Comparative analysis of complete chloroplast genome sequence in psammophytic *Haloxylon* species (Amaranthaceae) (#12337)

First submission

Please read the **Important notes** below, and the **Review guidance** on the next page.
When ready **submit online**. The manuscript starts on page 3.

Important notes

Editor and deadline

Thiago Venancio / 18 Aug 2016

Files	5 Figure file(s) 7 Table file(s) Please visit the overview page to **download and review** the files not included in this review pdf.
Declarations	**One or more DNA sequences were reported.** **Involves a field study on animals or plants.**

Review guidelines

Please in full read before you begin

How to review

When ready **submit your review online**. The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this **pdf** and upload it as part of your review

To finish, enter your editorial recommendation (accept, revise or reject) and submit.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to **PeerJ standard**, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (See **PeerJ policy**).

EXPERIMENTAL DESIGN

- Original primary research within **Scope of the journal**.
- Research question well defined, relevant & meaningful. It is stated how research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- Data is robust, statistically sound, & controlled.
- Conclusion well stated, linked to original research question & limited to supporting results.
- Speculation is welcome, but should be identified as such.

The above is the editorial criteria summary. To view in full visit **https://peerj.com/about/editorial-criteria/**

Comparative analysis of complete chloroplast genome sequence in psammophytic *Haloxylon* species (Amaranthaceae)

Wenpan Dong [1, 2] , **Chao Xu** [1] , **Delu Li** [3] , **Xiaobai Jin** [4] , **Qi Lu** [5] , **Zhili Suo** Corresp. [1]

[1] State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing, China
[2] Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China
[3] Gansu Desert Control Research Institute, Lanzhou, Gansu Provice, China
[4] Beijing Botanical Garden, Institute of Botany, Chinese Academy of Sciences, Beijing, China
[5] Institute of Desertification Studies, Chinese Academy of Forestry, Beijing, Beijing, China

Corresponding Author: Zhili Suo
Email address: zlsuo@ibcas.ac.cn

Haloxylon plants are members of the Amaranthaceae (formerly of the Chenopodiaceae). They are shrubs or small trees, known as the King of psammophytes and play an important role in environmental protection, such as sand fixation and wind control in the deserts. Similar to the chloroplast (cp) genome structure in other higher plants, the *Haloxylon* cp genome is a typical quadripartite, double-stranded circular DNA molecule of 151, 570 bp in length (*H. ammodendron*, HA) and 151, 586 bp in length (*H. persicum*, HP), and include a pair of inverted repeats (IR) [24,171 bp (HA)/24,177 bp (HP) each] that separate the genome into a large single copy region (LSC) of 84,214 bp (HA)/84,217 bp (HP) and a small single copy region (SSC) of 19,014 bp (HA)/19,015 bp (HP). There are totally 112 genes in the *Haloxylon* cp genomes, including 78 coding genes, 30 tRNA genes, and 4 ribosomal RNA genes. Fifty-nine different SSR loci were detected, including 44 mono-nucleotide repeat, 3 di-nucleotide repeat, 1 tri-nucleotide repeat, and 11 tetra-nucleotide repeat. Comparative analysis indicated that there are only 67 mutations including 44 substitutions, 23 indels, and two micro-inversions. The two inversionts with the fragment sizes of 14 bp and 3 bp occurred in the intergenic region of *petA-psbJ* and in the intron of *rpl16*, respectively. The two inverted sequences form hairpin structure, with repeat sequences of 27 and 19 bp at the two ends, respectively. The ratio of transitions (Ts) and transversions (Tv) was 0.76 in the *Haloxylon* cp genome. These results are valuable for studies on genetic diversity and will enhance our understanding about the phylogenetic evolution of the Amaranthaceae.

Comparative analysis of complete chloroplast genome sequence in psammophytic *Haloxylon* species (Amaranthaceae)

Wenpan Dong [1, 2], Chao Xu[1, 3], Delu Li [4], Xiaobai Jin [5], Qi Lu [6, *], Zhili Suo [1, *]

[1] State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, 20 Nanxincun, Haidian District, Beijing 100093, China
E-mail: wpdong@ibcas.ac.cn (W.P.D.); xuchao@ibcas.ac.cn (X.C.)

[2] Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, No.5 Yiheyuan Road Haidian District, Beijing 100871, China

[3] University of Chinese Academy of Sciences, 19 A Yuquan Road, Shijingshan District, Beijing 100049, China

[4] Gansu Desert Control Research Institute, No. 390 Beibinhe West Road, Anning District, Lanzhou, Gansu 730070, China
E-mail: lidlu2008@163.com

[5] Beijing Botanical Garden, Institute of Botany, Chinese Academy of Sciences, 20 Nanxincun, Haidian District, Beijing 100093, China.
E-mail: jinxiaobai@ibcas.ac.cn

[6] Institute of Desertification Studies, Chinese Academy of Forestry, No. 10, Huai-shu-ju Street, Haidian District, Beijing 100091, China.
E-mail: Luqi@caf.ac.cn

* Correspondence: zlsuo@ibcas.ac.cn; Tel.: +86-10-5163-6713

ABSTRACT

Haloxylon plants are members of the Amaranthaceae (formerly of the Chenopodiaceae). They are shrubs or small trees, known as the King of psammophytes and play an important role in environmental protection, such as wind control and sand fixation in the deserts. Similar to the chloroplast (cp) genome structure in other higher plants, the *Haloxylon* cp genome is a typical quadripartite, double-stranded circular DNA molecule of 151, 570 bp in length (*H. ammodendron*, HA) and 151, 586 bp in length (*H. persicum*, HP), and include a pair of inverted repeats (IR) [24,171 bp (HA)/24,177 bp (HP) each] that separate the genome into a large single copy region (LSC) of 84,214 bp (HA)/84,217 bp (HP) and a small single copy region (SSC) of 19,014 bp (HA)/19,015 bp (HP). There are totally 112 genes in the *Haloxylon* cp genomes, including 78 coding genes, 30 tRNA genes, and 4 ribosomal RNA genes. Fifty-nine different SSR loci were detected, including 44 mono-nucleotide repeat, 3 di-nucleotide repeat, 1 tri-nucleotide repeat, and 11 tetra-nucleotide repeat. Comparative analysis indicated that there are only 67 mutations including 44 substitutions, 23 indels, and two micro-inversions. The two inversionts with the fragment sizes of 14 bp and 3 bp occurred in the intergenic region of *petA-psbJ* and in the intron of *rpl16*, respectively. The two inverted sequences form hairpin structure, with repeat sequences of 27 and 19 bp at the two ends, respectively. The ratio of transitions (Ts) and transversions (Tv) was 0.76 in the *Haloxylon* cp genome. These results are valuable for studies on genetic diversity and will enhance our understanding about the phylogenetic evolution of the Amaranthaceae.

Keywords *Haloxylon*, Psammophytes, Chloroplast genome, Structure, Evolution, Amaranthaceae

INTRODUCTION

Haloxylon of the Amaranthaceae (formerly of the Chenopodiaceae) are psammophytic small trees or shrubs. Two species of the genus, distributed in the deserts of northwest China, known as the King of psammophytes and play an important role in environmental protection, such as wind control and sand fixation in the deserts

(*Zhu et al.*, *2004*; *Jia & Lu*, *2004*). However, populations of *Haloxylon* plants have been threatened in China in past decades, due to the decrease of underground water, overgrazing and over exploitation of agriculture. Genetic research on *Haloxylon* germplasm resources has been paid great attention (*Song & Jia*, *2000*; *Sheng et al.*, *2004, 2005*; *Zhang et al.*, *2006a, 2006b*). The *Haloxylon* plants only possess fine green assimilating shoots, without leaves. It is difficult to conduct evaluation on phenotypic diversity of them. The progress has been extremely slow in detecting genetic diversity of the *Haloxylon* germplasm resources, due to lack of morphological and molecular markers (*Sheng et al.*, *2004, 2005*; *Zhang et al.*, *2006a, 2006b*; *Wang et al.*, *2009*; *Suo et al.*, *2012a*).

Within the land plants, the eudicot clade comprises approximately 75% of all the flowering plant species, including major subclades: rosids, asterids, Saxifragales, Santalales, and Caryophyllales (*APG III*, *2009*). *Haloxylon* species are positioned phylogenetically in the Amaranthaceae Juss of the Caryophyllales Perleb among core eudicots (*APG III*, *2009*; *Pyankov et al.*, *2001*; *Akhani et al.*, *2007*). Relatively speaking, *Haloxylon* plants are advanced rather than primitive in evolution of flowering plants. Genus *Haloxylon* has about 11 species, with a distribution from the Mediterranean through Central Asia to China (*Zhu et al.*, *2004*). Among the plant cp genomes, *Haloxylon* cp genome is a representative type of precious psammophytic trees and shrubs that adapt to harsh environmental conditions, such as drought, desert, high temperature and sand storms.

