DOI: 10.1007/s10535-017-0759-y BIOLOGIA PLANTARUM 62 (2): 231-240, 2018

The complete chloroplast genome sequence of *Pseudoroegneria libanotica*, genomic features, and phylogenetic relationship with *Triticeae* species

D.D. WU[1,2], L.N. SHA[1], C. TANG[1], X. FAN[1], Y. WANG[1], H.Y. KANG[1], H.Q. ZHANG[1], and Y.H. ZHOU[1,2]*

Triticeae Research Institute, Sichuan Agricultural University, Wenjiang 611130, Sichuan, P.R. China[1]
Key Laboratory of Crop Genetic Resources and Improvement, Ministry of Education, Sichuan Agricultural University, Wenjiang 611130, Sichuan, P.R.China[2]

Abstract

Pseudoroegneria libanotica is an important herbage diploid species possessing the St genome. The St genome participates in the formation of nine perennial genera in *Triticeae* (*Poaceae*). The whole chloroplast (cp) genome of *P. libanotica* is 135 026 bp in length. The typical quadripartite structure consists of one large single copy of 80 634 bp, one small single copy of 12 766 bp and a pair of inverted regions (20 813 bp each). The cp genome contains 76 coding genes, four ribosomal RNA and 30 transfer RNA genes. Comparative sequence analysis suggested that: *1*) the 737 bp deletion in the cp of *P. libanotica* was specific in *Triticeae* species and might transfer into its nuclear genome; *2*) hot-spot regions, indels in intergenic regions and protein coding sequences mainly led to the length variation in *Triticeae*; *3*) highly divergence regions combined with negative selection in *rpl2*, *rps12*, *ccsA*, *rps8*, *ndhH*, *petD, ndhK, psbM, rps3, rps18*, and *ndhA* were identified as effective molecular markers and could be considered in future phylogenetic studies of *Triticeae* species; and *4*) *ycf3* gene with rich cpSSRs was suitable for phylogeny analysis or could be used for DNA barcoding at low taxonomic levels. The cpSSRs distribution in the coding regions of diploid *Triticeae* species was shown for the first time and provided a valuable source for developing primers to study specific simple sequence repeat loci.

Additional key words: hot-spot regions, phylogenetic tree, sequence transfer, simple sequence repeats.

Introduction

Chloroplasts are multifunctional organelles possessing own genetic material. The function of chloroplasts is to perform photosynthesis, to storage starch, to metabolize nitrogen and sulfate, and synthesize chlorophyll, carotenoids, fatty acids, and nucleic acids (Neuhaus and Emes 2000, Rodríguez-Ezpeleta *et al*. 2005). Chloroplast (cp) genome consists of two copies of inverted repeats (IRs) separating the large single-copy region (LSC, 80 - 90 kb) and small single-copy region (SSC, 16 - 27 kb) (Palmer 1991). The gene content and structure of the majority of angiosperm cp genomes are highly conserved (Wicke *et al*. 2011), but expansion and contraction of the IR, as well as gene inversion, loss and rearrangements, even transferring into nuclear genome have been reported (Matsuo *et al*. 2005). Because of the highly conservative nature and slow evolutionary rate of the cp genome,

Submitted 17 December 2016, *last revision* 13 May 2017, *accepted* 6 June 2017.
Abbreviations: BS - bootstrap value; CNS - conserved noncoding sequences; cp - chloroplast; cpSSR - chloroplast simple sequence repeat; d_N/d_S - ratio of nonsynonymous and synonymous substitution rates; indel - insertion and deletion; IR - inverted region; LSC - large single copy; ML - maximum likelihood; rRNA - ribosomal RNA; SSC - small single copy; TE - transposable element; tRNA - transfer RNA.
Acknowledgments: The authors are thankful to the National Natural Science Foundation of China (Grant Nos. 31470305, 31270243, and 31200252), the National Key Research and Development of China (2016YFD0102000), and the fund from the Science and Technology Bureau and Education Bureau of Sichuan Province, China. We are very grateful to the American National Plant Germplasm System (Pullman, Washington, USA) for providing seed material for this study. The first two authors contributed equally to this work.
* Corresponding author; fax: (+86) 28 82650350, e-mail: zhouyh@sicau.edu.cn

extensive molecular evolution data demonstrated that cpDNA sequences are suitable and invaluable tool for phylogenetic relationships and the maternal donors studies in polyploids (Mason-Gamer *et al*. 2002, Zhang *et al*. 2009, Sha *et al*. 2010).

The tribe *Triticeae*, an economically important plant group, includes not only economic crop species (*e.g*., wheat, rye, triticale, and barley) but also significant forage resources (*e.g*., *Pseudoroegneria strigosa*, *Elymus sibiricus*, *Leymus chinensis*). *Pseudoroegneria* is a perennial genus of *Triticeae*, including 11 species (Yen and Yang 2011). The basic genomic constitution was designed as St genome. Extensive cytological studies have suggested that the St genome with the other basic genomes W, H, Y, P, and E constitute in various combinations different genera such as *Douglasdeweya* (StP), *Elymus* (StH), *Roegneria* (StY), *Trichopyrum* (StE), *Anthosachne* (StYW), *Campeiostachys* (StYH), *Kengyilia* (StYP), *Psammopyrum* (EStP), and *Pascopyrum* (StHNsXm). The P, H, and W genomes are derived from *Agropyron*, *Hordeum*, and *Australopyrum*, respectively (Jensen 1989, Torabinejad and Mueller 1993, Lu 1994, Yen and Yang 2011). Phylogenetic analysis based on limited cpDNA sequences (*rpoA*, *ndhF, trnH-psbA, etc*.) have concluded that *Pseudoroegneria* species served as the maternal donor during the speciation of the St-containing polyploid species in *Triticeae* (Redinbaugh *et al*. 2000, Mason-Gamer *et al*. 2002, Liu *et al*. 2006, Sha *et al*. 2010, Dong *et al*. 2015). Nevertheless, on the basis of *trnL-F* region Zhang *et al*. (2009) demonstrated that the maternal donor of *Kengyilia melanthera* and *K. thoroldiana* may originate from *Agropyron* (P genome) species. Although these studies add to our understanding of phylogeny and maternal donor of *Triticeae* species, discordance still exists in St-containing species due to incomplete and different cp molecular markers. A robust phylogenetic analysis requires a substantial sequence length and primarily uniparental inheritance, which makes the cp genome a valuable source for genetic markers (Provan *et al*. 2001). The complete plastid genome information is very useful within maternal donors and it helps to decipher species evolution (Gornicki *et al*. 2014, Ma *et al*. 2014).

In this work, *Illumina* reads from a partial genome sequence of *Pseudoroegneria libanotica* were used to assemble the cp genome. The objectives in this study were: *1*) to determine the complete cp genome of *P. libanotica*, and *2*) to compare whole cp genomes among eight diploid *Triticeae* species. Based on those data sets, we revealed a solid resource for phylogenetic studies and comparative genomics of the *Triticeae* species.

