Analysis of population structure and genetic diversity in a Southern African soybean collection based on single nucleotide polymorphism markers

A. Tsindi[1,2], J. S. Y. Eleblu[1], E. Gasura[3], H. Mushoriwa[4], P. Tongoona[1], E. Y. Danquah[1], L. Mwadzingeni[2], M. Zikhali[2], E. Ziramba[2], G. Mabuyaye[2] and J. Derera[5*]

Abstract

Soybean is an emerging strategic crop for nutrition, food security, and livestock feed in Africa, but improvement of its productivity is hampered by low genetic diversity. There is need for broadening the tropical germplasm base through incorporation and introgression of temperate germplasm in Southern Africa breeding programs. Therefore, this study was conducted to determine the population structure and molecular diversity among 180 temperate and 30 tropical soybean accessions using single nucleotide polymorphism (SNP) markers. The results revealed very low levels of molecular diversity among the 210 lines with implications for the breeding strategy. Low fixation index (F_{ST}) value of 0.06 was observed, indicating low genetic differences among populations. This suggests high genetic exchange among different lines due to global germplasm sharing. Inference based on three tools, such as the Evanno method, silhouette plots and UPMGA phylogenetic tree showed the existence of three sub-populations. The UPMGA tree showed that the first sub-cluster is composed of three genotypes, the second cluster has two genotypes, while the rest of the genotypes constituted the third cluster. The third cluster revealed low variation among most genotypes. Negligible differences were observed among some of the lines, such as Tachiyukata and Yougestu, indicating sharing of common parental backgrounds. However large phenotypic differences were observed among the accessions suggesting that there is potential for their utilization in the breeding programs. Rapid phenotyping revealed grain yield potential ranging from one to five tons per hectare for the 200 non-genetically modified accessions. Findings from this study will inform the crossing strategy for the subtropical soybean breeding programs. Innovation strategies for improving genetic variability in the germplasm collection, such as investments in pre-breeding, increasing the geographic sources of introductions and exploitation of mutation breeding would be recommended to enhance genetic gain.

Keywords Glycine max, Molecular diversity, Phenotyping, Population structure, SNP, Soybean

*Correspondence:
J. Derera
j.derera@cgiar.org
Full list of author information is available at the end of the article

Introduction

Soybean is an important nutritious crop used for food, feed and industrial oils, worldwide. Its high utility is explained by the high protein content of about 40% and high oil content reaching and exceeding 20% for some genotypes (Bellaloui et al. 2010; Orf 2010). In 2019, the worldwide production was over 300 million metric tons produced on 120 million hectares of land (FAOSTAT 2021), which translate to a global average yield of 2.5 tons per hectare. Production is dominated by a few countries. The world's leading soybean producers are Brazil, United States of America, Argentina and China. Africa, contributes only 0.9% to the total world production (FAOSTAT 2021), which is negligible and does not match the regional demand for soybean products. The major producers are South Africa, Nigeria, Ghana, Uganda, Ethiopia, Zambia, Malawi and Zimbabwe. All these countries fail to meet their national demand. As a result, Africa imports soybean.

There is need to develop varieties that are highly productive and adapted to the tropical and subtropical ecologies in Africa. Efforts are underway to identify such varieties through the regional soybean breeding network that employs the Pan African Trials (PAT) under the leadership of the Soybean Innovation Lab (SIL), in collaboration with the International Institute of Tropical Agriculture (IITA), national public programs and private seed companies (https://www.soybeaninnovationlab.illinois.edu/). The PAT shows a general low level of productivity due to limited genetic improvements. However, genetic improvement efforts are challenged by the low genetic base of soybean (Cornelious and Sneller 2002; Lee et al. 2014; Li et al. 2013) owing to several domestication bottlenecks (Gwinner et al. 2017; Hyten et al. 2007; Rafalski 2002).

The baseline genetic diversity of the soybean germplasm pool and introductions should be established in order to devise a viable breeding strategy. Genetic improvement of any crop rests upon the diversity present within and among the breeding populations (Biyeu et al. 2010). Knowledge of genetic variability helps in selection of parental lines to be used when making crosses, establishment of core collections and enhanced utilization of the germplasm in breeding programs (Abebe et al. 2021; Bandillo et al. 2017). While there is limited diversity among cultivars within country or regional breeding programs because of sharing of common parents (Gwinner et al. 2017; Hahn and Würschum 2014; Tiwari et al. 2019), introduction of exotic germplasm plays a crucial role in widening the genetic base from which parents can be selected for use to make bi-parental crosses.

The tropical and subtropical soybean breeding programs in Africa utilizes temperate germplasm to improve local varieties. The major sources of soybean germplasm lines and populations for Africa have been China, Japan, Korea and USA (Grieshop and Fahey 2001; Jeong et al. 2019a, b) where greater genetic diversity has been reported (Oliveira et al. 2010). China and the USA maintain large collections in their gene banks. There are about thirty thousand accessions in the Chinese Gene bank, while the USDA gene banks contain about 15000 accessions (Liu et al. 2017). This germplasm cannot be directly used in the breeding programs in Africa. There is need to characterize the germplasm before crossing is done. For example, 93% of the Chinese germplasm accessions are primitive cultivars but highly diverse (Chen and Nelson (2005). These collections are important sources of favorable alleles which can enhance breeding in Africa. However, when such introductions are to be used for breeding purposes, they need to be screened for their usefulness (Jeong et al. 2019a, b; Li et al. 2014) and inform the breeding strategy.

A survey of the literature indicates that germplasm diversity characterization can be conducted following two approaches. Both morphological or phenotypic and molecular genetic diversity studies have been used to assess variation in soybean (Abebe et al. 2021; Bandillo et al. 2015; Chander et al. 2021; Malik et al. 2011; Jeong et al. 2019a, b; Ma et al. 2006; Nawaz et al. 2021; Ojo et al. 2012; Valliyodan et al. 2021; Wang et al. 2012; Mihaljević et al. 2020). The advantages and limitations of both approaches have been discussed.

While morphological or phenotypic methods have been successful for discriminating soybean genotypes, their efficiency is compromised by complications which are caused by the genotype by environment interactions (GxE) effects. GxE masks genotypic differences among the germplasm entries. The high levels of GxE effects requires that genotypes are evaluated at many sites. However, due to the exorbitant costs for conducting multi-location trials, a few sites are often used resulting in a low resolution due to few data points. There are also challenges of waiting for a long time to get results. The length of the cycle from seed to seed is a hindrance as it is time consuming, labor intensive and costly (Chander et al. 2021; Nadeem et al. 2018). As a result, use of molecular markers has increased. They are not affected by GxE interactions, not growth specific and are abundant within the genome (Nadeem et al. 2018). Although molecular markers were initially expensive, there have been improvements such as invention of single nucleotide polymorphism (SNPs) DNA markers and their amenability to automation that have brought the costs per data point to a very competitive level compared to phenotypic data.

Currently, SNPs are among the most widely used markers (Zhu et al. 2003; Edwards et al. 2007; Nadeem et al. 2018).

The SNPs are the markers of choice for molecular diversity studies. SNP markers have been successfully used for diversity studies for several crops including soybean (Abebe et al. 2021; Chander et al. 2021; Liu et al. 2017), cowpea (Fatokun et al. 2018; Qin et al. 2016; Sodedji et al. 2021), pigeon pea (Yang et al. 2006; Zavinon et al. 2020) and common bean (Blair et al. 2013; Cortés et al. 2011; Nemlı et al. 2017). Assessment of the genetic diversity among elite lines and varieties developed by IITA using SNPs revealed high diversity within the germplasm and grouped the germplasm into three clusters based on genetic relatedness (Abebe et al. 2021). Similarly, broad genetic base among tropical soybean lines with a genetic diversity index of 0.414 using SNP markers has been reported (Chander et al. 2021). However, previous studies cited low genetic diversity among the germplasm from Brazil, China, Europe and North America. Low genetic diversity was reported among Brazilian (Gwinner et al. 2017), USA and Chinese germplasm (Liu et al. 2017). Central European lines were reported to be closely related to the Swiss and Canadian lines, but distantly related to the Chinese (Hahn and Würschum 2014). These findings suggest the need for breeders to know the molecular diversity in the germplasm to guide breeding strategies.

Improvement of soybean varieties for adaptation and productivity ranks quite high on the product profile for the Southern Africa region. Early maturity in response to climate change, which has rendered growing seasons short, is one of the important traits for soybean lines for deployment in sub-Saharan Africa (Ziervogel et al. 2014). This requires sourcing of exotic germplasm with the favorable alleles for early maturity. Temperate germplasm is less sensitive to latitude, which is a major determinant of flowering and maturity time in soybean. The soybean breeding programs in Africa have collected both temperate and tropical germplasm for utilization in breeding. However, the levels of molecular diversity in this collection has not been established. The present study was therefore conducted to assess the population structure and genetic diversity of the temperate and tropical soybean accessions using SNP markers.

