Plant Mol Biol Rep (2012) 30:666–678
DOI 10.1007/s11105-011-0373-7

The Chloroplast Genome Sequence of Date Palm (*Phoenix dactylifera* L. cv. 'Aseel')

Asifullah Khan · Ishtiaq A. Khan · Berthold Heinze · M. Kamran Azim

Published online: 23 November 2011
© Springer-Verlag 2011

Abstract Date palm (*Phoenix dactylifera* L.) is an economically important and widely cultivated palm of the family Arecaceae. We sequenced the complete date palm chloroplast genome (cpDNA) from Pakistani cv. 'Aseel', using a combination of Sanger-based and next-generation sequencing technologies. Being very similar to a sequence from a Saudi Arabian date palm cultivar 'Khalas' published recently, the size of the genome was 158,458 bp with a pair of inverted repeat (IR) regions of 27,276 bp that were separated by a large single-copy (LSC) region of 86,195 bp and a small single-copy (SSC) region of 17,711 bp. Genome annotation demonstrated a total of 138 genes, of which 89 were protein coding, 39 were tRNA, and eight were rRNA genes. Comparison of cpDNA sequences of cultivars 'Aseel' and 'Khalas' showed following intervarietal variations in the LSC region; (a) two SNPs in intergenic spacers and one SNP in the *rpoc1* gene, (b) polymorphism in two mono-nucleotide simple sequence repeats (SSR), and (c) a 4-bp indel in the *accD-psaI* intergenic spacer. The SSC region has a polymorphic site in the mono-nucleotide SSR located at position 120,710. We also compared cv. 'Aseel' cpDNA sequence with partial *P. dactylifera* cpDNA sequence entries deposited in Genbank and identified a number of potentially useful polymorphisms in this species. Analysis of date palm cpDNA sequences revealed a close relationship with *Typha latifolia*. Occurrence of small numbers of forward and inverted repeats in date palm cpDNA indicated conserved genome arrangement.

Keywords Intervarietal polymorphism · Plastid · Next-generation sequencing · Genetic diversity

A. Khan · I. A. Khan · M. K. Azim (✉)
International Center for Chemical and Biological Sciences,
University of Karachi,
Karachi 75270, Pakistan
e-mail: kamran.azim@iccs.edu

M. K. Azim
e-mail: mkamranazim@yahoo.co.uk

B. Heinze
Department of Genetics, Federal Research Centre for Forests,
Hauptstraße 7,
1140 Vienna, Austria

Introduction

Date palm (*Phoenix dactylifera* L.) is an important fruit crop of family *Arecaceae* mostly grown in the arid regions of Africa, the Middle East, and South Asia (Al-Farsi and Lee 2008). It is one of the oldest known fruit trees cultivated for at least 5,000 years and reported to be originated from southern Iraq or the western Indian subcontinent (Zohary and Hopf 2000). The economic importance of date palm is due to its nutritionally valuable fruit which consists of 72–88% of sugar, minerals (i.e., iron, potassium, calcium, chlorine, copper, magnesium, sulfur, and phosphorus), amino acids, and vitamins (Al-Shahib and Marshall 2003). Moreover, antioxidant and antimutagenic activities of date fruit have also been reported (Vayalil 2002). The date palm tree can grow well in deserts with harsh climatic and soil conditions where the growth of other crops could be relatively difficult. Hence, the date palm offers high nutritive food in such areas (Al-Farsi et al. 2005). In 2006, the world date production was about seven million tonnes (www.faostat.fao.org).

Recently, a number of studies have addressed issues of genetic diversity among fruit-bearing plants including date palm (Zhang et al., 2011; He et al., 2011; Tanya et al.,

2011; Xie et al., 2011). Younis et al. identified sex-specific DNA markers for date palm using RAPD and ISSR techniques (Younis et al. 2008). Similarly, the utilization of the RAPD-PCR approach has been proposed for date palm cultivar identification (Sedra et al. 1998; Al-Khalifah and Askari 2003; Abdulla and Gamal 2010). DNA polymorphism studies of selected cultivars revealed high genetic diversity in date palm (Elshibli and Korpelainen 2008; 2009a; 2009b). Marqués et al. (2008) have identified a set of RNAs transcribed from chloroplast genome reported to be involved in brittle leaf disease of date palm (Marqués et al. 2008). The chloroplast is an essential organelle of photosynthetic cells. In angiosperms, cpDNA is a highly conserved, double-stranded, circular molecule with size ranges from 120–220 kb (Gao et al. 2010; Khan et al. 2010). Typical chloroplast DNA consists of large and small single-copy regions (denoted as LSC and SSC, respectively) which are separated by two inverted repeat regions (denoted as inverted repeats, IRA and IRB; Ravi et al. 2008). The availability of complete plastid genome sequences from different clades of autotrophs has greatly resolved the organization and evolution of this interesting cellular organelle. Moreover, comparative chloroplast genomics can provide new knowledge regarding phylogenetics of green plants. Here we report the chloroplast genome sequence of the date palm cv. ‘Aseel’ grown in Pakistan, using Sanger-based and next-generation sequencing technologies. Initially in June 2009, we submitted the sequence of the inverted repeat region of cv. ‘Aseel’ cpDNA to GenBank (accession number FJ212316). The complete sequence was submitted in the GenBank in April 2010. While this paper was in preparation, the date palm chloroplast genome sequence from another cultivar, cv. ‘Khalas’ grown in the Saudi Arabian peninsula, was published (Yang et al. 2010). Hence, the detailed comparison of the cpDNA sequences from both cultivars of date palm, as well as a comparison with shorter date palm chloroplast DNA sequences available in GenBank, is presented here. A comparison of date palm cpDNA with available monocot species has also been carried out.

Materials and Methods

Plant Material

Fresh leaves from young cultivar ‘Aseel’ of date palm, cultivated in the botanical garden of the University of Karachi, Karachi, Pakistan were collected. The leaves were collected from a single date palm tree. Voucher specimens are kept at the Herbarium, Department of Botany, University of Karachi under voucher specimen number: 02 and General Herbarium number: 75539.

DNA Isolation and Sequencing

A combination of Sanger-based and next-generation sequencing strategies were used for DNA sequencing. The date palm cv ‘Aseel’ leaves (3.0 g) were processed for isolation of total DNA (20 μg) using a modified CTAB method (Porebski et al. 1997) and the commercially available Biopsin plant genomic DNA extraction kit (Bioer Technology, Hangzhou, PR China). Initially, a primer walking strategy termed as “ASAP: Amplification, sequencing & annotation of plastomes” (Dhingra and Folta 2005) was used for amplification and Sanger-based sequencing of the inverted repeat region of cpDNA. Briefly, purified date palm DNA (20 μg) was used for generation of 6.0 kb amplicons with consensus set of primers (Dhingra and Folta 2005). The 6.0 kb amplicons were then used for generation of 1.0 kb fragments using internal sets of primers corresponding to 6.0 kb amplicons. Later on, gap filling primers were designed to fill the gaps within the inverted repeat region (Table 1). The Sanger-based sequencing of the abovementioned fragments was carried out using a CEQ8000 Genetic Analyzer (Beckman Coulter Inc. USA) and an ABI3130 Analyzer (Applied Biosystem, USA). For cycle sequencing reactions, the DTCS kit (Beckman Coulter Inc. USA) and Big Dye Terminator kit (Applied Biosystem Inc, USA) were used, with conditions as recommended by the suppliers.

