**Research Article**  **Open Access**

# Genetic Diversity and the Impact of Geographical Location on the Relationships Between *Phoenix dactylifera* L. Germplasms Grown in Saudi Arabia

**Widad S Aljuhani\***

*Biological Science Department, College of Applied Science, Umm Al-Qura University, Saudi Arabia*

## Abstract

*Phoenix dactylifera* L. is an important crop in the Middle East and North Africa. Date palm diversity faces risks including salinity and lack of rain. It is important to understand the genetic diversity of these germplasms. The objectives of this study were to investigate the degree of dissimilarity and to examine the impact of location on the genetic relationship between local cultivars in Saudi Arabia. The current study included 91 famous date palm accessions and a group of non-famous cultivars, collected from three main regions that are important in date palm cultivation in Saudi Arabia, and twenty-four nuclear microsatellite loci were tested. High polymorphism content was detected in some loci, making it possible to identify and distinguish strains using four markers. Examination of the genetic variation between local date palm germplasms showed a wide range of genetic dissimilarity (0-0.950). For the neighbor-joining algorithm tree, structural analysis and discrimination analysis for principal component, local genotypes can be classified within three main clusters. Cultivars of some regions were grouped according to their geographical location. The results of this study will be useful for those interested in strain identification, classification and conservation as well for those interested in improving fruit production.

## Introduction

Identifying and understanding the genetic diversity of germplasms is an important factor in breeding and conservation. Discrimination amongst date cultivars is extremely difficult. The first attempts were based on the morphological characteristics of the fruit, including shape, weight, colour and texture. However, fruits' characteristics are often unreliable for identifying cultivars due to the influence of environmental and agricultural conditions [1]. Recent developments in molecular techniques have led to an increased interest in identifying and studying the relationships between date palm cultivars. DNA fingerprinting applications have become key instruments for identifying and discriminating between date cultivars. A range of molecular markers, including restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), inter-simple sequence repeats (ISSRs) and simple sequence repeats (SSRs), have been used for this purpose [2].

Early attempts, e.g., Corniquel et al. [3] used RFLP to identify four date palms obtained from the United Arab Emirates. RFLP markers were applied in date palm cultivars from Morocco [4]. RAPD has been applied widely to date palm cultivars in different countries. For example, it was used by Sedra et al. [5] to screen genetic variations amongst thirty-seven date accessions from Morocco and six cultivars from Iraq and Tunisia. Furthermore, Soliman et al. [6] used the RAPD technique to identify date cultivars and to compare four females and four unknown males from Egypt. The most famous date palm cultivars in the Kingdom of Saudi Arabia (KSA) were analysed using the RAPD method [6-10].

Each of these studies revealed a high level of genetic similarity between the study samples. However, the studies used RAPD, which has been criticised for several reasons. Indeed, the RAPD method suffers from a number of major drawbacks. Many authors have pointed out problems related to the primers' action mechanism for amplification. For example, Lowe et al. [11] argued that the 'RAPD technique has been extensively criticised on technical and theoretical grounds; these criticisms include issues associated with reproducibility, primer structure, marker dominance, product competition, product

homology, allelic variation, genome sampling and non-independence of loci' (pp. 38-39).

Adawy et al. [12] used AFLP to screen the variations within and between 14 accessions from Egypt. The results showed a low level of polymorphism. AFLP was able to separate the cultivars based on their locations. In contrast, high polymorphism was detected using AFLP in 39 cultivars from Iraq, where genetically distinct varieties were detected [13]. In addition, Coe et al. [14] used AFLP to measure the variation among 21 cultivars growing in the US state of California. They confirmed the high polymorphism of markers and were able to distinguish between these germplasms. There has been a clear increase in the application of the microsatellite technique (SSR) for date palm germplasms. For example, Elshibli et al. [15] studied 37 females and 23 males from Sudan and Morocco.

Al-Ruqaishi et al. [16] examined the variation among 21 date palm accessions from Oman, Bahrain, Iraq and Morocco. Zehdi et al. [17] used microsatellite markers to study variations in 101 Tunisian date palm accessions. They found high polymorphism in the SSR markers and confirmed their discriminatory power. In fact, microsatellite markers have recently become the most used molecular method. This is because it is a simple technique with high polymorphic content, only a small amount of DNA is required and the process is relatively inexpensive. Hence, SSR plays an important role in studying the differences between closely related individuals as well as intra-population among crops. SSR has been used increasingly as a tool to assess genetic distances and genetic diversity [18,19] to identify cultivars [20,21].

**\*Corresponding author:** Widad S Alijuhani, Biological Science Department, College of Applied Science, Umm Al-Qura University, Saudi Arabia, Tel: 966595553791; E-mail: wsjuhani@uqu.edu.sa

**Figure 1:** Map of Saudi Arabia showing locations of sampling.

In the past five years, there has been significant progress in date palm research after the deployment of more than one a draft of the palm genome. Al-Dous et al. [22] published the first draft assembly of the sequences of the nuclear date palm genome as a project developed by Weill Cornell Medical College in Qatar (WCMCQ), identifying a region strongly linked to sex determination. Another draft included the date palm chloroplast genome [23]; the project was a joint venture between the Beijing Genomic Centre in China and the King Abdul-Aziz City for Science and Technology in Riyadh, Saudi Arabia. More recently, the first complete date palm genomic map [24] was constructed based on sequences taken from three males and three females from both Asia and Africa. This genetic map was combined with the genomic data from three commercially important palms- date, oil and coconut- and the application of this genomic information is expected to improved production and save these commercial cultivars.

In the KSA there are around 450 cultivars of date palm [25], it is large country and features wide variations in geographical and environmental conditions, there are distinctive types of dates in each region. Previously, Al-Bakr et al. and Hussein et al. [26,27] divided the local cultivars in the KSA into three main divisions based on the geographic regions in which they were observed. Currently, local date palm cultivars are divided within 13 regions [28]. The most famous regions in production are the Western, Central and Eastern regions of the KSA. However, no study has aimed to verify the existence of genetic relations between local germplasms from the same geographical area or with the same geographical origin.

The objectives of this paper are: 1) To investigate the degree of dissimilarity and genetic diversity between local varieties in the KSA using highly polymorphic SSR markers; 2) To compare cultivars according to geographical locations; and 3) To examine the possibility of classifying local samples according to their geographical location.

## Materials and Methods

### Study area

The female samples for this study were collected from three main regions of date palm production in the KSA: Al-Ahsa, in the Eastern region, from the National Date Palm Research Centre (NDPC); Al-Riyadh and Qassim in the Central region; and AL-Madinah, Makkah and Bisha in the Western region (Figure 1). Male samples were collected from the agricultural research station at King Saud University, Riyadh. Local female cultivars were subdivided into groups, which were based on the East, Central and Western geographical areas, in addition to the group of males in the Central region.

### Plant material

Young leaves were selected from 91 accessions of date palms representing 41 female cultivars and 17 males, as shown in Table S1 (supplementary tables). The cultivars represent the most common cultivars in the plantations and were chosen because of the quality of their fruit, their popularity with consumers and their popularity in the regions included in this study, In addition to a group of less-famous varieties. Accessions of cultivars were collected from more than one location. Expert agricultural engineers identified the date palms via study samples. Leaves representing three individuals per cultivar were randomly selected. The leaflets of each leaf were dried and preserved in silica gel.

