

Clustering Analysis of Ethiopian Safflower (*Carthamus tinctorius*) Using ISSR Markers

Baye Wodajo¹, Feysal Bushira Mustefa², Kassahun Tesfaye³

¹Woldia University, P.O. box 400, North Wollo, Ethiopia

E-mail: bavewodajo27@vmail.com

²Haramaya University; E-mail: bushfey@gmail.com

³Addis Ababa University, P.O. box 1176, Addis Ababa, Ethiopia

E-mail: kassahuntesfave@yahoo.com

Abstract- Safflower, *Carthamus tinctorius*, L. is an oilseed crop that belongs to the family Asteraceae. The genus *Carthamus* is comprised of 25 species including the only cultivated species of *Carthamus tinctorius*. So far, the characterization of safflower using molecular markers has been limited. The objective of this study was to examine the cluster analysis of safflower accessions collected from different regions of Ethiopia using ISSR molecular markers. For this purpose, seeds of seventy land race accessions collected from four administrative regions of Ethiopia (Amhara, Oromia, Tigray and SNNPR) were obtained from the EIB and grown in green house at Addis Ababa University, Faculty of Life Science. DNA was extracted from a bulk leaf sample of five randomly selected plants per accession using a triple CTAB extraction technique. Four primers were selected. The four selected ISSR primers produced a total of 43 bands across the 70 safflower accessions. The number of amplified fragments with ISSR primers ranged from 6 to 15 per primer with varied in size of 100 to 1000 base pairs. The cluster analysis based on ISSR data Safflower individuals assembled from different localities and regions observed to be spread all over the trees without forming strict grouping based on their geographic origin. However, some individual from Amhara and Oromia tends to form separate groups. The majority of the groups observed in UPGMA and NJ trees were intermixed individuals from SNNPR and Tigray with Oromia and Amhara populations. The results reveal the genotypes of safflower collected from different localities and regions of Ethiopia which had slightly grouping among accession.

Index Terms- *Carthamus tinctorius*, Safflower, genetic diversity, ISSR, molecular marker, cluster analysis, Ethiopia.

I. INTRODUCTION

Safflower, *Carthamus tinctorius*, L. is an oilseed crop that belongs to the family Asteraceae. The genus *Carthamus* has 25 species, of which *C. tinctorius* is the only cultivated one, and has $2n = 24$ chromosomes (Helm *et al.*, 1991). It is dicotyledonous, herbaceous, and annual plant. It has colorful flower heads, a deep taproot, and the production of white, oil-bearing seeds. It has adapted to grow in hot, dry climates, and well-drained soil (Smith, 1996). Safflower is believed to have been domesticated somewhere in the Fertile Crescent region over

4000 years ago (Ashri, 1975; Knowles, 1969). According to Vavilov (1951) considered Ethiopia, Afghanistan and India as the primary centers of origin of safflower. Currently, it is grown as an oilseed crop in over 60 countries worldwide, and India is the largest producer of safflower (Weiss, 2000).

Safflower is mainly cultivated for its seed, which is used primarily for edible oil. In the past, the crop is grown for its flowers used for coloring and flavoring foods, making dyes and medicine (Zhang, 1997). In Ethiopia, boiled and finely pounded safflower kernels are mixed with water and the supernatants is used to prepare the so called 'fitfit', which is used as fasting-food. Roasted seeds mixed with roasted chickpeas, barley or wheat, are eaten as a snack food in Ethiopia and Sudan (Dajue and Mündel, 1996). Ethiopian safflower is very much neglected and it is cultivated only as a minor oil crop with very limited information available on its genetic resources. In Ethiopia, safflower cultivation is mostly done by small scale farmers in well fertile and drained field, usually around homesteads (Edwards, 1991).

