Pak. J. Bot., 55(3): 1013-1021, 2023. DOI: http://dx.doi.org/10.30848/PJB2023-3(19)

MOLECULAR CHARACTERIZATION OF DOUM PALM *(HYPHAENE COMPRESSA)* FROM SELECTED REGIONS OF KENYA USING CHLOROPLAST SIMPLE SEQUENCE REPEATS (cpSSR) MARKERS

ALLEN JOHNNY BORLAY[1,5*], CECILIA MBITHE MWEU[2], STEVEN GER NYANJOM[3], KEVIN MBOGO OMOLO[3] AND AGNES OMIRE[4]

[1]*Department of Molecular Biology and Biotechnology, Pan African University Institute for Basic Sciences, Technology and Innovation, P. O. Box 62000-00200, Nairobi, Kenya*
[2]*Institute for Biotechnology Research, Jomo Kenyatta University of Agriculture and Technology, P. O. Box 62000-00200, Nairobi, Kenya*
[3]*Department of Biochemistry, Jomo Kenyatta University of Agriculture and Technology, P. O. Box 62000-00200, Nairobi, Kenya*
[4]*Department of Botany, Jomo Kenyatta University of Agriculture and Technology, P. O. Box 62000-00200, Nairobi, Kenya*
[5]*Department of Biological Sciences, University of Liberia, P. O. Box 10-9020, Monrovia, Liberia*
[*]*Corresponding author: allenborlay@gmail.com*

Abstract

Doum palm (*Hyphaene compressa*) is an enduring member of the Arecaceae family of plants, an essential multipurpose plant with exceptional features that serve as a socioeconomic resource for people in the East Africa region. The unavailability of genetic and genomic data leads to difficulties enhancing such a crop, thereby preventing the actualization of its agronomic and breeding potential, hence the need for characterization. This study included 64 genotypes of doum palm collected from four locations in Kenya and ten polymorphic chloroplast microsatellite markers. Locus Among the ten polymorphic cpSSRs studied, the OPCG13 locus exhibited the highest observed and expected heterozygosity. Across all loci, we detected the mean value of 10.145 for the gene flow parameter. The mean number of significant allele frequencies for the 79 alleles detected was 0.75, with a range of 0.531 to 0.89. The median was 0.0036, 0.341, and 0.275 for observed heterozygosity, predicted heterozygosity, and PIC, respectively. All 64 genotypes were sorted into three main categories using neighbor-joining clustering and STRUCTURE analysis. At the population level, PCoA analysis explained 51.66 percent of the variation. As a result of these findings, cpSSR markers were used for the first time to characterize selected doum palm germplasm, which signifies that such a marker helps study doum palms at the molecular level. Our findings indicate that doum palm species in Kenya have a reasonable genetic variation coupled with considerable heterozygosity; consequently, these doum palm genetic variations are essential for the genetic improvement, breeding, and conservation initiatives of doum palm genotypes in Kenya.

Key words: Doum palm (*Hyphaene compressa*), Rural pastoralists, Arid and semi-arid Kenyans, Chloroplast SSR (cpSSR), Genetic diversity, Molecular characterization.

Introduction

Doum palm is one of the well-known species of the Hyphaene, which is endemic to Africa, the Middle East, and Asia (El-Beltagi *et al.*, 2018) *Hyphaene* comprises multiple species, including *Hyphaene compressa,* a vital socioeconomic resource for rural pastoralists and small-scale farmers in Eastern Africa (Omire *et al.*, 2020). Arecaceae (palm) is a family of plants that includes the Doum palm (Lee & Balick, 2008; Balslev *et al.*, 2016). It is among the most notable and financially essential plants globally (Balslev *et al.*, 2016). Besides serving rural communities, palms have worldwide economic value as products of biological origin other than timbers in the Western part of Africa (Stauffer *et al.*, 2018). Multiple studies prove that it offers various applications, including non-timber building materials, food, medicine, and weaving items(Sukanya *et al.*, 2009). It flourishes in hot, dry environments, and each part has a distinct feature. People living in Kenya's arid and semi-arid regions rely on the *Hyphaene compressa* for economic empowerment and sustenance (Amwatta, 2004).

The doum palm is a poorly understood genus with insufficient information on its semi-domestication or domestication (Omire *et al.,* 2020). There is no data on genetic resources, despite their economic importance, contributions to the palm family's diversity in Africa, and status as a multifunctional crop. Genetic resources are necessary for varietal development, and their diversity levels determine the utility of genetic resources. Additionally, there is only a little research on the phenotypic variety, applications, and geographic location of the *Hyphaene compressa* (Amwatta, 2004; Omire *et al.*, 2020). The doum palm's essential quality, use, and improvement have limitations due to a lack of information at the molecular level, and the genome is yet to be sequenced.

Molecular techniques help characterize germplasm, discover cultivars, investigate phylogeny, and analyze diversity in flora (Yang *et al.*, 2015). In species with ambiguous/unaccounted for early breeding records and crossing unrelated individuals, where the controlled crossover is not always viable, molecular approaches become the necessary technique. One such approach is the molecular markers technique (Mayes *et al.*, 2008). Defining the genetic diversity of plant species at a more substantial level has resulted in using simple sequence repeats (SSRs) or Microsatellite markers worldwide (Vieira *et al.*, 2016). However, the stages leading to the production of functional SSR markers, such as initial

microsatellite locus identification, primer selection, and amplification/polymorphism detection assessment, necessitate the availability of molecular data; for example, the genome of the species of interest is costly and physically demanding (Squirrell *et al.*, 2003).

Chloroplast simple sequence repeat (cpSSR) markers are complementary to their nuclear counterpart due to their small size, relatively conserved gene order, low mutation rate, global accessibility of universal primers, absence of heteroplasmy, and absence of genetic material exchange between two different molecules (Diekmann *et al.*, 2012). Conservation objectives for threatened species can significantly assist through measures of genetic variation and differentiation derived from these markers (Petit *et al.*, 2008). Given the importance of the doum palm, attempts to improve genotypes are required. Therefore, its germplasm should undergo molecular characterization, allowing breeders to uncover genes of interest, link them to agronomic features, and provide genetic variety. Genetic diversity knowledge helps use germplasm efficiently, identify and eliminate duplicates, and establish nucleus collections (Ghafoor *et al.*, 2001). Genetic diversity is a vital part of whether cultivated species should progress genetically.