### DNA extraction

DNA was extracted from dry young leaves according to the hexadecyltrimethylammonium bromide (CTAB) method [29], with modifications (supplementary files methods). Extracted DNA concentrations were determined using a NanoDrop light spectrophotometer (Thermo Scientific). The quality of DNA was checked using 1% agarose gel electrophoresis and visualised under UV light.

| No. | Primer name | Primer's sequence | Optimal | Expected | Motif |
|---|---|---|---|---|---|
| 1 | *mPdCIR010 | F:ACCCCGGACGTGAGGTG<br>R: CGTCGATCTCCTCCTTTGTCTC | 55.9 | 118-161 | (GA)22 |
| 2 | *mPdCIR015 | F: AGCTGGCTCCTCCCTTCTTA<br>R: GCTCGGTTGGACTTGTTCT | 51.6 | 120-156 | (GA)15 |
| 3 | *mPdCIR016 | F: AGCGGGAAATGAAAAGGTAT<br>R:ATGAAAACGTGCCAAATGTC | 51.7 | 130-138 | (GA)14 |
| 4 | *mPdCIR025 | F: GCACGAGAAGGCTTATAGT<br>R:CCCCTCATTAGGATTCTAC | 51.6 | 120-156 | (GA)15 |
| 5 | *mPdCIR032 | F: CAAATCTTTGCCGTGAG<br>R:GGTGTGGAGTAATCATGTAGTAG | 51.5 | 284-305 | (GA)19 |
| 6 | *mPdCIR035 | F: ACAAACGGCGATGGGATTAC<br>R:CCGCAGCTCACCTCTTCTAT | 53.9 | 175-221 | (GA)15 |
| 7 | *mPdCIR044 | F: ATGCGGACTACACTATTCTAC<br>R:GGTGATTGACTTTCTTTGAG | 51.7 | 130-138 | (GA)14 |
| 8 | *mPdCIR048 | F: CGAGACCTACCTTCAACAAA<br>R:CCACCAACCAAATCAAACAC | 51.4 | 156-192 | (GA)20 |
| 9 | *mPdCIR050 | F: CTGCCATTTCTTCTGAC<br>R:CACCATGCACAAAAATG | 48.5 | 154-208 | (GA)21 |
| 10 | *mPdCIR057 | F: AAGCAGCAGCCCTTCCGTAG<br>R:GTTCTCACTCGCCCAAAAATAC | 49.3 | 199-231 | (GA)22 |
| 11 | *mPdCIR070 | F: CAAGACCCAAGGCTAAC<br>R: GGAGGTGGCTTTGTAGTAT | 48.7 | 182-208 | (GA)17 |
| 12 | *mPdCIR078 | F: TGGATTTCCATTGTGAG<br>R: CCCGAAGAGACGCTATT | 49.6 | 117-152 | (GA)13 |
| 13 | *mPdCIR085 | F: GAGAGAGGGTGGTGTTATT<br>R:TTCATCCAGAACCACAGTA | 51.5 | 284-305 | (GA)19 |
| 14 | *mPdCIR090 | F: GCAGTCAGTCCCTCATA<br>R: TGCTTGTAGCCCTTCAG | 48.6 | 142-175 | (GA)26 |
| 15 | *mPdCIR093 | F: CCATTTATCATTCCCTCTCTTG<br>R:CTTGGTAGCTGCGTTTCTTG | 51.8 | 153-184 | (GA)16 |
| 16 | **DP151 | F: TTGCTGGTTGAAATGGTGTT<br>R: GCAACAGATGCTCTTGCTCA | 53.9 | 175-221 | (GA)15 |
| 17 | **DP157 | F: TGGACAATGACACCCCTTTT<br>R: GCCCACACAACAACCTCTCT | 54.6 | 180-244 | (TC)19 |
| 18 | **DP159 | F: AGCTCCAATTTGCTGCAGAG<br>R: GCTGACCTGGAGTCCAAAAC | 54.3 | 120-156 | (TC)27 |
| 19 | **DP160 | F: AAGAGCGACAATCATGACCA<br>R: GGAAATTGAAGGGCATCTTG | 51.7 | 281-332 | (GA)19 |
| 20 | **DP168 | F: GCAGCAAAGCCCTTAGGC<br>R: GGTGTTATGTGCAGCCAATG | 54.3 | 163-175 | (CAT)8 |
| 21 | **DP169 | F: GCATGGACTTAATGCTGGGTA<br>R: GGTTTTCCTGCCAACAACAT | 57.1 | 129-223 | (AAT)12 |
| 22 | **DP170 | F: TCTTTGGGCTTACGACAACC<br>R: GTATGGCCCAAGATGCAGAT | 51.4 | 156-192 | (GA)20 |
| 23 | **DP171 | F: GTGGGAGTAGCGAGGTATGG<br>R: GTCCGGCACTTTAGGAAGTT | 56 | 197-218 | (TTC)10 |
| 24 | **DP172 | F: GGTGTTTGGGCCTATTTCCT<br>R: GTCCTCCTCCTCCTCTGTCC | 54.2 | 199-235 | (AGG)11 |

**Table 1:** Forward and reverse primer sequences, repeat motifs and expected sizes of microsatellite loci.

## Microsatellite primers

The 24 SSR primer pairs were selected from the literature and to reflect a high degree of polymorphism among previously sampled date palm cultivars. These included the dinucleotide repeat GA [30] in addition to di- and trinucleotides (TC, CAT, AAT, TTC, AGG) [31] (Table 1).

## Polymerase chain reaction (PCR)

PCR was performed following the method described by Billotte et al. [30], using a total volume of 10 μl: 5 μl BioMix (2×Master Mix, containing ultra-stable Taq DNA, BioLine, UK), 2.8 μl of molecular grade water, 2 μl (25 ng) of total genomic DNA and 0.1 μl of each primer. Amplification was performed in a Veriti 96-Well Fast Thermal Cycler (Applied Biosystems/USA), with an initial denaturation at 95°C for 1 min, followed by 35 cycles at 94°C for 30 s and an annealing temperature depending on the SSR locus (Table 1) for 60 s, followed by 72°C for 120 s and a final elongation at 72°C for 8 min.

## Fragment analysis and scoring SSR

The successful amplifications of PCR products of three or four markers were mixed into each well of a plate and then sent to Source BioScience/UK for fragment analyses. GeneMapper (Applied Biosystems, USA) software version 4.0 was used to detect the allele sizes at each SSR loci. MsatAlleses software version 1.02 from R software [32] was used to unify the size of alleles [33].

## Data analysis

Initially, the capability of the SSR markers used in this study was assessed by calculating the number of alleles of each locus and the mean alleles of all loci. In addition, number of genotype of each locus, major allele frequency and polymorphic information content (PIC) were

calculate using PowerMarker software version 3.25 [34]. Genepop software version 4.0.7 [35] was used to estimate the maximum likelihood of null allele frequency.

Based on the results of the genetic diversity of genotypes and allele numbers, a manual molecular identification key was constructed for 91 date palm accessions using SSR, which reflected a high amount of polymorphic information content. This key was used to identify the local genotypes with the fewest possible number of microsatellite markers and to facilitate the task of identifying the varieties. The SSR loci were initially arranged according to the largest number of alleles and the group of SSR loci that best discriminated all varieties. Thus, each cultivar was detected according to its distinctive genotype in the selected SSR loci.

Data gathered based on the evaluated microsatellite markers were used to implement all microsatellite analyses. The expected ($H_{exp}$) and observed heterozygosity ($H_{obs}$) were calculated for the mean and each SSR locus and per four sub-populations, the groups of disparate gender and the geographical location of female cultivars, based on the work of Nei et al. [36], to measure genetic variations within a population using the Genetix software program [37].

Wright's fixation indices were determined for differentiated hierarchical levels of a population structure [38]: $F_{is}$ (inter-individual), $F_{st}$ (sub-groups) and $F_{it}$ (total population) for each SSR locus, for three groups of females from the West, East and Central regions and for the male group, using Genetix [37].

Analysis of molecular variance (AMOVA) and pairwise $F_{st}$ values were calculated among the four groups of date palm genotypes, and the significance of the statistics was calculated based on 999 permutations using Genalex ver. 6.5 [38,39].