In current scenario, the DNA markers become the marker of choice for the study of crop genetic diversity (Karp *et al.*, 1997). However, agro-morphological markers were the first to be used for genetic diversity study of some safflower germplasm (Maryam *et al.*, 2011). RAPD markers were also applied to study the genetic diversity of *C. tinctorius* landraces germplasm collected from Iran (Mahasi *et al.* 2009; Pooran *et al.*, 2011) while ISSR marker has been used by Yang *et al.* (2007) and Pooran *et al.* (2011) to investigate relationship among *C. tinctorius* germplasm collected from different part of the world. None of the above studies either focused on Ethiopian germplasm or include samples from Ethiopia. This study provided valuable data to conservation attention through its ability to detect variation at the DNA level. Therefore, the main objective of this study was examining the cluster analysis of safflower (*C. tinctorius*) collected from Ethiopia using ISSR molecular markers.

II. MATERIALS AND METHODS

2.1. Plant materials and Genomic DNA extraction

Seeds of 70 safflower accessions collected from four different administrative regions of Ethiopia were obtained from Ethiopian Institute of Biodiversity (EIB), Addis Ababa, Ethiopia.

All the 70 safflower accessions seeds samples were grown in a greenhouse and fresh leaves from five week old plants were used for genomic DNA extraction. Five young leaves were collected separately from five randomly selected individual plants of each accession just before flowering and equal proportions of leaves were grinded for genomic DNA extractions. Genomic DNA was extracted from leaf samples according to Borsch *et al.* (2003) and later diluted to concentration of 1:5 .

2.2. Primer selection and optimization

A total of 10 ISSR primers (UBC 900) available at Genetics Laboratory, Addis Ababa University (AAU) were used for the initial variability test and reproducibility. Optimization and screening of primers was carried with one individual that was selected from each population. All pre-selected 10 primers were screened for reproducibility and polymorphism. Finally, two di (primer 812 and 818), one tetra- (primer 873) and one penta- (primer 880) –nucleotide primers (Table 1) were selected based on polymorphism and reproducibility.

Table 1: List of primers, annealing temperature, sequence motifs and amplification pattern

Primers	Annealing T°	Sequence	Amplification pattern
810	45	5'-GAGAGAGAGAGAGAT-3'	Reproducible but not polymorphic
812	45	5'-GAGAGAGAGAGAGAA-3'	Good
818	48	5'-CACACACACACACAG-3'	Good
824	48	5'-TCTCTCTCTCTCTCG-3'	Polymorphic but not reproducible
834	45	5'-AGAGAGAGAGAGAGYT-3'	Reproducible but not polymorphic
844*	48	5'-CTCTCTCTCTCTCTRC-3'	Polymorphic but not reproducible
872*	38	5'-GATAGATAGATAGATA-3'	Polymorphic but not reproducible
873*	45	5'-GACAGACAGACAGACA-3'	Good
874*	48	5'-CCCTCCCTCCCTCCCT-3'	Polymorphic but not reproducible
880**	45	5'-GGAGAGGAGAGGAGA-3'	Good

Source: Primer kit 900 (UBC 900); Single-letter abbreviations for mixed base positions: R= (A, G) Y = (C, T)

*Tetra-nucleotide primer, **Penta-nucleotide primer and the rest are di-nucleotides.

2.4. DNA amplification

ISSR PCR amplification based on four primers was carried out using the reaction mixture below. Amplification was carried out in a 25 µl reaction mixture containing 1µl template DNA, 13.45µl ddH₂O, 5.6µl dNTP (1.25mM), 2.6µl Taq buffer(10xThermopol reaction buffer), 1.25µl MgCl₂ (50mM), 0.6µl primer (20pmol/µl) and 0.5µl Taq Polymerase (5u/µl). The amplification was performed in a Biometra 2003 version 3.10 TPersonal using 48 well plates under 2 minutes preheating and initial denaturation at 94°C, followed by a regular cycling event of 40 cycles of 20 seconds denaturation at 94°C, 1 minute primer annealing at (45°C/ 480C) based on primers used, 1.30 minutes extension at 72°C and final extension for 7 minutes at 72°C. The PCR products were stored at 4°C until loading on gel electrophoresis.

