Vol. 14(1), pp. 1-11, January-March 2022
DOI: 10.5897/JPBCS2021.0986
Article Number: 73C9C2668625
ISSN 2006-9758

http://www.academicjournals.org/JPBCS

ACADEMIC JOURNALS
expand your knowledge

Journal of Plant Breeding and Crop Science

Full Length Research Paper

Genetic diversity assessment among 18 elite cashew tree genotypes (*Anacardium occidentale* L) selected in Western Burkina Faso

Konate Moumouni[1*], Tarpaga Windpouiré Vianney[1], Bourgou Larbouga[2] and Wonni Issa[1]

[1]Institut de l'Environnement et de Recherche Agricole (INERA), Centre National de Spécialisation en Fruits et Légumes (CNS-FL), DRREA-Ouest, 01 BP 910 Bobo Dioulasso 01, Burkina Faso.
[2]Institute de l'Environnement et de Recherches Agricoles (INERA), Programme Coton. 01 BP: 208 Bobo-Dioulasso 01, Burkina Faso.

Received 1 December, 2021; Accepted 5 January, 2022

The development of cashew orchards increased in recent years in Burkina Faso, due to high nut demand on international markets. However, little activities have been oriented toward cashew cultivar development, and farmers often use seed plants with little or no information about genetic characteristics of the material at hand. Therefore, a description of cashew diversity is needed to provide guidance to farmers and identify elite material for the crop improvement. In this study, phenotypic attributes of cashew accessions collected in Western Burkina Faso were recorded. Then, Gower's distance was used to show phenotypic relationship among accessions. Furthermore, four microsatellite markers were used to assess the genetic diversity of accessions, based on amplicon sequences. Sequence patterns across samples were converted into 75 binary alleles, combining sequence length and nucleotide polymorphisms. The 75 binary markers were 100% polymorphic and provided high average alleles per primer of 18.75. The polymorphic information content (PIC) varied between 0.003 and 0.895, averaging 0.534 per primer. Gower's distance between cultivars varied between 0.151 and 0.894. Dendrograms based on Nei's distance and Gower's distance revealed three main clusters each, although group compositions were different. These results were discussed with an outlook on future cashew tree breeding in Burkina Faso.

Key words: horticulture, breeding, molecular markers, phenotype, nut.

INTRODUCTION

The genus *Anacardium* contains eight species native to tropical America, of which *Anacardium occidentale* L, the cashew tree, is the most economically important. The cashew tree (2n = 42), is cultivated in many countries (FAO, 2020), with Western Africa contributing about 45% of global production (Audouin, 2014; Monteiro et al., 2017). Cashew tree is an ecologically flexible plant growing in a wide range of rainfall (500 - 3700 mm/year) and soils (Gupta, 1993). Cashew is grown for diverse reasons, including but not limited to land protection against erosion (Kouakou et al., 2020) and nutritional and health benefits. Many reviews highlighted the health

*Corresponding author. E-mail: mouni.konate@gmail.com, Tel: +226 70049504.

outcomes of regular cashew nut consumption such as the beneficial impacts on memory, obesity, diabetes mellitus, hypertension and cardiovascular diseases (Ros, 2010; Rico et al., 2016; de Souza et al., 2017; Rusu et al., 2019; Arsian et al., 2020). Furthermore, the same authors showed that nut-rich diets prevent oxidative stress, inflammation, visceral adiposity, hyperglycemia, insulin resistance, endothelial dysfunction, and metabolic syndrome, which are precursors of chronic diseases. Such health benefits of cashew nuts conferred prime importance to the crop worldwide.

In Burkina Faso, more than 250,000 ha are cultivated in cashew, primarily in the Western part of the country, with an annual production averaging 100,000 tonnes (Gecit, 2020; INSD, 2020). Cashew nuts represent the third largest export product, after gold mining and cotton, accounting for about 200 million USD, say near 5% of the total value of exports (INSD, 2020). In recent years, many efforts were deployed to enhance the economic value of the cashew industry in Burkina Faso by increasing domestic production and modernising cashew processing. The increasing commercial importance of cashew nut (Savadi et al., 2020; Tarpaga et al., 2020) makes us expect a significant expansion of the production to new regions. Consequently, the high demand for seed plants will go with the need for market desirable traits to meet domestic and international demands.

Despite its broad adaptation to even marginal environmental conditions (Gupta, 1993), cashew tree's productivity is currently low (Tarpaga et al., 2020), far below the cashew crop's potential (>1000 kg/ha) (Kouakou et al., 2020). Plant fruit setting and overall yield can be affected by many factors such as maintenance conditions, genetics, pollinators and environmental conditions (Yaman and Uzun, 2020, 2021). Low productivity could be also attributable to the use of genetically heterogeneous cultivars by farmers, the weakness of breeding activity to improve yields and nut quality (Tarpaga et al., 2020) and the presence of diseases (Wonni et al., 2017), which are exacerbated by increased climatic instability (Ben Zaied and Ben Cheikh, 2015). Therefore, the selection and breeding of high performing varieties is both important and urgent, to meet market demand for cashew nut. Currently, little is known about agro-morphological and genetic characteristics of locally produced cashews (Tarpaga et al., 2020).

One prerequisite to propose good performing varieties to farmers, and thus increase the quality and quantity of the production, is a comprehensive knowledge of agronomic potentials of planted cultivars (Tarpaga et al., 2020). However, the breeding process is tedious due to the perennial nature of the plant and its prolonged juvenile phase (Syed et al., 2005; Savadi et al., 2020). Additionally, genetic improvement of cashew by traditional approaches is slow and unpredictable due to high degree of heterozygosity and genetic complexity of most agronomic traits (Syed et al., 2005).

Therefore, the application of molecular markers to assist the pre-breeding and breeding processes is highly desirable to accelerate varietal selection and genetic gain.DNA markers are known to be highly efficient in diversity assessment of plant species (Yildiz et al., 2021). DNA analysis can be performed with only small amounts of material from any part of the plant, and markers are independent of environmental conditions and plant phenological stages (Lande and Thompson, 1990; He et al., 2014). Many marker systems have been applied in cashew, including microsatellites, Random Amplified Polymorphic DNA (RAPD) (Santhosh et al., 2009; Jena et al., 2016), single nucleotide polymorphisms (SNPs) (Savadi et al., 2020), Amplified fragment length polymorphism (AFLP) (Archak et al., 2003), or sequence-specific amplification polymorphism (SSAP) (Syed et al., 2005). These marker systems could provide sufficient polymorphism to discriminate within and between populations (Archak et al., 2003) and even construct genetic maps (Syed et al., 2005). Of these, microsatellites have been very popular due to their codominance and high polymorphism (Croxford et al., 2006; Guichoux et al., 2011; Makueti et al., 2015; Ngure et al., 2021) for fingerprinting and diversity estimation. The aim of this study was to assess the genetic relationship among 18 cashew genotypes collected in western Burkina Faso, based on microsatellite markers associated with a few phenotypic attributes.