Genetically shared allele distance ($D_{sa}$) [40] was calculated to summarize the genetic dissimilarity between each pair of individuals. PowerMarker software version 3.25 [34] was used to perform the genetic distance analysis, based on the mean allele frequencies in each SSR locus. A matrix obtained from $D_{sa}$ was used to set up the dendrogram using the neighbor-joining algorithm (NJ) [41] to summarize the relationships between the germplasms. Bootstrap analysis (1000 iterations) was conducted to detect the confidence and validity of clusters, also using PowerMarker 3.25. The resulting tree was generated using MEGA software version 4 [42].

A heat map was constructed using a matrix comprised of the calculated $D_{sa}$ among each pair of samples and used to facilitate visualised differentiations between and within the cultivars. The heat map was created using the R software program Reshape2 version 1.2.2 [43].

A Bayesian clustering analysis was also applied to estimate palms in this study. This analysis detected the number of sub-clusters based on the frequency of data gathered for alleles from multilocus genotypes, without initial information regarding population, where the number of the population (K) was unknown, using Structure software version 2.3.3 [44]. The length of the burn-in period was 7,000 steps, which was sufficient for observing data convergence in which K reached equilibrium in the values produced without excessive variations in parameters. After burn-in, the number of Monte Carlo Markov Chain (MCMC) iterations was 70,000. Several runs for each K value were performed, each of which included a different number of MCMC steps, in order to confirm that the results were consistent. The 20 iterations were tested to obtain reliable estimates of the proportions of ancestry membership within a population. The possible number of K

clusters was tested in a range from 1 to 20. The K defined by Evanno et al. [45] was used to detect the most the potential number of genetic sub-clusters that existed in the samples of the date likely number of populations (Figure 2).

A discriminant analysis of principal components (DAPC) was conducted to determine the maximum number of clusters that could be divided into local varieties and to gain an overview of the relationships between the clusters. This was performed using Adegenet software version 1.4-2 software [46] and R software version 3.0.2 (R Development Core Team, 2013). The optimal number of clusters was considered the smallest number, after increasing the number of clusters, that did not lead to a decrease in the Bayesian information criterion (BIC) value [46]. The model was run for $1 \times 10^{(6)}$ iterations to detect relationships between the individuals (convergence), maintaining the principal components that explained 95% of all variances.

A Mantel test [47] was conducted to determine the correlations between two matrices (AB) of geographical distance (Km) and $D_{sa}$ between each pair of genotypes. The p-values were calculated from the correlation coefficient for AB, r(AB), estimated from 10,000 permutations. The null hypothesis was that the matrices were not correlated; the alternative hypothesis was that the matrices were correlated. The Mantel test was performed using XLSTAT software version 2014.6.02.

## Results

Most microsatellite markers use amplified clear bands. In the present results, 16 out of 24 SSR primer pairs produced clear bands in all specimens. These pairs were mpdcir015, mpdcir016, mpdcir025, mpdcir032, mpdcir035, mpdcir050, mpdcir070, mpdcir085, mpdcir090, mpdcir093, pd160, pd168, pd169, pd170, pd171 and pd172. Primer pairs pd151, pd157 and pd159 failed to amplify under various PCR conditions, while mpdcir010, mpdcir44, mpdcir048, mpdcir057
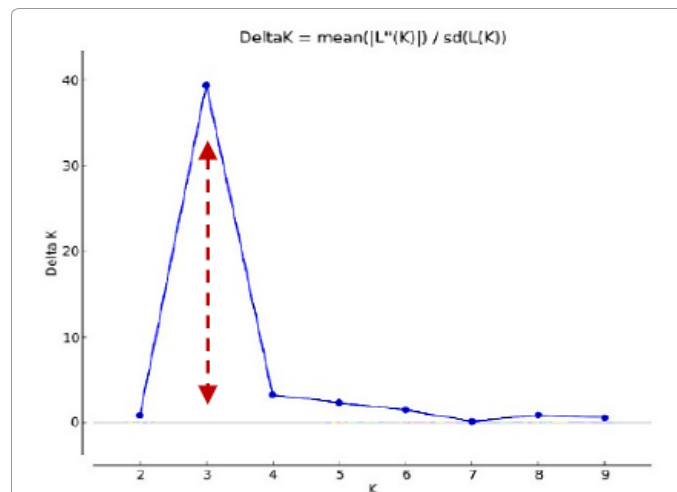


**Figure 2:** The inference of K, the most probable number of clusters, using Structure software, based on microsatellite analysis of 91 total samples of *phoenix dactylifera* L., changes the log-likelihood of the data (Δ K) as a function of K, calculated over twenty replicates, with the best number shown for the three sub-groups. Figure 4: Inferred clusters in the date palm cultivars using Structure, showing those where K=3, according to the Bayesian analysis. Each bar represents one individual genotype. Individuals with multiple colours have admixed genotypes from multiple clusters. Each colour represents the most likely ancestry of the cluster from which the genotype or partial genotype derived. Local date palm genotypes could be classified within three sub-groups: Western region females, central and Eastern females, and males. There is a clear separation for the local Western region group and for local varieties. F: Female. M: Male. W: West. C: Central. E: East.

| MarkersSSR | Major Allele Frequency | NoGenotype | No Allele | PIC | expH | obsH | isF | itF | stF | Null Allele Frequency |
|---|---|---|---|---|---|---|---|---|---|---|
| mpdcir015 | 0.200 | 28.000 | 15.000 | 0.820 | 0.745 | 0.588 | 0.175 | 0.236 | 0.070 | 0.101 |
| mpdcir016 | 0.330 | 10.000 | 7.000 | 0.390 | 0.625 | 0.500 | 0.168 | 0.232 | 0.077 | 0.157 |
| mpdcir025 | 0.430 | 7.000 | 6.000 | 0.370 | 0.710 | 0.546 | 0.081 | 0.145 | 0.070 | 0.111 |
| mpdcir032 | 0.260 | 16.000 | 8.000 | 0.740 | 0.703 | 0.626 | 0.115 | 0.177 | 0.070 | 0.098 |
| mpdcir035 | 0.380 | 9.000 | 7.000 | 0.580 | 0.616 | 0.586 | 0.046 | 0.092 | 0.072 | 0.126 |
| mpdcir050 | 0.300 | 16.000 | 10.000 | 0.700 | 0.699 | 0.891 | -0.218 | -0.109 | 0.089 | 0.002 |
| mpdcir070 | 0.360 | 9.000 | 6.000 | 0.540 | 0.671 | 0.643 | 0.196 | 0.237 | 0.051 | 0.170 |
| mpdcir085 | 0.403 | 19.000 | 12.000 | 0.700 | 0.628 | 0.637 | 0.079 | 0.212 | 0.144 | 0.030 |
| mpdcir090 | 0.310 | 18.000 | 9.000 | 0.750 | 0.709 | 0.662 | 0.190 | 0.232 | 0.052 | 0.136 |
| mpdcir093 | 0.830 | 9.000 | 6.000 | 0.300 | 0.340 | 0.232 | 0.127 | 0.202 | 0.085 | 0.067 |
| pd160 | 0.230 | 15.000 | 10.000 | 0.500 | 0.603 | 0.668 | -0.084 | -0.030 | 0.049 | 0.088 |
| PD168 | 0.700 | 6.000 | 4.000 | 0.320 | 0.615 | 0.296 | 0.637 | 0.678 | 0.111 | 0.187 |
| pd169 | 0.450 | 8.000 | 5.000 | 0.500 | 0.257 | 0.211 | 0.190 | 0.232 | 0.052 | 0.125 |
| pd170 | 0.440 | 9.000 | 5.000 | 0.500 | 0.625 | 0.500 | 0.168 | 0.232 | 0.077 | 0.067 |
| pd171 | 0.360 | 11.000 | 7.000 | 0.450 | 0.641 | 0.550 | 0.111 | 0.160 | 0.055 | 0.141 |
| pd172 | 0.330 | 19.000 | 11.000 | 0.600 | 0.710 | 0.546 | 0.081 | 0.145 | 0.070 | 0.118 |
| Mean | 0.395 | 13.063 | 8.000 | 0.548 | 0.619 | 0.543 | 0.129 | 0.192 | 0.075 | 0.108 |
| Sum | | 209 | 128 | | | | | | | |

**Table 2:** Summary of genetic diversity of 91 date palm accessions grown in Saudi Arabia, using 16. Microsatellite markers. PIC: Polymorphism Information Content, Hexp: expected heterozygosity, Hobs: observed heterozygosity, Fis, Fit, Fst: Wright's analysis of hierarchical F-statistics. Values calculated per locus.

and mpdcir078 loci were amplified in some cultivars or amplified a weak band in some cases. The last group was excluded from the next steps of the analysis.

