



Citation: Somayeh Saboori, Masoud Sheidai, Zahra Noormohammadi, Seyed Samih Marashi, Fahimeh Koohdar (2021) Genetic (SSRs) versus morphological differentiation of date palm cultivars: Fst versus Pst estimates. *Caryologia* 74(3): 151-168. doi: 10.36253/caryologia-1089

Received: September 20, 2020

Accepted: September 09, 2021

Published: December 21, 2021

Copyright: © 2021 Somayeh Saboori, Masoud Sheidai, Zahra Noormohammadi, Seyed Samih Marashi, Fahimeh Koohdar. This is an open access, peer-reviewed article published by Firenze University Press (<http://www.fupress.com/caryologia>) and distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

ORCID
ZN: 0000-0003-3890-9001

Genetic (SSRs) versus morphological differentiation of date palm cultivars: Fst versus Pst estimates

SOMAYEH SABOORI¹, MASOUD SHEIDAI², ZAHRA NOORMOHAMMADI^{1,*}, SEYED SAMIH MARASHI³, FAHIMEH KOOHDAR²

¹Department of Biology, Science and Research Branch, Islamic Azad University, Tehran, Iran

²Faculty of Life Sciences and Biotechnology, Shahid Beheshti University, Tehran, Iran

³Date Palm & Tropical Fruits Research Center, Horticultural Sciences Research Institute, Agricultural Research, Education and Extension Organization (AREEO), Ahwaz, Iran

*Corresponding authors. E-mail; marjannm@yahoo.com, z-nouri@srbiau.ac.ir

Abstract. Date Palm (*Phoenix dactylifera* L.) is one of the oldest domesticated fruit trees. For future breeding program, knowledge on genetic structure of cultivars is necessary. Therefore, the present study was performed with the following aims: 1- To provide data on genetic diversity and genetic structure of 36 date palm cultivars, 2- To provide data on the association between fruit characteristics and the genetic features of the cultivars. We used nine SSRs and EST-SSR loci for our genetic investigation. The most of SSR loci obtained have a high G_{st} value (0.70), and therefore have a good discrimination power for date palm cultivar differentiation task. K-Means clustering grouped date palm cultivars either in two broad clusters, or in 16 smaller genetic groups. This was supported by delta K = 2 of the STRUCTURE analysis. AMOVA produced significant genetic difference among date palm cultivars (PhiPT = 0.70, P = 0.001). New genetic differentiation parameters estimated also produced significant difference among date palm cultivars (G_{st} (Nei) = 0.673, P = 0.001; G_{st} (Hed) = 0.738, P = 0.001). Test of assignment revealed that some of the cultivars have 33-66% misassignment, probably due to genetic admixture. Heatmaps of genetic versus morphological and agronomical characters in date palm cultivars differed from each other showing the cultivars morphological changes is not merely related to their genetic content. It points toward the potential role played either by environmental conditions or local selection practice. The new findings can be utilized in future conservation and breeding of date palms in the country.

Keywords: date palm, genetic variability, genetic structure, PST index, population assignment.

INTRODUCTION

Plant species of the family Palmae/ or Arecaceae are distributed mainly in tropical and subtropical areas, but a few species grow at higher latitudes in

the southern hemisphere. The main diversification centers of these taxa are the equatorial coast of Africa, Oceania, the Brazilian coast, the Amazon, Indonesia and the Antilles (Moore & Uhl, 1982).

The palm trees greatly contribute to the economy of the people around the world. Different sort of fruits, seeds, the 'palmito', honeys, 'sagu' (material with starch extracted from the centre of the trunks), different drinks from the sap or the fruits, and crystallized sugar from the sap, are only some of the palm tree products consumed by mankind (Rivas et al. 2012). Among date palm tree species, African oil palm (*Elaeis guineensis*), the coconut tree (*Cocos nucifera*), the date palm (*Phoenix dactylifera*) and the betel nut palm (*Areca catechu*), are considered as the main cultivated plant species. They are cultivated in about 14,585,811, 11,208,072, 1,264,611 and 834,878 hectares respectively (FAO, 2010).

The Date Palm (*Phoenix dactylifera* L.) is one of oldest domesticated fruit trees, which its wild plants records date back to 5000-6000 BC in Iran, Egypt and Pakistan (El Hadrami & El Hadrami, 2009). This important food plant produced about 7,048,089 tons of date only in Algeria, Saudi Arabia, Egypt, the Arab Emirates, Iraq, Iran, Morocco, Oman, Pakistan and Tunis (FAO, 2010).

Successful future development of date palm industry and cultivation depends on proper evaluating, utilizing, and conserving date palm genetic resources, as well as efficient assessment of the present and potential future cultivars (Jaradat, 2014).

One of the main tasks in plant genetic resources investigation is evaluation of available genetic diversity. Genetic diversity of date palm would be studied at different levels, including between cultivars, populations or individual clones, as well as between different geographical regions. Genetic variability may be measured at the morphological, physiological, biochemical or molecular levels (Jaradat, 2014).

The degree and distribution amount of genetic diversity may vary among different oases and populations, due to historical, geographical, ecological and anthropogenic factors (Jaradat, 2014). Mankind can also influence the genetic diversity of date palms by his activities like cultivation practice, social behavior, artificial selection as well as spatiotemporal exchange and movement of germplasm (Jaradat, 2014).

Date palm cultivars are reported to have a common genetic back-ground and therefore, proper differentiation of the cultivars and individual plant assignments in each cultivar is a difficult task and mistakes are inevitable in that. This may also be due to genetic admixture of the date palms (Sharifi et al. 2018, Saboori et al. 2019, 2021 a,b, Gros-Balthazard et al. 2020).

"In general, the question of individual assignment to population samples resulted in the development of different statistical methods distinguishing between resident individuals that are "mis-assigned" (have a genotype that is most likely to occur in a population other than the one in which the individual was sampled) by error from real immigrant individuals (i.e., type I error, Piry et al. 2004). "In assignment investigation, Monte Carlo resampling methods have been proposed to identify a statistical threshold beyond which individuals are likely to be excluded from a given reference population sample. The principle behind these resampling methods is to approximate the distribution of genotype likelihoods in a reference population sample and then compare the likelihood computed for the to-be-assigned individual to that distribution (Piry et al. 2004)".

A combination of stable morphological characters and molecular markers may be used in date palm genetic diversity studies and discrimination among closely related date palm cultivars and clones (Johnson et al. 2015). Different molecular markers (neutral, multilocus and DNA-sequence based markers) have been utilized in date palm genetic diversity investigations as well as cultivar phylogeny analyses (see for example, Sharifi et al. 2018, Saboori et al. 2019, Saboori et al. 2020). Among these molecular markers, the nuclear microsatellite markers (simple sequence repeat, SSRs) are known to be precise and accurate in genetic finger printing of date palm cultivars (Ahmed et al., 2013, Johnson et al. 2015, Zehdi-Azouzi. et al. 2015). Moreover, Zhao et al. (2013) developed several EST-SSR (Expressed sequence tag-SSR) gene based markers to investigate date palm (*Phoenix dactylifera* L.) genetic finger printing. These genetic markers may provide a valuable genetic and genomic tool for further genetic research and varietal development in date palm, such as diversity study, QTL mapping, and molecular breeding.

Date palm comprises one of the most important horticultural crops of Iran which is cultivates in several parts of the country but it is mainly in southern parts of Iran (Fig. 1). They have about 400 date palm cultivars, currently under cultivation. Although domestic date palm identification started by 1960s in Iran, it was basically relied on morphological features. However, recent genetic investigations utilize molecular approaches (Hajia et al. 2015).

The genetic investigations on Iran date palms, are mainly focused on cultivar identification and evaluation, genetic diversity analyses and cultivars relationships, as well as male and female cultivars discrimination (see for example, Hajian, 2007, Marsafari and Mehrabi, 2013, Hasanzadeh Khankahdani and Bagheri, 2019, Saboori et al.

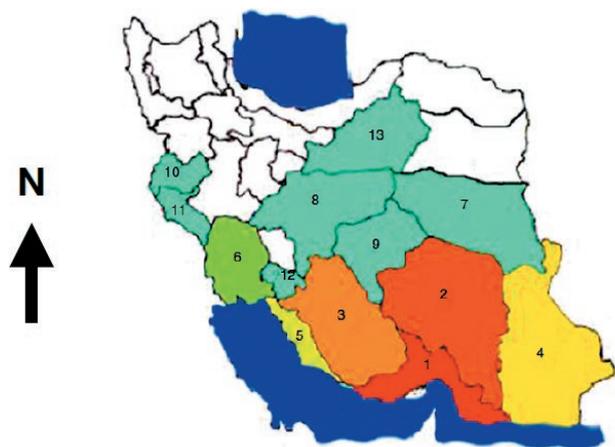


Figure 1. The provinces that are under date palm cultivations in Iran. Numbers 1- 13 are: Hormozgan, Kerma, Fars, Sistan & Baluchestan, Bushehr, Khuzestan, South Khorasan, Isfahan, Yazd, Kermanshah, Eilam, Kohgiluyeh and Boyer-Ahmad, and Seman, respectively (Hajian 2007, Hajian et al. 2015).

2020). However, with regard to 400 date palm cultivars and different geographical areas of their cultivation, we need a lot more detailed genetic studies in these cultivars.