In plants, chloroplasts (cp) with to 10, 000 copies per cell are key organelles for photosynthesis and other biochemical pathways such as biosynthesis of starch, fatty acids, pigments and amino acids (*Dong et al.*, *2013b*; *Raman and Park*, *2016*). Since the first cp genome of *Nicotiana tabacum* was sequenced in 1986, around 800 complete cp genome sequences have been made available in the National Center for Biotechnology Information (NCBI) organelle genome database. These data are valuable sources of genetic markers for phylogenetic analyses, genetic diversity evaluation, and plant molecular identification (*Dong et al.*, *2012, 2013a, 2013b, 2014*; *Ni et al.*, *2016*; *Suo et al.*, *2012b*).

There are two published complete cp genome sequences (*Spinacia oleracea* and *Beta vulgaris* subsp. *vulgaris*) in the Amaranthaceae (*Li et al.*, *2014*; *Schmitz-Linneweber et al.*, *2001*). Deeper study on cp genome is of significance for better understanding of the adaptability of *Haloxylon* plants to the severe desert environmental conditions and their genomic evolution in the Amaranthaceae. In this paper, we report patterns

of nucleotide substitutions, microstructural mutation, and simple sequence repeats (SSRs) in the cp genomes of two *Haloxylon* plants, *H. ammodendron* and *H. persicum*. Additionally, we performed genomic comparative analyses on four representative Amaranthaceae species.

MATERIALS & METHODS

Sampling and DNA extraction

Fresh young shoots of *Haloxylon ammodendron* (HA) and *H. persicum* (HP) were collected in May, 2011 from Minqin Eremophytes Botanical Garden, Gansu Province, China. The plants of HA and HP were introduced originally from Turpan Desert Botanical Garden of Chinese Academy of Sciences, Xinjiang Uygur Autonomous Region. The shoots of each accession were dried immediately using silica gel for future DNA extraction. Total gDNAs were extracted using the Plant Genomic DNA Kit (DP305) from Tiangen Biotech (Beijing) Co., Ltd., China.

Chloroplast genome sequencing

The cp genomes of HA and HP were sequenced using the short-range PCR method reported by Dong *et al*. (2012, 2013). The PCR protocol were as follows: preheating at 94°C for 4 min, 34 cycles at 94°C for 30 s, annealing at 55°C for 30 s and elongation at 72°C for 1.5 min, followed by a final extension at 72°C for 10 min. PCR amplification was performed in an Applied Biosystems VeritiTM 96-Well Thermal Cycler (Model#: 9902, made in Singapore). The amplicons were sent to Shanghai Majorbio Bio-Pharm Technology Co., Ltd (Beijing) for Sanger sequencing from both forward and reverse directions using a 3730xl DNA analyzer (Applied Biosystems, Foster City, CA, USA). The DNA regions with poly structures or difficulties to amplify or sequence were further sequenced using newly designed primer pairs for confirming reliable and high quality sequencing results.

Chloroplast genome assembling and annotation

The DNA sequences were manually confirmed using Sequencher (v4.6) software and then cp genome assembling was conducted. Chloroplast genome annotation was accomplished using the Dual Organellar Genome Annotator (DOGMA) (*Wyman et al., 2004*). BLASTX and BLASTN searches were utilized to

accurately annotate the genes encoding proteins and the locations of the transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs). Gene annotation information of other closely related plant species was also used for confirmation when the boundaries of the introns or exons could not be precisely determined due to the limited power of BLAST in cp genome annotation, e.g., some short exons of 6-9nt in length, such as in cases of *rps16*, *petB* and *petD*. Boundaries of promoters and stop codons of the protein-encoding genes, and the boundaries of introns and exons have been identified accurately. The cp genome map was drawn using Genome Vx software (*Conant & Wolfe*, *2008*) (http://wolfe.ucd.ie/GenomeVx/). The cp genome sequences have been deposited to GenBank (GenBank accession numbers KF534478 for *Haloxylon ammodendron* and KF534479 for *H. persicum*) (https://www.ncbi.nlm.nih.gov/nuccore/?term=Haloxylon+chloroplast+genome).

Repeat structure analysis

Gramene software (http://www.gramene.org/db/markers/ssrtool)(*Benson*, *1999*) was employed to search simple sequence repeat loci in the cp genome sequence with the threshold value of repeat number as ≥10 for mononucleotide repeats, ≥5 for dinucleotide repeats, ≥4 for trinucleotide repeats, and ≥3 for tetranucleotide repeats, pentanucleotide repeats, or hexanucleotide repeats.

Gene content analysis and comparative genomics

The mVISTA program was employed in Shuffle-LAGAN mode (*Frazer et al.*, *2004*) to compare the complete cp genome between *Haloxylon ammodendron* and *H. persicum*. The nucleotide sequences of all protein coding genes were aligned using MUSCLE software (*Thompson et al.*, *1997*) and were adjusted manually using Se-Al 2.0 (*Rambaut*, *1996*). Variable sites in the cp genome were calculated using DnaSP (DNA Sequences Polymorphism version 5.10.01) software (*Librado & Rozas*, *2009*). The genetic distance (p-distance) was computed using MEGA 6.0 software (*Tamura et al.*, *2011*). Based on the aligned sequence matrix, the micro-structure events were checked manually and were further divided into three categories, i.e., microsatellite-related indels, none-microsatellite-related indels and inverted sequences. Taking the cp genome sequence of *H. ammodendron* as standard reference, the size, location and evolutionary direction of the micro-structure events were counted up. The proposed secondary structure of the inverted regions in the cp genomes of HA and HP were analyzed using mfold software (*Zuker*, *2003*). The complete cp genome sequences of *Spinacia oleracea*

(GenBank accession number AJ400848.1, *Spinacia* L.) (*Schmitz-Linneweber et al.*, *2001*) and *Beta vulgaris* subsp. *vulgaris* (GenBank accession number KJ081864.1, *Beta vulgaris* subsp. *vulgaris*) (*Li et al.*, *2014*), two closely related species in the Amaranthaceae were downloaded from GenBank databases (www.ncbi.nlm.nih.gov). They are used for comparison of complete cp genomes with *H. ammodendron* and *H. persicum* within the Amaranthaceae.

RESULTS & DISCUSSION

Genome features

Similar to typical cp genome structure in other higher plants, the *Haloxylon* cp genome is a double-stranded circular DNA molecule of 151, 570 bp in length (HA) and 151, 586 bp in length (HP), and include a large single copy region (LSC) of 84,214 bp (HA)/84,217 bp (HP) and a small single copy region (SSC) of 19,014 bp (HA)/19,015 bp (HP) that were separated by a pair of inverted repeats (IR) [24,171 bp (HA)/24,177 bp (HP) each]. GC content in IR region is 43.0% in HA and 42.7% in HP. GC content in LSC and SSC regions is 34.4% (LSC) and 29.7% (SSC) in HA and is 34.5% (LSC) and 29.7% (SSC) in HP (Fig. 1, Table 1, Table S1).

Among the four Amaranthaceae species of the three genera, the longest cp genomes (151,570 bp, *H. ammodendron* and 151,586 bp, *H. persicum*) are 1935 bp to 1951 bp larger than the shortest one (149,635 bp, *Beta vulgaris* subsp. *vulgaris*) (*Li et al.*, *2014*), the size of *Spinacia oleracea* cp genome (150,725 bp) (*Schmitz-Linneweber et al.*, *2001*) is intermediate (Table 1). The cp genome of *H. persicum* is only 16 bp longer than that of *H. ammodendron* with minor difference between them.

There are a total of 112 genes in the *Haloxylon* cp genome, including 78 coding genes with 18 duplicated genes in the IRs region, 30 tRNA genes, and 4 ribosomal RNA genes (16S, 23S, 5S, 4.5S) (Fig. 1, Table S1). According to their functions, these genes can be divided into three categories, 1) genes related to transcription and translation; 2) genes related to photosynthesis; 3) protein-coding genes related to biosynthesis of amino acids, fatty acids, etc, and some functionally unknown genes (Table S1).

The number of genes in the cp genomes of HA, HP and *S. oleracea* in Amaranthaceae is the same (78 coding genes), while it is 79 in *B. vulgaris* with an additional gene of *rpl*23, which is a pseudogene in other species. There are seventeen genes harboring introns in the cp genomes of the four species in the Amaranthaceae (one class I intron, *trn*L^{UUA}, and 16 class II introns), and two of these genes, *ycf*3 and *clp*P,

containing two introns (Table 2).

Several lineages of angiosperms have lost introns from the *rpl*2 gene independently (*Downie et al.*, *1991*), which could also be regarded as a characteristic feature of the core members of Caryophyllales (*Logacheva et al.*, *2008*). In the four cp genomes in Amaranthaceae, *rpl*2 gene has lost its only intron. Some authors proposed that intron losses are not always dependable markers of phylogenetic relationships (*Millen et al.*, *2001*; *Dong et al.*, *2013b*; *Raman & Park*, *2016*). It is worthy of further study in more detail on the relationship between the intron loss and phylogenetic significance by sampling more taxa.

