BIOSCIENCES BIOTECHNOLOGY RESEARCH ASIA, March 2020. Vol. **17**(1), p. 27-44

Comparison of Plastome SNPs/INDELs among different Wheat (*Triticum* sp.) Cultivars

Shahira A. Hassoubah[1], Reem M. Farsi[1], Jehan S. Alrahimi[1], Nada M. Nass[1] and Ahmed Bahieldin[1,2,*]

[1]Department of Biological Sciences, Faculty of Science, King Abdulaziz University (KAU), Jeddah, Saudi Arabia.
[2]Department of Genetics, Faculty of Agriculture, Ain Shams University, Cairo, Egypt.

http://dx.doi.org/10.13005/bbra/2807

(Received: 25 March 2020; accepted: 10 April 2020)

Wheat is the most important cereal crop in the world as compared to other grain crops in terms of acreage and productivity. Based on next-generation sequencing data, we sequenced and assembled chloroplastid (cp) genomes of nine Egyptian wheat cultivars in which eight of them are hexaploid (*Triticum aestivum*, 2n=6x) and one is tetraploid (*T. turgidum* sub sp. *durum*, 2n=4x). Sequencing reads were first filtered in which all sequencing reads that mapped to mitochondrial (mt) genome were removed. Preliminary results indicated no intra-cultivar heteroplasmy for the different cultivars. Size of the resulted chloroplast wheat genome across different cultivars is 133,812 bp, which is less than the cp genome of "Chinese Spring" cultivar partially due to the presence of three large sequences in the later genome belonging to rice cp genome. Three new non-coding tRNA gene sequences were also found and function of one conserved ORF namely *ycf5*is shown. The protein-coding genes represent 67.26% of the total plastid genes. In the non-coding regions, a number of 5 tandem and 31 long repeats were found. Codon usage in the wheat cp genome has the same trend as that published for wheat mitochondrial genome. Assembled cp genomes after filtering out the gaps (≤ 5 bp) generated in the nine cultivars were also used for SNPs and INDELs analyses. Across different cultivars, 564 SNPs and 160 INDELs were identified, of which 230 and 4 were in the protein-coding regions, respectively. Five and nine cultivar-specific SNPs and INDELs were found, respectively. One SNP, while none for INDELs, was found in the genic regions unique to one of the two inverted repeats (IRA) in the coding sequence of *ndhB* gene. Two SNPs were non-synonymous substitutions in the two protein-coding genes *rpoA*and *rpl16,* while one was synonymous substitution in the protein coding gene *rpl23*. Three INDELs exist in *rpl2* gene. The first is 12-nucleotide that starts at nucleotide 4 of the gene and encodes for four amino acids. Two other INDELs starts from nucleotide 160 of the gene and are 19-nt apart. These two INDELs resulted in a frameshift of six amino acids, with a glycine amino acid in the middle that remained unchanged, then the default frame was restored. Results of dendrogram aligned with known relationships among cultivars. In conclusion, SNPs and INDELs analyses of wheat plastome were successfully used for detecting polymorphism among wheat cultivars.

Keywords: Hexaploid, tetraploid, frameshift, dendrogram, linage, polymorphism.

Chloroplast is a cell organelle that provides energy for plants and algae via the process of photosynthesis. Other biological processes occur in chloroplast including the production of starch, lipids, amino acids, vitamins, and key pathways of sulfur and nitrogen metabolism.[1] During evolution, chloroplasts were thought to arise from endosymbiosis between photosynthetic

*Corresponding author E-mail: abmahmed@kau.edu.sa

bacterium and non-photosynthetic host.[2] Plant plastid (cp) contains highly conserved genomes in terms of structure and gene content compared to those of mitochondrial and nuclear genomes.[3,4] Individual chloroplast contains up to 1,600 copies of cp genome or plastome.[5] In angiosperm, ex., monocots, cp DNA is circular and genome size ranges between 120-160 kb and featured with a quadripartite organization of two copies of inverted repeats (IRs) (20-28 kb), and a large (80-90 kb) and a small (16-27 kb) single-copy region, namely LSC and SSC, respectively. The cp genome mostly harbors ~4 rRNAs, ~30 tRNAs and ~80 protein-coding genes in addition to introns and intergenic spacers (IGS).[6] Chloroplast genome is maternally inherited and studies of its structure, sequence variation, and diversity are useful in cytoplasmic breeding and non-inherited transgene insertions.[5] Differences in gene content have been detected among angiosperm cpgenomes,[7-9] however, no records were made at the plant species level.

In the past, the advent of Sanger sequencing method has enabled the elucidation of genetic information, however, it was hampered by technical details, costs, time and data resolution. The next-generation sequencing (NGS) technology has overcome these problems and revolutionized the science of genomics more appropriately. NGS revealed unlimited insights into genomes and transcriptomes of many species during the last few years.

Wheat is among the most widely cultivated field crops worldwide. Cultivated wheats can be either hexaploid (*T. aestivum,*AABBDD, 2n=6x) or tetraploid (*Triticum durum*, AABB, 2n=4x). Complexity of wheat nuclear genome in terms of genome types and size makes it difficult to be sequenced and assembled. The draft genome of the A-genome progenitor (e.g., *T. urartu*, AA) has been assembled and assigned as a reference genome for further comparison with polyploidgenomes.[10]

A number of studies used the whole genome approach in order to detect SNPs and INDELs in the mitochondrial (mt) and cpgenomes.[5,11-13] Nonetheless, utilization of SNP/INDELs of plastome in detecting genetic distances is a challenging task. With the possibility that half of the cp genome has analogue sequences in mitochondrial genome and due to the incidence of intra-varietal heteroplasmy, drawing dendrograms to describe the relationships among cultivars based on organellar SNPs/INDELs is a challenging task. Although heteroplasmy has been reported as a rare event in cpgenomes,[14] earlier studies indicated higher probabilities.[15,16] We speculate that polymorphism due to partial genome transfer and heteroplasmy should be removed before we approach to detect SNPs/INDELs among genotypes.

The available reference cp genome of the hexaploid "Chinese Spring" cultivar was previously sequenced based on the constructed genomic library and the assembled clone-contigs.[3] In the present study, we have detected the structure and gene content of wheat plastome based on the new era of NGS with nine wheat cultivars. Eight of these cultivars are hexaploids and one is a tetraploid. We also attempted to detect genetic distance within hexaploid species or between the two wheat species based on SNPs/INDELsofcp genomes.

METHODS

Sampling and DNA Isolation

Nucleic acids were isolated from leaf tissues (~ 1 g) of 14-day-old etiolated seedlings of nine wheat cultivars (Table 1)using the modified procedure of Gawel and Jarret[17]. DNAs were treated with RNase A (10 mg/ml) and incubated at 37^{o}C for 30 min to remove RNA contaminants. Then, DNAs were shipped in liquid nitrogen to BGI, China for deep sequencing using the Illumina HiSeq 2000 platform.

Mapping of reads to reference cp genome

Between 101.34 to 195.28 million 100-bp paired-end reads were generated for each cultivarfrom 500-bp insert library. Adapter sequences in the raw data were deleted, and reads with 50% low quality bases (quality value d" 5) or more were discarded. The remaining sequences of different cultivars were first mappedto the published wheat mtgenome (acc. no. AP008982) before mapping to cpgenome (acc. no. AB042240) using CLC Genomics Workbench (version 3.0, http://www.clcbio.com/user manuals). All cp reads that aligned to mt genome were removed before cp genome assembly.

Sequence Annotation

Annotation was carried out by mapping

cp genome sequences with BLAST hits (identity 90% and overlap 90%)[18] to known plastid genes. Then, sequences were tested for consistency of the ORFs using NCBI online tool of the ORF finder (http://www.ncbi.nlm.nih.gov/projects/gorf/, the standard genetic code was applied). Gene and exon boundaries were determined by alignment of homologous genes from wheat and several other common plastid angiosperm genomes. The tRNA genes were identified by using BLAST search tools,[18] and the tRNAscan-SE program (version 1.4 with default parameters).[19] Repetitive sequences were identified using the REPuter (version 2.74; length e"50 bp; mismatch d" 3 mismatches).[20]Then, information on tandem repeats were identified using a tandem repeat finder (http://tandem.bu.edu/trf/trf.html,Benson[21]).

Identification of SNP and INDELs and phylogenetic analyses

As extra step of filtering was made by the removal of sequences in the reference cp genome corresponding to the gaps of e" 5 bp in all the nine wheat cultivars to avoid bias in the resulted INDELs analysis. Gaps in the cp genome of the nine cultivars that generated by the reference cp genome with less than 5 bp were considered insertions.However, gaps generated during alignment only in the reference cp genome were all considered as deletions. The mapping results after the third filtering were, then, used for SNPs/INDELs identificationbased on a Bayesian algorithm according to the BioScope software (version 1.3) guideused as visual double-check. Only SNPs/INDELs with a read depth ofe" 30, mapping quality of e" 30and SNPs/INDELs quality of e" 20 were retained.

Data matrices of different cultivar pairs were entered into TFPGA (version 1.3) and analyzed using qualitative routine and dissimilarity coefficients were utilized in drawingdendrogram using unweighted pair group method with arithmetic average (UPGMA) and Neighbor Joining (NJ) routine using NTSYSpc (version 2.10, Exeter software).The bootstrap value was set to 100. All other parameters are set as default.