Along with genetic diversity, significant difference in morphological and agronomic characters of date palm cultivars is important for breeding purpose. Q_{ST} , is a quantitative genetic analog of Wright's F_{ST} (Spitze 1993, Prout and Barker 1993). The F_{ST} gives provides a standardized measure of the genetic differentiation among presumed populations, while the Q_{ST} provides the amount of genetic variance among populations relative to the total genetic variance. In fact, the average Q_{ST} of a neutral additive quantitative trait is expected to be equal to the mean value of F_{ST} for neutral genetic loci. The F_{ST} can be readily measured on commonly available genetic markers, and Q_{ST} can be measured by an appropriate breeding design in a common garden setting. Therefore, Q_{ST} is an index of the effect of selection on the quantitative trait. If Q_{ST} is higher than F_{ST} , it is taken as evidence of spatially divergent selection on the studied quantitative trait. If Q_{ST} is much smaller than F_{ST} then this has been taken as evidence of spatially uniform stabilizing selection, which makes the trait diverge less than expected by chance.

According to Leinonen et al. (2006) and Brommer (2011) "when Q_{ST} estimates are not available, P_{ST} can be justified as a substitute." According to Brommer (2011) "divergence across populations of species that are less amenable for proper Q_{ST} estimation may still be of considerable evolutionary or conservation interest" and it can be assessed by using P_{ST} . This in turn estimates the quantitative genetic differentiation (i.e., additive

genetic variance) using quantitative trait measurements within populations (Brommer, 2011). The P_{ST} index assesses the local adaptation through natural selection of wild populations and is an approximation of the quantitative genetic differentiation index (Q_{ST}), obtained in common garden experiments (Gentili et al. 2018).

The relationship between the values of P_{ST} and F_{ST} can be used to estimate the relative importance of genetic processes and selection: (a) $P_{ST} = F_{ST}$ indicates that divergence is compatible with a scenario of genetic drift; (b) $P_{ST} > F_{ST}$ indicates directional selection (i.e., when one extreme phenotype (Gentili et al. 2018).

The quantification of population differentiation based on neutral genetic markers and quantitative traits can highlight the relative role of evolutionary processes such as natural selection, genetic drift and gene flow for patterns of local adaptation (Brommer, 2011; Leinonen et al., 2013).

Fixation index (F_{ST}) is widely used to estimate genetic differentiation with neutral loci (SSR, ISSR, AFLP) by analyzing the variance in allele frequency (Wright, 1965). In contrast, phenotypic differentiation index (P_{ST}) is an estimate of quantitative genetic differentiation (i.e., additive genetic variance) using quantitative trait measurements within populations (e.g., plant size, growth rate, etc.; Brommer, 2011).

MATERIAL AND METHODS

Plant materials and morphological features

We used 36 cultivars including 122 trees were collected from Ahwaz germplasm collection (Omol-tomair station of Date Palm & Tropical Fruits Research Center, Ahwaz, Iran) and different date palm orchards located in Hormozgan and kerman provinces, Iran (Saboori et al. 2019, Saboori et al. 2020).

The fruit characters were used based on Saboori et al. 2020. They were including weight of fruit and seed, length and width of fruit, length, and width of the seed.

EST-SSR and SSR markers

Genomic DNA of fresh leaves were extracted from date palm cultivars collected by modified CTAB protocol (Saboori et al. 2020). For genetic investigation we used three EST-SSR and six SSR loci. Two primers EST-PDG3119-rubisco and EST-DPG0633-Laccase were selected (Zhao et al., 2013), while EST-GTE primer was designed by Primer3 and Gene Runner software. They were then checked for accuracy by BLAST algorithm.

Six primers MPdCIR078, MPdCIR085, PdCUC3-ssr2, MPdCIR090, MPdCIR048 and MPdCIR025 were selected for SSR marker (Bodian et al, 2014). The sequences of primers of EST-SSR and SSR markers are listed in Table S1.

PCR reaction for EST-SSR and SSR loci were performed as following; a 25 μ L volume containing 20 ng genomic DNA and 5 U of *Taq* DNA polymerase (Bioron, Germany), 2X PCR buffer (50 mM KCl; 10 mM Tris-HCl, pH:8), 1.5mM MgCl₂; 0.2 mM of each dNTP (Bioron, Germany); 0.2 μ M of each primer.

The PCR program for EST-SSR and SSR markers were followed:

The reactions for EST-SSR were amplified in T100 thermal cycler (BioRad, USA) using the following procedure, 5 min at 94 °C, 35 cycles of 30 sec at 95 °C, 30 sec at 50-60 °C (EST-PDG3119-rubisco 50 °C, EST-GTE 52 °C, EST-DPG0633- Laccase 60 °C) and 1 min and 30 sec at 72 °C followed by 5 min at 72 °C as final extension.

The PCR program for SSR markers were performed as touch-up PCR; 94°C for 5 min, initial 10 cycles at 95°C for 30 sec, annealing step (MPdCIR078 51°C, MPdCIR085 47.5 °C, PdCUC3-ssr2 62 °C , MPdCIR090 47.5 °C , MPdCIR048 46.9 °C , MPdCIR025 45 °C)for 1 min, 72°C for 1 min and 30 sec, followed by 40 cycles at 95°C for 30 sec, annealing step (MPdCIR078 52°C, MPdCIR085 49.9 °C, PdCUC3-ssr2 65 °C , MPdCIR090 49.9 °C , MPdCIR048 48.8 °C , MPdCIR025 48 °C) for 1 min, 72°C for 1 min and 30 sec, a final cycle of 72 °C for 15 min. The PCR amplifications were separated on a 12% PAGE (poly acrylamide gel electrophoresis) with a 100-kb gene ruler (Parstous, Iran).

Data analyses

Genetic diversity analyses

The SSR and EST-SSR bands obtained were treated as binary characters (Podani 2000) and used for further analyses. DCA (Dentrented correspondance analysis) was used to evaluate suitability of SSR and EST-SSR bands obtained. Discriminant power of the bands obtained was determined by POPGENE program. Genetic diversity parameters in the date palm cultivars were estimated by GeneAlex 4.2. A heat map was produced on these parameters by R package.

Genetic grouping of the cultivars

In order to find the proper number of genetic groups within date palm studied, we followed two different sta-

tistical approaches. 1- We used K-Means clustering as performed in Genodive program, which is based on likelihood method. 2- Delta K was obtained from STRUTURE analysis which is a Bayesian-based method. Details of these methods are according to Sharifi et al. (2018).

GenoDive provides two different statistics that can determine the number of clusters. These are pseudo-F-statistic; (the optimal clustering is the one with the highest value for the pseudo-f statistic), and the Bayesian Information Criterion (BIC, calculated using sum of squares and the optimal clustering is the one with the lowest value) (Meirmans2020). Both these criteria work well for clustering populations and individuals, especially when there is random mating within populations but BIC has the benefit that it can be used to determine whether there actually is any population structure at all (Meirmans 2020).

Based on the number of Ks obtained we performed Ward clustering as performed in PAST and STRUTURE analysis as implemented in STRUCTURE program.

The genetic differentiation of the studied cultivars was determined by AMOVA as performed in GeneAlex, as well as by Gst- Nei and Gst-Hederick as performed in Genodive.

Correlation between morphological characters studied was determined by Pearson coefficient of correlation. In order to compare groups of the cultivars based on both molecular and morphological characters, heat maps were constructed by related commands in R package.

Population assignment was performed by two different methods: 1- By discriminant analysis (DA) as performed in SPSS program. In this analysis a summary table was produced which indicates relatedness of each case to its presumed population, and finally provide a percentage value for each population membership based on likelihood method. 2- By using Assignment test in GeneAlex, which is also based on likelihood method and provides a total membership percentage for all data in question and also provide pairwise populations graph.

Phenotypic versus genetic differentiation

PST index was used to estimate the role of local adaptation through natural selection in date palm populations, compared to that of genetic differentiation. For each population pair, pairwise PST values were calculated for each trait (and for an average PST), using the following formula:

$$PST = c\sigma^2_B / (c\sigma^2_B + 2h^2\sigma^2_W)$$

In this formula, δ_B^2 and δ_W^2 are between-population and within population variance components for a trait, respectively; h^2 expresses the heritability (the proportion of phenotypic variance that is due to additive genetic effects); the scalar c expresses the proportion of the total variance that is presumed to be due to additive genetic variance across populations (Brooker, 2011; Leinonen et al., 2013).

In the wild, the estimation of the additive genetic variance components is challenging as breeding design is impossible. Therefore, Q_{ST} is often approximated by P_{ST} (Leinonen et al., 2006), which is directly calculated from the total phenotypic variance components with no distinction between the relative contribution of genetic and environmental variations. Therefore, the phenotypic divergence between populations was estimated by the parameter P_{ST} as follows:

$$P_{ST} = \frac{c\sigma_b^2}{c\sigma_b^2 + 2h^2\sigma_w^2} \quad \text{or} \quad P_{ST} = \frac{\frac{c}{h^2}\sigma_b^2}{\frac{c}{h^2}\sigma_b^2 + 2\sigma_w^2},$$

In this formula, δ_B^2 and δ_W^2 are the respective phenotypic variances between and within populations, c is an estimate of the proportion of the total variance due to additive genetic effects across populations, and h^2 is heritability, the proportion of phenotypic variance due to additive genetic effects (Brommer, 2011). In present

study P_{ST} was estimated by P_{stat} of R package (Da Silva and Da Silva, 2018).

RESULTS

SSR and EST-SSR analyses

We obtained in total 40 SSR bands in 122 date palm trees studied. The lowest number of bands (13) occurred in cultivar "Wardi" (male, No. 32), while the highest number of bands was observed in cultivars Halili (No. 4), Male (No. 11), Khezrawi (No. 17), and Barhi (No. 20). The cultivars investigated did not have private band.

The suitability of SSR and EST-SSR bands for date palm population genetic studies was determined by DCA plot (Fig. 2). The plot shows a well-scattered distribution of SSR loci, which indicated that these loci are from different regions of the genome and are not clustered to each other. Such loci are useful in genetic diversity analyses of the populations.