Indels and SNPs

The Indel and SNP sites are important molecular features for plant identification, they have been proved to be valuable for development of DNA markers for plant identification and genetic analysis of population structure (*Dong et al.*, *2012*, *2013a*, *2013b*, *2014*; *Suo et al. 2012*, *2015*, *2016*).

Twenty-three indels were detected in the cp genome sequence alignment of HA and HP, including 16 indels caused by microsatellite repeat variations and 7 indels of non-microstellite-related types (Table 3). Most of the indel events occurred in non-coding regions (21/23). Most of the indels related to microsatellite repeat variations are characterized by single base mutation, six insertions of this mutation type were observed in the cp genome of HA. The size of ordinary indels of other mutation type is mostly 5 to 6 variable base sites, two insertions of them were detected in the cp genome of HA. There are 65 indels in the cp genomes between *Machilus yunnanensis* and *M. balansae* (*Song et al.*, *2015*), 156 indels in the cp genomes between *Panax ginseng* and *P. notoginseng* (*Dong et al.*, *2014*).

Forty-four SNPs in the cp genome were detected between *Haloxylon ammodendron* and *H. persicum* (Table 4), which are much less than those found between *Oryza sativa* and *O. nivara* (159 SNPs, Masood et al., 2004), between *Machilus yunnanensis* and *M. balansae* (231 SNPs, *Song et al.*, *2015*), between *Citrus sinensis* and *C. aurantiifolia* (330 SNPs, *Su et al.*, *2014*), between *Panax ginseng* and *P. notoginseng* (464 SNPs, *Dong et al.*, *2014*), and between *Solanum tuberosum* and *S. bulbocastanum* (591 SNPs, *Chung et al.*, *2006*).

The indel and SNP mutation events in the genome were not random but clustered as “hotspots” (*Shaw et al.*, *2007*; *Worberg et al.*, *2007*). Such mutational dynamics created the highly variable regions in the genome (*Suo et al.*, *2012*; *Song et al.*, *2015*).

195

196 **Repeat structure feature**

197 SSR loci are valuable for assessment of plant genetic diversity at population and intraspecific levels (*Dong et*

198 *al.*, *2014*; *Sumathi & Yasodha*, *2014*; *Choi & Park*, *2015*; *Kaur et al.*,*2015*; *Suo et al.*, *2016*).

199 Simple sequence repeats (SSRs) are also called microsatellites. Within the cp genome of HA and HP, 59

200 different SSR loci were detected. Of these, 44 loci are mono-nucleotide repeats, three are di-nucleotide repeats,

201 one is tri-nucleotide repeat, eleven are tetra-nucleotide repeats, while, pentanucleotide repeats, or higher-

202 numbered nucleotide repeats were not detected. Among the SSRs loci detected, the most frequently observed

203 repeats are A/T or AT/TA, accounting for 77.97% of the total number of SSR loci (Table 5). In the cp genomes

204 between *Machilus yunnanensis* and *M. balansae*, 36 SSR loci were identified (*Song et al.*, *2015*).

205

206 **Inversions**

207 Inversions are important events in evolution of plant cp genomes. The sequence alignment of the *Haloxylon* cp

208 genomes indicates that the fragment sizes of the two inversion events are 14 bp and 3 bp, occurring in the

209 intergenic region of *petA-psbJ* and in the intron of *rpl16*, respectively. The two inverted sequences form

210 secondary neck-ring structure, with repeat sequences of 27 bp and 19 bp at the two ends, respectively (Fig. 4).

211 Smaller inversions are less frequent in cp genome, and these inversions are generally associated with

212 hairpins (Fig. 4). Most of inversions are in spacers and introns. In most cases, the presence/absence of

213 inversions is highly homoplastic during cp genome evolution (*Kim & Lee*, *2005*; *Catalano et al.*, *2009*), even

214 at the population level (*Quandt & Stech*, *2004*).

215

216 **Pseudogenes**

217 Pseudogenes have been defined as nonfunctional sequences of genomic DNA that originally derived from

218 functional genes (*Balakirev & Ayala*, *2003*). The *rpl*23 is a pseudogene in *Haloxylon* cp genomes. The *rpl*22

219 and *rps*18 are found to be putative pseudogenes in the Paeoniaceae (*Dong et al.*, *2013b*). The *atp*B gene is a

220 pseudogene in *Aster sphathulifolius*. But, these three genes of *rpl*22, *rps*18 and *atp*B are found to be normal

221 and functional in the *Haloxylon* species. The pseudogenes are evolutionary relics of functional components in

222 the genome that provide important information regarding the history of the gene and genome evolution

(*Balakirev & Ayala*, *2003*; *Zou et al.*, *2009*; *Choi & Park*, *2015*).

Patterns of nucleotide substitutions

The difference between the cp genomes of HA and HP is minor. Forty-four variable nucleotide sites and a genetic distance of 0.00029 were found between them (Table 4). According to the locations of the variable nucleotide sites in the cp genome, 23 of them occurred in the intergenic regions (IG region), 6 of them in introns, and 15 of them in protein-encoding regions.

Probability of occurrence is different among the mutation patterns of the four kinds of nucleotides (A, G, C and T) as shown in Fig. 5. The most frequently occurred mutation is from A to C and from T to G, (12 times each); the mutation from A to T and from T to A exhibits the lowest frequency (only 1 time each). The ratio of transitions (Ts) and transversions (Tv) was 0.76 in the cp genome of *Haloxylon* species.

In the gene coding regions of the cp genomes of HA and HP, a total of 15 variable base sites were detected in eleven protein-encoding genes, i.e., there is one mutation site in each of *atp*A, *atp*I, *mat*K, *ndh*F, *ndh*I, *psb*C, *rpo*B, *rps*15 and *rps*3, and there are three mutation sites in each of *rpo*C2 and *ycf*1 (Table 6) among which 6 transitions (Ts) and 9 transversions (Tv) were detected, and 10 nonsynonymous substitutions occurred simultaneously in seven genes (Table 6).

Expansion and contraction of the border regions in *Haloxylon* cp genomes

To analyze these Amaranthaceae species at the genome-level, the sequences of all the four Amaranthaceae cp genomes were plotted using the VISTA program (*Frazer et al.*, *2004*) with the annotation of *Haloxylon ammodendron* as reference (Fig. 2). Similar to other angiosperms, the IR region is more conserved in these species than the LSC and SSC regions.

The expansion and contraction of the border regions between the two IR regions and the single copy region have contributed to genome size variations among plant lineages (*Dong et al.*, *2013b*; *Goremykin et al.*, *2003*; *Ni et al.*, *2016*). Therefore, we compared the exact IR border positions and their adjacent genes among the four Amaranthaceae cp genomes of the three genera (Fig. 3).

The IRa/LSC border is generally located upstream of the *trn*H^{GUG} gene. The distance between the IRa/LSC border and the *trn*H^{GUG} gene showed that, there is 1 bp separation at the upper stream of *trn*H^{GUG} gene in

Haloxylon cp genomes, 2 bp separations in *Beta* genus, no separation in *Spinacia* (Fig. 3). IRs region expanded 763 bp and entered to the 5' end of *ycf*1 gene in *Haloxylon* species, but expanded 1427 bp and 1492 bp respectively in *Spinacia* and *Beta*. Except for the expansion of *ycf*1 gene, IRs region extended to *rps*19 gene in all of the four Amaranthaceae cp genomes. The *rps*19 pseudogene was not observed in this study. Although there are expansions or contractions of IR regions observed among the investigated species of the Amaranthaceae, they contribute little to the overall size differences in the cp genomes. The 5'- end exon of the *rps*12 gene is located in the LSC region, and the intron and 3'- end exon of the gene are situated in the IR region in the four Amaranthaceae species.

CONCLUSIONS

The *Haloxylon* cp genomes were sequenced and characterized for the first time. The *Haloxylon* cp genome shares the same overall organization and gene contents of most of the unreorganized angiosperm cp genomes, including that of its closest *Spinacia* and *Beta* species. The location and distribution of repeat sequences were detected, and the nucleotide mutation sites of the two cp genomes were identified. The LSC/IRB/SSC/IRA boundary regions of the Amaranthaceae cp genomes were compared and no intense variations were identified within the genus *Haloxylon*. The complete cp genome sequences of the *Haloxylon* species reported here enhances the genomic information of the Amaranthaceae and contributes to the study of germplasm diversity. These data represent a valuable source of markers for future researches on *Haloxylon* population genetics.

Competing Interests

The authors declare that they have no competing interests.

Funding

The study was financially supported by China National Science and Technology Plan Project (2012BAD16B0101), "948 Program" (No. 2008-4-47), Identification and Screening of *Haloxylon* Germplasm Resources (80117B1001), State Forestry Administration "Public Welfare Research Foundation" (No. 201004010), the Scientific and Technological Support Project from the Chinese Ministry of Science and Technology (2006BAD26B0101), the National Natural Science Foundation of China (No. 30972412), and the

joint projects No. 70009C1020 and 70009C1036.

