Chapter

Sequences Evolution and Population Structure of Tunisian Date Palm (*Phoenix dactylifera* L.) Revealed by Chloroplast DNA Markers

Rhouma-Chatti Soumaya, Choulak Sarra and Chatti Khaled

Abstract

Date palm is present among the vital crops of arid and semiarid countries of North Africa and the Middle East. Chloroplast DNA is the best molecule for finding the evolutionary history of plant species. In the present study, cpDNA variation in date palm was estimated using the *trn*L-*trn*F intergenic spacer and *psb*Z-*psb*C region. The high AT values in both molecular markers may clarify the high proportion of transversions observed in this species. The neutrality tests, expansion parameter estimation (mismatch distribution), and haplotype network patterns proposed that demographic expansion had occurred in recent times. Furthermore, the taxa distribution is not related to geographical origins; neighbor-joining trees are clustered independently either from their geographic origin or from the sex of trees, suggesting a common genetic basis between different cultivars. Statistical analysis of chloroplast germplasm provides a means of assessing cytoplasmic gene flow, which occurred in Tunisian *Phoenix dactylifera* L. In fact, N_m was important between Tunisia and Eastern Arabic region (N_m = 2.57), which reflects high levels of connectivity between these population pairs. In conclusion, genomic studies prove date palm domestication happened in the Arabian Peninsula and showed an important gene flow with North African palm populations.

Keywords: *Phoenix dactylifera* L., *trn*L-*trn*F spacer, *psb*C-*psb*Z region, population expansion, molecular evolution, gene flow

1. Introduction

Date palms (*Phoenix dactylifera*, Arecaceae) are of major economic and ecological importance to the oasis agriculture of arid and semiarid zones. The past distribution area of this fruit crop covers Mauritania in the west to Pakistan in the east and northwestern India [1]. It is moreover existing in sub-Saharan Africa and has been hosted in California, Peru, Australia, and other countries [2]. The presence of this crop in these areas gives it an undeniable ecological role by limiting the progression of steppe areas and the silting of farmland.

P. dactylifera was brought to Tunisia by Phoenicians before the Roman occupation and has played a vital part in establishing oases. In Tunisia, date palm cultivation covers over 46 thousand ha, with a total number of palm trees of approximately 5.4 million [3]. One hectare of land host an average of 120 date palm trees [4]. Further, the Tunisian date palm germplasm is distinct by a remarkable richness, represented by the high number of cultivars (over 250) [5, 6]. Throughout the plant domestication procedure, breeding actions, selection, migration, and admixture have given growth to cultivated populations distinct from the inherited gene pools [7]. Humans have particularly selected qualities associated with production, fruit quality, and fertility [8].

Understanding the population genetics and domestication history of cultivated species is certainly important for the genetic improvement of crops relying on the conservation and usage of the germplasm [9]. In order to contribute to the varietal improvement of date palms and to offer novel perceptions on the influence of geographic origins and human action on the genetic structure of the date palm, this study investigated the diversity of the species using chloroplast DNA.

The small, sensibly constant size, and conservative evolution of chloroplast DNA (cpDNA) make it an ultimate molecule for finding the evolutionary history of plant species. In angiosperm, the use of chloroplast DNA sequences for intra-specific phylogenetic research is now routine [10]. These investigations have been simplified by the abundant number of whole chloroplast genome sequences that are accessible from an extensive variety of angiosperms and the development of universal PCR primers in conserved coding as well as noncoding regions.

Commonly, in terms of its size, organization, and sequence, cpDNA is the most recognized conservatively evolving genome. It has been used for genetic analysis in plants and provides an accessible and well-characterized source of comparative sequence data. Moreover, identifying the footmark of positive selection is an imperative assignment in evolutionary genetic studies [11]. Otherwise, different modes of selection may result in divergent patterns of the nature and extent of genetic variation. The neutral theory affirms that all observable mutations in populations have little or no effect on an organism's fitness, and their evolutionary dynamics are entirely measured by genetic drift [12, 13]. For such mutations, the evolution continues as equilibrium among the forces of mutation pressure, natural selection, and genetic drift. New molecular techniques based on single and combined sequences data sets have provided a vast understanding of the evolution of flowering plants [14]. Noncoding regions are usually less sensitive to natural selection than coding regions and then may be more beneficial for studying plant evolution.

The chloroplast genome can be classified into three functional categories: (1) protein-coding genes, (2) introns, and (3) intergenic spacers. It has greater phylogenetic potential than nuclear DNA because it is sufficient variable but conserves to be less variable within than between species [15]. The noncoding regions are leading systematic molecular, phylogeographic, and DNA barcoding studies for plants [16, 17]. Chloroplast DNA markers are used for systematic studies of plant species [18–23] and are particularly used to study phylogeny.

In *Paris* genus, Song et al. [24] described eight most variable barcodes, to discriminate between the different species, including psbC-trnS-psbZ region. Moreover, the psbC-trnS intergenic spacer was successfully used for phylogeographic study in *Lolium* species [25]. In 2013, Ballardini et al. confirmed that the psbZ-trnfM (CAU) region could be considered a good basis for the establishment of a DNA barcoding system in *Phoenix*, and is potentially useful for the identification of the female parent in *Phoenix* hybrids.

The date palm chloroplast genome is a typical circular double-stranded DNA molecule, and it shares a common quadri-partite structure: a pair of IRs (27.276 bp) separated by the LSC region (86.198 bp) and the SSC regions (17.712 bp) [26]. An important set of primer pairs improved for PCR amplification and covering sequencing in monocotyledons were optimized, which are distributed throughout the whole chloroplast genome, including exons, introns, and intergenic spacers (IGS).

Thus, in the present study, we undertake to find out the level of cpDNA variation in date palm using the *trn*L-*trn*F intergenic spacer and *psb*Z-*psb*C region and to accurate evolution process, which controls the pattern of polymorphism among the species.

2. Material and methods

2.1 Plant material and DNA extraction

Twenty-four accessions (20 cultivars and four male trees) of Tunisian date palm, listed in **Table 1**, were used in this study. Each cultivar was represented by one tree. Five varieties accommodated in Tunisian plantations ("Ghars Mettig" and "Tantabecht" from Algeria, "Berhi" and "Khadhraoui" from Iraq, and "Abou Meaan" from the United Arabic Emirates) were used in this survey. Young leaves were frozen until their use for DNA purification. Extraction of the total DNA was determined as stated by Dellaporta et al. [27] protocol. DNA concentration and integrity were checked by 0.8% agarose gel electrophoresis according to Sambrook et al. [28].