2.5. Gel electrophoresis and visualization

The quantity and quality of genomic DNA were tested using agarose gel electrophoresis. Stock solution of 10X Tris Borate EDTA (TBE), commonly used electrophoresis buffer: 108 gm Trisbase; 55 gm Boric acid; 40ml EDTA, pH 8.57 components per liter was prepared and stored at room temperature. From the stock, working solution of 1X TBE prepared and used to prepare the gel as well as fill the electrophoresis tank. The ISSR gel with 1.67% was prepared by boiling 1xTBE in 500 ml Erlenmeyer flask in micro-oven for 2:30 minutes. After the agarose solution cooled down at room temperature, 2µl of ethidium bromide was added for better visualization of the gel. Then poured on a gel tray and the comb was inserted to the gel and left for about 30 minutes to set properly. The comb was carefully removed and

put the gel tray in to the electrophoresis tank properly filled with electrophoresis buffer, and then the amplified ISSR product of 8µl was loaded onto ISSR gel with 2µl loading dye. DNA ladder 100 bp was used to estimate molecular weight. The electrode was connected and the power supply was turned on; the voltage was adjusted at 100 V and left for three hours. The electrophoresis gel was photographed using UV dual-intensity transilluminator, Zenith, canon camera connected with computer. But gel picture was taken after staining with 25µl Ethidium bromide (10mg/ml) which was mixed with 250 ml double distilled water for 30 minutes and washed for 30 minutes using double distilled water. Different photographs using different lens aperture, were taken and saved for later scoring.

2.6. Data scoring and statistical analysis of diversity

Although a large number of fragments were generated from each primer, only clearly distinguishable and reproducible bands were selected and data were entered in a computer file as a binary matrix “0” for absence and “1” for presence of a band. Data from the ISSR studies was analyzed using various statistical programs. The un-weighted pair group method with arithmetic mean (UPGMA) was used to analyze and compare the individual genotypes (Sneath and Sokal, 1973) and generates phenogram using NTSYS- pc version 2.02 (Rohlf, 2000). The neighbor joining (NJ) method (Saitou and Nei, 1987) was used to compare individual genotypes and evaluate patterns of genotype clustering using Free Tree 0.9.1.50 Software (Pavlicek *et al.*, 1999). To further examine the patterns of variation among individual samples, a principal coordinated analysis (PCO) was performed based on Jaccard’s coefficient (Jaccard, 1908). The calculation of

Jaccard's coefficient was made with PAST soft ware version 1.18 (Hammer *et al.*, 2001). The first three axes were later used to plot three dimensional graph with STATISTICA version 6.0 software (Statistica soft, Inc.2001).

III. RESULTS AND DISCUSSIONS

Clustering analysis and relationships of safflower populations

The four safflower populations from Ethiopia (Amhara, Oromia, SNNPRs and Tigray) were studied using ISSR markers. Out of the 10 primers, four with clear and reproducible bands were selected and used in this study (Table 2). The molecular weight of the bands amplified using the four primers were in the range of 100 to 1000bp (Figure 1). The generated ISSR marker data were used to construct a dendrogram using Un-weighted Pair-Group Method using Arithmetic Averages (UPGMA) and Neighbor joining (NJ) using NTSYS and free tree program, respectively. Both UPGMA and NJ based dendrogram for individual safflower appeared to show different grouping. Safflower individual assembled from different localities and regions observed to spread all over the trees without forming strict grouping based on their geographic origin. However, in few cases some individual from Amhara and Oromia tends to form their own groups. The majority of the groups observed in UPGMA and NJ trees were intermixed individuals from SNNPR and Tigray with Oromia and Amhara populations. Generally, different groupings was clearly in both UPGMA and neighbor joining analysis except for eight individuals assembled from

Amhara forming separate group in both UPGMA and NJ. Both trees recovered almost the same tree topology with similar groupings, although few individuals appeared to escape from groups in both cases. Yang *et al.* (2007) has got similar result, whereby the 24 accession from Asia were scattered all over the tree and comparatively, the accessions originated from Europe were relatively grouped together. In this study genotypes that were grouped in the same cluster could be because of common seed sources, similar selection criteria by local farmers, and sharing a common parentage and convergent evolution. Accordingly, Amhara and Oromia, SNNPR and Tigray have clustered in a same group (figure 2). Individuals from these regions could have originated from a common ancestor and/or could also have the same seed sources or duplicates (Khan *et al.*, 2008). This could be due to extensive seed exchange across the regions during domestication and cultivation of safflower (Khan *et al.*, 2008).

