

RESEARCH

Open Access



Whole mitochondrial and chloroplast genome sequencing of Tunisian date palm cultivars: diversity and evolutionary relationships

Hammadi Hamza^{1*}, Sara Villa^{2†}, Sara Torre², Alexis Marchesini^{2,3,4}, Mohamed Ali Benabderrahim¹, Mokhtar Rejili^{5,6} and Federico Sebastiani^{2*}

Abstract

Background Date palm (*Phoenix dactylifera* L.) is the most widespread crop in arid and semi-arid regions and has great traditional and socioeconomic importance, with its fruit well-known for its high nutritional and health value. However, the genetic variation of date palm cultivars is often neglected. The advent of high-throughput sequencing has made possible the resequencing of whole organelle (mitochondria and chloroplast) genomes to explore the genetic diversity and phylogenetic relationships of cultivated plants with unprecedented detail.

Results Whole organelle genomes of 171 Tunisian accessions (135 females and 36 males) were sequenced. Targeted bioinformatics pipelines were used to identify date palm haplotypes and genome variants, aiming to provide variant annotation and investigate patterns of evolutionary relationship. Our results revealed the existence of unique haplotypes, identified by 45 chloroplastic and 156 mitochondrial SNPs. Estimation of the effect of these SNPs on genes functions was predicted *in silico*.

Conclusions The results of this study have important implications, in the light of ongoing environmental changes, for the conservation and sustainable use of the genetic resources of date palm cultivars in Tunisia, where monoculture threatens biodiversity leading to genetic erosion. These data will be useful for breeding and genetic improvement programs of the date palm through selective cross-breeding.

Keywords Cultivars, Intra-specific genetic diversity, Organellar genome sequencing, Phylogenetic relationships, *Phoenix dactylifera* L., Single nucleotide polymorphisms

[†]Hammadi Hamza and Sara Villa contributed equally to this work.

*Correspondence:

Hammadi Hamza
hamza.hammedi@ira.rnrt.tn
Federico Sebastiani

federico.sebastiani@ips.cnr.it

¹Arid and Oases Cropping Laboratory, Arid Regions Institute, Route du Djorf, Medenine 4119, Tunisia

²Institute for Sustainable Plant Protection, National Research Council of Italy (CNR), via Madonna del Piano 10, Sesto Fiorentino, Florence 50019, Italy

³Present address: Research Institute on Terrestrial Ecosystems (IRET), National Research Council of Italy (CNR), via Marconi 2, Porano, Terni 05010, Italy

⁴NBFC, National Biodiversity Future Center, Piazza Marina 61, Palermo 90133, Italy

⁵Laboratory of Biodiversity and Valorization of Arid Areas Bioresources (BVBAA) – Faculty of Sciences of Gabes, University of Gabes, Erriadh, Gabes 6072, Tunisia

⁶Department of Biology, College of Sciences, Al Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11623, Saudi Arabia



Background

Date palm (*Phoenix dactylifera* L.) belongs to the Arecaaceae family and is the most widely grown crop in arid and semi-arid regions, contributing to feed the population of several countries [1]. It is cultivated especially across North Africa, the Middle East, Arabian Gulf and western parts of South Asia [2, 3], with a number of known cultivars ranging between 2,000 and 5,000 [4], for a global production of date fruits amounting to about 9.8 million tons in 2021 [5]. Due to its high nutritional values, many agri-food technologies use date fruit byproducts as food ingredients as well [6]. This species is one of the earliest domesticated fruit tree crops, with archaeological findings from the Persian Gulf showing evidence of cultivation dating back to around 6,000 years ago [7, 8]; however, in spite of many years of research in a variety of disciplines, the details of its domestication history and diversification has remained a question without a definite answer [2].

In the southern Tunisia, the total number of date palm trees growing is about 6 million, with more than 200 cultivars represented. Such genetic heritage is the result of centuries of artificial and natural selection processes involving domestication, migration, breeding and environmental effects [9]. Consequently, date palm cultivars diverged developing specific agronomic traits, among which the most prevalent are fruit yield and quality. Nevertheless, national and international market pressure has driven date palm cultivation in Tunisia towards monoculture, with many cultivars that become unfamiliar even to the rural population: this threatens date palm biodiversity and can potentially lead to genetic erosion and loss of valuable genetic resources. In the light of ongoing climate change and other potential pressures (e.g. pathogens and pests), adaptation measures for crops, e.g. the selection and use of varieties with different environmental (or diseases) tolerances, are becoming increasingly essential and therefore the preservation of germplasm diversity is an urgent priority [10].

The date palm is a dioecious species, i.e. with naturally cross-pollination between male and female trees. Artificial pollination is required in commercial date production for ensuring good fruit yields and quality [11]. Date palms can be propagated either from seeds or vegetative offshoots. However, seed propagation implies that half of the progeny are males and there are currently no effective methods for sex determination of young date palm seedlings: it is therefore unsuitable for commercial production, which relies on offshoots planting. Palms are long-living plants and both seeds and vegetative offshoots are characterized by an extremely slow growth, with female palms starting to produce fruits only after 5–10 years, depending to the cultivar [12]. Furthermore, female cultivars have been so far identified based on

fruit morphology and given local names mostly according to geographical origin, with the result of potentially different names corresponding to the same genotype. On the other hand, since variety/cultivar information is closely linked to the marketing of the fruit produced, this is traditionally neglected for male plants. All the above-mentioned peculiar features make genetic improvement (e.g. to enhance fruit production) difficult with the use of classical breeding techniques [13]. In angiosperms, both chloroplasts and mitochondria genomes are usually inherited from maternal parent, leading to a non-Mendelian inheritance pattern [14]. While chloroplast genomes are highly conserved, generally organized into circular DNA molecules varying between 120 and 160 kb in length [15, 16], plant mitochondrial genomes are highly variable in terms of both length and gene content [17]. The full-length organelle genomes sequencing can unveil DNA markers for cultivars discrimination and may allow the identification of protein-coding regions potentially involved in agronomically important traits. Within the species of agronomic interest, the date palm was among the first whose complete mitochondria and chloroplast genomes were sequenced and published (715,001 bp and 158,462 bp, respectively) [18–20]. The release of reference genomes enables the identification of Single Nucleotide Polymorphisms (SNPs) markers, which in turn can be used for assessing whole-genome diversity and genetic relationships between cultivars. Whole organellar genomes have been used in a great variety of studies to investigate genetic differentiation and evolution of plant species [21, 22]. Their widespread use could be attributed to their usual uniparental inheritance, high copy-number, and low frequency of recombination [23, 24]. Over the past decade, there have been numerous efforts to use molecular markers to characterize date palm biodiversity. However, most of these studies have been based on a limited set of traditional genetic markers (i.e. RAPD, SSR and AFLP or short chloroplast regions) [25, 26], including the so far performed genetic surveys focusing on Tunisian cultivars [27, 28]. Novel genomic approaches, allowing an in-depth knowledge and functional characterization of the genome, have proven to outperform traditional markers in assisted selection for plant breeding, also considering that many agronomically important traits are complex and affected by multiple genes [29, 30]. The advent of next-generation sequencing (NGS) has made possible the resequencing of whole genomes to explore the genetic diversity of plants. Among various NGS technologies, the use of rapid and cost-effective genome-skimming strategy, which consists of sequencing the genomic DNA of an individual at low nuclear genome coverage, has become increasingly popular in plant phylogenomics [21, 31–33]. This approach, capable of generating high-copy fractions of total genomic DNA

(including organelle genomes, nuclear ribosomal DNA and other multi-copy elements), was chosen for our study because it facilitates the creation of extensive phylogenetic datasets at relatively low cost and without the need for massive bioinformatics resources. To date, little research has been carried out based on whole genome sequencing (WGS) of date palm organelle genomes. The intraspecific diversity of Tunisian date palm cultivars was previously investigated by using nuclear and organelle markers [34–37], but variation at whole genome sequence remained largely unexplored. The phylogenetic relationships among nine Saudi Arabian cultivars were analyzed in 2014 [38] by means of organelle genome re-sequencing. Mohamoud et al. [3] used organelle WGS in a broad scale study across the main date palm growing regions (North Africa, the Middle East, Arabian Gulf and western parts of South Asia) to investigate the history and origin of date palm cultivation through variation in organelle genomes of different cultivars. Specifically, the latter study used only five Tunisian cultivars (i.e. Allig, Hamrata, Grenja, Nefred and Deglet Nour) and found that they could be divided into two neighboring haplotypes: NA1 comprising Allig, Hamrata and Grenja, and NA2 comprising Nefred and Deglet Nour, respectively.