There is no research on the molecular characterization of doum palm species, genotypes, or genetic diversity that we know. Understanding the genetic diversity and linkages among doum palm germplasm in local and regional breeding programs is critical for researchers. This study aims to use cpSSR markers to molecularly characterize and generate the first set of genetic data on the doum palm (*Hyphaene compressa*) to understand its genetic diversity, which will lead to the establishment of breeding and conservation programs for the doum palm.

Material and Methods

Plant materials: This study comprises sixty-four samples from four counties in Kenya: 16 from Tharaka Nithi, Tana River, Kwale, and Turkana counties. With the assistance of local farmers/pastoralists, young leaves were harvested from these plants using the GPS gadget. The leaves were collected and placed in falcon tubes filled with silicon gel before being brought to the lab for DNA.

SSR locus search and primer design: We screened the cpSSR loci against the complete chloroplast genome sequence of Elaeis guineensis (Accession # NC_017602) using the web-based software BatchPrimer3 v1.0 with its default setting of repeat thresholds in constructing our polymerase chain reaction (PCR) primer pairs from the sequences flanking the SSR loci. We then generated the primer sets using the following criteria: PCR product sizes ranged from 100 to 300 base pairs (bp), with an annealing temperature of 50 to 65 degrees Celsius and a percentage between 40 to 60 for GC content (Table 1).

Doum palm genomic DNA Isolation, qualification, and quantification: The plant DNA kit BIOLINE ISOLATE II (meridian BIOSCIENCE, Canada) was utilized to extract genomic DNA from leaves of the doum palm collected from several locations in Kenya using the manufacturer's guidelines. Spectrophotometers were used to quantify the DNA purity, while a 1.5 percent agarose gel with a 1kb DNA ladder was used to evaluate the DNA quality.

Chloroplast SSR amplification/genotyping: 2-µL of template DNA, 0.5 µL of My Taq DNA polymerase (meridian BIOSCIENCE, Canada), 6µL of My Taq reaction buffer, 9.5 µL of deionized water, and 1µL of forward and reverse primers (20 µM) were used in a 20 µL reaction mixture the amplification of the cpSSR. The PCR cycling conditions were as follows: initial denaturation began at 95^{o}C for 3 minutes, followed by 15 seconds at 95^{o}C of denaturation, and 30 seconds for melting temperature varies depending on the primer sets indicated in Table 2. The initial amplification extension lasted for 30 seconds at 72^{o}C, and the final extension lasted for 3 minutes at 72^{o}C. The obtained PCR products were validated using 1.5 percent agarose electrophoresis.

Table 1. cpSSR primers utilized in this research.

No.	Primer ID	Primer sequence	Length	Tm	Size	Motif	SSR length
1.	OPCG10 F	ACTCACTATTCATCGAACCAT	21	54	165	ATGTA	20
	OPCG10 R	GTCGAAAATGCTCTTCATTC	20				
2.	OPCG7 F	AATACAAATCAAACCCGATT	20	55	150	AT	14
	OPCG7 R	TCCATGGTTTATTCCTTATCTC	22				
3.	OPCG9 F	TATTAACCCGAGGCTCTTTA	20	60	163	TTAT	16
	OPCG9 R	AAGGCAGTGTGATAAAGCAT	20				
4.	OPCG11 F	GCCAACAGTCAAGTTTTTAGT	21	60	142	TATTT	15
	OPCG11 R	CTCGGATCGACTTCAATATC	20				
5.	OPCG12 F	AATAAGAAGAAGCGGAGAGAG	21	58	159	TCCCCA	18
	OPCG12 R	TTACTTCACGTGTCACATCAC	21				
6.	OPCG13 F	AAGAAATAGATGGAACGGAAC	21	62	145	AGGAAG	18
	OPCG13 R	CGGGTTTGATTTGTATTGTAT	21				
7.	OPCG1 F	GTGATGTGACACGTGAAGTAA	21	53	241	AT	14
	OPCG1 R	TATCCGTTTCGAATCATTATC	21				
8.	OPCG3 F	CGCAATCAAATCAATCAATA	20	52	136	AT	16
	OPCG3 R	GTGGTTTACGTTATGGAAGAA	21				
9.	OPCG6 F	AAAGGACGGCCCTATATAAT	20	53	155	AT	14
	OPCG6 R	ATTCTTCCTTCAACTTCGATT	21				
10.	OPCG4 F	TAAAGAATTCCTTGGTTTCG	20	57	219	AT	16
	OPCG4 R	CATCTACTTGTATGGGTTTCG	21				

Table 2. Diversity analysis of sixty-four doum palm genotypes using ten cpSSR markers.

Locus	MAF	NG	Na	Ne	Ho	He	I	Nei	Nm	PIC
OPCG10	0.891	64	7	4.998	0	0.196	0.345	0.305	5	0.176
OPCG7	0.891	64	8	4.998	0	0.196	0.345	0.375	3.444	0.176
OPCG9	0.671	64	8	7.118	0.002	0.444	0.632	0.469	20.273	0.344
OPCG11	0.531	64	8	7.13	0.02	0.502	0.691	0.492	1.571	0.374
OPCG12	0.781	64	8	6.077	0	0.344	0.525	0.429	3.727	0.283
OPCG13	0.593	64	8	7.507	0.011	0.486	0.676	0.429	6.611	0.366
OPCG1	0.891	64	8	4.984	0	0.196	0.345	0.305	8.818	0.176
OPCG3	0.843	64	8	5.466	0	0.266	0.433	0.375	5.5	0.231
OPCG6	0.781	64	8	6.076	0.001	0.345	0.525	0.375	43.5	0.283
OPCG4	0.672	64	8	6.953	0.002	0.444	0.633	0.429	5	0.344
Mean	**0.7545**	**64**	**7.9**	**6.131**	**0.0036**	**0.3419**	**0.5152**	**0.3984**	**10.145**	**0.275**
Sum		**640**	**79**							

MAF: major allele frequency; NG: number of genotypes; Na: number of observed alleles; Ne: effective number of alleles; Ho: observed heterozygosity; He: expected heterozygosity; I: Shannon information index; Nei: Nei's genetic identity; Nm: gene flow and PIC: polymorphism information content

Statistical analysis: Bands were rated 0 for absence and 1 for presence to perform genotypic analysis. Genetic data and statistical analysis were analyzed using multiple software combinations. GenAlEx version 6.502 assisted in calculating the number of observed alleles (Na), the effective number of alleles (Ne), the number of private alleles (Np), observed heterozygosity (Ho), expected heterozygosity (He), and Shannon information index (I) (Banks & Peakall, 2012). Power Marker software version 3.25 calculated the major allele frequency (MAF), the number of genotypes (NG) per locus, and the polymorphic information content (PIC) of each locus. We utilized the genotype distance matrix to perform the allelic frequency and the principal coordinate analysis (PCoA). The total of square and variance components within and between populations were discovered using GenAlEx software and molecular variance analysis (AMOVA). Using the POPGENE program version 1.32, we calculated Nei's genetic identity, gene flow (Nm), and genetic distances between populations.