Materials and methods

Plant material and sampling

Public (belonging to government/ national research institutions) and private (from private institutions) germplasm collection which comprised 210 lines from South Africa (10), Malawi (1), Zimbabwe (19), and USA (180) was used for the study. All the genotypes were planted in plastic sleeves in a screen house in 2019. The 10 genotypes from South Africa were planted in South Africa while the other 200 were planted in Zimbabwe. An average of six leaf discs was sampled from a single plant from each of the genotypes at 3 weeks after emergence using the LGC genomics plant sample collection kit. The leaf discs were placed in 96 well plates and sealed with perforated strip caps. A desiccant sachet was placed on top of the sealed tubes and a rack lid was fixed on top. The samples were placed in a sealable bag and shipped to LGC genomics, Germany, for genotyping using the targeted genotyping-by-sequencing (SeqSNP) method.

Rapid phenotypic screening

A total of 200 non-genetically modified accessions (temperate and Tropical) were planted in Zimbabwe. The ten accessions from South Africa could not be evaluated in Zimbabwe because they are genetically modified (contain the Roundup-ready herbicide resistance trait). The rapid screening was conducted at the Rattray Arnold Research Station (RARS) (17°38′60" S 31°14′24"E), near Harare. Rapid phenotypic screening for yield was done in an observation trial without replication in two row plots which were 1.5 m long and a spacing of 0.45 m inter row and 0.05 within row. Grain yield was recorded from the whole plot at maturity.

DNA extraction, SNP marker genotyping and data pre-processing

DNA extraction was done using magnetic bead chemistry (sbeadex™ mini plant kit from LGC, Biosearch Technologies, Berlin, Germany) on KingFisher Flex. SNP marker genotyping was performed using SeqSNP, a targeted genotyping by sequencing service offered by LGC, which allows for genotyping of SNPs and small insertions/deletions using a single primer enrichment technology (LGC Bioscience Technologies 2019). In order to design a SeqSNP assay, a total of 500 informative markers were selected from a panel of 1 082 markers in the LGC database (https://www.biosearchtech.com/products/pcr-kits-and-reagents/genotyping-assays/kasp-genotyping-chemistry/kasp-snp-libraries/soybean-genotyping-library), which were designed from an original set of 1 536 SNP markers, the "Universal Soy Linkage Panel" (USLP 1.0) described in Hyten et al. 2010. These SNP markers were selected based on the even distribution throughout each of the 20 consensus linkage groups, and for optimum allele frequency in diverse germplasm. The physical starting and end positions of the markers for the construction of a BED file for use in sequencing were taken from the Soybase database (https://www.soybase.org/) with the reference genome as Williams 82.

The total number of targets that passed design was 496 covered by a total of 984 oligo probes, i.e. the number

Tsindi *et al. CABI Agriculture and Bioscience* (2023) 4:15

Page 4 of 14

of oligo probes per target being ~ 1.98. The total number of targets which passed the quality criteria, that is, those that were successfully genotyped in at least 85% of all samples, was 485 (97.8%). NextSeq 500 sequencing was performed, with the number of pre-processed reads being 35 397 796 reads which is approximately 168 561 reads per sample. The percentage reads effectively used in genotyping was 83.4% and the average effective target SNP coverage 283x. The SNP genotyping pipeline and settings involved diploid genotyping with minimum coverage of 8 reads per sample and locus using Free Bayes (Garrison & Marth 2012). A total of 437 (87.1%) of the targets were polymorphic, 98.5% of all calls were homozygous and 1.5% heterozygous. Missing data was reported with 1.4%.

Demultiplexing of all library groups was done using the Illumina bcl2fastq 2.17.1.14 software. One or two mismatches or Ns were allowed in the barcode read when barcode distances between all libraries on the lane allowed for it. Clipping of sequencing adapter remnants was then done from all reads. Reads with final length < 65 bases were discarded. Quality trimming of adapter clipped illumina reads was performed for the removal of reads containing Ns and trimming of reads at 3` end to get a minimum average Phred quality score of 30 over a window of ten bases. Reads with final length < 65 bases were discarded. FastQC reports for all FASTQ files were then created. Read counts containing all read counts for all samples at a glance were then generated.

Data analysis

Alignment of quality trimmed reads against target genome using Bowtie2 was done followed by variant discovery and genotyping of samples with Freebayes V1.0.2−16 (https://github.com/ekg/freebayes#readme). Ploidy was set at 2 and genotypes were filtered for a minimum coverage of 8 reads. SNP marker diversity and profile were analyzed using the Powermarker and GenAlEx software. SNP data quality check was done by filtering, where SNPs with call rate greater than 90% were retained and those with minor allele frequency (MAF) of < 0.05 were discarded. The polymorphic information content (PIC), observed heterozygosity (Ho), expected heterozygosity (He), allele frequency and Shannon Information Index (I) were computed in Powermaker (Liu and Muse 2005) and GenAlEx (Peakall and Smouse 2012).

Genetic diversity analyses were conducted using the R software. The genotypes were subjected to Silhouette plot analysis in R Statistics 3.5.1 version (Team R Core 2015) to determine the probable number of clusters formed. Coefficients of similarity showing genetic distances among the soybean lines (Matrix of similarities) were calculated in R Statistics following the Gower's Distance model (Gower 1971). The similarity matrix was then used to group the soybean genotypes using the Unweighted Pair Group Method using Arithmetic average (UPGMA) algorithm in R Statistics (Team R Core 2015) giving an annotated phylogenetic tree (Rambaut 2016). The 30 tropical and 180 temperate genotypes were isolated and subjected to diversity analysis and a Dendogram was drawn in R Statistics separately for each group of genotypes.

Population structure analysis was performed using the Bayesian clustering approach in STRUCTURE v2.3.4 (Porras-Hurtado et al. 2012). Structure analysis was run using an Admixture model with 5 000 burning period and 50 000 Markov-chain Monte Carlo replications. The number of clusters (k) was set to range from 1 to 10 with 3 iterations. The output from STRUCTURE was then imported to Structure harvester (Earl and VonHoldt 2012) to visualize the delta K value which forms a distinct peak, using the Evanno Method. Analysis of molecular variance (AMOVA) was done using GenAlEx (Peakall and Smouse 2012) to determine the variance components and the molecular diversity between and within populations. Bases were coded A = 1, C = 2, G = 3, T = 4 and missing data 0. Clone Identification was also done in GenAlEx. The Nei's nucleotide distance and the fixation Index (F_{ST}) were also computed. The fixation index is a measure of genetic variation that can be explained by population structure and ranges from 0 (identical) to 1 (completely different with no common alleles shared) (Mohammadi and Prasanna 2003) calculated as;

$$\text{FST} = \frac{\delta_s^2}{\bar{p}(1 - \bar{p})}$$

where δ_s^2 is the variance in the frequency of the allele between different subpopulations, weighted by the sizes of the subpopulations, and $\bar{p}$ is the average frequency of an allele in the total population.

Results

Phenotypic yield data

The yield data showed that the tropical lines yielded more than the temperate genotypes in Zimbabwe. The top ten performing genotypes were all tropical genotypes while all the bottom 10 were temperate genotypes (Table 1). The frequency of the performance data of the genotypes is shown in Fig. 1. Only 15 genotypes were able to give yield that was above 4000 kg/ha and these were mainly of tropical origin. Out of the 49 genotypes which yielded between 3000 and 4000 kg/ha, 46 are of temperate origin. Most of the genotypes (70) were in the yield range of 2000–3000 kg/ha while no genotype gave a yield that was below 1000 kg/ha (Fig. 1).