Finally, the analysis of chloroplast genomes not only allows to solve the deeper branches of plants phylogeny and reconstruct phylogeographic patterns, but also contributes to the development of more effective DNA barcode markers for molecular characterization of germplasm and food products [39], and to the screening of genetic resources for breeding and genetic engineering [40].

This work includes the first massive resequencing of both organelle genomes of date palm accessions from different cultivars collected from three Tunisian continental oases, which represent the most important Tunisia's date palm production sites [34, 41]. Targeted bioinformatics pipelines were used to identify haplotypes, genome variants and annotation and to study patterns of evolutionary relationship. Finally, functional annotation of genomic variants was carried out *in silico* to estimate the SNPs effects on genes functions. Overall, the genetic resources developed in this research could be used for further phylogeographic and population genetics studies of date palms in different geographical areas. The long-term goal is to contribute to the creation of a valuable platform for assisting genetic breeding and industrial exploitation of this economically important monocot tree.

Results

Out of the 171 samples analyzed, 135 represent female plants and 36 represent males (Table S1). Among females, only 68 accessions (~50%) were successfully attributed to a certain cultivar based on fruit traits. In particular, the

cultivars Ammary, Hissa and Gosbi are characterized by early-ripening soft fruits (ripening period July 15th - August 15th), while Besser Helou, Kintichi and Kenta produce dry fruits with mid-season (August 15th – September 30th) or late (after September 30th) maturation. Deglet Nour is characterized by semi soft fruits and late maturation, which historically took place in December in the Tunisian oases but has been brought forward to October in the last 20 years due to rising temperatures [42]. Each accession was identified as a different cultivar, with the following exceptions: six samples attributed to the cultivar Alig, three samples identified as Besser Helou, two samples as Deglet Nour, three samples as Ghars Souf, two samples as Gondi, 11 samples as Gosbi and two samples identified as Khwat Alig, for a total of 42 different cultivars (Table S1).

Chloroplast and mitochondrial genomic haplotype analysis

We generated ~658 Gb of raw data with ~4.4 billion reads. After filtering low quality bases, ~652 Gb of clean data were obtained, with a clean base rate of 99.02% (Table S1). We obtained an average sequence depth of 761x for chloroplast genomes and 665x for mitochondrial genomes and no sample resulted to have less than 96x of average depth across the two organelle genomes. After filtering and cleaning, we identified a total of 201 SNPs in the 171 individuals, 45 SNPs in the chloroplast genome and 156 in the mitochondrial genome (Fig. S1). This set of markers and individuals was used in subsequent analyses.

Analysis of SNPs from the chloroplast genome identified 9 haplotypes (Table 1; Fig. 1A): H1 included most of the samples (88, 51.46%), followed by H3 (35, 20.47%), H6 (21, 12.28%), H2 (17, 9.94%), H7 (5, 2.92%) and H8 (2, 1.17%); H4, H5 and H9 were found only in single samples. As regards the cultivars represented by multiple accessions, the correspondence between cultivar and haplotype is not always linear. For Deglet Nour, Ghars Souf, Gondi and Khwat Alig, all accessions of the same cultivar were actually attributed to the same haplotype (H3, H2, H1 and H1, respectively; Table S1). Conversely, out of the six accessions identified as Alig cultivar, four were attributed to H1 and two to H6. The three accessions identified as Besser Helou were attributed to three different haplotypes (H1, H3 and H6). Finally, out of the 11 Gosbi accessions, two were attributed to H6, and the others to H1.

A total of 14 haplotypes were identified from the analysis of mitochondrial genome polymorphisms (Table 1; Fig. 1B): H1 haplotype was the most represented, with 75 individuals, followed by H3, with 40 individuals, H4 with 25 individuals, H2 with 19, and H5 with 3; all other haplotypes were represented by only one individual. The Alig samples were attributed to three different haplotypes

Table 1 Haplotype and nucleotide diversity in chloroplast and mitochondrial genomes found in sampled individuals, grouped for the oasis of origin of the sample

Oasis	Number of individuals	Chloroplast				Mitochondrial			
		Haplotypes	Private haplotypes	Haplotype diversity	Nucleotide diversity	Haplotypes	Private haplotypes	Haplotype diversity	Nucleotide diversity
Atillet	120	8	3	0.7164	1.32×10^{-4}	12	7	0.7287	8.75×10^{-5}
Jerid	22	3	0	0.3853	1.01×10^{-4}	6	2	0.7359	6.83×10^{-5}
Nefzaoua	29	6	1	0.5936	1.20×10^{-4}	5	0	0.6823	7.92×10^{-5}
TOTAL	171	9				14			

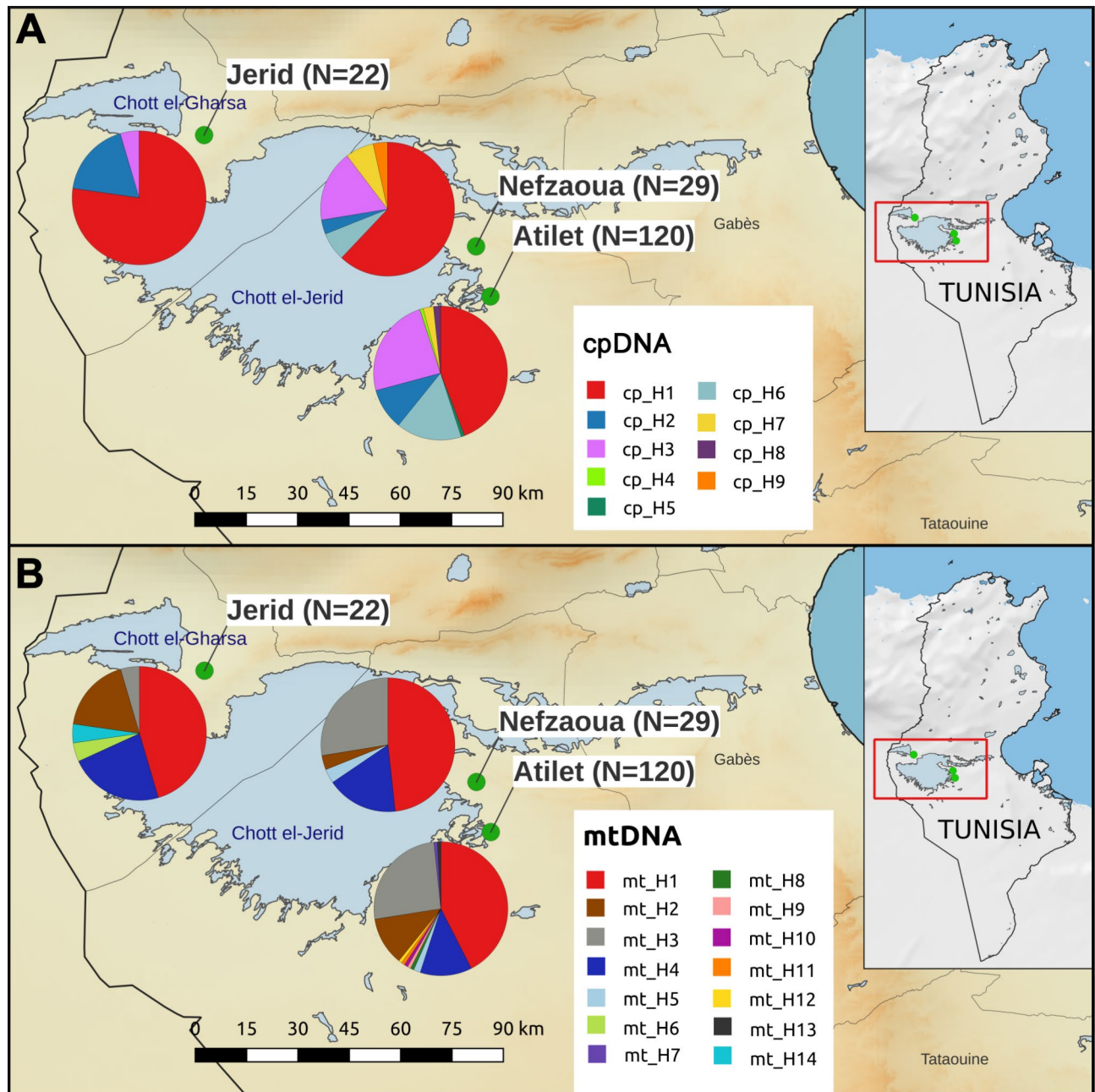


Fig. 1 Haplotypes map. Map of the three sampled Tunisian date palm oases (green dots), with the detected haplotype frequencies for both (A) chloroplast (cpDNA) and (B) mitochondrial (mtDNA) genomes. The number of samples from each site is reported in brackets

(H4, H8 and H11) and Besser Helou samples to H1 and H3, while all the Deglet Nour, Ghars Souf, Gondi, Gosbi and Khwat Alig samples were attributed to the same haplotypes (H3, H2, H1, H1 and H4, respectively; Table S1). Overall, 20 haplotypes were resolved by combining information from the two organelle compartments.