MATERIALS AND METHODS

Plant material

A collection of 18 cashew trees selected from orchards in Western Burkina Faso (Table 1) was used to conduct phenotypic and genotypic investigations. We previously determined agro morphological characteristics of these trees (Tarpaga et al., 2020), which were selected among 820 candidate trees, based on specific performance criteria to make a collection of elite cultivars.

Genomic DNA extraction and amplification

Leaf samples were collected from each of the germplasm accessions for genotyping. Each sample was obtained from three to five young leaves, and were used for DNA extraction using the CTAB method (cetyltrimethylammonium bromide) as described previously by Doyle and Doyle (1990). Extracted DNA was quantified through the Nanodrop model N1000 spectrophotometer (Thermo Fisher Scientific) and then checked for genomic DNA integrity using 1% agarose gel electrophoresis.

Four microsatellite markers selected among those developed by Croxford et al. (2006) were used to estimate the genetic diversity of cashew cultivars (Table 2). The screening of accessions was performed using marker specific primers through Polymerase chain reaction (PCR) amplification. The PCR was carried out in a Bioreba thermocycler, using a reaction volume of 25 µl containing 5 µl of 5X Promega buffer, 0.5 µl of each primer, 0.5 µl of dNTPs, 0.05 µl of Taq polymerase and 2 ng of DNA. The PCR programme consisted

Table 1. List and collection sites of 18 germplasm accessions used in this study.

Accession_ID	Province	County of collection
ET02	Comoé	Soubakaniédougou
ET03	Comoé	Soubakaniédougou
ET04	Comoé	Soubakaniédougou
ET05	Comoé	Soubakaniédougou
ET06	Comoé	Banfora
ET07	Kénédougou	Orodara
ET08	Kénédougou	Orodara
ET09	Kénédougou	Orodara
ET10	Kénédougou	Orodara
ET13	Comoé	Mangodara
ET20	Léraba	Sindou
ET21	Léraba	Sindou
ET24	Léraba	Sindou
ET25	Léraba	Sindou
ET31	Léraba	Sindou
ET32	Léraba	Sindou
ET36	Kénédougou	Kangala
ET53	Comoé	Niambrigo

Table 2. List of microsatellite markers and primer sequences used in this study.

Locus	Forward primer (5'-3')	Reverse primer (5'-3')	Repeat motif	Taille (bp)	TM (^{o}C)
mAoR3 (AM085801)	CAGAACCGTCACTCCACTCC	ATCCAGACGAAGAAGCGATG	$(AC)_{12}(AAAAT)_2$	241–247	60.3
mAoR6 (AM085802)	CAAAACTAGCCGGAATCTAGC	CCCCATCAAACCCTTATGAC	$(AT)_5(GT)_{12}$	143–157	58.2
mAoR11 (AM085804)	ATCCAACAGCCACAATCCTC	CTTACAGCCCCAAACTCTCG	$(AT)_3(AC)_{16}$	234–236	60.3
mAoR55 (AM085819)	TGACTTTCAAATGCCACAAC	CTCAAGCTTTCATGGGGATT	$(AT)_6CT(AC)_5$	104–114	58.2

Source: Croxford et al. (2006).

of an initial denaturation at 94°C for 5 min, followed by 35 cycles of denaturation at 94°C for 1 min, hybridization for 1 min at primer-specific temperatures and extension at 72°C for 1 min, and final extension at 72°C for 7 min. Next, PCR products were run along with 100 bp ladder, through 1% agarose gel to check the presence of a unique amplicon for each marker. The PCR products were then sent to GenWIZ for sequencing.

Generation of marker profiles

The software MUSCLE (https://www.ebi.ac.uk/Tools/msa/muscle/) was used to align amplified DNA sequences with VNTR (variable number tandem repeat) loci as well as their flanking regions.
For greater accuracy, nucleotide variations between repeats were evaluated and the tandems were counted (Supplemental file S1). The number of repeats at each VNTR locus included in the MLVA scheme (Multiple Locus Variable-Number Tandem Repeat Analysis) was converted into binary marker (present or absent). Furthermore, sequence sizes and base composition of insertions were also evaluated across accessions to take into consideration indels that may occur within VNTRs.

In this way, sequence size, repeats, and inserted bases were used in combination to generate a binary matrix of markers to define allelic profile of individuals.

Assessment of genotypic diversity

The matrix generated with binary markers was used to compute genetic diversity indices using the software GenAlex 6.5 (Peakall and Smouse, 2012). These indices included (i) the number of alleles in the population (Na); (ii) allele frequencies per locus, assuming marker dominance and Hardy-Weinberg equilibrium; (iii) Shannon diversity index; and (iv) an estimate of genetic diversity based on percentage of polymorphic loci. Additionally, the polymorphic information content (PIC) was computed using the formula: PIC = $1-\sum fi^2$, where '*fi*' = frequency of i^{th} allele (Botstein et al., 1980). Principal Coordinate Analysis (PCoA) of marker data was performed using the software GenAlex 6.5 (Peakall and Smouse, 2012) and the first two coordinates were used to present a scatter plot. Then, the binary marker data were used to generate a pairwise-distance matrix based on Nei's genetic distance (Nei, 1973) with the package *poppr* in R. Then, using the package *cluster*, a dendrogram was constructed based on Nei's distance and UPGMA as by Sneath and Sokal (1973).