2.2 Amplification and DNA sequencing

Chloroplastic DNA was sequenced for the *trn*L-*trn*F (intergenic spacer) and *psb*C-*psb*Z (intergenic spacer + gene). Target regions were amplified using universal primers: (Fw1: 5′GGTTCAAGTCCCTCTATCCC3′; Rv1: 5′ATTTGAACTGGTGACACGAG3′) and (Fw2:5′CAACCTTGGCAAGAACG3′; Rv2: 5′TTGACCAACCATCAGRAGA3′) *trnL-trnF* spacer and for *psb*C-*psb*Z region, respectively. Amplification was carried out for 35 cycles, all comprising of a denaturation phase at 95°C for 1 min, annealing at 50°C for 1 min, and an extension step at 72°C for 2 min.

The total volume of PCR reaction was 25 μL, which contained 25 mM of $MgCl_2$, 2 mM of dNTP mix, 1.6 mM of each primer and 1 unit of DNA Taq polymerase, and 20 ng of DNA. Agarose-gel electrophoresis (1.5%) was used to check the PCR products.

The purified PCR products for the *trnL-trnF* and *psb*C-*psb*Z regions were sequenced in both strands according to the automated Sanger method [29] using automated sequencer ABI PRISM™ 310 Genetic Analyzer (Applied Biosystems). The process consists of the selective incorporation of chain-terminating dideoxy-nucleotides by DNA polymerases during in vitro DNA replication. In Sanger sequencing, target DNA is copied multiple times, producing fragments of varying lengths. Fluorescent "chain-terminating" nucleotides mark the ends of the fragments so that the sequence can be determined. It is the most widely used method for detecting single nucleotide variations.

2.3 Sequence analysis

The identity of both sequenced regions was confirmed through a BLASTN search in NCBI database [30]. Nucleotide sequences were aligned by Mega 5.2.2 [31].

Origin	Ecotype	Label	*psb*C-*psb*Z		Combined *Trn*L-*trn*F and *psb*C-*psb*Z	
			Length	%GC	Total length	%GC
Tunisia	Lagou	Lg	953	39.7	1353	37.88
	BesserHelou	Bh	951	40.1	1356	38.14
	Gasbi	Gb	967	40.1	1347	38.21
	Boufeggous	Bf	960	40.1	1360	38.21
	Hamra	Hm	971	40.0	1370	38.14
	Tazerzit Safra	Ts	952	39.8	1335	37.91
	Goundi	Gd	962	39.2	1358	38.18
	Menakher	Mk	967	40.0	1350	37.76
	Ammari	Am	969	39.9	1372	38.11
	Deglet Nour	Dn	971	40.0	1374	38.06
	Kentichi	Kt	964	39.3	1372	38.44
	Oum Laghlez	Ol	965	40.1	1359	38.14
	Arichti	Ar	973	40.1	1373	38.14
	Guelb Jemel	Gj	973	39.8	1355	37.91
	Kharroubi	Kb	967	39.7	1382	38.06
Algeria	Tantabecht	Tb	976	40.2	1357	38.18
	Ghars Mettigue	Gm	952	39.9	1352	37.99
Iraq	Berhi	Br	962	40.1	1345	38.11
	Khadhraoui	Kd	971	39.8	1373	37.99
UEA	Abou Meaan	Ab	969	40.2	1364	38.18
	Male trees					
Tunisia	Borhane2	B2	967	39.9	1368	38.02
	Borhane3	B3	963	39.9	1348	38.06
	CRPh1	C1	966	40.0	1354	38.21
	CRPh5	C5	962	40.0	1362	38.14

Table 1.
Date-palm accessions studied, their origin, accession numbers, and their variation in length, GC and AT contents of the psbC-psbZ region and the combined sequences.

Several genetic parameters were determined with DnaSP program [32]. Haplotype diversity (Hd) [33] and genetic diversity (Pi) [34] were calculated to evaluate genetic diversity. The average of nucleotide differences (k), the minimum number of recombination events (Rm), and the average number of nucleotide differences among cultivars were also detected. By means of selective neutrality tests, we checked the hypothesis of the mutation/drift equilibrium for a supposedly neutral polymorphism. Selection neutrality for the detected mutations was tested by both Tajima's D [35] and Fu and Li's D* and F* methods [36], using the DNAsp program. Demographic parameters were assessed using the distribution of pairwise sequence differences (mismatch distribution) of Rogers and Harpending [37] and site-frequency spectra of Tajima [35].

Moreover, Fu's Fs statistics [38] was used to confirm the assumption of population growth and range expansion as revealed by the mismatch distribution of Rogers and Harpending [37]. In addition, we calculated the Harpending's raggedness index (r) corresponding to an estimate of the fluctuation in the frequency of differences between haplotype pairs [39]. In a complementary way, the R2 statistic [40] was calculated based on the differences between the number of singleton-type mutations and the average of the nucleotide differences. These analyses were executed using coalescent simulations implemented in DnaSP software, with 1000 simulated resampling replicates.

To study the genetic relationships between the studied sequences (haplotypes), we used the reduced median network analysis available in the NETWORK software [41]. The phylogenetic relationships between the studied chloroplast haplotypes were reconstructed using the neighbor-joining (NJ) method [42]. NJ builds a tree from a matrix of pairwise evolutionary distances relating to the set of taxa being studied. Gene flow (*Nm*) was estimated with the mean number of migrants per generation among populations. N_m values were calculated with 1000 data permutations using the software DnaSP v 5.10.01. With reference to the standard for gene flow, we described genetic flow as low for Nm < 1, high for 1 < Nm < 4, and very high for Nm > 4 [43]. In gene flow analysis, we consider three different geographic regions of *P. dactylifera* L. Tunisia, Algeria, and Eastern Arabic (Iraq and UAE).

3. Results

3.1 Sequences variation in the psbC-psbZ region

The DNA sequencing of the generated bands has been successfully performed and the blast search allowed confirming the identity of the sequences. DNA sequence varied from 951 bp for "*Besser Helou*" cultivar to 967 bp for "*Gashi*" cultivar (**Table 1**) with an average of 964.5 pb length. The nucleotide composition frequencies are 0.274 (A), 0.325 (T), 0.204 (C), and 0.196 (G). In addition, the GC content of the amplified sequences varied from 39.7% to 40.7%, and the AT content from 59.3% to 60.3% (**Table 1**).