The group of accessions sampled in Atillet oasis (N=120) showed the highest number of haplotypes and private haplotypes for both the chloroplast (8 and 3, respectively) and mitochondrial genomes (12 and 7, respectively; Table 1). Accessions sampled in Jerid (N=22) showed 3 chloroplast haplotypes and 6 mitochondrial haplotypes, 2 of which were private. Nefzaoua samples (N=29) showed 6 chloroplast haplotypes, 1 of which was private, and 5 mitochondrial haplotypes. The level of haplotype diversity related to the mitochondrial genome was high and of the same magnitude in all oases (Table 1), while for the chloroplast genome it was high in Atillet oasis (0.7164), intermediate in Nefzaoua (0.5936) and lower in Jerid (0.3853). The highest level of nucleotide diversity was observed in the chloroplast genome, ranging from 1.32×10^{-4} (accessions from Atillet) to 1.01×10^{-4} (Jerid); accessions from Jerid also showed the lowest level of nucleotide diversity (6.83×10^{-5}) for the mitochondrial genome, however in the same order of magnitude of the other two oases, Nefzaoua and Atillet (7.92×10^{-5} and 8.75×10^{-5} , respectively).

Analysis of the chloroplast and mitochondrial haplotype networks identified two main groups, in both cases centered on the H1 and H3 haplotypes, separated by several mutations, while the other, less frequent haplotypes were arranged as satellites around these two (Fig. 2A,B). The number of SNPs found between haplotypes is detailed in Table S2. Among haplotypes identified in the chloroplast genome, a maximum variation of 34 nucleotides was detected (between H4 and H5) and, similarly to the network, the neighbor-joining tree grouped the haplotypes in two clusters with high branch support (bootstrap value=100%; Fig. 2C). Within haplogroup 1 (comprising H1, H5, H6), haplotypes are differentiated by 1 SNP; within haplogroup 2, a maximum variation of 7 nucleotides was detected between the remaining haplotypes (Fig. 2A; Table S2). Similarly, two highly supported clusters were observed in the neighbor-joining tree of the mitochondrial haplotypes (Fig. 2D), reflecting a maximum variation of 88 nucleotides (between H7 and H14). Within each mitochondrial haplogroup, a maximum variation of 37 nucleotides was detected; again, there are haplotypes distinguished by only one nucleotide, such as the pairs H5-H8, H5-H9, H1-H10, H1-H14, H1-H4 (Fig. 2B; Table S2).

In agreement with previous analysis, the phylogenetic relationships between organelle haplotypes led to the clustering of the 171 date palm samples in two

phylogroups (Fig. 3A,B; Fig. S2). The larger group showed more homogeneity with no or very short internal branching, while the smaller group showed more pronounced internal subgroups. The two phylogroups corresponded to the two main groups found in the haplotype networks (Fig. S2). Both haplogroups were well represented in the three oases, although haplogroup 2 was always lower in frequency: it comprised the 23%, the 31% and the 39% of the samples in Jerid, Nefzaoua and Atillet, respectively (Fig. S3).

Finally, the analysis of plastidial SNPs revealed an exact match in alleles and positions along the sequence between haplotypes H1 and H3 and haplotypes NA1 and NA2 already identified by Mohamoud et al. [3] (Fig. S1). As regards mitochondrial sequences, the high number of indels did not allow a direct comparison.

Functional annotation of variants

The mitochondrial genome had a variant rate of 1 in every 4583 bases. Upon annotation of 156 mt SNPs, we found a total of 371 effects (Fig. 4), as individual variants can exhibit more than one effect. Based on their impact, the analysis revealed that only 0.3% of SNPs have a low impact (i.e., are generally harmless or unlikely to modify protein behavior), while the great majority were non-coding variants or variants affecting non-coding genes. The gene with the highest number of variants having a modifier impact is *nad2*. The distribution of SNPs in the mitochondrial genome revealed a higher concentration in the intron regions (57% of SNPs were classified as intron variants). The effects of the remaining SNPs portion were identified as upstream, intergenic, intragenic, and downstream gene variants. Only one mutation was found in the exon region and produced a synonymous variant in an RNA polymerase gene.

There were 405 potential effects produced by the 45 chloroplast SNPs. The plastome had a SNPs density of 1 variant every 3521 bases. Consistent with the mt genome, the main class of impact identified was modifiers, but small percentages of SNPs with moderate, low, and high impact (1.5%, 2% and 0.5%, respectively) were also detected in the cp. genome. Concerning the effect types of SNPs, the results showed that most of the genomic variations were located in the upstream and downstream regions of the genes. A small percentage of SNPs (4%) was found in the exon regions, producing mainly synonymous and missense variants (causing an amino acid substitution). These variants, deemed to be of moderate impact, occurred in 4 genes: *ycf1*, *rpl16*, *rpoC1* and *rpoC2* genes. Among these, the *ycf1* gene appears to be the most affected, as it contains three missense SNPs. Furthermore, SNPs with a stop-gain effect (0.5%), which have a disruptive impact on the protein, were identified in the genomic variants of cp., involving the *petD* and