Materials and methods

Plant material, DNA extraction, and high-throughput sequencing: Genomic DNA was extracted from fresh young leaves of diploid *Pseudoroegneria libanotica* (Hackel) D.R. Dewey (PI228392) using the mCTAB method (Li *et al*. 2013a). This material was planted at the Greenhouse of Sichuan Agriculture University in 2015 and the voucher specimen was deposited in the herbarium of the Triticeae Research Institute, Sichuan Agricultural University, China. Genomic DNA was fragmented by ultrasound treatment, and gel extraction was used to obtain DNA fragments with sizes between 400 - 600 bp. The *NEBNext*® kit (*New England Biolabs*, Ipswich, MA, USA) was applied to construct a 500 bp DNA library, and *Hiseq 4000 PE150* (*Illumina*, CA, USA) was utilized for sequencing.

Genome assembly, annotation, comparison and sequence analyses: High-throughput sequencing data were assembled by *SPAdes 3.6.1* (Bankevich *et al*. 2012), *CLC Genomics Workbench 8* and *SOAPdenovo2* (Luo *et al*. 2012). Chloroplast contigs were screened from whole genome sequences using the *BLAST* process. Assembly of cp reads was conducted with the help of *Sequencher 4.10* (*Gene Codes*, Ann Arbor, MI, USA). *Geneious 8.1* (Kearse *et al*. 2012) was used to map all reads to the total cp sequence to confirm the validity of contigs. Specific primers were designed for PCR in order to repair gaps. PCR condition: at 94 °C for 4 min, 34 cycles at 94 °C for 30 s, at 55 °C for 30 s, at 72 °C for 1.5 min, then at 72 °C for 10 min. The PCR products were sequenced by *ABI 3730* (*Applied Biosystems*, Foster City, CA, USA).

The cp genome was annotated using the program *DOGMA* (Wyman *et al*. 2004), coupled with manual corrections for start and stop codons. The coding and RNA genes were identified using *BLASTX* and *BLASTN*. The border of exons and introns, and some short exons (such as *rps16*, *petB*, *petD*) which could not be identified by *DOGMA*, were referred to published sequences. *OrganellarGenomeDRAW* (http://ogdraw.mpimp-golm.mpg.de/index.shtml) was used to construct the total cp genome structure. *Adobe Illustrator CS5* was utilized for editing.

Chloroplast simple sequence repeats (cpSSRs) of *P. libanotica* were screened by *SSRHunter1.3* (Li and Wan 2005), with thresholds of eight repeat units for mononucleotide SSRs, four repeat units for di- and tri-nucleotide SSRs (8 and 12 bp, respectively), and three repeat units for tetra- and penta-nucleotide SSRs (12 and 15 bp, respectively). Eight complete plastid genomes of cereal species were downloaded from *NCBI* (http://www.ncbi.nlm.nih.gov/) (Table 1 Suppl.). To align the cp genome of *P. libanotica* with other cp genomes *mVISTA* (Frazer *et al*. 2004) was employed in *Shuffle-LAGAN*

mode. Alignment of the complete cp genome and coding regions of eight diploid *Triticeae* species were conducted by *MAFFT* (http://mafft.cbrc.jp/alignment/ server/) independently. Protein coding sequences alignments were carried out by *DNAMAN 6.0.3.48* (*Lynnon Biosoft*, Quebec, Canada). To detect selective constraints on the coding portions (introns were excluded) of all protein coding genes, the ratio of non-synonymous to synonymous substitution (d_N/d_S) were computed using the modified Nei-Gojobori method in *MEGA 6.0* software (Tamura *et al*. 2013). Pairwise sequence divergences were calculated by using the Kimura's two-parameter model and nucleotide polymorphisms estimates for each gene was completed by *DNASp 4.0* (Rozas *et al*. 2003).

Phylogenetic analyses: Phylogenetic analysis was conducted using maximum likelihood (ML) with two sets of the data-complete cp (non-coding and protein coding) and protein-coding sequence (cds). The evolutionary model used for the phylogenetic analysis was determined using *ModelTest v. 3.0* with Akaike Information Criterion (AIC) (Posada and Crandall 1998). The optimal model identified was generalized time-reversible + gamma distribution + proportion of invariable sites (GTR+G+I). The ML analysis was performed using *PAUP*4.0b10* (Swofford and Sinauer Associates, http://www.sinauer.com). The ML heuristic searches were performed with 100 random addition sequence replications and tree bisection and reconnection (TBR) branch swapping algorithm. The robustness of the trees was estimated by bootstrap support (BS). Trees were viewed in *FigTree v. 1.4.1* (http://tree.bio.ed.ac.uk/).

Results

The total reads obtained from *P. libanotica* by *Illumina* sequencing were 13 265 148 and 102 083 mapped reads were screened for the entire cp genome. The whole cp genome of *P. libanotica* was 135 026 bp in length. The typical quadripartite structure consisted of one LSC of 80 634 bp, one SSC of 12 766 bp, and a pair of IRs of 20 813 bp each (Fig. 1 Suppl.). The overall guanine/cytosine (GC) content of the *P. libanotica* cp genome was 38.35 %. The GC content of the IR regions (44.02 %) was higher than that of the LSC (36.39 %) and SSC regions (32.23 %). High GC content in IR regions is mainly due to the presence of the rRNA gene group.

The cp genome of *P. libanotica* contained 110 encoding genes, including 30 tRNA, four rRNA (23S, 16S, 5S, and 4.5S) and 76 protein coding genes. Thirteen genes included one intron and one gene (*ycf3*) included two introns. Based on the position of 76 protein coding regions, 60 of them located in LSC, 10 in SSC, and six in IR regions. The cp genes can be divided into three functional types (Table 1). The first type contains

Table 1. Genes in the chloroplast genome of *Triticeae.* Intron-containing genes are marked by asterisks.

Category for genes	Group of gene	Name of gene
Photosynthesis related genes	Rubisco	*rbcL*
	photosystem I	*psaA,B,C,I,J*
	assembly/stability of photosystem I	**ycf3,ycf4*
	photosystem II	*psbA, B, C, D, E, F, H, I, J, K, L, M, N, T, Z*
	ATP synthase	*atpA, B, E, *F, H, I*
	cytochrome *b*/*f* complex	*petA, *B, *D, G, L, N*
	cytochrome *c* synthesis	*ccsA*
	NADPH dehydrogenase	**ndhA, *B, C, D, E, F ,G, H, I, J,K*
Transcription and translation related genes	transcription transcription	*rpoA, B, C1, C2*
	ribosomal proteins	*rps2, 3, 4, 7, 8, 11, 12, 14, 15, 16, 18, 19; *rpl2, 14, *16, 20, 22, 23, 32, 33,36*
	translation initiation factor	*infA*
RNA genes	ribosomal RNA	*rrn5, rrn4.5, rrn16, rrn23*
	transfer RNA	**trnA-UGC, C-GCA, D-GUC, E-UUC, F-GAA, *G-UCC, *G-GCC, H-GUG, I-CAU, *I-GAU, *K-UUU, L-CAA, *L-UAA, L-UAG, fM-CAUI, M-CAU, N-GUU, P-UGG, Q-UUG, R-ACG, R-UCU, S-GCU, S-GGA, S-UGA, T-GGU, T-UGU, V-GAC, *V-UAC, W-CCA, Y-GUA*
Other genes	RNA processing	*matK*
	carbon metabolism	*cemA*
	proteolysis	*clpP*

60 genes, which are involved in transcription and translation. The second type contains 47 genes involved in photosynthesis. The third type contains three genes, responsible for the biosynthesis of amino acids and fatty acids and of unknown function.