Complete sequencing of the date palm cpDNA was carried out by next-generation sequencing technology. For this purpose, a chloroplast-rich fraction was prepared from 10 g of date palm cv. ‘Aseel’ leaves followed by DNA purification (Triboush et al. 1998). The date palm DNA (7.0 μg) was then used for construction of paired-end libraries with insert size of 250 bp according to the protocol provided by the supplier (Illumina Inc. San Diego, USA). The massively parallel sequencing was carried out by the “sequencing by synthesis” approach using the HiSeq2000 system (Illumina Inc., San Diego, USA) in BGI, Shenzhen, China.

Table 1 Primers used for gap filling while sequencing inverted repeat (IR) regions of cpDNA date palm cv. Aseel

No.	IR position (bp)	Primer sequence (5′ to 3′)
1	87,473–87,456	AATCTGCCGAATCACTCATG
2	89,489–89,507	GAACCGGATCGGAAAAAAG
3	96,904–96,924	AGATTTTGAGTCTCGCGTGTC
4	101,674–101,693	TTCTCGATCAATCCCTTTGC
5	102,961–102,982	TCGTACAAGATAAGCTATGACC
6	106,677–106,699	GTACATGGACGATAGTTGGAGTC
7	107,662–107,681	ATGATGCCCTTAGCGAAAAC
8	108,443–108,463	ATAAGCTTCATCGTCGAGAGG
9	109,398–109,419	TGTTCCTTAAAGCGATGGAATC
10	111,739–111,760	AATTGACTGACCAGCATCCAAG

Genome Assembling, Annotation, and Analysis

The contiguous sequences obtained from Sanger-based sequencing were assembled using the Lasergene package version 7.1 (DNASTAR Inc., Madison, WI, USA). The sequencing data from the HiSeq2000 system was assembled using CLC Genomics Workbench version 3.5.1 (CLC bio, Denmark). The assembled sequences were combined using CLC Genomics Workbench (CLC bio, Denmark). Genome annotation was performed through the DOGMA server (Dual Organellar Genome Annotator; Wyman et al. 2004), ORF Finder (http://www.ncbi.nlm.nih.gov/projects/gorf/), and BLAST (Altschul et al. 1990). In addition, annotation of some tRNAs was performed using tRNAscan-SE (Lowe and Eddy 1997) and after similarity searches with other annotated plastomes. The beginnings and ends of genes were manually adjusted. Repeat analysis was performed using the REPuter program (Kurtz et al. 2001). A circular genome map of date palm cpDNA was constructed using the GenomeVx online tool (Conant and Wolfe 2008). The GeneOrder server was used for gene-order analysis (Celamkoti et al. 2004). Construction of multiple alignments and phylogenetic trees of complete cpDNA sequences was carried out by the mVISTA comparative genomics tool (Frazer et al. 2004). The maximum parsimony (MP)-based phylogenetic analysis of 25 protein-coding genes, i.e., *matK*, *petA*, *petB*, *petD*, *petG*, *petN*, *psaB*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbN*, *psbT*, *rpoB*, *rpoC1*, *rpoC2*, *rps8*, *rps11*, *rps14*, and *ycf3* was done by MEGA4 (Tamura et al. 2007).

Results and Discussion

Genome Assembling and Organization

We carried out complete chloroplast genome sequencing of date palm (*P. dactylifera* L.) cv. 'Aseel' grown in Pakistan using Sanger-based and next-generation sequencing methods. Initially, 22,918 bp of the inverted repeat (IR) region were sequenced using the ASAP protocol (see Methods; Dhingra and Folta 2005; GenBank accession number FJ212316). The primers reported by Dhingra and Folta (2005) resulted in 84.0% coverage of the IR region. Subsequently, ten primers were designed (Table 1) to fill the gaps within the IR region, resulting in up to 96.5% sequence coverage of this region, i.e., 26,316 bp. The HiSeq2000 system (Illumina Inc. San Diego, USA) gave 2,197,575 high-quality paired-end reads with an average length of 73.5 bp. The Illumina reads were filtered (10% as default) so that no ambiguities remained. From this data, 267,669 reads (12.18% of all reads) assembled the complete date palm chloroplast genome cv. 'Aseel' with an average of 124*X* coverage, using the cpDNA of the Saudi Arabian date palm cv. 'Khalas' as reference (Yang et al. 2010). The unassembled reads (87.81%) were mostly from the nuclear genome due to nuclear DNA contamination during chloroplast DNA isolation. While 'chimeric' reads consisting of parts of nuclear and chloroplast DNA may be expected, due to the presence of nuclear copies of chloroplast DNA in all plants sequenced to date (e.g., Matsuo et al. 2005; Tuskan et al. 2006), and likewise for mitochondrial DNA (e.g., Hirai and Nakazon 1993; Tuskan et al. 2006), we understand that due to the short read length (72 bp) of the Illumina data, it would be difficult to identify such reads without ambiguity, as no complete nuclear or mitochondrial genomes of date palm are available to date (i.e., up to March, 2011). The reads matching cut-off was set to 90–95% during reference-based assembly via CLCBio workbench assembler. Yang et al. (2010) described difficulties in assembling of '454' (also called GS FLX technology, Roche Applied Science, Germany) next-generation sequencing data across mono-nucleotide stretches. The '454' technology has been reported to be error prone in sequencing mono-nucleotide repeats stretches. On the other hand, sequencing of cv. 'Aseel' was done by the Illumina technology, which has the advantage of precisely sequencing such homopolymer sequences (Mardis 2008).

The chloroplast genome sequence of date palm cv. 'Aseel' had a total length 158,458 bp with two IRs of 27,276 bp separated by a large single-copy region of 86,195 bp and a small single-copy region of 17,711 bp. The genome was 4 bp shorter than the chloroplast genome of Saudi Arabian cv. Khalas (total size 158,462 bp; Genbank accession GU811709). The genome contained 59% coding and 41% non-coding regions including pseudogenes, introns, and intergenic spacers. A total of 138 genes were present including pseudogenes *Ψycf15*, *Ψycf68*, and a *Ψycf1* (short pseudo copy of *ycf1* gene; Table 2). Out of 89 protein-coding open-reading frames, 16 genes contained introns. Among these the *clpP*, *rps12*, and *ycf3* genes contained two introns each. Date palm chloroplast DNA contained 38 genes for tRNAs (30 distinct genes), and of these, eight tRNA genes contained introns. Four rRNA genes were confined to and duplicated in the IR regions. As a whole, 20 complete genes (including the *ycf15* pseudogene) and one 3′-exon of the *rps12* trans-splicing protein were duplicated in the IR regions. Of these 20 genes, eight were tRNA, four were rRNA, and seven were protein-coding genes.