RESULTS AND DISCUSSION

Mapping of reads to reference genome

The number of reads mapped to the cp genomes of the nine wheat cultivars ranged between 281,499-2,169,718 with CG representing 38.31% and mapped reads average representing 1.1% of the total reads (Table 2, Supplementary Files 1-9). Mapping of the reads to the reference wheat cp genome (acc. no. AB042240,Ogiharaetal.,[3]) resulted in 100% coverage of the genome. Removal of reads aligned to the wheat mt genome reduced the number of cp reads to 219,147-1,440,201, which represents an average of 0.73% of the total reads with mean filtered coverage of 644-1,450 (Table 2). As all reads that mapped to mitochondrial genome were eliminated, we confidently declare that intra-cultivar heteroplasmy for the different cultivars does not exist in alignment with the results incp genomes of many other angiosperms, ex., *B. hygrometrica*, in which no intraSNPs were found.[22] The intraSNPs have been demonstrated to be present in both cp and mt genomes in rice.[23] Additionally, in our earlier study on date palm cp genome following the same approach of removal of reads mapped to mt genome, we detected a number of intraSNPs that reflects plastid heteroplasmy.[24]This data confirmed that date palm cp genomes are heteroplasmic and scoped the light on the necessity to be cautious when analyzing SNP from data generated from next generation sequencing of total genomic DNA of other crop plants.

Comparative analysis of plastomes of several angiosperms

Although the nuclear wheat genome (~16-17 Gb) is about 3-35 fold larger than other cereals, like rice (0.43 Gb) and barley (5.3 Gb), the plastid genome (133,812 bp) is the smallest among angiosperms including cereals, after *Marchantiapolymorpha* (121,024 bp), and the total number of gene types (97), either protein coding, tRNA or rRNA genes, is the least among angiosperms (Table 3). The detailed gene content of wheat plastome is shown in Table 4. The largest known cp genome among angiosperms is that of *Chara vulgaris* (184,933 bp).[22] Plastid genome of the latter species also has the highest AT% (73.8%) and repeats % (3.162%) among angiosperms. The coding percentage in wheat cp genome is intermediate among angiosperms; date palm cp genome has the highest (99.39%). The number of tandem repeats of wheat cp genome is the highest (5) among published cpgenomes of

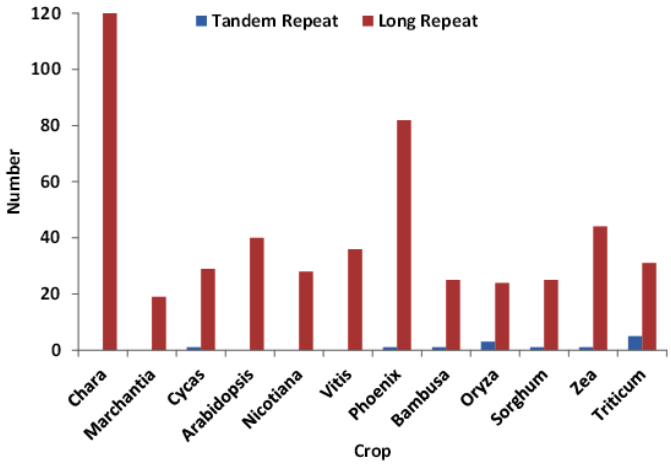

Fig. 1. Number of tandem and long repeats in plastomes of several angiosperms

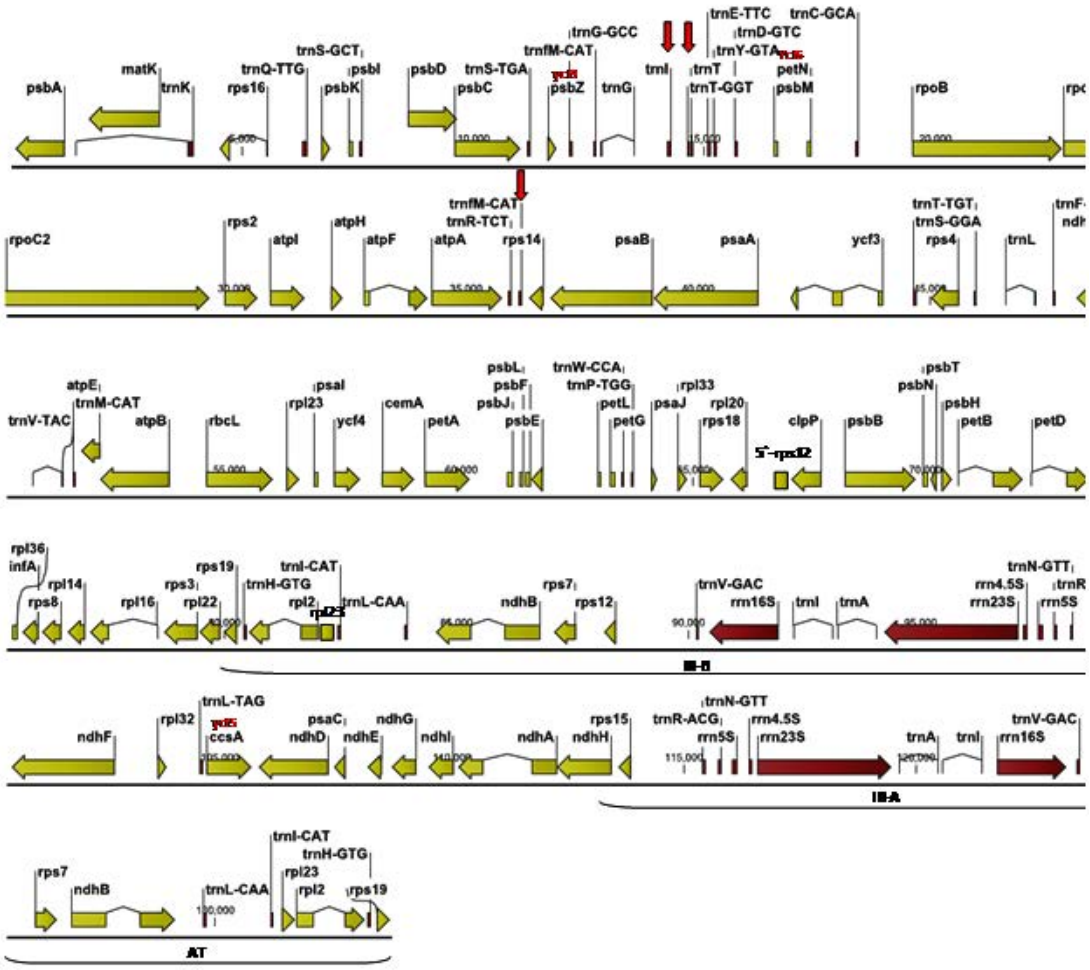

Fig. 2. Plastome of wheat cultivar G168 indicating the gene content. The arrow indicate three non-coding genes, e.g., *trnI*, *trnT* and *trnfM*missing in the plastome of the reference wheat cp genome(acc. no. AB042240)