Chloroplast SSRs are generally short mononucleotide tandem repeats and randomly spread throughout cp genomes, stretching from one to six nucleotide units. Perfect SSRs which are longer than 8 bp were detected in the *P. libanotica* cp genome. This threshold was set because SSRs of 8 bp or longer are prone to slipstrand mispairing, which is thought to be the primary mutational mechanism causing their high level of polymorphism (Rose and Falush 1998, Huotari and Korpelainen 2012). In *P. libanotica* complete cp genome, 128 cpSSRs repeats were found. Total cpSSRs consisted of five types: mononucleotides (91 SSRs: 71 %), dinucleotides (28 SSRs: 22 %), trinucleotides (two SSRs: 1 %), tetranucleotides (six SSRs: 5 %), and pentanucleotides (only one SSR: 1 %). There were 98 (76.6 %), 15 (11.7 %), and 15 (11.7 %) SSR repeats distributing in LSC, IR, and SSC regions, respectively (Fig. 1*A*). There were 38 (29.7 %) SSRs in exon regions, 11 (8.6 %) in intron regions, and 79 (61.7 %) in intergenic spacer (Fig. 1*B*). According to the distribution of SSR loci in coding genes of cp genome, the coding region of *rpoC2* contained the highest number of repeats with 6 SSRs rich in adenosine/thymine (A/T). Coding regions of *ndhF* and *ycf3* included five SSRs each (Table 2 Suppl.).

Structural mutations occur at hot-spot regions,

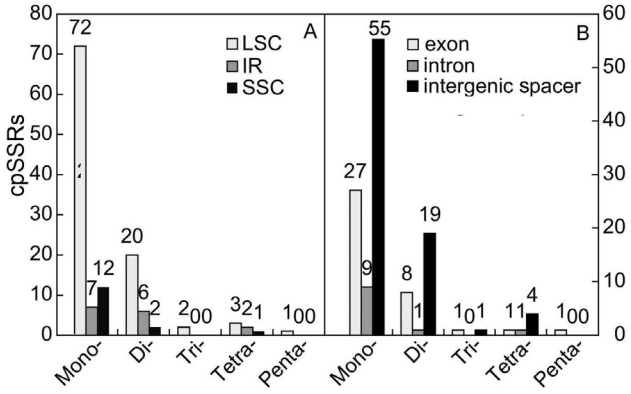

Fig. 1. The distribution of cp simple sequence repeats (cpSSRs) in *Pseudoroegneria libanotica*. *A* - SSR location according to type: mononucleotide, dinucleotide, trinucleotide, tetranucleotide, and pentanucleotide repeats. The number of SSR type in LSC, IR and SSC are displayed. *B* - Different SSRs type distribute among exon, intron, and intergenic spacer.

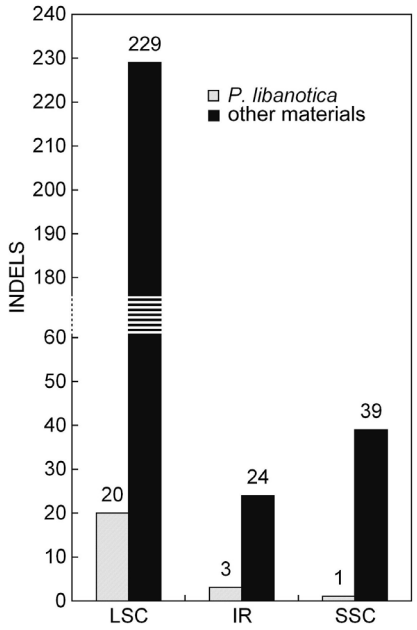

Fig. 2. Insertions and deletions in *Pseudoroegneria libanotica* and other species.

containing AT-rich content which contribute to the variation of genome size, rather than randomly. The cpDNA is well-known by AT-rich. Gene spacer regions are especially AT-rich. The overall AT content in the cp sequences of *Triticeae* diploid species is not as high as in liverwort (61.65 % and 80 %, respectively) whereas the intergenic regions of *rps4-trnL-trnL* (about 1 600 bp) and *rbcL-psaI* (about 1 200 bp) in *Triticeae* diploid species showed an AT content of more than 80 % (Fig. 2 Suppl.). *TrnL-F,* a widely used phylogenetic analyses marker, belongs to *rps4-trnL-trnL* hot-spot region. The hot-spots region *rbcL-psaI* includes the *rpl23* gene, which exhibits multiple mutations in different *Triticeae* diploid species. In the LSC region of *P. libanotica*, *rpl23* fulfils its function normally while it turned out to be a pseudogene or completely lost in the same position of other *Triticeae* diploid species.

Comparison of seven *Triticeae* cp genomes except the outgroup *Bromus vulgaris* (Table 1 Suppl.), revealed that *Triticum monococcum* (136 886 bp) and *T. urartu* (136 865 bp) possessed the biggest cp genome. *Hordeum jubatum* (136 826 bp) and *H. vulgare* subsp. *vulgare* (136 462 bp) were second in size. *Aegilops speltoides* var. *ligustica* (135 660 bp), *Ae. tauschii* (135 568 bp), and

Secale cereale (135 604 bp) had a cp genome of similar size. *P. libanotica* (135 026 bp) exhibited the smallest cp genome. *P. libanotica* displayed the highest GC content (38.35 %), followed by *Ae. speltoides* var. *ligustica* (38.31 %) and *Ae. tauschii* (38.34 %). The lowest GC content showed *T. monococcum* (38.27 %) and *T. urartu* (38.2 8%). Moreover, the total number of cp genes differed among the *Triticeae* species. Most of *Triticeae* species contained about 76 coding genes. *H. vulgare* subsp. *vulgare* had 75 coding regions in all and lacked *ycf4*, *S. cereale* encoded 75 genes. In *S. cereale*, 73 of the protein coding regions were the same as in other species while it was short for *rpl20* and *rpl23*. However, *S. cereale* possessed two unique *ycf* regions (*ycf2* and *ycf15*). Four conserved rRNAs were identified in all diploid *Triticeae* species.