The 16 SSR loci produced 128 alleles, and the mean number of alleles per marker was 8, with numbers ranging from 4 (pd168) to 15 alleles (mpdcir015). The polymorphic information content (PIC) ranged from 0.300 (mpdcir093) to 0.820 (mpdcir015), with a mean of 0.548 (Table 2). Two hundred and nine genotypes were identified, and the genotypes per locus ranged from 6 (locus pd168) to 28 (locus pdcir015), with a mean of 13. The null alleles' frequency ranged from low (0.002) to intermediate (0.187), with a mean of 0.108. No locus recorded a high null allele frequency level ($r \geq 0.20$).

Four loci, mpdcir015, mpdcir032, mpdcir050 and mpdcir085, were high polymorphic, and it was possible to discriminate the study samples successfully (Figure S1, supplementary figures).

The expected heterozygosis ($H_{exp}$) was 0.619 on average, ranging from 0.257 (pd169) to 0.745 (mpdcir015). The observed heterozygosis ($H_{ops}$) ranged from 0.211 (pd169) to 0.891 (mpdcir050), with an average value of 0.543. Low genetic variation within the four populations (Table A2, supplementary tables) was found between Western cultivars as well, with low values for expected ($H_{exb}=0.441$) and observed Heterozygosis ($H_{obs}=0.440$) and high values for diversity within males ($H_{exb}=0.732$ and $H_{obs}=0.643$).

The results obtained from Wright $F_{is}$ values are presented in Table S2 (supplementary tables). The highest value between germplasms ($F_{is}=0.364$) was found between the females of the Western specimens, while the lowest Fis value was found between the males ($F_{is} = -0.089$).

The results obtained from pairwise sub-groups ($F_{st}$ values) are presented in Table S3 (supplementary tables). The lowest genetic variation between local germplasms ($F_{st}=0.014$, p-value=0.016, permutations=999) was found between the males and females of

the Central region specimens. The largest variation was found between the Western and Eastern local cultivars ($F_{ST}=0.116$, p=0.001, permutations=999).

Wide range of dissimilarities between the cultivars, ranged from 0.250 to 0.950 (Figure S3, supplementary figures). A group of samples was gathered under the same cultivar and in one branch. However, variations within cultivars (in the range of 0.05-0.1) and between individuals in the 'Ajwa', 'Ruziez' and 'Naboot Saif' cultivars were also observed, and these were collected at different locations.

The relationships between local date palm genotypes are presented in an NJ dendrogram (Figure 3). The 91 accessions were divided into three main clusters and more than seven small sub-branches. Only the Western accessions appeared in the first cluster (branches with red colour), which contained 28 accessions (Figure 3). All Western samples were grouped within this cluster with the exception of 'Ruthana'. The second NJ cluster contained 39 accessions, including a mixture of Central and Eastern genotypes. The third cluster included 24 accessions and contained all males, with two cultivars from Eastern samples and two as well from Central samples. No Western specimens were included in the third cluster.

The heat map showed a wide range of differences (Figure S2, supplementary figures). The greatest frequency of dissimilarity was 0.5-0.7 (68%; Figure S3, supplementary file). The lowest dissimilarity between the cultivars was 0.250. The heat map showed less difference between the groups of cultivars in the Western region, as indicated in the blue square on the bottom-right portion of the map (Figure S2, supplementary figures).

Based on the K-values (Figure 2) and bar plots (Figure 4), the results of the Bayesian analyses suggested locating these cultivars in three sub-clusters. The first sub-group (red bar) includes all Western cultivars except 'Ruthana'. Most of the Central region cultivars are located in
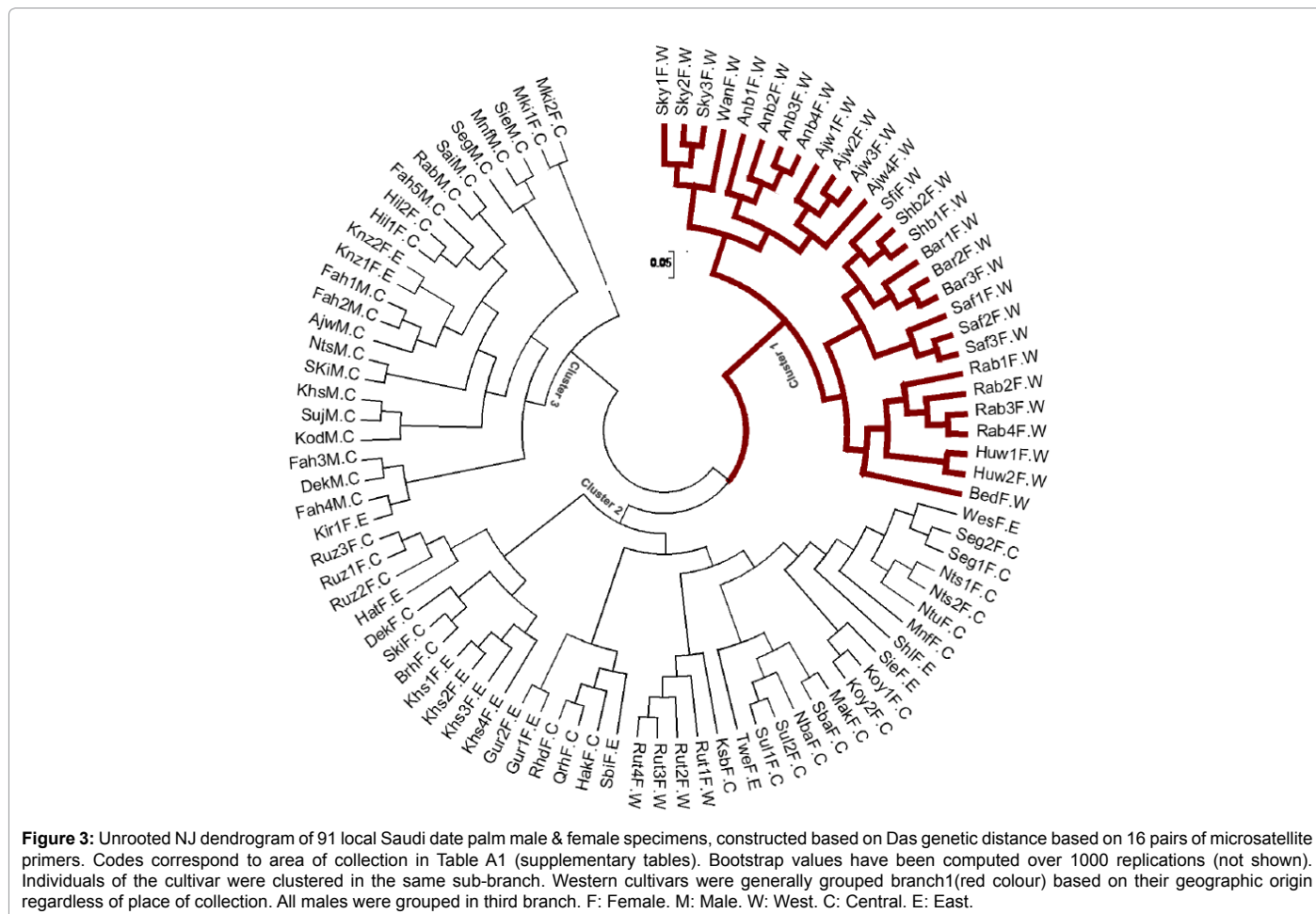
**Figure 3:** Unrooted NJ dendrogram of 91 local Saudi date palm male & female specimens, constructed based on Das genetic distance based on 16 pairs of microsatellite primers. Codes correspond to area of collection in Table A1 (supplementary tables). Bootstrap values have been computed over 1000 replications (not shown). Individuals of the cultivar were clustered in the same sub-branch. Western cultivars were generally grouped branch1(red colour) based on their geographic origin regardless of place of collection. All males were grouped in third branch. F: Female. M: Male. W: West. C: Central. E: East.

the second cluster (green bar), mixed with the Eastern region cultivar. The third sub-group (blue bar) includes all males (Figure 4). However, there were a group of females (9 cultivars) mixed with females in cluster three.