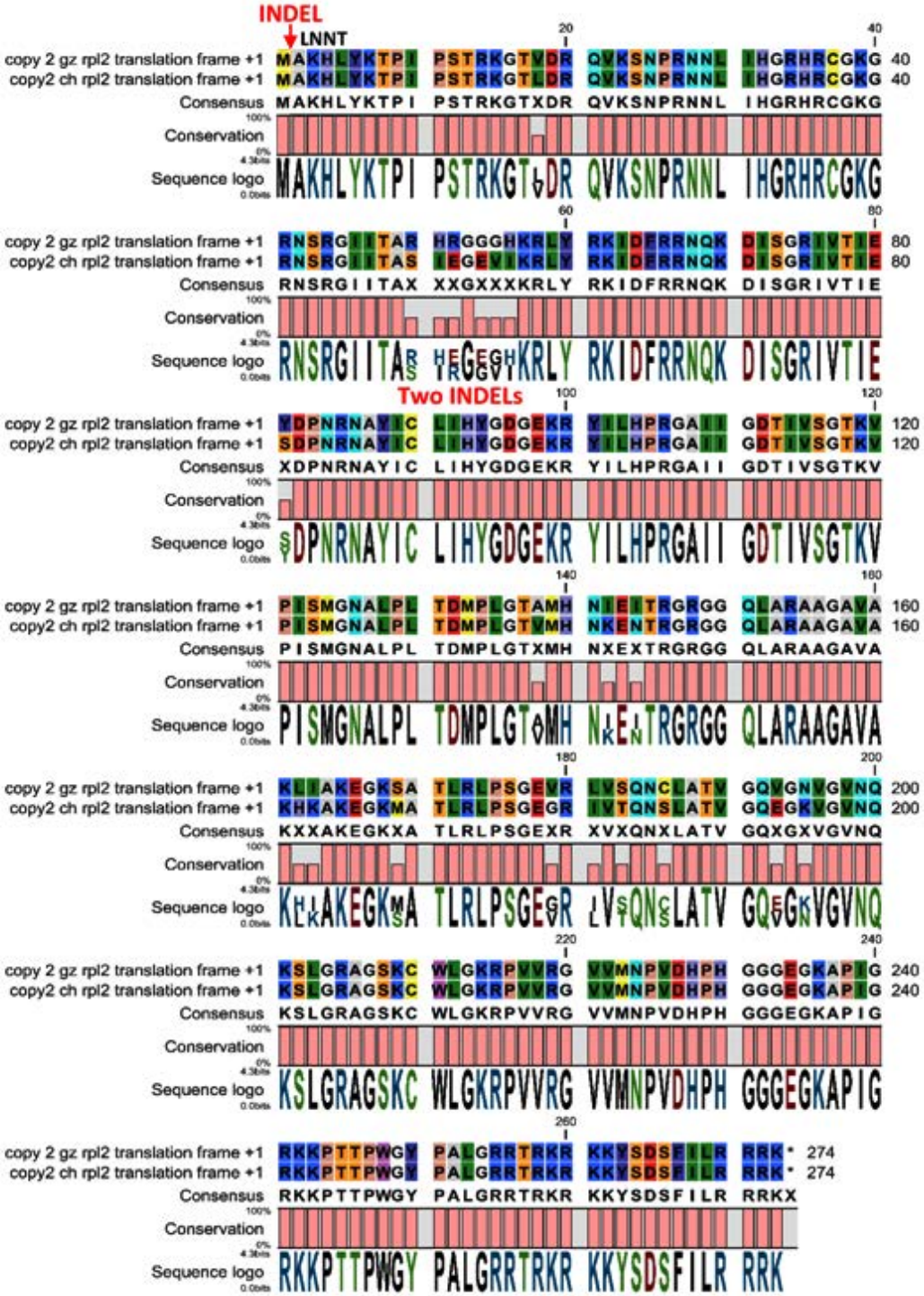

Fig. 3. Alignment of the *rpl2*-encoded amino acids sequenceof the cultivar G168 (KJ592713) to that of Chinese Spring cultivar(AB042240) indicating the changes due to SNPs (14 amino aacids) and INDELs (3).A long INDEL (LNNT) of 12 nucleotides is located right after the first amino acid (M). The other two INDELs are one-nucleotide each that changed the sequence of six amino acids, three amino acids each, with a G in the middle. Position of these six amino acids is indicated in red

other angiosperm. However, cp genome of *Chara vulgaris* possesses the highest number of long repeats (120) among angiosperms (Figure 1).

Plastome structure

Plastid nucleotide sequence of G168, as a model, was submitted to the NCBI and received the accession no. KJ592713. Plastome

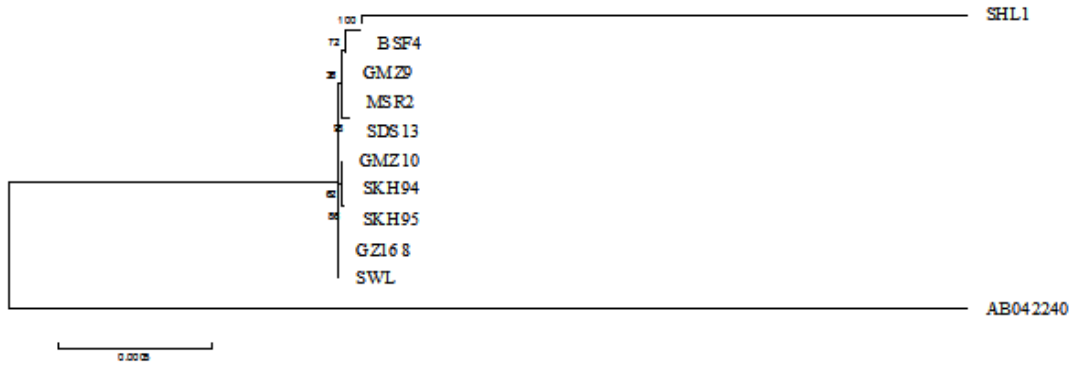

Fig. 4. Phylogenetic analysis using chloroplast sequences from nine wheat cultivars and the reference chloroplastid genome (acc. no. AB042240)with neighbor joining using routine using NTSYSpc

Table 1. Wheat cultivars examined along with their geographic locations, ploidy levels and pedigrees

No.	Name	Abbrev.	Geographic location	Ploidy level	Pedigree
1	Giza 168	GZ168	Delta, Egypt	Hexaploid	MRL/BUC//SERT
2	Shandweel	SWL	Upper Egypt	Hexaploid	SITE//MO/4/NAC/TH.AC//3*PVN/3/MRL/ BUC
3	Gemiza 10	GMZ10	Delta, Egypt	Hexaploid	MAYA74"S"/ON//1160-147/3/BB/GLL/4/ CHAT"S"/5/CROW"S"
4	Sakha 95	SKH95	Delta, Egypt	Hexaploid	N/A
5	Sakha 94	SKH94	Delta, Egypt	Hexaploid	OPATA/RAYON// KAUZ"S"
6	Misr 2	MSR2	Sinai, Egypt	Hexaploid	KAUZ"S"//BAV92
7	Sids 13	SDS13	Delta, Egypt	Hexaploid	KAUZ"S"/TSI//TSI/SNB"S"
8	Gemiza 9	GMZ9	Delta, Egypt	Hexaploid	ALD"S"/HUAC"S"//CMH74A.630/SX
9	BeniSweif 4	BSF4	Upper Egypt	Tetraploid	AUSL/5/CANDO/4/BY*2/TACE//II27655/3/ TME/ZB/W*2

Table 2. Statistics of DNA numerical data analysis for the nine wheat cultivars aligned to the chloroplast reference genome (acc. no. AB042240)

No.	Total read no.	GC (%)	No. reads mapped	No. filtered reads	Coverage	Filtered coverage	% reads mapped	% filtered reads
GZ168	107,565,480	38.31	1,195,172	803,643	1,229	799	1.11	0.75
SWL	121,447,620	38.31	1,349,418	852,158	1,394	902	1.11	0.70
GMZ10	25,334,910	38.31	281,499	219,147	864	644	1.11	0.87
SKH95	58,380,660	38.31	648,674	423,249	1,345	866	1.11	0.73
SKH94	110,930,580	38.31	1,232,562	824,490	1,279	824	1.11	0.74
MSR2	153,813,240	38.31	1,709,036	1,136,796	1,788	1,142	1.11	0.74
SDS13	121,444,110	38.31	1,349,379	895,590	1,368	902	1.11	0.74
GMZ9	195,274,620	38.31	2,169,718	1,440,201	2,243	1,450	1.11	0.74
BSF4	161,609,580	38.31	1,795,662	1,074,662	1,892	1,202	1.11	0.67

of the wheat cultivar along with gene content were generated earlier by our group.[25]Our results indicateda number of three new non-coding genes, e.g.,*trnI*, *trnT* and *trnfM* (Figure 2) located in the LSC region; of which the first two areshownin a cluster. Additionally, function of one out of three conserved ORFs, namely *ycf5* was assigned after annotation (Figure 2).The latter gene, also called*ccsA*, functions as a cytochrome c-type biogenesis protein required for heme attachment to chloroplast cytochromes.[26] Functions of the two other conserved ORFs, namely *ycf6* and *ycf9* were also deciphered.[22] Respectively, they are named *pbsZ*and *petN* genes. The first functions in photosystem II, while the second functions as a cytochrome in the generation of ATP via electron transport.

A total of 19,770 codons representing the coding capacity of all protein-coding genes of wheat cp genome were scored (Table 5). Among them, as high as 2,118 (10.71%) codons encode for leucine, while as low as 214 (1.08%) codons encode for cysteine. Yang et al.[5] indicated that isoleucine and cysteine are the most and least amino acids in plastid genome in terms of number of codons in date palm cp genome, respectively, (see Table 1, Yang et al.[5]). The most frequent codon (825) was scored for AUU encoding isoleucine. Similar conclusion was reached by Yang et al.[5] in their study on date palm cp genome. Our results also indicated that nucleotide frequencies vary at different codon positions. At the first position, "A" nucleotide is found the most frequent nucleotide (29.59%), followed by "G" (28.55%). The nucleotide "C" is the least (18.66%) at the first position. This indicates that purine is favored at the first position.At the second position, "U" is found as the most frequent nucleotide (32.70%), followed by "A" (27.61%). The nucleotide "G" scores the least (18.55%). At the third position, "U" also is the most frequent nucleotide (37.64%), followed by "A" (32.57%). The nucleotide "C" is the least frequent nucleotide (14.26%). These results indicate that "U" is favored for change at the second and third positions of the codon. Similar tendency of resultswasfound when studying codon usage in mitochondrialgenome of wheat.[11]This indicates that AT-rich genes in cp genome might be less conserved that CG-rich genes. Date palm also showed the same trend of results, except that nucleotide "C", not "G", is the least frequent at the first position of the codon in its plastide genome (Calculated from data in Table 1 of Yang et al.,[5]). The results of the relative synonymous codon usage (RSCU) indicated that UUA codon coding for leucine is the most common(2.07) compared to

Table 3. Comparative analysis of genomic features among 12 chloroplast genomes of angiosperms

Species	Size (bp)	AT (%)	No. genes*	Coding sequence (%)	Repeats (%)
Chara vulgaris	184,933	73.8	148/105/37/6	62.26	3.162
Marchantiapolymorpha	121,024	71.2	134/89/37/8	79.74	0.766
Cycastaitungensis	163,403	60.5	169/122/38/8	74.13	0.785
Arabidopsis thaliana	154,478	63.7	129/85/37/7	72.43	1.577
Nicotianasylvestris	155,941	62.2	149/101/37/8	74.99	0.878
Vitisvinifera	160,928	62.6	138/84/45/8	64.17	1.128
Phoenix dactylifera	158,462	62.8	149/95/44/8	99.39	2.729
Bambusaemeiensis	139,493	61.1	131/84/39/8	64.74	1.481
Oryza sativa/*indica*group	134,496	61.0	65/64/27/6	42.89	1.333
Sorghum bicolor	140,754	61.5	140/84/48/8	58.63	1.468
Zea mays	140,384	61.5	158/111/38/8	69.36	1.919
Triticumaestivum	133,812**	61.7	97/66***/27****/4	67.26	1.651

* Total/protein coding/tRNA/rRNA

** This size was corrected (Bahieldin et al. 2014), which is 728 bp shorter than the published wheat plastome(Ogihara et al. 2000)

*** A number of 74 protein-coding genes and two unidentified ORFs (ycf3 & ycf4)

**** A number of 30 tRNA genes plus three new sequences detected in the present study

the other codons of leucine or for any other amino acids (Table 5). This indicates that cp genes display a non-random usage of synonymous codons. The results also indicated that UAA is the most frequently-used stop codon (54.9%). A number of 28, out of the sense 61 codons, covering all the 20 amino acids have tRNAs existed in wheat plastome. Interestingly, most of the tRNAs are specific for less frequent codons. Therefore, the phenomenon of codon preference in wheat plastomeis not only explained by the frequency by which a certain codon of a given amino acid exists, but also by the availability of the cognate tRNAof such a codon (Table 5).