ACKNOWLEDGEMENTS

The authors thanks Prof. Borong Pan for instructions and discussion.

REFERENCES

Akhani H., Edwards G., Roalson E.H. 2007. Diversification of the old world Salsoleae *s.l.* (Chenopodiaceae): molecular phylogenetic analysis of nuclear and chloroplast data sets and a revised classification. *International J Plant Sci* 168: 931–956

APG III. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical J Linnean Society* 161: 105–121

Balakirev E.S., Ayala F.J. 2003. Psuedogenes: are they "junk" or functional DNA? *Annual Rev Genet* 37: 123–151

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acid Research* 27(2): 573–580

Catalano S.A., Saidman B.O., Vilardi J.C. 2009. Evolution of small inversions in chloroplast genome: a case study from a recurrent inversion in angiosperms. *Cladistics* 25, 93–104

Choi K.S., Park S.J. 2015. The complete chloroplast genome sequence of *Aster spathulifolius* (Asteraceae): genomic features and relationship with Asteraceae. *Gene* 572: 214–221

Chung H.J., Jung J.D., Park H.W., Kim J.H., Cha H.W., Min S.R., Jeong W.J., Liu J.R. 2006. The complete chloroplast genome sequences of *Solanum tuberosum* and comparative analysis with *Solanaceae* species identified the presence of a 241-bp deletion in cultivated potato chloroplast DNA sequence. *Plant Cell Rep* 25: 1369–1379 DOI:10.1007/s00299-006-0196-4.

Conant G.C., Wolfe K.H. 2008. GenomeVx: simple web-based creation of editable circular chromosome maps. *Bioinformatics* 24(6): 861–862.

Cosner M.E., Jansen R.K., Palmer J.D., Downie S.R. 1997. The highly rearranged chloroplast genome of *Trachelium caeruleum* (Campanulaceae): multiple inversions, inverted repeat expansion and contraction,

306 transposition, insertions/deletions, and several repeat families. *Curr Genet* 31: 419–429 DOI:
307 10.1007/s002940050225
308 **Dong W.P., Liu H., Xu C., Zuo Y.J., Chen Z.J., Zhou S.L. 2014.** A chloroplast genomic strategy for
309 designing taxon specific DNA mini-barcodes: a case study on ginsengs. *BMC Genetics* 15:138
310 **Dong W.P., Xu C., Cheng T., Lin K., Zhou S.L. 2013a.** Sequencing angiosperm plastid genomes made easy:
311 a complete set of universal primers and a case study on the phylogeny of Saxifragales. *Genome Biol Evol*
312 5(5): 989–997
313 **Dong W.P., Xu C., Cheng T., Zhou S.L. 2013b.** Complete chloroplast genome of *Sedum sarmentosum* and
314 chloroplast genome evolution in Saxifragales. *PLoS ONE* 8(10): e77965 DOI:10.1371/journal.
315 pone.0077965.
316 **Dong W.P., Liu J., Yu J., Wang L., Zhou S.L. 2012.** Highly variable chloroplast markers for evaluating
317 plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS ONE* 7(4): e35071 DOI:10.1371/
318 journal.pone.0035071.
319 **Doyle J.J., Doyle J.L., Palmer J.D. 1995.** Multiple independent losses of two genes and one intron from
320 legume chloroplast genomes. *Syst Bot* 20: 272–294 DOI:10.2307/2419496.
321 **Downie S.R., Olmstead R.G., Zurawski G., Soltis D.E., Soltis P.S., Watson J.C., Palme J.D. 1991.** Six
322 independent losses of the chloroplast DNA *rpl*2 intron in dicotyledons: molecular and phylogenetic
323 implications. *Evolution* 45: 1245–1259 DOI:10. 2307/2409731.
324 **Downie S.R., Palmer J.D. 1992.** "Use of chloroplast DNA rearrangements in reconstructing plant
325 phylogeny" in Molecular Systematics of Plants, eds P.S. Soltis, D.E. Soltis, and J.J. Doyle (NewYork,
326 NY: Chapman and Hall), pp14–35
327 **Goremykin V.V., Hirsch-Ernst K.I., Wolfl S., Hellwig F.H. 2003.** Analysis of the *Amborella trichopoda*
328 chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm. *Mol Biol Evol* 20: 1499–
329 1505 DOI: 10.1093/molbev/msg159.
330 **Frazer K.A., Pachter L., Poliakov A., Rubin E.M., Dubchak I. 2004.** VISTA: computational tools for
331 comparative genomics. *Nucleic Acids Res* 32: W273-W279 DOI:10.1093/nar/gkh053. PubMed: 15215394.
332 **Jansen R.K., Kaittanis C., Saski C., Lee S.B., Tomkins J., Alverson A.J., Daniell H. 2006.** Phylogenetic
333 analyses of *Vitis* (Vitaceae) based on complete chloroplast genome sequences: effects of taxon sampling and

phylogenetic methods on resolving relationships among rosids. *BMC Evol Biol* 6: 32 DOI:10.1186/1471-2148-6-32.

Jia Z.Q., Lu Q. 2004. *Haloxylon* Bunge. China Environmental Science Press, Beijing, China (in Chinese).

Kaur S., Panesar P.S., Bera M.B., Kaur V. 2015. Simple sequence repeat markers in genetic divergence and marker-assisted selection of rice cultivars: A review. *Crit Rev Food Sci Nutr* 55: 41–49

Kim K.J., Lee H.L. 2005. Wide spread occurrence of small inversions in the chloroplast genomes of land plants. *Molecules and Cells* 19:104–113

Li H., Cao H., Cai Y.F., Wang J.H., Qu S.P., Huang X.Q. 2014. The complete chloroplast genome sequence of sugar beet (*Beta vulgaris* ssp. *vulgaris*). *Mitochondrial DNA* 25: 209–211

Librado P., Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452

Logacheva M.D., Samigullin T.H., Dhingra A., Penin A.A. 2008. Comparative chloroplast genomics and phylogenetics of *Fagopyrum esculentum* ssp. *ancestrale* –a wild ancestor of cultivated buckwheat. *BMC Plant Biology* 8: 59 DOI:10.1186/1471-2229-8-59 PMID: 18492277.

Masood M.S., Nishikawa T., Fukuoka S., Njenga P.K., Tsudzuki T., Kadowaki K. 2004. The complete nucleotide sequence of wild rice (*Oryza nivara*) chloroplast genome: first genome wide comparative sequence analysis of wild and cultivated rice. *Gene* 340: 133–139 DOI:10.1016/j.gene. 2004.06.008.

Millen R.S., Olmstead R.G., Adams K.L., Palmer J.D., Lao N.T., Heggie L., Kavanaghd T.A., Hibberde J.M., Graye J.C., Mordenf C.W., Calieg P.J., Jermiinh L.S., Wolfe K.H. 2001. Many parallel losses of *inf*A from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell* 13: 645–658 DOI:10.1105/tpc.13.3.645.

Ni L.H., Zhao Z.L., Xua H.X., Chen S.L., Dorje G. 2016. The complete chloroplast genome of *Gentiana straminea* (Gentianaceae), an endemic species to the Sino-Himalayan subregion. *Gene* 577: 281–288

Pyankov V.I., Artyusheva E.G., Edwards G.E., Black C.C. JR, Soltis P.S. 2001. Phylogenetic analysis of tribe Salsoleae (Chenopodiaceae) based on ribosomal ITS sequences: implications for the evolution of photosynthesis types. *Am J Bot* 88(7): 1189–1198

Quandt D., Stech M. 2004. Molecular evolution and phylogenetic utility of the chloroplast *trn*T-*trn*F region in bryophytes. *Plant Biol* 6: 545–554