Discriminating power of SSR and EST-SSR bands versus migration (Nm) is provided in Table S2. The result shows that most of SSR loci obtained have a high G_{st} value (0.70), and therefore have a good discrimination power for date palm cultivar differentiation task. This is also evidenced with the high mean G_{st} value = 0.81 obtained.

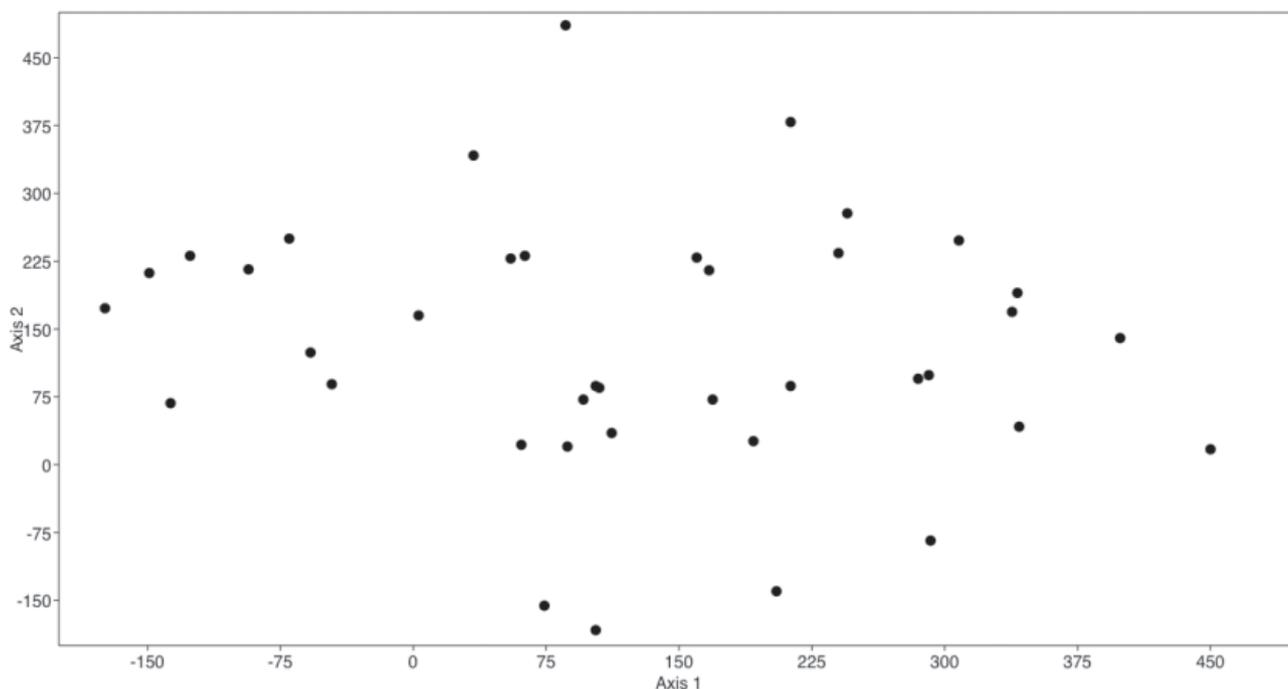


Figure 2. DCA plot of SSR and EST-SSR bands/loci in date palm cultivars showing well-scattered distribution of loci obtained.

Genetic diversity of Date palm cultivars

Data with regard to genetic diversity parameters determined in 122 individual trees of 36 date palm cultivars are presented in Table S3.

The range of polymorphism percentage varied from 2.5 in cultivar Kharook (No. 13), to 25 in cultivar Khadhrawi (No. 17). The mean value for polymorphism

was 13.07%. Usually, date palm cultivars show similar genetic contents, and therefore, about 13% genetic polymorphism is yet appreciable for further breeding studies if accompanied by some degree of morphological and agronomical desirable traits variation.

Heat-map constructed based on genetic diversity parameters (Fig. 3), reveals that based on percentage of genetic polymorphism (P), Nei' gene diversity (He) and

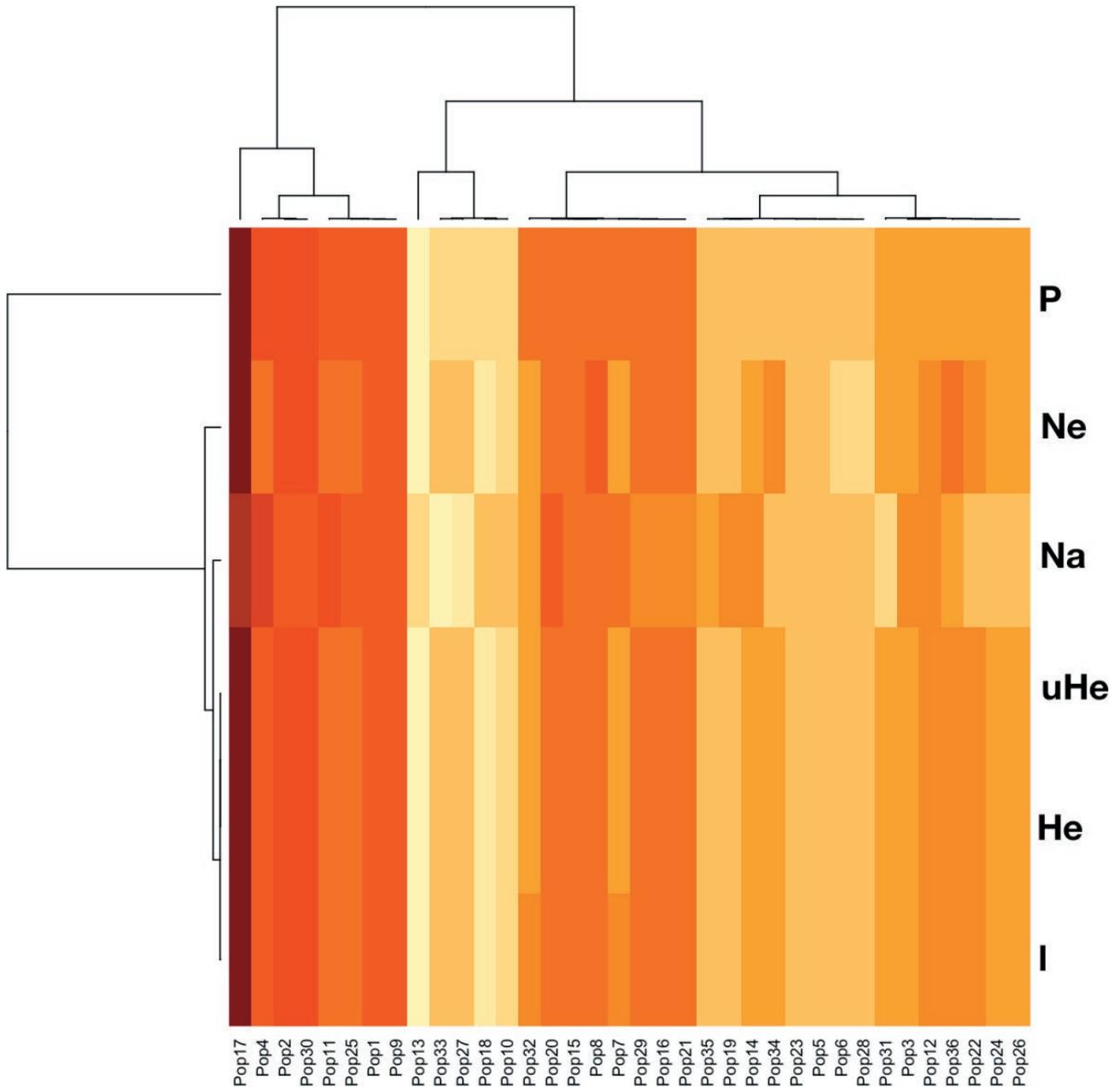


Figure 3. Heatmap of date palm cultivars based on genetic diversity parameters. Abbreviations: Na = No. of different alleles, Ne = No. of Effective alleles, I = Shan Information. index, He = Expected Heterozygosity, uHe = Unbiased Expected Heterozygosity, and P% = Polymorphism percentage.

Shanon Information Index (I), date palms may be classified in 5 or 6 genetic groups. This classification is sharper by considering only genetic polymorphism parameter.

Grouping of the cultivars

The Nei genetic distance determined in the cultivars studied varied from 0.067 between cultivars 1 and 2, to 0.46 between cultivars Estameran (No. 19) and Mashtoom (No. 28). These low values of genetic distance, indicates a high degree of genetic alikeness in date palm cultivars cultivated in the country.

For grouping of the cultivars based on SSR markers, we first performed K-Means clustering by Genodive program (Table S4). The results indicated that these cultivars can be grouped either in two broad clusters according to Calinski & Harabasz' pseudo-F: $k = 2$, or in 16 smaller genetic groups according to Bayesian Information Criterion: $k = 16$.

Ward clustering of the date palm cultivars based on SSR and EST-SSR data (Fig. 4), also grouped the genotypes in two major clusters and about 16 sub-clusters which is in agreement with K-Means clustering.

WARD dendrogram produced two main clusters or genetic groups in accord with K-Means clustering result. The cultivars 1-13 comprise the first genetic group and

form the first main cluster, while the other cultivars form the second major cluster or genetic group.

In the first main cluster, the cultivars are distributed in three sub-clusters A-C. Replicates of the cultivars 1-4 show a higher level of genetic similarity and are placed in a single sub-cluster, (A). Replicates of the cultivars 9-13 comprise the second sub-cluster B, while replicates of the cultivar 5-9 form the sub-cluster C. Replicates of the cultivar 4, were admixed in two sub-clusters A and C. Few date palm plants of these cultivars also show some degree of admixture.

