



GLOBAL JOURNAL OF SCIENCE FRONTIER RESEARCH: D
AGRICULTURE AND VETERINARY
Volume 20 Issue 6 Version 1.0 Year 2020
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 2249-4626 & Print ISSN: 0975-5896

Genetic Diversity and Population Structure Analysis of Tropical Soybean (*Glycine Max* (L.) Merrill) using single Nucleotide Polymorphic Markers

By Tonny Obua, Julius P. Sserumaga, Stephen O. Opiyo,
Phinehas Tukamuhabwa, Thomas L. Odong, Josiah Mutuku & Nasser Yao

Makerere University Kampala

Abstract- Soybean (*Glycine max* (L.) Merrill) is among the most important crops worldwide due to its numerous uses in feed, food, biofuel, and significant atmospheric nitrogen fixation capability. To understand the genetic diversity and population structure of tropical soybean germplasm, 89 genotypes from diverse sources were analyzed using 7,962 SNP markers. The AMOVA results showed low diversity among and high within the populations, while the polymorphism information content (PIC) was 0.27. Both phylogenetic and principal component analysis grouped the 89 soybean genotypes into three major clusters, while population structure grouped the soybean genotypes into two subpopulations. On the other, the average Roger genetic distances within the study population was 0.34. The low diversity reported in the studied soybean germplasm pool is particularly worrying, considering the new trends of climate change and the emergence of new pests and diseases of soybean. Therefore, in order to address these challenges and develop soybean varieties with desirable traits, there is a need to broaden the genetic base of tropical soybean through the importation of germplasm from other countries.

Keywords: *genetic diversity, population structure, single nucleotide polymorphism (SNP), soybean, tropical soybean genotypes.*

GJSFR-D Classification: FOR Code: 070199



Strictly as per the compliance and regulations of:



RESEARCH | DIVERSITY | ETHICS

© 2020. Tonny Obua, Julius P. Sserumaga, Stephen O. Opiyo, Phinehas Tukamuhabwa, Thomas L. Odong, Josiah Mutuku & Nasser Yao. This is a research/review paper, distributed under the terms of the Creative Commons Attribution 4.0 Generic License <https://creativecommons.org/licenses/by/4.0/>, permitting all commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Genetic Diversity and Population Structure Analysis of Tropical Soybean (*Glycine Max* (L.) Merrill) using single Nucleotide Polymorphic Markers

Tonny Obua ^α, Julius P. Sserumaga ^σ, Stephen O. Opiyo ^ρ, Phinehas Tukamuhabwa ^ω, Thomas L. Odong [¥], Josiah Mutuku [§] & Nasser Yao ^x

Abstract- Soybean (*Glycine max* (L.) Merrill) is among the most important crops worldwide due to its numerous uses in feed, food, biofuel, and significant atmospheric nitrogen fixation capability. To understand the genetic diversity and population structure of tropical soybean germplasm, 89 genotypes from diverse sources were analyzed using 7,962 SNP markers. The AMOVA results showed low diversity among and high within the populations, while the polymorphism information content (PIC) was 0.27. Both phylogenetic and principal component analysis grouped the 89 soybean genotypes into three major clusters, while population structure grouped the soybean genotypes into two subpopulations. On the other, the average Roger genetic distances within the study population was 0.34. The low diversity reported in the studied soybean germplasm pool is particularly worrying, considering the new trends of climate change and the emergence of new pests and diseases of soybean. Therefore, in order to address these challenges and develop soybean varieties with desirable traits, there is a need to broaden the genetic base of tropical soybean through the importation of germplasm from other countries.

Keywords: genetic diversity, population structure, single nucleotide polymorphism (SNP), soybean, tropical soybean genotypes.

1. INTRODUCTION

Soybean is among the most valuable crops in Uganda and across the East African region due to the high protein content that makes it an important ingredient in the diets of the people and livestock (Tukamuhabwa, 2001). Several soybean processing plants have been established in Uganda and across the East African region with large processing capacities to develop different products from soybean.

This new development has motivated the farmers to produce more grains to supply these plants (Tukamuhabwa et al., 2019). The three leading African countries in soybean production are South Africa (1,540,000 MT), Nigeria (758,033 MT), and Zambia (302,720 MT) (FAO 2018). Uganda is 11th in Africa and 1st in East Africa, with a production of 29,000 MT (FAO 2018). Hence soybean production and consumption have led to increased farmers' income, improved food and nutrition security, and poverty eradication at the rural household level (Ssengendo et al., 2010; SNV, 2011; Tukamuhabwa & Obua, 2015). Accordingly, soybean has the potential to contribute to poverty alleviation in Uganda and across the East African region.

Despite the contribution of soybean to smallholder farmers in Uganda and across the East African region, development of new varieties has been hindered by the low genetic diversity of the crop that have been observed in other countries (Gupta & Manjaya, 2017; Kumawat et al., 2015; Liu et al., 2017; Maldonado dos Santos et al., 2016; Torres et al., 2015). Kumawat et al. (2015) investigated the diversity of 82 Indian soybean accessions using SSR markers and identified three major clusters. In another study, Torres et al. (2015) found that both Principal Component Analysis (PCA) and STRUCTURE, clustered 191 soybean accessions in Brazil into two groups. Similarly, Gwinner et al. (2017) in another study that aimed at understanding the genetic diversity and population structure of 77 commercial soybean varieties in Brazil using 35 SSR markers, reported low genetic diversity in soybean germplasm.