Phenotypic diversity in the cashew population

The diversity of cashew accessions was evaluated using

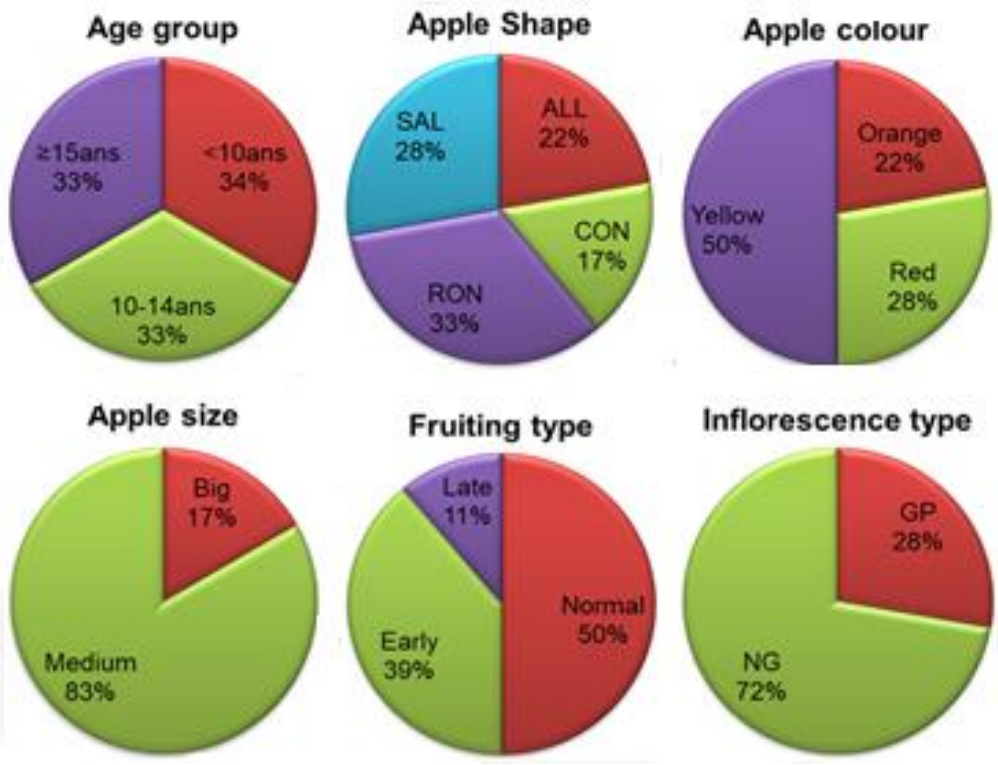

Figure 1. Summary statistics of qualitative variables in the 18 sample cashew trees. Apple shape includes SAL = hemi-elongate; ALL = elongate; CON = conical; RON = round. Inflorescence type includes NG = clustered; GP = sparce.

quantitative and qualitative traits, such as inflorescence type (Inflow), type of fruiting time (FruT), nut yield per tree (NutY), nut weight (NutW), shelling percentage (ShellP), apple size (ApSi), shape (ApSh) and colour (Col). Furthermore, the age of each subject tree was recorded, and then classified in three age groups: below 10 years; between 10 and 14, and 15 years and above.

Using the package *pastecs* in R, summary statistics of the data were computed. Quantitative data were summarised in a table. Then the Shapiro-Wilks test was applied to test for normality of the collection for quantitative traits, whereas qualitative data (ApSi, ApSh, Col, age range and geographic origin) were visualised using pie charts drawn in Microsoft Excel.

To produce pairwise-dissimilarity matrices based on phenotypic variables, we computed Gower's distance (Gower, 1971), which handles well data with different units in cluster analysis (Hummel *et al.,* 2017). The overall estimate of genetic dissimilarity was performed based on both qualitative and quantitative phenotypic data to generate a single data matrix. To sketch the relationship between cultivars, we performed a factor analysis for mixed data (FAMD) in R software, which output was used to construct a dendrogram.

Differentiation indices

Genetic distances between phenotypic and/or geographic groups in the cashew population were computed by performing a pairwise Phi statistic (Phi-St) (Michalakis and Excoffier, 1996) using *msap* package in R (Perez-Figueroa, 2013).

The significance of the Phi-St values was evaluated by the analysis of molecular variance (AMOVA), and the probability of a null hypothesis (Phi-St = 0) was estimated over 9999 permutations.

RESULTS

Phenotypic variability

The assessment of phenotypic diversity among 18 cashew tree genotypes revealed that this collection of elite cultivars presents noticeable variability in both qualitative (Figure 1) and quantitative traits. Although the coefficient of variation was low for nut weight (9%) and shelling percentage (6.5%), it was high (50.5%) for nut yield (Table 3). The Shapiro-Wilks test suggested that data were normally distributed for nut weight (P = 0.639) and shelling percentage (P = 0.123); whereas, data of nut yield were not (P = 0.019). The cashew tree ET06 had the poorest nut yield (23.5), but the best shelling percentage (33.9). The sample under investigation had three age groups equally shared between young, middle and old trees (Figure 1). Four apple shapes were present at proportions between 17% conical and 33% round, and half of the trees had semi-elongate or elongate apples. Most of trees presented yellow apples (50%) with predominant medium sizes (83%). While most individuals have normal fruiting time (50%), 39 and 11% of them have early and late fruiting schemes, respectively. Two

Table 3. Summary statistics of three quantitative variables in the 18 sample cashew trees.

Variable	NutW	NutY	ShellP
Number of values	18	18	18
Number of null values	0	0	0
Number of missing data	0	0	0
Minimum	6.51 (ET21)	23.5 (ET06)	27.13 (ET07)
Maximum	8.79 (ET10)	115.5 (ET31)	33.9 (ET06)
Range	2.28	92	6.77
1st Quartile	6.95	34.54	27.95
Median	7.425	48.325	29
Mean	7.457	57.725	29.451
3rd Quartile	8.008	71.88	30.46
SE mean	0.159	6.870	0.449
CI mean (0.95%)	0.335	14.495	0.947
Variance	0.455	849.598	3.628
Std.dev	0.674	29.148	1.905
Coefficient of variation	0.090	0.505	0.065
W	0.962	0.872	0.919
P-value (W)	0.639	0.019	0.123

NutW = nut weight; NutY = nut yield; ShellP = shelling percentage. SE = standard error; CI = confidence interval; Std.dev = standard deviation; W = Shapiro-Wilks normality test; P-value (W) = probability associated to Shapiro-Wilks test.