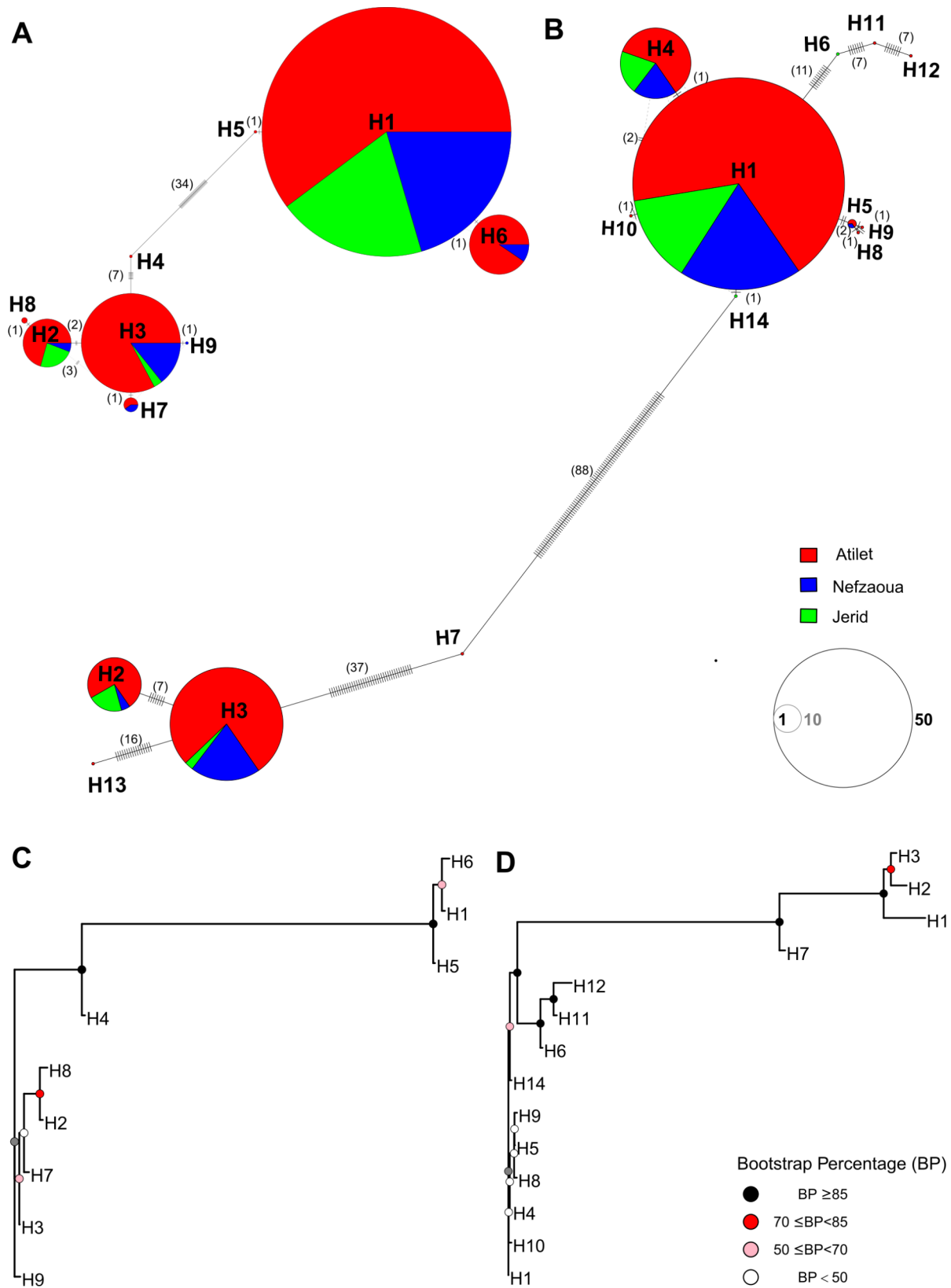


Fig. 2 Haplotype networks and neighbor-joining trees. Chloroplast and mitochondrial haplotype networks (**A** and **B**, respectively) and neighbor-joining trees (**C** and **D**, respectively). In the haplotype network colors indicate the sampled oases and circle size is proportional to the number of individuals (circle sizes for 1, 10 and 50 individuals are shown in different shades of grey). Black bars on the branches and numbers in brackets indicate the number of SNPs between haplotypes

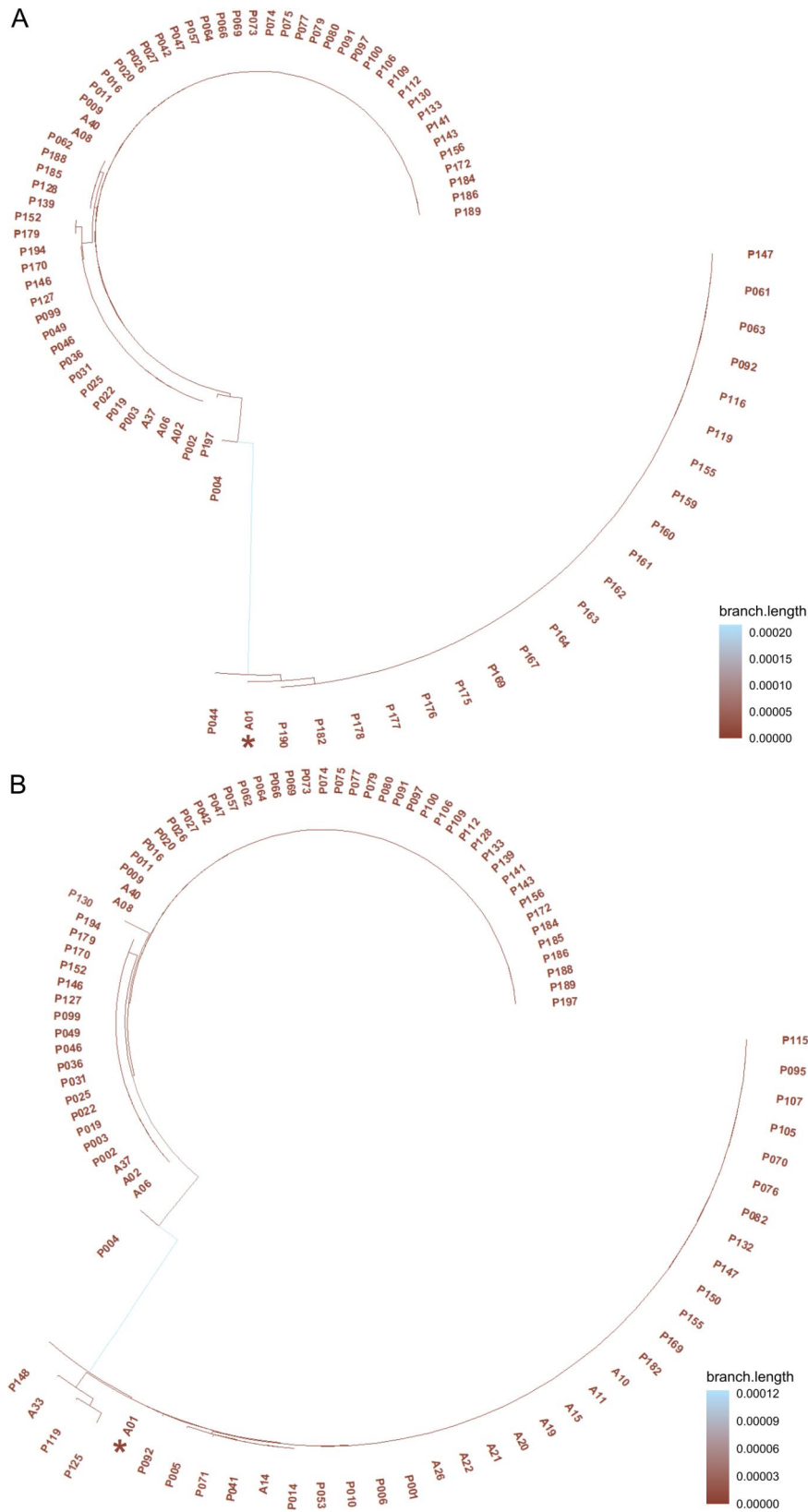


Fig. 3 Phylogenetic trees. Phylogenetic tree of date palm samples based on chloroplast **(A)** and mitochondrial **(B)** sequences, built with PEGAS following Toparslan et al. [43]. To improve readability, the entire H1 haplotype is represented by sample A01 only, highlighted by an asterisk. Both the analyses showed the existence of two main phylogroups: one larger and homogeneous, the other smaller and more variable

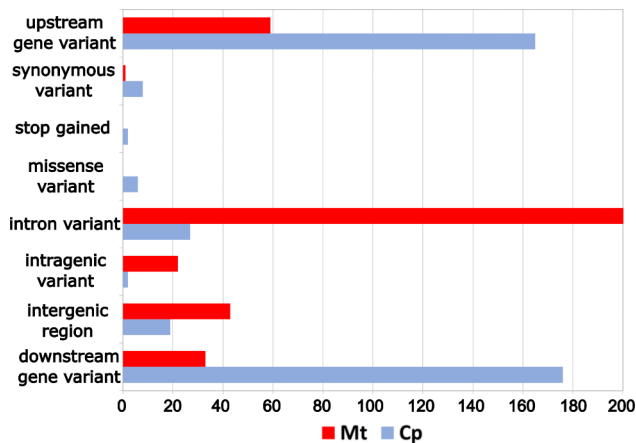


Fig. 4 SNP classes. SNPs classification according to their effect and distribution across date palm chloroplast (Cp) and mitochondrial (Mt) genomes. The x axis displays the number of predicted effects, and the y axis the different effects specified using the sequence ontology terms

rpl16 genes. Finally, the Ts/Tv ratio was 0.11 for the mitochondrial genome and 1.24 for the chloroplast genome. To verify this statistic, we calculated the organellar Ts/Tv ratio for date palm with data from Mohamoud et al. [3]. We obtained a difference between Ts/Tv of cp. and mt of the same magnitude as our data: mitochondrial Ts/Tv=0.03 and plastidial Ts/Tv=1.6.