Although some work based on morphological traits has been conducted to assess genetic variation in safflower genotypes, the results are still ambiguous, because phenotypic traits are affected by developmental stages and environmental conditions. The results of the present study are consistent with those of Yang *et al.* (2007), who emphasized that ISSR, is an effective marker system for detecting genetic diversity among safflower genotypes and provides useful information about the phylogenetic relationships. In line with the present study, the dispersion into the various groups appeared to be at random through a few accessions formed distinct clusters based on geographic origin.

Table 2: Fingerprint patterns generated using four ISSR primers; selected for this study.

Selected Primer	Repeated motif	No. of clear scorable bands	Amplification Pattern	ISSR gel observation
812	(GA) ₈ A	12	Good	Good with small smear
818	(CA) ₈ G	6	Good with smear	Good with smear
873	(GACA) ₄	15	Best	Good
880	(GGAGA) ₃	10	Good	Good
		43		

Source: primer kit 900 (UBC 900); Single-letter abbreviations for mixed base positions.

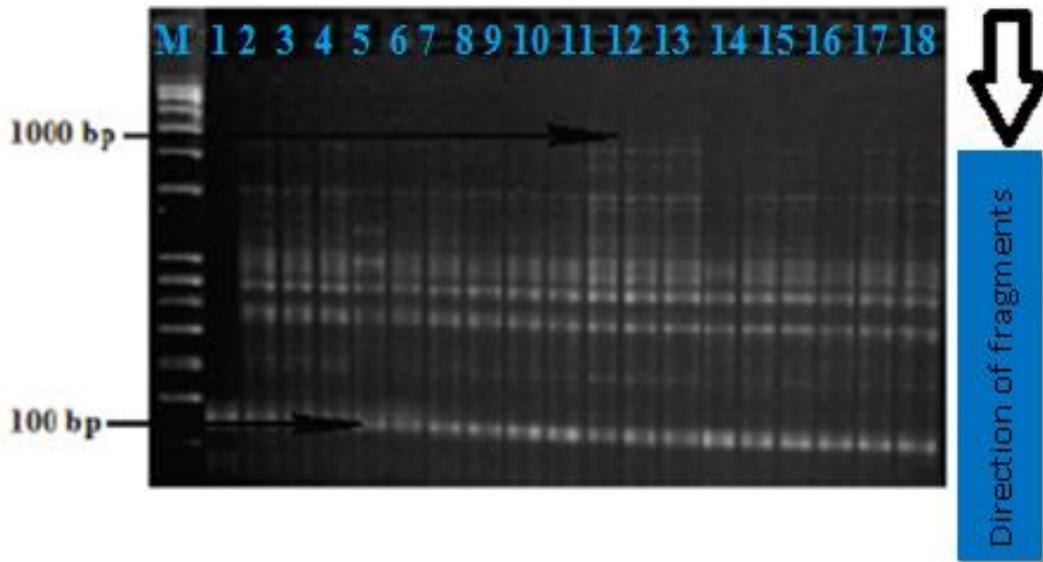


Fig. 1 ISSR amplification profile generated using primer 873 for *Carthamus tinctorius* accessions. Numbers represent the accessions from Amhara region. M: 100bp DNA ladder