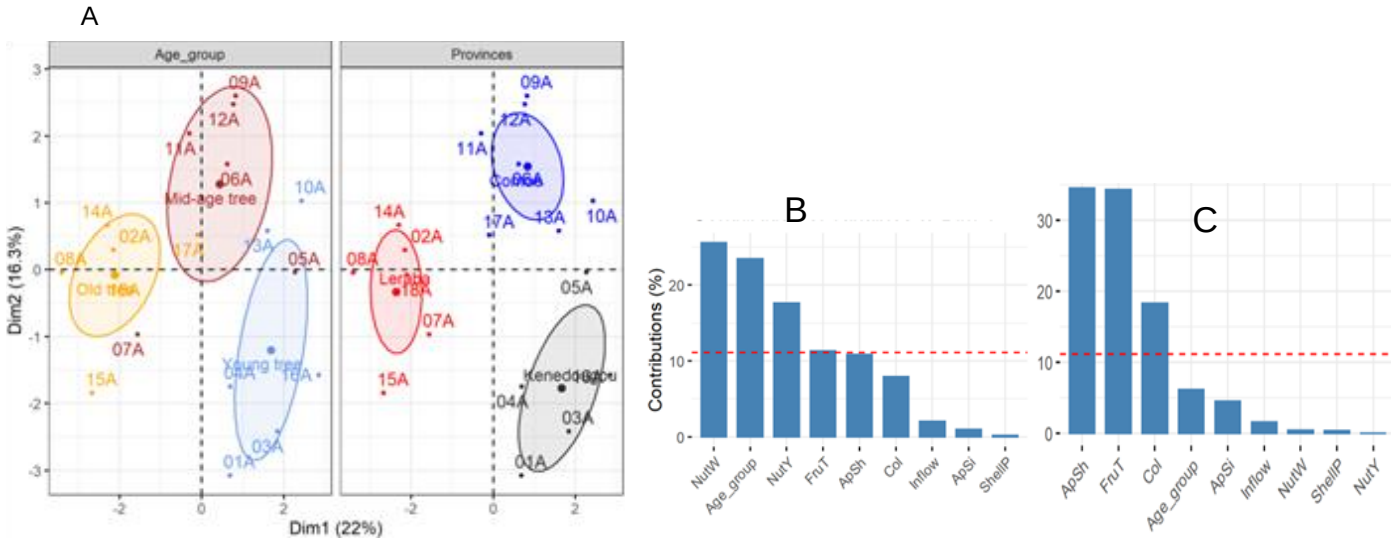

Figure 2. Scatter plots of sample cashew cultivars based on the FAMD and presenting trees' age groups and provinces of origin (A). Contribution of quantitative and qualitative variables to the dimensions 1 (B) and 2 (C). The red dashed lines on the graphs (B and C) show the expected average value, if the contributions were uniform.

inflorescence types were detected, of which 72% were clustered (NG) and 28% were sparce (GP) (Figure 1).

Multivariate analyses

FAMD-based scatter plots of sample cashew trees revealed that clusters correlate with age groups and provinces of origin (Figure 2). This grouping was mainly influenced by such traits as nut weight, tree age group and nut yield on the first dimension (Figure 2), and apple shape, fruiting time and apple colour on the second dimension (Figure 2).

The estimates of Gower's distance distribution were

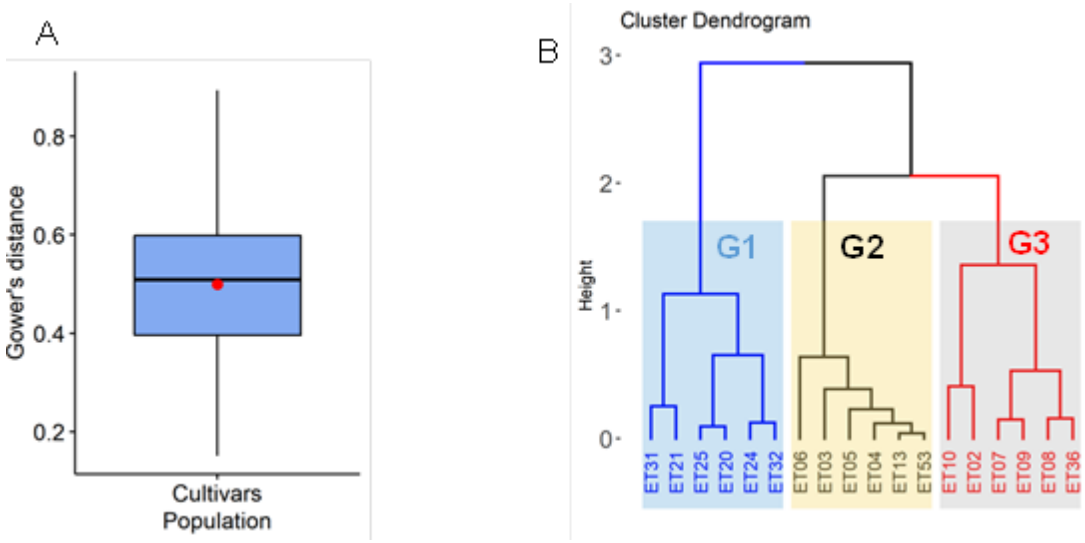

Figure 3. (A) Boxplot of pairwise Gower's distances between cashew cultivars. The horizontal line in the box shows the median whereas the red dot indicates the mean. (B) Dendrogram depicting genetic relationship among cashew cultivars based on morphometric characters using Gower's distance.

Table 4. Summary of alleles obtained from amplified sequences across the 18 cashew accessions. N = population size; Na = number of alleles; PIC = polymorphic information content.

Allele base	Loci				Total/mean
	maoR3	maoR6	maoR11	maoR55	
N	18	18	18	18	18
Na repeat number	5	8	7	9	31
Na sequence size	9	11	9	12	41
Na inserted motif	0	0	0	5	5
Na total	14	19	16	26	75
Polymorphic alleles	14	19	16	26	75
PIC value	0.895	0.592	0.648	0.003	
Shannon index	0.116	0.193	0.227	0.254	0.217±0.014
Mean Na per locus					18.75
Mean PIC value					0.535

presented in a boxplot which showed that cultivars are concentrated around the mean (Figure 3A). The genetic dissimilarity matrix generated based on phenotypic descriptors, showed that the mean value of Gower's distance was 0.5 (Figure 3A, red dot), with 0.151 and 0.894 as lower and upper extremes, respectively (Figure 3A). ET31 and ET36 were the most distant cultivars (0.894); whereas, ET20 and ET25 were the closest (0.151). The dendrogram obtained from correspondence analysis revealed that the population under investigation comprised three groups (G1, G2 and G3), each composed of equal number of individuals (Figure 3B). A close look at group composition showed that the clusters correlated with nut yield, which in turn seemed to depend on the age of trees.