The organization of the cp genome is rather conserved within diploid *Triticeae* species (Fig. 3 Suppl.) but 316 indels were found in total and 250, 27, and 40 indels in the LSC, IR and SSC regions, respectively (Fig. 2). Most of indels were located in intergenic spacer. In contrast, six indels exist in protein coding regions. In *Ae. speltoides* var. *ligustica* and *Ae. tauschii*, the *rpl2* gene lost a large proportion of the intron (total intron is 651 bp, lost 586 bp) and the complete exon-2 with respect to *P. libanotica* (Table 3 Suppl.). Similarly, *petB*, *petD*, *rpl16*, and *rps12* (in IRa) genes lost their introns in *Ae. speltoides* var. *ligustica*, *Ae. tauschii*, *S. cereale*, *T. monococcum*, and *T. urartu*, respectively, whereas they retained in *P. libanotica*. The *infA* gene length in *P. libanotica* (342 bp) was larger than in *Ae. speltoides* var. *ligustica*, *Ae. tauschii*, and *S. cereale*. The *P. libanotica* retained one copy of *rpl23* in LSC. Nevertheless, it showed early termination and formed a pseudo-gene in *Ae. speltoides* var. *ligustica*. The *rpl23* is absent in the LSC region of *H. vulgare* subsp. *vulgare*, *Ae. tauschii*, and *H. jubatum* while it was completely lost in cp genome of *S. cereale*.

P. libanotica contains 20 indels, 16 of them situated in LSC, three in IR, and one in SSC region (Fig. 2). Most of those indels were less than 100 bp in length. The longest sequences change was 737 bp absences in the *trnI-trnL* intergenic spacer of IRa and IRb regions in spite of IR regions were considered as more stable than single copy regions (LSC and SSC).

The non-synonymous (d_N) and synonymous (d_S)

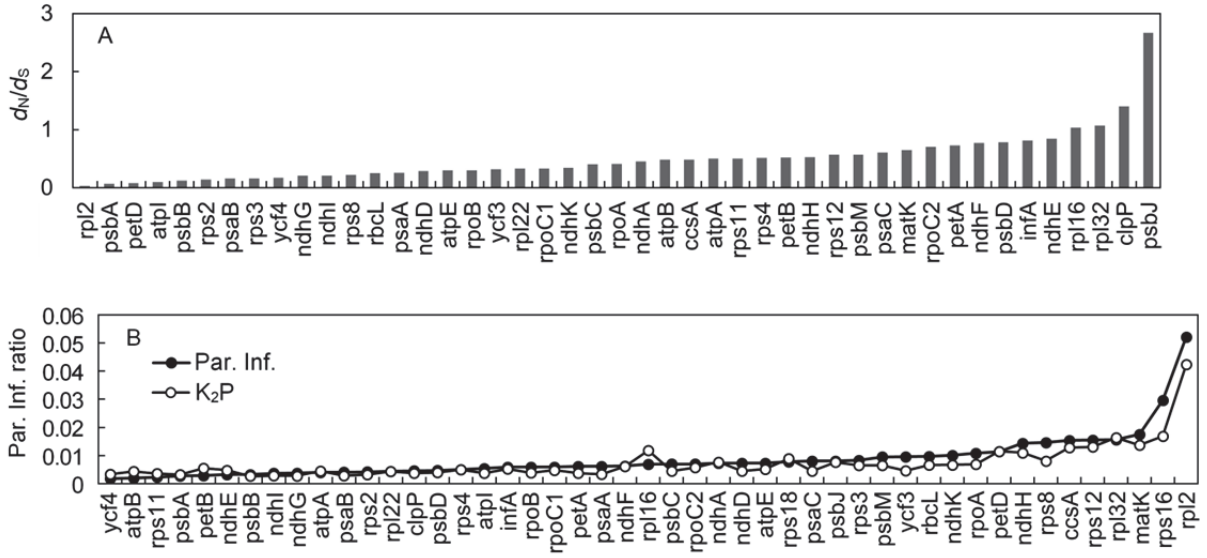

Fig. 3. Selective pressures in protein coding genes of eight diploid *Triticeae* species. *A* - d_N/d_S in protein coding genes in eight diploid *Triticeae* species (excluding genes where d_N and/or $d_S = 0$). *B* - Evolutionary rate of chloroplast genes in eight diploid *Triticeae* species (excluding genes where d_N and/or $d_S = 0$). Par. Inf. - parsimony-informative, K2P = Kimura's two-parameter model.

substitution rates ($\omega = d_N/d_S$) of >1, =1 and, <1 indicate positive selection, neutral evolution and purifying selection on the coding portions, respectively (Yang and Bielawski 2000). Parsimony-informative (Par. Inf.) ratios and evolutionary rate of coding regions were intended to analyse the most divergent coding regions in our species.

No changes were observed in the synonymous rate of *psaI*, *rpl33*, *rpl36* and the nonsynonymous rate of the *atpF*, *cemA*, *ndhB* (IR region), *ndhC*, *ndhJ*, *petL*, *psbE*, *psbH*, *psbI*, *psbK*, *psbL*, *rpl14*, *rpl20*, *rps14*, *rps15* (IR region) and *rps19* (IR region). Furthermore, *atpH*, *petG*, *petN*, *psbF*, *psbN*, *psbT*, *rpl23* (IR region), *rps7* (IR region) and *rps12* (IR region) showed no synonymous and non-synonymous changes. Most of the genes in IR (*ndhB*, *rps15*, *rps19*, *rpl23*, *rps7* and *rps12*) had no changes in the synonymous rate and/or nonsynonymous rate. The d_N/d_S ratio indicates most of the protein coding sequences got through purifying selection, except four genes (*clpP*, *psbJ*, *rpl16*, and *rpl32*) underwent positive selection (Fig. 3*A*).

Parsimony-informative characters and pairwise sequence divergences in coding regions showed similar changing tendency (Fig. 3*B*). Most cp genes contained less mutation. The *rpl2*, located in IRs, possessed the highest percentage of Par. Inf. characters and evolutionary rate (0.0521 and 0.0424, respectively), followed by *rps16* (0.0296 Par. Inf. and 0.0169 pairwise sequence divergences). In addition, analysis of *matK*, *rps12*, *ccsA*, *rps8*, *ndhH*, *petD*, *ndhK*, *rpoA*, *rbcL*, *psbM*, *ndhA*, *rps18*, *rps3*, and *ndhF* coding regions analysis showed a high Par. Inf. ratio (ranging from 0.0175 to 0.0063) and evolutionary rate (0.0164 to 0.0061).