To assess the genetic diversity of soybean and other plants, various methods such as morphological markers, geographic origins, pedigree information, isozymes, and DNA markers have been applied (Dayaman, 2007; Appiah-Kubi, 2012; Ojo et al., 2012; Malek et al., 2014; Villela et al., 2014). The use of morphological trait has remained a powerful taxonomic tool for preliminary grouping of germplasm before their classification using more precise marker techniques.

Author α ω ¥: College of Agricultural and Environmental Sciences, Makerere University Kampala, P. O. Box 7062 Kampala, Uganda.
e-mail: obuatonny@gmail.com

Author σ: National Agricultural Research Organization; National Livestock Resources Research Institute, Nakyesasa, 5704, Kampala, Uganda.

Author ρ: Molecular and Cellular Imaging Center – Columbus, The Ohio State University, Columbus, Ohio, 43210, University of the Sacred Heart Gulu, P. O. Box 374, Gulu, Uganda.

Author § x: Biosciences eastern and central Africa - International Livestock Research Institute, P.O Box 30709 Nairobi, 00100 Kenya.

In fact several studies involving the classification of plants still rely on the use of morphological traits (Khalid et al., 2010). Additionally, the use of morphological markers in classification is easy to score, cheaper and fast. However, the disadvantage of using morphological markers is that it's less robust compared to most molecular markers and outcomes can be easily influenced by environmental factors. In the case of pedigree information, limitations such as uncertain and incomplete data errors are likely, while for isozymes, chances of limited data are more prominent (Li & Nelson, 2001; Wang et al., 2006). So far, DNA markers remain the most precise method of genetic diversity analysis that have been complemented with morphological trait analysis. Among different DNA markers, random amplified polymorphic DNAs (RAPDs), amplified fragment length polymorphisms (AFLPs), restriction fragment length polymorphisms (RFLPs), simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs) have been widely used in understanding the diversity of soybean; each with its advantages and disadvantages (Chauhan et al., 2015; Chen et al., 2017; Doldi et al., 1997; Hipparagi et al., 2017; Ojo et al., 2012; Ren et al., 2013; Singh et al., 2013; Tantasawat et al., 2011; Tanya et al., 2001; Torres et al., 2015). SSR markers have been widely used to determine genetic diversity in many crops because they are easy to use, reasonably low price, and high level of polymorphism (Vignal et al., 2002). However, recently SNP markers have been widely utilized for assessment of diversity in plants because they occur much more frequently in the genome than SSR markers, and their genotyping can be easily automated (Mammadov et al., 2012). In the current study, we used Genotype By Sequencing (GBS) technology to study a collection of 89 tropical soybean germplasm collected from different countries. Therefore, the objective of the current study was to understand the genetic diversity and population structure of tropical soybean germplasm using SNP markers. Since the genotypes included in the current study are parental lines, land races, released varieties, and advanced lines, they are representative of the existing germplasm in tropical Africa.

II. MATERIALS AND METHODS

a) Plant materials

In this study, we used a total of 89 tropical soybean genotypes; these included collections from different sources that possess high genetic diversity (45 genotypes were from Uganda, 13 from Japan, six from the USA, 12 from World Vegetable Center (AVRDC) in Taiwan and 13 from Seed Co; a seed Company from Zimbabwe (Supplementary table).

b) DNA extraction, Determination of DNA Quality and Quantity

Seeds from each genotype were grown under controlled greenhouse conditions at Biosciences eastern and central Africa - International Livestock Research Institute (BecA - ILRI) Hub, Kenya. Twelve days after germination, one young leaf from one plant from each genotype was harvested, and DNA extracted using ZR Plant / Seed DNA Mini Prep™ according to manufacturer's protocol with minor modifications.

The DNA quality was first checked on 0.8% (w/v) agarose gel in 1 X Tris-acetate EDTA buffer and ran at 80V for 45 Minutes. The run gels were photographed using GelDoc-It™ Imager(UVP) and the picture image interpreted for DNA quality. The DNA was quantified using Thermo Scientific Nanodrop2000C Spectrophotometer and stored at 4 °C.

c) SNP Genotyping

The soybean genotypes were genotyped using the Illumina HiSeq 2500. Genotyping was conducted at Diversity Arrays Technology (DARTSeq™) in Australia. The genotypic process of the samples followed an integrated DAr T and genotyping-by-sequencing (GBS) methodology that involved complexity reduction of the genomic DNA, and repetitive sequences were eliminated using methylation-sensitive restrictive enzymes before sequencing on next-generation sequencing platforms (Kilian et al., 2012). The soybean reference genome was downloaded from ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v7.0/Gmax. The sequence data generated were then aligned to the soybean reference genome sequence, Soybean_v7, to identify single nucleotide polymorphisms (SNPs) markers.

d) Data analysis

GBS data from a total of 16,688 SNPs, distributed across all the 20 soybean chromosomes was received from Diversity Arrays Technology (DARTSeq™), Australia. The genotype data was filtered using a minor allele frequency (MAF) of 0.05 and a minimum count of 80% of the sample size using TASSEL v.5.2.43 software (Bradbury et al., 2007). Genetic distance was calculated between a pair of inbred lines in dataset using the identity by state similarity (IBS) method implemented in TASSELv.5.2.43. A marker-based kinship matrix was then calculated between a pair of inbred lines in data set using TASSELv.5.2.43.