Examining the links between populations and all individuals, a cluster dendrogram based on Nei's genetic distance was constructed in MEGA7 version 10.26 using an unweighted pair group method with arithmetic mean (UPGMA)-based cluster analysis (Kumar *et al.*, 2016). We used (Pritchard *et al.*, 2000) STRUCTURE program version 2.3.4 to investigate the population's genetic structure using Bayesian clustering to determine how many clusters exist in the dataset (K1 to K10). We investigated mixing ancestry and associated allele frequency using the Markov chain Monte Carlo simulation technique (MCMC) with a burn-in period of 10,000. For each K of collected doum palm genotypes, MCMC was set to 10,000 runs and done ten times to approximate K scores. Structure Harvester's online website determined the best K rate at each of the four K categories (Evanno *et al.*, 2005). An online software program called CLUMPP (Jakobsson & Rosenberg, 2007) was used to associate and show groups across duplicates, and display sets, with each K represented by color-coded, scatter plots.

Results

Allelic diversity of the cpSSR marker: Seventy-nine alleles were detected at the ten cpSSR loci across 64 individuals from 4 populations (Tana River, Tharaka, Turkana, and Kwale), with 7.9 accounting for the total observed alleles per locus. Alleles displayed a wide range of allelic variations (Table 2). The mean major allele frequency (MAF) was 0.754, with values ranging from 0.531 (OPCG11) to 0.891 (OPCG10, OPCG7, and OPCG1). The average number of effective alleles was 6.1, with a range of 4.984 (locus OPCG1) to 7.507 (for locus OPCG2 and OPCG13. For the first allele, Tharaka genotypes had the highest allele frequency rate of 1.00 (locus OPCG10) in the population and the lowest rate of 0.313 for locus OPCG11 (Fig. 1). Also, in Figure 1, Tharaka genotypes had the highest allele frequency of 0.688 (locus OPCG11) and the lowest (locus OPCG11).

Analysis of genetic variation and population structure differentiation: There was a lot of genetic variation among the 64 genotypes of doum palm accessions sampled in our current research. Table 2 and Figure 2 illustrate the genetic diversity parameters. The highest expected heterozygosity (He) value was 0.502 (OPCG11), the lowest was 0.196 (OPCG10 and OPCG7), and the average value was 0.341 for all loci, with 0.0036 as the observed heterozygosity (Ho). Most of the cpSSR markers in this investigation were significantly informative on average, with a mean PIC value of 0.275, with locus OPCG11 having the highest PIC (0.374) and locus OPCG10, OPCG7, and OPCG1 having the lowest of 0.176.

Furthermore, the doum palm populations collected from Kwale and Tharaka had high and low values of 0.398 and 0.289 heterozygosity. In addition, the highest mean numbers of alleles in doum palm populations were about 2.00 (Kwale, Tana River, and Turkana), with Tharaka having a slightly lower value of 1.90. (Fig. 2). Detection of locally common and private alleles failed in our analysis of the four populations (Fig. 2). Fixation indices were considered a measure of population differentiation for each locus in the doum palm population.

Only 1% of overall variation occurred between populations, according to AMOVA's genetic differentiation (Table 3), while 99% occurred within populations. The genetic differentiation coefficient (FST) between the populations was 0.195, and all results were highly significant (P 0.001) (Table 3). Using Nei's genetic identity and distance, the degree of genetic differentiation among the four populations of doum palms was examined (Table 3). Tables 2,3, and 4 give the AMOVA, Fst, and Nm values.

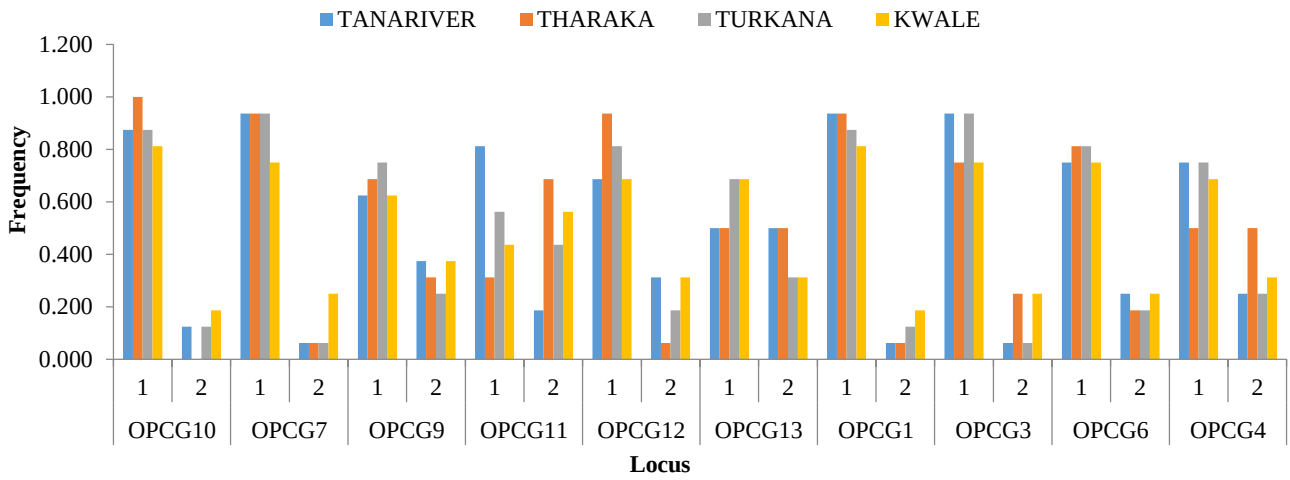

Fig. 1. Frequency of allele by locus for the four regions/populations.