Table 1 Top ten and bottom ten yield data for the soybean genotypes evaluated in Zimbabwe

Rank and trial statistics	Genotype name	Adaptation	Grain yield (kg/ha)
Top Ten performing genotypes	Saga	Tropical	4817.09
	Safari	Tropical	4761.03
	Serenade	Tropical	4734.82
	Saxon	Tropical	4730.93
	Mwenezi	Tropical	4501.17
	Solitaire	Tropical	4387.07
	Spike	Tropical	4375.53
	S722-6-1E	Tropical	4266.46
	S1440-5-2E	Tropical	4265.78
	Squire	Tropical	4243.98
Bottom Ten performing genotypes	Ozark	Temperate	1138.72
	NC-Tinius	Temperate	1119.99
	Spencer	Temperate	1112.50
	UI.San	Temperate	1107.63
	HF93-035	Temperate	1086.65
	HF93-083	Temperate	1075.04
	Defiance	Temperate	1052.57
	Clifford	Temperate	1042.83
	LN83-2356	Temperate	1011.36
	UA 4805	Temperate	1010.24
Statistics	Mean		2552.00
	SE mean		64.84
	STD		917.00
	P value		< 0.001
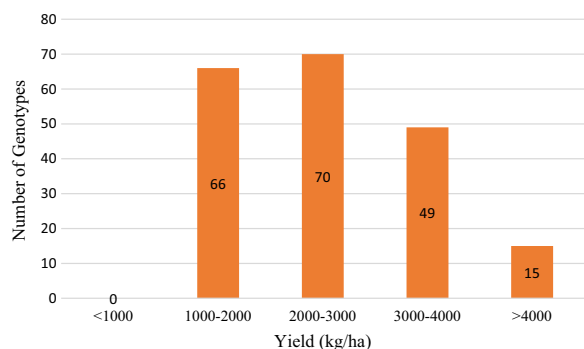
Fig. 1 Frequency distribution of 200 non-genetically modified soybean genotypes for grain yield

Table 2 SNP marker diversity for genotyping 210 diverse temperate and tropical soybean lines

	Mean	Min	Max
Major allele frequency	0.76	0.00	1.00
Minor Allele Frequency (MAF)	0.24	0.05	0.50
Expected Heterozygosity (H_e)	0.31	0.00	0.94
Observed Heterozygosity (H_o)	0.02	0.00	1.00
Polymorphic Information Content (PIC)	0.24	0.01	0.37
Allele number	1.88	1.00	3.00
Shannon information index	0.45	0.03	0.98

SNP marker diversity and profile

After filtering, 403 SNP markers remained with minor allele frequency > 0.05. The SNP marker profiles are presented in Table 2. The average minor allele frequency was 0.24. The number of alleles ranged from 1 to 3 with an average of 1.88. The Shannon Information index ranged from 0.03 to 0.98 with a mean of 0.45. The mean expected heterozygosity (He) was 0.31, whilst the mean observed heterozygosity was 0.02. The mean polymorphic information content (PIC) was 0.24.

Population structure

The silhouette plots showed that considering two clusters will produce one genotype with a negative silhouette value (Fig. 2a). When three clusters were considered, all the genotypes fitted perfectly into the three clusters (Fig. 1b). Having more clusters produced several genotypes with negative values on the silhouette plots.
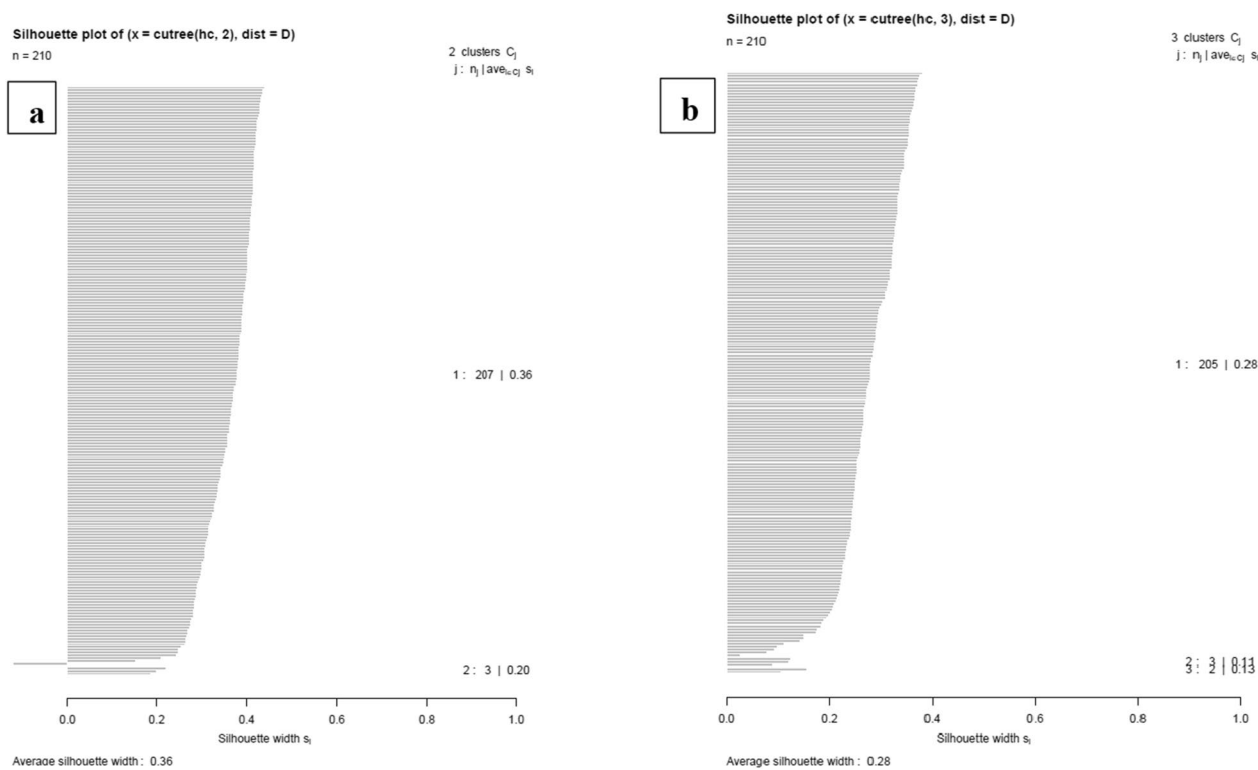
Fig. 2 Silhouette plots showing the number of possible clusters formed from 210 genotyped soybean lines **a**. considering 2 clusters **b.** considering 3 clusters

Therefore, three clusters were perfect in grouping all the genotypes (Fig. 2b) thus three clusters were the best fit for all genotypes. In the first cluster, 205 individuals were identified whilst cluster two and three had three and two lines, respectively. The average genetic distances (GD) were 0.28, 0.11 and 0.13 for the Clusters 1, 2 and 3, respectively.

According to the Gower's genetic distances calculated in R statistics, all the 210 genotypes were also grouped into three clusters as shown in the phylogenetic tree drawn using UPGMA cluster analysis (Fig. 3). The first cluster consisted of three temperate genotypes, Nitchuu 47, Tara and Tousan, while the second cluster consisted of two lines, namely Forrest and Fowler. The five genotypes in cluster one and two are all from USA. The third cluster consisted of 205 genotypes. The genotypes in this cluster consisted of all tropical genotypes from Zimbabwe, South Africa, Malawi and several temperate genotypes from the USA. There were genotypes which had short genetic distances (Fig. 3) between them such as Pudou 426 and Usada Zairai (0.02); Yougestu and Tachiyukata (0.02), UI. San and IC. San (0.05), Saga and Santee (0.07), Stanza and Mwenezi (0.08). Most of the lines from Zimbabwe are fitted in the third cluster. Three of the

South African genotypes clustered together. Several USA genotypes also clustered close to each other.

When only tropical lines were analysed three clusters were formed where all the Zimbabwean lines clustered together in the first cluster, while all the South African lines also clustered together in the second cluster (Fig. 4). The third cluster had Tikolore, the only line from Malawi. Sister lines clustered close to each other, for example S1440-5-1E and S1440-5-2E, as well as LDC-5-3 and LDC-5-9. Shortest genetic distance existed between Stanza and Mwenezi (0.08) and Solitaire and Pan 1867 with a genetic distance of 0.09. Greatest genetic distances were observed between Tikolore and Stanza (0.24), Tikolore and Mwenezi (0.17) and Tikolore and Serenade (0.12).