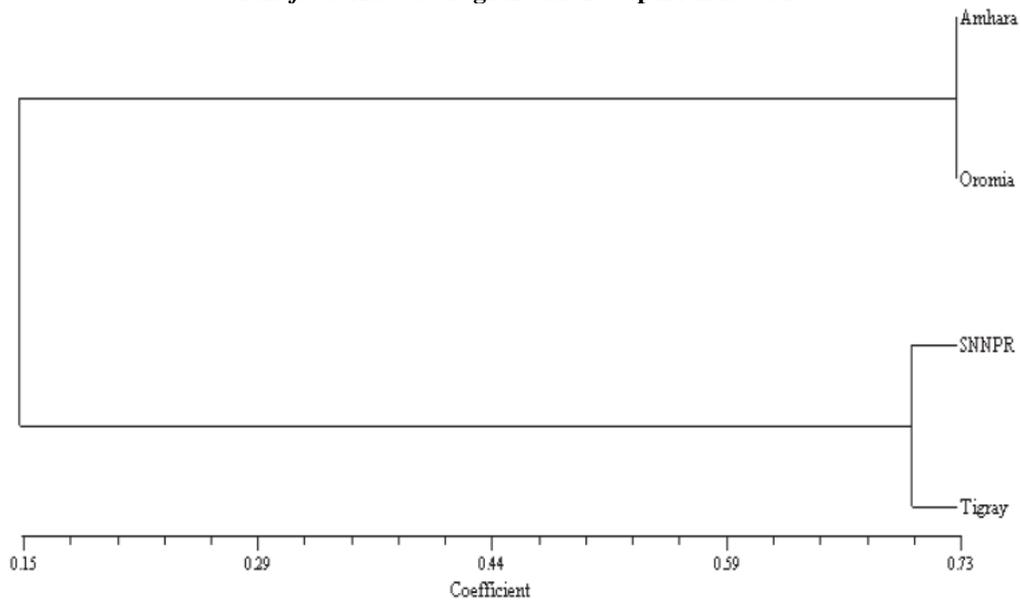


Figure 2: UPGMA based dendrogram for four safflower populations using Jaccard's similarity coefficient

Key:-AM=Amhara, OR= Oromia and TI=Tigray

Figure 3: UPGMA tree of 70 safflower individuals based on Jaccard's coefficients

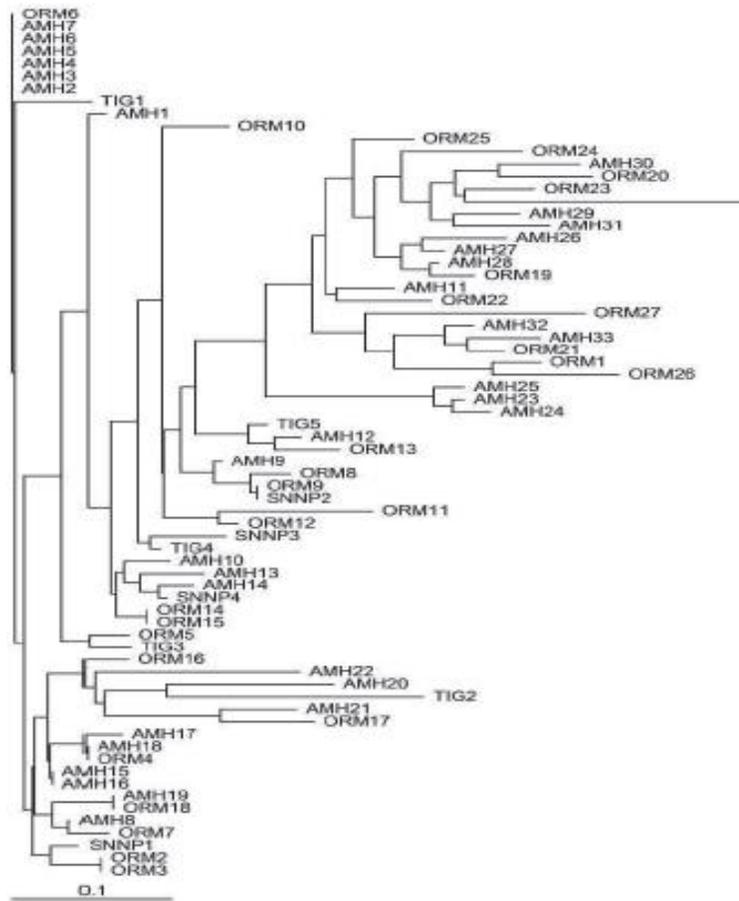


Figure 4: Neighbor-joining analysis of 70 individuals based on Jaccard's coefficients comparison of the presence-absence fingerprinting.