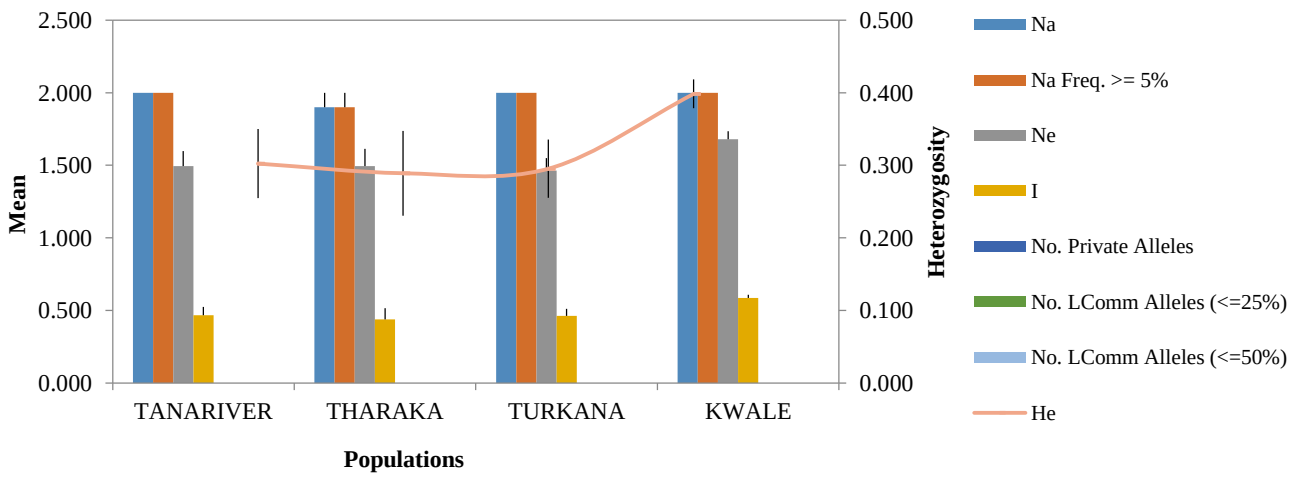

Fig. 2. Patterns of Allele of the doum palm genotypes.

Table 3. Summary of AMOVA.

Source of variation	df	SS	MS	Est. Var.	%	P-Value
Among populations	3	11.594	3.865	0.014	1%	0.001
Within populations	60	205.63	3.427	1.714	99%	0.001
Total	63	217.22		1.727	100%	
Fixation index (FST)	**0.195**					

Note: df: Degrees of freedom; SS: Sum of squares; MS: Means of a square; Est. Var: Variance component estimate; %: Percentage of total variation

Clustering and genetic structure of populations: The structural similarities amongst and within the doum accessions and their genetic differences were detected using a phylogenic tree approach and PCoA. All genotypes were categorized into three major clusters (Fig. 3). Six genotypes (9.375%) were considered as members in cluster 1, fifty genotypes (78.125%) in cluster 2 accounted were accessions, and eight genotypes (12.5%) in cluster 3. Cluster 2 was subdivided further into three subclusters. All subclusters in cluster 2 were further divided into many subgroups, and each constituted a clade of mixed genotype populations.

The dendrogram revealed that at least one genotypic representation from each of the four counties/populations exists in each cluster. In this study, Pritchard *et al*., (2000) STRUCTURE software version 2.3.4 investigated the population structure and genetic relationships among the 64 genotypes of doum palms from four Kenyan counties. Agarose gel information facilitated the number of genotyping clusters using the K-value. A strong peak of K at 3 indicated the perfect K value, which states the number of clusters in the studied population (Fig. 5).

The optimal K-value indicates that the three subpopulations showed the highest probability of population clustering, and these three subpopulations consisted of 64 genotypes. It was, therefore, necessary to do a population genetics analysis on the result at K = 3. Blue, red, and green bar plots were displayed in the online output (Fig. 6), indicating an intermixture pattern within each doum palm population.

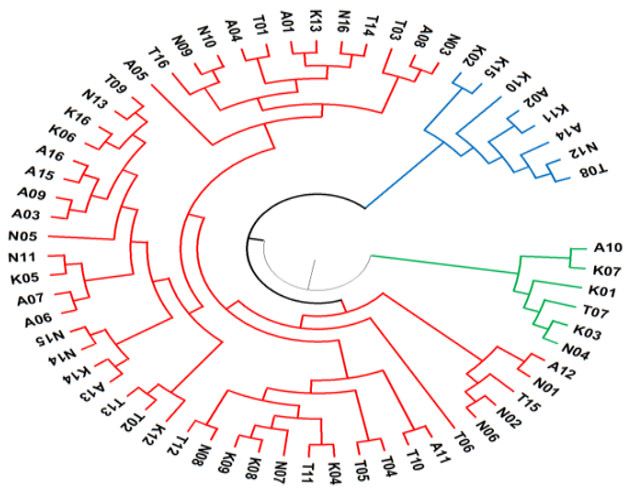

Fig. 3. Dendrogram represents the genetic relationship of the 64 doum palms.

Table 4. The STRUCTURE results of 64 doum palm genotypes for the fixation index (Fst) and average distance (expected heterozygosity/Exp.Het).

Population	Inferred clusters	Mean Fst	Exp. Het
Population 1	0.365	0.5939	0.1638
Population 2	0.358	0.0243	0.4565
Population 3	0.277	0.5977	0.1659

There was substantial divergence within each of the three populations, as measured by the fixation index (Fst), according to these findings (Table 4). Populations 1, 2, and 3 had Fst values of 0.589, 0.569, and 0.021, respectively (Table 3). The first three axes explained 51.66 percent of the variation observed in the principal coordinates analysis (PCoA) of all 64 genotypes (Fig. 4). The genotype distributions on the PCoA graph were not categorized geographically; rather, they demonstrated the mixing of samples among populations, as indicated by the NJ tree. We found that clustering (Fig. 3) and genotype distributions were consistent with genetic structure findings (Fig. 4).

Discussion

Plant genetic resources and levels of genetic diversity in tree species have significantly benefited from using molecular marker techniques (Porth & El-Kassaby, 2014). Among the several types of molecular marker approaches available for identifying and characterization of genetic variation, simple sequence repeat (SSRs) markers have extensively been utilized to identify the genetic diversity of plant species worldwide (Vieira *et al.*, 2016). However, developing SSR markers is still inefficient, time-consuming, and expensive, especially in organisms with large genomes (Squirrell *et al.*, 2003), and it is even more difficult in plants with no genome sequence, such as the doum palm. This work assessed the genetic characteristics of 64 doum palm populations using trans-specific amplification of chloroplast microsatellite markers from the oil palm chloroplast genome. The results showed that cpSSR primer pairs from other *Arecaceae* family species might be successfully transferred to the doum palm and used for the genetic investigation of this species. No data on nuclear SSR markers is yet to be generated for the doum palm species.