The polymorphism pattern of the *psb*C-*psb*Z sequence in *P. dactylifera* L. reveals a high level of mutation with 55 polymorphic sites (**Table 2**). We observed 33 singleton variable sites and 22 parsimony-informative ones. Moreover, a low transitional/transversional ratio (ti/tv = 0.68) occurs in the *psbC-psbZ* sequence. In addition, 22 haplotypes were detected among 24 cultivars analyzed, yielding a haplotype diversity of 0.989 (**Table 2**). Sequence variation observed between cultivar groups and nucleotide diversity was estimated (Pi = 0.00862) (**Figure 1a**). The average number of nucleotide differences "k" was estimated to be 8.264. Furthermore, seven regions of conserved DNA were detected using DnaSP program yielding a value of sequence conservation C = 0.913. Mismatch distribution within the species was unimodal for the *psb*C-*psb*Z marker (**Figure 2a**), suggesting demographic expansion had occurred in recent times.

A phylogenetic tree was constructed using *NJ* methods (**Figure 3a**). The evolutionary distances were estimated by the maximum composite likelihood method [44]. Haplotypes of the *psb*C-*psb*Z regions made it possible to make moderately sustained phylogenetic trees (Bootstrap values ≤74%). This dendrogram supported the varieties' organization into two main clusters, which were subdivided into different subclusters. Clearly, the obtained clustering is made independently from the geographical origin of cultivars since the foreign varieties, newly introduced in the

	psbC-psbZ **sequence**	**Combined sequence**
Number of accessions (N)	24	24
Lengths (pb) (Average)	964	1360
GC (%) (Average)	40	37.38
AT(%) (Average)	60	62.66
Number of polymorphic sites (S)	55	69
Parsimony informative sites (Pis)	22	31
Number of haplotypes (H)	22	24
Nucleotide diversity (Pi)	0.00862	0.00861
Haplotype diversity (H_d)	0.989	1
Average of pairwise nucleotide differences (k)	8.264	11.471
Tajima's D	−1.96519 (0.10 > P > 0.05)	−1.74847 (0.10 > P > 0.05)
Fu and Li's D*	−2.345430 (0.10 > P > 0.05)	−1.97901 (0.10 > P > 0.05)
Fu and Li's F*	−2.61261 (0.10 > P > 0.05)	−2.23846 (0.10 > P > 0.05)
Fu's Fs	−12.927 (0.10 > P > 0.05)	−15.474 (0.10 > P > 0.05)
R2 (Ramos-Onsins and Rozas's test)	0.000 P < 0.05	0.000 P < 0.05
Raggedness (r)	0.000 P < 0.05	0.000 P < 0.05

Table 2.
Sequences polymorphism and divergence within date palm cultivars.

Tunisian growing areas, did not remarkably separate from the autochthonous ones. In addition, the male and female trees did not diverge from each other.

In a complementary way, a genetic network based on the *psb*C-*psb*Z sequences was reconstructed (**Figure 3b**). All haplotypes are connected to the most frequent haplotype H1. It represented the ancestor sequence of the other ecotypes, which mutate during evolution. Several putative haplotypes, corresponding to intermediate evolutionary steps, were detected in our network (mv in **Figure 3b**). The haplotype patterns (star-like pattern) reflect a signature of a recent expansion in Tunisian date palm.

3.2 The combined region trnL-trnF and psbC-psbZ spacers

The combined sequence varied from 1347 bp for "*Gasbi*" cultivar to 1382 bp for "*Kharroubi*" cultivar (**Table 1**). Current sequences do not contain insertions or deletions (indels). These sequences revealed 69 polymorphic sites and defined 24 haplotypes. Among the 69 variable sites, 31 were parsimoniously informative and 38 were singletons sites. The means of the haplotype (*Hd*) and nucleotide (*Pi*) diversity (**Figure 1b**) were higher than for *trn*L-*trn*F spacer [16] and the *psb*C-*psb*Z region taken separately. In fact, these values were 1 ± 0.014 and 0.00861 ± 0.001, respectively. The average of pairwise

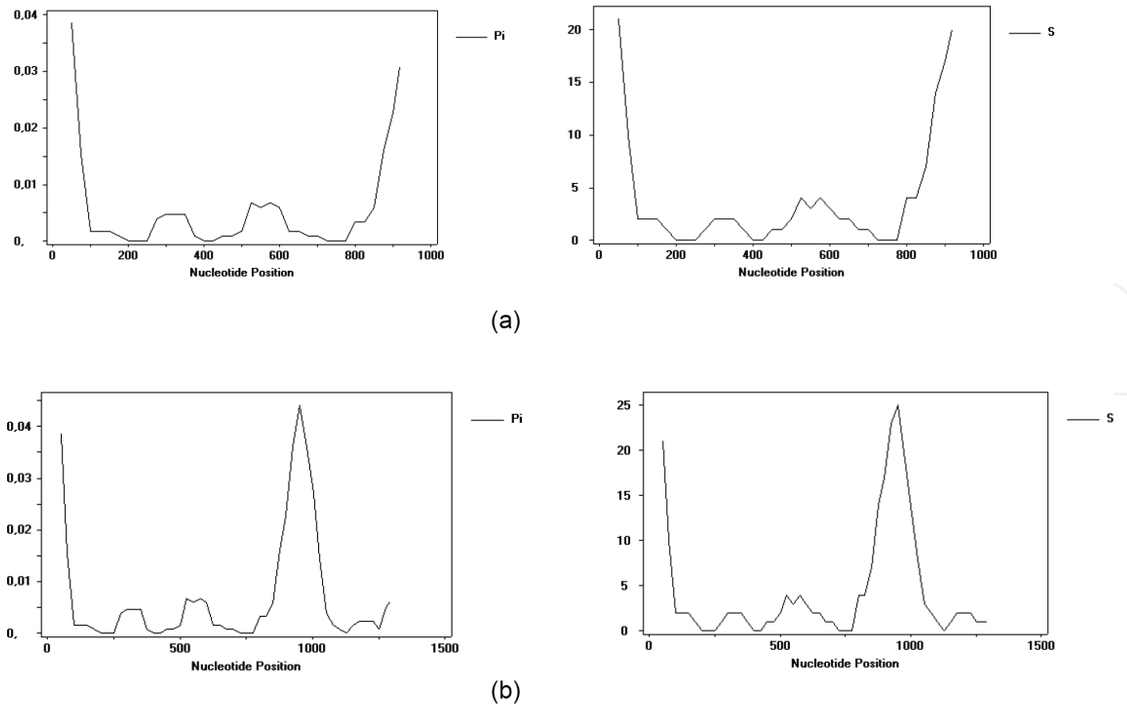

Figure 1.
Variability of nucleotide diversity (Pi) and segregating sites (S) for the psbC-psbZ chloroplastic DNA (a) and the combined sequences (b).