Figure 5 shows that the DAPC of 91 genotypes separated the genetic data into three groups. A clear discrimination between the three clusters was facilitated by two discriminant functions. The first represented 67% of the variations, while the second explained 33% of the variations. Of the 91 individuals, 42.8% were assigned to cluster 1, 39.5% to cluster 2 and 17.5% to cluster 3. Clusters 1 and 2 exhibited the highest density of samples.

There were similar results gained by DAPC; all the observed Western specimens (except Ruthana) were located in cluster 1. Most of the Central Eastern region cultivars were in cluster 2, which had been separated with PC2. All males were located in cluster 3, which was separated with PC1. Individuals in the same cultivar were gathered in the same cluster, which raises confidence in the results. Table S4 (supplementary tables) shows the cultivars alongside their clusters. The DAPC results were consistent with the NJ tree as well as for relationships between the local germplasms in terms of the mixes between samples from the Central and Eastern specimens in clusters, which were greater than the potential for overlap with the samples from the Western region. The loading plots for the discrimination function of PC1 and PC2 (Figures S4 & S5, supplementary figures) showed the greatest contribution loci in segregation for mpdcir015, mpdcir050, mpdcir085 and pd172. The loci also reflected high PIC (Table 2).

In the Mantel test (Figure 6), since the p-value (0.0001) was lower than the significance level (alpha=0.05), we could reject the null hypothesis (matrices are not correlated) and accept the alternative hypothesis (matrices are correlated). The value of the correlation coefficient of two matrixes genetic and geographical distance r(AB) was 0.121 from a maximum of +1, indicating that there was a relatively positive correlation between the two matrices.

## Discussion

The initial objective of this study was to detect the degrees of genetic variation between local date palm cultivars and verify the effects of geographic location on the relationships amongst them. The results showed SSR primer pairs, developed by Billotte et al. [30] and used in this study, had a high PIC. There were 128 and 209 alleles and genotypes, respectively, when screening 91 accessions from the KSA. Zehdi et al. [17] reported similar results in their study on 101 date palm accessions from Tunisia, recording 134 alleles and 311 genotypes. However, the number of alleles in the current study was much lower (343) than those identified by Elshibli et al. [15]. These differing results might be explained by variations in the degree of genetic diversity in the samples, resulting from a background of high sexual or backcross reproduction, which increases the level of diversity [48].

The PIC was high, with a mean of 0.536. In the study conducted by 49 Sharma et al. [49], SSR markers were classified as informative when PIC was ≥ 0.5. Most of the 16 SSR loci in the current study (11 out of 16) were ≥ 0.5, which is sufficient for discriminating between the 91

accessions. Four loci had high polymorphic information in SSR loci ≥ 0.7, which are mpdcir015, mpdcir032, mpdcir050 and mpdcir085 (PIC=0.82, 0.74, 0.70 and 0.70, respectively), which was sufficient for discriminating between the local date palm germplasms, using fewest possible number of microsatellite markers, we determined that each cultivar had a distinctive genotype in the selected SSR loci. The number of required markers compared to the number of cultivars was much lower in this study than in other studies, which was useful for cultivar identification; Al-Khalifa et al. [10] used 37 RAPD markers to identify 13 cultivars, El-Tarras et al. [9] used 10 markers to detect 6 cultivars, Askari et al. [7] used 21 RAPD primers to detect 100 local cultivars and Munshi et al. [50] used 10 RAPD and ISSR primers to detect four varieties. On the other hand, no locus recorded a high null allele level ($r ≥ 0.20$), instead ranging from low (0.02) to intermediate (0.187). This indicates the validity of the finding of homozygotes and strengthens the results.

The mean values of expected ($H_{exp}$) and observed ($H_{obs}$) heterozygotes were 0.558 and 0.508, respectively. High heterozygosity values for a breed may be due to long-term natural selection for adaptation, to the mixed nature of the breeds or to the historic mixing of strains of different populations. A low level of heterozygosity may be due to isolation and a subsequent loss of unexploited genetic potential [51]. Based on the mean value of the observed heterozygotes in the present study, the local strains represent a moderate level of diversity. Observed $H_{obs}$ heterozygotes were ≥ 0.5 for the mpdcir015, mpdcir032, mpdcir050 and mpdcir085 SSR loci. According to Kotze et al. [52], 'A high level of average heterozygosity at a locus could be expected to correlate with high levels of genetic variation at loci with critical importance for adaptive response to environmental changes' (p. 413-416).

The fewest heterozygotes were found within genotypes from the Western region (Table S2, supplementary tables). These agree with results, which are shown in the heat map. The heat map facilitated the visualisation of the genetic variations among these genotypes. Cultivars from the Western region showed less variance or were closer together in terms of genetic distance than those from the Eastern and Central regions. However, they differed from the genotypes of other regions, as clearly shown in the blue square at the bottom portion of the heat map. This might be because these strains have a long vegetative background with the desired qualities of reservation fruiting. At the same time, however, these qualities lead to low genetic diversity and strains that are very close together [51]. Another possibility is that these specimens did not overlap with cultivars of other regions or in isolation. The high Fis value (0.364) recorded between samples of Western cultivars compared to those from other geographical groups indicates the high degree of inbreeding among the local strains in this region. On the contrary, the negative values that were recorded between the male samples could be explained by the fact that these males were propagated via seed, with a high level of variation in the outcome of the backcross process.

The average $F_{st}$ value for the study samples was 0.075 (Table 2), indicating that there was a degree of variance between groups of germplasms. The variation between groups was higher than the value recorded between geographical groups of local date palm cultivars from Tunis ($F_{st}$=0.036) [17] and that recorded between gender groups for local male and female group genotypes from Oman (Fst = 0.02) [48], compared with the values from those two countries allow us to conclude that; there is a pattern of geographical classification between the specimens in the KSA.

In this study, there was a wider range of dissimilarity ($D_{sa}$=0.250 to 0.950) than in a previous study on local cultivars, which found that the percentage of similarity was 59-85% [7]. In addition, [9] reported

that the similarity was 85% to 96%, and Munshi et al. [50] reported an average similarity of 55-85%. The samples in the present study showed a wide range of variation, which is contrary to earlier studies, but this might be due to the discrimination ability and high polymorphism of the SSRs [53].

The current findings contradict the results of previous studies, which demonstrated the narrow genetic basis of local date palm strains in the KSA [6-10]. For example, the genetic distance between two local varieties, 'Anbara' from the Western region and 'Sheashee' from the Eastern region, was found to be 0.800 in the current study.

NJ, structure and DAPC analysis were used to separate individual cultivars into three clusters. The accessions from the Central region overlapped with those from the Eastern region. All Western germplasms were separated into one cluster, and all male specimens were clustered together. Cultivars within the clusters exhibited a similar (mostly) arrangement to the results of the structure analysis. Given the geographic position of the Central and Eastern regions, it is not uncommon to find varieties mixed between these two main groups. Breeders' activities are clear in this area in terms of improving cultivars and long-term natural selection for adaptation, which leads to the mixed nature of the breeds or to historic mixing of strains of different populations, and this is a clear pattern for most cultivars in these two regions.