Since clustering is based on distance parameter only, we also tried STRUCTURE analysis for genotype grouping, which is a Bayesian-based method. For this, we first obtained K value by Evanno method, which produced delta K = 2. This is in agreement with major growing obtained by K-Means clustering. However, to obtain a better and more detailed picture on the cultivars genetic grouping, we carried out STRUCTURE analysis based on K values 2-5 (Fig. 5). The best genetic grouping obtained seems to be K = 5.

Based on K = 5, the cultivars 1-4 show genetic affinity and comprise the first genetic group. This is followed by the cultivars 5-13, then 14-2, 23-30, and finally the cultivars 31-36, form the fifth genetic group. All these five genetic groups show a low degree of genetic admixture with the other groups.

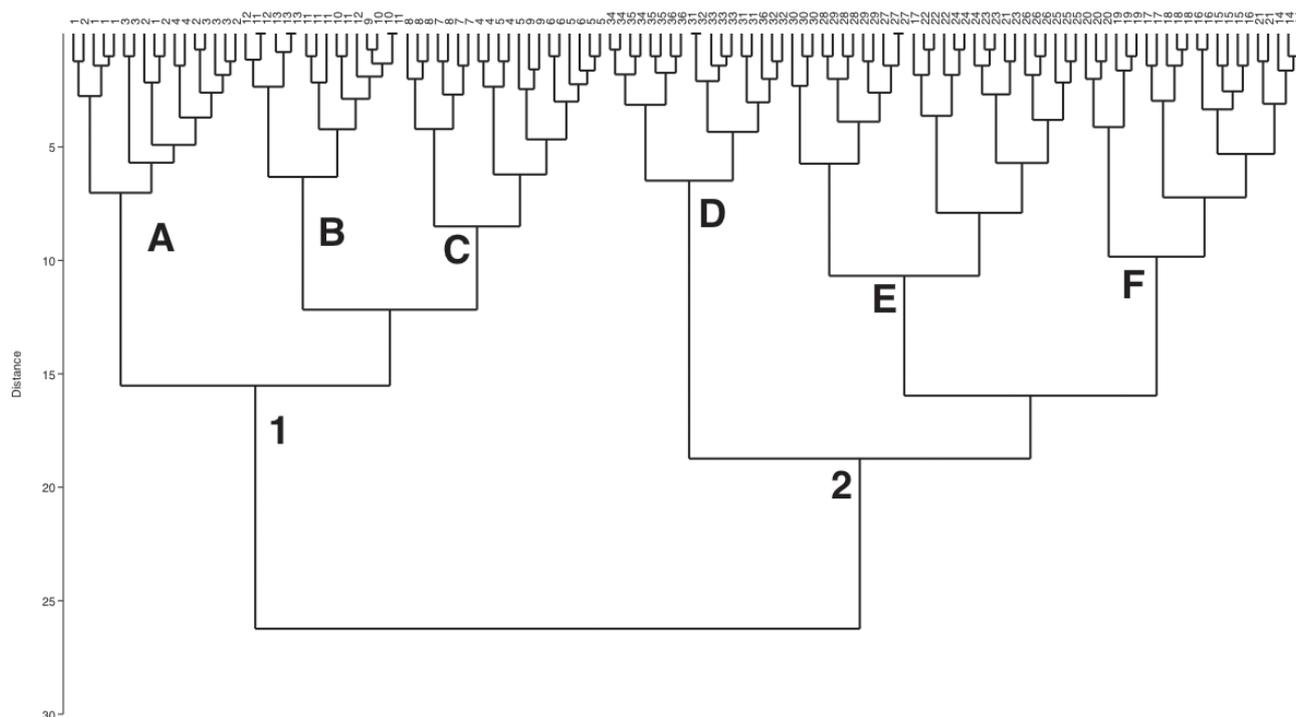


Figure 4. Ward dendrogram of date palm cultivars based on SSR and EST-SSR data.

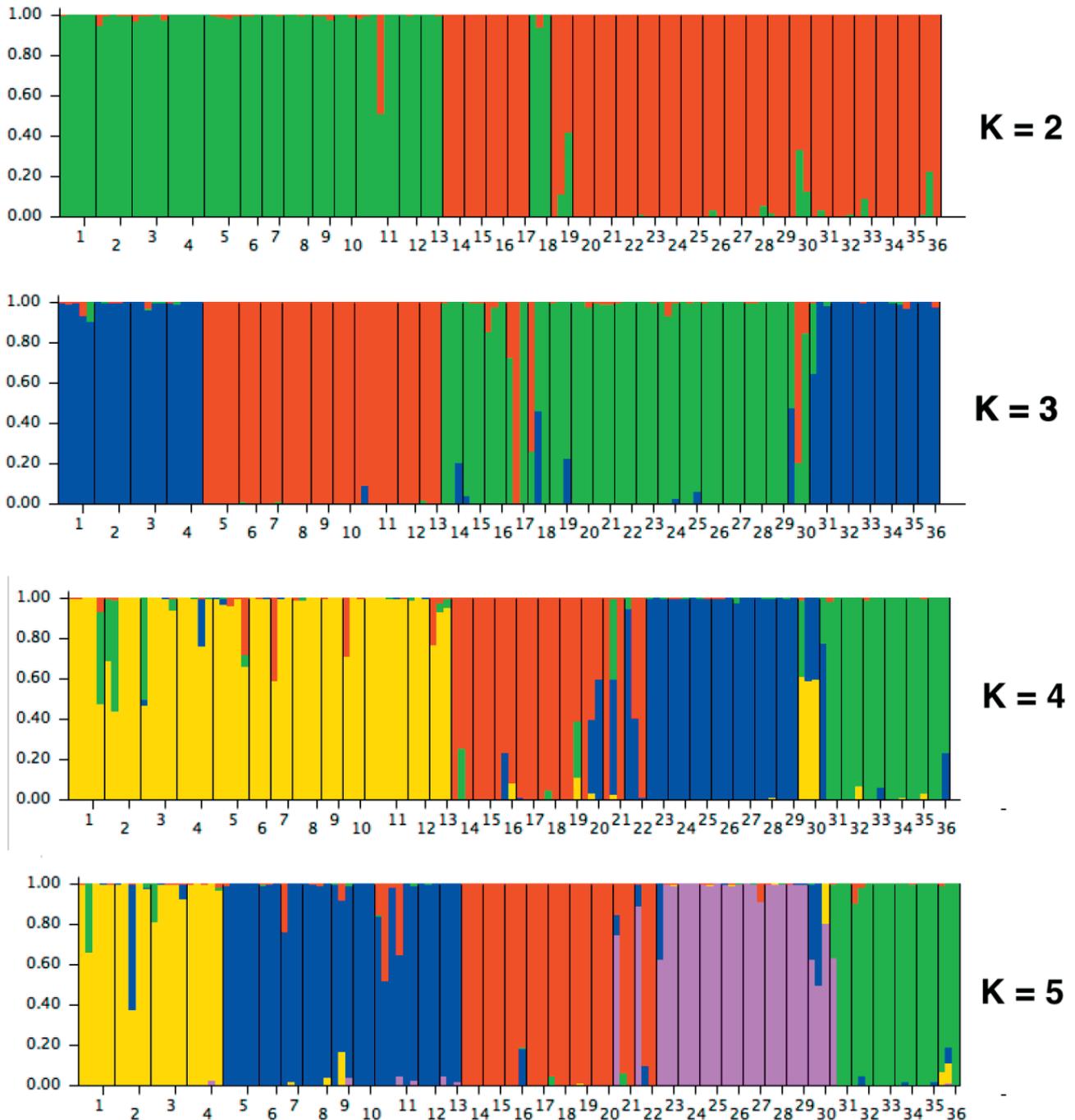


Figure 5. STRUCTURE plot of date palms studied based on K = 2-5.

Genetic difference of the cultivars

AMOVA produced significant genetic difference among date palm cultivars ($\Phi_{PT} = 0.70$, $P = 0.001$). It also revealed that 70% of total genetic variability occurs due to among cultivar difference, while 31% occurs due to within population genetic variability. Moreover, pair-

wise AMOVA (Table S5) produced significant genetic difference between the cultivars of the two main clusters as well as the cultivars of different sub-clusters in UPGMA dendrogram. New genetic differentiation parameters estimated produced significant difference among date palm cultivars ($G'_{st}(\text{Nei}) = 0.673$, $P = 0.001$; $G'_{st}(\text{Hed}) = 0.738$, $P = 0.001$). These results indicate the presence of genetic