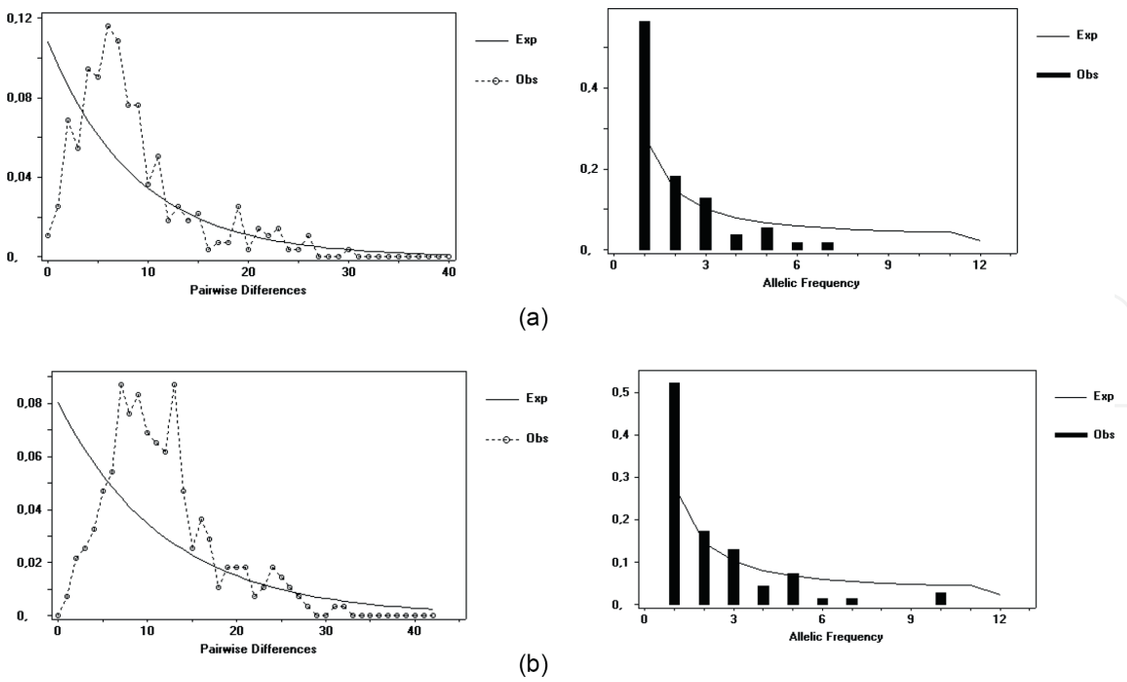

Figure 2.
Mismatch distribution of chloroplastic DNA sequences of the date palm cultivars based on pairwise nucleotide differences in the psbC-psbZ region (a) and the combined trnL-trnF spacer and psbC-psbZ (b). In addition, graphical representation of the site frequency spectrum of chloroplast DNA sequences. Solid lines in the site-frequency spectra indicate the expected distributions under neutrality and at equilibrium.