Males are produced by sexual reproduction and grown from seeds, whereas females result from vegetative reproduction via offshoots. A few female cultivars were grouped with the males in cluster 3, as can be seen in the structure analysis (Figure 4). Either they represented the females of the cultivar, or when there has been seed propagation in a cultivar's ancestry [54]. The number of admixed females cultivars in cluster 3 decreased in NJ tree (Figure 2) and DAPC (Figure 5) when they were moved to cluster 2 with Central and Eastern female cultivar specimens. The DAPC findings confirmed the ability of this analysis to identify the affiliations in sets of individuals that showed uncertainty between population sub-groups. DAPC was proposed as an alternative structure for analysing complex genetic data and detecting admixed individuals by determining the probability that each individual belonged to each cluster. Thus, it was possible to classify individuals with complicated relationships and hybrids such as Mimulus (Phrymaceae; [55].

The correlation coefficient r(AB) was 0.121 of the Mantel test for correlation between genetic and geographical distance falls in the range of −1 to +1, where being close to −1 indicates a strong negative correlation while +1 indicates a strong positive correlation. An r-value of 0 indicates no correlation [56]. Since the p-value of 0.0001 was lower than the alpha significance level (0.05), we could reject the null hypothesis (matrices are not correlated) and accept the alternative hypothesis (matrices are correlated). Moreover, a set of specimens with different geographical distances showed a gradient tendency in the curve of the relationship between the genetic and geographic distance matrices (Figure 6).

## Conclusion

This is the first time that microsatellite markers have been used to explore genetic diversity between and within local date palm cultivars. Moreover, this is the first study to investigate the effect of geographical location on genetic dissimilarity between date palms grown in Saudi Arabia. The conclusions that can be drawn from this study are described below.

The results showed a wide range of genetic dissimilarity (up to $D_{sa}$=0.950). The SSR markers used in the present study, mpdcir015,
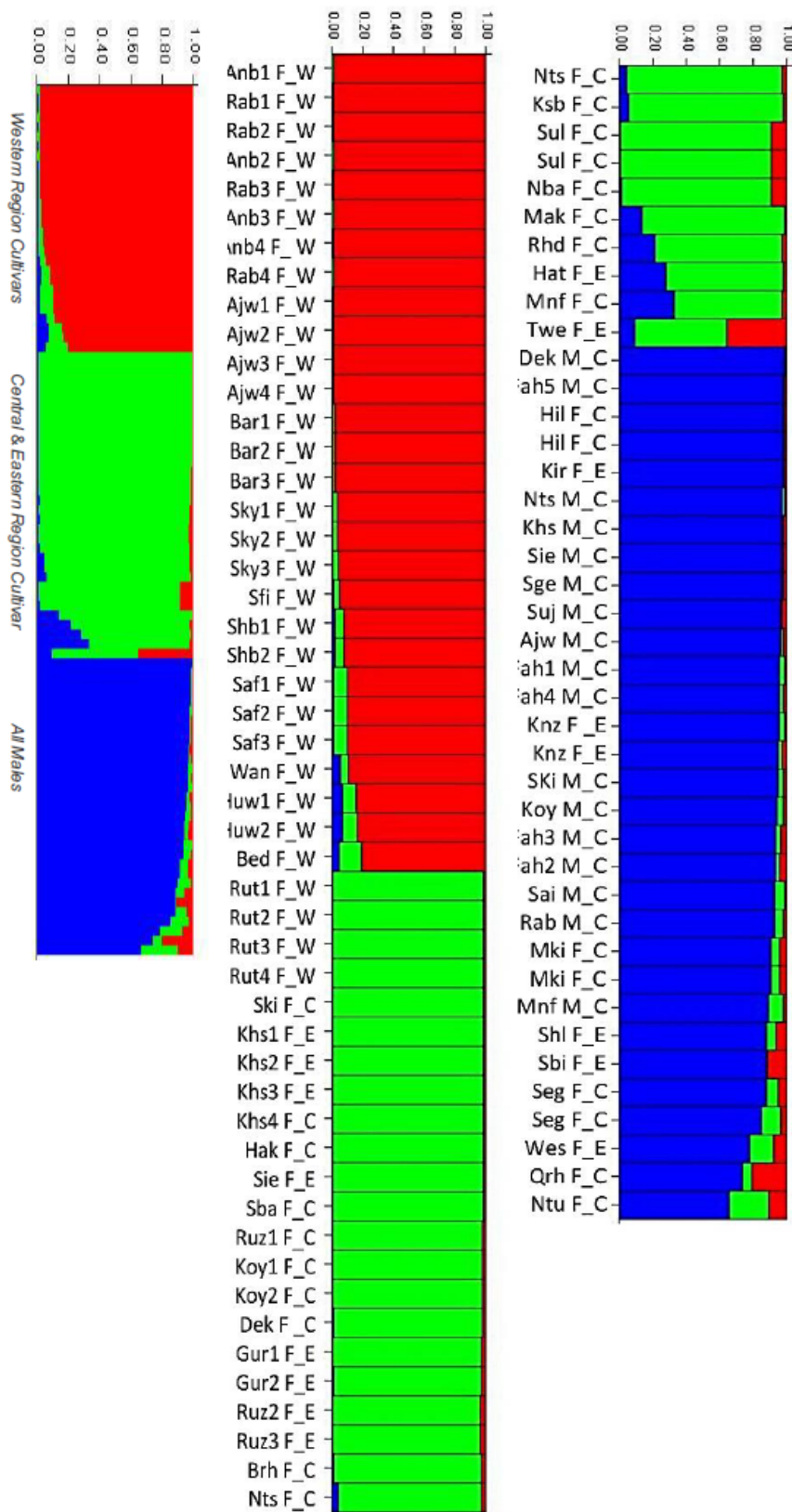
**Figure 4:** Inferred clusters in the date palm cultivars using Structure, showing those where K=3, according to the Bayesian analysis. Each bar represents one individual genotype. Individuals with multiple colours have admixed genotypes from multiple clusters. Each colour represents the most likely ancestry of the cluster from which the genotype or partial genotype derived. Local date palm genotypes could be classified within three sub-groups: Western region females, central and Eastern females, and males. There is a clear separation for the local Western region group and for local varieties. F: Female. M: Male. W: West. C: Central. E: East.
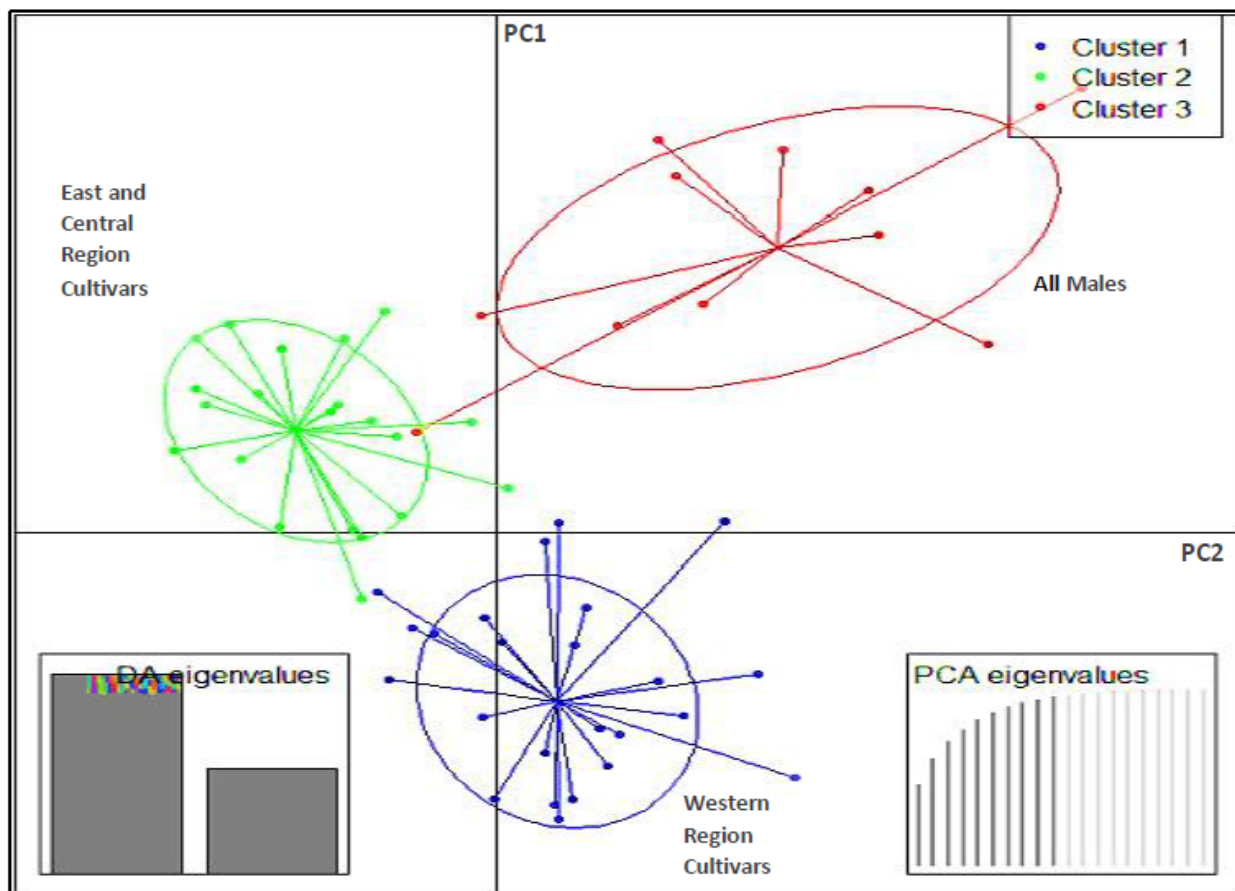
**Figure 5:** Scatter plot of the DAPC analysis for 91 date palm genotypes using 16 SSR loci. The local date palm genotypes are shown, clustered within 3 groups. Varieties of Western region females of the KSA were gathered in cluster 1 according to their geographical origin, while Eastern & central females were gathered in cluster 2. All males were gathered in cluster 3.
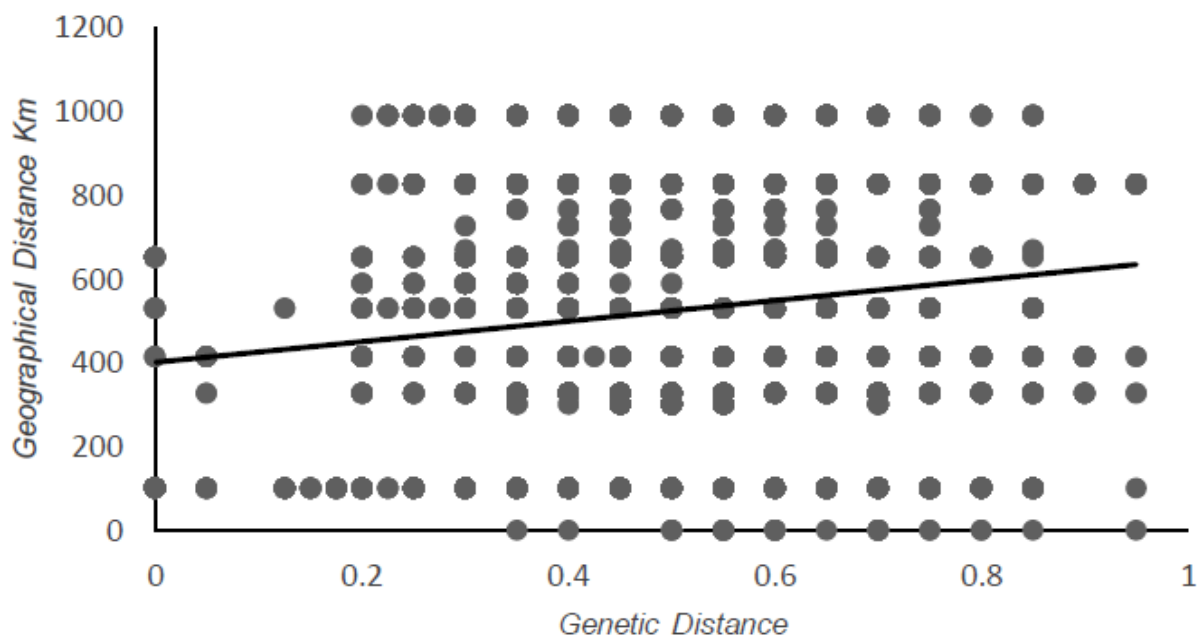