Allelic diversity of the cpSSR marker: Detection of 79 alleles explains significant genetic variation among doum palm populations from 64 individual samples. The number of alleles per locus was relatively constant in our investigation, with a mean value of 7.9. The allelic pattern and genetic diversity indices revealed genetic variety within the study's subpopulations. Population one and three had similarly expected heterozygosity, but population 3 was slightly higher than population 1. Population 2 had the highest expected heterozygosity, indicating that population 2 was more diverse than populations 1 and 3; the expected heterozygosity is affected by the number of alleles (richness) and their abundance (or evenness) in a population. Our findings on the allelic pattern at the population level are comparable to the results found on Camelina sativa Spring Panel by Luo *et al.*, (2019). Additionally, (Mir *et al.*, 2021) found an average mean allelic variety comparable to the one obtained in the current study in genetic diversity and population structure investigation on Himalayan bean germplasm using SSR markers.

Genetic diversity and population differentiation analysis: PIC values and expected heterozygosity (He) are genetic diversity among genotypes in a population, highlighting the evolutionary pressure on alleles and the mutation rate a locus might have undergone through time (Botstein *et al.*, 1980; Shete *et al.*, 2000). When used with haploid markers, "He" reflects genetic diversity and estimates the average heterozygosity and genetic distance between individuals in a group (Shete *et al.*, 2000; Salem & Sallam, 2016). According to (Nei, 1972) and (Shete *et al.*, 2000) good PIC values indicate the usefulness of markers for gene-hunting and genetic testing to detect the inheritance among genotypes. In the current investigations, the overall genetic diversity value was more significant than the PIC value (Fig. 3), as expected, because PIC values are always lower than genetic diversity and will increase closer to genetic diversity if more alleles are present. Allele frequencies become more even, i.e., individuals with identical heterozygote genotypes have a lower chance of surviving (Shete *et al.*, 2000).

Findings from this analysis revealed that the majority of the cpSSR markers had PIC values less than 0.5, with an average PIC of 0.27, implying that the majority of the cpSSR markers were moderately informative based on the (Botstein *et al.*, 1980) principle. The diversity across the Arecaceae family genomes, which primarily represent changes in the number of repeat motifs in the SSR area paired with indels and base substitutions, could explain the differences identified among the ten cpSSR in our analysis (Zaki *et al.*, 2012). This study's polymorphic information content (PIC) is higher than that reported by (Khalil *et al.*, 2020) in the assessment of molecular diversity in doum (*Hyphaene thebaica* L.) using the ISSR marker.

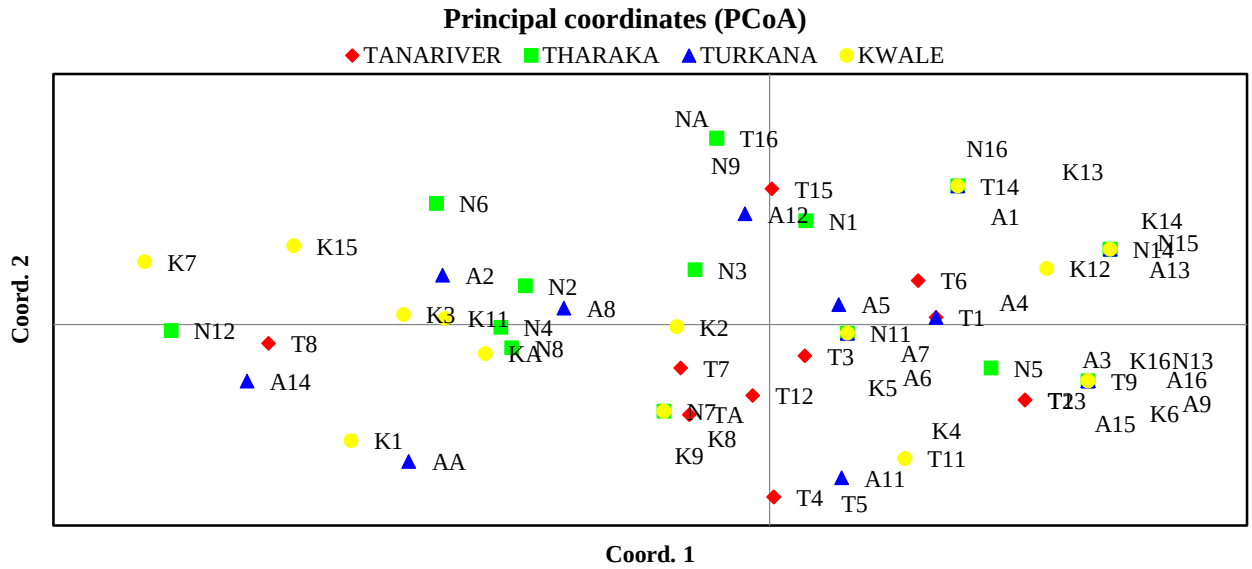

Fig. 4. PCoA results of 64 doum palm samples.