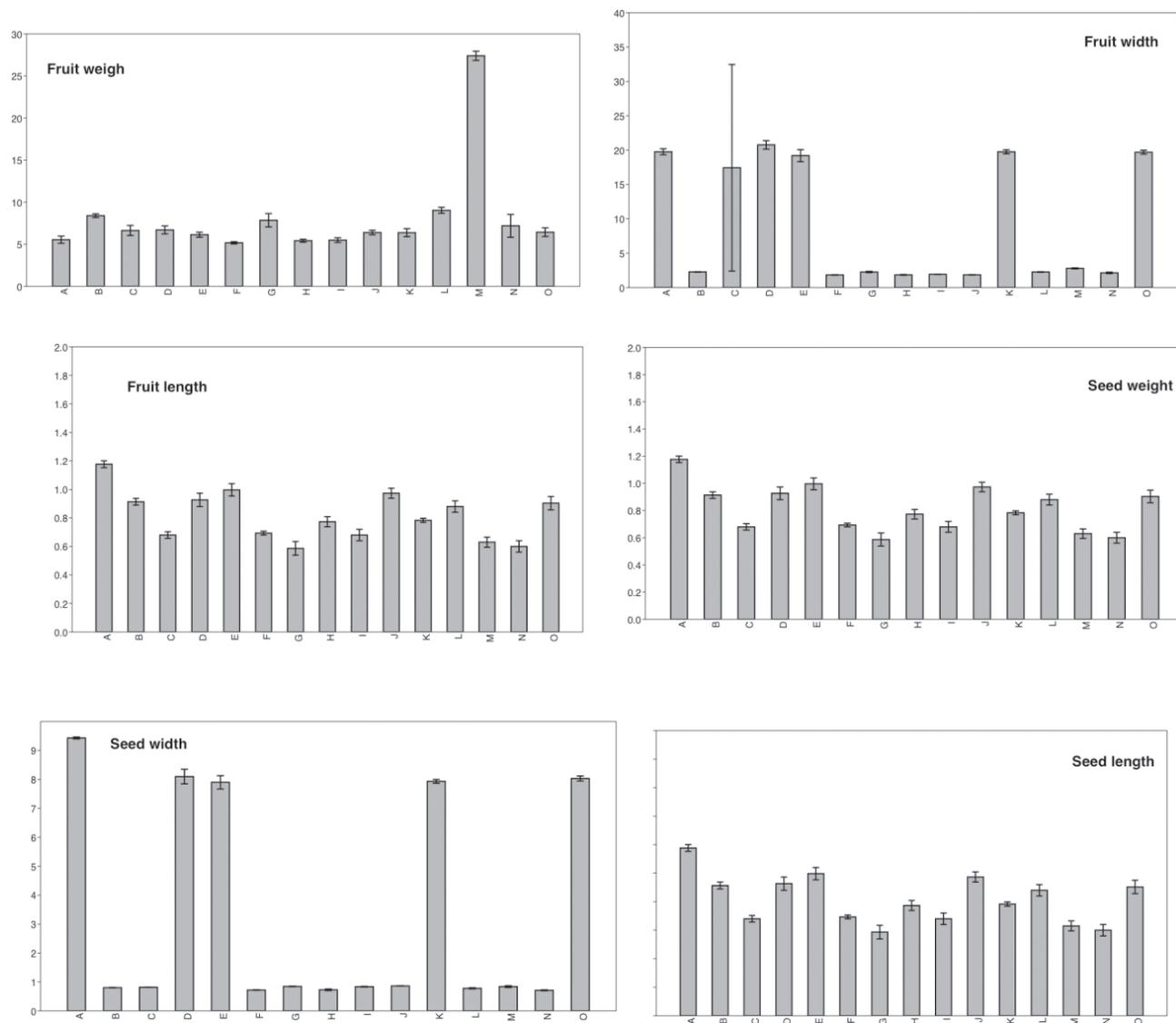


Figure 6. ANOVA of morphological characters studied in date palm cultivars.

variability within date palm cultivar germplasm, which can be used in future breeding program.

Assignment of date palms

Assignment of individual date palm plants by discriminant analysis revealed that the cultivars 2, 9, 10, 13, 20, 21, 25, 29, 31 and 32, have 33% mis-assignment, while cultivar 4 has 66% mis-assignment. GeneAlex also revealed 67% self population assignment and 33% of other population assignment.

In Table S6, some parts of assignment result for 122 date palms have been given (only those samples inferred to be from other population are given). Assignment is

based on positive likelihood, and therefore the lower the value shows the correct assignment (inferred population).

Fst versus Pst estimates

Details of morphological characters studied are given in Fig. 6. ANOVA produced significant difference ($P < 0.01$), for these characters among the studied cultivars.

Most of these characters show significant correlation ($P < 0.01$) (see for example, Fig. 7).

Heat-maps of the 45 date palm trees based on morphological versus genetic (SSRs), data are presented in Fig. 8. Comparison of the groupings obtained reveals difference in the clustering results.

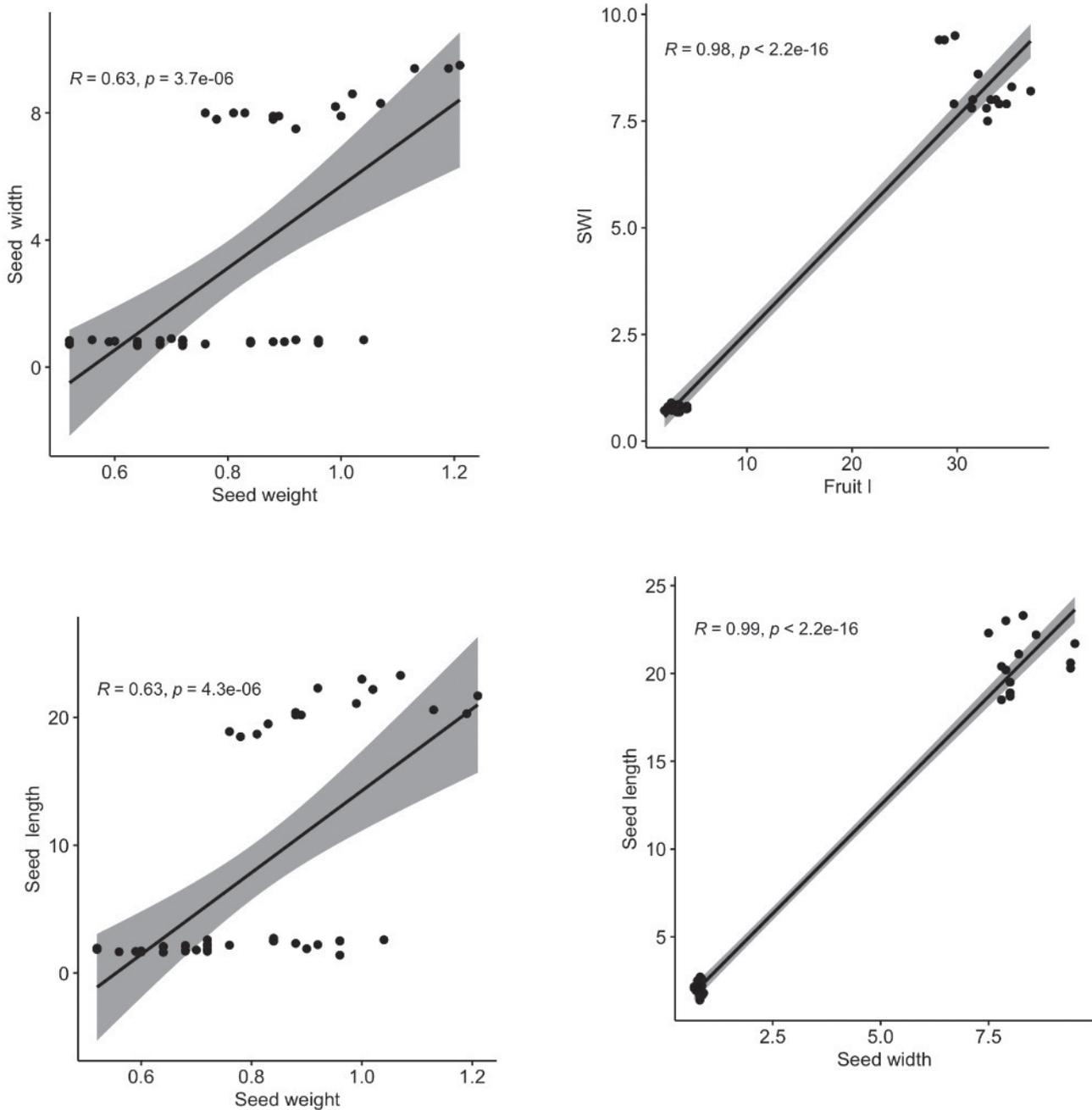


Figure 7. Representative Pearson coefficient of correlation among morphological characters studied in date palm cultivars.

Moreover, the Mantel test performed between the two clustering results did not produced a significant association between the two markers ($r = 0.057, P = 0.16$), supporting the heat maps. Therefore, grouping and cultivar relationship illustrated by morphological characters studied do not accord with genetic relationship of the same date palm cultivars.

Fst versus Pst analyses, revealed that in most of the studied morphological characters, the Pst value greatly

exceeds that of genetic Fst value. For example, some of the pair-wise comparison between cultivar No. 3 and the others are provided in Table S7.

Therefore, $PST > FST$ indicates directional selection in quantitative fruit and seed characteristics has been occurred in the studied date palm cultivars. Different factors may be responsible for these directional changes, like ecological and environmental conditions in which the cultivars grow, selection practiced by the breeders