362 **Raman G., Park S. 2016.** The complete chloroplast genome sequence of Ampelopsis: gene organization,
363 comparative analysis, and phylogenetic relationships to other angiosperms. *Front Plant Sci* 7: 341 DOI:
364 10.3389/fpls.2016.00341.
365 **Rambaut A. 1996.** Se-Al: sequence alignment editor. version 2.0. Oxford: University of Oxford, Department
366 of Zoology
367 **Schmitz-Linneweber C., Maier R.M., Alcaraz J.P., Cottet A., Herrmann R.G., Mache R. 2001.** The
368 plastid chromosome of spinach (*Spinacia oleracea*): complete nucleotide sequence and gene organization.
369 *Plant Mol Biol* 45: 307–315 DOI:10.1023/A:1006478403810. PubMed: 11292076.
370 **Shaw J., Lickey E.B., Schilling E.E., Small R.L. 2007.** Comparison of whole chloroplast genome sequences
371 to choose non-coding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Am J Bot*
372 94: 275–288 DOI:10.3732/ajb.94.3.275.
373 **Sheng Y., Zheng W.H., Quan P.K., Ma K.P. 2004.** Population genetic structure of a dominant desert tree,
374 *Haloxylon ammodendron* (Chenopodiaceae), in the southeast Gurbantunggut desert detected by RAPD
375 and ISSR markers. *Acta Botanica Sinica* 46: 675–681 (in Chinese with English Abstract)
376 **Sheng Y., Zheng W.H., Quan P.K., Ma K.P. 2005**. Genetic variation within and among populations of a
377 dominant desert tree *Haloxylon ammodendron* (Amaranthaceae) in China. *Ann Bot London* 96:
378 245–252.
379 **Song C.S., Jia K.F. 2000.** Scientific survey of Wulate *Haloxylon ammodendron* forest nature reserve (The
380 series of nature reserve). China Forestry Publishing House, Beijing, China (in Chinese with English
381 Overview)
382 **Song Y., Dong W., Liu B., Xu C., Yao X., Gao J., Corlett R.T. 2015.** Comparative analysis of complete
383 chloroplast genome sequences of two tropical trees *Machilus yunnanensis* and *Machilus balansae* in the
384 family Lauraceae. *Front. Plant Sci.* 6: 662 DOI:10.3389/fpls.2015.00662.
385 **Su H.J., Hogenhout S.A., Al-Sadi A.M., Kuo C.H. 2014.** Complete chloroplast genome sequence of omani
386 lime (*Citrus aurantiifolia*) and comparative analysis within the Rosids. *PLoS ONE* 9: e113049 DOI:
387 10.1371/journal.pone.0113049.
388 **Sumathi M., Yasodha R. 2014.** Microsatellite resources of *Eucalyptus*: Current status and future perspectives.
389 *Botanical Studies* 55: 73

Suo Z.L., Jia Z.Q., Lu Q., Pan B.R., Jin X.B., Xu G., Peng X.Q., Sun H.B., Tao Y.H. 2012a. Distinguishing *Haloxylon persicum* and *H. ammodendron* (*Haloxylon* Bunge, Amaranthaceae) using DNA Marker. AASRI Procedia 1: 305– 310

Suo Z.L., Zhang C.H., Zheng Y.Q., He L.X., Jin X.B., Hou B.X., Li J.J. 2012b. Revealing genetic diversity of tree peonies at micro-evolution level with hyper-variable chloroplast markers and floral traits. *Plant Cell Reports* 31: 2199–2213

Suo Z.L., Chen L.N., Pei D., Jin X.B., Zhang H.J. 2015. A new nuclear DNA marker from ubiquitin ligase gene region for genetic diversity detection of walnut germplasm resources. *Biotechnology Reports* 5: 40–45

Suo Z.L., Li W.Y., Jin X.B., Zhang H.J. 2016. A new nuclear DNA marker revealing both microsatellite variations and single nucleotide polymorphic loci: a case study on classification of cultivars in *Lagerstroemia indica* L. *J Microb Biochem Technol* 8: 266–271 DOI:10.4172/1948–5948.1000296 (in press).

Tamura K., Stecher G., Peterson D., Filipski A., Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30: 2725–2729. DOI:http://dx.doi.org/10.1093/molbev/mst197.

Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F., Higgins D.G. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25: 4876–4882 DOI:10.1093/nar/25.24.4876. PubMed: 9396791.

Wang X.M., Yang D.Y., Tian Y.Z., Zhang B.W., Tu P.F., Sun Q.S., Li X.B. 2009. Inter-simple sequence repeats analysis of *Haloxylon ammodendron* from seeds carried back by "Shenzhou No.4" spaceship. J Northwest University 39: 259–263 (in Chinese with English Abstract)

Worberg A., Quandt D., Barniske A.M., Lohne C., Hilu K.W., BorschT. 2007. Phylogeny of basal eudicots: insights from non-coding and rapidly evolving DNA. *Organ Diver Evol* 7: 55–77 DOI:10.1016/ j.ode.2006.08.001.

Wyman S.K., Jansen R.K., Boore J.L. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20: 3252–3255 DOI: 10.1093/bioinformatics/bth352. PubMed: 15180927.

Zhang P., Dong Y.Z., Wei Y., Hu C.Z. 2006a. ISSR analysis of genetic diversity of *Haloxylon ammodendron* (C. A. Mey.) Bunge in Xinjiang. *Acta Botanica Boreali-Occidentalia Sinica* 26: 1337–1341 (in Chinese with English Abstract)

Zhang P., Dong Y.Z., Wei Y., Hu C.Z. 2006b. Analysis of genetic diversity of *Haloxylon persicum* (Chenopodiaceae) in Xinjiang by ISSR. *Acta Bot Yunnaica* 28: 359–362 (in Chinese with English Abstract)

Zhu G.L., Mosyakin S.L., Clemants S.E. 2004. *Haloxylon* Bunge (Chenopodiaceae). In Flora of China Editorial Committee (eds) Flora of China (English ed.). Sci. Press, Beijing/Missouri Botanic Garden Press, St. Louis. 5: 395–396

Zou S.H., Lehti-Shiu M.D., Thibaud-Nissen F., Prakash T., Buell C.R., Shiu S.H. 2009. Evolutionary and expression signatures of pseudogenes in Arabidopsis and rice. *Plant Physiol* 151: 3–15

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* 31: 3406–3415

Table and Figure Legends

Figure 1 Representative map of the two *Haloxylon* chloroplast genomes. The annotation of the genome was performed using DOGMA. The genes that are drawn outside of the circle are transcribed clockwise, while those inside are transcribed counterclockwise. Small single copy (SSC), large single copy (LSC), and inverted repeats (IRa, IRb) are indicated.

Figure 2 Identity plot comparing the chloroplast genomes of four Amaranthaceae species using *Haloxylon ammodendron* as a reference sequence. The vertical scale indicates the percentage of identity, ranging from 50% to 100%. The horizontal axis indicates the coordinates within the chloroplast genome. Genomic regions are color coded as protein-coding, rRNA, tRNA, intron, and conserved non-coding sequences (CNS). Abbreviations HP: *H. persicum*; SO: *Spinacia oleracea*; BV: *Beta vulgaris* subsp. *vulgaris*.

Figure 3 Comparison of junction positions between single copy and IR regions among four Amaranthaceae genomes.

Figure 4 The hairpin loops of inversions in the chloroplast genome of *Haloxylon*.

Figure 5 The patterns of nucleotide substitutions among the two *Haloxylon* chloroplast genomes. The patterns were divided into 6 types as indicated by the six non-strand-specific base-substitution types (i.e., numbers of G to A and C to T sites for each respective set of associated mutation types).

446 The chloroplast genome of *H. ammodendron* was used as a standard.

447 **Table 1 Summary of complete chloroplast genome of *Haloxylon.***

448 **Table 2 Genes with introns in *Haloxylon ammodendron* (*H. persicum*) and length of exons and introns.**

449 **Table 3 Forms and numbers of indel mutation events in the chloroplast genome between two *Haloxylon***

450 **species.**

451 **Table 4 The patterns of nucleotide substitutions among the two *Haloxylon* chloroplast genomes.**

452 **Table 5 Location of repeats in the *Haloxylon ammodendron* chloroplast genomes.**

453 **Table 6 Comparisons of mutational changes, number of transitions (Ts) and transversions (Tv),**

454 **synonymous (S), and nonsynonymous (N) substitutions per gene of protein coding chloroplast**

455 **genes between *Haloxylon ammodendron* and *H. persicum.***

456

457 Supplemental Information

458 **Table S1 Genes contained in the chloroplast genome of *Haloxylon ammodendron* and *H. persicum.***

Figure 1(on next page)

Figure 1 Representative map of the two *Haloxylon* chloroplast genomes.

The annotation of the genome was performed using DOGMA. The genes that are drawn outside of the circle are transcribed clockwise, while those inside are transcribed counterclockwise. Small single copy (SSC), large single copy (LSC), and inverted repeats (IRa, IRb) are indicated.