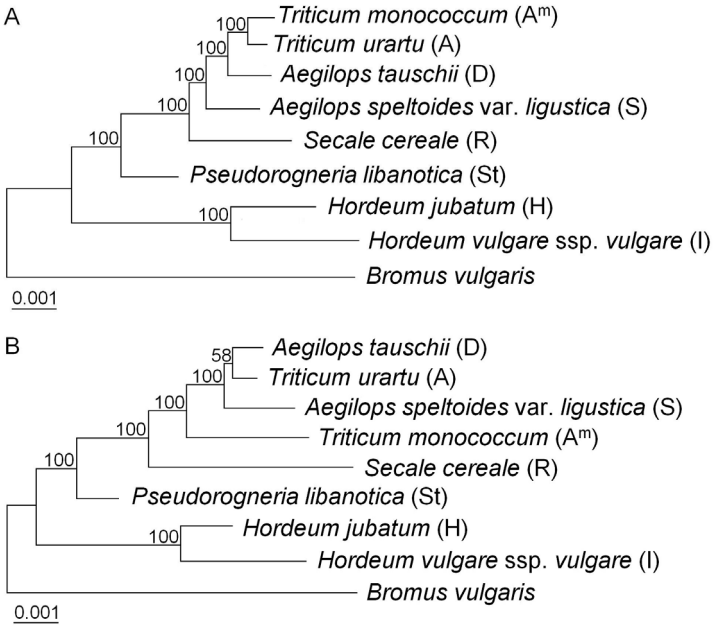

Fig. 4. Phylogenetic analyses based on complete cp sequences and coding regions, using the maximum likelihood (ML) tree. Numbers at nodes indicate bootstrap values (BS). *A* - ML tree based on the whole cp genomes; *B* - ML tree based on the coding region of cp genomes.

Bromus vulgaris was used as outgroup. The ML analysis of whole cp sequences yielded a single phylogenetic tree (-Ln likelihood = 220179.81568), with the assumed nucleotide frequencies A = 0.31140, C = 0.18900, G = 0.18980, T = 0.30980, α shape parameter = 0.7954, and the proportion of invariable sites = 0.778. Numbers above ML tree branches are bootstrap support (BS). The phylogenetic tree showed that *Hordeum* was firstly separated from the other *Triticeae* diploid species. *P. libanotica* had a relatively far relationship with *Triticum*. *S. cereale* was aparted from *Triticum*/*Aegilops* complex clade gradually with high statistical support. Based on the whole cp genome phylogenetic analysis, *Ae. speltoides* var. *ligustica* (S) separated first, followed by *Ae. tauschii* (D), *T. monococcum* (A^m), and *T. urartu* (A) with 100 % BS in *Triticum*/*Aegilops* complex (Fig. 4*A*).

The ML analysis of coding region sequences resulted in a single phylogenetic tree (-Ln likelihood = 93615.16235), with the assumed nucleotide frequencies A = 0.29990, C = 0.18110, G = 0.20810, T = 0.31090, α shape parameter = 0.8795, and the proportion of invariable sites = 0.7704. *Hordeum* indicated more far relationship with other diploid species and was followed by *P. libanotica*. *S. cereale* formed a group with *Triticum*/*Aegilops* complex. But the genetic differentiation of *Triticum*/*Aegilops* complex was more complex. In *Triticum/Aegilpos* complex sub-clade, *T. urartu* (A), *Ae. tauschii* (D), and *Ae. speltoides* var. *ligustic* (S) clustered together with low support rate and deviated from *T. monococcum* (A^m) based on the protein coding regions genetic analysis (Fig. 4*B*).

Phylogenetic analyses based on cp protein coding sequences yielded no well-supported contradicting topologies. In contrast, complete cp sequences resulted in quite high node support values (BS = 100 %) and identical topologies.

Discussion

Cp genomic features of *P. libanotica*: Previous studies have shown an association between repeats and genome rearrangement breakpoints (Lee *et al*. 2007, Guisinger *et al*. 2011). The cpSSRs and indels were mainly

distributed in the LSC region of *P. libanotica*. This observation could be explained by the large genome size, low GC content, and higher point mutation rate in comparison to the IR region. LSC occupies 59.72 % of the complete cp genome of *P. libanotica*. The GC content of single copy regions (36.39 and 32.23 % in LSC and SSC regions, respectively) were less than in IR regions (44.02 %). Li *et al*. (2013b) observed that IR regions of cp genome were more conserved in land plants compared to the single copy regions. Perry and Wolfe (2002) showed that the nucleotide substitution rate is 2.3 times higher in the single-copy regions relative to the inverted repeats. The cpSSRs and indels mainly distributed in the non-coding regions of the cp genome, indicating an uneven distribution of SSRs and indels within the cp genomes. The distribution preference of cpSSRs and indels in *P. libanotica* is in accordance with previous studies in cultivated olives and *Salvia miltiorrhiza* (Mariotti *et al*. 2010, Qian *et al*. 2013).

It is worth pointing out that a 737 bp deletion was located in the *trnI-trnL* intergenic spacer of IRa and IRb region in *P. libanotica*. Interspecies differentiation, transposable elements (TE), or gene transfer might explain this deletion. We preferred the possibility of transfer of the 737 bp sequence from the cp into its nuclear genome. The hypothesis of interspecific differentiation was rather unlike because the 737 bp absence existed in all cp genomes of *Triticeae* species. The specific deletion even showed high similarity (identify = 98 %) with *Pooideae* species. Consequently, the 737 bp deletion might be a relatively conserved sequence existing in all of *Pooideae*. The TEs were out of the question because of the 737 bp failed to match with any TE in the *Triticeae* repeat (http://wheat.pw.usda.gov/ITMI/repeats/). Following evidences support our hypothesis that the loss of the 737 bp fragment was caused by gene transfer. Primarily, numerous researches demonstrated that the nuclear genomes of higher plants harbour a significant amount of promiscuous DNA of plastid origin (Ayliffe *et al*. 1998, Timmis *et al*. 2004). Moreover, high sequence similarity exists in the horizontal transfer from different host species (Schaack *et al*. 2010). *BLAST* analysis was carried out and we found that the 737 bp deletion was exactly similar (identities = 99 %) to a genomic scaffold of chromosome 3B of *T. aestivum* and it possess high similarity (identify = 98 %) with cp sequences in *Pooideae* species.

Variation and phylogenetic relationship within diploid *Triticeae* species: Generally, the organization of the cp genome is relatively conserved within *Triticeae*, whereas length variation is caused by the presence of some hot-spot regions and sequences losses. The whole cp genome size of eight *Triticeae* species varied from 135 026 bp (*P. libanotica*) to 136 886 bp (*T. monococcum*). On one hand, the genome size varies along with the changes of the AT-rich content in hot-spot regions (Kim and Lee 2004). *TrnL-F* sequence, a part of an AT-rich region in *rps4-trnL-trnL*, has been extensively applied to evolutionary investigation in *Triticeae* owing to its abundant informative-sites (Zhang *et al*. 2009, Dong *et al*. 2015). Various *rpl23* genes in *rbcL-psaI* hot-spot region are presumed to be translocated by illegitimate recombination mediated by these short repeat sequences in diploid *Triticeae* species (Ogihara *et al*. 1992). It should be emphasized that the region around the sites where rearrangements took place, such as deletion/insertions with multiple nucleotides, translocation, and truncation, are highly AT-rich.