SNPs and INDELs analyses and cultivars relationships

Across the different Egyptian cultivars, 564 SNPs and 160 INDELs were identified in the study, of which 230 and 4 are in the protein-coding regions, respectively (Table 6). The number of monomorphic SNPs and INDELs are 553 and 154, respectively. A number of 212 SNPs were found in the long inverted repeat (IR) regions, of which 104 were found in the IRA and 108 were found in theIRBregion. One SNP, while none for INDELs, was found in the genic regions unique to one of the two inverted repeats (IRA) in the coding sequence of *ndhB* gene. The similarity of SNPs patterns in both IR regions is due to the fact that cp genome is conserved. However, there is a possibility that one single read within these regions might be mapped to either region. This possibility reduces the chance to detect the different patterns, if existed, of the IR region. Therefore, SNPs analysis using next generation sequencing of total genomic DNA should be taken cautiously. It is likely that the duplication of the IR region took place way after the occurrence of point mutations during evolution. Numbers of inter-cultivar polymorphic and cultivar-specific SNPs were nine and five, respectively (Table 6). The latter number was scored only in the intergenic spacers (IGS) region for cultivar BSF4. Among the polymorphic SNPs, six were found in the IGS regions, while only one was found in the introns (IN) of *atpF* gene and two SNPs were found in the GN regions of *proA* and

Table 4. The gene content across the nine assembled *Triticumaestivum* chloroplast genomes

Category	Gene name	No.
Ribosomal RNA	*rrn23S (x2), rrn16S (x2), rrn5S (x2), rrn4.5S (x2)*	8
Transfer RNAs	*trnA-UGC(x2), trnC-GCA, trnD-GTC, trnE-TTC, trnF-GAA, trnfM-CAT(x2), trnG-GCC, trnG-TCC, trnH-GTG(x2), trnI-GAT(x2), trnK-TTT, trnL-CAA(x2), trnL-TAA, trnL-TAG, trnM-CAT, trnN-GTT(x2), trnP-TGG, trnQ-TTG, trnR-ACG(x2), trnR-TCT, trnS-GCT, trnS-GGA, trnT-GGT, trnT-TGT, trnV-GAC(x2), trnW-CCA, trnY-GTA*	35
Photosystem I	*psaA, psaB, psaC, psaI, psaJ*	5
Photosystem II	*psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ (ycf9)*	15
Cytochrome b/f complex	*petA, petB, petD, petG, petL, petN (ycf6)*	6
ATP synthase	*atpA, atpB, atpE, atpF, atpH, atpI*	6
NADH dehydrogenase	*ndhA, ndhB(x2), ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*	12
RubisCO large subunit	*rbcL*	1
RNA polymerase	*rpoA, rpoB, rpoC1, rpoC2*	4
Ribosomal proteins (SSU)	*rps2, rps3, rps4, rps7(x2), rps8, rps11, rps12(x2), rps14, rps15(x2), rps16, rps18, rps19(x2)*	16
Ribosomal proteins (LSU)	*rpl2(x2), rpl14, rpl16, rpl20, rpl22, rpl23(x2), rpl32, rpl33, rpl36*	11
Other genes	*clpP, matK, ccsA (ycf5), infA, cemA*	5
hypothetical chloroplast reading frames	*ycf3, ycf4*	2
Total no.	116	

rpl16 genes. Numbers of 15 and ninepolymorphic and cultivar-specific INDELs were also found of which 10 and eight INDELs, respectively, are located in the IGS regions, while fivepolymorphic and one cultivar-specific INDEL, respectively, are located in the IN regions of the *rpl16* gene.

The highest number of SNPs in the protein coding regions was scored for gene *ndhB* (31 and 30 in the IRA and IRB regions, respectively), followed by *rpl2* (18 in each IR region), then *matK* (34) and *psbC* (25). One long INDEL of 12 nucleotide exists in the *rpl2* gene and starts at nucleotide 4 of the gene (Figure 3). Nucleotide sequence of this INDEL encodes for four amino acids (LNNT). Two other INDELsthat are 19-nt apart starting from nucleotide 160 of the gene were detected in the *rpl2* gene (Figure 3). The first is an inserted nucleotide in the nine wheat cultivars, while the second was a deleted nucleotide compared to Chinese Spring cultivar. The latter two INDELs resulted in a frameshift of six amino acids,with a glycine amino acid in the middle remains unchanged, then default frame wasregained(Figure 3). We concluded that *rpl2* gene in the reference genome is 12-nt shorter than that of the Egyptian cultivars. It is unlikely that the change in these amino acids has posed

Table 5. Codon usage and codon-anticodon recognition pattern for tRNA in nine assembled wheat chloroplast genomes

Amino acid	Codon	No.	RSCU*	tRNA	Amino acid	Codon	No.	RSCU	tRNA
Phe	UUU	730	1.33		Ser	UCU	402	1.71	
	UUC	368	0.67	trnF-GAA		UCC	255	1.08	trnS-GGA
Leu	UUA	731	2.07	trnL-TAA		UCA	242	1.03	trS-TGA
	UUG	385	1.09	trnL-CAA		UCG	116	0.49	
	CUU	443	1.26		Pro	CCU	343	1.61	trnP-TGG
	CUC	145	0.41			CCC	189	0.89	
	CUA	308	0.87	trnL-TAG		CCA	225	1.06	
	CUG	106	0.30			CCG	96	0.45	
Ile	AUU	825	1.52		Thr	ACU	457	1.71	
	AUC	297	0.55	trnI-GAT		ACC	184	0.69	trnT-GGT
	AUA	502	0.93			ACA	305	1.14	trnT-TGT
Met	AUG	456	1.00	trnfM-CAT		ACG	121	0.45	
Val	GUU	425	1.45		Ala	GCU	548	1.76	
	GUC	144	0.49	trnV-GAC		GCC	185	0.60	
	GUA	450	1.54	trnV-TAC		GCA	378	1.22	trnA-TGC
	GUG	150	0.51			GCG	133	0.43	
Tyr	UAU	567	1.58		Cys	UGU	164	1.53	
	UAC	152	0.42	trnY-GTA		UGC	50	0.47	trnC-GCA
Stop	UAA	45	1.65		Stop	UGA	17	0.62	
Stop	UAG	20	0.73		Trp	UGG	343	1.00	trnW-CCA
His	CAU	334	1.49		Arg	CGU	282	1.39	trnR-ACG
	CAC	115	0.51	trnH-GTG		CGC	110	0.54	
Gln	CAA	513	1.56	trnQ-TTG		CGA	252	1.24	
	CAG	144	0.44			CGG	84	0.42	
Asn	AAU	595	1.50		Ser	AGU	290	1.23	
	AAC	201	0.51	trnN-GTT		AGC	107	0.46	trnS-GCT
Lys	AAA	745	1.46	trnK-TTT	Arg	AGA	362	1.79	trnR-TCT
	AAG	278	0.54			AGG	125	0.62	
Asp	GAU	556	1.56		Gly	GGU	480	1.30	
	GAC	155	0.44	trnD-GTC		GGC	163	0.44	trnG-GCC
Glu	GAA	779	1.50	trnE-TTC		GGA	584	1.58	trnG-TCC
	GAG	259	0.50			GGG	255	0.69	

*RSCU: Relative synonymous codon usage

Table 6. SNPs and INDELs within plastid genomes of the nine Egyptian wheat cultivars as sorted by position and region of the genome. Plastid genome of Chinese Spring cultivar was used as the reference genome (acc. no. AB042240).GN refers to protein-coding genic regions, IN refers to intron regions and IGS refers to intergenic spacer regions, S refers to synonymous substitution, NS refers to non-synonymous.Letters in INDELs refer to insertions and (-) refers to deletions.Red blocks refer to SNPs in the protein-coding regions. Green blocks indicate SNPs unique to one of the two inverted repeats (IR) regions. Blue block indicates the unique SNP to one IR (IRa) region.Orange blocks indicate INDELs within the IR region that showed similar patterns in the two regions.