**Figure 6:** Mantel test determined genetic distance and geographic distance for date palm cultivars collected from the West, Central and East region of Saudi Arabia.

mpdcir032, mpdcir050 and mpdcir085, were informative and yielded high PIC. The results of this study are significant in regard to 1) the identification of genotypes and 2) a reduction in the number of markers required to distinguish variations. These aspects are important for the identification process, especially in dealing with large numbers of strains, as they minimise cost, time and effort and facilitate the ratification, exchange and storage of offshoots.

Therefore, these loci were highly heterozygous, which might also reflect adaptations to environmental changes. It is recommended that they should be used in future studies to screen the response to environmental conditions, such as salinity and drought.

The results of this study will be useful for those interested in the production of fruits and date palm breeding in terms of selecting female cultivars that show a high degree of genetic variation for use in programmes aimed at expanding genetic diversity in the local varieties, as cultivars with low genetic diversity are vulnerable to biotic and abiotic stresses.

This study confirmed the existence of a pattern of geographic distribution between date palm strains, shown clearly in Western region samples, which made it easy to genetically distinguish them from the rest of the samples.

On the other hand, a great similarity between Western region cultivars might be indicating to suffering from isolation, too. It is important to conserve and encourage the breeding of varieties that represent unique patterns.

## Acknowledgment

## References

1. Al-Khalifah NS, Askari E (2006) Detection of genetic diversity in date palm (Phoenix dactylifera L.). Nova Science Publishers, New York.

2. Agarwal M, Shrivastava N, Padh H (2008) Review: Advances in molecular marker techniques and their applications in plant sciences. Plant Cell Rep 27: 617-631.

3. Corniquel B, Mercier L (1997) Identification of date palm (Phoenix dactylifera L.) cultivars by RFLP: partial characterization of a cDNA probe that contains a sequence encoding a zinc finger motif. Int J Plant Sci 158: 152-156.

4. Trifi M, Rhouma A, Marrakchi M (2000) Phylogenetic relationships in Tunisian date palm (Phoenix dactylifera L.) germplasm collection using DNA amplification fingerprinting. Agron 20: 665-671.

5. Sedra MH, Lashermes P, Trouslot P, Combes MC, Hamon S (1998) Identification and genetic diversity analysis of date palm (Phoenix dactylifera L.) varieties from Morocco using RAPD markers. Euphytica 103: 75-82.

6. Soliman SS, Ali BA, Ahmed MMM (2003) Genetic comparisons of Egyptian date palm cultivars (Phoenix dactylifera L.) by RAPD-PCR. African Journal of Biotechnology 2: 86-87.

7. Askari E, Al-Khalifa NS, Ohmura T, Al-Hafedh YS, Khan FA, et al. (2003) Molecular phylogeny of seven date palm (Phoenix dactylifera L.) cultivars by DNA fingerprinting. Pakistan Journal of Botany 35: 323-330.