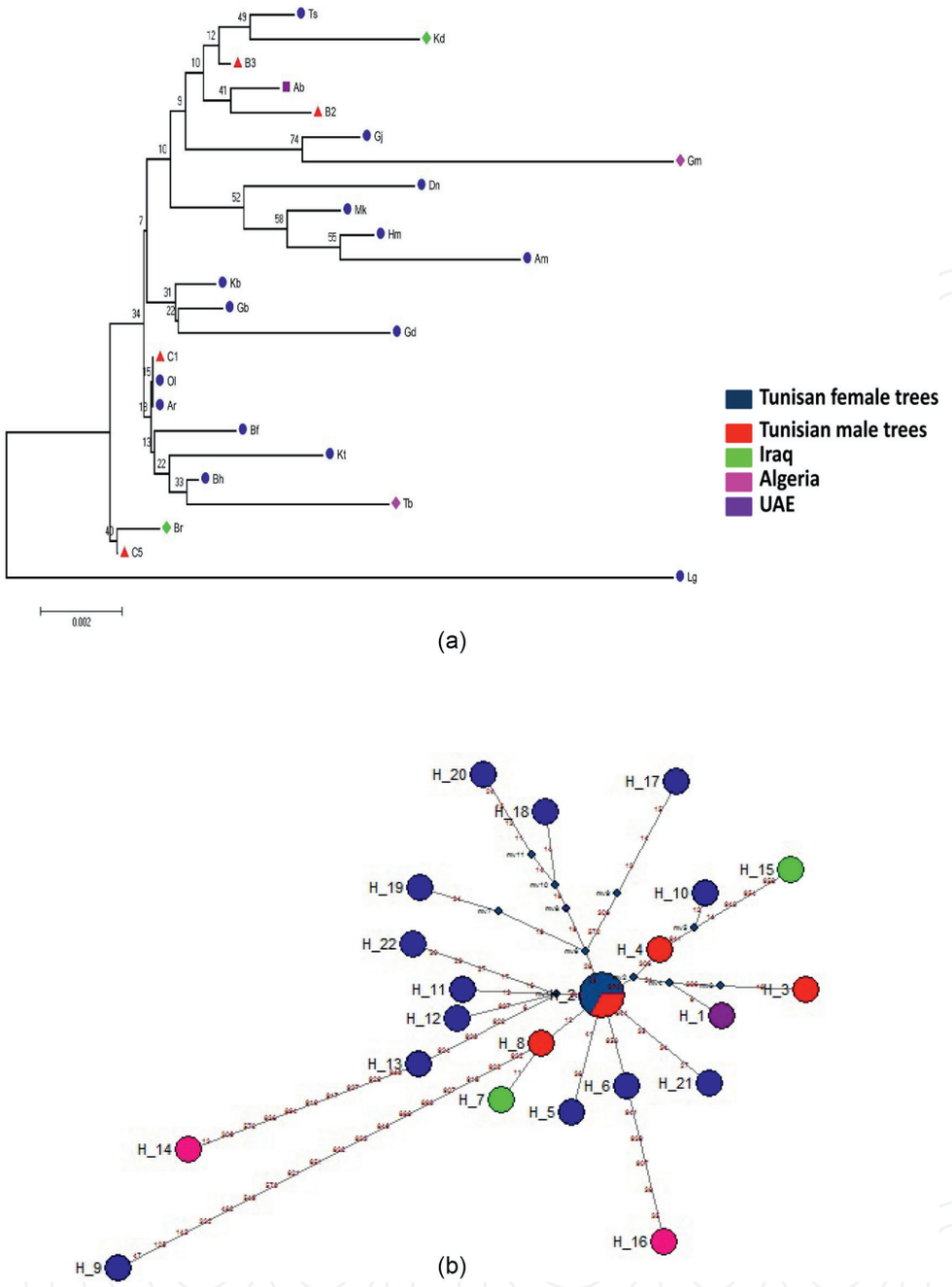

Figure 3.
Summary of phylogenetic analysis within Phoenix dactylifera based on psbC-psbZ sequences: (a) neighbor-joining tree of 24 Tunisian date-palm cultivars. (b) Median-joining network of the haplotypes inferred from the analyzed sequences. Nodes are proportional to haplotype frequencies and branch length is proportional to the number of mutations. The circles represent haplotypes; circle diameters are proportional to the frequencies. Black bands correspond to mutational steps.

nucleotide differences (k) was equal to 11.471 (**Table 2**). In addition, the GC content of the combined sequences varied from 37.88% to 38.44% (**Table 1**). The polymorphism pattern in *P. dactylifera* L. revealed an elevated level of mutation points with a low transitional/transversional ratio (ti/tv = 0.63).

Mismatch distributions within the species were generally unimodal for markers, suggesting population expansions (or past selective sweeps) (**Figure 2b**). Numerous

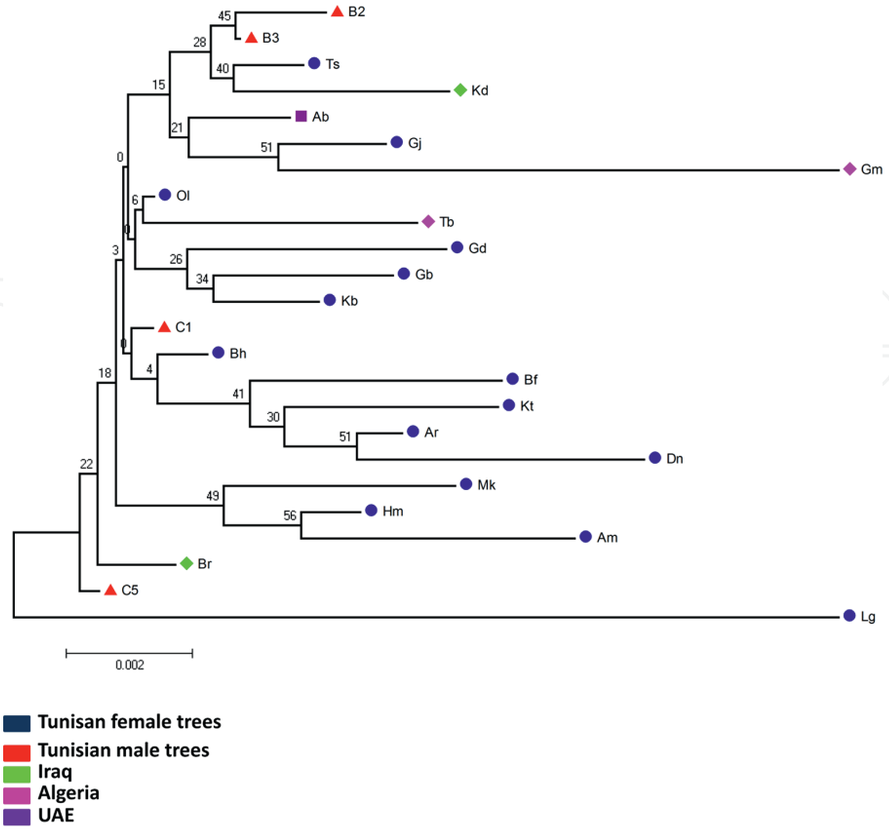

Figure 4.
Neighbor-joining tree based on trnL-trnF spacer and psbC-psbZ gene sequences.

significant negative values of *Fs* and Tajima's *D* in *trn*L-*trn*F [16], *psb*C-*psb*Z spacers, and the combined region (**Table 2**) rule out the hypothesis of a constant population and propose either selective sweeps or past demographic expansion (**Table 2**). In addition, low values of *R2* and Harpending's raggedness index (*r*) will be characteristic of a recent expansion.

For gene flow estimation, N_m was high ($N_m > 1.0$) between Tunisia and Algeria ($N_m = 1.8$) and Tunisia/Eastern Arabic accessions ($N_m = 2.57$) indicating that few differentiations could be established among them. On the other hand, *Nm* was weak between Algeria and Eastern Arabic ($N_m = 0.66$), this reflects low levels of connectivity between these population pairs.

Genetic relationships among date palm cultivars are investigated, also, using the variation observed in the combined (*trn*L-*trn*F and *psb*C-*psb*Z) sequences. As illustrated by **Figure 4**, the phylogeographical patterns were very similar between the *psb*C-*psb*Z gene and the combined sequences. The *NJ* dendrogram supported the variety organization into three monophyletic branches and one cluster. The first branch is composed of "*Lagou*" (Lg), the second one *CRPh5* (C5) Tunisian male tree, while the third is made of Iraquian variety "*Berhi*" (Br).

The remaining cultivars are ranged in the unique cluster, which is divided into four subgroups. In fact, genotypes' clustering is independent of the sex of trees, and it is not structured according to geographical origin. Indeed, with the exception of the foreign cultivar "*Berhi*" (Br) divergence, the other foreign cultivars did not significantly diverge from the Tunisian plantations (**Figure 4**).

4. Discussion

In the present survey, we tested the reliability of the noncoding chloroplast markers to identify date palm cultivars. Therewith, we sought an indication of the level of genetic variation and genetic classification within the date palm cultivars grown in Tunisia.