No.	Position	1-9[1]	REF	Region	Gene	No.	Position	1-9	REF	Region	Gene
SNPs											
1	1160	T	A	IGS	-	283	11335	T	C	GN	*psbC*
2	1186	T	C	IGS	-	284	11374	A	T	GN	*psbC*
3	1223	A	G	IGS	-	285	11395	A	G	GN	*psbC*
4	1275	T	C	IGS	-	286	14971	G (1,2)[2]	C	IGS	-
5	1282	A	C	IGS	-	287	29930	T (4,5)	C	IGS	-
6	1283	A	G	IGS	-	288	32015	A (9)	G	IGS	-
7	1285	G	C	IGS	-	289	32020	C (9)	G	IGS	-
8	1287	T	G	IGS	-	290	32025	G (9)	A	IGS	-
9	1289	A	T	IGS	-	291	32041	A (9)	G	IGS	-
10	1301	T	C	IGS	-	292	32052	C (4,5)	G	IGS	-
11	1305	T	A	IGS	-	293	32077	T (9)	A	IGS	-
12	1311	A	T	IGS	-	294	33103	T (6,7)	C	IGS	-
13	1322	T	A	IGS	-	295	33518	T (6,7)	C	IN	*atpF*
14	1325	T	A	IGS	-	296	60528	C	T	GN	*petA*
15	1350	C	G	IGS	-	297	60541	T	A	GN	*petA*
16	1354	G	A	IGS	-	298	60542	T	A	GN	*petA*
17	1355	T	A	IGS	-	299	60544	T	G	GN	*petA*
18	1357	A	C	IGS	-	300	60547	T	C	GN	*petA*
19	1364	T	C	IGS	-	301	60551	C	A	GN	*petA*
20	1365	T	A	IGS	-	302	60578	T	G	GN	*petA*
21	1369	G	T	IGS	-	303	60580	G	T	GN	*petA*
22	1371	G	C	IGS	-	304	60582	T	C	GN	*petA*
23	1374	T	C	IGS	-	305	60583	C	T	GN	*petA*
24	1440	C	A	IN	*trnK*	306	60584	T	C	GN	*petA*
25	1464	T	A	IN	*trnk*	307	60585	C	A	GN	*petA*
26	1507	T	A	IN	*trnK*	308	60586	A	T	GN	*petA*
27	1525	C	G	IN	*trnK*	309	60587	A	C	GN	*petA*
28	1527	A	T	IN	*trnK*	310	60590	A	C	GN	*petA*
29	1539	A	G	IN	*trnK*	311	60909	T	C	IGS	-
30	1566	A	G	IN	*trnK*	312	60912	T	C	IGS	-
31	1584	C	T	IN	*trnK*	313	60981	T	A	IGS	-
32	1588	T	A	IN	*trnK*	314	61022	G	C	IGS	-
33	1589	C	A	IN	*trnK*	315	61088	A	T	IGS	-
34	1599	A	G	IN	*trnK*	316	61129	A	T	IGS	-
35	1602	A	C	IN	*trnK*	317	61140	G	T	IGS	-
36	1604	A	G	IN	*trnK*	318	61141	A (1,2)	C	IGS	-
37	1621	A	G	IN	*trnK*	319	61167	T	G	IGS	-
38	1625	A	C	IN	*trnK*	320	6117	T	A	IGS	-
39	1627	A	G	IN	*trnK*	321	61200	T	C	IGS	-
40	1628	G	T	IN	*trnK*	322	61239	T	C	IGS	-
41	1629	A	G	IN	*trnK*	323	61544	C	A	GN	*psbJ*
42	1638	T	A	IN	*trnK*	324	61573	C	A	GN	*psbJ*
43	1639	T	C	IN	*trnK*	325	61722	A	C	GN	*psbL*

44	1656	C	T	IN	*trnK*	326	61736	A	G	GN	*psbL*
45	1658	C	A	IN	*trnK*	327	61833	A	G	GN	*psbF*
46	1663	C	T	IN	*trnK*	328	61834	A	G	GN	*psbF*
47	1664	T	A	IN	*trnK*	329	61931	A	G	GN	*psbF*
48	1669	C	A	IN	*trnK*	330	62074	G	A	GN	*psbE*
49	1673	A	G	IN	*trnK*	331	62121	A	T	GN	*psbE*
50	1677	C	T	IN	*trnK*	332	73770	G	A	GN	*petD*
51	1680	C	T	IN	*trnK*	333	74736	A (4,5)	G	GN	*proA*
52	1699	G	A	GN	*matK*	334	77436	T	G	GN	*rpl14*
53	1702	G	C	GN	*matK*	335	77693	T (4,5)	C	GN	*rpl16*
54	1708	G	A	GN	*matK*	336	81245	A	T	GN	*rpl2*
55	1720	T	G	GN	*matK*	337	81248	A	T	GN	*rpl2*
56	1722	T	G	GN	*matK*	338	81255	A	T	GN	*rpl2*
57	1748	G	C	GN	*matK*	339	81274	A	G	GN	*rpl2*
58	1753	T	C	GN	*matK*	340	81277	A	T	GN	*rpl2*
59	1759	A	C	GN	*matK*	341	81286	A	T	GN	*rpl2*
60	1761	A	G	GN	*matK*	342	81292	A	T	GN	*rpl2*
61	1771	A	G	GN	*matK*	343	81297	A	C	GN	*rpl2*
62	1772	A	T	GN	*matK*	344	81305	A	G	GN	*rpl2*
63	1773	G	A	GN	*matK*	345	81327	G	A	GN	*rpl2*
64	1785	T	C	GN	*matK*	346	81328	A	T	GN	*rpl2*
65	1817	A	C	GN	*matK*	347	81344	A	T	GN	*rpl2*
66	1838	G	A	GN	*matK*	348	81345	A	T	GN	*rpl2*
67	1851	G	A	GN	*matK*	349	81348	A	T	GN	*rpl2*
68	1863	T	C	GN	*matK*	350	81395	A	T	GN	*rpl2*
69	1886	C	G	GN	*matK*	351	81402	A	T	GN	*rpl2*
70	1889	G	C	GN	*matK*	352	82408	A	T	GN	*rpl2*
71	1943	C	T	GN	*matK*	353	81420	G	A	GN	*rpl2*
72	1944	G	T	GN	*matK*	354	81446	A	G	IN	*rpl2*
73	1945	T	C	GN	*matK*	355	81482	A	T	IN	*rpl2*
74	1951	C	T	GN	*matK*	356	81483	A	T	IN	*rpl2*
75	1963	A	G	GN	*matK*	357	82445	T	G	GN	*rpl23*
76	1999	T	C	GN	*matK*	358	82324	C	A	GN	*rpl23*
77	2111	G	T	GN	*matK*	359	82596	G	T	GN	*rpl23*
78	2610	G	T	GN	*matK*	360	82599	C	T	GN	*rpl23*
79	2611	A	G	GN	*matK*	361	82608	T	G	GN	*rpl23*
80	2616	A	T	GN	*matK*	362	82611	T	C	GN	*rpl23*
81	2673	A	G	GN	*matK*	363	82623	T	C	GN	*rpl23*
82	2674	G	A	GN	*matK*	364	82629	A	G	GN	*rpl23*
83	2692	A	G	GN	*matK*	365	82647	A	C	GN	*rpl23*
84	3127	G	A	GN	*matK*	366	82656	C	T	GN	*rpl23*
85	3128	T	G	GN	*matK*	367	83205	A	C	IGS	-
86	3335	T	C	GN	*trnK*	368	83324	A	G	IGS	-
87	3340	A	G	GN	*trnK*	369	83346	G	A	IGS	-
88	3347	G	A	IN	*trnK*	370	83443	G	A	IGS	-
89	3362	A	G	IN	*trnK*	371	83448	A	G	IGS	-
90	3373	T	C	IN	*trnK*	372	83474	T	G	IGS	-
91	3386	C	T	IN	*trnK*	373	83481	C	T	IGS	-
92	3393	A	T	IN	*trnK*	374	83529	G	A	IGS	-
93	3413	C	A	IN	*trnK*	375	83566	C	T	IGS	-
94	3414	A	C	IN	*trnK*	376	83575	A	C	IGS	-
95	3419	A	G	IN	*trnK*	377	83577	G	T	IGS	-
96	3434	G	A	IN	*trnK*	378	83657	A	G	IGS	-
97	3436	T	C	IN	*trnK*	379	83755	A	G	IGS	-
98	3437	T	C	IN	*trnK*	380	83791	C	G	IGS	-
99	3457	C	T	IN	*trnK*	381	83801	G	T	IGS	-