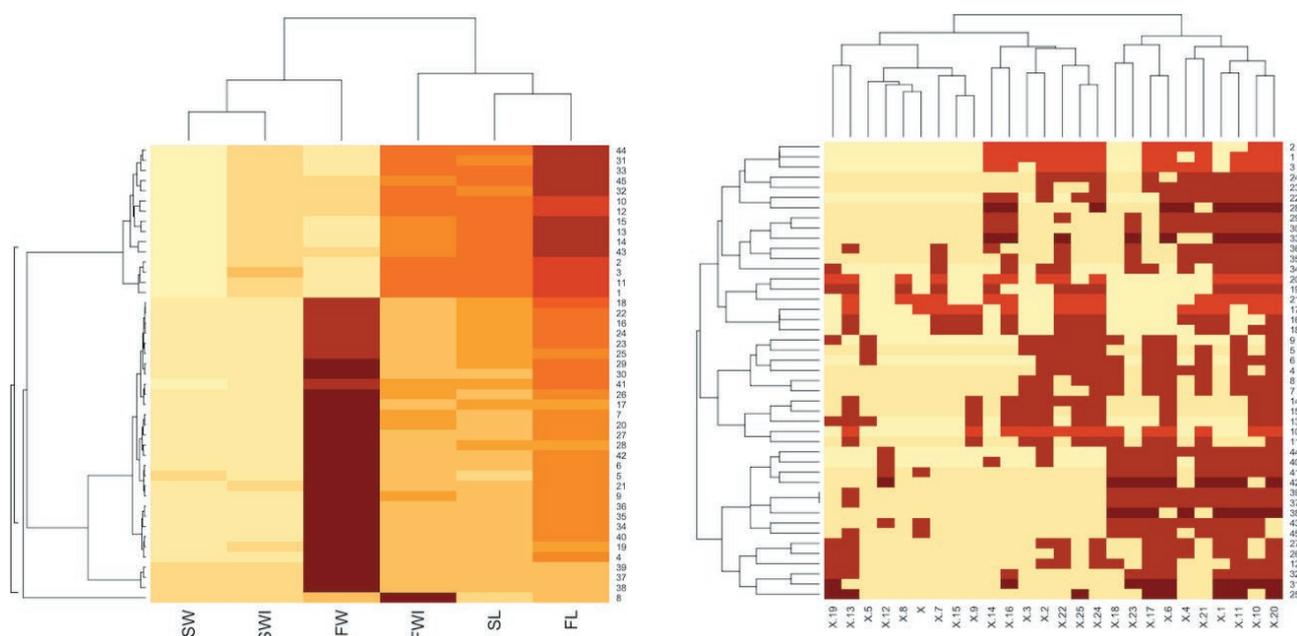


Figure 8. Heat maps of 45 date palm cultivars based on morphological and genetic (SSRs) data, showing different groupings of these cultivars. (Abbreviations in morphological heat map are: SW = Seed weight, SWI = Seed width, FW = Fruit weigh, SL = Seed. length, and. FL = Fruit length).

or locals, etc. In general, morphological difference along with genetic diversity present in the studied cultivars may contribute in future breeding of date palm.

DISCUSSION

Genetic diversity

Present study revealed the presence of a low to moderate genetic diversity within date palm cultivars studied. This is in accord with the studies performed in Iraq and Tunesian date palms by Jubrael et al. (2005) and Zehdi et al. (2015), who suggested a common genetic basis among date palm genotypes despite the differences in fruit characters and tree morphology. Low genetic diversity within date palm germplasm was revealed but both neutral molecular markers like, ISSRs and SSRs (see for example, Sharifi et al. 2018, Saboori et al. 2020), and sequence-based marker, like chloroplast DNA (Sharifi et al. 2018).

Cultivars genetic grouping

The cultivars studied were placed in two major genetic groups by both K-Means and Bayesian-based delta K estimation more detailed analysis, revealed that

they can be classified in 5 different genetic groups. Such data may be used in future breeding program. Cultivar grouping based on STRUCTURE analysis were also utilized by the other researchers in date palms (see for example, Sharifi et al. 2018). It is important in plants with almost common genetic background like date palms to classify them in different genetic classes.

Population assignment

Population assignment seems to be a prerequisite step in selecting plant individuals and breeding date palm, as these plants have a common genetic background and show overlapping genetic structure. This may also happen due to genetic admixture of the date palms (Sharifi et. al. 2018, Saboori et al. 2020). We obtained about 33% of incorrectly assigned date palms in respect to their presumed populations. This may be either due to improper plant sampling or identification within the germplasm, or due to gene flow and admixture among these cultivars. In any case, such cases should be considered in future breeding program.

In a similar study concerned with genetic structure of Tunesian date palms, Zehdi et al. (2015) reported the presence of admixed cultivars too. They considered that the gene flows between eastern and western origins mostly from east to west following a human-mediated

diffusion of the species, is the reason for the formation of mixed genotypes.

Saboori et al. (2020), investigated the genetic structure of 13 date palm cultivars by SCoT molecular markers and reported some degree of genetic admixture among the cultivars. Though, they did not study specifically assignment of the plants to their populations, by looking at the clustering result of their samples, it becomes evident that some of the plants a presumed cultivar has been placed intermixed with plants of another cultivar. However, Sharifi et al (2018) investigated the gene flow and assignment in 16 date palm cultivars by using ISSR molecular markers and observed some degree of population admixture and few cases of incorrectly assigned date palms.

In an elaborate and precisely studied report by Gros-Balthazard et al. (2020), they used a joint ethnographic study and genetic analysis of date palms to test whether named date palm types are true-to-type cultivars versus incorrectly assigned samples in desert nearby Siwa (also known as “feral” in Battesti in Egypt). They recognized three categories of genotypes within their extensive collection namely, true-to-type cultivar samples, ethnovarieties and samples of local categories. Therefore, there is a huge mistake in assigning date palms to their respective population or named cultivar.

Genetic versus phenotypic differential

Aljuhani (2016), studied the degree of dissimilarity and the impact of location on the genetic relationship between local cultivars in Saudi Arabia by using and twenty-four nuclear microsatellite loci. He reported a high level of genetic polymorphism in some of the loci, and could differentiate the studied cultivars by these markers. Some of these cultivars were grouped according to their geographical area in which they were cultivated. We obtained a higher value for Pst versus Fst, almost in all date palm cultivars studied and for most of the fruit and seed characters. The Pst is taken as index for morphological local adaptation through natural selection, but influenced by environment (Brommer, 2011). If $Pst = Fst$, it indicates that divergence is due to genetic drift; and if $Pst > Fst$, it indicates the role of directional selection (i.e., when one extreme phenotype is favored over other ones) among populations; and finally, if $Pst < Fst$, it indicates that the same phenotypes are favored in different populations due to stabilizing selection. We may therefore, suggest that, due to some local environmental face or local practice of cultivation or selection, some adaptive changes have occurred in date palm cultivars in the country. QST–FST compari-

son has shown that trait divergence due to natural selection, as opposed to genetic drift have occurred in many taxa (Leinonen et al. 2013).

In present study, the Mantel test did not produce significant association between the cultivar grouping and morphological grouping, in other words we did not see co-variation between genetic and morphological traits. However, in Qst-Fst investigation carried out by S`urinova` et al. (2018), in 11 populations of *Festuca rubra*, they reported the existence of adaptive differentiation in phenotypic traits and their plasticity across the climatic gradient and observed statistically significant co-variation between markers and phenotypic traits, which is likely caused by isolation by adaptation.

In a similar study, Caré et al. (2018) investigated the high morphological differentiation in crown architecture in contrasts with low population genetic structure of German Norway Spruce Stands by using Pst-Fst method and 11 nuclear SSR molecular markers.

Norway spruce trees have narrow crown phenotypes, whereas lowland trees have broader crowns. Narrow crown phenotypes are likely the result of adaptation to heavy snow loads combined with high wind speeds. They observed a high differentiation of morphological traits ($Pst = 0.952–0.989$) between the neighboring autochthonous and allochthonous stands of similar age contrasts with the very low neutral genetic differentiation ($Fst = 0.002–0.007$; $G^{*}st = 0.002–0.030$), suggesting that directional selection at adaptive gene loci was involved in phenotypic differentiation.

It has been suggested that “the QST–FST method is still underused in ‘omics’ contexts, in which it may be useful for identifying evolutionary significance in large data sets in the absence of evolutionary models (Leinonen et al. 2013)”.

In conclusion we may say that considering different molecular studies in date palm genotypes both around the world and in our country, and irrespective of molecular marker used (neutral versus sequence based markers), a low to moderate genetic diversity is present in limited number of cultivars investigated till now. We need to carry one further detailed population genetics analysis in much more number of accessions and cultivars to possibly broaden the genetic variability of date palm for future breeding.

AUTHORS’ CONTRIBUTIONS

Z.N. and M.Sh: conceptualization of the project; M.Sh.: analyses of data; S.S and F.K data collection and lab work; S.M.: providing samples

ACKNOWLEDGEMENTS

We acknowledge Science and Research Branch, Islamic Azad University for providing laboratory. We thank the Iran National Science Foundation (INSF), for partial financial support of this project (No.97010700).