Comparison with Saudi Arabian cv. 'Khalas'

The chloroplasts in plant cell are considered by some as a population with genetic heterogeneity (e.g., Bendich 1987; Johnson and Palmer 1989; Fitter et al. 1996; Wolfe and Randle 2004). Analyses of high-quality sequence reads may therefore reveal polymorphic sites in chloroplast genomes. The cpDNA sequence variations can be partitioned into intravarietal polymorphisms (intraSNPs), i.e.,

Table 2 Genes in the chloroplast genome of *Phoenix dactylifera* L

Group of genes								
Transfer RNA genes	trnQ-UUG	trnS-GCU	trnG-GCC*	trnR-UCU	trnC-GCA	trnD-GUC	trnY-GUA	trnE-UUC
	trnT-GGU	trnS-UGA	trnG-UCC	trnM-CAU	trnS-GGA	trnT-UGU	trnL-UAA*	trnF-GAA
	trnV-UAC*	trnM-CAU	trnW-CCA	trnP-UGG	trnH-GUG (×2)	trnl-CAU (×2)	trnL-CAA (×2)	trnV-GAC (×2)
	trnI-GAU* (×2)	trnA-UGC* (×2)	trnR-ACG (×2)	trnN-GUU (×2)	trnL-UAG	trnK-UUU*	trnI-UAU*	
Ribosomal RNA genes	rrn16 (×2)	rrn5 (×2)	rrn4.5 (×2)	rrn23 (×2)				
Ribosomal protein genes	rps16*	rps2	rps14	rps4	rpl33	rps18	rps12_5′ end	rps 12_3′end* (×2)
	rps11	rpl36	rps8	rpl14	rpl16*	rps3	rpl22	rps19 (×2)
	rpl2* (×2)	rpl23(x2)	rps7 (×2)	rpl32	rps15	rpl20*		
Photosystem proteins genes	psbA	psbK	psbI	psbM	psbD	psbC	psbZ	psaB
	psaA	psaI	psbJ	psbL	psbF	psbE	psaJ	psbB
	psbT	psbN	psbH	psaC				
ATP synthase subunit genes	atpA	atpF*	atpH	atpI	atpE	atpB		
NADH subunits genes	ndhJ	ndhK	ndhC	ndhB* (×2)	ndhF	ndhD	ndhE	ndhG
	ndhI	ndhA*	ndhH					
Cytochrome-related genes	petN	petA	petL	petG	petB*	petD	ccsA	
RNA polymerase genes	rpoC2	rpoC1*	rpoB	rpoA				
Genes of unknown function (YCF)	ycf4	ycf2 (×2)	ycf3**	ycf68* (×2)	ycf1	ycf1 (truncated)		
Rubisco subunit gene	rbcL							
Envelop membrane protein gene	cemA							
Acetyl-CoA-carboxylase subunit gene	accD							
Maturase gene	matK							
Proteosome like protease gene	clpP**							
Translation initiation factor gene	infA							
Pseudogenes	ycf15 (×2)							

One or two asterisks indicate one or two intron(s) containing genes respectively. Genes duplicated in IR regions represented by (×2)

sequence variations within a variety (or cultivar and subspecies), and inter-subspecific polymorphisms, i.e., sequence variations between different varieties of a species. These types of variations can be further characterized when one of the alleles becomes unique to a certain variety or subspecies (Tang et al., 2004; Yang et al., 2010).

The present study provided an opportunity to shed light on inter-subspecific polymorphisms in two "ecotypes" of date palm, i.e., the cultivars 'Khalas' and 'Aseel' grown in Saudi Arabia and Pakistan, respectively. The date palm chloroplast genome of the Saudi Arabian cv. 'Khalas' was reported by Yang et al. (2010). They achieved a sequence draft with 1,081*X* coverage using GS FLX ('454') next-generation sequencing technology. They observed intra-varietal single nucleotide polymorphisms (intraSNPs) in date palm cpDNA. We carried out detailed sequence comparison of cpDNA from the Saudi Arabian and Pakistani date palm cultivars to determine "inter-subspecific" variations.

The comparison of IR regions indicated no sequence variations. It has been well established that mutation rate in the IR region is lower than single-copy regions of chloroplast genomes (Wolfe et al. 1987; Maier et al. 1995). In this IR region, however, Yang et al. (2010) found one intravarietal SNP in *ycf2*, harboring a T-G mutation at position 92,696 of Saudi Arabian date palm cv. 'Khalas'. This type of intravarietal SNPs has been suggested to be considered as intervarietal variation among date palm cultivars (Yang et al. 2010). However, we could not find this mutation in cv. 'Aseel'.

The following polymorphic sites were detected in the non-coding sequences of the LSC region of the two cultivars. (1) At positions 9,218 and 9,221, the cv. 'Aseel' data showed G(kh) → A(as) and C(kh) → T(as) mutations respectively with >50*X* coverage (kh = Khalas and as = Aseel). (2) Closer to the abovementioned position, a mono-nucleotide SSR (simple sequence repeat) with 17 poly-A repeat was detected at position 9,263–9,279 in cv. 'Aseel' compared to 15 poly-A repeat in cv. 'Khalas'. (3) In the case of cv. 'Aseel' cpDNA, the *rbcL–accD* intergenic spacer region contained consecutive mono-nucleotide SSRs of 14 poly-C and 11 poly-A repeat units. However, cv. 'Khalas' cpDNA contains 13 poly-C and 12 poly-A repeats. (4) Yang et al. (2010) noted a characteristic 4-bp insertion of 'TAGA' at the position 61,482–61,485 in the *accD-psaI* intergenic region as a genotype in cv. 'Khalas' compared to other monocots with known cpDNA sequences. However, cv. 'Aseel' cpDNA sequence did not show this tetra-nucleotide insertion (at >100*X* coverage; Fig. 1). Hence, this site can be considered as a DNA marker for characterization of date palm cultivars.

In the intergenic sequences of the SSC region of cv. 'Aseel', a mono-nucleotide SSR with nine poly-T repeat was detected at position 120,710 at 60*X* coverage compared to ten poly-T repeat in cv. 'Khalas'. Comparative analysis of chloroplast genomes of cultivars 'Aseel' and 'Khalas' identified a G–T SNP at position 21,747 in the coding region of the *rpoc1* gene, which results in a degenerate codon. In cv. 'Aseel', the sequence data gave >100*X* coverage of 'T' at this locus. This SNP has been identified as an intravarietal SNP in cv. 'Khalas' (Yang et al. 2010). This comparison showed that the variation occurred in non-coding regions, except for a SNP in the *rpoc1* gene.

```
DP cv. Khalas     61464 CTTTCTCTTTCTAATAGATAGAAATATTAAATCGAG 61499
DP cv. Aseel      61465 CTTTCTCTTTCTAATAGA----AATATTAAATCGAG 61496
T. latifolia      63618 CTTTCTAATGGATACAAA----TATATAGATACAAA 63649
D. elephantipes   58627 CTTTATAAGAGGTACAA------ATATTAAATCGAG 58656
A. calamus        59375 CCTTATAACAGGGACAA------ATATTAAACCGAG 59404
```

Fig. 1 Multiple alignment of the position 61,464–61,499 (date palm cv. 'Khalas' numbering) in the *accD-psaI* intergenic region of chloroplast DNA sequences from date palm cultivars 'Khalas' and 'Aseel', *T. latifolia*, *D. elephantipes*, and *A. calamus* showing four base-pair insertion in cv. 'Khalas'

Comparison with Further GenBank Entries

A search in GenBank revealed 27 more short chloroplast DNA sequences of date palm; many of them analyzed in the frame of phylogenetic studies; and many of them as yet unpublished. Comparing these sequences to our genome sequence, a surprisingly high number of polymorphisms showed up (Table 3). However, the sequences at all these positions are identical in cv. 'Aseel' and 'Khalas'. Among those, there are quite large indels (e.g., 51 and 53 bp), some of them shared among voucher specimen (e.g., the 12 bp deletion in nos. 14, 15, and 18 in Table 3). A number of polymorphisms in the usually highly conserved 16S ribosomal RNA gene were also observed (nos. 3–8 in Table 3). A surprisingly high number of polymorphisms (given the fewer polymorphisms that distinguish 'Khalas' and 'Aseel') resulted in amino acid changes in proteins (nos. 21–28 in Table 3), including a 17 amino acid deletion in the *ndhI* gene.