Discussion

As in other North African regions [43], the Tunisian oases system is characterized by a large number of cultivars that derive from selection operated within the oases, aiming to optimize fruit production and quality under different environmental contexts. Identification of date palm cultivars based on a morphological approach is muddled by the high adaptive flexibility of the species, and relies on the expertise and the traditional knowledge of local growers [44]. This is further complicated by the many names that characterize differences related to fruit quality and morphology. Moreover, the lack of variety information for male individuals and consequently the reliance on the local expertise for the production of quality dates through selective crosses, makes these techniques non-reproducible and the results from different sites non-comparable, also increasing the risk of errors and loss of both genetic and cultural heritage. This study is among the first exploring extensively the Tunisian date palm germplasm, based on whole chloroplast and mitochondrial whole-genome sequencing as a reliable method to assess genetic variation of date palm. The lack of a univocal correspondence between cultivars identified using phenotypic traits and haplotypes is evident from our results, with multiple accessions for the same cultivar attributed to different haplotypes. Moreover, these results do not seem to be related to the sampling

location: considering for example the accessions identified as Gosbi cultivar, some samples from the three different oases were attributed to the same haplotype H1 according to cp. sequences, while samples from the same oasis Atilet were attributed to both haplotypes H1 and H6 (Table S1).

We found that the chloroplast genome is slightly more variable than the mitochondrial genome in terms of SNPs density. Annotation of the identified SNPs in both organellar genomes showed that almost all mutations fall in non-coding regions, as expected due to lower sequence conservation and reduced selection pressure in non-coding regulatory regions compared to coding regions. In mitochondria, most SNPs were identified in intergenic regions, whereas in the cp. genome, SNPs were much more abundant in the upstream and downstream regulatory regions of genes. However, some mutations have been identified in the cp. genome that could have a high or moderate impact on gene function. This small percentage of SNPs with a relevant effect on genes (4%) could be useful for developing DNA markers. Further studies on these polymorphisms could open up prospects for marker-assisted selection, taking into account nuclear-cytoplasmic interactions for improving date palm cultivars and studying adaptation.

The striking difference in the Ts/Tv ratio between cp. and mt genomes in the date palm was confirmed by the same analysis performed with another published date palm dataset [3]. A difference of the same magnitude, although of opposite direction, was observed in soybean [45] and Ginkgo [46] where Ts/Tv were 0.99 and 2.29 in mtDNA and 0.2 and 0.61 in cpDNA, respectively. It has already been reported that the rate of transversions can be slightly higher than that of transitions in the date palm mitochondrial genome [18]. The unusual transition transversion ratio was already observed in other species [47] and again in *Phoenix dactylifera* [38] where ~70% of the mitochondrial substitutions were transversions, probably due to the overwhelming ratio of intergenic regions.

The number of haplotypes observed in our study is small compared to other work on Tunisian cultivars done with specific loci e.g., of the chloroplast genome [27, 28]. Indeed, the applied method is highly conservative, eliminating indel-type and microsatellite polymorphisms, but it is however required to reduce the risks of intra-individual polymorphisms due to the potential phenomena of chloroplastic (nupt) and mitochondrial (numt) DNA transfer to the nucleus [48].

The most frequent haplotype in the Tunisian cultivars was H1, considering both cp. and mt genomes: together with its (less frequent) satellite haplotypes, H1 also constituted the main haplogroup in terms of haplotype frequencies. The second most abundant cp. and mt haplotype was H3, which is close to the haplotypes

characteristic of the eastern date palm growing area [3] and significantly divergent from H1, as shown by the haplotype networks and phylogenetic trees (Figs. 5 and 2). It is worth noting that the polymorphic positions identified in our study correspond exactly to those defining the two North-African main haplotypes identified in the study by Mohamoud et al. [3], with our cp. H1 and H3 haplotypes corresponding to their NA1 and NA2 haplotypes, respectively (Fig. S1). In particular, among the Tunisian cultivars considered, only Alig, Deglet Nour and possibly Hamra (indicated as “Hamrata” by Mohamoud et al. [3]) are present in both the studies. Alig and Hamra(ta) cultivars were grouped into the NA1 haplotype by Mohamoud et al. [3], and into H1 in the present study (with the only exception of samples P119 and P155, attributed to H6), while Deglet Nour was attributed to a significantly different haplotype (NA2 and H3, respectively). In addition, having analyzed more accessions, we found additional haplotypes that are phylogenetically closely related (i.e. separated only by a few mutations) to the main two, from which they probably originated.

Most female accessions (96/135) belong to the H1 haplogroup, supporting the hypothesis that Tunisian date palm production relies on a relatively homogenous group, in terms of organellar genomic variation (Fig. 3, Table S1). The membership of Deglet Nour to the haplogroup centered in H3 supports the hypothesis of [49] that this cultivar was introduced to the Jerid oasis in the 16th century from Algeria. Moreover, our results on the divergence of Deglet Nour from the haplogroup 1 (including the most frequent haplotype) agree with

both Zehdi-Azouzi et al. [36] and Mohamoud et al. [3] that Deglet Nour is genetically closer to the Arabian Gulf haplotype than to the major North African haplotypes. However, most of the male trees belong to H3, along with the Deglet Nour and Besser Helou cultivars. This suggests that the male trees may have been developed from the elite cultivar Deglet Nour (the most ubiquitous cultivar in Tunisia). In fact, Tunisian agricultures do not have a recognized male cultivar and obtain their pollinators from germination of seeds [50] of unknown cultivars. So, it is very likely that seeds of Deglet Nour fruits were the origin of the majority of male trees in Tunisia. In Tunisia, date palm has been cultivated for more than 3,000 years [51]. The domestication center of Tunisian date palm is generally believed to be located in the south-west of the country, next to the Berber and Roman settlements. Then, the distribution and selection of cultivars in different geographic regions were affected by the local environmental conditions such as climate and soil [52]. The date palm propagation patterns gave rise to and maintained a peculiar genetic diversity, and our finding based on whole chloroplast and mitochondrial genomes support this scenario.

The date palm gene pool observed in southwestern Tunisia and the information provided by our annotated organellar genome offer valuable resources for the breeding and genetic improvement of this important fruit tree crop. Indeed, organellar genome sequencing allowed the genetic characterization of both male accessions, whose variety is usually disregarded by growers, but contribute significantly in the offspring phenotype, and female

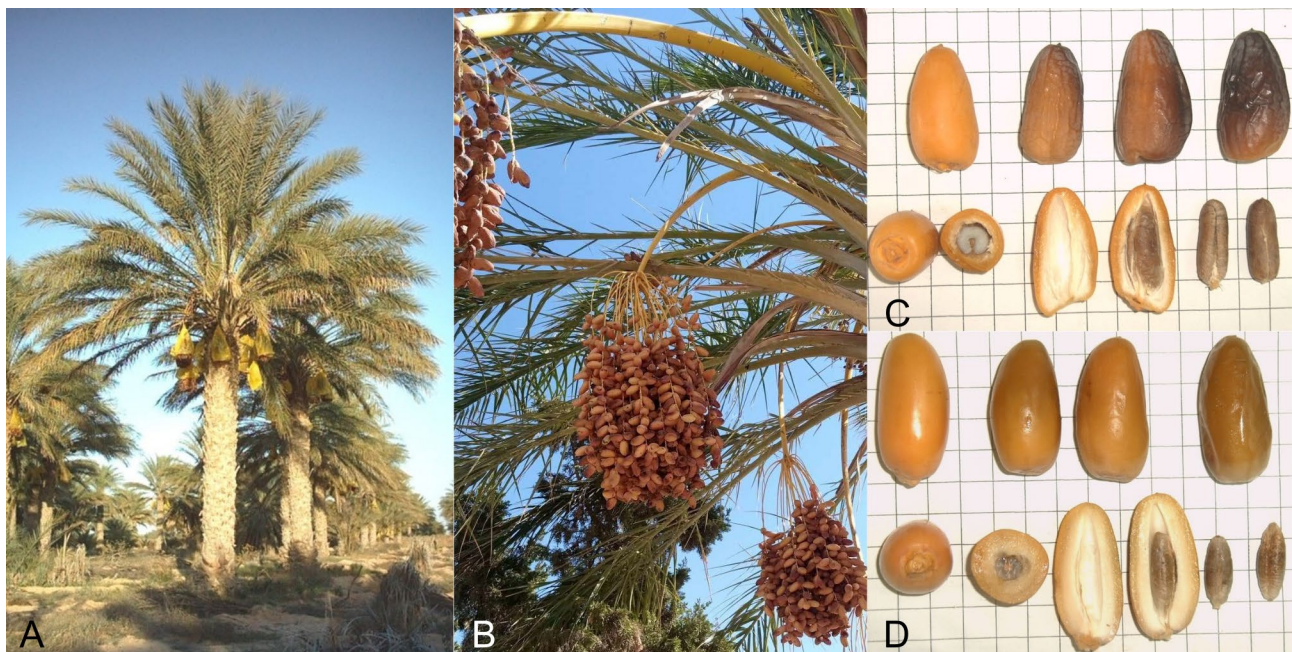


Fig. 5 *Phoenix dactylifera*. Deglet Nour tree (A) and detail of the fruits (B). Morphological differences between Hamra (C) and Deglet Nour (D) dates

accessions whose cultivar that could not be identified in the present study through phenotypic traits. This study is therefore a valuable contribution to the genetic improvement of date palm cultivation through selective cross-breeding.