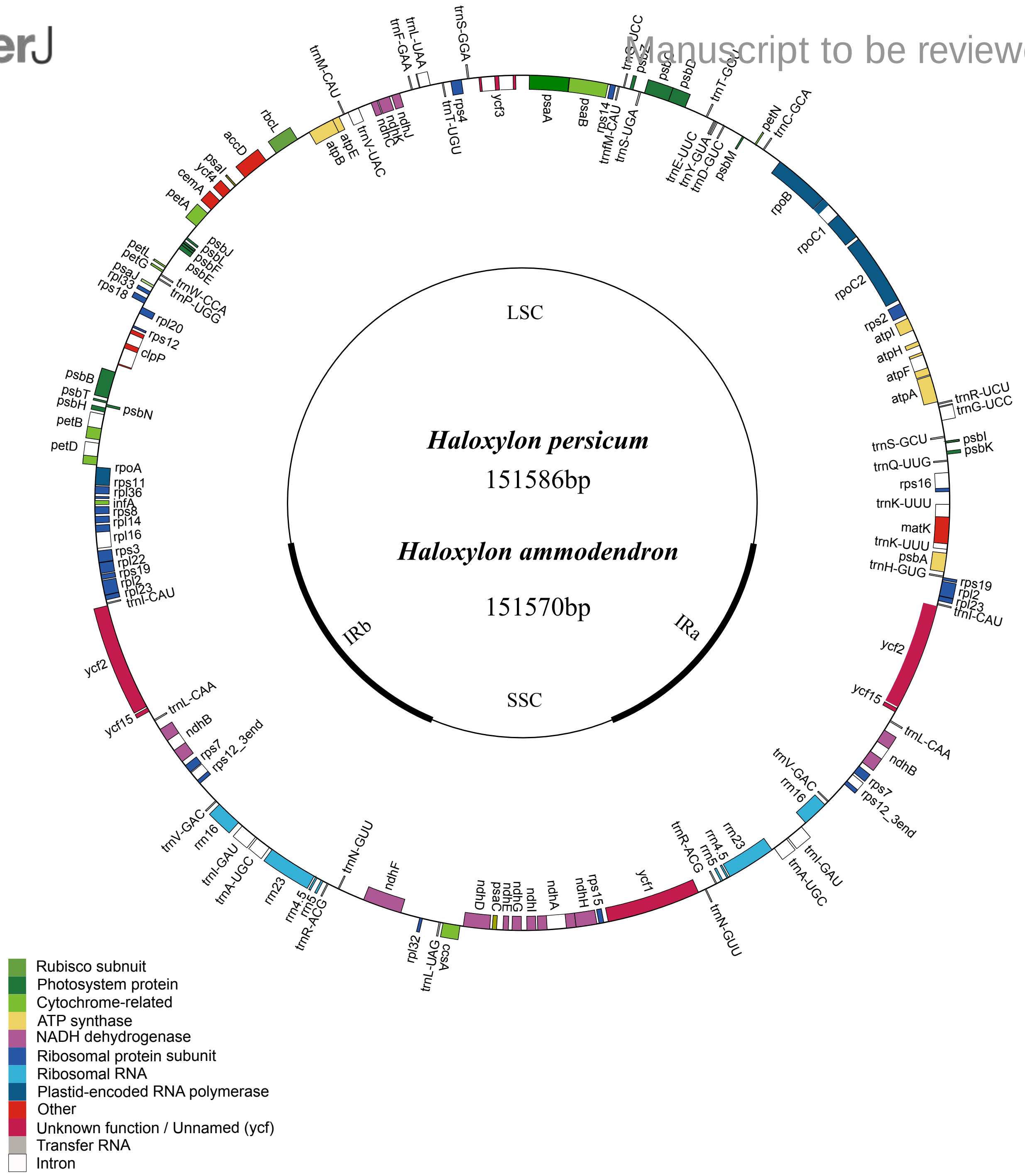

LSC
Haloxylon persicum
151586bp
Haloxylon ammodendron
151570bp
IRb
IRa
SSC
trnS-GGA
trnL-UAA
trnF-GAA
trnM-CAU
rbcL
accD
psaI
ycf4
cemA
petA
psbJ
psbL
psbF
psbE
petL
petG
psaJ
rpl33
rps18
trnW-CCA
trnP-UGG
rpl20
rps12
clpP
psbB
psbT
psbH
psbN
petB
petD
rpoA
rps11
rpl36
infA
rps8
rpl14
rpl16
rps3
rpl22
rps19
rpl2
rpl23
trnI-CAU
ycf2
ycf15
trnL-CAA
ndhB
rps7
rps12_3end
trnV-GAC
rrn16
trnI-GAU
trnA-UGC
rrn23
rrn4.5
rrn5
trnR-ACG
trnN-GUU
ndhF
rpl32
trnL-UAG
ccsA
ndhD
psaC
ndhE
ndhG
ndhI
ndhA
ndhH
rps15
ycf1
atpE
atpB
trnV-UAC
ndhC
ndhK
ndhJ
rps4
trnT-UGU
ycf3
psaA
psaB
rps14
trnfM-CAU
trnS-UGA
trnG-UCC
psbZ
psbC
psbD
trnT-GGU
trnE-UUC
trnY-GUA
trnD-GUC
psbM
petN
trnC-GCA
rpoB
rpoC1
rpoC2
rps2
atpI
atpH
atpF
atpA
trnR-UCU
trnG-UCC
trnS-GCU
psbI
psbK
trnQ-UUG
rps16
trnK-UUU
matK
psbA
trnH-GUG
Rubisco subunit
Photosystem protein
Cytochrome-related
ATP synthase
NADH dehydrogenase
Ribosomal protein subunit
Ribosomal RNA
Plastid-encoded RNA polymerase
Other
Unknown function / Unnamed (ycf)
Transfer RNA
Intron

Figure 2(on next page)

Figure 2 Identity plot comparing the chloroplast genomes of four Amaranthaceae species using *Haloxylon ammodendron* as a reference sequence.

The vertical scale indicates the percentage of identity, ranging from 50% to 100%. The horizontal axis indicates the coordinates within the chloroplast genome. Genomic regions are color coded as protein-coding, rRNA, tRNA, intron, and conserved non-coding sequences (CNS). Abbreviations HP: *H. persicum*; SO: *Spinacia oleracea*; BV: *Beta vulgaris* subsp. *vulgaris*.

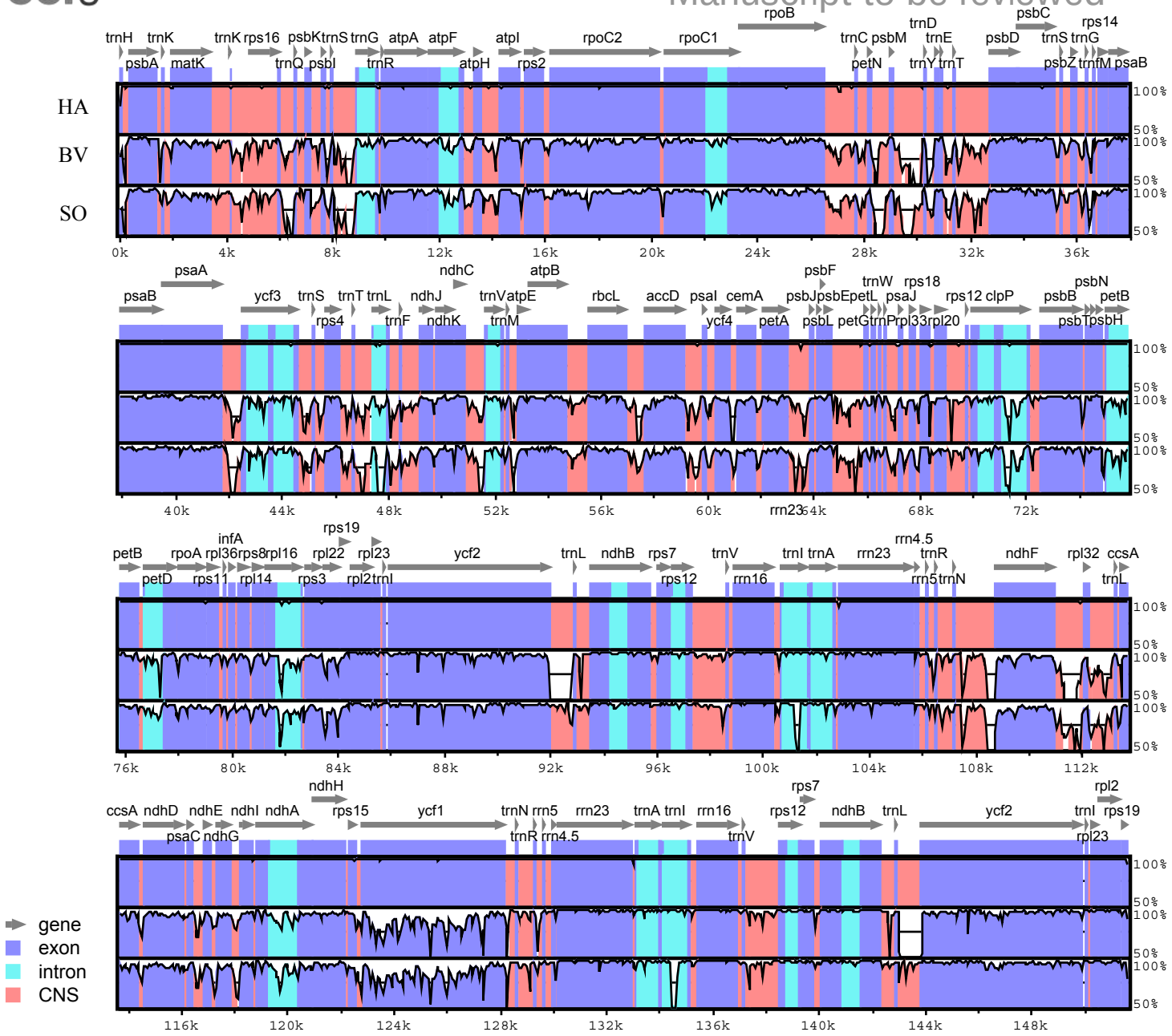

HA
BV
SO
gene
exon
intron
CNS

Figure 3(on next page)

Figure 3 Comparison of junction positions between single copy and IR regions among four Amaranthaceae genomes.