It would be worth investigating and confirming these polymorphisms in a wider array of date palm accessions. We cannot exclude, however, the possibility that some polymorphisms may turn out as sequencing errors, or due to poor sequence quality (no. 28 in Table 3). Our detailed comparison of the two complete genomes, sequenced with high coverage, suggests this. On the other hand, studies with nuclear DNA markers have revealed high diversity among date palm cultivars (e.g., Elshibli and Korpelainen 2008, 2009a, 2009b), and it would not be uncommon then to find high levels of polymorphisms in its chloroplast DNA as well.

Table 3 A comparison of GenBank date palm cpDNA entries with cv. 'Aseel' cp genome sequence

No.	NCBI entry/voucher specimen/cultivar/ reference	Regions	length (bp)	cv. 'Aseel' cpDNA position	Mutations compare to 'Aseel'	Origin
tRNA genes						
1	AM114567.1 specimen voucher Barrow 77 (K)	trnK gene intron	1,819	1,543	G → A	Not mentioned
2	AY690324.1	tRNA-Ala intron	273	No mutation		Not mentioned
16S rDNA						
3	EF470302.1 cultivar Ferey (unpublished)	16S ribosomal RNA gene	308	139,708–140,020	51 base substitutions, 4 deletions and a C insertion	Egypt
4	EF470303.1 cultivar Tagtagt (unpublished)	16S ribosomal RNA gene	310	139,708–140,020	53 base substitutions, 4 deletions, and a C insertion	Egypt
5	EF470306.1 cultivar Barhee (unpublished)	16S ribosomal RNA gene	311	140,004 140,016 140,018	T → C G → A Deletion of C	Egypt
6	EF470304.1 Cultivar Oshikagbil (unpublished)	16S ribosomal RNA gene	309	140,004 140,016	T → C G → A	Egypt
7	EF470305.1 cultivar Zaghloul (unpublished)	16S ribosomal RNA gene	309	140,004 140,016	T → C G → A	Egypt
8	HM753598.1 (unpublished)	16S ribosomal RNA gene	1,593	103,624–25	AC deletion	Saudi Arabia
				103,711	A → G	
				103,722	T → C	
				103,804	C → T	
				103,985	T → C	
				104,012	A → G	
				104,114	C → T	
				104,221	T → C	
				104,300	A → G	
				104,404	Deletion of C	
				104,507	G → A	
				104,657	G → T	
				104,673	A → G	
				104,734	T → C	
				104,763	C → A	
				104,994	G → A	
				105,083–105,088	Deletion of 6 bases (ATGCTT)	
				105,131	G → T	
				105,132	C → G	
				105,147	C → T	
				105,151	G → A	
				105,158	T → G	
				105,162	G → A	
				105,193–105,197	Deletion of 5 bases (GAAGT)	
				105,212	G → A	
Intergenic spacers						
9	EU053865.1	trnV_UAC-trnM_CAU spacer	378	110,852	T → A	Tunisia
10	AY044480.1	trnD-trnY spacer	882	31,599	A → C	USA
	Specimen voucher no. Hahn 6899			31,654	C → G	

Table 3 (continued)

No.	NCBI entry/voucher specimen/cultivar/reference	Regions	length (bp)	cv. 'Aseel' cpDNA position	Mutations compare to 'Aseel'	Origin
11	EU004213.1	trnD-trnY spacer	887	31,452	C insertion	France
	Specimen voucher no. JCP578 (unpublished)			31,716	C insertion	
12	AY044578.1	trnQ-rps16 spacer	1,057	6,003	Deletion of T	USA
	Specimen voucher no. Hahn 6899			6,418	G → C	
13	EF605522.1	trnQ-rps16 spacer	825	No mutation		France
	Voucher JCP578 (unpublished)					
14	EF690458.1 Voucher JCP578 (P) (unpublished)	trnS-UGA and trnS-psbZ spacers	1,168	36,063	A deletion	France
				36,159	T→C mutation	
				37,076–37,087	12 bases (CTAACTACTATA) deletion	
				37,146	G insertion	
15	EU043486.1 Voucher SZ10 (unpublished)	trnS-UGA and trnS-psbZ spacers	1,133	36,159	T→C mutation	France
				37,076–37,087	12 bases (CTAACTACTATA) deletion	
16	EU043484.1 Voucher SZ1 (unpublished)	trnS-UGA and trnS-psbZ spacers	1,160	No mutation		France
17	EU053866.1 (unpublished)	trnM_CAU-atpE spacer	293	No mutation		Spain
18	EU043485.1 Voucher SZ2 (unpublished)	trnS-UGA and trnS-psbZ spacers	1,148	36,159	T→C mutation	France
				37,076–37,087	12 bases (CTAACTACTATA) deletion	
19 (see No. 23)	AY166799.1	atpB-rbcL spacer	3,993	56,473	G insertion	Spain
	(unpublished)			56,480	G insertion	
				56,504	G insertion	
				56,512	G insertion	
				56,516	C insertion	
				56,521	T insertion	
20 (see No. 27)	AY166803.1	ndhG-ndhI spacer	2,787	121,917	Deletion of A	Spain
	(unpublished)			121,925	A→C	
				121,959	A→T	
				121,985	A→T	
Protein-coding genes						
21	AB040211.1	matK	1,570	2,055	A → T (Leu243 → Val)	Japan
				2,557	A → C (Ile410 → Asn)	
22	GQ422122.1	matK (partial sequence)	527	2,298	T → G	UAE
				2,311	T → C (Two mutations in the partial protein sequence probably due to sequence error.)	
23	EU053867.1	atpA	535	11,610	A → G (Val74→Ile)	Spain
				11,613	C → T (silent mutations)	
				12,125	C → T	

Table 3 (continued)

No.	NCBI entry/voucher specimen/cultivar/reference	Regions	length (bp)	cv. 'Aseel' cpDNA position	Mutations compare to 'Aseel'	Origin
					(silent mutations)	
				12,126	G → C	
					(silent mutations)	
24	AY166799.1	atpE	3,993	54,293	G → A	Spain
	(unpublished)	atpB			(silent mutation in atpE)	
		rbcL				
25	AY012411.1	atpB	1,515	54,876	A → C	USA
					(IIe427→Met)	
26	AY012468.1	rbcL	1,401	57,606	G → T	USA
	specimen_voucher: Hahn s.n., WIS, cult				(Val219→Leu)	
				58,076	G → A	
					(silent mutation)	
27	AY044529.1	ndhF	2,095	113,577	A → G	USA
					(silent mutation)	
				113,886	G → T	
					(silent mutation)	
				113,947	G → T	
					(Pro578→Gln)	
				114,155	T → G	
					(Lys509→Gln)	
28	AY166803.1 (unpublished)	ndhA (partial) and ndhI (complete)	2,787	122,305–122,355	51 bp deletion in ndhI gene resulted in 17 amino acid deletion	Spain
					Following mutations are in the intron of ndhA gene	
				123,414	Deletion of C	
				123,419	Deletion of C	
				123,426	Deletion of A	
				123,944	Insertion of G	
				123,953	C → A	
29	U96630.1	(psbA)	911	Poor sequence. Not analyzed.		USA

Comparison with Other Monocot Species

Currently, chloroplast genome sequences from six monocot families (i.e. Dioscoreaceae, Acoraceae, Orchidaceae, Araceae, Typhaceae, and Poaceae) are available in nucleotide databases. We compared date palm cpDNA sequences with ten species from six monocot families: one species each from Dioscoreaceae (Hansen et al. 2007), Orchidaceae

Table 4 Comparison of the main features of date palm chloroplast genome with representative species of six monocots