Population structure was estimated using the model-based clustering approach implemented in STRUCTURE v2.3.4 software (Pritchard, Stephens & Donnelly, 2000). To estimate the posterior probabilities (qK) a 100,000 burn-in period was used, followed by 100,000 iterations; with the hypothetical number of subpopulations (K) ranging from 1 to 10, with ten replicates for each K. The number of subpopulations was determined when ΔK reached its highest value (Evanno, Regnaut & Goudet, 2005). The Delta K was

calculated for each value of K using Structure Harvester (Earl, Cruz, and Vonholdt 2012; Evanno et al. 2005). A line was assigned to a given cluster when the proportion of its genome in the cluster (qK) was higher than a standard threshold value of 70 %. For the chosen optima value of K, membership coefficient matrices of replicates from STRUCTURE were integrated to generate a Q matrix using the software CLUMPP (Jakobsson and Rosenberg, 2007) and the STRUCTURE bar plot was drawn using the DISTRUCT software (Rosenberg, 2004). Principal coordinate analysis was performed based on the genetic distance matrix using the Dissimilarity Analysis and Representation for windows (DARwin) v.6.0.013 (<http://darwin.cirad.fr>). To validate and gain more insight into the genetic diversity of the soybean germplasm panel used in this study, we generated a phylogenetic

tree by the neighbor-joining method. Analysis of molecular variance (AMOVA) was performed using GenAIEX V6.5 software.

III. RESULTS

a) Genotype Diversity analysis

A total of 16,688 SNP markers were identified in the 89 genotypes of soybean; of those 7,962 polymorphic and non-redundant SNP markers, with greater than 5% minor allele frequency (MAF) and missing data lower than 20% were used for subsequent analysis. These 7,962 SNPs detected a total of 15,924 alleles as expected. The average PIC was 0.27, ranging from 0.01 to 0.50, and heterozygosity ranged from 0.0 to 0.35 of individuals and 0.0 to 0.8 of markers (Fig. 1).

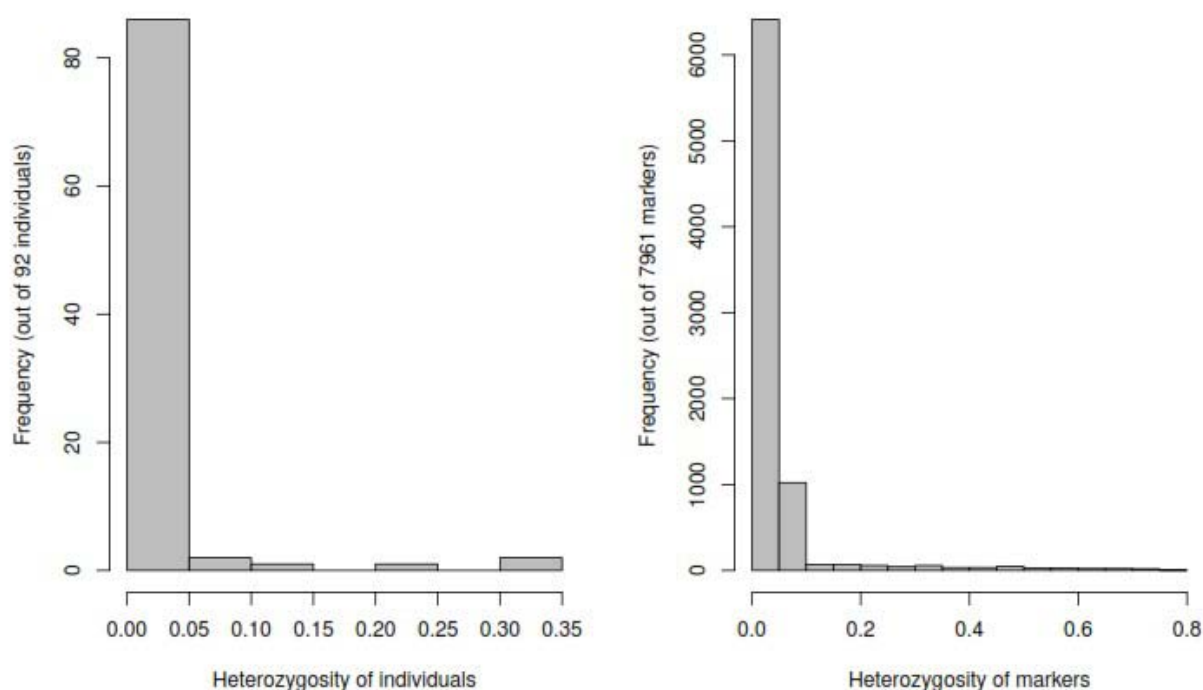


Fig. 1: Levels of heterozygosity of individual soybean genotypes and SNPs markers

b) Genetic distance and relationship

The average Roger genetic distances within the study population was 0.34. From a total of 89 genotypes, 18.1% of the distance values were between 0.0 and 0.05, while 20.7% were between 0.35 and 0.40 (Fig. 2). Relative kinship reflects the approximate degree of identity between two given genotypes. For combined analysis of all 89 genotypes, the kinship coefficients ranged from 0 to 1.04, with an overall average of 0.51; only 1.6% of the pair wise kinship estimates had values of 0.0 – 0.05 while 76.1% had values ranging from 0.5 – 0.550, indicating that most of the genotypes were in one way or another related and very few genotypes were not related (Fig. 3).