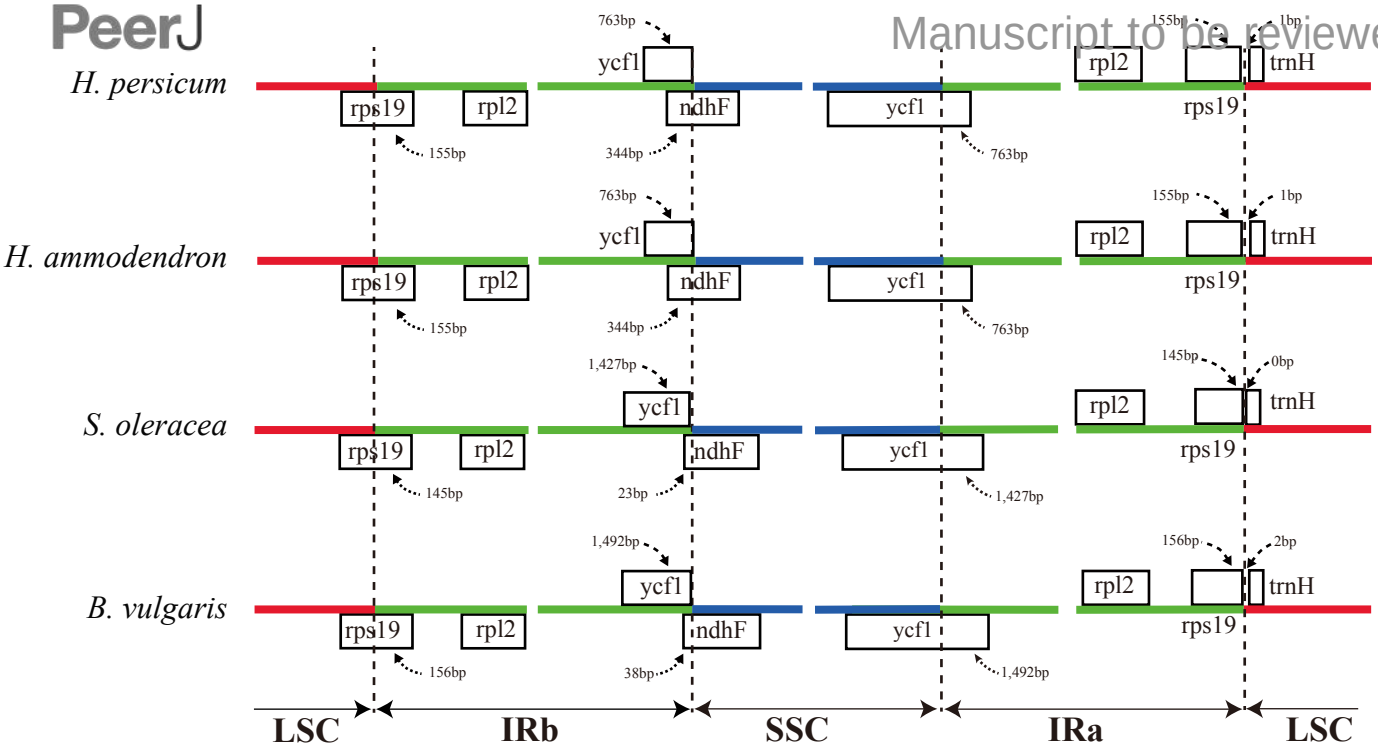
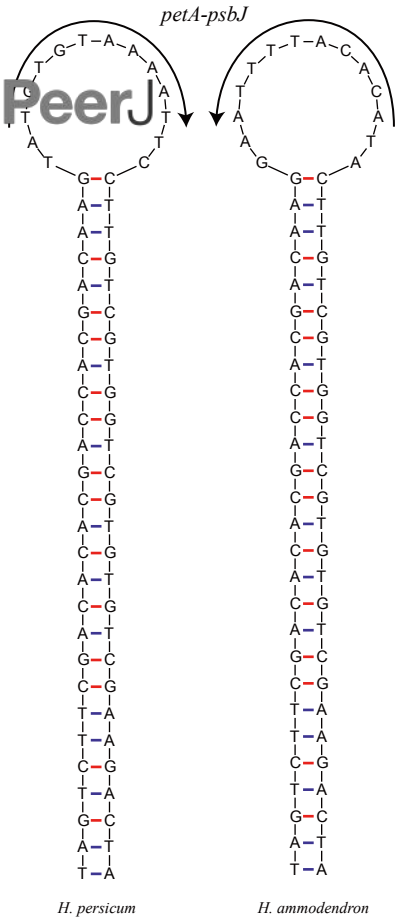


Figure 4(on next page)

Figure 4 The hairpin loops of inversions in the chloroplast genome of *Haloxylon*.



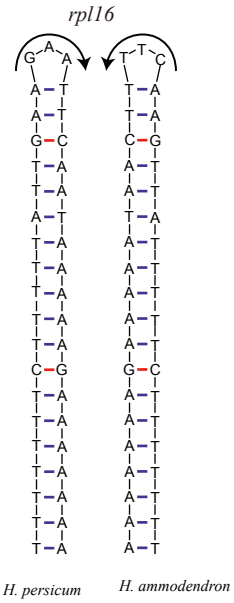


Figure 5(on next page)

Figure 5 The patterns of nucleotide substitutions among the two *Haloxylon* chloroplast genomes.

The patterns were divided into 6 types as indicated by the six non-strand-specific base-substitution types (i.e., numbers of G to A and C to T sites for each respective set of associated mutation types). The chloroplast genome of *H. ammodendron* was used as a standard.

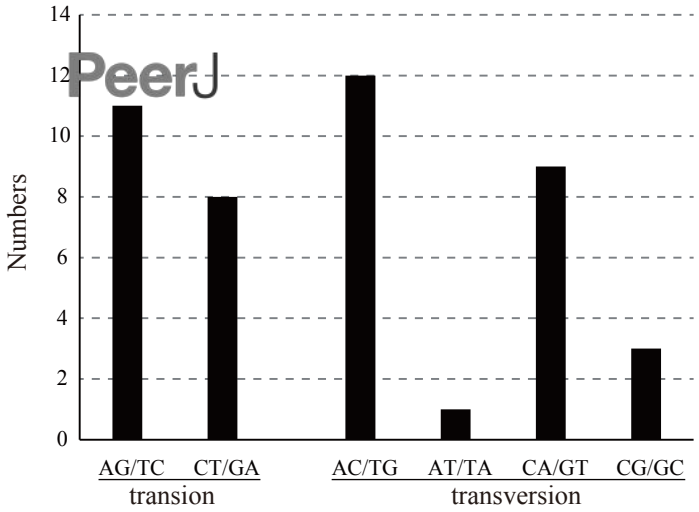


Table 1(on next page)

Table 1 Summary of complete chloroplast genome of *Haloxylon.*

1

2 **Table 1 Summary of complete chloroplast genome of *Haloxylon*.**

	H. ammodendron	*H. persicum*	*Spinacia oleracea*	*Beta vulgaris*
Total cpDNA size	151,570	151,586	150,725	149,635
Length of LSC region	84,214	84,217	82,719	83,057
Length of IR region	24,171	24,177	25,073	24,439
Length of SSC region	19,014	19,015	17,860	17,701
Total GC content (%)	36.6	36.6	36.9	36.4
LSC	34.4	34.5	34.8	34.1
IR	43.0	43.0	42.7	42.2
SSC	29.7	29.7	29.8	29.2
Total number of genes	112	112	112	113
protein encoding	78	78	78	79
tRNA	30	30	30	30
rRNA	4	4	4	4
Pseudogenes	2	2	2	1

3

Table 2(on next page)

Table 2 Genes with introns in *Haloxylon ammodendron* (*H. persicum*) and length of exons and introns.