REFERENCES

- Ahmed J, Al-Jasass FM, Siddiq M (2014) Date Fruit Composition and Nutrition. In: Siddiq M, Aleid S, Kader A (Eds) Dates: Postharvest Science, Processing Technology and Health Benefit. Wiley Blackwell, Chichester, UK, pp 261-283
- Aljuhani W S (2016) Genetic Diversity and the Impact of Geographical Location on the Relationships Between *Phoenix dactylifera* L. Germplasm Grown in Saudi Arabia. *Hereditary Genet* 5(3):1-11. <https://doi.org/10.4172/2161-1041.1000172>
- Bodian A, Nachtigall M, Frese L, Elhoumaizi M A, Hasnaoui A, Ndir KN, Sané D (2014) Genetic diversity analysis of date palm (*Phoenix dactylifera* L.) cultivars from Morocco using SSR markers. *J Biodivers Biopros Dev* 1(3): 2376-0214
- Brommer JE (2011) Whither PST? The approximation of QST by PST in evolutionary and conservation biology. *J Evol Biol* 24:1160-1168. <https://doi.org/10.1111/j.1420-9101.2011.02268.x>
- Caré O, Müller, M, Vornam, B, Höltnen, A, Kahlert K, Krutovsky K V, Gailing, O, Leinemann, L (2018) High Morphological Differentiation in Crown Architecture Contrasts with Low Population Genetic Structure of German Norway Spruce Stands. *Forests* 9: 752. <https://doi.org/10.3390/f9120752>
- Da Silva B, Anne Da Silva A (2018) Pstat: An R Package to Assess Population Differentiation in Phenotypic Traits by Stéphane. *The R Journal* 10(1): 447-454
- El hadrami I, El hadrami A (2009) Breeding date palm. In: Jain SM, Priyadarshan PM (eds) Breeding plantation tree crops: tropical species, Berlin, Springer pp191-216
- FAO (2010) Statistics Division. <<http://faostat.fao.org/default.aspx>>. Online. Accessed: Dez. 01, 2010.
- Gros-Balthazard M, Battesti V, Ivorra S, Paradis L, Aberlenc F, Zango O, Zehdi S, Moussouni S, Abbas Naqvi S, Newton C, Terral JF (2020) Integration of ethno botany and population genetics uncovers the agrobiodiversity of date palms of *Siwa Oasis* (Egypt) and their importance to the evolutionary history of the species. *Evol Appl*. <https://doi.org/10.1111/eva.12930>
- Gentili R, Solari A, Diekmann M, Duprè C, Monti GS, Armiraglio S, Assini S, Sandra Citterio S (2018) Genetic differentiation, local adaptation and phenotypic plasticity in fragmented populations of a rare forest herb. *Peer J* 6:e4929 <https://doi.org/10.7717/peerj.4929>.
- Hajia S, Hamidi-Esfahani Z (2015) Date Palm Status and Perspective in Iran. In: Al-Khayri J, Jain S, Johnson D. (eds) Date Palm Genetic Resources and Utilization, Springer, Dordrecht
- Hajian S (2007) Quantity and Quality Comparison of Offshoot and Tissue Cultured Barhee Date Palm Trees. *Acta Hort* 736:293-297. <https://doi.org/10.17660/ActaHortic.2007.736.27>
- Hassanzadeh Khankahdani H, Bagheri A (2019) Identification of Genetic Variation of Male and Female Date Palm (*Phoenix dactylifera* L.) Cultivars Using Morphological and Molecular Markers. *Int J Hort Sci* 6(1): 63-76. <https://doi.org/10.22059/ijhst.2019.276013.278>
- Jaradat A A (2014) Synthesis and assessment of date palm genetic diversity studies. *Emir J Food Agric* 26 (11): 934-952. <https://doi.org/10.9755/ejfa.v26i11.18977>
- Jubrael JMS, Udupa S, Baum M (2005) Assessment of AFLP based genetic relationships among date palm (*Phoenix dactylifera* L.) varieties of Iraq. *J Amer Soc Hort Sci* 130:442-447.
- Johnson, DV, Al-Khayri JM, Mohan Jain S (2015) Introduction: Date Production Status and Prospects in Africa and the Americas. In: Al-Khayri JM, Jain S M, Johnson D V. (Eds.) Date Palm Genetic Resources and Utilization, 3 rd. Africa and the Americas, pp: 3-18
- Leinonen T, McCairns RJS, O'Hara RB, Merilä J (2013) QST/FST comparisons: evolutionary and ecological insights from genomic heterogeneity. *Nature Reviews Genetics* 14:179190. <https://doi.org/10.1038/nrg3395>
- Leinonen T, Cano J M, Makinen H, Merilä J (2006) Contrasting patterns of body shape and neutral genetic divergence in marine and lake populations of three spine sticklebacks. *J Evol Biol* 19:1803-1812. <https://doi.org/10.1111/j.1420-9101.2006.01182.x>.
- Meirmans PG (2020) GENODIVE version 3.0: Easy-to-use software for the analysis of genetic data of diploids and polyploids. *Mol Ecol Resour* <https://doi.org/10.1111/1755-0998.13145>
- MOORE H, UHL N (2011) Major trends of evolution in Palm. *Botanical Review*. <http://www.springerlink.com/content/t637574148146075/>. Accessed 04 May 2011
- Marsafari M, Mehrabi AA (2013) Molecular identification and genetic diversity of Iranian date palm (*Phoenix dactylifera* L.) cultivars using ISSR and RAPD markers. *AJCS* 7(8):1160-1166

- Prout T, Barker J S F (1993) *F* statistics in *Drosophila buzzatii*: selection, population size and inbreeding. *Genetics* 134: 369–375
- Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L, Estoup A (2004) GENECLASS2: A Software for Genetic Assignment and First-Generation Migrant Detection. *J HERED* 95(6):536–539. <https://doi.org/10.1093/jhered/esh074>.
- Podani J (2000) Introduction to the Exploration of Multivariate Data [English translation], Leide, Netherlands: Backhuyes.
- Rivas M, Barbieri R L, da Maia C (2012) Plant breeding and in situ utilization of palm trees. *Ciência Rural*, Santa Maria 42: 261–269.
- Saboori S, Noormohammadi Z, Sheidai M, Marashi S (2019) SCoT molecular markers and genetic fingerprinting of date palm (*Phoenix dactylifera* L.) cultivars. *Genet Resour Crop Evol* 67(1): 73–82. <https://doi.org/10.1007/s10722-019-00854-x>
- Saboori S, Noormohammadi Z, Sheidai M, Marashi SS (2021 a) Date Palm (*Phoenix dactylifera* L.) Cultivar Relationships Based on Chloroplast Genotyping. *Iranian J Sci Technol Transact A: Sci* 45: 833–840.
- Saboori S, Noormohammadi Z, Sheidai M, Marashi SS (2021 b) Insight into date palm diversity: Genetic and morphological investigations *Plant Mol Biol Rep* 39: 137–145
- Sharifi M, Sheidai M, Koozdar F (2018) Genetic fingerprinting of date palm (*Phoenix dactylifera* L.) by using ISSR and cpDNA sequences. *Indian J Genet* 78: 507–514. <https://doi.org/10.31742/IJGPB.78.4.13>
- Stojanova B, Šurinová M, Klápště J, Kolářiková V, Hadincová V, Münzbergová Z (2018) Adaptive differentiation of *Festuca rubra* along a climate gradient revealed by molecular markers and quantitative traits. *PLoS ONE* 13(4): e0194670. <https://doi.org/10.1371/journal.pone.0194670>
- Spitze K (1993) Population structure in *Daphnia obtusa*: quantitative genetic and allozymic variation. *Genetics* 135: 367–374.
- Zehdi-Azouzi S, Cherif E, Moussouni S, Gros-Balthazard M, Naqvi S A, Ludeña B, Castillo K, Chabrillange N, Bouguedoura N, Bennaceur M et al. (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. <https://doi.org/10.1093/aob/mcv068>
- Wright S (1965) The Interpretation of Population Structure by F-Statistics with Special Regard to Systems of Mating. *Evol* 19: 395–420.
- Zhao Y, Williams R, Prakash CS, He G (2012) Identification and characterization of gene-based SSR markers in date palm (*Phoenix dactylifera* L.). *BMC Plant Biol* 12(1): 237. <https://doi.org/10.1186/1471-2229-12-237>

Table S1. EST-SSR and SSR primer name, sequences and references.

Locus name	EST- SSR primer sequence	Ref
EST-PDG3119-rubisco-F	CATACTGATTATTGGCACACC	(Zhao et al. 2012)
EST-PDG3119-rubisco-R	GTACCATACCGTACCAGTTCA	
EST-DPG0633- Laccase -F	AGACTGGTTAAGTTGGTGGAG	(Zhao et al. 2012)
EST-DPG0633-Laccase-R	CTACAAAACCTGATGTGGTGGT	
EST-GTE-F	GCTTGGCCATCTATGAAAC	--
EST-GTE-R	ACTCTGAGCATCCATATCG	--
SSR primer sequence		
MPdCIR025(GA)22-F	GCACGAGAAGGCTTATAGT	(Bodian et al. 2014)
MPdCIR025(GA)22-R	CCCCTCATTAGGATTCTAC	
MPdCIR048(GA)32-F	CGAGACCTACCTTCAACAAA	(Bodian et al. 2014)
MPdCIR048(GA)32-R	CCACCAACCAAATCAAACAC	
MPdCIR078(GA)13-F	TGGATTTCATTGTGAG	(Bodian et al. 2014)
MPdCIR078(GA)13-R	CCCGAAGAGACGCTATT	
mPdCIR085(GA)29-F	GAGAGAGGGTGGTGTATT	(Bodian et al. 2014)
mPdCIR085(GA)29-R	TTCATCCAGAACCACAGTA	
MPdCIR090(GA)26-F	GCAGTCAGTCCCTCATA	(Bodian et al. 2014)
MPdCIR090(GA)26-R	TGCTTGTAGCCCTTCAG	
PdCUC3-ssr2(GA)22-F	ACATTGCTCTTTTGCCATGGGCT	(Bodian et al. 2014)
PdCUC3-ssr2(GA)22-R	CGAGCAGGTGGGGTTCGGGT	

Table S2. Discrimination power (Gst value), of SSR loci obtained.