Monocot family	Arecaceae	Dioscoreaceae	Acoraceae	Orchidaceae	Araceae	Poaceae	Typhaceae
Representative species	*Phoenix dactylifera*	*Dioscorea elephantipes*	*Acorus calamus*	*Phalaenopsis aphrodite*	*Lemna minor*	*Zea mays*	*Typha latifolia*
Total size (bp)	158,462	152,609	153,821	148,964	165,955	140,387	161,572
LSC region (bp)	86,199	82,777	84,148	85,957	89,906	82,355	89,140
Inverted repeat (bp)	27,276	25,513	25,698	25,732	31,223	22,748	26,390
SSC region (bp)	17,711	18,806	18,277	11,543	13,603	12,536	19,652
GC content (%)	37.23	37.15	38.60	36.65	35.72	38.46	36.61
AT content (%)	62.77	62.85	61.40	63.35	64.28	61.54	63.39

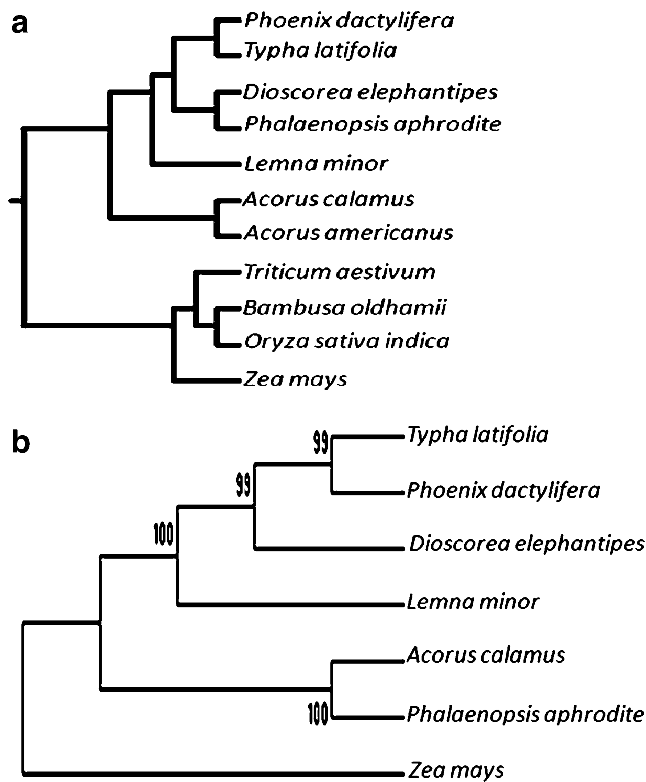


Fig. 2 (**a**) Phylogenetic tree of chloroplast genome sequences from ten monocot species including date palm (*P. dactylifera* L.) based on full-length multiple alignments. (**b**) Maximum-parsimony-based phylogenetic tree, derived from 25 concatenated chloroplast protein-coding genes sequences from representative species of different monocot families. *Numbers at node* indicate maximum parsimony bootstrap values

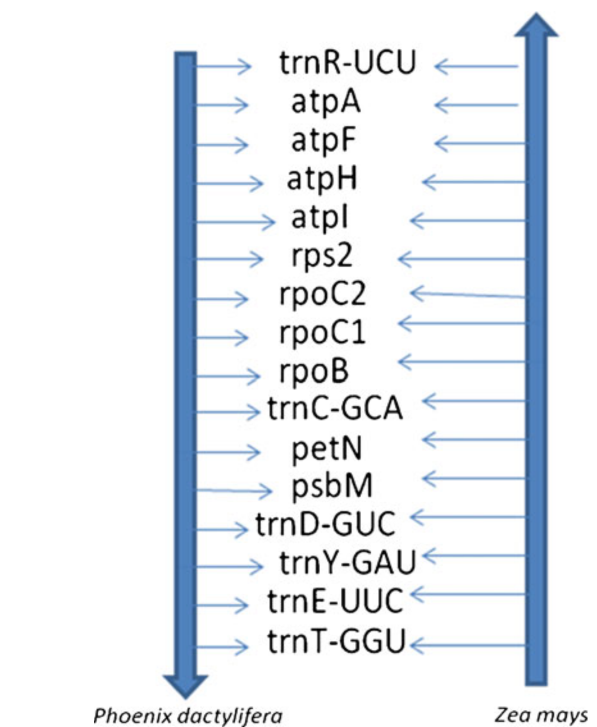
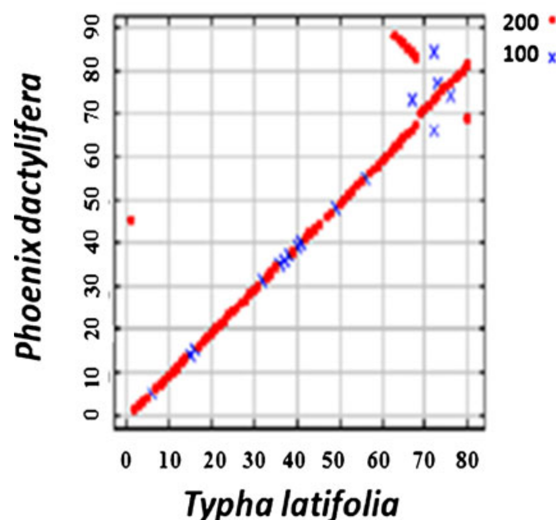


Fig. 4 The diagrammatic representation of large inversion in LSC region of *Z. mays*. As a result of this inversion the orientation of 16 genes of *Z. mays* are inverted in comparison to *P. dactylifera* and other monocots plastomes

(Chang et al. 2006), Araceae (Mardanov et al. 2008), and Typhaceae (Guisinger et al. 2010); two species from Acoraceae (GoremykinVV et al. 2005); and four species from Poaceae (Maier et al. 1995; Ogihara et al. 2000; Masood et al. 2004; Wu et al. 2009). Analysis showed that



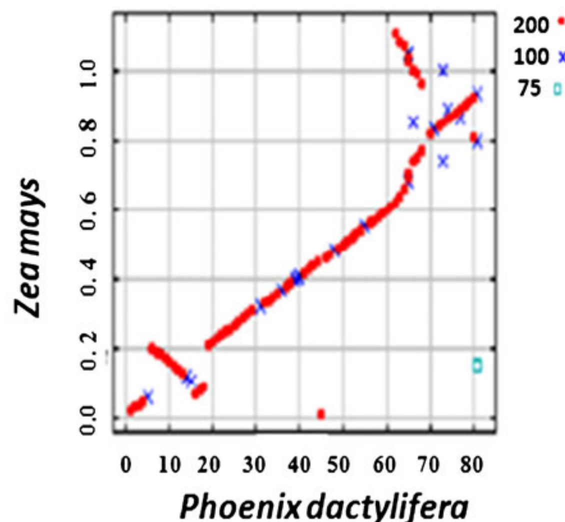


Fig. 3 (**a**) Dot plot of gene order between *P. dactylifera* and *T. latifolia* chloroplast genome. The straight diagonal line represents the synteny between the comparing genome. The upper small counter diagonal line is because of genes to be inverted within inverted repeat regions. (**b**) Dot plot of *P. dactylifera* and *Z. mays* chloroplast genome showing small counter diagonal in lower region of plot indicating the genes inverted in LSC region of *Z. mays* chloroplast genome. The number "*200*" and "*100*" with *sign of red box* and *blue cross*, respectively, indicate the gene similarity score

the total size of the date palm cpDNA was larger than the others, with the exception of *Lemna minor* (Araceae) and *Typha latifolia* (Typhaceae) (Table 4). The 'AT' and 'GC' percentages of the date palm genome are in close range with the other monocots (Table 4). Multiple alignments of full-length cpDNA sequences from 11 monocot species followed by phylogenetic tree construction using the mVISTA server (Frazer et al. 2004; Brudno et al. 2003) revealed a grouping of date palm cpDNA sequences with *T. latifolia*, *Dioscorea elephantipes* and *Phalaenopsis aphrodite* (Fig. 2a). However, the most closely related sequence was *T. latifolia* cpDNA (Guisinger et al. 2010). Furthermore, a maximum parsimony-tree based on 25 chloroplast protein-coding genes found in seven monocots (the sequence alignment that was used for phylogenetic analysis comprised 21,255 characters) showed that date palm and *T. latifolia* form a single clade with high bootstrap values, i.e., ≥95% (Fig. 2b).