In this object, we represented the evolution of the *psb*C-*psb*Z gene and the combined sequences (*trn*L-*trn*F spacer/*psb*C-*psb*Z) to reveal polymorphisms in *P. dactylifera* L. The high AT values in both molecular markers may clarify the high proportion of transversions (ti/tv = 0.67 and ti/tv = 0.63) observed in this species. This result corroborates the discovery in angiosperm chloroplast noncoding regions with a ratio ti/tv not exceeding one for any of the examined taxa [45–50]. Base content explains the relatively high proportion of transversions [45].

The use of coalescent theory [51, 52] for the *P. dactylifera* L. sequences has permitted inferences about past and present population size. It has been confirmed that demographic expansions conduct to star-shaped genealogies [53], an excess of rare mutations [54], and a unimodal mismatch distribution [37]. Haplotype network patterns, the neutrality tests, and the expansion parameter estimations (mismatch distribution) proposed that demographic expansion had occurred in recent times as indicated by the excess of mutations type singletons in the *P. dactylifera* sample.

Furthermore, there was no geographical structure in the relationships among the haplotypes; neighbor-joining trees are clustered independently either from their geographic origin or from the sex of trees, indicating a common genetic basis between different cultivars. A low genetic structure is usually related to natural habitat change or to human activities that increase gene flow between populations [55]. Human impact on these regions may be the reason for these findings. Actuality, in Tunisian localities, cultivars are manipulated by farmers after continuous selection, cloning, and exchange of varieties. Similar results have been observed using other molecular markers, where *P. dactylifera* ecotypes clustering made independently from either the tree's sex or their geographic origin in spite of their great distribution [16, 56–60]. This result is contrasted with Zehdi-Azzouzi et al. (2015), where *NJ* classification was clearly coherent with a geographical structuring into two clusters: Eastern pool (Djibouti, Oman, Iraq, the UAE) and the Western pool (North Africa accessions). Also, using three chloroplast regions of rbcL, matK, and trnH-psbA, Iranian date palm cultivars were separated into clades corresponding to their geographical distribution [61].

Statistic analysis of chloroplast germplasm provides a means of assessing cytoplasmic gene flow, which occurred in Tunisian *P. dactylifera* L. The pattern observed confirmed that gene flow is designed to be a significant evolutionary factor in date palm history, allowing a variability of the genetic diversity at different scales (local and foreign). In fact, N_m was important between Tunisia and Eastern Arabic region (N_m = 2.57), which reflects high levels of connectivity between these population pairs. Zehdi-Azzouzi et al. [9] proved the movement of gene flows between Eastern and Western origins, mostly from east to west, following a human-mediated diffusion of the species.

In addition, genomic studies prove date palm domestication has happened in the Arabian Peninsula and that the North African population has mixed ancestry with components from Middle Eastern *P. dactylifera* [62]. In this subject, Gros-Balthazard et al. [63] noted that a first domestication event was shown to be improbable, given the very similar form of seeds in cultivars from Africa and the

Middle East. In fact, travelers and pilgrims brought many cultivars of date palms grown in Tunisian oases, particularly from the east [64]. The origins of domestication, the direction of germplasm flows, and the breeding history within the traditional cultivation are sincerely the reason for the observed exchange.

In addition, the mean number of migrants per generation between Tunisia and Algeria (N_m) was important; this could be the result of the relative proximity of geographical sites. In fact, *Deglet Nour* cv. was introduced into Tunisia plantations four centuries ago from western Algeria [64, 65], and became the most valuable cultivar. After this period, the human migrations between Algeria and the south of Tunisia were perturbed by the French occupation of Algeria [64]. Despite that, other works prove that in Algeria, chloroplast diversity presents about 70% of the eastern Arabic chloroplast. In Tunisia and Morocco, this proportion only ranges from 11–42%, although Algerian nuclear diversity is similar to that obtained in Tunisia [66].

In conclusion, environmental change, over-exploitation of water reserves, and an increasing tendency toward *Deglet Nour* cv. Monoculture consists of some of the difficulties of date palm culture in Tunisia. Those are significant issues, which need in-depth research. Considering these, it is decisive to comprehend the genetic diversity and variation among and within accessions, and extensive programs must be prepared, to include collection, estimation, and preservation of plant genetic resources. Determining the gene flow and the demographic expansion of this population should be possible with our two molecular markers, nevertheless, the exploitation of additional nuclear markers clearly provides more accurate results. Actually, with the availability of the date palm nuclear and chloroplast genome sequences, more molecular markers are accessible to explore systematic, cultivar relatedness, and genetic map structure. Furthermore, a webpage regrouping all information about *P. dactylifera* germplasm conservation and utilization has an obligation to be planned in collaboration with all producer countries in order to engender bases for inhibiting its genetic erosion.

Conflict of interest

The authors declare that they have no conflict of interest.

Abbreviations

cpDNA	chloroplast DNA
dNTP	deoxynucleoside triphosphate
Hd	Haplotype diversity
IR	inverted repeats
IGS	Intergenic Spacers
k	average of nucleotide differences
LSC	large-single-copy region
$MgCl_2$	magnesium chloride
Nm	Gene flow
PCR	polymerase chain reaction
Pi	genetic diversity
Rm	minimum number of recombination events
SSC	small-single-copy regions
ti/tv	transition/ transversion

Author details

Rhouma-Chatti Soumaya*, Choulak Sarra and Chatti Khaled
Laboratory of Genetics Biodiversity and Valorisation of Bioresources (LR11ES41),
Higher Institute of Biotechnology of Monastir, University of Monastir, Tunisia

*Address all correspondence to: rhoumasoumaya@yahoo.fr

IntechOpen

References

[1] Pintaud JC, Ludena B, Aberlenc-Bertossi F, Zehdi S, Gros-Balthazard M, Ivorra S. Biogeography of the date palm (*Phoenix dactylifera* L., Arecaceae): Insights on the origin and on the structure of modern diversity. Acta Horticulturae. 2013;**994**:19-38

[2] Barrow S. A monograph of Phoenix L. (Palmae: Coryphoideae). Kew. Bulletin. 1998;**53**:513-575

[3] Makhlouf-gafsi I, Mokni-ghribi A, Bchir B, Razafindralambo H, Danthine S, Attia H, et al. Foamability and foam stability of male and female date palm sap (*Phoenix dactylifera* L.) during the collection period. Food Biophysics. 2015;**10**(3):360-367