Locus	Sample Size	Ht	Hs	Gst	Nm
Locus1	122	0.0526	0.0385	0.2676	1.3687
Locus2	122	0.3742	0.0354	0.9053	0.0523
Locus3	122	0.4580	0.1177	0.7430	0.1730
Locus4	122	0.2066	0.0438	0.7882	0.1343
Locus5	122	0.3214	0.0792	0.7536	0.1635
Locus6	122	0.2845	0.0083	0.9707	0.0151
Locus7	122	0.1859	0.0333	0.8209	0.1091
Locus8	122	0.1721	0.0136	0.9212	0.0428
Locus9	122	0.2731	0.0625	0.7710	0.1485
Locus10	122	0.0429	0.0302	0.2961	1.1888
Locus11	122	0.4963	0.0490	0.9013	0.0548
Locus12	122	0.1975	0.0000	1.0000	0.0000
Locus13	122	0.0759	0.0136	0.8214	0.1227
Locus14	122	0.1600	0.0469	0.7072	0.2071
Locus15	122	0.2133	0.0906	0.5751	0.3694
Locus16	122	0.4945	0.0678	0.8629	0.0794
Locus17	122	0.3883	0.1199	0.6913	0.2233
Locus18	122	0.4910	0.0604	0.8770	0.0702
Locus19	122	0.2330	0.0271	0.8836	0.0659
Locus20	122	0.0316	0.0136	0.5705	0.3765
Locus21	122	0.4727	0.1042	0.7796	0.1413
Locus22	122	0.2527	0.0552	0.7816	0.1397
Locus23	122	0.0636	0.0083	0.8691	0.0753
Locus24	122	0.4800	0.0521	0.8915	0.0609
Locus25	122	0.2209	0.0250	0.8869	0.0637
Locus26	122	0.4694	0.1230	0.7380	0.1775
Locus27	122	0.2788	0.0354	0.8729	0.0728
Locus28	122	0.0331	0.0219	0.3391	0.9744
Locus29	122	0.3906	0.0604	0.8453	0.0915
Locus30	122	0.4979	0.0469	0.9059	0.0519
Locus31	122	0.4457	0.1334	0.7006	0.2136
Locus32	122	0.1456	0.0271	0.8138	0.1144
Locus33	122	0.0651	0.0438	0.3276	1.0262
Locus34	122	0.4045	0.1593	0.6060	0.3250
Locus35	122	0.0621	0.0271	0.5633	0.3876
Locus36	122	0.3333	0.0219	0.9343	0.0351
Locus37	122	0.2922	0.0521	0.8217	0.1225
Locus38	122	0.3959	0.0438	0.8895	0.0621
Locus39	122	0.4788	0.0761	0.8411	0.0945
Locus40	122	0.4900	0.1020	0.7930	0.1340
Mean	122	0.2804	0.0530	0.8109	0.1166

* Nm = estimate of gene flow from Gst or Gcs. E.g., $Nm = 0.5(1 - Gst)/Gst$.

Abbreviations: Hs = inbreeding due to sub-population, Ht = Hnbreeding in total population, Gst = Discrimination power, and Nm = Number of migration.

Table S3. Genetic diversity parameters in date palm cultivars.

No	Cultivar name	Na	Ne	I	He	uHe	P
1	Mazafati	0.625	1.127	0.104	0.071	0.086	17.50
2	Kalooteh	0.625	1.138	0.116	0.079	0.095	20.00
3	Khalezohrei	0.650	1.138	0.116	0.079	0.095	20.00
4	Holeileh	0.700	1.112	0.106	0.069	0.083	20.00
5	Mordarsang	0.500	1.069	0.058	0.039	0.047	10.00
6	Khazab	0.500	1.056	0.053	0.035	0.042	10.00
7	Holoo	0.600	1.077	0.077	0.050	0.060	15.00
8	Khenizi	0.600	1.117	0.092	0.064	0.077	15.00
9	Negar	0.625	1.127	0.104	0.071	0.086	17.50
10	Shahani	0.475	1.058	0.046	0.032	0.038	7.50
11	Male isolate	0.675	1.101	0.094	0.062	0.074	17.50
12	Alimehtari	0.525	1.069	0.058	0.039	0.047	10.00
13	Kharook	0.450	1.024	0.017	0.012	0.015	2.50
14	Gantar	0.550	1.082	0.063	0.044	0.053	10.00
15	Zahidi	0.600	1.104	0.087	0.059	0.071	15.00
16	Jowzi	0.575	1.104	0.087	0.059	0.071	15.00
17	Khadhrawi	0.750	1.186	0.150	0.103	0.124	25.00
18	Shekkar	0.475	1.032	0.036	0.022	0.027	7.50
19	Istamaraan	0.525	1.045	0.041	0.027	0.033	7.50
20	Barhi	0.650	1.104	0.087	0.059	0.071	15.00
21	Hallawi	0.600	1.104	0.087	0.059	0.071	15.00
22	Braim	0.500	1.069	0.058	0.039	0.047	10.00
23	Dayri	0.525	1.080	0.070	0.047	0.056	12.50
24	Beliani	0.650	1.101	0.094	0.062	0.074	17.50
25	Owaidi	0.525	1.080	0.070	0.047	0.056	12.50
26	Sowaidani	0.450	1.071	0.051	0.037	0.044	7.50
27	Owaimri	0.525	1.056	0.053	0.035	0.042	10.00
28	Mashtoom	0.600	1.104	0.087	0.059	0.071	15.00
29	Fersee	0.600	1.114	0.099	0.067	0.080	17.50
30	SabzParak	0.475	1.080	0.070	0.047	0.056	12.50
31	GhannamiSabz	0.500	1.053	0.060	0.037	0.045	12.50
32	Wardi	0.400	1.071	0.051	0.037	0.044	7.50
33	GhannamiSorkh 1	0.525	1.095	0.068	0.049	0.059	10.00
34	GhannamiSorkh2	0.500	1.045	0.041	0.027	0.033	7.50
35	Foreign male 1	0.525	1.106	0.080	0.056	0.068	12.50
36	Foreign male 2	0.525	1.106	0.080	0.056	0.068	12.50

Abbreviations: Na = No. of different alleles, Ne = No. of Effective alleles, I = Shanon Information. index, He = Expected Heterozygosity, uHe = Unbiased Expected Heterozygosity, and P% = Polymorphism percentage.

Table S4. K-Means clustering of date palm cultivars based on SSR and EST-SSR data.

SSD(T)	SSD(AC)	SSD(WC)	r-squared	pseudo-F	BIC
1376.672	202.482	1174.191	0.147	20.693	871.945
1376.672	319.18	1057.492	0.232	17.959	863.978
1376.672	411.173	965.499	0.299	16.751	857.679
1376.672	498.76	877.912	0.362	16.618	850.881
1376.672	566.45	810.222	0.411	16.22	845.896
1376.672	620.841	755.831	0.451	15.744	842.222
1376.672	660.327	716.346	0.48	15.012	840.48
1376.672	695.563	681.109	0.505	14.425	839.13
1376.672	724.476	652.196	0.526	13.824	838.642
1376.672	753.245	623.427	0.547	13.411	837.943
1376.672	780.048	596.625	0.567	13.074	837.385
1376.672	807.604	569.068	0.587	12.891	836.42
1376.672	832.456	544.217	0.605	12.708	835.777
1376.672	853.797	522.876	0.62	12.48	835.7
1376.672	874.463	502.209	0.635	12.305	835.584
1376.672	891.686	484.986	0.648	12.066	836.131
1376.672	909.553	467.12	0.661	11.912	836.356
1376.672	925.794	450.878	0.672	11.75	836.842
1376.672	940.632	436.04	0.683	11.581	837.564

* Best clustering according to Calinski & Harabasz' pseudo-F: k = 2 & Best clustering according to Bayesian Information Criterion: k = 16
Abbreviations: SSD(T) = Total sum of squares, SSD(AC) = Among clusters sum of squares, and SSD(WC) = Within clusters sum of squares.

Table S5. Pair-wise AMOVA showing significant genetic difference between the studied date palm cultivars (cultivar numbers are according to Table S3).

Cultivar1	Cultivar2	Pvalue
3	21	0.001
4	8	0.001
4	16	0.001
5	10	0.001
5	35	0.001
6	7	0.001
6	11	0.001
6	27	0.001
6	29	0.001
6	1	0.001
7	12	0.001
8	22	0.001
8	30	0.001
11	23	0.001
14	15	0.001
15	36	0.001
21	23	0.001
18	24	0.001
20	24	0.001
20	36	0.092
21	36	0.001
24	26	0.001
25	34	0.001
27	31	0.001

Table S6. Assignment result of date palms (only samples inferred to be from other populations are given) based on positive likelihood. (cultivar numbers are according to Table S3).

Home cultivar	Inferred cultivar1	cultivar	1	2	3	4	5	6	7
1	3	4.432	4.15	3.516	6.055	10.076	12.431	14.59	12.748
1	2	6.034	4.099	4.789	5.549	10.356	13.829	15.386	14.793
2	3	4.592	4.533	3.789	4.453	8.18	10.812	12.845	13.6
2	1	3.373	4.724	3.947	5.708	13.331	12.732	10.891	9.901
2	3	3.704	3.413	3.227	4.453	10.414	14.352	14.289	14.969
3	2	3.579	2.798	3.617	4.152	8.812	12.13	12.067	12.873
3	2	3.771	2.798	3.77	4.328	11.59	14.051	11.988	13.094
4	2	5.057	4.439	5.537	5.554	8.796	10.431	10.368	9.997
8	7	8.698	9.4	9.588	6.106	7.683	7.414	4.429	5.641
9	5	8.379	9.044	9.093	10.565	7.239	7.271	11.271	9.776
9	10	8.328	7.597	8.093	7.935	10.96	12.829	12.908	11.316
10	11	10.93	9.09	10.19	8.759	11.106	11.13	14.085	13.316
11	10	9.93	8.898	9.792	9.537	12.437	12.829	14.607	12.89
11	12	11.555	11.032	10.588	10.236	11.692	11.829	10.243	8.043
11	10	12.708	12.713	13.588	10.333	10.68	11.607	15.306	13.316
15	16	11.437	14.442	14.257	13.379	16.294	17.574	15.431	15.288
15	16	9.592	10.965	11.081	9.981	13.118	14.352	15.908	13.714

Table S7. Fst versus Pst values in cultivar No.3 with others.

Character	Pst	Fst	Pst	Fst	Pst	Fst	Pst	Fst
Fruit weight	0.053	0.16	0.43	0.16	0.36	0.16	0.053	0.16
Fruit width	0.001	0.16	0.09	0.16	0.09	0.16	0.001	0.16
Fruit length	0.99	0.16	0.46	0.16	0.01	0.16	0.99	0.16
Seed weight	0.99	0.16	0.44	0.16	0.98	0.16	0.98	0.16
Seed width	0.98	0.16	0.44	0.16	0.99	0.16	0.99	0.16
Seed length	0.99	0.16	0.44	0.16	0.98	0.16	0.99	0.16