Key:-AMH=Amhara, ORO=Oromia, SNN=SNNPRs and TIG=Tigray

PCO Analysis (Principal Coordinate Analysis)

All the data obtained using four ISSR primers were used in PCO analysis using Jaccard's coefficient of similarity. The first three coordinates of the PCO with Eigen-values of 8.46, 2.92 and 1.92 with variance of 27.4%, 9.5% and 6.2%, respectively (Table 3) used to show the grouping of individuals using two and three dimensional coordinates. The four populations observed to form separate cluster and few of the individuals were dispersed all over both 2D and 3D space. However, some individuals from SNNPRs observed to be intermixed with Oromia and Amhara populations (Figure 5). In this investigation, there are weak

regional differentiations between safflower populations were observed and similarly patterns of grouping was also recovered in PCO UPGMA and NJ. The fact that they are grouped in the same cluster implies their close relationship in genetic characters and landrace lineages. This could also be due to the exchange of plant materials across the regions during the safflower cultivation and long distance seed transport for marketing. Therefore, their difference in cluster could imply their being originated from different sources, while grouping in the same cluster would mean genetic affinity among individuals in the same group.

Table 3: Eigen value and percent variance in the PCO

Principal Coordinate	Eigen value	Variance (%)
1	8.46	27.4
2	2.92	9.5
3	1.92	6.2
Total		43.1%

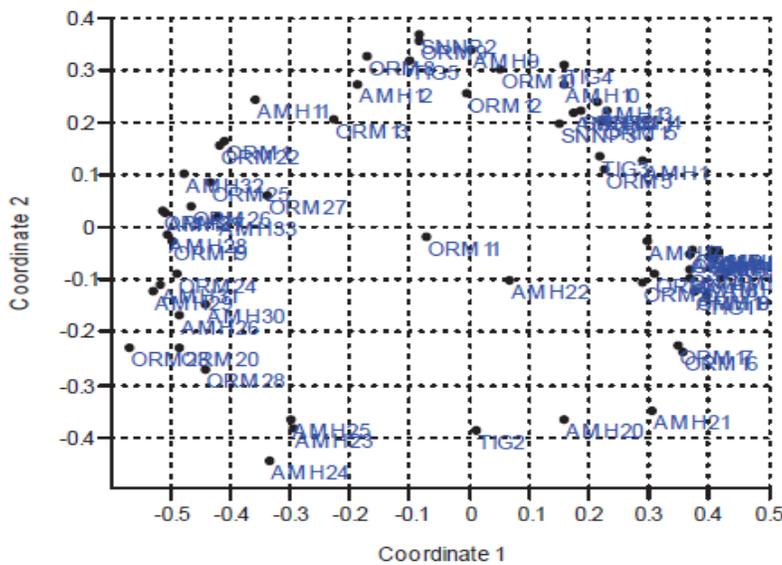


Figure 5: Two dimensional representations of principal coordinate analysis of genetic relationships among 70 individuals from four populations of safflower based on Jaccard's similarity coefficients.

Key: - AMH=Amhara, ORO=Oromia, SNN=SNNPRs and TIG= Tigray

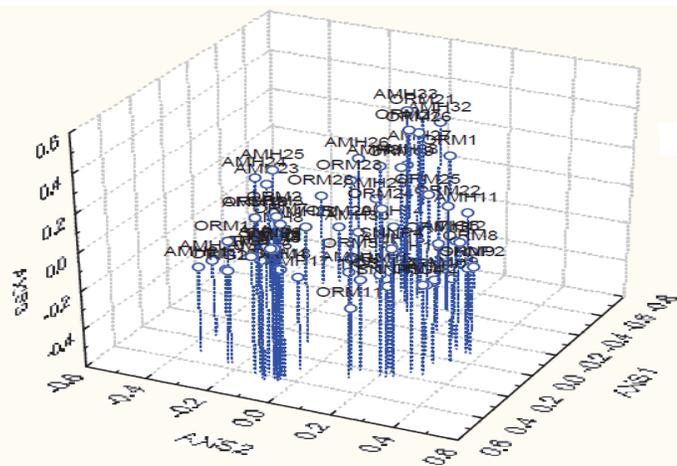


Figure 6: Three dimensional (3D) representation of principal coordinate analysis of genetic relationships among 70 individuals of four population safflower inferred from similarity matrix using the Jaccard's index.