On the other hand, numerous indels in intergenic spacer regions and length variations in protein coding regions result in the size differences in diploid *Triticeae* species. The *rpl2* (in IR regions), *petB*, *petD*, *rpl16*, and *rps12* genes retained their intron in *P. libanotica*, while some of them were absent in the *Triticum*/*Aegilops* complex. These intron losses are not unique in diploid *Triticeae* species. It have been demonstrated that the *rpoC1* intron is absent in some *Poaceae* lineages (Katayama and Ogihara 1993). The *rpl16* and *rpl2* lost their introns in *Geraniales* (Weng *et al*. 2014). In addition, protein coding sequences can be responsible for those gene length changes. Most coding regions length variations occurred at the beginning and the end of the sequences. *InfA*, ranging from 342 to 294 bp in length in *Triticeae* species, possessed the maximum gene length in *P. libanotica*. It is absent in the *Arabidopsis*, *Lotus*, and *Medicago* cp genomes and present as a truncated pseudogene in ginseng cp genome (Kim and Lee 2004).

The phylogenetic tree inferred from total cp genomes and protein coding regions showed a similar topology while incongruence still existed in the *Triticum*/*Aegilops* complex. The reason may be inferred as the whole cp genome possessed more variation sites than coding regions. Parks *et al*. (2009) illustrated that the noncoding sequences should be especially considered when inferring phylogenies at lower taxonomic levels, resulting from numerous variation sites. Phylogenetic relationship inferred from entire cp sequences in this study was consistent with two single-copy nuclear genes (*Acc*1 and *Pgk*1) (Fan *et al*. 2013) as well as the cp genomes from wheat, barley, rye, and their relatives (Middleton *et al*. 2014).

Evolution of cp genes in *Triticeae*: Erixon and Oxelman (2008) indicated that positive selection of the *clpP* gene in various plant lineages correlates with repeated duplication of the *clpP* gene and surrounding regions, repetitive amino acid sequences, and increasing synonymous substitution rates. The *clpP* gene in *Triticeae* contains no repeat but display numerous indels on surrounding regions, which may result in adaptive evolution. However, c*lpP* and *rpl16* show a purified pressure in *Asteraceae* family and Chinese *Apiaceae*

subfamily, respectively (Curci *et al*. 2015). Although a majority of those genes are conserved and essential for cell survival and plant development in some taxa, some genes were disappeared during long evolution. The *cplP* genes are totally absent in algae *Euglena*, *Odontella,* and *Porphyra* (Wakasugi *et al*. 1997). Millen *et al*. (2001) revealed that *rpl22*, r*ps16*, and *infA* genes partly disappered during angiosperm evolution or they were completely lost in *Legume* species.

In accordance with previous investigations, most of genes located in the LSC or SSC regions displayed a more rapid evolution rate compared to sequences in the IR regions (Perry and Wolfe 2002, Qian *et al*. 2013, Li *et al*. 2013b, Weng *et al*. 2014). Some genes in the IR region, however, have a faster divergence than those genes from the LSC or SSC regions. The *rpl2* gene is the fastest evolving gene in diploid *Triticeae* species. It showed more variations than several genes in the LSC or SSC regions. Usually, majority of the genes in IR (*ndhB*, *rps15*, *rps19*, *rpl23*, *rps7*, and *rps12*) regions show no synonymous rate and/or nonsynonymous rate changes while *rps* and *rpl* genes in single-copy regions possess higher divergence. Therefore, the divergence levels of cp genes are influenced not only by the location of the genes on the cp genome but also by the functional constraints of the genes.

Li *et al*. (2015) recently has proposed the use of taxon-specific barcodes for species identification using dedicated cp DNA regions that have a sufficiently high mutation rate. We confirmed the informative values for *rps16*, *matK*, *rpoA*, *rbcL* and *ndhF* regions previously adopted for the phylogenetic evolution of *Triticeae* species (Mason-Gamer 2002, Sun 2007, Hodge *et al*. 2010, Dong *et al*. 2015). Moreover, *rpl2*, *rps12* (LSC), *ccsA*, *rps8*, *ndhH*, *petD, ndhK, psbM, rps3, rps18*, and *ndhA*, exhibiting K_2P values, which means faster gene divergence, are identified in this work as highly Par. Inf. regions and thus can be considered in future phylogenetic studies, especially *rpl2*. Although those protein coding genes displayed a fast evolutionary rate and high Par. Inf. characters, genes which possess positive d_N/d_S ratios were also included. As a result, a gene used for phylogeny analysis should have fast evolutionary rate, high Par. Inf. characters and negative d_N/d_S ratio. Meanwhile, those informative genes selected for the phylogeny of *Triticeae* might not adapt to other higher taxonomic level plants. For example, *clpP* gene which is influenced by positive selective pressure in *Triticeae*, but it was seemed to be a promising gene which could be utilized to phylogeny analyses in the globe artichoke cp genome with a fairly high Par. Inf. ratio (Curci *et al*. 2015).

Utility of complete cp genome and cpSSRs in *Triticeae* phylogenetic analysis: The CpDNA can be used as an effective tool of genealogic studies in *Triticeae* species (Liu *et al*. 2006, Zhang *et al*. 2009, Hodge *et al*. 2010, Dong *et al*. 2015). However, a robust phylogenetic analysis requires a substantial sequence length. Next-generation sequencing technologies, which have been developed in recent years, enable the determination of the complete nucleotide sequence of both cp and mitochondrial DNAs of many higher plants, including wheat and its relatives. Entire cp genomes from the *Triticum*/*Aegilops* complex species demonstrated that the cp genomes of polyploid wheats were inherited from *Speltoides*. The *Sitopsis* lineage was not the source of the B or G genomes (Gornicki *et al*. 2014). The whole cp sequences of *Ae. tauschii* and *Ae. cylindrica* were used for shedding light on the D genome evolution (Gogniashvili *et al*. 2016). Additionally, whole cp genomes provided a more detailed insight into the evolution of the *Triticum* tribe and its relative diploid species than nuclear genes due to cp divergence date might be closer to the time of species separation (Gornicki *et al*. 2014, Middleton *et al*. 2014).

The CpSSRs often display high variation within the same species (Echt *et al*. 1998) and are considered as valuable markers for population genetics and phylogenetic analyses (Timme *et al*. 2007, Melotto-Passarin *et al*. 2011). In the *P. libanotica* cp genome, *rpoC2* displayed the highest number of repeats with six SSRs, *ndhF*, and *ycf3* with five SSRs independently. The *rpoC2* and *ndhF* coding genes have been successfully used in *Poaceae* genetic relationships (Igloi *et al*. 1990, Redinbaugh *et al*. 2000). It is possible that the highly variable *ycf3* coding regions are also suitable for phylogenetic analyses or DNA barcoding at low taxonomic levels.