100	3474	A	G	IN	*trnK*	382	83991	C	T	IGS	-
101	3481	C	T	IN	*trnK*	383	84260	A	G	IGS	-
102	3530	C	T	IN	*trnK*	384	84354	A	G	IGS	-
103	3543	T	A	IN	*trnK*	385	84365	T	G	IGS	-
104	3585	A	C	IN	*trnK*	386	84367	C	T	IGS	-
105	3588	G	A	IN	*trnK*	387	84368	T	C	IGS	-
106	3593	C	T	IN	*trnK*	388	84449	A	C	IGS	-
107	3611	A	G	IN	*trnK*	389	84463	A	G	IGS	-
108	3622	C	T	IN	*trnK*	390	84464	G	A	IGS	-
109	3777	A	C	IN	*trnK*	391	84504	C	G	IGS	-
110	4339	T	C	IGS	-	392	84545	A	C	IGS	-
111	4345	A	T	IGS	-	393	84555	A	G	IGS	-
112	4606	C	A	GN	*rps16*	394	84594	C	T	GN	*trnL*
113	4618	C	T	GN	*rps16*	395	84658	G	A	IGS	-
114	4694	T	C	GN	*rps16*	396	84938	A	C	IGS	-
115	4889	T	G	IN	*rps16*	397	85090	T	C	IGS	-
116	4891	A	G	IN	*rps16*	398	85918	A	G	GN	*ndhB*
117	4922	A	C	IN	*rps16*	399	85921	G	A	GN	*ndhB*
118	4930	G	T	IN	*rps16*	400	85922	A	G	GN	*ndhB*
119	4949	A	G	IN	*rps16*	401	85924	T	A	GN	*ndhB*
120	4954	C	G	IN	*rps16*	402	85925	A	T	GN	*ndhB*
121	4955	G	A	IN	*rps16*	403	85946	A	G	GN	*ndhB*
122	4959	G	A	IN	*rps16*	404	85972	C	T	GN	*ndhB*
123	4960	C	A	IN	*rps16*	405	85977	A	T	GN	*ndhB*
124	5147	A	T	IN	*rps16*	406	85990	C	A	GN	*ndhB*
125	5317	G	T	IN	*rps16*	407	85991	T	G	GN	*ndhB*
126	5325	C	T	IN	*rps16*	408	85992	G	T	GN	*ndhB*
127	5359	A	G	IN	*rps16*	409	85994	A	G	GN	*ndhB*
128	5364	C	A	IN	*rps16*	410	85995	G	A	GN	*ndhB*
129	5462	G	T	IN	*rps16*	411	85996	T	G	GN	*ndhB*
130	5492	C	A	IN	*rps16*	412	85997	A	T	GN	*ndhB*
131	5498	T	C	IN	*rps16*	413	85998	G	A	GN	*ndhB*
132	5506	A	G	IN	*rps16*	414	86017	G	A	IN	*ndhB*
133	5520	G	A	IN	*rps16*	415	86018	A	G	IN	*ndhB*
134	5561	A	G	IN	*rps16*	416	86019	G	A	IN	*ndhB*
135	5587	A	G	IN	*rps16*	417	86021	A	G	IN	*ndhB*
136	5641	C	A	GN	*rps16*	418	86522	A	T	IN	*ndhB*
137	5677	G	T	IGS	-	419	86671	A	T	IN	*ndhB*
138	5683	A	G	IGS	-	420	86804	T	G	GN	*ndhB*
139	5722	G	A	IGS	-	421	86838	G	T	GN	*ndhB*
140	5727	A	C	IGS	-	422	86927	T	C	GN	*ndhB*
141	5746	A	C	IGS	-	423	86951	T	A	GN	*ndhB*
142	5771	C	A	IGS	-	424	86954	T	A	GN	*ndhB*
143	5778	G	A	IGS	-	425	86957	T	A	GN	*ndhB*
144	5789	G	T	IGS	-	426	86962	T	C	GN	*ndhB*
145	5802	C	G	IGS	-	327	86984	T	C	GN	*ndhB*
146	5805	G	A	IGS	-	428	87296	T	A	GN	*ndhB*
147	5809	C	A	IGS	-	429	87337	T	A	GN	*ndhB*
148	5816	C	A	IGS	-	430	87353	T	A	GN	*ndhB*
149	5821	T	G	IGS	-	431	87390	T	C	GN	*ndhB*
150	5867	C	T	IGS	-	432	87435	T	C	GN	*ndhB*
151	5874	T	C	IGS	-	433	87447	T	A	GN	*ndhB*
152	5881	G	A	IGS	-	434	87521	A	G	IGS	-
153	5882	A	G	IGS	-	435	87543	T	A	IGS	-
154	5883	C	G	IGS	-	436	87620	T	C	IGS	-
155	5886	C	A	IGS	-	437	87645	T	C	IGS	-

156	5904	A	G	IGS	-	438	87761	C	T	IGS	-
157	5916	C	A	IGS	-	439	87782	A	C	IGS	-
158	5918	T	G	IGS	-	440	87877	T	C	GN	*rps7*
159	5919	T	G	IGS	-	441	94621	A	C	IN	*trnA*
160	5928	C	T	IGS	-	442	97095	C	T	IGS	-
161	5936	G	A	IGS	-	443	101188	C	G	IGS	-
162	5939	T	G	IGS	-	444	101241	C	A	GN	*ndhF*
163	5941	G	A	IGS	-	445	101328	C	G	GN	*ndhF*
164	5968	G	A	IGS	-	446	101355	C	A	GN	*ndhF*
165	5972	G	T	IGS	-	447	101606	C	G	GN	*ndhF*
166	5981	G	A	IGS	-	448	102640	G	T	GN	*ndhF*
167	5993	C	T	IGS	-	449	105635	T	C	GN	*ccsA*
168	5994	T	C	IGS	-	450	105859	T	G	GN	*ccsA*
169	5998	T	C	IGS	-	451	105865	T	C	GN	*ccsA*
170	6018	A	C	IGS	-	452	105868	T	C	GN	*ccsA*
171	6052	C	T	IGS	-	453	105869	T	C	GN	*ccsA*
172	6056	A	C	IGS	-	454	105876	T	C	GN	*ccsA*
173	6058	T	A	IGS	-	455	106112	T	G	GN	*ccsA*
174	6063	A	T	IGS	-	456	106116	T	G	GN	*ccsA*
175	6066	A	T	IGS	-	457	106123	T	G	GN	*ccsA*
176	6082	A	C	IGS	-	458	106156	T	G	GN	*ccsA*
177	6086	C	A	IGS	-	459	106176	T	C	GN	*ccsA*
178	6092	A	C	IGS	-	460	106237	A	C	GN	*ccsA*
179	6093	C	T	IGS	-	461	117992	G	A	IGS	-
180	6112	A	G	IGS	-	462	120466	T	G	IN	*trnA*
181	6113	A	T	IGS	-	463	127210	A	G	GN	*rps7*
182	6114	A	T	IGS	-	464	127305	T	G	IGS	-
183	6123	C	G	IGS	-	465	127326	G	A	IGS	-
184	6140	A	G	IGS	-	466	127442	A	G	IGS	-
185	6141	G	T	IGS	-	467	127467	A	G	IGS	-
186	6168	C	T	IGS	-	468	127537	A	C	IGS	-
187	6175	A	G	IGS	-	469	127561	T	C	IGS	-
188	6192	G	A	IGS	-	470	127640	A	T	GN	*ndhB*
189	6235	T	A	IGS	-	471	127652	A	G	GN	*ndhB*
190	6236	G	A	IGS	-	472	127697	A	G	GN	*ndhB*
191	6238	C	T	IGS	-	473	127734	A	T	GN	*ndhB*
192	6250	A	C	IGS	-	474	127750	A	T	GN	*ndhB*
193	6276	G	A	IGS	-	475	127791	A	T	GN	*ndhB*
194	6277	T	A	IGS	-	476	128103	A	G	GN	*ndhB*
195	6281	G	A	IGS	-	477	128125	A	G	GN	*ndhB*
196	6558	A	T	IGS	-	478	128130	A	T	GN	*ndhB*
197	6559	G	T	IGS	-	479	128133	A	T	GN	*ndhB*
198	6577	A	G	IGS	-	480	128136	A	T	GN	*ndhB*
199	6579	G	C	IGS	-	481	128160	A	G	GN	*ndhB*
200	6583	T	G	IGS	-	482	128249	C	A	GN	*ndhB*
201	6584	T	G	IGS	-	483	128283	A	C	GN	*ndhB*
202	6601	A	C	IGS	-	484	128416	T	A	IN	*ndhB*
203	6606	T	A	IGS	-	485	128565	T	A	IN	*ndhB*
204	6608	A	T	IGS	-	486	128923	A	G	IN	*ndhB*
205	6611	T	A	IGS	-	487	129089	C	T	GN	*ndhB*
206	6612	A	C	IGS	-	488	129090	T	A	GN	*ndhB*
207	6613	T	G	IGS	-	489	129091	A	C	GN	*ndhB*
208	6684	A	G	IGS	-	490	129092	C	T	GN	*ndhB*
209	6686	G	A	IGS	-	491	129093	T	C	GN	*ndhB*
210	6693	T	A	IGS	-	492	129095	C	A	GN	*ndhB*
211	6697	T	C	IGS	-	493	129096	A	C	GN	*ndhB*