Conclusion

In this study, we performed an extensive genetic survey of Tunisian date palm accessions by scanning polymorphisms of the complete chloroplast and mitochondrial genomes. Our results unveiled a previously ignored haplotype richness and framed the evolutionary relationships of Tunisian date palm organellar genetic resources within the date palm growing area. Phylogenomics analyses strongly supported two well-separated haplogroups corresponding to the Western including North African and the Eastern range of date palm cultivation. The polymorphisms highlighted in this work would contribute to further studies on genetic diversity and phylogeography of date palm and would help fine-tune DNA barcodes in *Phoenix* species. However, although the abundant genetic resources developed in this work would be valuable for genetic breeding programs and germplasm exploitation of this economically important species, an additional effort is needed at the nuclear genome level to fully explore date palm genetic diversity.

Methods

Plant material collection and DNA extraction

In the present study, sequences from 171 date palm accessions were produced and analyzed. We sampled individual leaves from the main Tunisian continental oases, Jerid and Nefzaoua, and from Ailet, a conservation plot located in the Institute of Arid Regions (IRA, Kébili, Tunisia) which maintains endangered cultivars and is therefore representative for the genetic diversity of Tunisian date palm [44] (Fig. 1). Cultivar identification for female plants, obtained by clonal multiplication, was determined by experienced taxonomists and according to fruit traits (e.g. fruit color and consistency, ripening period) as described by Ferchichi and Hamza [53] (Fig. 5). No cultivar name was assigned to male trees, being derived from seed germination. All voucher specimens with individual labels ("Sample ID" in Table S1) were deposited and are stored permanently in the Laboratory of Dry Land Farming and Oasis Cropping, Institute of Arid Regions (IRA, Kébili, Tunisia). One sample per individual was processed for total genomic DNA extraction, using the Invisorb Spin Plant Mini Kit (Invitex, Berlin, Germany). DNA purity was checked using agarose gel electrophoresis, while DNA concentration was measured using Qubit® DNA Assay Kit in Qubit® 2.0 Fluorometer (Life Technologies, CA, USA). High-quality genomic DNA was stored at -20°C until sequencing.

Sequence analysis: data cleaning, SNP calling

We used a genome skimming approach, sequencing the whole genomic DNA in paired-end mode (2×150 bp reads) at low nuclear genome coverage on an Illumina HiSeq 4000 platform. Library construction and sequencing were conducted by Novogene Co., Ltd. (Beijing, China). Sequencing reads were deposited in GenBank (BioProject ID: PRJNA951931). Raw reads were cleaned using Trimmomatic [54] with the following options: trailing: 10; leading: 10; sliding window: 4:20; seed mismatches: 2; palindrome clip threshold: 30; simple clip threshold: 10 and minlen: 40. Sequence analysis was performed following Scarcelli et al. [48]. The mapping step was performed using BWA MEM 0.7.17-r1188 [55] with the -M and -B 4 options against NCBI reference GU811709.2 for mapping plastidial reads and NC_016740.1 for mapping mitochondrial reads. Samtools 1.10 [55] with -B option was used to generate a mpileup file. We used Varscan 2.3.7 mpileup2snp [56] to call SNPs (options -min-var-freq of 0.5, a -min-freq-for-hom of 0.5, a minimum base quality of 30, and a min coverage of 5). The resulting Variant Call Format file (VCF) was filtered using VCFtools v0.1.16 [57] to exclude indels, missing site positions, heterozygote positions, and individuals with more than 20% of missing data; only biallelic sites with a minimum depth of 5 and a minimum genotype quality of 10 were kept. VCF files were then converted into fasta files for downstream analysis.

Analysis of organellar genomes variation

Chloroplast and mitochondrial haplotypes were inferred with the package PEGAS [58, 59] in R 4.2.1 [60], following the workflow proposed by Toparslan et al. [61]. Genetic relatedness between haplotypes were estimated and visualized through minimum spanning networks. For both organellar genomes, the genetic structure of haplotype diversity and the maximum-likelihood phylogenetic trees of the inferred haplotypes were inferred using the R package *fastbaps* [62]. Finally, alleles in polymorphic sites for cp. and mt haplotypes were visualized converting the VCF file into PHYLIP formatted sequence using vcf2phylip v1.5 [63]. Results were then compared with haplotypes from Mohamoud et al. [3]. Maps of the geographic distribution of detected haplotypes and haplogroups were produced using QGIS v3.20 (<http://www.qgis.org>) and the publicly available NASADEM Digital Elevation Model [64] and Tunisia Inland Waters map [65].

Annotation of SNPs

Functional annotation of chloroplast and mitochondrial genomic variants was performed using SnpEff (version 5.0e; [66]) using the sequence ontology. Since annotations for the date palm genomes are not available in the built-in SnpEff databases, custom annotation databases

were created for each organelle, based on data available in GenBank (NC_016740 and GU811709, for mitochondria and chloroplast, respectively). Variants were annotated according to their effect (high, moderate, modifier and low), functional class (missense, nonsense, silent) and genomic region (downstream, upstream, exon, intron, transcript, intragenic and intergenic). Furthermore, DNA substitution mutations (transitions and transversions) ratio (Ts/Tv) as well as changes in amino acids were identified. We calculated the organellar (Ts/Tv) ratio also for date palm haplotypes reported by Mohamoud et al. [3] using SnpSift [67].

Abbreviations

H	Haplotype (Chloroplast and Mitochondria)
cp	Chloroplast
mt	Mitochondria
NCBI	National Center for Biotechnology Information
SNP	Single Nucleotide Polymorphism
VCF	Variant Call Format file
WGS	Whole Genome Sequencing

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09872-7>.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3

Acknowledgements

We would like to express our very great appreciation to the Arid Regions Institute of Kébili, Tunisia for its help in collecting the vegetal material and all the technicians who helped us in the sampling.

Author contributions

Hammadi Hamza: Conceptualization, Resources, Writing - Original Draft, Project administration, Funding acquisition. Sara Villa: Formal analysis, Data curation, Writing - Review & Editing, Visualization. Sara Torre: Methodology, Formal analysis, Data curation, Writing - Review & Editing, Visualization. Alexi Marchesini: Methodology, Formal analysis, Data curation, Writing - Review & Editing, Visualization. Mohamed Ali Benabderrahim: Resources, Writing - Review & Editing. Mokhtar Rejili: Conceptualization, Resources, Writing - Review & Editing. Federico Sebastiani: Conceptualization, Methodology, Formal analysis, Validation, Investigation, Resources, Data curation, Writing - Review & Editing, Visualization, Supervision, Project administration, Funding acquisition.

Funding

The work was supported by the Italian Ministry of University and Research (legal name: Ministero dell'Università e della Ricerca, MUR) in the framework of the research program "PRIMA (Partnership for Research and Innovation in the Mediterranean Area) - bando Multitopic 2019 - Sect. 2" - "GREENPALM - Development of sustainable date palm-based agro systems by preserving their biodiversity".

Data Availability

The datasets generated during the current study are available in the NCBI repository under BioProject accession PRJNA951931 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA951931/>).