A UPMGA phylogenetic tree for temperate genotypes only is shown in Fig. 5. While this tree shows three clusters for these lines, the same lines that clustered close together when all 210 lines were included (including temperate lines), still clustered close to each other when these temperate lines were used in the analysis. Most of the LD lines clustered together just like when the temperate lines and tropical lines where used. Moreso, lines like Benning and Bingnan, Yougestu and Tachiyutaka and IC-San and
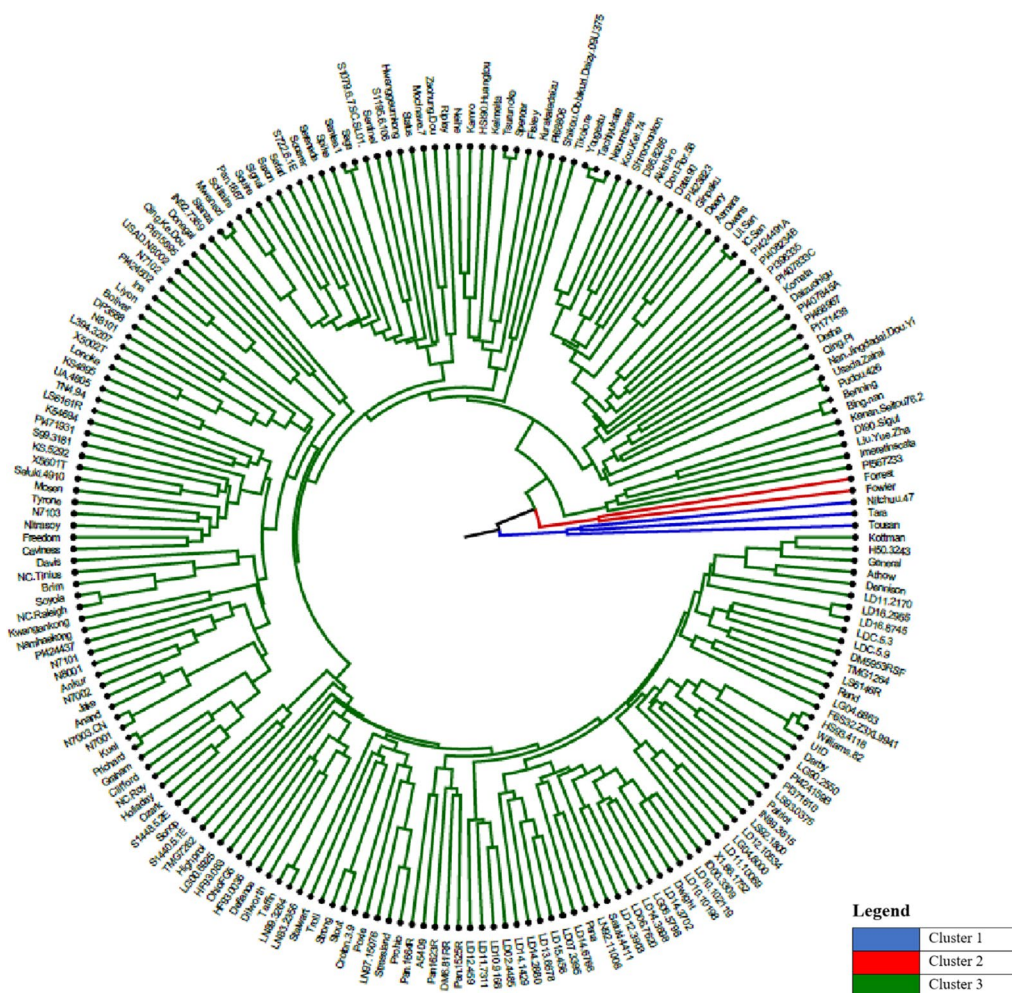
Fig. 3 UPMGA phylogenetic tree showing three clusters for all the 210 soybean lines drawn using the Gower's similarity distances

UI-San clustered close to each other with short genetic distances of 0.08, 0.02 and 0.05, respectively.

The Evanno method was used to reveal the optimum *k* value for the genotyped soybean lines in STRUCTU RE Harvester. The results of delta *k* (Δ*k*) curve show that the *k* peaked at 3 with a mean value of ln likelihood of -46516.5 and variance of ln likelihood of 3407.0 meaning a total of three clusters or subpopulations contributed to the total variation in the soybean lines under study (Fig. 6).

Population structure was constructed to reveal the architecture within the population. In agreement with the Evanno method, three sub populations were recognised (Fig. 7). Each of the colors (red, green and blue) in the population struture represents each cluster. The lines Fowler and Forrest (188 and 180 respectively) clustered close to each other while these are also closely clustered to Tousan (102), Tara (147) and Nutchu 47 which were in another cluster according to the UPMGA. Several other

genotypes consisted of genomes made of at least two of the subpopulations (Fig. 7).

Duplications

Clone analysis was done in GenAIEx to identify duplications. Table 3 shows the results. Two groups of duplicates were identified. Pudou-426 and Usada-Zairai were identified as duplicates while Tachiyukata and Yougestu were also identified as duplicates. The duplicate groups were labeled as A and B, respectively.

Genetic diversity among soybean lines

Analysis of molecular variance (AMOVA) was performed using the GenAIex for the three subpopulations identified in STRUCTURE. The AMOVA showed that total variation within the population can be partitioned into among- and within population sources, accounting for 4% and 96% of the total variation, respectively (Table 4). The F_{ST} value of 0.06 was low.
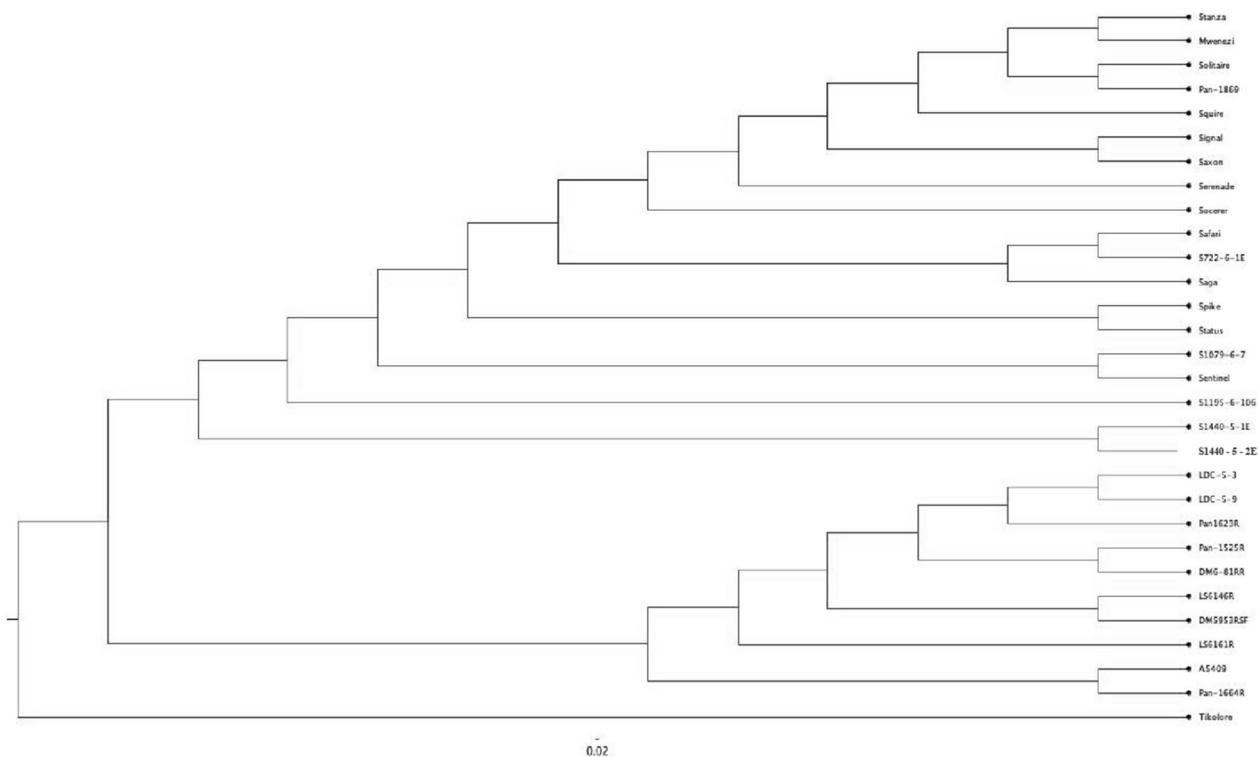
Fig. 4 Dendogram showing clustering of the 30 tropical soybean lines

Table 5 shows genetic variability among and within populations and the fixation index (F_{ST}) for the soybean lines. The Nei's net nucleotide distance ranged from 0.06 between cluster 1 and cluster 2 to 0.12 between cluster 2 and cluster 3. Cluster 1 and cluster 3 had a nucleotide distance of 0.09. This means that cluster 2 and 3 were furthest apart, whereas cluster 1 and 2 were closer to each other. The least within population variation was recorded in cluster 3 with an expected heterozygosity (H_e) of 0.21, whilst cluster 2 had the highest within population variation of 0.31. The fixation index (F_{ST}) were 0.06 (Cluster 1), 0.29 (cluster 2) and 0.02 (cluster 3). Cluster 3 had the lowest genetic variance proportion of 0.02 (Table 5).

Discussion
Phenotypic yield data
The results showed that the tropical lines yielded more than the temperate lines which indicates the tropical lines are well adapted to the Zimbabwean environment. This is usually expected especially when lines are introduced from a different region with different environmental conditions in terms of rainfall, latitude, altitude and temperatures. While the temperate genotypes yielded less than the tropical, 46 temperate genotypes yielded relatively better above 3000 kg/ha, indicating their potential utility for tropical and subtropical breeding programs.

These accessions can be utilized in soybean breeding programs for introgression of important traits, such as rust resistance and phenotypic maturity date if screened for such traits as this would reduce linkage drag effects on productivity (Abebe et al. 2021).