Key: - AMH=Amhara, ORO=Oromia, SNN=SNNPRs and TIG= Tigray

IV. CONCLUSION

According to the results obtained from cluster analysis, genotypes of safflower collected from different localities and regions had no clear grouping based on geographic location. This could be due to the exchange of safflower seed across the regions during safflower cultivation and introduction from and to various geographic locations in Ethiopia. The clustering analysis based on ISSR marker was useful in revealing genetic relatedness between the safflower accessions in Ethiopia. However, other marker system such as SSR and AFLP along with sequence based markers should be utilized to further reveal the level of diversity in the Ethiopia cultivated safflower.

ACKNOWLEDGMENT

This research was supported by laboratory resources supplied from International Foundation of Sciences (IFS) under grant C/4790-1.

REFERENCES

- [1] Ash, G. J., Raman, R. and Crump, N. S. (2003). An investigation of genetic variation in *Carthamus lanatus* in New South Wales, Australia, using inter simple sequence repeats (ISSR) analysis. *Weed Res.*, 43:208-213.
- [2] Ashri, A. (1975). Evaluation of the germplasm collection of safflower (*Carthamus tinctorius* L.) Distribution and regional divergence for morphological characters. *Euphytica*, 24:651-659.
- [3] Borsch, T., Hilu, K.W., Quandt, D., Wilde, V., Neinhuis, C. and Barthlott, W. (2003). Noncoding plastid trnT-trnF sequences reveal a well resolved phylogeny of basal angiosperms. *J. Evol. Biol.* 16: 558-576.