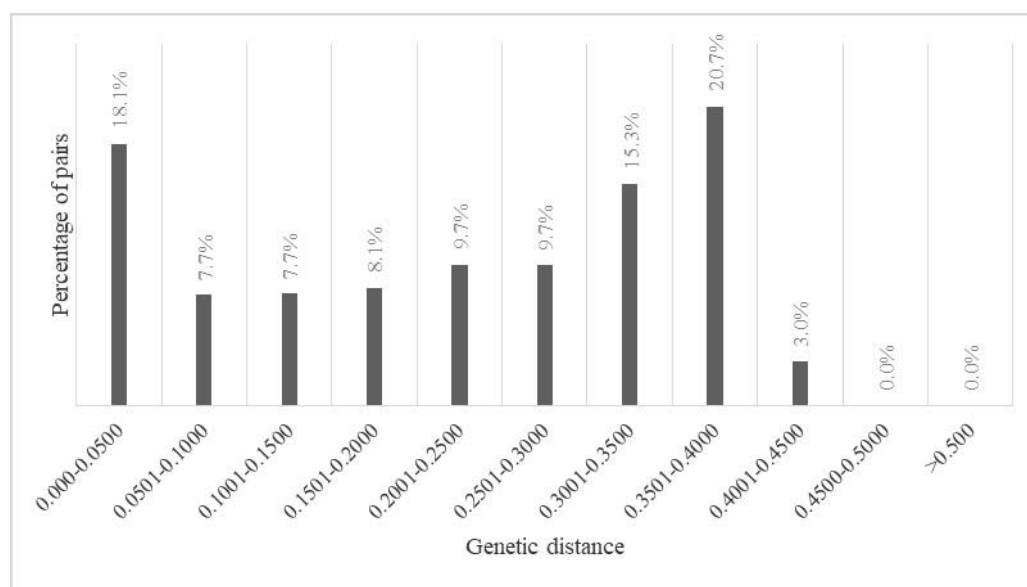


Fig. 2: Distribution of pairwise Roger's genetic distance calculated for 89 soybean genotypes

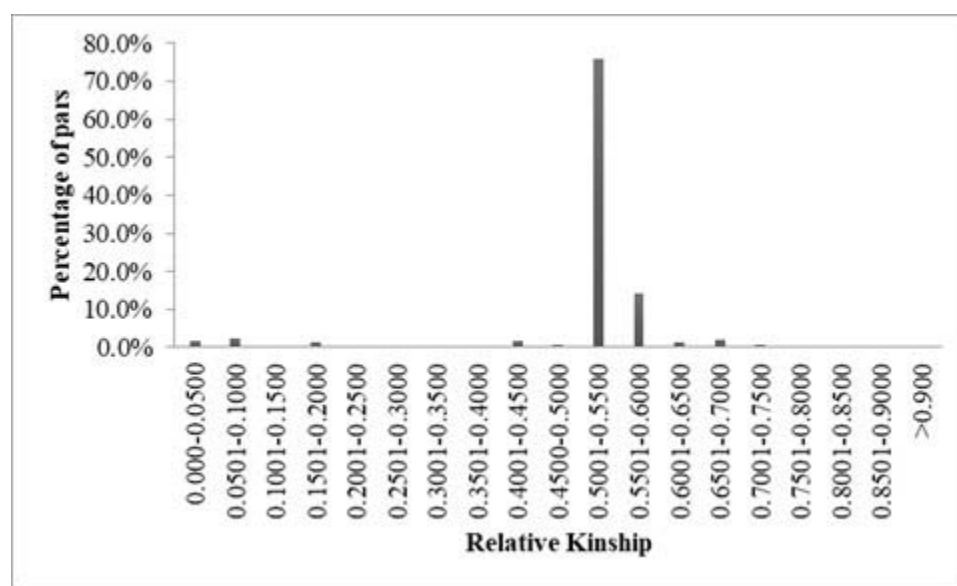


Fig. 3: Distribution of pair-wise kinship coefficients among 89 soybean genotypes from different sources based on 7,692SNPs

c) Population structure analysis

The log probability of the data $\ln P(D)$ increased continuously with increasing K (number of groups or populations). The ad hoc statistic ΔK showed a higher likelihood value at $K = 2$ as the highest level of structure (Fig. 4). This pattern was also observed in the population structure, where two groups were formed (Fig. 5).