SNP marker diversity and profile
The SNPs used were quite informative and desirable for differentiating the soybean genotypes under study. The allelic number ranging from 1 to 3 can be attributed to the crop being self-pollinated, which is consistent with previous reports for low allelic diversity and heterozygosity levels for soybean (Abebe et al. 2021; Wright 1921). The mean minor allele frequency (MAF) value of 0.24, which is above 0 reflects the SNPs were informative. The MAF values measures the ability of markers to discriminate genotypes. With SNP markers due to their bi-allelic nature, a value above 0 is considered informative or discriminating. In the present study, 60% of the markers had a MAF between 0.3 and 0.5 which is comparable to values reported on soybean in previous studies (Chander et al. 2021; Abebe et al. 2021). The mean PIC value of 0.24 also indicates that the markers were informative. Considering the bi-allelic nature of SNPs where the PIC cannot exceed 0.5 (Singh et al. 2013), the PIC values obtained in this study were desirable
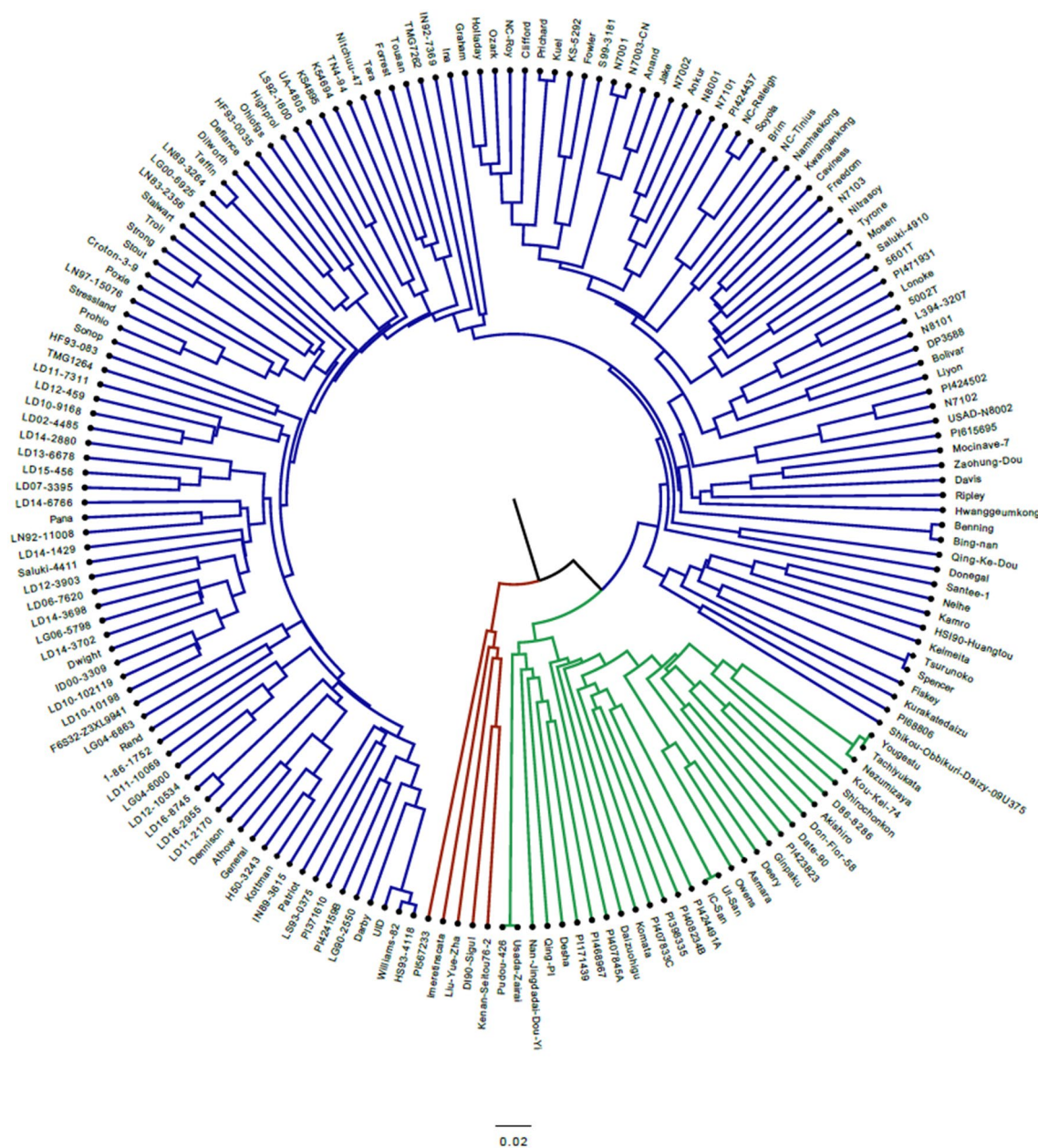
Fig. 5 UPMGA phylogenetic tree showing clusters of the 180 temperate soybean lines only

for differentiating the 210 soybean genotypes. Similar results were reported in soybean by Abebe et al. (2021) who reported a mean PIC value of 0.25 among elite lines developed by the IITA. In other self-pollinated crops, Singh et al. (2013) reported a mean PIC value of 0.23 in rice. The observed heterozygosity (H_o) of 0.02 was lower than the expected heterozygosity (H_e) in this study. This implies high possibilities of inbreeding and fixation at most of the loci (Nawaz et al. 2021). Overall, the SNPs used in this study were informative and

discriminating hence they can be recommended for diversity studies in other soybean populations.

Population structure and genetic diversity

The study was effective for determining the population structure and level of diversity in the germplasm collection. There was consistency in the outcome from the Silhouette plots, UPMGA and Evanno method in STRUCTURE used to discriminate the 210 soybean genotypes into clusters based on genetic similarity. The silhouette
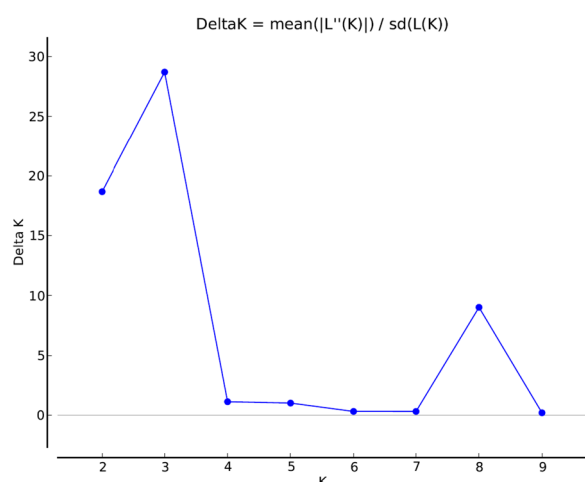
Fig. 6 Graph showing the best *k* value using the Evanno method

Table 3 Duplications of the soybean lines derived from clone analysis

Sample No	Sample	Pop	Number of duplications	Label of duplication
160	Pudou-426	2	2	A
125	Usada-Zairai	2	0	A
55	Tachiyukata	2	2	B
7	Yougestu	2	0	B

plots grouped the genotypes into three clusters perfectly, indicating that these were the effective number of clusters which could be formed from the germplasm used in this study. The silhouette plots are generally used to visualize how well the data points belongs to the cluster. The silhouette scores which range from -1 to 1 measure how similar an object is to its own cluster compared to other clusters (Menardi 2011; Pant et al. 2008; Rousseeuw 1987; Thinsungnoen et al. 2015). This finding was confirmed by two additional tools used in the study.

The Unweighted Pair Group Method using Arithmetic average (UPGMA) produced a phylogenetic tree with three populations which corroborated the findings from the silhouette plots and the Evanno method. While five genotypes from the USA (Nitchuu 47, Tara, Tousan, Forrest and Fowler) were grouped in clusters one and two, all other genotypes were grouped in the third cluster. The genotypes included in the third cluster were from different sources, from the USA, Zimbabwe, South Africa and Malawi. This means that there was limited molecular variation among the genotypes used in this study. This could be attributed to exchange of genetic material across the different breeding programs in the Southern Africa

region and external sources from other regions, such as Asia and America. An analysis of seed shipments indicates that there is a lot of germplasm exchange between the soybean breeding programs in Southern Africa and the USA. This implies that the soybean lines were derived from shared backgrounds and were selected for the same market requirements leading to utilization of the same set of alleles. According to the literature and actual pedigree analysis of this germplasm set, most soybean lines were developed from a narrow genetic base derived from a few ancestral lines. A survey of the literature indicates extensive utilization of external germplasm from different countries, such as China, Japan and Korea (Abebe et al. 2021; Bruce et al. 2019; Jeong et al. 2019a, b; Kim et al. 2014). It is a standard and recommended industry practice for breeders to continuously incorporate and integrate external germplasm in their breeding programs.