[4] Santoro A, Venturi M, Ben Maachia S, Benyahia F, Corrieri F, Piras F, et al. Agroforestry heritage systems as agrobiodiversity hotspots. The case of the mountain oases of Tunisia. Sustainability. 2020;**12**(10):4054

[5] Rhouma A. Le palmier dattier en Tunisie. Le patrimoine génétique Vol I. Tunis, Tunisie: Arabesques Editions et Créations; 1994

[6] Rhouma A. Le palmier dattier en Tunisie. Rome, Italy: IPGRI; 2005

[7] Doebley John F, Brandon S, Bruce Smith D. The molecular genetics of crop domestication. Cell. 2006:1309-1321

[8] Zohary D, Hopf M, Weiss E. Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in Southwest Asia, Europe, and the Mediterranean Basin. 4th ed. Oxford: Oxford University Press; 2012. DOI: 10.1093/acprof:osobl/9780199549061.001.0001

[9] Zehdi-Azouzi S, Cherif E, Moussouni S, Gros-Balthazard M, Abbas Naqvi S, Ludeña B, et al. Genetic structure of the date palm (*Phoenix dactylifera*) in the Old World reveals a strong differentiation between eastern and western populations. Annals of Botany. 2015;**116**(1):101-112

[10] Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, et al. The tortoise and the hare II: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. American Journal of Botany. 2005;**92**(1):142-166

[11] Zeng K, Fu YX, Shi S, Wu CI. Statistical tests for detecting positive selection by utilizing high-frequency variants. Genetics. 2006;**174**:1431-1439

[12] Kimura M. Evolutionary rate at the molecular level. Nature. 1968;**217**:624-626

[13] King JL, Jukes TH. Non-darwinian evolution. Science. 1969;**164**(3881):788-798

[14] Shen P, Wang F, Underhill PA, Franco C, Yang WH, Roxas A, et al. Population genetic implications from sequence variation in four Y chromosome genes. Proceedings of the National Academy of Sciences USA. 2000;**97**(13):7354-7359

[15] Filiz E. Phylogeny of some *Solanum* species (*Solanaceae*) based on complete chloroplast genomes (cpDNA) and individual chloroplast genes. Research in Biotechnology. 2012;**3**(6):33-41

[16] Rhouma-Chatti S, Choulak S, Zehdi-Azzouzi S, Chatti K, Said K. Molecular polymorphism and phylogenetic relationships within

Tunisian date palm (*Phoenix dactylifera* L.): Evidence of non-coding *trn*L-*trn*F regions of chloroplast DNAs. Scientia Horticulturae. 2014;**170**:32-38

[17] Shew J, Lickfy BE, Schilling EE. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms. American Journal of Botany. 2007;**94**(3):275-288

[18] Al-Qurainy F, Khan S, Al-Hemaid FM, Ali MA, Tarroum M, Ashraf M. Assessing molecular signature for some potential date (*Phoenix dactylifera* L.) cultivars from Saudi Arabia, based on chloroplast DNA sequences rpoB and psbA-trnH. International Journal of Molecular Sciences. 2011;**12**:6871-6880

[19] Ahmed A, Enan RM. DNA barcoding based on plastid matK and RNA polymerase for assessing the genetic identity of date (*Phoenix dactylifera* L.) cultivars. Genetics and Molecular Research. 2014;**13**(2):3527-3536

[20] Ahmed A, Enan RM. Cultivar-level phylogeny using chloroplast DNA barcode psbK-psbI spacers for identification of Emirati date palm (*Phoenix dactylifera* L.) varieties. Genetics and Molecular Research. 2016;**15**(3)

[21] Chase MW, Salamin N, Wilkinson M, Dunwell JM. Land plants and DNA barcodes: Short-term and long-term goals. Philosophical Transactions of the Royal Society B: Biological Sciences. 2005;**360**:1889-1895

[22] Chen S, Yao H, Han J, Liu C. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS One. 2010;**5**:e8613

[23] Hagar-A H, Gadalla EG, Haggag-S Z, Abu El-Maaty S, Mona HH. Identification of some cultivars of Egyptian date palm (*Phoenix dactylifera* L.) using DNA barcoding. Plant Archives. 2020;**20**:1807-1181

[24] Song Y, Wang S, Ding Y, Xu J, Li MF, Zhu S, et al. Chloroplast genomic resource of *Paris* for species discrimination. Scientific Reports. 2017;**7**(1):3427

[25] Balfourier F, Imbert C, Charmet G. Evidence for phylogeographic structure in Lolium species related to the spread of agriculture in Europe. A cpDNA study. Theoretical and Applied Genetics. 2000;**101**(1):131-138

[26] Yang M, Zhang X, Liu G, Yin Y, Chen K, Yun Q, et al. The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). PLoS One. 2010;**5**(9):e12762

[27] Dellaporta SL, Wood J, Hicks JB. A plant DNA mini preparation: Version II. Plant Molecular Biology Reporter. 1983;**1**:19-21

[28] Sambrook J, Frithsch EF, Maniatis T. Molecular Cloning: A Laboratory Manual. second ed. Cold Spring: Cold Spring Harbor Laboratory; 1989

[29] Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences USA 1977;**74**(12):5463-7. DOI: 10.1073/pnas.74.12.5463

[30] Altschul SF, Thomas LM, Alejandro AS, Jinghui Z, Zheng Z, Webb M, et al. Gapped BLAST and PSI-BLAST a new generation of protein database search programs. Nucleic Acid Research. 1997;**25**:3389-3402

[31] Kumar S, Dudley J, Nei M, Tamura K. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinformatics. 2008;**9**:299-306

[32] Librado P, Rozas J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;**25**:1451-1452

[33] Nei M, Tajima F. Maximum likelihood estimation of the number of nucleotide substitutions from restriction sites data. Genetics. 1983;**105**:207-217

[34] Jukes TH, Cantor CR. Evolution of protein molecules. In: Munroled HN, editor. Mammalian Protein Metabolism. New York: Academy Press; 1996. pp. 31-132

[35] Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989;**123**:585-595

[36] Fu YX, Li WH. Maximum likelihood estimation of population parameters. Genetics. 1993;**133**:693-709

[37] Rogers AR, Harpending H. Population growth makes waves in the distribution of pairwise genetic differences. Molecular Biology and Evolution. 1992;**9**:552-569

[38] Fu YX. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics. 1997;**147**:915-925

[39] Harpending HC. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. Human Biology. 1994;**1994**:591-600