1

2 **Table 2 Genes with introns in *Haloxylon ammodendron* (*H. persicum*) and length of exons and introns.**

	Exon I (bp)	Intron I	Exon II	Intron II	Exon III
atpF	145(145)	785(784)	410(410)		
clpP	71(71)	951(951)	292(292)	601(601)	228(228)
ndhA	553(553)	1090(1090)	533(533)		
ndhB	777(777)	675(675)	756(756)		
petB	6(6)	801(801)	642(642)		
petD	8(8)	722(722)	475(475)		
rpl16	399(399)	913(913)	9(9)		
rpl2	393(393)	668(668)	435(435)		
rpoC1	432(432)	780(780)	1602(1602)		
rps12	114(114)	–	231(231)	–	27(27)
rps16	40(40)	881(881)	197(197)		
trnA-UGC	38(38)	831(831)	42(42)		
trnG-GCC	23(23)	722(722)	58(58)		
trnI-GAU	42(42)	942(941)	35(35)		
trnK-UUU	35(35)	2909(2909)	37(37)		
trnL-UAA	35(35)	557(557)	50(50)		
trnV-UAC	39(39)	602(602)	35(35)		
ycf3	126(126)	772(772)	229(229)	812(812)	152(152)

rps12 is trans-spliced with the 5’ end located in the LSC region and the duplicated 3’ end in the IR regions.

3

Table 3(on next page)

Table 3 Forms and numbers of indel mutation events in the chloroplast genome between two *Haloxylon* species.

Table 3 Forms and numbers of indel mutation events in the chloroplast genome between two *Ha*

Gene	location	Types	*H. ammodendron*	*H. persicum*	Length (bp)
accD-psaI	Intergenic	homopolymeric indel	AA	-	2
atpA-atpF	Intergenic	homopolymeric indel	T	-	1
atpF	intron	homopolymeric indel	-	T	1
ndhI-ndhA	Intergenic	homopolymeric indel	-	A	1
ndhJ-ndhK	Intergenic	homopolymeric indel	-	T	1
psbI-trnS	Intergenic	homopolymeric indel	-	T	1
psbI-trnS	Intergenic	homopolymeric indel	-	A	1
rbcL-accD	Intergenic	homopolymeric indel	-	A	1
rps18-rpl20	Intergenic	homopolymeric indel	T	-	1
trnE-trnT	Intergenic	homopolymeric indel	-	A	1
trnK-rps16	Intergenic	homopolymeric indel	A	-	1
trnK-rps16	Intergenic	homopolymeric indel	A	-	1
trnL	intron	homopolymeric indel	-	A	1
trnL	intron	homopolymeric indel	A	-	1
trnL	intron	homopolymeric indel	-	T	1
trnR-aptA	Intergenic	homopolymeric indel	-	T	1
atpH-atpI	Intergenic	Indel	TTATT	-	5
clpP-psbB	Intergenic	Indel	-	GTCTT	5
petL-petG	Intergenic	Indel	-	G	1
rpoB-trnC	Intergenic	Indel	-	TGTAT	5
rpoB-trnC	Intergenic	Indel	TACAA	-	5
rrn23	coding	Indel	-	AATTAA	6
rrn23	coding	Indel	-	TTAATT	6

[a] The chloroplast genome of *H. ammodendron* was used as a standard

***loxylon* species.**

Driection [a]
insertion
insertion
deletion
deletion
deletion
deletion
deletion
deletion
insertion
deletion
insertion
insertion
deletion
insertion
deletion
deletion
insertion
deletion
deletion
deletion
insertion
deletion
deletion

Table 4(on next page)

Table 4 The patterns of nucleotide substitutions among the two *Haloxylon* chloroplast genomes.

Table 4 The patterns of nucleotide substitutions among the two *Haloxylon* chloroplast gen

Gene	location	*H. ammodendron*	*H. persicum*
atpA	coding	G	A
atpI	coding	T	C
matK	coding	C	A
ndhF	coding	C	T
ndhI	coding	G	T
psbC	coding	A	C
rpoB	coding	C	T
rpoC2	coding	C	A
rpoC2	coding	C	G
rpoC2	coding	G	T
rps15	coding	A	G
rps3	coding	T	G
ycf1	coding	A	G
ycf1	coding	G	C
ycf1	coding	G	T
atpB-rbcL	Intergenic	A	C
atpF-atpH	Intergenic	G	C
atpH-atpI	Intergenic	G	A
ndhF-rpl32	Intergenic	G	T
psaJ-rpl33	Intergenic	C	T
psaJ-rpl33	Intergenic	T	A
psbE-petL	Intergenic	C	A
psbM-trnD	Intergenic	A	G
rpl14-rpl16	Intergenic	T	G
rpl20-rps12	Intergenic	G	T
rpl33-rps18	Intergenic	T	C
rpoA-rps11	Intergenic	A	G
rpoA-rps11	Intergenic	T	C
rpoB-trnC	Intergenic	G	T
rpoB-trnC	Intergenic	T	G
rps18-rpl20	Intergenic	T	G
rps8-rpl14	Intergenic	G	A
trnG-trnR	Intergenic	A	C
trnH-psbA	Intergenic	T	G
trnK-matK	Intergenic	A	C
trnK-rps16	Intergenic	A	C
trnP-psaJ	Intergenic	C	T
trnP-psaJ	Intergenic	C	T
clpP	intron	T	G
ndhA	intron	T	C
rpl16	intron	T	C
rps16	intron	T	G
trnV	intron	T	C
ycf3	intron	T	C

Table 5(on next page)

Table 5 Location of repeats in the *Haloxylon ammodendron* chloroplast genomes.

Table 5 Location of repeats in the *H. ammodendron* chloroplast genome.

No.	Location	Motif	No.of Repeats	SSR start	SSR end
1	trnK-matK	A	11	1658	1668
2	trnK-rps16	A	12	4210	4221
3	rps16-trnQ	A	10	6461	6470
4	trnQ-psbK	A	10	6957	6966
5	psbK-psbI	A	10	7578	7587
6	psbI-trnS	A	12	7854	7865
7	atpF intron	A	10	12476	12485
8	rpoC1 intron	A	10	22386	22395
9	trnE-trnT	A	10	31169	31178
10	trnL-intron	A	12	47464	47475
11	trnF-ndhJ	A	10	48982	48991
12	rbcL-accD	A	12	57323	57334
13	accD-psaI	A	10	59584	59593
14	psbF	A	10	64309	64318
15	clpP intron	A	10	71717	71726
16	petB intron	A	18	75505	75522
17	ndhI-ndhA	A	10	118705	118714
18	psaA	C	10	40165	40174
19	trnK-rps16	T	10	4464	4473
20	psbI-trnS	T	10	7745	7754
21	trnR-atpA	T	11	9948	9958
22	atpA-atpF	T	10	11532	11541
23	atpF intron	T	11	12457	12467
24	rps2-rpoC2	T	11	15957	15967
25	rps2-rpoC2	T	11	18156	18166
26	rpoB	T	10	25865	25874
27	trnD-trnY	T	10	30323	30332
28	trnL-trnF	T	10	48029	48038
29	ndhJ-ndhK	T	10	49646	49655
30	trnV intron	T	15	52214	52228
31	trnM-atpE	T	10	52658	52667
32	rbcL-accD	T	14	57377	57390
33	petL-petG	T	10	66141	66150
34	psaJ-rpl33	T	12	67499	67510
35	rps18-rpl20	T	10	68447	68456
36	rpoA	T	10	78219	78228
37	rps11-rpl36	T	12	79577	79588
38	rpl32-trnL	T	11	112371	112381
39	ndhA intron	T	12	119581	119592
40	ndhA intron	T	10	119793	119802
41	ycf1	T	12	125285	125296
42	ycf1	T	10	125890	125899
43	ycf1	T	14	126895	126908
44	ycf1	T	10	127195	127204
45	rps16-trnQ	at	5	6277	6286
46	trnS-trnG	at	5	8177	8186
47	trnS-trnG	at	5	8300	8309
48	trnN-ndhF	taa	4	109380	109391
49	psbA-trnK	ttgt	3	1522	1533
50	matK-trnK	ttct	3	3873	3884
51	atpI-rps2	atta	3	15121	15132
52	trnE-trnY	atta	3	31084	31095
53	accD-psaI	taat	4	59721	59736
54	rps18-rpl20	ttta	3	68474	68485

55	clpP intron	tttc	3	71598	71609
56	rrn23	aggt	3	104481	104492
57	trnL-ccsA	aacc	3	113312	113323
58	ycf1	taat	3	124297	124308
59	rrn23	ctac	3	131310	131321

Table 6(on next page)

Table 6 Comparisons of mutational changes, number of transitions (Ts) and transversions (Tv), synonymous (S), and nonsynonymous (N) substitutions per gene of protein coding chloroplast genes between *Haloxylon ammodendron* and *H. persicum.*

Table 6 Comparisons of mutational changes, number of transitions (Ts) and transversions (Tv), synonymous (S), an

Gene	Ts	Tv	S	N
atpA	1	0	1	0
atpI	1	0	1	0
matK	0	1	0	1
ndhF	1	0	0	1
ndhI	0	1	0	1
psbC	0	1	1	0
rpoB	1	0	1	0
rpoC2	0	3	0	3
rps15	1	0	0	1
rps3	0	1	0	1
ycf1	1	2	1	2
Total	6	9	5	10

d nonsynonymous (N) substitutions per gene of protein coding chloroplast genes between H. ammodendro

n and H. persicum.