The chloroplast genomes of date palm and *T. latifolia* have the same gene contents. Unlike *P. aphrodite*, the date palm cpDNA contained a full set of *ndh* genes. Moreover, like *L. minor* and *P. aphrodite*, the *rps12* gene in date palm cpDNA was uniquely divided into a 5′-exon located in the LSC region and two 3′-exons located in duplications within IRs. The monocot family Poaceae has lost three genes (i.e., *accD*, *ycf1*, and *ycf2*) and several introns within the *clpP* and *rpoC1* genes; however, these are present in date palm and *T. latifolia* (Guisinger et al. 2010).

Conservation in gene order or synteny between date palm and *T. latifolia* plastomes was shown by the plot generated by GeneOrder 3.0 (Celamkoti et al. 2004) (Fig. 3a). However, due to a specific inversion within the LSC of family Poaceae, the pattern of some genes of *Zea mays* (Maier et al. 1995) were inverted in comparison to date palm (Figs. 3b and 4).

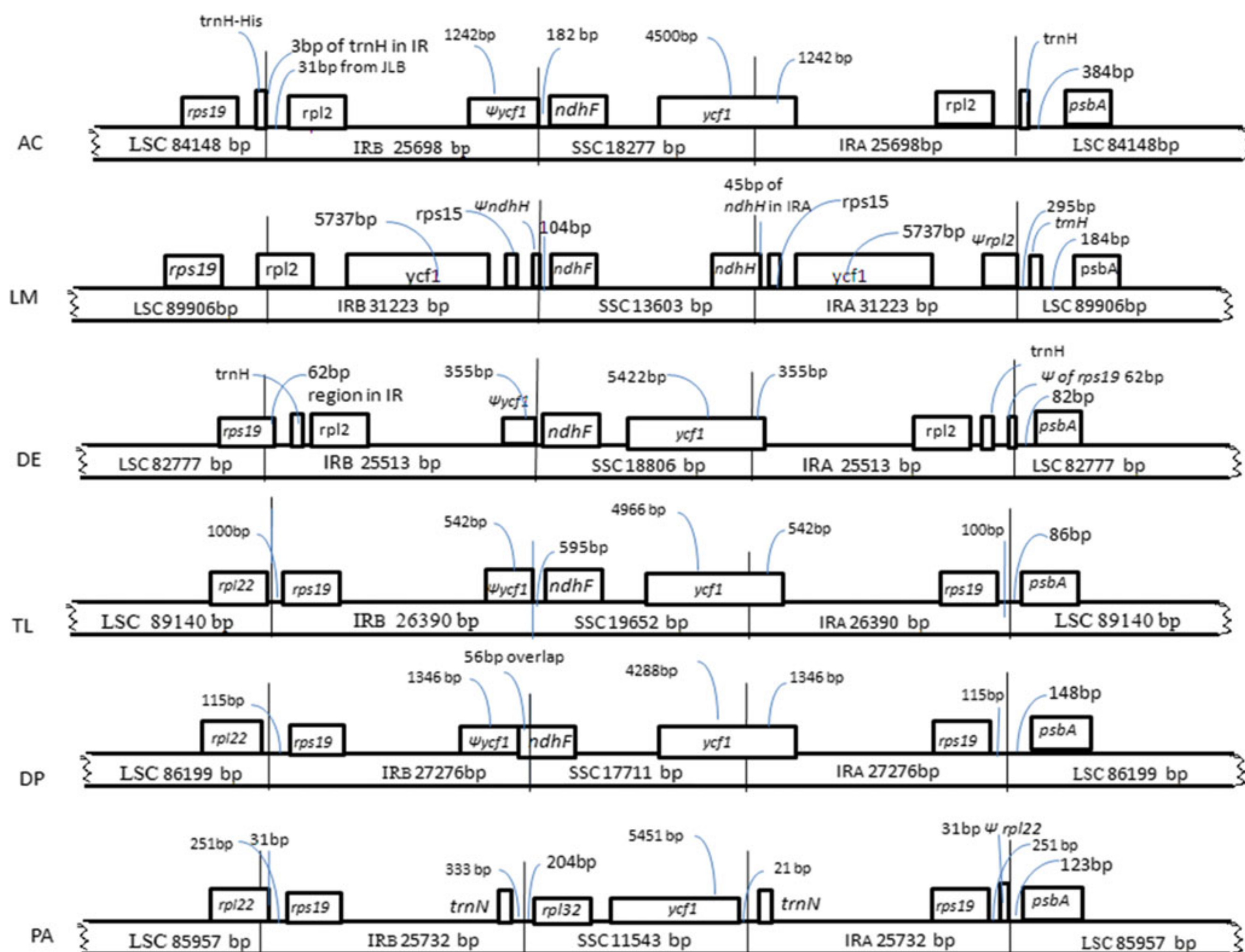


Fig. 5 Comparison of border positions of LSC, SSC, and IR among date palm and closely related monocot species. *Boxes above the main line* indicate the predicted genes while the pseudogenes at the borders are shown by *Ψ* (*letter*). The figure is not the scale and just shows relative changes at or near the IR–SC borders. *AC: Acorus calamus*, *LM: Lemna minor*, *DE: Dioscorea elephantipes*, *TE: Typha latifolia*, *DP: Phoenix dactylifera*, *PA: Phalaenopsis aphrodite*

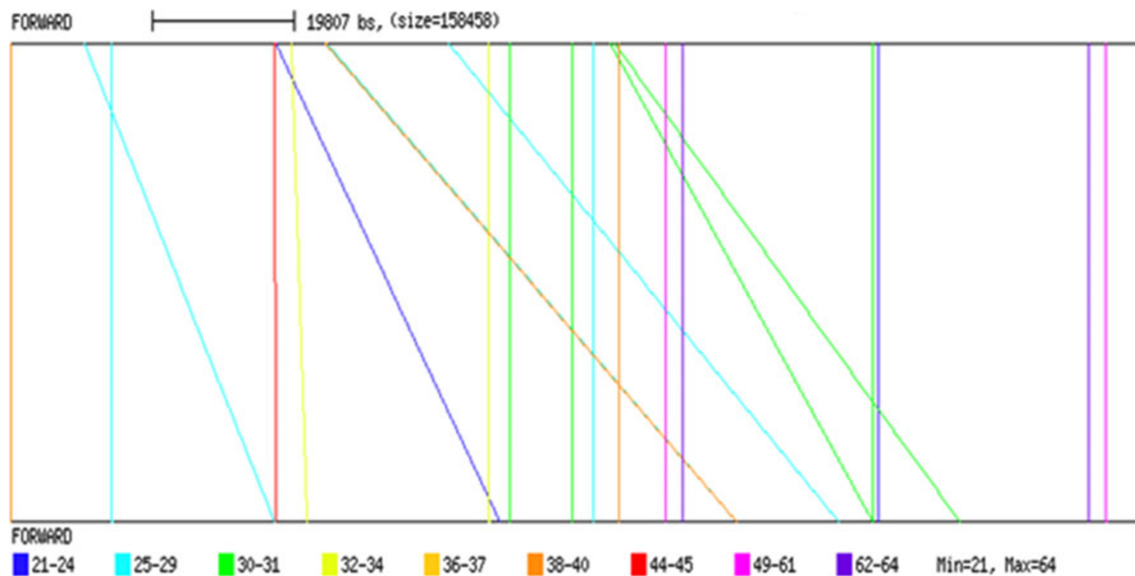


Fig. 6 A graphical output of repeats in date palm (*P. dactylifera* L.) chloroplast genome. The graph gives an overview of number, length, and location of repeats. The *lines* indicating repeats are *colored* according to respective repeat length. To keep the starting position information visible, each part of a repeat is displayed on a separate strand