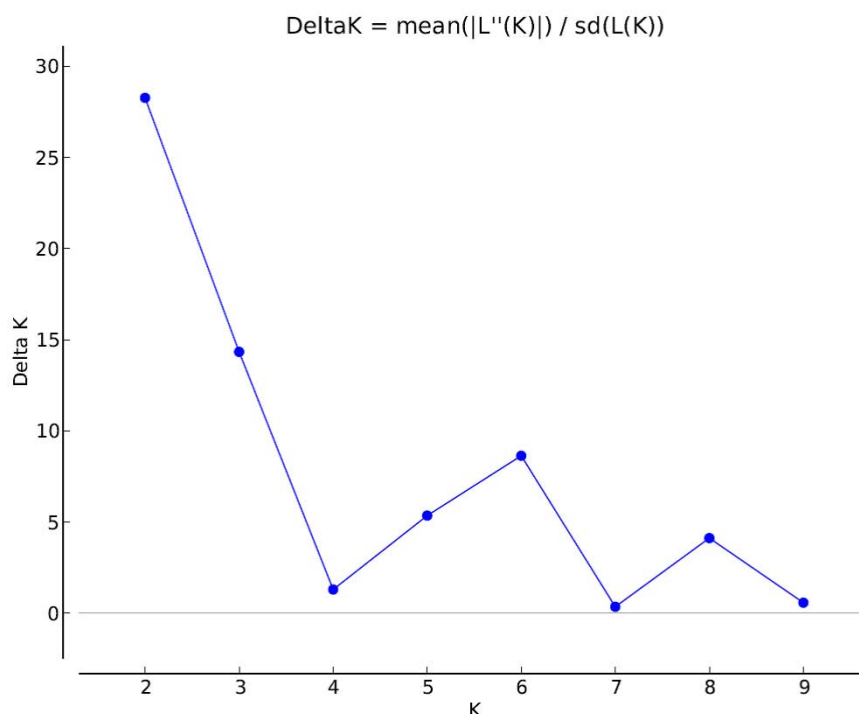


Fig. 4: Estimation of population using LnP(D) derived ΔK with K ranged from 1 to 10 with 7,692SNPs

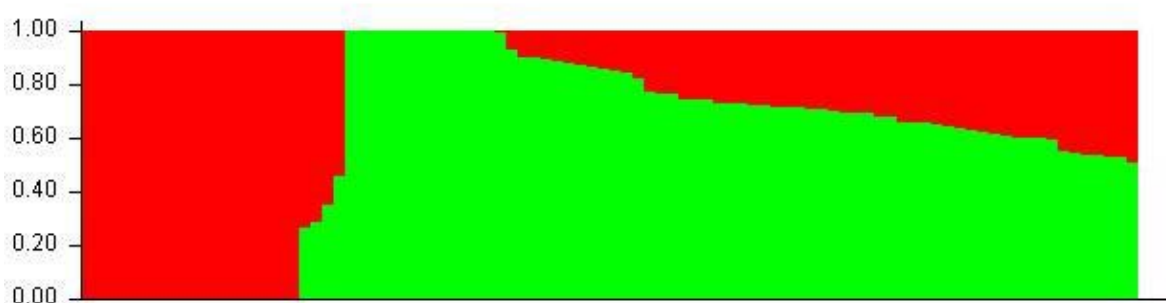


Fig. 5: Population structure (K = 2) inferred from STRUCTURE analysis for the 89 soybean genotypes based on 7,692SNPs

d) Neighbor-joining Phylogenetic Tree

The phylogenetic tree grouped the 89 soybean genotypes into three major clusters (Fig 6). The genotypes were separated into three distinct sub-clusters: There were 40 genotypes in sub-cluster 1, which included Nam II and GC00138-29, and 13 progenies derived from a cross between these two genotypes. Nam II is a Ugandan variety, which is a selection from TGM 79; obtained from IITA while GC00138-29 is a variety from AVRDC in Taiwan. This sub-cluster also included released varieties in Uganda; Namsoy 3 which is a cross between Kabanyolo 1 and Nam 1 (selection from ICAL 131 from the USA), and Maksoy 5N that is a progeny of Nam II and GC00138-29. The second sub-cluster had 26 genotypes, among which 13 genotypes were from Seed co in southern

Africa and eight genotypes from AVRDC, Taiwan. It was surprising that Namsoy 4M, a released Ugandan variety that is a progeny of Nam II and GC00138-29, was clustered in this sub-cluster. By comparison, the other remaining 23 genotypes belonged to sub-cluster 3, among which seven genotypes were progenies from a cross between Duiker and TGx 1835-10E while nine were from a cross between Duiker and GC00138-29. This sub-cluster also included released Ugandan varieties, Maksoy 1N (selection from TGx 1835-10E), Maksoy 2N (Duker X TGx 1835-10E), Maksoy 3N and Maksoy 4N (Duker X GC00138-29). However, few S lines and AVRDC genotypes were scattered in all three major clusters.



e) *Principal Component Analysis (PCA)*

2006). A PCA of the 89 genotypes with the 7,962 SNPs also showed a clear separation of the same three major groups that were identified by the phylogenetic tree (Fig. 7).



IV. DISCUSSION

One of the requirements for a successful breeding program is a high level of genetic diversity among the germplasm used for the development of new crop varieties. Over the years, most soybean breeding programs have replaced traditional varieties or

landraces with more modern varieties with desirable attributes that have led to increased yields. However, in the current study, to compare the genetic diversity of tropical soybean genotypes, we studied fairly diverse sets of genotypes from Uganda, Zimbabwe, Japan, Taiwan, and the USA. These genotypes included parental lines, land races, released varieties, and advanced lines that are representative of the existing germplasm in tropical Africa.