According to the phylogenetic tree of the 210 genotypes and a separate analysis of the tropical lines only, Zimbabwean and South African lines are clustered together separately. These lines were bred to satisfy the same market requirements with common trait preferences and common allelic constitutions. Several other genotypes clustered close to each other in accordance with their origin, adding credence to the possibility of utilizing common genetic background in breeding programs. Similar results of soybean genotypes that were clustered in accordance with the place of origin have been reported (Lee et al. 2014; Liu et al. 2017). This has also been reported for other legume crops, such as cowpea (Fatokun et al. 2018; Sodedji et al. 2021) and sesame (Basak et al. 2019). In the analysis involving tropical lines only, Tikolore was classified alone in its own cluster showing its potential for use in the tropical breeding programs for introgression of important traits.

Duplications show high level of genetic similarities (Makore et al. 2021) which was revealed in this study which is consistent with the findings from the phylogenetic tree that shows low genetic distances between some lines. Seemingly, the observations of duplications and minimal genetic distances indicates that there are introductions that were given different names by different breeders.

The results from analysis of molecular variance (AMOVA) supports the possibility of high gene flow as shown by the variation among populations that accounted for just 4% of the total variation, whilst within populations variation was about 96% of the total variation. The F_{ST} value of 0.06 indicated that there is low genetic difference among populations, suggesting high gene exchange. This observation is consistent with the literature. Wang et al. (2012) reported that most populations were exhibiting the effects of genetic bottlenecks.
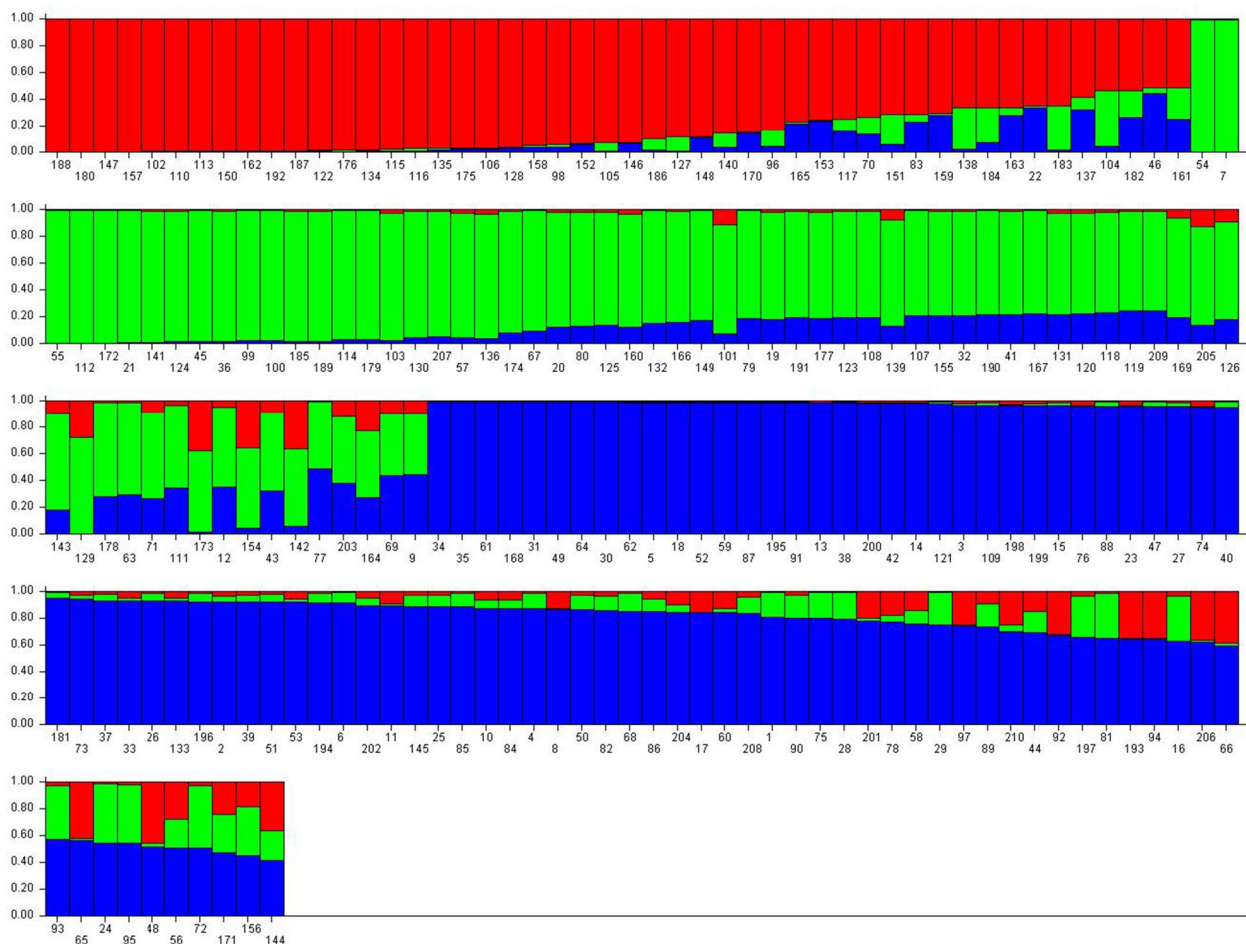
Fig. 7 Population structure of the 210 soybean lines

Table 4 Analysis of molecular variance (AMOVA) for the 210 soybean lines

Source	df	SS	MS	Est. Var	%	F_{ST}
Among pops	2	194.021	97.010	3.451	4	0.06
Within pops	207	16574.232	80.069	80.069	96	
Total	209	16768.252	80.230	83.520	100	

Table 5 Allele-frequency divergence among populations (Nei's Net nucleotide distance) and within populations (expected heterozygosity) and Fixation Index (F_{ST}) for 210 soybean lines

Population	Nei's nucleotide distance		Expected Heterozygosity	F_{ST}
	Cluster 2	Cluster 3		
1	0.06	0.09	0.30	0.06
2	–	0.12	0.31	0.29
3	–	–	0.21	0.02

Basak et al. (2019) also reported similar results in sesame. Abebe et al. (2021) cited moderate genetic variation and that 11% of the total variation was attributed to among clusters and 71% was due to individual genotypes and an F_{ST} value of 0.11 in soybean. Generally, low F_{ST} values close to 0 indicate that subpopulations are similar in almost all alleles or there is little divergence within the population, whilst F_{ST} value of 1 means the subpopulation is fixed at all alleles (Basak et al. 2019; Mohammadi and Prasanna 2003). In the current studies, the low F_{ST} values has an implication in breeding in that little

improvement can be done through simple hybridization in some traits of economic importance, for example yield. However, the low diversity can be utilized in conservation of such important traits by crossing the related genotypes. For example, crossing genotypes within cluster 3 to maintain high yields in some of the genotypes while taking advantage of some rare or minor alleles found in other genotypes. Minor alleles that can be leveraged on in such germplasm could be for earliness found in most USA genotypes. Genotypes from cluster 2 and 3 can be hybridized for improved varieties although the improvement has a certain ceiling because of the low genetic variation within the whole germplasm used in this study.

Conclusions and recommendations

The SNP markers used were informative and displayed high discrimination capacity, hence the results from this study were useful for molecular characterization of this soybean collection in Southern Africa. The 210 germplasm lines were consistently grouped into three clusters using three tools. Low molecular diversity was evident. These findings have serious implications for the breeding programs that aim to improve soybean varieties by utilizing this germplasm collection. Innovation strategies for improving variability in the germplasm collection, such as investments in pre-breeding, increasing the geographic sources of introductions and exploitation of mutation breeding would be recommended to enhance genetic gain.

Author contributions
AT conceptualization of the research, field work, data analysis, writing of the original draft, reviewing and editing of the final manuscript, EG data analysis, reviewing and editing, HM reviewing and editing, JFYE supervision, reviewing and editing, PT supervision, reviewing and editing, EYD supervision and reviewing, LM reviewing and editing, MZ selection of SNP markers, reviewing and editing, EZ reviewing and editing, JD supervision, reviewing and editing. All authors read and approved the final manuscript.

Availability of data and materials
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]West Africa Centre for Crop Improvement, College of Basic and Applied Sciences, University of Ghana, PMB 30, Legon, Accra, Ghana. [2]Seed Co Limited, Rattray Arnold Research Station, Chisipite, P. O. Box CH142, Harare, Zimbabwe. [3]University of Zimbabwe, MT Pleasant, P. O. Box MP167, Harare, Zimbabwe. [4]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Matopos Research Station, P.O. Box 776, Bulawayo, Zimbabwe. [5]International Institute of Tropical Agriculture (IITA), PMB 5320, Ibadan 200001, Nigeria.