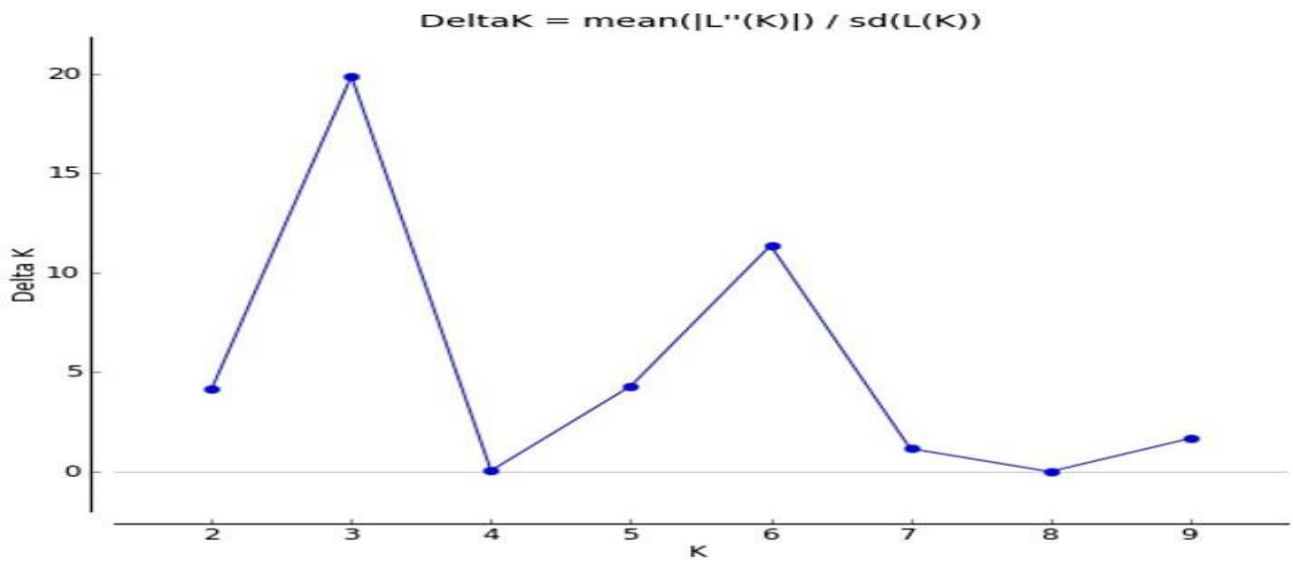

Fig. 5. Population estimation utilizing a cluster number (K) ranging from 1 to 10, K = mean (|L"(K)|) / sd(L(K)), and K = 3 being the excellent K value.

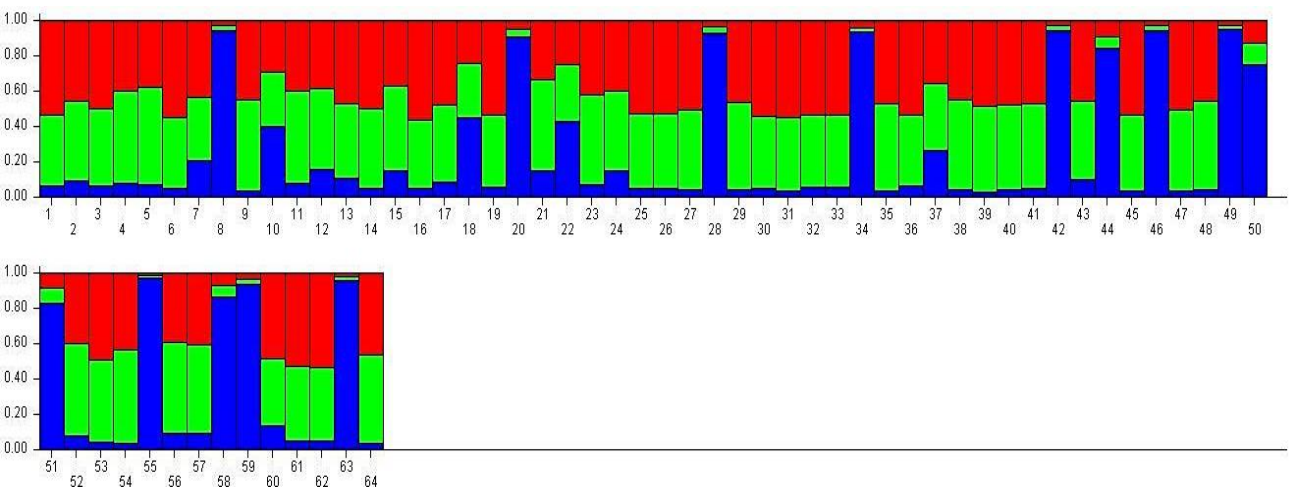

Fig. 6. STRUCTURE evaluation result based on cpSSR information and approximation of the genetic structure of the 64 doum palm. Vertical bars account for individuals, while the colors are assigned based on the structure software grouping (K = 3).

The level of polymorphism determined by the PIC value (mean= 0.275) in this investigation corroborated with the reported mean PIC value in previous work by (Ochieng *et al.*, 2015). A better understanding of the genetic variation among doum palm populations will help breeders plan and keep track of the genetic diversity needed for successful cultivation initiatives (Eltaher *et al.*, 2018). Fst is a genomic structure-based measure of population differentiation. Population differentiation may be substantial if the Fst value is more than 0.15 (Frankham *et al.*, 2002). Each of the three subpopulations of the doum palm indicated a considerable divergence in Fst values, as shown in Table 3. The AMOVA findings showed that within-population variation accounted for 99% of the total variation, whereas among-population variation accounted for only 1%. According to Wright, (1965) Nm values of less than one suggest little gene interchange between subpopulations. In contrast, our analysis's Nm score of 10.145 was high, which aligns with (Eltaher *et al.*, 2018), finding that many genetic interchanges or gene flow occurred, resulting in low genetic differences amongst subpopulations.

However, at the highest K level (K=3), the results of the phylogenic tree, PCoA, and structure investigation in this research are overshadowed by alloy structural forms among populations. As expected, numerous clustered populations seem helpful in capturing the actual population structure, as shown in the Baysian application results. This phenomenon, as suggested by (Meirmans, 2015), is because the germplasm in some geographical locations is diverse, and a pattern in the *genetic makeup is already* in existence. Even if a species spreads across many territories, it may be organized into ecotypes or host races because of gene migration or similar lineage (Meirmans, 2015). Hierarchically structuring populations at various levels are complicated and sometimes unnecessary.

Clustering and genetic structure of populations: According to the work of (Eltaher *et al*., 2018) an informative population structure assessment can help researchers better comprehend the genetic variation and identify specific functional genetic variants (loci, alleles) linked to a trait investigation more efficiently. This study established a clear genetic relationship between doum palm genotypes in NJ (neighbor-joining tree) and the results of the principal coordinate analysis. The dendrogram produced three significant clusters, with Cluster 1 (red branch) containing more than 90% of the accessions or individuals. This study discovered no visible difference between genotypes based on geographical location. Three individuals from Kwale and one from Tharaka, Turkana, and Tana River Counties were in cluster 3 (green color branch). Cluster 2 (blue color branch) had the most individuals, with mixed populations in multiple subclusters. Genotypes from the Kwale population were mainly observed in large numbers in the three clusters. NJ, PCoA, and population studies showed that genotype allocations on the principal component lines and mix structures are common in combining genotypes.