- [4] Dajue, L. and Mündel, H. (1996). Safflower, *Carthamus tinctorius* L. Promoting the Conservation and use of Underutilized and Neglected Crops. Institute of Plant Genetics and Crop Plant Research, International Plant Genetic Resources Institute, Rome Italy.
- [5] Edwards, Sue B. (1991). Crops with wild relatives found in Ethiopia. In: Plant Genetic Resources of Ethiopia. Pp., 54-63 (Engels, J. M. M., Hawkes, J. G. and Melaku Worede, eds). Cambridge Univ. Press, Cambridge.
- [6] Hammer, O., Harper, D. A. T. and Ryan, P. D. (2001). PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*.4:9.
- [7] Hamrick, J. L. and Godt, M. J. W. (1989). Allozyme diversity in plant species. In: Plant Population Genetics, Breeding and Germplasm Source. Pp. 43-63, (Brown, A. H. D., Clegg, M. T., Kahler, A. L. and Weir, B. S. eds.)Sunderland, MA: Sinauer.
- [8] Helm, J. L., Schneiter, A. A., Riveland, N. and Bergman, J. (1991). Safflower Production. North Dakota State University, Fargo, USA.
- [9] Jaccard, P. (1908). Nouvelles recherches Sur la distribution florale. *Bull.Soc.Vaud.Sci. Nat.* 44: 223-270.
- [10] Jaradat, A. A. and Shahid, M. (2006). Patterns of phenotypic variation in a germplasm collection of *Carthamus tinctorius* L. from the Middle East. *Euphytica*, 53:225-244.
- [11] Khan, M.A., von Witzke-Ehbrecht, S., Maass, B.L. and Becher, H.C. (2008) Relationships among different geographical groups, agro-morphology, fatty acid composition and RAPD marker diversity in safflower (*Carthamus tinctorius*). *Genet. Resour. Crop Evol.*, 56: 19–30.
- [12] Karp, A., Kresovich, S., Bhat K. V., Ayand, W. G. and Hodgkin, T. (1997). Molecular Tools in Plant Genetic Resources Conservation: A guide to the technologies. International Plant Genetic Resources Institute, Rome, Italy.
- [13] Knowles, P.F. (1969). Centers of plant diversity and conservation of crop germplasm: Safflower. *Econ. Bot.*, 23:324.329.
- [14] Lynch, M. and Milligan, B.G. (1994). Analysis of population genetic structure with RAPD markers. *Mol. Ecol.*, 3: 91-99.
- [15] Mahasi, M. J., Wachira, F. N., Pathak, R. S. and Riungu T. C. (2009). Genetic polymorphism in exotic safflower (*Carthamus tinctorius* L.) using RAPD markers. *J. Plant Bre. and Crop Sci.*, 1: 8-12.
- [16] Maryam, S., Hossein, Z. and Ehdi, Y. (2011). Investigation of Genetic Variation in *Carthamus Tinctorius* L. Genotypes using Agro-Morphological Traits. *Journal of Research in Agricultural Science*, 2: 101- 108.
- [17] Pavlicek, A., Hrda, S. and Flegr, J. (1999). Free tree free ware program for construction of phylogenetic trees on the basis of distance data and bootstrap/Jack Knife analysis of the tree robustness. Application in the RAPD analysis of genus *Frenkelia*. *Folia Biologica*. 45:97-99.
- [18] Pooran G., Ahmad A. and Abdolmajid M. R. (2011). Genetic Variation in Safflower (*Carthamus Tinctorius* L.) For Seed Quality-Related Traits and Inter-Simple Sequence Repeat (ISSR) Markers. *Int. J. Mol. Sci.* 12: 2664-2677.
- [19] Rohlf, F.J. (2000). NTSYS-pc, Numerical Taxonomy and Multivariate Analysis System, Version 2.1. Exeter Software, New York.
- [20] Saitou, N. and Nei, M. (1987).The neighbor joining method: a new method for reconstructing phylogenetic trees. *Mol.Biol.Evol.*, 4: 406-425.
- [21] Seyed, A. S., Seyyed, S. P., Mohmmad, T. and Mahmoud K. (2010). Assessment of genetic variation among safflower (*Carthamus tinctorius* L.) accessions using agro- morphological traits and molecular markers. *J. Food, Agri. Enviro.*, 8: 616 . 625.
- [22] Sneath, P. H. A. and Sokal, R. R. (1973). Numerical Taxonomy. Freeman. San Francisco.
- [23] Stat soft, Inc (2001). Statistica data analysis system, Statistica software.
- [24] Vavilov, N. I. (1951). The Origin, Variation, Immunity and Breeding of Cultivated Plants. The Ronald Press Co., New York.
- [25] Weiss, E. A. (2000). Oil Seed Crops. Blackwell Science Ltd, Oxford, UK.
- [26] Yamasaki, M. , M. I., Tenaillon , I. V., Bi , S. G., Schroeder , H., Sanchez-Villeda , J. F., Doebley, B. S. and McMullen, M. D. (2005). A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. *Plant Cell*, 17: 2859. 2872.
- [27] Yang, Y.X., Wu, W., Zheng, Y. L., Chen, L., Liu, R. J. and Huang, C.Y. (2007). Genetic diversity and relationships among safflower (*Carthamus tinctorius* L.) analyzed by inter-simple sequence repeats (ISSRs). *Genet. Resour. Crop.Evol.*, 54:1043.1051.

AUTHORS

First Author – Baye wodajo, Woldia University, P.O. box 400, North Wollo, Ethiopia, E-mail: bayekobo@gmail.com, bayewodajo27@ymail.com
Second Author- Feysal Bushira Mustefa, Haramaya University, E-mail: bushfey@gmail.com

Third Author – Kassahun Tesfaye, Addis Ababa University, P.O. box 1176, Addis Ababa, Ethiopia, E-mail: kassahuntesfaye@yahoo.com