Although among land plants, the chloroplast genomes are highly conserved in gene sets and order, the borders between IRs (IRA and IRB) and the two single-copy regions (LSC and SSC) are known to vary among species (Kim and Lee 2004). Considerable expansion and contraction of the IR region is mostly responsible for size variation in the chloroplast genome (Chung et al. 2006, Ravi et al. 2006). We here compare the position of IR borders in date palm and other monocot species. Due to a characteristic expansion of IRB sequences into the LSC region, a specific rearrangement was acquired by monocot chloroplast genomes early in evolution. This expansion resulted in the inclusion of *trnH* and *rps19* genes in the IR region. Among monocots, *Acorus calamus* shows similarity to dicots and contains a single copy of *rps19* in the LSC region, while in case of *D. elephantipes* only a 62-bp portion of *rps19* has been found in IRb and seems to be in the mid of this evolutionary implication (Fig. 5). The *L. minor* plastome shows a contrary feature compared to other monocots due to location of *rpl2* gene at the border of IRB/LSC, which resulted in a pseudo copy of *rpl2* gene, i.e., *Ψrpl2* in IRA region (Mardanov et al. 2008) (Fig. 5). Our analysis has shown that like most other monocot species, the date palm chloroplast genome has followed the same pattern, and IRB sequences have expanded into the LSC region. This expansion was also observed by Yang et al. (2010) for date palm cv. 'Khalas' cpDNA. This expansion resulted in two copies of *trnH* and *rps19* genes in the IR regions. In the case of date palm, this IRB expansion was 15 bp more than in the closely related *T. latifolia* genome (Fig. 5). An extreme expansion of IRB was found for *P. aphrodite*, where a 31-bp inclusion of the *rpl22* gene also occurred in the IR region. Furthermore, like other monocot plastomes, the date palm IRA is extended deep into the *ycf1* gene and resulted in the 1,346-bp *ycf1* pseudogene in IRB. The IRB/SSC border of the date palm chloroplast genome, located within the coding region of the *ndhF* gene, was not found in other monocot plastomes. Careful sequence analysis revealed 57 bp overlap between *ndhF* gene and *ycf1* pseudogene at the IRB/SSC border in both date palm

Table 5 Distribution of repeats of sizes ≥30 bases (hamming distance=3.0) in genes encoded by chloroplast genomes of *P. dactylifera* and selected monocot species

Species	ycf1	ycf2	ycf3	psaB-psaA	accd	rps18	rpoc2
P. dactylifera		12	–	1	–	–	–
A. calamus	1	13		2	–	–	–
D. elephantipes	1	5	–	1	–	–	–
L. minor	–	2		1	–	–	–
P. aphrodite	–	2	–	1	2	–	–
T. latifolia	–	8	2	2	–	–	–
Zea mays	–	–	–	1	–	5	12

cultivars. However, Yang et al. (2010) observed a 55-bp overlapping region between these two genes. These expansions at IR/SC borders increase the length of the IR region of the date palm chloroplast genome compared to other monocot plastomes, except for *L. minor*.

Higher number of repeats and larger repeat sequences are associated with extensive chloroplast genome rearrangement (Haberle et al. 2008). Small forward and inverted repeats in cpDNA sequences from date palm and six other monocots, i.e., *Z. mays*, *T. latifolia*, *P. aphrodite*, *L. minor*, *D. elephantipes*, and *A. calamus* was computed using the REPuter program (Kurtz et al. 2001). Repeats of ≥30 bases were calculated with a Hamming distance of 3.0 (Kurtz et al. 2001). In date palm cpDNA, 64 repeats of ≥30 bases were found, of which 28 were inverted, while 36 were direct repeats (Fig. 6). The number of forward and inverted repeats of ≥30 bases in plastomes of other monocot species (i.e., *A. calamus*, *D. elephantipes*, *L. minor*, *P. aphrodite*, *T. latifolia*, and *Z. mays*) were 122, 17, 22, 37, 80, and 80, respectively. In date palm cpDNA, 12 repeats were located in the *ycf2* gene and one repeat was in the *psaB* gene (Table 5), while the rest of the repeats belonged to non-coding regions. Table 5 contains the data about the repeats distribution in protein regions coding in cpDNA of date palm and six selected monocots. The gene order and repeat analyses supported the view of conserved arrangement of genes within the date palm chloroplast genome (Haberle et al. 2008).

As a next step, the polymorphisms we describe between the two fully sequenced cultivars, and those between 'Aseel' and the GenBank entries, should be analyzed on a wide panel of date palm accessions—across its natural and cultivated ranges. We expect that some of the polymorphisms will turn out to be useful for a preliminary (chloroplast-based) phylogeography of the species, and some of them may even be useful for cultivar identification (most probably in combination with nuclear DNA markers, as those described in previous studies).

References

Abdulla M, Gamal O (2010) Investigation on molecular phylogeny of some date palm (*Phoenix dactylifra* L.) cultivars by protein, RAPD and ISSR markers in Saudi Arabia. Aust J Crop Sci 4:23–28

Al-Farsi M, Alasalvar C, Morris A, Baron M, Shahidi F (2005) Compositional and sensory characteristics of three native sun-dried date (*Phoenix dactylifera* L.) varieties grown in Oman. J Agric Food Chem 53:7586–7591

Al-Farsi MA, Lee CY (2008) Nutritional and functional properties of dates: a review. Crit Rev Food Sci Nutr 48:877–887

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

Al-Shahib W, Marshall RJ (2003) The fruit of the date palm: its possible use as the best food for the future? Int J Food Sci Nutr 54:247–259

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Bendich AJ (1987) Why do chloroplasts and mitochondria contain so many copies of their genome? BioEssays 6:279–282

Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, Green ED, Sidow A, Batzoglou S (2003) LAGAN and Multi-LAGAN: Efficient Tools for Large-Scale Multiple Alignment of Genomic DNA. Genome Research 13(4):721–731

Celamkoti S, Kundeti S, Purkayastha A, Mazumder R, Buck C, Seto D (2004) GeneOrder 3.0: software for comparing the order of genes in pairs of small bacterial genomes. BMC Bioinforma 5:52

Chang C, Lin H, Lin I et al (2006) The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. Mol Biol Evol 23:279–291

Chung HJ, Jung JD, Park HW, Kim JH, Cha HW, Min SR, Jeong WJ, Liu JR (2006) The complete chloroplast genome sequences of *Solanum tuberosum* and comparative analysis with Solanaceae species identified the presence of a 241-bp deletion in cultivated potato chloroplast DNA sequence. Plant Cell Rep 25:1369–1379