At K = 3, STRUCTURE Harvest generated the highest mean similarity score, indicating the most likely population clustering and admixture structure. Doum palms may share a genetic history and alleles, and structural data suggest they have a common ancestor. The current PCoA results matched previous findings on the date palm (Zehdi-Azouzi *et al.*, 2015; Chaluvadi *et al.*, 2019). Using a model-based clustering technique, they found three genetically distinct subpopulations not established by the date palm locations. The current study's intermixture score of K = 3 agrees with research published by (Aljuhani, 2016), and (Moussouni *et al.*, 2017) also used date palms in their investigations. Similarly, (He *et al*., 2017) discovered three distinct date palm groups at K = 3 using various SSRs and four separate categories at K = 4, while Single nucleotide polymorphism makers.

When genotypes were sorted into inferred clusters (pop1: 0.365, pop2: 0.358, and pop3: 0.277), the exact threshold can be assumed for this combination (pop1: 0.365, pop2: 0.358, and pop3: 0.277). Cluster one contained most of the genotypes. Smaller sample sizes and genetic interchange between regions may be to blame, as well as a lack of power in this study. Intermixing results from the structural analysis of doum palm accessions in the present study are consistent with findings on date palm genetic diversity at three different K groups explored by employing distinct SSR markers (Chaluvadi *et al*., 2019; Flowers *et al.*, 2019).

Results from this study's phylogenic tree, PCoA, and structure analysis show that mixing the structural arrangement of the population is the essential factor. There is a possibility that this discovery is related to the codominant character of microsatellite markers, which contributes significantly to allele variability per locus. Furthermore, the significant heterozygosity within the doum palm population reported by the AMOVA findings could be attributed to DNA slippage during the replication process. These phenomena are a modification in microsatellite (SSR) size caused by the addition and omission of repetitions in DNA molecules during enzymatic replication of said molecules (Hosseinzadeh-Colagar *et al.*, 2016).

Other Arecaceae species, including the Doum Palm (Arecaceae), are dioecious plants that may be cross, and pollen-borne. Due to the species' uniqueness, Doum palm samples have more genetic variants. Cross-pollinating species' breeding technique allows for considerable genetic variation among plants (Rao & Hodgkin, 2002; Charlesworth, 2006). The study's doum palm collection locations lacked a detailed crop management plan. As a result, this condition may make it more likely for doum palms to reproduce through seeds rather than offshoots. Because doum palm seeds germinate dioeciously, we expected genetic divergence to rise.

Conclusion

This work is the first to provide the molecular characterization and genetic diversity of Hyphaene compressa. We have made it available to the scientific community to facilitate increasing studies on this palm

species. Therefore, it is worth noting that with a PIC value of 0.275, the doum palms we evaluated are considerably genetically diverse. Genetic clustering was not linked to geographic origin, meaning that individuals shared a common genetic background due to the exchange of germplasm. Using genome-wide association studies, efforts to enhance the doum palm breeding and conservation initiatives predicated on the future discovery of alleles/ genes will utilize these findings. The moderate genetic diversity of doum palms and the low genetic differentiation among populations suggest that the existing natural populations in Kenya and East Africa need to be maintained to protect the genetic diversity of this species. Given the importance of the doum palm (Hyphaene compressa), a large-scale study is required involving the collection of accessions from across Kenya and adjacent countries to develop a sound evaluation system and examine genetic diversity.

Acknowledgments

The authors are grateful to Dr. Cecilia Mbithe Mweu, Ms. Agnes Omire, and Dr. Johnstone Omukhulu Neondo for assisting in collecting doum palm samples. The African Union Commission, through the Pan African University Institute for Basic Sciences, Technology, and Innovation, and the University of Liberia deserve credit for funding this study.

References

Aljuhani, W.S. 2016. Genetic Diversity and the impact of geographical location on the relationships between *Phoenix dactylifera* L. *Germplasms Grown in Saudi Arabia. Hered. Genet.,* 5: 1041-2161.

Amwatta, C.J.M. 2004. Diversity of use of Doum palm (*Hyphaene compressa*) leaves in Kenya. *Palms,* 48: 184-190.

Balslev, H., R. Bernal and M.F. Fay. 2016. Palms–emblems of tropical forests. *Bot. J. of the Linn. Soc.,* 182: 195-200.

Banks, S.C. and R.O.D. Peakall. 2012. Genetic spatial autocorrelation can readily detect sex-biased dispersal. *Mol. Ecol*., 21: 2092-2105.

Botstein, D., R.L. White, M. Skolnick and R.W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Amer. J. Hum. Genet.,* 32: 314.

Chaluvadi, S.R., P. Young, K. Thompson, B.A. Bahri, B. Gajera, S. Narayanan, R. Krueger and J.L. Bennetzen. 2019. *Phoenix phylogeny*, and analysis of genetic variation in a diverse collection of date palm (*Phoenix dactylifera*) and related species. *Plant Divers,* 41: 330-339.

Charlesworth, D. 2006. Evolution of plant breeding systems. *Curr. Biol.,* 16: 726-735.

Diekmann, K., T.R. Hodkinson and S. Barth. 2012. New chloroplast microsatellite markers suitable for assessing genetic diversity of *Lolium perenne* and other related grass species. *Ann. Bot.,* 110: 1327-1339.

El-Beltagi, H.S., H.I. Mohamed, H.N. Yousef and E.M. Fawzi. 2018. Biological activities of the Doum palm (*Hyphaene thebaica* L.) extract and its bioactive components. *Antiox. in Foods and Its Appl.,* InTech.

Eltaher, S., A. Sallam, V. Belamkar, H.A. Emara, A.A. Nower, K.F.M. Salem, J. Poland and P.S. Baenziger. 2018. Genetic diversity and population structure of F3: 6 Nebraska winter wheat genotypes using genotyping-by-sequencing. *Front. Genet.,* 9: 76.

Evanno, G., S. Regnaut and J. Goudet. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.,* 14: 2611-2620.

Flowers, J.M., K.M. Hazzouri, M. Gros-Balthazard, Z. Mo, K. Koutroumpa, A. Perrakis, S. Ferrand, H.S.M. Khierallah, D.Q. Fuller and F. Aberlenc. 2019. Cross-species hybridization and the origin of North African date palms. *Proceed. of the Nat. Acad. of Sci.,* 116: 1651-1658.

Frankham, R., S.E.J.D. Ballou, D.A. Briscoe and J.D. Ballou. 2002. *Introduction to conservation genetics*. Cambridge university press.