The level of genetic diversity observed in this study is lower compared to previously reported results based on SNP data (Li et al. 2010; Hao et al. 2012; Zhou et al. 2015). The observed low diversity is because the genotypes used in the present study were mainly released varieties and advanced breeding lines. In contrast, the genotypes used in Li et al. (2010), Hao et al. (2012) and Zhou et al. (2015) included mainly wild relatives and landraces of soybean. On the other hand, previous studies that involved improved soybean varieties also observed low genetic diversity (Liu et al. 2017; Maldonado dos Santos et al. 2016). These improved varieties tend to have low genetic diversity because of the high selection pressure subjected to the genotypes during evaluation and selection (Gwinner et al. 2017). This was also confirmed by genetic distance and kinship analysis that showed that majority of the genotypes in this study are related to each other in one way or another.

The phylogenetic tree and PCA analyses indicated the existence of three major sub-clusters among the 89 genotypes of our study. On the other hand, population structure clustered the genotypes into two major subpopulations. Sub-cluster 1 included Nam II and GC00138-29 and 13 progenies derived from a cross between these two genotypes. Nam II is a Ugandan variety, which is a selection from TGM 79; obtained from IITA while GC00138-29 is a variety from AVRDC in Taiwan. This subpopulation also included Maksoy 5N, released in 2013 and NII X GC 44.2 that was released in 2017 as Maksoy 6N and are progenies of Nam II and GC00138-29 cross. Since TMG 79 and GC00138-29 were introduced to Uganda through Consultative Group for International Agricultural Research (CGIAR) institutions that usually collect germplasm from different countries, there is a possibility that they share the same geographical origin. On the other hand, genotypes from Seed Co and AVRDC, Taiwan were grouped in the second sub-cluster. This implies that soybean varieties from Seed Co share very similar parents and geographical origin with genotypes from AVRDC.

By comparison, the third sub-cluster mainly consisted of progenies from two crosses; Duiker X TGx 1835-10E and Duiker X GC00138-29. The sub-cluster also included released Ugandan varieties; Maksoy 1N (selection from TGx 1835-10E), Maksoy 2N (Duiker X TGx 1835-10E), Maksoy 3N and Maksoy 4N (Duiker X

GC00138-29). Duiker originated from Zimbabwe and was used as a female parent during generation of the two crosses.

V. CONCLUSION

Genetic variation and population structure of the core germplasm available for soybean breeding in Uganda and the East African region were assessed using high-density SNP markers. The results of the study showed a low level of heterogeneity within most of the genotypes studied, suggesting that the current generation of inbreeding has fixed lines. The observed low diversity in the germplasm pool is particularly worrying; considering the vulnerability of agriculture under the impact of climate changes. For example, in Uganda, we have observed the emergence of two new soybean pests (groundnut leaf miner and bruchids) that previously were not main production constraints. This is coupled with breakdown of soybean rust resistance in the existing soybean varieties in Uganda that were previously resistant to the disease due to several virulent races of soybean rust pathogen. Therefore addressing these challenges and developing soybean varieties with the desirable traits, requires diversification of the genetic background of the current breeding population by incorporating new genetic resources from other countries.

ACKNOWLEDGMENTS

This project was supported by the BecA-ILRI Hub through the Africa Biosciences Challenge Fund (ABCF) program. The ABCF Program is funded by the Australian Department for Foreign Affairs and Trade (DFAT) through the BecA-CSIRO partnership; the Syngenta Foundation for Sustainable Agriculture (SFSA); the Bill & Melinda Gates Foundation (BMGF); the UK Department for International Development (DFID) and; the Swedish International Development Cooperation Agency (Sida).

REFERENCES RÉFÉRENCES REFERENCIAS

1. Appiah-Kubi, D. (2012). Diversity Studies in Soybean (*Glycine Max* (L.) Merrill) and Validation of Shattering Resistant Markers for Marker Assisted Selection. Kwame Nkrumah university of science and technology.
2. Bradbury, Peter J., Zhiwu Zhang, Dallas E. Kroon, Terry M. Casstevens, Yogesh Ramdoss, and Edward S. B. (2007). TASSEL: Software for Association Mapping of Complex Traits in Diverse Samples. 23(19):2633–35.
3. Chauhan, Devendra K., J. A. Bhat, A. K. Thakur, S. Kumari, Z. Hussain, and C. T. Satyawathi. (2015). Molecular Characterization and Genetic Diversity Assessment in Soybean [*Glycine Max* (L .) Merr .] Varieties Using SSR Markers. 14(October):504–10.