Allelic variation and genetic diversity in the cashew population

Multiple alignments allowed us to estimate the exact number of repeats for each locus and the number of different patterns per VNTR locus. The four microsatellite loci used to screen the population of 18 cashew individual trees were all polymorphic (Table 4). Apart from the maoR11 locus which showed a single dinucleotide (CA) tandem repeat, the other three loci (maoR6, maoR3 and maoR55) presented two dinucleotide motifs of tandem repeats. The Conversion of patterns in these four loci into binary markers provided a total of 75 markers, based on sizes, repeat number, and insertion motifs (Table). With this approach, high numbers of alleles were obtained per

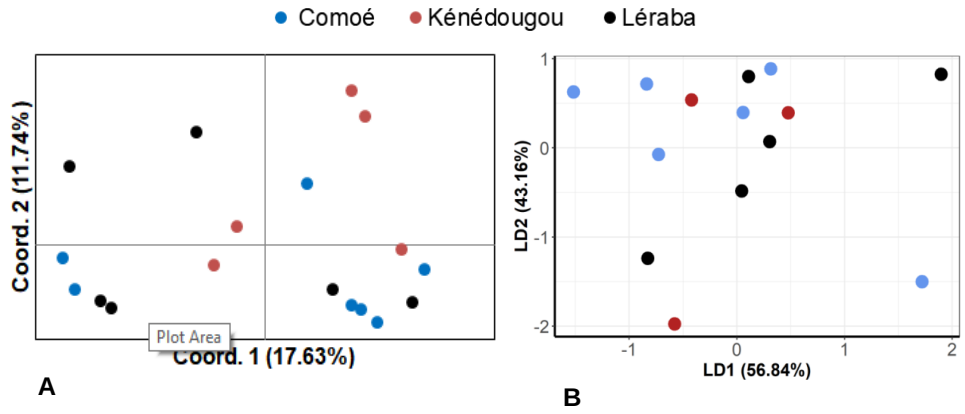

Figure 4. Scatter plot showing the relationships among the 18 cashew cultivars based on marker data. A = Principal Coordinates Analysis (PCoA) constructed using the first two coordinates (Coord 1 and Coord 2). B = Discriminant analysis of geographic origins of cashew cultivars. Coord. = coordinate; Pop = population; LD = Least discriminant factor.

Table 5. Pairwise PhiST values (genetic distances) based on genotypic profiles and calculated using msap package in R.

Correlation	Comoe	Kenedougou	Leraba	Phi-ST	P value
Comoe	-	0.134	0.270		
Kenedougou	0.041	-	0.297		
Leraba	0.016	0.019	-		
Total				0.026	0.176

PhiST between provinces, shown below the diagonal and P-values based on 9999 permutations are shown above the diagonal.

locus. Thus, the marker maoR55 was the most polymorphic with 26 alleles in total; followed by maoR6, 19 alleles and the remaining two loci (maoR3 and maoR11) had 16 and 14 alleles, respectively (Table 4). The allelic richness per locus (Na) was 18.75 and varied between 14 and 26 depending on individual markers. Marker primers used in this study provided high PIC values of 0.59 or above, except for the locus maoR55 which PIC value was as low as 0.003. Additionally, the overall Shannon index was about 0.217 ±0.074 among the cashew cultivars (Table 4).

Genetic structure of the cashew population

The principal coordinates analysis of genotypic data did not present any grouping related to geographic origins of cultivars (Figure 4A). Further discriminant analysis revealed that there is no geographic clustering of cultivars (Figure 4B). This lack of geographic structure was confirmed by AMOVA (analysis of molecular variance) and pairwise Phi-ST between provinces, which did not exhibit any significant spatial distinction (Table 5). Thus, the overall genetic distance between provincial samples was trivial (Phi-ST = 0.026, P = 0.176). None of the pairwise Phi-ST between provinces were significant (Table 5).

However, the AMOVA table showed a significant overall variability (Phi-ST = 0.224; P<0.0001) among the trees (Table 6). Additionally, Nei's pairwise distance-based dendrogram resulted in three genotypic clusters (Figure 5), which distinctiveness were revealed by AMOVA (pairwise P-values = 0.0001, 0.0135, or 0.0373; Table 7).

DISCUSSION

Phenotypic variability of cashew accession from western Burkina Faso

Assessment of agro-morphological variations, including qualitative and quantitative traits is commonly used toestimate crop genetic diversity (Sunita et al., 2021; Sousa et al., 2019; Tarpaga et al., 2020). Such traits are important in the breeding process, as they are easy to evaluate (Ngure et al., 2021). The most dominant qualitative traits were medium size of apples (83%), the NG inflorescence type (72%) and the yellow colour of

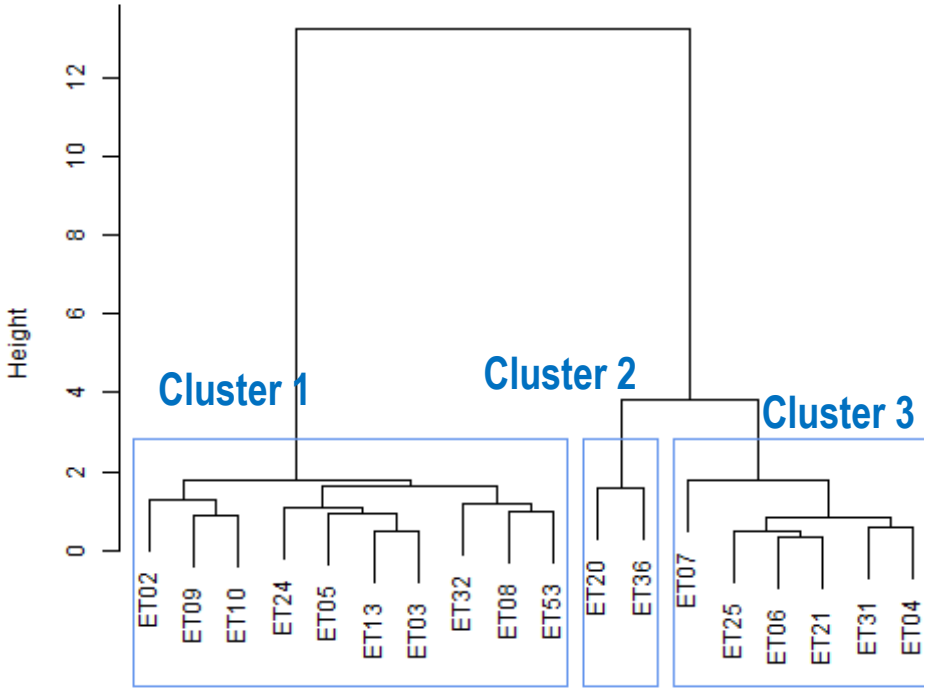

Figure 5. An UPGMA dendrogram constructed using Nei's genetic distance based on genotypic data.