References
Abebe AT, Kolawole AO, Unachukwu N, Chigeza G, Tefera H, Gedil M. Assessment of diversity in tropical soybean (Glycine max (L.) Merr.) varieties and elite breeding lines using single nucleotide polymorphism markers. Plant Genet Resour Charact Util. 2021;19(1):20–8. https://doi.org/10.1017/S1479262121000034.

Bandillo N, Jarquin D, Song Q, Nelson R, Cregan P, Specht J, Lorenz A. A population structure and genome-wide association analysis on the USDA soybean germplasm collection. Plant Genome. 2015. https://doi.org/10.3835/plantgenome2015.04.0024.

Bandillo NB, Anderson JE, Kantar MB, Stupar RM, Specht JE, Graef GL, Lorenz AJ. Dissecting the genetic basis of local adaptation in soybean. Sci Rep. 2017;7(1):1–12. https://doi.org/10.1038/s41598-017-17342-w.

Basak M, Uzun B, Yol E. Genetic diversity and population structure of the Mediterranean sesame core collection with use of genome-wide SNPs developed by double digest RAD-Seq. PLoS ONE. 2019;14(10):1–15. https://doi.org/10.1371/journal.pone.0223757.

Bellaloui N, Bruns HA, Gillen AM, Abbas HK, Zablotowicz RM, Mengistu A, Paris RL. Soybean seed protein, oil, fatty acids, and mineral composition as influenced by soybean-corn rotation. Agric Sci. 2010;1(3):102–9. https://doi.org/10.4236/as.2010.13013.

Biyeu K, Ratnaparkhe MB, Kole C. Genetics, genomics and breeding of soybean. New Hampshire: CRC Press; 2010. p. 1–18.

Blair MW, Cortés AJ, Penmetsa RV, Farmer A, Carrasquilla-Garcia N, Cook DR. A high-throughput SNP marker system for parental polymorphism screening, and diversity analysis in common bean (Phaseolus vulgaris L.). Theor Appl Genet. 2013;126(2):535–48. https://doi.org/10.1007/s00122-012-1999-z.

Bruce RW, Torkamaneh D, Grainger C, Belzile F, Eskandari M, Rajcan I. Genome-wide genetic diversity is maintained through decades of soybean breeding in Canada. Theor Appl Genet. 2019. https://doi.org/10.1007/s00122-019-03408-y.

Chander S, Garcia-Oliveira AL, Gedil M, Shah T, Otusanya GO, Asiedu R, Chigeza G. Genetic diversity and population structure of soybean lines adapted to sub-saharan africa using single nucleotide polymorphism (Snp) markers. Agronomy. 2021. https://doi.org/10.3390/agronomy11030604.

Chen Y, Nelson RL. Relationship between origin and genetic diversity in Chinese soybean germplasm. Crop Sci. 2005;45(4):1645–52. https://doi.org/10.2135/cropsci2004.0071.

Core TR. RStudio: Integrated development for R. RStudio, Inc., Boston. 2015. http://www.rstudio.com/. Accessed 15 Sept 2021.

Cornelious BK, Sneller CH. Yield and molecular diversity of soybean lines derived from crosses of Northern and Southern Elite parents. Crop Sci. 2002;42:642–7.

Cortés AJ, Chavarro MC, Blair MW. SNP marker diversity in common bean (Phaseolus vulgaris L.). Theor Appl Genet. 2011;123(5):827–45. https://doi.org/10.1007/s00122-011-1630-8.

Earl DA, VonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 2012;4(2):359–61. https://doi.org/10.1007/s12686-011-9548-7.

Edwards D, Forster JW, Chagné D, Batley J. What are SNPs? Assoc Mapp Plants. 2007. https://doi.org/10.1007/978-0-387-36011-9_3.

Tsindi *et al. CABI Agriculture and Bioscience*   (2023) 4:15

Page 13 of 14

FAOSTAT. Food and agriculture data. 2021. http://www.fao.org/faostat/en/#data/QC. Accessed 21 May 2022.

Fatokun C, Girma G, Abberton M, Gedil M, Unachukwu N, Oyatomi O, Yusuf M, Rabbi I, Boukar O. Genetic diversity and population structure of a mini-core subset from the world cowpea (Vigna unguiculata (L.) Walp.) germplasm collection. Sci Rep. 2018;8(1):1–10. https://doi.org/10.1038/s41598-018-34555-9.

Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. 2012. http://arxiv.org/abs/1207.3907. Accessed 10 Apr 2019.

Gower JC. A general coefficient of similarity and some of its properties. Biometrics. 1971;27(4):857–74.

Grieshop CM, Fahey GC Jr. Comparison of quality characteristics of soybeans from Brazil, China, and the United States. J Agric Food Chem. 2001;49:2669–73. https://doi.org/10.1021/jf0014009.

Gwinner R, Alemu Setotaw T, Pasqual M, Dos Santos JB, Zuffo AM, Zambiazzi EV, Bruzi AT. Genetic diversity in Brazilian soybean germplasm. Crop Breed Appl Biotechnol. 2017;17(4):373–81. https://doi.org/10.1590/1984-70332017v17n4a56.

Hahn V, Würschum T. Molecular genetic characterization of Central European soybean breeding germplasm. Plant Breed. 2014;133(6):748–55. https://doi.org/10.1111/pbr.12212.

Hyten DL, Choi IY, Song Q, Shoemaker RC, Nelson RL, Costa JM, Specht JE, Cregan PB. Highly variable patterns of linkage disequilibrium in multiple soybean populations. Genetics. 2007;175(4):1937–44. https://doi.org/10.1534/genetics.106.069740.

Hyten Dl, Choi I, Song Q, Specht JE, Carter TE, Shoemaker RC, Hwang EY, Matukumalli LK, Cregan PB. A high density integrated genetic linkage map of soybean and the development of a 1536 universal soy linkage panel for quantitative trait locus mapping. Crop Sci. 2010;50:960–8.

Jeong N, Kim KS, Jeong S, Kim JY, Park SK, Lee JS, Jeong SC, Kang ST, Ha BK, Kim DY, Kim N, Moon JK, Choi MS. Korean soybean core collection: genotypic and phenotypic diversity population structure and genome-wide association study. PLoS ONE. 2019a;14(10):1–16. https://doi.org/10.1371/journal.pone.0224074.

Jeong SC, Moon JK, Park SK, Kim MS, Lee K, Lee SR, Jeong N, Choi MS, Kim N, Kang ST, Park E. Genetic diversity patterns and domestication origin of soybean. Theor Appl Genet. 2019b;132(4):1179–93. https://doi.org/10.1007/s00122-018-3271-7.

Kim KH, Lee S, Seo MJ, Lee GA, Ma KH, Jeong SC, Lee SH, Park EH, Kwon YU, Moon JK. Genetic diversity and population structure of wild soybean (Glycine soja Sieb. and Zucc.) accessions in Korea. Plant Genet Resour Charact Util. 2014;12:48–51. https://doi.org/10.1017/S1479262114000239.

Lee G-A, Choi Y-M, Yi J-Y, Chung J-W, Lee M-C, Ma K-H, Lee S, Cho J, Lee J-R. Genetic diversity and population structure of korean soybean collection Using 75 microsatellite markers. Korean J Crop Sci. 2014;59(4):492–7. https://doi.org/10.7740/kjcs.2014.59.4.492.

LGC Bioscience Technologies. SeqSNP targeted GBS as alternative for array genotyping in routine breeding programs. 2019. https://biosearch-cdn.azureedge.net/assetsv6/seqsnp-tgbs-alternative-genotyping-routine-breeding-programs.pdf. Accessed 12 Feb 2020.

Li Y, Zhao S-C, Ma J-X, Li D, Yan L, Li J, Qi X, Guo X, Zhang L, He W, Chang R, Liang Q, Guo Y, Ye C, Wang X, Tao Y, Guan R, Wang J, Liu Y, Jin L, Zhang X, Liu Z, Zhang L, Chen J, Wang K, Nielsen R, Li R, Chen P, Li W, Reif J, Purugganan M, Wang J, Zhang M, Wang J, Qiu L-J. Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing. BMC Genomics. 2013. https://doi.org/10.1186/1471-2164-14-579.

Li YH, Reif JC, Jackson SA, Ma YS, Chang RZ, Qiu LJ. Detecting SNPs underlying domestication-related traits in soybean. BMC Plant Biol. 2014;14(1):1–8. https://doi.org/10.1186/s12870-014-0251-1.

Liu K, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics. 2005;21:2128–9. https://doi.org/10.1093/bioinformatics/bti282.