212	6705	C	T	IGS	-	494	129097	G	T	GN	*ndhB*
213	6707	G	T	IGS	-	495	129110	T	A	GN	*ndhB*
214	6710	T	C	IGS	-	496	129115	G	A	GN	*ndhB*
215	6724	C	A	IGS	-	497	129141	T	C	GN	*ndhB*
216	6829	C	G	GN	*trnQ*	498	129162	T	A	GN	*ndhB*
217	6832	G	T	GN	*trnQ*	499	129163	A	T	GN	*ndhB*
218	6844	G	C	GN	*trnQ*	500	129165	T	C	GN	*ndhB*
219	6857	A	C	IGS	-	501	129166	C	T	GN	*ndhB*
220	6859	C	A	IGS	-	502	129169	T	C	GN	*ndhB*
221	6861	T	C	IGS	-	503	129237	G	A	GN	*ndhB*
222	6878	A	T	IGS	-	504	129997	A	G	IGS	-
223	6881	A	G	IGS	-	505	130149	T	G	IGS	-
224	6896	C	G	IGS	-	506	130217	C (3,8,9)	T	IGS	-
225	6900	T	G	IGS	-	507	130428	C	T	IGS	-
226	6907	C	G	IGS	-	508	130492	G	A	GN	*trnL*
227	6915	G	A	IGS	-	509	130531	T	C	IGS	-
228	6929	A	C	IGS	-	510	130541	T	G	IGS	-
229	6989	A	C	IGS	-	511	130582	G	C	IGS	-
230	7069	T	G	IGS	-	512	130622	T	C	IGS	-
231	7070	G	T	IGS	-	513	130623	C	T	IGS	-
232	7091	A	G	IGS	-	514	130637	T	G	IGS	-
233	7105	G	T	IGS	-	515	130718	A	G	IGS	-
234	7106	A	C	IGS	-	516	130719	G	A	IGS	-
235	7120	T	C	IGS		517	130721	A	C	IGS	-
236	7134	T	C	IGS	-	518	130732	T	C	IGS	-
237	7139	T	C	IGS	-	519	130826	T	C	IGS	-
238	7176	C	T	GN	*psbK*	520	131095	G	A	IGS	-
239	7192	C	A	GN	*psbK*	521	131331	T	C	IGS	-
240	7200	T	G	GN	*psbK*	522	131429	T	C	IGS	-
241	7215	T	C	GN	*psbK*	523	131509	C	A	IGS	-
242	7224	A	G	GN	*psbK*	524	131511	T	G	IGS	-
243	7239	T	C	GN	*psbK*	525	131520	G	A	IGS	-
244	7261	A	T	GN	*psbK*	526	131557	C	T	IGS	-
245	7272	T	C	GN	*psbK*	527	131605	G	A	IGS	-
246	7494	A	G	IGS	-	528	131612	A	C	IGS	-
247	7880	A	T	IGS	-	529	131638	T	C	IGS	-
248	8642	C	A	IGS	-	530	131643	C	T	IGS	-
249	8649	T	C	IGS	-	531	131740	C	T	IGS	-
250	8656	G	A	IGS	-	532	131762	T	C	IGS	-
251	8657	A	T	IGS	-	533	131881	T	G	IGS	-
252	8693	T	C	IGS	-	534	132430	G	A	GN	*rpl23*
253	8858	G	T	IGS	-	535	132439	T	G	GN	*rpl23*
254	8883	T	G	IGS	-	536	132457	T	C	GN	*rpl23*
255	8900	G	T	IGS	-	537	132463	A	G	GN	*rpl23*
256	8913	T	G	IGS	-	538	132475	A	G	GN	*rpl23*
257	8954	T	C	IGS	-	539	132478	A	C	GN	*rpl23*
258	9184	T	G	GN	*psbD*	540	132487	G	A	GN	*rpl23*
259	9185	G	T	GN	*psbD*	541	132490	C	A	GN	*rpl23*
260	9229	T	G	GN	*psbD*	542	132641	G	T	GN	*rpl23*
261	10296	A	G	GN	*psbC*	543	132832	A	C	GN	*rpl2*
262	10357	G	A	GN	*psbC*	544	133603	T	A	IN	*rpl2*
263	10373	T	C	GN	*psbC*	545	133604	T	A	IN	*rpl2*
264	10536	T	C	GN	*psbC*	546	133640	T	C	IN	*rpl2*
265	10537	T	C	GN	*psbC*	547	133666	C	T	GN	*rpl2*
266	10555	T	C	GN	*psbC*	548	133678	T	A	GN	*rpl2*
267	10566	T	C	GN	*psbC*	549	133684	T	A	GN	*rpl2*

268	10596	G	C	GN	*psbC*	550	133691	T	A	GN	*rpl2*
269	10627	G	A	GN	*psbC*	551	133738	T	A	GN	*rpl2*
270	10663	T	C	GN	*psbC*	552	133741	T	A	GN	*rpl2*
271	10666	C	T	GN	*psbC*	553	133742	T	A	GN	*rpl2*
272	10687	A	C	GN	*psbC*	554	133758	T	A	GN	*rpl2*
273	10694	T	C	GN	*psbC*	555	133759	C	T	GN	*rpl2*
274	10784	T	C	GN	*psbC*	556	133781	T	C	GN	*rpl2*
275	10848	C	G	GN	*psbC*	557	133789	T	G	GN	*rpl2*
276	10879	T	C	GN	*psbC*	558	133794	T	A	GN	*rpl2*
277	10978	T	C	GN	*psbC*	559	133800	T	A	GN	*rpl2*
278	11041	T	C	GN	*psbC*	560	133809	T	A	GN	*rpl2*
279	11308	C	T	GN	*psbC*	561	133812	T	C	GN	rpl2
280	11327	T	G	GN	*psbC*	562	133831	T	A	GN	*rpl2*
281	11329	A	G	GN	*psbC*	563	133838	T	A	GN	*rpl2*
282	11330	A	T	GN	*psbC*	564	133841	T	A	GN	*rpl2*
INDELs											
1	1292	-	A	IGS	-	81	70935	T	-	IN	*petB*
2	1329	G	-	IGS	-	82	71498	-	T	IN	*petB*
3	1330	T	-	IGS	-	83	78533	A (4,5)[2]	-	IN	*rpl16*
4	1331	A	-	IGS	-	84	78534	A (4,5)	-	IN	*rpl16*
5	1332	A	-	IGS	-	85	78535	A (4,5)	-	IN	*rpl16*
6	1333	A	-	IGS	-	86	78536	A (4,5)	-	IN	*rpl16*
7	1467	-	T	IN	*trnK*	87	78537	A (4,5)	-	IN	*rpl16*
8	1550	T	-	IN	*trnK*	88	78538	A (9)	-	IN	*rpl16*
9	1568	T	-	GN	*trnK*	89	82328	-	T	GN-NS	*rpl2*
10	3084	T	-	GN	*trnK*	90	82348	C	-	GN-NS	*rpl2*
11	3085	A	-	GN	*trnK*	91	83171	G	-	IGS	-
12	3086	A	-	GN	*trnK*	92	83763	C (3,8,9)	-	IGS	-
13	3252	-	G	IN	*trnK*	93	83877	T	-	IGS	-
14	3323	-	T	IN	*trnK*	94	83878	T	-	IGS	-
15	3336	-	T	IN	*trnK*	95	83879	C	-	IGS	-
16	3337	-	G	IN	*trnK*	96	83880	C	-	IGS	-
17	3421	A	-	IN	*trnK*	97	83881	T	-	IGS	-
18	3422	A	-	IN	*trnK*	98	83882	C	-	IGS	-
19	3423	G	-	IN	*trnK*	99	84160	T	-	IGS	-
20	3424	A	-	IN	*trnK*	100	84161	T	-	IGS	-
21	3425	A	-	IN	*trnK*	101	84162	G	-	IGS	-
22	3426	C	-	IN	*trnK*	102	84163	A	-	IGS	-
23	3427	A	-	IN	*trnK*	103	84164	T	-	IGS	-
24	3533	A	-	IN	*trnK*	104	84174	A	-	IGS	-
25	3534	T	-	IN	*trnK*	105	84262	T	-	IGS	-
26	3609	C	-	IN	*trnK*	106	84419	A	-	IGS	-
27	3757	-	A	IN	*trnK*	107	84420	T	-	IGS	-
28	4965	-	A	IN	*rps16*	108	84421	A	-	IGS	-
29	4966	-	A	IN	*rps16*	109	84422	T	-	IGS	-
30	5157	-	C	IN	*rps16*	110	84867	- (9)	A	IGS	-
31	5158	-	T	IN	*rps16*	111	84868	- (9)	A	IGS	-
32	5649	A	-	IGS	-	112	84869	T (8)	-	IGS	-
33	5650	A	-	IGS	-	113	84870	A (8)	-	IGS	-
34	5651	C	-	IGS	-	114	84872	- (3)	A	IGS	-
35	5652	A	-	IGS	-	115	84873	- (3)	A	IGS	-
36	5660	-	A	IGS	-	116	86022	-	A	IN	*ndhB*
37	5661	-	A	IGS	-	117	86023	-	T	IN	*ndhB*
38	5735	-	G	IGS	-	118	86105	-	A	IN	*ndhB*
39	5749	A	-	IGS	-	119	86163	T	-	IN	*ndhB*