References

Ayliffe, M.A., Scott, N.S., Timmis, J.N.: Analysis of plastid DNA-like sequences within the nuclear genomes of higher plants. - Mol. Biol. Evol. **15**: 738-745, 1998.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G.,Alekseyev, M.A., Pevzner, P.A.: *SPAdes*: a new genome assembly algorithm and its applications to single-cell sequencing. - J. comp. Biol. **19**: 455-477, 2012.

Curci, P.L., De Paola, D., Danzi, D., Vendramin, G.G., Sonnante, G.: Complete chloroplast genome of the multifunctional crop globe artichoke and comparison with other *Asteraceae*. - PLoS ONE **10**: e0120589, 2015.

Dong, Z.Z., Fan, X., Sha, L.N., Wang, Y., Zeng, J., Kang, H.Y., Zhang, H.Q., Wang, X.L., Zhang, L., Ding, C.B., Yang, R.W., Zhou, Y.H.: Phylogeny and differentiation of the St genome in *Elymus* L. *sensu lato* (*Triticeae*; *Poaceae*) based on one nuclear DNA and two chloroplast genes. - BMC

Plant Biol. **15**: 179-193, 2015.

Echt, C.S., DeVerno, L., Anzidei, M., Vendramin, G.: Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait. - Mol. Ecol. **7**: 307-316, 1998.

Erixon, P., Oxelman, B.: Whole-gene positive selection, elevated synonymous substitution rates, duplication, and indel evolution of the chloroplast *clpP1* gene. - PLoS ONE **3**: e1386, 2008.

Fan, X., Sha, L.N., Wang, X.L., Zhang, H.Q., Kang, H.Y., Wang, Y., Zhou, Y.H.: Phylogeny and molecular evolution of the *Acc1* gene within the StH genome species in *Triticeae* (*Poaceae*). - Gene **529**: 57-64, 2013.

Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., Dubchak, I.: VISTA: computational tools for comparative genomics. - Nucl. Acids Res. **32** (Suppl.): W273-W279, 2004.

Gogniashvili, M., Jinjikhadze, T., Maisaia, I., Akhalkatsi, M., Kotorashvili, A., Kotaria, N., Beridze, T., Dudnikov, A.J.: Complete chloroplast genomes of *Aegilops tauschii* Coss. and *Ae. cylindrica* Host sheds light on plasmon D evolution. - Curr. Genet. **62**: 791-798, 2016.

Gornicki, P., Zhu, H.L., Wang, J.W., Challa, G.S., Zhang, Z.Z., Gill, B.S., Li, W.L.: The chloroplast view of the evolution of polyploid wheat. - New Phytol. **204**: 704-714, 2014.

Guisinger, M.M., Kuehl, J.V., Boore, J.L., Jansen, R.K.: Extreme reconfiguration of plastid genomes in the angiosperm family *Geraniaceae*: rearrangements, repeats, and codon usage. - Mol. Biol. Evol. **28**: 583-600, 2011.

Hodge, C.D., Wang, H., Sun, G.L.: Phylogenetic analysis of the maternal genome of tetraploid StStYY *Elymus* (*Triticeae*: *Poaceae*) species and the monogenomic Triticeae based on *rps16* sequence data. - Plant Sci. **178**: 463-468, 2010.

Huotari, T., Korpelainen, H.: Complete chloroplast genome sequence of *Elodea canadensis* and comparative analyses with other monocot plastid genomes. - Gene **508**: 96-105, 2012.

Igloi, G.L., Meinke, A., Döry, I., Kössel, H.: Nucleotide sequence of the maize chloroplast *rpo B/C1/C2* operon: comparison between the derived protein primary structures from various organisms with respect to functional domains. - Mol. gen. Genet. **221**: 379-394, 1990.

Jensen, K.B.: Cytology, fertility, and origin of *Elymus abolinii* (Drob.) Tzvelev and its F1 hybrids with *Pseudoroegneria spicata*, *E. lanceolatus*, *E. dentatus* ssp. *ugamicus*, and *E. drobovii* (*Poaceae*: *Triticeae*). - Genome **32**: 468-474, 1989.

Katayama, H., Ogihara, Y.: Structural alterations of the chloroplast genome found in grasses are not common in monocots. - Curr. Genet. **23**: 160-165, 1993.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A.: Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. - Bioinformatics **28**: 1647-1649, 2012.

Kim, K.J., Lee, H.L.: 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, 2004.

Lee, H.L., Jansen, R.K., Chumley, T.W., Kim, K.J.: Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (*Oleaceae*) are due to multiple, overlapping inversions. - Mol. Biol. Evol. **24**: 1161-1180, 2007.

Li, J.L., Wang, S., Jing, Y., Wang, L., Zhou, S.L.: A modified CTAB protocol for plant DNA extraction. - Chin. Bull. Bot. **48**: 72-78, 2013a.

Li, Q, Wan, J.M.: *SSRHunter*: development of a local searching software for SSR sites. - Hereditas **27**: 808-810, 2005.

Li, R., Ma, P.F., Wen, J., Yi, T.S.: Complete sequencing of five *Araliaceae* chloroplast genomes and the phylogenetic implications. - PLoS ONE **8**: e78568, 2013b.

Li, X.W., Yang, Y., Henry, R.J., Rossetto, M., Wang, Y.T., Chen, S.L.: Plant DNA barcoding: from gene to genome. - Biol. Rev. **90**: 157-166, 2015.

Liu, Q.L., Ge, S., Tang, H.B., Zhang, X.L., Zhu, G.F., Lu, B.R.: Phylogenetic relationships in *Elymus* (*Poaceae*: *Triticeae*) based on the nuclear ribosomal internal transcribed spacer and chloroplast *trnL-F* sequences. - New Phytol. **170**: 411-420, 2006.

Lu, B.R.: The genus *Elymus* L. In: Proceedings of the 2nd International Triticeae Symposium. Pp. 219-233. Logan 1994.

Luo, R.B., Liu, B.H., Xie, Y.L., Li, Z.Y., Huang, W.H., Yuan, J.Y., He, G.Z., Chen, Y.X., Pan, Q., Liu, Y.J., Tang, J.B., Wu, G.X., Zhang, H., Shi, Y.J., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C.L., Cheung, D.W., Yiu, S.M., Peng, S.L., Zhu, X.Q., Liu, G.M., Liao, X.K., Li, Y.R., Yang, H.M., Wang, J., Lam, T.W., Wang, J.: *SOAPdenovo2*: an empirically improved memory-efficient short-read *de novo* assembler. - GigaScience **1**: 18, 2012.

Ma, P.F., Zhang, Y.X., Zeng, C.X., Guo, Z.H., Li, D.Z.: Chloroplast phylogenomic analyses resolve deep-level relationships of an intractable bamboo tribe *Arundinarieae* (*Poaceae*). - Syst. Biol. **63**: 933-950, 2014.