Liu Z, Li H, Wen Z, Fan X, Li Y, Guan R, Guo Y, Wang S, Wang D, Qiu L. Comparison of genetic diversity between Chinese and American soybean (Glycine max (L.)) accessions revealed by high-density SNPs. Front Plant Sci. 2017. https://doi.org/10.3389/fpls.2017.02014.

Ma YS, Wang WH, Wang LX, Ma FM, Wang PW, Chang RZ, Qiu LJ. Genetic diversity of soybean and the establishment of a core collection focused on resistance to soybean cyst nematode. J Integr Plant Biol. 2006;48(6):722–31. https://doi.org/10.1111/j.1744-7909.2006.00256.x.

Makore F, Gasura E, Souta C, Mazarura U, Derera J, Zikhali M, Kamutando CN, Magorokosho C, Dari S. Molecular characterization of a farmer-preferred maize landrace population from a multiple-stress-prone subtropical lowland environment. Biodiversitas. 2021;22(2):769–77. https://doi.org/10.13057/biodiv/d220230.

Malik MFA, Ashraf M, Qureshi AS, Khan MR. Investigation and comparison of some morphological traits of the soybean populations using cluster analysis. Pak J Bot. 2011;43(2):1249–55.

Menardi G. Density-based Silhouette diagnostics for clustering methods. Stat Comput. 2011;21(3):295–308. https://doi.org/10.1007/s11222-010-9169-0.

Mohammadi SA, Prasanna BM. Analysis of genetic diversity in crop plants—salient statistical tools and considerations. Crop Sci. 2003;43(4):1235–48. https://doi.org/10.2135/cropsci2003.1235.

Nadeem MA, Nawaz MA, Shahid MQ, Doğan Y, Comertpay G, Yıldız M, Hatipoğlu R, Ahmad F, Alsaleh A, Labhane N, Özkan H, Chung G, Baloch FS. DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. Biotechnol Biotechnol Equip. 2018;32(2):261–85. https://doi.org/10.1080/13102818.2017.1400401.

Nawaz MA, Lin X, Chan TF, Lam HM, Baloch FS, Ali MA, Golokhvast KS, Yang SH, Chung G. Genetic architecture of wild soybean (Glycine soja Sieb. and Zucc.) populations originating from different East Asian regions. Genet Resour Crop Evol. 2021;68(4):1577–88. https://doi.org/10.1007/s10722-020-01087-z.

Nemlı S, Kaygisiz Aşçioğul T, Ateş D, Eşıyok D, Tanyolaç MB. Diversity and genetic analysis through DArTseq in common bean (Phaseolus vulgaris L.) germplasm from Turkey. Turkish J Agric For. 2017;41(5):389–404. https://doi.org/10.3906/tar-1707-89.

Ojo DK, Ajayi AO, Oduwaye OA. Genetic relationships among soybean accessions based on morphological and RAPDs techniques. J Trop Agric Sci. 2012;35(2):237–48.

Oliveira MF, Nelson RL, Geraldi IO, Cruz CD, de Toledo JFF. Establishing a soybean germplasm core collection. Field Crop Res. 2010;119(2–3):277–89. https://doi.org/10.1016/j.fcr.2010.07.021.

Orf J. Introduction. In: Biyeu K, Ratnaparkhe MB, Kole C, editors. Genetics, gonomics and breeding of soybean. New Hampshire: CRC Press; 2010. p. 1–18.

Pant M, Radha T, Singh VP. Particle swarm optimization using Gaussian inertia weight. Proceedings—international conference on computational intelligence and multimedia applications, ICCIMA 2007, 2008; 1, 97–102. https://doi.org/10.1109/ICCIMA.2007.328.

Peakall R, Smouse PE. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics. 2012;28:2537–9.

Porras-Hurtado L, Ruiz Y, Santos C, Phillips C, Carracedo Á, Lareu MV. An overview of STRUCTURE: Applications, parameter settings, and supporting software. Front Genet. 2012;4(MAY):1–13. https://doi.org/10.3389/fgene.2013.00098.

Qin J, Shi A, Xiong H, Mou B, Motes D, Lu W, Miller JC, Scheuring DC, Nzaramba MN, Weng Y, Yang W. Population structure analysis and association mapping of seed antioxidant content in USDA cowpea (Vigna unguiculata L. Walp.) core collection using SNPs. Can J Plant Sci. 2016;96(6):1026–36. https://doi.org/10.1139/cjps-2016-0090.

Rafalski A. Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol. 2002;5(2):94–100. https://doi.org/10.1016/S1369-5266(02)00240-6.

Rambaut A. FigTree: molecular evolution, phylogenetics and epidemiology. 2016. http://tree.bio.ed.ac.uk/software/figtree/. Accessed 15 Sept 2021.

Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math. 1987;20:53–65. https://doi.org/10.1016/0377-0427(87)90125-7.

Singh N, Choudhury DR, Singh AK, Kumar S, Srinivasan K, Tyagi RK, Singh NK, Singh R. Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. PLoS ONE. 2013;8(12):1–14. https://doi.org/10.1371/journal.pone.0084136.

Sodedji FAK, Agbahoungba S, Agoyi EE, Kafoutchoni MK, Choi J, Nguetta SPA, Assogbadjo AE, Kim HY. Diversity, population structure, and linkage disequilibrium among cowpea accessions. Plant Genome. 2021. https://doi.org/10.1002/tpg2.20113.

Thinsungnoen T, Kaoungku N, Durongdumronchai P, Kerdprasop K, Kerdprasop N. The Clustering Validity with Silhouette and Sum of Squared

Tsindi *et al. CABI Agriculture and Bioscience*   (2023) 4:15

Page 14 of 14

Errors. In proceedings of the 3rd international conferance on industrial application engineering. Japan: The Institute of Industrial applications Engeineers. 2015; 44–51. https://doi.org/10.12792/iciae2015.012

Tiwari S, Tripathi N, Tsuji K, Tantwai K. Genetic diversity and population structure of Indian soybean (Glycine max (L.) Merr.) as revealed by microsatellite markers. Physiol Mol Biol Plants. 2019;25(4):953–64. https://doi.org/10.1007/s12298-019-00682-4.

Valliyodan B, Brown AV, Wang J, Patil G, Liu Y, Otyama PI, Nelson RT, Vuong T, Song Q, Musket TA, Wagner R, Marri P, Reddy S, Sessions A, Wu X, Grant D, Bayer PE, Roorkiwal M, Varshney RK, Liu X, Edwards D, Xu D, Joshi T, Cannon SB, Nguyen HT. Genetic variation among 481 diverse soybean accessions, inferred from genomic re-sequencing. Sci Data. 2021;8(1):1–9. https://doi.org/10.1038/s41597-021-00834-w.

Wang Y, Guo J, Liu Y, Wang Y, Chen J, Li Y, Huang H, Qiu L. Population structure of the wild soybean (Glycine soja) in China: Implications from microsatellite analyses. Ann Bot. 2012;110(4):777–85. https://doi.org/10.1093/aob/mcs142.

Wright S. Systems of mating. II. The effects of inbreeding on the genetic composition of a population. Genetics. 1921;6:124–43.

Yang S, Pang W, Ash G, Harper J, Carling J, Wenzl P, Huttner E, Zong X, Kilian A. Low level of genetic diversity in cultivated Pigeonpea compared to its wild relatives is revealed by diversity arrays technology. Theor Appl Genet. 2006;113(4):585–95. https://doi.org/10.1007/s00122-006-0317-z.

Zavinon F, Adoukonou-Sagbadja H, Keilwagen J, Lehnert H, Ordon F, Perovic D. Genetic diversity and population structure in Beninese pigeon pea [Cajanus cajan (L.) Huth] landraces collection revealed by SSR and genome wide SNP markers. Genet Resour Crop Evol. 2020;67(1):191–208. https://doi.org/10.1007/s10722-019-00864-9.

Zhu YL, Song QJ, Hyten DL, Van Tassell CP, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB. Single-nucleotide polymorphisms in soybean. Genetics. 2003;163(3):1123–34. https://doi.org/10.1093/genetics/163.3.1123.

Ziervogel G, New M, van Garderen EA, Midgley G, Taylor A, Hamann R, Stuart-Hill S, Myers J, Warburton M. Climate change impacts and adaptation in South Africa. Wiley Interdiscip Rev Clim Ch. 2014. https://doi.org/10.1002/wcc.295.

Žulj Mihaljević M, Šarčević H, Lovrić A, Andrijanić Z, Sudarić A, Jukić G, Pejić I. Genetic diversity of European commercial soybean [Glycine max (L.) Merr.] germplasm revealed by SSR markers. Genet Resour Crop Evol. 2020;67(6):1587–600. https://doi.org/10.1007/s10722-020-00934-3.

Publisher's Note