Declarations

Ethics approval and consent to participate

Dr. Hammadi Hamza undertook the formal identification of the plant material used in our study. All voucher specimens with individual labels ("Sample ID" in Table S1) were deposited and are stored permanently in the Laboratory of Dry Land Farming and Oasis Cropping, Institute of Arid Regions (IRA, Kébili, Tunisia). The current study complies with relevant institutional, national, and international guidelines and legislation for experimental research and field studies on plants, including the collection of plant material.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 7 July 2023 / Accepted: 4 December 2023

Published online: 13 December 2023

References

- Johnson DV, Al-Khayri JM, Jain SM. Introduction: Date Production Status and Prospects in Africa and the Americas. In: Al-Khayri J, Jain S, Johnson D, editors. *Date Palm Genetic Resources and Utilization: Volume 1: Africa and the Americas*. Springer Netherlands, Dordrecht; 2015. p. 1–18. https://doi.org/10.1007/978-94-017-9694-1_1
- Gros-Balthazard M, Hazzouri KM, Flowers JM. Genomic insights into date Palm origins. *Genes*. 2018;9:502. <https://doi.org/10.3390/GENES9100502>
- Mohamoud YA, Mathew LS, Torres MF, Younusunju S, Krueger R, Suhre K, Malek JA. Novel subpopulations in date palm (*Phoenix dactylifera*) identified by population-wide organellar genome sequencing. *BMC Genom*. 2019;20:1–7. <https://doi.org/10.1186/s12864-019-5834-7>
- Jaradat AA. 2015. Biodiversity, genetic diversity, and genetic resources of date palm. In: Al-Khayri J, Jain S, Johnson D, editors. *Date Palm Genetic Resources and Utilization: Volume 1: Africa and the Americas*. Springer Netherlands, Dordrecht; 2015. p. 19–71. https://doi.org/10.1007/978-94-017-9694-1_2
- FAO. The State of Food and Agriculture. 2021. <https://doi.org/10.4060/cb4476en>
- Mrabet A, Hamza H, Rodríguez-Gutiérrez G, Jiménez-Araujo A, Sindic M. Date palm fruits as a potential source of functional Dietary Fiber: a review. *Food Sci Technol Res*. 2019;25:1–10. <https://doi.org/10.3136/fstr.25.1>
- Ashraf Z, Hamidi-Esfahani Z. Date and date Processing: a review. *Food Rev Int*. 2011;27:101–33. <https://doi.org/10.1080/87559129.2010.535231>
- Pérez-Escobar OA, Bellot S, Przelomska NAS, Flowers JM, Nesbitt M, Ryan P, et al. Molecular Clocks and Archeogenomics of a late period Egyptian date Palm Leaf Reveal Introgression from Wild relatives and add timestamps on the domestication. *Mol Biol Evol*. 2021;38:4475–92. <https://doi.org/10.1093/molbev/msab188>
- Doebley JF, Gaut BS, Smith BD. The Molecular Genetics of Crop Domestication. *Cell*. 2006;127:1309–21. <https://doi.org/10.1016/j.cell.2006.12.006>
- Snook LK, Dulloo ME, Jarvis A, Scheldeman X, Kneller M. Crop germplasm diversity: the role of Gene Bank Collections in Facilitating Adaptation to Climate Change. In: Yadav SS, Redden RJ, Hatfield JL, Lotze-Campden H, Hall AE, editors. *Crop adaptation to Climate Change*. Oxford, UK: Wiley-Blackwell; 2011. pp. 495–506. <https://doi.org/10.1002/9780470960929.ch34>
- Hamza H, Mrabet A, Jiménez-Araujo A. Date palm parthenocarpic fruits (*Phoenix dactylifera* L.) cv. Deglet Nour: chemical characterization, functional properties and antioxidant capacity in comparison with seeded fruits. *Sci Hortic*. 2016;211:352–7. <https://doi.org/10.1016/j.scienta.2016.09.031>
- Al-Mssallem I. Date palm. *Arab Global Encyclopedia*. 1996;7:182–7.
- Hadrami IE, Hadrami AE. Breeding date palm. In: Jain SM, Priyadarshan PM, editors. *Breeding Plantation Tree crops: Tropical species*. New York, NY: Springer New York; 2009. pp. 191–216. https://doi.org/10.1007/978-0-387-71201-7_6
- Greiner S, Sobanski J, Bock R. Why are most organelle genomes transmitted maternally? *BioEssays*. 2015;37:80–94. <https://doi.org/10.1002/bies.201400110>
- Dong W, Xu C, Cheng T, Zhou S. Complete chloroplast genome of *Sedum sarmentosum* and Chloroplast Genome Evolution in Saxifragales. *PLoS ONE*. 2013;8:e77965. <https://doi.org/10.1371/journal.pone.0077965>

16. Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol*. 2011;76:273–97. <https://doi.org/10.1007/s11103-011-9762-4>
17. Knoop V. The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. *Curr Genet*. 2004;46(46):3. <https://doi.org/10.1007/S00294-004-0522-8>
18. Fang Y, Wu H, Zhang T, Yang M, Yin Y, Pan L, et al. A complete sequence and transcriptomic analyses of date palm (*Phoenix dactylifera* L.) mitochondrial genome. *PLoS ONE*. 2012;7:e37164. <https://doi.org/10.1371/JOURNAL.PONE.0037164>
19. Khan A, Khan IA, Heinze B, Azim MK. The Chloroplast Genome sequence of date palm (*Phoenix dactylifera* L. Cv. 'Aseel'). *Plant Mol Biol*. 2012;30:666–78. <https://doi.org/10.1007/S11105-011-0373-7>
20. Khan AL, Asaf S, Lee IJ, Al-Harrasi A, Al-Rawahi A. First chloroplast genomics study of *Phoenix dactylifera* (var. Naghal and Khanezi): a comparative analysis. *PLoS ONE*. 2018;13:e0200104. <https://doi.org/10.1371/JOURNAL.PONE.0200104>
21. Lu R, Chen M, Feng Y, Yuan N, Zhang Y, Cao M, et al. Comparative plastome analyses and genomic resource development in wild rice (*Zizania* spp., Poaceae) using genome skimming data. *Ind Crops Prod*. 2022;186:115244. <https://doi.org/10.1016/J.IJINDCROP.2022.115244>
22. Zhang N, Wen J, Zimmer EA. Congruent Deep relationships in the grape family (Vitaceae) based on sequences of Chloroplast genomes and mitochondrial genes via genome skimming. *PLoS ONE*. 2015;10:e0144701. <https://doi.org/10.1371/JOURNAL.PONE.0144701>
23. Galtier N, Nabholz B, Glémin S, Hurst GDD. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol Ecol*. 2009;18:4541–50. <https://doi.org/10.1111/J.1365-294X.2009.04380.X>
24. Palmer JD, Jansen RK, Michaels HJ, Chase MW, Manhart JR. Chloroplast DNA variation and plant phylogeny. *Ann Mo Bot Gard*. 1988;75:1180. <https://doi.org/10.2307/2399279>
25. Sabir JSM, Abo-Aba S, Bafeel S, Zari TA, Edris S, Shokry AM, et al. Characterization of ten date palm (*Phoenix dactylifera* L.) cultivars from Saudi Arabia using AFLP and ISSR markers. *C R Biol*. 2014;337:6–18. <https://doi.org/10.1016/J.CRV.2013.11.003>
26. Saboori S, Noormohammadi Z, Sheidai M, Marashi SS. Date palm (*Phoenix dactylifera* L.) Cultivar relationships based on Chloroplast genotyping. *Iran J Sci Technol Trans A: Sci*. 2021;45:833–40. <https://doi.org/10.1007/S40995-021-01085-5>
27. Sakka H, Baraket G, Dakhloui 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 trnL (UAA) intron and trnL (UAA) trnF (GAA) intergenic spacer. *Sci Hortic*. 2013;164:256–69. <https://doi.org/10.1016/J.SCIHORT.2013.09.038>
28. Soumaya RC, Sarra C, Salwa ZA, Khaled C, Khaled S. Molecular polymorphism and phylogenetic relationships within Tunisian date palm (*Phoenix dactylifera* L.): evidence of non-coding trnL-trnF regions of chloroplast DNAs. *Sci Hortic*. 2014;170:32–8. <https://doi.org/10.1016/J.SCIHORT.2014.02.027>
29. Cerrudo D, Cao S, Yuan Y, Martinez C, Suarez EA, Babu R, et al. Genomic selection outperforms marker assisted selection for grain yield and physiological traits in a maize doubled haploid population across water treatments. *Front Plant Sci*. 2018;9:366. <https://doi.org/10.3389/FPLS.2018.00366>
30. Jannink JL, Lorenz AJ, Iwata H. Genomic selection in plant breeding: from theory to practice. *Brief Funct Genom*. 2010;9:166–77. <https://doi.org/10.1093/BFGP/ELQ001>
31. Dodsworth S. Genome skimming for next-generation biodiversity analysis. *Trends Plant Sci*. 2015;20:525–7. <https://doi.org/10.1016/j.tplants.2015.06.012>
32. Muraguri S, Xu W, Chapman M, Muchugi A, Oluwaniyi A, Oyebanji O, Liu A. Intraspecific variation within Castor bean (*Ricinus communis* L.) based on chloroplast genomes. *Ind Crops Prod*. 2020;155:112779. <https://doi.org/10.1016/J.IJINDCROP.2020.112779>
33. Nevill PG, Zhong X, Tonti-Filippini J, Byrne M, Hislop M, Thiele K, et al. Large scale genome skimming from herbarium material for accurate plant identification and phylogenomics. *Plant Methods*. 2020;16:1–8. <https://doi.org/10.1186/S13007-019-0534-5>
34. Hamza H, Elbekkay M, Ben Abederrahim MA, Ferchichi A. Molecular and morphological analyses of date palm (*Phoenix dactylifera* L.) subpopulations in southern Tunisia. *Span J Agric Res*. 2011;9:484–93. <https://doi.org/10.5424/SJAR/20110902-271-10>
35. Zehdi S, Cherif E, Rhouma S, Santoni S, Hannachi AS, Pintaud JC. Molecular polymorphism and genetic relationships in date palm (*Phoenix dactylifera* L.): the utility of nuclear microsatellite markers. *Sci Hortic*. 2012;148:255–63. <https://doi.org/10.1016/J.SCIHORT.2012.10.01163>
36. Zehdi-Azouzi S, Cherif E, Moussouni S, Gros-Balthazard M, Naqvi SA, 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. *Ann Bot*. 2015;116:101–12. <https://doi.org/10.1093/AOB/MCV068>
37. El Kadri N, Mimoun MB, Hormaza JI. Genetic diversity of Tunisian male date palm (*Phoenix dactylifera* L.) genotypes using morphological descriptors and molecular markers. *Sci Hortic*. 2019;253:24–34. <https://doi.org/10.1016/j.scienta.2019.04.026>
38. Sabir JSM, Arasappan D, Bahieldin A, Abo-Aba S, Bafeel S. Whole mitochondrial and Plastid Genome SNP Analysis of Nine Date Palm Cultivars Reveals Plastid Heteroplasmy and close phylogenetic relationships among cultivars. *PLoS ONE*. 2014;9:94158. <https://doi.org/10.1371/journal.pone.0094158>
39. Dong W, Liu J, Yu J, Wang L, Zhou S. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS ONE*. 2012;7:e35071. <https://doi.org/10.1371/journal.pone.0035071>
40. Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol* 17, 134. <https://doi.org/10.1186/S13059-016-1004-2>
41. Dhaouadi L, Karbout N, Zammel K, Ahmed M, Wahba C, Abdeyem S. Local adaptations under Tunisian oasis climatic conditions: characterization of the best practices in water agricultural sector. *J Res Environ Earth Sci*. 2021;10:294–300.
42. Faci M, Benziouche SE. Contribution to monitoring the influence of air temperature on some phenological stages of the date palm (cultivar 'Deglet Nour') in Biskra. *J Saudi Soc Agric Sci*. 2021;20:248–56. <https://doi.org/10.1016/j.jssas.2021.02.004>
43. Racchi ML, Bove A, Turchi A, Bashir G, Battaglia M, Camussi A. Genetic characterization of Libyan date palm resources by microsatellite markers. *3 Biotech*. 2014;4:21–32. <https://doi.org/10.1007/s13205-013-0116-6>
44. Hamza H, Jemni M, Benabderrahim MA, Mrabet A, Touil S, Othmani A, Salah MB. Date palm status and perspective in Tunisia. In: Al-Khayri J, Jain S, Johnson D, editors. *Date palm genetic resources and utilization: volume 1*. Dordrecht: Springer Netherlands; 2015. pp. 193–221. https://doi.org/10.1007/978-94-017-9694-1_6
45. Yue Y, Li J, Sun X, Li Z, Jiang B. Polymorphism analysis of the chloroplast and mitochondrial genomes in soybean. *BMC Plant Biol*. 2023;23:15. <https://doi.org/10.1186/s12870-022-04028-3>
46. Kan S, Liao X, Wu Z. The roles of mutation and selection acting on mitochondrial genomes inferred from intraspecific variation in seed plants. *Genes*. 2022;13:1036. <https://doi.org/10.3390/genes13061036>
47. Donnelly K, Cottrell J, Ennos RA, Vendramin GG, A'Hara S, King S, Perry A, Wachowiak W, Cavers S. Reconstructing the plant mitochondrial genome for marker discovery: a case study using *Pinus*. *Mol Ecol Resour*. 2017;17(5):943–54. <https://doi.org/10.1111/1755-0998.12646>
48. Scarcelli N, Mariac C, Couvreur TLP, Faye A, Richard D, Sabot F, et al. Intra-individual polymorphism in chloroplasts from NGS data: where does it come from and how to handle it? *Mol Ecol Res*. 2016;16:434–45. <https://doi.org/10.1111/1755-0998.12462>
49. Najeh D, Taher T, Kacem B. Tunisian deglet noor dates ripening, processing and storage. In: Gerasopoulos D, editor. *Post-harvest losses of Perishable Horticultural products in the Mediterranean Region*. Chania: CIHEAM; 1999. pp. 179–84.
50. Kadri K, Karim A, Hamza H, Manar A, Karmous C, Bettaieb T. Effect of pollination technique on agronomic and physicochemical characteristics of date palm fruits (*Phoenix dactylifera* L.) cv. 'Deglet Nour'. *Hortic Sci Technol*. 2022;40(2):121–33. <https://doi.org/10.7235/HORT.20220012>
51. Soumaya RC, Sarra C, Maha M, Khaled C, Noureddine C. Gene-targeted markers to assess genetic diversity and population structure within Tunisian *Phoenix dactylifera* L. cultivars. *Silvae Genetica*. 2020;69:29–35. <https://doi.org/10.2478/SG-2020-0005>
52. Hamza H, Rejili M, Elbekay M, Ferchichi A. New approach for the morphological identification of date palm (*Phoenix dactylifera* L.) cultivars from Tunisia. *Pak J Bot*. 2009;41:2671–81.
53. Ferchichi A, Hamza H. Le Patrimoine génétique phoenicicole des oasis continentales tunisiennes. Institut des régions arides; 2008.
54. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinform*. 2014;30:2114–20. <https://doi.org/10.1093/bioinformatics/btu170>

55. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26(5):589–95. <https://doi.org/10.1093/bioinformatics/btp698>
56. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012;22:568–76. <https://doi.org/10.1101/GR.129684.111>
57. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinform*. 2011;27:2156–8. <https://doi.org/10.1093/BIOINFORMATICS/BTR330>
58. Paradis E. Analysis of haplotype networks: the randomized minimum spanning tree method. *Methods Ecol Evol*. 2018;9:1308–17. <https://doi.org/10.1111/2041-210X.12969>
59. Paradis E. Pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics*. 2010;26:419–20. <https://doi.org/10.1093/bioinformatics/btp696>
60. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021.
61. Toparslan E, Karabag K, Bilge U. A workflow with R: phylogenetic analyses and visualizations using mitochondrial cytochrome b gene sequences. *PLoS ONE*. 2020;15:1–17. <https://doi.org/10.1371/JOURNAL.PONE.0243927>
62. Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J. Fast hierarchical bayesian analysis of population structure. *Nucleic Acids Res*. 2019;47:5539–49. <https://doi.org/10.1093/NAR/GKZ361>
63. Ortiz EM. vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis. 2019. <https://doi.org/10.5281/zenodo.1257057>
64. NASA JPL. 2020. NASADEM Merged DEM Global 1 arc second V001 [Data set]. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MEASURES/NASADEM/NASADEM_HGT.001. Accessed 13 Apr 2023.
65. ISCGM., 2008. Tunisia Inland Waters [map]. In: International Steering Committee for Global Mapping (ISCGM). https://geodata.lib.utexas.edu/catalog/sde-columbia-iscgm_tunisia_2008_water. Accessed 13 Apr 2023.
66. Cingolani P. Variant annotation and functional prediction: SnpEff. *Methods mol biol*. 2012;2493:289–314. https://doi.org/10.1007/978-1-0716-2293-3_19
67. Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, Lu X. Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program. *SnpSift Front Genet*. 2012;3:35. <https://doi.org/10.3389/fgene.2012.00035>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.