4. Chen, Wu, Lu Hou, Zhiyong Zhang, Xiaoming Pang, and Yingyue Li. (2017). Genetic Diversity, Population Structure, and Linkage Disequilibrium of a Core Collection of *Ziziphus Jujuba* Assessed with Genome-Wide SNPs Developed by Genotyping-by-Sequencing and SSR Markers. 8(April):1–14.
5. Dayaman, V., N. Senthil, M. Raveendran, N. Nadarajan, P. Shanmugasundaram, and P. Balasubramanian. (2009). Diversity Analysis in Selected Indian Soybean [*Glycine Max* (L.) Merrill] Using Morphological and SSR Markers. *International Journal of Integrative Biology* 5(2).
6. Doldi, M. L., J. Vollmann, and T. Lelley. (1997). Genetic Diversity in Soybean as Determined by RAPD and Microsatellite Analysis. *Plant Breeding* 116(4):331–35.
7. Earl, Dent A., Santa Cruz, and Bridgett M. Vonholdt. (2012). Earl DA, VonHoldt BM. Structure Harvester : A Website and Program for Visualizing STRUCTURE Output and Implementing the Evanno Method . Cons STRUCTURE HARVESTER: A Website and Program for Visualizing STRUCTURE Output and Implementing the Evanno Method. (June):359–61.
8. Evanno, Guillaume, Sebastien Regnaut, and Jerome Goudet. (2005). Detecting the Number of Clusters of Individuals Using the Software STRUCTURE : A Simulation Study. (August).
9. FAO. (2018). Online at <http://www.fao.org/faostat> (Accessed on 5th June, 2020)
10. Gupta SK, Manjaya JG. (2017). Genetic Diversity and Population Structure of Indian Soybean (*Glycine Max* (L.) Merr.) as Revealed by Microsatellite Markers. *Physiology and Molecular Biology of Plants* 25(4):953–64.
11. Gwinner, R., T. A. Setotaw, M. Pasqual, J. B. dos Santos, A. M. Zuffo, E. Z. Vinicius, and A. T. eodor. Bruzi. (2017). Genetic Diversity in Brazilian Soybean Germplasm. *Crop Breeding and Applied Biotechnology* 17(17):373–81.
12. Hao D, Cheng H, Yin ZT, Cui SY, Zhang D, Wang H, Yu D. (2012). Identification of Single Nucleotide Polymorphisms and Haplotypes Associated with Yield and Yield Components in Soybean (*Glycine Max*) Landraces across Multiple Environments. *Theor Appl Genet* 124(447–458).
13. Hipparagi, Yegappa, Rakesh Singh, Debjani Roy Choudhury, and Veena Gupta. (2017). Genetic Diversity and Population Structure Analysis of Kala Bhat (*Glycine Max* (L .) Merrill) Genotypes Using SSR Markers. 1–11.
14. Jakobsson, Mattias and Noah A. Rosenberg. (2007). CLUMPP: A Cluster Matching and Permutation Program for Dealing with Label Switching and Multimodality in Analysis of Population Structure. 23(14):1801–6.
15. Khalid, Muhammad, F. Farhatullah, Naqib Ullah Khan, Raziud Din, M. Yasir Khan, M. Akmal, and Nasir Ali. (2010). Linkage of Morphological Markers in *Brassica*. *Pakistan Journal of Botany* 42(5):2995–3000.
16. Kilian A. , Wenzl P., Huttner E., Carling J., Xia L., Blois H., Vanessa Caig, Katarzyna Heller-Uszynska, Damian Jaccoud, Colleen Hopper, Malgorzata Aschenbrenner-Kilian, Margaret Evers, Kaiman Peng, Cyril Cayla, Puthick Hok, Grzegorz Uszynski. (2012). Diversity Arrays Technology: A Generic Genome Profiling Technology on Open Platforms. In: Pompanon F., Bonin A. (Eds) *Data Production and Analysis in Population Genomics. Methods in Molecular Biology (Methods and Protocols)*, Vol 888. Humana Press, Totowa, NJ.
17. Kumawat, Giriraj, Gourav Singh, C. Gireesh, M. Shivakumar, Mamta Arya, Dinesh K. Agarwal, and Syed Masroor Husain. (2015). Molecular Characterization and Genetic Diversity Analysis of Soybean (*Glycine Max* (L.) Merr.) Germplasm Accessions in India. *Physiology and Molecular Biology of Plants* 21(1):101–7.
18. Li, Z. L., Nelson, R. L. (2001). Genetic Diversity among Soybean Accessions from Three Countries Measured by RAPDs. *Crop Sci.* 41:1337– 1347.
19. Li, Ying Hui, Wei Li, Chen Zhang, Liang Yang, Ru Zhen Chang, Brandon S. Gaut, and Li Juan Qiu. (2010). Genetic Diversity in Domesticated Soybean (*Glycine Max*) and Its Wild Progenitor (*Glycine Soja*) for Simple Sequence Repeat and Single-Nucleotide Polymorphism Loci. *New Phytologist* 188(1):242–53.
20. Liu, Zhangxiong, Huihui Li, Zixiang Wen, Xuhong Fan, Yinghui Li, Rongxia Guan, Yong Guo, Shuming Wang, Dechun Wang, and Lijuan Qiu. (2017). Comparison of Genetic Diversity between Chinese and American Soybean (*Glycine Max* (L.)) Accessions Revealed by High-Density SNPs. *Frontiers in Plant Science* 8(November).
21. Maldonado dos Santos, João Vitor, Trupti Joshi, Saad M. Khan, Yang Liu, Juexin Wang, Tri D. Vuong, Marcelo Fernandes de Oliveira, Francimar Corrêa Marcelino-Guimarães, Dong Xu, Henry T. Nguyen, and Ricardo Vilela Abdelnoor. (2016). Evaluation of Genetic Variation among Brazilian Soybean Cultivars through Genome Resequencing. *BMC Genomics* 17(1):1–18.
22. Malek, M. A., Mohd Y. Rafii, Most Shahida, Sharmin Afroz, Ujjal Kumar Nath, and M. Monjurul Alam Mondal. (2014). Morphological Characterization and Assessment of Genetic Variability , Character Association, and Divergence in Soybean Mutants. *The Scientific World Journal* 2014(article ID 968796):12.
23. Mammadov, Jafar, Rajat Aggarwal, Ramesh Buyyarapu, and Siva Kumpatla. (2012). SNP