Table 6. Analysis of molecular variance (AMOVA) of 18 cashew cultivars.

Variable	d.f.	SSD	MSD	Variance	Phi_ST	P value
among-groups	2	32.13	16.07	1.873		
within-groups	15	97.37	6.491	6.491		
Total	17	129.5	7.618	-	0.224	0.0001

d.f. = degrees of freedom; SSD = sum of squared differences; MSD = mean squared differences.

Table 7. Pairwise Phi-ST between the clusters obtained from the principal components analysis (PCA) of Nei's distance.

Comparison	Phi-ST	P value
Cluster 1 vs. Cluster 2	0.2317	0.0001
Cluster 1 vs. Cluster 3	0.1931	0.0135
Cluster 2 vs Cluster 3	0.2523	0.0373

apple (50%). However, at the western Burkina Faso scale, these proportions may be distorted by sampling bias (Panzeri et al., 2008; Michelangeli et al., 2015) due to the initial conditional selection of elite trees for this study. Nevertheless, the presence of diverse variants of apple characteristics (shape, size and colour), fruiting times, and inflorescence types, indicate that the diversity within this population is substantial. However, apart from nut yield which presented a significant variability between individual trees (cv >50%), there was low variability for quantitative traits such as nut weight and shelling percentage. The lack of sufficient variability for these traits was a direct consequence of the trees' selection process, which only retained individuals with nut weight and shelling percentage above 6 g and 25%, respectively (Tarpaga et al., 2020). Furthermore, nut yield showed a

rough correlation with geographic origin of cultivars. This could be attributed to space-specific performance; however, as trees from the same province were nearly in the same age group, such a conclusion may appear erroneous, because old trees usually produce more than their younger counterparts (Aliyaman and Indradewa, 2019; Tarpaga et al., 2020). This is probably explanatory of the lack of normal distribution for nut yield. Furthermore, multivariate analysis of phenotypic data presented significant genetic diversity in the population under investigation. As variability is essential to sustainable crop improvement (Cobb et al., 2019), genetic improvement can be undertaken from these local cultivars, relying on phenotypic selection. However, although phenotype-based selection usually prevails during orchard implantation (Tarpaga et al., 2020), this may be subject to environmental biases which can be circumvented (Lande and Thompson, 1990; He et al., 2014) by the use of molecular markers.

Marker scoring method and polymorphism

The marker system used in this study exploited sequence data from only forward strands of DNA. Though this is not common, in the absence of reverse strand sequences, codominant SSR markers were conveniently converted into binary markers. This approach had the advantage of detecting many alleles, which increase markers' efficiency in revealing genetic variability (Croxford et al., 2006). However, the lack of codominance appeared as a major drawback of this method. Viz., dominant markers often provide theoretical measures of heterozygosity, which has limited usefulness (Peakall and Smouse, 2012). Nonetheless, the utility of molecular markers in genetic diversity assessment is dependent on their polymorphism and power of discrimination among the genotypes under investigation (Croxford et al., 2006; Santhosh et al., 2009). In this study, the generated binary markers were not only polymorphic but were sufficiently informative (mean PIC = 0.535) to discriminate cashew individual trees. This suggests that the SSRs developed in cashew (Croxford et al., 2006) were of great quality, as testified by many other studies using all or part of the original set of these markers (Sika et al., 2015; Kouakou et al., 2020). This result corroborates with a previous diversity assessment of cashew trees in Benin, in which a set of eight markers presented a polymorphism of 100% (Sika et al., 2015). Microsatellite markers usually present high PIC values, due, at least partially, to higher mutation rates in SSRs compared with other marker types such as SNPs (Chen et al., 2017). However, the low PIC value of the polymorphic SSR maoR55 (0.003) shows that this locus presented many, but rare, alleles (up to 29). It is important to recognise that allelic frequencies and related metrics are sensitive to sample size (Müller et al., 1996). This, therefore, must be kept in mind when evaluating allele frequencies in a population.

Genotypic diversity and population structure

The level of genotypic diversity present among the studied cultivars was substantial (I = 0.217) for an introduced species. This diversity index was higher than that reported (I = 0.04) from the analysis of sixty accessions collected in central Benin (Sika et al., 2015). As a crop only recently introduced in West Africa (Aliyu, 2012; Kouakou et al., 2020), a high genetic variability was not expected in the studied cashew accessions, compared to that present in regions of the species origin. Thus, most of the studies conducted in this sub-region using the SSR markers showed a low genetic differentiation among local germplasms, such as that in Nigeria (Aliyu and Awopetu, 2007; Aliyu, 2012), Benin (Sika et al., 2015) and Cote d'Ivoire (Kouakou et al., 2020). Except Brazil where cashew originates, the crop's diversity is often low to moderate in other parts of the world (Kouakou et al., 2020). However, the detection of such a high diversity in the population under this study may be attributable to the polymorphism obtained from the marker system applied. Therefore, diversity metrics need to be taken with caution as they are influenced by many factors, including sample size, marker system, and number of alleles (Müller et al., 1996; Powell et al., 1996). There was no geographic structure of cultivars as shown by the AMOVA and Phi-St on genotypic data. The lack of correlation between clusters and geographic distribution of samples could be attributable to the sampling of the trees, which was based solely on their performance. Thus, the selection of elite individuals across provinces cannot capture neither the entire diversity of the crop, nor the spatial distribution of such a diversity. Nevertheless, the grouping revealed by the PCA shows that there is substantial genetic variability among the selected cultivars independently of geographic origins, as supported by the AMOVA among the three PCA-based groups (overall Phi_ST = 0.224; $P < 0.0001$). This result suggests that several introductions occurred over the years in Burkina Faso, and then, cultivars were probably disseminated through the same route around producing areas, without any cultivar confinement to a particular producing region. This hypothesis is consistent with conclusions drawn from previous studies in Cote d'Ivoire and India, where the authors thought that orchards in the respective countries were planted with material from a single-entry source (Archak et al., 2009; Kouakou et al., 2020). The detected diversity might also arise from the predominance of entomophilous crosspollination (Vanitha and Raviprasad, 2019) and self-incompatibility (Eradasappa and Mohana, 2019) in cashew.