40	5750	A	-	IGS	-	120	86164	C	-	IN	*ndhB*
41	5751	A	-	IGS	-	121	87525	-	A	IGS	-
42	5752	A	-	IGS	-	122	87526	-	G	IGS	-
43	5753	T	-	IGS	-	123	87548	-	T	IGS	-
44	5754	T	-	IGS	-	124	87549	-	T	IGS	-
45	6018	A	-	IGS	-	125	87550	-	G	IGS	-
46	6019	A	-	IGS	-	126	87653	G	-	IGS	-
47	6020	A	-	IGS	-	127	87664	-	T	IGS	-
48	6021	A	-	IGS	-	128	93885	-	C	IGS	-
49	6022	A	-	IGS	-	129	103716	-	A	IGS	-
50	6023	A	-	IGS	-	130	104196	T	-	IGS	-
51	6080	- (2-9)	T	IGS	-	131	121203	-	G	IGS	-
52	6104	A	-	IGS	-	132	125930	G	-	IGS	-
53	6105	A	-	IGS	-	133	127424	-	A	IGS	-
54	6135	T	-	IGS	-	134	127436	C	-	IGS	-
55	6136	T	-	IGS	-	135	127542	-	A	IGS	-
56	6137	G	-	IGS	-	136	127543	-	A	IGS	-
57	6551	T	-	IGS	-	137	127544	-	T	IGS	-
58	6675	-	A	IGS	-	138	127565	-	T	IGS	-
59	6885	A	-	IGS	-	139	127566	-	C	IGS	-
60	7058	-	G	IGS	-	140	128924	G	-	IN	*ndhB*
61	7081	C	-	IGS	-	141	128925	A	-	IN	*ndhB*
62	8337	-	A	IGS	-	142	128984	-	T	IN	*ndhB*
63	8338	-	G	IGS	-	143	129064	-	A	IN	*ndhB*
64	8339	-	C	IGS	-	144	129065	-	T	IN	*ndhB*
45	8340	-	A	IGS	-	145	130218	A (3-8,9)	-	IGS	-
66	8520	-	G	IGS	-	146	130669	T	-	IGS	-
67	8643	-	A	IGS	-	147	130670	A	-	IGS	-
68	8737	T	-	IGS	-	148	130671	T	-	IGS	-
69	8738	T	-	IGS	-	149	130672	A	-	IGS	-
70	32046	T (9)	-	IGS	-	150	130824	A	-	IGS	-
71	32047	A (9)	-	IGS	-	151	130913	T	-	IGS	-
72	33773	A	-	IN	*atpF*	152	130927	A	-	IGS	-
73	37273	A	-	IGS	-	153	130928	T	-	IGS	-
74	61183	- (1-5, 8)	T	IGS	-	154	130929	C	-	IGS	-
75	63063	T (6-7)	-	IGS	-	155	130930	A	-	IGS	-
76	65596	A (4,5)	-	IGS	-	156	130931	A	-	IGS	-
77	65597	A (4,5)	-	IGS	-	157	131323	T	-	IGS	-
78	65598	C (4,5)	-	IGS	-	158	131916	C	-	IGS	-
79	65599	A (4,5)	-	IGS	-	159	132739	G	-.	GN-NS	*rpl2*
80	65600	A (4,5)	-	IGS	-	160	132760	-	A	GN-NS	*rpl2*

[1]See Table 1
[2]A SNP or an INDEL that is unique to these cultivars

any functional constraints on proteins encoded by either versions of the gene as they were proven to be effectively functioning.

Based on the SNPs of the different nine cultivars in addition to the reference plastid genome, dendrogramwas constructed (Figure 4). The tree was well-resolved withhigh bootstrap support for resolved nodes. This might be due to the fact that the Egyptian cultivars are closely related on one hand, and genetically distant from the reference genome, on the other hand. The results indicated the correspondence between tree topology and linage of eight out of the nine cultivars. The cultivar pairs G168/SWL,SHK94/SKH95 and MSR2/SDS13 are closely related. In other words, the cultivars with shared ancestors showed genetically closer relationships. As no information is available on the lineage of SKH95, it is likely that it shares a common ancestor with SKH94. Interestingly, the tetraploid cp genome was closely related to

the other Egyptian hexaploid wheat cultivars as compared to the reference hexaploid wheat cultivar Chinese Spring. The SNPs/INDELs tree was not resolved and bootstrap support values were low (data provided upon request). This is due to the fact that some INDELs might be artifacts rather than real. The INDELsinside the IR region are more reliable as they should show similar patterns in the two inverted regions.

There are no intra-cultivar polymorphic SNPs were detected. This might be due to the fact that sequences ofthe mt genome mapped to the cp genome were filtered out and artifacts were removed before cp genome assembly. Generally speaking, intra-varietal heteroplasmy in the wheat cp genome within the studied cultivars does not exist in contradiction with previous reportsin other plants.[5,27]

CONCLUSION

We conclude that plastome SNPs and INDELs successfully separated wheat cultivars and results aligned with the known ancestral information of the different genotypes.

REFERENCES

1. Bausher,M.G, Singh,N.D,, Lee,S.B,, Jansen,R.K., Daniell, H. The complete chloroplast genome sequence of *Citrus sinensis*(L.) Osbeckvar 'Ridge Pineapple': organization and phylogenetic relationships to other angiosperms. *BMC Plant Biol.* 2006; **6**:21.
2. Howe,C.J., Barbrook,A.C., Koumandou,V.L., Nisbet,R.E., Symington,H.A., et al. Evolution of the chloroplast genome. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 2003; **358**: 99–106; discussion 106-7.
3. Ogihara, Y., Isono, K., Kojima, T., Endo, A., Hanaoka, M., et al. Chinese Spring wheat (*Triticum aestivum* L.) chloroplast genome: Complete sequence and contig clones. *Plant Mol. Biol. Rep.* 2000; **18**: 243-53.
4. Ogihara, Y., Yamazaki, Y., Murai, K., Kanno, A., Terachi, T., et al. Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. *Nucleic Acids Res.* 2005; **33**: 6235-50.
5. Yang, M., Zhang, X., Liu, G., Yin, Y., Chen, K., et al. The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). *PLoS ONE* 2010; **5**:e12762.
6. Chumley,T.W., Palmer, J.D., Mower, J.P., Fourcade, H.M., Calie,P.J., et al. The complete chloroplast genome sequence of *Pelargonium* x *hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Mol. Biol.Evol.* 2006; **23**: 2175-90.
7. Hansen, A.K., Escobar, L.E., Gilbert, L.E., Jansen, R.K. Paternal, maternal, and biparental inheritance of the chloroplast genome in *Passiflora* (Passifloraceae): Implications for phylogenetic studies. *Am. J. Bot.* 2007a; **94**:42-6.
8. Hansen,D.R., Dastidar,S.G., Cai, Z., Penaflor, C., Kuehl,J.V., et al. 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*. 2007b; **45**: 547-63.
9. Mardanov,A.V., Ravin, N.V., Kuznetsov,B.B., Samigullin,T.H., Antonovm A.S., et al. Complete sequence of the duckweed (*Lemna minor*) chloroplast genome: structural organization and phylogenetic relationships to other angiosperms. *J. Mol.Evol.* 2008; **66**: 555-64.
10. Ling, H.-Q., Zhao, S., Liu, D., Wang, J., Sun, H., et al. Draft genome of the wheat A-genome progenitor *Triticumurartu*. *Nature* 2013; **496**: 87-90.
11. Cui, P., Liu, H., Lin, Q., Ding, F., Zhuo, G., et al. A complete mitochondrial genome of wheat (*Triticum aestivum* cv. Chinese Yumai), and fast evolving mitochondrial genes in higher plants. *J. Genet*. 2009; **88**: 299-307.
12. Fang, Y., Wu, H., Zhang, T., Yang, M., Yin, Y., et al. A complete sequence and transcriptomic analyses of date palm (*Phoenix dactylifera* L.) mitochondrial genome. *PLoS ONE* 2012; **7**: e37164.
13. Khan, A., Khan,I.A., Heinze, B., Azim,M.K. The chloroplast genome sequence of Date palm (*Phoenix dactylifera* L. cv. 'Aseel'). *Plant Mol. Biol. Rep*. 2012; **30**: 666–78.
14. Birky,C.W. Relaxed cellular controls and organelle heredity. *Science* 1983; **222**: 468-75.
15. Chat, J., Decroocq, S., Decroocq, V., Petit, R.J. A case of chloroplast heteroplasmy in Kiwifruit (*Actinidiadeliciosa*) that is not transmitted during sexual reproduction. *The J.Hered*. 2002; **93**: 293-300.
16. Frey, J.E., Frey, B., Forcioli, D. Quantitative assessment of heteroplasmy levels in *Senecio vulgaris* chloroplast DNA. *Genetica* 2005; **123**: 255-61.

17. Gawel, N.J., Jarret,R.L. A modified CTAB DNA extraction procedure for Musa and Ipomoea. *Plant Mol. Biol. Rep.* 1991; **9**: 262-66.
18. Altschul, S.F., Madden,T.L., Schaffer, A.A., Zhang, J., Zhang, Z., et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997; **25**: 3389-3402.
19. Lowe,T.M., Eddy, S.R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.*1997; **25**: 955-64.
20. Kurtz, S., Choudhuri,J.V., Ohlebusch, E., Schleiermacher, C., Stoye, J., et al. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 2001; **29**: 4633-42.
21. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 1999; **27**: 573-80.
22. Zhang, T., Fang, Y., Wang, X., Deng, X., Zhang, X., et al. The complete chloroplast and mitochondrial genome sequences of *Boeahygrometrica*: Insights into the evolution of plant organellar genomes. *PLoS ONE* 2012; **7**: e30531.
23. Tang, J., Xia, H., Cao, M., Zhang, X., Zeng, W., et al. A comparison of rice chloroplast genomes. *Plant Physiol.* 2004; **135**:412-20.
24. Sabir, J.S.M., Arasappan, D., Bahieldin, A., Abo-Aba, S., Bafeel, S., et al. Whole mitochondrial and plastid genome SNP analysis of nine date palm cultivars reveals plastid heteroplasmy and relationships among cultivars. *PloS ONE* 2014; **9**: e94158.
25. Bahieldin, A., Al-Kordy, M.A., Shokry, A.M., Gadalla,N.O., Al-Hejin,A.M.M., et al. Corrected sequence of the wheat plastid genome. *C. R. Biol.* 2014; **337**: 499-502.
26. Feissner,R.E., Beckett,C.S., Loughman, J.A., Kranz, R.G. Mutations in cytochrome assembly and periplasmic redox pathways in *Bordetella pertussis*. *J.Bacteriol.* 2005; **187**: 3941-9.
27. Straub, S.C.K., Parks, M., Weitemier, K., et al. Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. *Am. J. Bot.* 2012; **99**: 349-64.