- Markers and Their Impact on Plant Breeding *International Journal of Plant Genomics* Article ID 728398.
24. Ojo, D. K., Ajayi, A. O., Oduwaye, O. A. (2012). Genetic Relationships among Soybean Accessions Based on Morphological and RAPDs Techniques. *Pertanika J. Trop. Agric. Sci.* 35(2):237–48.
 25. Villela Otvia, Tiago, Helena Unda-Trevisoli Sandra, Mota Da Silva Fabiana, Souza Brbaro Junior Laerte, and Orlando Di Mauro Antonio. (2014). Genetic Divergence of Roundup Ready (RR) Soybean Cultivars Estimated by Phenotypic Characteristics and Molecular Markers. *African Journal of Biotechnology* 13(26):2613–25.
 26. Patterson N., Price AL, Reich D. (2006). Population structure and Eigenanalysis. *PLoS Genet.* 2(12).
 27. Pritchard, Jonathan K., Matthew Stephens, and Peter Donnelly. (2000). Inference of Population Structure Using Multilocus Genotype Data.
 28. Ren, Jing, Daokun Sun, Liang Chen, Frank M. You, Jirui Wang, and Yunliang Peng. (2013). Genetic Diversity Revealed by Single Nucleotide Polymorphism Markers in a Worldwide Germplasm Collection of Durum Wheat. (February):7061–88.
 29. Rosenberg N.A. (2004). DISTRUCT: A Program for the Graphical Display of Population Structure. *Molecular Ecology Resources* 4:137–38.
 30. Singh, Nivedita, Debjani Roy Choudhury, Amit Kumar Singh, Sundeep Kumar, Kalyani Srinivasan, R. K. Tyagi, N. K. Singh, and Rakesh Singh. (2013). Comparison of SSR and SNP Markers in Estimation of Genetic Diversity and Population Structure of Indian Rice Varieties. 8(12):1–14.
 31. SNV. (2011). Increased Competitiveness of the Oilseed Value Chain through Improved Information on the Markets for Soybean in Uganda. Final Report Submitted to SNV Rwenzori Portfolio, Kampala, Uganda.
 32. Ssengendo M, Mayende J, Mubiru R. (2010). *Soybean Fact Sheet: Export Potential of the Soybean Sub-Sector in Uganda. A Survey Carried out by Participants of the COMESA Training Done by International Trade Center (ITC) to Asses the Export Potential of Soybean Industry. Uganda Export Promotion Board.*
 33. Tantasawat, P, J. Trongchuen, T. Prajongjai, S. Jenweerawat, and W. Chaowiset. (2011). SSR Analysis of Soybean (*Glycine Max* (L .) Merr .) Genetic Relationship and Variety Identification in Thailand. 5(3):283–90.
 34. Tanya, P., P. Srinives, T. Toojinda, A. Vanavichit, B. K. Ha, J. S. Bae, and S. H. Lee. (2001). Evaluation of Genetic Diversity among Soybean Genotypes Using SSR and SNP. *Korean Journal of Crop Science* 46(January):334–40.
 35. Torres, Adalgisa Ribeiro, Anna Karolina Grunvald, Talita Busulini Martins, Maria Aparecida Dos Santos, No lle Giacomini Lemos, Luis Ant nio Stabile Silva, and Mariangela Hungria. (2015). Genetic Structure and Diversity of a Soybean Germplasm Considering Biological Nitrogen Fixation and Protein Content. *Scientia Agricola* 72(1):47–52.
 36. Tukamuhabwa P., Obua T., Namara N. Okii D., Kabayi P., Yiga G. (2019). *Soybean Research and Development in Uganda: Highlights 2002-2018*. Makerere University, Kampala, Uganda.
 37. Tukamuhabwa P., Obua T. (2015). Production Guide In Uganda. Makerere University, Kampala, Uganda.
 38. Tukamuhabwa P. (2001). Soybean (*Glycine Max* L.). In Mukiibi, J. B. (Ed.). *Agriculture in Uganda*. Volume 2, Crops. National Agricultural Research Organization,. Pp. 133–41 in.
 39. Vignal, A., Milan, D., Cristobal S. M., and Eggen A. (2002). A Review on SNP and Other Types of Molecular Markers and Their Use in Animal Genetics. 34:275–305.
 40. Wang, L. X., Guan, R. X., Liu, Z. X., Chang, R. Z., Qiu, L. J. (2006). Genetic Diversity of Chinese Cultivated Soybean Revealed by SSR Markers. *Crop Sci.* 46:032–1038.
 41. Zhou, Zhengkui, Yu Jiang, Zheng Wang, Zhiheng Gou, Jun Lyu, Weiyu Li, Yanjun Yu, Liping Shu, Yingjun Zhao, Yanming Ma, Chao Fang, Yanting Shen, Tengfei Liu, Congcong Li, Qing Li, Mian Wu, Min Wang, Yunshuai Wu, Yang Dong, Wenting Wan, Xiao Wang, Zhaoli Ding, Yuedong Gao, Hui Xiang, Baoge Zhu, Suk-ha Lee, Wen Wang, and Zhixi Tian. (2015). Resource Resequencing 302 Wild and Cultivated Accessions Identifies Genes Related to Domestication and Improvement in Soybean. *Nature Biotechnology* 33(4).

Compliance with Ethical Standards

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors

Conflict of interest: The authors declare no conflict of interest.