Conclusion

This is a preliminary investigation to the study of cashew tree diversity in Burkina Faso, and can be useful evidence to subsequently engage selection and breeding

efforts. Characterization of the crop genetic diversity is key to the conservation of superior lines and the development of desirable cultivars by means of genetic improvement. This study provided a picture of genetic diversity among elite lines, and revealed that there is no geographic structure in the genotypes' distribution, which is consistent with the crop introduction history. The analysis of SSR target sequences including indels and their conversion into binary data provided sufficient markers to perform a conclusive study. This marker system revealed high polymorphism among cultivars. However, the present study could be improved by the use of a higher number of markers, to increase the accuracy of genetic diversity metrics. Since cashew production is correlated with the tree's age, comparison of yield between trees of different age groups holds inherent bias, which should be addressed in future studies. Future research should identify a set of markers with high discriminatory power to be used for quality control of cultivars.

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

ACKNOWLEDGMENTS

The authors appreciate the financial support of the Matching Fund initiative and the West African cashew value chain capacity building project.

REFERENCES

Aliyaman A, Indradewa D (2019). Growth and Yield Characters of Two Cashew (Anacardium occidentale L.) Cultivars at Different Ages in Baubau City, South-East Sulawesi Province. Ilmu Pertanian (Agricultural Science) 4(1):1-8.

Aliyu OM (2012). Genetic Diversity of Nigerian Cashew Germplasm, Genetic Diversity in Plants, Mahmut Çalişkan, IntechOpen, DOI: 10.5772/32892. Available from: https://www.intechopen.com/chapters/31477(163-184).

Aliyu OM, Awopetu JA (2007). Assessment of genetic diversity in three populations of cashew (Anacardium occidentale L.) using protein-isoenzyme-electrophoretic analysis. Genetic Resources and Crop Evolution 54(7):1489-1497..

Archak S, Gaikwad A, Swamy K, Karihaloo J (2009). Genetic analysis and historical perspective of cashew (*Anacardium occidentale* L.) introduction into India. Genome 52(3):222-230.

Archak S, Gaikwad AB, Gautam D, Rao EVVB, Swamy KRM, Karihaloo JL (2003). Comparative assessment of DNA fingerprinting techniques (RAPD, ISSR and AFLP) for genetic analysis of cashew (*Anacardium occidentale* L.) accessions of India. Genome 46(3):362-369.

Arslan J, Gilani AU, Jamshed H, Khan SF, Kamal MA (2020). Edible Nuts for Memory. Current Pharmaceutical Design 26(37):4712-4720. doi: 10.2174/1381612826666200806095649. PMID: 32767923.

Audouin S (2014). Systèmes d'innovation et territoires: un jeu d'interactions. Les exemples de l'anacarde et du Jatropha dans le Sud-Ouest du Burkina Faso (Doctoral dissertation, Université Panthéon-Sorbonne-Paris I; Institut International d'Ingénierie de l'Eau et de l'Environnement).

Ben Zaied Y, Ben Cheikh N (2015). Long-Run Versus Short-Run Analysis of Climate Change Impacts on Agricultural Crops. Environmental Modeling Assessment 20(3):259-271.

Botstein D, White RL, Skolnick M, Davis RW (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics 32(3):314.

Chen W, Hou L, Zhang Z, Pang X, Li Y (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. Frontiers in Plant Science 8:575-575.

Cobb JN, Juma RU, Biswas PS, Arbelaez JD, Rutkoski J, Atlin G, Hagen T, Quinn M, Ng, EH (2019). Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation. Theoretical and Applied Genetics 132(3):627-645.

Croxford AE, Robson M, Wilkinson MJ (2006). Characterization and PCR multiplexing of polymorphic microsatellite loci in cashew (*Anacardium occidentale* L.) and their cross-species utilization. Molecular Ecology Notes 6:249-251.

de Souza RGM, Schincaglia RM, Pimentel GD, Mota JF (2017). Nuts and Human Health Outcomes: A Systematic Review. Nutrients 9(12):1311. doi: 10.3390/nu9121311. PMID: 29207471; PMCID: PMC5748761.

Doyle JJ, Doyle JL (1990). Isolation of plant DNA from fresh tissue. Focus 12:39-40.

Eradasappa E, Mohana G (2019). Investigations on self-compatibility and extent of self and cross-pollination in cashew. Journal of Plantation Crops 47(2):72-81

Food and Agriculture Organization (FAO) (2020). Food and agricultural commodities production [Online]. FAO. Available: http://faostat.fao.org/site/339/default.aspx [Accessed 15-04 2021].

Gecit J (2020). Burkina Faso, Cajou: La campagne de commercialisation 2020 de la noix de cajou ouverte au Burkina [Online]. 75017 Paris, France: Commodafrica. Available: http://www.commodafrica.com/06-03-2020-la-campagne-de-commercialisation-2020-de-la-noix-de-cajou-ouverte-au-burkina [Accessed 20/03/2021 2021].

Gower JC (1971). A General Coefficient of Similarity and Some of Its Properties. Biometrics 857-871.

Guichoux E, Lagache L, Wagner S, Chaumeil P, Léger P, Lepais O, Lepoittevin C, Malausa T, Revardel E, Salin F, Petit RJ (2011). Current trends in microsatellite genotyping. Molecular Ecology Resources 11(4): 591-611.

Gupta RK (1993). Multipurpose trees for agroforestry and wasteland utilisation. Oxford IBH Publishing Co.

He J, Zhao X, Laroche A, Lu Z-X, Liu H Li Z (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Frontiers in Plant Science 5:484. https://doi.org/10.3389/fpls.2014.00484.