8. Al-Moshileh AM, Motawei MI, Al-Wasel A, Abdel-Latif T (2004) Identification of some date palm (Phoenix dactylifera L.) cultivars in Saudi Arabia using RAPD fingerprints. Agricultural and Marine Sciences 9: 1-3.

9. El-Tarras A, Al-Tawatti N, Al-Malki F (2007) Genetic fingerprint of some KSA date palm cultivars using modern biotechnological techniques. Biotechnology 6: 263-267.

10. Al-Khalifa N, Askari E, Shanavaskan P (2013) AE Date palm culture and genetic identification cultivars grown in Saudi Arabia. Riyadh, KSA: National Centre for Agriculture Technologies, King Abdulaziz City for Science and Technology (KACST).

11. Lowe A, Harris S, Ashton P (2004) Ecological genetics: Design, analysis and application. Oxford: Blackwell Publishing, UK.

12. Adawy SS, Hussein EHA, Ismail SEME, El-Itriby HA (2004) Genomic diversity in date palm (Phoenix dactylifera L.) as revealed by AFLPs in comparison to RAPDs and ISSRs. Arab Journal of Biotechnology 8: 99-114.

13. Jubrael JSM, Udupa SM, Baum M (2005) Assessment of AFLP-based genetic relationships among date palm (Phoenix dactylifera L.) varieties of Iraq. J Am Soc Hortic Sci 130: 442-447.

14. Coe E, Cone K, McMullen M, Chen S, Davis G, et al. (2002) Access to the maize genome: an integrated physical and genetic map. Plant Physiol 128: 9-12.

15. Elshibli S, Korpelainen H (2008) Microsatellite markers reveal high genetic diversity in date palm (Phoenix dactylifera L.) germplasm from Sudan. Genetica 134: 251-260.

16. Al-Ruqaishi IA, Davey M, Alderson P, Mayes S (2008) Genetic relationships and genotype tracing in date palms (Phoenix dactylifera L.) in Oman, based on microsatellite markers. Plant Genetic Resources: Characterization and Utilization 6: 70-72.

17. Zehdi S, Cherif E, Rhouma S, Santoni S, Hannachi AS, et al. (2012) Molecular polymorphism and genetic relationships in date palm (Phoenix dactylifera L.): The utility of nuclear microsatellite markers. Sci Hortic 148: 255-263.

18. Struss D, Plieske J (1998) The use of microsatellite markers for detection of genetic diversity in barley populations. Theor Appl Genet 97: 308-315.

19. Huang XQ, Borner A, Roder MS, Ganal MW (2002) Assessing genetic diversity of wheat (Triticum aestivum L.) germplasm using microsatellite markers. Theor Appl Genet 105: 699-707.

20. Galli Z (2005) Molecular identification of commercial apple cultivars with microsatellite markers. HortScience 40: 1974-1977.

21. Dangl GS (2005) Characterization of 14 microsatellite markers for genetic analysis and cultivar identification of Walnut. Journal of the American Society for Horticulture Science 130: 348-354.

22. AI-Dous EK, George B, Al-Mahmoud ME, Al-Jaber MY, Wang Y, et al. (2011) De novo genome sequencing and comparative genomics of date palm (Phoenix dactylifera L). Nat Biotechnol 29: 521-527.

23. Yang M, Zhang X, Liu G, Yin Y, Chen, K, et al. (2010) The complete chloroplast genome sequence of date palm (Phoenix dactylifera L.). PLoS One 5: e12762.

24. Mathew LS, Spannagl M, Al-Malki A, George B, Torres MF, et al. (2014) A first genetic map of date palm (Phoenix dactylifera L) reveals long-range genome structure conservation in the palms. BMC Genomics 15: 285-295.

25. Al-Khalifah NS, Askari E (2003) Molecular phylogeny of date palm (Phoenix dactylifera L.) cultivars from Saudi Arabia by DNA fingerprinting. Theor Appl Genet 107: 1266-1270.

26. Al-Bakr AL (1972) Date palm in past, present, and the new in their cultivation. Al-Ani Press, Baghdad, Iraq.

27. Hussein F, Al-Zaid A (1975) Field Studies on Date Varieties of the Kingdom of Saudi Arabia. Publications of the Ministry of Agriculture and Water Riyadh.

28. Ministry of Agriculture, Kingdom of Saudi Arabia (2006) The famous date varieties in the Kingdom of Saudi Arabia. Vol. First edition (2006), Riyadh, Saudi Arabia: Public Relations and Agricultural Information.

29. Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19: 11-15.

30. Billotte N, Marseillac N, Brottier P, Noyer JL, Jacquemoud-Collet, et al. (2004) Nuclear microsatellite markers for the date palm (Phoenix dactylifera L.): Characterization and utility across the genus Phoenix and in other palm genera. Molecular Ecology Notes 4: 256-258.

31. Hamwieh A, Farah J, Moussally S, Al-Shamaa K, Elmeer K, et al. (2010) Development of 1000 microsatellite markers across the date palm (Phoenix dactylifera L.) genome. Acta Hortic 882: 269-277.

32. Alberto F (2013a) MsatAllele_1.0: an R package to visualize the binning of microsatellite alleles. J Hered 3: 394-397.

33. Alberto F (2013b) MsatAllele1.02 R package notes.

34. Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21: 2128-2129.

35. Raymond M, Rousset F (1995) GENEPOP (ver1.2): a population genetics software for exact test and ecumenicism. J Hered 86: 248-249.

36. Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York.

37. Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) GENETIX 4.05, logiciel sous WindowsTM pour la génétique des populations. Laboratoire Génome, Populations, Interactions. Université de Montpellier II: Montpellier, France.

38. Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19: 395-420.

39. 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.

40. Chakraborty R, Jin L (1993) A unified approach to study hypervariable polymorphisms: statistical considerations of determining relatedness and population distances. DNA Fingerprinting: State of the Science, Basel: Birkhauser Verlag.

41. Saitou N, Nei M (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4: 406-425.

42. Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform 5: 150-163.

43. Wickham H (2007) Reshaping data with the reshape package. J Stat Softw 21: 12.

44. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

45. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14: 2611-2620.

46. Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of structured populations. BMC Genetics 11: 94.

47. Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Research 27: 209-220.

48. Al-Mamari AH (2013) Application of genomics and molecular genetics in date palm (Phoenix dactylifera L.). Nottingham University Sutton Bonington, Campus, Loughborough.

49. Sharma MV, Kantartzi SK, Stewart JM (2009) Molecular diversity and polymorphism information content of selected Gossypium hirsutum accessions. Summaries of Arkansas Cotton Research. AAES Research Series 582: 124-127.

50. Munshi A, Gamal O (2010) Investigation on molecular phylogeny of some date palm (Phoenix dactylifera L.) cultivars by protein, RAPD and ISSR markers in Saudi Arabia. Aust J Crop Sci 4: 23-28.

51. Ojango JM, Mpofu N, Marshall K, Andersson-Eklund L (2011) Quantitative methods to improve the understanding and utilisation of animal genetic resources. International Livestock Research Institute, Swedish University of Agricultural Sciences: Uppsala, Sweden.

52. Kotze A, Muller GH (1994) Genetic relationship in South African cattle breeds, in Proceedings of the 5th world congress on genetics applied to livestock production. University of Guelph: Guelph, Ontario, Canada.

53. Ganino T, Bartolini G, Fabbri A (2006) The classification of olive germplasm—a review. Journal of Horticultural Science and Biotechnology 81: 319-334.

54. El Hadrami A, Daayf F, El Hadrami A (2011) Date palm genetics and breeding. Springer, Heidelberg, New York.

55. Vallejo-Marin M, Lye GC (2013) Hybridisation and genetic diversity in introduced Mimulus (Phrymaceae). Heredity 110: 111-122.

56. Legendre P, Legendre L (1998) Numerical ecology (Developments in Environmental Modelling). Elsevier, Amsterdam, Netherlands.