Ghafoor, A., A. Sharif, Z. Ahmad, M.A. Zahid and M.A. Rabbani. 2001. Genetic diversity in blackgram (*Vigna mungo* L. Hepper). *Field Crops Res.,* 69: 183-190.

Hosseinzadeh-Colagar, A., M.J. Haghighatnia, Z. Amiri, M. Mohadjerani and M. Tafrihi. 2016. Microsatellite (SSR) amplification by PCR usually led to polymorphic bands: Evidence which shows replication slippage occurs in extend or nascent DNA strands. *Mol. Biol. Res. Comm.,* 5: 167.

Jakobsson, M. and N.A. Rosenberg. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23: 1801-1806.

Khalil, O.A., R.A. Ibrahim and M. Youssef. 2020. A comparative assessment of phenotypic and molecular diversity in Doum (*Hyphaene thebaica* L.). *Mol. Biol. Reports,* 47: 275-284.

Kumar, S., G. Stecher and K. Tamura. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.,* 33: 1870-1874.

Lee, R. and M.J. Balick. 2008. Palms, people, and health. *Explore,* 4: 59-62.

Luo, Z., J. Brock, J.M. Dyer, T. Kutchan, D. Schachtman, M. Augustin, Y. Ge, N. Fahlgren and H. Abdel-Haleem. 2019. Genetic diversity and population structure of a Camelina sativa spring panel. *Front. Plant Sci.,* 10: 184.

Mayes, S., F. Hafeez, Z. Price, D. MacDonald, N. Billotte and J. Roberts. 2008. Molecular research in oil palm, the key oil crop for the future. *Genom. Trop. Crop Plants,* 371-404.

Meirmans, P.G. 2015. Seven common mistakes in population genetics and how to avoid them. Wiley Online Library.

Mir, R.R., N. Choudhary, V. Bawa, S. Jan, B. Singh, M.A. Bhat, R. Paliwal, A. Kumar, A. Chitikineni and M. Thudi. 2021. Allelic diversity, structural analysis, and Genome-Wide Association Study (GWAS) for yield and related traits using unexplored common bean (*Phaseolus vulgaris* L.) germplasm from Western Himalayas. *Front. Genet.,* 1797.

Moussouni, S., J.C. Pintaud, Y. Vigouroux and N. Bouguedoura. 2017. Diversity of Algerian oases date palm (*Phoenix dactylifera* L., Arecaceae): Heterozygote excess and cryptic structure suggest farmer management had a major impact on diversity. *PLoS One*, 12.

Nei, M. 1972. Genetic distance between populations. *The Amer. Naturalist,* 106: 283-292.

Ochieng, L.A., S.M. Githiri, B.A. Nyende, L.K. Murungi, N.C. Kimani, G.K. Macharia and L. Karanja. 2015. Analysis of the genetic diversity of selected East African sweet potato (*Ipomea batatas* [L.] Lam.) accessions using microsatellite markers. *Afr. J. Biotechnol.,* 14: 2583-2591.

Omire, A., J.O. Neondo, N. Budambula, R. Gituru and C. Mweu. 2020. Phenotypic Diversity of Doum Palm (Hyphaene compressa), a Semi-Domesticated Palm in the Arid and Semi-Arid Regions of Kenya. Scientifica (Cairo) 2020.

Petit, R.J., A. el Mousadik and O. Pons. 2008. Identifying populations for conservation on the basis of genetic markers. *Conser. Biol.,* 12: 844-855.

Porth, I. and Y.A. El-Kassaby. 2014. Assessment of the genetic diversity in forest tree populations using molecular markers. *Diversity, (Basel),* 6: 283-295.

Pritchard, J.K., M. Stephens and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics,* 155: 945-959.

Rao, V.R. and T. Hodgkin. 2002. Genetic diversity and conservation and utilization of plant genetic resources. *Plant Cell Tissue Organ Cult.,* 68: 1-19.

Salem, K.F.M. and A. Sallam. 2016. Analysis of population structure and genetic diversity of Egyptian and exotic rice (*Oryza sativa* L.) genotypes. *C. R. Biol.,* 339: 1-9.

Shete, S., H. Tiwari and R.C. Elston. 2000. On estimating the heterozygosity and polymorphism information content value. *Theor. Popul. Biol.,* 57: 265-271.

Squirrell, J., P.M. Hollingsworth, M. Woodhead, J. Russell, A.J. Lowe, M. Gibby and W. Powell. 2003. How much effort is required to isolate nuclear microsatellites from plants? *Mol. Ecol.,* 12: 1339–1348.

Stauffer, F.W., D. Roguet, C. Christe, Y. Naciri, M. Perret, D.N. Ouattara and F. Stauffer. 2018. A multidisciplinary study of the doum palms ('Hyphaene'Gaertn.): origin of the project, current advances and future perspectives. *Nous*., 97: 115.

Sukanya, S.L., J. Sudisha, P. Hariprasad, S.R. Niranjana, H.S. Prakash and S.K. Fathima. 2009. Antimicrobial activity of leaf extracts of Indian medicinal plants against clinical and phytopathogenic bacteria. *Afr. J. Biotechnol.,* 8.

Vieira, M.L.C., L. Santini, A.L. Diniz and C. de F. Munhoz. 2016. Microsatellite markers: what they mean and why they are so useful. *Genetics & Mol. Biol.,* 39: 312-328.

Wright, S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution,* 395-420.

Yang, X., W. Su, L. Wang, J. Lei, S. Chai and Q. Liu. 2015. Molecular diversity and genetic structure of 380 sweetpotato accessions as revealed by SSR markers. *J. Integ. Agri.,* 14: 633-641.

Zaki, N.M., R. Singh, R. Rosli and I. Ismail. 2012. Elaeis oleifera genomic-SSR markers: exploitation in oil palm germplasm diversity and cross-amplification in Arecaceae. *Int. J. Mol. Sci.,* 13: 4069-4088.

Zehdi-Azouzi, S., E. Cherif, S. Moussouni, M. Gros-Balthazard, S. Abbas Naqvi, B. Ludeña, K. Castillo, N. Chabrillange, N. Bouguedoura and M. Bennaceur. 2015. Genetic structure of the date palm (*Phoenix dactylifera*) in the old world reveals a strong differentiation between eastern and western populations. *Ann. Bot.,* 116: 101-112.

(Received for publication 11 December 2021)