Hummel M, Edelmann D, Kopp-Schneider A (2017). Clustering of samples and variables with mixed-type data. PLoS ONE 12(11):e0188274.

INSD (2020). Annuaire statistique 2019 de l'Institut national de la statistique et de la démographie Burkina Faso. Ouagadougou, Burkina Faso: INSD.

Jena RC, Samal KC, Pal A, Das BK, Chand PK (2016). Genetic Diversity Among Some Promising Indian Local Selections and Hybrids of Cashew Nut Based on Morphometric and Molecular Markers. International Journal of Fruit Science 16(1):69-93.

Kouakou CK, N'da Adopo A, Djaha AJ-B, N'da DP, N'da HA, Bi IaZ, Koffi KK, Djidji H, Minhibo MY, Dosso M (2020). Genetic characterization of promising high-yielding cashew (Anacardium occidentale L.) cultivars from Côte d'Ivoire. BASE.

Lande R, Thompson R (1990). Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124(3):743-756.

Makueti JT, Otieno G, Tchoundjeu Z, Muchugi A, Tsobeng A, Asaah E Kariba R (2015). Genetic diversity of Dacryodes edulis provenances used in controlled breeding trials. Journal of Plant Breeding and Crop Science 7(12):327-339.

Michalakis Y, Excoffier L (1996). A Generic Estimation of Population Subdivision Using Distances Between Alleles With Special Reference for Microsatellite Loci. Genetics 142(3):1061-1064.

Michelangeli M, Wong BBM, Chapple DG (2015). It's a trap: sampling bias due to animal personality is not always inevitable. Behavioral Ecology 27(1):62-67.
Monteiro F, Catarino L, Batista D, Indjai B, Duarte MC, Romeiras MM (2017). Cashew as a High Agricultural Commodity in West Africa: Insights towards Sustainable Production in Guinea-Bissau. Sustainability 9(9):1666.
Müller K, Mcdermott JM, Wolfe MS, Limpert E (1996). Analysis of diversity in populations of plant pathogens: the barley powdery mildew pathogen across Europe. European Journal of Plant Pathology 102(4):385-395.
Nei M (1973). Analysis of Gene Diversity in Subdivided Populations. Proceedings of the National Academy of Sciences 70(12):3321-3323.
Ngure D, Kinyua M, Kiplagat O (2021). Morphological and microsatellite characterization of improved Lablab purpureus genotypes. Journal of Plant Breeding and Crop Science 13(2):23-34.
Panzeri S, Magri C, Carraro L (2008). Sampling bias. Scholarpedia 3(9):(4258).
Peakall R, Smouse PE (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics 28:2537-2539.
Perez-Figueroa A (2013). msap: a tool for the statistical analysis of methylation-sensitive amplified polymorphism data. Molecular Ecology Resources 13(3):522-527.
Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Molecular Breeding 2(3):225-238.
Rico R, Bulló M, Salas-Salvadó J (2016). Nutritional composition of raw fresh cashew (*Anacardium occidentale* L.) kernels from different origin Food Science and Nutrition 4(2):329-338.8.
Ros E (2010). Health benefits of nut consumption. Nutrients *2*(7):652–682. https://doi.org/10.3390/nu2070652.
Rusu ME, Mocan A, Ferreira ICFR, Popa DS (2019). Health Benefits of Nut Consumption in Middle-Aged and Elderly Population. Antioxidants 8(8):302.https://doi.org/10.3390/antiox8080302.
Santhosh W, Shobha D, Melwyn G (2009). Assessment of genetic diversity in cashew germplasm using RAPD and ISSR markers. Scientia Horticulturae 120(3):411-417.
Savadi S, Muralidhara BM, Preethi P (2020). Advances in genomics of cashew tree: molecular tools and strategies for accelerated breeding. Tree Genetics and Genomes 16(5):1-15.
Sika K, Adoukonou-Sagbadja H, Ahoton L, Saidou A, Ahanchede A (2015). Genetic characterization of cashew (*Anacardium occidentale* L) cultivars from Benin. Journal of Horticultural 2:3-17.
Sneath PH, Sokal RR (1973). Numerical taxonomy. The principles and practice of numerical classification. 1st Edition, W. H. Freeman, San Francisco.
Sousa CM, Pereira W, Da Silveira MG, Rocha MC, Almeida VG (2019). Genetic dissimilarity and growth of coffee in Cerrado. Journal of Plant Breeding and Crop Science 11(9):213-224
Sunita K, Kumar R, Chaudhary R, Chaturvedi R (2021). Morpho-agronomic study of some varieties of Golden Sweet Potato for their utilization as food and forage. Tropical Plant Research 8(1):22-30.
Syed NH, Sureshsundar S, Wilkinson MJ, Bhau BS, Cavalcanti JJV, Flavell AJ (2005). Ty1-copia retrotransposon-based SSAP marker development in cashew (*Anacardium occidentale* L.). Theoretical and Applied Genetics 110(7):1195-1202.
Tarpaga WV, Bourgou L, Guira M, Rouamba A (2020). Agro morphological characterization of cashew trees (*Anacardium occidental* L.), in improvement for the high yield and high quality of raw nuts in Burkina Faso. International Journal of Biological and Chemical Sciences 14(9):3188-3199
Vanitha K, Raviprasad TN (2019). Diversity, species richness and foraging behaviour of pollinators in cashew. Agricultural Research 8(2):197-206.
Wonni I, Sereme D, Ouedraogo I, Kassankagno A, Dao I, Ouedraogo L, Nacro S (2017). Diseases of cashew nut plants (*Anacardium Occidentale* L.) in Burkina Faso. Advances in Plants and Agriculture Research 6(3):6.
Yaman M, Uzun A (2020). Evaluation of Superior Hybrid Individuals with Intra and Interspecific Hybridization Breeding in Apricot. International Journal of Fruit Science, 20(sup3):S2045-S2055.
Yaman M, Uzun A (2021). Effects of Different Pollinators on Fruit Set and Quality Attributes of Texas Almond (*Prunus dulcis* L.) Cultivar . Horticultural Studies 38(1):46-49 . DOI: 10.16882/hortis.886517.
Yildiz E, Pinar H, Uzun A, Yaman M, Sumbul A, Ercisli S (2021). Identification of genetic diversity among *Juglans regia* L. genotypes using molecular, morphological, and fatty acid data. Genetic Resources and Crop Evolution 68(4):1425-1437.