
71	15552	Bale	3
72	230809	Harerge	7
73	244989	Wellega	11
74	228428	Gojjam	6
75	243795	Gojjam	6
76	207969	Jimma	9
77	234123	Tigray	4
78	237927	Bale	3
79	202455	Wollo	12
80	208392	Gojjam	6
81	200441	Arsi	1
82	17467	SNNP	13
83	15060	Bale	3
84	15535	Wollo	12
85	15113	Harerge	7

86	15579	B/G	2
87	29478	B/G	2
88	15119	Shewa	10
89	15035	Wellega	11
90	17468	SNNP	13
91	19048	Bale	3
92	9296	Wollo	12
93	15166	Wollo	12
94	15133	SNNP	13
95	15019	Harerge	7
96	15061	Shewa	10
97	237518	Tigray	4
98	15562	Bale	3
99	15016	Illubabor	8
100	9297	Wollo	12
101	23583	B/G	2
102	207968	Jimma	9
103	202452	Wollo	12
104	229141	Shewa	10
105	237515	Tigray	4
106	15003	Shewa	10
107	237928	Bale	3

108	15502	Arsi	1
109	9977	Gojjam	6
110	15784	Wollo	12
111	15516	Gojjam	6
112	15776	Wollo	12
113	15503	Arsi	1
114	15094	Gonder	5
115	236370	Wellega	11
116	15715	Gonder	5
117	15547	Gonder	5
118	213076	Wollo	12
119	29937	Bale	3
120	202459	Wollo	12
121	29474	B/G	2
122	29473	B/G	2
123	15501	SNNP	13
124	15193	SNNP	13
125	212839	Bale	3
126	202451	Wollo	12
127	202454	Wollo	12
128	207967	Illubabor	8
129	207966	B/G	2

130	23572	B/G	2
131	229778	B/G	2
132	35241	Gonder	5
133	245007	Illubabor	8
134	15006	Illubabor	8
135	23580	B/G	2
136	236188	Tigray	4
137	35242	Gonder	5
138	202456	Wollo	12
139	15595	Wollo	12
140	29475	B/G	2
141	202276	Gojjam	6
142	229140	Shewa	10
143	245009	Illubabor	8
144	211509	Wellega	11
145	35235	Gonder	5
146	15109	Gojjam	6
147	29472	B/G	2
148	15005	Gojjam	6
149	15117	Shewa	10
150	228408	Gojjam	6
151	211508	Wellega	11

152	15124	Shewa	10
153	229775	B/G	2
154	23570	B/G	2
155	23569	B/G	2
156	15585	Gonder	5
157	15001	Gonder	5
158	15086	Tigray	4
159	15085	Tigray	4
160	15118	Shewa	10
161	15087	Tigray	4