Conant GC, Wolfe KH (2008) GenomeVx: simple web-based creation of editable circular chromosome maps. Bioinformatics 2:861–862

Dhingra A, Folta KM (2005) ASAP: amplification, sequencing & annotation of plastomes. BMC Genomics 6:176

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

Elshibli S, Korpelainen H (2009a) Excess heterozygosity and scarce genetic differentiation in the populations of *Phoenix dactylifera* L.: human impact or ecological determinants. Plant Genet Res 7:95–104

Elshibli S, Korpelainen H (2009b) Biodiversity of date palms (*Phoenix dactylifera* L.) in Sudan: chemical, morphological and DNA polymorphisms of selected cultivars. Plant Genet Res 7:194–203

Fitter JT, Thomas MR, Rose RJ, Steelescott N (1996) Heteroplasmy of the chloroplast genome of *Medicago sativa* L cv 'Regen S' confirmed by sequence analysis. Theor Appl Genet 93:685–690

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comparative genomics. Nucleic Acids Res 32:W273–W279

Gao L, Su Y-J, Wang T (2010) Plastid genome sequencing, comparative genomics, and phylogenomics: current status and prospects. J Syst Evol 48:77–93

GoremykinVV HB, Hirsch-Ernst KI, Hellwig FH (2005) Analysis of *Acorus calamus* chloroplast genome and its phylogenetic implications. Mol Biol Evol 22:1813–1822

Guisinger MM, Chumley TW, Kuehl JV, Boore JL, Jansen RK (2010) Implications of the plastid genome sequence of *Typha* (Typhaceae, Poales) for understanding genome evolution in Poaceae. J Mol Evol 70:149–166

Haberle R, Fourcade H, Boore J, Jansen R (2008) Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. J Mol Evol 66:350–361

Hansen DR, Dastidar SG, Cai Z et al (2007) Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early-diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae), and *Illicium* (Schisandraceae). Mol Phylogenet Evol 45:547–563

He Q, Li XW, Liang GL, Ji K, Guo QG, Yuan WM, Zhou GZ, Chen KS, Van de Weg WE, Gao ZS (2011) Genetic diversity and identity of Chinese loquat cultivars/accessions (Eriobotrya japonica) using apple SSR markers. Plant Mol Biol Rep 29:197–208

Hirai A, Nakazon M (1993) Six percent of the mitochondrial genome in rice came from chloroplast DNA. Plant Mol Biol Rep 11 (2):98–100

Johnson L, Palmer J (1989) Heteroplasmy of chloroplast DNA in *Medicago*. Plant Mol Biol 12:3–11

Khan A, Khan IA, Asif H, Azim MK (2010) Current trends in chloroplast genomics. Afric J Biotech 9:3494–3500

Kim KJ, Lee HL (2004) Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. DNA Res 11:247–261

Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res 29:4633–4642

Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964

Maier RM, Neckermann K, Igloi GL, Kossel H (1995) Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. J Mol Biol 251:614–628

Mardanov AV, Ravin NV, Kuznetsov BB et al (2008) Complete sequence of the duckweed (*Lemna minor*) chloroplast genome: structural organization and phylogenetic relationships to other angiosperms. J Mol Evol 66:555–564

Mardis ER (2008) Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet 9:387–402

Marqués J, Fadda ZG, Duran-Vila N, Flores R, Bové JM, Daròs JA (2008) A set of novel RNAs transcribed from the chloroplast genome accumulates in date palm leaflets affected by brittle leaf disease. Phytopathol 98:337–344

Masood MS, Nishikawa T, Fukuoka S, Njenga PK, Tsudzuki T, Kadowaki K (2004) The complete nucleotide sequence of wild rice (*Oryza nivara*) chloroplast genome: first genome wide comparative sequence analysis of wild and cultivated rice. Gene 340:133–139

Matsuo M, Ito Y, Yamauchi R, Obokata J (2005) The rice nuclear genome continuously integrates, shuffles, and eliminates the chloroplast genome to cause chloroplast-nuclear DNA flux. Plant Cell 17:665–675

Ogihara Y, Isono K, Kojima T et al (2000) Chinese spring wheat (*Triticum aestivum* L.) chloroplast genome: complete sequence and contig clones. Plant Mol Biol Rep 18:243–253

Porebski S, Bailey LG, Baum BR (1997) Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Mol Biol Rep 15:8–15

Ravi V, Khurana J, Tyagi A, Khurana P (2006) The chloroplast genome of mulberry: complete nucleotide sequence, gene organization and comparative analysis. Tree Genet Genomes 3:49–59

Ravi V, Khurana JP, Tyagi AK, Khurana P (2008) An update on chloroplast genomes. Plant Syst Evol 271:101–102

Sedra MyH, Lashermes P, Trouslot P, Combes M-C (1998) Identification and genetic diversity analysis of date palm (*Phoenix dactylifera* L.) varieties from Morocco using RAPD markers. Euphytica 103:75–82

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Tang J, Xia H, Cao M, Zhang X, Zeng W, Hu S, Tong W, Wang J, Wang J, Yu J, Yang H, Zhu L (2004) A comparison of rice chloroplast genomes. Plant Physiol 135:412–420

Tanya P, Taeprayoon P, Hadkam Y, Srinives P (2011) Genetic diversity among Jatropha and Jatropha-related species based on ISSR markers. Plant Mol Biol Rep 29:252–264

Triboush SO, Danilenko NG, Davydenko OG (1998) A method for isolation of chloroplast DNA and mitochondrial DNA from sunflower. Plant Mol Biol Rep 16:183–189

Tuskan GA, DiFazio S, Jansson S et al (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313:1596–1604

Vayalil PK (2002) Antioxidant and antimutagenic properties of aqueous extract of date fruit (*Phoenix dactylifera* L. Arecaceae). J Agric Food Chem 50:610–617

Wolfe AD, Randle CP (2004) Recombination, heteroplasmy, haplotype polymorphism, and paralogy in plastid genes: implications for plant molecular systematics. Syst Bot 29:1011–1020

Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci 84:9054–9058

Wu F, Kan D, Lee SB et al (2009) Complete nucleotide sequence of *Dendrocalamus latiflorus* and *Bambusa oldhamii* chloroplast genomes. Tree Physiol 29:847–856

Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20:3252–3255

Xie R-J, Zhou J, Wang G-Y, Zhang S-M, Chen L, Gao Z-S (2011) Cultivar identification and genetic diversity of Chinese bayberry (*Myrica rubra*) accessions based on fluorescent SSR markers. Plant Mol Biol Rep 29:554–562

Yang M, Zhang X, Liu G, Yin Y, Chen K, Yun Q, Zhao D, Al-Mssallem IS, Yu J (2010) The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). PLoS One 5: e12762

Younis RAA, Ismail OM, Soliman SS (2008) Identification of sex-specific DNA markers for date palm (*Phoenix dactylifera* L.) using RAPD and ISSR techniques. Res J Agric Biol Sci 4:278–284

Zhang Q, Li J, Zhao Y, Schuyler SK, Han Y (2011) Evaluation of genetic diversity in Chinese wild apple species along with apple cultivars using SSR markers. Plant Mol Biol Rep Early online. doi:10.1007/s11105-011-0366-6

Zohary D, Hopf M (2000) Domestication of plants in the old world: the origin and spread of cultivated plants in West Asia, Europe, and the Nile Valley, 3rd edn. Oxford University Press, Oxon