[40] Ramos-Onsins SE, Rozas J. Statistical properties of new neutrality tests against population growth. Molecular Biology and Evolution. 2002;**23**(8):1642

[41] Bandelt HJ, Forster P, Rohl A. Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution. 1999;**16**:37-48

[42] Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular Biology and Evolution. 2007;**24**:1596-1599

[43] Boivin T, Bouvier JC, Beslay D, Sauphanor B. Variability in diapause propensity within populations of a temperate insect species: Interactions between insecticide resistance genes and photoperiodism. Biological Journal of the Linnean Society. 2004;**83**(3):341-351

[44] Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proceedings of the National Academy of Sciences USA. 2004;**101**:11030-11035

[45] Bakker FT, Culham A, Gomez Martinez R, Carvalho J, Compton J, Dawtrey R, et al. Patterns of nucleotide substitution in angiosperm cpDNA*trn*L (UAA)-*trn*F (GAA) regions. Molecular Biology and Evolution. 2000;**17**:1146-1155

[46] Choulak S, Marzouk Z, Chatti K, Rhouma-Chatti S, Ouni R, Chatti N. Genetic diversity, phylogeography and population gene flow of Tunisian *Pistacia vera* L. Turkish Journal of Botany. 2019;**43**(6):737-748

[47] Choulak S, Rhouma-Chatti S, Marzouk Z, Said K, Chatti N, Chatti K. Chloroplast DNA analysis of Tunisian pistachio (*Pistacia vera* L.): Sequence variations of the intron *trn*L (UAA). Scientia Horticulturae. 2015;**191**:57-64

[48] Hoot SB, Douglas AW. Phylogeny of the Proteaceae based on *atp*B and *atp*B-*rbc*L intergenic spacer region sequences. Australian Systematic Botany. 1998;**11**:301-320

[49] Manen JF, Natali A. Comparison of the evolution of ribulose-1,5-biphosphate carboxylase (*rbc*L) and *atp*B-*rbc*L non-coding spacer sequences in a recent plant group, the tribuRubieae (Rubiaceae). Journal of Molecular Evolution. 1995;**41**:920-927

[50] Morton BR, Clegg MT. Neighboring base composition in strongly correlated with base substitution bias in a region of the chloroplast genome. Journal of Molecular Evolution. 1995;**41**:597-603

[51] Kingman JFC. The coalescent. Stochastic Processes and their Applications. 1982a;**13**:235-248

[52] Kingman JFC. On the genealogy of large populations. Journal of Applied Probability. 1982b;**19**:27-43

[53] Slatkin M, Hudson RR. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics. 1991;**129**:555-562

[54] Harpending H, Rogers A. Genetic perspectives on human origins and differentiation. Proceedings of the National Academy of Sciences USA. 2000;**96**:10597-10602

[55] Zhang LJ, Cai WZ, Luo JY, Zhang S, Wang CY, Lv LM, et al. Phylogeographic patterns of *Lyguspratensis* (Hemiptera: Miridae): Evidence for weak genetic structure and recent expansion in Northwest China. PLoS One. 2017;**12**(4):e0174712

[56] Rhouma S, Dakhlaoui-Dkhil S, Ould Mohamed Salem A, Zehdi-Azouzi S, Rhouma A, Marrakchi M, et al. Genetic diversity and phylogenic relation-ships in date palms (*Phoenix dactylifera*L.) as assessed by Random Amplified Microsatellite Polymorphism markers (RAMPOs). Scientia Horticulturae-Amsterdam. 2008;**117**:53-57

[57] Rhouma S, Zehdi S, Ould Mohamed Salem A, Rhouma A, Marrakchi M, Trifi M. Genetic diversity in ecotypes of Tunisian date palm (*Phoenix dactylifera* L.) assessed by AFLP markers. Journal of Horticultural Science and Biotechnology. 2007;**82**(6):929-933

[58] Sakka H, Baraket G, Dakhlaoui Dkhil S, Zehdi-Azzouzi S, Salhi-Hannachi A. Chloroplast DNA analysis in Tunisian date-palm cultivars (*Phoenix dactylifera* L.): Sequence variations and molecular evolution of *trn*L (UAA) intron and *trn*L (UAA) *trn*F (GAA) intergenic spacer. Scientia Horticulturae. 2013;**164**:256-269

[59] Sakka H, Zehdi S, Ould Mohamed SA, Rhouma A, Marrakchi M, Trifi M. Genetic polymorphism of plastid DNA in Tunisian date-palm germplasm (Phoenix dactylifera L.) detected with PCR-RFLP. Genetic Resources and Crop Evolution. 2004;**51**(5):479-487

[60] Zehdi S, Cherif E, Rhouma-Chatti S, Santoy S, Salhi Hannachi A, Pintaud JC. Molecular polymorphism and genetic relationships in date palm (*Phoenix dactylifera* L.): The utility of nuclear microsatellite markers. Scientia Horticulturae- Amsterdam. 2012;**148**:255-263

[61] Saboori S, Noormohammadi Z, Sheidai M, Marashi S. Date palm (*Phoenix dactylifera* L.) cultivar relationships based on chloroplast genotyping. Iranian Journal of Science and Technology, Transactions A: Science. 2021;**45**(3):833-840

[62] Flowers JM, Hazzouri KM, Gros-Balthazard M, Mo Z, Koutroumpa K,

Perrakis A, et al. Cross-species hybridization and the origin of North African date palms. Proceedings of the National Academy of Sciences. 2019;**116**(5):1651-1658

[63] Gros-Balthazard M, Newton C, Ivorra S, Pierre MH, Pintaud JC, Terral JF. The domestication syndrome in *Phoenix dactylifera* seeds: Toward the identification of wild date palm populations. PLoS One. 2016;**11**:e0152394

[64] Hamza H, Jemni M, Benabderrahim MA, Mrabet A, Touil S, Othmani A, et al. Date palm status and perspective in Tunisia. In Date Palm Genetic Resources and Utilization. 2015;**2015**:193-221

[65] Kearney TH. Date Varieties and Date Culture in Tunisia. Washington: USDA Bureau of Plant Industry Bulletin; 1906

[66] Moussouni S, Pintaud JC, Vigouroux Y, Bouguedoura N. Diversity of Algerian oases date palm (*Phoenix dactylifera* L., Arecaceae): Heterozygote excess and cryptic structure suggest farmer management had a major impact on diversity. PLoSONE. 2017;**12**(4):e0175232