Mariotti, R., Cultrera, N.G., Díez, C.M., Baldoni, L., Rubini, A.: Identification of new polymorphic regions and differentiation of cultivated olives (*Olea europaea* L.) through plastome sequence comparison. - BMC Plant Biol. **10**: 211-223, 2010.

Mason-Gamer, R.J., Orme, N.L., Anderson, C.M.: Phylogenetic analysis of North American *Elymus* and the monogenomic *Triticeae* (*Poaceae*) using three chloroplast DNA data sets. - Genome **45**: 991-1002, 2002.

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

Melotto-Passarin, D., Tambarussi, E., Dressano, K., De Martin, V., Carrer, H.: Characterization of chloroplast DNA microsatellites from *Saccharum* spp and related species. - Genet. mol. Res. **10**: 2024-2033, 2011.

Middleton, C.P., Senerchia, N., Stein, N., Akhunov, E.D., Keller, B., Wicker, T., Kilian, B.: Sequencing of chloroplast genomes from wheat, barley, rye and their relatives provides a detailed insight into the evolution of the *Triticeae* tribe. - PLoS ONE **9**: e85761, 2014.

Millen, R.S., Olmstead, R.G., Adams, K.L., Palmer, J.D., Lao, N.T., Heggie, L., Kavanagh, T.A., Hibberd, J.M., Gray, J.C., Morden, C.W., Calieg, P.J., Jermiinh, L.S, Wolfe, K.H.: Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. - Plant Cell **13**: 645-658, 2001.

Neuhaus, H., Emes, M.: Nonphotosynthetic metabolism in plastids. - Annu. Rev. Plant Biol. **51**: 111-140, 2000.

Ogihara, Y., Terachi, T., Sasakuma, T.: Structural analysis of length mutations in a hot-spot region of wheat chloroplast DNAs. - Curr. Genet. **22**: 251-258, 1992.

Palmer, J.D.: Plastid chromosomes: structure and evolution. - Mol. Biol. Plastids **7**: 5-53, 1991.

Parks, M., Cronn, R., Liston, A.: Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. - BMC Biol. **7**: 84-100, 2009.

Perry, A.S., Wolfe, K.H.: Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. - J. mol. Evol. **55**: 501-508, 2002.

Posada, D., Crandall, K.A.: Modeltest: testing the model of DNA substitution. - Bioinformatics **14**: 817-818, 1998.

Provan, J., Powell, W., Hollingsworth, P.M.: Chloroplast microsatellites: new tools for studies in plant ecology and evolution. - Trends Ecol. Evol. **16**: 142-147, 2001.

Qian, J., Song, J.Y., Gao, H.H., Zhu, Y.J., Xu, J., Pang, X.H., Yao, H., Sun, C., Li, X.E., Li, C.Y., Liu, J.Y., Xu, H.B., Chen, S.L.: The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. - PloS ONE **8**: e57607, 2013.

Redinbaugh, M.G., Jones, T.A., Zhang, Y.: Ubiquity of the St chloroplast genome in St-containing *Triticeae* polyploids. - Genome **43**: 846-852, 2000.

Rodríguez-Ezpeleta, N., Brinkmann, H., Burey, S.C., Roure, B., Burger, G., Löffelhardt, W., Bohnert, H.J., Philippe, H., Lang, B.F.: Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. - Curr. Biol. **15**: 1325-1330, 2005.

Rose, O., Falush, D.: A threshold size for microsatellite expansion. - Mol. Biol. Evol. **15**: 613-615, 1998.

Rozas, J., Sánchez-DelBarrio, J.C., Messeguer, X., Rozas, R.: DnaSP, DNA polymorphism analyses by the coalescent and other methods. - Bioinformatics **19**: 2496-2497, 2003.

Schaack, S., Gilbert, C., Feschotte, C.: Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. - Trends Ecol. Evol. **25**: 537-546, 2010.

Sha, L.N., Fan, X., Yang, R.W., Kang, H.Y., Ding, C.B., Zhang, L., Zheng, Y.L., Zhou, Y.H.: Phylogenetic relationships between *Hystrix* and its closely related genera (*Triticeae*; *Poaceae*) based on nuclear *Acc1*, *DMC1* and chloroplast *trnL-F* sequences. - Mol. Phylogenet. Evol. **54**: 327-335, 2010.

Sun, G.: Genetic diversity of *rbcL* gene in *Elymus trachycaulus* complex and their phylogenetic relationships to several *Triticeae* species. - Genet. Resour. Crop Evol. **54**: 1737-1746, 2007.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S.: *MEGA6*: molecular evolutionary genetics analysis *version 6.0*. - Mol. Biol. Evol. **30**: 2725-2729, 2013.

Timme, R.E., Kuehl, J.V., Boore, J.L., Jansen, R.K.: A comparative analysis of the *Lactuca* and *Helianthus* (*Asteraceae*) plastid genomes: identification of divergent regions and categorization of shared repeats. - Amer. J. .Bot. **94**: 302-312, 2007.

Timmis, J.N., Ayliffe, M.A., Huang, C.Y., Martin, W.: Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. - Natur. Rev. Genet. **5**: 123-135, 2004.

Torabinejad, J., Mueller, R.: Genome constitution of the Australian hexaploid grass *Elymus scabrus* (*Poaceae*: *Triticeae*). - Genome **36**: 147-151, 1993.

Wakasugi, T., Nagai, T., Kapoor, M., Sugita, M., Ito, M., Ito, S., Tsudzuki, J., Nakashima, K., Tsudzuki, T., Suzuki, Y., Hamada, A., Ohta, T., Inamura, A., Yoshinaga, K., Sugiura, M.: Complete nucleotide sequence of the chloroplast genome from the green alga *Chlorella vulgaris*: the existence of genes possibly involved in chloroplast division. - Proc. Nat. Acad. Sci. USA **94**: 5967-5972, 1997.

Weng, M.L., Ruhlman, T.A., Gibby, M., Jansen, R.K.: Phylogeny, rate variation, and genome size evolution of *Pelargonium* (*Geraniaceae*). - Mol. Phylogenet. Evol. **64**: 654-670, 2012.

Wicke, S., Schneeweiss, G.M., Müller, K.F., Quandt, D.: The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. - Plant mol. Biol. **76**: 273-297, 2011.

Wyman, S.K., Jansen, R.K., Boore, J.L.: Automatic annotation of organellar genomes with *DOGMA*. - Bioinformatics. **20**: 3252-3255, 2004.

Yang, Z.H., Bielawski, J.P.: Statistical methods for detecting molecular adaptation. - Trends Ecol. Evol. **15**: 496-503, 2000.

Yen, C., Yang, J.L.: Biosystematics of *Triticeae*. Vol. 4. - China Agricultural Press, Beijing 2011.

Zhang, C., Fan, X., Yu, H.Q., Zhang, L., Wang, X.L., Zhou, Y.H.: Different maternal genome donor to *Kengyilia* species inferred from chloroplast *trnL-F* sequences. - Biol. Plant. **53**: 759-763, 2009.